FAMILIAL CORRELATIONS OR THE MULTIVARIATE GENERALISATIONS OF THE INTRACLASS CORRELATION

THE characters of individuals in a family or a group are influenced by two types of factors (i) the common factors which are characteristic of the group and (ii) the random factors which are independent of the group. The former type causes resemblance among the individuals of a group, while the latter brings about their variation. One of the problems, in the study of heredity, is to measure the strength of group characteristics. If we assume that there are a finite number of group factors effecting the characters of the individuals in a group, then suitable measures may be obtained by studying only a few characters for the individuals in a group. In this note, familial correlations, obtained by the multivariate generalisation of the intraclass correlation, have been introduced as suitable measures of the strength of group characteristics and their sampling distributions have been obtained.

2. Let x_{ijm} ($i = 1, 2, \ldots n$; $j = 1, 2, \ldots k$; $m = 1, 2, \ldots p$) represent the observation on the m-th character of the j-th individual in the i-th family. Following Fisher's technique, we replace the observations on the p characters of an individual by their linear combination and choose the compounding coefficients so as to maximise the intraclass correlation coefficient. If we maximise the function $t^2 = [1 + (k-1)r]/[1-r]$ instead of r, we get t as a root of the determinental equation.

$$| b_{rs} - t^2 w_{rs} | = 0 \qquad (2\cdot 1)$$

where
$$b_{rs} = \sum_i k\,(\bar{x}_{i.r} - \bar{x}_{..r})\,(\bar{x}_{i.s} - \bar{x}_{..s})$$

$$w_{rs} = \sum_i \sum_j (x_{ijr} - \bar{x}_{i.r})(\bar{x}_{ijs} - \bar{x}_{i.s})$$

and
$$k\,\bar{x}_{i.r} = \sum_j x_{ijr}, \text{ etc.}$$

There are, in general, p roots $t_1, t_2, \ldots t_p$ for the determined equation giving rise to p correlations $r_1, r_2, \ldots r_p$ which may be called the *familial correlations*. The following results have been obtained.

(a) The familial correlations are invariant under linear transformations of the variates.

(b) There exist familial correlations $\rho_1, \rho_2, \ldots \rho_p$, which are constants of the population obtained by employing the above principle to the population at large.

(c) The joint distribution of the familial correlations $r_1, r_2, \ldots r_p$, obtained from samples tends to the p variate independent normal distribution with

$$E(r_i) = \rho_i \text{ and}$$

$$V(r_i) = (1-\rho_i)(1 + \overline{k-1}\,\rho_i) / \sqrt{\tfrac{1}{2} n k (k-1)}$$

when the size of the sample (n) is increased.

(d) To carry on suitable large-sample tests the following transformations are suggested. Instead of r_i, we construct the statistic

$$z_i = \tfrac{1}{2} \log \frac{1 + \overline{k-1}\, r_i}{1 - r_i}$$

whose variance being $k/2\,(k-1)(n-2)$ is independent of ρ_i.

(e) To test whether the familial correlations are simultaneously zero we use the statistic $\Sigma\, z_i^2/V(z_i)$ which is distributed as χ^2 with p degrees of freedom in large samples.

(f) The exact sampling distribution of $t_1^2, t_2^2 \ldots t_p^2$ on the non-null hypothesis involves only $\tau_1, \tau_2, \ldots \tau_p$ defined by

$$\tau_i = \frac{1 + (k-1)\rho_i}{1 - \rho_i} \qquad (2\cdot 2)$$

and is given by

$$\text{const. } A \prod_{i,j=1}^{p} \frac{c_{ij}}{(\tau_i^2 + t_j^2)^{\frac{1}{2}}} \prod_{j=1}^{p} t_j^{n-p-1} \prod_{i=1}^{p} dt_i \qquad (2\cdot 3)$$
$$\times \left| \prod_{i<j} (t_i^2 - t_j^2) \right|$$

where A $\left[\sum_{i=p}^{p} \frac{D_{ij}^2}{(r_i^2 + t_j^2)} \right]^{\frac{nk-p-2}{2}}$

and D_{ij} is the coefficient of $c_{ij}/(r_i^2 + t_j^2)^{\frac{1}{2}}$ in the determinant $| c_{ij}/(r_i^2 + t_j^2)^{\frac{1}{2}} |$ and c_{ij}'s are defined as

$$c_{ij}^{2m} = 0, \quad c_{ij}^{2m+1} = 2^{\frac{2m+1}{2}} \Gamma\left(\frac{2m+1}{2}\right)$$

$$c_{ij}^{2m+1} \, c_{i'j'}^{2n+1} = 2^{\frac{m+n+1}{2}} \Gamma\left(\frac{2m+1}{2}\right) \times \Gamma\left(\frac{2n+1}{2}\right)$$

where $i \neq i'$ and $j \neq j'$

These substitutions are to be made only after expanding out A and multiplying it with other factor in (2.3). The distribution of r_1, r_2, ... r_p are obtained from (2.3) by making the transformations

$$t_i^2 = \frac{1 + (k-1)\, r}{1 - r_i} \quad (i = 1, 2, \ldots p) \qquad (2{\cdot}4)$$

(3) It is interesting to observe that the distribution (2.3) is similar to the distribution of the p-statistics of Roy (1942) on the non-null hypothesis. A fuller discussion of this subject will be attempted in a paper to be published in *Sankhya* shortly.

Statistical Laboratory,
Presidency College,
Calcutta, C. RADHA KRISHNA RAO.
November 18, 1944.

1. Roy, S. N., *Sankhya*, 1942, **6**, 16–34.