In: M. Greenacre & J.Blasius (Eds.),
Multiple correspondence analysis and related methods.
Boca Raton, FL: Chapman & Hall/CRC, 2006, 455-486.

Additive and Multiplicative Models for Three-Way Contingency Tables: Darroch (1974) Revisited

Pieter M. Kroonenberg & Carolyn J. Anderson

Last Revised December 22, 2004

1 Introduction

In an only occasionally referenced paper, Darroch (1974) discussed the relative merits of additive and multiplicative definitions of interaction for higher-order contingency tables. In particular, he compared the following aspects: partitioning properties, closeness to independence, conditional independence as a special case, distributional equivalence, subtable invariance, and constraints on the marginal probabilities. On the basis of this investigation, he believed that multiplicative modelling is preferable over additive modelling, "but not by so wide a margin as the difference in the attention that these two definitions have received in the literature" (p. 213). One important aspect of modelling contingency tables did not figure in this comparison: interpretability.

The potential systematic relationships in multivariate categorical data becomes progressively more complex as the number of variables and/or the number categories per variable increase. In turn, interpretation becomes increasingly difficult. We consider techniques for data that can be logically formatted as three-way contingency tables. This does not limit us to three variables but rather it limits us to, at most, three types or modes of variables.

The focus in this chapter lies with the interpretation of the dependence present in three-way tables and how insight can be gained into complex patterns of different types of dependence. Since the major aim in most empirical sciences is to apply (statistical) models to data and to obtain a deeper insight into the subject matter, we consider it worthwhile to take up Darroch's comparison and extend his set of criteria by considering the interpretational possibilities (and impossibilities) of multiplicative and additive modelling of contingency tables. The investigation will primarily take place at the empirical level guided by a particular data set that consists of four variables but is best analysed as a three-way table.

Similarities between additive and multiplicative modelling techniques for two-way tables have been discussed by Escoufier (1982), Goodman (1985, 1996), and Van der Heijden, Mooijaart and Takane (1994). Limited discussions of the three-way case can be found in Van der Heijden and Worsley (1988) and Green (1989). Neither of the latter two consider three-way correspondence analysis (three-way CA) nor multiple correspondence analysis (MCA).

The major empirical comparisons made in this chapter are between three-way CA, which uses an additive definition of interaction (Carlier & Kroonenberg, 1996, 1998), extensions of Goodman's multidimensional row-column association model (Goodman, 1979, 1985; Clogg & Shihadeh, 1994; Anderson, 2002; Anderson & Vermunt, 2000), which use a multiplicative definition of interaction, and categorical PCA which is related to MCA, the major method discussed in this book. These methods will be compared empirically in terms of how and to what extent they succeed in bringing out the structure in a data set.

2 Data and design issues

We hope to show that although our attention is focused on three-way tables, the methods discussed are more general than they may appear at first glance. After describing a data set from Wickens and Olzak (1989) that guides our empirical comparison of methods for three-way tables, we briefly discuss general design issues that help to identify potential data analytic problems to which the methods presented in this chapter can be applied.

2.1 The Wickens & Olzak data

Wickens and Olzak (1989) report the data from a single subject on a concurrent signal detection task (Table 1). In the experiment, a signal abruptly appeared on a computer screen and disappeared again after 100 milliseconds. The signal consisted of either one of two sine curves or both together. One curve had more cycles and is referred to as the high frequency signal ($\mathcal{H}$), while the other with fewer cycles is referred to as the low frequency signal ($\mathcal{L}$), however in the analyses they are mostly combined into a single interactively coded factor ($\mathcal{J}$) with four categories. On each trial of the experiment, the subject was presented with either both signals, only the high frequency signal, only the low frequency signal, or neither and each of these was repeated 350 times. Thus, the independent variables consisted of a fully-crossed 2×2 design, with a total of $4 \times 350 = 1400$ trials. However, one observation went astray and only 1399 trials are present in the data used here.

The subject's task on each trial was to rate his confidence on a scale from 1 to 6 regarding the presence of each of the two possible stimuli. The lower the rating, the more confident the subject was that the signal was absent, and the higher the rating, the more confident the subject was that the signal was present. Thus, the response variables were the two confidence ratings (H) and (L) .

When a signal was present, the subject was expected to express more confidence that it was presented than when the signal was absent. This is generally confirmed by the data (Table 1), but there appear to be interactions as well. Wickens and Olzak's (1989) analyses and those by Anderson (2002), both of which use a multiplicative definition of interaction, confirm the presence of a three-way interaction between the presence of the two signals and the confidence rating of the low signal. However, Anderson (2002), who used a more focused test, also detected a three-way interaction between the two signals and the rating of the high signal. Since there are interactions in the data involving the two stimuli and one or

Table 1: Concurrent Detection Data (Wickens-Olzak, 1989)

			High Signal absent ($-\mathcal{H}$) Confidence in High Signal (H)						High Signal present ($\mathcal{H}$) Confidence High Signal (H)					
			1	2	3	4	5	6	1	2	3	4	5	6
		1	44	4	9	7	6	7	7	4	5	5	14	69
Low		2	13	30	20	8	14	7	5	7	13	15	38	37
Signal	Confidence in	3	9	23	17	17	3	0	6	7	8	10	10	15
absent	Low Signal (L)	4	16	17	10	20	2	2	4	12	5	13	6	14
($-\mathcal{L}$)		5	5	4	9	10	4	0	2	3	1	1	3	5
		6	3	3	0	1	4	1	0	0	1	1	1	3
		1	8	2	2	1	0	4	4	1	2	0	4	37
Low		2	5	5	5	5	5	3	0	4	0	1	8	25
Signal	Confidence in	3	8	10	7	4	1	1	1	3	3	7	8	15
present	Low Signal (L)	4	12	17	15	13	2	2	4	4	8	17	12	21
($\mathcal{L}$)		5	12	17	19	18	10	4	3	12	8	11	20	20
		6	31	29	25	24	12	12	11	8	12	11	12	33

Note. An (i, j)-entry in the table indicates the number of times the subject expressed confidence level i in the presence/absence of the low frequency signal and his confidence level j in the presence/absence of the high frequency signal.

the other response, in this paper we format the data as a stimulus condition or joint signal ($\mathcal{J}$) by high rating (H) by low rating (L) cross-classification, in particular the data form a $I \times J \times K$-three-way table with $I = 4$, $J = 6$, and $K = 6$. The aim of the present analyses is to highlight and describe the nature of the interactions in the data.

2.2 Design issues

Several designs yield data that are naturally organised as entries in three-way contingency tables. Two aspects of designs are relevant in this context (i) the role of the variables and (ii) the mutual exclusiveness of the entries in the table.

With respect to the first design aspect, in a study variables typically take on the role of response or factor (explanatory variable). In a three-way contingency table, there can be three response variables or *responses* in which case we speak of a *response design*. When there two responses and one factor, or when there is one response variable and two factors, we speak of *response-factor designs*. In the Wickens and Olzak (1989) experiment, the two types of ratings are the two responses and the joint signal is the factor. Especially for multiplicative modelling, it is important to consider this response-factor distinction, while for additive modelling it is primarily relevant in interpretation.

With respect to the second design aspect, in most statistical approaches to analysing contingency tables, it is assumed that observational units are independent and each observation's "score" is a triplet (i, j, k). The design is fully-crossed if each observation falls into one and only one cell of the table. Most log-multiplicative models require fully-crossed designs, while additive modelling via CA is indifferent to the design as it does not explicitly use stochastic assumptions. In the case of the Wickens and Olzak data (1989), even though all of the observations come from a single subject, the responses over trials are considered to be independent as is common in this type of experiment.

3 Multiplicative and additive modelling

Since our methods for three-way tables are characterised as being either an additive or multiplicative model, we review the defining features of the models under study in this chapter. The exposition in this section leans heavily on that by Darroch (1974).

Let π_{ijk} be the probability and p_{ijk} the proportion of observations that fall into categories i, j, and k of variables A, B, and C, respectively, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$, and $\sum_{ijk} \pi_{ijk} = \sum_{ijk} p_{ijk} = 1$. As usual, marginal probabilities and proportions will be indicated by replacing the index over which is summed by a dot. For example, $p_{ij.}$ is the marginal proportion for the i-th category of A and the j-th category of B. All summations will run from 1 to the capital letter of the index, i.e. k will run from 1 to K.

In the *multiplicative definition*, the absence of three-way interaction is defined as

$$H_m : \pi_{ijk} = \xi_{jk}\phi_{ik}\psi_{ij} \tag{1}$$

for all i, j, and k and for some ξ_{jk}, ϕ_{ik}, and ψ_{ij}. One way to express this is that the interaction between two variables does not depend on the values of the third variable (see section 4.1). By taking logarithms of (1), we get models that are additive in the log-scale (Roy and Kastenbaum, 1956).

An equivalent expression for (1) is

$$\frac{\pi_{ijk}\pi_{i'j'k}\pi_{i'jk'}\pi_{i'j'k}}{\pi_{i'j'k'}\pi_{i'jk}\pi_{ij'k}\pi_{ijk'}} = 1 \tag{2}$$

for all $i \neq i'$, $j \neq j'$ and $k \neq k'$ (Darroch, 1974). The expression in (2) is the ratio of odds ratios for two variables given the third variable. Deviations from no three-way interaction are reflected by the extent to which expression (2) differs from the value of one (or equivalently the logarithm of (2) differs from zero).

The *additive definition* of no three-way interaction, introduced by Lancaster (1951), is

$$H_a : \frac{\pi_{ijk}}{\pi_{i..}\pi_{.j.}\pi_{..k}} = \alpha_{jk} + \beta_{ik} + \gamma_{ij} \tag{3}$$

for all i, j, and k and for some α_{jk}, β_{ik}, and γ_{ij}, and this is equivalent to

$$H_a : \Pi_{ijk} = \frac{\pi_{ijk}}{\pi_{i..}\pi_{.j.}\pi_{..k}} - 1 = \left(\frac{\pi_{.jk}}{\pi_{.j.}\pi_{..k}} - 1\right) + \left(\frac{\pi_{i.k}}{\pi_{i..}\pi_{..k}} - 1\right) + \left(\frac{\pi_{ij.}}{\pi_{i..}\pi_{.j.}} - 1\right) \tag{4}$$

(Darroch, 1974, p. 209). Thus according to the additive definition of no three-way independence (i.e. according to equation (4)), the deviation from complete independence between A,

B and C equals the sum of the deviations from (two-way) marginal independence between A and B, A and C, and B and C . For cell (i, j, k) the three-way interaction term which is hypothesised to be zero has the form

$$\frac{\pi_{ijk} - \tilde{\pi}_{ijk}}{\pi_{i..}\pi_{.j.}\pi_{..k}} \tag{5}$$

where

$$\tilde{\pi}_{ijk} = \pi_{i..}\pi_{.j.}\pi_{..k} + (\pi_{ij.}\pi_{..k} - \pi_{i..}\pi_{.j.}\pi_{..k}) + (\pi_{i.k}\pi_{.j.} - \pi_{i..}\pi_{.j.}\pi_{..k}) + (\pi_{.jk}\pi_{i..} - \pi_{i..}\pi_{.j.}\pi_{..k}).$$

or

$$\tilde{\pi}_{ijk} = \pi_{ij.}\pi_{..k} + \pi_{i.k}\pi_{.j.} + \pi_{.jk}\pi_{i..} - 2\pi_{i..}\pi_{.j.}\pi_{..k}. \tag{6}$$

Apart from the distinction between multiplicative and additive modelling, there is another sense in which the word "multiplicative" crops up in discussing models for contingency tables. The former distinction refers to the different ways that interactions are defined. A different question is how the interactions themselves are treated (regardless of the definition of interaction). In both types of modelling, the interactions are decomposed into multiplicative terms via two-way or three-way SVDs to provide a lower-rank representation of the systematic patterns in the interactions. The use of the SVD allows a separation of the interaction into a systematic part and an uninterpretable remainder. In the additive framework, there is only one decomposition for the overall three-way dependence from which the decompositions of the two-way interactions are derived, while in the multiplicative modelling definition, the decompositions are carried out either for each interaction separately, or for a group of interactions jointly, but in the latter case it is not possible to separate out which part belongs to which interaction (unless restrictions are placed on the parameters).

4 Multiplicative models

Log-linear models, extensions of Goodman's $RC(M)$ association model to multi-way tables, and association models with latent variable interpretations all use a multiplicative definition of dependence. Since the last two types of models are special cases of log-linear models, we start with a brief introduction and discussion of log-linear models using the Wickens and Olzak data. We use these models to show how the multiplicative definition of dependence manifests itself. Then we turn to extensions of the $RC(M)$ association model for three-way tables.

4.1 Log-linear models

In this discussion, we initially treat the Wickens and Olzak (1989) data as a four-way cross-classification: presence of a high frequency signal (i.e., "HIGH SIGNAL", $\mathcal{H}$); presence of a low frequency signal (i.e., "low signal", $\mathcal{L}$); confidence rating in the high frequency signal (i.e., "HIGH RATING", H); and confidence rating in the low frequency signal (i.e., "low rating", L). Later in the paper, we switch to a more advantageous format of a three-way table by fully crossing the two factors into a single "Joint Signal", $\mathcal{J}$, as will be justified by the results of preliminary analysis of the data.

Any table that has some form of independence (i.e., complete independence, joint independence or conditional independence) can be expressed as the product of marginal probabilities. For our data, ideally the subject's rating of the high signal should only depend on the presence of the high signal, and the rating of the low signal should only depend on the presence of the low signal; that is, the ratings are *conditionally independent* given the stimuli. This specific hypothesis can be expressed as the product of various marginal probabilities; thus the logarithm is an additive function of the logarithm of the marginal probabilities. Log-linear models are typically parameterised not in terms of logarithm of probabilities, but in such a way as to allow dependence structures that cannot be expressed as a product of marginal probabilities (i.e., some form of non-independence). The parameterisation of our hypothesis regarding the subject's behaviour is

$$\log(\pi_{hljk}) = \lambda + \lambda_h^{\mathcal{H}} + \lambda_l^{\mathcal{L}} + \lambda_j^{H} + \lambda_k^{L} + \lambda_{hj}^{\mathcal{H}H} + \lambda_{lk}^{\mathcal{L}L} + \lambda_{hl}^{\mathcal{H}\mathcal{L}} \tag{7}$$

where λ is a normalisation constant that ensures $\sum_{hljk} \pi_{hljk} = 1$, $\lambda_h^{\mathcal{H}}$, $\lambda_l^{\mathcal{L}}$, λ_j^{H} and λ_k^{L} are marginal effect terms for high signal h, low signal l, high rating j and low rating k, respectively, and $\lambda_{lj}^{\mathcal{H}H}$, $\lambda_{lk}^{\mathcal{L}L}$ and $\lambda_{hl}^{\mathcal{H}\mathcal{L}}$ are bivariate interaction terms. The presence of the marginal effect terms ensures that the fitted one-way marginal probabilities equal the observed margins (e.g., if $\lambda_h^{\mathcal{H}}$ is in the model, then $\pi_{h\cdots} = p_{h\cdots}$) and the presence of the interaction terms ensures that the two-way fitted probabilities equal the observed ones (e.g., if $\lambda_{hl}^{\mathcal{H}\mathcal{L}}$ is in the model, then $\pi_{hl\cdot\cdot} = p_{hl\cdot\cdot}$). In sum, the presence of a constant, marginal effect term and interaction terms guarantee that the fitted probabilities from the model reproduces the corresponding observed proportions. Since the experimenter determined the high by low signal margin (and it is inherently uninteresting), the term $\lambda_{hl}^{\mathcal{H}\mathcal{L}}$ should always be in the model. Location constraints are required on the log-linear model parameters to identify them (e.g., $\sum_j \lambda_j^{H} = 0$ and $\sum_h \lambda_{hj}^{\mathcal{H}H} = \sum_j \lambda_{hj}^{\mathcal{H}H} = 0$).

Besides algebraic representations, log-linear models also have schematic or graphical representations (Edwards, 2000; Whittaker, 1990). For example, model (7) is represented by the graph labelled 1a in the top part of Figure 1. The boxes represent the variables, lines connecting two boxes (i.e., variables) represent possible dependence, and the absence of a line between two variables represents conditional independence. According to model (7), the ratings on the two signals are conditionally independent given the stimuli, so there is no line connecting the high and low ratings in the graph. The line connecting the high signal and high rating and the one connecting the low signal and low rating indicate dependence between these variables may exist. Whether dependence actually exists depends on whether $\lambda_{hj}^{\mathcal{H}H}$ equals 0 for all h and j and $\lambda_{lk}^{\mathcal{L}L}$ equals 0 for all l and k.

A slightly more complex log-linear model for the data allows for a possible dependence between the ratings themselves (e.g., perhaps a response strategy on the part of the subject). The graph representing this model is labelled 1b in Figure 1. The corresponding algebraic model is

$$\log(\pi_{hljk}) = \lambda + \lambda_h^{\mathcal{H}} + \lambda_l^{\mathcal{L}} + \lambda_j^{H} + \lambda_k^{L} + \lambda_{hj}^{\mathcal{H}H} + \lambda_{lk}^{\mathcal{L}L} + \lambda_{jk}^{HL} + \lambda_{hl}^{\mathcal{H}\mathcal{L}}. \qquad (8)$$

Model (8) contains a sub-set of all possible bivariate interactions between pairs of variables. Deleting one or more set of interaction terms from model (8) yields a model that represents some form of independence.

When dealing with hierarchical log-linear models (i.e., models which include all lower order terms that comprise the interaction terms), the (in)dependence between two variables, say the high signal $\mathcal{H}$ and the high rating H, given fixed levels the other variables, can be expressed through the odds ratios in the 2×2 subtables of $\mathcal{H}$ and H. For instance, if the relationship between the high signal and high rating is independent of the presence and rating of the low signal, then the odds-ratios in these sub-tables for high stimuli h and h' and high ratings j and j' for fixed values of k and l do not depend on the specific values of k and/or l.

To show that model (8) (and all models that only include bivariate interaction terms) use a multiplicative definition of no three-way interaction, we first examine the *conditional* or *partial* odds ratios for two variables given a fixed level of the third variable. According to model (8), the odds ratio in the 2×2 subtables of variables $\mathcal{H}$ and H given level l and k of variables $\mathcal{L}$ and L equals

$$\begin{aligned} \theta_{hh',jj'(lk)} &= \frac{\pi_{hljk}/\pi_{hlj'k}}{\pi_{h'ljk}/\pi_{h'lj'k}} \\ &= \exp\left((\lambda_{hj}^{\mathcal{H}H} - \lambda_{hj'}^{\mathcal{H}H}) - (\lambda_{h'j}^{\mathcal{H}H} - \lambda_{h'j'}^{\mathcal{H}H})\right), \qquad (9) \end{aligned}$$

where the subindices in parentheses indicate the categories of the conditioning variables. The odds ratios $\theta_{ll',kk'(hj)}$ and $\theta_{hh',ll'(jk)}$ are also functions of their corresponding interaction terms. The dependence between two variables in partial sub-tables does not depend on the value of the other variables.

The definition of no three-way interaction that is lurking behind the scenes can be seen by considering the *ratio of odds ratios* for two variables given different levels of a third variable for a fixed level of the remaining variables (if there are any). For example, the ratio of odds ratios for the high signal conditions (i.e., h, h') and high ratings (i.e., j, j') given different low signal conditions (i.e., l, l') and a specified low rating (i.e., k) based on model (8) is

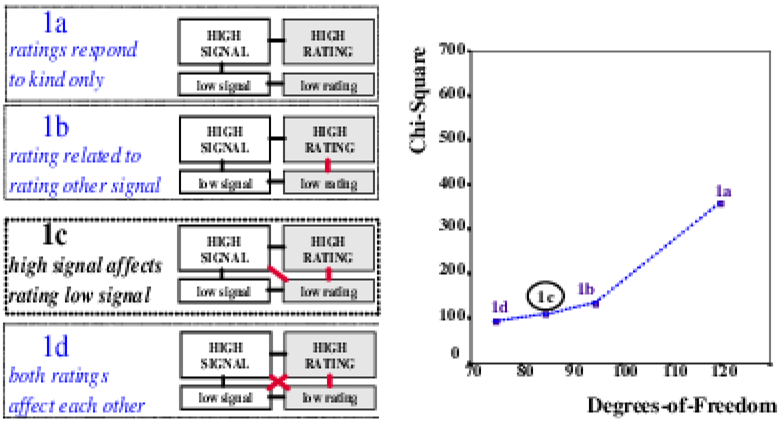
$$\frac{\theta_{hh',jj'(lk)}}{\theta_{hh',jj'(l'k)}} = \frac{\pi_{hljk}\pi_{h'lj'k}\pi_{h'l'jk}\pi_{hl'j'k}}{\pi_{h'ljk}\pi_{hlj'k}\pi_{hl'jk'}\pi_{h'l'j'k}} = 1. \qquad (10)$$

Equation (10) is equivalent to the multiplicative definition given earlier in equation (2). To see this, replace h in equation (10) by i and note that k is constant. In other words, models that include only bivariate interactions such as (8) use a multiplicative definition of no three-way interaction. An implication is that a measure of three-way association is the ratio of the odds ratios and the extent to which equation (2), or specifically in this case equation (10), departs from 1 indicates the amount of three-way dependence in data.

A hierarchical log-linear model that includes a three-way interaction is

$$\log(\pi_{hljk} = \lambda + \lambda_h^{\mathcal{H}} + \lambda_l^{\mathcal{L}} + \lambda_j^{H} + \lambda_k^{L} + \lambda_{hj}^{\mathcal{H}H} + \lambda_{lk}^{\mathcal{L}L} + \lambda_{jk}^{HL} + \lambda_{hl}^{\mathcal{H}\mathcal{L}} + \lambda_{hk}^{mathcal HL} + \lambda_{hlk}^{\mathcal{H}\mathcal{L}L} \quad (11)$$

The graphical representation of this model is labelled 1c in Figure 1. According to this model, the ratios of odds ratios, $\theta_{ll',kk'(hj)}/\theta_{ll',kk'(h'j)}$, are functions of the three-way interaction parameters $\lambda_{hlk}^{\mathcal{H}\mathcal{L}L}$. These ratios are not equal to 1 unless $\lambda_{hlk}^{\mathcal{H}\mathcal{L}L} = 0$ for all (h, l, k).



(a) Models (b) Deviance Plot

Figure 1: Concurrent Detection Data: Selection of loglinear models.

The log-linear models given in equations (7), (8) and (11), which correspond to graphs 1a, 1b and 1c in Figure 1, respectively, were fit to the data, as well as a model that includes the two three-way interactions $\lambda_{hlj}^{\mathcal{H}\mathcal{L}H}$ and $\lambda_{hlk}^{\mathcal{H}\mathcal{L}L}$, which has graphical representation 1d. The results for these models are presented here in the form of a deviance plot (Fowlkes, Freeny, & Landwehr, 1988). The 'convex hull' connecting the outer points on the lower edge is drawn to identify candidate models for selection. Ideally one wants to have a low χ^2 with as many degrees-of-freedom as possible. Based on the deviance plot, log-linear models 1c and 1d in Figure 1 are candidates for selection.

There are three-way interactions between the two signals and each of the ratings; therefore, the data are reformatted as a three-way table such that there is one factor, which consists of the four possible combinations of signals (i.e., a "joint signal" or "signal condition"). In our data, we will use $\mathcal{J}$ to denote joint signal and index it as $i = 1$ (both signals absent), 2 (high present, low absent), 3 (high absent, low present), 4 (both present). Therefore, in the next section, our starting log-linear model is

$$\log(\pi_{ijk}) = \lambda + \lambda_i^{\mathcal{J}} + \lambda_j^{H} + \lambda_k^{L} + \lambda_{ij}^{\mathcal{J}H} + \lambda_{ik}^{\mathcal{J}L} + \lambda_{jk}^{HL}, \quad (12)$$

which has bivariate interaction terms that represent the three-way interactions between the two stimuli and each of the responses (i.e., $\lambda_{ij}^{\mathcal{J}H}$ and $\lambda_{ik}^{\mathcal{J}L}$).

4.2 Log-multiplicative association models

In this section, we turn to those extensions of log-linear models that provide more interpretable representations of dependence in three-way tables. In particular, we seek to separate the systematic patterns from the unsystematic parts of the interactions so as to gain insight into the nature of the association between the variables. The main tool for this will be the SVD (for a good introduction to the SVD, see Greenacre, 1984, pp. 340ff).

Interaction terms in log-linear models are unstructured in the sense that they equal whatever they need to equal such that the corresponding margins of the data are fit perfectly. Goodman (1979, 1985) proposed that the two-way interaction terms in a saturated log-linear model for a two-way table be replaced by a lower rank approximation based on a SVD of the unstructured interaction terms (Clogg, 1986). The resulting model, known as the multidimensional row-column or $RC(M)$ association model, is log-multiplicative rather than log-linear, because it includes multiplicative terms for the interaction. Given the success of the $RC(M)$ model at providing interpretable representations of dependence in two-way tables, numerous extensions to three- and higher-way tables were proposed. Many of these proposals use bilinear terms (e.g., Becker, 1989), trilinear terms (e.g. Anderson, 1996), or both bilinear and trilinear terms (e.g., Choulakian, 1988a). Wong (2001) provides an extensive summary of most of these proposals including the required identification constraints.

We present a subset of possible log-multiplicative models for three-way tables (i.e., those which prove useful for the data at hand). Since our starting log-linear model, equation (12), only includes bivariate interaction terms, we present a general log-multiplicative model that only includes SVDs of two-way interactions. Following a brief review of the essential elements of log-multiplicative association models, we present a general strategy for modelling data that connects substantive research hypotheses to models through the use of a latent variable interpretation of log-multiplicative models.

Bivariate association models

We start with a general log-multiplicative model for three-way tables discussed by Becker (1989) that only includes bivariate interactions. In this model, each of the two-way interaction terms is replaced by a sum of bilinear terms, which can be computed via the SVD. In the case of our signal detection data, we replace the interaction terms $\lambda_{ij}^{\mathcal{J}H}$, $\lambda_{ik}^{\mathcal{J}L}$ and λ_{jk}^{HL} in equation (12) with separate bilinear terms; that is,

$$\begin{aligned}\log(\pi_{ijk}) &= \lambda + \lambda_i^{\mathcal{J}} + \lambda_j^H + \lambda_k^L + \sum_{r=1}^{R} \sigma^2_{\mathcal{J}H(r)} \omega_{ir}^{\mathcal{J}H} \nu_{jr}^{\mathcal{J}H} \\ &\quad + \sum_{s=1}^{S} \sigma^2_{\mathcal{J}L(s)} \omega_{is}^{\mathcal{J}L} \eta_{ks}^{\mathcal{J}L} + \sum_{t=1}^{T} \sigma^2_{HL(t)} \nu_{jt}^{HL} \eta_{kt}^{HL}, \qquad (13)\end{aligned}$$

where $\omega_{ir}^{\mathcal{J}H}$ and $\nu_{jr}^{\mathcal{J}L}$ are scale values for joint signal i and high rating j on dimension r representing the $\mathcal{J}H$ association, $\omega_{is}^{\mathcal{J}L}$ and $\eta_{ks}^{\mathcal{J}L}$ are scale values for joint signal i and low rating k on dimension s representing the $\mathcal{J}L$ association, ν_{jt}^{HL} and η_{kt}^{HL} are the scale values for high rating j and low rating k representing the HL association on dimension t, and $\sigma^2_{\mathcal{J}H(r)}$, $\sigma^2_{\mathcal{J}L(s)}$ and $\sigma^2_{HL(t)}$ are association parameters that measure the strength of the $\mathcal{J}H$, $\mathcal{J}L$ and HL relationships on dimensions r, s and t, respectively. For identification, location constraints are required on all parameters (e.g., $\sum_i \lambda_i^{\mathcal{J}} = \sum_i \omega_{ir}^{\mathcal{J}H} = 0$), and additional scaling and orthogonality constraints are required for the scale values (e.g., $\sum_i \omega_{ir}^{\mathcal{J}H}\omega_{ir'}^{\mathcal{J}H} = 1$ if $r = r'$ and 0 otherwise).

When $R = \min(I, J) - 1$, $S = \min(I, K) - 1$ and $T = \min(J, K) - 1$, model (13) is equivalent to log-linear model (12). Models where $R < \min(I, J) - 1$, $S < \min(I, K) - 1$ and/or $T < \min(J, K) - 1$ are all special cases of the log-linear model (12). Additional special cases can be obtained by placing equality restrictions on the scale values and association parameters across interaction terms, such as $\omega_{ir}^{\mathcal{J}H} = \omega_{is}^{\mathcal{J}L}$ and $\sigma^2_{\mathcal{J}H(r)} = \sigma^2_{\mathcal{J}L(s)}$ for $r = s$. Based on our experience and published applications, models with one or two dimensions often fit data well (i.e., R, S and/or T equal 1 or 2).

Given (13), the conditional odds ratios are functions of the scale values and association parameters. Based on model (13),

$$\begin{aligned}
\log(\theta_{ii',jj'(k)}) &= \sum_r \sigma^2_{\mathcal{J}H(r)}(\omega_{ir}^{\mathcal{J}H} - \omega_{i'r}^{\mathcal{J}H})(\nu_{jr}^{\mathcal{J}H} - \nu_{j'r}^{\mathcal{J}H}) \\
\log(\theta_{ii',kk'(j)}) &= \sum_s \sigma^2_{\mathcal{J}L(s)}(\omega_{is}^{\mathcal{J}L} - \omega_{i's}^{\mathcal{J}L})(\eta_{ks}^{\mathcal{J}L} - \eta_{k's}^{\mathcal{J}L}) \\
\log(\theta_{jj',kk'(i)}) &= \sum_t \sigma^2_{HL(t)}(\nu_{jt}^{HL} - \nu_{j't}^{HL})(\eta_{kt}^{HL} - \eta_{k't}^{HL}).
\end{aligned}$$

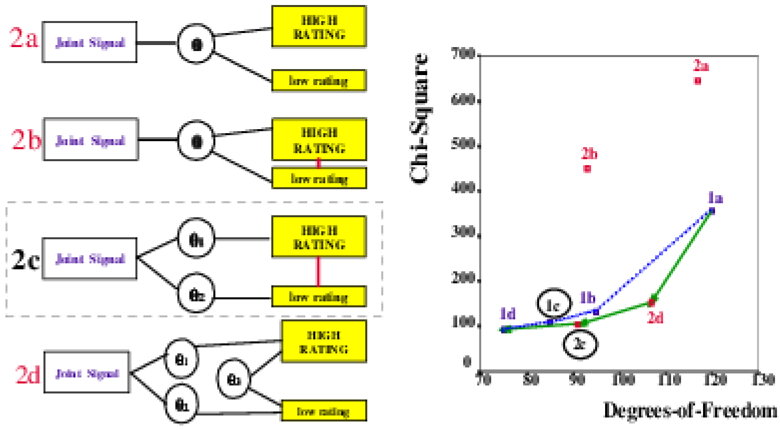
Using the fact that odds ratios are functions of scale values and association parameters, plots of scale values provide visual displays of the dependence structure in the data as measured by a multiplicative definition of interaction (i.e., odds ratios).

With higher-way tables, not only does one have the flexibility to decide what interactions to represent by multiplicative term(s), but also what restrictions should be placed on parameters. Given the large number of possibilities, we turn to an approach that provides guidance for the construction of an appropriate model or a subset of models for a given application.

Log-multiplicative latent variable models

The approach advocated here starts with a researcher's substantive theories about underlying processes. We describe in this section a latent variable model from which we derive log-multiplicative association models (Anderson & Vermunt, 2000; Anderson & Böckenholt, 2000; Anderson, 2002). The latent variable model described here is based on statistical graphical models for discrete and continuous variables (Lauritzen & Wermuth, 1989; Whittaker, 1989; Wermuth & Lauritzen, 1990). Just as we have graphical representations of log-linear models, adopting the latent variable perspective provides graphical representations of log-multiplicative association models.

In the latent variable model, the discrete variables are observed and the continuous variables are unobserved. In this chapter, the observed discrete variables, or a subset of them, are conditionally independent given the latent continuous variables. Provided a number of assumptions are met, the model derived for the observed discrete variables is a log-multiplicative association model (for details, see Anderson & Vermunt, 2000, or Anderson, 2002).



(a) Models (b) Deviance Plot

Figure 2: Concurrent Detection Data: Selection of latent association models. Boxes are observed variables; circled θ indicate latent variables. Model 2c is the preferred one.

Examples of graphical representations for latent variable models implying log-multiplicative models are given in Figure 2. In these figures, observed (discrete) variables are represented by boxes, and the latent (continuous) variables are represented by circles. Lines connecting variables indicate that the variables may be conditionally dependent and the absence of a line between two variables indicates that the variables are conditionally independent. Note that in graphs 2a and 2d in Figure 2, the three of the observed variables (boxes) are conditionally independent of each other given the latent variable(s). The other two graphs in Figure 2 (i.e., 2b and 2c) permit conditional dependence between two of the observed discrete variables (i.e., the high and low ratings). Unstructured two-way interaction terms are included if discrete (observed) variables are connected by a line. A multiplicative term is included if there is a path between two observed discrete values that passes through one latent variable. For example, consider graph 1a in Figure 2, which postulates one latent variable (perhaps a subjective impression of the joint stimulus) where each of the ratings is related to this latent variable. The log-multiplicative model with this graphical representation is

$$\log(\pi_{ijk}) = \lambda + \lambda_i^{\mathcal{J}} + \lambda_j^H + \lambda_k^L + \sigma^2\omega_i^{\mathcal{J}}\nu_j^H + \sigma^2\omega_i^{\mathcal{J}}\eta_k^L, \tag{14}$$

where σ^2 is the variance of latent variable θ, and given the additive model assumptions for the mean of the latent variable in cell (i, j, k) is

$$\mu_{ijk} = \sigma^2(\omega_i^2 + \nu_j^H + \eta_k^L).$$

The term $\sigma^2\omega_i^{\mathcal{J}}\nu_j^H$ is included in the model because there is a path between the "Joint signal" and the "HIGH RATING" that goes through the latent variable, and the term $\sigma^2\omega_i^{\mathcal{J}}\eta_k^L$ is included in the model because there is a path between the "Joint Signal" and the "low rating" that goes through the latent variable. A slightly more complex model is given in graph 2b, which has a line connecting the high and low ratings. The algebraic model for graph 2b is equation (14) with the (unstructured) interaction term λ_{jk}^{HL} added to represent a possible association between the high and low ratings.

From the deviance plot in Figure 2, the models corresponding to graphs 2a and 2b clearly do not fit very well relative to the log-linear models. The latent variable structure for these models is too simple; therefore, we add an additional latent variable. It may be the case there are separate internal (subjective) impressions of the signals and each of the ratings are based on the their subjective impressions. Graphs 2c and 2d in Figure 2 are diagrams for this conjecture. The log-multiplicative model corresponding to graph 2c is

$$\log(\pi_{ijk}) = \lambda + \lambda_i^{\mathcal{J}} + \lambda_j^H + \lambda_k^L + \lambda_{jk}^{HL} + \sigma_1^2\omega_i^{\mathcal{J}H}\nu_j^H + \sigma_2^2\omega_i^{\mathcal{J}L}\eta_k^L \qquad (15)$$

where σ_1^2 and σ_2^2 are the variances of latent variables θ_1 and θ_2, respectively. Since there is a line connecting the box labelled "HIGH RATING" and "low rating", we included the term λ_{jk}^{HL}. We include the multiplicative term $\sigma_1^2\omega_i^{\mathcal{J}H}\nu_j^H$ because there is a path between the box labelled "Joint Signal" and "HIGH RATING" that passed through the circle labelled "θ_1". The second multiplicative term $\sigma_2^2\omega_i^{\mathcal{J}L}\nu_k^L$ was included because there is a path between the box labelled "Joint Signal" and "low rating" that passes through the circle labelled "θ_2". In the model corresponding to graph 2d, the multiplicative term $\sigma_3^2\nu_j^{HL}\eta_k^{HL}$ replaces λ_{jk}^{HL}.

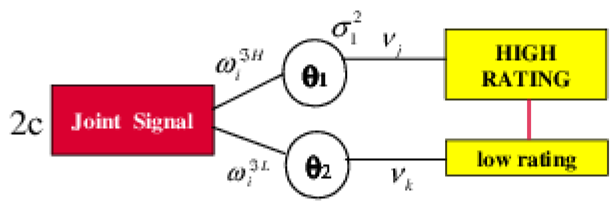


Figure 3: Concurrent Detection Data: Final latent association model with a joint signal variable as factor and two latent variables

Comparing the fit of the log-linear models from Figure 1 (i.e., models 1a - 1d) and the log-multiplicative models corresponding to graphs 2c and 2d in Figure 2, the latter are more parsimonious. Based on a detailed analysis, Anderson (2002) considered Model 2c to be adequate for describing the patterns in the data. Note that although the fits of Model 1c and Model 2c are almost equal, the latter model retains more degrees of freedom. This model has two latent variables, one for each signal. We can further refine this model by placing restrictions on the parameters, which correspond to specific hypotheses. For example, the way the subject "translates" the subjective impressions of a signal into a response may be the same regardless of whether rating the high or low signal. This conjecture would be represented by restricting $\nu_j^H = \eta_j^L = \nu_j$ for all $j = 1, \ldots, 6$. Specific hypotheses regarding the differential effect of the design variables on ratings tested by placing restrictions on ω_i^H and/or ω_k^L. Additional restrictions were placed on the interactively coded design variables to test, for example, whether the subjective confidence of the low signal when it is absent is the same regardless of whether the high signal is present (i.e., $\omega_1^{\mathcal{J}L} = \omega_2^{\mathcal{J}L}$).

Figure 3 represents the final model with the paths between variables labelled by parameters of the model representing the association. Table 2 contains the parameters estimates of the final model that include restrictions on the parameters.

Table 2: Concurrent Detection Data: Parameters of the Final Latent Association Model and Quantifications of Categorical PCA (CatPCA) Analysis.

Variable	Level	Parameter Estimate	CatPCA Quantifications[a] High	Low
Ratings	1	−0.41	−0.47	−0.44
(high confidence	2	−0.37	−0.38	−0.38
=	3	−0.22	−0.33	−0.21
low confidence)	4	−0.05	−0.11	0.06
(ν_j)	5	0.31	0.28	0.41
	6	0.74	0.66	0.66
Low ratings	HIGH – & low –	−0.50		
	HIGH + & low –	−0.50		
($\omega_i^{\mathcal{J}L}$)	HIGH – & low +	0.58		
	HIGH + & low +	0.44		
	variance	3.40	dim. 1	dim. 2[b]
HIGH RATINGS	HIGH – & low –	−0.58	−0.49	−0.74
	HIGH – & low +	−0.39	−0.51	−0.04
($\omega_i^{\mathcal{J}H}$)	HIGH + & low –	0.41	0.58	0.10
	HIGH + & low +	0.58	0.41	0.67
	variance	2.69		

Note: [a]Quantifications of confidence scores from CatPCA; scaled to unit length.[b]45^o rotated versions of category quantifications from a categorical PCA on the three-way table (see section 6 with Joint Signal ($\mathcal{J}$) as one of the variables. These quantifications are independent from the confidence ratings.

Note that the differences between the successive pairs of scale values increase with the values of the ratings. Variances can be interpreted as measures of strength of the relationships between stimuli and responses. The estimated cell means of the subjective confidence for the high signal, θ_1, equal $\mu_{1(ij)} = \sigma_1^2(\omega_i^{\mathcal{J}H} + \nu_j)$, and those for the low signal, θ_2, equal $\mu_{2(ik)} = \sigma_2^2(\omega_i^{\mathcal{J}L} + \nu_k)$.

5 Additive models: three-way correspondence analysis

In Section 3 it was pointed out that models using the multiplicative definition of interaction decompose each (group of) interactions separately, while models using the additive definition of interaction decompose the complete dependence with a single multiplicative model. The decomposition of the separate interactions is then derived from the global one. The most prominent model using the additive definition is three-way correspondence analysis (CA) and, in contrast with the association models, no distributional assumptions are made or used. In this sense, three-way CA is an exploratory technique. Its most important properties are that (1) the dependence between the variables in a three-way table is additively decomposed such that the relative size of each interaction term can be assessed with respect to each other, (2) a single multiplicative model for dependence operating at the level of proportions rather than on the log-scale is used to model global, marginal, and partial dependence allowing for additively assessing the contributions of these interactions to the modelled dependence, and (3) graphical comparisons between all interactions can be made within a single graph. Full details can be found in Carlier and Kroonenberg (1996, 1998), earlier technical results are contained in Dequier (1974), Choulakian (1988b) and Kroonenberg (1989).

5.1 Measuring dependence

Measuring global dependence

The global dependence between the joint signal $\mathcal{J}$ and the confidence in the low signal L and the confidence in the high signal H is measured by the *mean squared contingency* Φ^2, defined as

$$\begin{aligned}\Phi^2 &= \sum_i \sum_j \sum_k \frac{(p_{ijk} - p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k})^2}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}} = \sum_i \sum_j \sum_k p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k} \left[\frac{p_{ijk}}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}} - 1\right]^2 \\ &= \sum_i \sum_j \sum_k p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}P^2_{ijk}. \end{aligned} \tag{16}$$

Φ^2 is based on the deviations from the three-way independence model, and it contains the global dependence due to all two-way interactions and the three-way interaction.

For instance, the two-way marginal total of the two confidence ratings is defined as the sum over the joint signal weighted by its marginal proportion. If the conditional proportions for all values of the joint signal i are equal, then $p_{jk|i} = p_{\cdot jk}$. Then $P_{ijk} = P_{\cdot jk}$, and the three-way table can be analysed with ordinary CA between the two confidence ratings. The symmetric statement after permutation of the indices holds as well. One-way marginal totals are weighted sums over two indices and they are zero due to the definition of P_{ijk}, and thus the overall total is zero as well.

Measuring marginal and three-way dependence

The global dependence of cell, P_{ijk}, can be split into additive contributions of the two-way interactions and the three-way interaction (see also equation (5)).

$$P_{ijk} = \frac{p_{ij\cdot} - p_{i\cdot\cdot}p_{\cdot j\cdot}}{p_{i\cdot\cdot}p_{\cdot j\cdot}} + \frac{p_{i\cdot k} - p_{i\cdot\cdot}p_{\cdot\cdot k}}{p_{i\cdot\cdot}p_{\cdot\cdot k}} + \frac{p_{\cdot jk} - p_{\cdot j\cdot}p_{\cdot\cdot k}}{p_{\cdot j\cdot}p_{\cdot\cdot k}} + \frac{p_{ijk} - \tilde{p}_{ijk}}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}}\ , \tag{17}$$

where $\tilde{p}_{ijk} = p_{ij\cdot}p_{\cdot\cdot k} + p_{i\cdot k}p_{\cdot j\cdot} + p_{\cdot jk}p_{i\cdot\cdot} - 2p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}$ (see equation (6)). The last term of equation (17) measures the contribution of three-way interaction to the dependence for cell (i,j,k)(see also section 3, equation (5)).

Using the definition of global dependence of cells, the measure of global dependence of a table is defined as the sum over all cell dependencies, $\Phi^2 = \sum_{ijk} P_{ijk}$. Due to the additive splitting of the dependence of individual cells, Φ^2 can also be partitioned additively

$$\begin{aligned}
\Phi^2 &= \sum_i \sum_j p_{i\cdot\cdot}p_{\cdot j\cdot}\left(\frac{p_{ij\cdot} - p_{i\cdot\cdot}p_{\cdot j\cdot}}{p_{i\cdot\cdot}p_{\cdot j\cdot}}\right)^2 + \sum_i \sum_k p_{i\cdot\cdot}p_{\cdot\cdot k}\left(\frac{p_{i\cdot k} - p_{i\cdot\cdot}p_{\cdot\cdot k}}{p_{i\cdot\cdot}p_{\cdot\cdot k}}\right)^2 + \\
&\quad \sum_j \sum_k p_{\cdot j\cdot}p_{\cdot\cdot k}\left(\frac{p_{\cdot jk} - p_{\cdot j\cdot}p_{\cdot\cdot k}}{p_{\cdot j\cdot}p_{\cdot\cdot k}}\right)^2 + \sum_i \sum_j \sum_k p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}\left(\frac{p_{ijk} - \tilde{p}_{ijk}}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}}\right)^2 \\
&= \Phi^2_{IJ} + \Phi^2_{IK} + \Phi^2_{JK} + \Phi^2_{IJK}.
\end{aligned} \tag{18}$$

The importance of decomposition (18) is that it provides measures of fit for each of the interactions and thus their contributions to the global dependence.

The left-hand panel of Table 3 shows this partitioning for the Wickens-Olzak data. The two-way margin of the two confidence ratings $L \times H$ is distinctly smaller (18%) than the two-way interaction of the joint signal with the high ratings $\mathcal{J}H$ (28%) and that between the joint signal and the low ratings $\mathcal{J}L$(33%), with the three-way interaction in between (21%). The $L \times H$-interaction is not really interpretable given the presence of the other interactions because it is the sum of the four tables in Table 1, and summing the graphs virtually eliminates all systematic patterns present in the data. The other two two-way interactions have straightforward interpretations as they indicate to what extent there is a high (low) confidence in the presence of a particular combination of signals. Finally, the three-way interaction represents about a fifth of the dependence. Due to the presence of error, generally only a small part of this interaction contains interpretable information about the mutual influence of the ratings and the joint signal.

5.2 Modelling dependence

Up to this point, the decomposition of equation (17) is the additive parallel of the log-linear model of equation (11). In particular, it has separate terms for each of the interactions, but no decomposition of the interaction terms themselves has been considered yet. In ordinary CA, the SVD is used to acquire a more detailed insight into the nature of the dependence, and in three-way CA a three-way analogue of the SVD is needed.

Table 3: Concurrent Detection Data: Partitioning Fit for the $2 \times 2 \times 2$-Model

		Global Dependence		Residual Dependence		Residual
Source	*df*	χ^2	%	χ^2	%	% of Interaction
Joint Signal($\mathcal{J}$)	3	0	0	0	0	
HIGH RATINGS (H)	5	0	0	9	2	
low ratings (L)	5	0	0	1	0	
$\mathcal{J} \times H$	15	357	28	18	4	5
$\mathcal{J} \times L$	15	416	33	15	4	4
$L \times H$	25	222	18	199	46	90
Three-way interaction	75	261	21	193	44	74
Total dependence	130	1256	100	436	100	35

Note: % of Interaction = Residual χ^2/Global $\chi^2\times$ 100%; e.g. 90% = 199/222 $\times$ 100%

Modelling global dependence

There are several candidates for the three-way generalisation of the SVD, in particular Tucker's three-mode decomposition (Tucker, 1966) and Harshman's (1970) PARAFAC. In this paper we will only discuss the Tucker3 model, i.e. the global dependence is decomposed as

$$P_{ijk} = \sum_r \sum_s \sum_t g_{rst} a_{ir} b_{js} c_{kt} + e_{ijk}, \tag{19}$$

where a_{ir} are scale values for the joint signal, and they are orthogonal with respect to their weights $p_{i..}$ (i.e. $\sum_i p_{i..} a_{ir} a_{ir'} = 1$ if $r = r'$ and 0 otherwise). Similarly, the b_{js} are the orthogonal scale values for the confidence in the presence of the high signal, and the c_{kt} those for the confidence in the presence of the low signal, with dimensions r, s, and t, respectively. The g_{rst} are the three-way association parameters or analogues of the singular values, and the e_{ijk} represent the errors of approximation. In three-way CA, a weighted least-squares criterion is used: the parameters g_{rst}, a_{ir}, b_{js} and c_{kt} are those which minimise

$$\sum_i \sum_j \sum_k p_{i..} p_{.j.} p_{..k} e_{ijk}^2.$$

Thus the global measure of dependence, Φ^2, can be split into a part fitted with the three-way SVD and a residual part.

Modelling marginal dependence

The marginal dependence of the joint signal i and high ratings j is equal to

$$P_{ij.} = \left(\frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}} \right)$$

with similar expressions for the other two marginal dependencies. The elements $P_{ij.}$ are derived via a weighted summation over k from the global dependence of the cells,

$$P_{ij.} = \sum_k p_{..k} P_{ijk} \; . \tag{20}$$

Given we have modelled the global dependence P_{ijk} with the Tucker3 model in equation (19), we can use equation (20) to find the model for the marginal dependence,

$$P_{ij.} = \sum_r \sum_s \sum_t g_{rst} a_{ir} b_{js} c_{.t} + e_{ij.} \; , \tag{21}$$

with $c_{.t} = \sum_k p_{..k} c_{kt}$ and $e_{ij.} = \sum_k p_{..k} e_{ijk}$. Inspecting this formula leads to the conclusion that the marginal dependence between the joint signal and the high ratings is derived from the overall model by averaging the low ratings components.

Whereas in standard CA an optimal common dimension has to be determined for the rows and the columns, three-way CA requires a determination of the numbers of components for all three ways. For the Wickens-Olzak data all models with dimensions less than or equal to 3 for each way were investigated and displayed in a deviance plot. The most relevant models are included in Figure 4 together with the relevant log-linear and log-multiplicative models. From this figure we can see that the confidence ratings each needed two dimensions, and the joint-signal mode needed either two or three dimensions. Given the desire for relative simplicity and the fact that the proportional χ^2 values of the signal dimensions for the $2\times2\times2$ model were .39 and .26, and for the $3 \times 2 \times 2$ model .39, .26, and .07, led us to prefer the former. Thus of the total χ^2 in the three-way table, 65% is fitted by the three-way CA.

From Figure 4 it can be seen that the multiplicative models clearly outperform the additive models; they combine the same degrees-of-freedom with a lower chi-square. It is not quite clear why the CA fits so dramatically worse. One reason might be that the additive models try to fit the complete interaction (i.e. the global dependence), while the log-multiplicative models only fit the three-way interaction. A thorough investigation into this aspect will have to be made in the future.

The results for the $2 \times 2 \times 2$ three-way CA are summarised in the right-hand panel of Table 3. The $L \times H$ interaction has a bad fit of only 10%, but as argued above it contains very little interesting information about the problem at hand anyway. The other two two-way interactions are fitted very well by the model with 95% and 96% of their interaction explained by the model. Finally, only 26% of the three-way interaction is captured by the three-way model, but this is as expected. The fitted χ^2s of $\mathcal{J} \times H$ and $\mathcal{J} \times L$ together account for 90% of the total fitted χ^2 of 820 (=1256-436), in other words it is primarily these two interactions that are being fitted by the three-way CA.

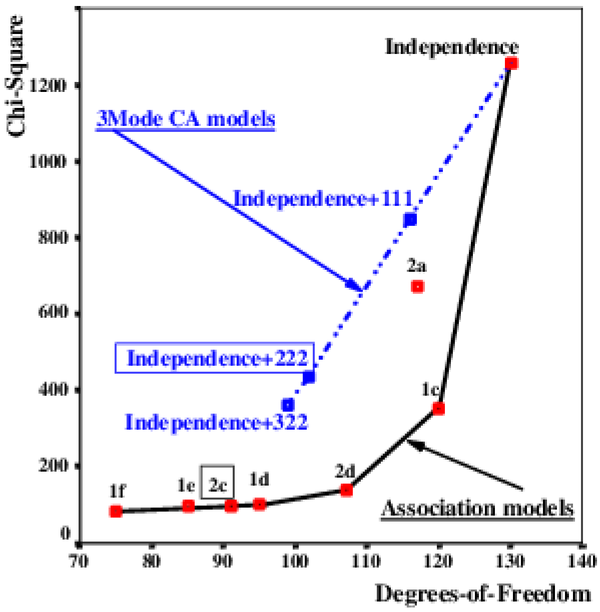


Figure 4: Concurrent Detection Data: Deviance Plot containing both multiplicative and additive models. "Independence" indicates the model of three-way independence; "Independence + rst" indicates the independence model with a decomposition of the global dependence by a $r \times s \times t$-Tucker3 model. "Independence + 222" is the preferred additive model and Model 2c the preferred multiplicative model.

5.3 Plotting dependence

The three-way CA model is symmetric in its three ways. However, this symmetry cannot be maintained when graphing the dependence, because no spatial representations exist to portray all three ways simultaneously in one graph. A strict parallel with ordinary CA can therefore not be maintained. To display the dependence or its approximation in three-way CA, we will make use of a *nested-mode biplot*, which was called the *interactive biplot* in Carlier and Kroonenberg (1998), but the present term is more explicit and avoids overusing the term interactive.

Plotting global dependence: Nested-mode biplot

The *nested-mode biplot* aims to portray all three ways in a single biplot. As a biplot has only two type of markers, two ways have to be combined into a single one. In the Wickens-Olzak data we have combined the two confidence ratings, indexed by (j, k) and represented it by a single marker ℓ. Thus for the construction of the plot the confidence ratings are coded interactively. The remaining mode, i.e. the joint signal, defines the plotting space and it also supplies a set of markers; it will be referred to as the *reference mode*.

The construction of the biplot for the fitted part of global dependence, designated as $\widehat{P}_{ijk}$ follows directly from three-way SVD of the global dependence.

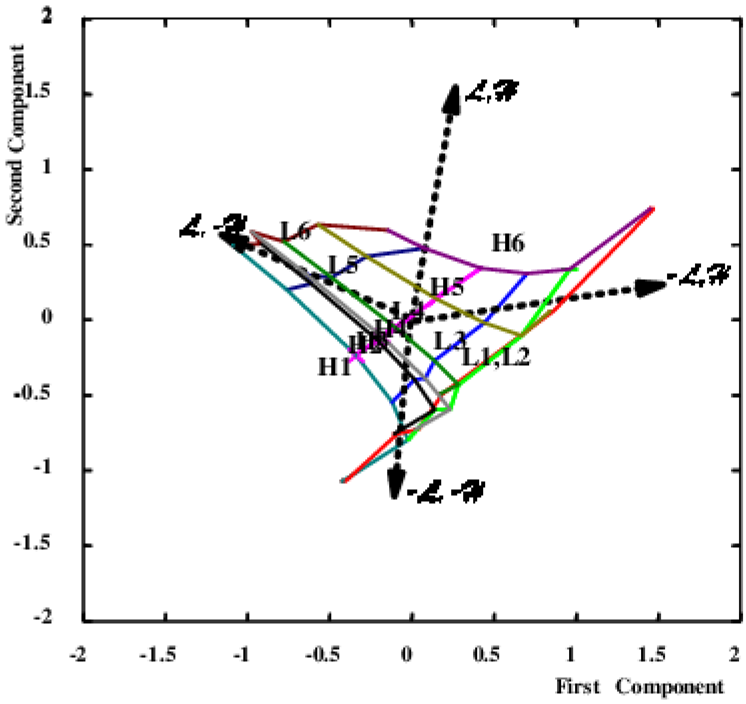


Figure 5: Concurrent Detection Data: Nested-mode biplot from three-mode correspondence analysis. *Hi* (*Li*) *i*-th level of confidence in presence of high (low) signal, i.e. co-ordinates of the HIGH RATINGS×Joint Signal (low ratings×Joint Signal) two-way margins. The script letters with associated arrows indicate the co-ordinates of the Joint Signal categories.

$$\begin{aligned}\widehat{P}_{ijk} &= \sum_r \left[\sum_s \sum_t g_{rst} b_{js} c_{kt}\right] a_{ir} = \sum_r d_{(jk)r} a_{ir} \\ \widehat{P}_{\ell i} &= \sum_r d_{\ell r} a_{ir} \ . \end{aligned} \tag{22}$$

By replacing the (jk) with a new index ℓ we see that the co-ordinates of the combined response variables on component r are the $d_{\ell r}$, and the a_{ir} are the co-ordinates of the joint signals. Given this construction, the combined response variables are in principal co-ordinates and the joint signals in standard co-ordinates. When plotting these co-ordinates in the nested-mode biplot we will portray the combined response variables as points and the joint-signal co-ordinates as arrows.

The interpretation of the nested-mode biplot can be enhanced by exploiting the ordinal nature of both response variables. In particular, we can draw a grid by connecting in their natural order both the high confidence ratings for each value of the low confidence rating and vice versa. The result is the grid in Figure 5, which shows the nested-mode biplot for the $2 \times 2 \times 2$ three-way CA. As 90% of the global dependence consists of the $\mathcal{J} \times H$ and $\mathcal{J} \times L$ dependence, these interactions primarily determine the shape of the plot.

What can one learn from such a plot? First, note that the co-ordinates of the signals $(\mathcal{H}, \mathcal{L}, \mathcal{H}, -\mathcal{L}, -\mathcal{H}, \mathcal{L}, -\mathcal{H}, -\mathcal{L})$, form nearly a rectangular cross indicating the relative independence of the high and low signals. The slight upward turn of the arrows for the condition where only one signal is present, indicates that slightly more often the subject judged the not-presented signal to be present as well. A characteristic of the confidence grid is that the presence (absence) of a signal generates the corresponding appropriate confidence score. In addition, the lack of signal generally leads to low confidence scores but there is not much differentiation between the lower levels of the confidence scores, especially in the case of the absence of a low signal. The longer grid lines for the high-signal confidence when the confidence in the presence of the low signal is at its lowest, indicate that in those circumstances the confidence in the presence of the high signal is much more marked.

Plotting Marginal Dependence

In section 5.2 we saw that the marginal dependence of the joint signal i and the high ratings j was derived by summing over the mode with low ratings k

$$\widehat{P}_{ij.} = \sum_r \left[\sum_s \sum_t g_{rst} b_{js} c_{\cdot t} \right] a_{ir} \ = \sum_r d_{(j\cdot)r} a_{ir} \ , \tag{23}$$

with $c_{\cdot t} = \sum_k p_{\cdot\cdot k} c_{kt}$, the weighted average of c_{kt} over k. This leads to the conclusion that the marginal co-ordinates for the high ratings can be derived from the overall model by averaging the appropriate components (here: the c_{kt}), and similarly for those of the low ratings by averaging the component scores d_{jkr} over j. Therefore, the $d_{j\cdot r}$ are the co-ordinates on the joint-signal axes for the marginal dependence of the high ratings. These marginal co-ordinates are indicates by $H1, \ldots, H6$ in Figure 5, and similarly $L1, \ldots, L6$ are the co-ordinates for the marginal dependence of the low ratings.

6 Categorical principal component analysis

As explained in Chapter xx, a categorical principal component analysis (PCA) in which all variables have a multiple nominal measurement level is equivalent to multiple correspondence analysis (MCA). Meulman and Heiser (1998) investigated to what extent MCA is able to portray higher-order interactions in contingency tables using a $2 \times 2 \times 2$ data set as an example.

The Wickens-Olzak data are a four-way data set with two numerical or ordinal six-point confidence rating variables and two binary signal variables. If we want to apply the Meulman and Heiser conceptualisation, we could carry out a MCA by performing a categorical PCA on the four variables whithout respecting their measurement levels. However, given that in categorical PCA one can do justice to the measurement levels, it seemed a better choice for the ordinal confidence ratings. Commonly, ordinal variables in the context of categorical principal component analysis are modelled with second-order splines with two interior knots (see e.g. Hastie, Tibshirani, and Friedman, 2001, chap. 5), and this is what was done here.

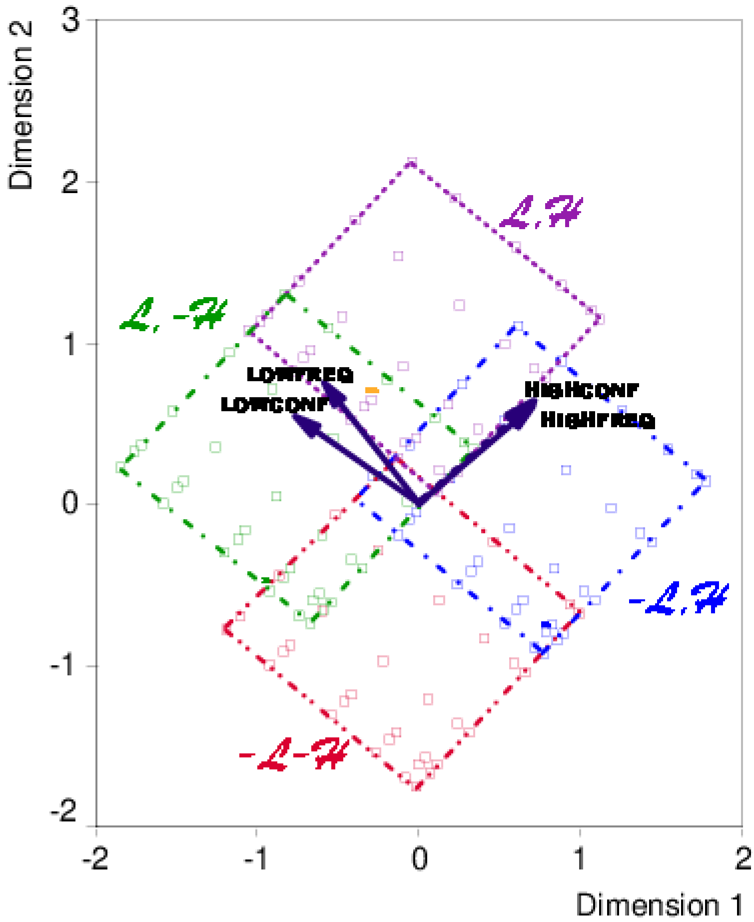


Figure 6: Concurrent Detection Data: Symmetric two-dimensional biplot for categorical PCA with profile points connected per Joint Signal, $\mathcal{L}, \mathcal{H}$ = both high and low signal present, $\cdots$, $-\mathcal{L}, -\mathcal{H}$ both absent; arrows for two ordinal confidence variables and the two binary signal-present variables.

A categorical PCA of the Wickens-Olzak data with two nominal variables (presence or absence of a high or low signal) and two ordinal variables (measure of confidence in presence in high and low signal) was fitted in two dimensions, providing a fit of 75%. Also the results for the three-way table with the joint signal were examined, but these are essentially the same due to the completely balanced design. The grids shown in the biplot of the profiles and the variables (Figure 6) have been made in the same way as in the three-mode CA by connection the confidence ratings in their natural orders. But here we have a separate grid for each signal presence-absence combination (cf. similar figures in Meulman and Heiser (1998). Each point in the plot represents a cell of the four-way contingency table. For instance, the topmost point is the cell with both signals present and both confidence ratings equal to 6. The lowest point is the cell with both signals absent and both confidence ratings equal to 1. In the present graph, the four grids are practically equal and translated versions of each other, illustrating additivity of the quantifications. The results suggest that the theory developed by Meulman and Heiser (1998) can be extended to variables with more and ordered categories as is suggested by them (p. 296), but this subject needs further investigation.

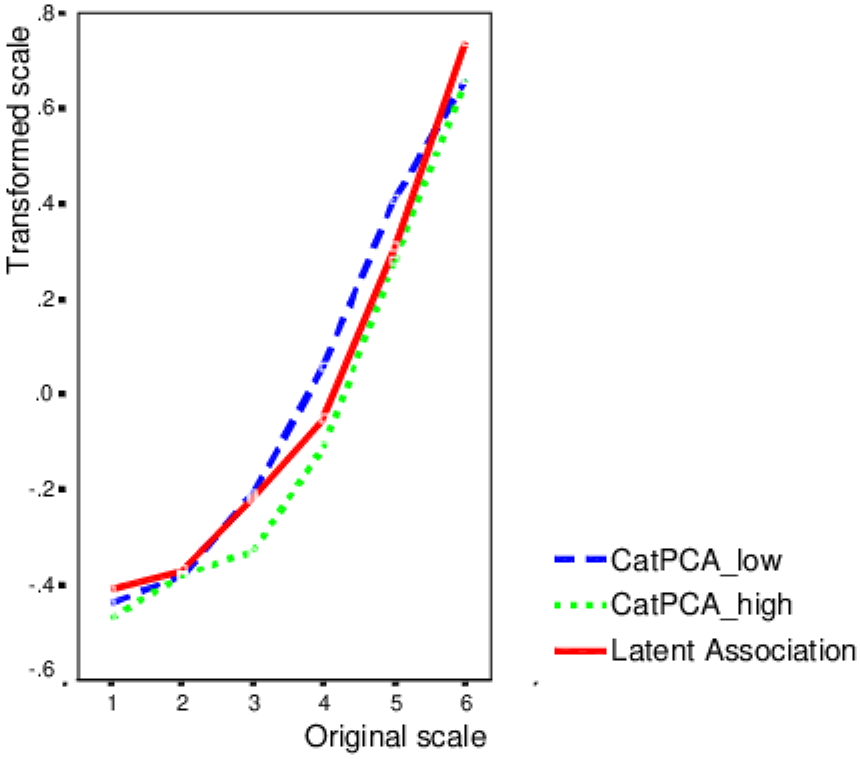


Figure 7: Concurrent Detection Data: Transformed confidence scores versus original scores for the latent association model (joint transformation for both scores; see section 4.2) and for categorical PCA (separate transformations for each confidence score). The transformations for the ordinal scores in the categorical PCA are based on B-spline transformations of degree two and two interior knots (see section 6).

The optimal scaling transformation plots are given in Figure 7 and they show how the original categories of the ordinal confidence variables were transformed from their original numbering of 1 through 6. These graphs are based on the values reported in Table 2, and it can be observed that the categorical PCA parameters are very similar to the jointly estimated values of the latent association model. The values for the joint signal variable are also given in Table 2 where is can be seen that the first dimension of the categorical PCA concurs very well with the scale values from the latent variable association model. These outcomes concur with an observation for the two-way case by Van der Heijden et al. (1994), "The conclusion is that in CA the interaction is decomposed approximately in a log-multiplicative way [..] This close relation between CA and models with log-bilinear terms holds also for more complex models" (p. 106).

7 Discussion and conclusions

The additive and multiplicative approaches to modelling are characterised by first fitting a model to the original frequencies. In the additive case, one can choose an appropriate model by examining the relative contributions of the interaction terms and then combining the chosen ones, while in the multiplicative case one has to fit series of models to assess the necessity of including each interaction term.

When dealing with variables with more than a few categories, straightforwardly fitting the models does not suffice because of the large number of parameters in the interaction terms. To model the interaction terms in both cases, multiplicative decompositions based on two-way and/or three-way SVD, are used. Again the two approaches differ in the way a model is "assembled". Within the additive approach, the complete dependence is decomposed and the contributions of the interactions are assessed to come to an appropriate model. In the multiplicative approach, first a model consisting of several log-additive terms and one or more multiplicative decompositions of log-interactions is constructed which is then assessed with respect to its goodness-of-fit. Several models constructed in this way are then compared to find the preferred model. There is an additional phase of sophistication for the multiplicative model in which latent variables are used to develop more parsimonious and powerful models.

Furthermore, methods have been developed to include restrictions on the estimated parameters for further simplification of the model and improved interpretation, as was shown in the Wickens-Olzak example. Such sophistication is at present not readily available for the additive approach to three-way tables, but several papers have appeared which introduce constraints in (M)CA (e.g. Böckenholt & Takane, 1994; Hwang & Takane, 2002). A sophisticated archeological example for the two-way case was presented by Groenen and Poblome (2002), but three-way extensions of this approach are not yet available.

With the multiplicative decomposition, several graphical representations are available for log-multiplicative models that can aid interpretation, such as graphs and schematic diagrams that represent the model itself, plots of scale values which sometimes are one-dimensional, and thus easy to interpret. In the case of latent variable models, one can fruitfully plot estimates of the means on the latent variables and use these to create insightful plots. With three-way CA all dependence can be displayed in a single plot so that an integrated overall view of the global, marginal and partial dependence can be obtained. However, such displays require careful inspection and it is necessary to develop a fairly detailed understanding of what can and what cannot be read in them.

The main interpretational conclusion of the paper is that both the multiplicative and additive models for three-way tables have much to offer compared to simple significance tests and numerical inspection of the interactions or modelling of them via log-linear models. When the variables have few categories an inspection or log-linear modelling might still be an option, but for larger tables it is becomes almost impossible. If the interactions have a good structure and stochastic assumptions can be met, sophisticated multiplicative modelling can lead to parsimonious models with good interpretability. Moreover, it becomes possible to zoom in to test detailed hypotheses. When the assumptions are more difficult to meet or the structure is messy, the additive approach has much to recommend itself, especially in supplying a grand overview of all aspects of the dependence, and as such it can be even useful when more precise modelling is possible. In the future, with more sophisticated machinery for the additive approach, it may be possible to model in a more refined way as well. For the present example, the log-multiplicative models outpreformed the three-way CA in terms of deviance-degrees of freedom ratio, but why this is so is not yet properly understood.

Categorical PCA (and MCA) can also be used as an alternative when there are problems with log-multiplicative models, and it is expected that it will give similar results. However, a detailed theoretical analysis is required to establish differences and similarities between the models discussed in this chapter.

To summarise our attempt to introduce an interpretational element in the decision whether one should prefer multiplicative or additive modelling, we have extended Darroch's (1974) original table with properties of the models (Table 4). Its top part lists the properties of additive and multiplicative methods as given in Darroch (1974), while the lower part shows the interpretational properties of the models of the models presented here. Unfortunately, our investigations have not led us to a more unequivocal conclusion than Darroch's. This is primarily due to the fact that interpretability is very much data dependent and no model-based arguments can solve content-based limitations or preferences.

8 Software notes

The first author has written a program for three-way CA as part of his general suite of programs for three-way analysis, 3WayPack, information about which can be found on the website of The Three-Mode Company[1]. The suite comes with a Pascal Interface and the program itself is written in Fortran90. An Splus program (with mainly French commentary and output) also exists. It was written by the late André Carlier and distributed via the Laboratoire de Statistique et Probabilité (LSP), Université Paul Sabatier, Toulouse, France. Further information can be obtained from the LSP website[2].

[1]http://three-mode.leidenuniv.nl/

[2]http://www.lsp.ups-tlse.fr/index.html; search for MULTIDIM

Table 4: Properties of multiplicative and additive models after Darroch

Property	Multiplicative Definition		Additive Definition
	Log-linear	Log-multiplicative	Thee-mode CA
Partition properties	No	No	Yes
Closest to independence	Yes	Yes	Yes
Conditional independence as special case	Yes	Yes	No
Distributional equivalence	No	No	Yes
Subtable invariance	Yes	Yes	No
No constraints on marginal probabilities	Yes	Yes	No
Statistical tests	Yes	Yes	No
Estimation	Usually possible	Can be difficult	Always possible
Modelling of data	Exploratory	Confirmatory	Exploratory
Ease of interpretation	Difficult	Simple given model found	Good overview; can be complex

Log-multiplicative modelling (including latent variables) was carried out with ℓ_{EM}(Jeroen Vermunt, University of Tilburg, Tilburg, The Netherlands)[3]. The output and details of the log-multiplicative models of the Wickens-Olzak data as presented in the paper by Anderson (2002) can found on her website[4].

Categorical PCA was carried out with software developed by the Leiden Data Theory Group as implemented in the program module *Categories* of SPSS 11.0[5] (Meulman, Heiser, & SPSS Inc., 2001).

Acknowledgements

This chapter was written while the first author was Fellow-in-Residence at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS). This work was supported by grant #SES-0351175 from the National Science Foundation awarded to the second author.

[3] http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html
[4] http://www.psych.uiuc.edu/~canderso/abstracts/lem/list.html
[5] http://www.spss.com/spssbi/categories/index.htm

References

Anderson, C.J. (1996). The analysis of three-way contingency tables by three-mode association models. *Psychometrika*, **61**, 465–483.

Anderson, C.J. (2002). Analysis of multivariate frequency data by graphical models and generalizations of the multidimensional row-column association model. *Psychological Methods*, **7**, 446–467.

Anderson, C.J. & Böckenholt, U. (2000). *RC* regression models for polytomous variables. *Psychometrika*, **65**, 479–509.

Anderson, C.J. & Vermunt, J.K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, **30**, 81–121.

Becker, M.P. (1989). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association*, **84**, 1014–1019.

Carlier, A. & Kroonenberg, P.M. (1996). Decompositions and biplots in three-way correspondence analysis, *Psychometrika*, **61**, 355–373.

Carlier, A. & Kroonenberg, P.M. (1998). The case of the French cantons: An application of three-way correspondence analysis. In J. Blasius and M.J. Greenacre (eds), *Visualization of Categorical Data*, pp. 253–275. New York: Academic Press.

Choulakian, V. (1988a). Exploratory analysis of contingency tables by loglinear formulations and generalizations of correspondence analysis. *Psychometrika*, **53**, 235–250. Errata., 593.

Choulakian, V. (1988b). Analyse factorielle des correspondances de tableaux multiples [Factorial correspondence analysis of multiway tables]. *Revue de Statistique Appliquée*, **36**, 33–41.

Clogg, C.C. (1986). Statistical modeling versus singular value decomposition. *International Statistical Review*, **54**, 284–288.

Clogg, C.C. & Shihadeh, E.S. (1994). *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Darroch, J.N. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika*, **61**, 207–214.

Dequier, A. (1974). Notes sur les tables de contingence entre trois caractères [Notes on three-way contingency tables]. Unpublished Document Nr. 1033. Arcueil, France: C.I.R.O.

Edwards, D. (2000). *Introduction to Graphical Modelling*, 2nd edition. New York: Springer-Verlag.

Escoufier, Y. (1982). L'analyse des tableaux de contingence simples et multiples [The analysis of simple and multiple contingency tables]. *Metron*, **40**, 53–77.

Fowlkes, E.B., Freeny, A.E. & Landwehr, J.M. (1988). Evaluating logistic models for large contingency tables. *Journal of the American Statistical Association*, **83**, 611–622.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537–552.

Goodman, L. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Annals of Statistics*, **13**, 10–69.

Goodman, L.A. (1996). A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association*, **91**, 408–428.

Green, M.E. (1989). In discussion of Van der Heijden, P.G.M., de Falguerolles, A., and de Leeuw, J.: A combined approach to contingency table analysis using correspondence analysis and loglinear analysis. *Applied Statistics*, **38**, 249–292.

Greenacre, M.J. (1984). *Theory and Applications Correspondence Analysis*. London: Academic Press.

Groenen, P. & Poblome, J. (2002). Constrained correspondence analysis for seriation in archaeology applied to Sagalassos ceramic tableware. In O. Opitz & M. Schwaiger (eds), *Exploratory data analysis in empirical research. Proceedings of the 25th annual conference of the Gesellschaft für Klassifikation*, pp. 90–97. Heidelberg: Springer.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1–84

Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning. Data mining, Inference, and Prediction.* New York: Springer.

Hwang, H., & Takane, Y. (2002). Generalized constrained multiple correspondence analysis. *Psychometrika*, **67**, 211–224.

Kroonenberg, P.M. (1989). Singular value decompositions of interactions in three-way contingency tables. In R. Coppi and S. Bolasco (eds), *Multiway Data Analysis*, pp. 169–184. Amsterdam: Elsevier.

Lancaster, H.O. (1951). Complex contingency tables treated by the partition of χ^2. *Journal of the Royal Statistical Society B*, **13**, 242–249.

Lauritzen, S.L. & Wermuth, N. (1989). Graphical models for association between variables of which some are qualitative and some quantitative. *The Annals of Statistics, 17*, 31–57.

Meulman, J.J. & Heiser, W.J. (1998). Visual display of interaction in multiway contingency tables by use of homogeneity analysis: The $2 \times 2 \times 2 \times 2$ case. In J. Blasius and M.J. Greenacre (eds), *Visualization of Categorical Data*, pp. 253–275. New York: Academic Press.

Meulman, J.J., Heiser, W.J. & SPSS Inc. (2001). *SPSS Categories 11.0.*. Chicago, IL: SPSS Inc.

Roy, S.N. & Kastenbaum, M.A. (1956). On the hypothesis of 'no interaction' in a multi-way contingency table. *Annals of Mathematical Statistics*, **27**, 749–757.

Tucker, L.R. (1966). Some mathematical notes to three-mode factor analysis. *Psychometrika*, **31**, 279–311.

Van der Heijden, P.G.M., Mooijaart, A. & Takane, Y. (1994). Correspondence analysis and contingency table models. In M.J. Greenacre & J. Blasius (eds), *Correspondence Analysis in the Social Sciences*, pp. 79–111. New York: Academic Press.

Van der Heijden, P.G.M. & Worsley, K.J. (1988). Comment on "Correspondence analysis used complementary to loglinear analysis". *Psychometrika*, **53**, 279–311

Wermuth, N. & Lauritzen, S.L. (1990). Discussion of papers by Edwards, Wermuth and Lauritzen. *Journal of the Royal Statistical Society*, Series B, **52**, 51–72.

Whittaker, J. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, **38**, 278–279.

Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Statistics.* NY: Wiley.

Wickens, T.D. & Olzak, L.A. (1989). The statistical analysis of concurrent detection rating. *Perception & Psychophysics*, **45**, 514–528

Wong, R.S.-K. (2001). Multidimensional association models: A multilinear approach. *Sociological Methods & Research*, **30**, 197–240.

Pieter Kroonenberg is a professor holding the special chair in multivariate analysis, in particular three-mode analysis, at the Department of Education and Child Studies, Leiden University, Leiden, The Netherlands. His major interest is in three-mode analysis in all its facets, but he is also interested in other multivariate data-analytic methods and applying such methods to data from a wide variety of fields. He is presently working on a book on the practice of three-mode analysis. Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands. E-mail: kroonenb@fsw.leidenuniv.nl

Carolyn Anderson is an associate professor in the departments of Educational Psychology, Psychology and Statistics at the University of Illinois at Urbana-Champaign. Her major interests are in the analysis of multivariate categorical data and latent variable models. University of Illinois, 1310 South Sixth Street, MC-708, Champaign, IL, 61820. E-mail: cja@uiuc.edu