A NOTE ON THE ESTIMATION OF SOME MEAN VALUES FOR A BIVARIATE DISTRIBUTION

By Edward Paulson[1]

Columbia University

In this paper two problems are discussed which were suggested by the theory of representative sampling [1], but which also occur in several other fields. The first problem is to set up confidence limits for $\frac{m_x}{m_y}$, the ratio of the mean values of the variates x and y. This comes up in the following situation. Let a population π consist of N units $x_1, x_2, \cdots x_N$ and suppose we wish to set up confidence limits for the mean $X = \frac{\sum_{i=1}^{N} x_i}{N}$. Also assume the population π has been divided into M groups, let v_j be the number of individuals in the j^{th} group and u_j be the sum of the values of x for the v_j individuals in the j^{th} group, so $X = \frac{u_1 + u_2 \cdots u_M}{v_1 + v_2 \cdots v_M} = \frac{Mm_u}{Mm_v}$. Now if a random sample of n out of the M groups is taken, yielding observations $(u_1, v_1), (u_2, v_2) \cdots (u_n, v_n)$ and N is unknown, the determination of confidence limits for X clearly becomes a special case of the first problem. The distribution of a ratio, discussed by Geary [2], does not seem to be well adapted for this purpose.

The second problem, which is of greater practical interest, arises when we again have a random sample $(u_1, v_1) \cdots (u_n, v_n)$ of n out of M groups and N and M are known. The standard estimate of X that has usually been made is $\hat{X} = \frac{M\bar{u}}{N}$, where $\bar{u} = \frac{\sum_{j=1}^{n} u_j}{n}$. This estimate does not utilize the fact that the n observations on v can be used to increase the precision of the estimate of the numerator of X. This is a special case of problem 2, which we can now formulate as how to best estimate m_x (the mean value of a trait x) both by a point and by an interval, when for each unit in the sample observations both on x and on a correlated variate y are obtainable, and m_y is known a priori. Situations of this type occur fairly often. It is possible to reduce the second problem to the first by using $\frac{\bar{x}}{\bar{y}} \cdot m_y$ as the estimate of m_x, and by multiplying the confidence limits for $\frac{m_x}{m_y}$ by m_y to secure limits for m_x, but this will not usually be the most efficient procedure.

In both problems two cases will be distinguished: (a) when σ_x^2, σ_y^2 and ρ are known a priori, and (b) when they are unknown. To determine confidence

[1] Work done under a grant-in-aid from the Carnegie Corporation of New York.