NOTES ON A PROBLEM OF MULTIPLE CLASSIFICATION*

Frederic M. Lord

EDUCATIONAL TESTING SERVICE

A solution is developed in implicit form for the problem of assigning N men to n jobs, the proportion of men to be assigned to each job being specified in advance.

Suppose that it is desired to assign N men to n jobs, the proportion of men to be assigned to each job being specified in advance. It is desired to maximize the average weighted productivity of the men, the productivity of each man being weighted according to the importance of the job to which he is assigned. It is assumed that the productivity of each man for each job is known in advance and can be used as a basis for assignment. If $x_{ia}$ is the productivity of man $a$ for job $i$, we can indicate the productivity of all men assigned to job $i$ by $\sum x_{ia}$. The quantity to be maximized is then

$$Q = w_1 \sum x_{1a} + w_2 \sum x_{2a} + \cdots + w_n \sum x_{na},$$

where $w_i$ is the weight assigned to job $i$.

A solution to an almost identical problem, assuming all weights to be unity, has been given by Hubert E. Brogden (1); these and related problems have more recently been treated by Thorndike (2); Votaw (3) has quite recently made important contributions toward a rapidly converging successive approximation method for the practical solution of such problems. The present paper is primarily concerned with developing in analytic form an implicit solution that is effectively the same as Brogden's; it is not immediately concerned with the problem of obtaining a practical solution by successive approximations.

**The Two-Dimensional Case**

Consider first the case when $n = 2$. The contribution of any individual to $Q$ will be $w_1 x_{1a}$ if he is assigned to job 1, $w_2 x_{2a}$ if he is assigned to job 2. The differential effect of the job assignment on $Q$ is in this case $w_1 x_{1a} - w_2 x_{2a}$. All individuals for whom the quantity $w_1 x_{1a} - w_2 x_{2a}$ has a given value are interchangeable with each other in their effect upon $Q$, and consequently are

*The author wishes to thank Dr. Hubert Brogden and Dr. Paul Horst for their helpful discussion and criticism.

†Dr. Paul S. Dwyer has recently developed important practical methods of solution, as yet unpublished.
interchangeable with each other for purposes of job assignment. Furthermore, if certain individuals characterized by some specified value of \( w_1x_{1a} - w_2x_{2a} \) are properly assigned to job 1, then all individuals with higher values must also be assigned to job 1. The same relation obtains between \( w_2x_{2a} - w_1x_{1a} \) and assignment to job 2.

Suppose a scatter diagram is plotted with \( x_1 \) and \( x_2 \) as axes, each individual being represented by a point corresponding to his values of \( x_1 \) and \( x_2 \). The reasoning just given shows that the optimum assignment of people to jobs corresponds to the division of the scatter diagram into two regions by the line \( w_1x_1 = w_2x_2 + k \), where \( k \) is a constant to be determined so that the required proportions of individuals fall in the two regions created.

We thus have the result for \( n = 2 \) that the optimum assignment corresponds to a region bounded by a straight line with a slope of \( w_2/w_1 \). This conclusion holds irrespective of the shape of the bivariate frequency distribution represented by the scatter diagram. The intercept of the line must be determined so that the proportions of cases cut off by the line are equal to the predetermined proportions of people to be assigned to the two jobs.

**The Three-Dimensional Case**

Let us next consider the case where \( n = 3 \). Using \( x_1 \), \( x_2 \), and \( x_3 \) as axes, a three-dimensional scatter plot may be prepared representing the data. The region of space containing the individuals to be assigned to job 1 will necessarily include all positive values on the \( x_1 \) axis above a certain point. Consider the surface bounding this region (region 1) from the region containing the individuals to be assigned to job 2 (region 2).

Suppose that the desired regions have already been set up. There will be of necessity certain individuals who could equally well be assigned to jobs 1 or 2, but who should not be assigned to job 3. Such individuals lie on the boundary surface between jobs 1 and 2. For such people, the \( x_3 \) score can have no effect on the decision as to whether they should be assigned to job 1 or to job 2. Consequently, the boundary between region 1 and region 2 is parallel to the \( x_3 \) axis. By the same reasoning used for the case where \( n = 2 \), it follows that this boundary must be represented by the equation \( w_1x_1 = w_2x_2 + k_{12} \) where \( k_{12} \) is a constant to be determined so as to obtain the proper proportion of people in each region. Similarly the boundaries between regions 1 and 3 and between regions 2 and 3 are respectively \( w_1x_1 = w_3x_3 + k_{13} \) and \( w_2x_2 = w_3x_3 + k_{23} \). In the present case these three equations define 3 planes, each of which is functionally independent of one of the variables and consequently parallel to the corresponding coordinate axis.

Let us see if any further conditions should be imposed on these planes. Let \( p_i \) be the proportion of cases to be assigned to job \( i \). Since \( \sum p_i = 1 \), only \( n - 1 = 2 \) of the values of \( p_i \) can be determined arbitrarily. Since the values of the \( k \)'s must be adjusted so that each region contains the proper