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

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

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

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

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

marginal totals calculated from other information or from theoretical considerations, and c_{ij} a set of constants known or estimated to be proportional to the reciprocals of the weights of the n_{ij}, i.e. proportional to their error variances. Since the weights are positive, the c_{ij} are non-negative and finite. It is assumed that the set of weights is such that for the given data an adjustment exists.

The least squares adjusted frequencies m_{ij} can be computed from the given numbers c_{ij}, n_{ij}, $m_{i.}$, and $m_{.j}$ by a series of approximate adjustments in a manner now to be explained. Let $m_{ij}^{(p)}$ be the pth approximation to m_{ij}. In conformity with this notation $m_{ij}^{(0)} = n_{ij}$. Let

$$(3) \quad d_{ij}^{(p)} = m_{ij} - m_{ij}^{(p)}, \qquad d_{i.}^{(p)} = m_{i.} - m_{i.}^{(p)}, \qquad d_{.j}^{(p)} = m_{.j} - m_{.j}^{(p)},$$

be corrections that must be added to the $m^{(p)}$ to produce the least squares adjusted frequencies. As $d \to 0$, $m^{(p)} \to m$. Let $\lambda_{i.}^{(p)}$ and $\lambda_{.j}^{(p)}$ be constants determined arbitrarily between the limits set by equations (5) to (7). Any one λ may be fixed arbitrarily and kept constant through successive approximations. Note that $\lambda_{i.}^{(0)} = \lambda_{.j}^{(0)} = 0$ and that, if at every step we set $\lambda_{.s}^{(p)} = 0$, the $\lambda^{(p)}$ are approximations to the Lagrange multipliers in the normal equations. After p steps in the iterative process the approximate adjusted frequencies will be

$$(4) \qquad m_{ij}^{(p)} = n_{ij} + c_{ij}(\lambda_{i.}^{(p)} + \lambda_{.j}^{(p)}).$$

The following conditions, derived from (19), (23), and (24), are sufficient to make the successive approximations converge to the least squares adjusted frequencies:

$$(5) \qquad \begin{aligned} \lambda_{i.}^{(p)} &= \lambda_{i.}^{(p-1)} + \theta_{i.}^{(p)} d_{i.}^{(p-1)}/c_{i.}, \\ \lambda_{.j}^{(p)} &= \lambda_{.j}^{(p-1)} + \theta_{.j}^{(p)} d_{.j}^{(p-1)}/c_{.j}, \end{aligned}$$

$$(6) \qquad 0 \le \theta_{i.}^{(p)}, \qquad 0 \le \theta_{.j}^{(p)}, \qquad \theta_{i.}^{(p)} + \theta_{.j}^{(p)} \le 2,$$

and, for at least one pair ij,

$$(7) \qquad \theta_{i.}^{(p)}(d_{i.}^{(p-1)})^2 + \theta_{.j}^{(p)}(d_{.j}^{(p-1)})^2 > 0; \qquad \theta_{i.}^{(p)} + \theta_{.j}^{(p)} < 2.$$

The θ's are introduced because in actual computations the successive approximations $\lambda^{(p)}$ can only be calculated to a limited number of digits and because the adjustment may progress more rapidly if the computer is permitted to use his judgment in determining the approximations as he observes the course of previous approximations.

The process of adjustment is continued until the $d_{i.}^{(p)}$ and $d_{.j}^{(p)}$ become sufficiently small to provide the desired degree of agreement between the adjusted and expected row and column totals.

3. Example. The following example shows the steps in the adjustment for a table of 3 rows and 4 columns with $\theta_{i.}^{(p)} \doteq \theta_{.j}^{(p)} \doteq 1$:

ij	n_{ij}	m_{ij}	$d_{ij}^{(0)}$	c_{ij}	$\lambda^{(1)}$	$m_{ij}^{(1)}$	$d_{ij}^{(1)}$	$d_{ij}^{(1)}/c_{ij}$	$\lambda^{(2)}$	$m_{ij}^{(2)}$	$d_{ij}^{(2)}$	m_{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
11	783	—	—	75	—	777.5	—	—	—	772.9	—	771
12	7426	—	—	455	—	7505.6	—	—	—	7496.3	—	7497
13	4709	—	—	358	—	4712.6	—	—	—	4709.6	—	4711
14	2145	—	—	176	—	2055.8	—	—	—	2051.3	—	2049
21	517	—	—	52	—	528.9	—	—	—	529.5	—	529
22	928	—	—	95	—	973.4	—	—	—	978.3	—	979
23	622	—	—	56	—	639.5	—	—	—	643.1	—	644
24	703	—	—	70	—	688.7	—	—	—	691.9	—	692
31	207	—	—	19	—	200.3	—	—	—	201.1	—	201
32	373	—	—	38	—	369.1	—	—	—	372.3	—	373
33	337	—	—	31	—	328.7	—	—	—	331.7	—	332
34	425	—	—	39	—	394.5	—	—	—	397.5	—	397
.1	1507	1501	−6	146	−.041	1506.7	−5.7	−.0390	−.0800	1503.5	−2.5	1501
.2	8727	8849	+122	588	+.208	8848.1	+0.9	+.0015	+.2095	8846.9	+2.1	8849
.3	5668	5687	+19	445	+.043	5680.8	+6.2	+.0139	+.0569	5684.4	+2.6	5687
.4	3273	3138	−135	285	−.474	3139.0	−1.0	−.0035	−.4775	3140.7	−2.7	3138
1.	15063	15028	−35	1064	−.033	15051.5	−23.5	−.0221	−.0551	15030.1	−2.1	15028
2.	2770	2844	+74	273	+.27	2830.5	+13.5	+.0495	+.3195	2842.8	+1.2	2844
3.	1342	1303	−39	127	−.31	1292.6	+10.4	+.0819	−.2281	1302.6	+0.4	1303
	19175	19175	0	1464	—	19174.6	+0.4	—	—	19175.5	−0.5	19175

Columns (1), (2) and (4) are given. Columns (3) and (6) to (11) are calculated in succession using equations (3), (4), and (5). It is not necessary in practice to record the θ's or even determine their values since the $\lambda^{(p)}$ may be determined directly at convenient values approximately equal to their corresponding $\lambda_{i.}^{(p-1)} + d_{i.}^{(p-1)}/c_{i.}$ and $\lambda_{.j}^{(p-1)} + d_{.j}^{(p-1)}/c_{.j}$. The final adjusted frequencies given in column (12) are derived by another repetition of the adjustment process but the amounts involved are so small that they can be calculated mentally and the n_{ij} rounded at the same time.

4. **Computing procedure.** The computing procedure may be set up in any of a number of ways to meet the preferences of the computer and the characteristics of the problem. Ordinarily it is desirable to make every number positive and the procedure as nearly routine as possible.

For two-dimensional adjustments the following procedure of computing alternately by columns and by rows is convenient:

(a) Set up a table of the c_{ij} in r rows and s columns. Enter the $c_{i.}$ in the $s+1$ column, the $c_{.j}$ in the $r+1$ row, and $c_{..} = \sum_i c_{i.} = \sum_j c_{.j}$ in the common cell.

(b) Calculate the quantities $A_{i.} \doteq (d_{i.}^{(0)}/c_{i.}) + a$ and $A_{.j} \doteq (d_{.j}^{(0)}/c_{.j}) + a$ and enter them in the $s + 2$ column and $r + 2$ row. The constant a is selected at some value that will make all quantities in the computations positive and may be any convenient integer greater than $2 \max | d_{i.}^{(0)}/c_{i.} |$ or $2 \max | d_{.j}^{(0)}/c_{.j} |$.

(c) Calculate the factors $\mu_{i.}^{(1)}$ approximately equal to the $A_{i.} - \frac{1}{2}a$ and enter each on its corresponding row in the $s + 3$ column. Throughout the computations the μ^{p} are merely $\lambda_{ij} + \frac{1}{2}a$.

(d) Take column j and multiply each c_{ij} by its corresponding $\mu_{i.}^{(1)}$ accumulating the products in the calculating machine. Divide the sum of products by $c_{.j}$, subtract the quotient from $A_{.j}$, and record the difference $\mu_{.j}^{(2)}$ in the jth column on the $r + 3$ row. Repeat for each of the other columns.

(e) Take row i and multiply each c_{ij} by its corresponding $\mu_{.j}^{(2)}$ accumulating the products in the calculating machine. Divide the sum of products by $c_{i.}$, subtract the quotient from $A_{i.}$, and record the difference $\mu_{i.}^{(3)}$ on the ith row in the $s + 4$ column bordering the table on the right. Repeat for each of the other rows.

(f) Repeat steps (d) and (e) alternately until a satisfactory degree of stability is reached in the $\mu_{i.}$ and $\mu_{.j}$. Then compute each adjusted frequency as follows:

$$m_{ij}^{(p)} = c_{ij}(\mu_{i.}^{(p)} + \mu_{.j}^{(p)} - a) + n_{ij}, \tag{8}$$

taking either $\mu_{i.}^{(p)} = \mu_{i.}^{(p-1)}$ or $\mu_{.j}^{(p)} = \mu_{.j}^{(p-1)}$ as the case may be.

(g) The computations may be checked at any step by computing

$$\sum_j \mu_{.j}^{(p)} c_{.j} = \sum_j A_{.j}c_{.j} - \sum_i \mu_{i.}^{(p-1)} c_{i.} = ac_{..} - \sum_i \mu_{i.}^{(p-1)} c_{i.}, \tag{9}$$

or

$$\sum_i \mu_{i.}^{(p)} c_{i.} = \sum_i A_{i.} c_{i.} - \sum_j \mu_{.j}^{(p-1)} c_{.j} = ac_{..} - \sum_j \mu_{.j}^{(p-1)} c_{.j}. \tag{10}$$

(h) At any step a constant may be added to all the $\mu_{i.}^{(p)}$ and subtracted from all the $\mu_{.j}^{(p)}$; this may be necessary to keep the μ's all positive. It has no effect on the value of a to be used in (8).

(i) If it is desired to "inflate" the adjusted frequencies $(\sum_{i,j} m_{ij} \neq \sum_{i,j} n_{ij})$ first multiply each n_{ij}, $n_{i.}$, and $n_{.j}$ by the factor $\sum_{i,j} m_{ij}/\sum_{i,j} n_{ij}$ and then proceed as above using the products in place of their corresponding n_{ij}, $n_{i.}$ and $n_{.j}$.

(j) If before the iterative process has reached an acceptable adjustment it is desired to force a satisfaction of the condition equations, compute:

$$m_{ij}^{(p+1)} = c_{ij}(\mu_{i.}^{(p)} + \mu_{.j}^{(p)} - a) + n_{ij} + (d_{i.}^{(p)}c_{.j} + d_{.j}^{(p)}c_{i.})/c_{..}, \tag{11}$$

in which either the $d_{i.}^{(p)}$ or the $d_{.j}^{(p)}$ are all zero.

5. **Adjustments in three dimensions.** If the sample is cross-tabulated in a three-way frequency distribution, there are two cases that are not reducible to

two-way distributions. These are designated Case III and Case VII in the earlier paper [1]. The adjustment equations are, respectively,

$$\begin{aligned} m_{ijk}^{(p)} &= n_{ijk} + c_{ijk}(\lambda_{i..}^{(p)} + \lambda_{.j.}^{(p)} + \lambda_{..k}^{(p)}) \\ m_{ijk}^{(p)} &= n_{ijk} + c_{ijk}(\lambda_{ij.}^{(p)} + \lambda_{i.k}^{(p)} + \lambda_{.jk}^{(p)}), \end{aligned} \tag{12}$$

subject to conditions on the choice of the λ corresponding to equations (5), (6), and (7). For Case III, the conditions are that

$$0 \le \theta_{i..}^{(p)}, \quad 0 \le \theta_{.j.}^{(p)}, \quad 0 \le \theta_{..k}^{(p)}, \quad \theta_{i..}^{(p)} + \theta_{.j.}^{(p)} + \theta_{..k}^{(p)} \le 2, \tag{13}$$

and for at least one triple ijk, $\theta_{i..}^{(p)}(d_{i..}^{(p-1)})^2 + \theta_{.j.}^{(p)}(d_{.j.}^{(p-1)})^2 + \theta_{..k}^{(p)}(d_{..k}^{(p-1)})^2 > 0$ and $\theta_{i..}^{(p)} + \theta_{.j.}^{(p)} + \theta_{..k}^{(p)} < 2$. Similar conditions apply to Case VII.

The computing procedure described in Section 4 can be extended readily to the three-dimensional case. For example, in Case VII calculate $\mu_{ij.}^{(1)}$ approximately equal to $(d_{ij.}^{(0)}/c_{ij.}) + \frac{1}{3}a$ and $\mu_{i.k}^{(1)}$ approximately equal to $(d_{i.k}^{(0)}/c_{i.k}) + \frac{1}{3}a$. Then multiply each c_{ijk} in the column jk by its corresponding $(\mu_{ij.}^{(1)} + \mu_{i.k}^{(1)})$ accumulating the products in the calculating machine. Divide the sum of the products by $c_{.jk}$ and subtract the quotient from $(d_{.jk}^{(0)}/c_{.jk}) + a$. Record the difference as $\mu_{.jk}^{(2)}$ and repeat the process for every other jk column. Take $\mu_{ij.}^{(2)} = \mu_{ij.}^{(1)}$ and repeat for each ik column to obtain $\mu_{i.k}^{(3)}$; then take $\mu_{.jk}^{(3)} = \mu_{.jk}^{(2)}$ and repeat for each ij column to obtain $\mu_{ij.}^{(4)}$ and so on. The final adjusted frequencies are

$$m_{ijk}^{(p)} = n_{ijk} + c_{ijk}(\mu_{ij.}^{(p)} + \mu_{i.k}^{(p)} + \mu_{.jk}^{(p)} - a). \tag{14}$$

6. **The general case.** The iterative method can be extended readily to more than three dimensions and to various systems of condition equations. A simple general notation may now be introduced. Let the cells be numbered in any order from 1 to t and for the ith cell let n_i be the value given by the sample, c_i a finite positive constant known or estimated to be inversely proportional to the weight of n_i, m_i the least squares adjusted value to be determined, $m_i^{(p)}$ the pth approximation to m_i, $d_i^{(p)} = m_i - m_i^{(p)}$, and $m_i^{(0)} = n_i$. Assume that the values m_σ of certain linear combinations of the m_i are given, i.e. there is a system of consistent linear equations of condition numbered in any order, the σth equation being

$$\sum_i b_{i\sigma} m_i = m_\sigma, \qquad \sum_i b_{i\sigma}^2 > 0, \tag{15}$$

$b_{i\sigma}$ and m_σ being known a priori. The corresponding linear combinations of the n_i and $d_i^{(p)}$ define

$$n_\sigma = \sum_i b_{i\sigma} n_i, \qquad d_\sigma^{(p)} = \sum_i b_{i\sigma} d_i^{(p)}. \tag{16}$$

Let

$$c_\sigma = \sum_i b_{i\sigma}^2 c_i. \tag{17}$$

The pth approximation to m_i is

$$(18) \qquad m_i^{(p)} = n_i + c_i \sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)},$$

where

$$(19) \qquad \lambda_\sigma^{(p)} = \lambda_\sigma^{(p-1)} + \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma, \qquad \lambda_\sigma^{(0)} = 0,$$

the $\theta_\sigma^{(p)}$, and therefore the $\lambda_\sigma^{(p)}$, being arbitrary for a finite number of steps, say p', but determined thereafter so that

$$(20) \quad 2 \sum_\sigma \theta_\sigma^{(p)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma)^2 \geq (d_\tau^{(p-1)})^2/(c_\tau H),$$

τ being a value of σ, chosen at the pth step, for which $(d_\sigma^{(p-1)})^2/c_\sigma$ is a maximum and H a finite number greater than 1 fixed prior to the first step as large as one will. That this condition can be satisfied may be shown by putting $\theta_\tau^{(p)} = 1$ and $\theta_\sigma^{(p)} = 0$ $(\sigma \neq \tau)$.

A weighted average of several of the possible selections of $\theta_\sigma^{(p)}$ satisfying (20) will also satisfy (20), positive "weights" being assumed. Let k added to the superscript represent the kth such selection and let $\alpha^{(p,k)} > 0$ be a constant for "weighting" the kth selection in the weighted average which may be chosen arbitrarily except that $\sum_k \alpha^{(p,k)} = 1$. Then, if the kth selection of $\theta_\sigma^{(p)}$ is represented by $\theta_\sigma^{(p,k)}$, the weighted averages are $\theta_\sigma^{(p,0)} = \sum_k \alpha^{(p,k)} \theta_\sigma^{(p,k)}$. Substitute them in the left-hand side of (20),

$$2 \sum_\sigma \theta_\sigma^{(p,0)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p,0)} d_\sigma^{(p-1)}/c_\sigma)^2$$

$$(21) \qquad = 2 \sum_\sigma \sum_k \alpha^{(p,k)} \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma \sum_k b_{i\sigma} \alpha^{(p,k)} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2$$

$$= \sum_k \alpha^{(p,k)} (2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma) - \sum_i c_i (\sum_k \alpha^{(p,k)} \sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)})/c_\sigma)^2,$$

which by the Cauchy-Schwarz inequality

$$\geq \sum_k \alpha^{(p,k)} (2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma)$$

$$- \sum_i c_i (\sum_k a^{(p,k)}) \{ \sum_k a^{(p,k)} (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2 \}$$

$$= \sum_k \alpha^{(p,k)} \{ 2 \sum_\sigma \theta_\sigma^{(p,k)} (d_\sigma^{(p-1)})^2/c_\sigma - \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p,k)} d_\sigma^{(p-1)}/c_\sigma)^2 \}$$

$$\geq \sum_k \alpha^{(p,k)} (d_\tau^{(p-1)})^2/(c_\tau H) = (d_\tau^{(p-1)})^2/(c_\tau H).$$

A simpler and more convenient but somewhat more restrictive condition may be derived as a special case of (20). Let $\theta_\sigma^{(p)} = 0$ except for a set of one or more σ so selected that $b_{i\sigma'} b_{i\sigma''} = 0$ for every i and every pair σ' and σ'' in the set. Then (20) becomes

$$\sum_{\sigma} \{2\theta_\sigma^{(p)} - (\theta_\sigma^{(p)})^2\} (d_\sigma^{(p-1)})^2/c_\sigma \geq (d_\tau^{(p-1)})^2/(c_\tau H). \tag{22}$$

Differentiating partially for a maximum with respect to one of the $\theta_\sigma^{(p)}$, we find that this special case of the condition will be satisfied if for one σ in the set, say π, such that

$$(d_\pi^{(p-1)})^2/c_\pi \geq (d_\tau^{(p-1)})^2/(c_\tau\sqrt{H}), \tag{23}$$

the value of $\theta_\pi^{(p)}$ is chosen in the range,

$$1/(2\sqrt{H}) \leq \theta_\pi^{(p)} \leq 2 - 1/(2\sqrt{H}) \tag{24}$$

and for every other σ in the set

$$0 \leq \theta_\sigma^{(p)} \leq 2, \tag{25}$$

all $\theta_\sigma^{(p)}$ not in the set being zero. A weighted average of such values of θ will satisfy (20) whence (6) and (7) follow.

In practice values of $\theta_\sigma^{(p)}$ satisfying (20) may be selected conveniently by the following procedure:

(a) Select a set of σ for at least one of which $\theta^{(p)}$ satisfies (23) and for every pair of which $b_{i\sigma'} b_{i\sigma''} = 0$. In so far as this restriction permits choose the σ corresponding to the larger values of $(d_\sigma^{(p-1)})^2/c_\sigma$.

(b) Determine values for each $\theta_\sigma^{(p)}$ in the set approximately equal to 1. Until other values are assigned to them assume all other $\theta_\sigma^{(p)} = 0$.

(c) Choose a σ not in the set, say ρ, for which $(d_\rho^{(p-1)})^2/c_\rho$ is relatively large and select a value for $\theta_\rho^{(p)}$ such that

$$\theta_\rho^{(p)} \doteq \{d_\rho^{(p-1)} - \sum_i \sum_{\sigma \neq \rho} c_i b_{i\rho} b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma\}/d_\rho^{(p-1)}. \tag{26}$$

(d) Having changed $\theta_\rho^{(p)}$ from 0 to a value approximately satisfying (26), continue with other σ not in the set letting ρ in (26) represent each in turn. The work may be terminated at any stage leaving some $\theta_\sigma^{(p)} = 0$.

7. **Convergence of the adjustment.** The condition equations may be written in the following form

$$\sum_i b_{i\sigma} d_i^{(0)} = d_\sigma^{(0)}, \tag{27}$$

as a system of consistent, but not necessarily independent, linear equations. They may also be written as conditions on the m_i. The least squares adjustment minimizes the quadratic form

$$S^{(0)} = \sum_i (d_i^{(0)})^2/c_i \tag{28}$$

subject to the restraints (27). Since the c_i are positive, $S^{(0)}$ is positive definite, and therefore a minimum exists and is non-negative. The values of the $d_i^{(0)}$ that minimize $S^{(0)}$ while satisfying (27) are $m_i - n_i$, the n_i being known and the m_i being the least squares adjusted values that are to be calculated.

If r is the rank of the matrix $\| b_{i\sigma} \|$, then from (15) and (16) it follows that r of the $d_i^{(0)}$ may be expressed as linear functions of the $t - r$ other $d_i^{(0)}$. The latter then constitute a set of $t - r$ independent variables. The normal equations

$$(29) \qquad \partial S^{(0)}/\partial d_h^{(0)} = 0,$$

are obtained by differentiating $S^{(0)}$ with respect to each one of them in turn, one equation resulting for each value of h corresponding to a d_i in the set of independent variables. The normal equations (29) are a system of $t - r$ independent linear equations and can be written in the form

$$(30) \qquad \Sigma\alpha_{i(h)}d_i^{(0)} = \Sigma\beta_{\sigma(h)}d_\sigma^{(0)},$$

where the first summation is over the set of independent variables, and the second over the $d_\sigma^{(0)}$ in the r selected condition equations. The right-hand terms are constants. Since a least squares adjustment exists the equations are consistent and the rank of the matrix $\| \alpha_{i(h)} \|$ is $t - r$. Any $d_i^{(0)}$ in the set, say $d_{i'}^{(0)}$, is the quotient of two determinants the divisor being the determinant $| \alpha_{i(h)} |$ and the dividend being the determinant obtained by replacing the $\alpha_{i'(h)}$ by $\sum_\sigma \beta_{\sigma(h)} d_\sigma^{(0)}$. Consequently each $d_i^{(0)}$ whether in the set or not is a linear combination of the $d_\sigma^{(0)}$ and the sum of the absolute values of the coefficients of the $d_\sigma^{(0)}$ is finite. Therefore

$$(31) \qquad \max | d_i^{(0)}/\sqrt{c_i} | \leq G \max | d_\sigma^{(0)}/\sqrt{c_\sigma} |$$

where G is $(\max c_\sigma/\min c_i)^{\frac{1}{2}}$ times the sum of the absolute values of the coefficients of the $d_\sigma^{(0)}$ in the linear combination for which such sum is a maximum. From (28)

$$(32) \qquad S^{(0)} \leq t \max \{(d_i^{(0)})^2/c_i\} \leq G^2 t \max \{(d_\sigma^{(0)})^2/c_\sigma\}$$

whence

$$(33) \qquad (d_\tau^{(0)})^2/c_\tau \geq S^{(0)}/(G^2 t).$$

Consider now the discrepancies

$$(34) \qquad d_i^{(p)} = m_i - m_i^{(p)} = d_i^{(p-1)} - c_i \sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)}/c_\sigma$$

between the m_i and the corresponding approximations $m_i^{(p)}$ and write the quadratic form

$$(35) \qquad S^{(p)} = \sum_i (d_i^{(p)})^2/c_i.$$

From (16), (18), and (34)

$$d_i^{(0)} = d_i^{(p)} + c_i \sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)}, \tag{36}$$

and

$$d_\sigma^{(0)} = d_\sigma^{(p)} + \sum_i \sum_v b_{i\sigma} b_{iv} c_i \lambda_v^{(p)}. \tag{37}$$

Hence the substitution of (36) in (27) merely changes (0) to (p) in the superscripts, the new equations being consistent by definition and the corresponding r of the $d_i^{(p)}$ being expressible as linear functions of the other $t - r$. Further (35) is positive definite and hence has a minimum, in fact substituting (36) in (28) we find that

$$\begin{aligned}\frac{\partial S^{(0)}}{\partial d_h^{(0)}} = \frac{\partial S^{(0)}}{\partial d_h^{(p)}} &= \frac{\partial}{\partial d_h^{(p)}} \{S^{(p)} + 2 \sum_i \sum_\sigma d_i^{(p)} b_{i\sigma} \lambda_\sigma^{(p)} + \sum_i c_i (\sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)})^2\} \\ &= \frac{\partial}{\partial d_h^{(p)}} (S^{(p)} + 2 \sum_\sigma d_\sigma^{(p)} \lambda_\sigma^{(p)}) = \frac{\partial S^{(p)}}{\partial d_h^{(p)}} = 0.\end{aligned} \tag{38}$$

Hence a least squares solution for the $d_i^{(p)}$ exists and it leads by (34) to the same values for the m_i as does the solution for the $d_i^{(0)}$. Since the coefficients $\alpha_{i(h)}$ and $\beta_{\sigma(h)}$ and the number G are functions of the $b_{i\sigma}$ and c_i they are invariant for the substitution. Consequently (30), (31), (32), and (33) may also be written with (p) in place of (0) in the superscripts (33) becoming

$$(d_\tau^{(p)})^2 / c_\tau \geq S^{(p)} / (G^2 t). \tag{39}$$

From (20), (34), and (35)

$$\begin{aligned} S^{(p)} &= \sum_i (d_i^{(p)})^2 / c_i \\ &= \sum_i (d_i^{(p-1)})^2 / c_i - 2 \sum_i \sum_\sigma d_i^{(p-1)} b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)} / c_\sigma \\ &+ \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)} / c_\sigma)^2 \\ &= S^{(p-1)} - 2 \sum_\sigma \theta_\sigma^{(p)} (d_\sigma^{(p-1)})^2 / c_\sigma + \sum_i c_i (\sum_\sigma b_{i\sigma} \theta_\sigma^{(p)} d_\sigma^{(p-1)} / c_\sigma)^2 \\ &\leq S^{(p-1)} - (d_\tau^{(p-1)})^2 / (c_\tau H), \quad p > p' \end{aligned} \tag{40}$$

and from (39)

$$S^{(p)} \leq S^{(p-1)} - S^{(p-1)} / M \leq S^{(p')} \{1 - 1/M\}^{p-p'} \tag{41}$$

where

$$M = G^2 H / t. \tag{42}$$

Therefore, as $p \to \infty$, $p - p' \to \infty$, $S^{(p)} \to 0$, $d_i^{(p)} \to 0$, $m_i^{(p)} \to m_i$ and consequently the successive adjusted frequencies obtained by an iterative process in which condition (20) is satisfied converge to the adjusted frequencies that are obtained by solving the normal equations.

8. **Rate of convergence.** The computer is not as much interested in the proof of convergence as he is in how rapidly the successive adjustments reach a satisfactory degree of approximation. Equations (39) or (41) are of no help to him. The adjustment may be made in one step, with every $\theta = 1$, (a) if the condition equations are such that every $b_{i\sigma'}b_{i\sigma''} = 0$, $\sigma' \neq \sigma''$, i.e. if the adjustment can be separated into one-dimensional cases when redundant condition equations are ignored, or (b), in the two and three-dimensional cases, if the c_{ij} or c_{ijk} are proportional to the $c_{i.}$ and $c_{.j}$ or to the $c_{i..}$, $c_{.j.}$, $c_{..k}$ or $c_{ij.}$, $c_{i.k}$, and $c_{.jk}$ respectively. Except in these and other special cases the rapidity of convergence depends on the $d_\sigma^{(0)}$ as well as on the $\| b_{i\sigma}c_i \|$ matrix. However, it seems that one can make very little use of the $d_\sigma^{(0)}$ to determine the rapidity of convergence without actually computing the successive adjustments or making some equivalent calculation.

Certain results can be obtained from the $\| b_{i\sigma}c_i \|$ matrix alone. Returning to the two-dimensional case and its notation, consider the matrix $\| c_{ij} \|$ and define

$$\delta_{ij} = c_{ij} - c_{i.}\, c_{.j}/c_{..}\,, \quad c_{..} = \sum_j c_{.j}\,. \tag{43}$$

Let the adjustments be made with the restriction that $\theta_{i.}^{(p)} = 0$ and $\theta_{.j}^{(p)} = 1$ when p is even, and $\theta_{i.}^{(p)} = 1$ and $\theta_{.j}^{(p)} = 0$ when p is odd. Then if $p > 1$

$$\begin{aligned} d_{i.}^{(p)} &= -\sum_j (c_{ij}/c_{.j})d_{.j}^{(p-1)} = \sum_j \sum_f (c_{ij}/c_{.j})(c_{ij}/c_{f.})d_{f.}^{(p-2)} \\ &= \sum_j \sum_f (\delta_{ij}/c_{.j})(\delta_{fj}/c_{f.})\, d_{f.}^{(p-2)} \qquad (f = 1, 2, \cdots, r) \end{aligned} \tag{44}$$

The sum of the absolute values is

$$\sum_i | d_{i.}^{(p)} | \leq b_1^2 \sum_i | d_{i.}^{(p-2)} | \leq b_1^{p-2} \sum_i | d_{i.}^{(2)} | \tag{45}$$

where

$$b_1^2 = \sum_i \sum_j | \delta_{ij}/c_{.j} | \, \gamma_{.j} \tag{46}$$

$\gamma_{.j}$ being the greatest of the $| \delta_{ij}/c_{i.} |$ in the jth column. Similarly for $p > 2$

$$\sum_j | d_{.j}^{(p)} | \leq b_2^2 \sum_j | d_{.j}^{(p-2)} | \leq b_2^{p-2} \sum_j | d_{.j}^{(1)} | \tag{47}$$

where

$$b_2^2 = \sum_i \sum_j | \delta_{ij}/c_{i.} | \, \gamma_{i.} \tag{48}$$

$\gamma_{i.}$ being the greatest of the $| \delta_{ij}/c_{.j} |$ in the ith column.

Assume again the conditions just preceding (44). Let $u_{i.}$ be the minimum $c_{ij}/c_{.j}$ in the ith row. Likewise let $v_{.j}$ be the minimum $c_{ij}/c_{i.}$ in the jth column. Then since $\Sigma d_{i.}^{(p)} = \Sigma d_{.j}^{(p)} = 0$,

$$\Sigma \mid d_{.j}^{(p)} \mid = 2\Sigma^{+} d_{.j}^{(p)} = -2\Sigma^{-} d_{.j}^{(p)}, \tag{49}$$

the + and − signs indicating that the last two summations are over positive and negative values of $d_{.j}^{(p)}$ respectively. When p is even, of course, all values of $d_{.j}^{(p)} = 0$.

From (44)

$$\begin{aligned} d_{i.}^{(p)} &= -\sum_j c_{ij} d_{.j}^{(p-1)}/c_{.j} = \sum_j^{-} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - \sum_j^{+} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j}, \\ &= \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - 2 \sum^{+} c_{ij} \mid d_{ij}^{(p-1)} \mid /c_{.j} \\ &= 2 \sum_j^{-} c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} \end{aligned} \tag{50}$$

and by (49)

$$\mid d_{i.}^{(p)} \mid \le \sum_j c_{ij} \mid d_{.j}^{(p-1)} \mid /c_{.j} - u_{i.} \sum_j \mid d_{.j}^{(p-1)} \mid, \tag{51}$$

$$\sum_i \mid d_{i.}^{(p)} \mid \le \sum_j \mid d_{.j}^{(p-1)} \mid (1 - \sum_i u_{i.}). \tag{52}$$

Similarly

$$\sum_j \mid d_{.j}^{(p)} \mid \le \sum_i \mid d_{i.}^{(p-1)} \mid (1 - \sum_j v_{.j}) \tag{53}$$

Let $b_3 = 1 - \sum_i u_{i.}$ and $b_4 = 1 - \sum_j v_{.j}$, then

$$\sum_i \mid d_{i.}^{(p)} \mid \le b_3 b_4 \sum_i \mid d_{i.}^{(p-2)} \mid \le (b_3 b_4)^{\frac{1}{2}p-1} \sum_i \mid d_{i.}^{(2)} \mid. \tag{54}$$

Now b_3 or b_4 may be greater or less than b_1 or b_2 but, unlike b_1 and b_2, they can not exceed unity. Let b^2 be the lesser of b_1^2 and $b_3 b_4$. Then under the conditions stated with equation (44)

$$\Sigma \mid d_{.j}^{(p+1)} \mid \le \Sigma \mid d_{i.}^{(p)} \mid \le b^2 \Sigma \mid d_{i.}^{(p-2)} \mid \le b^{p-2} \Sigma \mid d_{i.}^{(2)} \mid \le b^{p-2} \Sigma \mid d_{.j}^{(1)} \mid. \tag{55}$$

It follows from (40) that

$$\begin{aligned} S^{(p)} &= S^{(p+1)} + \sum_i (d_{i.}^{(p)})^2/c_{i.} + \sum_j (d_{.j}^{(p)})^2/c_{.j} \\ &= \sum_{h=p}^{\infty} \{\sum_i (d_{i.}^{(h)})^2/c_{i.} + \sum_j (d_{.j}^{(h)})^2/c_{.j}\} \\ &\le \sum_{h=p}^{\infty} \{(\sum_i \mid d_{i.}^{(h)} \mid)^2/\min c_{i.} + (\sum_j \mid d_{.j}^{(h)} \mid)^2/\min c_{.j}\} \\ &\le (\sum_i \mid d_{i.}^{(p)} \mid + \mid d_{i.}^{(p-1)} \mid)^2 \{(1/\min c_{i.}) + (1/\min c_{.j})\}/(1 - b^4). \end{aligned} \tag{56}$$

The reduction in $S^{(p)}$ in g steps of the iterative process is

$$(57)\quad \begin{aligned} D = S^{(p)} - S^{(p+g)} &= \sum_{h=p}^{p+g-1} [\sum_i (d_{i.}^{(h)})^2/c_{i.} + \sum_j (d_{.j}^{(h)})^2/c_{.j}] \\ &\geq \sum_{h=p}^{p+g-1} [(\sum_i | d_{i.}^{(h)} |)^2/(r \max c_{i.}) + (\sum_j | d_{.j}^{(h)} |)^2/(s \max c_{.j})]. \end{aligned}$$

from which, by (55), if $g > 1$ is odd,

$$(58)\quad D \geq \frac{1 - b^{-4g}}{1 - b^{-4}} (\sum_i | d_{i.}^{(p+g)} | + | d_{i.}^{(p+g+1)} |)^2 \left(\frac{1}{r \max c_{i.}} + \frac{1}{s \max c_{.j}}\right).$$

The relative decrease in $S^{(p)}$ is, therefore, by (56),

$$(59)\quad \frac{D}{S^{(p)}} = \frac{D}{D + S^{(p+g)}} \geq \left\{1 + \frac{1/\min c_{i.} + 1/\min c_{.j}}{b^4 (b^{-4g} - 1) \left(\frac{1}{r \max c_{i.}} + \frac{1}{s \max c_{.j}}\right)}\right\}^{-1}.$$

If the g steps actually have been taken a better lower limit for the relative decrease in $S^{(p)}$ may be obtained by computing D from (57) and using (56) for $S^{(p+g)}$. Similar equations can be written using b_2.

These results can be shown to be valid for an adjustment in which $\theta_{i.}^{(p)} = \theta_{.j}^{(p)} = 1$ at the first and any of the subsequent steps. They also can be extended to the three-dimensional cases but not to three-dimensional adjustments with every $\theta = 1$.

9. **Improvement resulting from the adjustment.** The least squares adjustment eliminates a portion of the errors of sampling, i.e. a portion of χ^2, from the set of frequencies estimated from the sample. In fact any adjustment that satisfies the condition equations does this.

Let ϵ_i be the error in the ith value given by the sample and $\delta_i^{(p)}$ the error in the pth approximation to the least squares adjusted value. Then

$$(60)\quad \delta_i^{(p)} = \epsilon_i + c_i \sum_\sigma b_{i\sigma} \lambda_\sigma^{(p)},$$

and

$$(61)\quad \sum_i (\delta_i^{(p)})^2/c_i = \sum_i \epsilon_i^2/c_i + 2 \sum_\sigma \lambda_\sigma^{(p)} \delta_\sigma^{(p)} - \sum_i c_i (\sum_\sigma b_\sigma \lambda_\sigma^{(p)})^2.$$

The complete adjustment makes $\delta_\sigma^{(p)}$ vanish and therefore, since the last term is non-negative, $\Sigma\delta_i^2/c_i < \Sigma\epsilon_i^2/c_i$ except in the trivial case in which all $d_\sigma^{(0)} = 0$. From (37)

$$(62)\quad \sum_i (\delta_i^{(p)})^2/c_i = \sum_i \epsilon_\sigma^2/c_i + \sum_\sigma \lambda_\sigma (d_\sigma^{(p)} - d_\sigma^{(0)}).$$

The last term may be computed readily at any stage in the iteration. If the sampling is at random, $k\ \Sigma\epsilon_i^2/c_i$ is distributed approximately as χ^2 with $t - 1$ degrees of freedom, where k is the ratio of the c_i to the corresponding error variances of the n_i. Therefore it would seem appropriate to compute $k\ \Sigma\lambda_\sigma\, d_\sigma^{(0)}$, the reduction in χ^2, as a measure of the improvement achieved in the final adjustment.

10. The method of iterative proportions. The iterative proportions method described in the earlier paper [1] implicitly defines, in the two dimensional case,

$$m_{ij} = \mu_{i.}\mu_{.j}n_{ij}\,, \tag{63}$$

the $\mu_{i.}$ and $\mu_{.j}$ being given by the $r + s$ condition equations

$$m_{i.} = \sum_j \mu_{i.}\,\mu_{.j}n_{ij}\,, \qquad m_{.j} = \sum_i \mu_{i.}\,\mu_{.j}n_{ij}\,, \tag{64}$$

any $r + s - 1$ of which constitute a consistent system of independent equations in $r + s$ unknowns. One multiplier, say $\mu_{1.}$, may be fixed arbitrarily. Then for a $2 \times s$ table it is necessary to solve an equation of the sth degree. If $s = 2$, there is only one acceptable solution, given by the positive root; if $s = 3$, there is only one solution of the cubic for which all the adjusted frequencies are non-negative. For 3×3 and larger tables the adjustment appears to involve the solution of equations of the tenth or higher degree and there is then no choice but to use methods of approximation.

The adjusted frequencies given by the method of iterative proportions are not identical to those given by the method of least squares. When the adjustments are small relative to the frequencies adjusted, however, the results given by this method approximate those of least squares. For the two-dimensional case the successive adjustments converge to a set of frequencies that satisfy the condition equations. The author has not found a proof of convergence or divergence for more than two dimensions.

I wish to express my appreciation of many stimulating conversations with Dr. W. Edwards Deming on this and related problems, and of the helpful critical reading of certain portions of the manuscript by Dr. Joseph F. Daly.

REFERENCE

[1] W. Edwards Deming and Frederick F. Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *Annals of Math. Stat.*, Vol. 11 (1940), p. 427.