

ON THE THEORY OF SAMPLING FROM FINITE POPULATIONS

BY MORRIS H. HANSEN AND WILLIAM N. HURWITZ

Bureau of the Census

I—HISTORICAL BASIS FOR MODERN SAMPLING THEORY

The theory for independent random sampling of elements from a population where the unit of sampling and the unit of analysis coincide was developed by Bernoulli more than 200 years ago. The theory that would measure the gains to be had from introducing stratification into sampling was indicated by Poisson a century later. Subsequently, Lexis systematized previous work and provided the theoretical basis for sampling clusters of elements.¹ The adaptation of the work of Bernoulli and Poisson to sampling from finite populations was summarized by Bowley in 1926 [1] approximately a century after the work of Poisson.

An impetus to sampling advancement, following some fundamental statistical contributions of Pearson, Fisher, and others, resulted from the work of Neyman when he published his paper in 1934 on the two different aspects of the representative method [8]. In that paper he introduced new criteria of the optimum use of resources in sampling, including the concept of optimum allocation of sampling units to different strata subject to the restriction that the sample have a fixed total number of sampling units.

If, no matter how a sample be drawn, the cost were dependent entirely on the number of elements included in the sample, there would be little need for theory beyond the classical theories of Bernoulli and Poisson covering the independent random sampling of elements within strata, supplemented by the extension of the theory to finite populations, and the extension to optimum allocation of sampling units. Very often, however, in statistical investigations it is extremely costly, if not impossible, to carry out a plan of independent random sampling of elements in a population. Such sampling, in practice, requires that a listing identifying all the elements of the population be available, and frequently this listing does not exist or is too expensive to get. Even if such a listing is available, the enumeration costs may be excessive if the sample is too widespread. Frequently also, there are other restrictions on the sample design, such as the requirement that enumerators work under the close supervision of a limited number of supervisors, and as a consequence the field operations must be confined to a limited number of administrative centers. Techniques such as cluster sampling [2, 3, 4, 5, 6, 7, 8, 10], subsampling, and double sampling [9], have been

¹ The sampling of clusters of elements refers to the sampling of units that contain more than one element. Examples of cluster sampling include the use of the city block or the county as the sampling unit when the purpose of the survey is to determine the properties of the population made up of individual persons or individual households. In these instances, the city block or county is referred to as the cluster of elements, and the individual person or household is referred to as the element.