ON A LEAST SQUARES ADJUSTMENT OF A SAMPLED FREQUENCY TABLE WHEN THE EXPECTED MARGINAL TOTALS ARE KNOWN

By W. Edwards Deming and Frederick F. Stephan

1. Introduction. There are situations in sampling wherein the data furnished by the sample must be adjusted for consistency with data obtained from other sources or with deductions from established theory. For example, in the 1940 census of population a problem of adjustment arises from the fact that although there will be a complete count of certain characters for the individuals in the population, considerations of efficiency will limit to a sample many of the cross-tabulations (joint distributions) of these characters. The tabulations of the sample will be used to estimate the results that would have been obtained from cross-tabulations of the entire population.[1] The situation is shown in Fig. 1 in parallel tables for the universe and for the sample. For the universe the marginal totals $N_{i.}$ and $N_{.j}$ are known, but not the cell frequencies N_{ij}; for the sample, however, tabulation gives both the cell frequencies n_{ij} and the marginal totals $n_{i.}$ and $n_{.j}$.

In estimating any cell frequency of the universe, such as N_{ij}, three possibilities present themselves; from the sample one may make an estimate from the ith row alone, another from the jth column alone, and still another from the over-all ratio n_{ij}/n: specifically, the three estimates would be $n_{ij}N_{i.}/n_{i.}$, $n_{ij}N_{.j}/n_{.j}$, and $n_{ij}N/n$. As a result of sampling errors these will not be identical except by accident, and though any of them by itself may be considered accurate enough, still, if the whole $r \times s$ table of universe cell frequencies were so estimated, the marginal totals would not come out right. In this paper we present a rapid method of adjustment, which in effect combines all three of the estimates just mentioned, and at the same time enforces agreement with the marginal totals. The method is extended to varying degrees of cross-tabulation in three dimensions.

In any problem of adjustment where the conditions are intricate it is necessary to have a method that is straight-forward and self-checking; this becomes imperative when we realize that in the three-dimensional Case VII of the problem now at hand (*vide infra*), any adjustment in one cell must be balanced by adjustments in at least seven others. The method of least squares is one possible procedure for effecting an adjustment and at the same time enforcing certain conditions among the marginal totals. It is essentially a scheme for

[1] Examples will occur in the 1940 census publications. Further discussion of this problem and of the sampling procedure is given by the authors in "The sampling procedure of the 1940 population census," *Jour. Am. Stat. Assn.*, Vol. 35 (1940), pp. 615-630.