POST CLUSTER SAMPLING[1]

BY SAKTI P. GHOSH[2]

University of California, Berkeley

1. Introduction. The main difficulty often faced in cluster sampling is the lack of information relating to the composition of the clusters. In such situations the clusters can be built up on the basis of an initial random sample, then the final sampling can be done with these clusters as sampling units. Thus the name "Post Cluster", which essentially means that the clusters are formed afterwards and are not known beforehand, has been coined by T. Dalenius, who first introduced the idea in his book *Sampling in Sweden*, pp. 156–158.

The notion of forming groups of units (which may either be regarded as strata or clusters) after drawing an initial sample is due to Friedman and Wilcox. They queried Neyman at a Conference on Sampling Human Populations held in Washington, D. C., in April 1937 whether there was a solution to the problem of the optimum size of the initial sample and the smaller stratified sample selected from the initial sample. Neyman solved this problem subsequently and gave the results in a paper in 1938. In the Friedman-Wilcox method a certain number of units (indicated by the relevant optimum allocation theory) are sampled from every cluster (or strata); but in the method proposed in the paper a subset of entire clusters is selected. Later the Friedman-Wilcox method came to be known as phase sampling. David following Neyman (1938), among other things, gave a slightly more general treatment of the problem and also used the method of characteristic random variables. The common feature in the two methods is the regrouping of units after initial sampling according to some rule.

This technique of sampling differs from phase sampling in that it uses a hierarchy of sampling units, and differs from ordinary subsampling in that sampling units at the second stage are larger than the sampling units at the first stage.

The purpose of this paper is to develop a stochastic model for analysis of sampling problems that may arise in cluster sampling when the composition of the clusters is not at hand.

2. A stochastic model for the selection of the initial random sample. The type of sampling outlined above, will obviously depend on the "rule" which will be adopted in forming the clusters out of the sampled elements. We shall first consider the simplest situation where the clusters already exist in the population but the elements cannot be identified to the proper clusters until some auxiliary character is observed. The model behind the sampling procedure can