ON LINEAR ESTIMATION AND TESTING OF HYPOTHESIS

1. The object of the present note is to consider the most general problem in the theory of linear estimation and get suitable generalisations of Markoff's theorem and to derive suitable tests of significance connected with estimated functions.

2. There are n stochastic variates $T_1, T_2, \ldots T_n$ with the variance and covariance matrix $\Lambda = ((\lambda_{ij}))$ such that

$$E(T_i) = \theta_i = a_{i1}\tau_1 + a_{i2}\tau_2 + \ldots + a_{im}\tau_m \qquad (2\cdot 1)$$

and m not necessarily less than n. Some of the τ's may be regression constants in which case their coefficients will be functions of concomitant variates. We want a linear function $L(T) = b_1T_1 + b_2T_2 + \ldots + b_nT_n$ such that $E\{L(T)\} = \theta = l_1\tau_1 + l_2\tau_2 + \ldots + l_m\tau_m$, a given parametric function and $V\{L(T)\}$ is minimum. If L(T) can be found to possess such properties then it is called the best estimate of θ.

3. It has been shown that the set of equations

$$Q_i = L_i \Lambda^{-1} y' = \sum_{j=1}^{m} L_i \Lambda^{-1} L_j' \tau_j = h_i \qquad (3\cdot 1)$$

($i = 1, 2, \ldots m$), where L_i and y are the row matrices $(a_{1i}, a_{2i}, \ldots a_{ni})$ and $(T_1, T_2, \ldots T_n)$ and the dash represents their transpose obtained by minimising

$$L = \Sigma\Sigma\, \lambda^{ij}\, (T_i - \theta_i)\, (T_j - \theta_j), \qquad (3\cdot 2)$$

where λ^{ij} are the elements of Λ^{-1}, are such that Q_i is the best estimate of h_i and the best estimate of θ is given by $\Sigma\, c_i Q_i$ where c's are such that $\theta \equiv \Sigma\, c_i h_i$. If the rank of the matrix of the equations (3·1) is less than m, then all parametric functions are not estimable

4. The set of equations (3·1) possesses the following properties.

(i) Variance of Q_i is the coefficient of τ_i in the ith equation ;

(ii) Covariance of Q_i and Q_j is the coefficient of τ_j in the ith equation ;

(iii) If $\Sigma\, c_i Q_i \equiv 0$, then the parametric function $\Sigma\, c_i \tau_i$ is not estimable ;

(iv) If $\Sigma\, a_i Q_i$ and $\Sigma\, b_i Q_i$ are the estimates of $\Sigma\, l_i \tau_i$ and $\Sigma\, m_i \tau_i$, then their variances and covariances are given by

$$\begin{pmatrix} \Sigma\, al & \Sigma\, bl \\ \Sigma\, am & \Sigma\, bm \end{pmatrix}$$

(v) The number of estimable parametric functions is equal to the number of functionally independent Q's.

(vi) An intrinsic property is that all the above five properties hold good even with a subset of the equations obtained by eliminating one or more τ's.

(vii) The best estimate of any estimable parametric function $\Sigma\, l\tau$ is unique and is obtained by substituting any particular solution for τ's obtained by solving the equations (3·1) or any set of equations derived from it ; and so also the expressions for the variances and covariances.

5. If we want to test the hypothesis $\Sigma\, l\tau = \xi$ then we construct the statistic

$$w = \frac{\Sigma\, cQ - \xi}{\sqrt{\Sigma\, cl}},$$

where $\Sigma\, cQ$ is the estimate of $\Sigma\, l\tau$ and refer the normal tables for tests of significance. If we want to test the composite hypothesis $\Theta_i = \Sigma\, l_{ij}\tau_j = \xi_i, i = 1, 2, \ldots k$ we take a linear compound of these k relations

$$\Sigma\, \lambda_i\, (\Theta_i - \xi_i)$$

with the corresponding estimate

$$\Sigma\, \lambda_i\, (\Sigma\, c_{ij} Q_j - \xi_i) = \Sigma\, \lambda_i P_i$$

and variance $\Sigma \Sigma \lambda_i \lambda_j V_{ij}$ where $V_{ij} = \text{cov}(P_i P_j)$ and construct the above statistic

$$V = \frac{\Sigma \lambda_i P_i}{\sqrt{\Sigma \Sigma \lambda_i \lambda_j V_{ij}}}.$$

We choose the compounding coefficients such that V^2 is maximun. This leads to the determinental equation

$$| P_i P_j - V^2 V_{ij} | = 0.$$

The distribution of V^2 on the non-null hypothesis is obtained as

$$\text{Const} \cdot e^{-\frac{V^2}{2}} (V^2)^{\frac{k-1}{2}} I_{\frac{k}{2}} (V\phi)\, d V^2.$$

on the assumption that the y's form a multivariate normal system. The necessary statistics, when the variances and co-variances are not known are obtained by studentising the above statistics. Some of the important distributions will be discussed in a paper to be published in full elsewhere.

C. Radhakrishna Rao.

Statistical Laboratory,
Presidency College,
Calcutta.
April 4th, 1944.