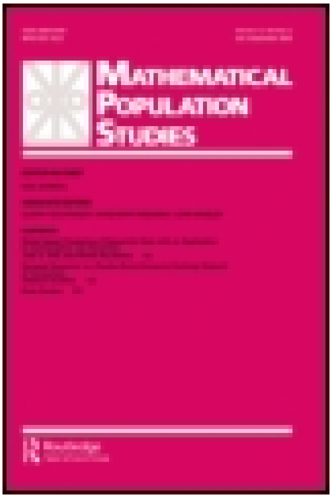

Mathematical Population Studies: An International Journal of Mathematical Demography

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmps20

Decomposition of the difference between two rates and its consistency when more than two populations are involved

Prithwis Das Gupta [a]

[a] Population Division , U.S. Bureau of the Census , Washington, DC, 20233

Published online: 21 Sep 2009.

To cite this article: Prithwis Das Gupta (1991) Decomposition of the difference between two rates and its consistency when more than two populations are involved, Mathematical Population Studies: An International Journal of Mathematical Demography, 3:2, 105-125, DOI: 10.1080/08898489109525329

To link to this article: http://dx.doi.org/10.1080/08898489109525329

Mathematical Population Studies, 1991, Vol. 3(2), pp. 105–125
Reprints available directly from the publisher
Photocopying permitted by license only

Printed in the United States of America

DECOMPOSITION OF THE DIFFERENCE BETWEEN TWO RATES AND ITS CONSISTENCY WHEN MORE THAN TWO POPULATIONS ARE INVOLVED

PRITHWIS DAS GUPTA

Population Division
U.S. Bureau of the Census
Washington, DC 20233

July 17, 1990, revised March 6, 1991

In recent years, many social scientists, in dealing with data of various kinds, have attempted to decompose the difference between two rates into effects of changes in the underlying factors. All these studies can be divided into two categories, viz., the rate as a function of two or more factors (their product being a special case), and the rate as computed from cross-classified data involving one or more factors. In the absence of general methods of decomposition to be used under these two circumstances, the researchers have devised their own ad hoc methods which in general produce meaningful results, but are sometimes less than satisfactory in terms of mathematical rigor and elegance. The objective of the present paper is to put together the general formulas that a researcher is expected to need for the two categories of decomposition problems for any number of factors. Applications of the use of the formulas under these two situations are illustrated by four examples. The paper also suggests a solution to the problem of possible internal inconsistencies of the factor effects when more than two populations are involved, and provides an illustrative example.

KEY WORDS: Decomposition, cross-classified data, consistency, life table, illegitimacy ratio, proportion childless.

1. INTRODUCTION

In recent years, many social scientists have attempted to decompose the difference between two indexes (rates, means, proportions, ratios, etc.) into effects of changes in the underlying factors, the latter depending on the scope of the available data. These studies include the decomposition of the change in intrinsic growth rate into effects of changes in birth rates and death rates (Keyfitz 1968, p. 189); fertility differentials of women by education and their relationships to age at marriage and labor force status (Janowitz 1976, Das Gupta 1984); contributions of different age groups to the difference between the expectations of life in two life tables (Arriaga 1984, Pollard 1988); the growth of households as a result of the changes in age and marital status composition (Sweet 1984); the effect of changes in occupational structure and occupational sex segregation on the index of occupational dissimilarity between sexes, and a technically similar study on spatial segregation of Blacks and

Whites across U.S. countries (Lichter 1985, Bianchi and Rytina 1986, Das Gupta 1987); the relationship between changes in age and education composition and the corresponding changes in out-migration rates (Wilson 1988); illegitimate ratio as a function of age and marital status structures and marital and nonmarital birth rates (Smith and Cutright 1988); the decomposition of the changes in unrelated individuals into effects of changes in age, sex, marital status, occupation, and mobility (Ruggles 1988); the disaggregation of changes in adolescent fertility into changes in marriage patterns, nonmarital sex, pregnancy, and birth (Nathanson and Kim 1989); the change in mean parity of women in terms of parity progression ratios (Pullum, Tedrow, and Herting 1989); and finally, the growth of female-headed families as a result of changes in demographic behavior such as marriage, fertility, divorce, and the propensity to live alone (Wojtkiewicz, McLanahan, and Garfinkel 1990).

All these studies can be divided into two broad categories. The first category deals with a rate that can be expressed as a *function* of two or more factors (e.g., Arriaga 1984; Pullum, Tedrow, and Herting 1989). A special case of this category is the situation in which the rate can be expressed as the *product* of two or more factors (e.g., Nathanson and Kim 1989; Wojtkiewicz, McLanahan, and Garfinkel 1990). The second and more common category is the case in which the decomposition is performed on *cross-classified data* involving one or more factors (e.g., Janowitz 1967; Ruggles 1988). Most of the technical papers on decomposition deal with this latter category (Kitagawa 1955; Cho and Retherford 1973; Kim and Strobino 1984; Das Gupta 1978, 1989; Liao 1989). The objective of the present paper is to discuss the two areas of decompositional problems in the next two sections and provide exact explicit general formulas that can be conveniently used for any number of factors. All these formulas appear in this paper for the first time, although their developments are linked to some earlier publications (Kitagawa 1955; Kim and Strobino 1984; Das Gupta 1989). Four examples are given to illustrate the applications of the formulas in the two categories of decomposition. The paper also develops correction terms to be added to the factor effects when more than two populations are involved, in order to make the results internally consistent, and illustrates their applications with an example.

2. RATE AS A FUNCTION OF FACTORS

If there are two factors α and β, the rate $F(\alpha,\beta)$ in this case is a function of these factors. If the factors take on values A,B and a,b in populations 1 and 2, respectively, then the difference between the rates can be expressed as (the formulas in this section are derived by proceeding as in section 1 in the Appendix)

$$F(a,b)-F(A,B)=\frac{[F(a,b)-F(A,b)]+[F(a,B)-F(A,B)]}{2}$$

$$+\frac{[F(a,b)-F(a,B)]+[F(A,b)-F(A,B)]}{2}$$

$$=\alpha\text{-effect}+\beta\text{-effect}. \qquad (1)$$

When there are three factors α, β, and γ, we can use similar notation to decompose the difference as

$$F(a,b,c) - F(A,B,C) = \alpha\text{-effect} + \beta\text{-effect} + \gamma\text{-effect},$$

where

$$\alpha\text{-effect} = \frac{[F(a,b,c) - F(A,b,c)] + [F(a,B,C) - F(A,B,C)]}{3} + \frac{[F(a,b,C) - F(A,b,C)] + [F(a,B,c) - F(A,B,c)]}{6},$$

$$\beta\text{-effect} = \frac{[F(a,b,c) - F(a,B,c)] + [F(A,b,C) - F(A,B,C)]}{3} + \frac{[F(a,b,C) - F(a,B,C)] + [F(A,b,c) - F(A,B,c)]}{6},$$

$$\gamma\text{-effect} = \frac{[F(a,b,c) - F(a,b,C)] + [F(A,B,c) - F(A,B,C)]}{3} + \frac{[F(a,B,c) - F(a,B,C)] + [F(A,b,c) - F(A,b,C)]}{6}. \quad (2)$$

When there are four factors α, β, γ, and δ, the difference $F(a,b,c,d) - F(A,B,C,D)$ can be written as the sum of four factor effects where the α-effect is given by

$$\alpha\text{-effect} = \frac{[F(a,b,c,d) - F(A,b,c,d)] + [F(a,B,C,D) - F(A,B,C,D)]}{4} + \frac{\begin{array}{c}[F(a,b,c,D) - F(A,b,c,D)] + [F(a,b,C,d) - F(A,b,C,d)] \\ +[F(a,B,c,d) - F(A,B,c,d)] + [F(a,B,C,d) - F(A,B,C,d)] \\ +[F(a,B,c,D) - F(A,B,c,D)] + [F(a,b,C,D) - F(A,b,C,D)]\end{array}}{12}. \quad (3)$$

Other effects have similar expressions.

We can similarly express the α-effects corresponding to five and six factor cases, respectively, as

$$\alpha\text{-effect} = \frac{[F(a,b,c,d,e) - F(A,b,c,d,e)] + [F(a,B,C,D,E) - F(A,B,C,D,E)]}{5} + \frac{\begin{array}{c}[F(a,b,c,d,E) - F(A,b,c,d,E)] + [F(a,b,c,D,e) - F(A,b,c,D,e)] \\ +[F(a,b,C,d,e) - F(A,b,C,d,e)] + [F(a,B,c,d,e) - F(A,B,c,d,e)] \\ +[F(a,B,C,D,e) - F(A,B,C,D,e)] + [F(a,B,C,d,E) - F(A,B,C,d,E)] \\ +[F(a,B,c,D,E) - F(A,B,c,D,E)] + [F(a,b,C,D,E) - F(A,b,C,D,E)]\end{array}}{20} + \frac{\begin{array}{c}[F(a,b,c,D,E) - F(A,b,c,D,E)] + [F(a,b,C,d,E) - F(A,b,C,d,E)] \\ +[F(a,b,C,D,e) - F(A,b,C,D,e)] + [F(a,B,C,d,e) - F(A,B,C,d,e)] \\ +[F(a,B,c,D,e) - F(A,B,c,D,e)] + [F(a,B,c,d,E) - F(A,B,c,d,E)]\end{array}}{30}, \quad (4)$$

$$\alpha\text{-effect} = \frac{[F(a,b,c,d,e,f) - F(A,b,c,d,e,f)] + [F(a,B,C,D,E,F) - F(A,B,C,D,E,F)]}{6}$$

$$+ \frac{\begin{array}{l}[F(a,b,c,d,e,F) - F(A,b,c,d,e,F)] + [F(a,b,c,d,E,f) - F(A,b,c,d,E,f)] \\ +[F(a,b,c,D,e,f) - F(A,b,c,D,e,f)] + [F(a,b,C,d,e,f) - F(A,b,C,d,e,f)] \\ +[F(a,B,c,d,e,f) - F(A,B,c,d,e,f)] + [F(a,B,C,D,E,f) - F(A,B,C,D,E,f)] \\ +[F(a,B,C,D,e,F) - F(A,B,C,D,e,F)] + [F(a,B,C,d,E,F) - F(A,B,C,d,E,F)] \\ +[F(a,B,c,D,E,F) - F(A,B,c,D,E,F)] + [F(a,b,C,D,E,F) - F(A,b,C,D,E,F)]\end{array}}{30}$$

$$+ \frac{\begin{array}{l}[F(a,b,c,d,E,F) - F(A,b,c,d,E,F)] + [F(a,b,c,D,e,F) - F(A,b,c,D,e,F)] \\ +[F(a,b,c,D,E,f) - F(A,b,c,D,E,f)] + [F(a,b,C,d,e,F) - F(A,b,C,d,e,F)] \\ +[F(a,b,C,d,E,f) - F(A,b,C,d,E,f)] + [F(a,b,C,D,e,f) - F(A,b,C,D,e,f)] \\ +[F(a,B,c,d,e,F) - F(A,B,c,d,e,F)] + [F(a,B,c,d,E,f) - F(A,B,c,d,E,f)] \\ +[F(a,B,c,D,e,f) - F(A,B,c,D,e,f)] + [F(a,B,C,d,e,f) - F(A,B,C,d,e,f)] \\ +[F(a,B,C,D,e,f) - F(A,B,C,D,e,f)] + [F(a,B,C,d,E,f) - F(A,B,C,d,E,f)] \\ +[F(a,B,C,d,e,F) - F(A,B,C,d,e,F)] + [F(a,B,c,D,E,f) - F(A,B,c,D,E,f)] \\ +[F(a,B,c,D,e,F) - F(A,B,c,D,e,F)] + [F(a,B,c,d,E,F) - F(A,B,c,d,E,F)] \\ +[F(a,b,C,D,E,f) - F(A,b,C,D,E,f)] + [F(a,b,C,D,e,F) - F(A,b,C,D,e,F)] \\ +[F(a,b,C,d,E,F) - F(A,b,C,d,E,F)] + [F(a,b,c,D,E,F) - F(A,b,c,D,E,F)]\end{array}}{60}. \quad (5)$$

The general formula for the α-effect for P factors (see section 2 in the Appendix) is derivable from the facts that (i) the number of terms for the α-effect is $P/2$ or $(P+1)/2$ depending on whether P is even or odd, (ii) the denominators in the terms are $P\binom{P-1}{0}, P\binom{P-1}{1}, P\binom{P-1}{2}, \ldots$, and (iii) the numerator of the rth term in the α-effect contains only those individual differences of functions F whose arguments (except the first) consist of either $(P-r)$ small and $(r-1)$ capital letters or $(P-r)$ capital and $(r-1)$ small letters. The general formula (A5) is given in Das Gupta (1990) in a somewhat different form.

2.1. Rate as the Product of Factors

As a special case, when the rate $F(\alpha,\beta)$ is the product $\alpha\beta$, then (1) reduces to

$$ab - AB = \left[\frac{b+B}{2}\right](a-A) + \left[\frac{a+A}{2}\right](b-B) = \alpha\text{-effect} + \beta\text{-effect}. \quad (6)$$

When the rate is the product of three factors α, β, and γ, (2) simplifies to

$$\begin{aligned} abc - ABC &= \left[\frac{bc+BC}{3} + \frac{bC+Bc}{6}\right](a-A) \\ &\quad + \left[\frac{ac+AC}{3} + \frac{aC+Ac}{6}\right](b-B) \\ &\quad + \left[\frac{ab+AB}{3} + \frac{aB+Ab}{6}\right](c-C) \\ &= \alpha\text{-effect} + \beta\text{-effect} + \gamma\text{-effect}. \end{aligned} \quad (7)$$

When the rate is the product of four, five, and six factors, the decomposition pertains to the differences $abcd - ABCD$, $abcde - ABCDE$, and $abcdef - ABCDEF$, respectively. The α-effects in these cases directly follow from (3)–(5) and are given by

$$\alpha\text{-effect} = \left[\frac{bcd + BCD}{4} + \frac{bcD + bCd + Bcd + BCd + BcD + bCD}{12}\right](a - A), \tag{8}$$

$$\alpha\text{-effect} = \left[\frac{bcde + BCDE}{5} + \frac{\begin{array}{c} bcdE + bcDe + bCde + Bcde \\ +BCDe + BCdE + BcDE + bCDE \end{array}}{20} + \frac{bcDE + bCdE + bCDe + BCde + BcDe + BcdE}{30}\right](a - A), \tag{9}$$

$$\alpha\text{-effect} = \left[\frac{bcdef + BCDEF}{6} + \frac{\begin{array}{c} bcdeF + bcdEf + bcDef + bCdef + Bcdef \\ +BCDEf + BCDeF + BCdEF + BcDEF + bCDEF \end{array}}{30} + \frac{\begin{array}{c} bcdEF + bcDeF + bcDEf + bCdeF + bCdEf \\ +bCDef + BcdeF + BcdEf + BcDef + BCdef \\ +BCDef + BCdEf + BCdeF + BcDEf + BcDeF \\ +BcdEF + bCDEf + bCDeF + bCdEF + bcDEF \end{array}}{60}\right](a - A). \tag{10}$$

The expressions for the other effects can be derived easily from those for the α-effects. For example, the β-effect is obtained from the α-effect by substituting b, a, B, and A for a, b, A, and B, respectively.

From (6)–(10), it is easy to extend the formulas when the rate is the product of any number of factors P (see section 3 in the Appendix) from the facts that (i) for any effect, the number of terms within the parentheses is $P/2$ or $(P+1)/2$ depending on whether P is even or odd, (ii) the denominators in the terms within the parentheses are $P\binom{P-1}{0}, P\binom{P-1}{1}, P\binom{P-1}{2}, \ldots,$ and (iii) the numerator of the rth term within the parentheses contains only those individual product terms whose letters consist of either $(P-r)$ small and $(r-1)$ capital letters or $(P-r)$ capital and $(r-1)$ small letters. For example, when $P = 6$, $r = 3$, the numerator contains terms whose letters consist of either 3 small and 2 capital letters or 3 capital and 2 small letters, as we can easily verify from (10). This exact general formula (A6) for P factors is given here for the first time, although its approximation for P greater than four is provided in Das Gupta (1989, formula A2). Since the exact formula is simple enough, it may no longer be necessary to use the approximate formula.

2.2. Examples

Lines 7–10 in Table 1 provide data for the application of the formulas in (7) when the rate is the product of three factors. The crude birth rate per person for a popu-

TABLE 1
Population and Birth Data Used for the Decomposition of the Difference Between the Crude Birth Rates and Between the Illegitimacy Ratios: United States, 1960 and 1980

Description	1960	1980
1. Total Population	179,323,175	226,545,805
2. Female Population	90,991,681	116,492,644
3. Female, 15-44 years	36,078,973	52,833,115
4. Unmarried Female, 15-44 years	10,274,373	24,918,036
5. Total Births	4,257,850	3,612,258
6. Births to Unmarried Female, 15-44 years	224,300	665,747
7. Crude Birth Rate [(5)/(1)] = $\alpha\beta\gamma$	.0237440	.0159449
8. General Fertility Rate [(5)/(3)] = α	.1180147 (a)	.0683711 (A)
9. Proportion of Female, 15-44 years [(3)/(2)] = β	.3965085 (b)	.4535318 (B)
10. Proportion of Female [(2)/(1)] = γ	.5074173 (c)	.5142123 (C)
11. Illegitimacy Ratio [(6)/(5)] = $\alpha\beta/[\alpha\beta+(1-\alpha)\gamma]$	.0526792	.1843022
12. Proportion of Unmarried Female, 15-44 years [(4)/(3)] = α	.2847745 (A)	.4716367 (a)
13. Illegitimacy Rate [(6)/(4)] = β	.0218310 (B)	.0267175 (b)
14. Marital Fertility Rate [{(5)-(6)}/{(3)-(4)}] = γ	.1563113 (C)	.1055527 (c)

Sources: U.S. Bureau of the Census (1964, Tables 46, 176; 1983, Tables 43, 46); U.S. National Center for Health Statistics (1963, Tables 1-A, 1-V; 1984, Tables 1-1, 1-71).

lation can be expressed as

$$R = \frac{B}{W_{15\text{–}44}} \cdot \frac{W_{15\text{–}44}}{W} \cdot \frac{W}{P} = \alpha\beta\gamma, \tag{11}$$

where α is the general fertility rate per woman aged 15–44, β is the proportion of women aged 15–44 among all women, and γ is the proportion of women in the total population. Treating the U.S. populations in 1980 and 1960 as populations 1 and 2, the difference between the crude birth rates is .00780. Using formula (11) and applying formulas in (7) to the data on lines 7–10 in Table 1, we decompose this difference into .01078, $-$.00271, and $-$.00027, which are, respectively, the effects (in the direction of 1980 to 1960) of the rise in the general fertility rate and the decline in the proportion of women aged 15–44 and in the proportion of women in the population. This means that the crude birth rate (per 1000 population) in 1960 was 7.80 points higher than that in 1980. However, this would have been 10.78 points *higher* if only the general fertility rate changed, 2.71 points *lower* if only the proportion of women aged 15–44 changed, and 0.27 point *lower* if only the proportion of women in the population changed. In other words, the latter two factors tend to depress the difference in the crude birth rates between 1960 and

1980. Had all the factor effects been positive, it would have been interesting to give the results in terms of percent contributions of the three factors to the total difference in the crude birth rates (as in some latter examples).

If the data are available by age groups, then we can modify Eq. (11) for the crude birth rate as

$$R(\alpha,\delta,\beta,\gamma) = \sum_i \frac{B_i}{W_i} \cdot \frac{W_i}{W_{15\text{–}44}} \cdot \frac{W_{15\text{–}44}}{W} \cdot \frac{W}{P} = \sum_i \alpha_i \delta_i \beta \gamma, \tag{12}$$

where α and δ are the vectors having the respective age-specific values as their elements. If all the elements in a vector assume values either for population 1 or for population 2 as a group, then this becomes a 4-factor case, and we can use the 4-factor formulas (8) to decompose

$$\sum_i (a_i d_i bc - A_i D_i BC), \tag{13}$$

by summing them over all age groups i in the childbearing years to obtain the final factor effects.

The application of the more general formulas (2) for three factors can be illustrated with the help of the data on lines 11–14 in Table 1. Births (I) to unmarried women and those (L) to married women can be expressed as

$$\begin{aligned} I &= W_{15\text{–}44} \cdot \frac{U_{15\text{–}44}}{W_{15\text{–}44}} \cdot \frac{I}{U_{15\text{–}44}} = W_{15\text{–}44} \cdot \alpha\beta, \\ L &= W_{15\text{–}44} \cdot \frac{M_{15\text{–}44}}{W_{15\text{–}44}} \cdot \frac{L}{M_{15\text{–}44}} = W_{15\text{–}44} \cdot (1-\alpha)\gamma, \end{aligned} \tag{14}$$

where $W_{15\text{–}44}$ is the number of women aged 15–44 of which $U_{15\text{–}44}$ women are unmarried and $M_{15\text{–}44}$ are married, α is the proportion of unmarried women in the childbearing years, β is the illegitimacy rate, and γ is the marital fertility rate. Defining illegitimacy ratio $F(\alpha,\beta,\gamma)$ as the ratio of births to unmarried women to total births, we have

$$F(\alpha,\beta,\gamma) = \frac{\alpha\beta}{\alpha\beta + (1-\alpha)\gamma}. \tag{15}$$

Treating the U.S. populations in 1960 and 1980 as populations 1 and 2, the difference between the illegitimacy ratios is .1316. Using formula (15) and applying formulas (2) to the data on lines 11–14 in Table 1, we decompose this difference into .0686, .0365, and .0265, which are, respectively, the effects of decline in nuptiality, increase in illegitimacy rate, and decrease in marital fertility. Converting these numbers into percentages, we can say that these three factors explain, respectively, 52.1, 27.7, and 20.2 percents of the difference between the illegitimacy ratios in 1960 and 1980.

As in the case of the first example, if the data are available by age groups (Smith and Cutright 1988), then I and L in (14) can be expressed as

$$I = \sum_i W_{15\text{–}44} \cdot \frac{W_i}{W_{15\text{–}44}} \cdot \frac{U_i}{W_i} \cdot \frac{I_i}{U_i} = \sum_i W_{15\text{–}44} \delta_i \alpha_i \beta_i,$$
$$L = \sum_i W_{15\text{–}44} \cdot \frac{W_i}{W_{15\text{–}44}} \cdot \frac{M_i}{W_i} \cdot \frac{L_i}{M_i} = \sum_i W_{15\text{–}44} \delta_i (1-\alpha_i)\gamma_i, \tag{16}$$

where α_i, β_i, γ_i, and δ_i are values that are specific for age group i. The illegitimacy ratio (15) in this case takes the form

$$F(\delta,\alpha,\beta,\gamma) = \frac{\sum_i \delta_i \alpha_i \beta_i}{\sum_i \delta_i \alpha_i \beta_i + \sum_i \delta_i (1-\alpha_i)\gamma_i}, \tag{17}$$

where δ, α, β, and γ are the vectors with respective age-specific values as their elements. If all the elements in a vector take on the values either for population 1 or for population 2, then the difference between the two illegitimacy ratios can be decomposed as a 4-factor case by using the 4-factor formulas in (3).

3. CROSS-CLASSIFIED DATA

When there is only one factor I, N_i and T_i are the number of persons and the rate for the ith category of I in population 1, $N_.$ and $T_.$ being the corresponding total number of persons and the crude rate. For population 2, analogous symbols are used with lower-case letters n and t. The difference between the two crude rates can be decomposed as (Kitagawa 1955)

$$t_. - T_. = \sum_i \left(\frac{t_i n_i}{n_.} - \frac{T_i N_i}{N_.} \right) = \sum_i \frac{t_i + T_i}{2} \left(\frac{n_i}{n_.} - \frac{N_i}{N_.} \right) + \sum_i \frac{\frac{n_i}{n_.} + \frac{N_i}{N_.}}{2} (t_i - T_i), \tag{18}$$

where the first and the second terms on the right-hand side are, respectively, the I-effect and the rate effect. This decomposition is similar to that in (6).

When there are two factors I and J, N_{ij} and T_{ij} are the number of persons and the rate for the (i,j)-category in population 1; $N_{i.}$ and $T_{i.}$ are the number of persons and the rate for the ith category of I, and $N_{.j}$ and $T_{.j}$ are the corresponding number of persons and the rate for the jth category of J. As before, $N_{..}$ and $T_{..}$ are the total number of persons and the crude rate. Analogous symbols are used for population 2 with lower-case letters n and t. As in (18), the difference between the two rates is first decomposed as

$$t_{..} - T_{..} = \sum_{i,j} \frac{t_{ij} + T_{ij}}{2} \left(\frac{n_{ij}}{n_{..}} - \frac{N_{ij}}{N_{..}} \right) + \sum_{i,j} \frac{\frac{n_{ij}}{n_{..}} + \frac{N_{ij}}{N_{..}}}{2} (t_{ij} - T_{ij}), \tag{19}$$

where the first and the second terms are, respectively, the combined IJ-effect and the rate effect.

In order to further decompose the combined IJ-effect in (19) into I-effect and J-effect, we write (this and other formulas in this section follow from the derivation in section 4 in the Appendix)

$$\frac{n_{ij}}{n_{..}} = \left(\frac{n_{ij}}{n_{.j}} \cdot \frac{n_{i.}}{n_{..}}\right)^{1/2} \cdot \left(\frac{n_{ij}}{n_{i.}} \cdot \frac{n_{.j}}{n_{..}}\right)^{1/2} = ab, \tag{20}$$

in which the two ratios in the first term represent only the I-effect and the two ratios in the second term represent only the J-effect. Expressing $N_{ij}/N_{..}$ as AB, as in (20), and using formula (6), we can finally decompose the combined IJ-effect in (19) into

$$\begin{aligned} I\text{-effect} &= \sum_{i,j} \frac{t_{ij}+T_{ij}}{2} \frac{\left(\frac{n_{ij}}{n_{i.}} \cdot \frac{n_{.j}}{n_{..}}\right)^{1/2} + \left(\frac{N_{ij}}{N_{i.}} \cdot \frac{N_{.j}}{N_{..}}\right)^{1/2}}{2} \\ &\quad \times \left[\left(\frac{n_{ij}}{n_{.j}} \cdot \frac{n_{i.}}{n_{..}}\right)^{1/2} - \left(\frac{N_{ij}}{N_{.j}} \cdot \frac{N_{i.}}{N_{..}}\right)^{1/2}\right], \\ J\text{-effect} &= \sum_{i,j} \frac{t_{ij}+T_{ij}}{2} \frac{\left(\frac{n_{ij}}{n_{.j}} \cdot \frac{n_{i.}}{n_{..}}\right)^{1/2} + \left(\frac{N_{ij}}{N_{.j}} \cdot \frac{N_{i.}}{N_{..}}\right)^{1/2}}{2} \\ &\quad \times \left[\left(\frac{n_{ij}}{n_{i.}} \cdot \frac{n_{.j}}{n_{..}}\right)^{1/2} - \left(\frac{N_{ij}}{N_{i.}} \cdot \frac{N_{.j}}{N_{..}}\right)^{1/2}\right]. \end{aligned} \tag{21}$$

The rate effect in (19) and the I-effect and J-effect in (21) together give the exact decomposition of the difference between the two rates. Obviously, the I-effect vanishes when the proportional distributions of I both in the population and within each j are identical in the two populations. A similar comment can be made for the J-effect. Unlike the *hierarchical* approaches by Cho and Retherford (1973) and Kim and Strobino (1984), the effects of the factors in this decomposition remain unchanged irrespective of which one of the factors is regarded as I and which one as J. This *symmetrical* treatment of I and J is also different from the symmetrical approaches suggested by Kitagawa (1955), Das Gupta (1978, 1989) and Liao (1989).

Table 2 compares the decompositions of the difference between the crude death rates of 1970 and 1985 for the United States in terms of two factor effects—age (I) and race (J)—and the rate effect corresponding to some of these methods. Although we do not see any difference between the results from Methods 4 and 6 for this particular set of two-factor data, in most cases, they are expected to be somewhat different, particularly for a higher number of factors (the rate effects are identical by definition). Both Methods 4 and 6 have the advantage of symmetrical treatment of factors without any interaction terms. Method 6, however, should be preferred because of its significant computational advantage over Method 4. The latter, for example, involves similar computations twice—once with age as factor I and race as factor J, and then age as factor J and race as factor I. This number of repetitions of similar computations increases drastically with the increase in the number of factors. For example, for five factors, this number is $5!(=120)$. As we

TABLE 2
The Decompositions of the Difference Between the Crude Death Rates by Six Methods: United States, 1970 and 1985[a]

Method	Difference (1970) -(1985)	Effects			
		Age (I)	Race (J)	I x J Interaction	Rate
Hierarchical Factors					
1. Cho and Retherford (1973)	.6828	-1.4847	-.0606	--	2.2281
2. Kim and Strobino (1984)	.6828	-1.4845	-.0607	--	2.2280
Symmetrical Factors					
3. Kitagawa (1955)	.6828	-1.5657	-.0606	.0813	2.2278
4. Das Gupta (1978)	.6828	-1.5250	-.0200	--	2.2278
5. Liao (Marginal CG) (1989)[b]	.683	-1.568	-.059	.079	2.231
6. Proposed Method (formulas 19 and 21)	.6828	-1.5250	-.0200	--	2.2278

[a]*Source of data*: Liao (1989), Table 1. The White population (in 000's) aged 35–44 in 1985 is corrected from 27,444 to 27,411.
[b]The results correct to 3 places of decimal are taken from the article.

have seen for two factors, the proposed method involves only one set of computations regardless of the number of factors considered.

For three factors I, J, and K, using appropriate notation, the difference $t_{...} - T_{...}$ is first decomposed into the combined IJK-effect and the rate effect as in (19). The combined IJK-effect is then decomposed by writing

$$\frac{n_{ijk}}{n_{...}} = abc, \qquad \text{where}$$

$$a = \left(\frac{n_{ijk}}{n_{.jk}}\right)^{1/3} \cdot \left(\frac{n_{ij.}}{n_{.j.}} \cdot \frac{n_{i.k}}{n_{..k}}\right)^{1/6} \cdot \left(\frac{n_{i..}}{n_{...}}\right)^{1/3},$$
$$b = \left(\frac{n_{ijk}}{n_{i.k}}\right)^{1/3} \cdot \left(\frac{n_{ij.}}{n_{i..}} \cdot \frac{n_{.jk}}{n_{..k}}\right)^{1/6} \cdot \left(\frac{n_{.j.}}{n_{...}}\right)^{1/3}, \tag{22}$$
$$c = \left(\frac{n_{ijk}}{n_{ij.}}\right)^{1/3} \cdot \left(\frac{n_{i.k}}{n_{i..}} \cdot \frac{n_{.jk}}{n_{.j.}}\right)^{1/6} \cdot \left(\frac{n_{..k}}{n_{...}}\right)^{1/3}.$$

The expressions a, b, and c contain only the I, J, and K effects, respectively. Expressing $N_{ijk}/N_{...}$ as ABC, as in (22), and using formula (7), we obtain the so-called α, β, and γ effects. These latter effects, when multiplied by $(t_{ijk} + T_{ijk})/2$ and then summed over i, j, and k, will give the final I, J, and K effects. The I-effect is zero when $a - A$ is zero, i.e., when the proportional distributions of I in the population, within each j, within each k, and within each (j,k) are identical in the two populations. Similar comments hold for the J and K effects.

The decomposition for more than three factors is performed in the same manner, i.e., the difference between the two rates is first decomposed into the rate effect and the combined factor effects. The latter is then decomposed into the individual factor effects by first writing the cell proportions as the product of several expressions representing the factors and then applying the formulas such as (8)–(10). The most crucial part in the decomposition is obtaining expressions such as those in (22) for a higher number of factors.

For four and five factors, the expressions for a are, respectively, as follows:

$$a = \left(\frac{n_{ijkl}}{n_{.jkl}}\right)^{1/4} \cdot \left(\frac{n_{ijk.}}{n_{.jk.}} \cdot \frac{n_{ij.l}}{n_{.j.l}} \cdot \frac{n_{i.kl}}{n_{..kl}}\right)^{1/12}$$
$$\cdot \left(\frac{n_{ij..}}{n_{.j..}} \cdot \frac{n_{i.k.}}{n_{..k.}} \cdot \frac{n_{i..l}}{n_{...l}}\right)^{1/12} \cdot \left(\frac{n_{i...}}{n_{....}}\right)^{1/4}, \tag{23}$$

$$a = \left(\frac{n_{ijklm}}{n_{.jklm}}\right)^{1/5} \cdot \left(\frac{n_{ijkl.}}{n_{.jkl.}} \frac{n_{ijk.m}}{n_{.jk.m}} \cdot \frac{n_{ij.lm}}{n_{.j.lm}} \cdot \frac{n_{i.klm}}{n_{..klm}}\right)^{1/20}$$
$$\cdot \left(\frac{n_{ijk..}}{n_{.jk..}} \cdot \frac{n_{ij.l.}}{n_{.j.l.}} \cdot \frac{n_{ij..m}}{n_{.j..m}} \cdot \frac{n_{i.kl.}}{n_{..kl.}} \cdot \frac{n_{i.k.m}}{n_{..k.m}} \cdot \frac{n_{i..lm}}{n_{...lm}}\right)^{1/30}$$
$$\cdot \left(\frac{n_{ij...}}{n_{.j...}} \cdot \frac{n_{i.k..}}{n_{..k..}} \cdot \frac{n_{i..l.}}{n_{...l.}} \cdot \frac{n_{i...m}}{n_{....m}}\right)^{1/20} \cdot \left(\frac{n_{i....}}{n_{.....}}\right)^{1/5}. \tag{24}$$

The expressions for b, c, d, etc., can be obtained easily from those for a. For example, b is obtained from a by substituting $.j$ and $i.$ for $i.$ and $.j$ in the subscripts.

The general expression for a for any number of factors P (see section 5 in the Appendix) can be derived by studying carefully the formulas in (20) and (22)–(24). It is evident that (i) the number of ratios involved in a is 2^{P-1}, (ii) the numbers of ratios with $1,2,3,\ldots,P$ dots in the denominators are, respectively, $\binom{P-1}{0},\binom{P-1}{1},\ldots,\binom{P-1}{P-1}$, and (iii) the exponents of the ratios with $1,2,3,\ldots,P$ dots in the denominators are, respectively, the reciprocals of $P\binom{P-1}{0},P\binom{P-1}{1},\ldots,P\binom{P-1}{P-1}$. Formula (A11) and its special cases in (20) and (22)–(24) are given here for the first time. As we mentioned earlier, the basic advantage of this approach is that it gives results for symmetrical factors by just one set of decomposition unlike the averaging method suggested previously by Das Gupta (1978, 1989) where the decomposition has to be carried out $P!$ times, i.e., for all possible permutations of the P factors.

3.1. Example

We illustrate the application of the formula in (22) for three factors by using the data in Table 3. It shows the number of women and the *proportion* childless among them cross-classified by family income, wife's labor force status,and wife's age at marriage for White and Black women (we may also use the *number* of childless women cross-classified by the three factors and compute the proportion childless). Treating these women by race, respectively, as population 1 and population 2, the difference between their overall rates of childlessness is .03953 (i.e., .10572 – .06619). Using the formula for three factors similar to that in (19) for two factors, we first decompose this difference into the combined IJK-effect of .01711 and the rate effect of .02242. In order to further decompose the former, we compute a,b,c, and A,B,C from (22) from the n's and N's for each (i,j,k) in Table 3, decompose $abc - ABC$ into three components by using formula (7), and multiply each of them by $(t_{ijk} + T_{ijk})/2$. Each one is then summed over 4 categories of i, 2 categories of j, and 2 categories of k, as in formula (21) for two factors, to finally obtain the I, J, and K effects. These

TABLE 3
Numbers and Proportion Childless by Family Income, Wife's Labor Force Status, and Wife's Age at Marriage for White and Black Women Aged 35–44: United States, 1970

Wife's Labor Force Status (J)	Wife's Age at Marriage (K)	Family Income (I)				
j=1 (not in LF) j=2 (in LF) j=. (Total)	k=1 (age 14 to 21) k=2 (age 22 & over) k=. (Total)	Under $6,000 i=1	$6,000-$9,999 i=2	$10,000-$14,999 i=3	$15,000 or more i=4	Total i=.
		N_{ijk} = Number of White Women				
1	1	191664	593308	785464	664681	2235117
2	1	77114	335607	828298	776816	2017835
.	1	268778	928915	1613762	1441497	4252952
1	2	115602	361741	490934	515182	1483459
2	2	35977	156338	371461	409726	973502
.	2	151579	518079	862395	924908	2456961
1	.	307266	955049	1276398	1179863	3718576
2	.	113091	491945	1199759	1186542	2991337
.	.	420357	1446994	2476157	2366405	6709913
		n_{ijk} = Number of Black Women				
1	1	41350	51720	24827	9973	127870
2	1	29195	55147	58366	38232	180940
.	1	70545	106867	83193	48205	308810
1	2	28734	33552	15256	5835	83377
2	2	19917	36644	43498	33441	133500
.	2	48651	70196	58754	39276	216877
1	.	70084	85272	40083	15808	211247
2	.	49112	91791	101864	71673	314440
.	.	119196	177063	141947	87481	525687
		T_{ijk} = Proportion Childless (White Women)				
1	1	.0463624	.0408826	.0318029	.0271092	.0340658
2	1	.0545037	.0481426	.0432948	.0426588	.0442846
.	1	.0486982	.0435056	.0377013	.0354888	.0389141
1	2	.1129392	.0897991	.0755112	.0691037	.0796867
2	2	.1844790	.1467973	.1528801	.1806573	.1647619
.	2	.1299191	.1069991	.1088364	.1185210	.1133954
1	.	.0714104	.0594106	.0486141	.0454460	.0522654
2	.	.0958520	.0794947	.0772238	.0903112	.0834928
.	.	.0779861	.0662387	.0624762	.0679419	.0661868
		t_{ijk} = Proportion Childless (Black Women)				
1	1	.0709311	.0585654	.0487373	.0358969	.0588879
2	1	.0858024	.0660235	.0649865	.0561833	.0668011
.	1	.0770855	.0624140	.0601373	.0519863	.0635245
1	2	.1479780	.1362065	.1128081	.0858612	.1324586
2	2	.1759301	.1906178	.1885834	.1860590	.1866217
.	2	.1594212	.1646105	.1689076	.1711732	.1657990
1	.	.1025198	.0891148	.0731233	.0543396	.0879255
2	.	.1223530	.1157630	.1177649	.1167804	.1176727
.	.	.1106916	.1029295	.1051590	.1054972	.1057188

Source: U.S. Bureau of the Census (1973, Table 57).

effects are, respectively, found to be .00394, .00768, and .00549. It takes only a few seconds for a computer to perform the entire computation. The effects, when converted to percentages, give 56.7, 10.0, 19.4, and 13.9 percents for the rate effect and the I, J, and K effects, respectively. This means that 56.7 percent of the difference between the rates of childlessness of White and Black women is the "real" difference that cannot be explained by the differences in family income, wife's labor force status, and wife's age at marriage. These three factors together, on the other hand, contribute 43.3 percent to the difference between the rates of childlessness, their individual contributions being 10.0, 19.4, and 13.9 percents, respectively. The factors are symmetrical, i.e., the results do not depend on how the factors are labeled I, J, and K.

4. CONSISTENCY OF THE EFFECTS IN THREE OR MORE POPULATIONS

Regardless of which of the two categories of the decomposition problem we are addressing, it cannot handle more than two populations at a time (Clogg and Eliason 1988). When there are more than two populations to be compared, these decompositions have to be carried out more than once by taking two populations at a time. This per se is not a problem because, if there are, say, three populations 1, 2, and 3, the decomposition can be performed three times—between 1 and 2, between 2 and 3, and between 1 and 3. However, this does become a problem if the results from the three separate applications of the method are not *internally consistent*.

In order to explain the problem of internal inconsistency, let us take the simplest case of the rate R depending on two factors α and β in three populations 1, 2, and 3 (for cross-classified data, the two effects are the effect of the only factor α, and the rate effect represented by β). Denoting the α-effect and the β-effect by α and β, the three sets of decomposition corresponding to the comparisons of populations 1 and 2, 2 and 3, and 1 and 3, are, respectively,

$$R_2 - R_1 = \alpha_{12} + \beta_{12}, \qquad R_3 - R_2 = \alpha_{23} + \beta_{23}, \qquad R_3 - R_1 = \alpha_{13} + \beta_{13}, \quad (25)$$

where α's and β's are computed from (1), (6), or (18). In order for these effects to be internally consistent, they should not only satisfy (25), but α_{13} and β_{13} should also be equal to $\alpha_{12} + \alpha_{23}$ and $\beta_{12} + \beta_{23}$, respectively. However, there is no mechanism in the formulas (1), (6), and (18) themselves that would guarantee these latter conditions. We, therefore, need some correction factors.

In order to derive these correction factors, let us denote by $\alpha_{xy.z}$ and $\beta_{xy.z}$ the internally consistent α-effects and β-effects when populations x and y are compared in the presence of population z. We, therefore, have

$$R_2 - R_1 = \alpha_{12.3} + \beta_{12.3}, \qquad R_3 - R_2 = \alpha_{23.1} + \beta_{23.1}, \qquad R_3 - R_1 = \alpha_{13.2} + \beta_{13.2}, \quad (26)$$

and

$$\alpha_{13.2} = \alpha_{12.3} + \alpha_{23.1}, \qquad \beta_{13.2} = \beta_{12.3} + \beta_{23.1}. \quad (27)$$

It is easy to verify that, in view of (25), the following 6 effects satisfy the 5 equations in (26) and (27):

$$\begin{aligned}
\alpha_{12.3} &= \alpha_{12} - \frac{\alpha_{12} + \alpha_{23} - \alpha_{13}}{3}, & \beta_{12.3} &= \beta_{12} - \frac{\beta_{12} + \beta_{23} - \beta_{13}}{3}, \\
\alpha_{23.1} &= \alpha_{23} - \frac{\alpha_{23} + \alpha_{31} - \alpha_{21}}{3}, & \beta_{23.1} &= \beta_{23} - \frac{\beta_{23} + \beta_{31} - \beta_{21}}{3}, \\
\alpha_{13.2} &= \alpha_{13} - \frac{\alpha_{13} + \alpha_{32} - \alpha_{12}}{3}, & \beta_{13.2} &= \beta_{13} - \frac{\beta_{13} + \beta_{32} - \beta_{12}}{3},
\end{aligned} \quad (28)$$

where $\alpha_{yx} = -\alpha_{xy}$, $\beta_{yx} = -\beta_{xy}$.

The effects in (28) are internally consistent and are expressed in terms of uncorrected effects when two factors and three populations are involved (these and other results in this section may be derived by proceeding as in section 6 in the Appendix).

It is possible to show that when there are three populations, then, regardless of the number of factors $\alpha,\beta,\gamma,\delta,\ldots$ involved, the corrected α-effect in the comparison of populations 1 and 2 is always given by the same expression as in (28), i.e., by

$$\alpha_{12.3} = \alpha_{12} - \frac{\alpha_{12} + \alpha_{23} - \alpha_{13}}{3}. \tag{29}$$

Other corrected effects follow directly from (29) and from the computed uncorrected effects, if we keep in mind that, for example, $\alpha_{32} = -\alpha_{23}$, etc.

Similarly, using analogous symbols, when there are four populations, the corrected α-effect in the comparison of populations 1 and 2 in the presence of populations 3 and 4, regardless of the number of factors, is

$$\alpha_{12.34} = \alpha_{12} - \frac{(\alpha_{12} + \alpha_{23} - \alpha_{13}) + (\alpha_{12} + \alpha_{24} - \alpha_{14})}{4}. \tag{30}$$

The general expression for corrected effects when there are N populations, regardless of the number of factors $\alpha,\beta,\gamma,\delta,\ldots$, is represented by

$$\alpha_{12.34\ldots N} = \alpha_{12} - \frac{\sum_{j=3}^{N}(\alpha_{12} + \alpha_{2j} - \alpha_{1j})}{N}. \tag{31}$$

4.1. Example

We now illustrate the computation of the general internally consistent factor effects in (31) corresponding to 5 populations and 4 factors. The five populations are the decennial life tables for White males in the United States for 1940, 1950, 1960, 1970, and 1980. The four factors α, β, γ, and δ are the mortality rates q_x in the age groups 0–25, 25–50, 50–75, and 75 and over. The rates to be compared are the expectations of life at birth ($\mathring{e}_0$'s) in the five life tables.

There are ten different ways the $\mathring{e}_0$'s in the five life tables can be compared taking *two life tables at a time.* These ten comparisons are shown in the upper half of Table 4. For example, the $\mathring{e}_0$'s for 1940 and 1950 are 62.812 and 66.314, respectively. The difference 3.502, when decomposed by formula (3) for α-effect and by similar formulas for other effects, gives 1.972, 0.807, 0.509, and 0.214 for α, β, γ, and δ effects. These effects along with other nine sets of effects are not internally consistent because they do not satisfy conditions such as (27). For example, the α-effects from the comparisons of $\mathring{e}_0$'s for (1940, 1950) and (1950, 1960) add up to 2.515, which is not equal to the α-effect of 2.520 for (1940, 1960).

The lower half of Table 4 shows the factor effects when the decompositions are performed by (31) by taking all *five life tables simultaneously*. In this case, for example, the difference of 3.502 between the $\mathring{e}_0$'s of 1940 and 1950 is decomposed into 1.978, 0.810, 0.499, and 0.215 as the α, β, γ, and δ effects, respectively. These and all other effects are internally consistent. For example, the α-effects corresponding to (1940, 1950) and (1950, 1960) add up to 2.523, which is identical with the α-effect

TABLE 4
Decomposition of the Difference Between the Expectations of Life at Birth in Two Decennial Life Tables into Factor (Age Group) Effects Using 4-Factor Data and 5 Populations: United States, White Males, 1940–1980

No. of Pops. at a Time	Life Tables for White Males for the Years	Comparison of Expectation of Life at Birth ($\mathring{e}_0$)					Effects of 4 Factors			
		Pop.1	Pop.2	Pop.1	Pop.2	Diff.	α(0–25)	β(25–50)	γ(50–75)	δ(75+)
2	1940, 1950	1940	1950	62.812	66.314	3.502	1.972	0.807	0.509	0.214
	1940, 1960	1940	1960	62.812	67.547	4.735	2.520	1.158	0.781	0.276
	1940, 1970	1940	1970	62.812	67.939	5.127	2.863	1.100	0.835	0.329
	1940, 1980	1940	1980	62.812	70.817	8.005	3.642	1.546	2.131	0.686
	1950, 1960	1950	1960	66.314	67.547	1.233	0.543	0.354	0.277	0.059
	1950, 1970	1950	1970	66.314	67.939	1.625	0.890	0.291	0.329	0.115
	1950, 1980	1950	1980	66.314	70.817	4.503	1.645	0.722	1.664	0.472
	1960, 1970	1960	1970	67.547	67.939	0.392	0.348	−0.064	0.051	0.057
	1960, 1980	1960	1980	67.547	70.817	3.270	1.099	0.359	1.395	0.417
	1970, 1980	1970	1980	67.939	70.817	2.878	0.746	0.427	1.350	0.355
5	1940,1950, 1960,1970, 1980	1940	1950	62.812	66.314	3.502	1.978	0.810	0.499	0.215
		1940	1960	62.812	67.547	4.735	2.523	1.165	0.773	0.274
		1940	1970	62.812	67.939	5.127	2.871	1.102	0.824	0.330
		1940	1980	62.812	70.817	8.005	3.626	1.533	2.159	0.687
		1950	1960	66.314	67.547	1.233	0.545	0.355	0.274	0.059
		1950	1970	66.314	67.939	1.625	0.893	0.292	0.325	0.115
		1950	1980	66.314	70.817	4.503	1.648	0.723	1.660	0.472
		1960	1970	67.547	67.939	0.392	0.348	−0.063	0.051	0.056
		1960	1980	67.547	70.817	3.270	1.103	0.368	1.386	0.413
		1970	1980	67.939	70.817	2.878	0.755	0.431	1.335	0.357

for (1940, 1960). Similarly, the four α-effects corresponding to the four decades in 1940–1980 add up to 3.626, which is identical with the α-effect for (1940, 1980).

5. CONCLUDING REMARKS

The formulas are shown in this paper without much emphasis on their derivations. The main objective of the paper is to put together all the formulas that a researcher is expected to need to deal with decomposition problems connected with various kinds of data. Some of these formulas are available in other works, but most of them appear here for the first time. The two broad categories of decomposition, along with their special cases and extensions provided in this paper, should cover virtually all cases of decomposition of the difference between two rates for any number of factors. Most technical papers on decomposition deal with two factors in the context of cross-classified data only. In the absence of general methods of decomposition to be used under various circumstances, social scientists have devised their own ad hoc methods to handle their respective problems. Although their approaches have produced meaningful results, sometimes they have been less than satisfactory in terms of mathematical rigor and elegance.

Interactions between factors have always been a sticky question to researchers, who have been troubled both by the inclusion and exclusion of interactions in their approaches. The present general methods are completely devoid of interaction terms and this can be justified in various ways. For example, it is possible to show (see section 7 in the Appendix) that when the rate is a function of several factors and when a linear saturated model involving interactions is used to represent this rate, the difference between the two rates is always free from the two-factor interaction terms. Since for any set of data, the three-factor and higher order interaction terms are expected to be negligible, it makes sense to find meaningful ways to decompose the difference into the main effects of the factors only.

When there are more than two populations to be compared, the decompositions can be carried out more than once by taking two populations at a time. However, this procedure may lead to internally inconsistent results. For example, the effects of the factor α in the comparisons of populations 1 and 2, and of populations 2 and 3, may not add up to the α-effect when populations 1 and 3 are compared. The solution to this problem suggested in this paper is particularly interesting for its applications to the rate effects (i.e., to the differences between standardized rates) from cross-classified data. It provides a unique way of achieving consistency in the rate effects based on the multiple populations being compared, without bringing in an exogenous population as standard.

ACKNOWLEDGMENT

The author is grateful to an anonymous referee for some helpful suggestions.

REFERENCES

Arriaga, E. E. (1984) Measuring and explaining the change in life expectancies. *Demography* **21:** 83–96.

Bianchi, S. M., and Rytina, N. (1986) The decline in occupational sex segregation during the 1970s: census and CPS comparisons. *Demography* **23:** 79–86.

Cho, L. J., and Retherford, R. D. (1973) Comparative analysis of recent fertility trends in East Asia. *International Population Conference*, Liege (International Union for the Scientific Study of Population), **2:** 163–181.

Clogg, C. C., and Eliason, S. R. (1988) A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review* **53:** 267–283.

Das Gupta, P. (1978) A general method of decomposing a difference between two rates into several components. *Demography* **15:** 99–112.

Das Gupta, P. (1984) Contributions of other socio-economic factors to the fertility differentials of women by education: a multivariate approach. *Genus* **40:** 117–127.

Das Gupta, P. (1987) Comment on Suzanne M. Bianchi and Nancy Rytina's 'The decline in occupational sex segregation during the 1970s: census and CPS comparisons'. *Demography* **24:** 291–295.

Das Gupta, P. (1989) Methods of decomposing the difference between two rates with applications to race-sex inequality in earnings. *Mathematical Population Studies* **2:** 15–36.

Das Gupta, P. (1990) Decomposition of the difference between two rates when the factors are nonmultiplicative with applications to the U.S. life tables. Paper presented at the 1990 Annual Meeting of the Population Association of America, Toronto, Canada.

Janowitz, B. S. (1976) An analysis of the impact of education on family size. *Demography* **13:** 189–198.

Keyfitz, N. (1968) *Introduction to the Mathematics of Population*. Reading, MA: Addison-Wesley.

Kim, Y. J., and Strobino, D. M. (1984) Decomposition of the difference between two rates with hierarchical factors. *Demography* **21:** 361–372.

Kitagawa, E. M. (1955) Components of a difference between two rates. *Journal of the American Statistical Association* **50:** 1168–1194.

Liao, T. F. (1989) A flexible approach for the decomposition of rate differences. *Demography* **26:** 717–726.

Lichter, D. T. (1985) Racial concentration and segregation across U.S. counties, 1950–1980. *Demography* **22:** 603–609.

Nathanson, C. A., and Kim, Y. J. (1989) Components of change in adolescent fertility. *Demography* **26:** 85–98.

Pollard, J. H. (1988) On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography* **25:** 265–276.

Pullum, T. W., Tedrow, L. M., and Herting, J. R. (1989) Measuring change and continuity in parity distributions. *Demography* **26:** 485–498.

Ruggles, S. (1988) The demography of the unrelated individual: 1900–1950. *Demography* **25**: 521–536.
Smith, H. L., and Cutright, P. (1988) Thinking about change in illegitimacy ratios: United States, 1963–1983. *Demography* **25**: 235–247.
Sweet, J. A. (1984) Components of change in the number of households: 1970–1980. *Demography* **21**: 129–140.
U.S. Bureau of the Census (1964) U.S. Census of Population: 1960. Vol. I. *Characteristics of the Population*. Part 1. United States Summary. Washington, DC: U.S. Government Printing Office.
U.S. Bureau of the Census (1973) U.S. Census of Population: 1970. *Women by Number of Children Ever Born*. Subject Report PC(2)-3A. Washington, DC: U.S. Government Printing Office.
U.S. Bureau of the Census (1983) U.S. Census of Population: 1980. Vol. 1B. *General Population Characteristics*. Part 1. United States Summary. Washington, DC: U.S. Government Printing Office.
U.S. National Center for Health Statistics (1963) *Vital Statistics of the United States 1960*. Vol. I—Natality. Washington, DC: U.S. Government Printing Office.
U.S. National Center for Health Statistics (1984) *Vital Statistics of the United States 1980*. Vol. I—Natality. Washington, DC: U.S. Government Printing Office.
Wilson, F. D. (1988) Components of change in migration and destination-propensity rates for metropolitan and nonmetropolitan areas: 1935–1980. *Demography* **25**: 129–139.
Wojtkiewicz, R. A., McLanahan, S. S., and Garfinkel, I. (1990) The growth of families headed by women: 1950–1980. *Demography* **27**: 19–30.

APPENDIX

1. DERIVATION OF FORMULA (2)

$$R = F(\alpha,\beta,\gamma).$$

α, β, and γ assume values A,B,C in population 1 and a,b,c in population 2, so that the difference $R_2 - R_1$ is

$$F(a,b,c) - F(A,B,C) = \alpha\text{-effect} + \beta\text{-effect} + \gamma\text{-effect}. \tag{A1}$$

We write the three effects as

$$\begin{aligned}\alpha\text{-effect} = {} & w[F(a,b,c) - F(A,b,c)] + x[F(a,b,C) - F(A,b,C)] \\ & + y[F(a,B,c) - F(A,B,c)] + z[F(a,B,C) - F(A,B,C)], \end{aligned} \tag{A2}$$

$$\begin{aligned}\beta\text{-effect} = {} & w[F(a,b,c) - F(a,B,c)] + x[F(a,b,C) - F(a,B,C)] \\ & + y[F(A,b,c) - F(A,B,c)] + z[F(A,b,C) - F(A,B,C)], \end{aligned} \tag{A3}$$

$$\begin{aligned}\gamma\text{-effect} = {} & w[F(a,b,c) - F(a,b,C)] + x[F(A,b,c) - F(A,b,C)] \\ & + y[F(a,B,c) - F(a,B,C)] + z[F(A,B,c) - F(A,B,C)], \end{aligned} \tag{A4}$$

where w,x,y,z are suitably chosen constants.

Substituting (A2)–(A4) on the right-hand side of (A1) and then equating the coefficients from both sides, we have

$$w = z = 1/3, \qquad x = y = 1/6.$$

Substituting these values in (A2)–(A4), we obtain the formulas in (2).

2. GENERAL FORMULA CORRESPONDING TO (1)–(5)

$$R = F(\alpha_1, \alpha_2, \ldots, \alpha_p).$$

α_i assumes values A_i and a_i in population 1 and population 2, respectively.

$$\begin{aligned} R_2 - R_1 &= F(a_1, a_2, \ldots, a_p) - F(A_1, A_2, \ldots, A_p) \\ &= \alpha_1\text{-effect} + \alpha_2\text{-effect} + \cdots + \alpha_p\text{-effect}, \end{aligned}$$

where

$$\alpha_1\text{-effect} = \sum_{r=1}^{S} \frac{\displaystyle\sum_{\substack{\alpha_i = A_i, a_i \\ [(P-r)\ a_i\text{'s and } (r-1)\ A_i\text{'s} \\ \text{or } (P-r)\ A_i\text{'s and } (r-1)\ a_i\text{'s,} \\ \text{not counting } a_1 \text{ and } A_1]}} [F(a_1, \alpha_2, \alpha_3, \ldots, \alpha_p) - F(A_1, \alpha_2, \alpha_3, \ldots, \alpha_p)]}{P\binom{P-1}{r-1}}, \tag{A5}$$

and

$$\begin{aligned} S &= P/2 \text{ when } P \text{ is even,} \\ &= (P+1)/2 \text{ when } P \text{ is odd.} \end{aligned}$$

3. GENERAL FORMULA CORRESPONDING TO (6)–(10)

$$R = \alpha_1 \alpha_2 \ldots \alpha_p.$$

α_i assumes values A_i and a_i in population 1 and population 2, respectively.

$$\begin{aligned} R_2 - R_1 &= a_1 a_2 \ldots a_p - A_1 A_2 \ldots A_p \\ &= \alpha_1\text{-effect} + \alpha_2\text{-effect} + \cdots + \alpha_p\text{-effect}, \end{aligned}$$

where

$$\alpha_1\text{-effect} = \sum_{r=1}^{S} \frac{\displaystyle\sum_{\substack{\alpha_i = A_i, a_i \\ [(P-r)\ a_i\text{'s and } (r-1)\ A_i\text{'s or} \\ (P-r)\ A_i\text{'s and } (r-1)\ a_i\text{'s}]}} \alpha_2 \alpha_3 \ldots \alpha_p}{P\binom{P-1}{r-1}} (a_1 - A_1), \tag{A6}$$

and

$$\begin{aligned} S &= P/2 \text{ when } P \text{ is even,} \\ &= (P+1)/2 \text{ when } P \text{ is odd.} \end{aligned}$$

4. DERIVATION OF FORMULAS IN (22)

$$\frac{n_{ijk}}{n_{...}} = abc, \tag{A7}$$

where a, b, and c involve ratios which represent, respectively, the I-effect, J-effect, and K-effect.

We write these three quantities as

$$a = \left(\frac{n_{ijk}}{n_{.jk}}\right)^x \cdot \left(\frac{n_{ij.}}{n_{.j.}} \cdot \frac{n_{i.k}}{n_{..k}}\right)^y \cdot \left(\frac{n_{i..}}{n_{...}}\right)^z, \tag{A8}$$

$$b = \left(\frac{n_{ijk}}{n_{i.k}}\right)^x \cdot \left(\frac{n_{ij.}}{n_{i..}} \cdot \frac{n_{.jk}}{n_{..k}}\right)^y \cdot \left(\frac{n_{.j.}}{n_{...}}\right)^z, \tag{A9}$$

$$c = \left(\frac{n_{ijk}}{n_{ij.}}\right)^x \cdot \left(\frac{n_{i.k}}{n_{i..}} \cdot \frac{n_{.jk}}{n_{.j.}}\right)^y \cdot \left(\frac{n_{..k}}{n_{...}}\right)^z, \tag{A10}$$

where x, y, z are suitably chosen exponents corresponding to the ratios with 0, 1, and 2 dots in the numerators, respectively.

Substituting (A8)–(A10) on the right-hand side of (A7) and then equating the exponents from both sides, we have

$$x = z = 1/3, \qquad y = 1/6.$$

Substituting these values in (A8)–(A10), we obtain the formulas in (22).

5. GENERAL FORMULA CORRESPONDING TO (20), (22)–(24)

Let the factors be $I_1, I_2, \ldots, I_p$. Using these factors as subscripts of n, as in (24), let us denote

$$\frac{n_{i_1 \text{ to } i_p}}{n_{....}} = a_1 a_2 \ldots a_p,$$

$$\frac{n_{i_1 i_2 \text{ to } i_p}}{n_{.i_2 \text{ to } i_p}} = m_{i_1}, \qquad \frac{n_{i_1 . i_3 \text{ to } i_p}}{n_{..i_3 \text{ to } i_p}} = m_{i_1 i_2},$$

$$\frac{n_{i_1 .. i_4 \text{ to } i_p}}{n_{...i_4 \text{ to } i_p}} = m_{i_1 i_2 i_3}, \quad \text{etc.}$$

then

$$a_1 = \prod_{r=1}^{P} \left[\prod_{\binom{P-1}{r-1} \text{ terms}} m_{i_1 i_2 \ldots i_r} \right]^{1/P\binom{P-1}{r-1}}. \tag{A11}$$

6. DERIVATION OF FORMULA (31) FOR $N = 4$

Let there be two factors α and β, and four populations. Let us denote the uncorrected factor effects by the corresponding letters so that

$$\begin{aligned} R_2 - R_1 &= \alpha_{12} + \beta_{12}, & R_3 - R_2 &= \alpha_{23} + \beta_{23}, \\ R_4 - R_3 &= \alpha_{34} + \beta_{34}, & R_4 - R_1 &= \alpha_{14} + \beta_{14}, \\ R_3 - R_1 &= \alpha_{13} + \beta_{13}, & R_4 - R_2 &= \alpha_{24} + \beta_{24}. \end{aligned} \tag{A12}$$

For corrected factor effects in the presence of other two populations, eqs. (A12) also hold so that

$$\begin{aligned} R_2 - R_1 &= \alpha_{12.34} + \beta_{12.34}, & R_3 - R_2 &= \alpha_{23.14} + \beta_{23.14}, \\ R_4 - R_3 &= \alpha_{34.12} + \beta_{34.12}, & R_4 - R_1 &= \alpha_{14.23} + \beta_{14.23}, \\ R_3 - R_1 &= \alpha_{13.24} + \beta_{13.24}, & R_4 - R_2 &= \alpha_{24.13} + \beta_{24.13}. \end{aligned} \tag{A13}$$

In addition, for internal consistency, equations similar to (27) should hold for all combinations of three populations out of four. In other words, we must have

$$\begin{aligned} \alpha_{13.24} &= \alpha_{12.34} + \alpha_{23.14}, & \alpha_{14.23} &= \alpha_{12.34} + \alpha_{24.13}, \\ \alpha_{14.23} &= \alpha_{13.24} + \alpha_{34.12}, & \alpha_{24.13} &= \alpha_{23.14} + \alpha_{34.12}, \end{aligned} \tag{A14}$$

and

$$\begin{aligned} \beta_{13.24} &= \beta_{12.34} + \beta_{23.14}, & \beta_{14.23} &= \beta_{12.34} + \beta_{24.13}, \\ \beta_{14.23} &= \beta_{13.24} + \beta_{34.12}, & \beta_{24.13} &= \beta_{23.14} + \beta_{34.12}. \end{aligned} \tag{A15}$$

In order to derive the expression for $\alpha_{12.34}$ in terms of the uncorrected α-effects in (A12), let us write

$$\alpha_{12.34} = \alpha_{12} + x[(\alpha_{12} + \alpha_{23} - \alpha_{13}) + (\alpha_{12} + \alpha_{24} - \alpha_{14})], \tag{A16}$$

so that when (A14) is satisfied by the uncorrected effects, $\alpha_{12.34}$ and α_{12} become identical.

Similarly,

$$\begin{aligned} \alpha_{13.24} &= \alpha_{13} + x[(\alpha_{13} + \alpha_{32} - \alpha_{12}) + (\alpha_{13} + \alpha_{34} - \alpha_{14})], \\ \alpha_{23.14} &= \alpha_{23} + x[(\alpha_{23} + \alpha_{31} - \alpha_{21}) + (\alpha_{23} + \alpha_{34} - \alpha_{24})], \end{aligned} \tag{A17}$$

where $\alpha_{32} = -\alpha_{23}$, etc.

Substituting eqs. (A16) and (A17) in the first equation in (A14), and simplifying, we obtain $x = -1/4$, so that from (A16),

$$\alpha_{12.34} = \alpha_{12} - \frac{\sum_{j=3}^{4}(\alpha_{12} + \alpha_{2j} - \alpha_{1j})}{4}. \tag{A18}$$

It is easy to show that, in view of (A12), the corrected factor effects represented by (A18) satisfy (A13)–(A15).

7. LINEAR SATURATED MODEL FOR 3 FACTORS IN FORMULA (2)

$$F(\alpha,\beta,\gamma) = K + E_\alpha + E_\beta + E_\gamma + E_{\alpha\beta} + E_{\alpha\gamma} + E_{\beta\gamma} + E_{\alpha\beta\gamma}, \tag{A19}$$

where

$$\sum_\alpha E_\alpha = \sum_\beta E_\beta = \sum_\gamma E_\gamma = 0, \tag{A20}$$

$$\sum_\alpha E_{\alpha\beta} = \sum_\beta E_{\alpha\beta} = \sum_\alpha E_{\alpha\gamma} = \sum_\gamma E_{\alpha\gamma} = \sum_\beta E_{\beta\gamma} = \sum_\gamma E_{\beta\gamma} = 0, \tag{A21}$$

$$\sum_\alpha E_{\alpha\beta\gamma} = \sum_\beta E_{\alpha\beta\gamma} = \sum_\gamma E_{\alpha\beta\gamma} = 0. \tag{A22}$$

There are 27 unknowns in (A19), which can be solved from 27 independent equations (8 in A19, 3 in A20, 9 in A21, and 7 in A22). Using these solutions, we have

$$\begin{aligned} F(a,b,c) - F(A,B,C) &= (E_a - E_A) + (E_b - E_B) + (E_c - E_C) + (E_{abc} - E_{ABC}) \\ &= \alpha\text{-effect} + \beta\text{-effect} + \gamma\text{-effect} + \alpha\beta\gamma\text{-interaction effect}, \end{aligned}$$

where

$$\alpha\text{-effect} = \frac{\begin{array}{c}[F(a,b,c) - F(A,b,c)] + [F(a,B,C) - F(A,B,C)] \\ +[F(a,b,C) - F(A,b,C)] + [F(a,B,c) - F(A,B,c)]\end{array}}{4},$$

$$\alpha\beta\gamma\text{-interaction effect} = \frac{\begin{array}{c}[F(a,b,c) - F(A,b,c)] + [F(a,B,C) - F(A,B,C)] \\ -[F(a,b,C) - F(A,b,C)] - [F(a,B,c) - F(A,B,c)]\end{array}}{4}.$$

β-effect and γ-effect have expressions similar to that for α-effect. If we distribute the $\alpha\beta\gamma$-interaction effect equally among the three main effects, we obtain the formulas in (2). All three two-factor interaction terms in (A19) turn out to be zero.

For any number of factors, the difference $F(a,b,c,\ldots) - F(A,B,C,\ldots)$ involves two-factor interaction terms such as $(E_{ab} - E_{AB})$, each of which vanishes because of the conditions similar to those in (A21). For example, the two equations $E_{AB} + E_{aB} = 0$ and $E_{aB} + E_{ab} = 0$ together give $E_{AB} = E_{ab}$. Thus, the two-factor interaction terms are always zero regardless of the number of factors involved.