Anatomy of Pearson's Chi-square Statistic in Three-way Contingency Tables

Yoshio Takane and Lixing Zhou

Abstract We consider orthogonal decompositions of Pearson's chi-square statistic in three-way contingency tables. We derive algebraic formulae for the decompositions, conditionally on given marginal frequencies. Results indicate that the order in which various effects are taken into account play a crucial role. This is analogous to multiple regression analysis with correlated predictor variables. Because of their orthogonality, terms in the decompositions follow independent asymptotic chi-square distributions under suitable null hypotheses. We also compare our results with partitions of the log likelihood ratio (LR) chi-square associated with log linear models for contingency tables.

1 Introduction

Research in psychology and other social sciences often involves discrete multivariate data. Such data are conveniently summarized in the form of contingency tables. There have been two widely used classes of techniques for analysis of such tables. One is log linear models (e.g.,Andersen, 1980; Bishop et al., 1975), and the other is correspondence analysis (CA; e.g., Greenacre, 1984; Nishisato, 1980). The former allow ANOVA-like decompositions of the log likelihood ratio (LR) statistic (also known as the deviance statistic or the Kullback-Leibler (1951) divergence). This statistic measures the difference in log likelihood between the saturated and independence models. When the latter model is correct, it follows the asymptotic

Yoshio Takane
Department of Psychology, University of Victoria, P. O. Box 3050 Victoria, BC V8W 3P5 Canada, e-mail: takane@uvic.ca

Lixing Zhou
Department of Psychology, McGill University, 1205 Dr. Penfield Ave. Montreal QC H3A 1B1 Canada, e-mail: lixing.zhou@mail.mcgill.ca

chi-square distribution with degrees of freedom (df) equal to the difference in the number of parameters in the two models.

In CA, on the other hand, an emphasis is placed on graphical representations of associations between rows and columns of contingency tables. This approach typically uses PCA-like (componentwise) decompositions of Pearson's (1900) chi-square statistic, measuring essentially the same thing as the log LR chi-square statistic. In this paper, we develop ANOVA-like decompositions of Pearson's chi-square statistic, similar to those for the log LR statistic.

These decompositions are useful in constrained CA, such as canonical correspondence analysis (CCA; ter Braak, 1986) and canonical analysis with linear constraints (CALC; Böckenholt & Böckenholt, 1990), in which the total association between rows and columns of contingency tables is decomposed into what can and cannot be explained by the constraints. Different terms in the decompositions highlight different aspects of the total association. The terms in the proposed decompositions are mutually orthogonal and follow independent asymptotic chi-square distributions under suitable null hypotheses. This is in contrast with the decompositions suggested by Lancaster (1951), in which individual terms do not necessarily follow asymptotic chi-square distributions (Placket, 1962). All terms in the proposed decompositions can be obtained in closed form unlike some of the terms in the decompositions of the log LR chi-square statistic.

Takane and Jung (2009b) proposed similar decompositions of the CATANOVA C-statistic (Light & Margolin, 1971), which also follows an asymptotic chi-square distribution. This statistic, however, has been developed for situations in which rows and columns of contingency tables assume asymmetric roles, that is, one is the predictor, and the other is the criterion. It thus represents the overall predictability of, say, rows on columns. Pearson's chi-square statistic, on the other hand, represents a symmetric association. It may be argued, however, that a symmetric measure of association may still be useful in the predictive contexts. There are many cases in which symmetric analysis methods (those that do not distinguish between predictors and criterion variables) are used for prediction purposes. For example, canonical correlation analysis (Hotelling, 1936) and its special cases, canonical discriminant analysis (Fisher, 1936), CCA and CALC (cited above), reduced rank regression analysis (Anderson, 1951; Izenman, 1975), maximum likelihood reduced-rank GMANOVA (growth curve models; Reinsel & Velu, 1998), and the curds and whey method (Breiman & Friedman, 1997) all involve some kind of symmetric analysis. This suggests that decompositions of a symmetric measure of association, such as Pearson's chi-square statistic, may well be useful in predictive contexts.

This paper is organized as follows. Section 2 briefly reviews basic facts about Pearson's chi-square statistic and its historical development. Section 3 presents our main results, the proposed decompositions, starting from elementary two-term decompositions to full decompositions. It will be shown that the order in which various effects are taken into consideration plays a crucial role in deriving the decompositions. Section 4 compares the proposed decompositions to those for the log LR statistic recently proposed by Cheng et al. (2006). Section 5 draws conclusions.

2 Preliminaries

We use upper-case Roman alphabets (e.g., A, B, ...) to designate variable names and the corresponding characters in italic (e.g., A, B, ...) to denote the number of categories (levels) in the variables. Categories of a variable are indexed by the corresponding lower case alphabets in italic (e.g., $a = 1, \cdots A$).

Let there be A mutually exclusive events with known probabilities of occurrence, p_a ($a = 1, \cdots, A$), and let f_a ($a = 1, \cdots, A$) denote the observed frequency of event a out of N replicated observations. Then the following statistic

$$\chi^2_A = \sum_{a=1}^{A} \left(\frac{f_a - Np_a}{\sqrt{Np_a}} \right)^2 \tag{1}$$

asymptotically follows the chi-square distribution with A df (Pearson, 1900). Here, Np_a is the expected value of f_a under the prescribed conditions. This is the generic form of Pearson's chi-square statistic, from which many special cases follow.

In one-way layouts (i.e., when there is only one categorical variable), we are typically interested in testing $\text{H}_0 : p_a = p$ for all a ($a = 1, \cdots, A$). We estimate p by $\hat{p} = 1/A$. If we insert this estimate in (1), we obtain

$$\chi^2_{A-1} = \sum_{a=1}^{A} \left(\frac{f_a - N/A}{\sqrt{N/A}} \right)^2. \tag{2}$$

This statistic follows the asymptotic chi-square distribution with $A-1$ df under H_0. Note that we lose 1 df for estimating p. When $A > 2$, the above statistic can be partitioned into the sum of $A-1$ independent chi-square variables each with 1 df. Let $\mathbf{g}$ denote the A-component vector of $(f_a - N/A)/\sqrt{N/A}$ ($a = 1, \cdots, A$). We may transform this vector by the Helmert type of contrasts for unequal cell sizes (Irwin, 1949; Lancaster, 1949). For $A = 3$, this contrast matrix looks like

$$\mathbf{T} = \begin{bmatrix} \sqrt{\frac{\hat{p}_2}{\hat{p}_1+\hat{p}_2}} & -\sqrt{\frac{\hat{p}_1}{\hat{p}_1+\hat{p}_2}} & 0 \\ \sqrt{\frac{\hat{p}_3\hat{p}_1}{(\hat{p}_1+\hat{p}_2)(\hat{p}_1+\hat{p}_2+\hat{p}_3)}} & \sqrt{\frac{\hat{p}_3\hat{p}_2}{(\hat{p}_1+\hat{p}_2)(\hat{p}_1+\hat{p}_2+\hat{p}_3)}} & -\sqrt{\frac{\hat{p}_1+\hat{p}_2}{\hat{p}_1+\hat{p}_2+\hat{p}_3}} \end{bmatrix}', \tag{3}$$

where $\hat{p}_a = f_a/N$. Define

$$\mathbf{h} = \mathbf{T}'\mathbf{g}. \tag{4}$$

Then each of the $A-1$ elements of $\mathbf{h}$ asymptotically follows the independent standard normal distribution under H_0, whose sum of squares (i.e., $\mathbf{h}'\mathbf{h}$) asymptotically follows the chi-square distribution with $A-1$ df under H_0. Note that $\mathbf{T}$ is not unique. It can be any columnwise orthogonal matrix with one additional requirement that it is also orthogonal to the vector with the square root of $\hat{p}_a$ as the a-th element for $a = 1, \cdots, A$. It can be easily verified that $\mathbf{T}'\mathbf{T} = \mathbf{I}_{A-1}$, and that $\mathbf{T}'\hat{\mathbf{p}} = \mathbf{0}$ for $\mathbf{T}$ defined in (3), where $\hat{\mathbf{p}} = (\sqrt{\hat{p}_1}, \cdots, \sqrt{\hat{p}_A})'$.

In two-way layouts, we assume that there is another variable B with B categories. Let f_{ba} denote the observed frequency of category b of variable B and category a of variable A. Let f_{ba} be arranged in a B by A contingency table $\mathbf{F}$. We are typically interested in testing the independence between the rows and columns of $\mathbf{F}$, i.e., $\mathrm{H}_0 : p_{ba} = p_b p_a$, where p_{ba} is the joint probability of row b and column a, and p_b and p_a are the marginal probabilities of row b and column a, respectively. Let $\hat{p}_b = \sum_a f_{ba}/N$ and $\hat{p}_a = \sum_b f_{ba}/N$ denote the estimates of p_b and p_a, and define

$$\chi^2_{(B-1)(A-1)} = \sum_{b=1}^{B} \sum_{a=1}^{A} \left(\frac{f_{ba} - N\hat{p}_b\hat{p}_a}{\sqrt{N\hat{p}_b\hat{p}_a}} \right)^2. \tag{5}$$

This statistic represents the total association (or the departure from independence) between the rows and columns of $\mathbf{F}$. It is sometimes referred to as the A by B interaction and is denoted as $\chi^2(\mathrm{AB})$. It follows the asymptotic chi-square distribution with $(B-1)(A-1)$ df under H_0. As before, it can be decomposed into the sum of $(B-1)(A-1)$ independent chi-square variables each with 1 df when $B > 2$ and/or $A > 2$. Let $\mathbf{G}$ represent the B by A matrix whose ba-th element is equal to $(f_{ba} - N\hat{p}_b\hat{p}_a)/\sqrt{N\hat{p}_b\hat{p}_a}$. We then pre- and postmultiply $\mathbf{G}$ by something analogous to $\mathbf{T}'$ and $\mathbf{T}$ defined in (3). The resultant matrix has $(B-1)(A-1)$ independent asymptotically standard normal variables under H_0, whose sum of squares follows the asymptotic chi-square distribution with $(B-1)(A-1)$ df.

It will be handy to have a matrix representation of the chi-square statistic given above. Let $\mathbf{K}$ and $\mathbf{L}$ denote the diagonal matrices whose diagonal elements are the row and the column totals of $\mathbf{F}$, and let $\mathbf{Q}_{1/K} = \mathbf{I}_B - \mathbf{1}_B\mathbf{1}_B'\mathbf{K}/N$, where $\mathbf{1}_B$ is the B-element vector of ones. Then, $\mathbf{G}$ can be expressed in terms of $\mathbf{F}$ by

$$\mathbf{G} = \sqrt{N}\mathbf{K}^{-1}\mathbf{Q}_{1/K}'\mathbf{F}\mathbf{L}^{-1} = \sqrt{N}\mathbf{Q}_{1/K}\mathbf{K}^{-1}\mathbf{F}\mathbf{L}^{-1}. \tag{6}$$

The $\chi^2_{(B-1)(A-1)}$ can then be rewritten as

$$\chi^2_{(B-1)(A-1)} = \mathrm{tr}(\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{L}) = \mathrm{SS}(\mathbf{G})_{K,L}. \tag{7}$$

In three-way layouts, we take into account a third variable C with C categories. Let f_{cba} denote the observed frequency of category c of variable C, category b of variable B, and category a of variable A, and define

$$\chi^2_{CBA-C-B-A+2} = \sum_{c=1}^{C} \sum_{b=1}^{B} \sum_{a=1}^{A} \left(\frac{f_{cba} - N\hat{p}_c\hat{p}_b\hat{p}_a}{\sqrt{N\hat{p}_c\hat{p}_b\hat{p}_a}} \right)^2. \tag{8}$$

This statistic represents the departure from independence among the three categorical variables. Under the independence hypothesis (i.e., H_0: $p_{cba} = p_c p_b p_a$), this statistic follows the asymptotic chi-square distribution with $CBA - C - B - A + 2$ df, which are always larger than 1. Consequently it can always be decomposed into the sum of $CBA - C - B - A + 2$ independent chi-square variables each with 1 df.

As in the case of two-way layouts, we can express the above chi-square in matrix notation. We first arrange a three-way table into a two-way format by factorially combining two of the three variables. Suppose that variables B and C are combined to form row categories. (Which two variables we choose to combine makes no difference for our immediate purpose. Note, however, that this will have a rather grave impact on the decompositions of Pearson's chi-square statistic that follow.) We may then take categories of A as columns. Suppose further that the row categories are ordered in such a way that the index for B categories moves fastest. (See Table 1 below for an example.) Let $\mathbf{F}$ denote the two-way table thus constructed. Let $\mathbf{K} = \mathbf{D}_C \otimes \mathbf{D}_B$, where $\mathbf{D}_C$ and $\mathbf{D}_B$ are diagonal matrices with marginal frequencies of categories of variables C and B, and $\otimes$ indicates a Kronecker product. Let $\mathbf{L} = \mathbf{D}_A$ denote the diagonal matrix of column totals of $\mathbf{F}$, and define

$$\mathbf{G} = N\mathbf{K}^{-1}(\mathbf{F} - \mathbf{K}\mathbf{1}_{CB}\mathbf{1}_A'\mathbf{L}/N^2)\mathbf{L}^{-1}. \tag{9}$$

Then

$$\chi^2_{CBA-C-B-A+2} = \mathrm{tr}(\mathbf{G}'\mathbf{K}\mathbf{G}\mathbf{L}) = \mathrm{SS}(\mathbf{G})_{K,L}. \tag{10}$$

Consider, as an example, the three-way contingency table given in Table 1. This is a 2 by 2 by 2 table arranged in a 4 by 2 two-way format according to the prescription given above. This is a famous data set used by Snedecor (1958) to illustrate the differences in the notion of the three-way interaction effect in a three-way contingency table given by several prominent statisticians, including Bartlett (1935), Mood (1950), and Lancaster (1951). According to Cheng et al. (2006), however, all of them made crucial mistakes in conceptualizing the three-way interaction effect. We are going to use this same data set to demonstrate our proposed decompositions of Pearson's chi-square statistic (Section 3) and compare them with those of the log LR statistic (Section 4). For the moment, however, we are satisfied with only calculating χ^2_4 for this data set using the formula given in (8) or (10). This value turns out to be 131.99.

Table 1 A three-way contingency table arranged in two-way format

		A_1	A_2	Total
C_1	B_1	79	177	256
	B_2	62	121	183
C_2	B_1	73	81	154
	B_2	168	75	243
Total		382	454	836

The χ^2_4 for this table reflects the joint effects of four sources, the A by B, A by C, B by C, and A by B by C interaction effects with the main effects of the three variables A, B, and C being eliminated by their marginal probabilities. Thus, χ^2_4 may also be written as χ^2(AB,AC,BC,ABC). Note, however, that these four effects are usually not mutually orthogonal due to unequal marginal frequencies, and consequently their joint effects cannot be obtained by their sum. In this paper, we develop systematic ways of orthogonalizing these effects to make them additive.

3 The Proposed Decompositions

In order to derive proper decompositions of Pearson's chi-square statistic for a three-way contingency table, its reduction to a two-way table seems essential. Table 1 shows one way of reduction. There are two other ways of reducing a three-way table into two, depending on which two of the three variables are combined to create a new variable. In Table 1, B and C were combined, but A and B, and A and C could likewise be combined. Generally, different decompositions result, depending on which reduction method is employed. In this section we start with the reduction method used in Table 1, and then expand our view to other situations.

If we look at Table 1 as purely a two-way table, we notice that the total association in this table excludes certain effects in the chi-square statistic for the original three-way table. The independence model for Table 1 implies that the expected cell frequency is estimated by $N\hat{p}_{cb}\hat{p}_a$, where $\hat{p}_{ba}$ is the estimate of the joint marginal probability of category c of variable C and category b of variable B. Following (5), Pearson's chi-square statistic representing the association between the rows and columns of Table 1 is given by

$$\chi^2_{(CB-1)(A-1)} = \sum_{cb=1}^{CB} \sum_{a=1}^{A} \left(\frac{f_{bca} - N\hat{p}_{cb}\hat{p}_a}{\sqrt{N\hat{p}_{cb}\hat{p}_a}} \right)^2 . \tag{11}$$

This is obviously different from (8), which further assumes $\hat{p}_{cb} = \hat{p}_c\hat{p}_b$.

How can we account for the difference? As noted toward the end of the previous section, $\chi^2_{CBA-C-B-A+2}$ reflects the joint effects of the AB, AC, BC, and ABC interactions, and thus it may be written as χ^2(AB,AC,BC,ABC). The $\chi^2_{(CB-1)(A-1)}$, on the other hand, reflects the joint effects of the AB, AC, and ABC interactions (i.e., $\chi^2_{(CB-1)(A-1)} = \chi^2$(AB,AC,ABC)) with the BC interaction effect excluded as the marginal effect of the rows of the table. The difference then must be due to the BC interaction effect. More specifically, we call this effect the BC interaction eliminating the joint effects of the AB, AC, and ABC interactions because it represents the portion of the AB,AC,BC,ABC effects left unaccounted for by AB,AC,ABC. This effect is denoted by BC|AB,AC,ABC, where the variables listed on the right of "|" indicate those eliminated from the effect listed on the left. The size of this effect is found by the difference between the two chi-squares, i.e.,

$$\chi^2(\text{BC}|\text{AB,AC,ABC}) = \chi^2(\text{AB,AC,BC,ABC}) - \chi^2(\text{AB,AC,ABC}). \tag{12}$$

An equivalent way of looking at the above equation is that AB,AC,BC,ABC is decomposed into the sum of the effects of AB,AC,ABC and BC|AB,AC,ABC, that is,

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AB,AC,ABC}) + \chi^2(\text{AB}|\text{AC,BC,ABC}). \tag{13}$$

For Table 1, we find $\chi^2_3(\text{AB,AC,ABC}) = 86.99$, so that $\chi^2_1(\text{BC}|\text{AB,AC,ABC}) = 131.99 - 86.99 = 45.00$.

If $\chi^2(\text{BC}|\text{AB,AC,ABC})$ has more than 1 df, it may be further decomposed into the sum of the effects each with 1 df. In the present case, it has only 1 df, so that no further decompositions are applicable. The $\chi^2(\text{AB,AC,ABC})$, on the other hand, has 3 df, which invites further decompositions. There are a number of (in fact, infinitely many) possible decompositions. For example, we may use the Helmert type of contrasts, as before, to decompose this chi-square. However, then each component χ^2 may be empirically less meaningful. We therefore focus on the decompositions that reflect the factorial structure among the rows of Table 1. This means that we are decomposing $\chi^2(\text{AB,AC,ABC})$ into separate effects of AB, AC, and ABC interactions. The problem is that these effects are usually not orthogonal to each other, and consequently must be orthogonalized to derive additive decompositions of the chi-square. As has been alluded to earlier, the order in which they are taken into account will have a crucial effect in this orthogonalization process. There are six possible ways of ordering three effects. We may, however, cut down this number by considering only those orderings in which lower-order interactions are always considered prior to higher-order interactions. We are then left with only two possibilities. One is in which AB is considered first, then AC, and then ABC, and the other is in which AC is considered first, then AB, and then ABC.

When we add a new effect, we only add its unique effect. For example, when we add AC in the first situation described above, we add only the portion of the AC not already explained by AB. This effect, called AC eliminating AB, is orthogonal to AB, and is denoted as AC|AB. The effect of AB considered first, on the other hand, ignores all other effects (AC and ABC), and is simply written as AB. The ABC effect considered last eliminates both AB and AC, and is written as ABC|AB,AC. In general, the effect taken into account first ignores all other effects, the effect considered last eliminates all other effects, and the effect taken into account in-between eliminates all the effects considered earlier, but ignores all the effects considered later. How to calculate the chi-square for these effects will be described shortly.

The two possible orderings of AB, AC, and ABC suggested above give rise to two orthogonal decompositions of the joint effects of AB, AC, and ABC. Symbolically, this is written as

$$\chi^2(\text{AB,AC,ABC}) = \chi^2(\text{AB}) + \chi^2(\text{AC}|\text{AB}) + \chi^2(\text{ABC}|\text{AB,AC}) \tag{14}$$

$$= \chi^2(\text{AC}) + \chi^2(\text{AB}|\text{AC}) + \chi^2(\text{ABC}|\text{AB,AC}). \tag{15}$$

Combining (13) and (14), we obtain the first decomposition of AB,AC,BC,ABC.

Decomposition (i):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AB}) + \chi^2(\text{AC}|\text{AB}) + \chi^2(\text{ABC}|\text{AB,AC}) + \chi^2(\text{BC}|\text{AB,AC,ABC}). \tag{16}$$

Combining (13) and (15), we obtain the second decomposition of AB,AC,BC,ABC.

Decomposition (ii):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AC}) + \chi^2(\text{AB}|\text{AC}) + \chi^2(\text{ABC}|\text{AB,AC}) + \chi^2(\text{BC}|\text{AB,AC,ABC}). \quad (17)$$

The $\chi^2(\text{AB})$, $\chi^2(\text{AC}|\text{AB})$, and $\chi^2(\text{ABC}|\text{AB,AC})$ are calculated as follows. We first set up contrast vectors,

$$\mathbf{t}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{t}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \text{ and } \mathbf{t}_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}. \quad (18)$$

The $\mathbf{t}_1$ represents the main effect of B among the rows of Table 1. When it is used as a linear constraint on the rows, it captures the portion of the association between the rows and columns that can be explained by the main effect of B, which is called the AB interaction effect. Similarly, $\mathbf{t}_2$ captures the AC interaction effect, and $\mathbf{t}_3$ captures the ABC interaction effect. Note that these contrast vectors assume that there are only two categories in all three variables. We will need more than one contrast to represent each of these effects if there are more than two levels in some of the variables. For example, if $B = 3$, $\mathbf{t}_1$ will be a matrix like

$$\mathbf{t}_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \\ 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{bmatrix}. \quad (19)$$

Note also that if we want to decompose the effects of AB,AC,ABC differently, for example, if AB,AC,ABC is decomposed into AB within C_1, AB within C_2, and AC, $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ would be:

$$\mathbf{t}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{t}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \mathbf{t}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}. \quad (20)$$

The following computations use $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ defined in (18). The χ^2 due to the AB interaction ignoring all other effects (AC and ABC) is calculated by first defining

$$\mathbf{H} = \sqrt{N}\mathbf{P}_{Q_{1/K}t_1}\mathbf{K}^{-1}\mathbf{F}\mathbf{L}^{-1}, \quad (21)$$

where

$$\mathbf{P}_{Q_{1/K}t_1/K} = \mathbf{Q}_{1/K}\mathbf{t}_1(\mathbf{t}_1'\mathbf{Q}_{1/K}'\mathbf{K}\mathbf{t}_1)^{-1}\mathbf{t}_1'\mathbf{Q}_{1/K}'\mathbf{K} \quad (22)$$

is the projector onto $\mathrm{Sp}(\mathbf{Q}_{1/K}\mathbf{t}_1)$ (the space spanned by $\mathbf{Q}_{1/K}\mathbf{t}_1$) along $\mathrm{Ker}(\mathbf{t}_1'\mathbf{Q}_{1/K}'\mathbf{K})$ (the space spanned by all vectors $\mathbf{y}$ such that $\mathbf{y}'\mathbf{Q}_{1/K}\mathbf{t}_1 = 0$). Recall that N is the total sample size, $\mathbf{K}$ and $\mathbf{L}$ are diagonal matrices of row and column totals of $\mathbf{F}$, respectively, and $\mathbf{Q}_{1/K} = \mathbf{I} - \mathbf{1}\mathbf{1}'\mathbf{K}/N$, where $\mathbf{1}$ is the CB-element vector of ones. Note that $\mathbf{Q}_{1/K}'\mathbf{K} = \mathbf{Q}_{1/K}'\mathbf{K}\mathbf{Q}_{1/K}$. We then calculate

$$\chi^2(\mathbf{H}) = \mathrm{SS}(\mathbf{H})_{K,L}. \tag{23}$$

This value turns out to be 24.10(1) for the data in Table 1 (the value in parentheses indicates the df). The $\chi^2(\mathbf{H})$ is equal to the chi-square representing the total association in the marginal two-way table obtained by collapsing the three-way table across the levels of C.

The $\chi^2(\mathrm{AC}|\mathrm{AB})$ (the AC interaction eliminating AB, but ignoring ABC) is calculated as follows: Let $\mathbf{T}_1 = [\mathbf{1}, \mathbf{t}_1]$, and define $\mathbf{Q}_{T_1/K}$ similarly to $\mathbf{Q}_{1/K}$ above, that is,

$$\mathbf{Q}_{T_1/K} = \mathbf{I} - \mathbf{T}_1(\mathbf{T}_1'\mathbf{K}\mathbf{T}_1)^{-1}\mathbf{T}_1'\mathbf{K}. \tag{24}$$

Then, define

$$\mathbf{P}_{Q_{T_1/K}\mathbf{t}_2/K} = \mathbf{Q}_{T_1/K}\mathbf{t}_2(\mathbf{t}_2'\mathbf{Q}_{T_1/K}'\mathbf{K}\mathbf{t}_2)^{-1}\mathbf{t}_2'\mathbf{Q}_{T_1/K}'\mathbf{K}, \tag{25}$$

and

$$\mathbf{E} = \sqrt{N}\mathbf{P}_{Q_{T_1/K}\mathbf{t}_2/K}\mathbf{K}^{-1}\mathbf{F}\mathbf{L}^{-1}. \tag{26}$$

Again, note that $\mathbf{Q}_{T_1/K}'\mathbf{K} = \mathbf{Q}_{T_1/K}'\mathbf{K}\mathbf{Q}_{T_1/K}$, and that $\mathbf{P}_{Q_{T_1/K}\mathbf{t}_2/K}$ is the projector onto $\mathrm{Sp}(\mathbf{Q}_{T_1/K}\mathbf{t}_2)$ along $\mathrm{Ker}(\mathbf{t}_2'\mathbf{Q}_{T_1/K}'\mathbf{K})$. Finally,

$$\chi^2(\mathbf{E}) = \mathrm{SS}(\mathbf{E})_{K,L}. \tag{27}$$

This value is found to be 55.83(1) for the data in Table 1. (There are other ways to calculate this quantity. See (37) and (38) in Takane and Jung (2009b).)

The $\chi^2(\mathrm{ABC}|\mathrm{AB,AC})$ (the ABC interaction eliminating both AB and AC) is calculated as follows: First let $\mathbf{T}_{12} = [\mathbf{1}, \mathbf{t}_1, \mathbf{t}_2]$, and define $\mathbf{Q}_{T_{12}/K} = \mathbf{I} - \mathbf{T}_{12}(\mathbf{T}_{12}'\mathbf{K}\mathbf{T}_{12})^{-1} \times \mathbf{T}_{12}'\mathbf{K}$. Then, define

$$\mathbf{P}_{Q_{T_{12}/K}\mathbf{t}_3/K} = \mathbf{Q}_{T_{12}/K}\mathbf{t}_3(\mathbf{t}_3'\mathbf{Q}_{T_{12}/K}'\mathbf{K}\mathbf{t}_3)^{-1}\mathbf{t}_3'\mathbf{Q}_{T_{12}/K}'\mathbf{K}, \tag{28}$$

and

$$\mathbf{J} = \sqrt{N}\mathbf{P}_{Q_{T_{12}/K}\mathbf{t}_3/K}\mathbf{K}^{-1}\mathbf{F}\mathbf{L}^{-1}. \tag{29}$$

Note that $\mathbf{Q}_{T_{12}/K}'\mathbf{K} = \mathbf{Q}_{T_{12}/K}'\mathbf{K}\mathbf{Q}_{T_{12}/K}$, and that $\mathbf{P}_{Q_{T_{12}/K}\mathbf{t}_3/K}$ is the projector onto $\mathrm{Sp}(\mathbf{Q}_{T_{12}/K}\mathbf{t}_3)$ along $\mathrm{Ker}(\mathbf{t}_3'\mathbf{Q}_{T_{12}/K}'\mathbf{K})$. Finally,

$$\chi^2(\mathbf{J}) = \mathrm{SS}(\mathbf{J})_{K,L}. \tag{30}$$

This value turns out to be 7.06(1) for the data in Table 1. Takane and Jung (2009b) showed that $\mathbf{J}$ above can also be calculated by

$$\mathbf{J} = \sqrt{N}\mathbf{K}^{-1}\mathbf{t}_3(\mathbf{t}_3'\mathbf{K}^{-1}\mathbf{t}_3)^{-1}\mathbf{t}_3'\mathbf{K}^{-1}\mathbf{F}\mathbf{L}^{-1}, \tag{31}$$

which is somewhat simpler.

It can be easily verified that 24.10(1), 55.83(1), and 7.06(1) add up to 86.99(3) calculated previously. The χ^2(AC) and χ^2(AB|AC) can be calculated similarly to the above. It turns out that the former is 68.66(1), and the latter is 11.27(1). These and 7.06(1) for the ABC interaction again add up to 86.99(3). So there are indeed two alternative decompositions of χ^2(AB,AC,ABC) depending on whether AB or AC is taken into account first. Corresponding to the two decompositions of AB,AC,ABC, there are two decompositions of χ^2(AB,AC,BC,ABC), as stated in (16) and (17).

As remarked earlier, there are two other possible arrangements of a three-way table into two. In Table 1, variables B and C were combined to form rows of the table. We may have also combined A and B, or A and C. In either case, the remaining variable constitutes the columns. Each of these two cases gives rise to two different decompositions of AB,AC,BC,ABC analogous to those given in (16) and (17).

Let us start with the case in which A and B are combined. In this case, (13) will become:

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AC,BC,ABC}) + \chi^2(\text{AB}|\text{AC,BC,ABC}), \tag{32}$$

and (14) and (15) become

$$\chi^2(\text{AC,BC,ABC}) = \chi^2(\text{AC}) + \chi^2(\text{BC}|\text{AC}) + \chi^2(\text{ABC}|\text{AC,BC}) \tag{33}$$
$$= \chi^2(\text{BC}) + \chi^2(\text{AC}|\text{BC}) + \chi^2(\text{ABC}|\text{AC,BC}). \tag{34}$$

The terms in these decompositions can be calculated similarly to the above. We find $\chi^2(\text{AC,BC,ABC}) = 93.73(3)$ (the df in parentheses), so that $\chi^2(\text{AB}|\text{AC,BC,ABC}) = 38.26(1) = 131.99(4) - 93.73(3) = \chi^2(\text{AB,AC,BC,ABC}) - \chi^2(\text{AC,BC,ABC})$. We also find $\chi^2(\text{AC}) = 68.66(1)$ (this is the same χ^2(AC) calculated previously), $\chi^2(\text{BC}|\text{AC}) = 18.44$, and $\chi^2(\text{ABC}|\text{AC,BC}) = 6.63$, so that 68.66(1) + 18.44(1) + 6.63(1) = 93.77(3) = χ^2(AC,BC,ABC), verifying (33). We also find $\chi^2(\text{BC}) = 31.80(1)$, and $\chi^2(\text{AC}|\text{BC}) = 55.30(1)$, so that 31.80(1) + 55.30(1) + 6.63(1) = 93.77(3), verifying (34). Combining (32) with (33) and (34), we respectively obtain

Decomposition (iii):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AC}) + \chi^2(\text{BC}|\text{AC}) + \chi^2(\text{ABC}|\text{AC,BC}) + \chi^2(\text{AB}|\text{AC,BC,ABC})), \tag{35}$$

and Decomposition (iv):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{BC}) + \chi^2(\text{AC}|\text{BC}) + \chi^2(\text{ABC}|\text{AC,BC}) + \chi^2(\text{AB}|\text{AC,BC,ABC}). \tag{36}$$

Similarly, when A and C are combined, we obtain

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AB,BC,ABC}) + \chi^2(\text{AC}|\text{AB,BC,ABC}), \quad (37)$$

and

$$\chi^2(\text{AB,BC,ABC}) = \chi^2(\text{AB}) + \chi^2(\text{BC}|\text{AB}) + \chi^2(\text{ABC}|\text{AB,BC}) \quad (38)$$
$$= \chi^2(\text{BC}) + \chi^2(\text{AB}|\text{BC}) + \chi^2(\text{ABC}|\text{AB,BC}). \quad (39)$$

For the illustrative data we have been using, we find $\chi^2(\text{AB,BC,ABC}) = 49.96(3)$, so that $\chi^2(\text{AC}|\text{AB,BC,ABC}) = 82.03(1) = 131.99(4) - 49.96(3) = \chi^2(\text{AB,AC,BC,}$ $\text{ABC}) - \chi^2(\text{AB,BC,ABC})$. We also find $\chi^2(\text{AB}) = 24.10(1)$ (this is the same $\chi^2(\text{AB})$ calculated previously), $\chi^2(\text{BC}|\text{AB}) = 19.18(1)$, and $\chi^2(\text{ABC}|\text{AB,BC}) = 6.35(1)$, so that 24.10(1) + 19.51(1) + 6.35(1) = 49.96(3) = $\chi^2(\text{AB,BC,ABC})$, verifying (38). We also find $\chi^2(\text{BC}) = 31.80(1)$ (this is the same $\chi^2(\text{BC})$ calculated before), and $\chi^2(\text{AB}|\text{BC}) = 1181(1)$, so that 31.80(1) + 11.81(1) + 6.35(1) = 49.96(3), verifying (39). Combining (37) with (38) and (39), we obtain the fifth and sixth decompositions of $\chi^2(\text{AB,AC,BC,ABC})$.

Decomposition (v):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{AB})$$
$$+ \chi^2(\text{BC}|\text{AB}) + \chi^2(\text{ABC}|\text{AB,BC}) + \chi^2(\text{AC}|\text{AB,BC,ABC}), \quad (40)$$

and Decomposistion (vi):

$$\chi^2(\text{AB,AC,BC,ABC}) = \chi^2(\text{BC})$$
$$+ \chi^2(\text{AB}|\text{BC}) + \chi^2(\text{ABC}|\text{AB,BC}) + \chi^2(\text{AC}|\text{AB,BC,ABC}). \quad (41)$$

Altogether we obtain (at least) six fundamental decompositions of Pearson's chi-square statistic for a three-way contingency table. Lancaster (1951) defined $\chi^2(\text{ABC}|\text{AB,AC,BC})$ by

$$\chi^2(\text{ABC}|\text{AB,AC,BC})$$
$$= \chi^2(\text{AB,AC,BC,ABC}) - \chi^2(\text{AB}) - \chi^2(\text{AC}) - \chi^2(\text{BC}). \quad (42)$$

Then, $\chi^2(\text{ABC}|\text{AB,AC,BC})$ is unique. However, as has been noted earlier, $\chi^2(\text{AB})$, $\chi^2(\text{AC})$, and $\chi^2(\text{BC})$ are usually not independent from each other, and consequently, $\chi^2(\text{ABC}|\text{AB,AC,BC})$ may not follow an asymptotic chi-square distribution (Placket, 1962).

4 Analogous Decompositions of the Log LR Statistic

In this section, we discuss decompositions of the log LR chi-square statistic analogous to Decompositions (i) through (vi). The log LR statistic for a three-way contingency table is defined as

$$LR_{CBA-C-B-A+2} = -2\sum_{c=1}^{C}\sum_{b=1}^{B}\sum_{a=1}^{A} f_{cba} \log \frac{f_{cba}}{\hat{p}_c\hat{p}_b\hat{p}_a}. \tag{43}$$

This statistic, like Pearson's chi-square statistic, represents the departure from the three-way independence model, and reflects the joint effects of AB, AC, BC, and ABC (i.e., AB,AC,BC,ABC). Similarly to the case of Pearson's chi-square statistic, these four effects are not mutually independent, and consequently their joint effects cannot be obtained by their sum. We find the effect of AB,AC,BC,ABC to be 120.59 for the data given in Table 1, using the above formula.

In this section, we first take a heuristic approach to get an intuitive idea about proper decompositions. We then present a theory due to Cheng et al. (2006) to back up our intuition. Our heuristic approach begins with analyzing the data in Table 1 by log linear models. In log linear analysis, no reduction of a three-way table into a two-way format is necessary in contrast to Pearson's statistic. The three variables are treated completely symmetrically.

We first ran the "Hiloglinear" procedure in SPSS. We obtained the three-way interaction effect of $LR(\text{ABC}|\text{AB,AC,BC}) = 6.82(1)$. We also obtained the joint effects of three two-way interactions of $LR(\text{AB,AC,BC}) = 113.77(3)$. The three individual two-way interaction effects (these were the two-way interactions eliminating all other two-way interactions) were found to be $LR(\text{AB}|\text{AC,BC}) = 12.22(1)$, $LR(\text{AC}|\text{AB,BC}) = 57.54(1)$, and $LR(\text{BC}|\text{AB,AC}) = 20.00(1)$. These effects do not add up to $LR(\text{AB,AC,BC})$, as $12.22+57.54+20.00 = 89.76 \neq 113.77$. Note that in log linear analysis, only the independence or conditional independence models can be fitted non-iteratively, which implies that none of the above quantities can be calculated in closed form.

In order to find proper constituents of the joint two-way interaction effects, we had to run another log liner analysis procedure in SPSS called "Loglinear", which provided individual two-way interaction effects ignoring the other two-way interaction effects. They were found to be $LR(\text{AB}) = 24.23(1)$, $LR(\text{AC}) = 69.54(1)$, and $LR(\text{BC}) = 32.04(1)$. These quantities can be calculated in closed form. They do not add up to $LR(\text{AB,AC,BC})$, either, as $24.23+69.54+32.02 = 125.79 \neq 113.77$. However, we find

$$\begin{aligned} &LR(\text{AB}) + LR(\text{AC}) + LR(\text{BC}|\text{AB,AC}) \\ &\quad = 24.23+69.54+20.00 = 113.77 = LR(\text{AB,AC,BC}), \end{aligned} \tag{44}$$

$$\begin{aligned} &LR(\text{AC}) + LR(\text{BC}) + LR(\text{AB}|\text{AC,BC}) \\ &\quad = 69.54+32.02+12.22 = 113.77 = LR(\text{AB,AC,BC}), \end{aligned} \tag{45}$$

and

$$\begin{aligned} &LR(\text{AB}) + LR(\text{BC}) + LR(\text{AC}|\text{AB,BC}) \\ &\quad = 32.02 + 24.23 + 57.54 = 113.77 = LR(\text{AB,AC,BC}). \end{aligned} \tag{46}$$

That is, we cannot add the three two-way interactions all ignoring the other two to obtain their joint effects. One of the three must be the two-way interaction eliminating the other two.

Adding one more term, $LR(\text{ABC}|\text{AB,AC,BC}) = 6.82$, to the above identities, we obtain three alternative decompositions of

$$\begin{aligned} &LR(\text{AB,AC,BC,ABC}) \\ &\quad = LR(\text{AB,AC,BC}) + LR(\text{ABC}|\text{AB,AC,BC}) = 113.77 + 6.82 = 120.59, \end{aligned} \tag{47}$$

namely, Decomposition (a):

$$\begin{aligned} &LR(\text{AB,AC,BC,ABC}) \\ &\quad = LR(\text{AB}) + LR(\text{AC}) + LR(\text{BC}|\text{AB,AC}) + LR(\text{ABC}|\text{AB,AC,BC}), \end{aligned} \tag{48}$$

Decomposition (b):

$$\begin{aligned} &LR(\text{AB,AC,BC,ABC}) \\ &\quad = LR(\text{AC}) + LR(\text{BC}) + LR(\text{AB}|\text{AC,BC}) + LR(\text{ABC}|\text{AB,AC,BC}), \end{aligned} \tag{49}$$

and Decomposition (c):

$$\begin{aligned} &LR(\text{AB,AC,BC,ABC}) \\ &\quad = LR(\text{AB}) + LR(\text{BC}) + LR(\text{AC}|\text{AB,BC}) + LR(\text{ABC}|\text{AB,AC,BC}). \end{aligned} \tag{50}$$

It is obvious that Decomposition (a) “corresponds” with Decompositions (i) and (ii), (b) with (iii) and (iv), and (c) with (v) and (vi) for Pearson’s chi-square statistic.

These three decompositions are consistent with Cheng et al.’s (2006) decompositions derived rigorously through information identities. Cheng et al., however, arrived at these decompositions via a somewhat different route. They first derived the sum of the last two terms in each of the above three decompositions. For example, they first obtained $LR^*(\text{BC}|\text{A}) \equiv LR(\text{BC}|\text{AB,AC}) + LR(\text{ABC}|\text{AB,AC,BC})$ for Decomposition (a). This quantity can be calculated in closed form using the information identities, whereas neither of the two terms on the right-hand side can. Cheng et al. (2006) called the quantity on the left-hand side, i.e., $LR^*(\text{BC}|\text{A})$, the conditional dependence between B and C across levels of A (or the simple two-way interaction between B and C across levels of A). They then split this into two additive terms on the right-hand side, $LR(\text{BC}|\text{AB,AC})$ ($LR(\text{BC}||\text{A})$ in their notation) and $LR(\text{ABC}|\text{AB,AC,BC})$, by way of log linear analysis. The first term was interpreted as the uniform part, and the second as the non-uniform part, of the conditional dependence between B and C across levels of A (or equivalently, the homogeneous and heterogenous aspects of the simple two-way interactions between B and C across levels of A). In our framework, the former is interpreted as the BC interaction eliminating the effects of AB and AC. It is interesting to find that this effect is equivalent to the uniform part of the simple two-way interactions. The latter is nothing but the three-way interaction among A, B, and C eliminating the joint effects of AB, AC, and BC. Similar remarks can be made for Decompositions (b) and (c).

5 Discussion

As has been observed in the previous section, the order in which two two-way interactions ignoring the other two are accounted for make no difference in the log LR statistic, while it does in Pearson's chi-square statistic. In fact, we have

$$LR(\mathrm{AB}) = LR(\mathrm{AB}|\mathrm{AC}) = LR(\mathrm{AB}|\mathrm{BC}) \neq LR(\mathrm{AB}|\mathrm{AC,BC}), \tag{51}$$

$$LR(\mathrm{AC}) = LR(\mathrm{AC}|\mathrm{AB}) = LR(\mathrm{AC}|\mathrm{BC}) \neq LR(\mathrm{AC}|\mathrm{AB,BC}), \tag{52}$$

and

$$LR(\mathrm{BC}) = LR(\mathrm{BC}|\mathrm{AB}) = LR(\mathrm{BC}|\mathrm{AC}) \neq LR(\mathrm{BC}|\mathrm{AB,AC}), \tag{53}$$

while the four versions of the AB interaction effects for Pearson's chi-square, $\chi^2(\mathrm{AB})$, $\chi^2(\mathrm{AB}|\mathrm{AC})$, $\chi^2(\mathrm{AB}|\mathrm{BC})$, and $\chi^2(\mathrm{AB}|\mathrm{AC,BC,ABC})$), are all distinct, and so are the four versions of AC and BC. Also, there is a single unique three-way interaction in the decompositions the log LR statistic ($LR(\mathrm{ABC}|\mathrm{AB,AC,BC})$), while there are three distinct versions of the three-way interaction effect for Pearson's chi-square, ($\chi^2(\mathrm{ABC}|\mathrm{AB,AC})$, $\chi^2(\mathrm{ABC}|\mathrm{AB,BC})$, and $\chi^2(\mathrm{ABC}|\mathrm{AC,BC})$). These differences stem from the fact that there is no way to evaluate $\chi^2(\mathrm{AB,AC,BC})$ in the latter, which in turn is more fundamentally caused by the fact that a three-way table must always be reduced to a two-way table to obtain the decompositions of Pearson's statistic. This prevents us from obtaining quantities such as $\chi^2(\mathrm{AB}|\mathrm{AC,BC})$, $\chi^2(\mathrm{AC}|\mathrm{AB,BC})$, $\chi^2(\mathrm{BC}|\mathrm{AB,AC})$, and $\chi^2(\mathrm{ABC}|\mathrm{AB,AC,BC})$.

Having fewer distinct terms in the decompositions of the log LR statistic may be a point in its favor over Pearson's statistic. However, there are still three alternative decompositions for the former. A choice among them may not be straightforward. This is particularly so because log linear analysis treats all variables symmetrically, yet the resultant decompositions are not symmetric.

The fact that Pearson's chi-square statistic has six alternative decompositions is surely a bit unwieldy. However, if one layout of a three-way table into a two-way format is in some sense more natural than the other two, this number is reduced to two, which differ from each other only in a minor way. Such is the case when analysis of contingency tables is conducted in predictive settings, and yet a symmetric measure of association such as Pearson's statistic is in order. In CCA, for example, one of the variables is typically taken as the criterion variable, while the others are used as predictor variables. There are also other considerations to be taken into account. Pearson's chi-square statistic is known to approach a chi-square distribution more quickly than the log LR statistic. It is also the case that all the terms in the decompositions of Pearson's chi-square can be calculated in closed form, whereas some of the terms in the log LR statistic must be obtained iteratively.

It may also be pointed out that there seems to be a "cultural" difference between log linear analysis (based on the log LR statistic) and CA (based on Pearson's statistic). The former tends to focus on residual effects (eliminating effects). If we fit the AB interaction effect, for example, we get the deviation chi-square of this model from the saturated model. It represents the effects of all variables not included in the model eliminating AB. To obtain the effect of AB ignoring all other variables we have to subtract this value from the independence chi-square representing the deviation of the independence model from the saturated model. To obtain the AB interaction effect eliminating some other effects, we have to fit the model with these "some other effects" only, and the model with the additional effect of AB, and take the difference in chi-square values between the two models. In CA, on the other hand, the chi-square value due to AB ignoring other effects is obtained directly by the difference between the fitted model and the independence model. We need an extra step to obtain a residual effect representing the effect of a variable not included in the fitted model, which amounts to taking the difference in chi-square between the saturated model (which is equal to Pearson's chi-square for the total association) and the fitted model. A notable exception is van der Heijden and Meijerink (1989), who attempted to analyze residual effects in constrained CA. In the present authors' view, both analyses (analyses of the fitted models and the residual effects) are equally important, as has been emphasized by Takane and Jung (2009a).

Cheng et al. (2007) attempts to extend their approach to higher-order designs, thereby generalizing their decompositions of the log LR statistic. Presumably, similar things could be done for Pearson's chi-square statistic.

Acknowledgements The work reported in this paper has been supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) grant 36952, and by the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant 10630 to the first author.

References

1. Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North Holland.
2. Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, **22**, 327–351.
3. Bartlett, M. S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society, Series B*, **2**, 248–252.
4. Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.
5. Böckenholt, U., & Böckenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, **55**, 633–639.

6. Breiman, L., & Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society, Series B*, **57**, 3–54.
7. Cheng, P. E., Liou, J. W., Liou, M., & Aston, J. A. D. (2006). Data information in contingency tables: A fallacy of hierarchical loglinear models. *Journal of Data Science*, **4**, 387–398.
8. Cheng, P. E., Liou, J. W., Liou, M., & Aston, J. A. D. (2007). Linear information models: An introduction. *Journal of Data Science*, **5**, 297–313.
9. Fisher, R. A. (1936). The use of multiple measurement in taxonomic problem. *Annals of Eugenics*, **7**, 179-188.
10. Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
11. Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, **28**, 417–441.
12. Irwin, J. O. (1949). A note on the subdivision of χ^2 into components. *Biometrika*, **36**, 130–134.
13. Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, **5**, 248–264.
14. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
15. Lancaster, H. O. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika*, **36**, 117–129.
16. Lancaster, H. O. (1951). Complex contingency tables treated by the partition of chi-square. *Journal of the Royal Statistical Society, Series B*, **13**, 242–249.
17. Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
18. Light, R. J., & Margolin, B. H. (1971). An analysis of variance for categorical data. *Journal of the American Statistical Association*, **66**, 534–544.
19. Mood, A. M. (1950). *Introduction to the theory of statistics* New York: McGraw-Hill.
20. Pearson, K. (1900). On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine*, **50**, 157–172.
21. Placket, R. L. (1962). A note on interactions in contingency tables. *Journal of the Royal Statistical Society, Series B*, **24**, 162–166.
22. Reinsell, G. C., & Velue, R. P. (1998). *Multivariate reduced-rank regression*. New York: Springer.
23. Snedecor, G. W. (1958). Chi-squares of Bartlett, Mood and Lancaster in a 2^3 contingency table. *Biometrics*, **14**, 560–562.
24. Takane, Y., & Jung, S. (2009a). Regularized nonsymmetric correspondence analysis. *Computational Statistics and Data Analysis*, **53**, 3159–3170.
25. Takane, Y., & Jung, S. (2009b). Tests of ignoring and eliminating in nonsymmetric correspondence analysis. *Advances in Data Analysis and Classification*, **3**, 315–340.
26. ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.
27. van der Heijden, P. G. M., & Meijerink, F. (1989). Generalized correspondence analysis of multi-way contingency tables. In R. Coppi, & S. Bolasco (eds.), *Multiway data analysis*, (pp. 185–202). Amsterdam: Elsevier.