No. 20

RODERICK J. A. LITTLE
and THOMAS W. PULLUM

The General Linear Model and Direct Standardization: A Comparison

OCCASIONAL PAPERS

AUGUST 1979

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office · Director: E. Lunenberg
428 Prinses Beatrixlaan, P.O. Box 950
2270 AZ Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35–37 Grosvenor Gardens
London SW1W OBS

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This publication is part of the WFS Publications Programme which includes the WFS Basic Documentation, Occasional Papers and auxiliary publications. For further information on the WFS, write to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, Voorburg, The Hague, Netherlands.

The General Linear Model and Direct Standardization: A Comparison

Reprinted from
Sociological Methods & Research,
Vol. 7, No. 4, May 1979, 475–501,
with permission of Sage Publications, Inc.

Prepared by:

RODERICK J. A. LITTLE,
World Fertility Survey,
and
THOMAS W. PULLUM,
University of Washington

A formal comparison is made between direct standardization and three cross-classified data structures: tables of means which are linear additive; tables of means which are log-linear additive; and tables of frequencies which are log-linear additive and can be converted to tables of proportions which are logit-linear additive. Standardization is an appropriate method of summarizing the data if the differences between standardized means and so on are not affected by the choice of standard distribution. This condition occurs when there is no interaction between the predictor and control variables in their impact on the dependent variable. It is shown that the condition may also be expressed in the form of the general linear model with the corresponding interaction terms absent. Then, when standardization is appropriate, differences between standardized quantities are estimates of differences between parameters in linear models. In some circumstances, e.g. when the cell sizes are small, if the specified interactions are believed absent then the cell entries may be fitted using the general linear model; standardization of the fitted entries would then be preferable to standardization of the observed entries.

THE GENERAL LINEAR MODEL AND DIRECT STANDARDIZATION

A Comparison

RODERICK J.A. LITTLE
World Fertility Survey
THOMAS W. PULLUM
Univeristy of Washington

The purpose of this article is to sharpen the distinction and to clarify the relationship between two alternative ways of analyzing a common data structure. The structure to be assumed is an array of rates, means, or proportions, cross-classified in a nonorthogonal design with two or more predictor or control variables. The frequencies or case bases in each cell are known, but the researcher may not have access to the within-cell sums of squares. (Of course, the sums of squares are calculable if the cell entries are proportions.)

The need to analyze such a structure is common, particularly in secondary analysis. For example, each Country Report issued through the World Fertility Survey (WFS) contains dozens of many-way tabulations of such a type, in which the cell entry is the mean number of children ever born, or the proportion wanting no more children, or the proportion using contraception, and so forth.

At one extreme of sophistication is some variant of direct standardization, such as the technique designated as Test Factor Standardization by Rosenberg (1962). Standardization has been used by demographers for decades and involves only simple calculations on a hand calculator. It is still occasionally used when more rigorous methods are accessible, as by Lieberson (1978) and by Clifford and Tobin (1977), but its use is easiest to justify when other methods are not available, as in many developing countries (Pullum, 1978).

At the other extreme, when computing facilities are at their best, many researchers would be inclined to use some variant of the General Linear Model (GLM) such as Multiple Classification Analysis. A discussion of these methods, some of which is repeated here, is given by Little (1978).

Our objective here is to reconsider the conditions under which standardization may be appropriate or inappropriate, and to reconsider the interpretation of its results, within the terminology of the general linear model. In no sense are we advocating increased use of standardization at the expense of more powerful models, but we seek to establish a formal linkage and to facilitate the correct use of a simple procedure.

In describing these approaches and in relating them we shall work with two main examples from the Fiji Fertility Survey, conducted in 1974. Table 2 gives the mean number of children ever born within categories of marital duration and education. This table is limited to Indian women, because there are pervasive ethnic differences in Fiji. Table 11 gives the proportion of Indian women who have ever used any efficient contraceptive method, within categories of current age, desire for more children, and education. For more details the reader may refer to the published report (Fiji, 1976) or to other WFS documentation.

STANDARDIZATION OF POPULATION QUANTITIES: IMPLICATIONS

Let us consider a two-way cross-classification of means. Denote the variables as A and B with category labels i running from 1 to I and j running from 1 to J, respectively. In this section we shall assume that population data are available, so that no sampling is involved. Let μ_{ij} be the mean, rate, or proportion in row i and column j, and let ν_{ij} be the population base frequency in that cell. Let $\nu_{i\cdot}$ be the marginal frequency for row i, and $\mu_{i\cdot}$ the mean for row i, using the usual dot notation. Thus

$$\mu_{i\cdot} = \sum_j \mu_{ij}(\nu_{ij}/\nu_{i\cdot}) \qquad (\text{where } \nu_{i\cdot} = \sum_j \nu_{ij}) \tag{1}$$

with weights $\nu_{ij}/\nu_{i\cdot}$ which sum to unity over j for all values of i.

Direct standardization with respect to variable B involves calculating row means with a new choice of weights $[\omega_j]$ which are the same for all values of i. Thus, the standardized mean for row i becomes

$$\mu_{i\cdot}(\omega) = \sum_j \mu_{ij}\omega_j, \tag{2}$$

with $\sum_j \omega_j = 1$. The set of weights $[\omega_j]$ is called the standard distribution and $\mu_{i\cdot}(\omega)$ represents the hypothetical mean that row i would have if B had this same distribution within each row.

A common choice of standard distribution is

$$\omega_j = \nu_{\cdot j} / \nu_{\cdot\cdot}, \tag{3}$$

which is the marginal distribution of B in the population. This is the form called Test Factor Standardization (TFS) and leads to standardized means

$$\mu^S_{i\cdot} = \sum_j \mu_{ij}(\nu_{\cdot j}/\nu_{\cdot\cdot}). \tag{4}$$

Another common choice of weights is $\omega_j = 1/J$ for all j, which leads to the unweighted row means.

As noted, the interpretation of the standardized means depends on the choice of standard distribution. However, often the analyst is more concerned with differences in the standardized means

$$\mu_{i.}(\omega) - \mu_{i'.}(\omega) = \Delta_{i,i'}(\omega) \quad \text{for } i \neq i' \qquad [5]$$

which represents the differences between levels of A when B has the same distribution $[\omega_j]$ within each level. Several authors, for example Kalton (1968) and Atchley (1969), have pointed out that the interpretation of these differences as the effects of A controlling B (or net of B) may be misleading when the differences are sensitive to the choice of standard distribution.

The artificial data in Table 1 (taken from Pullum, 1978) illustrate this phenomenon. In this table, when the effect of education on mean parity is controlled for marital duration using Test Factor Standardization, the difference between the low and high education groups is $(250 \cdot 4 + 250 \cdot 2)/500 - (250 \cdot 6 + 250 \cdot 2)/500 = -1.0$. That is, under this choice of standard distribution, higher-educated women would be said to have one child less than lesser educated women, controlling for marital duration. However, Table 1 shows that *within* levels of marital duration, the education effect is either 0.0 children (for low marital duration) or -2.0 (for high marital duration). Therefore the results of direct standardization will range from an education effect anywhere between 0.0 and -2.0, depending on choice of standard, is a misleading substitute for detailed examination of the table's interior.

Hence it is relevant to ask under what conditions the differences are the same for all choices of standard, that is,

$$\Delta_{i,i'}(\omega) = \Delta_{i,i'} \text{ for all } [\omega_j]. \qquad [6]$$

This condition is met if and only if the effects A and B are *linear additive,* according to the following definitions.

TABLE 1
Mean Parity of Women Having Specified Levels of Education and Marital Duration (hypothetical data)

		Education	
		Low	High
Marital Duration	Low	2.0 (50)	2.0 (200)
	High	6.0 (200)	4.0 (50)

NOTE: Base Frequencies in Parentheses.

Definition 1. The effects of A and B are linear additive if and only if one of the following conditions holds: (a) differences in the row means are the same for all columns j, that is,

$$\mu_{ij} - \mu_{i'j} = \Delta_{i,i'} \text{ for all j;} \tag{7}$$

(b) differences in the column means are the same for all rows i, that is,

$$\mu_{ij} - \mu_{ij'} = \Delta'_{j,j'} \text{ for all i;} \tag{8}$$

(c) there exist constants [μ, α_i, i = 1 to I, β_j, j = 1 to J] such that

$$\mu_{ij} = \mu + \alpha_i + \beta_j \text{ for all i and j.} \tag{9}$$

The equivalence of equations 7, 8, and 9 is well known and easily demonstrated. To show that equation 7 implies equation 6, note that if equation 7 holds then

$$\Delta_{i,i'}(\omega) = \sum_j \omega_j \mu_{ij} - \sum_j \omega_j \mu_{i'j} = \sum_j \omega_j \Delta_{i,i'} = \Delta_{i,i'} \text{ for all } [\omega_j].$$

To show that equation 6 implies equation 7, assume equation 6 and apply the particular set of weights $\omega_j = 1$, $\omega_j' = 0$ for $j \neq j'$ for each value of j in turn. We have thus established that equation 6 is equivalent to linear additivity. That is, differences in the standardized means can be intepreted unambiguously as effects of A net of, or controlling B if and only if A and B are linear additive. By contrast, if interaction terms γ_{ij} must be incorporated into equation 9, i.e., if A and B are not linear additive, then differences in standardized means will depend upon the magnitude of the interaction effects and the choice of weights. Such was the case for the hypothetical data in Table 1 discussed earlier.

There is one other situation where standardization can be used to calculate unambiguously the effects of A controlling B. Suppose that instead of calculating differences between the standardized means we calculate *ratios* of the standardized means $\delta_{i,i'}(\omega) = \mu_{i.}(\omega)/\mu_{i'.}(\omega)$, say. Alternative expressions of this quantity are $100\ [\delta_{i,i'}(\omega) - 1]$, the percentage difference between row i and row i', or $\log \delta_{i,i'}(\omega)$, which is the difference in the logarithm of the standaridized means of row i and row i', since $\log \delta_{i,i'}(\omega) = \log \mu_{i.}(\omega) - \log \mu_{i'.}(\omega) = \Delta_{i,i'}(\omega)$, say.

These quantities are useful representations of the net effect of A if they are constant for all choices of weights, $[\omega_j]$. This condition of invariance is met if and only if the effects of A and B are *multiplicative, or log-linear additive,* according to the following definition.

Definition 2. The effects of A and B are log-linear additive (or multiplicative), if and only if one of the following conditions holds: (a) differences in the logarithms of the row means are the same for all columns; that is,

$$\log \mu_{ij} - \log \mu_{i'j} = \Delta_{i,i'} \text{ for all } j; \qquad [10]$$

(b) differences in the logarithms of the column means are the same for all rows i; that is,

$$\log \mu_{ij} - \log \mu_{ij'} = \Delta'_{j,j'} \text{ for all } i; \qquad [11]$$

(c) there exist constants [μ, α_i, i = 1 to I, β_j, j = 1 to J] such that

$$\log \mu_{ij} = \mu + \alpha_i + \beta_j \text{ for all i and j.} \qquad [12]$$

It is easily shown, by analogy with the additive case, that log-linear additivity is equivalent to the invariance condition:

$$\log \mu_{i\cdot}(\omega) - \log \mu_{i'\cdot}(\omega) = \Delta_{i,i'} \text{ for all } [\omega_j]. \qquad [13]$$

The base of the logarithms in this definition is arbitrary. Equations 10 to 13 can be exponentiated to form a multiplicative analog to equation 9 and to show that if the effects of A and B are log-linear additive, the *ratios* between the standardized means of A are the same for all choices of standard and can be used to describe the effect of A controlling B.

SAMPLE ESTIMATION

So far we have considered only the structure of the population means of variables. We now apply these ideas to means from a sample of the population.

STANDARDIZATION OF SAMPLE MEANS

Suppose that we have a sample of size n_{ij} for the cell in row i, and column j, and an observed rate, mean, or proportion m_{ij} in that cell. The standardized mean for row i and standard distribution $[\omega_j]$ is

$$m_{i\cdot}(\omega) = \sum_j m_{ij}\omega_j. \qquad [14]$$

Test Factor Standardization corresponds to the choice of weights

$$\omega_j = n_j / n_{..} \cdot$$

The following results are relevant: (a) suppose m_{ij} is an unbiased estimate of μ_{ij}. Then the expected values of the differences in the standardized sample means of A (or B) are the same for all choices of standard distribution if and only if A and B are linear additive. (b) Suppose that m_{ij} is an unbiased estimate of μ_{ij}. Then the expected values of the ratios of the standardized sample means of A (or B) are approximately the same for all choices of standard distribution if and only if A and B are log-linear additive. The approximation involves replacing the expected value of ratios by the ratio of expected values.

The implications of (a) are as follows: for any cross-classification, the differences of the standardized means $m_{i.}(\omega)-m_{i'.}(\omega)$ will vary according to the choice of standard, ω. However, if A and B are linear additive, these differences are always unbiased estimates of the population difference, which represents the effect of A controlling B. The choice of standard affects the sampling variance of the estimate, and the results of applying different standards differ only because of sampling fluctuation. On the other hand, if A and B are not linear additive, the differences of standardized means estimate different quantities for each choice of weights. A similar interpretation of (b) is obtained by replacing differences by log differences or ratios in this statement.

Example 1. The data in Table 2, taken from the 1974 Fiji Fertility Survey, consist of a cross-classification of the Mean Number of Children Ever Born (or Mean Parity) by two factors: A = Education, with four levels (1 = No Education, 2 = Lower Primary, 3 = Upper Primary, 4 = Secondary or Higher), and B = Years Since First Marriage (or Marital Duration), with six levels (1 = 0-4 years, 2 = 5-9 years, 3 = 10-14 years, 4 = 15-19 years, 5 = 20-24 years, 6 = 25 or more years). The first entry in each cell is the mean parity m_{ij} and the second entry is the sample size n_{ij}.

Primary interest in the table concerns the relationship between education and fertility. It appears that much of the large differences in the raw mean parities between educational levels is attributable to the compositional effect of marital duration, that is, to the fact that better educated women tend to be younger and to marry later than less educated women. It is of interest to estimate the relationship between education and fertility after con-

TABLE 2
Mean Number of Children Ever Born Since First Marriage
(for ever-married women of Indian race)

	Years Since 0-4	First Marriage 5-9	10-14	15-19	(B) 20-24	25+	Mean
Educational Level (A) 1	.95[a)]	2.80	4.14	4.93	6.20	7.18	5.19
	82[b)]	93	118	151	160	288	892
2	.97	2.69	3.90	5.43	6.03	7.49	4.21
	150	184	202	159	111	131	937
3	.97	2.46	3.64	4.25	5.08	6.42	2.80
	188	149	99	63	48	31	578
4	.72	2.08	2.89	3.20	3.40	2.00	1.53
	149	64	36	10	10	1	270
Mean	.90	2.56	3.83	4.98	5.89	7.21	3.96
	569	490	455	383	329	451	2677

SOURCE: Fiji Fertility Survey, 1974.
a. First entry is mean number of children ever born.
b. Second entry is sample size.

trolling for marital duration. Hence we calculate means for each educational level standardized with respect to Marital Duration. We shall employ three standard distributions: (1) the marginal distribution of Marital Duration which corresponds to Test Factor Standardization; (2) the distribution giving the same weight to each level j of Marital Duration, which we call Equal Weights Standardization; and (3) the observed distribution of Marital Duration for the Secondary and Higher Educated women.

The weights for these distributions are given in Table 3 (rounded to two decimal places), and the standardized mean parities are given in Table 4. The last column of Table 4 gives the overall mean of the standardized education means, weighted by the total sample sizes in each education level. The low values for the third choice of standard reflects the prevalence of women with low marital durations and hence low parities among women with Secondary and Higher Education. When these means are sub-

TABLE 3
Weights for the Three Standard Distributions of Marital Duration (to be applied to Table 2)

	Marital Duration Category						
Standard	1	2	3	4	5	6	Total
1) Test factor	.21	.18	.17	.14	.12	.17	1.00
2) Equal Weights	.17	.17	.17	.17	.17	.17	1.00
3) Secondary and Higher	.55	.24	.13	.04	.04	.00	1.00

tracted from the standardized parities in the same row, we obtain the deviations given in parentheses. Result (a) implies that the deviations in Table 4 will have the same expectation for all standards if the effects of A and B are linear additive. However, we note that in this example these deviations appear to be sensitive to the choice of standard: for example, the women with secondary and higher education have 1.58 children less than the mean after TFS, but only .53 less than the mean after standardization with the marital duration distribution of the Secondary and Higher Education group.

Although no statistical test has been applied, this seems to suggest that the effects of A and B are not linear additive. In fact linear additivity can be discredited from theoretical considerations. The assumption implies that differentials in mean parity between education groups are the same for all levels of marital duration and this is clearly not appropriate; in the absence of premarital births, the mean parity of each group at marriage duration zero is zero, and hence differences are also zero. Differences between education groups emerge only as marital duration increases, and hence the effects of education and marital duration cannot be additive.

In contrast, the assumption that the effects are log linear additive, that is, *percentage* differences in mean parity between education groups are the same for all marriage durations, seems much more plausible as a working hypothesis. Accordingly we present

TABLE 4
Standardized Mean Parities (as a result of the application to Table 2 of the three standard distributions of marital duration)

Standard	Educational Level 1	2	3	4	Overall Mean (Weighted)
1) Test Factor	4.10 (.28)	4.14 (.32)	3.59 (-.23)	2.24 (-1.58)	3.82
2) Equal Weights	4.37 (.31)	4.42 (.36)	3.80 (-.26)	2.38 (-1.68)	4.06
3) Secondary and Higher	2.18 (-.12)	2.15 (-.09)	1.97 (-.09)	1.53 (-.53)	2.06

Deviations from the overall mean under each standard are given in parentheses.

the effects of education as percentage deviations of the standardized means from the overall standardized mean, with the results given in Table 5. It is clear that the percentage deviations are less sensitive to the choice of standard, thus lending support to this method of presentation and to the assumption of underlying log-linear additivity.

RELATIONSHIP TO MULTIPLE CLASSIFICATION ANALYSIS

We have noted that standardization has a particularly simple interpretation when the effects of A and B are linear additive. Thus in situations where this is plausible it makes sense to consider alternative estimates of the cell means μ_{ij} which exploit this additivity assumption. Recall that if A and B are linear additive then the population means can be written as $\mu_{ij} = \mu + \alpha_i + \beta_j$ for all i and j, for suitable choices of μ, α_i, and β_j. Hence we calculate estimates $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$ from the data and replace the sample means by fitted values

$$\hat{\mu}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \text{ for all i and j.} \qquad [15]$$

The constants $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$ are chosen so that the fitted values are as close as possible to the sample means. More precisely, they

TABLE 5
Standardized Mean Parities (when the standard distributions in Table 3 are applied to the logarithms of the parities in Table 2)

Standard	Educational Level 1	2	3	4	Overall Mean (Weighted)
1) Test Factor	4.10 (7.4)	4.15 (8.6)	3.59 (-5.9)	2.24 (-41.3)	3.82
2) Equal Weights	4.36 (7.5)	4.41 (8.7)	3.80 (-6.4)	2.38 (-41.4)	4.06
3) Secondary and Higher	2.18 (5.9)	2.15 (4.3)	1.98 (-4.1)	1.53 (-25.7)	2.06

NOTE: Percentage deviations from the overall means under each standard are given in parentheses.

are chosen so that the weighted sum of squares

$$SS = \sum_{i,j} n_{ij}(\hat{\mu}_{ij} - m_{ij})^2 \qquad [16]$$

is minimized. This procedure is the special case of additive analysis of variance called multiple classification analysis (MCA). It is optimal when the within cell variance is constant (of course, it will not be constant when the cell entry is a proportion, a case to be discussed below); a more general form is to minimize

$$\sum_{i,j} k_{ij} n_{ij}(\hat{\mu}_{ij} - m_{ij})^2$$

for some choice of constants k_{ij}.

This estimation of the cell means entails some extra computation. However, fitted values have certain advantages for the sample means, m_{ij}. (a) The fitted means can be calculated for cells with no observations; (b) the fitting process smooths the estimates for cells with small sample sizes, thus reducing the effect of sampling variance; (c) when within-cell sums of squares are known, the linear additivity assumption can be tested by an analysis of variance F-test, which essentially compares the minimized value of equation 16 with the average within-cell variance. Also, the statistical significance of the effects of A controlling B can be calculated.

Having obtained estimates of the cell means in this way, we can present the effects of A controlling B by standardizing the fitted values from equation 15. Then the standardized fitted values $\hat{\mu}_{i.}(\omega) = \sum_j \hat{\mu}_{ij}\omega_j$ clearly depend on the choice of standard distribution. However, the *differences* in the standardized fitted means are the same for any choice of standard, for

$$\hat{\mu}_{i.}(\omega) - \hat{\mu}_{i'.}(\omega) = \sum_j (\hat{\mu}_{ij} - \hat{\mu}_{i'j})\,\omega_j$$

$$= \sum_j (\hat{\alpha}_i - \hat{\alpha}_{i'})\,\omega_j = \hat{\alpha}_i - \hat{\alpha}_{i'} \quad \text{for all } \{\omega_j\}. \qquad [17]$$

Hence these estimates of the effects of A net of B are the same for any choice of standard, an attractive property not shared by estimates from the observed means.

Example 2. Consider again the data in Table 2 discussed in the previous section. Multiple classification analysis is *not* appropriate for this table since there are strong theoretical reasons for rejecting the assumption of linear additivity of effects. Despite this we shall apply it here for illustrative purposes. The fitted values are given in Table 6. The additive structure of the fitted values can be readily verified; for example, the difference between the first two rows is .03 for all columns (apart from some rounding). Using these fitted values, we obtain estimates of mean parity within educational levels, standardized by marital duration, as given in Table 7.

It is clear that the differences between the educational categories are the same for all choices of standard, as required by equation 17. Equivalently, the difference between each category and the overall (standardized) mean is the same for every choice of standard: .16, .19, −.27, and −.61 for educational levels 1, 2, 3, and 4, respectively.

RELATIONSHIP WITH LOG-LINEAR MODELS

We have noted that there are situations (such as the example above) when a linear additive model, with the consequent estima-

TABLE 6
Fitted Mean Parities for Table 2 (using MCA and the implausible assumption of linear additivity)

		Years Since First Marriage					
		0-4	5-9	10-14	15-19	20-24	25+
Educational Level	1	1.24	2.79	3.97	5.06	5.97	7.23
	2	1.27	2.81	4.00	5.09	6.00	7.26
	3	.81	2.35	3.53	4.63	5.54	6.79
	4	.46	2.00	3.19	4.28	5.19	6.45

tion of net effects as differences of standardized means, is not realistic but a log-linear additive model, with the consequent estimation of net effects as the ratios of standardized means, is appropriate. Under these circumstances it is natural to replace the cell means $m_{.j}$ by estimates of μ_{ij} which take the form

$$\log\hat{\mu}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \text{ for all i and j} \qquad [18]$$

where $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are chosen so that the $\hat{\mu}_{ij}$ are in some sense close to the sample means. The usual criterion is to minimize the chi-squared statistic

$$X^2 = \sum_{i,j} n_{ij} m_{ij} \log(\hat{\mu}_{ij}/m_{ij}). \qquad [19]$$

This is a variant of the generalized linear model, as discussed by Nelder and Wedderburn (1972). Note that despite formal similarities, the model equation 18 is conceptually distinct from the log-linear models for contingency tables developed by Goodman (1972), since here we are dealing with a cross-classification of means and not a two-way table of counts. Contingency tables will be discussed below.

As before, we can standardize these fitted values, forming $\hat{\mu}_{i.}(\omega) = \sum_j \hat{\mu}_{ij}\omega_j$. It is now clear that *ratios* of these quantities or differences in their logarithms are the same for all choices of

TABLE 7
Standardized Mean Parities: Three Standard Distributions in Table 3 Applied to the Fitted Data in Table 6

	Educational Level				Overall Mean
Standard	1	2	3	4	
1) Test Factor	4.12	4.15	3.69	3.35	3.96
2) Equal Weights	4.38	4.41	3.94	3.60	
3) Secondary and Higher	2.31	2.34	1.88	1.54	2.15

standard distribution. This follows from the equivalence of equations 12 and 13 applied to the fitted values $\hat{\mu}_{ij}$. Specifically,

$$\log \hat{\mu}_{i.}(\omega) - \log \hat{\mu}_{i'.}(\omega) = \hat{\alpha}_i - \hat{\alpha}_{i'} \text{ for all } [\omega_j],$$

so that the differences in the log standardized means are simply differences in the row parameters in equation 18.

If the effects are believed to be log-linear additive, the analyst is faced with a choice of presenting ratios of standardized means with or without smoothing by fitting the appropriate model. Fitting the model requires more work, but provides estimates for any empty cells, is statistically efficient and allows the assumption of log-linear additivity to be tested, for example by a chi-squared test on the minimized value of X^2.

Example 3. An additive log-linear model was fitted to the data in Table 2, using the computer package GLIM (General Linear Interactive Modeling). For those familiar with GLIM, this is achieved by specifying a Poisson error structure and a weight variable equal to the sample size in each cell. The fitted values are given in Table 8; the multiplicative structure can be verified by checking that ratios between rows are the same for all columns. (It is also worth noting that these fitted values are closer to the observed means in Table 2 than those obtained from the additive model in Table 6. This reflects the superiority of the log-linear

TABLE 8
Fitted Mean Parities for Table 2
(using the assumption of log-linear additivity)

		Years Since First Marriage					
		0-4	5-9	10-14	15-19	20-24	25+
Educational Level	1	1.02	2.75	4.00	5.09	6.03	7.24
	2	1.03	2.79	4.05	5.15	6.10	7.33
	3	.90	2.43	3.54	4.50	5.33	6.40
	4	.71	1.92	2.79	3.56	4.21	5.06

model for these data.) Using these fitted values, we obtain the estimates of mean parity within educational levels, standardized by marital duration, given in Table 9. It is readily verified that the percentage deviations of the standardized means from the overall mean are the same for all these standardizations: 5.6%, 6.8%, -6.8%, and -26.4% for educational levels 1, 2, 3, and 4, respectively. These estimates can be compared with the percent deviations derived from the observed table given above. The results from the Secondary and Higher standard are closest to those from the log-linear model, partly because the other standards appear to lead to an underestimate of the fertility of the secondary and higher educated group. The underestimate is caused by the low observed means for this group in the 20-24 and 25+ categories of Marital Duration, and the relatively high weights given to these cells in Test Factor and Equal Weights Standardization. One effect of fitting the model is to revise upwards the estimated fertility of these cells, resulting in an increase in the estimated fertility for this group from 41% below the mean (see above) to 26% below the mean.

It should be noted that the model fitting process also provides statistical evidence on the fit of the model and the significance of the effects. The log-linear additive model yielded a chi-squared of 16.07 on 15 degrees of freedom, indicating that the model fits the data. A log-linear model assuming no effect of Education yielded a chi-squared of 82.16 on 18 degrees of freedom, highly significant, indicating that this model does not fit the data. The conclu-

TABLE 9
Standardized Mean Parities: Three Standard Distributions in Table 3 Applied to the Fitted Data in Table 8

Standard	Educational Level 1	2	3	4	Overall Mean
1) Test Factor	4.09	4.14	3.62	2.86	3.88
2) Equal Weights	4.36	4.41	3.85	3.04	4.13
3) Secondary and Higher	2.19	2.21	1.93	1.53	2.07

sions are that the effect of Education, controlling for Marital Duration, is statistically significant, and the effect of Marital Duration and Education can be considered additive on the log-linear scale. Confidence limits for the percentage differences in parity between education levels can also be derived.

MULTIWAY CROSS-CLASSIFICATIONS OF MEANS

The results thus far are easily generalized to tables of any dimension. To simplify the notation we shall consider only three-way tables of means; the extension to higher order tables will be clear from this case.

Consider a three-way cross classification of means. Denote the variables as A, B, and C with category labels i running from 1 to I, j running from 1 to J and k running from 1 to K. Let μ_{ijk} be the population mean in row i, column j and panel k, and let ν_{ijk} be the population base frequency for that cell.

Let

$$\{\omega_{jk},\ j = 1 \text{ to } J,\ k = 1 \text{ to } K,\ \sum_{j,k} \omega_{jk} = 1\}$$

be a standard distribution over the factors B and C. Then the corresponding standardized mean for row i is

$$\mu_{i..}(\omega) = \sum_{j,k} \mu_{ijk}\omega_{ijk}.$$

Differences in the (log) standardized means between two rows are the same for all choices of standard distribution if and only if the effects of A and BC are (log) linear additive, where BC is the joint factor consisting of all pairs of levels of B and C. These comparisons (on the approprite scale) then represent the effect of A, controlling B and C.

Note that the joint control of B and C does not require that the effects of B and C are additive. In the notation of log-linear models, the condition is that the data fit the model [A, BC]. If the data fit the stronger model [A,B,C] which assumes the effects of A, B, and C are additive, then standardized means of B and C can also be interpreted as the net effects of those variables, since the model [A,B,C] implies the models [AC,B] and [AB,C].

CONTINGENCY TABLES

So far we have considered applications of standardization to cross-classification of means. In this section we consider contingency tables and in particular the relationship between standardization and the system of log-linear models for contingency tables given by Goodman (1970, 1972) and discussed in Davis (1974) or Bishop et al. (1975).

A contingency table presents the joint distribution of counts over a set of factors, whereas standardization concerns the relationship of one response variable to a set of other factors. Thus for our purposes it is necessary to define a response variable and to consider variation in the conditional distribution of that variable over the other variables. It will be sufficient to consider a dichotomous response variable. If there are more than two categories, say K, one can successively dichotomize each of K-1 categories against the remainder.

Accordingly consider the IxJx2 contingency table with three factors A, B, and C and category labels i running from 1 to I, j running from 1 to J and k taking the values 0 and 1, respectively. Let ν_{ijk} be the population count in row i, column j, and level k; that is, we assume that no sampling is involved. We are interested in the distribution of C within combinations of A and B. Accordingly, for each i and j we form the proportion with C = 1, referred to as π_{ij}, based on frequency $\nu_{ij.} = \nu_{ijo} + \nu_{ij1}$, so that

$$\pi_{ij} = \nu_{ij1}/\nu_{ij.} \qquad [20]$$

Then the conditional distribution of C given A and B can be studied using this two-way table of proportions.

Note that each proportion is also the mean of the variable C, and hence we have a table of means as discussed in previous sections, the only difference being that the response variable is dichotomous. Hence the results of the previous sections can be applied. That is, differences in the standardized proportions are appropriate if effects A and B are linear additive, and ratios are appropriate if the effects A and B are log linear additive.

However, neither of these conditions corresponds to a log linear model for the I X J X 2 contingency table. Consider the log linear model for the frequencies

$$\log \nu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad \text{for all i, j, k,} \qquad [21]$$

which is the model which fixes the AB, AC, and BC margins, and thus can be written [AB,AC,BC] in the notation of Goodman (1970). Note that equation 21 is the most general model short of the saturated model. The following comments apply to all log linear models for the frequencies which exclude three-way interactions between (a) the predictor, (b) the control, and (c) the dependent variable.

What model does equation 21 imply for the table of *proportions*? Note that from equations 20 and 21

$$\pi_{ij}/(1 - \pi_{ij}) = \nu_{ij1}/\nu_{ijo},$$

so

$$\log [\pi_{ij}/(1 - \pi_{ij})] = \log \nu_{ij1} - \log \nu_{ijo}$$

$$= (\lambda_1^C - \lambda_0^C) + (\lambda_{i1}^{AC} - \lambda_{i0}^{AC}) + (\lambda_{j1}^{BC} - \lambda_{j0}^{BC}).$$

Hence $\log [\pi_{ij}/(1 - \pi_{ij})] = \mu + \alpha_i + \beta_j$ [22]

for suitable choices of μ, α_i, and β_j. Thus the log linear model equation 21 corresponds to an additive model for log $[\pi_{ij}/(1 - \pi_{ij})]$ in the table of proportions. This transformation of the proportions is the familiar logit or log-odds function: logit $\pi_{ij} = \log[\pi_{ij}/(1-\pi_{ij})]$, and accordingly, equation 22 is called a *logit linear additive model,* and variables A and B satisfying this model are called *additive on the logit scale.*

In general, log linear models for contingency tables imply logit linear models for proportions. The relationship is discussed in more detail in Goodman (1970). Aside from this theoretical link, there is a practical reason for considering the logit transformation for tables of proportions: if a logit-linear model is calculated then the fitted values always lie between zero and one.

If equation 21 holds, then equation 22 implies that differences between standardized logits will be invariant with respect to the choice of standard. That is, differences of the form

$$\sum_j (\text{logit } \pi_{ij})\, \omega_j - (\text{logit } \pi_{i'j})\, \omega_j \qquad [23]$$

will not depend upon the choice of $[\omega_j]$. However, above it was the means themselves that were standardized, and then differences between the standardized means or the logs of the standardized means were found to be invariant according to whether the underlying structure was linear additive or log linear additive. We are motivated to ask whether, in the present case, any transformation of the standardized means (i.e., proportions) has the invariance property. For this purpose we shall consider an example.

Example 4. Consider the artificial table of proportions in Table 10. The effects of A and B are logit linear additive, as can be seen from the corresponding logits in Table 10 (ii). The difference between A = 1 and A = 2 is 2.0 on the logit scale within all the categories of B (controlling for B). As a result, standardization of the logits will be invariant with respect to the choice of standard: differences of the form equation 23 do not depend on the choice of $[\omega_j]$ and always equal 2.0.

TABLE 10
Standardization of a Logit-Linear Additive Table of Proportions

(i) Table of Proportions

		Factor B 1	2	3
Factor A	1	.119	.269	.500
	2	.018	.047	.119

(ii) Table of Logits

		1	2	3
Factor A	1	-2.0	-1.0	0.0
	2	-4.0	-3.0	-2.0

(iii) Comparisons of Standardized Proportions Within Categories of A, using Various Standard Distributions of B

	Standard Distribution			Standardized Proportions Factor A 1	2	Differences in Standardized Proportions Raw	Log	Logit
a)	1	0	0	.119	.018	.101	1.889	2
b)	3/5	2/5	0	.179	.030	.149	1.786	1.953
c)	3/5	1/5	1/5	.225	.044	.181	1.632	1.842
d)	1/3	1/3	1/3	.296	.061	.235	1.574	1.868
e)	1/5	1/5	3/5	.378	.084	.293	1.504	1.891
f)	0	0	3/5	.408	.090	.317	1.508	1.937
g)	0	0	1	.500	.119	.381	1.436	2

The last three columns of Table 10 (iii) give, for seven arbitrary standard distributions labelled (a) to (g), the differences between three functions of the standardized proportions, namely

$$\text{Raw:} \qquad \pi_{1.}(\omega) - \pi_{2.}(\omega), \qquad [24]$$

$$\text{Log:} \qquad \log \pi_{1.}(\omega) - \log \pi_{2.}(\omega), \qquad [25]$$

$$\text{Logit:} \qquad \operatorname{logit} \pi_{1.}(\omega) - \operatorname{logit} \pi_{2.}(\omega). \qquad [26]$$

Note that all three sets of differences are sensitive to the choice of standard distribution. This example is sufficient to prove that none of these transformations has the invariance property, and it can be shown that, in fact, *no transformation of the standardized proportions* has the invariance property if A and B are logit linear additive.

Nevertheless, Table 10 (iii) shows that differences in the logits (equation 26) are least sensitive to the choice of standard, varying

TABLE 11
Portion of Indian Women in Fiji Who Ever Used an Efficient Contraceptive Method According to Current Age, Desire for More Children and Educational Level

Education	Desire For More Children		INDIANS Age 1	2	3	4
LOW	YES	p	.60	.72	.58	.42
		n	247	163	135	19
LOW	NO	p	.82	.89	.90	.84
		n	67	160	497	306
HIGH	YES	p	.65	.82	.69	.67
		n	281	131	54	3
HIGH	NO	p	.84	.88	.91	.81
		n	50	101	133	47

p = proportion

n = sample size

SOURCE: Fiji Fertility Survey, 1974

in this example from 2.0 to 1.868, compared with variations of .101 to .381 in the raw differences (equation 24) and 1.889 to 1.436 in the logs (equation 25).

We have seen that when a table of proportions has a logit-linear additive structure, the constant difference in the logits between rows cannot be recovered from the standardized raw proportions for all choices of standard distribution. The relationship between logit-linear models and standardization for tables of proportions is at best approximate, resting in the replacement of equation 23 by equation 26.

This appears to limit the usefulness of standardization for observed tables of proportions where the logit-linear additive structure is believed to apply. The estimated effects from fitting the correct logit-linear model cannot be approximated from standardized proportions and are lost by standardizing the table of fitted proportions. Nevertheless, in some cases this theoretical limitation can be ignored, for the following reasons. (a) For proportions lying between 0.2 and 0.8, the logit scale is approximately linear and thus log-linear additivity is nearly the same as linear additivity. Thus if most observed proportions in a table lie within this range, the analyst can safely consider raw differences in the standardized proportions. (b) For proportions of less than 0.2, the logit scale closely resembles the log scale. Thus if all the observed proportions are small, the analyst may consider percentage differences in the standardized proportions. Similarly, if all the observed proportions are close to one, the analyst may replace each proportion by unity minus the proportion and again consider percentage differences in the standardized proportions, with appropriate interpretation. (c) For cases not covered by (a) or (b), the best approximate procedure is probably to consider the standardized logits, although the interpretation of these as log-odds is not quite as familiar as the other forms. We conclude the discussion by applying these suggestions to our final example.

Example 5. The proportions in Table 11 are derived from a 2x4x2x2 contingency table which gives the number of women who have never used modern contraception, within two categories of education, four categories of age, and two categories of fertility preference (whether the woman does or does not want children). As with Table 2, the data refer to Indian women in the Fiji Fertility Survey, 1974. Ever-use by women who want more children will have been primarily for spacing purposes.

These data are reorganized in Table 12 such that preferences and age are combined into an 8-category variable, and the interest is in the effect of education upon ever-use controlling for preferences and age simultaneously. From the preceding discussion, standardization will be an acceptable method for controlling so long as the differences (equation 23) do not depend upon the choice of standard. To check this, we have applied two alternative standards: the uniform distribution, in which one-eighth of the sample is in each of the preference x age categories, and TFS, which uses the overall sample frequencies in the final column of Table 12.

Under the uniform distribution, the standardized logit is 1.382 for the higher educated group and 1.108 for the less educated

TABLE 12
Proportion (and Logit) of Indian Women in Figi Who Ever Used an Efficient Contraceptive Method According to Current Age, Desire for More Children, and Educational Level

Desire for Another Child	Age Group	High Education Proportion	High Education Logit	Low Education Proportion	Low Education Logit	Sample Size
Yes	1	.65	.619	.60	.455	528
Yes	2	.82	1.516	.72	.944	294
Yes	3	.69	.800	.58	.323	189
Yes	4	.67	.709	.42	-.323	22
No	1	.84	1.658	.82	1.515	117
No	2	.88	1.992	.89	2.091	261
No	3	.91	2.314	.90	2.197	630
No	4	.81	1.450	.84	1.658	353

SOURCE: Table 11.
NOTE: Logits calculated from proportions before rounding.

group, for a difference in standardized logits of .274. Under Test Factor Standardization the quantities are 1.513 and 1.363, respectively, for a difference of .150. The differences (.274 and .150) depend heavily upon the choice of standard, and we conclude without making a formal test that standardization is not appropriate.

The application of logit linear models to the original 2x4x2x2 table shows that, as might have been expected from theoretical considerations, there is significant interaction between education and preferences. That is, education and the preference x age composite variable are not logit linear additive in their impact on ever-use, a formal confirmation that standardization is not appropriate.

Pursuing this illustration a bit further, we can focus on the women who state a desire for no more children, and consider (within this group) the effect of education controlling for age. There is only a small sensitivity to choice of standard in the difference between standardized logits. Under the uniform age distribution, the difference between the high and low education categories is −.012 on the logit scale; using the overall age distribution for these women, the difference is −.006. With the age distribution of the lower educated women as the standard, the difference is −.011, and with the age distribution of the higher educated women as the standard, the difference changes sign to .009. The variation is small. Calculation of the logit linear models for these data using ECTA (Everyman's Contingency Table Analysis) shows that (a) there is no significant education effect, controlling for age, and (b) the preceding four numbers are inside the 95% confidence interval for the education effect on the logit scale.

SUMMARY AND CONCLUSION

We have seen that standardization of a table of means is an efficient summarization of data if differences in the standardized means, on the raw or log scale, are insensitive to the choice of standard. This corresponds to certain linear or log-linear models for the table, which express that the effects of the factors on the response are additive, or in other words, that interactions are not present in the data. The first difficulty in using standardization is in deciding whether such significant interactions are present. Approximate tests of additivity could be developed but have not been given here. Even if this condition is met, there will be some statistical variation in the differences between (or ratios of) standardized quantities as different standards are used. This leads to a second difficulty, namely that standardization leads to statistically inefficient estimates of standardized differences and can be oversensitive to deviant cell means based on small numbers of observations.

An alternative procedure is to use the underlying additivity assumption to fit appropriate models to the data, such as the linear additive models of analysis of variance or log-linear additive models. Then the fitted means from these models can be standardized rather than observed means. The resulting standardized differences are completely insensitive to the choice of standard and are under certain assumptions statistically efficient. Also the fitting procedure can include formal statistical tests of the additivity assumption.

If a table of proportions is derived from a multidimensional contingency table, these can also be subjected to direct standardization. However, in such cases the condition of linear or log linear additivity is not the most natural; the conditions corresponding to log-linear models for the contingency table is that the table of proportions is additive on the logit scale. If logit-linear additivity applies, then no transformation of differences in the standardized proportions is entirely appropriate, although raw differences, or log differences, are approximately valid in certain situations, and in such cases can serve as approximations to the estimated effects from the logit-linear additive model.

REFERENCES

ATCHLEY, R. (1969) "A qualification of test factor standardization." Social Forces 47: 84-85.

BISHOP, Y.M.M., S. E. FIENBERG, and P. W. HOLLAND (1975) Discrete Multivariate Analysis. New York: John Wiley.

CLIFFORD, W. B. and P. L. TOBIN (1977) "Labor force participation of working mothers and family formation: some further evidence." Demography 14: 273-284.

DAVIS, J. A. (1974) "Hierarchical models for significance tests in multivariate contingency tables: an exegesis of Goodman's recent papers," in H. L. Costner (ed.) Sociological Methodology, 1973-1974. San Francisco: Jossey-Bass.

Fiji (1976) The Fiji Fertility Survey 1974—Principal Report. Suva, Fiji: Printing and Stationery Department.

GOODMAN, L. A. (1972) "A general model for the analysis of surveys." Amer. J. of Sociology 77: 1035-1086.

——— (1970) "The multivariate analysis of qualitative data: interactions among multiple classifications." J. of the Amer. Statistical Association 65: 225-256.

KALTON, G. (1968) "Standardization: a technique to control for extraneous variables." Applied Statistics 23: 118-136.

LIEBERSON, S. (1978) "A reconsideration of the income differences found between migrants and Northern-born blacks." Amer. J. of Sociology 83: 940-966.

LITTLE, R.J.A. (1978) "Models for cross-classified data from the World Fertility Survey." Technical Bulletin No. 5, World Fertility Survey.

NELDER, J. A. and R.W.M. WEDDERBURN (1972) "Generalized linear models." J. of the Royal Statistical Society, Series A 135: 370-384.

PULLUM, T. W. (1978) "Standardization." Statistical Bulletin No. 4, World Fertility Survey.

ROSENBERG, M. (1962) "Test factor standardization as a method of interpretation." Social Forces 41: 53-61.

Roderick J. A. Little completed his Ph.D. dissertation at Imperial College, University of London, and was for two years Research Associate at the University of Chicago. Since 1976 he has been a statistician with the World Fertility Survey, London. His research interests include multivariate analysis with incomplete data and log-linear models.

Thomas W. Pullum is Director of the Center for Studies in Demography and Ecology at the University of Washington. His interests include fertility and occupational mobility. He worked with the World Fertility Survey in London in 1976-1977, when much of the work leading to the present article was done.

INDEX

to

SOCIOLOGICAL METHOD AND RESEARCH

VOLUME 7

Number 1 (August 1978) pp. 1-120
Number 2 (November 1978) pp. 121-256
Number 3 (February 1979) pp. 257-376
Number 4 (May 1979) pp. 377-504

Authors:

WORLD FERTILITY SURVEY

OCCASIONAL PAPERS

N.V. DRUKKERIJ TRIO • THE HAGUE • THE NETHERLANDS