Psychological Test and Assessment Modeling, Volume 61, 2019 (4), 435-451

# Log-linear and configural analysis of tree structures

*Alexander von Eye*[1], *Wolfgang Wiedermann*[2] *& Stefan von Weber*[3]


## Abstract

In this article, two methods are proposed for the analysis of existing tree structures. The first method, log-linear modeling, is variable-oriented. The goals pursued with a log-linear analysis of tree structures concern the modeling of patterns of a sequence of decisions. Starting from a base model that is a standard hierarchical model, it adds special contrasts that represent the decisions that are made when progressing through the tree. Goal of analysis is to fit a model of the sequence of decisions. These contrasts render the model non-standard. The method of Schuster transformation is needed to create a design matrix with contrasts that can be interpreted as intended. The second method, Configural Frequency Analysis (CFA), is person-oriented. Starting from the final, fitting log-linear model, it removes the special contrasts from the model and asks whether there are individual patterns (configurations) that are discrepant from the base model and, thus represent the sequence of decisions. These patterns can be interpreted as types and antitypes in a CFA sense. In a data example, students' decisions and life satisfaction are examined. The intertwining of the proposed methods and possible extensions are discussed.


Key words: modeling decision trees, nonstandard log-linear models, Configural Frequency Analysis, parameter interpretation, Schuster transformation

[1] *Correspondence concerning this article should be addressed to:* Alexander von Eye, PhD, Michigan State University, 190 Allée du Nouveau Monde, 34000 Montpellier, France; email: voneye@msu.edu

[2] University of Missouri, Columbia

[3] University of Furtwangen

In this article, we discuss log-linear and configural approaches to the analysis of hypotheses that are compatible with tree-structures in cross-classified categorical data. These two approaches are intertwined at the level of estimation of expected cell frequencies. They both use log-linear modeling methods for this step of analysis. The approaches differ in that they are used to answer different questions. This difference is described in detail in later sections of this article. Specifically, this article is structured as follows. First we describe the underlying modeling perspective that is inherent in both, the log-linear and the configural methods. This is done in comparison to methods that help create classification trees. Second, we introduce log-linear modeling as a method that helps analyze existing trees. Third, the configural analysis of existing trees is introduced. These sections are followed by a real-world data example and a discussion.

## Classification trees and modeling

Tree-based models, developed both for categorical (classification trees) and continuous variables (regression trees; see Breiman, Friedman, Stone, Olshen, 1984), use known characteristics of cases to create profiles and to predict membership of these cases in classes. These classes are created either based on data or theory-guided, e.g., with respect to a targeted variable. In the present article, we focus on modeling categorical variables. The decisions concerning class membership are binary, that is, a case does versus does not possess a particular characteristic. The number of categories per variable is not necessarily two. Examples of decision tree methods and decision tree predictions can be found in standard textbooks on categorical data analysis (e.g., Agresti, 2013) and encyclopedias (e.g., Daniels, & Shi, 2005). Applications can be found in the literature on machine learning (e.g., Quinlan, 1986) and data mining (e.g., Murthy, 1998).

One characteristic that many methods share that create classification trees is that they are non-statistical in the sense that they do not allow one to test hypotheses. Instead, a series of binary decisions is made such that, given a profile of characteristics, probabilities are determined as to whether cases possess additional characteristics. Alternatively, methods of information theory have been used to determine the loss of information that comes with each decision. This approach has been used, for instance, in life-tree research (Müller, 1985; von Eye, & Brandtstädter, 1981). At the end, individual cases are described by a profile of characteristics or, equivalently, are members of the class that possesses this profile of characteristics. Variable relations are rarely taken into account, and neither are a priori probabilities (see Agresti, 2013; Daniels, & Shi, 2005).

When compared with logistic regression, other binary classification methods, or log-linear modeling, classification trees can easily be comprehended even by practitioners who possess no more than basic understanding of statistics. When, however, variable relations are considered important, hypotheses are to be tested, or the probability of individual class sizes is scrutinized, statistical methods are used. In this article, we discuss such methods.

To illustrate the modeling scenario, consider the three binary variables, A, B, and C. Crossed, they constitute a $2 \times 2 \times 2$ contingency table. Now, let the first decision be that a case possesses characteristic $a_2$, the second decision points at $b_2$, and the third at $c_2$. That is, this case exhibits the profile $a_2$ $b_2$ $c_2$. The corresponding classification tree is depicted in Figure 1.

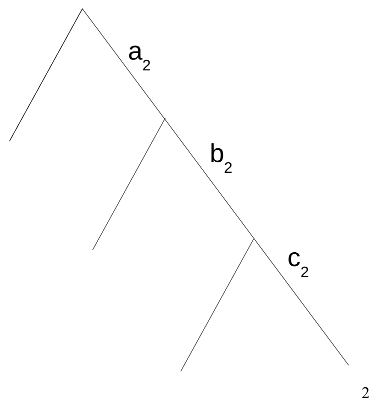


**Figure 1:**
Classification tree for the profile $a_2$ $b_2$ $c_2$

It should be noted that, in the present context, the order of variables and, therefore, the order of decisions is mostly irrelevant. Changing the order leads to the same profile and the same models. This issue is taken up again in the discussion section.

The resulting profile, $a_2$ $b_2$ $c_2$, is one out of eight possible profiles. The remaining seven profiles are not included in Figure 1. In contrast, Table 1 does display these eight profiles.

**Table 1:**
The eight three-variable profiles of the variables A, B, and C

| Variable | | | Targeted Profile |
|---|---|---|---|
| A | B | C | |
| 1 | 1 | 1 | |
| 1 | 1 | 2 | |
| 1 | 2 | 1 | |
| 1 | 2 | 2 | |
| 2 | 1 | 1 | |
| 2 | 1 | 2 | |
| 2 | 2 | 1 | |
| 2 | 2 | 2 | x |

The *targeted profile*, that is, the one depicted in Figure 1, is marked in Table 1. We now ask which effects must be included in a log-linear model that tests hypotheses that reflect the three decisions that result in Profile $a_2$ $b_2$ $c_2$.

## Log-linear modeling of decision trees

To develop a log-linear model for decision trees, we first continue the example given in the last section, that is, the example with the profile $a_2$ $b_2$ $c_2$. We use the model $log\ m = X\lambda$, where $m$ is the vector of model frequencies, $X$ is the design matrix, and λ is the parameter vector. In the example, three decisions were made. In the first, cases were found to exhibit Characteristic $a_2$. In the second, cases who exhibit Characteristic $a_2$ were found to also exhibit Characteristic $b_2$, and in the third, cases who exhibit Characteristics $a_2$ and $b_2$ were found to also exhibit Characteristic $c_2$, that is, in all, Profile $a_2$ $b_2$ $c_2$. The design matrix for this model, therefore, needs contrast variables that represent these three decisions. Table 2 displays the complete design matrix for this model.

**Table 2:**
Design matrix for log-linear model of the decision tree that leads to Profile $a_2$ $b_2$ $c_2$
(effect coding; constant vector implied)

| Profiles | Main Effects for Variables | | | Effects for Decisions | |
|---|---|---|---|---|---|
| | A/D1 | B | C | D2 | D3 |
| 111 | 1 | 1 | 1 | 0 | 0 |
| 112 | 1 | 1 | -1 | 0 | 0 |
| 121 | 1 | -1 | 1 | 0 | 0 |
| 122 | 1 | -1 | -1 | 0 | 0 |
| 211 | -1 | 1 | 1 | -1 | 0 |
| 212 | -1 | 1 | -1 | -1 | 0 |
| 221 | -1 | -1 | 1 | 1 | -1 |
| 222 | -1 | -1 | -1 | 1 | 1 |

1. *One vector each for the decisions* that are made during the climb through the decision tree; a decision leads to the assignment of a case to a category of the decision tree.

The first column of Table 2 contains the variable categories of the eight possible profiles in the present example. The following three columns contain the effect coding vectors for the main effects of the three variables A, B, and C. The first of these vectors contrasts the two categories of Variable A. It is, thus, equivalent with the one that reflects the first decision. Because of this equivalence, this vector is needed only once in the model specification. When the first variable has more than two categories, contrasts can be set up such that the first decision reflects one of the contrasts between the categories. Columns 5 and 6 reflect the second and the third decisions. The second decision is made only for those cases who exhibit the category that results from the first decision. Therefore, the patterns 1 1 1, 1 1 2, 1 2 1, and 1 2 2 are ignored in the comparison of frequencies. Accordingly, the third decision is made only for those cases that exhibit the categories that result from the first two decisions, that is, Patterns 2 2 1 and 2 2 2.

In general, a log-linear model that is specified to test whether hypotheses that are compatible with particular outcomes in decision trees can be retained, contains the following effects (the constant vector is implied):

1. *All main effect vectors*. Decision tree models usually do not propose hypotheses concerning marginal probabilities. Therefore, any marginal distribution is admissible (except those in which only one category has non-zero probability). These distributions must be taken into account by the model, because the model must not fail just because marginal probabilities are not reproduced. When, as in the above example, main effects are equivalent with effects that represent classification decisions, redundancy must be avoided and effect vectors must not be repeated.
2. *One vector each for the decisions* that are made during the climb through the decision tree; a decision leads to the assignment of a case to a category of the variable that the decision is made about.

Variable interactions are not part of a decision tree model. The reason for this is that the decisions reflect such interactions. Exceptions can be cases in which particular dependencies need to be taken into account, as in decisions that are made concerning observations over time. Please note that the vectors that represent the decisions, render the log-linear models non-standard (Mair, & von Eye, 2007) because these terms cannot be expressed in terms of hierarchical log-linear models. This issue is discussed in more detail in the next section.

**Identification and parameter interpretation**. Let $c_i$ be the number of categories of the $i$th variable, with $i = 1, \ldots, I$. Then, the number of main effects included in a decision tree model equals $\sum_i (c_i - 1)$. The number of decision tree-specific effect vectors is $I$. Then, when no covariates, special variable interactions, serial or temporal dependencies are taken into account (see Stemmler, von Eye, & Wiedermann, 2015), the total number of vectors in the decision tree design matrix is $V = 1 + \sum_i (c_i - 1) + I - 1 = \sum_i (c_i - 1) + I$. The number 1 is subtracted because of the redundancy of the first decision with one of the main effect vectors of the first binary variable. $V$ cannot exceed the number of cells, that is $V \le \prod_i c_i$. $V$ equals the number of cells only in 2 × 2 tables. Therefore, in tables that are larger than 2 × 2, decision tree models will always be identified and parameters can be estimated, when there are no covariates to be included in the model, or vectors that represent special variable interactions or dependencies.

Parameter interpretation of log-linear decision tree models can be complicated. As is well known (see Mair, & von Eye, 2007; von Eye, & Mun, 2013), log-linear parameters can be interpreted as coded in the design matrix only when the design matrix is orthogonal. The columns of orthogonal design matrices are uncorrelated. In the above example, and, in general, in log-linear tree models, this is not the case. Table 3 displays the correlation matrix for the columns of the design matrix in Table 2.

**Table 3:**
Pearson correlation matrix of the columns of the design matrix in Table 2

| | A/D1 | B | C | D2 | D3 |
|---|---|---|---|---|---|
| A/D1 | 1.000 | | | | |
| B | 0.000 | 1.000 | | | |
| C | 0.000 | 0.000 | 1.000 | | |
| D2 | 0.000 | -0.707 | 0.000 | 1.000 | |
| D3 | 0.000 | 0.000 | -0.500 | 0.000 | 1.000 |

Evidently, the design matrix in Table 2 is not orthogonal. Specifically, the second and the third decision tree effects correlate with main effects B and C, respectively. In general, decision tree design matrices for log-linear modeling will not be orthogonal because hypotheses are nested within categories. More specifically, beginning with the second decision, all following decisions are nested within categories of prior variables.

Two approaches have been discussed to overcome the issue of non-orthogonality (see von Eye, & Mun, 2013; von Eye, & Wiedermann, 2016). The first approach involves a stepwise procedure. Decision tree-specific effects are added to the model without these effects, one after the other. The resulting design matrices will be non-orthogonal, but they are nested. Therefore, the contribution made by adding a particular effect can be estimated and evaluated statistically, using, e.g., the likelihood ratio tests for nested models. The second approach involves performing the *Schuster transformation* (von Eye, Schuster, & Rogers, 1998). In brief, a more general formulation of this transformation can be found in von Eye & Wiedermann (2018 b), this transformation is based on the fact that the relation between the design matrix $X$ and the parameters is

$$\lambda = \left(X'X\right)^{-1} X'\mu$$

(see von Eye, & Mun, 2013), with $\mu = \log \hat{m}$ . When the columns of $X$ are not orthogonal, parameter interpretation will not correspond to the effects specified in $X$. The Schuster transformation has the effect that the transformed $X$ results in parameter estimates that can be interpreted as intended, that is, as originally specified in the design matrix $X$.

For the transformation, consider a matrix $H$ whose transpose, $H'$, has the same dimensions as $X$, and $H\mu$ specifies the contrasts of interest. Let, for the column spaces, $C$, of $X$ and $H$, the following relation hold:

$$C\left(X\right) = C\left(H\right),$$

and let

$$X^* = H'\left(HH'\right)^{-1}.$$

Substituting $X^*$ for $X$ when estimating $\lambda$, results in

$$\lambda = \left(\left(X^*\right)' X^*\right)^{-1} \left(X^*\right)' \mu.$$

In words, when $X^*$ is used instead of the originally specified matrix, $X$, the parameters correspond to the contrasts specified in $X$. Note that the Schuster transformation does not alter model fit. That is, model fit statistics of Schuster-transformed models are identical to the corresponding fit statistics obtained from non-transformed models. The only precondition for the application of this version of the Schuster transformation is that the inverse of $HH'$ exists.

## Configural analysis of decision trees

Configural Frequency Analysis (CFA; Lienert, 1968; Lienert, & Krauth, 1975; von Eye, 2002; von Eye, & Gutiérrez Peña, 2004) takes a person-oriented perspective. Specifically, instead of asking questions concerning main effects and interactions of variables, CFA asks whether there exist individual patterns (configurations), that (a) significantly contradict a base model and (b) can be interpreted with respect to a hypothesis of interest. Typically, these patterns emerge in an exploratory search, but CFA can also be applied in explanatory contexts (see, for example, von Eye, & Wiedermann, 2018 a).

Here, we propose an approach to CFA that is related to von Eye and Mair's (2008 a, b) functional CFA. In this approach vectors are systematically entered/removed from a design matrix to determine the effects that cause CFA types and antitypes to emerge. In the present context, consider the log-linear model $log\ m = X\lambda + T\lambda$, where the term $X\lambda$ represents the base model (it includes the constant vector), that is, a standard hierarchical log-linear model, and the term $T\lambda$ represents the special contrasts that are introduced to model the climb through the decision tree.

For the following considerations, suppose that the model $log\ m = X\lambda + T\lambda$ describes the existing decision tree in a satisfactory way. Now, if the term $T\lambda$ is really needed to explain the frequency distribution, removing it will result in an ill-fitting model. This lack of fit can come about because many cells show moderate deviations of the observed from the expected cell frequencies. However, it can also come about because individual cells – in the extreme cases, this will be just one cell – exhibit particularly strong deviations. These extreme cells are called *CFA types* if they contain more cases than expected, and they are called *CFA antitypes* when they contain fewer cases than expected (von Eye, & Gutiérrez Peña, 2004).

In the analysis of decision trees, types and antitypes play a particular role. Types describe decision sequences that occur more often than expected under the model that does not represent the decision sequence. Antitypes describe decision sequences that occur less often than expected under this model. In the next section, we illustrate the application of log-linear modeling and configural analysis to a real-world data set.

It is important to realize that types and antitypes from a simple base model, e.g., a main effect model, cannot be interpreted as reflecting sequences of decisions. There can be many reasons why these types and antitypes emerge. However, when these types and antitypes disappear when the decision tree-specific vectors are inserted into the design matrix, then, an interpretation in the context of a decision tree becomes possible (cf. von Eye and Mair's functional CFA, 2008 a, b).

## Data example

In the following data example, we re-analyze the data presented by von Eye and Brandtstädter (1981). A sample of 207 students at two German universities responded to a questionnaire that included demographic questions and questions concerning subjective control. The following four items are used in the following analysis:

1. Gender (G; 1 = male, 2 = female);
2. subjective length of time needed to decide on major (E; 1 = short, 2 = long);
3. satisfaction with current life (Z; 1 = satisfied, 2 = not satisfied); and
4. feeling of empowerment (K; 1 = above average, 2 = below average).

As von Eye and Brandtstädter (1981), we propose the following (weak) temporal order:

$$G \geq K \geq E \geq Z,$$

where the symbol ≥ can be read as ' is located in time before or is contemporaneous with.' In the following analyses, we focus on those 16 paths that involve all four variables. The total number of paths is 28. This includes those paths that are defined by fewer than four variables. We first present log-linear and, then, configural analyses. The observed univariate frequencies are G: 77, 130; K: 110, 97; E: 102, 105; and Z: 54, 153. The final profile that is hypothesized is *female students who needed subjectively short time spans to make a decision about their major are rather dissatisfied with their current life and express below average feelings of empowerment*.

## Log-linear decision tree models

To create a baseline for comparison, we first estimate a log-linear main effect model, that is, the model $log\ m = \lambda + \lambda^G + \lambda^K + \lambda^E + \lambda^Z$. Table 4 displays the cross-classification of the four variables, G × K × E × Z, and cell-specific results of the log-linear main effect model.

The overall goodness of fit values for this model appear in Table 6, in the 'main effect model' row. The estimated cell frequencies and the standardized residuals are displayed in Table 4, above.

**Table 4:**
G × K × E × Z cross-classification and estimates from log-linear main effect

| G K E Z | $m$ | $\hat{m}$ | $z$ |
|---|---|---|---|
| 1 1 1 1 | 5.000 | 5.260 | -0.113 |
| 1 1 1 2 | 11.000 | 14.903 | -1.011 |
| 1 1 2 1 | 14.000 | 5.414 | 3.690 |
| 1 1 2 2 | 15.000 | 15.341 | -0.087 |
| 1 2 1 1 | 2.000 | 4.638 | -1.225 |
| 1 2 1 2 | 17.000 | 13.141 | 1.064 |
| 1 2 2 1 | 2.000 | 4.775 | -1.270 |
| 1 2 2 2 | 11.000 | 13.528 | -0.687 |
| 2 1 1 1 | 7.000 | 8.880 | -0.631 |
| 2 1 1 2 | 27.000 | 25.160 | 0.367 |
| 2 1 2 1 | 7.000 | 9.141 | -0.708 |
| 2 1 2 2 | 24.000 | 25.900 | -0.373 |
| 2 2 1 1 | 5.000 | 7.831 | -1.012 |
| 2 2 1 2 | 28.000 | 22.187 | 1.234 |
| 2 2 2 1 | 12.000 | 8.061 | 1.387 |
| 2 2 2 2 | 20.000 | 22.839 | -0.594 |

In comparison with the main effect model, we also estimate a first decision tree effect model. We call it the 'simple decision tree model.' This model links time-adjacent elements in the weak temporal order proposed above. The design matrix for this model is given in Table 5.

Table 6 shows that neither the main effect nor the simple decision tree model can be retained. In fact, under the decision tree model, although it comes with numerically smaller goodness-of-fit values, the observed frequency distribution seems to be even less likely than under the main effect model (due to the smaller number of degrees of freedom). To obtain a model that fits the data, we now also include three of the six possible two-way interactions. For the selection of the interactions to be included, we take into account that Gender as well as the first decision might be related to all of the following decisions, not just the ones next to them in the decision tree. The interactions G × E, G × Z, and G × K are, therefore, included in the model. Including these terms results in the overall goodness-of-fit values in the bottom row of Table 5. Clearly, this model fits the frequency distribution in Table 4 well. In addition, this model represents a significant improvement over the simple decision tree model ($\Delta(L^2) = 10.27$; $\Delta(df) = 3$; $p = 0.016$).

**Table 5:**
Design matrix for simple decision tree model of the G × K × E × Z cross-classification (effect coding; constant vector implied)

| Variable | Effects | | | | | | |
|---|---|---|---|---|---|---|---|
| GKEZ | G/D1 | K | E | Z | D2 | D3 | D4 |
| 1111 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 1112 | 1 | 1 | 1 | -1 | 0 | 0 | 0 |
| 1121 | 1 | 1 | -1 | 1 | 0 | 0 | 0 |
| 1122 | 1 | 1 | -1 | -1 | 0 | 0 | 0 |
| 1211 | 1 | -1 | 1 | 1 | 0 | 0 | 0 |
| 1212 | 1 | -1 | 1 | -1 | 0 | 0 | 0 |
| 1221 | 1 | -1 | -1 | 1 | 0 | 0 | 0 |
| 1222 | 1 | -1 | -1 | -1 | 0 | 0 | 0 |
| 2111 | -1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 2112 | -1 | 1 | 1 | -1 | 1 | 0 | 0 |
| 2121 | -1 | 1 | -1 | 1 | 1 | 0 | 0 |
| 2122 | -1 | 1 | -1 | -1 | 1 | 0 | 0 |
| 2211 | -1 | -1 | 1 | 1 | -1 | 1 | 0 |
| 2212 | -1 | -1 | 1 | -1 | -1 | 1 | 0 |
| 2221 | -1 | -1 | -1 | 1 | -1 | -1 | 1 |
| 2222 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |

**Table 6:**
Goodness-of-fit for main effect and decision tree models of the G × K × E × Z cross-classification

| Model | $X^2$ | $p(X^2)$ | $L^2$ | $p(L^2)$ | df |
|---|---|---|---|---|---|
| Main effect | 25.37 | 0.008 | 21.97 | 0.025 | 11 |
| Simple Decision tree model | 21.40 | 0.006 | 18.09 | 0.021 | 8 |
| Decision tree plus G × E, G × Z, and G × K | 8.16 | 0.147 | 7.82 | 0.155 | 5 |

The cross-products of orthogonal vectors are zero. Instead of being zero, the cross-product of the fourth and the seventh columns of the matrix is +2. We, therefore, have to perform the Schuster transformation with the matrix in Table 5 and, then, re-estimate the model. Table 7 displays the matrix from Table 5 after Schuster transformation (for other applications of the Schuster transformation, see, e.g., von Eye and Wiedermann, 2017, 2018 b).

**Table 7:**
Design matrix for simple decision tree model of the G × K × E × Z cross-classification after Schuster transformation

| Variable | Effects | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| GKEZ | Constant | G/D1 | K | E | Z | D2 | D3 | D4 |
| 1111 | 1/16 | 1/16 | 1/8 | 1/12 | 1/14 | -(1/8) | -(1/12) | -(1/14) |
| 1112 | 1/16 | 1/16 | 1/8 | 1/12 | -(1/14) | -(1/8) | -(1/12) | 1/14 |
| 1121 | 1/16 | 1/16 | 1/8 | -(1/12) | 1/14 | -(1/8) | 1/12 | -(1/14) |
| 1122 | 1/16 | 1/16 | 1/8 | -(1/12) | -(1/14) | -(1/8) | 1/12 | 1/14 |
| 1211 | 1/16 | 1/16 | -(1/8) | 1/12 | 1/14 | -1/8 | -(1/12) | -(1/14) |
| 1212 | 1/16 | 1/16 | -(1/8) | 1/12 | -(1/14) | 1/8 | -(1/12) | 1/14 |
| 1221 | 1/16 | 1/16 | -(1/8) | -(1/12) | 1/14 | 1/8 | 1/12 | -(1/14) |
| 1222 | 1/16 | 1/16 | -(1/8) | -(1/12) | -(1/14) | 1/8 | 1/12 | 1/14 |
| 2111 | 1/16 | -(1/16) | 0 | 1/12 | 1/14 | 1/8 | -(1/12) | -(1/14) |
| 2112 | 1/16 | -(1/16) | 0 | 1/12 | -(1/14) | 1/8 | -(1/12) | 1/14 |
| 2121 | 1/16 | -(1/16) | 0 | -(1/12) | 1/14 | 1/8 | 1/12 | -(1/14) |
| 2122 | 1/16 | -(1/16) | 0 | (1/12) | -(1/14) | 1/8 | 1/12 | 1/14 |
| 2211 | 1/16 | -(1/16) | 0 | 0 | 1/14 | -(1/8) | 1/4 | -(1/14) |
| 2212 | 1/16 | -(1/16) | 0 | 0 | -(1/14) | -(1/8) | 1/4 | 1/14 |
| 2221 | 1/16 | -(1/16) | 0 | 0 | 0 | -(1/8) | -(1/4) | 1/2 |
| 2222 | 1/16 | -(1/16) | 0 | 0 | 0 | -(1/8) | -(1/4) | -(1/2) |

When this matrix is used, the parameters can be interpreted as given in Table 8.

Evidently, the rows in Table 8 represent the design matrix in Table 5 perfectly (please notice that the constant vector was implied in Table 5, but it is included in the Schuster transformation and it appears in the first row of Table 8). To give an example, the last column vector in Table 5 suggests that the last two entries of the frequency table in Table 4 be contrasted. In Table 8, the last row reflects exactly that.

We now re-estimate the decision tree model that includes the longer-term two-way interactions, that is, G × E, G × Z, and K × Z, but after a Schuster transformation of the corresponding design matrix. The overall goodness-of-fit of this model is unchanged ($L^2$ = 7.82; $df$ = 5; $p$ = 0.155). However, now, we can interpret the parameters. We find that only one of the special effect parameters is significant, the third (D4; $z$ = 3.069 $p$ = 0.001). Together, the three decision tree-specific parameters come with a Wald statistic of $X^2$ = 9.27, which, for $df$ = 3 also suggests a significant effect ($p$ = 0.0118). When a model fits and all decision tree-specific parameters are significant, the tree structure can be considered fully supported. When, however, a model fits and only some of the decision tree-specific parameters are significant, partial support of the tree structure hypothe-

**Table 8:**
Parameter interpretation of the design matrix in Table 7

| Parameter | Interpretation |
|---|---|
| Constant | m1111+m1112+m1121+m1122+m1211+m1212+m1221+m1222+m2111+ m2112+m2121+m2122+m2211+m2212+m2221+m2222 |
| G/D1 | m1111+m1112+m1121+m1122+m1211+m1212+m1221+m1222-m2111- m2112-m2121-m2122-m2211-m2212-m2221-m2222 |
| K | m1111+m1112+m1121+m1122-m1211-m1212-m1221-m1222+m2111+ m2112+m2121+m2122-m2211-m2212-m2221-m2222 |
| E | m1111+m1112-m1121-m1122+m1211+m1212-m1221-m1222+m2111+ m2112-m2121-m2122+m2211+m2212-m2221-m2222 |
| Z | m1111-m1112+m1121-m1122+m1211-m1212+m1221-m1222+m2111- m2112+m2121-m2122+m2211-m2212+m2221-m2222 |
| D2 | m2111+m2112+m2121+m2122-m2211-m2212-m2221-m2222 |
| D3 | m2211+m2212-m2221-m2222 |
| D4 | m2221-m2222 |

ses can be discussed. In the present case, the decision tree model is supported as a whole, but this result is carried predominantly by the third decision (D4 in Tables 7 and 8).

We conclude that the weak temporal order that was proposed for the four variables, according to which female students who needed subjectively short time spans to make a decision about their major are rather dissatisfied with their current life and express below average feelings of empowerment does explain significant portions of the variability in the frequency distribution of Table 4, in particular when the interactions G × E, G × Z, and K × Z are part of the model. To better understand the structure of this data set, we now take a configural perspective. Results of this analysis are reported in the next section.

## Configural analysis of a decision tree

In the last section, we realized that after weakening the temporal order of the decision tree model by including all longer-term interactions, we are able to explain the frequency distribution in Table 4 in a satisfactory manner. We now return to the original, the simple decision tree model and ask which individual patterns deviate the most from this model. This question can be answered by Configural Frequency Analysis.

For the following analysis, we perform a standard first order CFA. The base model for this variant of CFA is the log-linear main effect model, identical to the comparison model estimated in the last section. Cells will be evaluated using the z-test, and the significance threshold will be protected using the Holland-Copenhaver procedure.

As was indicated at the beginning of this article, log-linear modeling and configural analysis of tree structures are intertwined. Here, we see why this is the case: the base model for CFA and the comparison model for the development of a fitting log-linear model are the same. The observed and the expected cell frequencies as well as the cell-wise test statistics from Table 4 can, therefore, be used in this section again. From a CFA perspective, this model corresponds to the one that results from removing those vectors that represent the sequence of decisions. It should be noted that, for this analysis, the Schuster transformation is not necessary.

The results in Table 4 suggest that one configuration in particular is responsible for the poor fit of the base model. This is Configuration 1 1 2 1. This configuration describes male students who took a subjectively short time to make a decision on their major, are dissatisfied with their current life, but do feel empowered. Based on the model, 5.41 students had been expected to display this profile, but 14 did display it. This discrepancy comes with a standardized residual of 3.65. The tail probability of this value is more extreme than the Bonferroni-adjusted significance threshold (0.00018 < 0.00278). Configuration 1 1 2 1, thus, constitutes a *type*. That is, it describes more cases than can be defended based on the simple decision tree model.

With respect to the weak order assumption under which we had approached the G × K × E × Z cross-classification, we come to the following conclusion. It is not the case that the configuration *female students who needed subjectively short time spans to make a decision about their major are rather dissatisfied with their current life and express below average feelings of empowerment* describes the sequence of decision best. Instead, it is, assuming that the interactions G × E, G × Z, and K × Z do not exist, *male students who took a subjectively short time to make a decision on their major, are dissatisfied with their current life, but do feel empowered*. Almost three times as many students displayed this pattern as was expected under the null hypothesis that there is no typical decision tree in the G × K × E × Z cross-classification.

If only the terms are removed from the log-linear model that represent the climb through the decision tree, and the three interaction terms G × E, G × Z, and K × Z are left in the model, the model does still not fit. We calculate a $L^2 = 14.4694$, which, for $df = 8$, comes with a tail probability of $p(L^2) = 0.0124$. Table 9 displays the cell-specific results of this analysis.

As can be seen in Table 9, taking into account the interactions G × K, G × E, and G × Z weakens the type that had been found when these interactions were not considered. The standardized residual of 2.287 comes with a tail probability of 0.011. This value is still extreme enough to support the notion of a CFA type when Hommels' modified procedure for alpha protection is used (Hommel, 1989; Hommel, Lehmacher, & Perli, 1985). Interestingly, a second type emerges in this analysis. This type is constituted by Cell 2 2 2 1. Its standardized residual is 2.548 ($p = 0.005$). This type describes *female students who took a subjectively long time to make a decision concerning their major, are dissatisfied with their current life, but do feel empowered*. 5.84 students had been expected to display this pattern, but 12 did display it. This result shows that the simple, weak temporal order that was imposed in the model hypothesis does enable researchers to describe

**Table 9:**
G × K × E × Z cross-classification and estimates from the log-linear model that includes all two-way interactions with Gender

| G K E Z | $m$ | $\hat{m}$ | $z$ |
|---|---|---|---|
| 1 1 1 1 | 5.000 | 6.389 | -0.549 |
| 1 1 1 2 | 11.000 | 12.353 | -0.385 |
| 1 1 2 1 | 14.000 | 7.667 | 2.287 |
| 1 1 2 2 | 15.000 | 14.824 | 0.046 |
| 1 2 1 1 | 2.000 | 4.066 | -1.024 |
| 1 2 1 2 | 17.000 | 12.193 | 1.377 |
| 1 2 2 1 | 2.000 | 4.879 | -1.303 |
| 1 2 2 2 | 11.000 | 14.631 | -0.949 |
| 2 1 1 1 | 7.000 | 9.764 | -0.884 |
| 2 1 1 2 | 27.000 | 25.678 | 0.261 |
| 2 1 2 1 | 7.000 | 9.181 | -0.720 |
| 2 1 2 2 | 24.000 | 24.145 | -0.030 |
| 2 2 1 1 | 5.000 | 6.213 | -0.487 |
| 2 2 1 2 | 28.000 | 25.345 | 0.527 |
| 2 2 2 1 | 12.000 | 5.842 | 2.548 |
| 2 2 2 2 | 20.000 | 23.832 | -0.785 |

the frequency distribution of the G × K × E × Z cross-classification, but, evidently, there is more than one path through this decision tree. CFA suggests that there are two paths, 1 1 2 1 and 2 2 2 1.

## Discussion

In this article, two approaches are proposed to the analysis of existing tree structures in a categorical variable context. In contrast to classification attempts such as information theory, the methods proposed here examine decision trees from a modeling perspective. This perspective allows researchers to test hypotheses, to incorporate covariates, to take into consideration prior probabilities, and to estimate expected frequencies.

The methods proposed here for decision tree analysis differ in the perspectives they take when testing hypotheses. Log-linear modeling focuses on variable relations. In contrast, CFA is person-oriented. That is, it allows one to inspect individual patterns, that is, configurations, or groups of those.

As was discussed at the beginning of this article and in the last section, the intertwining of the two approaches becomes evident in two aspects. First, the models specified in the context of log-linear modeling also imply a person-oriented perspective. At least the last vector of the design matrix contrasts two configurations one with the other. CFA does not compare events but asks whether individuals or groups of individuals are observed at rates that differ from expectancy. Still, these are two versions of person-oriented analysis.

Second, CFA also implies a variable-oriented element. This is the way in which the CFA base model is specified. In such a base model, variable relations are included that are not of interest to the researcher. CFA types and antitypes can, therefore emerge only when relations exist beyond those considered in the base model. Still, the main focus of CFA is person-oriented, as types and antitypes constitute local violations of a base model.

Extensions of the models proposed here can be conceptualized along the lines of questions that are often posed by scholars. For example, one can ask whether series of decisions are unchanged across populations. Using the variables in the real-world data example in the last section, one could ask whether the series of decisions is the same across students of different disciplines. The models needed to answer this question are log-linear or configural moderator models.

Covariates can be incorporated in both, log-linear and configural base models. For example, one could ask whether disposable income results in different expected frequencies and also in different conclusions concerning the decision tree.

A third set of models is conceivable in which certain decisions are left open. This can imply one of at least two modeling options. The first is that the categories for a particular decision are hypothesized to come with equal probabilities. The second is that a particular decision is skipped. In the first case, a vector is included in the design matrix in which the entries for the equal probability categories are modeled so that they do not differ, given the other effects in the model. In the second case, vectors need to be specified that represent decisions that are farther apart from each other than just one step in the decision tree.

*Limitations*. Two limitations of the proposed methods are of note. First, tables must be of a certain size to be analyzable with the log-linear base model proposed here. In the simplest case, and as was indicated in the above section on *Identification*, the design matrix for the log-linear decision tree model contains $V = 1 + \sum_i (c_i - 1) + I - 1$ vectors. To be not-saturated, the cross-classification to be analyzed needs, therefore, at the least $V + 1$ cells. When covariates are part of the model, the table needs $V + 1$ cells plus the number of vectors needed to represent the covariates. To give an example, when two decisions are examined for three binary variables, the cross-classification contains eight cells, and $V = 6$, without covariates. This model is viable, even when a covariate is considered that consumes one degree of freedom. In contrast, a 2 × 2 table is not viable.

The second limitation of note is conceptual in nature. Implicit in the hypotheses discussed here is a sequence of decisions. In the example with the students, we posited the weak temporal order G ≥ K ≥ E ≥ Z, and we analyzed the G × K × E × Z cross-classification. Changing the temporal order or the order of variables in the cross-classification would have no effect on the results. Therefore, although the temporal order of decisions might be defensible, theoretically, it is not entirely part of the statistical analysis proposed here. The fact that time-adjacent transitions are modeled does help, but only to a certain degree. It helps because it models the entire sequence of decisions. However, when this order is reversed, results will stay the same. Therefore, when the whole order of decisions is of critical importance, researchers need to establish the position of the first (or the last) of decisions in the sequence before results from analyses as the ones proposed here can be interpreted as supporting all of the order.

*In sum*, this article presents two methods for the analysis of a priori existing decision trees. These methods allow researchers to answer variable-oriented questions concerning the structure of such trees as well as event-based, person-oriented questions. In addition, these methods differ by being predominantly explanatory (log-linear modeling) versus exploratory (CFA).

## References

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). *Classification and regression trees*. Milton Park, UK: Taylor and Francis.

Daniels. R., & Shi, D. (2005). Tree models. In B.S. Everitt, & D.C. Howell (eds.), *Encyclopedia of statistics in behavioral science* (*pp*. 2059 – 2060). Chichester, UK: Wiley.

Hommel, G. (1989). A comparison of two modified Bonferroni procedures. *Biometrika*, *75*, 383 – 386.

Hommel, G., Lehmacher, W., & Perli, H.-G. (1985). Residuenanalyse des Unabhängigkeitsmodells zweier kategorialer Variablen. In J. Jesdinsky, & J. Trampisch (eds.), *Prognose- und Entscheidungsfindung in der Medizin* (*pp* 494 – 503). Berlin: Springer.

Lienert, G. A. (1968). *Die "Konfigurationsfrequenzanalyse" als Klassifikationsmethode in der klinischen Psychologie.* Paper presented at the 26. Kongress der Deutschen Gesellschaft für Psychologie in Tübingen 1968.

Lienert, G.A., & Krauth, J. (1975). Configural Frequency Analysis as a statistical tool for defining types. *Educational and Psychological Measurement*, *35*, 231 – 238.

Mair, P., & von Eye, A. (2007). Application Scenarios for Nonstandard Log-Linear Models. *Psychological Methods*, *12*, 139 - 156.

Müller, N. (1985). Real structure modelling: a methodology for the description of large-scale social units. *Social Science Information*, *24*, 603 – 624.

Murthy S. (1998). Automatic construction of decision trees from data: A multidisciplinary survey. *Data Mining and Knowledge Discovery*, *2*, 345 – 389.

Quinlan, J. R., (1986). Introduction of Decision Trees. *Machine Learning*, *1*, 81-106.

Stemmler, M., von Eye, A., & Wiedermann, W. (Eds.) (2015). *Dependent data in social sciences research: Forms, issues, and methods of analysis*. New York: Springer.

von Eye, A., & Brandtstädter, J. (1981). Lebensbäume als entwicklungspsychologische Modelle: Ansätze zur Analyse von Lebensereignissequenzen. *Trierer Psychologische Berichte*, *8*(2).

von Eye, A. (2002). *Configural Frequency Analysis - Methods, Models, and Applications*. Mahwah, NJ: Lawrence Erlbaum.

von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. *Journal of Applied Statistics*, *31*, 981 – 997.

von Eye, A., & Mair, P. (2008). Functional Configural Frequency Analysis: Explaining types and antitypes. *Bulletin de la Société des Sciences Médicales*, *Luxembourg*, *144*, 35 – 52. (a)

von Eye, A., & Mair, P. (2008). A functional approach to Configural Frequency Analysis. *Austrian Journal of Statistics*, *37*, 161 – 173. (b)

von Eye, A., & Mun, E.-Y. (2013). *Log-linear modeling - Concepts, Interpretation and Applications*. New York: Wiley.

von Eye, A., Schuster, C., & Rogers, W.M. (1998). Modeling synergy using manifest categorical variables. *International Journal of Behavioral Development*, *22*, 537 - 557.

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.

von Eye, A., & Wiedermann, W. (2017). Direction of effects in categorical variables - looking inside the table. *Journal of Person-Oriented Research*, *3*, 11 - 27.

von Eye, A., Wiedermann, W. (2018). Locating event-based causal effects: A configural persepctive. *Integrative Psychological and Behavioral Science, 52*, 307-330. (a)

von Eye, A., Wiedermann, W. (2018). Reciprocal relations in binary variables. (in preparation) (b)