# 36. Multidimensional Quantification. II

By Chikio HAYASHI
Institute of Statistical Mathematics, Tokyo

(Comm. by Z. SUETUNA, M.J.A., March 12, 1954)

The present paper is a continuation of [3] and deeply related to [1], [2]. The construction of problems, the meanings of symbols, terms and definitions etc. are the same as in [3]. For simplicity we call the method of [3] "case 1".

§ 1. Case 2, where elements are classified into $S$ strata by an outside criterion which is not unidimensional. Each element has response patterns in $R$ items and the label of the stratum to which it belongs. This label is called an outside variable. We use the same symbols as in [3]. In this case we consider the $R$-dimensional Euclidean space, each dimension of which corresponds to each item. Now we set $n$ orthogonal axes corresponding to the items. Each element will be represented as a point in this space if sub-categories in items are to be quantified. Now we should like to quantify the sub-categories $C_{tm}$ so as to maximize the effect of stratification. Here it is not at all effective to use the method of the case 1. As the total variance $\sigma^2$, we take the generalized variance which is considered to be proportionate to square of the volume of the so-called ellipsoid of concentration, where $\sigma^2=|\rho_{jl}\sigma_j\sigma_l|$, $\rho_{jl}\sigma_j\sigma_l$ is covariance between the $j$-th and the $l$-th item (dimension) when the sub-categories are quantified and $|\cdots|$ expresses a determinant, the element of which is $\rho_{jl}\sigma_j\sigma_l$ $j, l=1, 2,\dots, R$. As the within variance, we take $\sigma_t^2=|\rho_{jl}(t)\sigma_j(t)\sigma_l(t)|$, where $\rho_{jl}(t)\sigma_j(t)\sigma_l(t)$ is covariance between the $j$-th and the $l$-th item (dimension) in the $t$-th stratum, which is deeply related, in the above sense, to the ellipsoid of concentration in the $t$-th stratum.

Thus we take $\mu=1-(\sum_{t=1}^{S} p_t\sigma_t^2/\sigma^2)$ as a measure of efficiency of stratification, i.e. a measure of discriminative power of items which will be an index related to the efficiency of classification (success rate of prediction) by quantified behaviour patterns, where $p_t$ is a weight assigned to the $t$-th stratum, $\sum_{t=1}^{S} p_t=1$; especially in this case, we take $p_t=\frac{n_t}{n}$, $n_t$ is the size of the $t$-th stratum, $\sum_{t=1}^{S} n_t=n$. $\sum_{t=1}^{S} p_t\sigma_t^2$ is considered to be a sort of so-called within variance in the whole. If $\sigma_t^2=0$, $\mu=1$, if $\sigma_t^2=\sigma^2$, $\mu=0$. Besides this, we can take several indices as a measure. One example of these will be shown later on.