*Marginal Standardization and Table Shrinking: Aids in the Traditional Analysis of Contingency Tables**

KENT W. SMITH, *Princeton University*

ABSTRACT

Marginal standardization is an iterative procedure for adjusting the marginal totals of cross-tabulations without affecting the core patterns of association as measured by odds-ratios. One-way distributions may obscure the patterns of association in tables and, in frequently unwanted ways, affect zero-order and partial coefficients and confound the comparisons of tables. The standardization procedure can help overcome these problems and, with slight modifications, has many other applications in table analysis. Before standardizing a table, "random" zeros should be removed by Fienberg and Holland's "table-shrinking" method of adding pseudocounts based on the table and a model of it.

In his 1967 presidential address to the American Statistical Association, Frederick Mosteller described several applications of an iterative procedure for fitting a contingency table to an arbitrary set of marginal frequencies. Thus far, this procedure, which I shall call marginal standardization, has received very little attention from social scientists using the traditional methods of table analysis such as percentaging and nominal and ordinal measures of association; however, it can be a very useful supplement to these methods. In the first part of this article, I shall discuss some of the occasional, undesirable effects of differing marginal frequencies in traditional table analysis and describe how standardizing marginals can help overcome them. The article by Fienberg addressed to historians, and the recent one by Romney et al., focusing on response and sample bias, have also attempted to draw attention to the wide applicability of the procedure. My discussion here complements theirs.

Marginal standardization is a multiplicative process that will not change the value of a cell frequency if it initially is zero. When a zero cell frequency is "random," meaning that it is not inherent in the design or the real world,[1] this fixity of the zero during standardization is undesirable. Fienberg and Holland (a, b) have proposed a Bayesian procedure, which they refer to as "table shrinking," to eliminate these random zeros. The second part of this article is devoted to a relatively non-technical description of this arithmetically simple procedure.

*I wish to thank Princeton University for their generous support of computer usage and the National Data Program for the Social Sciences at the National Opinion Research Center for their policy of placing their data in the public domain so that they could be used here.

MARGINAL STANDARDIZATION

I. THE EFFECTS OF MARGINALS ON ZERO-ORDER MEASURES OF ASSOCIATION

A table's cell frequencies and the measures of association calculated from them reflect in part the association between the variables forming the table, but they also in part are a consequence of the one-way frequency distributions that form the marginal totals of two-way tables. These effects of the one-way distributions often are artifacts of sampling, return rate, and other aspects of the research design that are irrelevant to the phenomenon being studied. Also, they can easily change from time to time or from population to population for reasons unrelated to the causal or explanatory relationship between the variables forming the table, which is generally what we are trying to summarize with a measure of association.

One way to describe the effects of marginals on measures of association is to consider whether their values are changed when every cell frequency in a given row or column of the table is multiplied by the same constant. For instance, both Guttman's lambda (λ) and Goodman and Kruskal's tau (τ) (Goodman and Kruskal, 740-5, 759-60) are always affected by such changes in the marginal distributions of either the row or column variables; Goodman and Kruskal's (748-50) gamma (γ) is affected by marginal changes in either variable if the variable that is changed has more than two categories; Somers' d is always affected by marginal changes in the dependent variable and is affected by changes in the independent variable if it has more than two categories; and Kendall's tau-b (τ_b) is always affected by marginal changes in either variable since it is the geometric mean of the two asymmetric ds.[2]

These effects on measures of association are illustrated by the hypothetical tables shown in Table 1. In Table 1-2, the second column of Table 1-1 has been multiplied by three, and the fourth column by five; in Table 1-3, the first row of Table 1-1 has been multiplied by five; and in Table 1-4, the first row of the transformed Table 1-2 has been multiplied by three. Below each table are the values of the five measures. For consistency, I have used the same double subscripting notation for the three asymmetric ones (λ, Goodman and Kruskal's τ, and d). The first subscript is for the dependent variable; the second, for the independent one.

Recently, Hawkes has reviewed the concept of the *variance* of a categorical variable, based on the comparisons of categorical memberships for all pairs of observations and calculated from the one-way frequency distribution. (Ploch has discussed the *variation* of a categorical variable, which is simply the variance of the variable times the number of observations.) This categorical variance is the direct analog of variance for interval variables; indeed, the variance of an interval variable can easily be calculated from the comparisons of the values of the variable for all possible pairs of observations. Given this close parallel, it seems reasonable to use what we know about the interpretation of interval-level statistics to help inform the analysis of categorical data. For interval-level data, Blalock, Schoenberg, and many others have argued forcefully that unstandardized regression coefficients are usually more substantively meaningful than standardized ones because the unstandardized

Table 1. HYPOTHETICAL TABLE 1-1 AND THREE MARGINAL TRANSFORMATIONS OF IT, WITH SELECTED MEASURES OF ASSOCIATION

Table 1-1

	X(1)	X(2)	X(3)	X(4)	Total
Y(1)	20	15	10	5	50
Y(2)	5	10	15	20	50
Total	25	25	25	25	100

$\lambda_{yx} = 0.400$	$\gamma = 0.625$
$\lambda_{xy} = 0.200$	$d_{yx} = 0.333$
$\tau_{yx} = 0.200$	$d_{xy} = 0.500$
$\tau_{xy} = 0.067$	$\tau_b = 0.408$

Table 1-3

	X(1)	X(2)	X(3)	X(4)	Total
Y(1)	100	75	50	25	250
Y(2)	5	10	15	20	50
Total	105	85	65	45	300

$\lambda_{yx} = 0.0$	$\gamma = 0.625$
$\lambda_{xy} = 0.077$	$d_{yx} = 0.191$
$\tau_{yx} = 0.130$	$d_{xy} = 0.500$
$\tau_{xy} = 0.038$	$\tau_b = 0.309$

Table 1-2

	X(1)	X(2)	X(3)	X(4)	Total
Y(1)	20	45	50	5	120
Y(2)	5	30	75	20	130
Total	25	75	125	25	250

$\lambda_{yx} = 0.250$	$\gamma = 0.510$
$\lambda_{xy} = 0.0$	$d_{yx} = 0.262$
$\tau_{yx} = 0.103$	$d_{xy} = 0.337$
$\tau_{xy} = 0.029$	$\tau_b = 0.297$

Table 1-4

	X(1)	X(2)	X(3)	X(4)	Total
Y(1)	60	135	150	15	360
Y(2)	5	30	75	20	130
Total	65	165	225	35	490

$\lambda_{yx} = 0.038$	$\gamma = 0.510$
$v_{xy} = 0.0$	$d_{yx} = 0.201$
$\tau_{yx} = 0.081$	$d_{xy} = 0.337$
$\tau_{xy} = 0.023$	$\tau_b = 0.260$

ones do not include the effects of the variables' sample variances while the standardized ones do. If we are interested in the basic patterns of association and effect between categorical variables independent of the happenstances of their sample variances, then we should by the same argument use measures of nominal and ordinal association that are not confounded by marginal distributions. But we have just seen that the five commonly used measures of association for categorical data are all, with some exceptions for two-category variables, affected by the variances—that is, the one-way distributions—of variables when they are calculated from the observed, unstandardized data.

For instance, if we assume that the integers identifying the categories of x and y in Table 1 are points on interval scales, then the values of the unstandardized regression coefficient (b_{yx}) for Tables 1-1 through 1-4 are, in order, 0.20, 0.20, 0.12, and 0.15; the corresponding values for x regressed on y (b_{xy}) are 1.00, 0.51, 1.00, and 0.51. Quite appropriately, the values are the same only for those pairs of tables that differ because of a direct transformation of just the independent variable's distribution (Tables 1-1 and 1-2 for b_{yx} and the two pairs, Tables 1-1 and 1-3 and Tables 1-2 and 1-4 for b_{xy}). Notice, though, that the only categorical measures that behave similarly are λ and d, and even they do so only under the special condition of a two-category independent variable.

Apart from this special case for τ and d, we cannot eliminate the often misleading effects of marginal distributions on the values of measures of nominal and ordinal association in the way we can for interval-level regression coefficients.

As an alternative, though, we can *standardize* the marginal distributions so that their effects are uniform across variables, samples, and populations. The conceptually simplest way to standardize marginals, and the one we shall employ here, is to standardize the marginal distribution of a variable so that there is the same proportion of observations in every category.

Actually, as we shall see in the next section, even regression does not always identify all equivalencies of association if we conceptualize association in terms of proportional risks of category membership, or odds-ratios. For example, the corresponding odds-ratios are the same for all four tables in Table 1. Such differences in conceptions of association, however, cannot be explored further in this article; I merely note in passing that they exist.

From a different perspective, Goodman and Kruskal (745-7, 753-4) have also argued for the frequent advantages of the type of marginal standardization discussed here. This standardization, though, should not be confused with Rosenberg's "test factor standardization," an entirely different procedure.

2. THE ODDS-RATIO

A premise of the argument in the previous section was that there is indeed a core pattern of association in a table, analogous in a general way to the regression slope for interval variables, that remains invariant under marginal transformations. If the traditional measures of associations do not measure this stable pattern of association across marginal transformations, then how can it be measured? Several writers have suggested measuring it with the odds-ratios for all 2×2 partitions of a table (see, for example, Mosteller, 3-6; Goodman, a).[3] A brief look at the basic notion of odds-ratios will be useful so that we can use them to confirm that this stable, core pattern of association is unaffected by the iterative standardization procedure.

At this point, I need to introduce some notation for describing tables. I shall use f_{ij} to denote the frequency count of the cell in row i and column j of the table. The marginal total for row i, summed across all columns, is f_{i+} ; similarly, f_{+j} represents the marginal total for column j. I shall use N to denote the total number of observations in the table (f_{++} would be equally appropriate). For multiway tables, we merely add additional subscripts. For instance, in a three-way table, f_{ijk} is the frequency count for the cell in row i, column j, and category k of the third dimension. The total number of observations in category i of the first variable, summed across all categories of the other two variables is denoted by f_{i++} .

In a two-way table, the odds of an observation's being in row one rather than row two, given that it is in column one, are f_{11} / f_{21}. In Table 1-1, these odds, for instance, are 20 / 5 or 4.0: an observation in column one is 4 times more likely to be in row one than in row two. For an observation in column two, the odds of being in row one rather than row two are f_{12} / f_{22}. For column two of Table 1-1, they are 15 to 10, or 1.5. The ratio of these two odds provides a good summary of the pattern of association between the two rows and the first two columns of Table 1-1:

$$\frac{f_{11}}{f_{21}} : \frac{f_{12}}{f_{22}} = \frac{f_{11}f_{22}}{f_{21}f_{12}} = 4 / 1.5 = 2.67. \qquad (1)$$

We can see from the second formula in (1) why this odds-ratio is often referred to as the *cross-product* ratio.

Some quick calculations will show that this odds-ratio is 2.67 for all four tables in Table 1, and that all other possible odds-ratios are also the same for all four tables. Thus, the core pattern of association, as measured by the odds-ratios, is invariant under these marginal transformations. Fortunately, marginal standardization helps us discover this equivalent core association. When we standardize the marginals of the last three tables to equal those of Table 1-1, the resulting standardized cell frequencies are all equal to the corresponding ones of Table 1-1.

While the odds-ratio is based on asymmetric conditional odds, the direction of the asymmetry can be interpreted as going either way, with the distribution for either of the variables conditional on that of the other. For instance, there are actually three other equivalent interpretations of the odds-ratio in equation (1): $f_{11}/f_{12} : f_{21}/f_{22}$; $f_{22}/f_{12} : f_{21}/f_{11}$; and $f_{22}/f_{21} : f_{12}/f_{11}$.

In measuring the pattern of association in a table, the odds-ratio can be calculated for any possible 2 $\times$ 2 partition of the table as long as the cross– product version is of the general form $f_{hj}f_{ik} \,/\, f_{hk}f_{ij}$, in which $h < i$ and $j < k$. This general form suggests why multiplicative transformations of one-way distributions do not affect odds-ratios: every category in the numerator of the ratio appears an equal number of times in the denominator so that any transformation constant cancels out.

In three-way tables, the general form of the odds-ratio is either $f_{hjs}f_{ikt} \,/\, f_{hks}f_{ijt}$ or $f_{hjs}f_{ikt} \,/\, f_{hkt}f_{ijs}$. The subscript s may equal the subscript t, in which case the odds-ratio is within a single category of the third, control variable. For instance, the odds-ratio for the 2 $\times$ 2 partition formed by the first and second rows and third and fourth columns within the first category of the control is $f_{131}f_{241} \,/\, f_{141}f_{231}$. Alternatively, s may not equal t, with the odds-ratio across categories of the control variable. An example is that for the first and second rows, the third column within the first control category, and the fourth column within the second control category: $f_{131}f_{242} \,/\, f_{142}f_{231}$. (Actually, the subscripts in any one of the three pairs h and i, j and k, or s and t may be equal as long as those in the other two pairs are not.) Again, any multiplicative transformation of one-way distributions cancels out of the odds-ratios.

We can also look at the odds-ratios for the two-way marginals formed by summing across the third variable. For instance, the odds-ratio for the two-way marginals of the first and second rows and third and fourth columns is $f_{13+}f_{24+} \,/\, f_{14+}f_{23+}$.

3. SOME OTHER FREQUENT PROBLEMS OF MARGINAL DISTRIBUTIONS

Besides the complications posed by marginal distributions for using measures of association to summarize the core pattern of association, there are at least two other frequently unwanted effects of marginal distributions.

I have already alluded to the first. Namely, differences in marginal distributions often interfere with substantively meaningful comparisons of tables for different variables, samples, or populations. If we are interested in comparing the core patterns of association independent of one-way variances, then we probably should standardize marginals before comparing tables (assuming, of course, that the tables are comparable in terms of the number of categories in each dimension). The hypothetical data in Tables 1-1 and 1-4 can provide an illustration. Let us suppose that the rows of the tables represent approval and disapproval of a proposed tax reform and that the columns represent four ordered categories of social class. If the tables are the results for parallel probability surveys of two towns, then all five measures suggest that social class has much more effect on attitudes toward the tax reform in the first town than in the second one. However, we have seen that the odds-ratios, the relative likelihoods of approving rather than disapproving for different social classes, are the same in both tables. The differences in column percentages (not shown) and in the values of the measures of association for the two tables are merely reflecting the fact that the class distributions and the overall proportions approving the tax reform are different in the two towns. When these marginal differences are eliminated by standardizing Table 1-4 to the marginals of Table 1-1, the two tables are identical. Clearly, for many analytic purposes these comparisons of the standardized tables are more meaningful than those of the observed tables.

Second is the effect of the one-way distributions of control variables on coefficients of partial association. The partial τ_b and ds proposed by Hawkes are calculated, on the one hand, from a matrix of zero-order τ_b's and, on the other hand, from a matrix of the variables' ordinal variances and covariances, with the covariance of two variables equalling their Kendall's S statistic (the numerator of γ, both ds and τ_b) divided by N^2 (Hawkes, 912). Since the distribution of a control variable determines its variance and affects the value of the zero-order τ_b and Kendall's S for any two-way table involving the control variable, its distribution obviously will affect the values of Hawkes' partial coefficients.

A control variable's distribution will also affect the values of those partial coefficients calculated from the conditional subtables in a multiway table, such as Goodman and Kruskal's (761) partial lambda, Davis's partial gamma, and Ploch's partial d. A similar partial for Goodman and Kruskal's tau can also be formed, although I do not know of any published mention of this partial. As Davis has shown for the partial gamma, each of these partial coefficients is a weighted sum of the conditional coefficients for the two-way subtables. The weight for a given conditional coefficient is a function of the distribution of observations within its subtable *and* of the overall proportion of observations that are in that subtable. In a three-way table, this latter, of course, is simply the one-way distribution of the control variable. Unless the distributions of the control variables have important substantive implications, it may often be best to eliminate this aspect of the weighting of the conditionals by standardizing the controls so that there is an equal proportion of observations in each category. Certainly such standardization should be seriously considered if one wishes to compare the partial coefficients for different samples or populations.

4. THE ITERATIVE STANDARDIZATION PROCEDURE

For the types of problems discussed so far, we generally want to eliminate only the effects of differential one-way distributions. In such cases, therefore, we arbitrarily fix only the category totals summed across all other dimensions of the table; all other aspects of the table will be allowed to change during the iterative procedure. I shall algebraically outline the procedure for the three-way case, since the generalization to other cases is straightforward. We begin by establishing the target frequency for each category of the variables in the table. Let us call these target frequencies t : t_{i++} represents, for example, the number of standardized observations we wish to have at the end of the iteration in category i of the first variable. For the purposes discussed so far, one probably wants to select these target values so that they are the same for every category of a given variable. This equalization of a variable's categories is a conceptually straightforward way to make marginal effects uniform; and, as Hawkes (913) has observed, it maximizes the variances of the variables within the limits set by their numbers of categories. However, the target values for the categories of a variable do not all have to be equal. The only restriction the iterative procedure sets on them is that those for the various variables must be consisent with each other so that they add up to the same total number of standardized observations for the table: if there are r rows, c columns, and a control categories,

$$\sum_{i=1}^{r} t_{i++} = \sum_{j=1}^{c} t_{+j+} = \sum_{k=1}^{a} t_{++k} = t_{+++}. \tag{2}$$

We then begin cycling through each variable in turn. At each stage of the iteration, we standardize the cell frequencies in the previous stage so that they add to the appropriate target figure, just as we do when we standardize cells to add to 100 percent. The superscripts in the following equations represent the stage of iteration. The notation $f_{ijk}^{(0)}$ represents the initial frequency of cell (i, j, k) before the first step of the iteration.

Step 1.
We standardize the initial cell frequencies to the target totals for the first variable:

$f_{ijk}^{(1)} = (f_{ijk}^{(0)} \ / \ f_{i++}^{(0)})t_{i++}$ for all i, j, and k.

At the end of this step, the first variable's marginal totals $f_{i++}^{(1)}$ will equal t_{i++}.

Step 2.
After using the new cell frequencies $f_{ijk}^{(1)}$ to calculate new totals for the second variable ($f_{+j+}^{(1)}$), we standardize the cell frequencies from step 1 to the target totals for the second variable:

$f_{ijk}^{(2)} = (f_{ijk}^{(1)} \ / \ f_{+j+}^{(1)}) \ t_{+j+}$ for all i, j, and k.

Step 3.

Following the same procedure, we standardize the new cell frequencies from step 2 to the target totals for the third variable:

$$f_{ijk}^{(3)} = (f_{ijk}^{(2)} \; / \; f_{++k}^{(2)}) \; t_{++k} \quad \text{for all } i, j, \text{ and } k.$$

Step 3 completes the *first cycle* of the iteration. We now check to see if all category totals differ from their respective targets by less than some arbitrary amount. If not, then we start a new cycle, beginning again with step one, only this time standardizing the frequencies from step 3 to the target totals for the first variable:

$$f_{ijk}^{(4)} = (f_{ijk}^{(3)} \; / \; f_{i++}^{(3)}) \; t_{i++} \quad \text{for all } i, j, \text{ and } k.$$

For the examples in this article, the iteration was continued until the standardized marginal totals were all within +0.01 of their target figures. Normally, the iteration converges to this degree of accuracy in less than seven iteration cycles; seldom does it take more than ten.

The extension of this procedure to tables with more than three variables is very simple: for each additional variable we merely add another step to each iteration cycle in which we standardize the cell frequencies from the previous step to the target figures for the additional variable.

A two-way numerical example would perhaps be useful. In Table 2 are data from the 1972 General Social Survey of the National Data Program for the Social Sciences conducted by the National Opinion Research Center under the direction of James A. Davis. Forming the rows of the table is a three-category scale of attitudes toward abortion practices,[4] which we shall interpret as the dependent variables; the independent column variable is the number of years of education completed by the respondents, collapsed into three categories. The actual cell frequencies and marginal totals are given in parentheses. In Table 3, these marginal frequencies have been standardized to equal proportions of 100 (33.33). Table 4 shows step by step

Table 2. ATTITUDES TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED, IN PERCENTAGES WITH INITIAL FREQUENCIES IN PARENTHESES*

	Years of Schooling			
Abortion	0–11	12	13+	Total
Generally disapprove	38.2% (209)	21.5% (151)	9.1% (16)	26.4% (376)
Middle position	18.5 (101)	17.9 (126)	12.0 (21)	17.4 (248)
Generally approve	43.3 (237)	60.6 (426)	78.9 (138)	56.2 (801)
Total	100.0% (547)	100.0% (703)	100.0% (175)	100.0% (1425)

$\gamma = 0.387 \quad d_{yx} = 0.229$

*Data from the 1972 General Society Survey of the National Data Program.

Table 3. FREQUENCIES FROM TABLE 2 AFTER STANDARDIZING MARGINALS TO EQUAL PROPORTIONS OF 100*

	Years of Schooling			
Abortion	*0–11*	*12*	*13+*	*Total*
Generally disapprove	16.5	10.9	5.9	33.3
Middle position	10.7	12.2	10.4	33.3
Generally approve	6.2	10.2	17.0	33.3
Total	33.3	33.3	33.3	100.0
	$\gamma = 0.418$	$d_{yx} = 0.286$		

*Some totals are incorrect in the last figure because of rounding.

the iterative procedures for transforming Table 2 into Table 3. In step one, the initial row marginals were standardized to 33.33 each; in step two, the column marginals from step one were standardized to 33.33 each; and so forth until, at the end of step eight, all the marginal totals were within 0.01 of their target values.

In passing, I might note the relative clarity of the pattern of association in Tables 2 and 3. While one would not draw different conclusions from the two tables, the positive association is visually much stronger in Table 3 because we have equalized the marginal distribution of the dependent variable. For instance, if we were to graph the tables using the heights of bars rising from the table to indicate the relative sizes of the cell frequencies, there would be a clear ridge down the main diagonal of Table 3 with the rest of the table falling away smoothly to both sides. In Table 2 there would not be this clear visual portrayal of the association.

A three-way example will help clarify some other aspects of interpreting standardized tables, especially in relation to odds-ratios. In Table 5, I have added a control variable to the two-way relationship of attitudes toward abortion and years of education completed that was reported in Table 2. This third variable is the respondents' answers to the question ''What do you think is the ideal number of children for a family to have?'' collapsed into three categories for illustrative purposes here.[5]

As the log-linear model of contingency tables (see, for example, Goodman, b, c) emphasizes, a direct analogy can be made between the effects underlying the cell means in multiway analysis of variance and the effects underlying cell frequencies in contingency tables. The cell frequencies f_{ijk} can each be viewed as a function of a three-way effect for that particular combination of three categories; of 3 two-way effects, one for each combination of the categories taken two at a time; and of 3 one-way effects, one for each of the separate categories; plus an overall factor for the total number of observations in the table. The odds-ratio for a 2 × 2 partition of these cell frequencies is determined by the three-way and two-way

Table 4. STANDARDIZING MARGINAL TOTALS OF TABLE 2 TO EQUAL PROPORTIONS

First Cycle

Step 1. Changing $f_{ij}^{(0)}$ to $f_{ij}^{(1)}$ to fit t_{i+}

	X(1)	X(2)	X(3)	Total
Y(1)	18.53	13.39	1.42	33.34
Y(2)	13.58	16.94	2.82	33.34
Y(3)	9.86	17.73	5.74	33.33
Total	41.97	48.06	9.98	100.01

Step 2. Changing $f_{ij}^{(1)}$ to $f_{ij}^{(2)}$ to fit t_{+j}

	X(1)	X(2)	X(3)	Total
Y(1)	14.72	9.29	4.74	28.75
Y(2)	10.79	11.75	9.42	31.96
Y(3)	7.83	12.30	19.17	39.30
Total	33.34	33.34	33.33	100.01

Second Cycle

Step 3. Changing $f_{ij}^{(2)}$ to $f_{ij}^{(3)}$ to fit t_{i+}

	X(1)	X(2)	X(3)	Total
Y(1)	17.07	10.77	5.50	33.34
Y(2)	11.25	12.26	9.82	33.33
Y(3)	6.64	10.43	16.26	33.33
Total	34.96	33.46	31.58	100.00

Step 4. Changing $f_{ij}^{(3)}$ to $f_{ij}^{(4)}$ to fit t_{+j}

	X(1)	X(2)	X(3)	Total
Y(1)	16.28	10.73	5.81	32.82
Y(2)	10.73	12.21	10.36	33.30
Y(3)	6.33	10.39	17.16	33.88
Total	33.34	33.33	33.33	100.00

Third Cycle

Step 5. Changing $f_{ij}^{(4)}$ to $f_{ij}^{(5)}$ to fit t_{i+}

	X(1)	X(2)	X(3)	Total
Y(1)	16.53	10.90	5.90	33.33
Y(2)	10.74	12.22	10.37	33.33
Y(3)	6.23	10.22	16.88	33.33
Total	33.50	33.34	33.15	99.99

Step 6. Changing $f_{ij}^{(5)}$ to $f_{ij}^{(6)}$ to fit t_{+j}

	X(1)	X(2)	X(3)	Total
Y(1)	16.45	10.90	5.93	33.28
Y(2)	10.69	12.22	10.43	33.34
Y(3)	6.20	10.22	16.97	33.39
Total	33.34	33.34	33.33	100.01

Fourth Cycle

Step 7. Changing $f_{ij}^{(6)}$ to $f_{ij}^{(7)}$ to fit t_{i+}

	X(1)	X(2)	X(3)	Total
Y(1)	16.48	10.92	5.94	33.34
Y(2)	10.69	12.22	10.43	33.34
Y(3)	6.19	10.20	16.94	33.33
Total	33.36	33.34	33.31	100.01

Step 8. Changing $f_{ij}^{(7)}$ to $f_{ij}^{(8)}$ to fit t_{+j}

	X(1)	X(2)	X(3)	Total
Y(1)	16.47	10.92	5.94	33.33
Y(2)	10.68	12.22	10.44	33.34
Y(3)	6.19	10.20	16.95	33.34
Total	33.34	33.34	33.33	100.01

effects on the four cells since the one-way effects and the overall factor appear equally in the numerator and denominator of the ratio and cancel out.

In an attempt to avoid terminological confusion in what follows, I shall use the term *association* to refer to the pattern of association as measured by odds-ratios and the term *interaction* to refer to the pattern of effects for a particular combination of variables: this use of "interaction" parallels that in analysis of variance. For instance, the three-way association in a three-way table is what we measure by the odds-ratios of cell frequencies; the three-way interaction is the pattern of only the three-way effects on the cell frequencies. Thus, following from the previous paragraph, the three-way interaction is only one of the determinants of the three-way association: it is also a function of 3 two-way interactions.

When we standardize the one-way distributions to equal proportions, we eliminate only the one-way effects of being in one category rather than another of a

Table 5. ATTITUDES TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH IDEAL NUMBER OF CHILDREN CONTROLLED, INITIAL FREQUENCIES IN PERCENTAGES WITH ACTUAL FREQUENCIES IN PARENTHESES*

	Ideal Number of Children											
	None to Two				*Three*				*Four or More*			
	Years of Schooling				*Years of Schooling*				*Years of Schooling*			
Abortion	*0–11*	*12*	*13+*	*Total*	*0–11*	*12*	*13+*	*Total*	*0–11*	*12*	*13+*	*Total*
Generally disapprove	28.3% (58)	15.7% (55)	1.7% (2)	17.1% (115)	33.8% (44)	24.1% (48)	21.1% (8)	27.2% (100)	50.5% (107)	31.2% (48)	28.6% (6)	41.6% (161)
Middle position	21.0 (43)	14.6 (51)	8.6 (10)	15.5 (104)	14.6 (19)	18.6 (37)	15.8 (6)	16.9 (62)	18.4 (39)	24.7 (38)	23.8 (5)	21.2 (82)
Generally approve	50.7 (104)	69.7 (244)	89.7 (104)	67.4 (452)	51.5 (67)	57.3 (114)	63.2 (24)	55.9 (205)	31.1 (66)	44.2 (68)	47.6 (10)	37.2 (144)
Total	100.0% (205)	100.0% (350)	100.0% (116)	100.0% (671)	100.0% (130)	100.0% (199)	100.0% (38)	100.0% (367)	100.0% (212)	100.0% (154)	100.0% (21)	100.0% (387)
	$\gamma = 0.478$		$d_{yx} = 0.239$		$\gamma = 0.148$		$d_{yx} = 0.087$		$\gamma = 0.286$		$d_{yx} = 0.189$	

*Data from the 1972 General Social Survey of the National Data Program. Some totals are incorrect in the last figure because of rounding.

variable and, therefore, do not change the pattern of association among the cells. For example, Table 6 is the standardized version of Table 5. The zero-order table is merely the table of the two-way marginals f_{ij+} summed across the subtables. As we can see from the row and column totals of the zero-order table and the overall totals of each of the subtables, every category of all three variables now has 33.3 standardized observations in it. All of the corresponding odds-ratios for cell frequencies, both within and between subtables, are the same in Tables 5 and 6, indicating that the two-way and three-way interactions have not been changed by standardization. For instance, the odds-ratio ($f_{112}f_{222} / f_{122}f_{212}$) is 1.8 in both tables.

Let us use the term "two-way faces" to refer to the tables of two-way marginals: the zero-order table is the face for the f_{ij+}, the table that can be formed from the column marginals of the subtables is the face for the f_{+jk}, and the table of the row marginals of the subtables is the face for the f_{i+k}. The factors affecting the association in one of these faces are more complicated than those affecting the association among three-way cells. Every two-way marginal is in part a function of the corresponding two-way effect of that pair of categories, of the 2 one-way effects of the individual categories, and of the overall factor for the total number of observations in the table. But since each two-way marginal is a sum of cell frequencies across the third dimension of the table, it is also a function of the three-way effects on the cell frequencies comprising it. Moreover, these three-way effects are weighted by the one-way proportional distribution of the third variable. In any odds-ratio for a 2 × 2 partition of a face, the one-way effects and the factor for the total number of observations again cancel out; and we are left with the fact that the pattern of association in a face of a three-way table is determined by the corresponding two-way interaction *and* the weighted three-way one.

Standardizing the one-way distributions not only will eliminate the differences among one-way effects, but will also change the relative weighting of the three-way effects' contributions to the two-way marginals. Consequently, while one-way standardization of a three-way table will not change the odds-ratios of cell frequencies, it will in general change the odds-ratios of the two-way marginals in the faces of the table. For illustration, we can use Table 2 (the zero-order table for Table 5) and the corresponding zero-order portion of the standardized Table 6. Using the initial frequencies in Table 2, the odds-ratio ($f_{12+} f_{33+} / f_{13+} f_{32+}$) is 3.1; but the corresponding odds-ratio for the standardized two-way marginals in Table 6 is 2.5.

These changes in the odds-ratios of the zero-order tables occurred only because we specifically allowed for three-way interaction by including a third variable. As we have already seen in Table 3, standardization did not change the odds-ratios when Table 2 was considered to be a simple two-way contingency table rather than a face of a three-way one. If there is no three-way interaction in a particular three-way table so that the proportional cell frequencies are determined entirely by the two-way and one-way effects, then one-way standardization will not change the odds-ratios for the faces of the table because, in this case, the association in a two-way face is determined by the corresponding two-way interaction.

We can also see the effect of retaining the three-way interaction in the standardization by comparing the standardized two-way Table 3 with the zero-order portion of the standardized three-way Table 6. Note particularly that the standardized frequencies in the upper right and lower left cells of the zero-order portion of Table 6 are larger and that the values of both the zero-order γ and d_{yx} for Table 6 are smaller. The reason lies in the three-way table, and here we see that one's opinion about the ideal number of children in a family greatly specifies the association between abortion attitudes and schooling, as shown by the large differences among the values of the conditional γ and d_{yx}'s for the individual subtables in both Tables 5 and 6. Because 47 percent of the people (671/1425) are in the first subtable of Table 5, the strong three-way effects in this subtable contribute more to the zero-order table than do the weaker three-way effects in the other two subtables. The standardization in Table 3 does not change this differential weighting; however, that in Table 6 equalizes the proportion of observations in each subtable, with the result that there is less zero-order association in Table 6.

The specification by the control variable of the relationship between abortion attitudes and schooling means that partial coefficients are inappropriate (using Ploch's W^2 statistic, the specification of the conditional d_{yx}'s is significant at the 0.025 level); however, for illustrative purposes, it is worth examining their values for Tables 5 and 6. As mentioned earlier, both Davis's partial γ and Ploch's partial d_{yx} can be interpreted as weighted sums of the conditionals, with a conditional's weight a function of both the distribution within the corresponding subtable and the overall proportion of observations in it. Because standardization in Table 6 has decreased the proportion of observations in the first subtable, the conditional measures for this subtable contribute less to the partials for Table 6 than they do to those for Table 5. The value of the partial γ for abortion attitude by schooling with the ideal number of children controlled decreases from 0.371 in Table 5 to 0.326 in Table 6, even though the first conditional γ increases from 0.478 to 0.561. Similarly, the partial d_{yx} for attitude dependent upon schooling changes from 0.187 in Table 5 to 0.214 in Table 6, which is a relatively small increase compared with the change from 0.239 to 0.369 for the first conditional d_{yx}.[6]

Looking at the actual cell frequencies and marginals, we probably would not draw different conclusions from the standardized Table 6 than we would from the percentaged Table 5. However, I think it is fair to say that the core pattern of association is visually sharper and more easily recognized in Table 6. While measures of association are important, we should not overlook the fact that the visual clarity of the association can also be very valuable in the analysis and interpretation of contingency tables, especially those with larger numbers of cells. What one loses in standardized tables is a sense of how the one-way distributions for this sample affect its multiway distributions; and for purposes of estimation and empirical description, this may be an important loss. It is not, however, an important loss for many analytic and theoretical purposes.

Table 6. FREQUENCIES AFTER ONE-WAY MARGINAL STANDARDIZATION FOR ATTITUDES TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH IDEAL NUMBER OF CHILDREN CONTROLLED, FROM TABLE 5*

Zero-order Table

	Years of Schooling			
Abortion	*0–11*	*12*	*13+*	*Total*
Generally disapprove	16.2	10.5	6.6	33.3
Middle position	10.5	12.3	10.5	33.3
Generally approve	6.7	10.5	16.2	33.3
Total	33.3	33.3	33.3	100.0
	$\gamma = 0.374$		$d_{yx} = 0.225$	

Three-way Table

Ideal Number of Children

	None to Two				*Three*				*Four or More*			
	Years of Schooling				*Years of Schooling*				*Years of Schooling*			
Abortion	*1–11*	*12*	*13+*	*Total*	*0–11*	*12*	*13+*	*Total*	*1–11*	*12*	*13+*	*Total*
Generally disapprove	3.0	2.6	0.5	6.1	4.1	4.1	3.6	11.8	9.1	3.8	2.5	15.4
Middle position	3.2	3.5	3.6	10.3	2.5	4.6	3.9	11.0	4.8	4.3	3.0	12.0
Generally approve	2.1	4.6	10.3	16.9	2.4	3.8	4.3	10.5	2.2	2.1	1.6	5.9
Total	8.3	10.7	14.4	33.3	9.0	12.5	11.8	33.3	16.0	10.2	7.1	33.3
	$\gamma = 0.561$		$d_{yx} = 0.369$		$\gamma = 0.149$		$d_{yx} = 0.100$		$\gamma = 0.270$		$d_{yx} = 0.173$	

*Some totals are incorrect in the last figure because of rounding.

5. SOME ADDITIONAL APPLICATIONS OF THE ITERATIVE PROCEDURE

So far, our main focus has been on the use of the iterative procedure to equalize one-way distributions; however, this is far from its only application. In this section, I shall just briefly mention some other ways it may be used as an adjunct to traditional table analysis.

First, in one-way standardization one does not have to set the categories of every variable to equal proportions. The only restriction of the procedure is that the target values must be consistent with one another (see equation [2]). Thus, for instance, if the proportional distribution of one variable had substantive implications that one wanted to preserve, while the distributions of the other variables were research artifacts that one wished to ignore, then the standardization could be toward target marginals that had the observed proportional distribution for the first variable and equal proportions for the others.

Second, for three-way tables, we can modify the procedure to eliminate some or all of the two-way interactions in addition to the effects of one-way distributions; in four-way tables, we can eliminate some or all of the three-way interactions; and so forth. We can eliminate a two-way interaction quite simply by equalizing the two-way marginals in the corresponding face of the table. For example, say we wished to eliminate one-way effects and the two-way interaction of schooling by ideal number of children in Table 5 while maintaining the other 2 two-way interactions and the three-way one. The face corresponding to this two-way interaction is shown in Table 5 as the column totals of the subtables, the marginals f_{+jk}. To eliminate the effect of these marginals, we assign the same target value t_{+jk} to each of them. When the two-way or higher marginals for some variables are equalized, all lower-order marginals involving only those variables are perforce also equalized. For instance, by equalizing the t_{+jk}'s, we are also equalizing the t_{+j+}'s and t_{++k}'s. Thus, our iterative procedure now needs to include only two steps per cycle, first fitting the table to the t_{i++}'s and then to the t_{+jk}'s. Notice that this standardization is the same as the one we would have performed if we had treated the three-way table as a two-way 3×9 one, combining schooling and ideal number of children into one typology. (See Mosteller, 10-11, for another example.)

Going further, we could eliminate both the two-way interaction of schooling by ideal number and that of attitude by ideal number, retaining only the two-way interaction of attitude and schooling and the three-way interaction. In this case we iterate to fit equal t_{+jk}'s and equal t_{i+k}'s. This procedure, of course, is the same as standardizing, to a common set of targets, each of the subtables as a separate two-way table.[7] Finally, we can eliminate all one-way effects and two-way interactions by setting equal targets t_{ij+} for the remaining two-way marginals. The iteration cycle in this case would involve three steps, transforming the f_{ijk}'s so that the table fits, in turn, the targets t_{ij+}, t_{i+k}, and t_{+jk}. The pattern of association in the resulting table is a function of only the three-way interaction.

As Mosteller (8-11) discusses, one can also use the iterative procedure to estimate the joint frequency distribution for some sample or population using that sample's or population's marginal distributions and the interactions from some other one. (Mosteller [10] reports that this use of the iterative procedure was evidently first proposed in 1940 by W. Edwards Deming and Frederick F. Stephan.) For instance, suppose that we have estimates of the one-way distributions of attitudes toward abortion, years of schooling, and ideal family size for a particular city but do not know their joint distributions because the estimates came from different sources. If we are willing to assume that the interactions among the three variables are the same in this city as in the national sample data in Table 5, then we can estimate the joint distribution for the city by iteratively transforming Table 5 to fit the one-way distributions for the city. If we are fortunate enough to know the city's two-way distribution of abortion attitudes and ideal family size but do not know for the city, either variable's relationship with schooling, then we can transform Table 5 to fit the city's two-way marginals of abortion attitude by ideal family size (t_{i+k}) and its one-way schooling distribution (t_{+j+}).

In all the applications discussed so far, we have used the iterative procedure to eliminate or change lower-order effects while preserving higher ones. It can also be used, however, to preserve lower-order effects while eliminating higher ones. Two good brief discussions of this application are by Mosteller (16-21) and Goodman (c, 1080-5). In this application, we begin with a table that has the same number of dimensions and categories as the observed table, but in which there is complete independence of distributions—specifically, a table with 1.0 in every cell.[8] Note that all odds-ratios in such a table are equal to 1.0. Then we merely iteratively transform this "independence table" of ones to fit the observed marginals corresponding to the interactions we wish to preserve. If we retain any two-way or higher interactions, we perforce also retain any lower-order interactions among the same variables and those variables' one-way effects. For instance, let us say that we want to know what the cell frequencies would be for a table with no three-way interaction but with all the two-way interactions and one-way effects of Table 5. Beginning with an initial $3 \times 3 \times 3$ table with every f_{ijk} equalling 1.0, we iteratively fit it to the two-way marginals f_{ij+}, f_{i+j}, and f_{+jk} from Table 5. An example of this application is discussed in the next section. If, instead, we wished to retain only the two-way interaction between abortion attitudes and schooling and the one-wav distribution of the ideal number of children, we would fit the independence table to the marginals f_{ij+} and f_{++k} of Table 5. Again, the convergence is quite rapid, taking usually only four to seven cycles for accuracy to one decimal place.

Before moving on to random zeros and table shrinking, I must enter a very important caveat about all of the applications of the iterative procedure, including standardization to equal proportions. If one is employing significance tests, they should be calculated on the sample data before any iteration. Significance tests are based on the observed characteristics of the sample, including sample size, and become meaningless after these characteristics are altered by the iterative procedure. At the same time, one must recognize that significance tests of measures of association for the initial tables are not tests for the core patterns of association, just as measures of association for the initial tables are not measures of these patterns.[9]

For instance, if the four tables in Table 1 were subtables of a three-way table, the probability of the value of the W^2 test for differences among the four conditional d_{yx}'s would be 0.15 even though all four subtables would have exactly the same set of odds-ratios.

TABLE SHRINKING TO ELIMINATE RANDOM ZEROS

Even with large surveys, contingency tables with more than two dimensions all too often contain zero cell frequencies, not because the cells are ipso facto empty, but simply because there did not happen to be any observations in them for the particular sample. For a different or larger sample, these cells might very well be nonzero. If we standardize a table containing these ''random'' zeros, we do not want them to unduly influence or distort the resulting picture of the patterns of association. Unfortunately, they often do. Since the iterative procedure only multiplies and divides an initial cell frequency by other values, a zero frequency will remain zero; and any change in the proportion of observations in a category containing a zero will have to be absorbed by the other cells in that category. Moreover, random zeros are most likely to occur in categories which contain proportionately small numbers of observations and which, consequently, must proportionately be transformed a great deal.

An example is given in Tables 7 and 8. In Table 7 we have our old friends abortion attitudes and years of schooling with a new control variable, religious affiliation.[10] In this sample, there were no Jews with more than a high-school education in either the generally disapprove or middle-position categories of the attitude variable. However, there is no reason to believe that there are no Jews in the United States belonging to these cells. Indeed, the expected frequencies of these cells for samples of this size from the national population are probably fractional values slightly larger than zero. (Here I am using ''expected frequencies'' in the sampling sense; I am not referring to the values calculated from the one-way distributions under the null hypothesis of independence.) Table 8 is the standardized version of Table 7, with the initial two random zeros unchanged. None of the subtables of Table 8 have the nice smooth monotonic relationship between abortion attitudes and schooling that we have seen in the previous tables. For the Roman Catholic and Protestant subtables, this fact may actually indicate the extent to which religious affiliation specifies and alters the effect of schooling in the population, although the frequency of the middle-position attitude for Catholics with a high-school education (6.0) seems surprisingly high. However, I think we can be less certain of the representativeness of the standardized picture in the Jewish subtable, quite apart from the smaller initial sample size. For instance, the value of γ for this subtable has been somewhat increased and that of d_{yx} considerably increased by standardization, not because one-way effects have been eliminated, but rather because zeros have been retained. Retaining the two off-diagonal zeros during the standardization has artificially increased the number of concordantly ordered pairs in

Table 7. ATTITUDE TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH RELIGIOUS AFFILIATION CONTROLLED, INITIAL FREQUENCIES*

	Religious Affiliation											
	Roman Catholic				Protestant				Jewish			
	Years of Schooling				Years of Schooling				Years of Schooling			
Abortion	*1–11*	*12*	*13+*	*Total*	*0–11*	*12*	*13+*	*Total*	*0–11*	*12*	*13+*	*Total*
Generally disapprove	65	68	16	149	189	92	6	287	1	2	0	3
Middle position	26	47	6	79	79	79	17	175	4	1	0	5
Generally approve	59	52	22	133	166	297	105	545	6	22	18	46
Total	150	167	44	361	434	468	105	1007	11	25	18	54
	$\gamma = 0.038$		$d_{yx} = 0.024$		$\gamma = 0.482$		$d_{yx} = 0.296$		$\gamma = 0.815$		$d_{yx} = 0.248$	

*Those in other religions and with no religious affiliation have been excluded. The data are from the 1972 General Social Survey of the National Data Program.

Table 8. FREQUENCIES AFTER ONE-WAY MARGINAL STANDARDIZATION FOR ATTITUDES TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH RELIGIOUS AFFILIATION CONTROLLED, FROM TABLE 7*

	Religious Affiliation											
	Roman Catholic				*Protestant*				*Jewish*			
	Years of Schooling			*Total*	*Years of Schooling*			*Total*	*Years of Schooling*			*Total*
Abortion	*1–11*	*12*	*13+*		*0–11*	*12*	*13+*		*0–11*	*12*	*13+*	
Generally disapprove	5.8	6.1	5.4	17.3	8.0	3.9	0.9	12.8	1.1	2.1	0.0	2.3
Middle position	3.3	6.0	2.9	12.2	4.8	4.8	3.8	13.4	6.2	1.5	0.0	7.7
Generally approve	1.2	1.0	1.6	3.8	1.5	2.7	2.8	7.1	1.4	5.2	15.8	22.4
Total	10.3	13.1	9.9	33.3	14.3	11.4	7.6	33.3	8.7	8.8	15.8	33.3
	$\gamma = 0.049$		$d_{yx} = 0.029$		$\gamma = 0.490$		$d_{yx} = 0.547$		$\gamma = 0.840$		$d_{xy} = 0.547$	

*Some totals are incorrect in the last figure because of rounding.

the subtables and has decreased both the number of discordantly ordered pairs and the number of pairs ordered on schooling but tied on attitude, the latter appearing only in the denominator of d_{yx}.

In general, standardization will more accurately portray the associations in the population from which the sample is drawn if random zeros are replaced by some *appropriate* nonzero values that can be changed by the iterative procedure. Typically, zero frequencies are eliminated by adding "pseudocounts" to some or all of the cells. The problem is deciding what these pseudocounts should be. Ideally, they should be close approximations of the expected sample frequencies of these cells.

Fienberg and Holland (a, b) have proposed a Bayesian procedure that uses the observed data and some prior model of the data to estimate these expected values. Here I shall briefly outline the basics of the procedure. For a fuller and more technical discussion of the theory, the reader should consult Fienberg and Holland's original articles. They present the method for two-way tables; however, as they point out, it can easily be extended to multiway ones. Following Fienberg and Holland's (a) argument closely, I shall outline the derivation of the calculating formula for the two-way case and then indicate how this formula may be applied to tables with more dimensions. At times, I shall need to refer to tables as single entities; following a conventional notation for matrices, I shall do so with roman upper-case letters.

Let us assume that the observed table, $F = [f_{ij}]$, with r rows and c columns is a sample table drawn from a multinormal distribution with cell probabilities $P = [p_{ij}]$ with

$$\sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} = 1.$$

If N is the total number of observations in F, then N P is the *expected* value of F. Let us also say that we have some *prior model* of P, $\Lambda = [\lambda_{ij}]$, such that

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \lambda_{ij} = 1$$

also.[11] Thus, $N\ \Lambda$ is our prior model of what N P is and of what F should be within the bounds of sampling fluctuations.

Given this information, a Bayesian esitmate, B $= [b_{ij}]$ of N P is

$$\mathrm{B} = \frac{N}{N + k}(\mathrm{F} + k\ \Lambda) \tag{3}$$

where k is selected to minimize the average squared difference between b_{ij} and $N\,p_{ij}$. This B is the "shrunk" table: F has been "shrunk" toward $k\ \Lambda$. As Fienberg and Holland (a, 238) observe, equation (3) is equivalent to adding k pseudocounts to the table, with $k\ \lambda_{ij}$ added to cell (i,j), and then adjusting all the cells by $N\ /\ (N + k)$ so that the total number of observations in the shrunk table still equals N.

The problem is that k is dependent on the unknown P; however, if we use F / N as a sample estimate of P, we can calculate an estimate of k, denoted here by $\hat{k}$ (it is $\hat{k}$ (λ) in Fienberg and Holland's notation). Replacing P with this sample estimate, the number of pseudocounts is determined by the observed data, F, and our prior model, $N\ \Lambda$:

$$\hat{k} = N^2 - \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij}^{\,2} \Big/ \sum_{i=1}^{r} \sum_{j=1}^{c} (N\ \lambda_{ij} - f_{ij})^2 \tag{4}$$

From the denominator of formula (4), we readily see that the closer $N\ \Lambda$ is to F, the larger is $\hat{k}$. Thus, the number of pseudo-counts $\hat{k}$ that we add to the table is dependent upon the accuracy of our model of the sample data. Indeed, by playing some algebra with (3) after substituting $\hat{k}$ for k, we see that the ratio of N to $\hat{k}$ indicates the relative weight given to F and to $N\ \Lambda$ in calculating the estimate B of N P:

$$\mathrm{B} = \frac{N}{N + \hat{k}}\,(\mathrm{F}) + \frac{\hat{k}}{N + \hat{k}}\,(N\ \Lambda). \tag{5}$$

Once we have $\hat{k}$, we can easily use formula (5) to calculate the shrunk table.

Since the procedure works with the values of individual cells, the number of dimensions in which these cells are arrayed really makes no difference. The $3 \times 3 \times 3$ Table 7 could, for purposes of the procedure, be considered a two-way 3×9 table as long as we maintained the correspondence between the cells of the observed table and those of our prior model.[12] Thus, the extension to more than two dimensions poses no problem. For a three-way table, we now have three subscripts for the observed cell frequencies f_{ijm} and the cell values of our prior model $N\ \lambda_{ijm}$.

(I have here changed the third subscript to avoid confusion with $\hat{k}$.) Formula (4) for the number of pseudocounts becomes

$$\hat{k} = N^2 - \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{m=1}^{a} f_{ijm}{}^2 / \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{m=1}^{a} (N\ \lambda_{ijm} - f_{ijm})^2, \tag{6}$$

and cell b_{ijm} of the shrunk table is

$$b_{ijm} = \frac{N}{N + \hat{k}} f_{ijm} + \frac{\hat{k}}{N + \hat{k}} (N\ \lambda_{ijm}). \tag{7}$$

What remains, of course, is calculating the model $N\ \Lambda$. Fienberg and Holland suggest three types of models. The first two are general purpose in that they require no substantive interpretations of the data; however, they generally provide relatively poor fits with the data and, hence, low values of $\hat{k}$.

1. The Constant Table.

In this model, every cell has an equal proportion of observations. With it, the same number of pseudocounts will be added to every cell of the observed table. In the three-way case, every $N\ \lambda_{ij\,m}$ equals $N / r\, c\, a$. This model will almost never have the same one-way distributions as the observed table. For Table 7, the constant model has a $\hat{k}$ of only 14.68, compared with an N of 1422.

2. The Independent Projection.

This is the model of completely independent distributions. Each cell is N times the product of the proportions of the one-way distributions in the categories forming the cell. In the three-way case,

$$N\ \lambda_{ijm} = N \cdot \frac{f_{i++}}{N} \cdot \frac{f_{+j+}}{N} \cdot \frac{f_{++m}}{N}.$$

In the two-way case, this model's $N\ \Lambda$ is the usual table of expected values used for chi-square tests. Since the general null hypothesis in most table analyses is independence, shrinking toward this independent projection is a conservative general-purpose procedure. On the other hand, interesting tables usually have some association in them and do not fit this model of independence; therefore, it generally gives fairly low values of $\hat{k}$. For Table 7, $\hat{k}$ has a value of 112.17 for the independent projection.

3. An Interaction Table.

If one has some substantive expectations about the data, either from theory or previous research, then he may be able to form a table whose interactions and associations are those expected in the observed table. The iterative procedure may, of course, be used to fit this "interaction table" to the one-way distributions and, as appropriate, the higher-order marginals of the observed table. In fact, we can use the iterative procedure to create a $N\ \Lambda$ model from the observed table itself by leaving out some of the observed interactions. For instance, to shrink Table 7, we can use the model of Table 7 without any three-way interaction by fitting a table of ones to the two-way marginals of Table 7. The cell frequencies of the resulting $N\ \Lambda$ table with only two-way interactions are given in parentheses in Table 9. While the differences among the associations in the three subtables of this $N\ \Lambda$ table are somewhat less than among those of Table 7, this is a far better model than the previous two, giving a $\hat{k}$ of 768.73. The result of shrinking Table 7 toward this model with no three-way interaction is shown in the unparenthesized portion of Table 9. From the conditional coefficients in Tables 7 and 9, we can see that the shrinking naturally has decreased the specification somewhat: there is more association between abortion attitudes and schooling for Roman Catholics in Table 9 than in Table 7 and less for the Jews. But the overall pattern is the same, and the changes in the values of the measures are generally small. The largest change, 0.180, for the gamma of the Jewish subtable is due in large part to the elimination of the two zeros. Examination of the individual cell frequencies and odds-ratios of the two tables also reveals that Table 7 has not been drastically altered by shrinking.

Table 9. AFTER SHRINKING TOWARD THE TABLE WITH NO THREE-WAY EFFECTS BUT WITH ALL TWO-WAY AND ONE-WAY ONES RETAINED: ATTITUDE TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH RELIGIOUS AFFILIATION CONTROLLED, FROM TABLE 7 (THE $N\lambda$ VALUES ARE IN PARENTHESES)*

	Religious Affiliation											
	Roman Catholic				Protestant				Jewish			
	Years of Schooling			Total	Years of Schooling			Total	Years of Schooling			Total
Abortion	1–11	12	13+		0–11	12	13+		0–11	12	13+	
Generally disapprove	70.7	64.4	14.0	149.0	183.3	95.9	7.8	287.0	1.0	1.7	0.3	3.0
	(81.2)	(57.6)	(10.2)		(172.7)	(103.2)	(11.1)		(1.1)	(1.2)	(0.8)	
Middle position	27.4	44.4	7.2	79.0	78.3	81.3	15.3	175.0	3.2	1.2	0.5	5.0
	(30.0)	(39.6)	(9.3)		(77.1)	(85.7)	(12.2)		(1.8)	(1.7)	(1.5)	
Generally approve	51.9	58.2	22.9	133.0	172.4	290.7	81.9	545.0	6.7	22.0	17.2	46.0
	(38.8)	(69.8)	(24.5)		(184.1)	(279.1)	(81.8)		(8.1)	(22.1)	(15.8)	
Total	150.0	167.0	44.0	361.0	434.0	468.0	105.0	1007.0	11.0	25.0	18.0	54.0
	$\gamma = 0.151$		$d_{yx} = 0.098$		$\gamma = 0.448$		$d_{yx} = 0.273$		$\gamma = 0.635$		$d_{yx} = 0.186$	

*The value of $\hat{k}$ used in the shrinking was 768.73. Some totals are incorrect in the last figure because of rounding.

Table 10 is the standardized version of the shrunk Table 9. Even though the data in Table 10 are more "artificial" than those in Table 8 (the standardization of Table 7 without shrinking), the pattern of association in the Jewish subtable probably better represents the population data now that the constraints of zeros have been removed from the standardization process. Notice, for instance, that the conditional γ and d_{yx} for the Jewish subtable are now much closer to those for the Protestant one. I should caution that the slight smoothing of the anomalies in the

Table 10. AFTER STANDARDIZING THE MARGINALS OF THE SHRUNK TABLE 9: ATTITUDE TOWARD ABORTION BY YEARS OF SCHOOLING COMPLETED WITH RELIGIOUS AFFILIATION CONTROLLED*

	Religious Affiliation											
	Roman Catholic				*Protestant*				*Jewish*			
	Years of Schooling			*Total*	*Years of Schooling*			*Total*	*Years of Schooling*			*Total*
Abortion	*1–11*	*12*	*13+*		*0–11*	*12*	*13+*		*0–11*	*12*	*13+*	
Generally disapprove	6.7	5.8	4.2	16.7	7.9	4.0	1.1	13.0	1.0	1.6	0.9	3.6
Middle position	3.6	5.6	3.1	12.3	4.7	4.7	3.0	12.5	4.5	1.7	2.4	8.6
Generally disapprove	1.2	1.3	1.8	4.3	1.9	3.1	2.9	7.8	1.7	5.4	14.1	21.2
Total	11.5	12.8	9.0	33.3	14.6	11.8	7.0	33.3	7.3	8.7	17.3	33.3
	$\gamma = 0.155$		$d_{yx} = 0.093$		$\gamma = 0.451$		$d_{yx} = 0.304$		$\gamma = 0.596$		$d_{yx} = 0.359$	

*Some totals are incorrect in the last figure because of rounding.

Roman Catholic subtable is at least in part because the model used in shrinking had no three-way interaction. Table shrinking is not a panacea for random zeros that can be used blindly without considering the implications of one's model. Nevertheless, there is little doubt that random zeros should be replaced by pseudocounts before standardization or other applications of the iterative procedure, and Fienberg and Holland's (b) research indicates that their way of using the data and a well-chosen model of it to select the pseudocounts is more accurate than other presently available ways.

CONCLUDING REMARKS

In this article, I have not attempted to advance the frontiers of table-analysis methodology. Instead, I have tried to bring marginal standardization and table shrinking to the attention of the typical sociologist working with percentaged tables and the traditional measures of association. I hope I have shown that these conceptually straightforward and easy-to-calculate techniques have a wide-ranging utility outside the advanced statistical approaches that fostered them. They should not remain within the limited domain of statisticians and methodological specialists.

NOTES

1. An example of *non*random, or fixed, zero cell count suggested by Fienberg and Holland (a, 235) is the number of "male obstetrical cases." Their example of a random one that might be zero in a particular sample but which is not ipso facto fixed to be zero is "Jewish farmers from Iowa."
2. For the sake of space and continuity, I have not included demonstrations of these effects. I shall be happy to supply them to the interested reader.
3. The methods for estimating and comparing various types of interaction, or association, in contingency tables that Goodman discusses in his 1969 article are specifically based on the logs of odds-ratios.
4. Respondents were asked if they approved or disapproved abortions performed for six different reasons. If a respondent disapproved more reasons than he approved, he was placed in the "generally disapprove" category ("don't know" and no responses were ignored); if he disapproved and approved an equal

number, he was placed in the "middle position" category; and if he approved more than he disapproved, he was placed in the "generally approve" category.
5. Those not answering this question or responding with either "don't know" or "as many as one wants" were eliminated from both Table 2 and Table 5.
6. Using Hawkes's terminology, the weight of a conditional d_{yx} is the variance of x in the subtable times the number of observations in the subtable, as a proportion of the sum of these N-times-variance products for all subtables. Standardization in this case has little effect on the variances of x (years of schooling) in the subtables. In Table 5, they are, from left to right, 0.302, 0.285, and 0.269; in Table 6, 0.325, 0.330, and 0.315. Clearly, in this case the change in the weighting of the conditional d_{yx}'s is predominantly due to the equalization of the number of observations in each subtable.
7. This is the procedure we use when the subtables represent cross-tabulated data for separate samples or populations which we wish to compare by standardizing each to the same equal marginals, the application discussed by Romney et al.
8. If a cell frequency is a nonrandom zero with no conceivable observation appearing in it, the cell can be assigned a zero in the independence table and will remain zero throughout the iteration.
9. For direct significance tests of various aspects of the core patterns of association, see Goodman's (a, b, c, d) articles and others on the log-linear model for contingency tables, such as those by Bishop (a, b) and Grizzle et al.
10. Those people responding with other religious affiliations and with no religious affiliation have been omitted for the illustrative purposes here.
11. Fienberg and Holland use X where I am using F, p instead of P, and λ instead of Λ.
12. In their 1973 article, Fienberg and Holland work with the cells arrayed in a *vector* rather than a matrix.

REFERENCES

Bishop, Y. M. M. a:1969. "Full Contingency Tables, Logits, and Split Contingency Tables." *Biometrics* 25:383-400.

______. b:1971. "Effects of Collapsing Multidimensional Contingency Tables." *Biometrics* 27:545-62.

Blalock, H. M., Jr. 1967. "Causal Inferences, Closed Populations, and Measures of Association." *American Political Science Review* 61:130-6.

Davis, J. A. 1967. "A Partial Coefficient for Goodman and Kruskal's Gamma." *Journal of the American Statistical Association* 62(March):189-93.

Fienberg, S. E. 1971. "A Statistical Technique for Historians: Standardizing Tables of Counts." *Journal of Interdisciplinary History* 1:305-15.

Fineberg, S. E., and P. W. Holland. a:1970. "Methods for Eliminating Zero Counts in Contingency Tables." In G. P. Patil (ed.), *Random Counts in Models and Structures*. University Park: Pensylvania State University Press.

______. b:1973. "Simultaneous Estimation of Multinomial Cell Probabilities." *Journal of the American Statistical Association* 68(September):683-91.

Goodman, L. A. a:1969. "How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables." *American Journal of Sociology* 75(July):1-40.

______. b:1972. "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables." *American Sociological Review* 37(February):28-46.

______. c:1972. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77(May):1035-86.

______. d:1973. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." *American Journal of Sociology* 78(March):1135-91.

Goodman, L. A., and W. H. Kruskal. 1954. "Measures of Association for Cross-Classifications." *Journal of the American Statistical Association* 49(December):732-64.

Grizzle, J. E., C. F. Starmer, and G. G. Koch. 1969. "Analysis of Categorical Data by Linear Models." *Biometrics* 25:489-504.

Hawkes, R. K. 1971. "The Multivariate Analysis of Ordinal Measures." *American Journal of Sociology* 76(March):903-26.

Mosteller, F. 1968. "Association and Estimation in Contingency Tables." *Journal of the American Statistical Association* 63(March):1-28.

National Data Program for the Social Sciences. 1972. "Codebook for the Spring 1972 General Social Survey." Chicago: National Opinion Research Center (multilithed).

Ploch, D. R. 1974. "Ordinal Measures of Association and the General Linear Model." In Hubert M. Blalock, Jr. (ed.), *Measurement in the Social Sciences: Theories and Strategies*. Chicago: Aldine.

Romney, A. K., M. Kieffer, and R. E. Klein. 1973. "A Normalization Procedure for Correcting Biased Response Data." *Social Science Research* 2(December): 307-20.

Rosenberg, M. 1962. "Test Factor Standardization as a Method of Interpretation." *Social Forces* 41(October):53-61.

Schoenberg, R. 1972. "Strategies for Meaningful Comparison." In Herbert L. Costner (ed.), *Sociological Methodology*, 1972. San Francisco: Jossey–Bass.

Somers, R. H. 1962. "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review* 27(December):799-811.