

Analyzing tree structures with Configural Frequency Analysis and the R-package *confreq*

Mark Stemmler¹, Jörg-Henrik Heine², Susanne Wallner³

Abstract

This article describes the use of Configural Frequency Analysis (CFA; von Eye, 2002) for detecting a tree structure in data that was analyzed with the SPSS module Answer Tree. At the same time, the application of the R package *confreq* (Heine, 2019) is demonstrated. The data example is taken from a longitudinal study on deviant and delinquent behavior in juveniles. The study is called 'Chances and Risks in the Life-Course (CURL)' (Reinecke et al., 2013). Here, the offender status of juveniles in 11th grade was predicted by several risk factors measured two years earlier. The CHAID analysis detected a two level tree with illegal substance use as the first node and with the nodes antisocial attitudes and living in a single parent home on the second level. Functional Configural Frequency Analysis (von Eye & Mair, 2008) was applied to model the data structure in several steps: First, a main effects model is performed, then a significant configuration addressing non-offending in juveniles is blanked out and finally, effect vectors modeling the decision tree sequence were introduced into the design matrix. This resulted in a fitting model which suggests that the tree structure was correctly modeled. The theoretical background for this data application can be found in this Special Issue (see von Eye, Wiedermann, & von Weber, 2019). Syntax and outputs based on *confreq* are presented to create a hands-on application.

Keywords: Configural Frequency Analysis (CFA), R-package *confreq*, tree-based methods, antisocial behavior, log-linear modeling (LLM)

¹Correspondence concerning this article should be addressed to: Mark Stemmler | Friedrich-Alexander University Erlangen-Nürnberg, Department of Psychology, Nägelsbachstrasse 49c | D-91054 Erlangen, Germany; email: mark.stemmler@fau.de

²Technical University Munich (TUM), School of Education, Germany

³Friedrich-Alexander University Erlangen-Nürnberg, Department of Psychology, Germany

Introduction and outline

This article demonstrates the use of Configural Frequency Analysis (CFA; von Eye, 2002) for detecting tree structures analyzed with CHAID. It is basically an illustration and application based on the more theoretical chapter by von Eye et al. (2019) in this Special Issue. The data analyses use the SPSS module Answer Tree and a dataset from a longitudinal study on deviant and delinquent behavior in juveniles. A special version of CFA, that is functional CFA (fCFA; von Eye & Mair, 2008), is explained and demonstrated to model the decision tree structure in terms of overfrequent cells while using the design matrix to model certain effects. At the same time, the application of the R package *confreq* (Heine, 2019) is demonstrated.

Brief description of CHAID and CART

Multiple categorical variables, one which defined as a dependent variable, can be analyzed using a statistical technique known as CHAID. This method is part of the SPSS module Answer Tree. "... *CHAID partitions the data into mutually exclusive, exhaustive, subsets that best describe the dependent variable*" (Kass, 1980, p. 119). "CHAID" stands for **Chi-Square Automatic Interaction Detection** (Lautsch & Plichta, 2005). It is frequently used in marketing research, where consumer decisions to buy or not to buy a particular product are modelled. The use of this SPSS module results in a graphical illustration called a tree diagram (Lautsch & Thöle, 2005).

CHAID is a stepwise procedure: First, the program searches for the best predictor of the dependent variable by partitioning the data. The χ^2 -statistic is used to pick the best predictor, in a fashion similar to how the F -value is used in stepwise regression to decide which variable should be included or excluded (Kass, 1980). Then the selection of significant predictors goes on until the best combination of categories, as measured by the Bonferoni adjusted p -value of the corresponding χ^2 or likelihood ratio test is found. The selection procedure continues until the resulting tree has reached a certain size which can be specified by the user, through defining the size of a parenting and child node (e.g., 50 and 30). The branches of the resulting tree can be interpreted in the sense of a Configural Frequency Analysis (CFA) as a Type or as an overfrequent cell (cf. von Eye et al., 2019, in this Special Issue).

One advantage of a tree diagram is the ability to investigate the effects of several independent variables on a dependent variable without any restriction on the level of measurement. If the variables are measured on the interval level CART (Classification and Regression Trees) is used for model building. With CART the best cut-offs are generated as part of the procedure (Haughton & Oulabi, 1997). CHAID is applicable to situations where all variables, the dependent and the independent variables are categorical. The dependent variable can involve more than two responses.

In this article, we first create a tree diagram using CHAID; subsequently Configural Frequency Analysis (CFA; von Eye, 2002) is applied to investigate the branches of the tree and the underlying structure of the data in further detail. For all CFA analyses, we will use the *R*-package *confreq* (Heine, 2019). The dependent variable in the following example was the status of a juvenile as 'offender versus a non-offender'. The independent variables were risk factors associated with committing an offense in one cohort of German students (9th graders).

Methodological details on the study data

Study subjects

The selection of independent variables, here possible risk factors, for the following analyses was based on a publication with the title 'risk factors for the development of antisocial behavior in childhood and youth' (German translation: Risikofaktoren für die Entwicklung dissozialen Verhaltens in der Kindheit und Jugend; Stemmler, Wallner, & Link, 2018). In this chapter, which included an introduction to the concept of risk factors and their characteristics, also data from the project 'Chances and Risks in the Life-Course' (CURL) were analyzed (for more information on the project the reader may turn to Reinecke et al., 2013; Reinecke, Stemmler, & Wittenberg, 2016). For the following analyses all bivariate associations were included that showed any significant correlation between the risk factors and antisocial behavior. The design of the study and a detailed description of the measures is described in Weiss and Wallner (2019).

The risk factors for time 1 (t_1) were taken from the 'Cracow Instrument' (Corrado, 2002), an assessment tool for the identification of risk factors for an early onset of antisocial behavior and violence for children and youth. The German translation, adaptation, and validation of the original Canadian psychometric instrument was part of the Erlangen-Nuremberg Development and Prevention Study (ENDPS; Lösel, Beelmann, Stemmler, & Jaursch, 2006; Lösel & Stemmler, 2012; Wallner, 2007). The Cracow Instrument encompasses risk factors for dissociality in five different domains: *Environment*: social and early medical risk factors; *Individual*: characteristics of the person, sorted by biological and psychological criteria as well as functional criteria like, for instance, low coping abilities; *Family*: external factors such as family characteristics, like for instance parenting behavior; *Interventions*: a list of early interventions, which tried to influence the developmental course of the individual to counterbalance an antisocial development; and *Externalizing Behavior*: of the child or juvenile under investigation. In this study, only those risk factors of the Cracow Instrument were included that were measured or available in our study at t_1 (cf. Stemmler et al., 2018; Wallner & Stemmler, 2006). The assessed constructs that were used to operationalize the risk factors can be found in Weiss and Wallner (2019) (see also Arnis, 2015). Each risk factor was dichotomized ('no/medium risk' versus 'high risk' with regard to the respective risk factor). Table 1 presents the list of risk factor variables that were selected for the tree-based analyses.

Table 1: Significant bivariate associations between the dichotomized risk factors and delinquency (last year's prevalence) at time 1 (t_1)

Delinquency (Prevalence of Last Year)	
Risk for Antisociality:	Youth (9th Grade)
<i>Environment</i>	
community disorganization	+
delinquent peers	++
low school cohesion	+
<i>Individual</i>	
hyperactivity/inattention	+
dark personality traits	+
antisocial attitudes	++
problems at school	++
<i>Family</i>	
corporal punishment	+
single-parent home	+
ineffective parenting	++
death of a parent	+
attachment deficits	+
<i>Externalizing Behavior</i>	
general behavior problems	++
violence/aggression	+++
alcohol and illegal drug use	++

Note. The plus signs were based on odds ratios. '+': $0 < d \leq .20$; '++': $.20 \leq d < .50$; '+++': $d \leq .50$.

Delinquent behavior encompassed behavior that is forbidden under the penal law; this includes property crime, vandalism and violence (cf. Reinecke et al., 2016).

An 'offender' was defined as a person that reported having committed at least one crime in the past year. A 'non-offender' was a study person who had not committed any crime in the past year. A detailed description on the assessment of delinquency can be found in Weiss and Wallner (2019). In a first step, odds ratios between the selected risk factors and last year's delinquency were calculated: The relationship between the binary risk factor for antisociality ('no/medium risk' versus 'high risk') and the dichotomized delinquency status ('non-offender' versus 'offender') for the juveniles (9th grade) were examined on cross-sectional analyses. Table 1 also lists the effect sizes in terms of Cohens' d (Cohen, 1988).

Data example

Analyses for 9th grade students

The data includes 1345 students from 9th grade; 494 (36.7 %) of whom had reported at least one crime in the last year. The longitudinal data for t_1 to t_3 (time gap: two years) included 606 juveniles with complete data with regard to delinquency. Of the 494 offenders included at t_1 , 195 (ca. 40 %) remained in the longitudinal data file, and about one third (33.8 %) reported of having committed another crime at t_3 . The predictors listed in Table 1 were entered in the CHAID analysis to predict the outcome, Offender Status at t_3 (i.e., 11th grade). Figure 1 presents the resulting tree diagram for the 9th graders. The tree diagram consists of two levels with four branches. The group

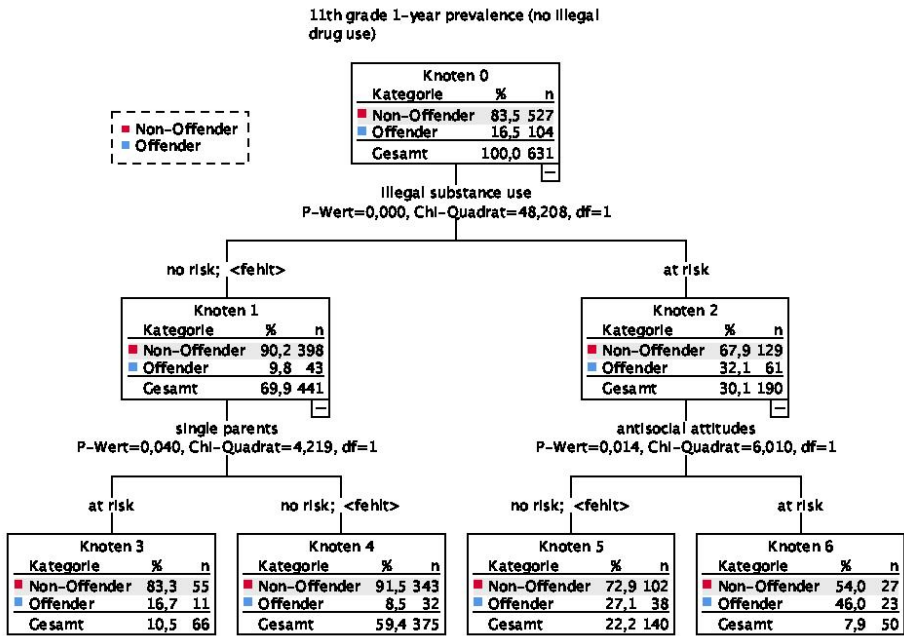


Figure 1: Graphical illustration of resulting tree diagram for the older cohort predicting the offender status at 11th grade

of offenders in 11th grade can be predicted by three risk factors in 9th grade: illegal substance use, antisocial attitudes and living in a single parent home. In a second step, we performed Configurational Frequency Analysis (CFA) based on a model of independence which consists of four main effects. Configurational Frequency Analysis (CFA; von Eye, 2002) is as person-centered analytic approach for the analysis of frequencies in multi-

way contingency tables. In such cross-tabulations individuals, animals or objects are grouped into disjunct categories based on their respective patterns or configurations. The comparison of observed frequencies (o_{ijk}) with expected cell frequencies (e_{ijk}), patterns (configurations), that occur significantly more often than predicted or expected under the null hypothesis are called *Types*. Configurations occurring significantly less often than predicted under the null hypothesis are called *Antitypes* (cf. Stemmler & Heine, 2017). Log-linear modeling (LLM) and CFA are closely related (von Eye & Mun, 2013). LLM is used to identify the structure among the categorical variables. It parametrizes the distribution of cell frequencies, or to put it differently, the logarithms of cell frequencies, in terms of main effects and interactions. Each CFA base model can be expressed as a LLM model, but not the other way round.

In our example, the CFA base model (i.e., the main effect model) can be expressed by the corresponding log-linear model,

$$\ln e_{ijkl} = \lambda_0 + \lambda_{i\dots} \textit{Offender}_i + \lambda_{.j\dots} \textit{Substance Use}_j \\ + \lambda_{\dots k} \textit{Antisocial Attitudes} + \lambda_{\dots l} \textit{Single Parent}. \quad (1)$$

\ln = the natural logarithm for base e .

Equation 1 represents the *log-linear model*, which estimates the e_{ijkl} as a linear combination consisting of a constant and four main effects. The λ s are parameters weights of the variables or variable interactions. They are not observed from the data but estimated and interpreted as regression effects in multiple regression. Equation 1 also reveals the linear as well as additive characters of log-linear modeling. The first order CFA results can be taken from the following Table 2:

A model fits or the observed frequencies do not deviate significantly from the expected cell frequencies, when the respective χ^2 - or LR -values have a p -value greater (!) than $\alpha = .05$. Here, both p -values of the main effects model indicate a non-fit: $\chi^2_{LR} = 102.26$, $df = 11$, $p = .000$; $\chi^2 = 140.02$, $df = 11$, $p = .000$. *confreq* also calculates the respective Information Criteria which can be used to compare two (or more) models; a better model fit leads to lower values for the Information Criteria: For this (main effect) model $AIC = 183.46$ and $BIC = 187.33$. Two *Types* concerning a juvenile offender emerged (here the observed frequencies were significantly greater than the expected frequencies): the configuration '2 2 2 1' represents an offender in 11th grade who used illegal substances, who also had antisocial attitudes but did not live in a single parent household at 9th grade, whereas the configuration '2 2 2 2' displays a typical offender who used illegal drugs, who had antisocial attitudes and who lived in a single parent home. Obviously, the attribute 'living with a single parent home' is not (significantly) related to the substance use and antisocial attitude. Therefore, this variable can be taken from the tree diagram, where 'a single parent home' was located on a different branch. The pattern '1 1 1 1' stands for a typical non-offender who had no risk-factors.

Table 2: Main effects model or first order Configural Frequency Analysis (CFA) for the 9th grade data with offender status and three risk factors

Configurations	f(o)	f(e)	z-statistic	p-value	Type/Antitype
1 1 1 1	291	235.96	3.58	0.00	T
1 1 1 2	42	49.86	-1.11	0.13	
1 1 2 1	21	38.38	-2.81	0.01	A
1 1 2 2	7	8.11	-0.39	0.35	
1 2 1 1	73	105.59	-3.71	0.00	A
1 2 1 2	21	22.31	-0.28	0.39	
1 2 2 1	19	17.17	0.44	0.33	
1 2 2 2	7	3.63	1.77	0.04	
2 1 1 1	24	48.07	-3.47	0.00	A
2 1 1 2	11	10.16	0.26	0.40	
2 1 2 1	4	7.82	-1.37	0.09	
2 1 1 2	0	1.65	-1.29	0.10	
2 2 1 1	27	21.51	1.18	0.12	
2 2 1 2	9	4.55	2.10	0.02	
2 2 2 1	19	3.50	8.29	0.00	T
2 2 2 2	4	0.74	3.79	0.00	T

Note. T = 'Type' and A = 'Antitype'.

The R-package *confreq*

Now we are going to introduce the R-package *confreq*, which is the abbreviation for **configural frequencies**. The package was written by Heine (2019). The name *confreq* avoids a mix-up with the statistical method called Confirmatory Factor Analysis. The package *confreq* is now available in version 1.5.4. from the repositories on CRAN (see <https://cran.r-project.org/>). R is an object oriented program which uses a syntax to run the respective analyses (R Core Team, 2019); a syntax line beginning with an hash key is a comment.

Within R one can read in a frequency table by typing in the pattern and their frequencies into an spreadsheet file. Such form of data is typically named as *tabulated data* where the rows represent all possible combinations of the variables and the rightmost column holds the respective frequencies.

Conducting a first-order Configural Frequency Analysis using the R-package *confreq*

To prepare the data to be imported into R, save the spreadsheet as an csv-file into your current R workspace directory by naming it for example as '*9thgrade_tabulated.csv*'. For correct processing the tabulated data with *confreq* the header of the rightmost column holding the pattern frequencies must be named 'Freq'. Depending on the language settings of your operating system or your spreadsheet-program, the field separator for the spreadsheet cells will either be a semicolon or a comma in the csv file. Either format (semicolon or a comma) will be fine, but always remember the respective format for a proper import of the file into R. After saving the data file, switch to your R editor and

type in and execute either of the following commands (assuming that the field separator is a semicolon – see argument: `sep=";"`).

The following R syntax will lead to the results of Table 2.

```
# reading in an spreadsheet file in csv-format
# do not use zeros as configural patterns!
dat.tab <- read.table("9thgrade_tabulated.csv", sep=";", header=TRUE)
dat.tab
# you need to load the R-package
library("confreq")
# convert the data to a pattern frequencies object
dat.pat<-dat2fre(fre2dat(dat.tab))
dat.pat
# first order CFA
# CFA includes the confreq function
res.01 <- CFA(dat_pat, alpha=0.05, form="~ Offender + Substance
+ Attitude + Singelpar")
summary(res_01)
# show the structure of resd3
str(res.01)
# show the desgin matrix
res.01$designmatrix
```

Functional Configural Frequency Analysis or blanking out cells

Because the typical non-offender Type with the configuration '1 1 1 1' does not contribute substantively to the search for offender Types, this Type will be neglected in the next step. This type is also rather trivial, because it says that juveniles with no risk factors do mostly belong to the non-offenders. Functional Configural Frequency Analysis (fCFA) is a variation of CFA which is an iterative approach while blanking out certain cells. fCFA was presented by von Eye and Mair (2008) and it is able to find extreme cells or outlier cells. Extreme cells are blanked out to test whether the rest of the table is independent, this is called **quasi-independence** in the presence of one Type (cf. Stemmler, 2014). fCFA uses the design matrix to blank out extreme cells. Because the presented research in our example is mostly interested in predicting offenders, the typical non-offender Type was blanked out. The results given in Table 3 were found.

The fit of the resulting model is much better, however, still not perfect and we lose one degree of freedom for the blanked out cell: $\chi^2_{LR} = 28.91$, $df = 10$, $p = 0.001$; $\chi^2 = 30.60$, $df = 10$, $p = 0.001$. The resulting Information Criteria are lower: $AIC = 112.11$ and $BIC = 116.74$. Pattern '2 2 2 1' is still significant. Blanking out cells in a design matrix is very easy to handle with *confreq*.

Table 3: CFA for the 9th grade data with offender status and three risk factors with one cell configuration being blanked out

Configurations	f(o)	f(e)	z-statistic	p-value	Type/Antitype
1 1 1 1	291	291	0.00	0.50	blanked out
1 1 1 2	42	29.14	2.38	0.01	
1 1 2 1	21	21.72	-0.15	0.44	
1 1 2 2	7	8.35	-0.47	0.32	
1 2 1 1	73	73.45	-0.05	0.48	
1 2 1 2	21	28.22	-0.16	0.09	
1 2 2 1	19	21.03	-0.44	0.33	
1 2 2 2	7	8.08	-0.38	0.35	
2 1 1 1	24	27.96	-0.75	0.23	
2 1 1 2	11	10.74	0.08	0.47	
2 1 2 1	4	8.01	-1.42	0.08	
2 1 1 2	0	3.08	-1.75	0.04	
2 2 1 1	27	27.08	-0.02	0.49	
2 2 1 2	9	10.40	-0.44	0.33	
2 2 2 1	19	7.75	4.04	0.00	T
2 2 2 2	4	2.98	0.59	0.27	

Note. T = 'Type'.

Let's have a look at the respective syntax:

```
# the argument 'blank = c(1)' will blank out the configuration number 1
res.02 <- CFA(dat.pat, blank = c(1))
summary(res.02)
# inspect the respective designmatrix
res.02$designmatrix
```

The design matrix underlying the results in Table 3 appears in Table 4 (the intercept vector is implied).

Table 4: Design matrix for the main effects model plus a blanked out cell

	Offender	Substance Use	Antisocial Attitude	Singelparent	Blanked Cell
1	1	1	1	1	1
2	1	1	1	1	0
3	1	1	1	-1	0
4	1	1	1	-1	0
5	1	-1	1	1	0
6	1	-1	1	-1	0
7	1	-1	-1	1	0
8	1	-1	-1	-1	0
9	-1	1	1	1	0
10	-1	1	1	-1	0
11	-1	1	-1	1	0
12	-1	1	-1	-1	0
13	-1	-1	1	1	0
14	-1	-1	1	-1	0
15	-1	-1	-1	1	0
16	-1	-1	-1	-1	0

Table 4 contains the effect coding vectors. The first four columns represent the main effects of the variables offender status, substance use, antisocial attitudes and living in a single parent home. The last vector represents the blanked out cell. The blanked out cell is coded with '1' and the rest of the configurations are coded '0'.

As a next step, the obtained tree structure in Figure 1 will be analyzed with CFA as proposed by von Eye et al. (2019) in this Special Issue. This approach is closely related to Functional Configural Frequency (fCFA) as presented above. Instead of blanking out certain cells, several vectors are entered to the design matrix (see Table 5) to model the effects of the tree structure. To obtain an appropriate design matrix to model the tree structure in Figure 1, the design matrix for the main effects model including a blanked out cell from above needs to be extended. The following *R*-syntax will do the necessary modifications and in turn will run the respective CFA model.

```
# create effect parameters which represent a decision tree model
tree.effect <- cbind(
Vector1=c(0,0,0,0, 0,0,0,0, 1,1,1,1, -1,-1,-1,-1),
Vector2=c(0,0,0,0, 0,0,0,0, 0,0,0,0, 1,1,-1,-1),
Vector3=c(0,0,0,0, 0,0,0,0, 0,0,0,0, 0,0,1,-1))
tree.effect
# add these 3 new columns to the design matrix
design.tree <- cbind(res.02$designmatrix,tree.effect)
# run CFA model with modified design matrix
res.03 <- CFA(dat.pat, form=design.tree)
summary(res.03)
```

Let's have a look at the respective design matrix depicted in Table 5 created by *confreq* that models the tree structure of the data of Table 1 (here again, the intercept vector was skipped).

Table 5: Design matrix for the main effects model, a blanked out cell and the effect coding vectors for the decision tree structure

	Offender	Substance Use	Attitude	Singlepar	Blanked Cell	Vector1	Vector2	Vector3
1	1	1	1	1	1	0	0	0
2	1	1	1	-1	0	0	0	0
3	1	1	-1	1	0	0	0	0
4	1	1	-1	-1	0	0	0	0
5	1	-1	1	1	0	0	0	0
6	1	-1	1	-1	0	0	0	0
7	1	-1	-1	1	0	0	0	0
8	1	-1	-1	-1	0	0	0	0
9	-1	1	1	1	0	1	0	0
10	-1	1	1	-1	0	1	0	0
11	-1	1	-1	1	0	1	0	0
12	-1	1	-1	-1	0	1	0	0
13	-1	-1	1	1	0	-1	1	0
14	-1	-1	1	-1	0	-1	1	0
15	-1	-1	-1	1	0	-1	-1	1
16	-1	-1	-1	-1	0	-1	-1	-1

The vectors in the last three columns in Table 5 model the decision tree structure or the decision sequences of the data. If the new fCFA model with the vectors representing the decision structure results in a fitting model and when all types have disappeared, the existing decision structure is obviously modeled correctly and 'an interpretation in the context of a decision tree becomes possible' (von Eye et al., 2019, p. 10).

Now, let's see whether the decision tree vectors lead to quasi independence for the rest of the table (see Table 6). And indeed no further *Types* or *Antitypes* evolved. The model fits very well: $\chi^2_{LR} = 7.26$, $df = 7$, $p = .40$; $\chi^2 = 5.71$, $df = 7$, $p = .57$. The resulting Information Criteria are also much lower: $AIC = 96.46$ and $BIC = 103.41$. Again, for the blanked out cell, the observed frequencies are equal to the expected cell frequencies, but this also happened to the last to configurations.

Table 6: CFA for the 9th grade data with offender status and three risk factors with one cell configuration being blanked out and the decision tree model

Configurations	f(o)	f(e)	z-statistic	p-value	Type/Antitype
1 1 1 1	291	291	0.00	0.50	blanked out
1 1 1 2	42	38.98	0.48	0.31	
1 1 2 1	21	23.06	-0.43	0.33	
1 1 2 2	7	7.96	-0.34	0.37	
1 2 1 1	73	74.06	-0.12	0.45	
1 2 1 2	21	25.58	-0.91	0.18	
1 2 2 1	19	15.13	0.99	0.16	
1 2 2 2	7	5.23	0.77	0.22	
2 1 1 1	24	24.07	-0.01	0.44	
2 1 1 2	11	8.31	0.93	0.17	
2 1 2 1	4	4.92	-0.41	0.34	
2 1 1 2	0	1.70	-1.30	0.10	
2 2 1 1	27	26.76	0.05	0.48	
2 2 1 2	9	9.24	-0.08	0.47	
2 2 2 1	19	19.00	0.00	0.50	
2 2 2 2	4	4.00	0.00	0.50	

Discussion

This article demonstrated the analysis of a tree structure based on CHAID with the help of Configural Frequency Analysis (CFA; cf. Stemmler, 2014). The branches of a tree can be modeled with the use of effect coding vectors. The CFA procedure used is similar to the approach of a functional CFA (fCFA; von Eye & Mair, 2008). fCFA uses an iterative procedure to either blank out extreme cells or, as demonstrated here, to add additional vectors which try to model the data. If the addition of vectors to the design matrix leads to a model fit, the rest of the table can be interpreted as quasi independence in the presence of *Types*. The presented procedure is valid for trees with branches of equal and unequal length. In addition, covariates can be entered into the model (cf. Glück & von Eye, 2000).

This paper also demonstrated the use of the open source *R*-package *confreq* (Heine, 2019). In *confreq* the frequency tables can be entered into a spread sheet and read in into the program. The package is easy to use and in addition, it allows a flexible modeling of the design matrix. In CFA, the expected frequencies are calculated according to the null model and different null models reflect different hypotheses. One may blank out cells in order to identify possible extreme cells or, as demonstrated, the decision sequence of a tree structure may be modeled. Also the use of additional covariates is straight forward.

CFA is a method to detect over- or under frequented cells. A significant pattern or configuration is a set of characteristics or attributes that belong together, as several symptoms that belong to a syndrome (e.g., the Leuner syndrome; see Krauth & Lienert, 1973). In the case of a tree structure, the detected *Types* can be interpreted differently. The original use of CHAID or CART was intended to investigate the decisions of consumers to buy or not to buy a certain item. The branches would therefore represent a temporal order of decisions. In our case one could conjecture that, when juveniles 'decide' to use illegal drugs in 9th grade and when they 'chose' to exhibit in addition, antisocial attitudes, then the likelihood for becoming an offender two years later is very high. If one has longitudinal data, like in our example, the temporal order cannot be changed, but with a cross-sectional data, any temporal order of a tree structure can be debated.

References

- Arnīs, M. (2015). *Die Entstehung und Entwicklung devianten und delinquenten Verhaltens im Lebensverlauf und ihre Bedeutung für soziale Ungleichheitsprozesse: Fragebogendokumentation der Schülerbefragung in Dortmund und Nürnberg. Band 2: Skalendokumentation* (Tech. Rep. No. 18). Bielefeld: DFG Research Center (SFB) 882 From Heterogeneities to Inequalities: Bielefeld University.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
- Corrado, R. (2002). An introduction to the risk/needs case management instrument for children and youth at risk for violence: The Cracow instrument. In R. Corrado, R. Roesch, S. D. Hart, & J. K. Gierowski (Eds.), *Multi-problem violent youth: A foundation for comparative research on needs, interventions and outcomes* (pp. 295–301). Amsterdam: IOS Press.
- Glück, J., & von Eye, A. (2000). Including covariates in Configural Frequency Analysis. *Psychologische Beiträge*, 42(3), 405–417.
- Haughton, D., & Oulabi, S. (1997). Direct marketing modeling with CART and CHAID. *Journal of Direct Marketing*, 11(4), 42–52. doi: 10.1002/(SICI)1522-7138(199723)11:4<42::AID-DIR7>3.0.CO;2-W
- Heine, J.-H. (2019). *confreq: Configural Frequencies Analysis Using Log-linear Modeling (V 1.5.4-3)*. Retrieved from <http://CRAN.R-project.org/package=confreq> (R package version 1.5.4-3)
- Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2), 119–127. doi: 10.2307/2986296
- Krauth, J., & Lienert, G. A. (1973). *Die Konfigurationsfrequenzanalyse (KFA) und ihre Anwendung in Psychologie und Medizin: ein multivariates nichtparametrisches Verfahren zur Aufdeckung von Typen und Syndromen; mit 70 Tabellen*. Freiburg: Alber Karl.
- Lautsch, E., & Plichta, M. M. (2005). Configural Frequency Analysis (CFA) and Latent Class Analysis (LCA): Are the outcomes complementary? *Psychology Science*, 47(3/4), 424.
- Lautsch, E., & Thöle, U. (2005). Identification and analysis of types of personality in a brief measure of the Big-Five personality domains. *Psychology Science*, 47(3/4), 424.
- Lösel, F., Beelmann, A., Stemmler, M., & Jaurisch, S. (2006). Prävention von Problemen des Sozialverhaltens im Vorschulalter: Evaluation des Eltern- und Kindertrainings EFFEKT. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35(2), 127–139.
- Lösel, F., & Stemmler, M. (2012). Preventing Child Behavior Problems in the Erlangen-Nuremberg Development and Prevention Study: Results from Preschool to Secondary School Age. *International Journal of Conflict and Violence*, 6(2), 214–224.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing. Retrieved from <https://www.R-project.org/>
- Reinecke, J., Stemmler, M., Arnīs, M., El-Kayed, N., Meinert, J., Pöge, A., ... Wittenberg, J. (2013). Entstehung und Entwicklung von Kinder- und Jugenddelinquenz: erste Ergebnisse einer Längsschnittstudie. *Neue Kriminalpolitik*, 25(3).
- Reinecke, J., Stemmler, M., & Wittenberg, J. (Eds.). (2016). *Devianz und Delinquenz im Kindes- und Jugendalter: Ungleichheitsdimensionen und Risikofaktoren*. Wiesbaden: Springer VS.
- Stemmler, M. (2014). *Person-centered methods: Configural Frequency Analysis (CFA) and other methods for the analysis of contingency tables*. New York: Springer Publishing Company.
- Stemmler, M., & Heine, J.-H. (2017). Using Configural Frequency Analysis as a Person-centered

- Analytic Approach with Categorical Data. *International Journal of Behavioral Development*, 41(5), 632–646. doi: 10.1177/0165025416647524
- Stemmler, M., Wallner, S., & Link, E. (2018). Risikofaktoren für die Entwicklung dissozialen Verhaltens in der Kindheit und Jugend. In D. Hermann & A. Pöge (Eds.), *Kriminalsoziologie: Handbuch für Wissenschaft und Praxis* (pp. 247–262). Baden-Baden: Nomos Verlagsgesellschaft.
- von Eye, A. (2002). *Configural Frequency Analysis: Methods, models, and applications*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- von Eye, A., & Mair, P. (2008). A functional approach to Configural Frequency Analysis. *Austrian Journal of Statistics*, 37(2), 161–173.
- von Eye, A., & Mun, E.-Y. (2013). *Log-Linear Modeling: Concepts, Interpretation, and Application*. New York: John Wiley & Sons.
- von Eye, A., Wiedermann, W., & von Weber, S. (2019). Log-linear and configural analysis of tree structures. *Psychological Test and Assessment Modeling*, Special Issue on Tree-based Methods.
- Wallner, S. (2007). *Die Prognose von Dissozialität mit dem Cracow Risk Assessment Instrument* (Monographie). University of Erlangen-Nuremberg, Institute of Psychology, Erlangen-Nuremberg.
- Wallner, S., & Stemmler, M. (2006). Profile physischer Aggression in Kindheit und Jugend. In S. Wallner, M. Weiss, J. Reinecke, & M. Stemmler (Eds.), *Devianz und Delinquenz in Kindheit und Jugend neue Ansätze der kriminologischen Forschung* (pp. 121–140). Wiesbaden, Germany: Springer VS.
- Weiss, M., & Wallner, S. (2019). Methodik der Studie. In S. Wallner, M. Weiss, J. Reinecke, & M. Stemmler (Eds.), *Devianz und Delinquenz in Kindheit und Jugend: Neue Ansätze der kriminologischen Forschung* (pp. 17–38). Wiesbaden: Springer VS.