V. 80
V 95

MIXED MODEL :

ANALYSIS OF VARIANCE AND COVARIANCE

&

TWO STAGE TEST PROCEDURE

By

A.K.P.C. SWAIN.
(1962 - 63)

Dissertation
submitted in fulfilment of
the requirements for the award of Diploma
in Agricultural and Animal Husbandry Statistics of the
Institute of Agricultural Research Statistics
(I.C.A.R.)
New Delhi.

ACKNOWLEDGEMENTS

I am much indebted to Dr. M.N. Ghosh, Professor of Statistics, Institute of Agricultural Research Statistics, Indian Council of Agricultural Research, New Delhi for his valuable guidance, keen interest and constant encouragement during the course of investigation and of preparation of the thesis.

I am also highly thankful to Dr. V.G. Panse, Statistical Adviser, Institute of Agricultural Research Statistics, Indian Council of Agricultural Research for providing me with adequate research facilities for this work.

A.K. P.C. Swain.

(A.K. P.C. Swain)

Part I

Mixed Model Analysis of Variance

Chapter 1

INTRODUCTION

In a program of research, agricultural experiments on a single factor or a group of factors are usually carried out at a number of places over a number of years. It is generally seen that the effect of most factors (fertilizers, varieties etc.) varies considerably from place to place and from year to year, owing to differences in soil, agronomic practices, climatic conditions and other variations in environment. Consequently the results obtained from experiments at a single place in a single year, however accurate and trustworthy may be are of limited utility for determining the most profitable variety, level of manuring and suitable agricultural practices. To find the response of manures and practices in different soil climatic regions and to select the best manure or practice in the region, it is essential that the experiments should be scattered over a region, without being concentrated at a single place. In India fertilizer trials have been extensively conducted at various centres including farmers' lands to set up optimum standards of manuring for different soil climatic regions of the country, so that the differential response or interaction of fertilizer and fields within a given soil-climatic region is small, if any. If this interaction is not small, the further subdivisions of a given region by the physical or chemical properties of the soil or the climatic factors including rainfall, temperature etc. will make the interaction small within the subregions and hence making further optimal use of fertilizers. Thus the problem reduces to test hypothesis that the component of variation due to interaction of treatment and locations within the given region is negligible.

Yates and Cochran (1938) analysed groups of experiments assuming a mixed model with treatments as fixed factors and locations or places as random factors. The places were selected randomly from the entire region representing the population. They assumed the model

$$Y_{ijk} = \mu + \alpha_i + b_j + \lambda_{ij} + f_{jk} + e_{ijk}$$

Where Y_{ijk} = observation on i^{th} treatment applied in k^{th} field or block of j^{th} place.

μ = general mean

α_i = effect of i^{th} treatment

b_j = effect of j^{th} place

λ_{ij} = effect of the interaction of i^{th} treatment with j^{th} place

f_{jk} = effect of k^{th} field or block in j^{th} place

and e_{ijk} = random error.

The underlying assumptions about the model were

$$\alpha_i = \text{constant for all places}$$

$$E(\lambda_{ij}) = 0; \quad V(\lambda_{ij}) = \sigma_\lambda^2 \text{ independently of } i \text{ \& } j$$

$$E(b_j \lambda_{ij}) = 0; \quad E(e_{ijk}) = 0; \quad V(e_{ijk}) = \sigma_e^2$$

and b_j, λ_{ij}, e_{ijk} are independent normal variables.

In their analysis, heterogeneity of interaction variance that is, a dependence of σ_{λ}^2 on place and treatment is likely to be of frequent occurrence. For example, at one place the experimental material may be unresponsive to all the treatments, while at another place, there may be little response to a few treatments and very large response to other treatments. Such situations make the usual test of treatment m.s. by treatment x place m.s. ineffective and moreover there will be little interest in an exact test of hypothesis that all α_i's are equal to a constant.

The assumption of independence of places and interaction of places x treatment postulated by Yates and Cochran does not seem to be justified in most of the situations. For example, consider two different regions A and B, so that A has a uniformly distributed rainfall and B is arid. It is well-known that the response to nitrogenous fertilizer is much better in A than in B. Thus comparing two treatments (1) no fertilizer, (2) 20 lbs. of nitrogen per acre, we generally have

$$b_A > b_B$$

$$\text{and} \quad \lambda_{2A} > \lambda_{2B},$$

showing that the b_j and λ_{ij} are correlated. Such a case was actually observed in experiment on cotton in India, where it was found that in many cases, use of nitrogenous fertilizers in high yielding soil gave relatively better response per unit of fertilizer than in poorer soil ("Manuring of Cotton in India" - Indian Central Cotton Committee Publication). To have realistic picture of the situation and not to confront with difficulties in the process of analysis encountered in the case of heterogeneity of interaction component, we shall consider a model suggested by Scheffe (1956) for an industrial experiment with machines and workers.

Scheffé considers an experiment involving I machines and J workers. The J workers are assumed to be a random sample from a large population of workers. Each worker is put on each machine for K days during the tenure of the experiment. If Y_{ijk} is a measurement of the output of the j^{th} worker on k^{th} day that he is assigned to i^{th} machine, he puts forward the model

$$Y_{ijk} = \mu + \alpha_i + b_j + C_{ij} + e_{ijk}$$

where the general mean μ and the machine effects $\{\alpha_i\}$ are constants and where the effects $\{b_j\}$ of workers, the interaction $\{C_{ij}\}$ and errors $\{e_{ijk}\}$ are random variables about whose joint distribution certain assumptions are made.

He, therefore, assumes $Y_{ijk} = m_{ij} + e_{ijk}$ where the errors $\{e_{ijk}\}$ are normally, independently and identically distributed with zero mean and variance σ_e^2 and independently of the true means $\{m_{ij}\}$. The I random variables $\{m_{ij}\}$ are the component of a vector random variable $m = (m_{1j}, \ldots\ldots m_{Ij})'$ whose multivariate normal distribution is the basic concept of the model.

Under the above assumptions, he calculates the expectation of mean squares (with suitable definitions of variance components) usually calculated in the analysis of variance and studies the behaviour of the analysis of variance test under this model. He shows that for the test of interaction component, the usual analysis of variance test has good properties but for testing the hypothesis of the equality of treatment means one has to use a Hotelling's T^2 - Statistic.

Many statisticians regard Scheffe's model as more suited for the analysis of agricultural experiments spread over large tracts. However, the assumption of equality of variance of the error term e_{ijk} in different regions may not possibly be realistic, a point already noted by Scheffe in his book "Analysis of Variance". M.R. Sampford while commenting on a paper entitled "Models in the Analysis of Variance" read by Plackett (1960) at a Research methods meeting of the Royal Statistical Society considered the applicability of Scheffe's model to an agricultural situation. He, however, does not consider the assumptions regarding error in Scheffe's model as quite realistic.

We shall consider Scheffe's model as a first approximation to the real nature of variation in such experiments and use them until better models are developed basing on empirical studies, e.g. uniformity trials. The present work concerns with the analysis of Factorial experiments and Incomplete block designs, when distributed over the different regions, under Scheffe's model.

Chapter 2

MIXED MODEL FACTORIAL EXPERIMENTS

2.1 - The model:

Suppose it is decided to carry a factorial experiment with three levels of nitrogen and three levels of phosphate, in a given region, to recommend the use of these two fertilizers to the farmers. For this a random sample of r villages is taken from the available villages of the region and further in each village a number of fields is taken for the experiment. For the simplicity of calculations an equal number of k fields is selected. In each field a factorial experiment constituting the above nine treatment combinations is carried out.

The mathematical model is then given by

$$(2.1.1) \quad Y_{ijkl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \vartheta_l + \lambda_{il}^A + \lambda_{jl}^B + \lambda_{ijl}^{AB} + f_{kl} + e_{ijkl}$$

where μ is the average yield, α_i^A is the average response of the i^{th} level of A(Nitrogen), α_j^B is the average response of the j^{th} level of B(Phosphate), α_{ij}^{AB} is the average interaction response of the i^{th} level of A with j^{th} level of B; $\mu, \alpha_i^A, \alpha_j^B, \alpha_{ij}^{AB}$ are fixed effects and $\vartheta_l, \lambda_{il}^A, \lambda_{jl}^B, \lambda_{ijl}^{AB}, f_{kl}$ and e_{ijkl} are random variables; ϑ_l being the response of l^{th} village, λ_{il}^A the interaction between i^{th} level of A with l^{th} village, similar meanings for λ^{B}_{jl} and λ^{AB}_{ijl}; f_{kl}, e_{ijkl} denote the experimental errors - e_{ijkl} arising from plot to plot variation and f_{kl} from field to field variation. In the classical analysis of variance theory ν_l is assumed to be independent of λ's. As has been shown before, the situation in the agricultural experiments over a region demands the use of Scheffe's theory of mixed model analysis which assumes ν_l is correlated with λ^{A}_{il}, λ^{B}_{jl} and λ^{AB}_{ijl}. Again λ^{A}_{il}'s are not all independent so also λ^{B}_{jl} and λ^{AB}_{ijl}.

As ν_l and λ's are not independent we may write

(2.1.2) $Y_{ijkl} = m_{ijl} + f_{kl} + e_{ijkl}$

where m_{ijl} is the true response per plot of the $(ij)^{th}$ treatment in l^{th} village and f_{kl} and e_{ijkl} are experimental errors, which are independent of m_{ijl} and are independently and identically distributed with mean zero and respective constant variances σ_f^2 and σ_e^2.

For each (ijl), m_{ijl} are random variables and since the villages are randomly selected in the region, m_{ijl}'s may be considered as random variables depending upon (ij). With different l, m_{ijl} may be considered as independent random variables but for different (ij), m_{ijl} need not be independent. Thus it shows that the response of treatment depends upon the village l. This is because of the fact that the fertilizer applied acts or reacts according to the varying amount and nature of existing nutrients in the soil of the villages.

Moreover the effect is not often additive and so we consider a more general assumption in which m_{ijl} follow a multivariate normal distn. for each pair (ij). Although a multivariate normal distn. presupposes the population of villages to be infinite, the assumption which is not true in the given case may be accepted to have an approximate description of the nature of variation.

2.2 - Definition of effects and variance components:

Labeling the village in the population by an index U with the population distn. P_u, we shall denote the true response of the i^{th} level of factor A with j^{th} level of factor B in the village labeled U by m(ij.u)

$$\text{where } m(ij.u) = \frac{1}{K} \sum_{k=1}^{K} m(ijku);$$

that is, when an index is replaced by a dot, it means averaging with respect to that index.

Define the true mean for (ij) treatment to be

$$(2.2.1)\quad \mu_{ij} = m(ij..) = \underset{u}{E} \left\{ m(ij.u) \right\} .$$

The general mean is defined as the arithmetic mean of μ_{ij} over all the treatments i.e.

$$(2.2.2)\quad \mu = \mu .. = m(....)$$

$$= \underset{u}{E} \left\{ \frac{m(00.u) + ... + m(22.u)}{9} \right\}$$

where (00), (01), (02).... (22) constitute all the possible treatment combinations with three levels of A and three levels of B.

Now the true mean of i^{th} level of A can be defined as (2.2.3) $\mu_{i.} = m(i...) = \underset{u}{E}\left[\frac{1}{3K}\sum_{jk} m(ijku)\right]$

similarly, (2.2.4) $\mu_{.j} = m(.j..) = \underset{u}{E}\left[\frac{1}{3K}\sum_{ik} m(ijku)\right]$

The amount by which the general mean is exceeded by the true mean of the i^{th} level of A is called the main effect of the i^{th} level of the factor A.

(2.2.5) that is, $\alpha_i^A = \mu_{i.} - \mu_{..} = m(i...) - m(....)$

(2.2.6) similarly $\alpha_j^B = \mu_{.j} - \mu_{..} = m(.j..) - m(....)$

(2.2.7) $\alpha_{ij}^{AB} = m(ij..) - m(i...) - m(.j..) + m(....)$

The true mean for the village labeled u is, then

(2.2.8) $\nu(u) = m(...u) - m(....)$ and may be called the main effect of the village labeled u in the population.

Again the interaction of the i^{th} level of A and the village labeled u is defined as

(2.2.9) $\lambda_i^A(u) = m(i..u) - m(i...) - m(...u) + m(....)$

Similarly (2.2.10) $\lambda_j^B(u) = m(.j.u) - m(.j..) - m(...u) + m(....)$

(2.2.11) $\lambda_{ij}^{AB}(u) = m(ij.u) - m(ij..) - m(i..u) - m(.j.u) +$

$$m(i...) + m(.j..) + m(...u) - m(....)$$

From (2.2.5), (2.2.6), (2.2.7),(2.2.8), (2.2.9), (2.2.10), (2.2.11), it is easily seen that

$$\sum_i \alpha_i^A = 0 \qquad\qquad \sum_j \alpha_j^B = 0$$

$$\sum_i \alpha_{ij}^{AB} = 0 \qquad\qquad \sum_j \alpha_{ij}^{AB} = 0$$

$\sum_i \lambda_i(u) = 0$ for all u

$\sum_j \lambda_j(u) = 0$ for all u

$\sum_i \lambda_{ij}(u) = 0$ for all u

$\sum_j \lambda_{ij}(u) = 0$ for all u

$E(\lambda_i^A(u)) = 0$ for all i

$E(\lambda_j^B(u)) = 0$ for all j

$E(\lambda_{ij}^{AB}(u)) = 0$ for all (ij)

From the above definitions of α_i^A, α_j^B, α_{ij}^{AB}, ν_l [or $\nu(u)$]

λ_{il}^A, λ_{jl}^B and λ_{ijl}^{AB}, we may write

$$Y_{ijkl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \nu_l + \lambda_{il}^A + \lambda_{jl}^B + \lambda_{ijl}^{AB} + f_{kl} + e_{ijkl}$$

where $i = 0,1,2;$ $\quad k = 1 \ldots K$

$j = 0,1,2;$ $\quad l = 1 \ldots r$

Now the random effects $\{\nu(u),\ \lambda_0^A(u),\ \lambda_1^A(u), \lambda_2^A(u)\}$ are not independent and their variances and covariances are the functions of the covariance matrix of the random variable m(i..u). Similarly the covariance matrices of the random variables m(.j.u) and m(ij.u) are associated with the random effect sets $\{\nu(u),\ \lambda_0^B(u),\ \lambda_1^B(u),\ \lambda_2^B(u)\}$ and $\{\nu(u),\ \lambda_{00}^{AB}(u), \ldots\ \lambda_{22}^{AB}(u)\}$. Define

(2.2.13) $\sigma_{ii'}^A = \text{cov}\{m(i..u)\ m(i'..u)\}$

(2.2.14) $\sigma^{B}_{jj'} = \text{Cov}\left\{m(.j.u)\ \ m(.j'.u)\right\}$

(2.2.15) $\sigma^{AB}_{iji'j'} = \text{Cov}\left\{m(ij.u)\ \ m(i'j'.u)\right\}$ and

$$\sigma^{AB}_{iji'j'} = \sigma^{AB}_{ij} \quad \text{if} \quad \begin{matrix} i = i' \\ j = j' \end{matrix}$$

Therefore we calculate $\vartheta(u) = \frac{1}{9K} \sum_{k} \sum_{ij} m(ijku)$

(2.2.16) $\text{Var}\ \vartheta(u) = \frac{1}{81K^2}\ E\left\{\sum_{k} \sum_{ij} m(ijku)\right\}^2$

$$= \sigma_{..} \quad \text{say}$$

Now $\lambda^{A}_{i}(u) = m(i..u) - m(...u) - m(i...) + m(....)$

$$= m(i..u) - m(...u) - \mu_{i.} - \mu_{..}$$

Since Cov $(\lambda^{A}_{i}(u),\ \lambda^{A}_{i'}(u)$ will not depend on the $\mu_{i.}$, we may assume in its calculation $\mu_{i} = 0$

Then Cov $\left\{\lambda^{A}_{i}(u),\ \ \lambda^{A}_{i'}(u)\right\}$

$$= E\left[\left\{m(i..u) - m(...u)\right\}\left\{m(i'..u) - m(...u)\right\}\right]$$

$$= \sigma^{A}_{ii'} - E\left\{m(i..u)\tfrac{1}{3}\sum_{i''} m(i''..u)\right\} - E\left\{m(i'..u)\tfrac{1}{3}\sum_{i''} m(i''..u)\right\} +$$

$$\sigma_{..} = \sigma^{A}_{ii'} - \tfrac{1}{3}\sum_{i''=0,1,2} \sigma^{A}_{ii''} - \tfrac{1}{3}\sum_{i''=0,1,2} \sigma^{A}_{i'i''} + \sigma_{..} = \hat{\sigma}_{ii'} - \hat{\sigma}_{i.} - \hat{\sigma}_{i'.} + \sigma_{..}$$

Similarly Cov $\left\{\lambda^{B}_{j}(u),\ \ \lambda^{B}_{j'}(u)\right\} = \sigma^{B}_{jj'} - \sigma^{B}_{j.} - \sigma^{B}_{j'.} + \sigma_{..}$

And Cov $\left\{\lambda^{AB}_{ij}(u),\ \lambda^{AB}_{i'j'}(u)\right\} = E\left[\left\{m(ij.u) - m(i..u)\right.\right.$

$\left.- m(.j.u) + m(...u)\right\} \quad x \left\{m(i'j'.u) - m(i'..u) - m(.j'.u)\right.$

$\left.\left. + m(...u)\right\}\right]$, since other terms do not contribute anything to the calculation of covariance.

$$= \sigma^{AB}_{iji'j} - \sigma^{AB}_{i.i'j'} - \sigma^{AB}_{.ji'j'} - \sigma^{AB}_{iji'.} - \sigma^{AB}_{ij.j'} + \sigma^{AB}_{ij..}$$
$$+ \sigma^{AB}_{..i'j'} + \sigma^{AB}_{i.i'.} + \sigma^{AB}_{.j.j'} + \sigma^{AB}_{i..j'} + \sigma^{AB}_{.ji'.} - \sigma^{AB}_{i...}$$
$$- \sigma^{AB}_{..i'.} - \sigma^{AB}_{.j..} - \sigma^{AB}_{...j'} + \sigma_{..}$$

Because of the symmetric property of the covariance matrices

$$\sigma^A_{ii'} = \sigma^A_{i'i} \qquad \sigma^A_{i.} = \sigma^A_{.i'}$$

$$\sigma^B_{jj'} = \sigma^B_{j'j} \qquad \sigma^{AB}_{iji'j'} = \sigma^{A\,B}_{i'j'ij}$$

$$\sigma^B_{j.} = \sigma^B_{.j'} \quad \text{and} \quad \sigma^{AB}_{ij..} = \sigma^{AB}_{..i'j'} \text{ etc.}$$

In the similar way we can have

$$\mathrm{Cov}\left\{\nu(u)\ \lambda^A_i(u)\right\} = E\left[m(\ldots u)\ \left\{m(i..u) - m(\ldots u)\right\}\right]$$

$$= E\left\{m(i..u)\ m(i..u)\right\} - E\left\{m(\ldots u)\right\}^2$$

$$= \sigma^A_{i.} - \sigma_{..}$$

$$\mathrm{Cov}\left\{\nu(u)\ \lambda^B_j(u)\right\} = \sigma^B_{j.} - \sigma_{..} \quad \text{and}$$

$$\mathrm{Cov}\left\{\nu(u)\ {}^{AB}_{ij}(u)\right\} = \sigma^{AB}_{ij..} - \sigma^A_{i.} - \sigma^B_{j.} + \sigma_{..}$$

Now define

(2.2.17)
$$\sigma^2_A = \frac{1}{2}\sum_i (\alpha^A_i)^2$$
$$\sigma^2_B = \frac{1}{2}\sum_j (\alpha^B_j)^2$$
$$\sigma^2_{AB} = \frac{1}{4}\sum_{ij} (\alpha^{AB}_{ij})^2$$

and

(2.2.18)
$$\sigma^2_L = \mathrm{Var}\ \nu(u)$$
$$\sigma^2_{AL} = \frac{1}{2}\sum_i \mathrm{Var}(\lambda^A_{il}) = \frac{1}{2}\sum_i V(\lambda^A_i(u))$$
$$= \frac{1}{2}\sum \sigma^2_{AL,i}$$
$$\sigma^2_{BL} = \frac{1}{2}\sum_j \mathrm{Var}(\lambda^B_{jl}) = \frac{1}{2}\sum_j V(\lambda^B_j(u))$$
$$= \frac{1}{2}\sum \sigma^2_{BL,j}$$
$$\sigma^2_{ABL} = \frac{1}{4}\sum_i\sum_j \sigma^2_{ABL,ij}$$

The quantities σ_L^2, σ_{AL}^2, σ_{BL}^2, σ_{ABL}^2 may be expressed in terms of the elements of covariance matrices as

$$\sigma_L^2 = \sigma_{..}$$

$$\sigma_{AL}^2 = \frac{1}{2}\sum_i (\sigma^A_{ii} - \sigma_{..})$$

$$\sigma_{BL}^2 = \frac{1}{2}\sum_j (\sigma^B_{jj} - \sigma_{..})$$

$$\sigma_{ABL}^2 = \frac{1}{4}\sum_{ij}[\sigma^{AB}_{ij} - 2\sigma^{AB}_{i.} - 2\sigma^{AB}_{.j} + \sigma^A_{ii'} + \sigma^B_{jj'} + \sigma_{..}]$$

It is clear to see that $\sigma_L^2 = 0$, if and only if $\vartheta(u) = 0$ for all u; that is if the basic vector $m(u) = \{m(00.u), \ldots m(22.u)\}$ has a degenerate distn, satisfying $m(00.u) + \ldots + m(22.u) = \text{constant} \doteq 9\mu$. Also $\sigma_{AL}^2 = 0$ if and only if var $\{\lambda^A_i(u)\} = 0$ for all i or $m(i..u) = m(...u) + \alpha^A_i$, that is except for additive constants $\{\alpha^A_i\}$, the random variables $m(i..u)$ are identical (not just identically distributed). Similar conditions hold good for σ_{BL}^2 and σ_{ABL}^2.

2.3 - Calculation of sum of squares:

According to usual Least square method of finding sum of squares

$$S.S.(A) = \sum_l \sum_k \sum_j \sum_i (Y_{i...} - Y_{....})^2$$

$$= 3rK \sum_i (Y_{i...} - Y_{....})^2$$

$$S.S.(B) = \sum_l \sum_k \sum_j \sum_i (Y_{.j..} - Y_{....})^2$$

$$= 3rK \sum_j (Y_{.j..} - Y_{....})^2$$

$$S.S.(AB) = \sum_{l}\sum_{k}\sum_{j}\sum_{i} (Y_{ij..} - Y_{i...} - Y_{.j..} + Y_{....})^2$$

$$S.S.(VA) = \sum_{l}\sum_{k}\sum_{j}\sum_{i} (Y_{i..l} - Y_{i...} - Y_{...l} + Y_{....})^2$$

$$S.S.(VB) = \sum_{l}\sum_{k}\sum_{j}\sum_{i} (Y_{.j.l} - Y_{.j..} - Y_{...l} + Y_{....})^2$$

Moreover S.S. due to villages = S.S.(L) = $\sum_{l}\sum_{k}\sum_{j}\sum_{i}(Y_{...l} - Y_{....})^2$

and S.S. due to fields within villages

$$= \sum_{l} \left\{ \sum_{k} \sum_{j} \sum_{i} (Y_{..kl} - Y_{...l})^2 \right\}$$

Total S.S. = $\sum_{l}\sum_{k}\sum_{j}\sum_{i} (Y_{ijkl} - Y_{....})^2$ ss(VAB) ?

Error sum of squares can be obtained by subtracting all the component of S.S. from Total S.S.

Partition of Degrees of freedom

Sources	d.f.
Villages (V)	r - 1
Fields within villages	r(K - 1)
A } Treatments	2
B } Treatments	2
AB } Treatments	4
V x A	2(r-1)
V x B	2(r-1)
V x AB	4(r-1)
Error	8r(K-1)
Total:	9rK - 1

Since the interaction AB contains 4 d.f., we can break up this into two orthogonal parts - one containing absolutely linear component and the other corresponding to quadratic components.

That is,	AB	A_1B_1	1 d.f.
		A_1B_q A_qB_1 A_qB_q	3 d.f.

where A_1B_q stands for the interaction of linear component of A with the quadratic component of B. Similar meanings can be attached to A_1B_1, A_qB_1 and A_qB_q.

Similarly VAB	VA_1B_1	(r-1) d.f.
	VA_1B_q VA_qB_1 VA_qB_q	3(r-1) d.f.

where VA_1B_q stands for the interaction of linear x quadratic component of AB with village and etc.

$$\text{Now S.S.}(A_1B_1) = \frac{1}{4rK}\left[\sum_{l}\sum_{k}(Y_{22kl} - Y_{20kl} - Y_{02kl} + Y_{00kl})\right]^2$$

$$\text{S.S.}(A_1B_q, A_qB_1, A_qB_q) = \sum_{l}\sum_{k}\sum_{j}\sum_{i}(Y_{ij..} - Y_{i...} - Y_{.j..} + Y_{....})^2$$

$$- \frac{1}{4rK}\left[\sum_{l}\sum_{k}(Y_{22kl} - Y_{20kl} - Y_{02kl} + Y_{00kl})\right]^2$$

$$\text{S.S.}(VA_1B_1) = \frac{1}{4K}\sum_{l=1}^{r}(Y_{22.l} - Y_{20.l} - Y_{02.l} + Y_{00.l} - Y_{22..}$$

$$+ Y_{20..} + Y_{02..} - Y_{00..})^2$$

Should K be in the numerator?

and $S.S.(VA_1B_q, VA_qB_1, VA_qB_q)$

$$= S.S.(VAB) - \frac{1}{4K} \sum_{\ell=1}^{r} (Y_{22.1} - Y_{20.1} - Y_{02.1} + Y_{00.1} - Y_{22..} + Y_{20..} + Y_{02..} - Y_{00..})^2$$

In agricultural experiments over a small region, the interaction of quadratic components of AB with village may be assumed to be small. Therefore appropriate test procedure should be worked out to test this interaction and if it is found not significant, it can be pooled with error.

2.4 - Expectation of Mean Squares:

$$E\{S.S.(A)\} = 3rK \; E\Big[\Sigma(\alpha_i^A - \alpha_.^A)^2 + \Sigma(\alpha_{i.}^{AB} - \alpha_{..}^{AB})^2 + \Sigma(\lambda_i^A - \lambda_.^A)^2 + \Sigma(\lambda_{i..}^{AB} - \lambda_{...}^{AB})^2 + \Sigma(e_{i...} - e_{....})^2\Big]$$

$$= 3rK \sum_i (\alpha_i^A)^2 + 3K \sum_i \sigma^2_{AL,i} + 2\sigma_e^2$$

Therefore $E(M.S.A) = 3rK\,\sigma_A^2 + 3K\,\sigma_{AL}^2 + \sigma_e^2$

Similarly $E(M.S.B) = 3rK\,\sigma_B^2 + 3K\,\sigma_{BL}^2 + \sigma_e^2$

$$\text{Again } E(S.S.AB) = E\,\Sigma\Sigma\Sigma\Sigma(\alpha_{ij}^{AB} - \alpha_{i.}^{AB} - \alpha_{.j}^{AB} + \alpha_{..}^{AB})^2 + E\,\Sigma\Sigma\Sigma\Sigma(\lambda_{ij.}^{AB} - \lambda_{i..}^{AB} - \lambda_{.j.}^{AB} + \lambda_{...}^{AB})^2 + E\,\Sigma\Sigma\Sigma\Sigma(e_{ij..} - e_{i...} - e_{.j..} + e_{....})^2$$

$$= rK\,\Sigma\Sigma(\alpha_{ij}^{AB})^2 + rKE\Sigma\Sigma(\lambda_{ij.}^{AB})^2 + 4\sigma_e^2$$

$$= rK\,\Sigma\Sigma(\alpha_{ij}^{AB})^2 + K\,\Sigma\Sigma\,\sigma^2_{ABL,ij} + 4\sigma_e^2.$$

$\therefore E(M.S.AB) = rK\,\sigma_{AB}^2 + K\,\sigma_{ABL}^2 + \sigma_e^2$

To find $E(S.S.Q) = E[S.S.(A_1B_q, A_qB_1, A_qB_q)]$, we have

Notation A? B?

$$4\ S.S.(A_1B_1) = rK(\alpha_{22} - \alpha_{20} - \alpha_{02} + \alpha_{00})^2$$

$$+ rK(\lambda_{22.} - \lambda_{20.} - \lambda_{02.} + \lambda_{00.})^2$$

? Cross-product terms?

$$+ rK(e_{22..} - e_{20..} - e_{02..} + e_{00..})^2$$

Therefore $E(S.S.Q) = rK\, E\sum\sum(\alpha^{AB}_{ij})^2 + rK\, E\sum\sum(\lambda^{AB}_{ij.})^2 + 4\sigma_e^2$

Notation, A'B ?

$$- \frac{rK}{4}E(\alpha_{22} - \alpha_{20} - \alpha_{02} + \alpha_{00})^2$$

?

$$- \frac{rK}{4}E(\lambda_{22.} - \lambda_{20.} - \lambda_{02.} + \lambda_{00.})^2$$

$$- \frac{rK}{4}E(e_{22..} - e_{20..} - e_{02..} + e_{00..})^2$$

$$= rK(\alpha_{01}^2 + \alpha_{10}^2 + \alpha_{11}^2 + \alpha_{12}^2 + \alpha_{21}^2)$$

$$+ \frac{3rK}{4}(\alpha_{22}^2 + \alpha_{20}^2 + \alpha_{02}^2 + \alpha_{00}^2) + \frac{rK}{2}(\alpha_{22}\alpha_{20} + \alpha_{22}\alpha_{02} + \alpha_{20}\alpha_{00} + \alpha_{02}\alpha_{00} - \alpha_{22}\alpha_{00} - \alpha_{20}\alpha_{02})$$

$$+ K(\sigma_{01L}^2 + \sigma_{10L}^2 + \sigma_{11L}^2 + \sigma_{12L}^2 + \sigma_{21L}^2)$$

Notation clumsy? GARLY

$$+ \frac{3K}{4}(\sigma_{22L}^2 + \sigma_{20L}^2 + \sigma_{02L}^2 + \sigma_{00L}^2)$$

$$+ \frac{K}{2}(\sigma_{22L,20L} + \sigma_{22L,02L} + \sigma_{20L,00L} + \sigma_{02L,00L}$$

$$- \sigma_{22L,00L} - \sigma_{20L,02L}) + 3\sigma_e^2$$

where $\sigma_{ijL}^2 = V(\lambda_{ijl})$

and $\sigma_{ijL,i'j'L} = \mathrm{Cov}(\lambda_{ijl};\ \lambda_{i'j'l})$

$$\therefore\ E\ (M.S.Q.) = \frac{rK}{3}(\alpha_{01}^2 + \alpha_{10}^2 + \alpha_{11}^2 + \alpha_{12}^2 + \alpha_{21}^2)$$

$$+ \frac{rK}{4}(\alpha_{22}^2 + \alpha_{20}^2 + \alpha_{02}^2 + \alpha_{00}^2) + \frac{rK}{6}(\alpha_{22}\alpha_{20} + \alpha_{22}\alpha_{02} + \alpha_{20}\alpha_{00} + \alpha_{02}\alpha_{00} - \alpha_{22}\alpha_{00} - \alpha_{20}\alpha_{02})$$

$$+ \frac{K}{3}(\sigma_{01L}^2 + \sigma_{10L}^2 + \sigma_{11L}^2 + \sigma_{12L}^2 + \sigma_{21L}^2)$$

$$+ \frac{K}{4}(\sigma_{22L}^2 + \sigma_{20L}^2 + \sigma_{02L}^2 + \sigma_{00L}^2)$$

$$+ \frac{K}{6}(\sigma_{22L,20L} + \sigma_{22L,02L} + \sigma_{20L,00L} + \sigma_{02L,00L}$$

$$- \sigma_{22L,00L} - \sigma_{20L,02L}) + \sigma_e^2$$

Therefore Now $E(M.S.VA) = \frac{3K}{2} \sum_i \sigma_{AL,i}^2 + \sigma_e^2$

$$= 3K \sigma_{AL}^2 + \sigma_e^2$$

$$E(M.S.VB) = 3K \sigma_{BL}^2 + \sigma_e^2$$

and $E(M.S.VAB) = \frac{K}{4} \sum_i \sum_j \sigma_{ABL,ij}^2 + \sigma_e^2 = K \sigma_{ABL}^2 + \sigma_e^2$

To find $E(M.S.VQ)$, we first calculate

$$E(S.S.VA_1B_1) = \frac{K}{4} E \sum_l (\lambda_{22l} - \lambda_{02l} - \lambda_{20l} + \lambda_{00l}$$

$$- \lambda_{22.} + \lambda_{20.} + \lambda_{02.} - \lambda_{00.})^2 + (r-1)\sigma_e^2$$

Then $E(S.S.VQ) = (r-1)K \sum_i \sum_j \sigma_{ABL,ij}^2 + 4(r-1)\sigma_e^2$

$$- \frac{K}{4}(r-1)\left\{\sigma_{22L}^2 + \sigma_{20L}^2 + \sigma_{02L}^2 + \sigma_{00L}^2\right\}$$

$$+ \frac{K}{2}(r-1)\left\{\sigma_{22L,20L} + \sigma_{22L,02L} + \sigma_{20L,00L}\right.$$

$$\left. + \sigma_{02L,00L} - \sigma_{22L,00L} - \sigma_{20L,02L}\right\} - (r-1)\sigma_e^2$$

$$\therefore E(M.S.VQ) = \frac{K}{3}(\sigma_{01L}^2 + \sigma_{10L}^2 + \sigma_{11L}^2 + \sigma_{12L}^2 + \sigma_{21L}^2)$$

$$+ \frac{K}{4}(\sigma_{22L}^2 + \sigma_{20L}^2 + \sigma_{02L}^2 + \sigma_{00L}^2)$$

$$+ \frac{K}{6}(\sigma_{22L,20L} + \sigma_{22L,02L} + \sigma_{20L,00L} + \sigma_{02L,00L}$$

$$- \sigma_{22L,00L} - \sigma_{20L,02L}) + \sigma_e^2$$

Lastly it can be easily proved

$$E(\text{Error M.S.}) = \sigma_e^2$$

2.5 - Tests of Significance:

In the present case the natural hypotheses to be tested are:

$$HA : \sigma_A^2 = 0 \qquad H_{VA} : \sigma_{AL}^2 = 0$$

$$HB : \sigma_B^2 = 0 \qquad HVB : \sigma_{BL}^2 = 0$$

$$HQ : \sigma_Q^2 = 0 \qquad HVQ : \sigma_{VQ}^2 = 0$$

Though MSA and MSVA are statistically independent and under the hypothesis HA: $\sigma_A^2 = 0$ i.e. all $\alpha_i^A = 0$ have the same expected values, their quotient does not in general have the 'F' distn. under HA. This is due to the fact that neither numerator nor denominator can be distributed as constant times the noncentral or central χ^2 variables. An exact test of this hypothesis can be obtained with the help of Hotelling's T^2 statistic (Scheffe, 1956). The same consideration can be applied to test HB : $\sigma_B^2 = 0$

The hypothesis HVA : $\sigma_{VA}^2 = 0$ and HVB : $\sigma_{VB}^2 = 0$ may be tested respectively with the statistic $\frac{(MS)VA}{(MS)e}$ and

Notation σ_{AL} ?
σ_{VA}

Analysis of Variance Table

Source	d.f.	S.S.	E(M.S.)
Village	$r - 1$	$\sum_l \sum_k \sum_j \sum_i (Y_{...l} - Y_{....})^2$	
Fields within villages	$r(K-1)$	$q \sum_l \sum_k (Y_{..kl} - Y_{...l})^2$	
Treatments:			
A	2	$\sum_l \sum_k \sum_j \sum_i (Y_{i...} - Y_{....})^2$	$3rK\,\sigma_A^2 + 3K\,\sigma_{AL}^2 + \sigma_e^2$
B	2	$\sum_l \sum_k \sum_j \sum_i (Y_{.j..} - Y_{....})^2$	$3rK\,\sigma_B^2 + 3K\,\sigma_{BL}^2 + \sigma_e^2$
AB A_1B_1	1 } 4	$\frac{1}{4}rK[\sum_l \sum_k (Y_{22kl} - Y_{20kl} - Y_{02kl} + Y_{00kl})]^2$	
A_1B_q, A_qB_1, A_qB_q } Q	3	$\sum\sum\sum\sum (Y_{ij..} - Y_{i...} - Y_{.j..} + Y_{....})^2 - \frac{1}{4}rK[\sum\sum(Y_{22kl} - Y_{20kl} - Y_{02kl} + Y_{00kl})]^2$	$\frac{rK}{3}(\alpha_{01}^2 + \alpha_{10}^2 + .. + \alpha_{21}^2) + \frac{rK}{4}(\alpha_{22}^2 + .. + \alpha_{00}^2) + \frac{rK}{6}(\alpha_{22}\alpha_{20} + \alpha_{22}\alpha_{02} + \alpha_{20}\alpha_{00} + \alpha_{02}\alpha_{00} - \alpha_{22}\alpha_{00} - \alpha_{20}\alpha_{02}) + \frac{K}{3}(\sigma_{01L}^2 + .. + \sigma_{21L}^2) + \frac{K}{4}(\sigma_{22L}^2 + .. + \sigma_{00L}^2) + \frac{K}{6}(\sigma_{22L,20L} + .. + \sigma_{02L,00L} - \sigma_{22L,00L} - \sigma_{20L,02L}) + \sigma_e^2$ $= \sigma_Q^2 + \sigma_{VQ}^2 + \sigma_e^2$
V x A	$2(r - 1)$	$\sum_l \sum_k \sum_j \sum_i (Y_{i..l} - Y_{i...} - Y_{...l} + Y_{....})^2$	$3K\,\sigma_{AL}^2 + \sigma_e^2$

(Table Continued ......)

Analysis of Variance Table (Contd....)

V x B	2(r - 1)	$\sum\sum\sum\sum (Y_{.j.l} - Y_{.j..} - Y_{...l} + Y_{....})^2$	$3K\,\sigma_{BL}^2 + \sigma_e^2$
VAB: VA_1B_1	(r - 1)	$\frac{1}{4K}\sum_{l=1}^{r} (Y_{22.l} - Y_{20.l} - Y_{02.l} + Y_{00.l} - Y_{22..} + Y_{20..} + Y_{02..} - Y_{00..})^2$	
VAB: VA_1B_q, VA_qB_1, VA_qB_q (VQ)	3(r - 1)	$\sum_i\sum_k\sum_j\sum_i (Y_{ij.l} - Y_{i..l} - Y_{.j.l} - Y_{ij..} + Y_{i...} + Y_{.j..} + Y_{...l} - Y_{....})^2$ - S.S. VA_1B_1	$\frac{K}{3}(\sigma_{01L}^2 + \sigma_{10L}^2 + \sigma_{11L}^2 + \sigma_{12L}^2 + \sigma_{21L}^2) + \frac{K}{4}(\sigma_{22L}^2 + .. + \sigma_{00L}^2) + \frac{K}{6}(\sigma_{22L,20L} + .. + \sigma_{20L,00L} - \sigma_{22L,00L} - \sigma_{20L,02L}) + \sigma_e^2 = \sigma_{VQ}^2 + \sigma_e^2$
Error	8r(K - 1)	By subtraction	σ_e^2
Total	9rK - 1	$\sum\sum\sum\sum (Y_{ijkl} - Y_{....})^2$	

$\frac{(M.S.)_{VB}}{(M.S.)_e}$, which under the hypothesis have the 'F' distn. with $2(r - 1)$ and $8r(K - 1)$ d.f. Since each of the above statistics is distributed as the quotient of a linear combination of independent χ^2 variables by another independent χ^2 variable, the power of the test is not expressible in terms of the non-central 'F' distn. but it can be approximated by central 'F' distn.

To test H_{VQ} : $\sigma_{VQ}^2 = 0$, the quantity

$$E(M.S.)_{VQ} - E(M.S.)_e = \frac{K}{3}(\sigma_{01L}^2 + \sigma_{10L}^2 + \sigma_{11L}^2 + \sigma_{12L}^2 + \sigma_{21L}^2)$$
$$+ \frac{K}{4}(\sigma_{22L}^2 + \sigma_{20L}^2 + \sigma_{02L}^2 + \sigma_{00L}^2) + \frac{K}{6}(\sigma_{22L,02L} + \sigma_{22L,20L}$$
$$+ \sigma_{20L,00L} + \sigma_{02L,00L} - \sigma_{22L,00L} - \sigma_{20L,02L})$$
$$\geq \frac{K}{3}(\sigma_{01L}^2 + \,..\, + \sigma_{21L}^2) + \frac{K}{4}(\sigma_{22L}^2 + \,..\, + \sigma_{00L}^2)$$
$$+ \frac{K}{6}(-\sigma_{22L,02L} - \sigma_{22L,20L} - \sigma_{20L,00L} - \sigma_{02L,00L} - \sigma_{22L,00L}$$
$$- \sigma_{20L,02L}) = \frac{K}{3}(\sigma_{01L}^2 + \sigma_{10L}^2 + \,..\, + \sigma_{21L}^2) + \frac{K}{12}(\sigma_{22L} - \sigma_{02L})^2$$
$$+ (\sigma_{22L} - \sigma_{20L})^2 + (\sigma_{20L} - \sigma_{00L})^2 + (\sigma_{02L} - \sigma_{00L})^2$$
$$+ (\sigma_{22L} - \sigma_{00L})^2 + (\sigma_{20L} - \sigma_{02L})^2$$

is not less than zero, it being zero only when $\sigma_{00L}^2, \ldots \sigma_{22L}^2$ are zero i.e. when $\sigma_{VQ}^2 = 0$.

Thus under H_{VQ} : $\sigma_{VQ}^2 = 0$, $\frac{(M.S.)_{VQ}}{(M.S.)_e}$ follows 'F' distn. with $3(r - 1)$, $8r(K - 1)$ d.f.

However the power of the test is not expressible in terms of the central or noncentral 'F' distn., since under alternative hypothesis $\sigma_{VQ}'^2 \neq 0$, $(M.S.)_{VQ}$ is not distributed as a constant component of χ^2 variable. Since $(M.S.)_{VQ}$ is distributed as a linear function of independent χ^2 variables and $(M.S.)_e$ is distributed as a constant component of χ^2 variable, the power may be calculated by using Box's (1954) result.

To test H_Q : $\sigma_Q^2 = 0$:-

When $\sigma_{VQ}^2 = 0$, that is when the interaction mean square $(M.S.)_{VQ}$ is not significant, we can pool it with error mean square and test $(M.S.)_Q$ against this pooled m.s. Under alternative hypothesis $\sigma_Q'^2 \neq 0$, $\frac{(M.S.)_Q}{(M.S.)_{pooled}}$ follows noncentral 'F' distribution and therefore power can be calculated easily.

Chapter 3

MIXED MODEL EXPERIMENTS WITH INCOMPLETE BLOCKS

The mixed model analysis could be extended to incomplete block experiments also. We shall in particular consider the problem of testing for the interaction of treatment x places when a number of similar incomplete block experiments are considered in different randomly chosen places.

Let there be J fields in each of the r villages and in j^{th} field of each village i^{th} treatment is applied n_{ij} times. Let the total number of treatments to be tested be I. Therefore we have n_{ij} cell frequencies in $(ij)^{th}$ cell and this design is replicated in r villages.

The mixed model assumed is

$$Y_{ijk} = \mu + \alpha_i + f_{jk} + \nu_k + \lambda_{ik} + e_{ijk}$$

$$i = 1 \; \ldots\ldots \; I$$

$$j = 1 \; \ldots\ldots \; J$$

$$K = 1 \; \ldots\ldots, \; r$$

Where μ and α_i's are fixed effects corresponding to general mean and treatments; ν_k and λ_{ik} are random effects corresponding to villages and interaction of village with treatment; f_{jk} and e_{ijk} are error components - f_{jk} representing the variation from field to field and e_{ijk} representing experimental errors.

The assumptions regarding the components of the model are alike to those assumed in the previous chapter.

Partition of Degrees of freedom

Source	d.f.
Villages	$r - 1$
Fields within villages	$r(J - 1)$
Treatments	$(I - 1)$
Treatments x villages	$(r - 1)(I - 1)$
Error (experimental)	$rn_{..} - r(I+J-1)$
Total	$rn_{..} - 1$

where $n_{..} = \sum_i \sum_j n_{ij}$

Treatment S.S. is obtained from all the r villages and is given by $\sum_{i=1}^{I} \hat{t}_i Q_i$

where $Q_i = \sum_{k=1}^{r} Q_{ik}$ and $Q_{ik} = T_{ik} - \sum_j \frac{n_{ij}}{n_{.j}} B_{jk}$

with T_{ik} = The total yield of i^{th} treatment from k^{th} village.

B_{jk} = The total yield of j^{th} field in k^{th} village

n_{ij} = The number of observations in $(ij)^{th}$ cell

$n_{.j} = \sum_i n_{ij}$, $n_{i.} = \sum_j n_{ij}$ and $\hat{t}_i$'s are the solutions of the equations

$$a_{11}\hat{t}_1 + a_{12}\hat{t}_2 + .. + a_{1I}\hat{t}_I = \frac{Q_1}{r}$$

$$a_{21}\hat{t}_1 + a_{22}\hat{t}_2 + .. + a_{2I}\hat{t}_I = \frac{Q_2}{r}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{I1}\hat{t}_1 + a_{I2}\hat{t}_2 + .. + a_{II}\hat{t}_I = \frac{Q_I}{r}$$

No punctuation

Where $a_{ii} = n_{i.} - \sum_j \frac{n_{ij}^2}{n_{.j}}$

and $a_{ii'} = - \sum_j \frac{n_{ij}n_{i'j}}{n_{.j}} \qquad (i \neq i')$

Since these I equations are not independent, we omit the last equation and again put $\hat{t}_I = 0$

Therefore we have $\sum_{i'=1}^{I-1} a_{ii'} \; t_{i'} = \frac{Q_i}{r} \quad (i = 1 \; \ldots \; I-1)$

$$\therefore \begin{pmatrix} t_1 \\ \vdots \\ t_{I-1} \end{pmatrix} = \frac{1}{r} \left[a_{ii'}\right]^{-1} \begin{pmatrix} Q_1 \\ \vdots \\ Q_{I-1} \end{pmatrix}$$

$\therefore$ Sum of squares for treatments

$$= \frac{1}{r}(Q_1, \ldots \; Q_{I-1}) \left[a_{ii'}\right]^{-1} \begin{pmatrix} Q_1 \\ \vdots \\ Q_{I-1} \end{pmatrix}$$

$$= \frac{1}{r} Q'CQ = \frac{1}{r} \sum_{i,i'=1}^{I-1} C_{ii'} Q_i Q_{i'}$$

The sum of squares for treatment x village interaction is calculated as

$$\sum_{k=1}^{r} \sum_{i,i'=1}^{I-1} C_{ii'} \; Q_{ik} Q_{i'k} - \frac{1}{r} \sum_{i,i'=1}^{I-1} C_{ii'} \; Q_i Q_{i'}$$

Now E (Treatment S.S.) = $r \sum_{i}^{I-1} \sum_{i'}^{I-1} a_{ii'} \; \alpha_i \alpha_{i'}$

$$+ \sum_{i=1}^{I-1} \sum_{i'=1}^{I-1} a_{ii'} \mathrm{Cov} \left\{ \lambda_i(u), \lambda_{i'}(u) \right\} + (I-1) \sigma_e^2$$

and E(treatment x village S.S.)

$$= (r-1) \sum_{i=1}^{I-1} \sum_{i'=1}^{I-1} a_{ii'} \mathrm{Cov} \left\{ \lambda_i(u), \; \lambda_{i'}(u) \right\} + (r-1)(I-1)\sigma_e^2$$

To prove the unbiasedness of the test of interaction m.s. by the error m.s., it will be sufficient to prove that

$$\text{Cov}\left\{\lambda_i(u)\,\lambda_{i'}(u)\right\} \text{ is always positive.}$$

Now $$\sum_{i=1}^{I-1}\sum_{i'=1}^{I-1} a_{ii'}\ \text{Cov}\left\{\lambda_i(u)\,\lambda_{i'}(u)\right\} = E\ \sum_{i=1}^{I-1}\sum_{i'=1}^{I-1} a_{ii'}\ \left\{\lambda_i(u)\lambda_{i'}(u)\right\}$$

Since $[a_{ii'}]^{-1}$ is a symmetric positive definite matrix (being variance - covariance matrix), $[a_{ii'}]$ is also positive definite. Therefore, the above expression can be reduced by a suitable transformation to the form $\epsilon_1 \lambda_1^2\ (u) + \ldots + \epsilon_{I-1}\ \lambda_{I-1}^2\ (u)$ where ϵ's are all positive. Therefore $E\left\{\epsilon_1 \lambda_1^2\ (u) + \ldots + \epsilon_{I-1}\ \lambda_{I-1}^2(u)\right\}$

$= \epsilon_1$ Var $\lambda_1(u) + \ldots +$ Var $\lambda_{I-1}(u)$, which is always positive.

Thus if the treatment x village interaction is not present, the usual F-test will have the same level of significance as the nominal one. When the interaction is present, the distribution of the ratio,

$$\frac{\text{S.S. due to interaction of treatment x village}}{\text{S.S. due to error}}$$

will not have a constant times noncentral 'F' distribution but will still yield an unbiased test.

The power of this test under the alternative hypothesis is indicated by the difference of the expected values of treatment x village S.S. and error S.S. Since the numerator and denominator of the test criterion are quadratic forms in normal variables we can find the approximate power of this test by using Satterthwaite's approximation for a linear function of independent χ^2 variables. Some calculations of this type in a different context has been done by Imhof (1961).

Part II

Mixed Model Analysis of Covariance

MIXED MODEL ANALYSIS OF COVARIANCE IN AGRICULTURAL EXPERIMENTS

1. Introduction:

The use of auxiliary information has been made extensively in the model I of Eisenhart (1947). The utility of these analyses in increasing the precision of an experiment has been verified in numerous occasions and one would imagine the same result to follow from the use of auxiliary information in Scheffe's mixed model. There are, however, some difficulties in setting up a model with auxiliary variables in this case unlike in the case of model I. A mixed model will be considered in the following for the analysis of covariance with certain assumptions about the covariance structure of the main and auxiliary variates. In the postulated model the components due to errors has been ? subdivided into two independent parts - unit errors corresponding to difference in fertility between plots within sites and technical errors corresponding to errors of measurements etc. During large scale fertilizer trials distributed over different places, it is usually seen that the yield of the previous year for the whole region (with no use of fertilizer) is correlated with the yield of the given year. Therefore it seems that the efficiency will be gained by the use of yield data of the previous year in the analysis problem. The main aim as mentioned in previous chapters rests in testing the interaction of treatment with region in order to demarcate effectively the soil-climatic zones.

2. Model :

Suppose Y_{ijk} is the response on the k^{th} replicate of i^{th} treatment in j^{th} region. Then we write

$$Y_{ijk} = m_{ij} + e_{ijk} + \epsilon_{ijk}$$

$$i = 1 \;\ldots\ldots\; I$$
$$j = 1 \;\ldots\ldots\; J$$
$$k = 1 \;\ldots\ldots\; K$$

Again suppose X_{ijk} is the response on the k^{th} replicate of i^{th} treatment in j^{th} region for the previous year. For this we assume the model

$$X_{ijk} = X_j + e_{ijk} + \eta_{ijk}$$

where e_{ijk} is the error due to differences in soil fertility between plots within the region and ϵ_{ijk}, η_{ijk} are specific errors of observations etc. X_j is the true mean yield in the j^{th} region.

We further assume $(m_{1j}, m_{2j} \ldots m_{Ij}, X_j)$ follow $(I + 1)$ variate normal distribution and $Cov(m_{1j}\ X_j)$ $= Cov(m_{2j}\ X_j) = \ldots. = Cov(m_{Ij}\ X_j)$. That means X_j is equally correlated with the variates $(m_{1j}, m_{2j} \ldots m_{Ij})$.

This assumption is based on the idea that X_j represents the fertility of the soil and should affect all m_{ij}'s equally.

Now $\{Y_{ijk}\}$, $\{X_{ijk}\}$ will also have a multivariate normal distribution. Let the vector

$$Y' = (Y_{111} \ldots Y_{IJK}) \quad \text{and} \quad X' = (X_{111} \ldots X_{IJK})$$

and let $T = \begin{bmatrix} Y \\ X \end{bmatrix}$, then the distribution of T is given by

$$p(T) = \text{Constant} \; |\Sigma_T|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(T-ET)' \; \Sigma_T^{-1}(T-ET)\right\}$$

where $\Sigma_T = \begin{bmatrix} \Sigma_{YY} & \vdots & \Sigma_{YX} \\ \cdots\cdots & \vdots & \cdots\cdots \\ \Sigma_{XY} & \vdots & \Sigma_{XX} \end{bmatrix}$

Σ_{YY}, Σ_{YX}, Σ_{XX} are partitioned matrices representing the variance-covariance matrices of Y and X. The conditional distribution of Y for given X is

$$N\left\{\mu_Y + \Sigma_{YX} \; \Sigma_{XX}^{-1} (X - u_X), \; \Sigma_{YY.X}\right\}$$

where $\Sigma_{YY.X} = \Sigma_{YY} - \Sigma_{YX} \, \Sigma_{XX}^{-1} \, \Sigma_{XY}$

$$\text{Now Cov}(Y_{ijk} \;\; Y_{i'j'k'}) = \delta_{jj'}(\sigma_{ii'} + \delta_{ii'} \, \delta_{kk'} \;\; \sigma_e^2 + \delta_{ii'} \;\; \delta_{kk'} \;\; \sigma_\epsilon^2)$$

$$\text{Cov}(X_{ijk} \;\; X_{i'j'k'}) = \delta_{jj'}(\sigma_X^2 + \delta_{ii'} \; \delta_{kk'} \; \sigma_e^2 + \delta_{ii'} \; \delta_{kk'} \; \sigma_\eta^2)$$

$$\text{Cov}(Y_{ijk} \;\; X_{i'j'k'}) = \delta_{jj'}(\sigma_{XY} + \delta_{ii'} \; \delta_{kk'} \; \sigma_e^2)$$

Where $\sigma_{ii'} = Cov(m_{ij}\ m_{i'j})$, $E(X) = \mu_X = 0$

$\sigma_X^2 = Var\ (X_j)$

$\sigma_{XY} = Cov(m_{ij}\ X_j)$

Thus $$\Sigma_{YY} = \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & A \end{bmatrix}_{J \times J}$$

Where A is a submatrix of size IK

$$A = ((\Lambda_{ii'}))_{I \times I} + (\sigma_e^2 + \sigma_\epsilon^2)\ U_{IK \times IK}$$

and $$\Lambda_{ii'} = \begin{bmatrix} \sigma_{ii'} & \cdots & \sigma_{ii'} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \sigma_{ii'} & \cdots & \sigma_{ii'} \end{bmatrix}_{K \times K}$$

Again $$\Sigma_{XX} = \begin{bmatrix} B & 0 & \cdots & 0 \\ 0 & B & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & B \end{bmatrix}_{J \times J}$$

Therefore $$\Sigma_{XX}^{-1} = \begin{bmatrix} B^{-1} & 0 & \cdots & 0 \\ 0 & B^{-1} & \cdots & 0 \\ 0 & 0 & \cdots & B^{-1} \end{bmatrix}_{J \times J}$$

Where $$B = \begin{bmatrix} \sigma_X^2 + \sigma_e^2 + \sigma_\eta^2 & \cdots\cdots & \sigma_X^2 \\ \cdot \quad \cdot \quad \cdot & \cdots\cdots & \cdot \\ \cdot \quad \cdot \quad \cdot & \cdots\cdots & \cdot \\ \sigma_X^2 & \cdots\cdots & \sigma_X^2+\sigma_e^2+\sigma_\eta^2 \end{bmatrix}_{IK \times IK}$$

$B^{-1} = ((b^{rs}))_{IK \times IK}$ is a symmetric matrix having the property

$$b^{rr} = b^{ss}$$

$$b^{rs} = b^{r's'}$$

$= \underset{\sim}{b}$ say.

$$\text{Again } \Sigma_{XY} = \begin{bmatrix} C & 0 & \dots & 0 \\ 0 & C & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & C \end{bmatrix}_{J \times J}$$

Where $C = ((\sigma_{XY}))_{IK \times IK} + \sigma_e^2 \; U_{IK \times IK} = \underset{\sim}{\sigma} + \sigma_e^2 \; U$

$$\text{Now : } \Sigma_{XY} \Sigma_{XX}^{-1} = \begin{bmatrix} CB^{-1} & 0 & \dots & 0 \\ 0 & CB^{-1} & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & CB^{-1} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\sigma}\underset{\sim}{b} + \sigma_e^2 \underset{\sim}{b} & \dots & 0 \\ . & . \dots . & . \\ . & . \dots . & . \\ 0 & 0 \dots & \underset{\sim}{\sigma}\underset{\sim}{b} + \sigma_e^2 \underset{\sim}{b} \end{bmatrix}$$

Therefore $E(Y|X) = \mu_Y + \Sigma_{XY} \Sigma_{XX}^{-1} X,$

$$\text{But } \underset{\sim}{\sigma}\underset{\sim}{b} = \sigma_{XY} \begin{bmatrix} 11 & \dots & 1 \\ 11 & \dots & 1 \\ .. & \dots & . \\ 11 & \dots & 1 \end{bmatrix}_{IK \times IK} \begin{bmatrix} b^{11} & b^{12} & \dots & b^{1,IK} \\ . & . & \dots & . \\ . & . & \dots & . \\ b^{IK,1} & & \dots & b^{IK,IK} \end{bmatrix}$$

$$= \sigma_{XY} \begin{bmatrix} \sum_{i=1}^{IK} b^{i1} & \sum_{i=1}^{IK} b^{i2} & \dots & \sum_{i=1}^{IK} b^{i,IK} \\ . & . & \dots & . \\ . & . & \dots & . \end{bmatrix}_{IK \times IK}$$

In the above matrix all the elements are the same i.e. b_* (say).

Therefore $\underset{\sim}{\sigma}\ \underset{\sim}{b} = \sigma_{XY}\ b_* \begin{bmatrix} 1\,1 & \cdots & 1 \\ 1\,1 & \cdots & 1 \\ \cdot\,\cdot & \cdots & \cdot \\ 1\,1 & \cdots & 1 \end{bmatrix}_{IK \times IK}$

$$\therefore E(Y|X) = \mu_Y + \sigma_{XY}\ b_* \begin{bmatrix} \begin{pmatrix} 1\,1 & \cdots & 1 \\ 1\,1 & \cdots & 1 \\ \cdot\,\cdot & \cdots & \cdot \\ 1\,1 & \cdots & 1 \end{pmatrix}_{IK \times IK} & \cdots\cdots & 0_{IK \times IK} \\ \cdots\cdots & & \cdots\cdots \\ 0_{IK \times IK} & \cdots\cdots & \begin{pmatrix} 1\,1 & \cdots & 1 \\ 1\,1 & \cdots & 1 \\ \cdot\,\cdot & \cdots & \cdot \\ 1\,1 & \cdots & 1 \end{pmatrix} \end{bmatrix}_{J \times J}$$

X missing ?

$$+ \sigma_e^2 \begin{bmatrix} ((b^{rs}))_{IK \times IK} & \cdots & 0_{IK \times IK} \\ \cdots & \cdots & \cdots \\ 0_{IK \times IK} & & ((b^{rs}))_{IK \times IK} \end{bmatrix}_{J \times J}$$

X missing ?

(R)
$$\therefore E(Y_{ijk}|X_{ijk}) = \mu + \alpha_i + \sigma_{XY}\ b_*\ X_{ijk} + \sigma_{XY}\ b_* \sum_{\substack{i' \neq i \\ k' \neq k}} X_{i'jk'} + \sigma_e^2\ b^{rr}\ X_{ijk} + \sigma_e^2\ b^{rs} \sum_{\substack{i' \neq i \\ k' \neq k}} (-1)^m\ X_{i'jk'}$$

Should be $(i',k') \neq (i,k)$?

* all have the same sign

Where m is determined by the sign of the co-factor in the inverse matrix $((b^{rs}))$

$$= \mu + \alpha_i + \beta_1\ X_{ijk} + \beta_2\ \delta_{ijk} + \beta_3\ Z_{ijk}$$

Where $\beta_1 = (\sigma_{XY}\ b_* + \sigma_e^2\ b^{rr})$, $\quad \beta_2 = \sigma_{XY}\ b_*$,

$\beta_3 = \sigma_e^2\ b^{rs}$, $\quad \delta_{ijk} = \sum_{\substack{i' \neq i \\ k' \neq k}} X_{i'jk'}$, $\quad Z_{ijk} = \sum_{\substack{i' \neq i \\ k' \neq k}} (-1)^m\ X_{i'jk'}$

Therefore we can assume model

$$Y_{ijk} = \mu + \alpha_i + b_j + C_{ij} + \beta_1 X_{ijk} + \beta_2 \delta_{ijk} + \beta_3 Z_{ijk} + e'_{ijk}$$

Where μ = general mean

α_i = i^{th} treatment effect

b_j = j^{th} region effect

C_{ij} = the interaction of i^{th} treatment with j^{th} region

$$= m_{ij} + \beta_1 X_{ijk} + \beta_2 \delta_{ijk} + \beta_3 Z_{ijk} + e'_{ijk}.$$

3. Structure of conditional covariance matrix:

$$\text{Now } \Sigma_{YY.X} = \begin{bmatrix} A - CB^{-1}C & \cdots\cdots & 0 \\ 0 \cdots\cdots & & A - CB^{-1}C \end{bmatrix}_{J \times J}$$

Where A, B, C are defined in accordance with the previous section.

$$\text{Write } B = \begin{bmatrix} \sigma_X^2 + \sigma_e^2 + \sigma_\eta^2 & \cdots\cdots & \sigma_X^2 \\ \cdot & \cdots\cdots & \cdot \\ \sigma_X^2 & \cdots\cdots & \sigma_X^2 + \sigma_e^2 + \sigma_\eta^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_X^2 + \lambda & \cdots\cdots & \sigma_X^2 \\ \cdot & \cdots\cdots & \cdot \\ \sigma_X^2 & \cdots\cdots & \sigma_X^2 + \lambda \end{bmatrix} \quad \text{where } \lambda = \sigma_e^2 + \sigma_\eta^2$$

(See backside of page 32) In fact this contradicts author's statement on page 34

It may be easily shown that

$$B^{-1} = \frac{1}{\lambda}\begin{bmatrix} 1-\frac{\sigma_X^2}{IK\,\sigma_X^2+\lambda} & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & (-1)^{r+s}\frac{\sigma_X^2}{IK\,\sigma_X^2+\lambda} & \cdots \\ \cdots\cdots & 1-\frac{\sigma_X^2}{IK\,\sigma_X^2+\lambda} & \end{bmatrix}$$

$$+ = \frac{1}{\lambda}(U_{IK \times IK} + \lambda_1 U'_{IK \times IK})$$

Where $\lambda_1 = \frac{\sigma_X^2}{IK\,\sigma_X^2+\lambda}$ and U is an unit matrix of order IK and U' is a matrix which has got (-1) as the diagonal elements and in the other positions is +1 or -1 according as the sum of the row and column indices are even or odd.

Therefore $CB^{-1}C = \frac{1}{\lambda}(\sigma_{XY} + \sigma_e^2 U)(U + \lambda_1 U')(\sigma_{XY} + \sigma_e^2 U)$

$$= \frac{1}{\lambda}\Big\{IK((\sigma_{XY}^2)) + 2\sigma_e^2((\sigma_{XY})) + \lambda_1(-2)\,IK((\sigma_{XY})) + (\lambda_1\sigma_e^2 + \lambda_1)(-2)((\sigma_{XY})) + \sigma_e^4 U + \lambda_1\sigma_e^4 U'\Big\}$$

if IK is even.

If IK is odd, (-2) in the above expression is replaced by (-1).

$$= \frac{1}{\lambda}\Big\{(IK - 2IK\lambda_1)((\sigma_{XY}^2)) + \left[2\sigma_e^2 - 2(\lambda_1 + \lambda_1\sigma_e^2)\right]\times ((\sigma_{XY})) + \sigma_e^4 U + \lambda_1\sigma_e^4 U'\Big\}$$

Now contribution of $CB^{-1}C$ to the covariance of Y_{ijk} $Y_{i'j'k'}$

is given by $\frac{1}{\lambda}\left\{(IK - 2IK\lambda_1)\sigma_{XY}^2 + \left[2\sigma_e^2 - 2(\lambda_1 + \lambda_1\sigma_e^2)\right]\sigma_{XY} + \sigma_e^4 - \lambda_1\sigma_e^4\right\}$ when $i = i'$, $j = j'$, $k = k'$

$$= \frac{1}{\lambda}\left\{(IK - 2IK\lambda_1)\sigma_{XY}^2 + \left[2\sigma_e^2 - 2(\lambda_1 + \lambda_1\sigma_e^2)\right]\sigma_{XY} \pm \lambda_1\sigma_e^4\right\}$$

for $i = i'$, $j = j'$, $k \neq k'$ or $i \neq i'$, $j = j'$, $k = k'$

and $= 0$ otherwise

Again the contribution of A to the covariance of Y_{ijk}, $Y_{i'j'k'}$

is given by $\sigma_{ii'} + \sigma_e^2 + \sigma_\epsilon^2$ if $i = i'$, $j = j'$, $k = k'$

$\sigma_{ii'}$ if $i \neq i'$, $j = j'$, $k = k'$

σ_{ii} if $i = i'$, $j = j'$, $k \neq k'$

and 0 otherwise

4. Test of significance of interaction :

Define Type I model $Y_{ijk} = \mu + \alpha_i + b_j$

$$+ \beta_1 X_{ijk} + \beta_2 \delta_{ijk} + \beta_3 Z_{ijk} + e_{ijk}$$

and Type II model $Y_{ijk} = \mu + \alpha_i + b_j + c_{ij}$

$$+ \beta_1 X_{ijk} + \beta_2 \delta_{ijk} + \beta_3 Z_{ijk} + e'_{ijk}$$

Residual sum of squares from the Type I model

$$= \sum_{ijk} (Y_{ijk} - Y_{i..} - Y_{.j.} + Y_{...})^2 - \hat{\beta}_1 \sum_{ijk} Y_{ijk}(X_{ijk} - X_{i..}$$

$$- X_{.j.} + X_{...}) - \hat{\beta}_2 \sum_{ijk} Y_{ijk} (\delta_{ijk} = \delta_{i..} - \delta_{.j.} + \delta_{...})$$

$$- \hat{\beta}_3 \sum_{ijk} Y_{ijk}(Z_{ijk} - Z_{i..} - Z_{.j.} + Z_{...})$$

where the symbol . represents the average over the particular suffix and $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are the estimates of β_1, β_2 and β_3 from Type I model.

Again Residual s.s. under Type II model

$$= \sum_{ijk} (Y_{ijk} - Y_{ij.})^2 - \hat{\beta}_{1*} \sum_{ijk} (Y_{ijk} - Y_{ij.})(X_{ijk} - X_{ij.})$$

$$- \hat{\beta}_{2*} \sum_{ijk} (Y_{ijk} - Y_{ij.})(\delta_{ijk} - \delta_{ij.})$$

$$- \hat{\beta}_{3*} \sum_{ijk} (Y_{ijk} - Y_{ij.})(Z_{ijk} - Z_{ij.})$$

where $\hat{\beta}_{1*}$, $\hat{\beta}_{2*}$ and $\hat{\beta}_{3*}$ are the estimates of β_1, β_2 and β_3 from Type II model.

To test the interaction H_0': $Var(C_{ij}) = 0$ for all i, we have

$$F = \frac{IJK - IJ - 3}{(I-1)(J-1)} \times \frac{\text{Res. s.s.(Type I model)} - \text{Res. s.s.(Type II Model)}}{\text{Residual s.s.(Type II Model)}}$$

To determine the distribution of Denominator and Numerator it can be argued in the following way:

Since $Y_{ijk} - Y_{ij.}$ does not involve m_{ij} the residual s.s. in the Denominator for fixed values of X_{ijk} is distributed as constant component of central χ^2 with (IJK - IJ - 3) d.f. and since it has the same distribution for every value of X_{ijk}, unconditionally also, the residual s.s. follows χ^2 distribution. Again under the hypothesis the contribution of m_{ij} to the s.s. in the numerator vanishes and in the result we have the numerator distributed as constant component of χ^2 with (I-1)(J-1) d.f.

Therefore the ratio 'F' follows Snedecor's 'F' distribution (under the null hypothesis) with (I-1)(J-1) and (IJK - IJ - 3) d.f.

5. The property of unbiasedness of test :

When X_{ijk}, m_{ij} are fixed the interaction s.s. (s.s. in the numerator of 'F' ratio) follows constant component of noncentral χ^2 distribution with (I-1)(J-1) d.f. and non-centrality parameter $\dfrac{\sum\sum C_{ij}^2}{Var(Y_{ijk}|X_{ijk}, m_{ij})}$.

Therefore the ratio of interaction m.s. to error m.s. follows a noncentral 'F' distribution under the alternative hypothesis that Var $(C_{ij}) > 0$. Due to the fact that 'F' test is unbiased, we shall get an unbiased test of the ratio under alternative hypothesis with the condition that X_{ijk}, m_{ij} are fixed. Since for every value of X_{ijk}, m_{ij} we get an unbiased test, unconditionally also the property of unbiasedness follows.

6. Gain in the use of auxiliary information :

The use of auxiliary information does not result in increasing the precision of the experiment if

$$\text{Var}(Y_{ijk} | m_{ij}) = \sigma_e^2 + \sigma_\epsilon^2$$

In the case of proposed model

$$\text{Var}(Y_{ijk} | X_{ijk}, m_{ij}) = \sigma_e^2 + \sigma_\epsilon^2 - \frac{(1-\lambda_1)}{\lambda} \sigma_e^4$$

Thus when the covariance technique is useless

$$\text{E(Error mean square)} = \sigma_e^2 + \sigma_\epsilon^2$$ and when the proposed model holds good

$$\text{E(Error mean square)} = \sigma_e^2 + \sigma_\epsilon^2 - \frac{(1-\lambda_1)}{\lambda} \sigma_e^4$$

Therefore the expected reduction in variance due to the use of auxiliary information is $\frac{(1-\lambda_1)}{\lambda} \sigma_e^4$

Where $\lambda_1 = \frac{\sigma_X^2}{IK\,\sigma_X^2 + \lambda}$ and $\lambda = \sigma_e^2 + \sigma_\eta^2$.

One will generally expect some gain by using the covariance technique. Also the assumption of the normality of the distribution of $\{X_j\}$ which is implied in assuming that the distribution of $(m_{1j}, m_{2j} \ldots . m_{Ij}, X_j)$ has a $(I + 1)$ variate normal distribution is not essential for the application of this technique as in the usual regression analysis. However, the specific form of the regression equation (R), depends on the assumption of the $(I + 1)$ variate normal distribution and it would not generally be possible to get a covariance analysis of this form without introducing a distribution of $\{X_j\}$. The reason for this is that E_{ijk} enters both $\{Y_{ijk}\}$ and $\{X_{ijk}\}$. If instead of taking the conditional distribution of $\{Y_{ijk}\}$ when $\{X_{ijk}\}$ is fixed, we consider the conditional distribution of $\{Y_{ijk}\}$ for $\{X_j\}$ fixed, then the above difficulty regarding the distribution of $\{X_j\}$ would not appear in usual regression analysis. However, this would not be appropriate here since $\{X_j\}$ corresponds to a random sample of experimental regions and to fix $\{X_j\}$ would be to fix the regions, so that the basic character of Scheffe's model would change. Thus some assumption regarding the distribution of $\{X_j\}$ has to be made and provisionally one may take a normal distribution as appropriate. The validity of this assumptions can be established through a uniformity trial over a number of randomly selected regions.

Part III

Two Stage Test Procedure

ON STEIN'S TWO SAMPLE THEORY TO TEST THE DIFFERENCE BETWEEN MEANS OF NORMAL POPULATIONS WITH COMMON UNKNOWN VARIANCE AND UNEQUAL AMOUNT OF SAMPLING FROM THE POPULATIONS

1. Introduction :

Stein (1945) developed two sample theory to test hypothesis concerning the mean of a normal population with power independent of population variance. He also considered a general hypothesis for independent homoscedastic normal observations and extended Danzig's (1940) theorem of non-existence of any non-trivial single sample test for the same case. Proceeding on the lines indicated by Stein, Chapman (1950) obtained tests for the equality of means from two normal populations under Fisher-Behren's set-up. Ruben (1961) found a slightly stronger form of two stage sampling to determine confidence intervals for the means of the normal populations with common unknown variance.

For the test of means, both Stein and Ruben considered an equal amount of sampling from the populations in question and initial samples were of same size. But in practical situations, this may not be optimum procedures from the point of view of minimising cost of taking samples, as the relative cost of selecting sampling units may be different and there may be other operational difficulties. Therefore in the following chapter a modification of Stein's procedure, which envisages unequal amount of sampling is considered for the test of means of the normal populations having common unknown variance.

2. Basic Theory :

Take a sample of n_0 observations and calculate

$$S_0^2 = \frac{1}{n_0-1} \sum_{1}^{n_0} (X_i - \bar{X})^2$$ as an estimate of σ^2 with n_0-1 d.f.

Let $n = \text{Max.}\left\{ \left[\frac{S_0^2}{Z}\right] + 1,\ n_0+1 \right\}$ where Z is a predetermined constant and it may be called "Studentized Scale Factor".

Take $n - n_0$ observations such that

$$\text{(i)} \quad \sum_{i}^{n} a_i = 1 \qquad \text{(ii)} \quad a_1 = \ldots = a_{n_0}$$

$$\text{(iii)} \quad \sum_{i}^{n} a_i^2 = \frac{Z}{S_0^2}$$

Now $\dfrac{\sum_{i}^{n} a_i X_i - \mu}{\sqrt{Z}}$ follows 't' distribution with (n_0-1) d.f. and hence a confidence interval for μ can be obtained independent of σ^2.

3. Test Procedure for H : $\mu_1 = \mu_2$:

Let X_{ij}, $i = 1,2$; $j = 1,2 \ldots$ be independent random variables distributed according to $N(\mu_i, \sigma)$. We wish to test hypothesis H : $\mu_1 = \mu_2$. Following Dantzig(1940) and Stein (1945) it can be proved that there does not exist any non-trivial single sample test for the hypothesis ($\mu_1 = \mu_2$), whose power is independent of σ.

Choose a sample of size n_0 and m_0 respectively from the given populations and compute

$$S^2 = \frac{\sum^{n_0}(X_{i1} - \bar{X}_1)^2 + \sum^{m_0}(X_{i2} - \bar{X}_2)^2}{m_0 + n_0 - 2}$$

Then $$n = \max.\left\{\left[\frac{S^2}{Z_1}\right] + 1,\ n_0 + 1\right\}$$

$$m = \max.\left\{\left[\frac{S^2}{Z_2}\right] + 1,\ m_0 + 1\right\}$$

Where Z_1, Z_2 are specified constants.

Additional observations $X_{n_0+1,1} \ldots X_{n,1}$ and $X_{m_0+1,2} \ldots X_{m,2}$ are taken from respective populations and real numbers $a_1, a_2 \ldots a_n$ and $b_1 \ldots b_m$ are chosen such that

$$\sum_{i=1}^{n} a_i = 1 \qquad a_1 = a_2 = \ldots = a_{n_0}$$

$$\sum_{j}^{m} b_j = 1 \qquad b_1 = b_2 = \ldots = b_{m_0}$$

$$S^2 \sum_{i=1}^{n} a_i^2 = Z_1 \qquad S^2 \sum_{j=1}^{m} b_j^2 = Z_2$$

This is possible since

$$\min. \sum a_i^2 = \frac{1}{n} \leq \frac{Z_1}{S^2}$$

$$\min. \sum b_j^2 = \frac{1}{m} \leq \frac{Z_2}{S^2}$$

Now define $t = \frac{\sum^{n} a_i X_{i1} - \sum^{m} b_j X_{j2}}{\sqrt{Z}}$

where $Z = Z_1 + Z_2$

$$= \frac{(\sum^{n} a_i X_{i1} - \sum^{m} b_j X_{j2}) - (\mu_1 - \mu_2)}{\sqrt{Z}} + \frac{\mu_1 - \mu_2}{\sqrt{Z}}$$

$$= u + \frac{\mu_1 - \mu_2}{\sqrt{Z}}$$

Then u follows student's 't' distribution with $n_0 + m_0 - 2$ d.f. Let α be the size of C.R. and

$$P\left\{ t_{n_0 + m_0 - 2} > t_{n_0 + m_0 - 2,\ \alpha/2} \right\} = \alpha/2$$

Then if we reject H, whenever

$$\left| \frac{\sum_i^{n} a_i X_{i1} - \sum_j^{m} b_j X_{j2}}{\sqrt{Z}} \right| > t_{n_0 + m_0 - 2,\ \alpha/2}$$

we get an unbiased test whose power function is given by

$$1 - \beta(\mu) = 1 - P\left\{ -t_{n_0+m_0-2,\ \alpha/2} + \frac{\mu_1 - \mu_2}{\sqrt{Z}} < t_{n_0+m_0-2} < t_{n_0+m_0-2,\ \alpha/2} + \frac{\mu_1 - \mu_2}{\sqrt{Z}} \right\}$$

?

A confidence interval for $\mu_1 = \mu_2$ of predetermined length L, confidence coefficient $1 - \alpha$ can be obtained by choosing Z such that

$$1 - \alpha = P\left\{ -\frac{L}{2\sqrt{Z}} < t_{n_0+m_0-2} < \frac{L}{2\sqrt{Z}} \right\}$$

$2t_{\alpha/2}\sqrt{3} = L$

$$= P\left\{ \sum_i^n a_i X_{i1} - \sum_j^m b_j X_{j2} - L/2\sqrt{Z} < \mu_1 - \mu_2 < \sum_{i=1}^n a_i X_{i1} - \sum_{j=1}^m b_j X_{j2} + L/2\sqrt{Z} \right\}$$

To find expected values of n and m:

Following Stein we can find E(n) and E(m) satisfying inequalities

$$(n_0 + 1)\, P\left\{ \chi^2_{n_0 + m_0 - 2} < y_1 \right\} + \frac{\sigma^2}{Z_1} \times P\left\{ \chi^2_{n_0 + m_0 - 1} > y_1 \right\} < E(n) < (n_0+1) \times P\left\{ \chi^2_{n_0 + m_0 - 2} < y_1 \right\} + \frac{\sigma^2}{Z_1} P\left\{ \chi^2_{n_0 + m_0 - 1} > y_1 \right\} + P\left\{ \chi^2_{n_0 + m_0 - 2} > y_1 \right\}$$

and

$$(m_0 + 1)\, P\left\{ \chi^2_{n_0 + m_0 - 2} < y_2 \right\} + \frac{\sigma^2}{Z_2} \times P\left\{ \chi^2_{n_0 + m_0 - 1} > y_2 \right\} < E(m) < (m_0 + 1) \times P\left\{ \chi^2_{n_0 + m_0 - 2} < y_2 \right\} + \frac{\sigma^2}{Z_2} P\left\{ \chi^2_{n_0+m_0-1} > y_2 \right\} + P\left\{ \chi^2_{n_0 + m_0-2} > y_2 \right\}$$

The maximum error involved by taking upper or lower limit of n and m is only unity.

From the limits of E(n) and E(m) it is clear that

$$\overline{\lim_{\sigma \to \infty}} \left\{ E(n) - \frac{\sigma^2}{Z_1} \right\} \leq 1$$

$$\underline{\lim_{\sigma \to \infty}} \left\{ E(n) - \frac{\sigma^2}{Z_1} \right\} \geq 0 \quad \text{and}$$

$$\overline{\lim_{\sigma \to \infty}} \left\{ E(m) - \frac{\sigma^2}{Z_2} \right\} \leq 1$$

$$\underline{\lim_{\sigma \to \infty}} \left\{ E(m) - \frac{\sigma^2}{Z_2} \right\} \geq 0$$

The approximations $E(n) \cong \frac{\sigma^2}{Z_1}$ and $E(m) \cong \frac{\sigma^2}{Z_2}$ hold provided that $\sigma^2 > \text{Max.} \left\{ Z_1 n_0, Z_2 m_0 \right\}$

4. Selection of Z_1 and Z_2 :

The problem of the optimum division of Z into Z_1 and Z_2 can be best solved by taking into consideration of the problem of cost. One method may be to minimise the cost of taking the second samples with fixed $Z = Z_1 + Z_2$.

Let $n - n_0$ be the size of second sample to be taken from the first population and $m - m_0$ ~~from~~ be the size of the second sample to be taken from the 2nd population. Let C_1, C_2 be the cost per unit of second samples from 1st and 2nd population respectively. Then for given Z, the expected cost of taking second samples is $C_1 E(n - n_0) + C_2 E(m - m_0)$ which is an expression involving Z_1 and Z_2. The problem is to minimise this for fixed $Z = Z_1 + Z_2$.

Now $\underline{\text{lim.}}\ E(n) = (n_0 + 1)P\left\{\chi^2_{n_0+m_0-2} < y_1\right\} + \frac{\sigma^2}{Z_1}P\left\{\chi^2_{n_0+m_0-1} > y_1\right\}$

$$\underline{\text{lim.}}\ E(m) = (m_0 + 1)P\left\{\chi_{n_0+m_0-2} < y_2\right\} + \frac{\sigma^2}{Z_2}P\left\{\chi^2_{n_0+m_0-1} > y_2\right\}$$

where $$y_1 = \frac{(n_0 + 1)\ (n_0 + m_0 - 2)\ Z_1}{\sigma^2}$$

$$y_2 = \frac{(m_0 + 1)\ (n_0 + m_0 - 2)\ Z_2}{\sigma^2}$$

The values of Z_1 and Z_2 (for fixed Z) which minimise the expected cost of taking second samples are given by the solution of the equation

$$\frac{C_1}{C_2} = \frac{\left(1 - \frac{\sigma^2}{Z-Z_1}\right) f_1(Z-Z_1) - \frac{f_2(Z-Z_1)}{(Z-Z_1)^2}}{\left(1 - \frac{\sigma^2}{Z_1}\right) f_1(Z_1) - \frac{f_2(Z_1)}{Z_1^2}}$$

where $$f_1(Z_1) = P\left\{\chi^2_{n_0+m_0-1} > y_1\right\}$$

$$f_1(Z - Z_1) = P\left\{\chi^2_{n_0+m_0-2} > y_2\right\}$$

$$f_2(Z_1) = e^{-\frac{1}{2}y_1}\ y_1^{\frac{1}{2}(n_0+m_0-4)}\ \frac{(n_0+1)(n_0+m_0-2)}{\sigma^2}$$

$$f_2(Z - Z_1) = e^{-\frac{1}{2}y_2}\ y_2^{\frac{1}{2}(n_0+m_0-4)}\ \frac{(m_0+1)(n_0+m_0-2)}{\sigma^2}$$

Since the solution of this equation, solving for Z_1 is not easy, trial and error method of solution will be adopted.

For the case when σ^2 is large, the value of Z_1 which minimises the expected cost is given by the equation

$$\frac{C_1}{Z_1^2} = \frac{C_2}{(Z - Z_1)^2}$$

$$\therefore \sqrt{C_1}/\sqrt{C_2} = Z_1/(Z - Z_1)$$

$$\therefore Z_1 = Z\sqrt{C_1}/(\sqrt{C_2} + \sqrt{C_1})$$

When $C_1 = C_2$, the minimising value of $Z_1 = Z_2 = Z/2$. But the procedure will be more general than Stein's as (n_0, m_0) and (n_1, m_1), where $n_1 = n - n_0$, $m_1 = m - m_0$ need not be equal as postulated by Stein. When preliminary samples from both the populations are equal, the above procedure may be modified by putting simply $n_0 = m_0$

5. Case of more than two populations :

Let $X_{ij} \left\{ \begin{array}{l} i = 1 \ldots t \\ j = 1, 2 \ldots \end{array} \right\}$ be independently normally

distributed with variance σ^2 and means $E(X_{ij}) = \mathcal{Y}_i$, $i = 1 \ldots u$

$E(X_{ij}) = 0$, $i = u+1 \ldots t$

we wish to test H_0 : $\mathcal{Y}_i = 0$ $i = 1 \ldots\ldots p \leq u$, the $\mathcal{Y}_i$'s for $i = p + 1 \ldots\ldots u$ and σ^2 being nuisance parameters.

Obtain preliminary samples as.

$$
\begin{array}{lll}
X_{11}, X_{12} & \cdots\cdots\cdots & X_{1,n_1} \\
X_{21}, X_{22} & \cdots\cdots\cdots & X_{2,n_2} \\
\cdot \quad \cdot & \cdots\cdots\cdots & \cdot \\
X_{i1}, X_{i2} & \cdots\cdots\cdots & X_{i,n_i} \\
\cdot \quad \cdot & \cdots\cdots\cdots & \cdot \\
X_{t1}, X_{t2} & \cdots\cdots\cdots & X_{t,n_t}
\end{array}
$$

Estimate the variance $S^2 = \dfrac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2}{(n_1 + n_2 + .. + n_t) - u}$

Now $n^{(1)} = \max. \left\{ \left[\frac{S^2}{Z_1}\right] + 1, \quad n_1 + 1 \right\}$

$$n^{(2)} = \max. \left\{ \left[\frac{S^2}{Z_2}\right] + 1, \quad n_2 + 1 \right\}$$

$$\cdot \qquad \cdots\cdots\cdots\cdots$$

$$n^{(i)} = \max. \left\{ \left[\frac{S^2}{Z_i}\right] + 1, \quad n_i + 1 \right\}$$

$$\cdot \qquad \cdots\cdots\cdots\cdots$$

$$n^{(t)} = \max. \left\{ \left[\frac{S^2}{Z_t}\right] + 1, \quad n_t + 1 \right\}$$

Where $n^{(1)}$, $n^{(2)}$ ..... $n^{(t)}$ are total sample sizes to be selected from each population and Z_1, Z_2 .... Z_t are predetermined constants.

Let the set of real numbers attached to each set of samples be $\sum_{j=1}^{n^{(i)}} a_{ij} = 1$ and $S^2 \sum_{j=1}^{n^{(i)}} a_{ij}^2 = Z_i$ $\quad i = 1, 2 \cdots t$

Then $$F' = \frac{\sum_{i=1}^{t}\left\{\sum_{j=1}^{n^{(i)}} a_{ij} \, X_{ij}\right\}^2}{Z(n_1 + \, .. \, + n_t - u)}$$

has the noncentral 'F' distribution with d.f. $\sum_{i=1}^{t} n_i - u$ and P

and with noncentrality parameter $\sum_{i=1}^{p} c_i^2 = \dfrac{\sum_{i=1}^{p} y_i^2}{Z(n_1 + n_2 + \cdots + n_t - u)}$

where $Z = Z_1 + Z_2 + \ldots + Z_t$

The test of significance and confidence regions are obtained as in the case of student's hypothesis mentioned by Stein.

The selection of $Z_1, Z_2 \ldots Z_t$ or the subdivision of Z into t components can be made in accordance with the rule mentioned above for the case of testing the means of two normal populations having common unknown variance.

SUMMARY

In the first part of the thesis, the analysis of groups of experiments located at different places has been considered in the line of Scheffe's mixed model (1956), postulated for an industrial experiment. The usual analysis, based on the independence of the effect of the places and treatment x places interaction effect, does not seem to be justified in the actual experimental conditions. Therefore, the analysis has been considered under the model in which the effect due to places depends upon the effect of the interaction between places and treatments and moreover their joint distribution has been assumed to follow a multivariate normal distribution. In the present thesis two types of experiment - factorial experiments and experiments conducted in Incomplete blocks, have been considered for the case of the model under consideration. The usual 'F' test for testing the component of variation due to interaction is valid for the model under consideration. When the interaction is present, the ratio of Interaction M.S. to error M.S. will not have a noncentral 'F' distribution but will still yield an unbiased test. The exact test for the test of treatment means can be made using Hotelling's T^2-statistic.

In the second part of the thesis an attempt has been made to use auxiliary information in Scheffe's mixed model for the analysis of groups of experiments. It has been seen that

the usual 'F' test for the test of the component due to interaction of places with treatment holds good. The property of unbiasedness of the test has also been justified for the test.

In the third part of the thesis Stein's two sample theory for the test of means of normal populations with common unknown variance has been considered for the case when the amount of sampling from the populations in question is unequal. This results in the division of "studentized scale factor" into components which minimise the cost of taking second samples.

REFERENCES

1. Box, G.E.P. (1954), "Some theorems on quadratic forms applied in the study of analysis of variance problems : I. Effect of inequality of variance in the one-way classification". Ann. Math. Stat. Vol. 25: 290 - 302.

2. Chapman, Douglas G. (1950), "Some two sample tests". Ann. Math. Stat. Vol. 21: 601 - 606.

3. Cochran, W.G. and Cox, G.M. (1957), "Experimental Designs", Second edition, John Wiley & Sons, New York, pp. 545 - 568.

4. Dantzig, George B. (1940), "On the non-existence of tests of student's hypothesis having power functions independent of σ", Vol. 11, pp. 186 - 192. Ann. Math. Stat.

5. Eisenhart, C. (1947), "The assumptions underlying the analysis of variance". Biometrics Vol. 3: 1 - 21.

6. Imhof, J.P. (1961), "Computing the distribution of quadratic forms in normal variables". Biometrike Vol. 48: 419 - 426.

7. Kempthrone, O. (1952), "The Design and Analysis of experiments". John Wiley & Sons, New York.

8. Plackett, R.L. (1960), "Models in the Analysis of Variance". J. Roy. Stat. Soc. Series B Vol. 22: 195-217.

9. Ruben, H. (1961), "Studentisation of Two-stage sample means from Normal populations with unknown common variance". Sankhya Series A, Vol. 23: 231 - 250.

10. Scheffe, H. (1956), "A mixed model for the analysis of variance". Ann. Math. Stat. Vol. 27: 23 - 36.

11. Scheffe, H. (1959), "The Analysis of Variance". John Wiley & Sons, New York.

12. Stein, C.M. (1945), "A two sample test for a linear hypothesis whose power is independent of variance". Ann. Math. Stat. Vol. 16: 243 - 258.

13. Satterthwaite F.E. (1946), "An approximate distribution of estimates of variance components". Biometrics Bull. Vol. 2: 110 - 114.

14. Yates, F. and Cochran, W.G. (1938), "The analysis of groups of experiments". J. Agri. Sci. Vol. 28: 556 - 580.

15. Indian Central Cotton Committee Publication (1962), "Manuring of cotton in India".