LETTERS TO THE EDITOR

STATISTICAL FORMULÆ

It is perhaps worth while for statistical computers to have some fresh formulæ in addition to the familiar ones found in the common text-books. These formulæ can at least serve to open up new methods of checking results obtained by older methods. Here is a variance formula for k varieties, $x_1, x_2, \cdots x_k$:

$$k^2\sigma^2=(k-1)\sum_{r=1}^{r=k}(x_r-m)^2$$

$$-\sum_{r=1}^{r=k}\sum_{s=1}^{r=k}(x_r-m)(x_s-m),\ (r\neq s) \qquad (1)$$

m being any number whatsoever.

It leads to the coefficient of intra-class correlation (R) for n families with varying number of members in each family, in a new form not noticed before:

$$R=1-\frac{\Sigma k_i^2\sigma_i^2}{\sigma^2\Sigma k_i(k_i-1)}, \qquad (2)$$

where k_i, σ_i^2 denote the number of members and the variance respectively in the i-th family and σ^2 is the general variance. The proof of this formula is immediate, if we set

$$P_i=\sum_{s=1}^{r=ki}(x_{ir}-m)(x_{is}-m),\ (r\neq s);$$

and
$$S_i=(k_i-1)\sum_{x=1}^{r=ki}(x_{ir}-m)^2,$$

where m is the general mean.

Then

$$R=\frac{\Sigma P_i}{\Sigma S_i}=1-\frac{\Sigma(S_i-P_i)}{\Sigma S_i}=1-\frac{\Sigma k_i^2\sigma_i^2}{\sigma^2\Sigma k_i(k_i-1)}.$$

It is readily seen that

$$R\geqslant -\Sigma k_i\sigma_i^2/\Sigma k_i(k_i-1)\sigma_i^2 \qquad (3)$$

Lastly, if we consider three groups of variates (x), (y), (z) arranged as in the following table:

	x_1	x_2	·	·	·	·	x_m	
y_1	z_{11}	z_{21}	·	·	·	·	z_{m1}	v_1
y_2	z_{12}	z_{22}	·	·	·	·	z_{m2}	v_2
·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·
y_n	z_{1n}	z_{2n}	·	·	·	·	z_{mn}	v_n
	u_1	u_2	·	·	·	·	u_m	

where $mv_i=z_{1i}+z_{2i}+\cdots+z_{mi}$, and $nu_j=z_{j1}+z_{j2}+\cdots+z_{jn}$, we can write down the mean (M) and the variance (σ^2), of mn variates of the form $ax_j+by_i+cz_{ji}$ ($j=1, 2, \cdots m$; $i=1, 2, \cdots n$), a, b, c being any arbitrary constants, thus:

$$M=a\bar{x}+b\bar{y}+c\bar{z}; \qquad (4)$$

$$\sigma^2=a^2\sigma_x^2+b^2\sigma_y^2+c^2\sigma_z^2+2r_{xu}\,ac\,\sigma_u\sigma_x + 2r_{yv}\,bc\,\sigma_v\sigma_y; \qquad (5)$$

in the usual notation of Statistics.

The proof follows by ordinary methods of expansion and summation. The last formula is useful in some genetic investigations.

Maharaja's College,
Mysore, A. A. Krishnaswami Ayyangar.
February 12, 1943.