Leo A. Goodman

Statistical Magic and/or Statistical Serendipity: An Age of Progress in the Analysis of Categorical Data

Leo A. Goodman

Department of Sociology and Department of Statistics, University of California, Berkeley, California 94720-1980; email: lgoodman@berkeley.edu

Annu. Rev. Sociol. 2007. 33:1–19

First published online as a Review in Advance on March 22, 2007

The *Annual Review of Sociology* is online at http://soc.annualreviews.org

This article's doi: 10.1146/annurev.soc.33.040406.131720

0360-0572/07/0811-0001$20.00

Key Words

social stratification and mobility, survey analysis, panel analysis, latent structures

Abstract

This essay describes in simple terms some of the major concepts of categorical data analysis (CDA) that have been and will continue to be useful in the analysis of sociological data, examples of which include data in the area of social stratification and mobility, and in many other areas that make use of survey data and/or panel studies data, and in empirical studies of latent types, latent variables, and latent structures. The exposition does not make use of any mathematical formulas, and the only arithmetic used is very simple multiplication, division, and addition. Simple numerical examples, constructed for expository purposes, are used as an aid in describing the concepts of categorical data analysis that are considered in the essay. These concepts include quasi-independence, quasi-symmetry, symmetric association, uniform association, and other related concepts useful in the analysis of mobility tables, and also other concepts that are useful in other areas of study.

INTRODUCTION

I would like to thank the Editors of the *Annual Review of Sociology*, Karen Cook and Doug Massey, and their Editorial Committee, for having invited me to write the lead article for the 2007 volume. This essay describes in simple terms some of the major concepts of categorical data analysis (CDA), and it does so in a way that will, I hope, interest *ARS* readers ranging from those who have had no interest in the analysis of categorical data to those who have had a strong interest in this subject. The concepts that are considered, or reconsidered, here have been developed over approximately the past 50 years. They have helped to change, in a dramatic way, how categorical data are analyzed now. The development of new related concepts in categorical data analysis continues unabated today.

The exposition in this essay does not make use of any mathematical formulas, and the only arithmetic used here is some very simple multiplication, division, and addition. (Even when log-linear models are considered, logarithms are not used.) Simple numerical examples, constructed for expository purposes, are used here as an aid in describing the concepts under consideration.

When the great Michelangelo was sculpting his colossal figure of *David*, he worked under the premise that the image of *David* was already in the block of marble that he had selected, and his task was to release the image from the block. (He worked under this kind of premise when he was sculpting some of his other statues as well.) Now, faced with a set of categorical data of interest, data analysts can work under the premise that there is an image, or more than one image, embedded in the set of data, and their task is to release that image, or those images, using suitable tools. The tools could be the kind developed for categorical data analysis and the kinds of concepts that are described here.

Because of serious space limitations, this essay considers only some of the major concepts that have been and continue to be useful in the analysis of sociological data. Each section and subsection here can be viewed as a new kind of introduction to one of the topics in categorical data analysis—a new introduction to one of the chapters, selected from a large collection of chapters, in a statistical autobiography. Reference is made in this essay primarily to my work; the work of others is cited in the publications included in the Related Resources section at the end.

THE CONCEPTS OF QUASI-INDEPENDENCE, QUASI-SYMMETRY, SYMMETRIC ASSOCIATION, UNIFORM ASSOCIATION, AND RELATED CONCEPTS

I describe here in simple terms the concepts of quasi-independence, quasi-symmetry, symmetric association, uniform association, and related concepts. For the sake of simplicity, I describe each of these concepts by noting how it can be applied in the analysis of data in a two-way 3×3 table of observed frequencies, a table that represents the relationship between two trichotomous categorical variables, a row variable and a column variable, but each of these concepts can also be applied much more generally. Comments on generalizations are also included (see On Some Generalizations, subsection below).

Among the various substantive areas in which these concepts have been applied is the field of social stratification and mobility. The introduction of these concepts, and other related concepts, in this field has changed the way that mobility tables are analyzed. In his description of the work that I have done, using the concept of quasi-independence and related concepts, in developing methods for analyzing mobility tables, Duncan (1974, 1975) wrote as follows:

> Goodman's work on methods for analyzing mobility tables solved a problem that had plagued research workers in this field for at least two decades.... In solving this

problem, he rendered a substantial corpus of previous work . . . obsolete—no softer word will do.

I now describe in simple terms an index that would have been calculated in the previous work rendered obsolete and then the corresponding index that can be calculated in the current work that has replaced it, making use of this quasi-independence concept.

Independence and Quasi-Independence

For the sake of simplicity, I consider now the three two-way tables (**Tables 1*a***, **1*b***, and **1*c***) in **Table 1**. The first two-way table (**Table 1*a***) can be viewed as an example of a simplified mobility table. The column categories might represent, say, three status categories (Upper, Middle, Lower) for contemporary male subjects, and the row categories might represent the corresponding status categories for their fathers. There are nine cells in the two-way 3×3 table, and each cell corresponds to one of the nine possible combinations of a subject's status category and his father's status category. The entry in each cell of the table is the number (i.e., the frequency) of subjects who have the corresponding combination of subject's status category and father's status category. We see in **Table 1*a*** that the entries in the cells on the main diagonal (i.e., the number of subjects who are in the same status category as their fathers, for categories U, M, and L) are, in some sense, relatively large; we can refer to this phenomenon, for lack of a better term, as "status inheritance." To measure the magnitude of the status inheritance in **Table 1*a***, the previous method, now rendered obsolete, would have been carried out in two steps: First, **Table 1*b*** would have been calculated from **Table 1*a***; second, the entries in the cells on the main diagonal in **Table 1*a*** would have been compared with the corresponding entries in **Table 1*b***.

Table 1 Examples of the concepts of independence (perfect mobility) and quasi-independence (quasi-perfect mobility). Comparison of previous category-inheritance ratios with current category-inheritance ratios (Table 1*d*). (U, Upper; M, Middle; L, Lower)

Table 1a Two-way table of observed frequencies

	U	M	L	Total
U	15	3	2	20
M	5	45	10	60
L	4	12	24	40
Total	24	60	36	120

Table 1b Expected frequencies under independence model for Table 1a

	U	M	L	Total
U	4	10	6	20
M	12	30	18	60
L	8	20	12	40
Total	24	60	36	120

Table 1c Main-diagonal entries in Table 1a replaced by independence-model entries

	U	M	L	Total
U	1	3	2	6
M	5	15	10	30
L	4	12	8	24
Total	10	30	20	60

Table 1d Previous versus current category-inheritance ratios for Table 1a

	Previous	Current
U	15/4 = 3.75	15/1 = 15
M	45/30 = 1.50	45/15 = 3
L	24/12 = 2.00	24/8 = 3

In the first step, **Table 1*b*** would have been calculated from **Table 1*a*** using the row totals (i.e., the row marginal) and column totals (i.e., the column marginal) in **Table 1*a*** to estimate the frequencies that would be expected under the model in which the column status variable is independent of the row status variable (i.e., the independence model). Note that the row and column marginals in **Table 1*b*** are equal to the corresponding marginals in **Table 1*a***, and the frequency distribution in each row of **Table 1*b*** is the same as the frequency distribution in the column marginal; viz., 0.2, 0.5, 0.3 (i.e., 24/120, 60/120, 36/120), for categories U, M, L, respectively. In other words, the frequency distribution in each row is independent of the particular row under consideration. Note also that, for each 2×2 subtable of **Table 1*b***, the corresponding cross-product ratio (frequently called the odds ratio) is equal to 1. [For example, for the 2×2 subtable formed from row categories U and M and column categories U and M of **Table 1*b***, the corresponding cross-product ratio is $(4 \times 30)/(10 \times 12) = 1$.] Whenever the cross-product ratio in each 2×2 subtable of a two-way table is equal to 1, the independence model holds true.

In the second step in the previous method rendered obsolete, as noted above in this subsection, the magnitude of the status inheritance in **Table 1*a*** would have been calculated by comparing the entries in the cells on the main diagonal in **Table 1*a*** with the corresponding entries in **Table 1*b***. The status-inheritance ratios in **Table 1*a*** thus obtained are presented in **Table 1*d***, and we see that they range from 1.50 to 3.75. By contrast, we also see in **Table 1*d*** that the status-inheritance ratios in **Table 1*a*** that would be calculated in current work range instead from 3.00 to 15.00.

The status-inheritance ratios in **Table 1*d*** that would be calculated in current work are obtained simply by comparing the entries in the cells on the main diagonal in **Table 1*a*** with the corresponding entries in **Table 1*c*** (rather than with the corresponding entries in **Table 1*b***). **Table 1*c*** was obtained simply by deleting the entries in the cells on the main diagonal in **Table 1*a*** and replacing them with appropriate entries obtained with the independence model. (Note that the frequency distribution in each row of **Table 1*c*** is the same as the frequency distribution in the column marginal for this table. Also note that, for each 2×2 subtable of **Table 1*c***, the corresponding cross-product ratio is equal to 1.) **Table 1*a*** fits the quasi-independence model; i.e., the entries in the cells on the main diagonal of **Table 1*a*** can be replaced by alternative entries in such a way that the revised table (**Table 1*c***) fits the independence model.

What do the current status-inheritance ratios tell us? First note that the 60 individuals in **Table 1*c*** fit the model of independence (or "perfect mobility") from father's status category to subject's status category; the status categories of these subjects are independent of their father's status categories. We can view these individuals as having been "perfectly mobile." The individuals who are in the cells on the main diagonal in **Table 1*c*** can be viewed as those perfectly mobile individuals who arrived in the same status category as did their fathers "by chance." Now note the additional 60 individuals who are in the cells on the main diagonal in **Table 1*a***. These additional individuals can be viewed as "stayers" (or "status inheritors"). The current status-inheritance ratios in **Table 1*d*** compare the total number of individuals in each of the cells on the main diagonal of **Table 1*a*** with the corresponding number of perfectly mobile individuals who arrived in the same status category as did their fathers by chance. The ratio of the number of stayers (or status inheritors) in each cell on the main diagonal in **Table 1*a*** compared with the corresponding number of perfectly mobile individuals who arrived in the same status category as did their fathers by chance is obtained simply by subtracting 1 from each of the current status-inheritance ratios in **Table 1*d***.

Symmetry, Quasi-Symmetry, and Symmetric Association

For expository purposes, I consider now the first two two-way tables (**Tables 2*a*** and **2*b***) in **Table 2**. Note that the entries in **Table 2*a*** are clearly not symmetric. In other words, the entries in the cells above the main diagonal in **Table 2*a***, in the upper-right triangle (viz., 3, 2, 10), are different from the corresponding entries below the main diagonal, in the lower-left triangle (viz., 5, 4, 12). **Table 2*b***, by contrast, is clearly an example of a symmetric table. **Table 2*a*** is defined as quasi-symmetric if the table can be transformed into a symmetric table by row multiplicative changes and column multiplicative changes (see, e.g., Caussinus 1966; Goodman 1979a, 2002a). **Table 2*c*** illustrates how the asymmetric **Table 2*a*** can be transformed into a symmetric table like the symmetric **Table 2*b***. (Note that the entries in the three cells on the main diagonal of **Table 2*a*** have been replaced by the letters A, B, and C in the first 3×3 table in **Table 2*c***. It is not necessary to make this replacement, but I have done so here for expository purposes to illuminate the row multiplicative changes and column multiplicative changes. The changes are made in **Table 2*c*** by dividing the entries in the L column by 2, then multiplying the entries in the U row by 4, and then multiplying the entries in the M column by 5/12.)

As for the symmetric association concept, the association in **Table 2*a*** is defined as symmetric if the cross-product ratios in the four 2×2 subtables presented on the left side of **Table 2*d*** are symmetric (see, e.g., Goodman 1979a, equation 3.3). Note that the upper-right cross-product ratio and the

Table 2 Examples of the concepts of symmetry, quasi-symmetry, and symmetric association

Table 2a Two-way table

	U	M	L
U	15	3	2
M	5	45	10
L	4	12	24

Table 2b Symmetric two-way table

	U	M	L
U	15	5	4
M	5	45	5
L	4	5	24

Table 2c Row and column multiplicative transformations of Table 2a, with main-diagonal entries deleted

	U	M	L
U	A	3	2
M	5	B	10
L	4	12	C

U	M	L
A	3	1
5	B	5
4	12	C/2

U	M	L
4A	12	4
5	B	5
4	12	C/2

U	M	L
4A	5	4
5	5B/12	5
4	5	C/ 2

Table 2d Association in 2×2 subtables of Table 2a, with symmetric cross-product ratios

	U	M
U	15	3
M	5	45

M	L
3	2
45	10

M	5	45
L	4	12

45	10
12	24

	U M	M L
U M	45	1/3
M L	1/3	9

corresponding lower-left cross-product ratio on the right side of **Table 2*d*** are equal. The symmetric association concept and the quasi-symmetry concept are different concepts, but they are equivalent. A special case of symmetric association is considered next.

Uniform Association and More on Quasi-Independence

For expository purposes, I now consider **Tables 3*a*** and **3*c*** in **Table 3**. The association in **Table 3*a*** is defined as uniform if the cross-product ratios in the four 2 × 2 subtables presented on the left side of **Table 3*c*** are equal (see, e.g., Duncan 1979; Goodman 1979a,b). Note that all the corresponding cross-product ratios on the right side of **Table 3*c*** are equal. As noted at the end of the preceding subsection, the uniform association concept considered here is a special case of symmetric association.

Now let us consider **Table 3*b***. As noted above in the quasi-independence analysis of **Table 1*a*** using **Table 1*c***, **Table 3*b*** can be used in a quasi-independence analysis of **Table 3*a***. **Table 3*b*** was obtained simply by deleting the entries in the cells on the main diagonal in **Table 3*a*** and replacing them with appropriate entries obtained with the independence model. In this quasi-independence analysis, the entries on the main diagonal in **Table 3*a*** are compared with the corresponding entries in **Table 3*b***; we see then

Table 3 Examples of the concepts of uniform association and quasi-independence with category inheritance/disinheritance ratios

Table 3a Two-way table of observed frequencies

	U	M	L
U	16	8	6
M	12	12	18
L	10	20	60

Table 3b Main-diagonal entries in Table 3a replaced by independence-model entries

	U	M	L
U	4	8	6
M	12	24	18
L	10	20	15

Table 3c Association in 2×2 subtables of Table 3a, with uniform cross-product ratios

	U	M
U	16	8
M	12	12

M	L
8	6
12	18

	U	M
M	12	12
L	10	20

M	L
12	18
20	60

	U M	M L
U, M	16/8 = 2	18/9 = 2
M, L	20/10 = 2	12/6 = 2

Table 3d Inheritance/disinheritance ratios for Table 3a

Category	Inheritance/ Disinheritance		
U	16/4	=	4
M	12/24	=	1/2
L	60/15	=	4

from **Table 3*d*** that there is both strong category-inheritance in the U and L categories of **Table 3*a*** and also strong category-disinheritance in the corresponding *M* category. So, we see that the quasi-independence model can be used to study both category-inheritance and category-disinheritance.

What do the inheritance/disinheritance ratios in **Table 3*d*** tell us? With respect to the ratios for the U and L categories of **Table 3*a***, because these ratios in **Table 3*d*** indicate that there is category-inheritance in these two categories, we can apply to these ratios the same kind of interpretation of the category-inheritance ratio that was applied in the final paragraph of the subsection above on Independence and Quasi-Independence. The interpretation was described there in terms of stayers (or category-inheritors) and the perfectly mobile. With respect to the ratio for the *M* category of **Table 3*a***, because this ratio in **Table 3*d*** indicates that there is category-disinheritance in this category, a different interpretation is needed. For expository purposes, the interpretation of this ratio presented now will be more easily understood if we modify **Table 3*a*** by replacing the entry 12 in the (*M*,*M*) cell with 8. With this modified **Table 3*a***, the corresponding disinheritance ratio in **Table 3*d*** is one-third (rather than one-half) for the *M* category. With respect to this category-disinheritance ratio for the *M* category of the modified **Table 3*a***, a different interpretation is now introduced, which is described here in terms of the "changers" (or the "category-disinherited") and the "perfectly mobile" in the present context: I first need now to define the "changers" (or the "category-disinherited"). These are the individuals in the row *M* category who are not in the (*M*,*M*) cell and who are not viewed as "perfectly mobile" individuals who arrived in the other cells in row *M* "by chance."

Now, with the disinheritance ratio of one-third for the *M* category of the modified **Table 3*a***, comparing the number of individuals in each of the cells in row M of the modified **Table 3*a*** with the corresponding number in **Table 3*b***, we can view one-third of each number in row *M* of **Table 3*b*** (namely, 4, 8, 6, in columns U, *M*, L, respectively) as having been "perfectly mobile;" the additional individuals in row *M* of the modified **Table 3*a*** (namely, 8, 0, 12, of them in columns U, *M*, L, respectively) can be viewed as "changers" (or the "category-disinherited"). With the changers (or the category-disinherited) and the perfectly mobile individuals in cells (*M*,U) and (*M*,L), the number of individuals in each of these two cells is three times the corresponding number of perfectly mobile in that cell. Note that the 3 (in "three times") is simply the reciprocal of one-third, the disinheritance ratio. The ratio of the number of changers (or the category-disinherited) in each of these two cells compared with the corresponding number of perfectly mobile in that cell can be obtained simply by subtracting 1 from this reciprocal.

To further clarify this new interpretation of the category-disinheritance ratio, I now present a second example of its application: If we modify **Table 3*a*** by replacing the entry 12 in the (*M*,*M*) cell with 20, the corresponding disinheritance ratio in **Table 3*d*** is 5/6 (rather than 1/2) for the *M* category. Now, with the disinheritance ratio of 5/6 for the *M* category of this modified **Table 3*a***, comparing the number of individuals in each of the cells of row M of the modified **Table 3*a*** with the corresponding number in **Table 3*b***, we can view 5/6 of each number in row *M* of **Table 3*b*** (namely, 10, 20, 15, in columns U, *M*, L, respectively) as having been perfectly mobile. And the additional individuals in row *M* of this modified **Table 3*a*** (namely, 2, 0, 3 of them in columns U, *M*, L, respectively) can be viewed as changers (or the category-disinherited). With the changers (or the category-disinherited) and the perfectly mobile individuals in cells (*M*,U) and (*M*,L), the number of individuals in each of these two cells is 6/5 times the corresponding number of perfectly mobile in that cell. Note that 6/5 is simply the reciprocal of 5/6, the disinheritance ratio. The ratio of the number of changers (or the category-disinherited) in each of these two cells compared with the corresponding number of perfectly mobile in that cell can be obtained simply by subtracting 1 from this reciprocal.

On Some Generalizations

The quasi-independence model described using **Tables 1*a*** and **1*c*** can be viewed as follows: The independence model frequencies in **Table 1*c*** can be expressed as a multiplicative model in which the entry in each cell in the table is obtained by multiplying a corresponding row effect (viz., 6, 30, 24, for row U, M, L, respectively) by a corresponding column effect (viz., 10/60, 30/60, 20/60, for column U, M, L, respectively). [For example, the entry in the (U,U) cell is obtained by multiplying the corresponding row effect (viz., 6) by the corresponding column effect (viz., 10/60).] Thus, the independence model in the 3×3 table has simply three multiplicative row effects and three multiplicative column effects. The quasi-independence model frequencies in **Table 1*a*** can be expressed simply as a multiplicative model that has the three row effects and the three column effects noted above as well as an additional three multiplicative effects pertaining to the three cells on the main diagonal [viz., the current inheritance-ratios in **Table 1*d***—15, 3, 3, for cells (U,U), (M,M), (L,L), respectively].

The quasi-symmetry and symmetry concepts described in **Tables 2*a*** and **2*b*** can also be expressed in similar terms, as multiplicative models with various kinds of appropriate multiplicative effects. The uniform association model in **Table 3*a*** can also be expressed in similar terms, but one of the appropriate multiplicative effects in this case is somewhat more complicated (see, e.g., Goodman 1979b, equation 14).

For the sake of simplicity, attention has been focused in the preceding subsections on the analysis of two-way 3×3 tables of observed frequencies, tables that can be viewed as describing the observed relationships between two trichotomous variables. Some of the concepts considered in the preceding subsections can be applied more generally to an $I \times J$ table with I row categories and J column categories ($I = 3,4, \ldots; J = 3,4, \ldots$); this table can be a rectangular table (when $I \neq J$) or a square table (when $I = J$). Some concepts can be applied only to square tables where there is a one-to-one correspondence between row categories and column categories. The quasi-symmetry (symmetric association) model can be applied only to square tables, and the quasi-independence model and uniform association model can be applied to rectangular and square tables. As noted above, when the uniform association model is applied to the square table, it can be viewed as a special case of the symmetric association model. On the other hand, the former model can also be applied more generally to rectangular tables, while the latter model cannot. The quasi-independence model considered above in the analysis of the 3×3 table simply deleted the three entries in the cells on the main diagonal of the table, but this model can also be applied more generally with the deletion of the entries in any specified subset of the cells in an $I \times J$ table.

The preceding subsections describe the concepts considered there in simple terms. Now, with respect to the statistical methods that are needed to apply these concepts in the analysis of categorical data, the methods that were developed for doing this using the simple multiplicative quasi-independence model, in which the deleted entries are in the cells on the main diagonal in the square $I \times J$ table (with $I = J$), then led me to be able to see how to develop the corresponding statistical methods for the more general multiplicative quasi-independence model, in which the deleted entries are in any specified subset of the cells in the $I \times J$ table (where I need not be equal to J). This then led me to be able to see how to develop the corresponding statistical methods for many other multiplicative models, including the collection of models for survey analysis in the next section.

My work on the simple quasi-independence model began with an interest in the analysis of some data that did not come from the field of social stratification and mobility (see Goodman 1961; 1963; 1964a,b). Then, with an interest in the analysis of mobility tables, the concept of quasi-perfect mobility (later called simple quasi-independence) was introduced, and additional multiplicative models together with the corresponding statistical methods, also of substantive interest in the study of mobility, were developed (see, e.g., Goodman 1965; 1969a,b; 1971b; 1972a; 1979b; 1981b; 1982; Goodman & Clogg 1992). In addition, other multiplicative models together with the corresponding statistical methods, of substantive interest in other areas of study, were also developed (see, e.g., Goodman 1968; 1975; 1979a,c,f; 1981a; 2002a).

Before concluding this subsection, note should be taken of the fact that the descriptions of a few of the models considered in some of the references cited above require the use of terms more complicated than those used here. This is the case, as was noted above in this subsection, in the description of the uniform association model as a multiplicative model; and a corresponding comment should also be made, for example, about the RC association model introduced in Goodman 1979a (see, e.g., equations 4.5b and 4.6b).

CONCEPTS USEFUL IN SURVEY ANALYSIS: LOGLINEAR MODELS

I describe here in simple terms some concepts useful in survey analysis, using log-linear models, but without using logarithms. I usually prefer to use the term multiplicative models, rather than loglinear models, for the kinds of models that are described in this section, but it turns out that the term loglinear models is a more popular way to describe these models for now. The general quasi-independence model, expressed in terms of multiplicative effects on the expected frequencies (see, e.g., Goodman 1968), led me to be able to see how to develop the still more general multiplicative models that are now called loglinear models (see, e.g., the multiplicative models in Goodman 1972b,c; 1973a,b; 1979a,b).

In the preceding section, the analysis of two-way tables is considered. Now, for the analysis of surveys, the analysis of m-way tables (for $m = 2,3,\ldots$) is of interest. In his description of my work in developing models for survey analysis, Duncan (1974, 1975) wrote as follows:

> Goodman's collection of models for survey analysis . . . has provided *for the first time* a set of statistical methods that are adequate to the tasks posed by the "language of social research" hitherto associated with the Columbia school and kindred approaches to survey analysis. The practiced user of Goodman's methods can accomplish with ease everything that this school attempted, and a great deal more . . . . It is . . . no doubt portentous that almost any complex body of data previously analyzed by even a skilled practitioner of survey analysis yields different conclusions by Goodman's methods . . . . It is easy, moreover, to see after the fact how the practitioner fell into . . . error . . . . Many survey researchers are not yet aware of the magnitude of the revolution that Goodman's methods are producing . . . .

I now describe in simple terms some concepts useful in survey analysis. For the sake of simplicity, I first consider the three-way table {A,B,C} of observed frequencies as noted on the left side of **Table 4*a*** in **Table 4**. Variables A, B, and C in this table can pertain, say, to three different questions in a survey; each survey respondent's response can be at either level 1 or level 2 (e.g., either a Yes response or a No response) on each of the three questions. We see, for example, in **Table 4*a*** that there are 25 respondents whose responses were at level 1 on all three questions, and there are 15 respondents whose responses were at level 1 on questions A and C, and at level 2 on question B. The three-way **Table 4*a*** can be viewed as a cube, with variable A described

Table 4 An example of a loglinear analysis of data in a three-way table {A,B,C} of observed frequencies and the corresponding two-way marginal tables, with the corresponding cross-product ratios

Table 4a

		C					
		1		2			
		B		B		B	
		1	2	1	2	1	2
A	1	25	15	12	16	37	31
A	2	8	9	10	25	18	34

Table 4b

		B					
		1		2			
		C		C		C	
		1	2	1	2	1	2
A	1	25	12	15	16	40	28
A	2	8	10	9	25	17	35

Table 4c

		A					
		1		2			
		C		C		C	
		1	2	1	2	1	2
B	1	25	12	8	10	33	22
B	2	15	16	9	25	24	41

Table 4d The corresponding cross-product ratios

a. [AB|C = 1] = 15/8 = 1.87, [AB|C = 2] = 15/8 = 1.87, [AB] = 2 .25
b. [AC|B = 1] = 125/48 = 2.60, [AC|B = 2] = 125/48 = 2.60, [AC] = 2 .94
c. [BC|A = 1] = 20/9 = 2.22, [BC|A = 2] = 20/9 = 2.22, [BC] = 2 .56

by the row categories, variable B described by the column categories, and variable C described by the layer categories. Similarly, in the three-way table {A,C,B} in **Table 4*b***, the row, column, and layer categories in the cube describe variables A, C, and B, respectively; in the three-way table {B,C,A} in **Table 4*c***, they describe variables B, C, and A, respectively. The corresponding two-way marginal tables {A,B}, {A,C}, and {B,C} are on the right side of **Tables 4*a***, **4*b***, and **4*c***, respectively.

From **Table 4*d***, we see that, for the cube on the left side of **Table 4*a***, the cross-product ratio in the 2×2 row by column table $\{A,B \mid C = 1\}$ at layer level 1 is equal to the corresponding cross-product ratio in the 2×2 table $\{A,B \mid C = 2\}$ at layer level 2. This is also the case for the cubes in **Tables 4*b*** and **4*c***, comparing table $\{A,C \mid B = 1\}$ with table $\{A,C \mid B = 2\}$, and table $\{B,C \mid A = 1\}$ with table $\{B,C \mid A = 2\}$. Thus, we see that the three-way table {A,B,C} in **Table 4** exhibits zero three-factor interaction; i.e., the cross-product ratio for any two of the variables in the cube is unaffected by the level of the third variable.

Also from **Table 4*d***, we see that for the two-way table on the right side of **Table 4*a*** the cross-product ratio, [AB] = 2.25, is larger than the corresponding cross-product ratio, [AB | C] = 1.87. Similarly, we see that the cross-product ratio [AC] is larger than the corresponding [AC | B], and [BC] is larger than [BC | A]. In any three-way table exhibiting zero three-factor interaction, the relationship between the cross-product ratios [AB] and [AB | C] is completely determined by whether (1) the two cross-product ratios [AC | B] and [BC | A] are both larger than 1 or both less than 1, or (2) one of the two cross-product ratios is larger than 1 and the other is less than 1, or (3) one or both of the two cross-product ratios is equal to 1 (see Goodman 1972b, 2004). A similar kind of statement also applies when comparing [AC] with [AC | B], and [BC] with [BC | A].

The model of zero three-factor interaction in the three-way table {A,B,C} is one example in the collection of models for survey analysis. All of the models in this collection are multiplicative models with various kinds of appropriate multiplicative effects. In **Figure 1**, a diagram describing the zero three-factor interaction model (**Figure 1*a***), and three other diagrams describing three other models in the three-way table (**Figures 1*b***, **1*c***, **1*d***), are presented. **Figure 1*b*** shows the simple model of conditional independence between variables B and C, given the level of variable A; **Figure 1*c*** shows the simple model of independence between the joint variable AB and variable C; and **Figure 1*d*** shows the simple model of mutual independence between variables A, B, and C. There are actually 8 different kinds of models in the collection of models for survey analysis applied to the three-way table (7 rather simple kinds of models and 1 somewhat less simple), and there are 27 different kinds of models that can be applied to the four-way table (17 rather simple kinds of models and 10 kinds that are somewhat less simple) (see Goodman 1970, tables 3 and 4). The collection of models for survey analysis can be applied to m-way tables, for $m = 2,3,4,\ldots$.

The three-way table {A,B,C} of observed frequencies on the left side of **Table 4*a*** can be viewed as describing the observed relationships among the three dichotomous variables A, B, and C. Table {A,B,C} here is a $2 \times 2 \times 2$ table, a cube. The multiplicative models considered for the $2 \times 2 \times 2$ table in this section can be applied more generally to the $I \times J \times K$ table (for $I = 2,3,\ldots$; $J = 2,3,\ldots$; $K = 2,3,\ldots$). The multiplicative models considered for the m-way table (for $m = 2,3,4,\ldots$), describing the observed relationships among m dichotomous variables, can be applied more generally to the m-way table describing the observed relationships among m polytomous variables (see, e.g., Goodman 1970, 1971a, 1973c).

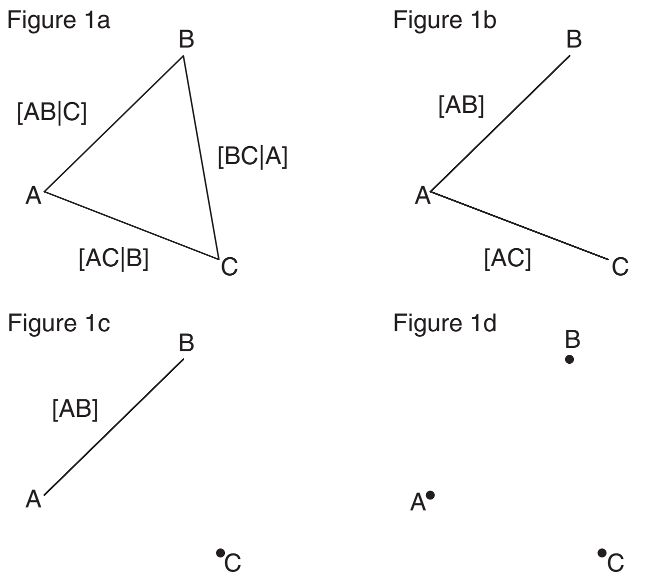

Figure 1

Diagrams of four possible relationships among the three variables in a three-way table {A,B,C}.

CONCEPTS USEFUL IN PANEL ANALYSIS: RECURSIVE MODELS AND MORE

In a two-wave panel study on two questions of interest, where each question has two possible responses, the data of interest can be described in a two-way 4×4 table. The four row categories in the table represent the four possible responses to the two questions (labeled, say, A and B) in the first wave (at, say, time 1), and the four column categories represent the four possible responses to the same two questions (labeled, say, C and D) in the second wave (at, say, time 2). The 16 cells of the 4×4 tables correspond to the 16 possible response patterns for each of the panel respondents on items A, B, C, and D; the entry in each cell of the table is the observed frequency of the corresponding response pattern. We let {AB,CD} denote the two-way 4×4 table pertaining to the two-wave panel data. More generally, for a T-wave panel study ($T = 2,3,\ldots$) on m questions of interest ($m = 1,2,3,\ldots$), where each question has two possible responses, the data on the frequency of the possible response patterns of the panel respondents can be described in a T-way $2^m \times 2^m \times \ldots \times 2^m$ table. In his description of my work on developing models for the analysis of panel data, Duncan (1974, 1975) wrote as follows:

> Goodman has put panel analysis on a sound footing for the first time and, as a consequence, we can now ignore a substantial body of misguided literature that provided erroneous, misleading, or merely useless procedures for manipulating panel data.

I now describe in simple terms some concepts useful in the analysis of panel data. For the sake of simplicity, I consider the two-way 4×4 table {AB,CD} of observed frequencies shown in **Table 5*a*** of **Table 5**, describing data obtained in a two-wave panel study on two questions, where each question has two possible responses. In analyzing these data, I first apply a "recursive" model to the data, with items A and B viewed as prior to items C and D, and items A, B, and C viewed as prior to item D. (Later in this section, I consider a recursive model in which items A and B are prior to C and D, and items C and D are simultaneously posterior to A and B.)

Table 5 An example of the analysis of panel data using a recursive model in analyzing the observed frequency of response patterns in a two-wave panel study {AB,CD} on two questions, where each question has two possible responses

Table 5a

AB	CD 11	12	21	22
11	144	18	36	18
12	27	54	3	24
21	16	2	36	18
22	12	24	12	96

Table 5b

AB	C 1	2
11	162	54
12	81	27
21	18	54
22	36	108

Table 5c

AB	Total
11	216
12	108
21	72
22	144

Table 5d

AC	Total
11	243
12	81
21	54
22	162

Table 5e

ABC	D 1	2
111	144	18
112	36	18
121	27	54
122	3	24
211	16	2
212	36	18
221	12	24
222	12	96

Table 5f Corresponding cross-product ratios

[AB]	=	4	[AC]	=	9
[AB|C = 1]	=	4	[AB|C = 2]	=	4
[AC|B = 1]	=	9	[AC|B = 2]	=	9
[BC|A = 1]	=	1	[BC|A = 2]	=	1
[AD|BC = 11]	=	1	[AD|BC = 12]	=	1
[AD|BC = 21]	=	1	[AD|BC = 22]	=	1
[BD|AC = 11]	=	16	[BD|AC = 12]	=	16
[BD|AC = 21]	=	16	[BD|AC = 22]	=	16
[CD|AB = 11]	=	4	[CD|AB = 12]	=	4
[CD|AB = 21]	=	4	[CD|AB = 22]	=	4

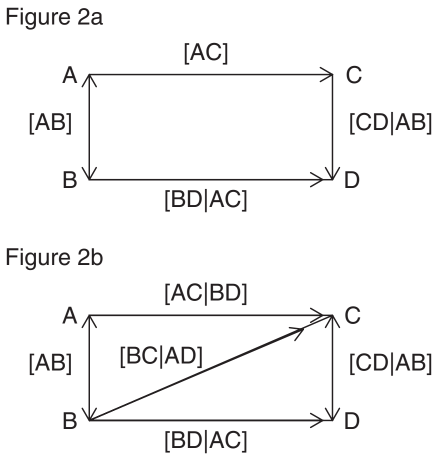

Figure 2

Diagrams of two possible relationships among the variables in a two-wave panel study on two questions.

Consider now **Tables 5*a*,*b*,*c*,*d***, and ***e***. **Tables 5*b*** and **5*e*** can be obtained directly from **Table 5*a***; and **Tables 5*c*** and **5*d*** can be obtained directly from **Table 5*b***. **Tables 5*c*** and **5*d*** are 4×1 tables, {AB} and {AC}, respectively, which can also be viewed as 2×2 tables, {A,B} and {A,C}, respectively; **Table 5*b*** is a 4×2 table {AB,C}, which can also be viewed as a three-way table {A,B,C}. For the 2×2 tables {A,B} and {A,C} in **Tables 5*c*** and **5*d***, we see from **Table 5*f*** that the corresponding cross-product ratios are $[AB] = 4$ and $[AC] = 9$, respectively; for the $2 \times 2 \times 2$ table {A,B,C} in **Table 5*b***, we see from **Table 5*f*** that $[AB \mid C] = 4$, $[AC \mid B] = 9$, and $[BC \mid A] = 1$. Thus, we also see that the response on item B and the response on item C are conditionally independent of each other, given the response on item A (see, e.g., **Figure 1*b*** presented in the preceding section).

Now let us focus our attention on **Table 5*e***, which can be viewed as an 8×2 table {ABC,D}. Considering the responses on items A, B, and C as possible predictors of the response on item D, we see from **Table 5*f*** that the corresponding cross-product ratios are $[AD \mid BC] = 1$, $[BD \mid AC] = 16$, and $[CD \mid AB] = 4$. Thus, we also see that, given the response on items B and C, the response on item D is conditionally independent of the response on item A.

Figure 2*a* in **Figure 2** shows the results obtained above using the recursive model considered there. **Figure 2*b*** illustrates the results obtained using the recursive model in which items C and D are simultaneously posterior to A and B. The relationship between items B and C included in **Figure 2*b*** is not included in **Figure 2*a*** because, in the three-way table {A,B,C} (as noted above), the response on item C is conditionally independent of the response on item B, given the response on item A. The relationship between items B and C is included in **Figure 2*b*** because we see from **Tables 5*a*** and/or **5*e*** that the corresponding cross-product ratio is $[BC \mid AD = 11] = (144 \times 3)/(36 \times 27) = 4/9$, and, more generally, that $[BC \mid AD] = 4/9$. The relationship between items A and D is not included in **Figures 2*a*** and **2*b*** because we see from **Table 5*f*** that the corresponding cross-product ratio is $[AD \mid BC] = 1$, and so the response on item D is conditionally independent of the response on item A, given the response on items B and C.

Figure 2*a* was obtained by using multiplicative models to analyze, in turn, the two-way 2×2 table {A,B}, the two-way 4×2 table {AB,C} (which can also be viewed as the three-way $2 \times 2 \times 2$ table {A,B,C}), and the two-way 8×2 table {ABC,D} (which can also be viewed as the four-way $2 \times 2 \times 2 \times 2$ table {A,B,C,D}), whereas **Figure 2*b*** was obtained by using multiplicative models to analyze, in turn, only the two-way 2×2 table {A,B} and the four-way table {A,B,C,D}. The statistical methods that were developed earlier to be able to apply appropriate multiplicative models in the analysis of survey data then led me to be able to see how to develop the corresponding statistical methods to be able to apply, in turn, the corresponding models appropriate for the analysis of panel data. For additional models for the analysis of panel data, the interested reader is referred to Goodman (1962; 1973a,b; 1979e) and Duncan (1985).

Before concluding this section, I include the following brief comment for those readers who use a loglinear computer program: The numerical values for the four cross-product ratios in **Figure 2*a***, and the additional two cross-product ratios in **Figure 2*b***, have been given either in **Table 5*f*** or above in this section. As discussed above, all these numerical values were obtained here using very simple multiplication and division. In contrast to this very simple arithmetic, a loglinear computer program does something different—something somewhat more complicated, using iterative computer algorithms. However, if we compare the numerical values of the cross-product ratios obtained here by this simple arithmetic with the corresponding estimated parameters obtained as the output of the loglinear computer program, we find that each estimated parameter in this computer output is simply equal to 0.25 times the natural logarithm of the corresponding cross-product ratio.

CONCEPTS USEFUL IN LATENT-STRUCTURE ANALYSIS: LATENT-CLASS MODELS

The statistical methods that were developed earlier to enable the application of the multiplicative models for survey analysis also helped to lead me to be able to see how to develop the corresponding statistical methods for latent-structure analysis. As we noted above in the survey analysis section, the methods considered there could be applied to analyze the m-way table of observed frequencies, for $m = 2,3,4,\ldots$. Now, for the sake of simplicity in our description of latent-class models, we again consider the three-way table {A,B,C}. The three observed variables (A, B, and C) pertaining to this table can be referred to as manifest variables. In latent-class analysis, the following question is considered: Is there an unknown latent (unobserved or unobservable) categorical variable X (a dichotomous or polytomous variable) that can explain the observed relationships among the three observed manifest variables (A,B,C) in the three-way table {A,B,C}? In other words, is there, for example, a latent dichotomous variable having two latent categories (latent classes), which are such that, for those respondents in each of the two latent classes, the responses on the manifest variables A, B, and C are mutually independent? We can think of this question in the following terms: For the three-way table {A,B,C}, is there a corresponding four-way table {A,B,C,X} in which the manifest variables A, B, and C are mutually independent for the respondents in each category (class) of variable X? Mutual independence of the kind described above is called conditional mutual independence. Recall that **Figure 1*b*** in **Figure 1** is a portrait of conditional independence—namely, the conditional independence between variables B and C, given the level of variable A—which we can denote as $[B \otimes C \mid A]$ in the three-way table {A,B,C} (see Goodman 1970, table 3). Now, in the four-way table {A,B,C,X} considered above, the multiplicative model of interest can be described as $[A \otimes B \otimes C \mid X]$. So our task is to develop a method for determining how many respondents in the three-way marginal table {A,B,C} of the four-way table {A,B,C,X} are in each category of variable X, in such a way so that the model $[A \otimes B \otimes C \mid X]$ is congruent with the data in the four-way table. In his description of my work on this subject, Duncan (1974, 1975) wrote as follows:

> Goodman has provided a substantial statistical foundation for the latent structure model of Lazarsfeld . . . . It is notorious that for the 25 or 30 years that these models have been discussed and applied . . . the estimation and testing procedures suggested for the models by their inventors and employed, faute de mieux, by research workers [have not been satisfactory] . . . . The statistical problems had defeated . . . some very eminent statisticians. Now, thanks to Goodman, [by using the methods presented in his statistical foundation] we can begin to understand

> correctly what is at stake.... Prestidigitation will no longer suffice as a legitimation for some ad hoc procedure..., nor will incantation of a rule of thumb....

I now illustrate, by example, a simple latent structure, a latent-class model that has two latent classes, applied to a three-way table of observed frequencies. Let us consider the three-way table {A,B,C} of observed frequencies in **Table 6*a*** in **Table 6**. This three-way table can be viewed as describing how respondents in a survey responded (giving either response 1 or response 2—e.g., either a Yes response or a No response) on questions A, B, and C. Is it possible that some of these respondents are in one latent (unobserved or unobservable) class and the remaining respondents are in the other latent class, and the three-way table describing the responses of those respondents in the first latent class is such that $[A \otimes B \otimes C \mid X = 1]$ and the corresponding three-way table for the respondents in the second latent class is such that $[A \otimes B \otimes C \mid X = 2]$? The answer to this question is: Yes, it is possible. **Tables 6*b*** and **6*c*** describe the three-way tables for the respondents in the two latent classes; **Table 6*d*** gives the proportion of respondents in each of the latent classes and the probabilities of a 1 response or a 2 response on questions A, B, and C, for the respondents in each latent class. We see from **Table 6*d***

Table 6 An example of a latent-class analysis of data in a three-way table {A,B,C} of observed frequencies, with the corresponding latent-class proportions and response probabilities

Table 6a

	C = 1		C = 2	
	B = 1	B = 2	B = 1	B = 2
A = 1	25	15	12	16
A = 2	8	9	10	25

Table 6b

	C = 1		C = 2	
	B = 1	B = 2	B = 1	B = 2
A = 1	24	12	8	4
A = 2	6	3	2	1

Table 6c

	C = 1		C = 2	
	B = 1	B = 2	B = 1	B = 2
A = 1	1	3	4	12
A = 2	2	6	8	24

Table 6d Latent-class proportions and response probabilities

	Latent-Class 1			Latent-Class 2		
Proportion	.50			.50		
Variables	A	B	C	A	B	C
Response	Probabilities			Probabilities		
1	.80	.67	.75	.33	.25	.20
2	.20	.33	.25	.67	.75	.80

that the respondents in the first latent class are more likely to give a 1 response (rather than a 2 response) on each of the three questions, whereas the reverse is true for those in the second latent class.

In Duncan's description of my work on statistical methods for latent structure analysis, which was quoted above in this section, he was referring implicitly to Goodman (1974a,b). Readers interested in this general topic are also referred to, e.g., Clogg & Goodman 1984, 1985; Goodman 1979d; 1987a,b; 2002b; 2004; 2005; 2007.

In concluding this section, let me again note that the three-way table latent-class analysis presented above in this section is used here for the sake of simplicity. Latent-class analysis can be applied, more generally, in the analysis of an m-way table (for $m = 2,3,4,\ldots$) of observed frequencies pertaining to m dichotomous or polytomous observed (manifest) variables, where there may be one or more latent (unobserved or unobservable) dichotomous or polytomous variables.

CONCLUDING COMMENTS: MAGIC AND/OR SERENDIPITY?

The idea of looking at a set of data and seeing the different images embedded in the set, and releasing them so that they have a comprehensible form, seems to me to be magical, although of course not as magical as Michelangelo releasing the form of *David* from the block of marble that he had selected. By magical, I do not mean here something supernatural, but rather a result obtained as if by magic.

The data analyst's ability to release a comprehensible form or comprehensible forms from the block of data, using the concepts and tools of categorical data analysis and/or whatever other appropriate concepts and tools that he or she might have, seems to me quite magical. For example, we are able to look at, say, a three-way table of observed frequencies pertaining to three dichotomous or polytomous variables and see whether these variables are related to each other, and if so, how. In addition, we are able to see whether there might be a latent (unobserved or unobservable) dichotomous or polytomous variable that could explain, or explain away, the observed relationships among the three observed variables. Similarly, we are able to look at an m-way table pertaining to m dichotomous or polytomous variables (for $m = 2,3,4,\ldots$) and see whether these variables are related to each other, and if so, how. In addition, we are able to see whether there might be one or more latent dichotomous or polytomous variables that could explain, or explain away, the observed relationships among the m observed variables.

Each of the concepts described in this essay was introduced in the earlier literature using a mathematical approach with formulas, and the application of these concepts also required the aid of a computer program constructed to apply appropriate iterative computer algorithms. (Strictly speaking, the statement above does not apply to latent-class analysis, whose initial introduction did use a mathematical approach with formulas, but it did not use iterative computer algorithms. This initial approach turned out to be unsatisfactory. However, with the later introduction of a different mathematical approach with different formulas, and an appropriate iterative computer algorithm, a satisfactory method of analysis was obtained; see, e.g., Goodman 1974a,b; 2002b.) In the description of each of these concepts (including the latent-class analysis concept) presented in this essay, by contrast, no use is made of mathematical formulas or iterative computer algorithms. As noted at the beginning of this essay, the only arithmetic used is very simple multiplication, division, and addition. Is this magic, or serendipity, or both, or neither?

The results obtained by applying the concepts described in this essay to substantive data of interest sometimes seem magical—the sudden release of form formerly hidden, embedded in a block of dense data—but perhaps "serendipity" better describes the way in which these concepts were developed. In this essay, I have noted that the information to which I was exposed in my work on one statistical problem, in one substantive area of interest, then led me to be able to look at a second substantive area of interest, and then be able to see what was the statistical problem that needed to be dealt with there, and how to proceed with work on it. And the information to which I was exposed in my work on a second statistical problem in a second substantive area of interest then led me to be able to look at a third substantive area of interest, and then be able to see . . . and so forth, and so on. By a serendipitous result, I do not mean here a result obtained simply by accident or chance, but rather a result obtained by an accidental exposure to information and a prepared mind.

In concluding this section, I refer the reader interested in topics covered in this essay to the list of ten books included in the Related Resources section at the end, and to the literature cited in these books. Categorical data analysis is a growth industry.

ACKNOWLEDGMENT

For helpful comments, the author is indebted to Mike Hout.

LITERATURE CITED

Caussinus H. 1966. Contribution à l'analyse statistique des tableaux de correlation. *Ann. Fac. Sci. Univ. Toulouse* 29:77–182

Clogg CC, Goodman LA. 1984. Latent structure analysis of a set of multidimensional contingency tables. *J. Am. Stat. Assoc.* 79:762–71

Clogg CC, Goodman LA. 1985. Simultaneous latent structure analysis in several groups. *Sociol. Methodol.* 16:81–110

Duncan OD. 1974. Stouffer award: Duncan requests reconsideration of award. *Am. Sociol. Assoc. Footnotes* 2:2–3

Duncan OD. 1975. The statistics of Leo Goodman. *Soc. Indic. Newsl.* 6:4–5

Duncan OD. 1979. How destination depends on origin in the occupational mobility table. *Am. J. Sociol.* 84:793–803

Duncan OD. 1985. New light on the 16-fold table. *Am. J. Sociol.* 91:88–128

Goodman LA. 1961. Statistical methods for the mover-stayer model. *J. Am. Stat. Assoc.* 56:841–68

Goodman LA. 1962. Statistical methods for analyzing processes of change. *Am. J. Sociol.* 68:57–78

Goodman LA. 1963. Statistical methods for the preliminary analysis of transaction flows. *Econometrica* 31:197–208

Goodman LA. 1964a. A short computer program for the analysis of transaction flows. *Behav. Sci.* 9:176–86

Goodman LA. 1964b. The analysis of persistence in a chain of multiple events. *Biometrika* 51:405–11

Goodman LA. 1965. On the statistical analysis of mobility tables. *Am. J. Sociol.* 70:564–85

Goodman LA. 1968. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Am. Stat. Assoc.* 63:1091–31

Goodman LA. 1969a. How to ransack social mobility tables and other kinds of cross-classification tables. *Am. J. Sociol.* 75:1–40

Goodman LA. 1969b. On the measurement of social mobility: An index of status persistence. *Am. Sociol. Rev.* 34:831–50

Goodman LA. 1970. The multivariate analysis of qualitative data: Interactions among multiple classifications. *J. Am. Stat. Assoc.* 65:226–56
Goodman LA. 1971a. The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics* 13:33–61
Goodman LA. 1971b. A simple simultaneous test procedure for quasi-independence in contingency tables. *J. R. Stat. Soc. Ser. C* 20:165–77
Goodman LA. 1972a. Some multiplicative models for the analysis of cross-classified data. In *Proc. 6th Berkeley Symp. Math. Stat. Probability*, ed. LM LeCam, J Neyman, EL Scott, 1:649–96. Berkeley: Univ. Calif. Press
Goodman LA. 1972b. A general model for the analysis of surveys. *Am. J. Sociol.* 77:1035–86
Goodman LA. 1972c. A modified multiple regression approach to the analysis of dichotomous variables. *Am. Sociol. Rev.* 37:28–46
Goodman LA. 1973a. Causal analysis of data from panel studies and other kinds of surveys. *Am. J. Sociol.* 78:1135–91
Goodman LA. 1973b. The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika* 60:179–92
Goodman LA. 1973c. Guided and unguided methods for the selection of models for a set of multidimensional contingency tables. *J. Am. Stat. Assoc.* 68:165–75
Goodman LA. 1974a. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *Am. J. Sociol.* 79:1179–259
Goodman LA. 1974b. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61:215–31
Goodman LA. 1975. A new model for scaling response patterns: An application of the quasi-independence concept. *J. Am. Stat. Assoc.* 70:755–68
Goodman LA. 1979a. Simple models for the analysis of association in cross-classifications having ordered categories. *J. Am. Stat. Assoc.* 74:537–52
Goodman LA. 1979b. Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classification tables. *Am. J. Sociol.* 84:804–19
Goodman LA. 1979c. Multiplicative models for square contingency tables with ordered categories. *Biometrika* 66:413–18
Goodman LA. 1979d. The analysis of qualitative variables using more parsimonious quasi-independence models, scaling models, and latent structures. In *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld*, ed. RK Merton, JS Coleman, PH Rossi, pp. 119–37. New York: Free Press
Goodman LA. 1979e. A brief guide to the causal analysis of data from surveys. *Am. J. Sociol.* 84:1078–95
Goodman LA. 1979f. On quasi-independence in triangular contingency tables. *Biometrics* 35:651–55
Goodman LA. 1981a. Three elementary views of loglinear models for the analysis of cross-classifications having ordered categories. *Sociol. Methodol.* 12:193–239
Goodman LA. 1981b. Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table. *Am. J. Sociol.* 87:612–50
Goodman LA. 1982. On a structural analysis of occupational mobility. In *Social Structure and Network Analysis*, ed. PV Marsden, N Lin, pp. 32–36. Beverly Hills, CA: Sage
Goodman LA. 1987a. New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *Am. J. Sociol.* 93:529–83

Goodman LA. 1987b. The analysis of a set of multidimensional contingency tables using log-linear models, latent-class models, and correlation models: The Solomon data revisited. In *Contributions to the Theory and Application of Statistics: A Volume in Honor of Herbert Solomon*, ed. AE Gelfand, pp. 429–84. New York: Academic

Goodman LA. 2002a. Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi-independence, log-linear models, log-bilinear models, and correspondence analysis models. *Ann. Fac. Sci. Univ. Toulouse, Math.* 11:525–40

Goodman LA. 2002b. Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In *Applied Latent Class Analysis*, ed. JA Hagenaars, AL McCutcheon, pp. 2–55. Cambridge, UK: Cambridge Univ. Press

Goodman LA. 2004. *Three different ways to view cross-classified categorical data: Rasch-type models, log-linear models, and latent-class models. The Inaugural Otis Dudley Duncan Honorary Lecture.* Presented at Annu. Winter/Spring Conf. Am. Sociol. Assoc. Methodol. Sect., Ann Arbor, MI

Goodman LA. 2005. *Some simple latent structure models, based on Rasch-type latent-trait models and on latent-class models, for the analysis of cross-classified categorical data. Keynote Lecture.* Presented at Conf. Latent Variable Models Soc. Sci., Stat. Appl. Math. Sci. Inst., Research Triangle Park, NC

Goodman LA. 2007. On the assignment of individuals to latent classes. *Sociol. Methodol.* In press

Goodman LA, Clogg CC. 1992. New methods for the analysis of occupational mobility and other kinds of cross-classifications. *Contemp. Sociol.* 21:609–22

RELATED RESOURCES

Agresti A. 1996. *An Introduction to Categorical Data Analysis*. New York: Wiley

Agresti A. 2002. *Categorical Data Analysis*. Hoboken, NJ: Wiley. 2nd ed.

Fienberg SE. 1980. *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: MIT Press. 2nd ed.

Goodman LA. 1978. *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis.* Cambridge, MA: Abt Books/Lanham, MD: Univ. Press Am.

Goodman LA. 1984. *The Analysis of Cross-Classified Data Having Ordered Categories*. Cambridge, MA: Harvard Univ. Press

Haberman SJ. 1978/1979. *Analysis of Qualitative Data*, Vols. 1, 2. New York: Academic

Hagenaars JA, McCutcheon AL, eds. 2002. *Applied Latent Class Analysis.* Cambridge, UK: Cambridge Univ. Press

Hout M. 1983. *Mobility Tables.* Beverly Hills, CA: Sage

Powers DA, Xie Y. 2000. *Statistical Methods for Categorical Data Analysis.* San Diego, CA: Academic