



Direction of Effects in Binary Variables

Wolfgang Wiedermann

University of Missouri, Columbia, MO, USA – wiedermannw@missouri.edu

Alexander von Eye

Michigan State University, MI, USA – voneye@msu.edu

Abstract

A new approach to determining direction of effects in binary variables is presented. In cross-tabulations of binary variables, the univariate probability distribution of variables is considered explained if omitting the univariate effects of this variable does not damage model fit. Log-linear models are proposed that allow statements concerning the direction of effect in binary data. The conceptual link between this log-linear approach and existing direction of effect methodology for metric variables is discussed. The new approach is applied to empirical data from research on the development of aggression in adolescence.

Keywords: Direction of effects; binary data; non-uniformity; non-hierarchical log-linear models

1. Introduction

In this work, we propose directional log-linear models for binary variables. These are models to determine direction dependence. In the domain of continuous variables, the development of direction dependence models has made considerable progress since the publication of the work by Dodge and Rousson (2000, 2001). Specifically, suppose that a valid linear relation exists between the two variables X and Y , that X is skewed, that the error distribution is symmetric, and that X is independent of the errors. Then, the equation $Y = \alpha_{YX} + X\beta_{YX} + \varepsilon_{YX}$ suggests that Y is a convolution of a symmetrically distributed error, ε_{YX} , and a possibly asymmetrically distributed predictor variable. By implication, the skewness of the outcome variable will always be less than the skewness of the explanatory variable (for proofs, see Dodge, & Rousson, 2000, 2001; Wiedermann, Hagmann, & von Eye, 2015; Wiedermann & von Eye, 2015a, 2015b). In the following sections, we propose log-linear models for cases in which researchers model the relations between binary variables, specifically one putative predictor, X , and one putative outcome, Y .

2. Bivariate Direction of Effects in Binary Variables

The models to be presented here are parallel to the models proposed by Dodge and Rousson (2000, 2001). Within the log-linear modeling framework, the distributional assumption of non-normality can be redefined as follows: Main effects in log-linear models reflect deviations from uniformity of marginal distributions. When a main effect is non-significant, one concludes that the marginal frequencies are uniformly distributed. It, thus, evinces a skewness of zero (for metric variables), the marginal distribution of the corresponding variable is symmetric. In contrast, when the main effect of a variable exists, the distribution of the marginal frequencies deviates from uniformity. In binary variables, non-uniformity implies asymmetry.

A log-linear model of direction dependence in binary variables that is parallel to the regression model for metric variables includes 1) the main effect term for the putative explanatory variable, and 2) the interaction term for the predictor and the outcome. The main effect term is needed to capture non-uniformity of the predictor categories. The interaction term is needed because the interaction between the two variables must exist. It establishes the relation between X and Y . The focus of the models discussed here is the univariate distribution of Y (focusing on interaction terms is discussed in a different context, see von Eye, & Wiedermann, 2016). If $X \rightarrow Y$ holds, explicitly modeling the univariate distribution of Y is not needed to explain the variability of observed Y values. If the corresponding model fits, X can be considered the explanatory variable. If the model fails, X is not explanatory of the



univariate distribution of Y . In this case, researchers may wish to re-specify the model to test the reverse-direction hypothesis according to which Y is the predictor and X the outcome variable. If the $X \times Y$ interaction is non-significant, neither variable is the explanatory one, because they are unrelated.

The saturated log-linear model for X and Y can be expressed as $m \sim \text{Poisson}(\mu)$ with $\log(\mu) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$; $i = 1, \dots, I$; $j = 1, \dots, J$ (cf. Nelder, 1974) where m are the observed cell frequencies, μ are the model frequencies, and the λ s are the model parameters; the superscripts indicate the variables involved in an effect (single indexes denote main effects and the double index denotes the interaction effect).

The log-linear models to test whether the univariate distribution of a putative outcome can be captured by the univariate distribution of the putative predictor and the predictor-outcome interaction are non-hierarchical (see Mair, & von Eye, 2007). Model selection is based on goodness of fit properties of competing non-hierarchical models. Let $X \rightarrow Y$ describe the true data generating process. Then, the log-linear model that reflects the above characteristics is

$$\log(\mu) = \lambda + \lambda_i^X + \lambda_{ij}^{XY}.$$

In this model, no parameter for the main effect of Y is estimated because the hypothesis is tested that Y has no effect on the joint frequency distribution of X and Y that cannot be explained by X . If this hypothesis can be retained, the parameter for this main effect, λ_j^Y , is redundant. The raw residuals for this model are defined as $\hat{r} = m - \hat{m}$ where $\hat{m} = \exp(\hat{\lambda} + \hat{\lambda}_i^X + \hat{\lambda}_{ij}^{XY})$ and $\hat{\lambda}$, $\hat{\lambda}_i^X$, and $\hat{\lambda}_{ij}^{XY}$ are the parameter estimators of λ , λ_i^X , and λ_{ij}^{XY} . Standardizing the raw residuals results in Pearson residuals, or

$$\hat{r}_p = \frac{\hat{r}}{\sqrt{\exp(\hat{\lambda} + \hat{\lambda}_i^X + \hat{\lambda}_{ij}^{XY})}}$$

The mis-specified log-linear model that erroneously treats X as the outcome and Y as the predictor can be written as

$$\log \mu = \lambda' + \lambda_j^Y + \lambda_{ij}^{XY}$$

where λ' describes the intercept, λ_j^Y denotes the main effect of Y , and λ_{ij}^{XY} denotes the interaction term. The corresponding raw residuals are $\hat{r}' = m - \hat{m}'$ with $\hat{m}' = \exp(\hat{\lambda}' + \hat{\lambda}_j^Y + \hat{\lambda}_{ij}^{XY})$. Inserting the estimated (true) model, $m = \hat{m} + \hat{r}$, the raw residuals of the mis-specified model are $\hat{r}' = \hat{r} + \exp(\hat{\lambda}_{ij}^{XY}) [\exp(\hat{\lambda} + \hat{\lambda}_i^X) - \exp(\hat{\lambda}' + \hat{\lambda}_j^Y)]$, and the corresponding Pearson residuals are

$$\hat{r}'_p = \frac{\hat{r} + \exp(\hat{\lambda}_{ij}^{XY}) [\exp(\hat{\lambda} + \hat{\lambda}_i^X) - \exp(\hat{\lambda}' + \hat{\lambda}_j^Y)]}{\sqrt{\exp(\hat{\lambda}' + \hat{\lambda}_j^Y + \hat{\lambda}_{ij}^{XY})}}$$

In words, the numerator of \hat{r}'_p is the sum of the raw residuals of the true model and a term that increases with the main effect of the true predictor X . The portion of unexplained variability in the mis-specified model increases with λ_i^X and can, thus, be expected to be larger than the portion of unexplained variability in the true model and model selection becomes possible based on model fit statistics of competing log-linear models. Thus, when $\lambda_i^X \neq 0$, the prerequisite of asymmetry is fulfilled. In contrast, when $\lambda_i^X \rightarrow 0$ (in addition to the assumption that $\lambda_j^Y = 0$), the term $\exp(\hat{\lambda}) - \exp(\hat{\lambda}')$ approximates zero, and the two models are no longer distinguishable. To establish model fit, we use, for example, the likelihood ratio (LR) statistic $\chi^2 = 2\sum m \log(m/\hat{m})$, with $df = (\# \text{ cells}) - (\# \text{ non-redundant parameters})$.

Deciding about direction dependence. Based on the above considerations, the decision that X is explanatory for Y can be defended if:

- 1) the model of $X \rightarrow Y$ explains the data well;
- 2) the main effect X , represented by λ_i^X , exists; this result, parallel to Dodge and Rousson's (2000, 2001) assumption of non-normal metric variables, refers to the non-uniformity assumption



- discussed above. It is important to realize that this result is not needed for a satisfactory model fit of the true model. However, this result increases the chance of rejecting the mis-specified model;
- 3) the $X \times Y$ interaction, represented by λ_{ij}^{XY} , exists; this result is needed because it makes little sense to consider one variable explanatory for the other if they are unrelated;
 - 4) the model $Y \rightarrow X$ does not fit. The hypothesis is not supported when the model does not fit or parameters of interest (the main effect of the putative predictor) are non-significant (i.e., $\lambda_j^Y = 0$ but $\lambda_{ij}^{XY} \neq 0$).

When variables are binary, the model that tests the opposite direction of effect is the only alternative. In contrast, it can be just one of a possibly large number of models when variables are polytomous. In addition, when variables are binary and models are hierarchical or non-hierarchical, parameter interpretation based on the relation $\lambda = (W'W)^{-1}W' \log m$ is always possible (von Eye & Mun, 2013), where λ is the vector of parameters, W is the design matrix, and m is the vector of observed cell frequencies (cf. Rindskopf, 1990). The rows of the matrix $(W'W)^{-1}W'$ specify the linear combinations of $\log m$ that lead to the model parameters.

3. Data Example: Development of Aggression in Adolescence

In a study on the development of aggressive behavior, Finkelstein, von Eye, and Preece (1994) asked 67 adolescent girls and 47 boys to self-rate their aggressive behavior at three points in time. For the following example, we use the responses given in 1987 on Verbal Aggression against Adults (V). In addition, we use the Tanner scores (T) of the same year. These are measures of physical pubertal development. On average, the adolescents were, in 1987, 15 years of age.

We model the hypothesis that physical pubertal development is explanatory of self-rated verbal aggression against adults. Thus, $T \rightarrow V$ is the target model (Model 1) and $V \rightarrow T$ is the reverse-direction model (Model 2; we do not really consider this hypothesis; however, this model should fail to describe the data well). We dichotomized the standardized Tanner scores and the standardized ratings of Verbal Aggression against Adults at the mean, with 1 = 'below average pubertal development' / 'below average verbal aggression' and 2 = 'above average pubertal development' / 'above average verbal aggression.'

Model 1

$$\log \mu = \lambda + \lambda^T + \lambda^{TV},$$

posits that physical pubertal development (T) is explanatory of V , where superscript T indicates the Tanner score and V indicates physical pubertal development and subscripts are implied. Model 2, that is, $V \rightarrow T$ posits that V is explanatory of physical pubertal development, or

$$\log \mu = \lambda' + \lambda^V + \lambda^{TV}.$$

The $T \times V$ cross-classification is given in Table 1 along with the expected cell frequencies for these two models.

Table 2. Cross-classification of Tanner scores (T) with Verbal Aggression against Adults (V) with expected cell frequencies from two competing log-linear models.

Cells ($T V$)	Observed Cell Frequencies	Estimated Cell Frequencies	
		Model 1 ($T \rightarrow V$)	Model 2 ($V \rightarrow T$)
1 1	31	29.12	35.93
2 1	27	25.12	22.07
1 2	16	17.89	20.93
2 2	39	40.89	34.07



The goodness of fit statistic, $LR-\chi^2 = 0.552$ ($df = 1$; $p = 0.458$), suggests that Model 1, $T \rightarrow V$, describes the data well. We can, therefore, interpret the parameters. Wald statistics were used for testing the significance of individual model parameters. For the main effect of T , we obtain a standardized estimate of -1.779 ($p = 0.075$). We also obtain a significant parameter for the $T \times V$ interaction (standardized estimate of 2.515 ; $p = 0.012$). In contrast, we conclude that Model 2 fails to describe the data well ($LR-\chi^2 = 3.6821$; $df = 1$; $p = 0.054$), and we refrain from interpreting the parameters. We are well aware that, for a nominal α of 5%, the main effect of T and the LR-test for Model 2, do not allow one to reject the null hypotheses in the strict sense. However, in the present context, we are in favor of interpreting 1) T as sufficiently non-uniform ($= 0.584$) and 2) the difference in LR-test statistics (3.68 versus 0.54) as substantively meaningful. Therefore, we propose the following decision about the directionality of effects:

- 1) The model $T \rightarrow V$, it locates physical pubertal development on the explanatory and verbal aggression against adults on the outcome side describes the data well. The alternative but implausible model is $V \rightarrow T$, it places verbal aggression against adults on the explanatory and physical pubertal development on the outcome side, cannot be used to describe the data well.
- 2) The main effect of T suggest a deviation from uniformity of the marginal distribution of physical pubertal development (58.4% of the respondents showed a Tanner score above the sample average).
- 3) The significance of the $T \times V$ interaction fulfills the requirement that the two variables in the model be related to each other.
- 4) Finally, and based on the specification of the model, we conclude that the main effect of T that is included in the model is sufficient to represent the marginal distribution of the outcome variable, V . Omitting λ^V does not result in significantly worse model fit. In sum, for respondents who are 15 years of age, physical pubertal development can be interpreted as explanatory of verbal aggression against adults.

4. Discussion

To study the effect of an explanatory variable on a putative dichotomous outcome variable, binary logistic regression models are routinely applied; polytomous outcomes are routinely analyzed using multinomial logistic regressions (Hosmer & Lemeshow, 2002). However, these methods have an important limitation in that they provide no information concerning direction of relations. This fact can easily be illustrated using log-linear models. Consider the dichotomous explanatory variable X and the dichotomous outcome variable Y . The logit-model has the form $logit(P[Y = 1]) = \alpha_{YX} + X\beta_{YX}$, where α_{YX} is the intercept and β_{YX} is the slope parameter. The equivalent hierarchical log-linear model, $\log(\mu) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$, (Agresti 2013; von Eye & Mun, 2013) shows that main effects of both, predictor and outcome, are implied in the logit-model. Now, the reversed-path logit-model is $logit(P[X = 1]) = \alpha_{XY} + Y\beta_{XY}$. The equivalent log-linear equivalent remains, unchanged, $\log(\mu) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$. Evidently, standard logit-models as well as hierarchical log-linear models are not suited to test directional hypotheses and to derive conclusions about the status of variables as explanatory and outcome variables. Using these type of models requires a priori decisions concerning the status of variables, decisions that are based on theory alone. In contrast, the directional log-linear models proposed here are non-hierarchical. Instead of explaining the observed relation between variables, the models explain the marginal distributions of outcome variables.

Directional log-linear models restrict the parameter space of marginal effects and, thus, require strong statistical assumptions to establish directionality statements. Future work is needed to derive directional log-linear models in which $\lambda_i^X \neq 0$, $\lambda_j^Y \neq 0$, $\lambda_{ij}^{XY} \neq 0$, and $\lambda_i^X \neq \lambda_j^Y$. This effort can be seen as parallel to the one in which models to test direction dependence hypotheses of the form $\Delta(\gamma) > 0$ with $\Delta(\gamma) = |\gamma_X| - |\gamma_Y|$ are developed in the metric case in which both skewnesses, γ_X and γ_Y , can deviate from zero.



Multivariate extensions. As in the bivariate case, direction dependence analysis requires the assumptions 1) that an a priori grouping of variables exists into the classes of putative explanatory and outcome variables, and 2) that, under certain conditions, the putative explanatory variables and the putative outcome variables change roles. Models for binary variables that allow one to test direction dependence hypotheses include the following terms:

- 1) All main effects of all explanatory variables;
- 2) All interactions among the explanatory variables; these terms are needed because the model does not make any assumptions about relations among the explanatory variables; therefore, an arbitrary interaction pattern for these variables is admissible (this can be viewed as parallel to log-linear models that are specified to be equivalent to logistic regression models; cf. Agresti, 2013; von Eye & Mun, 2013); the model is, thus, saturated in the explanatory variables;
- 3) All possible interactions among the explanatory and the outcome variables; these terms are required because direction of effect hypotheses can be retained only if the explanatory variables are related to the outcome variables; note that at least one of these interactions must exist for a hypothesis about directional dependence to be retained; however, to be considered explanatory, every putative explanatory variable must interact with at least one outcome variable; similarly, every putative outcome variable must interact with at least one explanatory variable, at any interaction level;
- 4) All possible interactions among the outcome variables; these terms are required for the same reasons as for the interactions among the explanatory variables; however, in contrast to 2), the model is not saturated in the outcome variables, because none of their main effects is part of the model.

To illustrate, consider the I putative explanatory variables X_i , and the K putative outcome variables Y_k . The log-linear model that contains the four required elements listed above is

$$\log \mu = \lambda + \sum_i \lambda^{X_i} + \sum_{i,j,\dots} \lambda^{X_{i,\dots} \times X_{j,\dots} \times \dots} + \sum_{i,j,\dots,k,l,\dots} \lambda^{X_{i,j,\dots} \times Y_{k,l,\dots} \times \dots} + \sum_{k,l,\dots} \lambda^{Y_{k,\dots} \times Y_{l,\dots} \times \dots},$$

where λ is the model constant, the terms with one X in the superscript are the main effects of the explanatory variables, the terms with more than one X in the superscript are the interactions among the explanatory variables, the terms with both X and Y in the superscript are the interactions among the explanatory and the outcome variables, and the terms with more than one Y in the superscript are the interactions among the outcome variables. The main effects of the outcome variables are omitted because the model is specified to test the hypothesis that the direction of effect goes from the X to the Y variables. If the model fits, the univariate distributions of the outcome variables can be considered explained by the explanatory variables. The following model can be used to test the hypothesis that the direction of effect is reversed:

$$\log \mu = \lambda + \sum_{i,j,\dots} \lambda^{X_{i,\dots} \times X_{j,\dots} \times \dots} + \sum_{i,j,\dots,k,l,\dots} \lambda^{X_{i,j,\dots} \times Y_{k,l,\dots} \times \dots} + \sum_k \lambda^{Y_k} + \sum_{k,l,\dots} \lambda^{Y_{k,\dots} \times Y_{l,\dots} \times \dots}.$$

Here, the main effects of X are not part of the model. The interaction terms are the same in both models.

When direction of research is the goal of analysis, the methods proposed here will help researchers make decisions about which variable can be considered explanatory and which can be considered as an outcome. Without a solid theoretical background, such decisions, however, can be futile. Still, they can, given the necessary caution, be used to establish first empirical evidence in the process of development of causal theories.



References

- Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New York: Wiley.
- Dodge, Y., & Rousson, V. (2000). Direction dependence in a regression line. *Communications in Statistics: Theory and Methods*, 29, 1957-1972.
- Dodge, Y., & Rousson, V. (2001). On asymmetric properties of the correlation coefficient in the regression setting. *The American Statistician*, 55, 51-54.
- Finkelstein, J. W., von Eye, A., & Preece, M. A. (1994). The relationship between aggressive behavior and puberty in normal adolescents: A longitudinal study. *Journal of Adolescent Health*, 15, 319-326.
- Hosmer, D. W., & Lemeshow, S. (2002). *Applied logistic regression* (2nd ed.). New York: Wiley.
- Mair, P., & von Eye, A. (2007). Application scenarios for nonstandard log-linear models. *Psychological Methods*, 12, 139-156.
- Nelder, J. A. (1974). Log linear models for contingency tables: A generalization of classical least squares. *Journal of the Royal Statistical Society C*, 23, 323-329.
- Rindskopf, D. (1990). Nonstandard log-linear models. *Psychological Bulletin*, 108, 150-162.
- von Eye, A., & Mun, E.-Y. (2013). *Log-linear modeling - Concepts, interpretation and applications*. New York: Wiley.
- von Eye, A., & Wiedermann, W. (2016). Direction of effects in categorical variables: A structural perspective. In W. Wiedermann & A. von Eye (eds.), *Statistics and Causality: Methods for applied empirical research*, pp. 107-130. Hoboken, NJ: Wiley & Sons.
- Wiedermann, W., & von Eye, A. (2015a). Direction-dependence analysis: A confirmatory approach for testing directional theories. *International Journal of Behavioral Development*, 39, 570-580.
- Wiedermann, W., & von Eye, A. (2015b). Direction of effects in multiple linear regression model. *Multivariate Behavioral Research*, 50, 23-40.
- Wiedermann, W., Hagmann, M., & Eye, A. (2015). Significance tests to determine the direction of effects in linear regression models. *British Journal of Mathematical and Statistical Psychology*, 68, 116-141.