Table-specific continuity corrections for
Configural Frequency Analysis

STEFAN VON WEBER[1], ERWIN LAUTSCH[2], ALEXANDER VON EYE[3]

Summary

This article presents new continuity corrections for six tests in Configural Frequency Analysis (CFA). For each table, the correction is a constant. The magnitude of this constant depends on characteristics of the table such as degrees of freedom, size, cell frequency, strength of type, and on the nominal level α. Strength of type is a quantitative descriptor of types. Simulation results suggest that three tests, specifically the test proposed by Perli et al., the new test proposed by Lautsch and von Weber, and the χ^2-component test, cover over 90% of the best solutions.

Key words: Contingency table, configural frequency analysis, CFA, search for types, statistical test, simulation, continuity correction

---

[1] Prof. Dr. Dr. Stefan von Weber, Fachhochschule Furtwangen, Fachbereich MuV: Maschinenbau und Verfahrenstechnik, PSF 3840, D-78027 Villingen-Schwenningen, E-mail: webers@fh-furtwangen.de

[2] Prof. Dr. Dr. Erwin Lautsch, Universität Gesamthochschule Kassel, Fachbereich 5: Gesellschaftswissenschaften, Nora-Platiel-Straße 1, D-34127 Kassel, E-mail: erla@uni-kassel.de

[3] Prof. Dr. Alexander von Eye, Michigan State University, Department of Psychology, 119 Snyder Hall, East Lansing, MI 48824-1117; E-mail: voneye@msu.edu

1. Introduction

Since the first presentation of the software program SICFA (**S**imulation **C**onfigural **F**requency **A**nalysis; Lautsch & von Weber, 1990, 1995), the speed and memory capacity of PCs have dramatically increased. When updating such a program, both old and new issues arise that can be addressed using the new technical possibilities. The most important issues include that the program should be able to recommend to users which local type test to use in a particular application.

For application in CFA, a selection of significance tests and test procedures is most popular whose characteristics have been studied extensively and are well known. For example, Lehmacher (1981) investigated and proposed exact and approximative hypergeometric tests, Herrendörfer et al. (1982) studied 2 x 2 tables (cf. von Eye, 2003a,b; von Eye & Mun, 2003), Disconis and Efron (1983) and Kareev (1992) discussed random sampling, Lindner (1984) discussed exact tests, Perli (1985) presented test procedures, Küchenhoff (1986) proposed a first continuity correction for Lehmacher's (1981) test, von Eye and Rovine (1988) studied the relative power of significance tests (cf. Indurkhya & von Eye, 2000; von Eye, 2002a,b, 2003b), Krauth (1993) discussed the selection of tests, Lautsch and von Weber (1995) discussed splitting of samples and random sampling, and von Weber (2000) discussed methods of estimating the β-error in CFA testing.

One result that emerged consistently is that all CFA tests can be both conservative and non-conservative, depending on the characteristics of the table under study. Early on, researchers tried to deal with non-conservative tendencies of tests in small samples, for example, by means of continuity corrections (Küchenhoff, 1986; Dunkl & von Eye, 1990). Typically, continuity corrections improve test performance substantially (for a counterexample, see von Eye, 2002b, pp. 71, 76). The increased power of PCs allows program developers to further develop methods of test improvement. These developments are needed in CFA applications, because preliminary simulation results suggested that *one* correction constant may not be optimal for all tables. Table characteristics and the selected nominal α determine the usefulness of this constant.

In this article, we present results from a broad investigation of the effects of continuity corrections for a selection of CFA tests. These tests are the χ-component test, the exact binomial test, Lehmacher's (1981) asymptotic hypergeometric test, the asymptotic test of Perli, Hommel, and Lehmacher (1984), and Lautsch and von Weber's (2002) new test. The investigation also takes into account two procedures that guarantee that the probability of detecting types in CFA is high if the nominal α is close to the factual α. In other words, these two procedures guarantee that the probability of committing a β-error is minimal. Here, the β-error is defined as the ratio of the number of missed types to the total number of existing types across all tables with the same characteristics (cf. von Weber, 2000). The number of existing types can be exactly determined only in a computer simulation.

The simulations that are reported in this article were performed under the following conditions:

- The continuity correction K is a constant for a specific table; in other words, there is only one K per table;
- The estimation of K is based only on the characteristics of a table and the nominal α;
- Both types and antitypes are considered; and
- Characteristics of tables are taken into account only if they can be varied independently of each other.

2. Data generation

The characteristics of tables include, for example, the number of variables, d, that span the table, the number ZZ of cells, the degrees of freedom, df, and the sample size, N (or the average number of cases per cell, $mf = N/ZZ$). The overall χ^2 of a table can be estimated as well as the maximum type strength, τ_{max}. There exist other table characteristics, for example the type of residual distribution (DT), that are hard to estimate, although sampling procedures often determine distributional characteristics (von Eye, Schuster, & Gutiérrez-Peña, 2000). The following paragraphs provide definitions of terms that are key for the following considerations.

The *type strength*, τ, describes the weight of a type or an antitype. τ is the main determinant for the estimation of the magnitude K of the continuity correction. The concept of *type strength* can be derived from Lienert's (1969) definition of a *contingency type*: a cell that constitutes a type contains more cases than expected based on the assumption of variable independence. Let n_i be the observed cell frequency and $\hat{e}^*_i$ the expected cell frequency that was estimated using Victor's log-linear models of quasi-independence (see Kieser and Victor, 1999; Victor 1989; cf. Lautsch & von Weber, 2002), then the ratio $n_i/\hat{e}^*_i$ is an estimate of $\tau + 1$. The difference $n_i - \hat{e}^*_i$ represents the surplus frequency in Cell i that constitutes a type. A cell with $n_i = \hat{e}^*_i$ thus carries a type strength of zero, and $\tau = 0$. If n_i is twice as large as $\hat{e}^*_i$, we obtain $\tau = (n_i - \hat{e}^*_i)/\hat{e}^*_i = 1$, etc. A Cell constitutes an antitype if $n_i < \hat{e}^*_i$. The strength τ of an antitype is defined by $\tau = (\hat{e}^*_i - n_i)/n_i$. If the strength of an antitype is $\tau = 1$, the observed frequency of Cell i is half the size of the expected frequency.

The maximum type strength, τ_{max}, of a contingency table is estimated by $\max_i(n_i - \hat{e}_i)/mf$, where i goes over all cells in the table, n_i is the observed cell frequency, $\hat{e}_i$ is the expected cell frequency, estimated under the appropriate CFA base model, and mf is the average cell frequency. In simulations, tables with a priori determined type strength can be created. The above measures can then be used to estimate the value of τ_{max} for an observed table.

The distribution type, DT, indicates whether the cell-specific residuals are normally ($DT = 0$) or binomially ($DT = 1$ or $DT = 2$; see below) distributed. Another distribution type that can be considered is the hypergeometric distribution. The hypergeometric and the binomial distributions approximate each other if the size of the population is much larger than the size of the sample. Empirical data, in which the population and the sample are of equal size, are extremely rare. However, to be able to accommodate cases in which not the total sample size was determined a priori but the marginals of a particular variable (product-multinomial sampling; for example, researchers may determine that they wish to examine the same, a priori specified number of male and female respondents), two forms of the binomial distributions were considered. The first, labeled with $DT = 1$, uses the sample size N to estimate the binomial probability $P_{n.k}$ (see Krauth, 1993). The second form of the binomial distribution considered here, labeled with $DT = 2$, uses the minimum of the marginal frequencies for a particular cell. The third of these distributions comes with the largest skewness (cf. von Eye's, 2003b, "CFA conditions").

The estimation of the residual distribution type from empirical data can be rather involved. However, simulation results suggest that the effect of the distribution type is minimal. Therefore, there is no need to estimate it.

In general, when data are collected under optimal conditions, the cell frequencies follow a hypergeometric distribution (Krauth, 1993; Lindner, 1984). This distribution is very close to a binomial distribution. This, however, does not apply to the residuals $n_{ij} - \hat{e}_{ij}$. Whenever the expected cell frequencies $\hat{e}_{ij}$ are estimated from the marginals, Gauß' limit theorem applies

and the residual distribution approximates the normal distribution. When, as is typical, the data are collected under suboptimal conditions, situation-specific variations of the a priori cell probability p can occur. These fluctuations also shift the a priori cell probabilities towards a normal distribution. Therefore, the residuals $n_{ij} - \hat{e}_{ij}$ follow a mixed distribution with varying contributions of the normal distribution.

The degrees of freedom, df and the number of cells, N_c, are highly correlated, in particular when one focuses on only one CFA base model. The following simulations used only the main effect base model. In this case, the correlation between df and N_c is perfect, and using the df renders the number of cells redundant, and vice versa. Another preliminary investigation concerned the dimension of the table d. We found that d does have an effect, but only a small one. Additional important factors include mf, τ_{max}, and α.

Data for the following simulations were generated in the following seven steps.

Step 1. Contingency tables with $d = 2, 3,$ and 4 were created. The four dimensions are termed *rows*, *columns*, *blocks*, and *pages*. The tables had I rows, J columns, K blocks (for $d > 2$), and L pages (for $d > 3$), each randomly drawn from the interval [2, 6]. Of the resulting tables, all those were eliminated that had degrees of freedom outside an a priori-specified interval. Intervals for the degrees of freedom were [4, 4], [18, 22], and [80, 90].

Step 2. This step involved the random generation of the probabilities of the marginal distributions under the usual constraints that $0 < p < 1$ and $\sum p = 1$. The cell frequencies were then estimated under the assumption of independent rows, columns, blocks, and pages, that is $p_{ijkl} = p_i p_j p_k p_l$. Thus far, the data conform exactly to the model of variable independence.

Step 3. This step involves determining the number of types and antitypes, Nt. The upper limit of this number is given by $1 \leq Nt \leq df$ and $1 \leq Nt \leq int(df^{1/2}/2)$. The second of these conditions was chosen arbitrarily. However, it is conform to Kieser and Victor's (1999) concepts of CFA. The df of a table indicates the maximum number of independent hypotheses in a contingency table (see Perli, 1985; Perli et al., 1985).

Step 4. The position of each of the Nt types and antitypes was selected randomly. However, one constraint was placed. No marginal sum could contain more than 50% of its cases from type cells or antitype cells. Thus, for example, in 2 x 3 tables, no column can contain more than one type and antitype.

Step 5. To create types and antitypes, the probabilities p_{ijkl} were now multiplied by factors of the form $(1 + \tau)$. Thereby, the type strength is an even random number between 1 and τ_{max} with $\tau_{max} = \{1, 2, 3\}$. In addition, the value of $\tau_{max} = 0$ was included to also have tables with no types or antitypes. Thus, for $\tau = 1$, the observed frequency is twice as large as the Victor-estimated expected frequency. As far as types are concerned, this definition conforms to Victor's definition of types. As far as antitypes are concerned, there is a dispute in the literature. However, this dispute focuses more on the interpretation of antitypes than on the statistical result that an antitype may exist.

The ratio of thus randomly created types and antitypes was set to be 2 : 1. The first type was assigned the value of a maximum type strength, τ_{max}. When there were more types, their weights were assigned linearly decreasing values, with the constraint that these values be greater than or equal to 1. For example, for a table with 4 types and antitypes and a maximum type strength of $\tau_{max} = 3$, the four τ-values are 3, 2, 1, 1. The reason for the lower limit for τ lies in the control of ß. When a weight is too small, the chance of reliably identifying a type, and - even more so - an antitype, is very small.

Step 6. This step is needed to make sure that the condition $\Sigma = 1$ holds. After adding a constant to each cell, this condition no longer holds. Therefore, a correction is needed. Consider the cell, for which the a priori probability is $p = 0.055$. When, for this cell, the weight

$\tau = 1$ is used, that is, when this cell is randomly selected to constitute a type, its probability changes to be $p^* = 0.055 \cdot 2 = 0.11$, and the sum of all cell probabilities increases from $\Sigma p = 1.0$ to $\Sigma p = 1.055$. Therefore, we reduce the probability for each cell proportionally by the factor $1/1.055$. The type-cell then has the probability of $p^{**} = 0.11/1.055 = 0.1043$, and the sum of all thus corrected cell probabilities is 1 again. This applies accordingly when more than one type or antitype exists in a table.

Step 7. The last step of the simulation involves calculating the estimated expected cell frequencies. Each cell is multiplied by the a priori determined sample size, $N = ZZ \cdot mf$. The values of mf used in this simulation were $mf = 5$ for small samples, $mf = 15$ for medium-size samples, and $mf = 50$ for large samples. The estimated expected cell frequencies were thus $\hat{e}_i = Np_i$, where i goes over all cells in the table. A drawing error[4] was then added to each cell, depending on the distribution type DT. The resulting values were rounded to be integers. Negative values were set to zero. It should be noted that this drawing error is the main source of errors in weakly frequented contingency tables. It prevents researchers from reliably identifying types, in particular types of low strength.

3. The tests and the test procedures

Each test in the simulation was done under the two-sided null hypothesis that Cell i constitutes neither a type nor an antitype, with the alternative hypothesis that Cell i does constitute a type or an antitype. Holm's (1979) sequential test procedure was used. However, the first test was not performed under the adjusted threshold $\alpha^* = \alpha/N_C$, but under $\alpha^* = \alpha/df$ (cf. Perli et al., 1985). The reason for this further adjusted first α^* in the sequence of tests is that the estimation of expected cell frequencies decreases the degrees of freedom and, thus, the number of independent hypotheses. For instance, in a table that is spanned by three variables, estimating the expected cell frequencies using the marginals reduces the degrees of freedom from

$$df = IJK \quad \text{to} \quad df = IJK - (I - 1) - (J - 1) - (K - 1) - 1.$$

A 2 x 2 table with $df = 1$ allows one to test only one independent hypothesis.

A two-stage search (also called *hybrid test procedure*) was implemented:

(1) The sample is divided randomly in two parts with each case having a probability of $p = 0.5$ of being placed in either group. Using this sample, the local tests for types and antitypes are performed without using the Holm-Perli multiple test procedure described above. This exploratory step yields a preliminary selection of type and antitype cells.
(2) The cells that emerged as possibly constituting types and antitypes in the first step are the only cells that are examined in the second step. This step is performed using the Holm procedure without the Perli-component. Thus, the adjusted significance threshold now is $\alpha^* = \alpha/N_H$, where N_H is the number of prospective type and antitype cells identified in the first step of the two-stage search.

---

[4] The term *drawing error* is used here to denote the discrepancy between observed and expected frequencies that can be observed even under optimal sampling conditions. The distribution of this error depends on the selected model. The term *sampling error* denotes additional errors that may be hard to quantify. These errors reflect discrepancies between model and reality.

The *continuity correction* constitutes an essential component of the present study. The correction factor, K, is selected so that the nominal level α prevails asymptotically, that is, for a large number of tables with the same characteristics. Küchenhoff's (1986) continuity correction involves subtracting the constant 0.5 from each difference between observed and estimated expected cell frequency. The effect of this correction is minimal for large cell frequencies, and large for small cell frequencies. The continuity correction proposed by Dunkl and von Eye (1990) increases the estimate of the standard error in the denominator, e^*_{ijk}, by the factor $(e^*_{ijk} + 0.5) / (e^*_{ijk} - 0.5)$. Here again, the effect of the correction is stronger for small frequencies e^*_{ijk}. In both cases, the magnitude of the resulting test statistic is reduced (note again, that von Eye, 2002b, showed (pp. 71, 76), that Küchenhoff's (1986) correction has the opposite effect when the difference between the observed and the estimated expected cell frequency is less than 0.5).

The exact binomial test (Krauth, 1973) and the exact hypergeometric test (Lindner, 1984; Lehmacher, 1981) do not calculate a test statistic. Instead, they calculate the exact probability of the observed cell frequency directly. A different correction than subtracting a constant is needed for these tests. We performed a number of experiments for the tests of Lienert, Krauth, and Lehmacher to study the effects of various correction formulas. The formulas presented in the following paragraphs seem to work efficiently, and, as important for software development, are numerically tractable. The constant *K* denotes the correction term in the following equations.

The constant K is estimated iteratively. Let α be the a priori specified, nominal Type I error level for the multiple level hypothesis concerning the existence of types and antitypes in a contingency table, and α^* the estimate of the factual error level for the M contingency tables that were created in the simulation. The estimate α^* is a function of the parameters *d, mfr*, τ_{max}, etc., but also of the constant K. If the simulation varies K while keeping all other parameters constant, the estimate is $\alpha^*(K) = f(K) + err$, where f is an unknown, typically nonlinear (and for $M \to \infty$ assumed to be monotonous and differentiable) function, and err is an error of unknown distribution. The iteration attempts to estimate the equation $\alpha^* = f(K)$ as precisely as possible, in spite of the error element.

The algorithm used in the present simulations employs the three α-levels α = 0.01, 0.05, and 0.1. For each of these levels, the constant $K = K(\alpha)$ is estimated. The estimation process itself begins by specifying the boundaries of the search interval $[K_{min}, K_{max}]$. The lower limit, K_{min}, is specified such that the estimate of α will be extremely conservative, that is, $\alpha^* \to 0$. The upper limit, K_{max}, is specified such that the estimate of α will be extremely non-conservative, that is, $\alpha^* \to 1$. During the iteration, this interval is reduced, and between 10 and 20 estimates $\alpha^*(K_i)$ are identified, each of which is located close to the a priori specified nominal level α. Because of computational constraints, the number M of tables that are generated in the simulation is $M \le 1000$, and because err > 0, the exact correspondence $\alpha = \alpha^*$ is extremely unlikely. The estimator of $K(\alpha)$ is the weighted sum of the 10 to 20 K_i identified above. A weight of 1 is assigned if $\alpha = \alpha^*_i$. Otherwise, the weights shrink exponentially with the square of the difference $\alpha - \alpha^*_i$.

To be able to specify the K-values for the three α-levels 0.01, 0.05, and 0.1, a linear regression is estimated for the three estimates $(K(\alpha_j), \ln(\alpha_j))$, with *j* = 1, 2, 3. Negative slopes are considered estimation errors, because they would contradict the assumption that large α-values imply large *K*-values. When a slope is negative, the mean of the three $K(\alpha_j)$ is used for all three α-levels.

The following six tests were examined. Each equation is given for the case of a three-dimensional table. The presentation for tables with different dimensions is straightforward.

1. The χ-component test of G. A. Lienert (1969) (in the following tables abbreviated **Li**)

$$\chi = \frac{n_{ijk} - \hat{e}_{ijk}}{\sqrt{e_{ijk}(1-K)}} \; .$$

2. The exact Binomial Test of Krauth (1973) (abbreviated with **Kr**)

$$b_{ijk} = \sum_{l=n_{ijk}}^{n} \binom{N}{l} p_{ijk}^{l}(1-p_{ijk})^{N-l} \;, \qquad\qquad K\text{-corrected: } p = \exp(\ln(b_{ijk})/(1-K))$$

3. The Hypergeometric Residual Test of Lehmacher (1981) and Krauth (1993; abbreviated with **LK**)

$$z_{ijk} = \frac{n_{ijk} - \hat{e}_{ijk}}{\sqrt{V_{ijk}(1-K)}} \qquad\qquad \text{with} \quad V_{ijk} = Np_{ijk}(1-p_{ijk}-(N-1)(p_{ijk}-p_{ijk}^{*}))$$

$$\text{and} \quad p_{ijk}^{*} = (N_{i..}-1)(N_{.j.}-1)(N_{..k}-1)/(N-1)^{3}$$

4. The Asymptotic Test of Perli, Hommel and Lehmacher (1995) (abbreviated with **Pe**)

$$W_{ijk} = \frac{n_{ijk} - \hat{e}_{ijk}}{\sqrt{N\sigma^{2}(1-K)}} \qquad\qquad \text{with} \quad \sigma^{2} = p_{ijk}(1+2p_{ijk}-(p_{i..}p_{.j.} + p_{i..}p_{..k} + p_{.j.}p_{..k}))$$

(The σ^{2} has this simple formula only under simplifying assumptions)

5. The Exact Hypergeometric Test of Lindner (1984) (abbreviated with **Ld**)

$$P(X \geq n_{ijk}) = \sum_{m \geq n_{ijk}} P(X = m) \quad \text{with} \quad P(X = m) = \sum_{i=r}^{s} H(m;i,c,N)(H(i;a,b,N)$$

with $\qquad\qquad r = \max(m, a+b-N)$, $\qquad$ a,b,c marginal sums of cell *ijk*,

$\qquad\qquad\qquad s = \min(a,b,m-c+N)$, $\qquad$ N total number of cases,

$$\text{and} \qquad\qquad H(m;a.b.N) = \frac{\binom{a}{m}\binom{N-a}{b-m}}{\binom{N}{b}} \; .$$

The *K*-correction is the same as for the exact Binomial Test.

6. The Test of Dunkl and von Eye with Victor-expected values $\hat{e}^{*}_{ijk}$ yielded by the new procedure of Lautsch and von Weber (2002) (abbreviated with **nPr**)

$$X = \frac{n_{ijk} - \hat{e}^{*}_{ijk}}{\sqrt{\sigma^{2*}_{ijk}(1-K)}} \quad \text{with variance} \qquad \sigma^{2*}_{ijk} = \hat{e}^{*}_{ijk}(\hat{e}^{*}_{ijk}+0.5)/(\hat{e}^{*}_{ijk}-0.5) \;.$$

4. Simulation results

Table 1 presents results for tables with *df* = 88 that are spanned by three variables with type/antitype strength values 1, 2, and 3. The information concerning the ß-values is based on several thousand contingency tables with the same characteristics. To generate frequencies, we used the RANDOM function available in Borland's Turbo Pascal 6.0. Each simulation started with different random seeds. These seeds were created by invoking the function RANDOM between 1 and 999 times. As compared to the subroutine RANDOMIZE which is also available in Turbo Pascal, this procedure has the advantage of creating reproducible quasi-random numbers. Table 1 presents results on the accuracy of the β-estimates in two panels. The simulations were run twice with the same parameters but with different seeds of the random number generator. The results of the first run appear in the top panel of the table, the results of the second run appear in the bottom panel. The comparison of the top with the bottom panels of Table 1 shows that the variability of the *K*-estimates is about 5%, and the variability of the β-estimates is about 2%.

Tab. 1:
Variability of *K*- und β-estimates (trials in top panel, trial two in bottom panel)

d	df	mf	α	τ	*K*	β
3	88	15	0.01	1	-0.2986	87.39
3	88	15	0.01	2	-1.0842	88.96
3	88	15	0.01	3	-1.9241	88.44
3	88	15	0.01	1	-0.2828	86.62
3	88	15	0.01	2	-1.0409	87.56
3	88	15	0.01	3	-1.8837	87.46

The results presented here are more informative than the previous ones presented by von Weber (2000), because here, each test keeps the a priori determined nominal α. The correction factor, *K*, is estimated together with ß. In addition, the present simulations process α in a way that is closer to real data analysis situations. α is the rate of false *multiple* hypotheses instead of *simple* hypotheses as simulated earlier (von Weber, 2000).

The following tables present selected results of the larger simulation. Specifically, the tables present the tests that display the smallest ß-estimates for a simulation condition. Results for τ = 2 are omitted[5]. For Lindner's test, not all variations of table parameters were evaluated because this test can be prohibitively computationally intensive, in particular when there are many degrees of freedom or when the table is spanned by more than 3 variables. For the new procedure of Lautsch and von Weber (2002), there exists no two-stage search. However, the new version of the program SICFA which can also be obtained from the lead author (see Footnote 5), generates the complete set of rejected null hypotheses for all nominal α-levels between 0.001 and 0.1.

---

[5] The complete results can be obtained from the lead author upon request. Please send an E-Mail to webers@fh-furtwangen.de

The following 7 tables use the same abbreviations and labels as before in this article. *df* indicates degrees of freedom, α is the a priori determined multiple type I error, τ is the a priori determined maximum type strength. *Test* indicates the test used, with an * indicating that the two-stage procedure was performed. ß is the Type II error (given in percent), and *K* is the estimate of the correction constant for the test that was employed. If *K* is positive, this test suggests conservative decisions about the existence of types or antitypes. Negative values of *K* indicate that a test suggests non-conservative decisions. The first column of the left-most result panel in each table presents the test with the smallest ß-error. The first column in the second result panel presents the test with the second-smallest ß-error, and the first column in the third result panel presents the test with the third-smallest ß-error. Note that, for a successful search for types, a ß no larger than 30% is desirable.

Table 2: ß for *df* = 4 and *d* = 2

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	Pe*	97.9	0.52	nPr	98.4	-0.59	Pe	98.8	-0.03
		3	Pe*	90.6	0.50	Pe	94.6	-0.12	nPr	98.2	-1.71
	0.05	1	Pe*	94.0	0.69	Pe	95.4	-0.05	nPr	96.3	-0.54
		3	Pe*	83.1	0.66	Pe	86.3	-0.16	nPr	91.7	-1.17
15	0.01	1	nPr	96.8	-1.40	Pe*	96.8	0.15	Pe	97.0	-0.26
		3	nPr	82.3	-1.97	Pe*	88.3	-0.05	Pe	88.8	-0.55
	0.05	1	nPr	88.7	-1.06	Pe	90.6	-0.31	Pe*	91.1	0.14
		3	nPr	66.1	-1.55	Pe	74.1	-0.61	Pe*	74.4	-0.10
50	0.01	1	nPr	94.4	-2.16	Kr	98.4	-1.35	Kr*	99.0	-0.63
		3	nPr	76.9	-4.21	Pe	94.5	-1.91	Pe*	96.6	-1.29
	0.05	1	nPr	80.7	-1.87	Pe*	88.6	-0.47	Pe	88.7	-0.97
		3	nPr	59.0	-3.58	Pe	67.2	-1.65	Pe*	69.7	-1.09

Table 3: ß for df = 4, d = 3

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	Pe*	97.6	0.52	Pe	98.3	-0.06	nPr	98.6	-0.64
		3	Pe*	91.9	0.46	Pe	92.1	-0.15	LK*	97.1	0.38
	0.05	1	Pe	93.8	-0.02	Pe*	94.0	0.66	nPr	94.5	-0.39
		3	Pe	80.8	-0.14	Pe*	84.9	0.70	nPr	89.9	-1.04
15	0.01	1	nPr	94.8	-1.72	Pe*	95.4	0.17	Pe	95.5	-0.29
		3	nPr	80.8	-1.72	Pe*	84.0	-0.06	Pe	89.5	-0.70
	0.05	1	nPr	85.2	-0.73	Pe	87.8	-0.32	Pe*	88.5	0.16
		3	nPr	60.3	-1.03	Pe*	71.6	-0.12	Pe	74.9	-0.77
50	0.01	1	nPr	90.8	-1.68	Li	95.5	0.23	Ld	95.6	-2.23
		3	nPr	53.2	-1.81	Ld	84.1	-6.21	Ld*	87.7	-2.03
	0.05	1	nPr	67.6	-1.18	Ld	86.5	-2.24	LK*	86.6	-0.51
		3	nPr	40.1	-1.20	LK*	72.6	-1.32	Ld	73.6	-5.54

Table 4: ß for df = 20, d = 2

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	Pe*	97.4	0.33	Pe	97.6	-0.05	LK	97.9	0.03
		3	Pe	83.1	-0.09	Pe*	84.6	0.31	LK	86.6	-0.10
	0.05	1	LK	94.3	0.02	Pe*	94.6	0.35	Pe	94.7	-0.03
		3	Pe	74.0	-0.08	LK	77.2	-0.07	Pe*	78.0	0.37
15	0.01	1	nPr	88.9	-0.45	Pe*	95.3	-0.03	LK*	96.8	-0.03
		3	nPr	66.7	-0.95	Pe*	69.1	-0.11	LK*	79.8	-0.30
	0.05	1	nPr	79.2	-0.35	Pe*	83.1	0.08	Pe	87.1	-0.28
		3	nPr	50.3	-0.40	Pe*	58.3	-0.08	Pe	58.7	-0.40
50	0.01	1	nPr	47.3	-0.35	Kr	97.5	-1.26	Kr*	99.1	-2.00
		3	nPr	27.2	-0.39	Kr*	85.1	-2.00	Ld	85.5	-8.72
	0.05	1	nPr	34.9	-0.31	Pe	79.3	-0.94	Pe*	79.4	-0.80
		3	nPr	18.7	-0.31	Pe*	49.8	-0.79	Pe	51.3	-1.12

Table 5: ß for df = 20, d = 3

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	LK	96.8	0.03	Pe*	97.0	0.32	Pe	97.2	-0.04
		3	Pe	79.5	-0.10	LK	82.4	-0.09	Li	83.5	0.40
	0.05	1	Pe	92.8	-0.02	LK	92.9	0.05	Pe*	94.2	0.33
		3	Pe	71.4	-0.06	LK	74.4	-0.08	Li	75.4	0.48
15	0.01	1	nPr	87.2	-0.40	Pe*	91.1	0.11	LK	91.4	-0.11
		3	Li	64.0	0.21	Pe	65.4	-0.16	Pe*	67.5	0.02
	0.05	1	nPr	75.4	-0.26	Li	80.6	0.44	Pe	82.0	0.05
		3	nPr	47.4	-0.31	Pe	55.0	-0.11	Li	55.1	0.32
50	0.01	1	nPr	42.6	-0.27	Pe*	71.5	0.03	Pe	71.6	-0.24
		3	nPr	24.7	-0.31	Pe	49.0	-0.48	Pe*	50.1	-0.22
	0.05	1	nPr	31.9	-0.25	Pe	46.9	-0.07	Li	47.8	0.31
		3	nPr	17.0	-0.26	Pe	42.5	-0.41	Pe*	44.4	-0.26

Table 6: ß for $df = 20$, $d = 4$

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	Li	97.4	0.43	Pe*	97.4	0.54	Pe	97.5	0.17
		3	Li	84.0	0.40	Pe	86.1	0.12	Pe*	87.0	0.52
	0.05	1	Pe	93.7	0.20	Li	93.8	0.52	nPr	94.3	-0.35
		3	Pe	75.4	0.17	Li	75.9	0.48	Ld	77.4	-0.19
15	0.01	1	nPr	86.6	-0.39	Li	90.1	0.34	Pe*	91.3	0.30
		3	Li	65.2	0.20	Pe*	67.6	0.18	Pe	71.1	-0.24
	0.05	1	nPr	74.8	-0.26	Pe*	80.2	0.30	Li	80.9	0.44
		3	nPr	48.8	-0.37	Li	56.1	0.32	Pe	57.8	-0.16
50	0.01	1	nPr	45.0	-0.31	Pe	65.9	-0.16	Li	68.1	0.16
		3	nPr	24.4	-0.31	Pe*	53.5	-0.30	Pe	53.6	-0.67
	0.05	1	nPr	31.3	-0.22	Li	48.3	0.31	Pe	49.5	-0.09
		3	nPr	16.2	-0.23	Li	44.8	-0.02	Pe	46.6	-0.59

Table 7: β for df = 80, d = 3

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	LK	96.4	-0.01	Li	96.9	0.27	Pe*	97.1	0.18
		3	Li	77.1	0.26	LK	78.1	-0.04	Pe	78.3	-0.02
	0.05	1	LK	92.3	0.03	Li	92.3	0.39	Pe	94.0	0.02
		3	Li	71.3	0.37	Pe	72.0	0.01	LK	72.7	0.01
15	0.01	1	Li	81.7	0.25	Pe	82.2	-0.03	Pe*	85.6	0.14
		3	Li	52.8	0.20	Pe	53.4	-0.07	Kr	59.3	-0.22
	0.05	1	Pe	71.8	0.01	Li	72.9	0.34	Kr	78.2	0.00
		3	Li	46.9	0.31	Pe	49.4	-0.06	Kr	50.8	-0.15
50	0.01	1	Li	39.2	0.23	Pe	44.4	-0.02	Kr	44.5	-0.15
		3	nPr	34.7	-0.64	Pe	37.2	-0.16	Li	38.2	-0.07
	0.05	1	Li	31.0	0.32	Kr	33.1	-0.12	nPr	35.4	-0.25
		3	nPr	23.7	-0.41	Li	27.4	0.12	Kr	30.4	-0.89

Table 8: ß for df = 80, d = 4

mf	α	τ	Test	ß	K	Test	ß	K	Test	ß	K
5	0.01	1	Pe	96.6	0.01	Pe*	96.9	0.27	Li	97.0	0.26
		3	Li	77.2	0.25	Pe	77.5	-0.01	Kr	80.8	0.10
	0.05	1	Li	92.5	0.38	Pe	92.7	0.04	Pe*	94.7	0.26
		3	Li	71.3	0.37	Pe	71.6	0.03	Kr	74.9	0.11
15	0.01	1	Li	80.6	0.25	Pe	81.9	-0.04	Pe*	85.0	0.14
		3	Li	52.9	0.20	Pe	53.7	-0.07	Kr	60.0	-0.24
	0.05	1	Li	70.9	0.35	Pe	72.6	-0.01	nPr	75.9	-0.19
		3	Li	46.3	0.30	Pe	49.4	-0.07	Kr	50.2	-0.17
50	0.01	1	Li	39.4	0.23	Kr	45.1	-0.18	Pe	45.3	-0.03
		3	nPr	36.6	-0.66	Li	38.3	-0.08	Pe	38.4	-0.20
	0.05	1	Li	30.9	0.32	Kr	33.9	-0.15	nPr	36.8	-0.26
		3	nPr	24.4	-0.43	Li	27.4	0.12	Kr	30.4	-0.89

5. Results, discussion and recommendations

Tables 2 through 8 show that the ß-estimates are often large, in particular for tables with small samples ($mf = 5$). Types are close to impossible to identify if type strength is no larger than $τ = 1$. This is surprising, because a type strength of $τ = 1$ implies that an expected frequency of 5 now is expected to be 10. We now ask what the reasons are for this surprising result.

The main causes for large α- or β-errors include:

- The drawing error of a cell with assumed normal distribution, $σ_{ijk} = \sqrt{\hat{e}_{ijk}}$, which can be substantial when the expected cell frequencies are small. Consider the following example. If the estimated expected cell frequency of a type cell is $\hat{e} = 9$, the drawing error is $σ = 3$. That is, with probability 95%, we draw a frequency from the interval [e-2σ, e+2σ], or between 3 and 15. In this case, the frequency of 3 is already below the mean frequency of 5, and thus points in the direction of a possible antitype. There is only one option to decrease this source of a large β-error, that is, to increase the total sample size.

- The drawing error, $\sqrt{Np_{i.}}$ of a row sum. Again, this error can be substantial when the sample size or the size of the table is small. The marginal sums are used to estimate the a-priori probabilities of the cells. Thus, drawing errors in the marginal sums cause errors in the estimates. Here again, the only option to reduce the β-error is to increase the sample size.
- Although antitypes appear more frequently than types when samples are large (von Eye, 2001a), they are still harder to detect reliably. Lautsch and von Weber (2002) present β-estimates that support this result. The present simulations created a proportion of only 33% antitypes, which contrasts with the expected proportion of 50%. Thus, the resulting β-estimates for antitypes are lower. Thus, if researchers focus on the search for types at the expense of antitypes, the β-estimates may be much better than indicated in Tables 2 - 8.

One reason for the increased -levels and, consequently, the large correction constants *K*, can be seen in the occurrence of so-called *phantom types* and *phantom antitypes* in the neighborhood of strong type cells. Phantom types and antitypes result for the following reason. Type cells come with large observed cell frequencies. Large individual cell frequencies imply that the marginals that describe this cell are also relatively large. The magnitude of these marginals may be a result of the existence of the type and thus of a violation of the main effect model itself. Now, in tables with more than two dimensions, whenever large marginals intersect, the estimated expected cell frequencies are large too. The types and antitypes that result for these intersections have been suspected to be phantoms, that is, *false types* or *false antitypes*. Kieser and Victor (1999) and Lautsch and von Weber (2002) have found satisfactory algorithms for the simultaneous identification of all type cells.

Further inspection of Tables 2 through 8 suggests:
- A large strength τ decreases β. This decrease can be declared by the fact that types are easier to identify if a cell frequency is high. The existence of phantom types diminishes this effect.
- Larger samples reduce β because the drawing errors will be relatively smaller and results become more reliable.
- Often, a small α implies a large β, and vice versa. A small number of tested null hypotheses (small α) implies that the number of rejected null hypotheses must be small (large β). Inversely, many tested hypotheses (large α) implies that many types can be detected (small β).
- The errors α and β do not vary greatly with the number of variables, *d*.

Finally, we ask whether the present simulation results allow us to make recommendations as to which test to use. We conclude:
- Perli, Hommel, and Lehmacher's test (1995) is obviously most capable, that is, comes with small α- and β-errors when the sample size is small. The two-stage procedure improves the test's characteristics even more.
- Lautsch and von Weber's new procedure which represents an algorithmic adaptation of Victor's CFA, in combination with Dunkl and von Eye's test (1990) is most useful for $df \le 20$ and $mf \ge 15$.
- Lienert's χ-component test is often viewed as a tool that can be used only in exploratory contexts. The results presented here may change this perspective. As Perli et al. (1985) point out, the number of independent hypotheses in a table is given by the model's degrees of freedom. Thus, each χ-component of a chosen set of cells carries a full degree of freedom (instead of a fraction of a *df*). In addition, the concern that the χ^2-distribution can be used only

when the drawing errors are normal seems to carry less weight when the present consideration concerning mixture distributions is taken into account. A third criticism of Lienert's test is that it may be too conservative. This objection is certainly correct, at least for tables of certain sizes (von Eye, 2002), because the present simulations show only positive K-scores for the χ-test. However, the presentation of a tailored continuity correction in the present article remedies this problem entirely. As a consequence, Lienert's test moves dramatically up in the performance rank order of tests. It is a well-performing test in cases with $df > 40$.

Future software developments as the anticipated upgrade of SICFA may, based on the present results, not only make the proper test available, but also information about which α to select. If the user chooses an α that is too small, chances of detecting types and antitypes are diminished unnecessarily. In addition, programs may make available estimates of the magnitude of β and the maximum type strength, τ. Finally, programs may be able to estimate the optimal continuity correction constant K to protect the user against overly conservative or non-conservative behavior of a particular test.

References

1. Diaconis, P., and Efron, B. (1983). Statistik per Computer: Der Münchhausen-Trick. Spektrum der Wissenschaft, Heft 7.
2. Dunkl, E. and von Eye, A. (1990). Kleingruppentests gegen Victor-Typen und -Syndrome. Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie, 44, 46 - 51.
3. Herrendörfer, G., et al. (1982). Robustheit von Tests für 2x2-Kontingenztafeln, Probleme der angewandten Statistik. Heft 7, Forschungszentrum für Tierproduktion Dummerstorf-Rostock.
4. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.
5. Indurkhya, A., and von Eye, A. (2000). The power of tests in Configural Frequency Analysis. Psychologische Beiträge, 42, 301 - 308.
6. Kareev, Y. (1992). Generation of random sequences. Journal of Experimental Psychology. Human Perception and Performance, 18, 1189-1194.
7. Kieser, M., and Victor, N. (1991). A test procedure for an alternative approach to configural frequency analysis. Methodika, 5, 87-97.
8. Kieser, M., and Victor, N. (1999). Configural Frequency Analysis (CFA) Revisited – a New Look at an Old Approach. Biometrical Journal, 41, 967-983.
9. Krauth (1993). Einführung in die Konfigurationsfrequenzanalyse. Weinheim: Beltz.
10. Krauth, J., and Lienert, G. A. (1973). KFA. Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin. Freiburg: Karl Alber.
11. Küchenhoff, H. (1986). A note on a continuity correction for testing in three-dimensional configural frequency analysis. Biometrical Journal, 28, 465-468.
12. Lautsch, E. und von Weber, S. (1990). Die Konfigurationsfrequenzanalyse - Methoden und Anwendungen. Berlin: Volk und Wissen.
13. Lautsch, E., und von Weber, S. (1995). Methoden und Anwendungen der Konfigurationsfrequenzanalyse. Weinhein: Beltz.
15. Lautsch, E., and von Weber, S. (2003). A new test procedure in the CFA. (in this volume).
16. Lehmacher, W. (1981). A more powerful simultaneous test procedure in configural frequency analysis. Biometrical Journal. 23, 429 - 436.

17. Lienert, G. A. (1969). Die Konfigurationsfrequenzanalyse als Klassifikationsmittel in der klinischen Psychologie, In: M. Irle (Hrsg.), Bericht über den 26. Kongress der Deutschen Gesellschaft für Psychologie 1968 in Tübingen (pp. 244 - 255). Göttingen: Hogrefe.

18. Lindner, K. (1984). Eine exakte Auswertungsmethode zur Konfigurationsfrequenzanalyse. Psychologische Beiträge, 26, 393 - 415.

19. Perli, H.-G. (1985). Testverfahren in der Konfigurationsfrequenzanalyse. Erlangen: Palm und Enke.

20. Perli, H.-G., Hommel, G. and Lehmacher, W. (1985). Sequentially rejective test procedures for detecting outlying cells in one- and two-sample multinomial experiments. Biometrical Journal, 27, 885-893.

21. Victor, N. (1989). An alternative approach to configural frequency analysis. Methodika, 3, 61-73.

22. von Eye, A. (2002a). The odds favor antitypes - A comparison of tests for the Identification of configural types and antitypes. Methods of Psychological Research - online, 7, 1-29.

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

24. von Eye, A. (2003a). A comparison of tests used in 2 x 2 tables and in two-sample CFA. Psychology Research (in press).

25. von Eye, A. (2003b). A comparison of residual measures under Configural Frequency Analysis conditions. In R. Perera, A. González, and K.A. Anaya (Eds.), Memorias del XVII Foro Nacional de Estadística. (in press)

26. von Eye, A., and Mun, E.-Y. (2003). Characteristics of measures for 2 x 2 tables (in preparation)

27. von Eye und Rovine, M. J. (1988). A comparison of significance tests for configural frequency analysis. EDV in Medizin und Biologie, 19, 6-13

29. von Eye, A., Schuster, C., and Gutiérrez-Peña, E. (2000). Configural Frequency Analysis under retrospective and prospective sampling schemes - frequentist and Bayesian approaches. Psychologische Beiträge, 42, 428 - 447.

29. von Weber, S. (2000). Ein Vergleich in der KFA verwendeter Tests mittels Simulationsrechnungen. Psychologische Beiträge, 42, 260-272.

E. Keitel, G. Süß, R. Gunzenhäuser, A. Hahn (Hrsg.)

Neue Medien im Alltag: Computerspiele – Eine Provokation für die Kulturwissenschaften?

E. Keitel: Einleitung: Zum Genre Computerspiel
G. Süß: Filmstudien und Computerspielanalyse
R. Gunzenhäuser: Kulturwissenschaften und Computerspiele
K. Wenz: Spiel, Computerspiel und digitale Medienkunst
R. Gunzenhäuser: Massenhafte Immersion: Rollenspiele im Netz
R. Gunzenhäuser: Computerspiele als Herausforderung an die Kulturwissenschaften