ON THE ANALYSIS OF VARIANCE IN CASE OF MULTIPLE CLASSIFICATIONS WITH UNEQUAL CLASS FREQUENCIES

By Abraham Wald[1]

Columbia University

In a previous paper[2] the author considered the case of a single criterion of classification with unequal class frequencies and derived confidence limits for σ'^2/σ^2 where σ'^2 denotes the variance associated with the classification, and σ^2 denotes the residual variance. The scope of the present paper is to extend those results to the case of multiple classifications with unequal class frequencies.

For the sake of simplicity of notations we will derive the required confidence limits in the case of a two-way classification, the extension to multiple classifications being obvious.

Consider a two-way classification with p rows and q columns. Let y be the observed variable, and let n_{ij} be the number of observations in the ith row and jth column. Denote by $y_{ij}^{(k)}$ the kth observation on y in the ith row and jth column ($k = 1, \cdots, n_{ij}$). Let the total number of observations be N. We order the N observations and let y_α be the αth observation on y in that order. Consider the variables:

$$t, t_1, \cdots, t_p, v_1, \cdots, v_q,$$

and denote by t_α the αth observation on t, by $t_{i\alpha}$ the αth observation on t_i and by $v_{j\alpha}$ the αth observation on v_j. The values of t_α, $t_{i\alpha}$ and $v_{j\alpha}$ are defined as follows:

$$\begin{aligned} t_\alpha &= 1 \ (\alpha = 1, \cdots, N), \\ t_{i\alpha} &= 1 \text{ if } y_\alpha \text{ lies in the } i\text{th row}, \\ t_{i\alpha} &= 0 \text{ if } y_\alpha \text{ does not lie in the } i\text{th row}, \\ v_{j\alpha} &= 1 \text{ if } y_\alpha \text{ lies in the } j\text{th column}, \\ v_{j\alpha} &= 0 \text{ if } y_\alpha \text{ does not lie in the } j\text{th column}. \end{aligned}$$

We make the assumptions

$$y_{ij}^{(k)} = x_{ij}^{(k)} + \epsilon_i + \eta_j,$$

where the variates $x_{ij}^{(k)}$, ϵ_i, η_j ($i = 1, \cdots, p; j = 1, \cdots, q; k = 1, \cdots, n_{ij}$) are independently and normally distributed, the variance of $x_{ij}^{(k)}$ is σ^2, the variance of ϵ_i is σ'^2, the variance of η_j is σ''^2, and the mean values of ϵ_i and η_j are zero.

---

[1] Research under a grant-in-aid from the Carnegie Corporation of New York.

[2] "A note on the analysis of variance with unequal class frequencies," *Annals of Math. Stat.*, Vol. 11 (1940).