A further Note on the Use of the K-Statistic for Testing a Certain Class of Hypotheses.

In a note published in the last issue of SCIENCE AND CULTURE,[1] I announced that I had defined and worked out the sampling distribution of certain 'Statistics' (the investigations being subsequently published in the current issue of *Sankhya*:[2] *the Indian Journal of Statistics*) which are designed to be appropriate tools for analysis of variance in multi-variate problems. For convenience I considered a problem closely allied and mathematically similar to the above problem of analysis of variance, which can be stated thus. Given two p-variate samples Σ and Σ' of sizes n and n' with dispersion matrices $||a_{ij}||$ and $||a'_{ij}||$ respectively, we want to test the hypothesis that the populations Π and Π' from which Σ and Σ' are supposed to have been drawn, have the same dispersion matrix, say, $||\alpha\ ||$. I briefly indicated how this could be done with the help of the sampling distribution (obtained under the assumption that the hypothesis to be tested is true) of certain members of the set of statistics K_i $(i=1, 2, \ldots p)$ defined by the following determinantal equation in K

$$|a_{ij} - K^2 a'_{ij}| = 0, (i, j = 1, 2, \ldots p) \quad .. \quad (1)$$

The K_i 's vary each from O to ∞; the distribution of the maximum statistic K_1 and the minimum one K_p were obtained respectively in the forms

$$\text{const } F(K_1) dK_1 \quad .. \quad .. \quad (1\cdot1)$$

and

$$\text{const } f(K_p) dK_p \quad .. \quad .. \quad (1\cdot2)$$

the general nature of this functions $F(K_1)$ and $f(K_p)$ were discussed in the note referred to above.

In the present note I propose to discuss in a general way some important additional points about the use of the set of statistics K $(i=1, 2, \ldots p)$ for purposes of answering two distinct types of questions connected with the testing of the hypothesis mentioned above.

(*i*) To accept the hypothesis *with safety* (on any level of significance) we use either K_1 and (1·1), or K_p and (1·2) according as K_1 is greater or less than $1/K_p$. The logic behind this procedure is that we look at the samples from a point of view which puts them farthest apart. If even from such a point of view we find that the difference is insignificant we are entitled to assert *with safety* that the difference is in fact insignificant (on a cerain level of course). But if this is significant we cannot really assert anything with safety.

(*ii*) To reject the hypothesis *with safety* on a given level of significance we note the statistic which or whose reciprocal is nearest to unity. Suppose it is K_r. Then it has been found that the distribution of K_r is of the general form

$$\phi(K_r) dK_r \quad .. \quad .. \quad .. \quad (1\cdot3)$$

where $\phi(K_r)$ is again a sum of terms each fo which is a product of incomplete Γ-functions. In the next issue of *Sankhya*[3] this will of course come out in a paper. The logic behind the step is that we now look at the samples from a point of view which puts them nearest to each other. If even from such a point of view the difference seems significant we can very well *safely* affirm that the difference is in fact really significant. For mathematical convinience a transformation to a new set of variates has been made for (1·1), (1·2) (1·3), the variates being defined by

$$K_i = e^{\lambda_i} \quad (i=1, 2, \ldots p) \quad .. \quad (1\cdot4)$$

In actual practice these new variates and their sampling distributions have been used in place of the K_i 's.

Statistical Laboratory, Presidency College, Calcutta, 15-8-39. — Samarendra Nath Ray.

[1] On the *K*-statistic, SCIENCE AND CULTURE, 5, 131, 1939.

[2] The *p*-statistic or certain generalisations in analysis of variance appropriate to multi-variate problems, *Sankhyā* 4, 3, 1939.

[3] On the Use and Sampling Distribution of the *K*-statistic, *Sankhyā*, 4, 4, 1939 (in the press).