AN ITERATIVE METHOD OF ADJUSTING SAMPLE FREQUENCY TABLES WHEN EXPECTED MARGINAL TOTALS ARE KNOWN

By Frederick F. Stephan

Cornell University and U. S. Census Bureau

1. **Introduction.** In a previous paper by W. Edwards Deming and the author [1] the method of least squares was applied to the adjustment of sample frequency tables for which the expected values of the marginal totals are known. From observations on a sample the frequencies n_{ij} for the cell in the ith row and jth column of a two dimensional table and the r row and s column totals, $n_{i.}$ and $n_{.j}$, are obtained. These frequencies are subject to the errors of random sampling and it is desired to adjust them so that the row and column totals will agree with their expected values, $m_{i.}$ and $m_{.j}$, which are known. The adjustment involves the solution of the $r + s - 1$ normal equations

$$
\begin{aligned}
n_{i.}\lambda_{i.} + \sum_j n_{ij}\lambda_{.j} &= m_{i.} - n_{i.}, \qquad i = 1, 2, \cdots, r \\
\sum_i n_{ij}\lambda_{i.} + n_{.j}\lambda_{.j} &= m_{.j} - n_{.j}, \qquad j = 1, 2, \cdots, s - 1
\end{aligned}
\tag{1}
$$

where the λ are Lagrange multipliers from which are calculated the adjusted frequencies

$$m_{ij} = n_{ij}(1 + \lambda_{i.} + \lambda_{.j}). \tag{2}$$

Similar equations arise in the three dimensional case.

A method of iterative proportions was presented for effecting the adjustments more conveniently than by solving the normal and condition equations, and it was stated that "the final results coincide with the least squares solution." This statement is incorrect, for although the adjusted values satisfy the condition equations, they do not satisfy the normal equations and hence they provide only an approximation to the solution. The method of iterative proportions has several interesting characteristics that will be discussed in a later section. This paper now presents a method that converges to the values given by the least squares adjustment and is self correcting. It can be used with any set of data and weights for which a least squares solution exists. The two-dimensional case will be considered first.

2. **The two-dimensional case; expected row and column totals known.** Assume that a sample of n items is drawn at random and cross-classified in a table of r rows and s columns. As in the previous paper, let n_{ij} be the frequency in the ith row and jth column of the two-way frequency distribution. Indicate summation by substituting a dot for the letter over which the summation is to be performed. Then $n_{i.}$ and $n_{.j}$ are the marginal totals for the ith row and jth column respectively. Let $m_{i.}$ and $m_{.j}$ be the expected values of these