The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1596–1603

A SELECTION PROBLEM

BY M. MAHAMUNULU DESU[1]

State University of New York at Buffalo

1. Introduction and formulation of the problem. In many fields of research, one is faced with the problem of selecting the better ones from a given collection. We consider such a selection problem. We assume that there are k populations ($k \geqq 2$) populations $\Pi_1, \Pi_2, \cdots, \Pi_k$ at our disposal from which we want to select a subset. These may be varieties of a grain or some treatments or some production methods. The quality of the ith population is characterized by a real-valued parameter θ_i. The population with the largest θ-value is called the best population. A population is considered as a superior one if its quality measure does not fall too much below that of the best population. If $d(\theta_i, \theta_j)$ is a suitable distance measure between θ_i and θ_j and if $\theta_{\max} = \max(\theta_1, \theta_2, \cdots, \theta_k)$, population Π_i is

$$\text{superior (or good)} \quad \text{if} \quad d(\theta_{\max}, \theta_i) \leqq \Delta,$$
$$\text{inferior (or bad)} \quad \text{if} \quad d(\theta_{\max}, \theta_i) > \Delta,$$

where Δ is a given positive constant. It must be emphasized that this definition is different from the usual one considered in the literature [6], where θ_i is compared with θ_0, the quality measure of the standard or control population. Our definition is appropriate to situations where comparisons with a standard or control population are not possible. As pointed out by Lehmann [6], such a situation arises when a new product is being developed and one is interested in selecting the most promising of a number of production methods. In such cases each method must be compared with the totality of the remaining methods. A population is then considered superior if it does not fall too much below the best. In such cases our definition is a natural (or appropriate) one.

In some cases, it is reasonable to assume that whenever $d(\theta_{\max}, \theta_i) = \Delta$ one is indifferent towards branding Π_i as superior or inferior. In view of such cases, we may assume that there exist two positive constants Δ_1, Δ_2 (both, presumably, small compared with Δ) such that considering Π_i as inferior when $d(\theta_{\max}, \theta_i) \leqq \Delta - \Delta_1$ and considering Π_i as superior when $d(\theta_{\max}, \theta_i) \geqq \Delta + \Delta_2$.

Further it is of no serious consequence in whatever way one classifies Π_i when $\Delta - \Delta_1 < d(\theta_{\max}, \theta_i) < \Delta + \Delta_2$. In view of these remarks, we modify our previous definition as follows: A population Π_i is said to be

$$\text{superior (or good)} \quad \text{if} \quad d(\theta_{\max}, \theta_i) \leqq \delta_1^*, \tag{1}$$
$$\text{inferior (or bad)} \quad \text{if} \quad d(\theta_{\max}, \theta_i) \geqq \delta_2^*,$$

where δ_1^*, δ_2^* are specified constants such that $0 < \delta_1^* < \delta_2^*$.

With this modified definition of superior and inferior populations, we are

Received March 1, 1968; revised April 25, 1969.

[1] The author has published previously under the name D. M. Mahamunulu.