SANKHYĀ
THE INDIAN JOURNAL OF STATISTICS

Edited By P. C. MAHALANOBIS

VOL. 7 PART 3 1946

ON THE LINEAR COMBINATION OF OBSERVATIONS AND THE GENERAL THEORY OF LEAST SQUARES

BY C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta.

INTRODUCTION

The general theory of linear combinations of observations as estimates of parametric functions has been treated by the author (Rao: 1943, 1945a, 1945b) under very general conditions. A unified method of approach to the problem of testing of linear hypotheses, involved in a variety of cases, has also been put forward and the necessary distributions worked out in the case when the variances and covariances are known and observations, whose expectations are linear functions of certain parameters, form a multivariate normal system. This problem which admits a simple solution leading to the use of published tables of χ^2 alone brings in fresh complications when the variances and covariances are not known. But by suitable 'Studentisation' it has been possible to show (Rao: 1945c) that many problems in testing of linear hypotheses, for some of which solutions were not available, can be solved with the help of two distributions viz Fisher's t and z alone.

The main aim of the present paper is to bring out the generality of the method of least squares of which all tests of linear hypotheses come out as special cases. Various aspects of this problem have been considered by Markoff (1912), Student (1908), Fisher (1922, 1925) Neyman and Pearson (1930, 1931) and David and Neyman (1939). A recent paper which gives a useful application of the general ideas is due to Kolodziejczyk (1935) who following the theories of Neyman and Pearson (1930,1931,1933) gave a general treatment of the problem but the conclusions arrived at are same as those of the above authors. *The important restrictions in the problem treated above are certain inequality relations connecting the number of parameters, the number of observations and the rank of the matrix of the compounding coefficients of parameters. These restrictions have now been withdrawn and it has been possible to show that by a certain rule of procedure the necessary statistics can be obtained from minimised sum of squares.* This problem for some particular cases of linear hypothesis has been treated by Bose (1944) from a different point of view.

The problem treated here is the theory of estimation and testing of hypothesis concerned with the linear functions of certain parameters and regression coefficients the compounding coefficients in the latter case being functions of concomitant variates. The theory of least squares may now be made secure on solid foundations and a unique principle, 'without which the solution of new problems has, as it were, to be picked out, afresh, by intution' leading to a 'generality in the customary methods' is made available.

As an application of this method the appropriate linear hypothesis and the analysis of variance and covariance in biological experiments have been considered and a general theory of statistical regression is given. Some examples are worked out to illustrate the theory.

In the general treatment of the problem some results are stated without proof as they can be easily constructed from the methods given by Rao (1945a, 1945b)

2. THE GENERAL PROBLEM WITH CONCOMITANT VARIATES.

Let $Y = (y_1\ y_2 \ldots y_n)$ be the vector of n stochastic variates $y_1, y_2, \ldots y_n$ and X be the matrix, with n rows and k columns, of k concomitant variates, the k observations $(i=1,2,\ldots,k)$ in the j-th row corresponding to the j-th stochastic variate y_j. Let $\Theta=(\theta_1,\theta_2\ldots\theta_n)$ be the vector of expectation of Y for the given set of observations on the concomitant variates. This is expressed as $E(Y : X)=\Theta$ or simply $E(Y)=\Theta$ when there is no ambiguity. It is given that

$$\Theta=FG'=TA'+RX' \qquad \ldots (2\cdot 1)$$

where F and G are partitional matrices defined by $F=(T|R)$ and $G=(A|X)$, $T=(\tau_1\ \tau_2\ldots\tau_m)$ and $R=(\rho_1\ \rho_2\ldots\rho_k)$ are the row matrices of m unknown parameters τ's and k unknown parameters ρ's and A is a known matrix with n rows and m columns.

The assumption made above is one of linear regression on the concomitant variates and this can be done, without loss of generality, for if the regression is a general polynomial or any other linear combination of functions of directly observable concomitant variates (such as squares, cubes etc. in the case of polynomial regression or logorithms, exponential etc.) each of these functions may be regarded as a separate concomitant variate. The general problem treated here is the case where, for a given set of observations on concomitant variates the expectation of stochastic variates are linear functions of m unknown parameters τ's and k unknown parameters ρ's, the compounding coefficients for the ρ parameters being single valued functions of concomitant variates.

If $y_1, y_2, \ldots, y_n$ do not have any functional relationship we can get a positive definite matrix Λ of rank n as $E[(Y-\Theta)'(Y-\Theta) : X]$ where E stands for the mathematical expectation, in which case Λ is referred to as the dispersion matrix of the stochastic vector Y for the observed concomitant matrix X. It is assumed that the elements of Λ are finite and are known apart from a constant multiplier. *We are not assuming any equality on inequality relations among n, k, m and the ranks of* Λ *and* X. The set of equations $E(Y) = FG'$ are known as observational equations.

There are two types of problems to be answered in the theory of linear estimation connected with concomitant variates.

(i) Given a linear parametric fonction such as LT' where $L=(l_1\ l_2\ldots l_m)$ is a row matrix, the problem is to find a vector B such that

$$\text{(a)}\quad E(BY') = LT' \text{ independently of T and R} \qquad \ldots (2.2)$$

$$\text{(b)}\quad V(BY') \text{ is minimum.} \qquad \ldots (2.3)$$

(ii) Given a linear parametric function such as MR' where $M=(m_1 m_2 \ldots m_k)$ is a row matrix the problem is to find a vector C such that

$$\text{(a)}\quad E(CY') = MR' \text{ independently of T and R} \qquad \ldots (2.4)$$

$$\text{(b)}\quad V(CY') \text{ is minimum.} \qquad \ldots (2.5)$$

In the case (a) is satisfied the corresponding parametric functions are said to be estimable and if (b) is also satisfied the parametric functions are said to have the best unbiassed estimates given by the corresponding linear functions of the variates.

The problems mentioned above are only particular cases of the general problem considered by Rao (1945a) but requires a special discussion in view of the importance of the role of concomitant variates which will be dealt with in a later section.

The definition of unbiassedness involved in (2·2) gives that if

$$LT' \equiv E(BY') \equiv BAT' + BXR' \quad \ldots \quad (2.6)$$

then

$$BA = L \text{ and } BX = 0 \quad \ldots \quad (2.7)$$

Also if B exists such that $BA=L$ and $BX=0$ then $LT'=E(BY')$ which gives the following theorem.

Theorem 2a. *The necessary and sufficient condition that LT' is estimable is that there exists a vector B such that $BA=L$ and $BX=0$.*

From the set of B's satisfying the condition (2.7) we have to choose B such that $V(BY') = B\Lambda B'$ is the least. Introducing the vectors $2D=(d_1\, d_2 \ldots d_m)$ and $2F=(f_1\, f_2 \ldots f_k)$ of Lagrangian multipliers and minimising the expression.

$$B\Lambda B' - 2D(A'B' - L') - 2FX'B' \quad \ldots \quad (2\cdot 8)$$

We get

$$B\Lambda - DA' - FX' = 0 \quad \ldots \quad (2\cdot 9)$$

$$BA = L, \qquad BX = 0 \quad \ldots \quad (2\cdot 10)$$

Multiplying both sides of (2.9) by Λ^{-1}, the inverse of Λ, we get the vector B leading to the best estimate as

$$B = DA'\Lambda^{-1} + FX'\Lambda^{-1} \quad \ldots \quad (2.11)$$

which with the help of (2.10) shows that the vectors D and F satisfy

$$DA'\Lambda^{-1}A + FX'\Lambda^{-1}A = L \quad \ldots \quad (2\cdot 12)$$

$$DA'\Lambda^{-1}X + FX'\Lambda^{-1}X = 0 \quad \ldots \quad (2\cdot 13)$$

the best estimate and the minimum variance being

$$BY' = DA'\Lambda^{-1}Y' + FX'\Lambda^{-1}Y' \quad \ldots \quad (2.14)$$

$$DA'\Lambda^{-1}AD' + FX'\Lambda^{-1}AD' = LD' \quad \ldots \quad (2\cdot 15)$$

Theorem 2b. *The necessary and sufficient condition that the parametric function LT' is estimable is that there exist vectors D and F satisfying the equations*

$$(D|F)(A'|X')\Lambda^{-1}A = L$$

$$(D|F)(A'|X')\Lambda^{-1}X = 0$$

and the best unbiassed estimate of LT', if estimable, and its variance are given by $(D|F)(A'|X')\Lambda^{-1}Y'$ and LD' respectively.

The necessary and sufficient condition given in theorem 2b is only another form of that given in theorem 2a and is more useful as it is directly connected with the best estimate. Similarly we have the following theorems regarding the estimability and the best estimate of the parametric functions.

Theorem 2c. *The necessary and sufficient condition that the parametric function MR' is estimable is that there exists a vector C such that $CA=O$ and $CX=M$.*

Theorem 2d. *The necessary and sufficient condition that MR' is estimable is that there exist vectors G and H satisfying the equations*

$$(G|H)(A'|X')\Lambda^{-1}A = O$$

$$(G|H)(A'|X')\Lambda^{-1}X = M$$

and the best estimate of MR' if estimable and its variance are given by $(G\,|\,H)(A'\,|\,X')\,\Lambda^{-1}Y'$ and MH' respectively.

The following theorem on uniqueness of the results can be proved as in (Rao : 1945a.)

Theorem 2e. *The best unbiassed estimates are unique independently of whatever may be the vectors F and D of theorem 2b and G and H of theorem 2d satisfying the corresponding equations and the variances of the estimates thus arrived at are the least.*

3. EXTENSION OF MARKOFF'S RESULTS

In this section a simple and a practical procedure for the problem of estimation of linear parametric functions, without laying much emphasis on the algebra given in the previous section has been developed. From the variates of the stochastic vector Y we construct the variates $Y\,\Lambda^{-1}\,(A\,|\,X)$ by a linear transformation so that

$$E[Y\Lambda^{-1}(A\,|\,X)]=(T\,|\,R)\,(A\,|\,X)'\Lambda^{-1}(A\,|\,X) \quad \ldots \quad (3.1)$$

leading to the equations

$$E(U)=(T\,|\,R)H'_1,\ E(W)=(T\,|\,R)H'_2 \quad \ldots \quad (3\cdot 2)$$

where

$$U=Y\Lambda^{-1}A,\ W=Y\Lambda^{-1}X \quad \ldots \quad (3.3)$$

and

$$(A\,|\,X)'\,\Lambda^{-1}(A\,|\,X)=(H_1\,|\,H_2)' \quad \ldots \quad (3.4)$$

From theorem 2b we get that if there exist vectors F and D such that

$$LT'\equiv(F\,|\,D)\,(H_1\,|\,H_2)\,\,(T\,|\,R)' \quad \ldots \quad (3.5)$$

then the best estimate of LT′ is given by $(F\,|\,D)\,\,(U'\,|\,W')$. If $(\bar{T}\,|\,\bar{R})=(t_1\,t_2\ldots t_m\,|\,r_1\,r_2\ldots r_k)$ be a solution of the equations $(U\,|\,W)=(T\,|\,R)\,(H_1\,|\,H_2)'$ then because of (3·5) we have

$$L\bar{T}'=(F\,|\,D)\,(H_1\,|\,H_2)\,(\bar{T}\,|\,\bar{R})'=(F\,|\,D)\,\,(U'\,|\,W') \quad \ldots \quad (3\cdot 6)$$

which shows that the best estimate of LT′ if estimable, can be obtained from $L\bar{T}'$ where $\bar{T}=(t_1t_2\ldots t_m)$ as defined above. This is unique independently of whatever may be the solution $(\bar{T}\,|\,\bar{R})$ satisfying the above equations. Following the proof given in (1945a) we can show that the result of substituting $(\bar{T}\,|\,\bar{R})$ in LT′ leads to the best estimate if and only if

(i) $L\bar{T}'$ is homogeneous in y's

(ii) $\hat{E}(L\bar{T}')=LT'$

are satisfied. This shows that the parametric function LT′ need not be tested for estimability before applying Markoff's principle of substitution. If it is not estimable either one or both of the conditions are violated in which case its estimate ceases to exist ; similar results can be derived in the case of the parametric function MR′. Hence we get the following theorem.

Theorem 3a. *The best estimate of the parametric function LT′ (or MR′) exists and is given by $L\bar{T}'$ (or $M\bar{R}'$) where $\bar{T}$ (or $\bar{R}$) is the vector giving a set of solutions of the equations defined in (3·1) $Y\Lambda^{-1}(A\,|\,X)=(T\,|\,R)(A'\Lambda^{-1}A+X'\Lambda^{-1}X)$ if (i) $L\bar{T}'$ (or $M\bar{R}'$) is homogeneous in the stochastic variates and (ii) $E(L\bar{T}')=LT'$ (or $E(M\bar{R}')=MR'$).*

Since $(\bar{T}\,|\,\bar{R})$ is any solution of the equations given above we get the following result.

Theorem 3b. *The vector $(\bar{T}\,|\,\bar{R})$ mentioned in theorem 3a can be obtained after adding any consistent and convinient set of equations in τ's and ρ's to the equations given in theorem (3a) viz. $Y\Lambda^{-1}(A\,|\,X)=(T\,|\,R)\,(A'\Lambda^{-1}\,A+\chi'\Lambda^{-1}\chi)$ and solving them.*

The equations $(U\,|\,W)=(T\,|\,R)\,(H_1\,|\,H_2)'$ are called normal equations and are readily obtained by equating to zeroes the partial derivatives of

$$\sum_i \sum_j \lambda^{ij}\,(y_i-\theta_i)\,(y_j-\theta_j) \quad \ldots \quad (3.7)$$

when λ^{ij} are the elements of Λ^{-1}, with respect to the parameters τ's and ρ's respectively.

The analysis is complete with the help of the following theorem which has been proved in Rao (1945a; 1945b).

Theorem 3c. *The set of equations* $Y\Lambda^{-1}(A \mid X)=(T \mid R)\,(A'\Lambda^{-1}A+X'\Lambda^{-1}X)$ *are always solvable for* τ*'s and* ρ*'s.*

4. REGRESSION PARAMETERS AND THE STRUCTURE OF CONCOMITANT VARIATES.

The parameters $\rho_1, \rho_2, \ldots, \rho_k$, whose compounding coefficients are functions of concomitant variates, are known as regression parameters. In this Section some properties of regression parameters, simpler methods of estimation and the structure of concomitant variates are discussed.

From the normal equations $(U \mid V)=(T \mid R)\,(H \mid H_2)'$ we can, by the method of sweep out, eliminate the set of parameters $\tau_1, \tau_2, \ldots \tau_m$ or $\rho_1, \rho_2, \ldots, \rho_k$ in which case the resulting equations may be represented as

$$P=TS' \quad \text{or} \quad Z=RN' \qquad \ldots (4.1)$$

It has been shown (Rao : 1945a) that the dispersion matrix of $(U \mid V)$ is $(H_1 \mid H_2)$ which is the matrix of the normal equations and this intrinsic property of the normal equations is preserved even when some of the parameters are eliminated by the method of sweep out. Hence we get the following theorem.

Theorem 4a. *The dispersion matrices of P and Z, of the equations* $P=TS'$ *and* $Z=RN'$ *obtained from the normal equations* $(U \mid V)=(T \mid R)(H_1 \mid H_2)'$ *by the method of sweep out, are the matrices of resulting equations S and N respectively.*

The theorem 4a gives us that if the best estimate of LT' (or MR') is BP' (or CZ') then

$$V\,(BP') = BSB' = LB' \qquad \ldots (4.2)$$

$$V\,(CZ') = CNC' = MC' \qquad \ldots (4.3)$$

and if L_1T' and L_2T' are estimated by B_1P' and B_2P' then

$$\text{Cov}\,(B_1P' \,.\, B_2P')=L_1B'_2=L_2B'_1 \qquad \ldots (4.4)$$

with similar results for the covariance of the best estimates of M_1R' and M_2R'.

We could by any other device eliminate the τ's and get the equations in ρ's but the above intrinsic properties are not usually preserved. But in this case we can construct another system of parameters which are linear functions of ρ's and adjust such that the matrix of the equations in the new variables is the dispersion matrix of the corresponding quantities which occur on one side of the equations. By certain devices we can preserve these intrinsic properties and these are particularily important as the further calculation of standard errors are simplified.

The normal equations from which the best estimates are to be obtained are

$$Y\Lambda^{-1}A = TA'\Lambda^{-1}A+RX'\Lambda^{-1}A \qquad \ldots (4.5)$$

$$Y\Lambda^{-1}X = TA'\Lambda^{-1}X+RX'\Lambda^{-1}X \qquad \ldots (4.6)$$

If the regression parameters are zero then the normal equations for the τ parameters corresponding to the observational equations $E(Y)=TA'$ are

$$Y\Lambda^{-1}A=TA'\Lambda^{-1}A \qquad \ldots (4.7)$$

Since the parametric functions TA' are all estimable there exists a matrix F with n rows and m columns such that

$$TA'=TA'\Lambda^{-1}AF' \qquad \ldots (4.8)$$

in which case the best estimates of the parametric functions TA′ are given by $Y\Lambda^{-1}AF'$. Since (4.8) holds identically

$$A' = A'\Lambda^{-1}AF' \quad \text{.. (4.9)}$$

Premultiplication of both sides of (4.9) by F and taking the transpose we get

$$FA' = FA'\Lambda^{-1}AF' \quad \text{.. (4.10)}$$

$$AF' = FA'\Lambda^{-1}AF' \quad \text{.. (4.11)}$$

which shows

$$AF' = FA' \quad \text{.. (4.12)}$$

Post multiplication of both sides of (4.5) by F′ we get

$$Y\Lambda^{-1}AF' = TA'\Lambda^{-1}AF' + RX'\Lambda^{-1}AF' \quad \text{.. (4.13)}$$

$$YN' = TA'N' + RX'N' \quad \text{.. (4.14)}$$

where

$$N' = \Lambda^{-1}AF'$$

Post multiplying both sides of (4.14) with $\Lambda^{-1}X$ and subtracting from (4.6) we get the equations giving the regression parameters as

$$Y(I-N')\Lambda^{-1}X = RX'(I-N')\Lambda^{-1}X \quad \text{.. (4.15)}$$

which is the same as

$$Y(I-N')\Lambda^{-1}(I-N)X = RX'(I-N')\Lambda^{-1}(I-N)X \quad \text{.. (4.16)}$$

since

$$(I-N')\Lambda^{-1}N = \Lambda^{-1}N - \Lambda^{-1}AF'\Lambda^{-1}FA'\Lambda^{-1}$$
$$= \Lambda^{-1}N - \Lambda^{-1}FA'\Lambda^{-1} = \Lambda^{-1}N - \Lambda^{-1}N = 0 \quad \text{.. (4.17)}$$

by the use of (4.10) and (4.12). Also the dispersion matrix of $Y(I-N')\Lambda^{-1}X$ is

$$X'\Lambda^{-1}(I-N)\Lambda(I-N')\Lambda^{-1}X = X'(I-N')\Lambda^{-1}X \quad \text{.. (4.18)}$$

for

$$X'\Lambda^{-1}N\Lambda(I-N')\Lambda^{-1}X = X'\Lambda^{-1}FA'(I-N')\Lambda^{-1}X$$
$$= X'\Lambda^{-1}(FA'-AF') = 0 \quad \text{.. (4.19)}$$

which shows that by the method adopted above the variance-covariance property is preserved in the equations (4.15) for the estimation of regression parameters. The set of equations in the form expressed in (4.16) shows that these are normal equations derivable from the set of observational equations $E(\eta) = R\xi'$ with the dispersion matrix of η as Λ where

$$\eta = Y(I-N') \text{ and } \xi = (I-N)X \quad \text{.. (4.20)}$$

which may be called the *residual* stochastic vector and the concomitant matrix respectively.

Let $X_1, X_2, \ldots, X_k$ represent the vectors of the 1st, 2nd, ... and k-th concomitant variates, the *i*-th element of each corresponding to y_i the i-th stochastic variate. We can construct the observational equations

$$E(Y) = T^{(o)}A' \text{ and } E(X_i) = T^{(i)}A' \quad \text{.. (4.21)}$$
$$(i = 1, 2, \ldots k)$$

where $T^{(i)} = (\tau_1^{(i)}\ \tau_2^{(i)} \ldots \tau_m^{(i)})$ are the parameters corresponding to the i-th concomitant vector and and $T^{(o)} = (\tau_1^{(o)}\ \tau_2^{(o)} \ldots \tau_m^{(o)})$ for the stochastic vector and obtain the best unbiassed estimates of the parametric functions $T^{(o)}A'$ and $T^{(i)}A'$. There are n variables in each case and n best estimates of their corresponding expectations. If we denote the best estimates by placing a bar over the vector of the parameters we get the residual vectors

$$\eta = Y - \bar{T}^{(o)}A' \text{ and } \xi_i = X_i - \bar{T}^{(i)}A' \quad \text{.. (4.22)}$$

mentioned in (4·18). Hence we get the theorem.

Theorem 4b. *The normal equations leading to the best estimates of the regression parameters are derivable from the observational equations* $E(\eta) = R\xi'$ *construsted with the help of residual vectors* η *and* ξ_i *derived in (4·22) and assuming the dispersion matrix of* η *to be* Λ. *The equations thus obtained for the estimation of regression parameters possess the variance-covariance properties.*

With the help of these equations $\eta \Lambda^{-1}\xi = R\xi' \Lambda^{-1}\xi$ we can solve for the ρ parameters and use it for the purpose of estimating linear functions of regression parameters and finding their variances and covariances. Having obtained the vector $\bar{R}$ of solutions we can substitute it in (4·5) to obtain the equations in the τ parameters.

$$Y \Lambda^{-1} A - \bar{R}X' \Lambda^{-1}A = TA' \Lambda^{-1}A \qquad .. \quad (4.23)$$

If LT′ is an estimable paramatric function then there exists a vector C such that $CA'\Lambda^{-1}AT' = LT'$ in which case the best estimate is

$$C\,A'\Lambda^{-1}Y' - CA'\,\Lambda^{-1}X\;\bar{R}' = L\bar{T}^{(0)'} - \Sigma r_i L\bar{T}^{(i)'} \qquad .. \quad (4.24)$$

where $\bar{T}^{(0)}$ and $\bar{T}^{(i)}$ are as defined above and r_i is solution for ρ_i the ith regression parameter. Hence we get the theorem.

Theorem 4c. *The best unbiassed estimate of an estimable parametric function LT′ is given by* $LT^{(0)'} - \Sigma r_i L\bar{T}^{(i)'}$ *where* $LT^{(0)'}$ *and* $L\bar{T}^{(i)'}$ *are the best estimates of* $LT^{(0)'}$ *and* $LT^{(i)'}$ *from the observational equations* $E(Y) = T^{(0)}A'$ *and* $E(X_i) = T^{(i)}A'$ *and* r_i *are the solutions of the equations* $\eta\Lambda^{-1}\xi = R\xi'\Lambda^{-1}\xi$ *where* η *and* ξ *are the residual vectors defined in* (4·22).

It appears as if the parametric functions connected with the stochastic variates and the concomitant variates are related with the parametric function LT′ in the manner

$$LT' = LT^{(0)'} - \Sigma\rho_i\, LT^{(i)'} \qquad .. \quad (4.25)$$

which may admit a physical interpretation with reference to particular problems. We can find the variances and covariances of the estimates given in (4.24) with the help of the following theorem. The expressions $L\bar{T}^{(i)'}$ are constants being only functions of the concomitant variates.

Theorem 4d. *The function* $L\bar{T}^{(0)'}$ *the estimate of LT′ when the regression parameters are zero is uncorrelated with the estimates of linear functions of regression parameters.*

The proof follows from the fact that $L\bar{T}^{(0)'}$ is a linear combination of $Y\Lambda^{-1}A$ or $Y\Lambda^{-1}AF' = YN'$ defined above, and the estimate of a linear function of the regression parameters is a linear combination of $Y\Lambda^{-1}(I-N')$ in which case the relation $(I-N')\Lambda^{-1}N=0$, proved in (4.17) showing that the elements of YN' and $Y\Lambda^{-1}(I-N')$ are uncorrelated, establishes the theorem. Hence we have

$$V[L\bar{T}^{(0)'} - \Sigma r_i\, L\bar{T}^{(i)'}] = V[L\bar{T}^{(0)'}] + V[\Sigma r_i\, L\bar{T}^{(i)'}] \qquad .. \quad (4\cdot 26)$$

for each of which the variances are known.

Similarly

$$\begin{aligned} &\mathrm{Cov}[L\bar{T}^{(0)'} - \Sigma r_i\, L\bar{T}^{(i)'}]\,[M\bar{T}^{(0)'} - \Sigma r_i\, M\bar{T}^{(i)'}] \\ &= \mathrm{Cov}[LT^{(0)'}\,.\,M\bar{T}^{(0)'}] + \mathrm{Cov}[\Sigma r_i\, LT^{(i)'}\,.\,\Sigma r_i\, MT^{(i)'}] \qquad .. \quad (4\cdot 27)\end{aligned}$$

The expressions (4.25) for the estimate of a parametric function LT′ and the expressions (4.26) and (4.27) for the variances and covariances consist of two portions (i) the corresponding expressions when the regression parameters are absent and (ii) a correction due to the concomitant variates. The calculations are so arranged that the expressions obtained by assuming the regression parameters to be zero are made use of in deriving the correction factors themselves.

It is interesting to note that the above analysis is not limited to concomitant variates alone but applicable whenever we want to distinguish two sets of parameters in the theory of linear estimation. Thus we could, in the above case, reverse the roles of τ's and ρ's and consider the elements of A as observations on certain concomitant variates.

5. THE CASE OF INDEPENDENT VARIATES

A general problem that occurs in the estimation of regression and other parameters is when the stochastic variates are independent and have the same common variance. If this common variance is σ^2 then on replacing Λ by σ^2 I, in the estimating equations we find that σ^2 cancells out in the normal equations giving the result that the best estimates are independent of the value of σ^2 whether known or unkonwn. In this section the practical procedure for estimation is dealt with at length and certain short cuts are suggested.

Let the observational equations be

$$E(y_i)=a_{i1}\tau_1+\ldots+a_{im}\tau_m+\rho_1x_{1i}+\ldots+\rho_kx_{ki} \qquad i=1,2\ldots,n \quad \ldots \quad (5\cdot1)$$

with certain restrictions on parameters

$$g_i=r_{i1}\tau_1+\ldots+r_{im}\tau_m \qquad i=1,2,\ldots,s \quad \ldots \quad (5\cdot2)$$

In the general treatment given above we have not considered linear restrictions on the parameters but as shown in (Rao : 1945b) this does not introduce any fresh difficulty but in deriving the normal equations the quadratic form :

$$\lambda^{ij}(y_i-\theta_i)(y_j-\theta_j) \quad \ldots \quad (5\cdot3)$$

has to be minimised subject to the linear restrictions. The expressions for the estimates, variances and covariances all follow from the general treatment given in (Rao : 1945b).

Assuming the regression parameters to be zero we obtain the best estimates of parametric functions. Let $c_{i0}+c_{i1}\,y_1+\ldots+c_{in}\,y_n=\bar{y}_i$ be the best estimate of $a_{i1}\;\tau_1+\ldots+a_{im}\;\tau_m$ when ρ's are zero but τ's subject to (5.2). Then the i-th residual is $(y_i-\bar{y}_i)$. Similarly the i-th residual for the j-th concomitant variate is $\xi_{ji}=x_{ji}-c_{i0}-c_{i1}x_{j1}-\ldots-c_{in}x_{jn}$ where c's are the same as those used in the case of y's. The residual observational equations are then

$$E(\eta_i)=\rho_1\xi_{1i}+\ldots+\rho_k\xi_{ki} \qquad i=1,2,\ldots,n \quad \ldots \quad (5\cdot4)$$

from which we get the normal equations giving ρ's as

$$\Sigma\eta_i\xi_{ji}=\rho_1\Sigma\xi_{1i}\xi_{ji}+\ldots+\rho_k\Sigma\xi_{ki}\xi_{ji} \qquad j=1,2,\ldots,k \quad \ldots \quad (5\cdot5)$$

Certain short cuts are available if solutions of the normal equations corresponding to each of the variates are available. Starting from the observational equations (5.1) with the restrictions (5.2) and assuming that the regression parameters are absent we get the normal equations whose solution we represent by $t_1^{(0)},t_2^{(0)},\ldots,t_m^{(0)}$. By replacing the n observations on the stochastic variate by those of the j-th concomitant variate we can get corresponding solutions which we represent by $t_1^{(j)},t_2^{(j)},\ldots,t_m^{(j)}$.

$$\Sigma\eta_i\xi_{ji}=\Sigma(y_i-a_{i1}t_1^{(0)}-\ldots)(x_{ji}-a_{i1}t_1^{(j)}-\ldots)$$

$$=\Sigma y_ix_{ji}-t_1^{(j)}\Sigma a_{i1}y_i-t_2^{(j)}\Sigma a_{i2}y_i-\ldots \quad \ldots \quad (5\cdot6)$$

$$=\Sigma y_ix_{ji}-t_1^{(0)}\Sigma a_{i1}x_{ji}-t_2^{(0)}\Sigma a_{i2}x_{ji}-\ldots \quad \ldots \quad (5\cdot7)$$

Similarily

$$\Sigma\,\xi_{im}\xi_{jm}=\Sigma x_{im}x_{jm}-t_1^{(j)}\Sigma a_{m1}x_{im}-\ldots \quad \ldots \quad (5\cdot8)$$

$$=\Sigma x_{im}\,x_{jm}-t_1^{(i)}\Sigma a_{m1}x_{jm}-\ldots \quad \ldots \quad (5\cdot9)$$

With the help of (5.6) to (5.9) we can set down the quations (5.5) from which the regression parameters are solved for. If $r_1, r_2, .., r_k$ are the solutions thus obtained then the best estimate of $l_1 \tau_1 + .. + l_m \tau_m$, if estimable, is given by

$$\Sigma l_i (t_i^{(o)} - r_1 t_i^{(1)} - r_2 t_i^{(2)} - \ldots \ldots) \quad .. \quad (5 \cdot 10)$$

The estimates $\Sigma l_i t_i^{(o)}$ and $\Sigma\Sigma l_i r_j t_i^{(j)}$ are uncorrelated as shown in theorem 4d, so that the variance of (5.10) is the sum of the variances of its two uncorrelated parts. The variance of each part is easily derivable from the normal equations from which the $t_i^{(o)}$ and r_j are obtained. Thus if $t_i^{(o)}$ are obtained from the equations.

$$Q_s = b_{1s} t_1^{(o)} + \ldots + b_{lm} t_m^{(o)} \quad .. \quad (5 \cdot 11)$$
$$s = 1, 2, \ldots, m$$

and (5.2) and if $\Sigma l_i t_i^{(o)} = b_o + \Sigma b_i Q_i$ then $V(\Sigma l_i \, t_i^{(o)}) = \Sigma l_i \, b_i \sigma^2$ and if the equations giving the regression parameters are

$$P_i = h_{io} + h_{i1} r_1 + \ldots + h_{ik} r_k \quad .. \quad (5 \cdot 12)$$
$$i = 1, 2, \ldots k$$

and if $\Sigma m_i r_i = n_o + \Sigma n_i P_i$ then $V(\Sigma m_i r_i) = \Sigma m_i n_i \sigma^2$ and so on for the covariances also. It is to be observed here that the solutions t's and r's represent the best estimates of the parameters τ's and ρ's only when they are estimable individually but they are helpful otherwise in giving the best estimates of the estimable parametric functions. Estimability can also be tested with the help of these solutions as indicated before. All the equations used above are always solvable (Rao : 1945a) and it is enough for our purpose to get any single solution.

6. THE GENERAL THEORY OF LEAST SQUARES

The hypothesis involved in the theory of linear estimation is the assignment of the value of a single parametric function or the values of a number of parametric functions. The method of construction of suitable statistics and their distributions when the stochastic variates form a multivariate normal system and the variances and covariances are known have been discussed fully in Rao (1945a, 1945b). If LT′ is a parametric function whose assigned value is ξ then denoting the estimate of LT′ by $L\bar{T}'$ we construct the statistic

$$v = (L\bar{T}' - \xi) / \sqrt{V(L\bar{T}')} \quad .. \quad (6 \cdot 1)$$

and use it as a normal variate. To test the hypothesis $L_1 T' = \xi_1, \ldots, L_r T' = \xi_r$ concerning r independent parametric functions we find the vector of best estimates $P = (P_1 \, P_2 \ldots P_r)$ where $P_i = L_i \bar{T}' - \xi_i$ of $L_1 T' - \xi_1, .., L_r T' - \xi_r$ with their dispersion matrix D and construct the generalised variance statistic V defined by the root of the determinental equation

$$|P'P - VD| = 0 \quad .. \quad (6 \cdot 2)$$

The method of deriving the generalised variance statistic is to take a linear compound $\lambda_1 P_1 + \, .. \, + \lambda_r P_r$ of $P_1, P_2, .. P_r$ and maximise the statistic $\Sigma \lambda_i P_i / \sqrt{V(\Sigma \lambda_i P_i)}$. The statistic V is distributed as χ^2 with r degrees of freedom as shown in Rao (1945a, 1945b)

Lemma 1. *The generalised variance statistic V designed to test the hypothesis $L_1 T' = \xi_1, .., L_r T' = \xi_r$ concerning r independent parametric functions and derived as a root of the determinental equation obtained by the principle of maximisation is invariant under linear transformations of the hypothesis or the variates and is distributed as χ^2 with r degrees of freedom when the variances and covariances are known.*

When the variances and covariances are not known the tests can be performed only after studentising the above statistics. The general method of studentisation and the derivation of statistics in the case the stochastic variates are uncorrelated and have the same variance have been discussed here and the other cases are reserved for a subsequent communication. We may now state the problem as follows.

Given $Y=(y_1\ y_2 \ldots y_n)$ *the vector of n stochastic variables, such that* $E(Y)=TA'$, *where A is a known matrix with n rows and m columns and* $T=(\tau_1\ \tau_2 \ldots \tau_m)$ *is the vector of m unknown parameters subject to the restrictions* $G=TR'$ *where G is a row vector and R is a matrix with q rows and m columns, and further that* Λ *the dispersion matrix of Y is of the form* $\sigma^2 I$, *where* σ^2 *is unknown, the problem is to test the hypothesis whether the r independent parametric functions* $L_1T', \ldots, L_rT'$ *have the assigned values* $\xi_1, \ldots, \xi_r$. *Some of the parameters* τ*'s may be regression parameters in which case their coefficients are functions of concomitant variates.*

Following the procedure given above let $P=(P_1 \ldots P_r)$ be the vector of best estimates of $L_1T'-\xi_1, \ldots, L_rT'-\xi_r$. Since the stochastic variates are all uncorrelated the estimating vector P is independent of the unknown parameter σ^2. The dispersion matrix of P can in this case be written as $\sigma^2 D$ where the matrix D is completely known. The generalised variance statistic then comes out as

$$|P'P-V\sigma^2 D|=0 \qquad \ldots \quad (6\cdot3)$$

$$V=\frac{PD^{-1}P'}{\sigma^2} \qquad \ldots \quad (6\cdot4)$$

where D^{-1} is the inverse of D and the unknown prameter σ^2 appears only in the denominator.

If the y's form a normal system then V is distributed as χ^2 with r degrees of freedom when the null-hypothesis is correct. On the other hand if we can get the totality of linear functions of y's whose expectations are known independently of the null hypothesis then the generalised variance statistic associated with them is distributed as χ^2 with certain number of degrees of freedom. If $Q=(Q_1 \ldots Q_s)$ are s independent functions of y's whose expectations ars known to be $\zeta_1, \ldots, \zeta_s$ and if the dispersion matrix of Q's is F then the generalised variance statistic

$$V_0=\frac{(Q-\zeta)\,F^{-1}(Q-\zeta)'}{\sigma^2} \qquad \ldots \quad (6\cdot5)$$

is distributed as χ^2 with s degrees of freedom. If V_0 is distributed independently of V which is distributed as χ^2 with r degrees of freedom on the null hypothesis then a studentised statistic can be constructed by forming a function of V_0 and V which will be independent of σ^2 the simplest of which is the ratio V/V_0. The statistic sV/rV_0 is distributed as the variance ratio with r and s degrees of freedom respectively. The tests of significance can in this case be carried out with exactitude. We will now show that such an exact test is available on the maximum possible degrees of freedom for V_0 and the exact expressions for V and V_0 can be obtained by the simple device of least squares.

Theorem 6a. *If the ranks of A, R and* $(A'|R')$ *are s, k and p respectively then there are* $n+k-p$ *independent functions of the form* c_0+CY' *whose expectations are identically zeroes.*

Evidently $(s+k)\geqslant p$ and $d=(s+k-p)$ gives the number of independent rows of R which can be derived as linear combinations of the rows in A. If $E(c_0+CY')=0$ then $c_0+CAT'=0$ subject to the condition $G=TR'$ which shows that there exists a vector D such that

$$CA+DR=0 \text{ and } c_0=DG' \qquad \ldots \quad (6\cdot6)$$

Since the rank of A is s, there are $n-s$ independent vectors C which satisfy (6.6) when D=0. When D is non-null there are d vectors satisfying (6.6). This follows from the definition of

d given above. The value of c_o for each of the C's is given by DG'. Hence the number of independent functions whose expectations are zeros is

$$n-s+d=n+k-p \quad .. \quad (6\cdot7)$$

Theorem 6b. *If c_o+CY' is any function whose expectation is zero and b_o+BY' is the function giving the best unbiassed estimate of a parametric function LT' then $BC'=0$ or the two functions are uncorrelated.*

It has been shown elsewhere (Rao : 1945b) that the best estimates of parametric functions are obtained from the normal equations

$$YA=TA'A+SR \text{ and } G=TR' \quad .. \quad (6\cdot8)$$

where $S=(\sigma_1\ \sigma_2 \ldots \sigma_q)$ is a vector of q pseudo-parameters. If LT' is estimable then there exist two vectors χ_1 and χ_2 such athat

$$\chi_1A'A+\chi_2R=L \quad .. \quad (6\cdot9)$$

$$\chi_1R'=0,\ b_o=\chi_2G' \quad .. \quad (6\cdot10)$$

in which case the best estimate of LT' is given by b_o+BY' where $B=\chi_1A'$. If $E(c_o+CY')=0$ then from (6.6) we get $CA+DR=0$. Premultiplying by χ_1 we get

$$\chi_2A'C'+\chi_1R'D'=0=\chi_1A'C'=BC' \quad .. \quad (6\cdot11)$$

Hence the theorem 6b follows.

Theorem 6c. *If the best estimates of n parametric functions TA' are represented by $\bar{T}A'$ then the expectation vector of the residual vector $Y-\bar{T}A'$ is null and the totality of the linear functions of y's (not necessarily homogeneous) whose expectations are zeroes are derivable by linear combinations of the elements of the residual vector alone.*

Let $Y\Gamma'+\Delta$ (where Γ' is a matrix with n rows and m columns and Δ is a vector with n elements) give the best estimates of TA'. Since $E(Y-Y\Gamma'-\Delta)=TA'-TA'=0$, we get that the expectation of the residual vector $Y(I-\Gamma')-\Delta$ is null. By theorem 6b we also get that $\Gamma(I-\Gamma')=0$. It can be easily shown that the number of estimable functions for which the estimates are not pure constants is $s-d$ where s and d are as defined in theorem 6a, in which case it follows that the rank of Γ is $(s-d)$ and because of the relation $\Gamma(I-\Gamma')=0$ it follows that the rank of $(I-\Gamma')$ is $(n-s+d)$. Further any vector D such that $D(I-\Gamma')=0$ makes $D\Delta'=0$ for otherwise the expectation of the residual vector is not zero. Thus the residual vector $Y(I-\Gamma')-\Delta$ constitutes $(n-s+d)$ independent functions whose expectations are zeroes and by theorem 6a this gives the totality of such functions.

Theorem 6d. *The generalised variance statistic V_o derived from these $(n-s+d)$ independent functions whose expectations are zeroes is given by $[Y(I-\Gamma')-\Delta][(I-\Gamma)Y'-\Delta']/\sigma^2$ and is distributed as χ^2 with $(n-s+d)$ degrees of freedom independently of the functions giving the best estimates of parametric functions.*

Instead of taking the number of independent functions we can take a linear function $B[(I-\Gamma)Y'-\Delta')]$ of the m functions in $Y(I-\Gamma')-\Delta$ with its variance $B(I-\Gamma)(I-\Gamma')B'\sigma^2$ and maximise the statistic V_o given by

$$B[(I-\Gamma)Y'-\Delta'][Y(I-\Gamma')-\Delta]=V_o\,B(I-\Gamma)B'\sigma^2 \quad \ldots \quad (6\cdot12)$$

Differentiating with B we get

$$B[(I-\Gamma)Y'-\Delta'][Y(I-\Gamma')-\Delta]=V_o\,B(I-\Gamma)\sigma^2 \quad .. \quad (6\cdot13)$$

Multiplying the above expressions by themselves we get

$$V_o\,\sigma^2=[Y(I-\Gamma')-\Delta]\,[(I-\Gamma)Y'-\Delta'] \quad .. \quad (6\cdot14)$$

This is evidently distributed as χ^2 with $(n-s+d)$ degrees of freedom as it is the generalised variance statistic constructed out of $(n-s+d)$ independent functions. This is distributed independently of the best estimates by theorem 6b.

Theorem 6e. *The expression $V_0\sigma^2$ (which is independent of σ^2) is the same as the minimum value of $(Y-TA')(AT'-Y')$ when minimised with respect to the τ's subject to the condition $G=TR'$.*

The expression $V_0\sigma^2$ in (6.14) is the sum of squares of the residuals $(Y-\bar{T}A')(A\bar{T}'-Y')$ where $\bar{T}$ is a any solution of the normal equations which are obtained by equating the partial derivatives of $(Y-TA')(AT'-Y')$ with respect to τ's subject to the condition $G=TR'$ to zeroes. Hence the theorem.

The above theorem gives an easy method of evaluating $V_0\sigma^2$. Now to test the hypothesis $L_1T'=\xi_1$, $L_2T'=\xi_2, \ldots, L_rT'=\xi_r$ we have to calculate the statistic $V\sigma^2=PD^{-1}P'$ given in (6.4) and use the statistic $V\sigma^2(n-s+d)/V_0\sigma^2 r$ as the variance ratio with r and $(n-s+d)$ degrees of freedom. We may now observe that if $L_1T'=\xi_1 \ldots L_rT'=\xi_r$, then r more linear functions of y's have zero expectations so that the generalised variance appropriate to totality of independent functions, $(n-s+d+r)$, in number, can be obtained with the help of theorem 6e by finding the minimum value of $(Y-TA')(AT'-Y')$ subject to the conditions $G=TR'$ and the given relations $L_1T'=\xi_1, \ldots, L_rT'=\xi_r$. Let V_1 be its value. This generalised variance can also be obtained starting from the best estimates $P_1, P_2, \ldots, P_r$ of $L_1T'=\xi_1, \ldots, L_rT'-\xi_r$ and the residual vectors $Y(I-\Gamma')-\Delta$ which are all uncorrelated with P's due to theorem 6b. The generalised variance of the latter alone is V_0. If the generalised variance due to two uncorrelated sets of functions are additive then we infer that the generalised variance due to P's alone is V_1-V_0. We now prove the following theorems.

Theorem 6f. *If $Q_1=(q_{11}\, q_{12} \ldots q_{1r})$ and $Q_2=(q_{21}\, q_{22} \ldots q_{2m})$ are two sets of functions such that no function of the first set is correlated with those of the other and have dispersion matrices $\sigma^2\Lambda_1$ and $\sigma^2\Lambda_2$ then V, the generalised variance of $(Q_1|Q_2)$ is equal to the sum of V_1 the generalised variance of Q_1 and V_2 the generalised variance of Q_2.*

If L_1 and L_2 are the vectors of compounding co-efficients we get V as the solution of

$$L_1Q'_1Q_1+L_2Q'_2Q_1=VL_1\Lambda_1 \quad \ldots \quad (6\cdot15)$$

$$L_1Q'_1Q_2+L_2Q'_2Q_2=V\,L_2\Lambda_2 \quad \ldots \quad (6\cdot16)$$

from which by proper multiplications and addition we get

$$(V_1+V_2)\,(L_1Q'_1+L_2Q'_2)=V\,(L_1Q'_1+L_2Q'_2) \quad \ldots \quad (6\cdot17)$$

or

$$V_1+V_2=V \quad \ldots \quad (6\cdot18)$$

Theorem 6g. *The generalised variance V appropriate to the best estimates of $L_1T'-\xi_1, \ldots, L_rT'-\xi_r$ is obtained as the difference between the minimum values of $(Y-TA')(AT'-Y')$ when minimised with respect to the parameters of the vector T subject to the restriction $G=TR'$ and when minimised with the further restrictions $L_1T'-\xi_1=0, \ldots L_rT'-\xi_r=0$.*

The proof of this theorem follows from the results of theorem 6f and the discussion above.

This leads us to the following procedure in the theory of least squares. The observational equations are $E(Y)=TA'$ with the restrictions $G=TR'$. The hypothesis to be tested may be put as $H=TF'$. The hypothesis can be tested when and only when the system of equations $H=TF'$ are not inconsistent with $G=TR'$ and the matrix F' can be obtained from $(A'|R')$ by linear combination of its row or columns. If the minimum values of $(Y-TA')(AT'-Y')$ subject to $G=TR'$ and $G=TR'$, $H=TF'$ are denoted by $V_0\sigma^2$ and $V_1\sigma^2$

then V_0 and V_1 are distributed as χ^2 with $(n-s+d_1)$ and $(n-s+d_1+d_2)$ degrees of freedom respectively where d_1 is the number of independent vectors in R derivable from A and d_2 is the additional number independent vectors in F derivable from A and s is the rank A. If the ranks of A, R, $(A'|R')$; $(R'|F')$ and $(A'|R'|F')$ are s, k, p, r and t then $d_1=s+k-p$ and $d_1+d_2=s+r-t$. In any particular problem these degrees of freedom can be obtained from other simpler considerations. To test the hypothesis $H=TF'$ the statistic $\frac{V_1-V_0}{d_2} \cdot \frac{n-s+d_1}{V_0}$ is used as the variance ratio with d_2 and $(n-s+d_1)$. degrees of freedom.

The minimum values can be uniqualy obtained from

$$(Y-\bar{T}A')\,(A\bar{T}'-Y') \qquad \text{.. (6·18)}$$

where $\bar{T}$ is any solution from the equations obtained by equating the partial derivatives of $(Y-TA')(AT'-Y')$ subject to the restraining conditions to zero. We thus arrive at the general theory of least squares without making any restrictions on the number of variables or the parameters or the rank of the matrix of observational equations. This covers the case of independent stochastic variates whose expectations are linear functions of regression and certain other parameters, the coefficients of regression parameters being functions of concomitant variates.

Since the generalised variance statistic V_0 appropriate to s independent linear functions whose expectations are known independently of the hypothesis is distributed as χ^2 with s degrees of freedom it follows that $E(V_0)=s$. This gives us another method of calculating the number of such independent linear functions or the degrees of freedom associated with V_0. We need calculate only the expectation value of the minimised sum of squares $\sigma^2V_0=(Y-\bar{T}A')(Y'-A\bar{T}')$ where $\bar{T}$ is the vector of solutions from normal equations. If the parametirc functions TA' are estimated by $QF'+\Delta$ where $Q=YA$, F is a matrix with n rows and m columns and Δ is a row matrix with n elements then

$$\begin{aligned}\sigma^2V_0&=(Y-QF'-\Delta)\,(Y-FQ'-\Delta)\\ E(\sigma^2V_0)&=E(Y-\Theta)\,(Y'-\Theta')-E[QF'-E(QF')]\,[FQ'-E(FQ')]\\ &=\sigma^2(n-\text{trace of}AF')\end{aligned} \qquad \text{.. (6·19)}$$

where the trace of AF' is the sum of the diagonal elements of the matrix AF'. This follows from the fact that the variance of $CQ'+D$ the estimate of LT', a single parametric function is given by $LC'\sigma^2$. Hence we get the following theorem.

Theorem 6*h*. *If the parametric functions TA' are estimated by $QF'+\Delta$ then the degrees of freedom of V_0 the minimised sum of squares $(Y-TA')(Y'-AT')$ subject to assigned restrictions* is given by $(n-\textit{trace of } AF')$

7. COVARIANCE TECHNIQUE IN FIELD AND BIOLOGICAL EXPERIMENTS

As an application of the previous analysis we may consider the analysis of data from biological and field experiments. Certain functions of observations from these experiments have their expectations as linear functions of unknown parameters. The nature of the linear hypothesis involved in such problems and the construction of suitable functions as observational equations amenable to reduction by the method of least squares have been discussed elsewhere (Rao : 1944). The mathematical treatment of the question is carried out after a translation of the particular problem into mathematical language with special reference to field experiments.

The problem is to test whether one variety is superior to another in a certain specified region. We cannot utilise the whole of the specified region nor is it desirable to choose certain number of plots at random for one variety and a certain number for the other. Evidently two varieties cannot be tried one after the other on identical plots without vitiating the comparisons. So we choose sets of plots adjacent to one another which are called blocks. Let us consider a block of k plots and imagine that the yields on these k plots with respect to a hypothetical variety are $\alpha_1, \alpha_2, \ldots, \alpha_k$. We define the quantities

$$\beta=(\alpha_1+\alpha_2+\ldots+\alpha_k)/k, \quad \eta_i=\alpha_i-\beta, \quad \Sigma\eta_i=0 \qquad \ldots \quad (7\cdot1)$$

If any given variety is tried on the i-th plot then the observed yield y may be supposed to have the composition

$$y=\alpha_i+\tau+\epsilon=\beta+\tau+\eta_i+\epsilon \qquad \ldots \quad (7\cdot2)$$

where ϵ is a technical error due to experimentation and is such that its average over repeated experimentation under similar conditions is zero. The parameter τ may be called the *effect* of the variety. If $x_1, x_2, \ldots, x_s$ are s concomitant observations besides the yield y then the composition of y may be written, assuming linear regression on the concomitant variates as

$$y=\beta+\tau+\eta_i+\rho_1x_1+\ldots+\rho_sx_s+\epsilon' \qquad \ldots \quad (7\cdot3)$$

The parameter τ in (7.3) may be called the effect of the variety corrected for concomitant variation. It is clear from these definitions of effects that these are deviations from a hypothetical standard so that the superiority of one treatment (or variety) over the other can be gauged by a comparison of the effects alone. If, for any treatment of a given set of treatments, variations in the concomitant variates alone produce significant variations in the yield then the comparison of the treatments must involve the elimination of variations of yield due to differences in the values of concomitant variates alone.

If the position of the plot for which the given variety is to be tried is determined at random from the plots of a block then the observed yield y becomes a stochastic variate in which case for the observed set of concomitant variates $x_1, x_2, \ldots, x_s$

$$\begin{aligned}E(y)&=\beta+\tau+E(\eta_i)+\rho_1x_1+\ldots+\rho_sx_s+E(\epsilon')\\ &\beta+\tau+\rho_1x_1+\ldots+\rho_sx_s\end{aligned} \qquad \ldots \quad (7\cdot4)$$

for

$$E(\eta_i)=(\Sigma\eta_i)/k=0 \text{ and } E(\epsilon')=0$$

If $y_1, y_2, \ldots, y_k$ are observations on k varieties tried on k plots with *random determination* of position for the varieties and $x_{1i}, x_{2i}, \ldots, x_{si}$ are the values of the s concomitant variates corresponding to the i-th variate then

$$\begin{aligned}E(y_i)&=\beta+\tau_i+\rho_1x_{1i}\ldots+\rho_sx_{si}\\ i&=1, 2, \ldots, k\end{aligned} \qquad \ldots \quad (7\cdot5)$$

constitute the observational equations corresponding to the k observed yields. The variables y_i are evidently correlated with equal correlation for any two and have equal variance. From the set of equations (7.5) we can construct k other equations.

$$\begin{aligned}B&=(\Sigma y_i)/\sqrt{k}, \quad R_i=\Sigma l_{ij}y_j\\ i&=1, 2, \ldots, (k-1)\end{aligned} \qquad \ldots \quad (7\cdot6)$$

such that

$$\sum_j l_{ij}=0 \text{ and } \sum_j l_{ij}l_{rj}=0, \quad \sum_j l^2_{ij}=1 \qquad \ldots \quad (7\cdot7)$$

with the consequent relations

$$E(B)=(\Sigma\tau)/\sqrt{k}+\beta\sqrt{k}+\rho_1(\Sigma x_{1i})/\sqrt{k}+\ldots \qquad \ldots \quad (7\cdot8)$$

$$\begin{aligned}E(R_i)&=\Sigma\, l_{ij}\,\tau_j+\rho_1\Sigma l_{ij}x_{1j}+\ldots\\ i&=1, 2, \ldots, (k-1)\end{aligned} \qquad \ldots \quad (7\cdot9)$$

In this scheme all the variables are uncorrelated but B and R_i are subject to different variances. The variance to which R_i is subject may be called the *intra block error* denoted by σ^2_{ia}. For all possible random arrangements the variation in B is caused by ε, the technical error, alone which may be represented by σ^2_ε.

A design may be defined as an arrangement of v kinds of elements, in b sets of $k_1, \ldots, k_b$ elements such that i-th kind of element is used r_i times and the i-th and j-th kind of elements occur together in λ_{ij} sets. Such arrangements when available can be used in actual experimentation which involves the testing of v varieties, or more generally treatments in the case of factorial experiments, by choosing the sets as blocks. If we assume that the intrablock error is of the same magnitude irrespective of the fertility of the block (in which case it is intrinsically connected with size and number of plots in a block) then it is theoretically advantageous to choose blocks of equal size so that all comparisons from the experiment may be of equal weight. The elements are identified with v treatments and the treatments (elements) belonging to a set are assigned at random with in a block. We discuss below only as an illustration the particular case when there are b blocks with k plots and each of the v treatments is used r times. We may set down the observational equations as.

$$E(B_i)=\beta_i/\sqrt{k}+(\Sigma\tau)/\sqrt{k}+\ldots \qquad (7\cdot10)$$
$$i=1, 2, \ldots, b$$

corresponding to the observed block totals, β_i being the value of β for the i-th block and $\Sigma\tau$ is the sum of treatment effects from the i-th block and b sets of $(k-1)$ equations of the type (7.9) from each block giving altogether bk equations. Estimation and testing of hypothesis is concerned only with parametric functions of τ's so that all linear combinations of observations involving the block totals are not permissable. Hence we may take the $b(k-1)$ equations of the type (7.9) alone as observational equations. This can also be observed after forming the normal equations. We thus arrive at the problem of estimation and testing of hypothesis involving $b(k-1)$ independent stochastic variates with equal variance and having their expectations as linear functions of parameters.

Denoting the effect of the i-th variety by τ_i and the regression parameters by $\rho_1, \rho_2, \ldots, \rho_s$ the normal equations leading to the best estimates are (observing that under conditions (7.7) $\sum_i l_{ij}^2=(k-1)/k$ and $\sum_i l_{ij}\, l_{ir}=-1/k$)

$$Q_i(y) = \frac{r(k-1)}{k}\, t_i - \sum_j \frac{\lambda_{ij}}{k}\, t_j - \sum_p r_p Q_i(x_p) \qquad (7\cdot11)$$

and

$$S(yx_i) - \frac{S[y(b_j)\, x_i(b_j)]}{k}$$
$$=\Sigma t_p Q_p(x_i) + \Sigma\, r_p\left\{(S(x_p x_i) - \frac{S[x_p(b_j) x_i(b_j)]}{k}\right\} \qquad (7\cdot12)$$
$$i=1, 2, \ldots, s$$

where $Q_i(z)=$ sum of the observations on the z variate for the i-th treatment minus the means of the z variates for blocks in which the i-th treatment occurs, $S(u\,v)=$sum of products of the variables u and v, $w(b_s)=$total of the w variates from the s-th block. Observing that, as discussed earlier, the solutions t_i are given by

$$t_i = t_i^{(0)} - \Sigma r_p t_i^{(p)} \qquad (7\cdot13)$$

where $t_i^{(o)}$ are the solutions of

$$Q_i(y) = r\frac{k-1}{k} t_i^{(o)} - \sum_j \frac{\lambda_{ij}}{k} t_j^{(o)} \qquad .. \quad (7\cdot14)$$

$$i=1, 2, \ldots, v$$

and $t_i^{(m)}$ are the solutions of

$$Q_i(x_m) = r\frac{k-1}{k} t_i^{(m)} - \Sigma \frac{\lambda_{ij}}{k} t_j^{(m)} \qquad .. \quad (7\cdot15)$$

and substituting for t_i its expression $t_i^{(o)} - \Sigma r_p t_i^{(p)}$ in (7 ·12) we get

$$T_{oi} - V_{oi} = \Sigma r_p (T_{pi} - V_{pi}) \qquad .. \quad (7\cdot16)$$

$$i=1, 2, \ldots, s$$

where

$$V_{ij} = \Sigma t_p^{(i)} Q_p(x_j) = \Sigma t_p^{(j)} Q_p(x_i) \qquad .. \quad (7\cdot17)$$

$$\left.\begin{aligned} T_{oi} &= S(yx_i) - \frac{S[y(b_p)x_i(b_p)]}{k} \\ T_{ij} &= S(x_i x_j) - \frac{S[x_i(b_p)x_j(b_p)]}{k} \end{aligned}\right\} \qquad .. \quad (7.18)$$

Defining $E_{oi} = T_{oi} - V_{oi}$, $E_{ij} = T_{ij} - V_{ij}$ which may be called the error sum of squares and products, we get the equations giving the regression parameters as

$$E_{oi} = \Sigma r_p E_{pi} \quad (i=1, 2, \ldots, s) \qquad .. \quad (7\cdot19)$$

To get the estimates of t_i we use the formula (7 ·13) with the values of regression parameters as obtained from (7 ·19). As shown earlier

$$V(t_i) = V(t_i^{(o)}) + V(\Sigma r_p t_i^{(p)}) \qquad (7\cdot20)$$

$$\text{Cov } (t_i t_j) = \text{Cov } (t_i^{(o)} t_j^{(o)}) + \text{Cov } (\Sigma r_p t_i^{(p)}) (\Sigma r_p t_j^{(p)}) \qquad .. \quad (2\cdot21)$$

for each of which the expressions can be easily deduced following the discussion in the previous Sections.

To obtain the generalised variance due to error we take the minimised sum of squares

$$\Sigma[S(ly) - S(lt) - \rho_1 S(lx_1) - \ldots]^2 \qquad .. \quad (7\cdot22)$$

which on reduction comes out as

$$S(y^2) - \frac{S[y(b_p)^2]}{k} - \Sigma t_i^{(o)} Q_i(y)$$

$$-\Sigma r_p \left\{ S(yx_p) - \frac{S[y(b_i)x_p(b_j)]}{k} - \Sigma t^{(p)}{}_m Q_m(y) \right\}$$

$$= E_{oo} - \Sigma r_p E_{po} \qquad .. \quad (7\cdot23)$$

Let the degrees of freedom associated with this be d equal to b (k−1) minus the number of estimable independent parametric functions in τ's and ρ's. The generalised variance appropriate to test the hypothesis whether certain parametric functions have assigned values we minimise (7 ·22) subject to the restrictions imposed by the hypothesis and subtract the generalised variance due to error, the degrees of freedom associated with it being equal to the number of independent parametric functions whose values are assigned.

Thus, if the hypothesis to be tested is H=TF′ where the rank of F is r we have to minimise (7 ·22) subject to H=TF′ and construct the generalised variance statistic by subtracting (7 ·23) from it. The formula (7 ·23) and the previous analysis suggest that this minimum can be calculated with the help of the residual vectors. In this case we find the least square solutions of

$$\Sigma[S(ly)-S(l\tau)]^2 \qquad .. \quad (7\cdot24)$$

and

$$\Sigma[S(lx_i)-S(l\tau)]^2, \qquad i=1,2,\ldots,s \qquad .. \quad (7\cdot25)$$

subject to H=TF′ and calculate the residuals by substituting the solutions t's for τ's. Denoting the residuals by $[S(ly)-S(lt)]$ and $[S(lx_i)-S(lt)]$ where t stands for the corresponding solutions we can construct the following sum of squares and products.

$$E'_{oo}=\Sigma[S(ly)-S(lt)]^2=\Sigma S(ly)\,[S(ly)-S(lt)] \qquad .. \quad (7\cdot26)$$

$$E'_{po}=\Sigma[S(ly)-S(lt)]\,[S(lx_p)-S(lt)]$$
$$=\Sigma S(ly)\,[S(lx_p)-S(lt)] \qquad .. \quad (7\cdot27)$$

$$E'_{pi}=\Sigma S(lx_p)\,[S(lx_i)-S(lt)] \qquad .. \quad (2\cdot28)$$

The minimum value is given by $E'_{oo}-\Sigma r'_p E'_{po}$ where r'_p are solutions of $E'_{oi}-\Sigma r'_p E'_{pi}=0$. In the case of general observational equations E(Y)=TA′+RX′, of which the above example is a complicated case, we may state the rule as follows.

Lemma 7a. *The minimum value of* $(Y-TA'-RX')(Y'-AT'-XR')$ *subject to* $H=TF'$ *is given by* $\eta\eta'-\Sigma r_i\xi_i\eta'$ *where r's are any solutions of* $\xi_i\eta'=\Sigma r_p\xi_i\xi_p (i=1,2,\ldots,s)$ *and* η *and* ξ_i *are residual vectors obtained from the observational equations* $E(Y)=TA'$ *and* $E(X_i)=TA'$ *subject to the restristions* $H=TF'$.

As a particular case we may consider the testing of the hypothesis $\tau_1=\tau_2=\ldots\tau_v$. In this case the normal equations become independent of τ's. The minimised sum of squares is

$$\Sigma\,[S(ly)-\rho_1 S(lx_1)-\ldots]^2$$
$$=T_{oo}-\Sigma r'_p T_{po} \qquad .. \quad (7\cdot29)$$

where r_p' are solutions of

$$T_{0i}=\Sigma r'_p\, T_{pi} \qquad .. \quad (7\cdot30)$$

where T_{ij}'s are as defined in (7 ·18). The generalised variance due to (v−1) independent parametric functions whose values are assigned is given by (7 ·29)−(7 ·23). The test is supplied by the statistic d[(7 ·29)−(7 ·23)]/(v−1)(7 ·23) which is distributed as the variance ratio with (v−1) and *d* degrees of freedom. The application to special cases is reserved for a subsequent communication.

We thus arrive at a scheme of computation for the covariance technique. The expressions that are to be evaluated require a logical subdivision of the arithmetical discussion into two sections, *the analysis of variance* which alone is sufficient for tests of significance of varietal effects when the regression parameters are not of importance and the *analysis of covariance* which supplies the necessary correction terms leading to precise tests of significance when the regression parameters are significant. An additional advantage to be gained in this is that they give us in a simple manner not only the systematic scheme of computation which can be carried out merely as a routine but also the structure of the experiment itself.

8. COMBINATION OF WEIGHTED OBSERVATIONS

The general theory of least squares as discussed in Section 6 can be extended to the general case of observations having unequal variances but known proportions. In this case the variances $\sigma^2_1,\ldots,\sigma_n^2$ of the observations $y_1,\ldots,y_n$ can be expressed as $\omega_1^2\sigma^2,\ldots,\omega_n^2\sigma^2$

where $\omega_1, \ldots, \omega_n$ are known quantities and σ is an unknown parameter. Let W denote a Kronecker matrix with $1/\omega_1, \ldots, 1/\omega_n$ as diagonal elements. If from Y we construct the vector Z by the transformation Z=YW then the observational equations E(Y)=TA' transform to E(Z)=E(YW)=TA'W and the dispersion matrix of Z becomes $\sigma^2 I$ where σ is the unknown parameter defined above.

This problem can thus be reduced to the case already discussed. Instead of the variate y_i we consider the variate $z_i = y_i/\omega_i$ and in its expectation we divide the corresponding coefficients of the τ parameters by ω_i and proceed to the problem of estimation and testing of hypothesis with new variates and new matrix for the observational equations, the rank of which remains the same. The theory of weighted least squares can thus be reduced to the simple theory of least squares.

As an example, we may consider the following problem which occurs in the theory of regression. For a given value x_i of the concomitant variate x the stochastic variate y has the mean value βx_i and variance $\sigma^2 x_i$ so that the coefficient of variation of y is constant for all x arrays. For estimation and testing of hypothesis concerning the parameter β from pairs of observations $(x_1, y_1), (x_2 y_2), \ldots (x_n y_n)$, we proceed as follows. If $E(y_i) = \beta x_i$ are observational equations then they transform to $E(z_i) = \beta$ $(i = 1, 2, \ldots, n)$ where $z_i = y_i/x_i$. The normal equation leading to the best estimate of β is $\Sigma z_i = n\beta$, so that the estimate of β and its variance are $b = \Sigma z_i/n$ and σ^2/n respectively where σ^2 is estimated from the least squares $\Sigma(z_i - b)^2$ with (n−1) degrees of freedom. To test the hypothesis that the value of $\beta = \beta_0$, we use the statistic

$$t = (b - \beta_0)\sqrt{n} \; / \; \sqrt{\Sigma(z_i - z)^2 \,/\, (n-1)} \qquad \ldots \quad (8 \cdot 1)$$

and refer it to the t distribution with (n−1) degrees of freedom.

This treatment of linear regression may find an application in some biological and other problems where the mean value increases in a linear relation with the concomitant variate and the variance increases in such a manner that the coefficient of variation remains steady. As for instance the weight of dry paddy is linearly related to the weight of green paddy and the variance increases with increase in the green weight. The constancy of coefficient of variation is a plausible hypothesis in which case the above estimate of the regression coefficient is the most efficient and any hypothesis concerning the conversion factor may be tested as above.

9. The theory of statistical regression.

In this section some tests of significance connected with the regression coefficients are derived with the help of the results derived in Sections 4,5 and 6. The observations on the stochastic vector $Y = (y_1 \ldots y_n)$ are taken to be distributed normally and independently about their expectations with a common variance σ^2. The expectation vector is E(Y)= TA'+RX' where A, R and X are as defined before, with some restrictions on the τ parameters. Following the method of analysis outlined in Section 5 we can write down the equations (5 ·5) giving the regression parameters as

$$\Sigma\, \eta_i \xi_{ji} = \rho_1 \Sigma \xi_{1i} \xi_{ji} + \ldots + \rho_k \Sigma\, \xi_{ki} \xi_{ji} \qquad (9 \cdot 1)$$
$$j = 1, 2, \ldots, k$$

where the residual vectors ξ, η and the methods of calculation are given in (5 ·6) to (5 ·9). Due to lemma 7a we get the residual sum of squares when minimised with the whole set of parameters r's and ρ's as given by $\eta\eta' - \Sigma r_i \xi_i \eta'$ where r_i's are the solutions of (9·1). The degrees of freedom associated with this can be obtained with the help of theorem 6a or theorem 6h. Let this be f.

The hypotheses to be tested in the theory of regression are certain linear hypotheses concerning the ρ parameters. If in particular the hypothesis is $\rho_i = 0$, then it will be testing the significance of the partial regression coefficient of the stochastic variate on the i-th concomitant variate. If the hypothesis to be tested is $\rho_1 = \rho_2 = \ldots\ldots = \rho_k = 0$ (i.e. all ρ parameters are zeroes) it will be testing the significance of the multiple correlation coefficient.

Two methods are open to us. If the linear functions of the ρ parameters are estimable from the set of equations (9·1) then from their best estimates and variances and covariances, we can build up the appropriate generalised variance. If $P_1, P_2 \ldots P_s$ are the estimates of s independent linear functions with assigned values $p_1, p_2 \ldots p_s$ and $\sigma^2 d_{ij}$ is the covariance between P_i and P_j then the generalised variance with s degrees of freedom appropriate to P's is

$$\sigma^2 V = \Sigma\Sigma\, d^{ij}\, (P_i - p_i)\, (P_j - p_j) \qquad (9{\cdot}2)$$

where d^{ij} are the elements of the matrix reciprocal to $((d_{ij}))$. If we denote the residual sum of squares by $\sigma^2 V_0$ then to test the present hypothesis we use the statistic fV/sV_0 as the F-statistic with s and f degrees of freedom.

The alternative method is to find the residual sum of squares subject to the additional restrictions imposed by the hypothesis and get V by subtracting V_0 from the above residual sum of squares. In the particular case when the hypothesis is $\rho_1 = \rho_2 = \ldots = \rho_k = 0$ we immediately observe that $\eta\eta'$ is the residual sum of squares under this hypothesis. Subtracting the error residual sum of squares from this we get $\Sigma r_i \xi_i \eta'$ as the appropriate generalised variance with s degrees of freedom. Hence to test the significance of the multiple correlation between the stochastic variate and the totality of the concomitant variates we use the statistic $f\Sigma r_i \xi_i \eta' / k[\eta\eta' - \Sigma r_i \xi_i \eta']$ as the F statistic with k and f degrees of freedom.

If s linear functions of the regression parameters have some assigned values, then to find the residual we have to minimise with these extra restrictions. If these restrictions are

$$\begin{aligned} c_{11}\rho_1 + \ldots, + c_{1k}\rho_k &= d_1 \\ \ldots\ldots \quad & \ldots\ldots \\ c_{s1}\rho_1 + \ldots, + c_{sk}\rho_k &= d_k \end{aligned} \qquad (9.3)$$

then introducing Lagrangian multiplies $\lambda_1, \ldots, \lambda_s$ we get the normal equations leading to the estimates of ρ's as

$$\Sigma\, \eta_i\, \xi_{ji} = \rho_1\, \Sigma\, \xi_{1i}\xi_{ji} + \ldots + \rho_k \Sigma \xi_{ki}\xi_{ji} + \lambda_1 c_{1i} + \ldots + \lambda_s c_{si} \qquad (9,4)$$
$$j = 1, 2, \ldots k$$

and
$$d_i = c_{i1}\rho_1 + \ldots + c_{ik}\rho_k \qquad (9.5)$$

If $r_1', \ldots, r_k'$ is a solution to the above equations then the residual sum of squares is given by

$$\begin{aligned} &\Sigma\, (\eta_i - r'_1 \xi_{1i} - r'_2 \xi_{2i} - \ldots)^2 \\ &= \Sigma \eta_i^2 - \Sigma r'_i\, (2\Sigma \eta_j \xi_{ij} - \Sigma r'_s \Sigma \xi_{sk} \xi_{ik}) \end{aligned} \qquad (9.6)$$

The generalised variance appropriate to these s functions is obtained by subtracting the error residual sum of squares from this. If this quantity be denoted by $\sigma^2 V$ then the statistic fV/sV_0 which is distributed as the F statistic with s and f degrees of freedom can be used to test the above hypothesis. If the hypothesis is $\rho_1=0$ then in the equations (9·4) all λ's and ρ_1 become zeroes. If $r_1',\ldots,r_k'$ are the solutions then the sum of squares appropriate to this is given by $\Sigma r_s'\xi_s\eta' - \Sigma r_1\xi_1\eta'$ with 1 degrees of freedom.

The distribution of the statistics used above on the null-hypothesis involve only the degrees of freedom and no other nuisance parameters so that tests of significance can be carried with exactitude. Because of this fact the distribution of these statistics is independent of the distribution of the concomitant variates so that the same tests of significance hold good independently of the distribution of the concomitant variates provided the relative distribution of the stochastic variates for given sets of concomitant variates are normal.

REFERENCES.

1. Bose, R.C. (1944). The fundamental theorem of linear estimation *Proc. Ind. Sc. Cong.* 4(3).

2. David, F. N. and Neyman, J. (1939). Extension of Markoff's theorem on least squares. *Biom. Res. Mem.* 2,238-249

3. Fisher, R. A. (1922). The goodness of fit of regression formula and the distribution of regression coefficients. *J. R. S. S.* 85,597 612

4. ————(1925). Application of Student's distribution. *Metron* 3, 30-104

5. Kolodziejczyk, St. (1935). On an important class of statistical hypotheses. *Biom.* 27, 161-190.

6. Markoff, A. A. (1904). Calculus of Probability. Russian Edition.

7. Neyman, J. and Pearson, E.S. (1930). On the problem of two samples. *Bull de Ac. pol. des Sc et des let.* 73-96.

8. ————(1931). On the problem of k samples. *ibid* 460-482.

9. ————(1933). On the problem of most efficient tests of statistical hypotheses. *Phil. Trans. of Roy. Soc.* 231*A*, 289-337.

10. Rao, C.R. (1943). Thesis submitted to the Calcutta University for the M.A. degree.

11. ————(1944). On the linear set up leading to intra and interblock informations. *Science and Culture* 10,259-260.

12. ————(1945a). Generalisation of Markoffs' theorem and tets of linear hypotheses. *Sankhya* 7(1), 9-16.

13. ————(1945b). Markoff's theorem with linear restrictions on parameters. *Sankhya* 7(1), 16-19.

14. ————(1945c). Studentised tests of linear hypotheses. *Science and culture* 11.

15. "Student" (1908). The probable error of the mean. *Biom.*. 6, 1-25.

Paper received 7 December, 1944.