



# Log-linear modeling

Alexander von Eye,<sup>1\*</sup> Eun-Young Mun<sup>2</sup> and Patrick Mair<sup>3</sup>

This article describes log-linear models as special cases of generalized linear models. Specifically, log-linear models use a logarithmic link function. Log-linear models are used to examine joint distributions of categorical variables, dependency relations, and association patterns. Three types of log-linear models are discussed, hierarchical models, nonhierarchical models, and nonstandard models. Emphasis is placed on parameter interpretation. It is demonstrated that parameters are best interpretable when they represent the effects specified in the design matrix of the model. Parameter interpretation is illustrated first for a standard hierarchical model, and then for a nonstandard model that includes structural zeros. In a data example, the relationships among race of defendant, race of victim, and death penalty sentence are examined using a log-linear model with all three two-way interactions. Recent developments in log-linear modeling are discussed. © 2011 Wiley Periodicals, Inc.

## How to cite this article:

WIREs Comput Stat 2012, 4:218–223. doi: 10.1002/wics.203

**Keywords:** log-linear modeling; hierarchical models; nonhierarchical models; nonstandard models; parameter interpretation

## INTRODUCTION

Researchers use log-linear models to explain the structure of frequency distributions in bi- or multivariate cross-classifications of categorical variables.<sup>1–3</sup> Goodman<sup>4</sup> discussed three elementary views of log-linear modeling in the context of  $I \times J$  tables. Log-linear models can be used

1. to examine the joint distribution of two variables,
2. to assess the possible dependence of a response variable upon an explanatory or regressor variable, and
3. to study the association between two response variables.

When cross-classifications of more than two variables are studied, these three aims carry over accordingly. In addition, as has been discussed, for example, by Hand and Vinciotti,<sup>5</sup> local models can

be considered,<sup>6</sup> which include only part of a table, exclude part of a table, or contain parameters that focus on only parts of a table.

When the *joint distribution* of variables is modeled, results can be stated in terms of a joint frequency distribution. To give an example, a cross-classification of two variables can be symmetric with respect to the main diagonal of the table.<sup>7</sup> Consider a square,  $I \times I$  cross-classification, with cell probabilities  $\pi_{ij}$ , with  $i, j = 1, \dots, I$ . This cross-classification is axial symmetric if  $\pi_{ij} = \pi_{ji}$ . Other concepts of symmetry have been discussed,<sup>2</sup> and there are many forms of joint distributions.

*Dependency relations* are typically expressed in terms of conditional probabilities, odds ratios, or regression parameters from logit models. Log-linear models can be used equivalently. Path models, moderator models, and mediator models can be estimated.

*Association patterns* are typically expressed in terms of associations or interactions which can involve two or more variables. To analyze associations and interactions, variables are not classified as dependent or independent. All variables have the same status.

This article is structured as follows. First, we identify log-linear models as members of the family

\*Correspondence to: voneye@msu.edu

<sup>1</sup>Michigan State University, Department of Psychology, East Lansing, MI, USA

<sup>2</sup>Rutgers University, Piscataway, NJ, USA

<sup>3</sup>WU Vienna University of Economics and Business, Wien, Austria

of generalized linear models (GLMs). We then discuss types of log-linear models, specifically hierarchical, nonhierarchical, and nonstandard models. In Section 3, we discuss the interpretation of the parameters of log-linear models. In the following three sections, we present residual analysis, data examples, and a brief discussion.

## GENERALIZED LINEAR AND LOG-LINEAR MODELS

The GLM<sup>8,9</sup> is defined by three components:

1. independent random response variates  $Y_1, \dots, Y_n$ ; these variates are assumed to follow a probability distribution from the *exponential family*, with expected value  $E[Y_i] = \mu_i$ , which, in log-linear models, is the mean of the logarithm of the expected cell frequencies; the response variable,  $Y$ , is the first part of the *random component* of the GLM; the second part is the *residual*;
2. a linear predictor that is based on the predictor variables  $X_{i,1}, \dots, X_{i,p-1}$  and the corresponding parameters,  $\beta$ : this is the *systematic component* of a GLM;
3. a monotone *link function*,  $g$ , that relates the linear predictor to the expected response,

$$x_i' \beta = g(\mu_i). \quad (1)$$

Examples of GLM include the *binomial logit model* for binary data, the *Poisson model*, and the *general linear model* for continuous outcome variables.<sup>1,3</sup> The natural link function for the Poisson distribution is the logarithmic function. It results in a linear component

$$\log E[Y_i] = x_i' \beta, \quad (2)$$

where  $x_i$  is the  $i$ th vector of the design matrix  $X$ .

Log-linear models are usually expressed by the parameters involved in a model. For example, a log-linear model for the cross-classification of variables  $A$  and  $B$  which considers only the main effects of both variables or, in other words, the model of independence of  $A$  and  $B$ , is expressed as

$$\log \hat{m} = \lambda + \lambda_r^A + \lambda_s^B, \quad (3)$$

where  $\hat{m} = \mu$ ,  $\lambda$  is the model constant,  $\lambda_r^A$  is the  $r$ th parameter of the  $k_A - 1$  parameters of variable  $A$  (with  $k_A$  indicating the number of categories of  $A$ ), and is

the  $s$ th parameter of the  $k_B - 1$  parameters of variable  $B$  (with  $k_B$  indicating the number of categories of  $B$ ). A model is typically evaluated using model-data discrepancies, and with respect to a base model. One often used base model is the null model,  $\log \hat{m} = \lambda$ . Model comparison can then be performed using the difference between the likelihood ratio coefficients  $G^2$  between the target model and the base model. The coefficient is

$$G^2 = 2 \sum_i m_i \log \frac{m_i}{\hat{m}_i}, \quad (4)$$

where  $m_i$  is the observed frequency for Cell  $i$ ,  $\hat{m}_i$  is the expected cell frequency for this cell, and  $i$  goes over all cells of the cross-classification. When just one model is evaluated, the Pearson  $X^2$  may be the better choice, which is

$$X^2 = \sum_i \frac{(m_i - \hat{m}_i)^2}{\hat{m}_i}. \quad (5)$$

Both  $G^2$  and  $X^2$  are approximately distributed as  $\chi^2$  with  $df = \text{number of cells} - \text{number of estimated parameters}$ .

## TYPES OF LOG-LINEAR MODELS

Three types of log-linear models have been considered.<sup>10</sup> The first and most common type is that of *hierarchical models*. Analogous to ANOVA models, these models include all lower-order terms that are relatives of a higher order interaction. For example, the hierarchical model that includes the three-way interaction  $[ABC]$  also includes all two-way interactions  $[AB]$ ,  $[AC]$ , and  $[BC]$ , as well as all main effects  $[A]$ ,  $[B]$ , and  $[C]$ . In other words, hierarchical log-linear models are created following the *effect hierarchy principle*.<sup>2,11,12</sup>

The second type of log-linear models includes *nonhierarchical models*. These models do not include all of the lower-order interactions. Nonhierarchical models follow the *effect heredity principle*, if they contain an interaction of order  $j$ , and include at least one effect of lower-order.

The third type of models is *nonstandard*. Log-linear models are nonstandard if they include terms that cannot be expressed in terms of hierarchical log-linear models. In addition, covariates and parameter constraints can render a model nonstandard. Examples of covariates include cell-specific means of observed variables and variables used to indicate structural zeros. As is discussed below, parameter interpretation is straightforward in hierarchical models.

Nonstandard models need to be specified such that issues with parameter interpretation are minimized.

Nonhierarchical models can be more parsimonious than hierarchical models. The main reason for this is that they contain fewer parameters because some of the effects that are not of interest to the researchers are not estimated. Nonstandard models can be attractive<sup>3,10,13</sup> because they:

1. Allow researchers to fix any parameter to zero.
2. Allow researchers to set two or more parameters equal.
3. Allow researchers to specify parameters in a fixed ratio to each other.
4. Can result in better models than hierarchical or nonhierarchical models.
5. Allow researchers to take scale characteristics of variables into account; for example, ordinal scales do not need to be treated as if they were at the nominal scale level.<sup>14,15</sup>
6. Allow researchers to take covariates into account.
7. Allow researchers to include custom contrasts in the design matrix.
8. Allow researchers to take structural zeros into account.
9. Can be more parsimonious than hierarchical or nonhierarchical models because they allow researchers to tailor specific hypotheses in a way parallel to planned contrasts in ANOVA.

## PARAMETER INTERPRETATION

The interpretation of parameters in log-linear modeling is performed in three ways. First, one asks questions concerning the magnitude of a parameter. Second, one asks whether a parameter differs from expectation. Third, and in parallel to OLS-estimated parameters in the general linear model, parameter interpretation is based on

$$\lambda = (X'X)^{-1}X' \log m, \quad (6)$$

where  $X$  is the design matrix and  $m$  is the vector of observed cell frequencies. This representation differs from the one for OLS-estimated parameters in that, in the general linear model, the parameters are actually calculated using this equation. In contrast, in log-linear modeling, parameters are not estimated using OLS or WLS. Instead, log-linear modeling uses maximum likelihood (ML) methods for parameter estimation. ML estimators are efficient but can come

with bias. More specifically, under regularity conditions such as:

- The first and second derivatives of the log-likelihood functions are defined,
- The Fisher information matrix,  $X'X/\sigma^2$ , is nonzero,
- The Fisher information matrix is continuous, and a function of the parameters,
- The ML estimator is consistent,

ML estimators are asymptotically optimal,<sup>16</sup> which implies that:

1. ML estimators are asymptotically unbiased (they can come with bias when distributional assumptions are not met).
2. ML estimators are asymptotically efficient (no other asymptotically unbiased estimator comes with a smaller standard error than the ML estimator).
3. ML estimators are asymptotically normally distributed (with a mean vector  $\lambda$  and a covariance matrix equal to the inverse of the Fisher information matrix).

We present two examples of parameter interpretation in log-linear models (from Ref 3). In the first example, we present the parameter interpretation for the main effect model of a  $2 \times 2$  cross-classification,  $\log \hat{m} = \lambda + \lambda^A + \lambda^B$ , where  $A$  is crossed with  $B$ . The design matrix for this model is given in the first matrix on the right hand side of the following equation (effects coding):

$$\begin{pmatrix} \log \hat{m}_{11} \\ \log \hat{m}_{12} \\ \log \hat{m}_{21} \\ \log \hat{m}_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \lambda \\ \lambda^A \\ \lambda^B \end{pmatrix}. \quad (7)$$

Inserting into the above equation gives us the design matrix-based parameter interpretation,

$$\lambda = \frac{1}{4}(\log m_{11} + \log m_{12} + \log m_{21} + \log m_{22}), \quad (8)$$

$$\lambda^A = \frac{1}{4}(\log m_{11} + \log m_{12} - \log m_{21} - \log m_{22}), \quad (9)$$

$$\lambda^B = \frac{1}{4}(\log m_{11} - \log m_{12} + \log m_{21} - \log m_{22}). \quad (10)$$

Clearly, these parameters express exactly the effects that had been specified in the design matrix.

Taking the exponent of  $\lambda^A$  yields

$$\begin{aligned} \exp\left(\frac{1}{4}(\log m_{11} + \log m_{12} - \log m_{21} - \log m_{22})\right) \\ = \frac{4\sqrt{m_{11}}4\sqrt{m_{12}}}{4\sqrt{m_{21}}4\sqrt{m_{22}}}. \end{aligned} \quad (11)$$

Taking the exponent of  $\lambda^B$  yields

$$\begin{aligned} \exp\left(\frac{1}{4}(\log m_{11} - \log m_{12} + \log m_{21} - \log m_{22})\right) \\ = \frac{4\sqrt{m_{11}}4\sqrt{m_{21}}}{4\sqrt{m_{12}}4\sqrt{m_{22}}}. \end{aligned} \quad (12)$$

These two expressions show that even main effect terms have an odds ratio-type interpretation. If there is no main effect,  $\exp(\lambda)$  will approach the value of 1. The more the estimate for  $\exp(\lambda)$  deviates from 1, the stronger a main effect. Naturally, this interpretation does not apply to the constant term, for which one obtains

$$\begin{aligned} \exp\left(\frac{1}{4}(\log m_{11} + \log m_{12} + \log m_{21} + \log m_{22})\right) \\ = 4\sqrt{m_{11}}4\sqrt{m_{21}}4\sqrt{m_{12}}4\sqrt{m_{22}}. \end{aligned} \quad (13)$$

In the second example, we illustrate the problems that arise when a cross-classification contains structural zeros. In contrast to empirical zero cells which, in a particular sample, do not contain any case but, under different conditions, possibly could, structural zeros indicate that, for design or logical reasons, no case can be found for a particular pattern of variable categories. Consider a  $4 \times 4$  cross-classification with structural zeros in cells 4 1, 4 2, and 4 3. The design matrix for the main effect model for this table (effects coding) is

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & -1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 \end{pmatrix}. \quad (14)$$

The first column block in this design matrix contains the vectors for the main effect model. A column vector of 1s, for the constant term, is implied. The second column block contains the three vectors that are used to specify the position of the structural zeros in the table. If only the main effect model is estimated, the parameters are interpretable as for any standard log-linear model of variable independence for a  $4 \times 4$  cross-classification. For example, one obtains for the first contrast for the row variable (in the above design matrix, this contrast is represented in the first column), the parameter estimate with the interpretation

$$\begin{aligned} \lambda_1 = \frac{1}{336}(61 \log m_{11} + 62 \log m_{12} + 67 \log m_{13} \\ + 23 \log m_{14} - 23 \log m_{21} - 22 \log m_{22} \\ - 17 \log m_{23} - 22 \log m_{24} - 20 \log m_{31} \\ - 23 \log m_{32} - 18 \log m_{33} - 23 \log m_{34} \\ - 23 \log m_{41} - 22 \log m_{42} - 17 \log m_{43} \\ - 22 \log m_{44}). \end{aligned} \quad (15)$$

When the model takes the structural zeros into account, this parameter interpretation changes. For the same parameter, one obtains the interpretation

$$\begin{aligned} \lambda_1^* = \frac{1}{6672}(1059 \log m_{11} + 1063 \log m_{12} \\ + 1263 \log m_{13} + 1619 \log m_{14}) \\ - \frac{1}{6672}(609 \log m_{21} + 605 \log m_{22} \\ + 405 \log m_{23} + 49 \log m_{24}) \\ - \frac{1}{1668}(150 \log m_{31} + 152 \log m_{32} \\ + 102 \log m_{33} + 13 \log m_{34}) - \frac{1}{4} \log m_{44}, \end{aligned} \quad (16)$$

when the second column block in the above design matrix is taken into account. Clearly, this representation is not easy to interpret and does not correspond to the contrast expressed in the design matrix.

## RESIDUAL ANALYSIS

The Pearson  $X^2$  and the  $G^2$  statistics presented above serve as overall goodness-of-fit measures for the entire log-linear model. If these measures suggest that the model must be rejected, cell-wise analysis of model-data discrepancies will show the researcher where the model fails in particular. In some instances, that is, for specific log-linear models, large cell-specific discrepancies come with substantive interpretations,

as has been discussed in the context of configural frequency analysis.<sup>17,18</sup> Among the most frequently used residual measures are Pearson's

$$X_i^2 = \frac{(m_i - \hat{m}_i)^2}{\hat{m}_i}, \quad (17)$$

and its standardized version,

$$r_i = \frac{X_i^2}{\sqrt{1 - h_i}}, \quad (18)$$

where  $i$  goes over the cells of the cross-classification, and  $h_i$  is the leverage which is the  $i$ th diagonal entry of the hat matrix,

$$H = W^{\frac{1}{2}} X (X' W X)^{-1} X' W^{\frac{1}{2}}. \quad (19)$$

For Poisson GLMs, the diagonal entries,  $w_{ii}$ , of matrix  $W$  ( $W$  is a diagonal matrix) are the model frequencies  $\hat{m}_i$ .

In addition to identifying where, in a table, the model fails, residuals can be used in residual plots. Plotted against their expectancies, they should reflect a normal distribution when a model fits (note that the individual residual values and the residual plots may be somewhat uninformative when variables are dichotomous and there are only few distinct values, or when the observed cell frequencies are small).

## DATA EXAMPLE

For the following example, we use data that were published by the New York Times on January 8, 2003. For 1311 murder cases on record in Maryland in the years from 1978 to 1999, the variables race of victim ( $V$ ),

race of defendant ( $D$ ), and penalty ( $P$ ) were crossed. The two race variables were coded as 1 = black and 2 = white; the penalty variable was coded as 1 = no death penalty issued and 2 = death penalty issued. In the following analyses, we ask whether penalty can be predicted from race of victim and race of defendant. For comparison, we estimate two models. The first is the main effect model, and the second is the model that includes all two-way interactions (this model is equivalent to a logistic regression model). The comparison models are

$$\log \hat{m} = \lambda + \lambda^P + \lambda^V + \lambda^D, \quad (20)$$

and

$$\log \hat{m} = \lambda + \lambda^P + \lambda^V + \lambda^D + \lambda^{PV} + \lambda^{PD} + \lambda^{VD}. \quad (21)$$

The second of these two models corresponds to a logistic regression model. Table 1 presents results for both models.

For the main effect model of the three variables penalty ( $P$ ), race of victim ( $V$ ), and race of defendant ( $D$ ), one obtains a  $G^2 = 389.07$ . The model is, therefore, rejected ( $df = 4$ ;  $p < 0.01$ ), and we estimate the model with all two-way interactions. For this model, we obtain a  $G^2 = 0.65$ . The model can, therefore, be retained ( $df = 1$ ;  $p = 0.42$ ). In addition to standing for itself, this model is significantly better than the base model ( $\Delta X^2 = 388.42$ ;  $\Delta df = 3$ ;  $p < 0.01$ ). The parameter estimates appear in Table 2.

Table 2 shows that, with the exception of  $\lambda^{VP}$ , all parameters are significant. Considering that each of the variables that span the table is involved in at least one of the interactions, we only interpret the interactions. The formal interpretation of the  $D \times P$  interaction can be found in von Eye et al.<sup>3</sup> Substantively, this

**TABLE 1** | Observed and Expected Frequencies for the Main Effect (Left Column Panel) and All Two-Way Interactions Models (Right Column Panel) Model of the Cross-classification of Penalty ( $P$ ), Race of Victim ( $V$ ), and Race of Defendant ( $D$ ).

$V$	$D$	$P$	Observed	Main Effect Model		All Two-Way Interactions	
				Estimated	Std. Res.	Estimated	Std. Res.
1	1	1	593	441.843	7.191	592.416	0.024
1	1	2	14	28.603	-2.730	14.584	-0.153
1	2	1	284	430.647	-7.068	284.584	-0.035
1	2	2	38	27.880	1.917	37.416	0.095
2	1	1	25	152.617	-10.333	25.584	-0.115
2	1	2	1	9.883	-2.826	0.416	0.905
2	2	1	272	148.812	10.089	271.416	0.035
2	2	2	23	9.633	4.307	23.584	-0.120

Penalty ( $P$ : 1 = no death penalty; 2 = death penalty), race of victim ( $V$ : 1 = black; 2 = white), and race of defendant ( $D$ : 1 = black; 2 = white); Std. Res. =  $\sqrt{X^2}$ .



**TABLE 2** | Parameter Estimates and Significance Tests for the Model with All Two-Way Interactions.

Parameter	Estimate	Std. Error	z	p(z)
$\lambda^V$	0.901	0.081	11.109	<0.01
$\lambda^D$	−0.826	0.090	−9.139	<0.01
$\lambda^P$	1.537	0.083	18.442	<0.01
$\lambda^{DP}$	0.419	0.078	5.391	<0.01
$\lambda^{VP}$	−0.104	0.068	−1.522	0.06
$\lambda^{VD}$	0.774	0.054	14.263	<0.01

interaction suggests that the ratio of Death Penalty to other outcome is different in black defendants than in white defendants. Specifically, the frequencies in Table 2 suggest that the odds of being issued the death penalty are over 4.5 times higher for white defendants than for black defendants. The  $D \times V$  interaction suggests that the odds are over 21.4 times higher that a black defendant is accused of having murdered a black victim than the odds that a white defendant is accused of having murdered a black victim.

## CONCLUSION

Log-linear models are superbly flexible. They allow researchers to test hypotheses about interactions of any order; take covariates into account; consider structural zeros and structural cells (these are cells with fixed frequencies); do justice to scale characteristics; test hypotheses concerning the entire cross-classification, but also parts of it<sup>5</sup>; estimate path models<sup>13</sup>; or estimate latent variable models, even for longitudinal observations.<sup>19</sup> The models allow one to consider symmetric relationships among variables (i.e., associations) as well as asymmetric variable relationships (as in logit models which can always be expressed in terms of log-linear models, but not vice versa). The relationships can be expressed formally using odds ratios. Recent work has shown that parameter interpretation can be used to select models such that only models are considered that come with parameters that reflect the researchers hypotheses, and methods have been devised to specify models that reflect the researchers' hypotheses, which is not always feasible using standard models.<sup>20</sup>

## REFERENCES

- Agresti A. *Categorical Data Analysis*. 2nd ed. New York: John Wiley & Sons; 2002.
- Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press; 1975.
- von Eye A, Mair P, Mun EY. *Applied Log-Linear Modeling*. New York: John Wiley & Sons; 2012 (in press).
- Goodman LA. Three elementary views of log-linear models for the analysis of cross-classifications having ordered categories. In: Leinhardt S, ed. *Sociological Methodology*. San Francisco, CA: Jossey-Bass; 1981, 180–195.
- Hand DJ, Vinciotti V. Local versus global models for classification problems: fitting models where it matters. *Am Stat* 2003, 57:124–131.
- Havránek T, Lienert GA. Local and regional versus global contingency testing. *Biometrical J* 1984, 26:483–494.
- von Eye A, Spiel C. Standard and non-standard log-linear symmetry models for measuring change in categorical variables. *Am Stat* 1996, 50:300–305.
- McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall; 1989.
- Dobson A. *An Introduction to Generalized Linear Models*. New York: Chapman & Hall; 2002.
- Mair P, von Eye A. Application scenarios for non-standard log-linear models. *Psychol Methods* 2007, 12:139–156.
- Wu CFJ, Hamada MS. *Experiments: Planning, Analysis and Optimization*. 2nd ed. New York: John Wiley & Sons; 2009.
- von Eye A. Fractional factorial designs in the analysis of categorical data. *InterStat* 2008.
- Vermunt JK. *Log-Linear for Event Histories*. Thousand Oaks, CA: Sage; 1997.
- Agresti A. *Analysis of Ordinal Categorical Data*. New York: John Wiley & Sons; 1984.
- Clogg CC, Shihadeh ES. *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage; 1994.
- Enders C. Direct maximum likelihood. In: Everitt BS, Howell DC, eds. *Encyclopedia of Behavioral Statistics, Vol. 3*. West Sussex: John Wiley & Sons; 2005, 1164–1170.
- von Eye A, Gutiérrez Peña E. Configural frequency analysis: the search for extreme cells. *J Appl Stat* 2004, 31:981–997.
- von Eye A, Mair P, Mun EY. *Advances in Configural Frequency Analysis*. New York: Guilford Press; 2010.
- Vermunt JK, Langeheine R, Böckenholt U. Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *J Ed Behav Stat* 1999, 24:179–207.
- von Eye A, Schuster C, Rogers WM. Modeling synergy using manifest categorical variables. *Int J Behav Dev* 1998, 22:537–557.