SIMULTANEOUS TESTS OF HYPOTHESES & CONFIDENCE INTERVALS

By

M.C. VERMA

Institute of Agricultural Research Statistics (I.C.A.R)
Library Avenue,
NEW DELHI - 12.

1961.

SIMULTANEOUS TESTS OF HYPOTHESES & CONFIDENCE INTERVALS

By

M. C. VERMA

Dissertation submitted in partial fulfilment of the requirements for the award of Diploma in Agricultural and Animal Husbandary Statistics of the Institute of Agricultural Research Statistics (ICAR) Library Avenue, New Delhi.

1961

"ACKNOWLEDGEMENT"

I have great pleasure in expressing my deep sense of gratitude to Dr. M.N. Ghosh, Professor of Statistics, Institute of Agricultural Research Statistics (I.C.A.R.) New Delhi for his valuable guidance, keen interest and constant encouragement throughout the course of investigation and immense help in writing up the thesis by critically going through the manuscript.

I am grateful to Dr. G.R. Seth, Statistical Adviser, I.C.A.R., New Delhi-12, for providing necessary facilities to carry out the investigation.

MCVerma

M. C. VERMA

CONTENTS

Chapter I

Introduction

A very common situation in the Analysis of Variance is that one makes a number of tests of several related or unrelated hypotheses. The 'simultaneous test' necessary in such cases is one which specifies which of the several hypotheses should be accepted or rejected, at a particular chosen level of significance. Very many times it is insisted in such tests that the estimates of parameters corresponding to various hypotheses should be uncorrelated with one another. This is the well known criterion of orthogonality of data. When the experiment is well planned in advance, the theory of design of experiments allows us to have such orthogonality in general, but in case of Survey data for example, such planning is not possible and the estimates of various parameters become correlated in various ways. Even for experiments put in the framework of a specific design such a situation arises when the cell frequencies are unequal or due to some mishap some part of data cannot be used or is unavailable. This sort of thing is very common in animal experiments where there is a paucity of satisfactory experimental units and the experimenter has to use whatever materials are made available to him. Suppose for instance, we have to test three hypotheses simultaneously: $H_1: \theta_1 = 0$, $H_2: \theta_2 = 0$, $H_3: \theta_3 = 0$ where θ's may be vector-valued parameters. If they have to be tested from the same body of data, usually the error variance will be common, so that even if the estimates of θ's are uncorrelated the three tests will not be independent. This shows the inadqquacy of orthogonality as the prerequisite of independent tests. But independence is too severe a restriction in view of the fact that the objects of our interest are only the first and second kinds of errors of the component tests. If those errors of one test do not depend on the parameters of the other hypotheses, it may be sufficient for our purpose. In fact, it will be seen later that from the point of view of decision theory, if we take certain types of loss functions, this property alone is sufficient to prove many optimum properties of the simultaneous decision procedure whereever the individual decision procedures have such optimum properties. Ghosh (1955) has called this modified criterion as 'quasi independence' and has defined quasi-independent test of multiple hypotheses as given below.

<u>Quasi independent tests</u>: Consider the hypotheses $H_i : \theta_i = 0$, and suppose for each i (i = 1,2,3) the test T_i of H_i is such that the first and second kinds of error for it do not involve any parameters other than θ_i; then T_i's are supposed to provide quasi independent tests of H_i's. Mathematically, for each i

$$\Pr[\text{Accept } H_i / \theta_i \neq 0, \theta_1 \text{ etc}] = \Pr[\text{Accept } H_i / \theta_i \neq 0, \theta_1 = 0 \ldots \text{ etc}]$$

and $\Pr[\text{Reject } H_i / \theta_i = 0, \theta_1 \cdots \text{etc}] = \Pr[\text{Reject } H_i / \theta_i = 0, \theta_1 = 0 \cdots \text{etc}]$

hold for all values of the other parameters θ_1 etc i.e. whether H_1 etc are true or not.

On the other hand, the criterion of independence of tests of H_1 and H_2 say, requires that for the critical regions C_1, C_2 for H_1, H_2 respectively we should have:

$$\Pr[X \subset C_1, X \subset C_2 / \theta_1, \theta_2] = \Pr[X \subset C_1 / \theta_1] . \Pr[X \subset C_2 / \theta_2]$$

for all θ_1 and θ_2, X reppesenting the sample observed.

<u>Control of Errors in the simultaneous tests of hypotheses</u>:

There are different points of view of assigning significance levels in the case of simultaneous tests of hypotheses depending on the convictions of the experimentor, considerations of cost etc. In some cases when the decisions regarding the hypotheses $H_1, H_2 \cdots$ etc are unrelated it is proper to consider the significance level of each hypothesis individually at 5% or 1% (say). But this is possible only when the tests are independent of each other. As seen earlier this is not usually so. Consider for example 2 tests of H_1 and H_2, which provide 2 sum of squares S_1 and S_2 along with an independent estimate of error, S_e^2 say. If S_1 and S_2 are distributed independently as $\chi_1^2\sigma^2$ and $\chi_2^2\sigma^2$ under the null hypotheses, as is the case in orthogonal data and if S_e^2 is distributed as $\chi_3^2\sigma^2$ where χ_i^2 denotes χ^2-distribution, then $G_1 = \frac{\chi_1^2}{S_e^2} \sim C_1F$ and $G_2 = \frac{\chi_2^2}{S_e^2} \sim c_2F$ where F denotes F-distribution. When one has made a test on H_1 by χ_1^2/S_e^2, the test of H_2 should be strictly speaking, a conditional test on observing certain value of $\chi_1^2/S_e^2 = \lambda_1$ say. But we should see below that this conditional test of H_2 does not provide a quasi-independent test. For, under the null hypotheses.

$$\Pr\left[\frac{\chi_2^2}{S_e^2} < \lambda_2 \Big/ \frac{\chi_1^2}{S_e^2} = \lambda_1 \right] = \int_0^{\lambda_2} p\left(\frac{\chi_2^2}{S_e^2} \Big/ \frac{\chi_1^2}{S_e^2} = \lambda_1 \right) d\left(\frac{\chi_2^2}{S_e^2} \right) \qquad \ldots(1\ldots)$$

$$= \int_0^{\lambda_2} f(G_1, G_2)\, d\, G_2 \,/\, f_1(G_1)$$

where $f(G_1\ G_2)$ is the joint distribution of G_1 and G_2 and $f_1(G_1)$ is the marginal distribution.

$$= \int_0^{\lambda_2} \frac{G_1^{\frac{K_1-2}{2}}\ G_2^{\frac{K_2-2}{2}}}{(1+G_1+G_2)^{\frac{K_1+K_2+n_e}{2}}}\, dG_2 \Bigg/ C_2\ G_1^{\frac{K_1-2}{2}} / (1+G_1)^{\frac{K_1+n_e}{2}} \qquad \ldots(2)$$

where χ_1^2 has K_1 d.f., χ_2^2 has K_2 and S_e^2 has n_e d.f.

$$= C(\lambda_1, K_1 K_2, n_e) \int_0^{\lambda_2} \frac{G_2^{\frac{K_2-2}{2}}}{(1+G_2/1+\lambda_1)^{\frac{K_1+K_2+n_e}{2}}} \cdot \frac{1}{(1+\lambda_1)^{\frac{K_1+n_e}{2}}}\, dG_2 \qquad \ldots(3)$$

$$= C'(\lambda_1, K_1\ K_2, n_e) \int_0^{\lambda_2/1+\lambda_1} \frac{x^{\frac{K_2-2}{2}}}{(1+x)^{\frac{K_1+K_2+n_e}{2}}}\, dx \qquad \text{where } x \sim \chi^2_{K_2} / \chi^2_{K_1+n_e}$$

comparing with F-distribution we note that i. the significance level for the conditional test of H_2 is 5% , then the limit λ_2 will be obtained from

$$\lambda_2/1+\lambda_1 = K_2/K_1+n_0\ \ F_{.95,k_2,k_1\ n_0} \qquad \ldots(4)$$

where $F_{1-\alpha, n_1, n_2}$ means the upper α % poin of a F-distribution with degrees of freedom (n_1, n_2). Hence,

$$n_0/K_2 \cdot \lambda_2 = n_0/K_2 .\ K_2/K_1 + n_e\ \ (1+\lambda_1)\ F_{.95,k_2,k_1+n_e} \qquad \ldots(5)$$

$$= n_0/K_1+n_0\ \left[1 + K_1/n_e \cdot F_{1,obs} \right] F_{.95,k_2,k_1+n_e}$$

i.e. $F_{2,condl.} = n_e + k_1\, F_{1,obs} / (n_e + k_1) .\ F_{.95,k_2,k_1\ n_e}$ $\ldots(6)$

where $F_{1,obs}$ denotes the observed F value for H_1 and F_2, condl. denotes the required 5% point of the conditional test of H_2.

The use of F_{2cond1} will give an exact 5% level for H_2 if S_1 is distributed on the null hypothesis $\theta_1 = 0$. But if $\theta_1 \neq 0$, S_1 gives a non central χ^2 and so the distribution of χ_1^2/S_e^2 becomes;

$$f_1(\chi_1^2/S_e^2) = C \sum_{\gamma=0}^{\infty} \left(\frac{\nu^2}{2}\right)^{\gamma} \frac{1}{\gamma!} e^{-\frac{\nu^2}{2}} \frac{G_1^{\frac{K_1}{2}+\gamma-1}}{(1+G_1)^{\frac{K_1+n_e}{2}+\gamma}} \, dG_1 \quad ....(7)$$

where ν^2 is a function of the vector θ_1 which is non null. The joint distribution of χ_1^2/S_e^2 and χ_2^2/S_e^2 is of the form ;

$$C \sum_{\gamma=0}^{\infty} \left(\frac{\nu^2}{2}\right)^{\gamma} \frac{1}{\gamma!} e^{-\frac{\nu^2}{2}} \frac{G_1^{\frac{K_1}{2}+\gamma-1} \; G_2^{\frac{K_2-2}{2}}}{(1+G_1+G_2)^{\frac{K_1+K_2+n_e}{2}+\gamma}} \quad ....(8)$$

hence from (3) we can no more get a λ_2 which is free of ν^2 so that whatever limit we choose for the conditional that the significance level is not free of the parameters of the other hypothesis H_1 i.e. not quasi-independent.

This shows that it is not possible to fix the significance levels individually at some preconceived value if we want to have quasi-independent test. But if we consider the first kind of error of the simultaneous test above, as the rejection of at least one of the hypotheses, when both of them are in fact true; then this can be fixed at some particular value and quasi-independent tests obtained as shown by Ghosh (1955). As we shall see this is proper specially when the decisions about the 2 hypotheses have joint import.

Simultaneous level of significance: It is defined as the probability of rejecting at least one of the several hypotheses considered, when all of them are true.

Suppose the experiment has not been planned in advance and the investigator inserts several parameters in the model to find whether any of them is important. If the number of p ramaters so inserted is high, it is quite likely that some of them will show significance if tested individually on some fixed level of significance even though it were not true; because the simultaneous significance level increases as number of groups increases. Such a situation occurs often when Survey data from

Medicinal, Sociological or even Biological fields is taken where the investigator was not able to put the data in the framework of some design. In fact the may not be even clear about what statistical hypotheses he is going to test, unless he has a look at the data itself. Since it is more likely in these cases to ascribe significance to one or other group of parameters, we should proceed to fix the simultaneous significance level of the various hypotheses instead of conducting each test in isolation with a fixed level of significance.

That the above considerations need not apply only to unplanned experiments is clear from the example of Quality control for acceptance of material. Here one may like to reject the material if it falls short of standard in any of the several characters in which the investigator is interested. It may be useful then, to determine the particular deficiency in any of these characters and so we must use a simultaneous test with the simultaneous level of significance to insure against too frequent rejections.

A similar situation arises in agricultural field trials. A new variety before it is introdused is compared with the existing varieties and will be recommended if it is superior in any of the characters, for example, Strength and Length of a Cotton fibre. Since one would be intrested to know the particular character in which the new variety is superior, a simultaneous test of the hypotheses concerning various characters should be done. In this case the significance level of this test should be naturally be the simultaneous significance level, as a caution against hastely introducing the new variety. This conservatism is natural for one would not like to declare the new variety superior when in fact it is not.

The necessity of using a safety device, like joint significance level, as it is above, would be even more apparent when the number of characters is large, in which case the chance of declaring a new variety different from the common variety may be much larger than 5% or 1%.

The common situation of testing for differences of means in the Analysis of Variance provides another example of multiple tests. The experimenter may like to use repeated t-tests on the difference of means in which he is interested. If his tests relate strictly to what he had originally planned or to obviously natural comparisons, he may make the different tests independently of each other at some fixed individual level say 5%. But, as what happens in general, some comparisons are suggested by the data themselves and so some device like the simultaneous significance level again becomes necessary for not arriving at too many significant results, which may actually be due to chance. For example, it is well known (Fisher; 'Design of Experiments') that if one does 20 tests independently of each other, one significant result from this set may be attributed to chance alone.

In Statistical literature, though it is mentioned many times that such a false conclusion may arise but the use of individual levels of significance for various tests derived from the same sample is as frequent. In the standard Analysis of Variance case of manifold classification a number of tests are made using the same error Variance in the denominator. As is mentioned by Kendall (1948): 'The ordinary z.test (or F test) given the the probabilities that the ratio of 2 Variances chosen at random does not exceed a given value.' If the number of Variance ratios is large, and if we deliverately pick out the largest to test for significance the chance that this is shown significant at the 'α-level' is a good deal greater than and we run into the danger of attributing significance to what may be pure sampling effect. But the problem of actual evaluation of the simultaneous significance level is quite intricate, sometimes defying explicit solution and this may be one reason why there is a discrepancy between the demands of statistical theory and current practice of those tests.

However, the control of Ist kind of error in a simultaneous test is achieved at the cost of increasing the IInd kind of error for individual hypotheses. In any particular problem, whether the point of view of individual or simultaneous significance level should be adopted depends, roughly, on how much a priori weight one attaches to the alternative hypotheses, either from theoretical expectations or from considerations of cost. Thus in case of the example of new variety vs a common variety, if the cost is small e.g. if varieties are grown only on an experimental scale one would be relatively free to decide on any of the varieties as superiof and so the individual significance levels say 5% each, will be proper. But, with an established variety the costs of replacement by a newer one will be enormous and so the statistician should take an attitude of caution. It is essentially this conservative attitude that requires the use of simultaneous significance level say 5% again.

We may also mention the standard case of factorial set of treatments in a standard design. If the decisions have joint import as is likely here, the simultaneous significance level should be preferred. But as we shall see in the sequel, the actual computation of this is not easy; the tables when formed, involves n_e, the error degrees of freedom along with $k_1, k_2 \ldots k_s$ the d.f.'s for the s hypotheses $H_1 \ldots H_s$ and so a $(s+1)$ way table is needed.

For the case of non orthogonal data, we know that it is difficult to fix even the individual levels of significance exactly and so the evaluation of simultaneous significance level is still more difficult. We shall be concerned herew mainly with this problem where we can separate out some groups of parameters representing different aspects of the problem.

Chapter II

Review of various methods for simultaneous tests:

In recent years a variety of methods have been proposed by many authors notably Scheffe, Tukey & Duncan for the special case of simultaneous tests of differences between means of the same set. They relate mostly to ranking of means with respect to one another and sometimes confidence intervals for particular contrasts of means. In many situations it may be of greater interest to get confidence bounds for some interesting contrasts of means rather than testing merely the equality of means as is possible by an ordinary F test in the standard Analysis of variance. A confidence interval always involves a significance statement while the reverse is not true. Hence it may be preferable to derive the confidence intervals first and compare various methods with respect to them.

In 1957 Lehmann developed a theory of multiple decision problems but the original idea of converting a problem of ranking of means into a set of 2-decision problems was given earlier in 1950 (Lehmann). Duncan in his 1955 paper utilizes this approach to compare many current methods of ranking of means, by the help of what he calls 'p-mean protection levels' $\gamma_p(p=2,3\ldots k)$ where

$$\gamma_p = \text{Min}^m \left[\text{Test shows that } \mu_1 = \cdots = \mu_p \,/\, \mu_1 = \mu_2 = \cdots = \mu_p \right]$$

This he does by making tests on different hypotheses one after another. Thus he defines 2-mean protection level, 3-mean protection level ..... k-mean protection level, where the first is usually fixed at '$1-\alpha$', the common level of significance and the rest are either kept constant or are in decreasing order. In fact it only means that first we are testing various hypotheses concerning the differences between pairs of means with a joint level of α, next we are testing for equality of means taken 3 per set, and so on upto the final test of equality of k means with a different joint or individual (here) level of significance. This is the approach of the tests of Duncan, Newman-Keuls and one test of Tukey.

We shall refer to Duncan's tests after and especially his exhausting paper in Biometrics 1955. Of course, in all such tests some restrictive assumptions are involved for example, equality of the variance of different means, independence and equal no. of observations per mean. Duncan's method for ranking cannot directly be used for setting confidence limits for differences of pairs or triads ...etc ; for his main concern is testing for the ranks of 2 or more means when they have already been arranged in an ascending order, say. That is partly why the reasons advanced by him for using lower protection levels for higher order ranking of means do not hold for getting confidence intervals for the means and some relevant contrasts of them.

Thus, he says on the basis of an analogy with independent tests of several hypotheses, that as there so in the case of k-means the higher order protection level may be smaller e.g. Take k-means $\mu_1 \cdots \mu_k$ and define the 3-mean protection level $\gamma(1,2,3) = \text{Min}^m$ (decision $\mu_1=\mu_2=\mu_3 \,/\, \mu_1=\mu_2=\mu_3$). This he fixed at $(1-\alpha)^{3-1} = (1-\alpha)^2$ whereas the protection level for 2 means is just $(1-\alpha)$. The exponent 2 represents the number of independent comparisons possible with '3' means here, for one can form 2 independent χ^2 of 1 d.f. each for these 2 comparisons and the joint level of significance for the simultaneous hypotheses $[\mu_1=\mu_2 \;,\; \mu_1-2\mu_2+\mu_3=0]$, say shall be $(1-\alpha)^2$ if each test is individually on 'α' level. But there are many cases where these two component tests should not be made on individual levels 'α' but on the simultaneous level 'α'. We should keep this difference in mind while comparing the methods which we are going to present with the methods advanced by Duncan, Newman-Keul and others, as also the restrictive assumptions under which these methods work, though in some recent contributions to this topic by C.Y. Kramer (1956 & 1957) some of them have been removed. Thus the means from the sample need not be based on equal number of observations and they may be correlated as well But the methods advanced for the extension suffer from the handicap, that though they are hoped to be conservative, the exact evaluation of specific significance levels is found to be extremely difficult. Only the lower bounds for the levels can be easily obtained.

For finding the confidence intervals for contrasts of a set of means the oldest methods is probably Fisher's 'least significant difference' method. Though it has been shown by many writers like Duncan, that this is not satisfactory for it gives too low protection levels, its use is common. Of course, if one can specify the contrast of interest before data is collected, the method gives the optimum confidence intervals. But in general many comparisons inside the set are suggested by the data themselves and then testing for difference between the highest mean and the lowest, for example, will be faulty for its level of significance will not be α but much higher, only a upper bound for which can be specified. Another variation of this method is using it in combination with the general F test of equality of the k means. The only effect of this is that the highest protection level viz. k-mean level is raised to $1-\alpha$ while others even now remain too low.

Various other multiple F tests have been proposed using the above idea, notably the 'multiple-comparisons' test of Duncan (1952). According to it, the difference between any two means in a set of k means is significant provided the variance of each and every subset which contains the given means is significant according to α_p-level F-test where $\alpha_p = 1-\gamma_p = 1-(1-\alpha)^{p-1}$. This can be generalized to test for a contrast $\sum^k \lambda_i \mu_i$ being different from zero where the F tests for all subsets containing means μ_i having $\lambda_i \neq 0$ along with a t-test on $\lambda_i \mu_i$ is done. Duncan suggested a compromise between the first and its 'conservative' generalization but the power properties of the test are still obscure, though some bounds can be put for the various protection levels. Again, it will be found that these tests do not lend themselves to confidence statements about differences of true means, being primarily for detecting significance.

Nandi (195$) has considered what he calls D-test which is in fact repeated t-tests but the actual gain in using D-test for small error degrees of freedom has not been calculated. Further he suggests another multiple F-test but with the difference that t-tests are to be made firstly and then the non-significant t's are to be combined to yield a F-test about those means. But again the power properties of this 'modified D-test' are left in obscurity and unless that is done the test has little to recommend itself. Actually those tests suggested, do not concern themselves with ranking of means only and may be expected to throw more light on the possible sources of discrepancy from the hypothesis of homogeneity of k means than is possible by the comprehensive F-test.

The test proposed by Scheffe (1953) essentially starts from confidence intervals for all possible linear contrasts of means. This again may use F-test of homogeneity of means as a first step to finding the confidence intervals. If F-test shows homogeneity, all the subsequent confidence intervals are not found. But as against other multiple F and 't' tests, there is no complication in power properties here; for it happens sometimes in other tests that though the overall test shows homogeneity, the test for some comparison shows significance. Many times it has been objected that the confidence intervals given by Scheffe's approach are too wide to be of any use; but except for one method of Tukey (1951), no other method gives confidence intervals for many complex contrasts. Scheffe has extensively considered his confidence intervals against Tukey's or repeated t-test method and has shown that the increase in length is not so large that one may prefer a biased method to the exact one.

Bose & Roy in their extensive (1953) paper have given a different method of arriving at the result of Scheffe, for confidence intervals. This approach essentially derives from Roy's Union-Intersection principle of test construction. Simultaneously with the test of hypothesis this approach directly gives the various simultaneous confidence intervals. Thus Tukey's method for finding confidence intervals for all pair-differences of $\binom{k}{2}$ pairs of true means, can be easily derived by considering the component hypothesis $\mu_i = \mu_j$ $(i, j = 1, \ldots k)$ and using the known optimum test. This gives

$$P_r \left[\max_{i,j} |\bar{x}_i - \bar{x}_j| / s\sqrt{\tfrac{2}{n}} < Q \right] = 1-\alpha$$

where $\sqrt{2}Q$ is the α % point of Studentised range for k-means. In this paper Bose and Roy have generalized the univariate approach to that of finding the confidence bounds of all 'double linear compounds of the p sets of means'. Thus in case of 2 correlated characters for a Cotton variety considered earlier, this method gives a conservative method of finding confidence intervals for differences of means of the varieties of the 2 characters considered. This is a wasteful way for, one may need only the confidence intervals for means of one variable and for the other variable separately but none for the joint differences of the means of the 2 varieties. They have also considered an extension of Tukey's method of finding the confidence intervals for all $\binom{K}{2}$ pair differences to multivariate ... also, but these problems are far too difficult to allow exact evaluation of levels.

Scheffe in his paper gives a method of Tukey more general than that of allowances for pairs of means' alone. But as above, here also Tukey's method is restrictive in the sense that means are supposed to have the same variance and ... are supposed to have equal covariances (if not zero as assumed in the test of allowances for pairs of means). Thus for a contrast $\sum_1^K c_i\mu_i$ with $\sum c_i = 0$, $V(\hat{\mu}_i) = a\,\sigma^2$ $Cov(\hat{\mu}_i, \hat{\mu}_j) = b\,\sigma^2$ for all $i, j = 1, \ldots, k$, this generalized method yields the confidence intervals $\sum c_i\hat{\mu}_i - T\hat{\sigma} \leq \sum c_i\mu_i \leq \sum c_i\hat{\mu}_i + T\hat{\sigma}$ where $T = \frac{1}{2}\sum|c_i|\; q_{\alpha, k, n_0}\cdot\sqrt{a-b}$ where q_{α, k, n_0} = upper α % point of studentized range (k, n_0).

The corresponding test of significance for the hypothesis $\mu_1 = \cdots = \mu_k$ is such that

$$Pr\left[\,|\sum c_i(\hat{\mu}_i - \mu_i)|/\hat{\sigma} < T \;/\; \text{all values of } c_i\right] = 1-\alpha$$

... that this modified test, when it can be used is better than Scheffe's, only for contrasts of the form

$$\left[1/t \sum_1^t \mu_i - 1/k-t \sum_{t+1}^k \mu_i\right] \quad \text{for } t = k/2 \text{ when } k \text{ even}$$
$$= k-1/2 \quad \text{when } k \text{ odd}$$

When applicable Tukey's above 2 methods have some good properties. In fact Ramachandran (1956,a) has shown the unbiased nature of the test of pairs of means. But actual comparison in terms of power function is very complex and has been hardly attempted. In fact, it can be seen that though Scheffe's & Tukey's generalized tests have the same level and test the same homogeneity hypothesis, the power function of Scheffe's test is maximum when alternatives form a well defined class as shown by Hsu (1941) and Hartley, while that of Tukey's Ist method is high for alternatives such that $\left\{ \mu_{(\max^m \hat{\mu}_i)} - \mu_{(\min^m \hat{\mu}_i)} \right\}$ is the only relevant thing. The comparison of 2 methods thus becomes very cumbersome and some prior judgement of possible alternatives is very necessary to choose between the two.

Our approach below concerns mainly with cases where treatments or parameters involved in the Analysis of variance fall into certain natural groups and make a simultaneous test of the various component hypotheses so formed, with a given simultaneous significance level. As is well known, if the various tests for individual hypotheses are all independent the joint level α will be attained if each individual level is fixed at $(1-\alpha)^{1/k}$. But in practice the tests are not independent both due to the use of same error S.S. in the F-ratio, as well due to non-orthogonality of the various groups of parameters. In case of orthogonality some results are available; due to Hartley (1938) as approximations and Cochran (1941) and Finney (1941) as exact expressions. It is suggested that if the error degrees of freedom are sufficiently large, the different ratios may be treated as independent.

When the number of degrees of freedom in the numerator are all equal, the evaluation of simultaneous level of significance involves only, what Ramachandran has termed 'Largest Studentized χ^2' (1956, d) on f_2 d.f., say. Ramachandran (1956, b), has given some calculations also for the above special case but for 2 χ^2's only. Some other results in this connection are available in Pillai & Ramachandran's paper of 1954.

The very special case of k tests in the ANOVA when each has only 1 d.f., has been solved by Nair (1948) who has given extensive calculations and method of computations. Though a special case this may be very useful in t^{n} factorial experiments where, as frequently happens, each main effect is broken up into (t-1) comparisons of 1 d.f. each. Ramachandran (1956, c) has solved a similar problem in simultaneous tests by a different approach.

But, in all that has been done in recent years there is an initial restriction - that of orthogonality, the independence of the estimates of one group of parameters of those of other groups. In practice non-orthogonality may be inherent in the design as in case of Survey data from Medical or Sociological fields or it may arise due to faults in a planned experiment, like missing observations or it may arise because we want to utilize whatever materials are available as in case of almost all animal experiments, with feeds, managemental conditions etc. Ghosh (1955) has extensively dealt with this problem and suggested 3 methods which may be useful under different conditions. The tests of hypotheses may be inverted to give confidence intervals and vice-versa. But these tests are not strictly independent, being only quasi-independent. We shall briefly give in the following chapters the methods given there and make comparisons among themselves and with a new method. Further we shall give some practical situations where these methods are necessary alongwith 2 actually worked out examples.

-***********-

Chapter- III

Simultaneous tests of linear hypotheses

Consider the linear model

$$E(y_i) = a_{i1} p_1 + \cdots + a_{im} p_m \qquad (i = 1, 2 \ldots N)$$

y_i being independent normal variables with unknown variance σ^2 and $p_1, p_2 \ldots p_m$ are unknown parameters. Let rank $(a_{ij}) = N_0$ and let $\pi_1, \ldots \pi_R$ be estimable linear functions of the parameters $p_1, \ldots p_m$.

$$\pi_i = l_{i1} p_1 + l_{i2} p_2 \cdots + l_{im} p_m \qquad (i = 1, \ldots R) \qquad \ldots(9)$$

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:

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

$$\vdots$$

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

For the simultaneous test of these 's' hypotheses we want to get some test procedure such that the Ist kind of error i.e. the simultaneous level of significance is $\leq \alpha$ and that the test has good properties against certain class of alternatives which depend on the experience and interests of the experimenter.

Let $Y_1 \ldots Y_{k_1}$; $Y_{k_1+1} \ldots Y_{k_1+k_2}$; $\ldots Y_{k_1+ \ldots + k_s}$ 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 Markoff's theorem we have an independent estimate of error variance say s_e^2 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,

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

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

Let $U_{b_{n-1}+1}, \ldots, U_{b_n}$ be orthonormal 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

$$E\left\{(U_i, y)\right\} = \sum_{b_{n-1}+1}^{b_n} \alpha_j\, E(Y_j, y) = \sum \alpha_j \pi_j \quad \text{from } E(Y_j, y) = \pi_j$$
$$= \phi_j \text{ say} \qquad \ldots\ldots(10)$$

On the hypothesis H_n, $\chi_n^2/\sigma^2 = \frac{1}{\sigma^2}\sum\left[(U_i, y) - E(U_i, y)\right]^2$ has a χ^2-distribution with k_n d.f. and so $\chi_n^2/k_n s_e^2$ has a F-distribution with d.f. (k_n, n_e). The IInd kind of error for this test of H_n depends on the parameters $\Pi_{b_{n-1}+1} \cdots \Pi_{b_n}$ only.

We shall now group the 's' hypotheses $H_1, \ldots, H_s$ into r groups such that the vectors corresponding to the parameters involved in each set are such that any vector for one set is orthogonal to any vector for the any of the other sets. Thus if hypotheses H_i & H_j are such that $(Y_{K_{i-1}+1}, \ldots, Y_{K_i})$ are not orthogonal to the vectors relating to H_j $(Y_{K_{j-1}+1} \ldots Y_{K_j})$, then they will be grouped together in a set say M_1 and so on. So vectors belonging to different sets are orthogonal and vice-versa. These sets are termed as orthogonal sets and the hypotheses belonging to 2 different sets, 'orthogonal hypotheses'.

Case-I. We shall first consider the case where each orthogonal set consists of only one hypothesis, and consider methods of exact evaluation of the simultaneous level of significance in the general case. Let $U_1 \ldots U_{K_1}$; $U_{K_1+1} \ldots U_{K_1+K_2}$; ……; $U_{K_1+\cdots+K_{s-1}+1} \ldots U_{K_1+\cdots+K_s}$ be the ortho-normal systems of vectors in the space determined by $Y_1 \ldots Y_{K_1}$; $Y_{K_1} \ldots Y_{K_1+K_2}$; … $Y_{K_1+\cdots+K_s}$ respectively. The method suggested by Ghosh consists in forming χ^2's like

$$\chi_1^2 = \frac{1}{\sigma^2}\sum_1^{K_1}\left[(U_i, y) - \phi_i\right]^2 \quad \ldots\ldots \quad \chi_s^2 = \frac{1}{\sigma^2}\sum_1^{K_s}\left[(U_{K_1+\cdots+K_{s-1}+i}, y) - \phi_{K_1+\cdots+K_{s-1}+i}\right]^2 \qquad \ldots\ldots(11)$$

$\chi^2_{error} = n_e S_e^2 / \sigma^2$ and finding the joint distribution of $G_1 = \chi_1^2 / n_e S_e^2$ .... $G_s = \chi_s^2 / n_e S_e^2$. From this distribution we should find constants $\lambda_1 \cdots \lambda_s$ such that the joint level 'α' is attained i.e.

$$\Pr \left[G_1 < \lambda_1 \;\ldots\; G_s < \lambda_s \,/\text{all H's are true} \right] = 1 - \alpha \qquad \ldots(12)$$

which means

$$1-\alpha = C(k_1 k_2 \ldots k_s; n_e) \int_0^{\lambda_1} \cdots \int_0^{\lambda_s} \frac{\prod_1^s G_i^{\frac{k_i-2}{2}}}{\left(1+\Sigma G_i\right)^{\frac{1}{2}(n_e+\Sigma k_i)}} \prod_1^s d\,G_i \qquad \ldots(13)$$

Consider now the region defined by $\left\{ G_1 < \lambda_1 \;\ldots G_s < \lambda_s \right\}$ which has the probability $P(\lambda_1 \cdots \lambda_s)$ say, under the simultaneous hypotheses $H_1 \ldots H_s$. When $\lambda_1 \cdots \lambda_s$ are fixed subject to (13).

We immediately get a system of simultaneous confidence regions $C_1, \ldots C_s$ say, for the sets of parameters with a confidence coefficient $1-\alpha$, as

$$C_1 : \pi_1 \cdots \pi_{k_1} \qquad \sum_1^{k_1} \left[(v_i, y) - \varphi_i \right]^2 < \lambda_1 n_e S_e^2$$

$$\cdots\cdots\cdots$$

$$C_s : \pi_{k_1 \cdots + k_{s-1}+1} \cdots \pi_{k_1 + \cdots + k_s} \qquad \sum_1^{k_s} \left[(v_{k_1+\cdots k_{s-1}+i}, y) - \varphi_{k_1+\cdots+k_{s-1}+i} \right]^2 < \lambda_s n_e S_e^2 \qquad \ldots(14)$$

Obviously $\lambda_1 \cdots \lambda_s$ should be so chosen as to minimize the volumes of these regions put together in some way. Ghosh had intuitively suggested $\lambda_i \propto k_i$, the d.f. Later Ramachandran (1956,b) further investigated the problem and gave some simplified formulae to evaluate the integral and hence fix λ's

Due to great complexity of evaluation of evaluation of integral (13), it is felt that some approximations be used. Ghosh had suggested the use of Kimball's inequality (1951), which will give an upper bound for the joint significance level. Thus if the numerator χ^2's are independent of each other but denominator is the same χ^2, then

$$\Pr \left[\chi_1^2 / n_e S_e^2 < \lambda_1 \;\ldots\; \chi_s^2 / n_e S_e^2 < \lambda_s \right] \geq \Pr \left[\chi_1^2 / n_e S_e^2 < \lambda_1 \right] \cdots \cdot \Pr \left[\chi_s^2 / n_e S_e^2 < \lambda_s \right] \ldots$$

The right hand side component probabilities may be taken to be all equal to get,

R.H.side $(1-\alpha')^{g}$ which we may equate with $1-\alpha$ to get 'α'', the individual significance level. In matters of practical interest α may be .05 and so the λ's for the individual hypotheses will have to be found from ordinary F-tables by interpolation from 1%, 5% and .1 % levels etc.

We shall find that the inequality is conservative, implying loss of power and increase in the confidence regions. Instead we try below to actually evaluate the integral (13) for the special case of 2 groups only. In this case it will be noted that if both k_1 and k_2 are even integers, the integral (13) can be explicitly put in the form of a sum of several Incomplete B-functions (Pearson).

Thus, $$C(\cdot\cdot)\int_0^{\lambda_1}\int_0^{\lambda_2} \frac{G_1^{k_1-2/2}\; G_2^{k_2-2/2}}{(1+G_1+G_2)^{k_1+k_2+n_e/2}}\, dG_1\, dG_2 = I\{\lambda_1\ \lambda_2\} \text{ say}$$

$$= C\int_0^{\lambda_1} G_1^{q}\left[\int_0^{\lambda_2} G_2^{p}\Big/(1+G_1+G_2)^{p+q+2+\frac{n_e}{2}}\right] dG_1 \qquad \text{where } k_1 = 2p+2,\ k_2 = 2q+2$$

$$= C\int_0^{\lambda_1} G_1^{q}\left[\left\{1-\frac{1}{p+q+1+\frac{n_e}{2}}\frac{G_2^{p}}{(1+G_1+G_2)^{p+q+1+\frac{n_e}{2}}}\right\}_0^{\lambda_2} + \frac{p}{p+q+1+\frac{n_e}{2}}\int_0^{\lambda_2}\frac{G_2^{p-1}}{(1+G_1+G_2)^{p+q+1+\frac{n_e}{2}}}dG_2\right] dG_1$$

— on integration by parts.

When G_2^{p-r} vanishes, we get terms like $\int_0^{\lambda_1}\frac{G_1^{q}}{(1+\lambda_2+G_1)^{v}}$ and $\int_0^{\lambda_1}\frac{G_1^{q}}{(1+G_1)^{v}}dG_1$ which are simply Incomplete B-integrals like $B\left\{\frac{\lambda_1}{1+\lambda_1+\lambda_2},\ q+1,\ v-q-1\right\}$ or $B\left\{\frac{\lambda_1}{1+\lambda_1},\ q+1,\ v-q-1\right\}$.

Of course, actual evaluation when one of the 2 integers k_1 & k_2 is odd runs into similar lines but that of odd k_1 & k_2 is not amenable to this treatment. Hence we shall only try interpolation between the even values to get that for odd pair (k_1, k_2). Ramachandran's table (1956, b) may be consulted for the special case of $k_1 = k_2 = k$ which deals with odd k also.

Case II All this obviously fails if a set M_t say, contains vectors of more than one hypothesis. For convenience we shall first consider the case where M_t contains vectors of only 2 hypotheses H_1 & H_2 say. In many practical situations of non-orthogonality, there shall be only one set M_t containing all the hypotheses.

In this case Ghosh has shown that there exists no non-singular transformation by which the linear functions can be transformed into mutually orthogonal sets corresponding to hypotheses H_1 and H_2 respectively.

Let $U_1 \cdots U_{k_1}$; $U_{k_1+1} \cdots U_{k_1+k_2}$ be an orthonormal basis of the vector-space formed by $Y_1 \ldots Y_{k_1}$; $Y_{k_1+1} \ldots Y_{k_1+k_2}$ so that $U_1 \cdots U_{k_1}$ is a basis of the vector space formed by $Y_1 \ldots Y_{k_1}$. As before let $E(U_i, y) = \phi_i$, then ϕ_i's are linear functions of $\pi_1 \cdots \pi_{k_1}$; $\pi_{k_1+1} \cdots \pi_{k_1+k_2}$. 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 let $\tilde{\pi}_i = \pi_i / |Y_i|$ then the relation between U-vectors and Y-vectors is expressed by

$$\begin{pmatrix} \tilde{Y}_I \\ \tilde{Y}_{II} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix} \qquad \ldots(15)$$

where $\alpha\ (k_1 \times k_1)$ and $\gamma\ (k_2 \times k_2)$ are two non-singular matrices, and $\beta(k_2 \times k_1)$ is non-null, since vectors Y_I and Y_{II} are not mutually orthogonal. Also,

$$\tilde{Y}_I = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_{k_1} \end{pmatrix}, \quad \tilde{Y}_{II} = \begin{pmatrix} \tilde{y}_{k_1+1} \\ \vdots \\ \tilde{y}_{k_1+k_2} \end{pmatrix}$$

and similarly U_I and U_{II}.

On the basis of relation (15) Ghosh has derived 3 different methods of finding confidence regions for the parameters $\tilde{\pi}_1 \cdots \tilde{\pi}_{k_1}$; $\tilde{\pi}_{k_1+1} \cdots \tilde{\pi}_{k_1+k_2}$. Method a) is essentially the same as that of Scheffe' & Roy-Bose, method b) gives confidence regions for $\pi_1 \cdots \pi_{k_1}$ and $\pi_{k_1+1} \cdots \pi_{k_1+k_2}$ separately and method c) employs a singular transformation of $\begin{Bmatrix} Y_I \\ Y_{II} \end{Bmatrix}$ to arrive at sets of linear functions which are mutually orthogonal. We shall briefly describe those methods below and proceed to their relative comparisons.

We shall first note some invariance properties of β and γ. Thus consider an orthogonal transformation of the form $\begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}$ where C_1 & C_2 are orthogonal matrices of order k_1 and k_2 respectively. Suppose the original orthonormal system of vectors U_I, U_{II} is transformed by this.

Then we shall get:

$$\begin{pmatrix} \tilde{Y}_I \\ \tilde{Y}_{II} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha^* & 0 \\ \beta^* & \gamma^* \end{pmatrix} \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha^* C_1 & 0 \\ \beta^* C_1 & \gamma^* C_2 \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix} \quad \ldots(16)$$

i.e.

$$\alpha \equiv \alpha^* C_I \qquad \alpha^* = \alpha C_I'$$
$$\beta \equiv \beta^* C_I \qquad \beta^* = \beta C_I'$$
$$\gamma \equiv \gamma^* C_2 \qquad \gamma^* = \gamma C_2'$$

Thus, since rank of β remains unchanged on multiplication by a non-singular matrix, β^* has the same rank as β. Further if α were an orthogonal matrix, as we shall get below, then α^* is again so. Also γ^* still remains a non singular matrix.

a) <u>Method of simultaneous confidence intervals of all linear contrasts.</u>

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.

So far, we do not know whether $Y_1 \ldots Y_{k_1}$ are mutually orthogonal or not. For convenience we shall now assume that the hypothesis H_1(or H_2) ; $\pi_1' = 0, \ldots \pi_{k_1}' = 0$ has been formulated in such a way that these linear functions $Y_1 \ldots Y_{k_1}$ become orthogonal and similarly those of other hypothesis viz: $Y_{k_1+1} \ldots \ldots y_{k_1+k_2}$

That this reformulation of the problem does not change it can be easily seen for, if all k_1 independent linear functions of $\pi_1' \ldots \pi_{k_1}'$ (which are expected values of $Y_1 \ldots Y_{k_1}$) are equal to zero so are $\pi_1 \ldots \pi_{k_1}$ and vice-versa.

Now from relation (15) we have

$$\sum_1^{k_1} \alpha_{ni}^2 = 1 \qquad (n = 1, \ldots k_1)$$

$$\text{and } \sum_1^{k_1} \beta_{ni}^2 + \sum_1^{k_2} \gamma_{ni}^2 = 1 \qquad (n = 1, \ldots k_2) \qquad \ldots(17)$$

Actually under the reformulation α becomes an orthogonal matrix and may even be taken as unit matrix I_{k_1}.

$$\text{Further} \quad (\tilde{Y}_n, y) = \sum_1^{k_1} \alpha_{ni} (U_i, y) \qquad n = 1, \ldots k_1$$

$$(\tilde{Y}_{k_1 n}, y) = \sum_1^{k_1} \beta_{ni} (U_i, y) + \sum_1^{k_2} \gamma_{ni} (U_{k_1+i}, y) \qquad n = 1, \ldots k_2 \qquad \ldots(18)$$

and similar equations hold between π's and φ's.

From (17) and (18) after application of Schwartz's inequality we get:

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

hence

$$\Pr\left[\,|(\tilde{Y}_n, y) - \tilde{\pi}_n| \leq \delta \text{ for all } n = 1, \ldots k_1 + k_2\right] \geq \Pr\left[\,\sum [(U_j, y) - \varphi_j]^2 \leq \delta^2\right] \qquad \ldots(19)$$

If $\Pr\left[\,\sum [(U_j, y) - \varphi_j]^2 \leq \delta^2\right]$ is fixed at $1-\alpha$, then with confidence coefficient $1-\alpha$, we get the set of simultaneous linear intervals.

$$(\tilde{Y}_n, y) - \delta_\alpha \leq \tilde{\pi}_n < (\tilde{Y}_n, y) + \delta_\alpha \qquad \text{for all } n = 1, \ldots k_1 + k_2. \qquad \ldots(20)$$

For testing H_1 and H_2 we now use the $\chi^2 = \frac{1}{\sigma^2} \sum [(U_j, y) - \varphi_j]^2$ and test against $n_e s_e^2 / \sigma^2$; whence δ^2 will be found to be equal to $(k_1 + k_2) F_{1-\alpha;\, k_1 + k_2, n_e} s_e^2$.

In fact, if we want to specify also as to which hypothesis, out of H_1 & H_2 is accepted (rejected), we may use two test statistics $T_1 = \sum_1^{k_1} (\tilde{Y}_n, y)^2 / n_e s_e^2$, and $T_2 = \sum_{k_1+1}^{k_1+k_2} (\tilde{Y}_n, y)^2 / n_e s_e^2$. From (18) by taking linear functions of the first k_1 $\tilde{Y}$'s and second k_2 $\tilde{Y}$'s separately and maximizing over the coefficients of linear functions we shall find that if $\pi_1 = 0, \ldots \ldots \pi_{k_1+k_2} = 0$

$$\left.\begin{array}{l} T_1(n_e s_e^2) \leq \sum_1^{K_1+K_2} (U_{j,y})^2 \\ \text{and } T_2(n_o s_e^2) \leq \sum_1^{K_1+K_2} (U_{j,y})^2 \end{array}\right\} \quad \ldots(21)$$

These 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 level of significance will be provided by the distribution of $\sum_1^{K_1+K_2} (U_{j,y})^2$ so that the critical limits for T_1 and T_2 both are the same i.e.

$$\lambda = k_1+k_2/n_e \quad F_{1-\alpha\,,\,k_1+k_2,\,n_e}$$

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

This method besides giving the confidence region from the inequality $\sum_1^{K_1+K_2} [(U_{j,y}) - \varphi_i]^2 \leq \delta^2$, also gives the confidence intervals for any linear function of Π's, with the same confidence coefficient $1-\alpha$. For, consider $\sum_1^{K_1+K_2} \alpha_i \tilde{\pi}_i$, a contrast involving parameters from both sets. We have $V(Y_i, y) = \sigma^2$ for all $i = 1, \ldots k_1+k_2$.

$$\text{Var}\left[\sum \alpha_i (Y_{i,y})\right] = \sum \alpha_i^2 \cdot \sigma^2 + \sum_{i\neq j=1}^{K_1+K_2} (\alpha_i \cdot \alpha_j) \; \text{Cov}\{(\tilde{Y}_i, y)(\tilde{Y}_j, y)\}$$

$$= \sigma^2\left[\sum \alpha_i^2 + \sum_{i\neq j=1}^{K_1+K_2} (\alpha_i \cdot \alpha_j) \; \text{Cov}\{\tilde{Y}_i \cdot \tilde{Y}_j\}\right] = \sigma^2 K \quad \text{say}$$

where $(\tilde{Y}_i, \tilde{Y}_j)$ is the scalar product. Then the confidence interval for $\sum_1^{K_1+K_2} \alpha_i \tilde{\Pi}_i$ shall be:

$$\sum \alpha_i(\tilde{Y}_i\, y) - \delta\sqrt{K} \leq \sum_1^{K_1+K_2} \alpha_i \tilde{\Pi}_i \leq \sum \alpha_i(\tilde{Y}_i, y) + \delta\sqrt{K} \quad (*) \quad \ldots\ldots(22)$$

If we confine only to contrasts in either $\tilde{\pi}_1 \ldots \tilde{\pi}_K$ or in $\tilde{\pi}_{K_1+1} \ldots \tilde{\pi}_{K_1+K_2}$, the expression simplifies considerably giving the confidence interval for $\sum_1^{K_1} l_i \tilde{\pi}_i$, say as

$$\sum l_i(\tilde{Y}_i, y) - \delta\sqrt{\sum l_i^2} \leq \sum_1^{K_1} l_i \tilde{\pi}_i < \sum l_i(\tilde{Y}_i, y) + \delta\sqrt{\sum l_i^2} \quad \ldots(23)$$

(*) The result given by Ghosh is true only for the particular case (23) and does not not work for the general situation as implied in the paper. It may be noted that much of the supposed simplicity of method a) is taken away by this correction.

Another method of Ghosh consists in using Schwartz's inequality in 2 stages:

Method b): For the first k_1 parameters, the confidence region, and hence the confidence intervals is given by

$$\sum_1^{k_1} [(U_n, y) - \varphi_n]^2 \leq c_1$$

For the IInd group of parameters we have

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

Hence $$[(\tilde{Y}_{k_1+n}, y) - \tilde{\pi}_{k_1+n}]^2 \leq \sum_1^{k_1+k_2} [(U_i, y) - \varphi_i]^2 \quad n = 1, \ldots k_2$$

which gives the confidence intervals for the parameters $\pi_{k_1+1}, \ldots \pi_{k_1+k_2}$.

Take a linear normalized function of $\tilde{Y}_{k_1+1}, \ldots, Y_{k_1+k_2}$; $\sum_1^{k_2} d_i \tilde{Y}_{k_1+i}$, say.

$$\tilde{Y}_d = \sum_1^{k_2} d_i \tilde{Y}_{k_1+i} = \beta_1' U_1 \ldots + \beta_{k_1}' U_{k_1} + \gamma_1' U_{k_1+1} + \ldots + \gamma_{k_2}' U_{k_1+k_2} \qquad \ldots\ldots(24)$$

As $\tilde{Y}_d$ is normalized (i.e. $\sum d_i^2 = 1$), we must have $\sum_1^{k_1} \beta_i'^2 + \sum_1^{k_2} \gamma_i'^2 = 1$ and application of Schwartz's inequality gives

$$\left[\sum_1^{k_2} d_i (\tilde{Y}_{k_1+i}, y)\right]^2 \leq \left[\sum_1^{k_1+k_2} (U_i, y)^2\right] \qquad \ldots\ldots(25)$$

If we take the maximum over the left hand side for varying d vectors we shall get

$$\left[\sqrt{\sum_1^{k_2} (Y_{k_1+i}, y)^2}\right]^2 \leq \sum_1^{k_1+k_2} (U_i, y)^2$$

i.e. $$\sum_1^{k_2} (Y_{k_1+i}, y)^2 \leq \sum_1^{k_1+k_2} (U_i, y)^2$$

and in general

$$\sum_1^{k_2} [(Y_{k_1+i}, y) - \tilde{\pi}_{k_1+i}]^2 \leq \sum_1^{k_1+k_2} [(U_i, y) - \varphi_i]^2 \quad (*) \qquad \ldots\ldots(26)$$

(*) Ghosh had actually given a more conservative inequality than (26), by starting from $\gamma^{-1} Y_{II}$ and normalizing the k_2 individual components. As is clear, that transformation is totally unnecessary. In fact the inequality given there for IInd group of parameters is so conservative that method b) would always be far inferior to method a) specially for the contrasts of IInd group of parameters and the corresponding test of hypothesis.

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

$$Pr\left[\sum_{1}^{K_1}\left\{(U_n, y) - \varphi_n\right\}^2 \le C_1, \sum_{1}^{K_1+K_2}\left\{(U_i, y) - \varphi_i\right\}^2 \le C_2\right] = 1-\alpha \quad \ldots\ldots(27)$$

Instead of solving this integral, Ghosh had suggested the use of the joint distribution of $\sum_{1}^{K_1}\left[(U_n, y) - \varphi_n\right]^2$ and $\sum_{1}^{K_2}\left[(U_{k_1+n}, y) - \varphi_{K_1+n}\right]^2$ i.e. those of G_1 and G_2, in evaluating λ_1 and λ_2 and then putting $\lambda_1 = C_1$ and $\lambda_2 + \lambda_1 = C_2$.

It will be noted that this gives an upper bound of the joint significance level. Further, as he suggested, if we use Kimball's inequality, C_1 and C_2 will merely become:

$$C_1 = k_1 s_e^2 \, F_{1-\alpha', k_1, n_e} \qquad C_2 = s_e^2\left[k_1 F_{1-\alpha', k_1, n_e} + k_2 F_{1-\alpha'', k_2, n_e}\right] \quad \ldots\ldots(28)$$

where $(1-\alpha')(1-\alpha'') = 1-\alpha$; α' and α'' may be taken to be equal.

As in method a) we could have started with an 'orthogonal set' consisting of 3 hypotheses H_1, H_2, H_3 say. The equation (27) will immediately be converted to:

$$Pr\left[\sum_{1}^{K_1}\left[(U_n, y) - \varphi_n\right]^2 \le C_1, \sum_{1}^{K_1+K_2}\left[(U_i, y) - \varphi_i\right]^2 \le C_2, \sum_{1}^{K_1+K_2+K_3}\left[(U_j, y) - \varphi_j\right]^2 \le C_3\right] = 1-\alpha \quad \ldots\ldots(29)$$

Further, if as suggested by Ghosh we use the joint distribution of G_1, G_2 and G_3 on k_1, k_2 and k_3 d.f.'s respectively, we shall get.

$$C_1 = k_1 s_e^2 \, F_{1-\alpha', k_1, n_e} \qquad C_2 = s_e^2\left[k_1 F_{1-\alpha', k_1, n_e} + k_2 F_{1-\alpha'', k_2, n_e}\right]$$

$$\text{and } C_3 = s_e^2\left[k_1 F_{1-\alpha', k_1, n_e} + k_2 F_{1-\alpha'', k_2, n_e} + k_3 F_{1-\alpha''', k_3, n_e}\right] \quad \ldots\ldots(30)$$

where $(1-\alpha')(1-\alpha'')(1-\alpha''') = 1-\alpha$

We shall have the following confidence regions for the 3 groups of parameters.

$$(\pi_1 \ldots \pi_{K_1}), \quad \sum_{1}^{K_1}\left[(\tilde{Y}_i, y) - \tilde{\pi}_i\right]^2 \le \sum_{1}^{K_1}\left[(U_i, y) - \varphi_i\right]^2 \le C_1$$

$$(\pi_{K_1+1} \ldots \pi_{K_1+K_2}); \quad \sum_{1}^{K_2}\left[(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{K_1+i}\right]^2 \le \sum_{1}^{K_1+K_2}\left[(U_i, y) - \varphi_i\right]^2 \le C_2$$

$$(\pi_{K_1+K_2+1} \ldots \pi_{K_1+K_2+K_3}); \quad \sum_{1}^{K_3}\left[(\tilde{Y}_{k_1+k_2+j}, y) - \tilde{\pi}_{K_1+K_2+j}\right]^2 \le \sum_{1}^{K_1+K_2+K_3}\left[(U_j, y) - \varphi_j\right]^2 \le C_3 \quad \ldots\ldots(31)$$

where C_1, C_2 and C_{33} are given by (30)

This method gives all confidence intervals for all linear combinations of $\pi_1 \cdots \pi_{k_1}$ or $\pi_{k_1+1} \cdots \pi_{k_1+k_2}$ or $\pi_{k_1+k_2+1} \cdots \pi_{k_1+k_2+k_3}$ with the joint confidence coefficient $1-\alpha$, but cannot give any confidence interval for a linear combination of π's of more than one group of parameters. But this drawback is not serious for in practice we rarely require a confidence statement concerning more than 1 group of parameters.

A modification of (31) is possible if the experimenter thinks that contrasts involving parameters of group I and II may be interesting but none involving those of group III with those of I or II. Method a) may then be applied to groups I and II and then this can be joined to group III by means of a simple application of method b). Thus the equation (31) will change to:

$$(\pi_1 \cdots \pi_{k_1+k_2}) : \sum_1^{k_1+k_2} \left[(U_i, y) - \varphi_i \right]^2 \leq C_2'$$

$$(\pi_{k_1+k_2+1} \cdots \pi_{k_1+k_2+k_3}) : \sum_1^{k_3} \left[(Y_{k_1+k_2+j}, y) - \pi_{k_1+k_2+j} \right]^2 \leq \sum_1^{k_1+k_2+k_3} \left[(U_i, y) - \varphi_i \right]^2 \leq C_3' \qquad \ldots\ldots(32)$$

Where C_2' and C_3' may be taken as (conservative):

$$C_2' = (k_1+k_2)\, s_e^2\, F_{1-\alpha_1,\, k_1+k_2,\, n_e} \qquad C_3' = s_e^2 \left[(k_1+k_2)\, F_{1-\alpha_1, k_1+k_2, n_e} + k_3 F_{1-\alpha_2, k_3, n_e} \right] \qquad \ldots\ldots(33)$$

Where $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$.

For the test of the hypotheses we can proceed from (31).

Thus for H_1: $\pi_1 = 0, \ldots\ldots\ldots, \pi_{k_1} = 0$.

H_2: $\pi_{k_1+1} = 0, \ldots\ldots\ldots\ldots, \pi_{k_1+k_2} = 0$.

H_3: $\pi_{k_1+k_2+1} = 0, \ldots\ldots\ldots\ldots, \pi_{k_1+k_2+k_3} = 0$.

We consider

$$\chi_1^2 = \frac{1}{\sigma^2} \sum_1^{k_1} (U_i, y)^2 = \frac{1}{\sigma^2} \sum_1^{k_1} (Y_i, y)^2 \qquad F_1 = \chi_1^2 / k_1 s_e^2$$

$$\chi_2^2 = \frac{1}{\sigma^2} \sum_{k_1+1}^{k_1+k_2} (U_i, y)^2 \qquad F_2 = \chi_2^2 / k_2 s_e^2$$

$$\chi_3^2 = \frac{1}{\sigma^2} \sum_{k_1+k_2+1}^{k_1+k_2+k_3} (U_i, y)^2 \qquad F_3 = \chi_3^2 / k_3 s$$

since ρ_i $(i=1,\ldots,k_1+k_2+k_3) = 0$ on the given hypotheses.

If we use $\chi_1'^2 = \sum_1^{k_1} (Y_i, y)^2$ $\chi_2'^2 = \sum_{k_1+1}^{k_1+k_2} (Y_i, y)^2$ and $\chi_3'^2 = \sum_{k_1+k_2+1}^{k_1+k_2+k_3} (Y_i, y)^2$ as the three test statistics, then we shall reject the hypotheses H_i $(i=1,2,3)$ when $\chi_i'^2 > C_i$.

The first kind of error of this simultaneous test will not be exactly 'α' but less, because of the inequalities (31) as also due to approximate solution of integral (29). But, as is clear from the definition of test-statistics $\chi_1'^2$, $\chi_2'^2$ and $\chi_3'^2$ the 3 component tests will be quasi-independent.

<u>Method c):</u> In this case if k_1 and k_2 are unequal, say $k_2 > k_1$ we can find orthogonal sets of k_1 linear functions from Ist set and (k_2-k_1) linear functions from the second set for the respective hypotheses H_1 and H_2. Then, we can either exactly evaluate the simultaneous level from the joint distribution of G_1 and G_2 or else use Kimball's Inequality to get an upper bound of the simultaneous level.

Let $U_1 \ldots\ldots U_{k_1}$; $U_{k_1+1} \ldots\ldots U_{k_1+k_2}$ and matrices Y_I and Y_{II} have same meaning as before. Since the rank of β is at most k_1, we can operate on $\tilde{Y}_{II} = \beta\, U_I + \gamma\, U_{II}$ in such a way that the matrix β is reduced to a Triangular matrix with its lower (k_2-k_1) rows containing all zeros. This will be done by linear transformations on the vectors of Y_{II} and we shall get like:

$$\tilde{Y}'_{k_1+1} = \beta'_{11} U_1 + \beta'_{12} U_2 \;\ldots\; \beta'_{1k_1} U_{k_1} + \gamma'_{11} U_{k_1+1} \;\ldots\; + \gamma'_{1k_2} U_{k_1+k_2}$$

$$\tilde{Y}'_{k_1+2} = 0 \quad \beta'_{22} U_2 \;\ldots\; \beta'_{2k_1} U_{k_1} + \gamma'_{21} U_{k_1+1} \;\ldots\; + \gamma'_{2k_2} U_{k_1+k_2} \;\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\tilde{Y}'_{2k_1+1} = 0 \quad\quad 0 + \gamma'_{k_1+1,1} U_{k_1+1} \;\ldots\; + \gamma'_{k_1+1,k_2} U_{k_1+k_2}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\tilde{Y}'_{k_1+k_2} = \quad\quad \gamma'_{k_2 1} U_{k_1+1} \;\ldots\; + \gamma'_{k_2 k_2} U_{k_1+k_2} \qquad \ldots\ldots\ldots (34)$$

Let the orthonormal basis of the vector-space formed by $Y'_{2K_1+1}, \dots, Y'_{K_1+K_2}$, the vectors free of first k_1 U's, be $W_1, \dots, W_{k_2-k_1}$. Then W's being linear functions of $Y'_{2K_1+1}, \dots, Y'_{K_1+K_2}$ viz. of the original $\tilde{Y}_{K_1+1}, \dots, \tilde{Y}_{K_1+K_2}$, the expectations will be :

$$E(W_i, y) = \sum_1^{K_2} \xi_{ij} \tilde{\pi}_{K_1+j} \qquad \text{say, } i = 1, \dots, k_2-k_1.$$

Now $\frac{1}{\sigma^2}\sum_1^{k_2-k_1} [(W_i, y) - \psi_i]^2$ has a χ^2 distribution on (k_2-k_1) d.f. and is independent of that of $1/\sigma^2 \sum_1^{K_1} [(U_i, y) - \varphi_i]^2$ arising from the first k_1 parameters. Then the confidence regions for the 2 sets of parameters will be given by,

$$(\pi_1 \dots \pi_{K_1}) : \sum_1^{K_1} [(\tilde{Y}_i, y) - \tilde{\pi}_i]^2 = \sum_1^{K_1} [(U_i, y) - \varphi_i]^2 \le C_1$$

$$\text{and } (\pi_{K_1+1} \dots \pi_{K_1+K_2}) : \sum_{j=1}^{K_2-K_1} [(W_j, y) - \psi_j]^2 \le C_2 \qquad \dots\dots(36)$$

where C_1 and C_2 are so determined that the probability,

$$\Pr [\chi^2_{K_1} / n_e S_e^2 \le \lambda_1 , \; \chi^2_{K_2-K_1} / n_e S_e^2 \le \lambda_2] = 1-\alpha \quad \text{whence}$$

$C_1 = \lambda_1 n_e S_e^2$ $C_2 = \lambda_2 n_e S_e^2$ for exact evaluation of levels; otherwise, using Kimball's inequality,

$$C_1 = K_1 S_e^2 \; F_{1-\alpha';k_1,n_e} \qquad C_2 = (k_2-k_1) S_e^2 \; F_{1-\alpha'';k_2-k_1,n_e} \qquad \dots(37)$$

where $(1-\alpha')(1-\alpha'') = 1-\alpha$.

Note that ψ_i's have rank less than k_2 and thus do not represent all linear functions of the $\tilde{\pi}_{K_1+j}$ $(j = 1, \dots, k_2)$. Thus from (36) one can get the confidence intervals only for certain linear combinations of $\tilde{\pi}_{K_1+1} \dots \tilde{\pi}_{K_1+K_2}$, the number of independent combinations being just (k_2-k_1). This may serve for certain types of experiments, but there is a serious drawback in that the linear combinations cannot be chosen by the experimenter but depend essentially on the matrices β and γ.

As for the test of the joint hypotheses H_1 & H_2, we see that H

-2499-

H_2: $\pi_{k_1+1} = 0 \dots\dots\dots \pi_{k_1+k_2} = 0$ implies

H_2': $\psi_{k_1+1} = 0 \dots\dots\dots \psi_{k_1+k_2} = 0$

Hence rejection of H_2' implies that of H_2 and so the power of the test is reduced at the cost of increased stringency in the Ist kind of error. Further the power of the simultaneous test, i.e. using the test statistics $\sum_1^{k_1} (U_i, y)^2$ and $\sum_1^{k_2-k_1} (W_i, y)^2$, will be such that it is insensitive to certain kinds of alternatives viz: Those which cannot be expressed as linear combinations of $\psi_1 \cdots \psi_{k_2-k_1}$. But it may be hoped that for (k_2-k_1) not too small, the above tests can be used for they are better for certain class of alternatives, whose rank is (k_2-k_1) instead of the full k_2.

Relative merits and certain comparisons of the 3 methods

Two methods of comparison immediately suggest themselves

i) In terms of 'volume' of the confidence regions in $\pi_1 \cdots \pi_{k_1}$; $\pi_{k_1+1} \cdots \pi_{k_1+k_2}$

ii) In terms of length of confidence interval of π's or any standard function of them.

In i) we shall find the ellipsoids of $(k_1 + k_2)$ dimensions in method a) and product of 2 ellipsoids of dimensions k_1 & k_2 in method b). But the ratio of those 2 volumes does not give a valid comparison for, as we know in case of 2 dimensions the area of a circle is smaller than that of a square of a side equal to the diameter. Volume is a overall measure of a confidence region while we need comparison in parts, for group I parameters alone and for group II parameters alone but not together. Thus it seems desirable to compare only in terms of confidence intervals for a linear function of either of the groups of parameters separately but not together.

a) Consider a linear function of $\pi_1 \cdots \pi_{k_1}$.

Method a) gives with $(1-\alpha)$ confidence the intervals

$(\tilde{Y}_n, y) - \delta_\alpha \leq \tilde{\pi}_n \leq (\tilde{Y}_n, y) + \delta_\alpha$

$$1/2 \text{ length } = \delta_\alpha = \sqrt{(k_1+k_2)\ F_{1-\alpha,k_1+k_2,n_e}\ s_e^2} \qquad \ldots\ldots(39)$$

and that of a normalized contrast of $\tilde{\pi}$'s, say $\sum_1^{k_1} \alpha_i \tilde{\pi}_i$ $(\sum \alpha_i^2 = 1)$ is the same.

From method b), equation (27) yields the confidence region $\sum_1^{k_1} [(\tilde{Y}_n y) - \hat{\pi}_n]^2 \leq C_1$ whence the confidence interval for $\tilde{\pi}_n$ or any normalized contrast of π's is known to have the 1/2 length,

$$\sqrt{C_1} = \sqrt{k_1\ F_{1-\alpha', k_1, n_e}\ s_e^2} \quad \text{(if Ghosh's suggestion is followed)}. \qquad \ldots\ldots(40)$$

The length as obtained in method c) is exactly as in method b) except that ' ' may change. Suppose $1-\alpha' = \sqrt{1-\alpha}$ and let $\alpha = 0.05$, then

$$\delta_\alpha / \sqrt{C_1} = \sqrt{k_1+k_2 / k_1 \cdot F_{.95, k_1+k_2, n_e} / F_{.975, k_1, n_e}} \qquad \ldots\ldots(41)$$

we are giving some representative values of this ratio in table I below.

Table 1 : Values of δ_α^2/C_1 for different triads of (k_1, k_2, n_e) and $\alpha = .05$

Constants																	
k_1	2	2	2	3	5	5	4	6	8	8	9	5	8	8	5	2	3
k_2	2	2	2	5	5	1	8	6	2	4	3	1	2	4	5	2	5
n_e	12	3	2	12	12	6	20	30	12	30	30	∞	∞	∞	∞	α	α
$F_{.95, k_1+k_2, n_e}$	3.26	9.12	19.25	2.85	2.76	4.28	2.28	2.09	2.76	2.09	2.09	2.09	1.83	1.75	2.37 (1.83)	1.9[illegible]	
$F_{.975, k_1, n_e}$	5.10	15.04	39.00	4.47	3.89	5.99	3.51	2.87	3.51	2.65	2.57	2.60	2.19	2.19	2.57	3.69	3.1[illegible]
Ratio $\frac{\delta_\alpha^2}{C_1}$	1.22	1.12	0.99	1.70	1.42	1.86	1.93	1.35	1.00	1.18	1.09	0.97	1.04	1.20	1.42	1.28	1.[illegible]

Thus we see that method a) is always inferior to b) or c) except for cases like (5,1). The superiority of b) or c) increases as k_2 increases faster than k_1, but we shall see later that when $k_2 > k_1$, our advantage in using the d) is reduced in overall measure. Further the ratio depends on n_e rather heavily which is not very good, specially for smaller k_1 and k_2 but is stabilized after n_e is moderately large e.g. > 30.

b) Consider a linear function which can be estimated by confidence intervals by all 3 methods. Obviously we should take ψ_i's of equation (31). Considering a normalized linear function of ψ_i's viz.

$$\sum_{1}^{k_2-k_1} l_i \psi_i \qquad (\Sigma l_i^2 = 1) \qquad \ldots\ldots(42)$$

equation (36) gives the half length of confidence interval as $\sqrt{C_2'}$ where C_2' may be taken from equation (37) or (38) i.e.

$$C_2' = \lambda_2(k_1, k_2-k_1) \cdot n_e S_e^2 \quad \text{or} \quad = (k_2-k_1)\, F_{\sqrt{1-\alpha}, k_2-k_1, n_e}. \qquad \ldots\ldots(43)$$

As $W_1 \ldots W_{k_2-k_1}$ can be expressed as linear functions of $\widetilde{Y}_{k_1+1} \ldots, \widetilde{Y}_{k_1+k_2}$ so also $\sum_{1}^{k_2-k_1} l_i W_i$ corresponding to (42). Now, as $\Sigma l_i W_i$ is a normalized vector, the corresponding linear function in $\widetilde{Y}_{k_1+1} \cdots \widetilde{Y}_{k_1+k_2}$ will also be normalized i.e. $\sum_{1}^{k_2} l_j'^2$ of $\sum_{1}^{k_2} l_j' \widetilde{Y}_{k_1+j} \equiv \sum_{1}^{k_2-k_1} l_i W_i$, should be equal to 1. Thus from method a) the half length of confidence interval for (42) shall be simply δ_α, while from method b) using equations (27) and (28) the half length shall be $\sqrt{C_2}$, where following Ghosh's suggestion

$$C_2 = S_e^2 \left[k_1 F_{\sqrt{1-\alpha}, k_1, n_e} + k_2 F_{\sqrt{1-\alpha}, k_2, n_e} \right] \qquad \ldots\ldots(44)$$

Comparing C_2 with $\delta_\alpha^2 = S_e^2 \left[(k_1+k_2) F_{1-\alpha, k_1+k_2, n_e}\right]$ we find that C_2 is always $\geq \delta_\alpha^2$ for the F-table values increase both as n_1, the d.f.'s of numerator decrease and percentage point (α) falls. Hence method b) will always give higher half-lengths than method a) and so it may be desirable in many cases to keep k_1 always $\geq k_2$ so that at least for the bigger group of parameters, the method b) be superior to a).

Comparison of methods a) and c) means comparison of δ_α^2 with $C_2' = \lambda_2\, n_e S_e^2$ (exactly)
$= (k_2-k_1)\, F_{\sqrt{1-\alpha}, k_2-k_1, n_e} S_e^2$

$$\text{Now,}\quad \left\{ \frac{\text{Half-length for method a}}{\text{Half-length for method c}} \right\}^2 = \frac{\delta_\alpha^2}{C_2^2} = \frac{k_1+k_2}{k_2-k_1} \cdot \frac{F_{1-\alpha,\,k_1+k_2,\,n_e}}{F_{1-\frac{\alpha}{2},\,k_2-k_1,\,n_e}} \qquad \ldots\ldots(45)$$

Table 2 shows the values of this ratio for some triads (k_1, k_2, n_e). These may be compared with those of Table 3 where exact value of $C_2^{\delta_\alpha}$ is used.

Table 2 : Values of δ_α^2/C_2 $\alpha = .05$

Constants														
k_1	6	4	3	3	2	1	1	1	1	2	2	3	4	3
k_2	10	8	6	5	10	11	11	2	7	18	10	6	8	5
n_e	30	20	12	12	4	12	30	30	14	12	∞	∞	∞	∞
$F_{.95,k_1+k_2,n_0}$	1.99	2.28	2.80	2.85	5.91	2.69	2.09	2.92	2.70	2.54	1.75	1.88	1.75	1.94
$F_{.975,k_2-k_1,n_0}$	3.25	3.51	4.47	5.10	8.98	3.37	2.59	5.57	3.50	3.16	2.19	3.12	2.79	3.69
Ratio δ_α^2/C_2^2	2.47	1.93	1.88	2.52	1.00	0.96	1.00	1.53	1.03	1.01	1.20	1.81	1.88	2.10

From this table it becomes clear that method c) is superior to a) provided we confine only to those parametric functions which can be estimated by c). Since a) is shown to be better than b) it means c) is the best under these circumstances. Further as is clear from the extention to $n_e = \infty$, the gain in method c) increases with increasing n_e.

If we use exact values of c_2^*, obtained from the joint distribution of $G_1(k_1)$ and $G_2(k_2-k_1)$ the advantage is further increased. Let the limits for G_1 and G_2 be the same, then this single 'λ' when compared with single δ_α will give the overall idea of the superiority of c) over a) whenever both are applicable in a problem.

$$\hat{\sigma}_a^2/\sigma_2' = (k_1+k_2)\,F_{1-\alpha,\,k_1+k_2,\,n_0}\cdot s_e^2 \,/\, \lambda\cdot n_e\theta_e^2 = k_1+k_2/n_e\cdot F_{1-\alpha,\,k_1 k_2,\,n_e}$$

$$= \lambda_a/\lambda_c \text{ say.} \qquad .....(46)$$

Table 3) gives the values of λ_a/λ_c for such values of k_1 & k_2 for which computation of λ_c is easy viz. where k_1 and k_2 are even integers.

Table 3 Values of λ_a/λ_c $\alpha = .05$

Constants	1	2	3	4	5	6	7	8	9	10	11	12
k_2	10	10	10	10	10	10	10	10	8	8	8	8
k_1	6	6	6	6	4	4	4	4	6	6	6	6
n_0	8	12	20	α	8	12	20	α	6	8	12	α
λ_a	6.41	3.47	1.74		5.65	3.08	1.55		9.22	5.65	3.08	
$n_e\lambda_a$				26.30				23.69				23.69
λ_c	2.98	1.62	0.83		2.98	1.62	0.83		4.40	2.75	1.52	
$n_e\lambda_c$				13.21				13.21				12.74
λ_a/λ_c	2.15	2.14	2.10	1.99	1.90	1.90	1.87	1.79	2.10	2.06	2.02	1.86

Continued ......

Constants	1	2	3	4	5	6	7	8	9
k_2	8	8	8	6	6	6	6	4	4
k_1	2	2	2	2	2	4	4	2	2
n_0	8	12	α	12	α	12	α	12	α
λ_a	4.19	2.29		1.90		2.31		1.50	
$n_e\lambda_a$			18.31		15.51		18.31		12.59
λ_c	2.75	1.52		1.15		1.15		0.84	
$n_e\lambda_c$			12.74		9.87		9.87		7.38
λ_a/λ_c	1.52	1.50	1.43	1.65	1.58	2.01	1.86	1.79	1.71

The ratio nowhere falls below 1.50 and is infact many times as high as 2.0. But this is not the optimum way of computing λ_c, for as suggested by Ghosh (1955). λ_1 may be taken to be proportional to k_1 and λ_2' proportional to

(k_2-k_1); but then some method has to be found for combining the 2 ratios λ_a/λ_1 and λ_a/λ_1 and getting an overall picture. This suggests that exact computations in case of method b) may also give good results compared to a) and we proceed to give below a modification of b).

Modification: i) We had the equation (27),

$$\Pr\left\{\sum_1^{k_1}\left[(\tilde{Y}_n,y)-\tilde{\pi}_n\right]^2=\sum_1^{k_1}\left[(U_i,y)-\varphi_i\right]^2\le C_1,\ \sum_1^{k_1+k_2}\left[(U_i,y)-\varphi_i\right]^2\le C_2\right\}=1-\alpha$$

along with inequality,

$$\sum_1^{k_2}\left[(\tilde{Y}_{k_1+n},y)-\tilde{\pi}_{k_1+n}\right]^2\le\sum_1^{k_1+k_2}\left[(U_i,y)-\varphi_i\right]^2=\chi^2_{k_1}+\chi^2_{k_2}\quad\text{say.}\qquad\ldots\ldots\ldots(47)$$

where $\chi^2_{k_1}=\sum_1^{k_1}\left[(U_i,y)-\varphi_i\right]^2$ etc.

Let $G_1=\chi^2_{k_1}/n_eS_e^2$ $\quad G_2=\chi^2_{k_2}/n_eS_e^2$, then the joint distribution of G_1 and G_2 is

$$C(k_1,k_2,n_e)\ \frac{G_1^{\frac{k_1-2}{2}}\ G_2^{\frac{k_2-2}{2}}}{(1+G_1+G_2)^{\frac{k_1+k_2+n_e}{2}}}\ dG_1\ dG_2$$

Transforming to $Z_1=G_1$, $Z_2=G_1+G_2$ we get

$$C(k_1,k_2,n_e)\ \frac{Z_1^{k_1-2/2}\ (Z_2-Z_1)^{k_2-2/2}}{(1+Z_2)^{\frac{k_1+k_2+n_e}{2}}}\ dZ_1\ dZ_2\qquad\ldots\ldots\ \ldots\ldots(48)$$

and we need constants λ_1 and λ_2 such that, according to equation (47)

$$\Pr\left[Z_1<\lambda_1,\ Z_2<\lambda_2\right]=1-\alpha\qquad\text{which gives}$$

$$\Pr\left\{\sum_1^{k_1}\left[(\tilde{Y}_n,y)-\tilde{\pi}_n\right]^2\le\lambda_1 n_eS_e^2,\quad\sum_1^{k_2}\left[(\tilde{Y}_{k_1+n},y)-\tilde{\pi}_{k_1+n}\right]^2\le\lambda_2 n_eS_e^2\right\}\ge 1-\alpha\qquad\ldots\ldots(49)$$

Thus even the exact evaluation of (48) gives only, an upper bound of the simultaneous significance level and chance of improvement still remains. The lengths of confidence intervals for a normalized linear function of $\pi_1,\ldots\pi_{k_1}$ and $\pi_{k_1+1},\ldots\ldots\pi_{k_1+k_2}$ separately are proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$ respectively. In many cases we do not want to have λ_1 and λ_2 very unequal for the distinction between 'first' group of parameters and the 'second' is only arbitrary. From the joint distribution it may be shown that as λ_1 increases λ_2, as a function of λ_1, decreases for constant $1-\alpha$. In many cases it may be possible to specify the relative ratio of λ_1 and $\lambda_2(\lambda_1)$, depending on the prior knowledge the experimenter has about the two groups. Thus let $\sqrt{\lambda_1} = q\sqrt{\lambda_2(\lambda_1 \cdot 95)}$. From the point of view of optimality of the 2 lengths it may be useful to fix 'q' at k_1/k_1+k_2 , borring any prior knowledge about the 2 groups. It will be seen that fixing 'q' in the beginning helps ease the heavy computations. Now consider the equation:

$$1-\alpha = C(k_1,k_2,n_e)\int_0^{\lambda_1}\int_{z_1}^{\lambda_2} \frac{z_1^{\frac{k_1-2}{2}}\;(z_2-z_1)^{\frac{k_2-2}{2}}}{(1+z_2)^{\frac{k_1+k_2+n_e}{2}}}\,dZ_1,\,dZ_2 \qquad \ldots\ldots(50)$$

Limits of Z_2 being from Z_1 to λ_2 for $Z_2 \geq Z_1$ always.

For convenience we shall take k_1 and k_2 both as even integers. In this case $(Z_2-Z_1)^{k_2-2/2}$ can be expanded to yield $k_2/2$ terms, each of which along with $Z_1^{k_1-2/2} \Big/ (1+z_2)^{\frac{k_1+k_2+n_e}{2}}$ will give an Incomplete B-function . Taking a typical term,

$$f(\lambda_1\,\lambda_2) = C\int^{\lambda_1} z_1^{r}\int_{z_1}^{\lambda_2} z_2^{s}\Big/(1+z_2)^{\frac{k_1+k_2+n_e}{2}}\;dZ_2\;dZ_1$$

$$= C\left[\frac{z_1^{r+1}}{r+1}\int_{z_1}^{\lambda_2} z_2^{s}\Big/(1+z_2)^{\frac{k_1+k_2+n_e}{2}}\,dz_2\right]_0^{\lambda_1} + C\int_0^{\lambda_1}\frac{z_1^{r+1}}{r+1}\cdot z_1^{s}\Big/(1+z_1)^{\frac{k_1+k_2+n_e}{2}}\,dZ_1 \qquad \ldots\ldots(51)$$

(on integrating by parts, taking the 'first function' as Z_1^r and noting that

$$\frac{d}{dz_1}\left[\int_{z_1}^{\lambda_2} z_2^{s}\Big/(1+z_2)^{\frac{k_1+k_2+n_e}{2}}\,dz_2\right] = -\,z_1^{s}\Big/(1+z_1)^{\frac{k_1+k_2+n_e}{2}}\Bigg)$$

$$f(\lambda_1\,\lambda_2) = C'\lambda_1^{r+1}\int_{\lambda_1}^{\lambda_2} z_2^{s}\big/(1+z_2)^{U}\,dZ_2 + C'\int_0^{\lambda_1} z_1^{r+s+1}\big/(1+z_1)^{U}\,dZ_1\;,\quad U = k_1+k_2+n_e/2$$

$$= C'\lambda_1^{r+1}\int_{\lambda_1/1+\lambda_1}^{\lambda_2/1+\lambda_2} x^{s}(1-x)^{u-s-2}\,dx + C'\int_0^{\lambda_1/1+\lambda_1} y^{r+s+1}(1-y)^{u-r-s-3}\,dy$$

$$= C''\lambda_1^{r+1}\left[B(\lambda_2/1+\lambda_2,\ s+1,\ u-s-1) - B(\lambda_1/1+\lambda_1,\ s+1,\ u-s-1)\right]$$

$$+\ C'''\,B(\lambda_1/1+\lambda_1,\ r+s+2,\ u-r-s-2) \qquad \ldots\ldots(52)$$

Note that if $\lambda_1 = \lambda_2 = \lambda$ which appears to be optimum for the case of $k_1 = k_2$ the first term in the brackets above vanishes giving only $CB(\lambda/1+\lambda,\ k_1+k_2/2,\ n_e/2)$ for $r+s = k_1-2/2 + k_2-2/2$. Hence in this particular case we reach the F-value at $(k_1\ k_2\ ;\ n_e)$ d.f.'s and this can be seen also from the equation (47). For

$$Pr\ [\ \chi^2_{k_1} \leq c_1\ ,\ \chi^2_{k_1} + \chi^2_{k_2} \leq c_2\ ]$$

$$Pr\ [\ \chi^2_{k_1} \leq \lambda_1\ n_e s_e^2,\ \chi^2_{k_1} + \chi^2_{k_2} \leq \lambda_2 n_e s_e^2\ ]$$

$$Pr\ [\ \chi^2_{k_1} + \chi^2_{k_2} \leq \lambda\ n_e s_e^2\ ] \quad \text{if} \quad \lambda_1 = \lambda_2 = \lambda$$

$$Pr\ [\ \chi^2_{k_1+k_2}\ /\ (k_1+k_2)\ s_e^2 \leq n_e/k_1+k_2\ \lambda\ ] = Pr\ [\ F_{k_1+k_2;\,n_e} \leq n_e/k_1+k_2\ \lambda\ ]$$

Thus $\lambda_{1-\alpha} \equiv k_1+k_2/n_e\ F_{1-\alpha;\,k_1+k_2;\,n_e}$ ....(53)

Thus the 2 methods a) and b) become identical, showing that the superiority of method b) lies only in using 2 different inequalities instead of one and so it is expected that when λ_1 and λ_2 differ λ_1 will be smaller than the 'λ' value and λ_2 higher. The computations given in Table 4 bear out this expectation. For convenience we shall compare λ_1 and λ_2 with 'λ' instead of $F_{1-\alpha;\,k_1+k_2;\,n_e}$.

To see how far this method is superior to that using Kimball's inequality we shall be giving also the values $\lambda_1' = k_1/n_e\ F_{1-\alpha';\,k_1;\,n_e}$ and $\lambda_2' = k_1+k_2/n_e\ F_{1-\alpha'';\,k_1+k_2;\,n_e}$ where $(1-\alpha')(1-\alpha'') = 1-\alpha$ and (α', α'') are so chosen that

$$F_{1-\alpha';\,k_1;\,n_e} = F_{1-\alpha'';\,k_1+k_2;\,n_e} \qquad \text{.....(54)}$$

It may be noted that if $\lambda_1 \propto k_1$ and $\lambda_2 \propto (k_1+k_2)$ as suggested before, the ratio of λ_1 to λ_1' is same as λ_2 to λ_2' and is shown in the Table 4. Further we are giving the corresp onding value for $n_e = \infty$, with λ's being replaced by $n_e\lambda$.

Table 4 : Values of λ_1/λ_a and λ_2/λ_a and $\lambda_1'/\lambda_1 = \lambda_2'/\lambda_2$ $\alpha = .05$

Constants										
k_1	2	2	2	2	4	4	2	2	4	4
k_2	2	2	4	4	4	2	2	4	4	2
n_0	2	12	4	12	12	8	α	-	-	-
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'	39.0	0.850	5.18	0.79	1.37	2.86	7.38	6.51	11.16	13.02
(α')	(.025)	(.025)	(.0269)	(.0325)	(.025)	(.0303)	(.025)	(.0387)	(.025)	(.0387)
λ_2'	78.0	1.700	15.54	2.37	2.74	4.29	14.76	19.53	22.32	19.53
(α'')	(.025)	(.025)	(.0231)	(.0175)	(.025)	(.0197)	(.025)	(.0118)	(.025)	(.0113)
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
$\lambda_1'/\lambda_1 = \lambda_2'/\lambda_2$	1.625	1.193	1.360	1.133	1.191	1.355	1.181	1.052	1.124	1.319
λ_1'/λ_a	1.01	0.79	0.56	0.53	0.72	1.07	0.78	0.52	0.72	1.03
λ_2'/λ_a	2.03	1.57	1.69	1.58	1.44	1.60	1.56	1.55	1.44	1.55

From the above it is clear that in many cases of practical significance λ_1 and λ_2 put together will give better results than method a). There is a tendency that for smaller degrees of freedom the advantage in using above modification of b) may be considerable, specially when $k_2 > k_1$. Also for the case of equal $k_1 = k_2$, the method is generally better.

The raw λ_1'/λ_1 shows the amount of loss incurred in using an approximation which is quite high in case of small error d.f.'s thus throwing away a large amount of gain which would accur in using b) over a) in general. Only when $k_2 > k_1$, this approximation still holds its utility andmay be used in practice for the sake of convenience. It will be noted in this connection that this comparison is not based on strict practical considerations for then it may be found more convenient and logical to use equal levels $\alpha' = \alpha''$ instead of the above somewhat artificial criterion used in (54) for computing λ_1' and λ_2'.

For the special case of n_e sufficiently large, say > 60, the computations of λ_1, λ_2 or λ_1', λ_2' run in a similar fashion except that some convenience is brought in due to Incomplete Γ-integrals which involves only 1 parameter instead of two. As is shown by the extension of the Table, the gain in using method b) or b') slowly reduces as error d.f. increases to 12 and then more or less is stabilised. As it is desirable to have at least 12 d.f. for error we may specify this gain which may serve as a lower bound for the gain obtained in any case.

d) <u>A new method:</u> We had the 2 sets of vectors $\tilde{Y}_1 \ldots \tilde{Y}_{k_1}$; $\tilde{Y}_{k_1+1} \ldots \tilde{Y}_{k_1+k_2}$ where $E(\tilde{Y}_i) = \tilde{\pi}_i$ and where $\tilde{Y}_1 \ldots \tilde{Y}_{k_1}$ are all mutually orthogonal as also $\tilde{Y}_{k_1+1} \ldots \tilde{Y}_{k_1+k_2}$.

The transformation from Y's to orthonormal basis $U_1 \ldots U_{k_1+k_2}$ was

$$\begin{pmatrix} \tilde{Y}_I \\ \tilde{Y}_{II} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix}$$, $\beta \neq 0$, α & γ non-singular matrices

$$\tilde{Y}_{II} = \beta_{k_2 \times k_1} U_I + \gamma_{k_2 \times k_2} U_{II} \qquad \ldots\ldots(55)$$

If γ is orthogonal matrix, it means $\beta \equiv 0$ which is a degenerate case. The relation (55) gives,

$$\left[\tilde{Y}_{k_1+i} - \tilde{\pi}_{k_1+i} \right]^2 \leq \sum_{j}^{k_1+k_2} \left[(U_j, \gamma) - \varphi_j \right]^2 \quad \text{for all } i = 1, \ldots k_2$$

Obviously we are conservative here, for we are not using the fact that the rank of β may not be k_1 but less. In fact if $k_1 > k_2$, rank of β is $\leq k_2$ and so it may be possible to get a better inequality from equation (55).

Supposing then, $k_1 > k_2$, we have,

$$\tilde{Y}_{k_1+1} = \beta_{11} U_1 + \ldots \beta_{1k_1} U_{k_1} + \gamma_{11} U_{k_1+1} + \ldots \gamma_{1k_2} U_{k_1+k_2}$$

$$\ldots\ldots\ldots\ldots$$

$$\tilde{Y}_{k_1+k_2} = \beta_{k_2 1} U_1 + \ldots \beta_{k_2 k_1} U_{k_1} + \gamma_{k_2 1} U_{k_1+1} + \ldots \gamma_{k_2 k_2} U_{k_1+k_2} \qquad \ldots\ldots(56)$$

The k_2 vectors $\sum_1^{k_1} \beta_{n,i} U_i$, $n = 1, \ldots k_2$ will form a vector space whose rank $\leq k_2$. We shall transform $U_1 \ldots U_{k_1}$ to another orhonormal 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 $\sum \beta_{n,i} U_i$ $n = 1, \ldots k_2$. 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 $\sum \beta_{n,i} U_i$, $n = 1, \ldots k_2$ can be expressed as linear functions of $U_1' \ldots U_r'$, equations (56) may be rewritten as:

$$\tilde{Y}_{k_1+1} = \sum_1^{r} \beta'_{1j} U_j' + \sum_1^{k_2} \gamma_{1j} U'_{k_1+j}$$

$$\ldots\ldots\ldots\ldots$$

$$\tilde{Y}_{k_1+k_2} = \sum_1^{r} \beta'_{k_2 j} U_j' + \sum_1^{k_2} \gamma_{k_2 j} U'_{k_1+j} \qquad \ldots\ldots(57)$$

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:

$$\sum_1^{r} \beta'^2_{ij} + \sum_1^{k_2} \gamma^2_{ij} = 1 \quad \text{for all } i = 1, \ldots k_2$$

Applying Schwartz's inequality on relations (57) we get;

$$[(\tilde{Y}_{k_1+n}, y) - \tilde{\Pi}_{k_1+n}]^2 \leq \sum_1^{r} [(U_i', y) - \phi_i']^2 + \sum_1^{k_2} [(U_{k_1+i}, y) - \phi_{k_1+i}]^2 \qquad \ldots\ldots(58)$$

for all $n = 1, \ldots k_2$, where $\phi_i' = E[(U_i', y)]$ a linear function of $\tilde{\Pi}_i$'s ($i = k_1 - 1, \ldots k_1+k_2$). For the first k_1 parameters, the inequality is the same as in method b) i.e.

$$[(\tilde{Y}_n, y) - \tilde{\Pi}_n]^2 \leq \sum_1^{k_1} [(U_i, y) - \phi_i]^2 = \sum_1^{k_1} [(U_j', y) - \phi_j']^2 \quad n = 1, \ldots k_1 \qquad \ldots\ldots(59)$$

since sum of squares is invariant under orthogonal transformation from $U_1 \ldots U_{k_1}$ to $U_1' \ldots U_{k_1}'$.

In general, nothing may be known about the exact rank of β in a xxxx practical case and so we shall use the upper bound of r viz k_2 in our computations below.

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

and $\sum_{1}^{k_2} [(U_{k_1+j}, y) - \phi_{k_1+j}]^2 = \chi_2^2$, then we shall use the joint distribution of

$G_1 = \chi_1^2 / n_e S_e^2$ $\quad G_2 = \chi_1'^2 / n_e S_e^2$ $\quad G_3 = \chi_2^2 / n_e S_e^2$; where χ_1^2, $\chi_1'^2$ and χ_2^2 are independent.

The inequalities (58) and (59) may be rew written as:

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

$$\text{and } [(\tilde{Y}_{k_1+j}, y) - \tilde{\pi}_{k_1+n}]^2 \leq (G_1 + G_3)\, n_e S_e^2 \qquad j = 1, \ldots, k_2 \qquad \ldots\ldots(60)$$

We require 2 constants λ_1 and λ_2 such that

$$\Pr[G_1 + G_2 < \lambda_1 \; ; \; G_1 + G_3 < \lambda_2] = 1 - \alpha \qquad \ldots\ldots(61)$$

Then the lengths of 100 $(1-\alpha)$% confidence intervals for any normalized linear function of $\tilde{\pi}_1 \cdots \tilde{\pi}_{k_1}$ or of $\tilde{\pi}_{k_1+1} \cdots \tilde{\pi}_{k_1+k_2}$ are respectively proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$. Since the distribution between Ist group of parameters and IInd group may be arbitrary, it may be desirable to put $\lambda_1 = \lambda_2 = \lambda$ say. This has another advantage in that the comparison with method a) is straight forward.

The joint distribution of G_1, G_2, G_3 is:

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

Transforming to $Z_1 = G_1$, $Z_2 = G_1 + G_2$, and $Z_3 = G_1 + G_3$, since Jacobian is 1, we have the joint distribution of Z_1, Z_2, Z_3 as:

$$C(k_2, k_1-k_2, k_2, n_e) \quad Z_1^{\frac{k_2-2}{2}} \quad (Z_2-Z_1)^{\frac{k_1-k_2-2}{2}} \quad (Z_3-Z_1)^{\frac{k_2-2}{2}} \; / (1+Z_2+Z_3-Z_1)^{u} \quad \text{where}$$

$U = k_1 + k_2 + n_e / 2$ and $Z_2 \geq Z_1$, and $Z_3 \geq Z_1$. Equation (61) then yields:

$$C(k_2, k_1-k_2, k_2; n_0)\int_0^{\lambda}\int_{z_1}^{\lambda}\int_{z_1}^{\lambda} \frac{Z_1^{\frac{k_2-2}{2}} (Z_2-Z_1)^{\frac{k_1-k_2-2}{2}} (Z_3-Z_1)^{\frac{k_2-2}{2}}}{(1+Z_2+Z_3-Z_1)^{U}} dZ_3 dZ_2 dZ_1 = 1-\alpha \qquad \ldots\ldots(62)$$

In general the integral $I(\lambda_1, k_1, k_2)$ on the left hand side of (62) can be evaluated by successive iteration and quadrature formulae; but if k_1 and k_2 both are 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.

Let a typical term of the expansion above be:

$$I = \int_0^{\lambda} Z_1^{t} \left[\int_{z_1}^{\lambda}\int_{z_1}^{\lambda} Z_2^{r}\; Z_3^{s} / (1+Z_2+Z_3-Z_1)^{u}\; dZ_2 dZ_3\right] dZ_1$$

$$= \int_0^{\lambda} \frac{z_1^{t}}{(1-z_1)^{v}} \left[\int_{z_1}^{\lambda}\int_{z_1}^{\lambda} \left(\frac{z_2}{1-z_1}\right)^{r}\left(\frac{z_3}{1-z_1}\right)^{s} \Big/ \left(1+\frac{z_2}{1-z_1}+\frac{z_3}{1-z_1}\right)^{u} d\left(\frac{z_2}{1-z_1}\right) d\left(\frac{z_3}{1-z_1}\right)\right] dz_1, \qquad \ldots\ldots(63)$$

$V = U-r-s-2$

$$= \int_0^{\lambda} \frac{z_1^{t}}{(1-z_1)^{v}} \left[\int_{z_1/1-z_1}^{\lambda/1-z_1}\int_{z_1/1-z_1}^{\lambda/1-z_1} x^{r}\, y^{s}/(1+x+y)^{u}\; dx\, dy\right] dz_1$$

It is known that under certain mild restrictions over the function $g(X)$, which are satisfied here, we have:

$$\frac{d}{dz}\int_{f(z)}^{\phi(z)} g(X)dX = \phi'(z)\; g\left[\phi(z)\right] - f'(z)\; g\left[f(z)\right]$$

Taking $Z_1^{t}/(1-Z_1)^{V}$ as the first function and the double integral in Z_2, Z_3 as the IInd function, we integrate the triple integral by parts:

Now,
$$\frac{d}{dz_1}\left[\int_{z_1/1-z_1}^{\lambda/1-z_1}\int_{z_1/1-z_1}^{\lambda/1-z_1} X^{r}\, Y^{s}/(1+X+Y)^{u}\; dXdY\right]$$

$$= \frac{d}{dz_1}\left(\frac{\lambda}{1-z_1}\right)\int_{z_1/1-z_1}^{\lambda/1-z_1} \frac{x^{r}}{\left(1+x+\frac{\lambda}{1-z_1}\right)^{u}}\cdot\left(\frac{\lambda}{1-z_1}\right)^{s} dx - \frac{d}{dz_1}\left(\frac{z_1}{1-z_1}\right)\int_{z_1/1-z_1}^{\lambda/1-z_1} \frac{x^{r}}{\left(1+x+\frac{z_1}{1-z_1}\right)^{u}}\cdot\left(\frac{z_1}{1-z_1}\right)^{s} dx$$

+ two exactly similar terms in 'Y' and 's' instead of 'X' and 'r'.

$$= \lambda^{s+1}/(1-z_1)^{s+2}\int_{z_1/1-z_1}^{\lambda/1-z_1} x^{r}/\left(1+x+\frac{\lambda}{1-z_1}\right)^{u} dx + z_1^{s}/(1-z_1)^{s+2}\int_{z_1/1-z_1}^{\lambda/1-z_1} x^{r}/\left(1+x+\frac{z_1}{1-z_1}\right)^{u} dx$$

$$+ \lambda^{r+1}/(1-z_1)^{r+2}\int y^{s}/\left(1+y+\frac{\lambda}{1-z_1}\right)^{u} dy + z_1^{r}/(1-z_1)^{r+2}\int y^{s}/\left(1+y+\frac{z_1}{1-z_1}\right)^{u} dy$$

Thus the triple integral can be reduced into a few double integrals, each of which can be separately evaluated. Consider for instance a typical term.

$$\int_0^{\lambda} Z_1^{t}/(1-Z_1)^{\nabla} \int_{Z_1/1-Z_1}^{\lambda/1-Z_1} X^{\theta}/(1+X+\tfrac{\lambda}{1-Z_1})^{\nabla}\, dX\, dZ_1$$

This will be converted into a few. Incomplete B-functions, by taking $Z_1^t/(1-Z_1)^{\nabla}$ as the first function and integrating by parts.

It will be noted that the method is some improvement over the previous modification of b) only when $k_1 > k_2$. As soon as $k_1 = k_2$, the above modification fails, and we reach the previous method. But here, it may be possible to get an idea of the actual rank of β and then again the method d) can work. We give below a lemma to show that the value of λ satisfying (61) decreases with decrease in the rank of β and attains the minimum value for $\beta \equiv 0$.

Lemma 1 Let χ_1^2, χ_2^2 and χ_3^2 be 3 independently distributed χ^2-variables with respective degrees of freedom 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 $\chi^2_{(n)}$ can be expressed as the 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:

$[\sum_1^r x_i^2 + \sum_{r+1}^{v_1} x_i^2 < \lambda,\ \sum_1^r x_i^2 + \sum_{v_1+1}^{v_1+v_2} x_i^2 < \lambda]$ which implies the Event:

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

The smallest value of r being '0', the maximum probability will be given by the event: $[\sum_1^{v_1} x_i^2 < \lambda,\ \sum_{v_1+1}^{v_1+v_2} x_i^2 < \lambda]$

Hence the result.

If we divide each x^2 by a non zero positive quantity, say S_0^2, 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 & χ_3^2/S_e^2 on integrating the probabilities with respect to S_e^2 over its whole range.

Since $\sum_1^r [(U_i', y) - \phi_i']^2/\sigma^2 = \chi^2_{(r)}$, $\sum_{r+1}^{k_1} [(U_i', y) - \phi_i']^2/\sigma^2 = \chi^2_{(k_1-r)}$ and $\sum_{k_1+1}^{k_1+k_2} [(U_j', y) - \phi_j]^2/\sigma^2 = \chi^2_{(k_2)}$ have the required distributions, we have

$$\Pr\left[\chi^2_{(r)}+\chi^2_{(k_1-r)}<\lambda\ ,\ \chi^2_{(r)}+\chi^2_{(k_2)}<\lambda\right]\le \Pr\left[\chi^2_{(s)}+\chi^2_{(k_1-s)}<\lambda\ ,\ \chi^2_{(s)}+\chi^2_{(k_2)}<\lambda\right]\ s<r \quad \ldots\ldots(64)$$

and on dividing by S_e^2/σ^2 and integrating over S_e^2 we get the same inequality. Then, corresponding to equation (61) we may say that $\Pr\left[G_1' + G_2' < \lambda, G_1' + G_3 < \lambda\right]$ increases as 'r' decreases, where dashes denote that the exact rank 'r' is used instead of 'k_2'

Since both sides of (64) increase as λ increases, it is clear that the value of λ necessary to make both sides equal to $1-\alpha$, will be such that $\lambda(\text{rank} = k_2) \geq \lambda(\text{rank} = r)$. The argument immediately applies to the special case $r = 0$, when we get two independent χ^2's for fixed S_e^2, and this gives the minimum value of λ for attaining a confidence level $1-\alpha$.

Table 5 below gives some computations of 'λ' needed in method d) when the rank of β is full, a well as when it is minimum i.e. zero. In any practical situation the actual value is expected to be between these 2 bounds. The 'λ' for orthogonal case has again been chosen to be the same for both sets of parameters for the sake of convenience in comparisons, though there are reasons as given earlier that it should correspond somewhat to the degrees of freedom on which $G_{1(k_1)}$ and $G_{2(k_2)}$ are based. These computations are based on the integral:

$$C(k_1, k_2; n_e) \int_0^{\lambda_1}\int_0^{\lambda_2} G_1^{\frac{k_1-2}{2}} \quad G_2^{\frac{k_2-2}{2}} \Big/ (1+G_1+G_2)^{\frac{k_1+k_2+n_e}{2}} dG_1\, dG_2 = 1-\alpha \quad \ldots\ldots(65)$$

which yields simple Incomplete B-functions on integration by parts. In the special case $k_1 = k_2$, when method d) fails, we give two values of λ from Table 4) along with the single 'λ' values from (63) for each $k_1 = k_2 = k$.

As in Table 4 we shall take k_1 and k_2 to be even integers. The values of odd k_1 or k_2 can be obtained by simple interpolation from Table 5)*

---

* If k_2 is even but not k_1, the integrals (62) and (65) can still be exactly evaluated by first integrating out G_3 in (62) and G_2 in (65) and then solving the resulting double integral or single integral. In many case it may be found easier to start directly with (62) instead of using simplification like (63)

Table 5: Showing upper and lower bounds of λ_d as compared with λ_a

Constants													
k_1	4	-	-	6	-	-	6	-	-	8	2	2	4
k_2	2	-	-	2	-	-	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	1.71	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 {18.0	.713 1.426	1.15 2.30
λ (orthogonal)	7.00	2.06	1.15	4.40	2.75	1.52	2.98	1.62	.83	1.32	28.8	.837	1.36
λ_a/λ_d	1.19	1.19	1.17	1.22	1.22	1.21	1.17	1.17	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

While observing the above table one or two points strike one. Firstly the advantage in using method d) is nearly always small (unless we have some knowledge of the actual rank of β) coming to nearly 10% decrease in lengths of confidence intervals.

Further, when rank $\beta = 0$ i.e. orthogonal, the advantage in using method b) as compared to method a) is quite considerable and is nearly unaffected by changes in the d.f.'s of error variance. This is important, for then we can specify the gain likely to accrue by using method b). This is so even for $k_1 = k_2$ when method d) coincides with modification b'). Also the last two rows between themselves give the upper and lower bounds of the gain by using method d);

<u>Large Sample Theory:</u> If n_e is quite large, S_e^2 may be taken as a constant fixed so that we get the joint distribution of two χ^2's. If further the hypotheses are 'orthogonal' ($\beta = 0$), this simply means combination of 2 independent tests, as mentioned by Finney (1941). In general, we have to start with the joint

distribution of $V_1 = \chi_1'^2/s_e^2$, $V_2 = \chi_1'^2/s_e^2$ and $V_3 = \chi_2'^2/s_3^2$ which is the joint distribution of 3 independent χ^2's on k_2, k_1-k_2 and k_2 d.f.'s respectively i.e.

$$\text{Const.}\quad e^{-\frac{V_1}{2}}(V_1)^{\frac{k_2-2}{2}} e^{-\frac{V_2}{2}}(V_2)^{\frac{k_1-k_2-2}{2}} e^{-\frac{V_3}{2}}(V_3)^{\frac{k_2-2}{2}} \quad dV_1 dV_2 dV_3$$

Now making the transformation $V_1 = Z_1$, $V_1 + V_2 = Z_2$ and $V_1 + V_3 = Z_3$ and noting that Z_2 and Z_3 both $\geq Z_1$, we want to get one single 'λ' such that $Pr\,[\,Z_2 < \lambda,\, Z_3 < \lambda\,] = 1-\alpha$ i.e.

$$1-\alpha = C\int_0^{\lambda}\left[\int_{z_1}^{\lambda}\int_{z_1}^{\lambda} e^{-\frac{1}{2}(Z_2+Z_3-Z_1)} (Z_2-Z_1)^{\frac{k_1-k_2-2}{2}} \quad (Z_3-Z_1)^{\frac{k_2-2}{2}} \quad Z_1^{\frac{k_2-2}{2}} \quad dZ_3 dZ_2\right] dZ_1 \qquad \ldots\ldots(66)$$

As before if, k_1 and k_2 are both even integers, the above can be easily evaluated by expanding $(Z_2-Z_1)^{\frac{k_1-k_2-2}{2}}$ and $(Z_3-Z_1)^{\frac{k_2-2}{2}}$ and integrating term by term. Otherwise we can transform to $Z_2-Z_1 = U$, $Z_3 - Z_1 = V$ in the inner integral to get a simple product of two indefinite integrals viz.,

$$\int_0^{\lambda-z_1} e^{-\frac{U}{2}} \; U^{\frac{k_1-k_2-2}{2}} \quad dU,\int_0^{\lambda-z_1} e^{-\frac{V}{2}} \quad V^{\frac{k_2-2}{2}} \quad dV \qquad \ldots\ldots(67)$$

Thus we shall finally get only a few Incomplete Gamma functions.

If k_1 is odd but k_2 even integer, we can still evaluate (66), for a typical term of double integral (after integrating out 'V') is like,

$$\int_0^{\lambda}\int_0^{\lambda-z_1} e^{-\frac{z_1}{2}} \;\; e^{-\frac{U}{2}} \; (Z_1)^{r} \quad f(Z_1).\; U^{\frac{k_1}{2}-r-2} \quad dUdZ_1 \qquad \ldots\ldots(68)$$

where $f(Z_1)$ comes from IInd part of (67). The integral (68) can be reduced to single integrals by taking $e^{-\frac{z_1}{2}} Z_1^{r} f(Z_1)$ as the first function, of which the indefinite integral can be explicitly written, and $\int_0^{\lambda-z_1} e^{-\frac{U}{2}}\, U^{\frac{k_1}{2}-r-2} \; dU$ as the second function and integrating by parts. This yields a few Incomplete Gamma functions.

Further from (64) it is clear that the value of satisfying (66) decreases as rank of β (assumed k_2 in 66), decreased and attains a minimum where $\beta = 0$, when we simply get 2 independent χ^2 for which the ordinary χ^2-tables for α = .01, .025, .05 will suffice.

By an approach similar to that of (67), and (68) we can reduce (62) also to some double integrals if k_2 is even and k_1 odd. But as soon as k_1 is odd (62) cannot be solved explicitly and we have to adopt quadrature methods (numerical integration). The most difficult case is of course, k_2 odd and k_1 even, when the iterated integrals of 3rd order have to be dealt with in (62) and (66). Unfortunately the procedure becomes very lengthy and time-involved in the numerical computations.

But, even if we are not willing to interpolate between the values of k_1 and k_2, both even, we can still get upper and lower bounds of λ. Consider the events E_1 and E_2:

E_1: $\left[\chi^2_{1(k_2)} + \chi'^2_{1(k_1-k_2)} < \lambda \ , \ \chi^2_{1(k_2)} + \chi^2_{2(k_2)} < \lambda \right]$ where $\chi^2_{i(r)}$ means a χ^2 variable at 'r' d.f's. This implies the events,

E_2: $\left[\chi^2_{1(k_2-1)} + \chi'^2_{1(k_1-1-\overline{k_2-1})} < \lambda \ , \ \chi^2_{1(k_2-1)} + \chi^2_{2(k_2-1)} < \lambda \right]$

Hence the values of λ for E_1 and E_2, such that $P(E_1) = P(E_2)$ .95 say, are such that $\lambda(E_1) \geq \lambda(E_2)$. Thus for odd k_1 and k_2, we can get the lower bound of $\lambda(E_1)$. If n_e is not large, the only modification will be that the lower bound of $\lambda(E_1)$ will become $\lambda_{n_e}(E_2)$.

The upper bound of $\lambda(E_1)$ can be found from:

E_3: $\left[\chi^2_{1(k_2+1)} + \chi'^2_{1(k_1+1-\overline{k_2+1})} < \lambda \ , \ \chi^2_{1(k_2+1)} + \chi^2_{2(k_2+1)} < \lambda \right] \Rightarrow E_1$

so that $\lambda(E_1) \leq \lambda(E_3)$. Hence,

$$\lambda(E_2) \leq \lambda(E_1) \leq \lambda(E_3)$$

In many cases another upper bound may be obtained for $\lambda(E_1)$, from another consideration. We give below the relevant lemma:

<u>Lemma 2</u>: If x, y, z are independent variables which take only values ≥ 0 and if x' is identically and independently distributed as x, then we have:

$$P\left[x+y < \lambda \ , \ x+z < \lambda \right] \geq P\left[x+y < \lambda \ \ x'+z < \lambda \right] \qquad \text{.....(70)}$$

Proof. L.H.side $= P(x<\lambda-z).\ P(x+y<\lambda / x<\lambda-z)$

R.H.side $= P(x<\lambda-z).\ P(x+y<\lambda)$

We shall compare only $P(x+y<\lambda / x<\lambda-z)$ and $P(x+y<\lambda)$.

Find the conditional probabilities for above 2 expressions when y and z are fixed,

i.e. $P(x<\lambda-y/x<\lambda-z)$ and $P(x<\lambda-y)$

There arise two cases according as $y<z$ or $y>z$

i) Let $y<z$, then

$$P(x<\lambda-y/x<\lambda-z) = P(x<\lambda-y,\ x<\lambda-z)/P(x<\lambda-z) = P(x<\lambda-z)/P(x<\lambda-z) = 1$$
$$\geq P(x<\lambda-y) \quad \text{for } \lambda-z<\lambda-y$$

ii) Let $y>z$, then,

$$P(x<\lambda-y/x<\lambda-z) = P(x<\lambda-y)/P(x<\lambda-z) \quad \text{as before}$$
$$\geq P(x<\lambda-y) \quad \text{for any } y.$$

Hence for fixed (y,z) we have

$$P(x<\lambda-y/x<\lambda-z) \geq P(x<\lambda-y)$$

Multiplying by $P(x<\lambda-z)$ on both sides, we get for <u>fixed (y,z)</u>:

$$P(x+y<\lambda,\ x+z<\lambda) \geq P(x+y<\lambda,\ x'+z<\lambda).$$

Now multiplying by the probability of any pair (y,z) and summing over all the admissible pairs (i.e. where $y,z<\lambda$), we shall get the same inequality for the unconditional probabilities. For continuous variables sum will be replaced by integrals. Hence the result (70).

Suppose now that $x, x' \sim \chi^2_{(k_2)}$, $y \sim \chi^2_{(k_1+k_2)}$, $z \sim \chi^2_{2(k_2)}$ then

$$P(\chi^2_{(k_2)}+\chi^2_{(k_1+k_2)}<\lambda,\ \chi^2_{(k_2)}+\chi^2_{2(k_2)}<\lambda) \geq P(\chi^2_{(k_2)}+\chi^2_{(k_1+k_2)}<\lambda,\ \chi^2_{1(k_2)}+\chi^2_{2(k_2)}<\lambda)$$
$$= P(\chi^2_{(k_1)}<\lambda,\ \chi^2_{(2k_2)}<\lambda),$$

for the sum of 2 independent χ^2's is again a χ^2. Further as the probabilities on both sides go on increasing as λ increases, it means that for some fixed '$1-\alpha$' to be obtained by both sides we shall have

$$\lambda_{1-\alpha}\,(\text{L.H.side}) \leq \lambda_{1-\alpha}\,(\text{R.H.side})$$

Since R.H.side λ can be easily calculated this provides an upper bound for $\lambda(E_1)$ of equation (69). This may or may not be better than $\lambda(E_3)$ of equation (69) depending on the bigness of k_1 and k_2.

Table 6: Extension of Table 5 for the case where n_0 is large

Constants						
k_1	4	6	6	8	2	4
k_2	2	2	4	4	2	4
λ_a	12.59	15.51	18.31	21.03	9.49	15.51
λ_d	10.73	12.96	16.00	17.26	{6.25 {12.50	9.93 19.87
λ(ortho.)	9.87	12.74	13.21	15.72	7.38	11.14
λ_a/λ_d	1.17	1.20	1.14	1.22	{1.52 {0.76	1.56 0.78
λ_a/λ(ortho)	1.28	1.22	1.39	1.34	1.29	1.39

Comparison with Table 5 suggests that as n_0 increases the gain in using method d) over a) decreases at first but is soon stabilized.

Generalization of method d) to the case of 3 hypotheses:

We have 3 groups of parameters $\pi_1 \ldots \pi_{k_1}$; $\pi_{k_1+1} \ldots \pi_{k_1+k_2}$; $\pi_{k_1+k_2+1} \ldots$; $\pi_{k_1+k_2+k_3}$ for whom the estimates are provided by the vectors $Y_1 \ldots Y_{k_1}$; $Y_{k_1+1} \ldots Y_{k_1+k_2}$; $Y_{k_1+k_2+1} \ldots Y_{k_1+k_2+k_3}$ where the 3 sets of vectors are such that within each set the k_i (i=1,2,3) vectors are mutually, orthogonal but not between one set and another. From Y's we arrive at the orthonormal basis $U_1, \ldots, U_{k_1}$; $U_{k_1+1} \ldots, U_{k_1+k_2}$; $U_{k_1+k_2+1} \ldots U_{k_1+k_2+k_3}$ which are such that U_I forms the basis of vector space generated by Y_I and U_I & U_{II} together form the basis of the vector space generated by $Y_1\ Y_2 \ldots\ Y_{k_1+k_2}$, etc.

The transformation Y's to U's will be like:

$$\begin{Bmatrix} Y_I \\ Y_{II} \\ Y_{III} \end{Bmatrix} = \begin{matrix} \\ k_1 \\ k_2 \\ k_3 \end{matrix} \begin{Bmatrix} k_1 & k_2 & k_3 \\ \alpha & 0 & 0 \\ \beta & \gamma & 0 \\ \delta & \epsilon & \xi \end{Bmatrix} \begin{Bmatrix} U_I \\ U_{II} \\ U_{III} \end{Bmatrix} \qquad \ldots\ldots(71)$$

Let $k_1 \geq k_2 \geq k_3$; the matrices α, γ, ξ are all non-singular . $\beta \neq 0$ and at least one of δ and ϵ is non-null. As before in developing the method we shall assume that β, δ and ϵ have their full ranks viz. k_2, k_3 and k_3 respectively. If any thing be known we can utilise the 3 ranks r_1, r_2 and r_3 as in method d) before . By the method employed in reaching (60) we shall form

$$G_1 = 1/n_e s_e^2 \sum_1^{k_3} [(U_i', y) - \phi_i']^2 \qquad G_2 = \sum_{k_3+1}^{k_2} [(U_i', y) - \phi_i']^2 / n_e s_e^2$$
$$G_3 = \sum_{k_2+1}^{k_1} [(U_i', y) - \phi_i']^2 / n_e s_e^2 \qquad G_4 = \sum_1^{k_3} [(U'_{k_1+j}, y) - \phi_{k_1+j}']^2 / n_e s_e^2$$
$$G_5 = \sum_{k_3+1}^{k_2} [(U'_{k_1+j}, y) - \phi_{k_1+j}']^2 / n_e s_e^2 \qquad G_6 = \sum_1^{k_3} [(U'_{k_1+k_2+j}, y) - \phi_{k_1+k_2+j}]^2 / n_e s_e^2$$
$$\ldots\ldots(72)$$

Where U_1' .... U_{k_1}' are an orthonormal transform of U_1 ... U_{k_1} in such a way that U_1' ... U_{k_2} form a basis of the k_2 rows of the matrix βU_I ; further U_1' ... U'_{k_2} have been selected in such a fashion (i.e. by orthogonal transformation) that U_1' .... U'_{k_3} (first k_3) form a basis of the vector space generated by 'k_3' rows of δU_I . That this is always possible may be seen by starting from the vector space of the vectors of δU_I, then of βU_I etc. Again, U'_{k_1+1} ... $U'_{k_1+k_2}$ represent an orthogonal transform of U_{k_1+1} ....$U_{k_1+k_2}$ such that U'_{k_1+1} .... $U'_{k_1+k_3}$ (first k_3) form a basis of the vector space generated by the rows of the ($k_3 \times 1$) matrix ϵU_{II}. This again is always possible. If $\delta = 0$ (or ϵ) G_1 will be merged with G_2 to form G_1^* say, and it will not occur in the inequalities in Y_{III}.

Using Schwarts inequality in 3 steps we shall get 3 sets of confidence regions(and intervals)like :-

(i) $\pi_1 \ldots \pi_{k_1}$: $\sum_1^{k_1} [(\tilde{Y}_i, y) - \tilde{\pi}_i]^2 \leq (G_1 + G_2 + G_3) n_e B_e^2$

Intervals being, $\quad |(\tilde{Y}_i, y) - \tilde{\pi}_i|^2 \leq (G_1 + G_2 + G_3)\, n_e s_e^2$

(ii) $\pi_{k_1+1} \cdots \pi_{k_1+k_2}$: $\sum_1^{k_2} [(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}]^2 \leq (G_1+G_2+G_4+G_5)\, n_e s_e^2$

Intervals being, $\quad |(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}|^2 \leq (G_1+G_2+G_4+G_5)\, n_e s_e^2$

(iii) $\pi_{k_1+k_2+1} \cdots \pi_{k_1+k_2+k_3}$: $\sum_1^{k_3} [(Y_{k_1+k_2+j}, y) - \tilde{\pi}_{k_1+k_2+j}]^2 \leq (G_1 + G_4+G_6)\, n_e s_e^2$

Intervals being $\quad |(\tilde{Y}_{k_1+k_2+j}, y) - \tilde{\pi}_{k_1+k_2+j}|^2 \leq (G_1+G_4+G_6)\, n_e s_e^2$ ....(73)

For $1-\alpha$ confidence level we want to find 3 constant λ_1, λ_2 & λ_3 such that
$Pr[G_1 + G_2 + G_3 < \lambda_1, G_1+G_2+G_4+G_5 < \lambda_2, G_1+G_4+G_6 < \lambda_3] = 1-\alpha$ ....(74)

Since the numerators of $G_1, \ldots, G_6$ are independent χ^2's with $k_3, k_2-k_3, k_1-k_2, k_3, k_2-k_3, k_3$ d.f.'s respectively and denominator is an independent on n_e d.f., (74) becomes

$$1-\alpha = \text{const.} \int \cdots_{6} \cdots \int \frac{G_1^{\frac{k_3-2}{2}} \; G_2^{\frac{k_2-k_3-2}{2}} \; G_3^{\frac{k_1-k_2-2}{2}} \; G_4^{\frac{k_3-2}{2}} \; G_5^{\frac{k_2-k_3-2}{2}} \; G_6^{\frac{k_3-2}{2}}}{(1+G_1+\cdots\cdots+G_6)^{\frac{k_1+k_2+k_3+n_e}{2}}}\, dG_6 \cdots dG_1 \qquad \text{.........(75)}$$

Integrated over the domain $\quad G_1+G_4+G_6 \leq \lambda_3 \quad G_1+G_2+G_4+G_5 \leq \lambda_2, \; G_1+G_2+G_3 \leq \lambda_1$

Since the grouping of parameters into Ist, IInd and IIIrd may be arbitrary we may take $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ which is also easier to deal with and easier to compare with other methods. Even then, the 6-ple integral looks formidable to evaluate and it may be desirable to make some approximations. This if we add 'G_3' in part ii) of relations (73) and $G_2 + G_3 + G_5$ in part iii) on the right hand side, we shall get an upper bound of joint significance level α ; for we shall have only 3 independent χ^2's to deal with $(G_1+G_2+G_3)$ at k_1 d.f., (G_4+G_5) at k_2 d.f. and G_6 at k_3 d.f. But this means returning to method b) given earlier and throwing away the advantages of method d). However, in practice, such a situation may be forced on us if $k_1 = k_2 = k_3$, when in the absence of any knowledge about the ranks of matrices ρ, δ and ϵ we have to use the inequalities suggested as approximations, and use $\lambda_1 \propto k_1$, $\lambda_2 \propto (k_1+k_2)$ and $\lambda_3 \propto (k_1+k_2+k_3)$. If k_1, k_2 and k_3 all are even integers the integral 75) can still be evaluated explicitly specially for many practical cases where k's are small. We give below an example to show how it it can be done.

Example: Take $k_1 = 2$ $k_2 = 2$ $k_3 = 2$ Such a case may arise from a 3^3 factorial experiment in non-orthogonal data

We get only a quadruple integral:

$$1-\alpha = C\int \cdots_4 \int \; G_1^0 \; G_3^0 \; G_4^0 \; G_6^0 \;/\; (1+G_1+G_3+G_4+G_6)^{4+\frac{n_e}{2}} dG_6 \ldots dG_1$$

domain of integration being $Z_4 = G_1+G_4+G_6 < \lambda$, $Z_3 = G_1+G_4 < \lambda$, $Z_2 = G_1+G_3 < \lambda$

$Z_1 = G_1 < \lambda$, and take $n_e = 16$.

$$= C\int_0^\lambda \int_{z_1}^\lambda \int_{z_1}^\lambda \int_{z_3}^\lambda 1/(1+Z_4+Z_2-Z_1)^{12} \; dZ_4\, dZ_3\, dZ_2\, dZ_1 \qquad \text{Jacobian being } 1.$$

$$= C/11\int_0^\lambda \int_{z_1}^\lambda \int_{z_1}^\lambda \left[1/(1+Z_3+Z_2-Z_1)^{11} - 1/(1+\lambda+Z_2-Z_1)^{11} \right] dZ_3\, dZ_2\, dZ_1$$

$$= C/11.10\int_0^\lambda \int_{z_1}^\lambda \left[1/(1+Z_2)^{10} - 1/(1+\lambda+Z_2-Z_1)^{10} - \lambda-Z_1/(1+\lambda+Z_2-Z_1)^{11} \right] dZ_2\, dZ_1$$

$$= C' \left[B^*(\lambda/1+\lambda, 2, 8) + 1/9 \int_0^\lambda \left[\{ 1/(1+2\lambda-Z_1)^9 - 1/(1+\lambda)^9 \} dZ_1 \right.\right.$$

$$\left. + 1/10 \int_0^\lambda \{ \lambda-Z_1/(1+2\lambda-Z_1)^{10} - \lambda-Z_1/(1+\lambda)^{10} \} dZ_1 \right]$$

Where $B^*(\cdots)$ represents actual value of Incomplete B- integral [Not the ratio of the actual integral to $\beta(2,8)$].

$$= B(\lambda/1+\lambda, 2, 8) - 1/9\; \lambda/(1+\lambda)^9 - 1/10\; \lambda^2/(1+\lambda)^{10} - 1/20\; \lambda^2/(1+\lambda)^2$$

$$+ 1/9.(1+\lambda)^8\, B^*(\lambda/1+2\lambda, 1, 8) + 1/10.(1+\lambda)^8\, B^*(\tfrac{\lambda}{1+2\lambda}, 2, 8)$$

$$= B(\lambda/1+\lambda, 2, 8) + 1/10.(1+\lambda)^8\, B(\lambda/1+2\lambda, 2, 8) - 1/20\; \lambda^2/(1+\lambda)^2 - 1/9.\; \lambda/(1+\lambda)^9$$

$$+ 1/9.(1+\lambda)^8\, B(\lambda/1+2\lambda, 1, 8).\; B(1,8)/\beta(2,8) \text{ where } \beta(1,8)/\beta(2,8) = 9$$

Values of $B(x,2,8)$ and $B(y,1,8)$ may be obtained from Pearson's Incomplete Be Function tables.

<u>A special case:</u> Sometimes we have to make tests of significance for a number of parameters individually, making all k_i's equal to 1. This is the case of Factorial experiments, where a number of contrasts of d.f. 1 each are isolated and tested individually, though in general the hypotheses have a joint importance. If we decide on a simultaneous level of significance, and if the data are non-orthogonal, the method b) can be immediately extended and gives:

$$Pr\left[G_1 < \lambda_1 \ , G_1 + G_2 < \lambda_2 \ , \ldots\ldots\ldots \ , \sum_1^t G_i < \lambda_t \right] = 1 - \alpha$$

If knowledge of the ranks of matrices like , β, δ, ...etc is available implying the knowledge about the orthogonality of some of the vectors Y_I, Y_{II} ...Y_t this inequality may be simplified.

The special case when all are orthogonal is essentially a genralization of the case studied before but is simpler. Nair (1938) studied this case and provided tables, while the other special case of 't' orthogonal hypotheses of 'k' parameters each was studied by Ramachandran through 'Studentized χ^2' and some tables for $t = 2$ were provided.

<u>Tests of hypotheses.</u>

In method d) we can immediately rach the corresponding tests of hypotheses H_1 and H_2 from the confidence intervals (60) i.e.

$$\sum_1^{k_1}\left[(\tilde{Y}_{n}, y) + \tilde{\pi}_n \right]^2 \leq (G_1 + G_2) \, n_e s_e^2$$

$$\sum_1^{k_2}\left[(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i} \right] \leq (G_1 + G_3) \, n_e s_e^2$$

For testing $H_1 : \ldots \pi_1 = 0 , \ldots\ldots \pi_{k_1} = 0$ and $H_2 : \pi_{k_1+1} = 0 , \ldots\ldots , \pi_{k_1+k_2} = 0$ we shall use the test statistics $T_1 = \sum_1^{k_1} (\tilde{Y}_n, y)^2 / n_e s_e^2$ and $T_2 = \sum_1^{k_2} (\tilde{Y}_{k_1+i}, y)^2 / n_e s_e^2$ and shall get the critical limits for T_1 and T_2 from the distribution of $G_1 + G_2$, $G_1 + G_3$. Table 11 of Appendix provides the limit λ_α for T_1 and T_2 for many representative values of (k_1, k_2, n_e) for $\alpha = .05$. The simultaneous test of the hypotheses H_1 & H_2 given by $(T_1 > \lambda, T_2 > \lambda)$ will have a simultaneous level bounded above by 'α' though the bound will be nearer to 'α' than in method a) or b). The two tests provide quasi-independent tests of H_1 and H_2 for the distributions of T_1 and T_2 involve only the parameters of Ist and IInd groups respectively, when the hypotheses H_1 and H_2 are not true. The power functions of tests T_1 and T_2 are better than those of methods a) or b) if we want quasi- independent tests for H_1 and H_2.

The generalization to the case of 3 hypotheses needs the additional computation of $T_3 = \sum_{1}^{k_3} \left(\tilde{Y}_{k_1+k_2+j} \cdot Y \right)^2 / n_e S_e^2$, but the Tables for this case are yet to be prepared.

<u>Computations in practice.</u> The new method d) does not require any more computations than needed in method b) or b'). The essential thing is, the computation of orthogonal vectors $\tilde{Y}_1, \ldots, \tilde{Y}_{k_1}$; $\tilde{Y}_{k_1+1}, \ldots$ $Y_{k_1+k_2}$ both sets being necessary in method d), though in method a) sometimes one can avoid the computation of one of the 2 sets of vectors above. But for making confidence statements about π's it is every where necessary to compute all the k_1 and k_2 vectors.

If $\alpha_1, \alpha_2, \ldots, \alpha_p$ are a set of vectors, Gram - Schmidt method gives the way to compute an orthogonal set starting from α_1 say. These are:

$$\alpha_1' = \alpha_1 \;,\; \alpha_2' = \alpha_2 - \alpha_1 \cdot \frac{(\alpha_2 \cdot \alpha_1)}{(\alpha_1 \cdot \alpha_1)} \;,\; \alpha_3' = \alpha_3 - \alpha_1 \cdot \frac{(\alpha_3 \cdot \alpha_1)}{(\alpha_1 \cdot \alpha_1)} - \alpha_2' \cdot \frac{(\alpha_3 \cdot \alpha_2')}{(\alpha_2' \cdot \alpha_2')}$$

and so on upto α_p'. Further as we need $(\alpha_2' \cdot \alpha_2')$, $(\alpha_3' \cdot \alpha_3')$ ...... it will be found preferable to start systematically from $(\alpha_1 \cdot \alpha_1)$ and get all the $p(p+1)/2$ scalar products viz:

$$(\alpha_1 \cdot \alpha_1) \,,\, (\alpha_1 \cdot \alpha_2) \,,\, (\alpha_1 \cdot \alpha_3) \;\ldots\; ; (\alpha_2 \cdot \alpha_2) \;\ldots\; (\alpha_2 \cdot \alpha_p) \;; \ldots$$

Actually all of these are needed in finding an ortho-normal system, and we have to do it this in 2 parts once for $Y_1, \ldots, Y_{k_1}$ then for $Y_{k_1+1}, \ldots$ $Y_{k_1+k_2}$.

Chapter IV

Some applications of the methods developed

Here we give some actual situations arising in course of investigations in Agriculture or Animal husbandry fields, which necessitate the use of methods a) .... to d). Both, orthogonal and non-orthogonal cases are considered.

1) : In a Randomised block design, an experiment was conducted at Rice Research, Station, Tirur. There were 8 treatments and the idea was to study the effect of direct application of Phosphatic manure, to Paddy and through green manure-cup proceeding the Paddy crop. They were:

A) No manure B) 45 lb/acre of P_2O_5 applied at the time of transplanting paddy.

C) Sun-hemp grown without P_2O_5 but 45 lb/acre of P_2O_5 ....transplantation

D) Sun-hemp grown with 45 lb/acre of P_2O_5.

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

F) Dhaincha grown with 45 lb/acre P_2O_5.

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

H) Sasbaria grown with 45 lb/acre P_2O_5.

Clearly there are two natural groupings; one of A & B, with direct application of P_2O_5; the other of C.....H, where phosphate is given through green manure. The experimenters' interest is confined to comparisons within each group (1 and 5 d.f.'s respectively) and finally one between the average of the 2 groups. Thus we have 3 groups of parameters.

i) $t_a - t_b$

ii) 5 contrasts between C.......H.

iii) 1 contrast of group comparison; $3(t_a+t_b) - (t_c+ \ldots\ldots t_h)$. where $t_a \ldots t_h$ represent the effect of that treatment and they have a restriction $\sum t_a = 0$.

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:

$y = \mu + \alpha_1 (t_a - t_b) + \sum_2^6 \alpha_i$ (5 orthogonal contrasts in $t_c \ldots t_h$)

$+\alpha_7 [1/2(t_a + t_b) - 1/6 (t_c + \ldots\ldots t_h)] + e$, where e is the error term, independently and normally distributed, μ is the general effect and are known constants taking values 1,2 ...., etc. according as the observation comes from one treatment or another. The 5 orthogonal contrasts may be formed from Fisher's Orthogonal Polynomials for n = 6 at equal steps or they may be t_c-t_d, t_e-t_f, t_g-t_h, $t_c+t_d-t_g-t_h$, $t_c+t_d-2t_e+2t_f-t_g-t_h$; which are probably of more immediate interest.

Method a) would have us using a F-test at 7 d.f. for testing the homogeniety 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-lenths of the contrasts of particular interest. If some observations were missing we would use method d) which would still be superior to method a) considerably.

In this particular case a method of Tukey, as given by Scheffe may also be used, if all $\hat{t}$'s have equal variance and convariance. For orthogonal data this obvously holds true and so the confidence interval for the contrast $\sum \lambda_a t_a$ will be

$$\sum \lambda_a \hat{t}_a - TS_e \leq \sum \lambda_a t_a \leq \sum \lambda_a \hat{t}_a + TS_e$$ where

$T = 1/2 \sum |\lambda_a| \, q_{\alpha, k, n_e} \cdot \sqrt{v_{11} - v_{12}}$ where $Var(t_a) = v_{11} \sigma_e^2$

$Cov.(\hat{t}_a \hat{t}_b) = v_{12} \sigma_e^2$ and q_{α, k, n_e} is the upper α % point of studentised range $(7, n_e)$.

Since the experimenter's interest is mostly in contracts like (t_a-t_b), (t_c-t_d).........., Tukey's method may be slightly better than Scheffe's for these confidence intervals.

ii) Feeding trials: We have a certain number of feeds distinguished by content of C.D.P. (crude digestible Protein) etc. Further we have the animals, cow say in different lactation stages. The breed may form one replication in which feeds and lactation stages form a two way classification, but with unequal number of observations in each cell, mainly due to lack of control on lactation stages for a given dairy farm, but also due to abnormalities in the individuals like sickness etc. If we are interested in a simultaneous test of the differences in feeds as well as in lactation stages, both of which are supposed to affect the milk yield, we shall be using, method d) for we are sure that contrasts involving parameters of 'feeds' and 'Lactation stages' both are not interesting. This absence of interaction is realistic in very many cases and we can afford to use method d) instead of a) and thus reduce the lengths of confidence intervals.

iii) Qualitative cum Quantitative trials and other factorial experiments

Suppose we have 3 qualities of nitrogen : Urea, Ammonium phosphate and Ammonium nitrate each at 3 doses $0, n_1, n_2$ lbs/ acre. The experiment is laid out as an ordinary factorial experiment 3^2 in r randomized blocks, the only difference in the analysis being that some more error d.f.'s can be obtained. The experimenter is interested in simultaneous tests of the effects of nitrogen - quantity, qualities as also their intraction. But in general only the tests of Quality and Quantity have joint importance while intraction is difficult to interpret. Actually the model, linear or proportional, to be used here, should be one that minimizes the 'intraction'; because by common sense considerations there should not be intraction and in fact, if some test shows it, still it will be difficult to interpret.

We have H_1 and H_2 each having 2 parameters where H_1: Effect of Quantities of nitrogen are zero H_2: Qualities are the same. For the simultaneous level Table 5 may be used to give a simultaneous significance level of .05. The gain in using method b) over method a) will be considerable, for the S.S. for Nitrogen and Qualities are orthogonal and our only interest is in the 2 contrasts of N and Q separately. In fact from Table 5 the reduction in the length of a normalized contrast will be nearly 13 per cent if error d.f. are ≥ 12.

In many other Factorial experiments we test for a hypothesis having 'v' parameters, but if the over all F-ratio is found significant, the 'v-1' degrees of freedom are broken up into 'v-1' orthogonal tests of 1 d.f. each. The properties of such a procedure are still unknown even though it may be argued to use 'α' % level for each of the 'v-1' tests made later; for these are conditional on the first homogeniety test. If instead simultaneous level of significance is deemed necessary for the second part, we should use Nair's Tables. It may be suggested here that if the 'v-1' contrasts of interest can be specified during the planning of experiment, the Nair's Tables should be used to give an overall level of 5% or 1% and not use the F-test at 'v-1' d.f. at all. This will not only be valid from the Probability view point, but also will not increase the computations for the test. Such situations are very common in Factorial type of experiments and attention needs be paid to the use of simultaneous significance level to avoid too many significant results.

IV) In factorial experiments, many times experimentor's interest is in response at a particular dose level and he will be greatly helped if he can get confidence intervals for a series of Responses with a given confidence. Suppose, for instance, we have 3 levels of Nitrogen n_0, n_1, n_2 and the experiments is done in simple randomized blocks. The estimates of the 3 effects (at n_0, n_1, n_2) will be obtained as linear functions of observations. Let $m = a' + b'n + c'n^2 = a + b(n-n_0) + C(n-n_0)^2$, be the response curve where 'm' represents the mean of observations, at dose 'n'. The estimates of a, b and c may be obtained from

$$a = m_0 \qquad m_1 - m_0 = b(n_1 - n_0) + C(n_1 - n_0)^2$$

$$m_2 - m_0 = b(n_2 - n_0) + C(n_2 - n_0)^2$$

b and c will also be found as linear functions of m_0, m_1 and m_2. But the estimates m_0 etc are linear functions of observations, and thus the three estimates a, b, and c will also be linear functions of observations. Thus the estimate of the mean (or response) at any level 'n' will always be a linear function of given observations.

Hence we can find the confidence intervals of the mean at any level 'n' of nitrogen with the joint confidence coefficient obtained by the F- test of homogeneity of the 3 nitrogen levels. And with the same confidence level we can give the confidence interval for the response for any level of of nitrogen, and this is of immediate utility to the practical man.

If the experiment had levels both of nitrogen and Phosphorus the question may be raised: How much the yield will increase by a certain amount of nitrogen and how much by another amount of phosphorus. Then we have two response curves, the estimates of means 'm' and 'M' respectively will always be linear functions of observations and will be mutually orthogonal as well. Hence we should use method b) to have a joint confidence level of 95% say, with $k_1=2$, $k_2=2$.

But if we consider the levels of Nitrogen and Phosphorus together, we shall get a response surface of the form (supposed quadratic)

$$m = a+b_1n+b_2p+c_1n^2+c_2p^2 + dnp$$

The estimates of a, ...d all are linear functions of observations; in fact, of the 2 nitrogen main-effects, 2 phosphorus main-effects and one interaction between N, P. We have to use method a) here with total number of parameters 5, for the response is a linear combination of the parameters of all the three groups.

Worked out Examples:

i) The data refers to two way classification with unequal number of observations in each cell. The observations are total milk yields of all the Beetal goats (from Indigenous Goat Breeding Scheme, Hissar) during the first lactation. Goats are arranged in 4 columns, IInd to fifth 'generation', and 5 rows corresponding to the total period of 15 years (1942 - 1959) divided into well defined periods on the basis of climatic and managemental conditions prevailing on the farm.

<u>Table 7</u>: giving the total Ist lactation yields (lbs) for each all and n_{ij}, the number of 'does' in each cell, is given in brackets:

Generations / Periods	II	III	IV	V	$N_{i.}$
April'42-March'47	8558.75 (24)	4958.25 (6)	- (0)	- (0)	30
April'47-March'50	4168.75 (16)	2204.75 (8)	1107.75 (5)	- (0)	29
April'50-March'52	1034.69 (6)	2190.29 (16)	1757.65 (11)	18.80 (1)	34
April'52-March'54	1675.19 (4)	3827.69 (10)	2863.00 (7)	542.50 (1)	22
April'54-March'57	- (0)	3542.88 (8)	6169.05 (19)	4977.00 (15)	42
$N_{.j}$	50	48	42	17	157

Total corrected S.S. = 5215313.0 $(lb)^2$

<u>Computations</u>: In general for 'p' row effects r_i and 'q' column effects 'c_j' and using the standard analysis of variance model without Interaction as:

$$Y_{ijk} = \mu + r_i + c_j + e_{ijk} \quad (i=1 \ldots, p;\ j = 1, \ldots, q,\ k = 1, \ldots, n_{ij})$$

we get the normal equations for μ, r_i and c_j s as : [Restrictions are : $\sum_i r_i = \sum_j c_j = 0$]

$$N_{..}\mu + \sum_{i=1}^{p} N_{i.} r_i + \sum_{j=1}^{q} N_{.j} c_j = \sum_i \sum_j Y_{ij.} = Y_{...} \text{ say} \qquad Y_{ij.} = \text{cell total}$$

$$N_{i.}\mu + N_{i.} r_i + \sum_j n_{ij} c_j = \sum_j Y_{ij.} = Y_{i..} \text{ say}$$

$$N_{.j}\mu + \sum_i n_{ij} r_i + N_{.j} c_j = \sum_i Y_{ij.} = Y_{.j.} \text{ say}$$

........(76)

Eliminating μ and r_i, equations for c_j are:

$$(N_{.j} - \sum_i n_{ij}^2 / N_{i.})\hat{c}_j - \sum_{k\neq j=1}^{q} (\sum_i n_{ij} n_{ik} / N_{i.})\hat{c}_k = Q_j = Y_{.j.} - \sum_i n_{ij} Y_{i..}/N_{i.}$$

......(77)

Since Q_j's are non-orthogonal linear functions we have to get $q-1$, orthogonalised functions by the help of Gram-Schimidt process. But if n_{ij}'s were all equal these $(q-1)$ linear functions may have been obtained from

$$Q_j - Q_{j'} = (Y_{.j.} - 1/q\, Y_{...}) - (Y_{.j'.} - 1/q\, Y_{...}) = Y_{.j.} - Y_{.j'.} \qquad j \neq j' = 1...q$$

e.g. '$q-1$' orthogonal contrast derived from Fisher's orthogonal Polynomials. For method d) we have to find all scalar products like $(\alpha_i \cdot \alpha_j)$ $i,j = 1,...,q$. In fact, this is easy for

$$(\text{length})^2 \text{ of vector} - Q_j \text{ say } \alpha_j = \sum_i n_{ij} (1 - n_{ij}/N_{i.})^2 + \sum_i \sum_{j' \neq j} n_{ij}^2 / N_{i.}^2 \cdot n_{ij'}$$

$$= \sum_i N_{ij} - 2 \sum_i n_{ij}^2/N_{i.} + \sum_i n_{ij}^3 / N_{i.}^2 + \sum_i n_{ij}^2 / N_{i.}^2 \cdot (N_{i.} - n_{ij})$$

$$= N_{.j} - \sum_i n_{ij}^2/N_{i.} = \text{coefficient of } c_j \text{ in equation (77)}$$

Similarly scalar product of vectors corresponding to Q_j and Q_k $(j \neq k)$ can be shown to be coefficient of c_k in the normal equation for c_j in (77) or vice versa.

So we have to write both sets of normal equations for r_i and c_j, find the corresponding orthogonalised vectors and the values of these orthogonalised linear functions. The S.S. to be used in tests in method d) is

$$\left[(\alpha_1 \cdot y)^2/(\alpha_1 \cdot \alpha_1) + (\alpha_2' \cdot y)^2/(\alpha_2' \cdot \alpha_2') + \ldots,(q-1) \text{ terms} \right]$$ and corresponding expression for rows.

This S.S* may be seen to coincide with the adjusted S.S. for $c_1 ... c_q$ as computed by $\sum_j \hat{c}_j Q_j$ from equations (77); similarly for the rows. The error S.S* will have to be obtained as usual by substraction from Total S.S*, the adjusted column S.S. and unadjusted row S.S* obtained from the $(p-1)$ orthogonal linear functions of $(\bar{y}_{i..} - \bar{y}_{i'..})$ $i \neq i'$, $1,...,p$, where $\bar{y}_{i..} = 1/N_{i.}\, Y_{i..}$

If we apply method a) we have to get the normal equations for one of the two sets of parameters, say columns and further the expected values of orthogonal linear functions formed out of $(Y_{i..}/N_{i.} - Y_{i'..}/N_{i'.})$ which are necessary for finding confidence intervals for contrasts of interest. Thus the labour of computations is not much increased in method d) compared to a).

<u>Method d)</u> <u>Tests of hypotheses</u>

We give below the 2 sets of normal equations in g's and p's. If in a similar case interaction were likely, that could be taken as third group of parameters and generalized method d) applied. The Error S.S. is here based on 149 d.f. which is very large and so large sample theory of method d) will be applied. This implies the assumption that actual population error variance is known.

Since $k_1 = 4$, $k_2 = 3$ we have either to interpolate between the λ-values for (4,4) and (4,2) or else to leave at giving these two bounds for λ. The interpolation has to be linear as a first approximation.

From actual computations for $\alpha = .05$, we have from Table 6

$$10.73 \le \lambda_\alpha \le 15.51 \quad \text{and so} \quad \lambda_d = 13.12 \text{ approx.}$$

$$S_e^2 \equiv \sigma_e^2 = 26,311.2 \text{, so that } 531.3 \le \sqrt{\lambda}\, S_e^2 \le 639.1$$

1) Parameters $p_1, \ldots, p_5$: Normal equations are:

p_1	p_2	p_3	p_4	p_5	P_i
17.73	-8.69	-4.88	-3.17	-1.01	1389.0
	21.96	-5.89	-3.78	-3.59	-1164.8
		25.01	-5.72	-8.52	- 4884.4
			18.38	-5.71	2505.6
				18.83	2134.6

Values of orthogonalised linear functions and lengths etc. are:

	(α_i', y)	$(\alpha_i' . \alpha_i')$	Length $= \sqrt{\alpha_i' \cdot \alpha_i'}$
$P_1' = P_1$	1389.0	17.73	4.211
P_2'	- 484 .2	17.70	4.207
P_3'	- 4708.7	19.80	4.450
P_4'	- 445 .1	12.03	3.468

The S.S.' $\sum_1^{k_1} (\tilde{Y}_i \; , \; y)^2 = 1,258,321.5$

ii) Parameters $g_1 \ldots g_4$: Normal equations are:

g_1	g_2	g_3	g_4	G_1
20.20	-13.86	-5.97	-0.36	392.88
	30.99	-13.36	-3.78	356.86
		26.75	-7.43	-489.95
			11.57	-259.79

Values of orthogonalised linear functions and lengths etc. are:

	(β_j', y)	$(\beta_j' \cdot \beta_j')$	Length $= \sqrt{\beta_j' \cdot \beta_j'}$
$G_1' = G_1$	392.88	20.20	4.494
G_2'	626.41	21.48	4.635
G_3'	134.99	10.81	3.288

The S.S. $\sum_{1}^{k_1} (\tilde{Y}_{k_1+1}, y)^2 = 27,594.7$

<u>Tests of Significance:</u> We want to test simultaneously the hypotheses H_1: (Period effects are zero) & H_2: (Generation effects are zero). The values of $T_1 = 1,258,521.8 / 26311.2 = 47.7$ and of $T_2 = 27594.7/26311.2 = 1.05$. On the simultaneous significance level of .05, the critical limit λ_α' is such that $10.73 \le \lambda_\alpha' \le 15.51$

Hence the hypotheses that period effects are zero is rejected, while the hypothesis for generation effects is accepted. It follows that the confidence intervals for any contrast of $g_1 \ldots g_4$ will always include '0' between the two ends of the intervals. Many times it may not then seem necessary to give these confidence intervals but at least for the sake of illustration we give them also:

<u>Confidence intervals:</u> For $p_1 \ldots p_5$,

Table 8 below gives the contrast of p's along with the value of linear, function estimating it as also the divisor for normalizing each 'contrast':

Table 8

Contrast	p_1	p_2	p_3	p_4	p_5	Estimate	Normalizing factor
1	17.73	-8.59	-4.88	-3.17	-1.01	-1389.0	4.211
2	0	17.70	-8.28	-5.33	-4.09	-484.2	4.207
3	0	0	19.80	-9.08	-10.71	-4708.7	4.450
4	0	0	0	-12.03	12.03	-445.1	3.468

For any normalized contrast of p's, which can always be expressed as a linear function of the 4 contrasts given above, we shall have the length as:

$$531.3 < 1/2 \text{ length} < 639.1$$

For instance: $-445.1/3.468 - l \leq 12.03/3.468\ (p_4 - p_5) \leq -445.1/3.468 + l$

where 'l' may be obtained from interpolation i.e. $l = \sqrt{13.12 \times 162.20} = 587.4$

For $g_1 \ldots g_4$:

Table 9.

Contrast	g_1	g_2	g_3	g_4	Estimate	Normalizing factor
1	20.20	-13.66	-5.97	-0.36	392.88	4.494
2.	0	21.48	-17.45	-4.03	626.41	4.635
3	0	0	10.81	-10.81	134.99	3.288

As before for any linear contrast of g's, when normalized we have the same approximate 1/2 length 587.4 with a confindence of 95%. For instance,

$134.99/3.288 - 587.4 \leq 10.81/3.288\ (g_3 - g_4) \leq 134.99/3.288 + 587.4$

For the sake of illustration we give method a) also.

Method a): Tests of hypotheses

For testing the hypotheses H_1 and H_2 simultaneously we need the S.S. $\sum_1^{k_1+k_2}(U_i, y)^2$. This my be obtained from the 'Adjusted S.S.' of $p_1 \ldots p_5'$, as it is called in ANOVA literature, which is given by the 4 orthogonal vectors already obtained in method d) above plus the S.S. due to vectors $\beta_1, \beta_2, \beta_3$ which are orthogonal to α's. This set of vectors may be taken to be that which gives rise to what is called 'unadjusted' S.S. for g's'. This test will only tell whether the joint hypothesis: $p_1 = p_2 \ldots = p_5 = 0 = g_1 = g_2 = g_3 = g_4 = 0$ is accepted or not but if we also want quasi-independent tests of the 2 component hypotheses we have to use as in the previous case the same statistics $T_1 = \sum_1^{k_1}(\bar{Y}_n, y)^2/S_e^2$ and $T_2 = \sum_1^{k_2}(\bar{Y}_{k_1+n}, y)^2/S_e^2$. The only difference here will be that the critical limits of T_1 will be found from the distribution of $\sum_1^{k_1+k_2}(U_i, y)^2/S_e^2$ which in general is F- with (k_1+k_2, n_e) d.f. In our case n_e being large $\sum(U_i, y)^2/S_e^2$ has a χ^2 distribution on 7 d.f. and so the limit is 14.07 at 5% level of significance.

i) For α's the S.S. from previous method is = 1,258,321.S. The S.S. for β's as obtained from the totals of the 4 columns of Table 7 Viz: 15432.38, 13723.86, 11897.45, 5538.30 based on 50, 48, 42 and 17 observations respectively is:

= 34,671.8

Hence $\sum_1^7(U_i, y)^2 =$ 1,292,993.3 and ratio $\sum(U_i, y)/S_e^2 =$ 49.14 > 14.07

So the hypothesis that both generation effects and period effects are simultaneously zero is rejected at 5% level.

ii) For quasi-independent tests of H_1 and H_2, we have $T_1 =$ 1,258,321.3/26311.2 = 47.7 and $T_2 =$ 27,594.7/26,311.2 = 1.05, as obtained before and again H_1 is rejected that H_2 accepted.

Confidence intervals: We already have 4 contrasts in p's from previous method and we require 3 more β_1, β_2 and β_3 which will involve both g's and p's. These in fact, can be 3 orthogonal contrasts from the 4 generation means $\bar{y}.1, \ldots \bar{y}.4$ and the grand mean $\bar{\bar{y}} \ldots$..

Taking $\beta_1 \equiv \bar{y}.1 - \bar{\bar{y}}..$, $\beta_2 \equiv \bar{y}.2 - \bar{y}.3$, $\beta_3 \equiv \bar{y}.4 - \bar{y}..$

we can find from Gram Schimidt process 3 orthogonalised functions:

$$\beta_1' = \beta_1 \qquad \beta_2' = \beta_2 \qquad \beta_3' = \beta_3 - \beta_1 \frac{(\beta_3 \cdot \beta_1)}{(\beta_1 \cdot \beta_1)} \qquad \text{for}$$

β_2 is orthogonal to both β_1 and β_3, but β_1 is not orthogonal to β_3. The lengths $(\beta_1 \cdot \beta_1)$, $(\beta_2 \cdot \beta_2)$, $(\beta_3 \cdot \beta_3)$ and $(\beta_1 \cdot \beta_3)$ have to be found by actually writing down the functions in terms of y's explicitly.

Table 10 gives the 7 orthogonal contrasts of $p_2 .. p_5, g_1 ... g_4$, along with their estimates and normalising factors:

Table 10

Contrast No.	p_1	p_2	p_3	p_4	p_5	g_1	g_2	g_3	g_4	Estimate	Normalizing factor
1	17.73	-8.69	-4.88	-3.17	-1.01	0	0	0	0	1389.0	4.211
2	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	
4	0	0	0	-12.03	12.03	0	0	0	0	-445.1	3.468
5	.2889	.1853	-.0966	-.0601	-.2675	.6815	-.3057	-.2675	-.1088	11.884	.01863
6	.1250	.0477	-.0714	.0416	-.2857	0	1.0000	1.0000	0	2.625	.04463
7	.1996	.0490	-.1703	-.0981	.0198	.5327	-.4486	-.3925	.3084	34.571	.04948

The confidence region is: $\sum_1^7 [(U_i, y) - \varphi_i]^2 \leq 14.07 \times 26311.2$ and the 1/2 length of any normalized contrast of $p_1 ... p_5, g_1 ... g_4$ is

$$\sqrt{14.07 \times 26311.2} = 608.41$$

For a contrast involving only p's or g's we may use the Tables already given for previous method. But for a contrast involving both groups of parameters we have to take a suitable linear function of the 7 contrasts given above. Compared with the previous method the relative length of a normalised contrast in p's or g's alone is:

608.41/587.4 = 1.037 i.e. the increase is hardly 4%. This is so because $k_1 - k_2$, is 1, very small; but if the difference were highest the loss in using this method would also be higher.

<u>Example 2:</u> With the object of evolving an efficient method of sampling wool from the body of sheep, an experiment was conducted at Poona with 6 cross-bred animals. The data below refer to Fleece-density, arranged for each sheep according to 2 classes: Methods and Regions on the body of a sheep. There are 3 methods relating to samples of size 50, 100 and 150 fibres. The body of the sheep provides 5 regions: Neck, wither, shoulder, side and back; on each of which 1 sq. cm cut as well as a 4 sq. cm. was made. The data refers to only 1 sq. cm. cut.

Sheep no.		Regions R_1	R_2	R_3	R_4	R_5
I	M_1	294	298	149	151	300
	M_2	296	302	164	138	295
	M_3	299	316	165	159	298
II	M_1	179	211	188	234	180
	M_2	167	244	146	164	180
	M_3	180	222	149	139	179
III	M_1	141	201	186	181	248
	M_2	159	185	181	122	281
	M_3	122	165	174	136	243
IV	M_1	128	139	149	91	157
	M_2	155	148	98	93	160
	M_3	134	136	125	84	163
V	M_1	390	268	236	159	418
	M_2	412	316	248	143	520
	M_3	476	274	255	146	429
VI	M_1	299	192	168	105	475
	M_2	317	348	182	98	413
	M_3	335	223	177	102	361

The 6 sheep form 6 repetitions of the Methods & Regions classification. As there appears to be a lot of variation from sheep to sheep it is preferable to first test for heterogeneity of error variances. Our interest is to test for differences in regions as well as in methods and we shall like to have confidence intervals both types of parameters separately. It thus relates to application of method b) for $k_1 = 4$ and $k_2 = 2$, but the hypotheses are 'orthogonal'.

The six separate Analyses for Methods x Regions classification are given below:

Sum of Squares
Sheep no.

Variation due to	d.f.	I	II	III	IV	V	VI
Regions	4	76446.3	9765.1	30178.6	18368.6	195247.3	178973.3
Methods	2	253.2	535.6	844.9	48.5	2905.7	2762.8
Error s_i^2	8	336.1	5415.7	1319.8	1796.2	9052.9	18165.9
Total	14	77035.6	15716.4	32343.3	20213 .3	207205.9	199902.0

$\bar{s}^2 = 1/6 \sum s_i^2 = 6047.7$

Bartlet's test of homogeneity: $\chi^2_5 = \{(\sum n_a)\log_e \bar{s}^2 - n_a \sum^6 \log_e s_i^2\}/c$

where $c = 1+1/3(k-1)[\sum 1/n_i - 1/\sum n_i] = 1+1/3.5 \times 35/48 = 1.049$

Hence $\chi^2_5 = 31.2$ which is highly significant at 1% level.

Thus the necessity of separate tests of significance for each sheep is clear.

Tests etc: For testing H_1: (Region effects are zero) and H_2: (Method effects are zero) simultaneously with the simultaneous significance level of .05, we use Table 5 for $k_1 = 4$, $k_2 = 2$ and find λ(ortho.) = 2.05. From the above table, it is clear that Methods never reach significance, though regions do so for all sheep except II.

Confidence intervals: Since for sheep II no set of parameters shows significance we need not find out confidence intervals for contrasts in parameters for this sheep. Now $\hat{m}_i$ (effect of i^{th} method) = $Y_{i}/5 - Y_{..}/15$ for each sheep. We may take the two orthogonal contrasts in m_i's as (m_1-m_2) and $(m_1+m_2-2m_3)$, as given by Fisher's orthogonal Polynomials and similarly for r's.

Table below shows these contrasts along with the normalising factors.

No.	Contrast	Normalizing Factor
1')	$m_1 - m_2$	$\sqrt{6/5}$
2')	$m_1 + m_2 - 2m_3$	$\sqrt{6/5}$
1)	$-2r_1 - r_2 + r_4 + 2r_5$	$\sqrt{10/3}$
2)	$2r_1 - r_2 - 2r_3 - r_4 + 2r_5$	$\sqrt{14/3}$
3)	$-r_1 + 2r_2 \; 2 \; -2r_4 + r_5$	$\sqrt{10/3}$
4)	$r_1 - 4r_2 + 6r_3 - 4r_4 + r_5$	$\sqrt{70/3}$

Below we give the estimates of 1'), 2') ,...,4) from the samples taken along with the ½-length of confidance intervals for each sheep separately. For, 1/2 length = $\sqrt{\lambda n_e s_e^2} = \sqrt{206} \cdot \sqrt{\text{Error S.S.}} = 1.435 \times \sqrt{n_1^2}$

Sheep no.

Contrasts		I	III	IV	V	VI
Estimates	1')	-0.949	-6.643	3.162	-53.125	-37.630
	2')	-15.884	-19.901	6.268	- 9.129	36.699
of	1)	-83.982	88.228	-5.298	-42.359	25.196
	2)	191.953	56.414	55.394	386.676	351.495
	3)	171.616	123.059	68.102	165.965	221.652
	4)	-55.619	46.925	25.188	128.010	75.219
1/2 length for any normalised contrast		26.308	55.943	60.858	196.388	196.047

Chapter V

Some multiple-decision problems for means of univariate & multivariate normal distns.

Sec. I : In recent years many authors have considered the theory of multiple decision problems, both in the most general form as an extension of Wald's theory of decision functions and the restricted form of the tests of several hypotheses simultaneously. Let X be a random variables with n coordinates whose probability density: $f(x/\theta_1,\ldots,\theta_k)$ involves k parameters, θ. Suppose a number of different hypotheses concerning θ are of interest say H_r: $\theta \varepsilon W_r$, $r \varepsilon T$. The class of alternatives k_r to H_r is that θ lies in the complement of W_r, say W_r^{-1}. This is the most general formulation, due to Lehmann (1957). Though in many practical cases W_r will be hyper-cubes in k dimensions. Following Nandi (1957) suppose that the admissible values Ω_i of each θ_i comprise an interval of the θ_i axis and divide it into m consicutive intervals Ω_{ij} (j = 1,2,....m); $\Omega_{ij} : \theta_i^{j-1} < \theta_i \leq \theta_i^j$ such that $\sum_j \Omega_{ij} = \Omega_i$. With each Ω_{ij} is associated a terminal decision $d_i = j$. The problem is to make terminal decision like $d_1,\ldots,d_k$ where each component d_i can take values from 1 to m. These may be called compound decision problems (Hannan & Robbins ; 1955). The formulation of Lehmann starts with the ordinary Neyman-Pearson formulation of 2 decisions but is structurally a little complicated because of the constant presence of incompatible decisions which have to be eliminated first.

Let the decision function be $\delta(d_1, d_2 \ldots d_k/x)$ which gives the probability of coming to a decision $d_1 d_2 \ldots d_k$ on the basis of sample observation X. Following Lehmann, Robbins and Nandi let the losses be additive that is the loss function $L(d_1,\ldots d_k/\theta_1 \ldots \theta_k)$ is such that

$$L(d_1 \ldots d_k/\theta) = \sum_1^k L(d_i/\theta_i) \quad d_i = 1, \ldots m \quad (i=1,\ldots k) \qquad \ldots\ldots (78)$$

Such additive loss functions may in fact be justified if we can assume that the total loss in terms of cost or Error variance or probabilities of error is affected by the individual decisions independently of each other. But if the decision $d_i = r$ affects the losses in decision $d_j = 1, \ldots m$, additivity does not hold and (78) will change to some function of d_i's and θ's not necessarily linear. Such a case may arise for example, if θ_1 and θ_2 relate to two different correlated characters and the total loss may not be the simple sum but something less or more. But in many important cases the losses are additive or at most a weighted sum, which is clearly included in the formulation (78) itself. The degree of importance of several decisions is thus taken into account.

It has been shown by Nandi that under the Additivity assumption, we need consider decision functions which are simple products of individual decision functions i.e.

$$\delta(d_1, \ldots, d_k / X) = \prod_i \delta_i(d_i/X) \qquad \ldots\ldots(79)$$

Nevertheless, the risk function $r(\theta/\delta)$ written as the sum of several components $r_i(\theta/\delta_i)$ viz: $r(\theta/\delta) = \sum_i r_i(\theta/\delta_i)$ is such that the components may contain all the parameters other than θ_i also. Thus it does not follow that a good compound decision function should be merely the product of the k good individual decision procedures δ_i about θ_i, $(i=1 \ldots k)$ derived independently of all the rest.

If the individual loss functions take only constant values then, assuming continuity of $E_\theta \phi_r(x)$ -- , the power function for H_r (Lehmann's formulation) -- Lehmann has given a powerful result about the unbiasedness and uniform smallness of risk. This relates to the class of tests which are similar on the boundaries of W_r and W_r^{-1} with a certain level $b_r/a_r + b_r$. Though the loss functions are essentially discontinuous this restriction is appropriate as is apparent from the corresponding necessity of unbiasedness in the ordinary Neyman-Pearson formulation. Nandi has also given a similar result for similar loss functions though it is less general for it concerns with only 2 decisions for each parameter. Lehmann's results essentially come from the results about the unbiasedness and uniform smallness of risk functions of each component but they leave out the methods of getting such individual tests. Nandi has given some results in this direction, notably if for each parameter there exists a sufficient statistic separately, then the component decisions depend only on that. Actually he has shown (1960) that under some restrictions about availability of similar - UMP tests, and for the class of functions for which risk functions $r_i(\theta/\delta_i)$ are continuous, the above mentioned decision functions based on separate sufficient statistics form an essentially complete class.

But here, the restriction of continuity, for the type of loss functions is considered, is quite big. Actually the proof for continuous loss functions (over the boundary of H_1 and $\bar{H}_1$, say) breaks down because the condition of continuity is no more a restriction. To overcome this it may be suggested to consider only that class of decision functions for which component risk function $r_i(\theta/\delta_i)$ only involves θ_i ($i=1,..k$), but this restriction is quite big and corresponds to the conditions of 'similarity' in the Neyman-Pearson formulation.

In the general case where $L_i(d_j/\theta_i)$'s are not constants but functions of θ_i, it is difficult to impose some, well recognized restriction like unbiasedness. In fact Lehmann (1951) shows that if L is of the form $L(d,\theta)=f(\theta)\,V(d,d')$ where d' represents the 'correct' decision for 'θ', then except for trivial procedures like those where $E_\theta V(\delta(x), d')=0$ for some d' and some value of θ, no unbiased procedure can exist unless $f(\theta)$ is a constant on each part of parameter space separately. If the problems exhibits certain symmetry, criterion of Invariance may instead be used to reduce the number of decision functions considered.

Sec. 2. In the sequel we shall be using a continuous loss function instead of the customary discontinuous functions. It has been recognized since long that losses like 0,1 or a,b do not represent the true losses for the resulting loss say, from incorrect acceptance of the hypothesis will not be the same for all alternatives. For instance, if we compare the mean of a new variety with respect to a standard one, the loss in terms of cost involved in wrong replacement will vary continuously as $\mu_{stand.} - \mu_{new}$ increases in the positive direction. In a certain sense a loss function proportional to this difference will represent the situation better and in general a monotonic increasing loss function should be used. The same arguments hold for the other wrong decision, that of non-replacement of existing standard variety and here again a monotonic increasing loss function will be generally desirable.

In statistical literature a quadratic loss function is often used in connection with theory of estimation. Even in the Analysis of Variance theory, the 'distance measure' is a quadratic function of several means, in fact simply $\Sigma \mu_i^2 / \sigma^2$. The same is true of 'distance in the theory of classificatory analysis. We shall be using below such a quadratic loss function for test of hypothesis. $\theta_1 \leq 0$ against $\theta_1 > 0$ and show how to derive a complete set of decision functions and minimax procedures. For the sake of simplicity we first consider the simplest case of 1 hypothesis and only two decisions.

Suppose we have n observations $x_1 \ldots x_n$ from a Normal population with mean θ and Variance σ^2 (known). For simplicity we may take $\sigma^2 = 1$. The two possible decisions about the mean θ are: $\theta \leq 0$ say and $\theta > 0$. Since the loss of wrong rejection of hypothesis $\theta \leq 0$ may be more important than that for wrong acceptance, we take the quadratic loss function as a θ^2 for the left (i.e. Reject when $\theta \leq 0$) and b θ^2 on the right.

The risk function has correspondingly two forms:

$$r(\theta/\delta) = a\,\theta^2 \ \ \text{Pr.(Reject } \theta \leq 0 / \ \theta \leq 0\text{)} \ \text{ for } \ \theta \leq 0$$

$$= b\,\theta^2 \ \ \text{Pr (Accept } \theta \leq 0 / \ \theta > 0\text{)} \ \text{ for } \ \theta > 0 \qquad \ldots\ldots\ldots\ldots(80)$$

The decision when $\theta = 0$ is immaterial, as also shown by loss and risk functions.

We shall find all Bayes solutions of the problem (80) and these will provide a complete class of decision functions. If further we can get an a priori distribution, say $\xi(\theta)$ such that the Bayes solution with respect to this, say δ_ξ is such that

$$\sup_\theta r(\theta, \delta_\xi) = R(\xi, \delta_\xi) \qquad \text{.........(81)}$$

where $R(\xi, \delta_\xi)$ is the average risk of δ_ξ with respect to the a priori distribution $\xi(\theta)$; then δ_ξ is also a minimax solution for the problem.

Considering then an a-priori distribution $d\xi(\theta)$ we have

$$R(\xi, \delta) = \int_{-\infty}^{0} a\,\theta^2 \; Pr\,(\text{Reject } H/H)\; d\xi(\theta) + \int_0^\infty b\,\theta^2 \; Pr\,(\text{Accept } H/H \text{ false})\; d\xi(\theta)$$

$$= \int_{-\infty}^{0} a\,\theta^2 \int_X [1 - \delta(x_1 .. x_n)] f(X)\, dX\, d\xi(\theta) + \int_0^\infty b\,\theta^2\, \delta(x_1 .. x_n) f(X)\, dX\, d\xi(\theta) \qquad \text{.........(82)}$$

where $\delta(x_1 .. x_n) \equiv \delta(X)$ takes the value 1 if H is rejected and 0 otherwise, and $dX \equiv dx_1 \,..\, dx_n$.

Interchanging the order of integration, which is permissible and integrating out θ we get,

$$R(\xi, \delta) = \int_X h(X) [1 - \delta(X)]\, dX + \int_X g(X)\ (X)\, dX \qquad \text{.........(83)}$$

This will be minimized (for any given ξ) if $\delta(X)$ is such that it takes values 0 & 1 as:

$$\delta(X) = 1 \quad \text{when } h(X) > g(X)$$
$$= 0 \quad \text{when } h(X) < g(X) \qquad \text{.........(84)}$$

where $h(X) = \int_{-\infty}^{0} a\,\theta^2 f(X/\theta)\, d\xi(\theta)$, $g(X) = \int_0^\infty b\,\theta^2 f(X/\theta)\, d\xi(\theta)$ ....(85)

Thus the decision function which is a Bayes solution with respect to $\xi(\theta)$ is non-randomized.

Now (84) and (85) yield in case of Normal distribution, where $f(X/\theta) = c(X) \exp(-1/2\, n\theta^2 + n\bar{x}\theta)$, the following equations:

$$\delta(X) = 1 \quad \text{if} \int_0^\infty b\,\theta^2 \exp(n\bar{x}\theta - 1/2\, n\theta^2)\, d\xi(\theta) < \int_{-\infty}^{0} a\,\theta^2 \exp(n\bar{x}\theta - 1/2\, n\theta^2)\, d\xi(\theta)$$
$$= 0 \qquad \text{otherwise.} \qquad \text{.........(86)}$$

Instead of working with $d\xi(\theta)$ as it is, we can work with $\xi'(\theta)$ where

$$\xi'(\theta) = a\,\xi(\theta) / [a \int_{-\infty}^{0} d\xi(\theta) + b \int_0^\infty d\xi(\theta)] \quad \text{where } \theta \le 0$$
$$= b\,\xi(\theta) / [a \int_{-\infty}^{0} d\xi(\theta) + b \int_0^\infty d\xi(\theta)] \quad \text{when } \theta \ge 0$$

Then 'a' and 'b' (86) will vanish.

From (86) it can be verified that there exists a constant c_0, depending on ξ (and n) such that the 2 inequalities above are equivalent to ;

$\bar{x} < a$ and $\bar{x} > c$ respectively ('a' and 'b' are positive), c may take irregular values $-\infty$ and ∞ in some cases e.g. if $\int_{-\infty}^{0} d\xi(\theta) = 0$ and $\int_{0}^{\infty} d\xi(\theta) = 0$ respectively. Exactly the same proof will work if instead of $a\theta^2$ and $b\theta^2$ we had $f(\theta)$ and $g(\theta)$ where $f(\theta)$ is monotonic increasing in $\theta \le 0$ and $g(\theta)$ monotinic increasing in $\theta > 0$.

Let $\delta_c(X)$ denote the decision function for any constant c given as $\delta_c = 1$ when $\bar{x} < c$ and equal to 0, otherwise.

Then from Wald's result it follows that any Baye's solution must be identical 'c with $\delta_c(X)$ for some 'c'. The converse is also true, so that for any given 'c' there always exists an a-priori distribution $\xi(\theta)$ such that $\delta(X)$ is a Bayes solution with respect to $\xi(\theta)$. Since the risk function is bounded though the loss function is not, we can say, from an extension of Wald's result of complete class of decision functions about Bayes solutions, that δ_c for all c gives a complete class.

Our main interest then, is to find a minimax solution of the problem, belonging to the complete class δ_c. We shall show below that each risk function corresponding to any δ_c has one single maximum and then proceed to get some δ_{c0} for which this maximum is minimized.

Consider a function $\phi(\theta) = \theta^2 \int_c c' \exp(-1/2\,(x+\theta)^2)\,dx$, $\theta \ge 0$

$$= \theta^2 \int_{c+\theta}^{\infty} c' \exp(-1/2\,y^2)\,dy.$$

$$\phi'(\theta) = 2\theta \int c' \exp(-1/2\,y^2)\,dy + \theta^2 c' \exp(-1/2\,(c+\theta)^2) \qquad \ldots\ldots\ldots(87)$$

Since $\int_{c+\theta}^{\infty} \exp(-1/2y^2)\,dy = \int 1/y.y \exp(-1/2\,y^2)\,dy = (-1/y\,e^{-\frac{y^2}{2}}]_{c+\theta}^{\infty} - \int 1/y^2 e^{-y^2/2}\,dy$

$= 1/c + \theta \cdot \exp(-1/2\,(c+\theta)^2) -$ 'a quantity essentially positive'

$$\therefore\ 2\theta \int \exp(1/2y^2)\,dy < 2\theta/c+\theta \exp(-1/2[c+\theta)^2) = 2/\theta(c+\theta).\ \theta^2 \exp(-1/2(c+\theta)^2) \qquad \ldots\ldots\ldots(88)$$

Hence as $\theta \to \infty$ (or when $2/\theta(c+\theta) < 1$), the value of $\phi'(0)$ becomes -ve i.e. the functional value of $\phi(0)$ decreases uniformly after a certain value of θ and so there exists at least one maximum of $\phi(\theta)$.

Similarly for $\theta \to -\infty$, we can show that there exists at least one maximum of the corresponding function,

$$\psi(\theta) = \theta^2 \int_{-\infty}^{c} c' \exp(-1/2\ (x-\theta)^2)\ dx.$$

Further differentiating $\phi(\theta)$ we have

$$\phi''(\theta) = 2 \int_{c+\theta}^{\infty} c \exp[-y^2/2]\ dy - 4\theta\ c \exp(-1/2\ (c+\theta)^2) + \theta^2(c+\theta)\ c \exp(-1/2(c+\theta)^2) \qquad \ldots\ldots\ldots(89)$$

Also from (87) we get

$$\phi'(\theta) \leq [2/\theta(c+\theta) - 1]\ \theta^2\ c \exp(-1/2(c+\theta)^2) \quad \text{when } \theta > 0$$

This shows that $\phi'(\theta)$ can vanish only at points 'θ' which satisfy

$$2/\theta(c+\theta) \geqslant 1 \quad \text{i.e.}\quad c\theta + \theta^2 \leq 2 \qquad \ldots\ldots\ldots(90)$$

Suppose $\theta = \theta_1$ is such a point, then,

$2 \int_{c+\theta}^{\infty} c \exp(-y^2/2)\ dy = \theta\ c \exp(-1/2(c+\theta)^2)$ which in conjunction with (84) gives: $\phi''(\theta_1) = [-3\theta + \theta^2(c+\theta)] \exp(-1/2(c+\theta)^2)$ at $\theta = \theta_1$ since $\theta \geqslant 0$ and since from (90) $c\theta^2 + \theta^3 \leq 2\theta$ ($\theta \geqslant 0$) we find that $\phi''(\theta_1)$ is -ve.

Thus at any point where $\phi'(\theta)$ may vanish, the second derivative is negative which means that there can exist at most one stationary point for the function $\phi(\theta)$.

Since our risk function for $\delta_c(X)$ is essentially similar to $\phi(\theta)$ we conclude that there is exactly one stationary point and that is a maximum. The argument for the risk function for $\theta \leq 0$ is exactly similar and shows the existance of another maximum.

Consider $r(\theta/\delta) = a\theta^2 \int_{c+\theta}^{\infty} \exp(-y^2/2)\ dy = g_\theta(c)$ say, $\theta \leq 0$.

The function $g_\theta(c)$ decreases monotonically as c increases. Hence the maximum of $g_\theta(c)$ over θ given any c monotonically goes on decreasing as 'c' increases. Conversely, for $r(\theta/\delta) = b\theta^2 \int_{-\infty}^{c-\theta} \exp(-y^2/2)\ dy = f_\theta(c)$, say, $\theta \geqslant 0$. $f_\theta(c)$ goes on increasing as c increases and so the maximum with respect to θ, also increases. This if 'c' increases the maximum risk on one side i..

i.e $Max_{\theta \leq 0}$ [$g_\theta(c)$] falls monotonically while that on the other side viz. $Max_{\theta \geq 0}$ [$f_\theta(c)$] rises monotonically. Considering the $R_{max} = max_\theta$ [$g_\theta(c), f_\theta(c)$], we find that the minimum of this R_{max} will be attained when the 2 components are equal i.e.

$$g_{\underset{\theta \leq 0}{max}}(c) \quad f_{\underset{\theta \geq 0}{max}}(c) \qquad \ldots\ldots\ldots(91)$$

This equation determines 'c_0' which will thus provide us the minimax solution δ_0 .

$$\delta_0 = 1 \quad \text{i.e Reject } \theta \leq 0 \quad \text{if } \bar{x} > c_0$$
$$= 0 \quad \text{i.e. Accept } \theta \leq 0 \quad \text{if } \bar{x} < c_0 \qquad \ldots\ldots\ldots(92)$$

Further since (91) provides essentially 1 point where f_{max} and g_{max} become equal, the solution δ_0 will be unique and so will also be admissible. These results can be stated in the form of a Theorem.

<u>Theorem 1:</u> With a quadratic loss function and corresponding risk functions (80) the minimax admissible decision function for making any of the two decisions $\theta \leq 0$, $\theta \geq 0$ from $f(x, \theta) = c \exp(-1/2(x-\theta)^2)$ is given by (92) where c_0 is determined from equation (91). It may be noted that if $a = b$; 'g' and 'f' being symmetrical functions the equation (91) will give the simple solution $c_0 = 0$.

Due to certain very big complications it is not possible to generalize this result for the case of Normal population with unknown population variance but it is hoped that student's t will provide such a procedure.

<u>Sec. 3</u> If we consider more than one parameter, say means of a multivariate normal population with covariance matrix known, the result obtained before can be generalized under certain restrictions. These relate to the absence of other parameters from the risk function of θ_1 say $r_1(\theta_1/\delta_i)$ where δ depends on all the observations of all the variates. So that

$$r(\theta/\delta) = \sum r_i(\theta/\delta_i) = \sum r_i(\theta_i/\delta_i) \qquad \ldots\ldots\ldots(93)$$

Nandi imposes this restriction from the consideration of 'Unrelatedness of parameters'.

Thus we consider only the class of decision function D_{1s}. We give below a simple result for such decision procedures.

<u>Theorem 2</u>: i) If the loss function is additive, say with respect to only 2 parameters θ_1 and θ_2 and we consider decision function of D_1 then, the compound decision (δ_1^o, δ_2^o) is minimax if δ_i^o are minimax procedures for component problems. The result may be easily extended to more than 2 parameters.

ii) Further, with the same loss function and again confining to the class of decision functions D_{1s} if the solutions δ_1^o and δ_2^o of the component problems are 'admissible', the compound decision procedure (δ_1^o, δ_2^o) is also admissible for the general problem.

<u>Proof</u>: i) Assume the existence of least favourable distributions $f_1(\theta_1)$ and $f_2(\theta_2)$ for the 2 component problems respectively. Then considering the 2-parameters a-priori distribution $[f_1(\theta_1)\, f_2(\theta_2)]_{\theta_1,\theta_2} = \lambda(\theta_1,\theta_2)$ say, we want to get a Baye's solution i.e. to minimize the risk of the product procedure $(\delta_1\ \delta_2)$ which is:

$$R(\delta, \lambda) = \int_{\theta_1}\int_{\theta_2}[r_1(\theta_1/\delta_1) + r_2(\theta_2/\delta_2)]\,[f_1(\theta_1) f_2(\theta_2)]\,d\theta_1\,d\theta_2$$
$$= \int_{\theta_1} r_1(\theta_1/\delta_1)\,f_1(\theta_1)d\theta_1 + \int_{\theta_2} r_2(\theta_2/\delta_2)\,f_2(\theta_2)\,d\theta_2 \qquad \ldots\ldots\ldots(94)$$

Since the first term on the right of (93) is minimized for $\delta_1 = \delta_1^o$ and the second term for $\delta_2 = \delta_2^o$; (δ_1^o, δ_2^o) minimize $R(\delta, \lambda)$ i.e. $(\delta_1^o\ \delta_2^o)$ is a Bayes solution for the compound decision problem. Now,

$$R'(\delta^o, g) = \iint [r_1(\theta_1/\delta_1^o) + r_2(\theta_2/\delta_2^o)]\,g(\theta_1,\theta_2)\,d\theta_1\,d\theta_2, \quad \text{for any a-priori distribution } g(\theta_1\ \theta_2)$$

$$= \int [r_1(\theta_1/\delta_1^o)\,[\int g(\theta_1\ \theta_2)\,d\theta_2]d\theta_1 + \int r_2(\theta_2/\delta_2^o)\,[\int g(\theta_1\theta_2)d\theta_1]d\theta_2$$
$$R(\delta^o, g) = \int r_1(\theta_1/\delta_1^o)\,h_1(\theta_1)\,d\theta_1 + \int r_2(\theta_2/\delta_2^o)h_2(\theta_2)d\theta_2, \text{ say} \qquad \ldots\ldots(95)$$

To maximize (95) with respect to 'g' i.e. with respect to $h_1(\theta_1)$ and $h_2(\theta_2)$

we note that $h_1(\theta_1)$ must be equal to $f_1(\theta_1)$ a.e. as also $h_2(\theta_2) = f_2(\theta_2)$. Hence $g \equiv f_1 \cdot f_2$ a.e. and so the Bayes solution above $(\delta_1^\circ, \delta_2^\circ)$, corresponds to the a least favourable distribution g and is therefore a minimax solution.

ii) Consider any other decision procedure (δ_1, δ_2) say and suppose it is uniformly better than $(\delta_1^\circ, \delta_2^\circ)$ viz over the whole domain of θ_1 and θ_2. We will show that this leads to a contradiction proving the admissibility of $(\delta_1^\circ, \delta_2^\circ)$.

Since δ_1 is not uniformly better than δ_1° for the component problem, there exists at least 1 point on the θ_1 axis where the risk for δ_1° viz: $r_1(\theta_1' / \delta_1^\circ)$ is less than $r_1(\theta_1' / \delta_1)$. (In the other case the 2 procedures δ_1 and δ_1° can only be essentially the same). Similarly considering δ_2, there exists at least one point, say θ_2' on the θ_2 axis suchthat $r_2(\theta_2' / \delta_2^\circ) < r_2(\theta_2' / \delta_2)$. Now consider the point (θ_1', θ_2') for the compound decision problem. On adding up we get,

$$r_1(\theta_1'/\delta_1^\circ) + r_2(\theta_2' /\delta_2^\circ) < r_1(\theta_1' /\delta_1) + r_2(\theta_2'/ \delta_2)$$

$$\text{Or} \quad r(\theta'/\delta^\circ) < r(\theta'/\delta)$$

But this contradicts the assumption that δ is uniformly better than δ°, leading to the result. As before the result can be immediately extended to the case of k parameters.

It should be noted that the results of this section are true for the general case of 'm' decisions for each parameter. Further the restriction to decision functions satisfying the condition $r(\theta / \delta) = \sum r_i(\theta_i/\delta_i)$ is less restrictive than making independent tests of hypotheses for each parameter θ, which is sometimes suggested. Actually the demand of obtaining independent tests is unnecessary for the above condition is much weaker and can be satisfied by a weaker concept of Quasi-independence. For example in the standard Analysis of Variance case, though the different tests are not independent they satisfy the above assumption and thus for additive loss function, quasi-independence should be at least as good as independence of tests. However, nothing can be said about the optimality of the compound decision procedures if the decision functions are not restricted to the class D_1.

Sec. 4. In the more general case where the assumption (93) need not hold, it is found that the Bayes solution are very difficult to find explicitly. For example, consider the 2 parameter case of 2 means of Bivariate normal distribution i.e. $\theta_1 \leq 0$, $\theta_1 \geq 0$ and $\theta_2 \leq 0$, $\theta_2 \geq 0$, the variances and covariance being known. Here, since $(\bar{x}, \bar{y})$ the sample means provide a sufficient statistic for (θ_1, θ_2), we need consider decision functions based on them alone. The loss functions are say:

i) $a\theta_1^2$ when $\theta_1 \leq 0$ and $b\theta_1^2$ when $\theta_1 \geq 0$ ii) $c\theta_2^2$ when $\theta_2 \leq 0$ and $d\theta_2^2$ when $\theta_2 \geq 0$

then risk function is:

$a\,\theta_1^2$ Pr (Reject H_1 that $\theta_1 \leq 0$) + $c\,\theta_2^2$ Pr (Reject H_2/that $\theta_2 \leq 0$) $(\theta_1 \leq 0,\ \theta_2 \leq 0)$

and similar other 3 forms for the other 3 parameteric regions.

Considering an a-priori distribution $d\,\xi(\theta_1, \theta_2)$ as before, it willbe found that Baye's solutions $\delta = \delta_1 \delta_2$ are such that:

$$\delta_1(\bar{x},\bar{y}) = 1 \text{ if } \int_0^\infty b\theta_1^2 \int_{-\infty}^{\infty} kf(\bar{x},\bar{y})\, d\xi(\theta_1\theta_2) < \int_{-\infty}^{0} a\theta_1^2 \int_{-\infty}^{\infty} kf(\bar{x},\bar{y})\, d\xi(\theta_1\theta_2)$$

$$= 0 \quad \text{otherwise}$$

$$\delta_2(\bar{x},\bar{y}) = 1 \text{ if } \int_0^\infty d\theta_2^2 \int_{-\infty}^{\infty} kf(\bar{x},\bar{y})\, d\xi(\theta_1\theta_2) < \int_{-\infty}^{0} c\theta_2^2 \int_{-\infty}^{\infty} kf(\bar{x},\bar{y})\, d\xi(\theta_1\theta_2)$$

$$= 0 \quad \text{otherwise.}$$

From these it appears that Baye's solutions are such that δ_1 involves both $\bar{x}$ and $\bar{y}$ in general. In fact if $d\,\xi(\theta_1\theta_2)$ be a simple 2-point distribution with probability a/a+b for the point (p, q) and b/a+b for $(-p, q)$ we get:

$$\delta_1 = 1 \text{ if } \operatorname{Exp.} -n/2\left[(\bar{x}-p)^2/\sigma_1^2 - 2\rho(\bar{x}-p)(\bar{y}-q)/\sigma_1\sigma_2 + \{(\bar{x}+p)^2/\sigma_1^2 - 2\rho(\bar{x}-p)(\bar{y}-q)/\sigma_1\sigma_2\right] < 1$$

$$\text{i.e. } 4p\left[\bar{x}/\sigma_1^2 - (\bar{y}-q)/\sigma_1\sigma_2\right] < 0$$

$$\text{or } \bar{x} - \rho\frac{\sigma_1}{\sigma_2}(\bar{y}-q) < 0 \text{ if } p > 0$$

If is very difficult to proceed with Baye's solutions in general, and it cannot be said that minimax and admissible solutions will involve the other set of observations or not.

Instead of trying Baye's solutions and minimal complete class of decision functions we proceed below to show that under certain restrictions with loss functions of the type 0 or 1, the two component problems will involve the corresponding variate only in the decision functions.

We shall use the language of testing of hypotheses instead of the more general decision functions. Suppose x and y come from the Bivariate normal distribution with known covariance matrix and suppose the hypotheses to be tested jointly are: $\theta_1 \leq 0$ and $\theta_2 \leq 0$. Since $(\bar{x}, \bar{y})$ form a sufficient statistic for $(\theta_1 \; \theta_2)$ the 'best' critical function will be of the form $\phi(\bar{x}, \bar{y})$ which may or may not be randomized critical functions.

In the Neyman-Pearson formulation we may like to impose the condition of similarity on the 2 components tests of θ_1 and θ_2. But this is a big restriction many times and it is more realistic to impose the lesser restriction that the Ist kind of error be Bounded above by a fixed quantity, say α. We then consider IInd kinds of error only, to find 'good' tests.

We have the distribution of sample as :

$$f_n(x, y) = \left(1/2\pi\sigma_1\sigma_2\sqrt{1-\rho^2} \right)^n \exp -1/2(1-\rho^2) \sum \left[(x-\theta_1)^2/\sigma_1^2 + (y-\theta_2)^2/\sigma_2^2 - 2\rho(x-\theta_1)(y-\theta_2)/\sigma_1\sigma_2 \right]$$

$$= f(\bar{x}, \bar{y}, \theta_1, \theta_2) \cdot \psi(x_1 \ldots x_n ; y_1 \ldots y_n) \text{ ——— (free of } \theta_1 \text{ and } \theta_2) \quad \text{where}$$

$$f(\bar{x}, \bar{y}) = \text{const.} \; \text{Exp} - n/2(1-\rho^2) \left[(\bar{x}-\theta_1)^2/\sigma_1^2 + (\bar{y}-\theta_2)^2/\sigma_2^2 - 2(\bar{x}-\theta_1)(\bar{y}-\theta_2/\sigma_1\sigma_2 \right]$$

$$= \text{const.} \; \text{Exp} - n/2 \left[(\bar{x}-\theta_1)^2/\sigma_1^2 + 1/\sigma_2^2(1-\rho^2) \left\{ \bar{y}-\theta_2 - \rho\frac{\sigma_2}{\sigma_1}(\bar{x}-\theta_1) \right\}^2 \right]$$

$$\text{Or } f_1(x, z) = c. \; \text{Exp} \left[-a(x-\theta_1)^2/2 - b(z-\theta_2+p\theta_1)^2/2 \right] \qquad \ldots\ldots(96)$$

where $x = \bar{x}$, $z = \bar{y} - \rho\frac{\sigma_2}{\sigma_1}\bar{x}$; a, b, p are constants.

The transformed variates x, z of (96) are independent. Considering first the test of hypothesis relating to θ_1 we find that the first restriction is:

Make the test for $H_1 : \theta_1 \leq 0$ such that,

$$\text{Const.} \iint \phi(x, z) \exp \left[-a(x-\theta_1)^2/2 - b(z-\theta_2+p\theta_1)^2 \right] dx\, dz \leq \alpha \qquad \ldots\ldots\ldots(97)$$

for all $\theta_1 \leq 0$ and all θ_2.

Further we want to choose the critical function $\phi(x,z)$ such that the power is maximized for all $\theta_1 > 0$ (and all θ_2). .... (B)

Consider a particular "simple alternative" (θ_1,θ_2) with $\theta_1 > 0$, against the null composite hypothesis $\theta_1 \leq 0$. Form an apriori distribution over $(\theta_1,\theta_2)_{\theta_1 \leq 0}$ such as it assigns probability 1 to the single parameter point (θ_1', θ_2') where $\theta_1' = 0$; $\theta_2' = \theta_2 - p\theta_1$.

Let the averaged null hypothesis (simple) be:

$$H_\lambda: \quad \text{Const.}\ \psi(X,Y) \int_{\theta'} \exp\left[-a(x-\theta_1')^2/2 - b(z-\theta_2' - p\theta_1')^2/2\right] d\lambda(\theta') \quad \theta=(\theta_1,\theta_2)$$
$$\text{Const.}\ \psi(X,Y) \quad \text{Exp}\left[-ax^2/2 - b(z-\theta_2-p\theta_1)^2/2\right] \qquad \text{.......(97)}$$

while the simple alternative is,

$$\bar{H}: \quad \text{Const.}\ \psi(X,Y)\ \exp\left[-a(x-\theta_1)^2/2 - b(z-\theta_2-p\theta_1)^2/2\right] \qquad \text{.......(98)}$$

From Neyman-Pearson lemma the most powerful level α - test of (97) against (98) is: Reject $\theta_1 \leq 0$ if, $x > z_{\alpha\%}$ - where z_α represents the α % right hand tail point of the normal distribution. Thus we have,

$$\phi^*(x,z) = 1 \quad \text{if} \quad x > z_{\alpha\%}$$
$$= 0 \quad \text{otherwise} \qquad \text{..........(99)}$$

The critical function (99) satisfies the condition (A) for the test of H_1 and is also the unique test (with respect to $\lambda(\theta)$). By Theorem 7 of Lehmann ("Testing of stat. hypotheses" p.91), it follows them, that (99) also gives the most powerful test for H_1 (composite hypothesis) against the simple alternative $(\theta_1,\theta_2)_{\theta_1>0}$

Further, this unique test is the same for any such simple alternative (θ_1,θ_2) and so it must also be the unique admissible test for $\theta_1 \leq 0$ against $\theta_1 > 0$ for our problem.

We shall give a lemma to show that the above uniformly most powerful test provides the only admissible solution for our decision problem.

<u>Lemma:</u> Consider the problem of testing a simple hypothesis p_0 against a simple, alternative p_θ say. Let ϕ_θ be the critical function of the most powerful test i.e. $E_0(\phi_\theta) = \alpha$ and $E_\theta(\phi_\theta)$ is maximum, where E_0 denotes expectation with respect to p_0 and E_θ that with respect to p_θ.

Let $\lambda(\theta)$ be a two point a-priori distribution, for which $Pr(\theta_1) = \gamma_1$, and $Pr(\theta_2) = \gamma_2$; $\gamma_1 + \gamma_2 = 1$. We have to find a critical function ϕ such as

$E_0(\phi) = \alpha$ and $E_\lambda(\phi)$ is Maximum, where E_λ denotes expectation with respect to $p_\lambda = \int p_\theta \, d\lambda(\theta)$.

Now $E_\lambda(\phi) = \gamma_1 E_{\theta_1}(\phi) + \gamma_2 E_{\theta_2}(\phi)$

If now $\phi_{\theta_1} = \phi_{\theta_2}$, then putting $\phi = \phi_{\theta_1}$ we get $E_0(\phi) = \alpha$ and

$$E_\lambda(\phi) = \gamma_1 E_{\theta_1}(\phi_{\theta_1}) + \gamma_2 E_{\theta_2}(\phi_{\theta_2}) \geq \gamma_1 E_{\theta_1}(\phi) + \gamma_2 E_{\theta_2}(\phi) = E_\lambda(\phi)$$

Hence E ( ) is maximized for $\phi = \phi_{\theta_1} = \phi_{\theta_2} = \phi_\theta$

Similarly this will hold for any a-priori distribution assigning positive probabilities to a finite number of points θ_i, and so in the limit, for all a-priori distributions which are Borel - measurable.

<u>Corollary:</u> When we have a composite null hypothesis we replace the equality

$E_0(\phi) = \alpha$ by $E_{\theta'}(\phi_\theta) \leq \alpha$ for all $\theta' \in w'$, if the composite hypothesis is $H : (\theta' \in w')$.

The above procedure of proof holds exactly as it is, if a unique most powerful test exists according to Lehmann's method of least favourable distribution referred to before.

From the above corollary we see that the test $x > Z_\alpha$ derived before is the only admissible solution for our restricted decision problem (at level α). The proof runs parallely for the test of corresponding hypothesis H_2: $\theta_2 \leq 0$, where we got the solution $y \equiv \bar{y} > Z_{\alpha\%}$.

Hence it follows that under the mild restrictions imposed on the risk functions, the only admissible solution of the multiple decision problem for θ_1 and θ_2

is simply the product of the two standard procedures $\bar{x} > Z_{\alpha_1\%}$ and $\bar{y} > Z_{\alpha_2\%}$ where α_1 and α_2 are suitably chosen, so that the Ist kind of error for the compound problem is $\leq \alpha$ a pre-assigned quantity.

Incidentally the tests obtained above are 'similar' on the boundary and are similar to what Nandi obtains by placing the condition of continuity on the risk functions.

<u>Generalization</u>: In case of p such correlated characters of a multi-variate normal population from which we have taken a sample of n observations, we impose the 'p' restrictions on the first kinds of errors of the p component hypotheses, i.e.

A') $E\left[\phi_i(x_1 \ldots x_p)\right] \leq \alpha_i$ for $\theta_i \leq 0$ $i = 1, \ldots p$ and,

B') $E\left[\phi_i(x_1 \ldots x_p)\right]$ is maximum for $\theta_i > 0$ (and all other θ's).

If the covariance matrix is completely known, the previous proof of H_λ against a simple alternative holds with some obvious changes. Instead of x and Z we shall get the i and p-1 variables- like:

$$x \equiv x_i$$

$$Z \equiv (Z_1, \ldots Z_{i-1}, Z_{i+1}, \ldots Z_p) \quad \text{where}$$

$Z_j = \bar{x}_j - \beta_{ji}\,\bar{x}_i$ w, β_{ji}'s being the regression coefficient of x_j on x_i and is known for all $j = 1 \ldots i-1, i+1, \ldots p$; $i=1, \ldots p$.

This transformation willyield the standard test for H_i; $\theta_i \leq 0$ and so in 'p' steps the final result comes. Thus the only admissible solution to the problem under restrictions A') and B') are the simple productsof the p standard procedures viz:

$(x_i > Z_{\alpha_i}$; $i=1, \ldots p)$ where α_i's are so chosen that the Ist kind of error for compound problem is $\leq \alpha$.

If, covariance matrix is partially or completely unknown, the previous method fails to give any most powerful test against a simple alternative. But it is hoped that the best test will again come out as the product of several t-tests. It should be noted in this connection that Lehmann (1948) has shown that for the single-variate Student's hypothesis (σ unknown) the most powerful test for a simple alternative (θ_1, σ) is Student's 't' only if $\alpha \geqslant 1/2$. But this result cannot be easily generalized.

-******************-

Summary

Simultaneous tests of significance are of very common occurrence especially in the Analysis of Variance. For performing such tests many times orthogonality of the estimates corresponding to the hypotheses is demanded. But this being insufficient for the criterion of independence of various tests, which itself is too strict, another notion of quasi-independence is presented. In many situations where orthogonality is an exception rather than, a rule, such quasi-independent tests can still be constructed and have been presented by Ghosh (1955). In carrying out such tests simultaneously on the same data we may consider only the individual levels of significance if the decisions are more or less unrelated. In the other case we have to combine such levels of significance in some meaningful way. One of these methods is based on the notion of simultaneous level of significance. Several examples have been discussed to show the necessity of using simultaneous level of significance under different conditions.

Treating the tests of several hypotheses in the frame work of multiple-decision theory of statistical decision functions, additive loss functions have been used. For convenience, minimax and Bayes solutions have been obtained for Normal distribution. Under certain conditions, the optimum properties of the product decision procedure can be derived from those of individual procedures. These conditions are satisfied if we consider quasi-independent tests showing that in such cases quasi-independence is at least as good a criterion for optimality as independence. Continuous loss functions are sought to be used for they are more realistic. Under certain less strict conditions product procedures for testing the means of a Bivariate normal distribution are obtained and can be generalized for the multivariate case.

Various methods proposed by several authors for orthogonal case have been discussed but the main attention is given to Ghosh's method for the case of non-orthogonality. The 3 methods proposed by Ghosh have been discussed and compared

on the basis of numerical computations because theoretical comparison is not possible. It has been shown further that, if there are only 2 hypotheses with k_1 and k_2 parameters with $k_1 > k_2$, then a new method can be evolved which is always better than the extension of Scheffe's method a) of constructing confidence intervals for all contrasts. Tables and methods of computation are given to determine exact simultaneous significance level. These involve Incomplete B - or γ - integrals and a table is provided to use this method in practice. The approach for more than 2 hypotheses has been indicated but it seems to be very cumbersome and so further work is needed in this field.

A number of situations arising out of Agricultural and Animal Husbandry fields have been discussed where these methods may be found useful and the necessity of using simultaneous level of significance shown, whether the data be orthogonal or not. For the sake of illustration of the methods presented two examples have been actually worked out, giving detailed calculations and methods for simplifying them. These examples have been taken from data collected in various I.C.A.R. Schemes.

-***********-

References.

1. Bose, R.C. & Roy, S.N. (1953) : Simultaneous confidence interval estimation, Ann.Math.Stats.24

2. Cochran, W.G. (1941) : The distribution of the largest of a set of Variances as a fraction of their total. Ann.Eng. 11.

3. Duncan, D.B. (1955) : Multiple range tests and comparison tests, Biometrics, 11.

4. — (1952) : On the properties of the Multiple comparisons tests, Virginia Jr. of Sc.

5. Finney, D.J. (1941) : The Joint distribution of variance ratios based on a common error mean square. Ann. Eng., 11

6. Ghosh, M.N. (1955) : Simultaneous tests of linear hypotheses, Biometrika, 42

7. Hannan & Robbins (1954) : Asymptotic solutions of the compound decision problem for 2 completely specified distributions, Ann. Math.Stats.,25.

8. Hartley, H.O. (1938) : Studentization and large sample theory; Supp.Jr.R.S.S.,5

9. Hsu, P.L. (1941) : Analysis of Variance from power-function stand-point, Biometrika, 32.

10. Kendall, M.G. (1948) : Advanced theory of statistics, vol II, Charles Griffin & Sons.

11. Kimball, A.W. (1951) : On dependent tests of significance in the Analysis of Variance, Ann. Math.Stats.,22

12. Kramer, C.Y. (1956) : Extension of Multiple range tests to group means with unequal number of replications, Biometrics, 12

13. — (1957) : Extension of Multiple range tests to group correlated adjusted means, Biometrics, 13.

14. Lehmann, E & Stein, C.(1948) : Most powerful tests of composite hypotheses, Ann.Math.Stats,19

15. Lehmann, E. (1950) : Some principles of the theory of testing hypotheses, Ann. Math. Stats.,21

16. — (1951) : General concept of unbiasedness, Ann.Math.Stats.,22

17. — (1957) : Theory of some multiple decision problems, Ann.Math.Stats.,28

18. — (1960) : Testing of Statistical Hypotheses, John Wiley & Sons, New York.

19. Nair, K.R. (1948) : The studentized form of the extreme mean square test in the ANOVA, Biometrika, 35

20. Nandi, H.K. (1951) : On the Analysis of Variance test, Bul.Calc.Stat.Assoc.3.

21. Nandi, H.K. (1957) : On some optimum compound decision procedures, Bul.Calc.Stat. Asson., 7

22. -- (1960) : A complete class of compound decision procedures, Bul.Calc. Stat. Asson., 9

23. Pearson, K. (1934) : Tables of Incomplete B-functions. The Biometrika office, London.

24. Pillai, K.C.S.& Ramachandran, K.V. (1954) : Distribution of a Studentized order statistic, Ann.Math. Stats.,25

25. Ramachandran, K.V. (1956,a) : On Tukey's test and Hartley's test, Ann.Math.Stats.,27

26. -- (1956,b) : On simultaneous ANOVA test, Ann.Math.Stats.,27

27. -- (1956,c) : Contributions to Simultaneous confidence interval estimation, Biometrics, 12.

28. -- (1956,d) : Studentized largest & smallest χ^2 (Abstract), Ann.Math. Stats., 27.

29. -- (1957) : Decision procedure based on Tukey's statistic, Ann.Math. Stats., 28.

30. Scheffe', H. (1953) : A method of judging all contrasts in the ANOVA, Biometrika,40.

31. Tukey, J. (1949) : Comparing individual means, Biometrics, 5

32. Wald, A. (1950) : Statistical Decision functions, John Wiley & Sons, New York.

****

APPENDIX

Table 11: Showing values of λ_α and $n_e \cdot \lambda_\alpha$ (in brackets) $\alpha = .05$

Pairs (k_1, k_2) \ n_e	4	6	8	10	12	30
(3,2)	6.93 (27.72)	3.25 (19.50)	2.06 (16.48)	1.485 (14.85)	1.156 (13.87)	0.378 (11.34)
(4,2)	7.77 (31.08)	3.60 (21.60)	2.26 (18.08)	1.639 (16.30)	1.275 (15.30)	0.411 (12.33)
(5,2)	8.85 (35.40)	4.03 (24.18)	2.53 (20.24)	1.816 (18.16)	1.411 (16.93)	0.442 (13.26)
(6,2)	9.87 (39.48)	4.53 (27.18)	2.83 (22.64)	2.040 (20.40)	1.574 (18.84)	0.498 (14.94)
(7,2)	11.1 (44.4)	5.11 (30.66)	3.16 (25.28)	2.268 (22.68)	1.737 (20.84)	0.571 (17.13)
(8,2)	12.4 (49.6)	5.66 (33.96)	3.50 (28.00)	2.501 (25.01)	1.929 (23.15)	0.611 (18.33)
(5,4)	12.3 (49.2)	5.62 (33.72)	3.49 (27.92)	2.489 (24.89)	1.924 (23.09)	0.616 (18.48)
(6.4)	12.9 (51.6)	5.82 (34.92)	3.59 (28.72)	2.561 (25.61)	1.975 (23.70)	0.625 (18.75)
(7,4)	13.7 (54.8)	6.08 (36.48)	3.75 (30.00)	2.666 (26.66)	2.066 (24.79)	0.646 (19.38)
(8,4)	14.1 (56.4)	6.33 (37.98)	3.94 (31.52)	2.804 (28.04)	2.160 (25.92)	0.674 (20.22)
(8,6)	17.8 (71.2)	8.14 (48.84)	5.01 (40.08)	3.541 (35.41)	2.705 (32.46)	0.844 (25.32)

NB: 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.