Psychology Science, Volume 47, 2005 (3/4), p. 342-355

# Prediction models for Configural Frequency Analysis

ALEXANDER VON EYE[1], PATRICK MAIR[2], G. ANNE BOGAT[1]

## Abstract

In this article, we extend the discussion of prediction models of Configural Frequency Analysis (CFA). We build on the correspondence of such prediction models to logit models that was shown by von Eye and Bogat (in this issue), and derive additional possible CFA prediction models. These models can, as is illustrated, include multiple predictors and multiple criteria. Design matrices are presented for each of the models that are discussed, and the corresponding logit models are specified. Data examples are presented.

Key words: Prediction Configural Frequency Analysis; logit models; design matrix

---

[1] Address correspondence concerning this article to Alexander von Eye, Michigan State University, 107 D Psychology Building, East Lansing, MI 48824-1116; voneye@msu.edu; G. Anne Bogat, Michigan State University, Department of Psychology

[2] Patrick Mair, Universität Wien, Department of Statistics

## Prediction models for Configural Frequency Analysis

One of the more attractive models of Configural Frequency Analysis (CFA; Lienert, 1968; von Eye, 2002; von Eye & Gutiérrez-Peña, 2004) is that of Prediction CFA. This model allows one to identify those predictor patterns that go hand-in-hand with particular patterns on the criterion side. Thus far in the literature, only a very limited range of prediction models for CFA has been discussed and applied (Lienert & Krauth, 1973; von Eye, 1985). In this article, we extend the approach proposed by von Eye and Bogat (2006), derive prediction models for CFA from (1) corresponding models of logistic regression (cf. Fienberg, 1980), and (2) the concept of a CFA base model. We focus on models that can be expressed in terms of hierarchical log-linear models.

## 1. Prediction CFA

Configural Frequency Analysis (CFA; Lienert, 1968; von Eye, 2002; von Eye, & Gutiérrez Peña, 2004) takes a perspective of data analysis that differs from the perspective taken by most statistic methods. Most methods analyze data by relating variables to each other. For example, the amount of alcohol consumed and body mass are used to predict blood alcohol content, or severity of disease is used to predict life expectancy. The variables used in these examples are continuous. In categorical variables, the same approach is typically taken. For example, graduation odds are predicted from gender and race of student athletes. Methods of logistic regression and log-linear modeling are typically used to analyze such data. These methods also relate variables to each other. Local deviations from a model are held against the model, and steps are undertaken to improve model fit.

In contrast, Prediction CFA (PCFA) does not ask whether variables are associated with each other, interact, or are predictive of each other. PCFA asks whether a particular pattern of categorical predictor variables allows one to predict the above or the below expectancy occurrence rate of a particular criterion pattern.

As all other models of CFA, PCFA asks for each cell of the cross-classification of all variables under study, whether the observed cell frequency, $m_r$, differs significantly from the expected cell frequency, $m_r$, that was estimated under a suitable base model (more detail concerning base models follows below), where $r$ goes over all cells of the cross-classification. The cell-specific null hypothesis is $H_0$: $E[m_r] = m_r$. If this hypothesis can be rejected, Cell $r$ constitutes a *CFA type*, if $m_r > m_r$, and a *CFA antitype*, if $m_r < m_r$ (for details concerning testing and α protection in CFA, see von Eye, 2002).

Standard CFA does not distinguish between groups of variables that differ in status. In contrast, PCFA distinguishes between predictors and criteria. To introduce PCFA, let $A$ and $B$ be the predictor variables and $C$ the criterion. Crossed, the three variables span a table with $R$ cells, indexed as $r = 1, ..., R$. Let the probability of cell $r$ be $\pi_r$, the observed cell frequency $m_r$, and the expected cell frequency, $m^*_r$, estimated under the PCFA *base model*.

The main characteristic of CFA base models is that they contain all terms that are *not* of interest to the researcher. If this model is contradicted, at least one of the terms that the researcher is interested in will exist. The base model of the version of PCFA that was proposed by Lienert and Krauth (1973) has the following characteristics:

1. It is *saturated in the predictors*. This characteristic is needed to prevent types and antitypes from emerging just because associations among predictors exist.
2. It is *saturated in the criteria*. This characteristic is needed to prevent types and antitypes from emerging just because associations among criteria exist.
3. It proposes *independence among predictors and criteria*.

Based on this set of characteristics, PCFA types and antitypes indicate predictor - criteria relationships at the level of patterns of variable categories. Specifically, if the cell-specific null hypothesis $H_0$: $E[m_r] = m_r$ can be rejected, cell $r$ constitutes a *prediction type*, if $m_r > m^*_r$, and a *prediction antitype*, if $m_r < m_r$. In different words, if, for the pattern of predictor and criterion variables in cell $r$, more cases are found than expected, the predictor pattern of cell $r$ is said to predict the occurrence of the criterion pattern of cell $r$. If fewer cases are found than expected, the predictor pattern of cell $r$ is said to predict the non-occurrence of the criterion pattern of cell $r$. Depending on theory, prediction types are sometimes interpreted as causal, for instance by stating that a particular predictor pattern causes a particular criterion pattern to occur. Antitypes sometimes are interpreted as causal as well, for instance by stating that a particular predictor pattern prevents a particular criterion pattern from occurring.

Consider the following data example. On January 8, 2003, the New York Times published a table in which the three variables Race of Victim (V), Race of Defendant (D), and Penalty (P) issued were crossed for a total of 1311 murder cases that had been death penalty-eligible in Maryland from 1978 to 1999[3]. For the present purposes, we scale the two race variables as 1 = black, and 2 = white, and the Penalty variable as 1 = no death penalty and 2 = death penalty issued. In the following configural analysis, we ask whether there is a relationship between these three variables such that patterns of the two race variables predict the penalty issued. Specifically, we use the PCFA base model [DV][P]. For the cell-wise tests, we use the $z$-test, and we protect α using the Bonferroni procedure which results in the adjusted α* = 0.00625. Table 1 presents the results of PCFA.

Table 1:
PCFA of the predictors Race of Victim (V), Race of Defendant (D) and the criterion Penalty (P)

| Configuration VDP | $m$ | $m^*$ | $z$ | $p$ | Type/antitype? |
|---|---|---|---|---|---|
| 111 | 593 | 570.094 | .959 | .169 | |
| 112 | 14 | 36.906 | -3.770 | .000 | Antitype |
| 121 | 284 | 302.422 | -1.059 | .145 | |
| 122 | 38 | 19.578 | 4.164 | .000 | Type |
| 211 | 25 | 24.419 | .118 | .453 | |
| 212 | 1 | 1.581 | -.462 | .322 | |
| 221 | 272 | 277.064 | -.304 | .380 | |
| 222 | 23 | 17.936 | 1.196 | .116 | |

[3] The frequencies in the following analyses sum to 1250. This discrepancy to the sample size reported in the New York Times is the result of rounding (the paper only reported % values), and the newspaper's omission of 5% of cases that did not fall in the above eight patterns (mostly other-race cases).

The overall likelihood ratio $X^2$ for the CFA base model is 35.34. This value suggests significant data - model discrepancies ($df = 3$; $p < 0.01$). We thus can expect types and antitypes to emerge. Indeed, PCFA suggests that one type and one antitype exist. The antitype indicates that the pattern black defendant - black victim is less likely to result in the death penalty than one would expect under the assumptions specified in the PCFA base model. The type indicates that the pattern white defendant - black victim is more likely to result in the death penalty than expected from the PCFA base model. From these results, we conclude that there exists a predictors - criterion relationship such that two cells violate the PCFA base model.

## 2. Logistic regression

Logistic regression is a method to estimate the occurrence probability of one of the categories of a dependent variable using one or more predictors. The predictors can be categorical or continuous. In the present context in which we create parallel logistic regression and PCFA models, we focus on categorical variables only.

To illustrate, consider the example from Section 1 again. Let the probability of the death penalty being issued be denoted by $p_{P1}$. Then, the logistic regression model for the two predictors, V, and D, and the criterion, P, is

$$\log\left(\frac{p_{P=1}}{1-p_{P=1}}\right)\alpha + \beta_i^V + \beta_j^D$$

where α is the model constant, and the β are the parameter estimates (for alternative representations of this model see, e.g., Agresti, 2002). It is well known that this model can be equivalently expressed in terms of a log-linear model. Specifically, the log-linear model that is equivalent to the above logit model is

$$\log m^* = \lambda_0 + \lambda_i^V + \lambda_j^D + \lambda_k^P + \lambda_{ik}^{VD} + \lambda_{jk}^{VP} + \lambda_{ij}^{DP},$$

where the λ are the unknown model parameters, the subscripts index the parameters, and the superscripts indicate the variables. In bracket notation, this model can be expressed by [VD][VP][DP].

We now estimate this model for the data in Table 1. Table 2 displays the observed and the expected cell frequencies for this logistic regression analysis. To appraise overall goodness-of-fit, the Hosmer and Lemeshow test was performed.

Table 2:
Observed and expected cell frequencies for the logistic regression model for the data in Table 1

| | | $p_{P=1.00}$ | | $p_{P=1.00}$ | | |
|---|---|---|---|---|---|---|
| V | D | Observed | Expected | Observed | Expected | Total |
| 1 | 1 | 593 | 592.42 | 14 | 14.58 | 607 |
| 1 | 2 | 284 | 284.58 | 38 | 37.42 | 322 |
| 2 | 1 | 25 | 25.58 | 1 | .42 | 26 |
| 2 | 2 | 272 | 271.42 | 23 | 23.58 | 295 |

Obviously, the discrepancies between the observed and the expected cell frequencies are much smaller in Table 2 than in Table 1. Accordingly, the Hosmer and Lemeshow goodness-of-fit $X^2 = 0.88$ suggests that the model - data discrepancies are non-significant ($df = 1$; $p = 0.35$). Table 3 shows that Race of Defendant is the only significant predictor. Both predictors were entered simultaneously.

Table 3:
Regression table for logistic regression of the data in Table 1

| | b | S.E. | Wald | df | Sig. | Exp(b) |
|---|---|---|---|---|---|---|
| V(1) | -.414 | .272 | 2.316 | 1 | .128 | .661 |
| D(1) | 1.675 | .311 | 29.058 | 1 | .000 | 5.341 |
| Constant | 2.443 | .213 | 131.859 | 1 | .000 | |

## 3. Prediction models for CFA

Models of logistic regression with one dependent measure have the following characteristics:

a) They are saturated in the predictors;
b) They contain the terms that relate individual predictors and their interactions to the criteria.

An example of such a model for three variables was given above. It was the model [VD][VP][DP]. This model is saturated in the predictors, because it contains all main effects and the interaction between the two predictors, V and D. It also contains all bivariate predictor-criterion relationships. Including the three-way interaction would render this model saturated.

This model is the standard model of logistic regression of one categorical criterion and two categorical predictor variables. Other models are conceivable (see Agresti, 2002). We now discuss these models in the context of PCFA. For each model, we discuss the terms of interest in logistic regression, and the terms (not) of interest in PCFA. We begin with the simplest model, [VD][P], in which no relationship between the predictors and the criterion is assumed. This and the following models form a hierarchy. At the higher levels of the hierarchy, increasingly complex predictor-criterion relationships are modeled.

The simplest model, [VD][P], states that the criterion is jointly independent of the two predictors. This is equivalent to the logit model that only has an intercept. It can be contradicted only if predictor-criterion relationships exist. Each of the terms not included in the model relates one or both predictors to the criterion. These are the terms [VP][DP], and [VDP]. From the perspective of PCFA, this logistic regression model could be a base model also. If V and D are unrelated to P, this base model will fit, and there will be no prediction types or antitypes. A graphical representation of this model appears in Figure 1.

At the next hierarchical level, we find two models. These are the models [VD][VP] and [VD][DP]. In each of these models, the criterion is related to one of the predictors, and the relationship of this predictor to the respective other one is taken into account. Specifically, using the model [VD][VP], P is predicted from V only. From the perspective of PCFA, the relationship of interest is the one between the predictor V and the criterion P. This relationship must not be part of the base model. The relationship between the second predictor, D, and the criterion, P, is assumed to be non-existing (or not of interest). Therefore, it is part of the PCFA base model. In all, the PCFA base model for the logistic regression model [VD][VP] is [VD][DP]. A graphical representation of this model appears in the left panel of Figure 2. Accordingly, the PCFA base model for the logistic regression model [VD][DP] is [VD][VP]. This model is depicted in the right panel of Figure 2. The models in the left panel of Figure 2 (Model 2a in Table 4, below) and in the right panel of Figure 2 (Model 2b in Table 4) are hierarchically unrelated to each other.

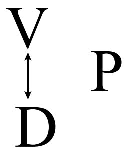


Figure 1:
Graphical representation of the logistic regression and PCFA model [VD][P]; the double-headed arrow indicates an association


Figure 2:
Graphical representation of (1) the logistic regression model [VD][VP] and the PCFA base model [VD][DP] (left panel), and the logistic regression model [VD][DP] and the PCFA base model [VD][VP] (right panel)

The most complex, non-saturated model is the one in which both predictors are related to the criterion variable. This is the logistic regression model [VD][VP][DP]. The only term not included in this model is the three-way interaction term which would make the model saturated. The two relationships of interest from the perspective of predicting the dependent variable, P, are [VP][DP]. Therefore, these terms as well as the three-way interaction are not part of the PCFA base model. This model is thus [VD][P]. This model is illustrated in the first data example, in Section 1, and depicted in Figure 3.

The model that also includes the interaction among all three variables, [VDP], is saturated. If the outcome variable is binary, this model is equivalent to a logit model with an interaction between the predictors, V and D. This model describes the data well if, for instance, the VP odds ratio varies across the categories of D. In the present context, this model is of lesser interest. Because it is saturated, the corresponding PCFA base model would be the same as for Models 1 and 3. This model can be contradicted by any predictor - criterion interaction. Table 4 summarizes the logistic regression and the PCFA base models that were discussed in this section. The design matrices for each of the models are given in the appendix. The base models are hierarchical log-linear.

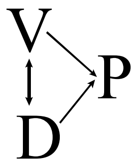


Figure 3:
Graphical representation of the logistic regression model [VD][VP] [DP] and the PCFA base model [VD][P]

Table 4:
Logistic regression models and corresponding PCFA base models

| Model in hierarchy | Log-linear representation | Logit model | Corresponding PCFA base model | Terms not part of PCFA base model[a] |
|---|---|---|---|---|
| 1 | [VD][P] | $\alpha$ | [VD][P] | [VP][DP][VDP] |
| 2a | [VD][VP] | $\alpha + \beta_i^V$ | [VD][DP] | [VP][VDP] |
| 2b | [VD][DP] | $\alpha + \beta_j^D$ | [VD][VP] | [DP][VDP] |
| 3 | [VD][VP][DP] | $\alpha + \beta_i^V + \beta_j^D$ | [VD][P] | [VP][DP][VDP] |
| 4 | [VDP] | $\alpha + \beta_i^V + \beta_j^D + \beta_{ij}^{VD}$ | [VD][P] | [VP][DP][VDP] |

*[a]In this column, the lower order terms are included in addition to the higher order terms, although they are redundant; this was done to illustrate the predictor-criterion relationships not included in the PCFA base models.*

Table 4 shows that, from a PCFA perspective, the first and the third models in the hierarchy are equivalent. The first model states that neither predictor is related to the criterion. The third model states that both predictors are related to the criterion. The base model for both states that neither predictor is related to the criterion. Both models can be contradicted only if predictor-criterion relationships exist. Thus, if types and antitypes emerge, the first model can be interpreted as contradicted. In contrast, types and antitypes for the third model confirm the model, suggesting that, at least locally, the predictors and the criterion are related to each other.

## 4. Data examples

In this section, we present a data example for the models 1 through 3 in Table 4. We use the example from Section 1. For each model, we present results from logistic regression and PCFA.

The PCFA results for Model 1 were already given in Table 1. Therefore, we start with the equivalent logistic representation.

Table 5:
Observed and expected cell frequencies for the logistic regression Model 1

| | | $p_{P=1.00}$ | | $p_{P=1.00}$ | | |
|---|---|---|---|---|---|---|
| V | D | Observed | Expected | Observed | Expected | Total |
| 1 | 1 | 593 | 570.09 | 14 | 36.91 | 607 |
| 1 | 2 | 284 | 302.42 | 38 | 19.58 | 322 |
| 2 | 1 | 25 | 24.42 | 1 | 1.58 | 26 |
| 2 | 2 | 272 | 277.06 | 23 | 17.94 | 295 |

Only the intercept parameter α is included in the model. This representation leads to a goodness-of-fit statistic of $X^2 = 35.95$ ($df = 3$; $p < 0.01$), and a rejection of the model. The parameter statistics are given Table 6.

Table 6:
Regression table for Model 1

| | b | S.E. | Wald | df | Sig. | Exp(b) |
|---|---|---|---|---|---|---|
| Const | 2.737 | .118 | 535.997 | 1 | .000 | 15.446 |

The next models in the hierarchy emanating from Table 1 are the Models 2a and 2b. The PCFA base Model 2a is [VD][DP], which leads to the frequencies given in Table 7.

Table 7:
PCFA of Model 2a

| Configuration VDP | *m* | *m** | *z* | *p* | Type/antitype? |
|---|---|---|---|---|---|
| 111 | 593 | 592.62 | .016 | .494 | |
| 112 | 14 | 14.38 | .100 | .460 | |
| 121 | 284 | 290.17 | .362 | .359 | |
| 122 | 38 | 31.83 | 1.093 | .137 | |
| 211 | 25 | 25.38 | .075 | .470 | |
| 212 | 1 | .62 | .482 | .315 | |
| 221 | 272 | 265.83 | .378 | .353 | |
| 222 | 23 | 29.17 | 1.142 | .127 | |

Since this base model fits the data well ($X^2$ = 3.03; *df* = 2; *p* = 0.220), no types and antitypes are detectable. The logistic regression representation and the parameter statistics are given in Table 8 and Table 9, respectively. The model fit indicated by a $X^2$-value of 36.43 (*df* = 2; *p* < 0.01) is again not tenable.

Table 8:
Observed and expected cell frequencies for the logistic regression Model 2a

| | | $p_{P=1.00}$ | | $p_{P=1.00}$ | | |
|---|---|---|---|---|---|---|
| V | D | Observed | Expected | Observed | Expected | Total |
| 1 | 1 | 593 | 573.02 | 14 | 33.98 | 607 |
| 1 | 2 | 284 | 303.98 | 38 | 18.02 | 322 |
| 2 | 1 | 25 | 24.06 | 1 | 1.94 | 26 |
| 2 | 2 | 272 | 272.94 | 23 | 22.06 | 295 |

Table 9:
Regression table for Model 2a

| | b | S.E. | Wald | df | Sig. | Exp(b) |
|---|---|---|---|---|---|---|
| V (1) | .310 | .256 | 1.464 | 1 | .226 | 1.363 |
| Const | 2.737 | .118 | 535.997 | 1 | .000 | 15.446 |

The effect parameter $\beta_1^V$ provides no significant contribution for the model. Concerning PCFA, Model 2b differs from Model 2a in that it includes the interaction [VP] instead of [DP]. Regarding the logistic regression, the effect of Race of Defendant, $\beta_1^D$, is included. The corresponding observed and expected frequencies for PCFA and logistic regression are given in Table 10 and Table 11, respectively, and the regression statistics appear in Table 12.

Table 10:
PCFA of Model 2b

| Configuration VDP | *m* | *m** | *z* | *p* | Type/antitype? |
|---|---|---|---|---|---|
| 111 | 593 | 573.02 | .834 | .202 | |
| 112 | 14 | 33.98 | 3.428 | .000 | Antitype |
| 121 | 284 | 303.98 | 1.146 | .126 | |
| 122 | 38 | 18.02 | 4.707 | .000 | Type |
| 211 | 25 | 24.06 | .192 | .576 | |
| 212 | 1 | 1.94 | .675 | .250 | |
| 221 | 272 | 272.94 | .570 | .284 | |
| 222 | 23 | 22.06 | .200 | .420 | |

The model fit is poor ($X^2$ = 36.44; *df* = 2; $p < 0.01$) and, as expected, types and antitypes are detected: A black victim and a black defendant allow one to predict that the death penalty is not issued. In contrast, a black victim and a white defendant allow one to predict that the death penalty is issued.

Table 11:
Observed and expected cell frequencies for the logistic regression Model 2b

| | | $p_{P=1.00}$ | | $p_{P=1.00}$ | | |
|---|---|---|---|---|---|---|
| V | D | Observed | Expected | Observed | Expected | Total |
| 1 | 1 | 593 | 592.62 | 14 | 14.38 | 607 |
| 1 | 2 | 284 | 290.17 | 38 | 31.83 | 322 |
| 2 | 1 | 25 | 25.38 | 1 | .62 | 26 |
| 2 | 2 | 272 | 265.83 | 23 | 29.17 | 295 |

Table 12:
Regression table for Model 2b

| | b | S.E. | Wald | df | Sig. | Exp(b) |
|---|---|---|---|---|---|---|
| V (1) | 1.509 | .294 | 26.317 | 1 | .000 | 4.522 |
| Const | 2.210 | .135 | 268.304 | 1 | .000 | 9.116 |

The goodness-of-fit value of $X^2$ = 3.03 ( *df* = 2; $p = 0.22$) indicates that the observed and expected frequencies match and the contribution of the effect parameter $\beta_1^D$ is significant.

As obvious from Table 4, the PCFA representation of Model 3 is equivalent to Model 1. Therefore, the corresponding expected frequencies are as given in Table 1. Furthermore, the logistic regression outputs for Model 3 are already given in Table 2 and 3.

*Extensions*. The ordinary binary logistic approach is only applicable for a binary response variable. For a polytomous response, the generalization of this model is straightforward: The multinomial logistic regression model. If more than one response variable is part of a model, the logistic regression approach is not appropriate anymore, since the approach to regression used above allows only one dependent variable on the left side of the linear model equation. Thus, only the log-linear and the PCFA representations are feasible. An extension of the current example could be an additional variable that indicates whether the trial went into retrial, denoted by R (1 = yes, 2 = no), which acts as additional response variable. Thus, we use V and D to predict P and R. The models corresponding to the model hierarchy given in Table 4 are given in Table 13.

Table 13:
log-linear representations, logit models, and corresponding PCFA base models

| Model in hierarchy | Log-linear representation | Logit model | Corresponding PCFA base model |
|---|---|---|---|
| 1 | [VD][PR] | - | [VD][PR] |
| 2a | [VD][PR][VP][VR] | - | [VD][PR][DP][DR] |
| 2b | [VD][PR][DP][DR] | - | [VD][PR][VP][VR] |
| 3 | [VD][PR][VP][DP][VR][DR] | - | [VD][PR] |

These models are consistent with the figures in Section 3. It should be noted, however, that on the response side, the interaction between the criteria must be part of the model since the aim of PFCA with more response variables is to predict patterns of the response variables (e.g., patterns of PR) from predictor patterns (e.g., patterns of VD).

## 5. Discussion

Thus far, the prediction models that were discussed for CFA were rather simple, and the correspondence to other prediction models for categorical variables was not known. The article by von Eye and Bogat (2006, in this issue) was the first to specify the relationship between PCFA and logistic regression. The present work extends this discussion by defining various models of PCFA and by showing the correspondence of these models with logit models.

It is only natural to ask whether this approach to PCFA can be carried even further. For example, one can ask whether mediator and moderator models can be distinguished at the level of PCFA, and if they can be translated into CFA base models. Such models would carry the idea of local relationships (Havránek & Lienert, 1984) further, and apply it to the level of relations among three or more variables, none of which shares status with any other. In the extreme, three-variable case, one variable could be external, the second variable could be both external and internal, and the last variable could be internal. Based on this classification, mediator models can be defined. In these models, variable relations are defined at the level of local associations that manifest in types and antitypes. Accordingly, chains of variable relations can be considered, and patterns of causal relationships (as originally proposed by von Eye & Brandtstädter, 1997).

We conclude that the correspondence between log-linear models and CFA base models carries farther than just stating that the original CFA base model is a log-linear main effect model. We see that more complex models can be derived. As will be seen in the discussion of mediator models (von Eye, 2006), some of these base models will even be non-hierarchical.

## References


1. Agresti, A. (2002). Categorical data analysis, 2nd ed. New York: Wiley.
2. Fienberg, S. E. (1980). The analysis of cross-classified categorical data. (2nd ed.). Cambridge, MA: MIT Press.
3. Havránek, T., & Lienert, G. A. (1984). Local and regional versus global contingency testing. Biometrical Journal, 26, 483 - 494.
4. 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.
5. Lienert, G. A., & Krauth, J. (1973). Die Konfigurationsfrequenzanalyse als Prädiktionsmodell in der angewandten Psychologie. In H. Eckensberger (Ed.), Bericht über den 28. Kongress der Deutschen Gesellschaft für Psychologie in Saarbrücken 1972 (pp. 219 - 228). Göttingen: Hogrefe.
6. von Eye, A. (1985). Konfigurationsfrequenzanalyse bei gerichteten Variablenbeziehungen (GKFA). EDV in Medizin und Biologie, 16, 37-40.
7. von Eye, A. (2002). Configural Frequency Analysis - Methods, Models, and Applications. Mahwah, NJ: Lawrence Erlbaum.
8. von Eye, A. (2006). Configural mediator analysis. (in preparation)
9. von Eye, A., & Bogat, G.A. (2006). Logistic regression and prediction Configural Frequency Analysis - A comparison. (This issue)
10. von Eye, A., & Brandtstädter, J. (1997). Configural Frequency Analysis as a searching device for possible causal relationships. Methods of Psychological Research - Online, 2, 2, 1 - 23.
11. von Eye, A., & Gutiérrez-Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. Journal of Applied Statistics, 31, 981 - 997.


## APPENDIX

Here, the design matrices for the Models 1- 3 are presented. We use dummy coding with Category 2 as reference, for every variable in the model. For each model, just one design matrix is produced. The vectors needed for logistic regression (and its corresponding log-linear representation), and for PCFA, will be indicated.

*Model 1*

| Const | V | D | P | VD |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 0 | 0 | 0 | 0 |

For model 1, the design matrix is the same for the log-linear and the PCFA representation.

*Model 2a*

| Const | V | D | P | VD | VP | DP |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |

The log-linear and the PCFA model differ with respect that for the latter DP is required instead of VP. The logistic regression approach needs the intercept, as always, the main effects of all three variables, and the interactions between V and D and between V and P.

*Model 2b*

| Const | V | D | P | VD | VP | DP |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |

The design matrix as a whole is the same as in Model 2a. In Model 2b however, the log-linear and logistic regression design vectors are the same as the PCFA representation for Model 2a. In turn, the PCFA Model 2b requires the same vectors as the log-linear Model 2a.

*Model 3*

| Const | V | D | P | VD | VP | DP |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Again, the design matrix as a whole is the same for Model 3. The log-linear and logistic regression representations need the whole design matrix, whereas the PCFA model is defined by the same vectors (i.e., main effects plus VD interaction) as the PCFA Model 1.