POLYCHOTOMY SAMPLING[1]

By Sakti P. Ghosh

Thomas J. Watson Research Center, New York

1. Introduction. In sampling, one sometimes has to deal with a particular type of binary character known as semi-observation-destructive (SOD) i.e., the action of observing the particular character in the unit may result in destroying the unit or leaving the unit unchanged. Such situations are encountered very often in quality control or biological experiments, etc. So long as one is interested in only one character of the unit, the situation does not present much of a problem; but when one is interested in more than one SOD character, the situation becomes more complex because no further observations can be made on units which are destroyed while a character is observed. Thus some combination of characters cannot be observed and yet an estimate of their proportion may be of interest. The aim of this paper is to illustrate how by splitting a single sample into a number of subsamples and then making different types of observations on different subsamples and by using the properties of Boolean algebra, estimates of all possible combinations of characters can be built up. This problem with two characters was first treated by Dalenius (1959). In that paper, the bivariate case was discussed in detail, and it was also pointed out that the design could be generalized to the case with three or more variates. The aim of the present paper is to derive the variances of the different estimates possible. In order to make the paper self-contained, an account is given of the content of Dalenius (1959), which is written in Swedish and thus not generally available. Here the bivariate situation will be treated first and the variances of the estimates given by Dalenius will be presented. Some other new estimates with their variances will also be presented. The sampling scheme will then be generalized to trivariate situations. The multivariate generalization will not be treated because it will be a simple extension of trivariate.

2. Bivariate sampling scheme and the estimates. Suppose A and B are two binary characters of a bivariate population (A, B), i.e., there are only four different types of units viz $(1, 1)$, $(1, 0)$, $(0, 1)$ and $(0, 0)$. When one character is being observed, the outcome will be denoted by $A(0)$ or $A(1)$ or $B(0)$ or $B(1)$ as the case may be. If $A(0)$ is observed, then the unit is destroyed and no further observation can be made on B. Similarly if $B(0)$ is observed then the unit is destroyed and no further observation can be made on A. The problem is to estimate the proportion of $(0, 0)$ in the population.

Suppose the total sample size available is n. n is divided into two parts n_1 and n_2, i.e., $n_1 + n_2 = n$. Then two samples of sizes n_1 and n_2 are drawn from the population without replacement. In the sample n_1, the character A is first ob-