On the Degree of Freedom Associated with the Sum of Squares.

By Motosaburō Masuyama.

Let $x_1, x_2, \cdots, x_N$ be a random sample of size N from the normal population with the population mean $m = 0$ and the finite population variance σ^2 under k linear homogeneous restrictions

(1) $\sum q_{\alpha i} x_i = a_\alpha$, $\alpha = 1, 2, \cdots, k (< N)$,

where $q_{\alpha i}$ are constants and the rank of matrix $(q_{\alpha i})$ is equal to k. Σ means the summation by the repeated latin suffix from 1 to N.

Let us suppose N-dimensional euclidean space R_N whose orthonormal fundamental vectors are $\{e_i\}$, and consider the following vectors, viz.

$$X = \Sigma X_i e_i,$$

(2) $q_\alpha = \Sigma q_{\alpha i} e_i, \quad \alpha = 1, 2, \cdots, k,$

$$p_\beta = \Sigma p_{\beta i} \cdot e_i, \quad \beta = k+1, k+2, \cdots, N.$$

Then the vectorial equation of (1) is

(3) $q_\alpha \cdot X = a_\alpha$

We shall select the constant vectors p_β in (2) as follows:

(3) $q_\alpha \cdot p_\beta = 0, \quad p_\beta \cdot p_\gamma = \delta_{\beta\gamma};$

where $\delta_{\beta\gamma}$ means the so-called Kronecker's delts.

As the rank of $(q_{\alpha i})$ is equal to k, the rank of the matrix

(4) $(W_{\alpha\beta}) = (q_\alpha \cdot q_\beta)$

is equal to k, and there exists its reciprocal matrix $(W^{\alpha\beta})$

If we put

(5) $\overset{*}{q_\alpha} = S W^{\alpha\beta} q_\beta,$

where S means the summation by the repeated greek suffix from 1 to k, then we obtain

(6) $\overset{*}{q}_\alpha \cdot q_\beta = \delta_{\alpha\beta}$, $\quad \overset{*}{q}_\alpha \overset{*}{q}_\beta = W_{\alpha\beta}$, $\quad \overset{*}{q}_\alpha p_\beta = 0$.

As the gramian

(7) $$\begin{vmatrix} \overset{*}{q}_\alpha \cdot \overset{*}{q}_\beta & \overset{*}{q}_\alpha \cdot p_\gamma \\ p_\varepsilon \cdot \overset{*}{q}_\beta & p_\varepsilon \cdot p_\gamma \end{vmatrix} = \begin{vmatrix} W_{\alpha\beta} & 0 \\ 0 & \delta_{\varepsilon\gamma} \end{vmatrix} = |W_{\alpha\beta}| \neq 0,$$

the N vectors $\overset{*}{q}_1, q_2, \cdots, \overset{*}{q}_k, p_{k+1}, p_{k+2}, \cdots, p_N$ are linearly independent.

Accordingly an arbitrary vector X in R_N can be represented uniquely as follows:

(8) $\quad X = S\, y_\alpha \overset{*}{q}_\alpha + S'\, z_\beta p_\beta$,

where S' means the summation by the repeated greek suffix from $k+1$ to N.

Multiplying q_α, p_β on X, we have

(9) $\quad y_\alpha = q_\alpha \cdot X = a_\alpha, \quad z_\beta = p_\beta \cdot X$.

Substituting these values in (8), we have

(10) $\quad G = (X - S\, a_\alpha \overset{*}{q}_\alpha)^2 = S'\, z_\beta^2$.

Lemma. The population mean of z_β^2 is equal to σ^2.

Proof. $E(Z_\beta^2) = E(\Sigma p_{\beta i} X_i)^2 = E(\Sigma\Sigma p_{\beta i} p_{\beta j} X_i X_j)$

$$= \Sigma\Sigma p_{\beta i} p_{\beta j} E(X_i X_j)$$

$$= \Sigma\Sigma p_{\beta i} p_{\beta j} \delta_{ij} \sigma^2$$

$$= \Sigma p_{\beta i}^2 \sigma^2 = \sigma^2 p_\beta^2 = \sigma^2$$

Theorem. The population mean of G is equal to $(N-E)\sigma^2$.

Proof. $E(G) = E(S' Z_\beta^2) = S' E(Z_\beta^2) = (N-E)\sigma^2$

Definition. $(N-k)$ is called the number of the degrees of freedom of the square G or of the unbiassed variance $V = \bar{G}/(N-k)$. This definition is more general than the one given by A. T. Craig,[1] and the theorem shows the explicit form of the sum of squares which is given by the scalar transcription of the square G. The applications of this theorem shall be published separately.[2]

1) A.T. craig: Certain Mean-Value problem in Statistics, Bull. Amer. Math. Soc., 42(1938), 670-674.

2) Tōkeisūrikenkyūsyo Kōkyūroku.