A new test procedure in the CFA

Erwin Lautsch[1], Stefan von Weber[2]

Summary

One of the aims of the Configural Frequency Analysis (CFA) is the identification of symptom configurations as types or anti-types. Following the pattern of an efficient regression algorithm of Cierzynski and von Weber the authors developed a gradient method, which minimizes iteratively the total χ^2 of a contingency table, reducing by a small amount the frequency of the most suspicious cell in the step of iteration (in the case of an anti-type the frequency will be increased). The final result is a table, which fulfills perfectly the hypothesis of independence. The expectation values one can calculate from this table are known as Victor-expectation values in the literature. With these values and the original cell frequencies the trustworthy small-group test of Dunkl and von Eye is performed and then confirmed by Holm's procedure. With a Bayesian ansatz (types are hidden by preference in highly frequented cells) a further improvement of the results has been reached. Numerical simulations are showing the correctness of the method. A comparison with the results of the "new approach" of Kieser and Victor is showing a large measure of conformity.

Key words: Contingency table, Configural Frequency Analysis, CFA, search for types, test, test procedure, simulation

---

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

[2] 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

1. Introduction

Since the foundation of the Configural Frequency Analysis (CFA) by G. A. Lienert (1969) a lot of publications have been done concerning an improvement of the CFA methodology. J. Krauth developed the exact binomial test (Krauth and Lienert 1973). Habermann published his residual test in 1978, Lehmacher his hypergeometrical cell test in 1981. Also in 1981 Lehmacher developed (contemporarily with Krauth) another exact test both using the Freeman-Halton principle. In 1986 Bergman and von Eye published a numerical approximation for the binomial test using Stirling's formula for $\ln(N!)$. Perli et al. (1985) published an asymptotical cell test. In 1984 Lindner published a generalisation for 2^d-tables of Fisher's exact hypergeometrical test for 2x2-tables. This enumeration is far away from to be complete, but it wants to recall only to important mile-stones of the development of the CFA.

As a result of computer simulations (e.g. Lindner 1984) one knows that under real conditions (i.e. the expectation values of the cells are estimated from the marginal sums of the contingency table) some tests are showing a conservative behaviour, other tests an anti-conservative behaviour. *Conservative* means here that existing types are not recognized. *Anti-conservative* means here that harmless symptom configurations are identified falsely as types. One reacted on such behaviour with different methods. So, e.g., in 1986 Küchenhoff proposed a continuity procedure for the residual test of Lehmacher. It reduces in the case of small expectation values the danger arising from anti-conservative behaviour. Other authors avoid on principle cells with an expectation value $\hat{e}_{ij} < 5$.

N. Victor published in 1989 an essential contribution to the methodology of the CFA. He criticized the common practice to use type cells (or anti-type cells) in the calculation of the expectation values without change of their frequency. One calculates expectation values under the hypothesis of independence H_o, but type cells (or anti-type cells) are outliers in the sense of H_o. Victor proposed the use of the Deming-Stephan algorithm for the estimation of the independence expectation values of the cells (we call these values Victor-expectation values). In 1991 M. Kieser and N. Victor made another step. They published a test procedure for the contemporary search for all types (and anti-types). This procedure was used also with good success in the CFA-software SICFA (as an appendix in Lautsch and von Weber 1990, 1995) of the authors. Two studies concerning the test effectiveness (von Weber 2000, Lautsch and von Weber 2003 (in preparation)) gave the results, e.g., that just the procedure of Victor and Kieser often yielded the best values.

In 1999 Kieser and Victor published an improved version of their test procedure. New are the following points:

- The CFA is now based on the analysis of a loglinear model with outliers,
- There test statistics exist which test the hypothesis of independence after elimination of the outliers (types or anti-types) on the one hand, and which test the significance of the chosen set of types on the other hand,
- There are proposed two choice strategies, whereby the "two-stage forward inclusion" gives the most practicable impression to us, the finding of type cells too. Therefore we should correct conservative test behaviour just as anti-conservative behaviour - the first to diminish the β-error, the second to keep the demanded α.

Error type I (or α) is in this study the probability that a multiple hypothesis is proved to be wrong, i.e., in the selected set of types is one (or some) false type or anti-type. Error type II (or β) is the probability that a type cell or an anti-type cell is not recognized. In the simulation studies α is the quotient of the number of tables which failed and the total number of all analysed tables. Error β is the quotient of the number of types (and anti-types) not found to the number of all existing types (and anti-types).

2. A new test procedure

The authors are in agreement with Kieser and Victor (1999) in that sense that only few symptom configurations of a contingency table can be really types or anti-types. If one estimates the expectation values of the cell frequencies of a contingency table from the marginal sums (what is true in nearly all practical cases), then the maximum number of possible independent hypotheses is equal to the degree of freedoms DF of the table (c.f. Perli et al. 1985). But to find such a number of types one should try only in extreme situations. By the definition of G. A. Lienert a *contingency type* is a symptom configuration which appears with a frequency higher than what is expected by chance or according to an assumed distribution. But this expectation value used in the comparison one can calculate only if enough non-type cells exist. These cells define the level, which the type cells exceed and the anti-type cells do not reach.

The new procedure of the authors has been proved in computer simulations. The procedure consists of two main steps:

- the estimation of the Victor-expectation values $\hat{e}^*_{ij}$ with implicit choice of types,
- the calculation of the local test statistics and the confirmatory proof of the set of chosen types by Holm's procedure.

During the estimation of the $\hat{e}^*_{ij}$ the aim is to minimize the χ^2-statistic of a contingency table $\{n^*_{ij}\}$ cleared from outliers, i.e., to make the table really *independent*. It is the same aim which Kieser and Victor (1999) followed. The crux is to find an indicator which original cell frequency n_{ij} we had to increase and which cell frequency n_{ij} we had to decrease. The only objective criterion is the quote of correctly identified types under observance of the given α. We know similar issues and their resolution from the theory of non-linear optimization, but also from the variable selection strategies of the multiple regression analysis (c.f. Goldstein und Dillon 1984 or Weisberg 1986 or Cierzynski and von Weber 1992).

Our algorithm starts with a fictitious value of $\chi^2_o = \infty$ and with the original contingency table $\{n^o_{ij}\} = \{n_{ij}\}$. The upper index is a consecutive enumeration of the steps. Similarly to the Deming-Stephan algorithm our algorithm works iteratively, too.

At the start of the loop: From the n^k_{ij} we calculate the marginal sums $n^k_{i.}$, $n^k_{.j}$ and the total sum n^k_{tot}. These numbers must not be integers in the case of higher iteration step numbers $k > 1$. From the marginal sums and from n^k_{tot} we estimate the (preliminary) expectation values $\hat{e}^k_{ij}$ as usual.

$$(1) \qquad \hat{e}^k_{ij} = \frac{n^k_{i.} n^k_{.j}}{n^k_{tot}}$$

(The formulas are given here for the case $d = 2$ by reasons of simplicity. Extension to cases $d > 2$ is easy.) We calculate a differential Δn^k_{lm}, i.e., a small variation of a selected cell (l,m) in the sense that we maximize the decrease of the χ^2 (maximum decrease). At the first attack we chose the cell (l,m) with maximum decrease $\Delta \chi^2_{k,lm}$ of its χ^2-component:

$$(2) \qquad \Delta \chi^2_{k;lm} = \overset{max}{\underset{ij}{}} \left(\chi^2_{k;ij} - \chi^2_{k;ij,\Delta} \right)$$

with the χ^2-component without variation of its frequency n^k_{ij}

$$(3) \qquad \chi^2_{k;ij} = \frac{(n^k_{ij} - \hat{e}^k_{ij})^2}{\hat{e}^k_{ij}}$$

and a χ^2-component with variation of its frequency n^k_{ij}

(4) $\chi^k_{k;ij} = \dfrac{(n^k_{ij} - \Delta n^k_{ij} - \hat{e}^k_{ij})}{\hat{e}^k_{ij}}$

The simulations have shown that an additional appraisement is very important for the differential Δn^k_{ij} in the sense of Bayes. The absolute amounts of the expectation values correlate strongly with the probability that these cells are real existing types. The most simple expression which yields such a weight is

(5) $\Delta n^k_{ij} = \Delta n^o \, \dfrac{\hat{e}^k_{ij}}{\hat{e}^k_{max}} \, sign(\hat{e}^k_{ij} - n^k_{ij})$.

The sign-function gives the direction of the change of frequency, i.e., the frequency n^k_{ij} will be decreased if the expectation value $\hat{e}^k_{ij}$ is less than the frequency n^k_{ij}, and it will be increased if $\hat{e}^k_{ij} > n^k_{ij}$. Δn^o is a stepwidth we had to chose properly.

In all cases of cells with a higher type probability, i.e., if $n_{ij} > \bar{N}$ and $n^k_{ij} > \hat{e}^k_{ij}$, then Δn^k_{ij} will be multiplied additionally with a factor F_t

(6) $F_t = 1 + \dfrac{n_{ij} - \bar{N}}{n_{ij,max} - \bar{N}}$.

Here $\bar{N}$ is the averaged cell frequency (or mean frequency mfr) and $n_{ij,max}$ is the maximal frequency of the original contingency table. In all cases of cells with a higher anti-type probability, i.e., if $n_{ij} < \bar{N}/5$ and $n^k_{ij} < \hat{e}^k_{ij}$, then the Δn^k_{ij} will be calculated in the sense of Bayes and independently on the formulas above by the following relation:

(7) $\Delta n^k_{ij} = \Delta n^o \left(\dfrac{\hat{e}^k_{ij} - n^k_{ij}}{\hat{e}^k_{ij}} \right)$.

The selected cell frequency n^k_{lm} will be increased or decreased, respectively, by the value of Δn^k_{lm}. Cells with an expectation value of $\hat{e}^k_{ij} < 0.6$ are taken only if the relation $\Delta n^k_{ij} > 0$ holds, i.e., if the cell frequency will be increased (and together with it the expectation value $\hat{e}^k_{ij}$). Now we increase the number k of the iteration step and calculate χ^2_k as sum of the new χ^2-components $\chi^2_{k,ij}$. The *end of the iteration cycle* is reached.

The loop between *start* and *end* will be repeated until the stop-relation holds $\chi^2_{k+1} \geq \chi^2_k$. After stop of the iteration we have got an estimation of the Victor-expectation values $\hat{e}^*_{ij}$, similarly to the use of the Deming-Stephan algorithm. But the above description of the new algorithm shows that there is not selected explicitly a set of types (and/or anti-types). The second part of the test procedure is known routine, i.e. the calculation of the local test statistics of Dunkl and von Eye (1990) and the confirmatory proof of the set of hypotheses found.

3. Simulation experiments and results

One can find the effectiveness of a statistical test or of a statistical test procedure only by means of well constructed test data with known properties. So, as a clock will not become more precise if we read the time often, so a test will not become better if one uses it often. Therefore, generating test data, one had to consider the following points:

- The number and position of types (or anti-types, respectively) must be given (known)
- Types (and anti-types) should be scattered randomly over the contingency table and should be of different strength
- The distribution of the frequencies n_{ij} around their expectation values $ê_{ij}$ has to be in accordance with the reality. In our opinion the supposition of a hypergeometrical or binomial distribution is a fiction. Caused by a lot of different influences during data sampling, but also caused by the simple fact of calculation of marginal sums from the observed frequencies, we expect a blended distribution from binomial and normal distribution
- Observed parameters of the contingency table, as dimension d, mean frequency mfr, or degrees of freedom dg, but also parameters which we put, as the given error type I α, or the maximal type strength τ_{max}, should be considered during simulation.

The authors developed a powerful table generating program. It produces contingency tables until dimension d = 5 and degrees of freedom with a value of df = 4 to df = 90. We resolved the issue of the blended distribution so that the program uses three different distributions with equal probability - the normal distribution and two different binomial distributions. The last differ from another by the filling of the urn - once it is the smallest observed marginal sum, in the other case the total number of probands N_{tot}.

Table 1 shows first results of the new procedure together with results of well known tests. Only error type II (β) is printed, i.e. the probability that an existing type (or anti-type) will not be found.

Table 1:
ß or error type II for different tests and different contingency tables

d	df	mfr	α	τ_{max}	Li	Kr	LK	Pe	Ld	Vi	nP	nP$_{Typ}$	nP$_{At}$
2	4,4	5	0,05	2	94,61	96,49*	95,43	**88,17***	95,99	96,79	92,13	89,21	99,34
2	21,7	15	0,05	2	75,24	78,29*	72,13	66,35*	77,68*	80,93	**59,39**	44,21	98,31
3	21,7	15	0,05	2	62,91	70,29	69,96*	62,68	68,08	62,04	**55,28**	39,48	96,14
4	86,0	50	0,05	2	24,47	27,29	88,92*	34,63	fehlt	21,87	**20,75**	7,20	52,52

In table 1 **d** is the dimension of the contingency table, **df** are the degrees of freedom, **mfr** is mean frequency, α is the given error type I, τ_{max} is the given maximal strength of type, **Li** is Lienert's χ^2-component test, **Kr** is Krauth's exact binomial test, **LK** is Lehmacher's hypergeometrical test with continuity correction by Küchenhoff, **Pe** is Perli's et al. asymptotical test in its simple form, **Ld** is Lindner's exact hypergeometrical test (here in the cases of great tables the results were not computed by the reason of computing time), **Vi** is an old test procedure induced by Victor and Kieser with χ^2-controlled stepwise selection of a set of types with local cell tests by Dunkl and von Eye, **nP** is the new procedure of the authors with local cell tests by Dunkl and von Eye. The last two columns both are concerning **nP** - once the probability of not finding types (**nP$_{Typ}$**) and otherwise of not finding anti-types (**nP$_{At}$**). An asterisk (*) marks results with randomly halvened sample (with part 1 of the halvened sample one seeks exploratively a set of hypotheses, with part 2 one proves confirmatorily this set). All results are proved confirmatorily by Holm's procedure. An additional correction to all tests ensures that the given α will be exhausted asymptotically, but it will not be exceeded. (The authors will give some more information to this correction in another study.)

One can see that the new procedure wins in three of four of the given table situations. Only in the case of small tables (df = 4 or df = 5) together with low frequencies (mfr = 5) the asymptotical test of Perli et al. is clearly better. In all, the results suggest that there exists only a small probability to find a type or anti-type (cf. Weber 2000). In the case of anti-types this is true. The finding of anti-types is not easy, since frequencies $n_{ij} < 0$ can not appear contrary to high frequencies in the case of types. Another reason is that the given strength of type τ changes much less the expectation value of an anti-type cell, compared with a type cell. If a type cell has under independence the expectation value $\hat{e}^*_{ij} = 10$, then a given strength of type $\tau = 2$ causes an increase by 20 and so a new value of $\hat{e}_{ij} = 10 \cdot (1+\tau) = 30$. In the case of an anti-type cell a given strength of type $\tau = 2$ causes a decrease by 6.7 only and a new value of $\hat{e}_{ij} = 10/(1+\tau) = 3.33$. The last two columns of table 1 show (by reason of space only for the new procedure) the different β-values of non-finding a type and of non-finding an anti-type. In the case of middle-sized tables with dimension d=3 the probability of finding a type is about 60% (i.e. 100-39.48). An anti-type will be detected by the new procedure with a probability of less than 4%. The authors will give a broad comparison of all tests used here in another study.

A simple example shall show how the new procedure works. Three symptoms M1, M2, M3, each of them with two categories, have got the marginal probabilities (0.4, 0.6), (0.7, 0.3) and (0.45, 0.55). Under independence and with $N_{tot} = 40$ we calculate for the eight cells (111, 112, ..., 222) the following expectation values (5.04, 7.56, 2.16, 3.24, 6.16, 9.24, 2.64, 3.96). For cell (111), e.g., we find $p1..*p.1.*p..1*N_{ges} = 0.4*0.7*0.45*40 = 5.04$. Table 2 shows the calculated results. A type of strength 2 at cell (112) increases the frequency of this cell (112) to a value of $7.56*(1+2) = 7.56*3 = 22.68$ and a type of strength 6 increases the frequency of cell (121) to a value of $2.16*(1+6) = 2.16*7 = 15.12$. The algorithm estimates the maximal type strength nearly exact with a value of $\tau_{max} = 5.93$. (1000 contingency tables of the same dimension, size and mean frequency are generated per one value of a given maximal type strength. From this 1000 tables the statistic $Max_{ijk}((n_{ijk} - \hat{e}_{ijk})/mfr)$ is calculated and compared with the same statistic of the original table.) If we set the error type I to $\alpha=0.05$, then we expect $\beta = 36\%$, using a correction constant K = -0.40. The Nijk are the rounded frequencies, the Eijk the common expectation values $\hat{e}_{ijk}$ calculated from the marginal sums, and the Vijk are the Victor-expectation values calculated by the new procedure. The Vijk are not far from the frequencies under independence. Type 121 is ensured at the multiple level 0.001, but the types 112 and 121 together we can ensure only at the multiple level 0.02. "Kor" is the value of the correction constant. It ensures asymptotically that we do not exceed the given α. Kor is yielded by a simple linear regression. The regression coefficients we get from a simulation. We compute the simulations for three α-levels 0.01, 0.05 and 0.25 (run-time about 4 minutes with 350 MHz and the given size of the table). The simulation is used to estimate both - the β-values and the regression coefficients. Both of the value-tripels are given in table 2.

Table 2:
Results of an artifically constructed example with types at 112 and 121

```
Dimension...........      3
Degrees of freedom...     4
Maximal type strength    5.93

alfa=  0.010    beta=   41.9%   Korrektur=  -0.47
alfa=  0.050    beta=   36.0%   Korrektur=  -0.40
alfa=  0.250    beta=   29.9%   Korrektur=  -0.34
```

i j k l m	Nijk	Eijk	Vijk	Kor	Tw	KIW	Typ/Antityp
1 1 1 . .	5	12.4	6.9	-0.34	0.483	1.000	0
1 1 2 . .	23	16.8	10.5	-0.44	2.491	0.020	1
1 2 1 . .	15	7.2	2.9	-0.56	3.862	0.001	1
1 2 2 . .	3	9.7	4.4	-0.34	0.377	1.000	0
2 1 1 . .	6	5.9	6.1	-0.34	0.012	1.000	0
2 1 2 . .	9	8.0	9.4	-0.34	0.027	1.000	0
2 2 1 . .	3	3.4	2.6	-0.34	0.022	1.000	0
2 2 2 . .	4	4.6	3.9	-0.34	0.004	1.000	0

Table 3 shows the results after smearing the frequencies of the constructed example with formula $N_{ijk}^* = N_{ijk} \pm z\sqrt{N_{ijk}}$. z are normal distributed random numbers with expectation 0 and variance 1.

Table 3:
Results of the constructed example after smearing the frequencies

i j k l m	Nijk	Eijk	Vijk	Kor	Tw	KIW	Typ/Antityp
1 1 1 . .	8	18.3	12.4	-0.38	0.790	1.000	0
1 1 2 . .	29	20.5	12.3	-0.66	2.737	0.010	1
1 2 1 . .	17	8.8	4.4	-0.73	3.208	0.005	1
1 2 2 . .	3	9.8	4.4	-0.38	0.512	1.000	0
2 1 1 . .	8	6.8	8.2	-0.38	0.055	1.000	0
2 1 2 . .	8	7.7	8.1	-0.38	0.079	1.000	0
2 2 1 . .	3	3.3	2.9	-0.38	0.021	1.000	0
2 2 2 . .	2	3.7	2.9	-0.38	0.177	1.000	0

One can see that type 121 is ensured still with 0.005 only, but both types together now with 0.01. Additionally, the Victor-expectation values are clearly more distant from the given frequencies under independence.

4. Examples with original data sets

We illustrate the working of our new test procedure using first the classic LSD data of G. A. Lienert (1970). The experiment investigated the incidence of Leuner's Psychotoxic Base Syndrome after the administration of lysergic acid diethylamide (LSD) to 65 volunteers. According to Leuner the reaction to LSD is dominated by a combination of the syndroms M1 = "clouded consciousness", M2 = "disturbed thinking" and M3 = "altered affectivity". For the participating volunteers, these three syndroms were rated as being either present ("2") or absent ("1"). The data of the experiment and the results calculated by the new procedure are shown in table 4.

Table 4:
LSD data (Lienert 1970) and CFA results of the new procedure

```
Dimension............          3
Degrees of freedom...          4
Maximal type strength      5.70


alfa=  0.010   beta=    46.7%   Korrektur=  -0.57
alfa=  0.050   beta=    37.7%   Korrektur=  -0.40
alfa=  0.250   beta=    31.0%   Korrektur=  -0.32
```

i j k l m	Nijk	Eijk	Vijk	Kor	Tw	KIW	Typ/Antityp
1 1 1 . .	20	12.5	2.2	-0.74	5.441	0.001	1
1 1 2 . .	1	6.8	2.0	-0.30	0.416	1.000	0
1 2 1 . .	4	11.4	4.0	-0.30	0.005	1.000	0
1 2 2 . .	12	6.2	3.6	-0.50	2.568	0.020	1
2 1 1 . .	3	9.5	3.0	-0.30	0.006	1.000	0
2 1 2 . .	10	5.2	2.7	-0.50	2.454	0.020	1
2 2 1 . .	15	8.6	5.4	-0.50	2.488	0.020	1
2 2 2 . .	0	4.7	4.9	-0.30	1.525	1.000	0

The null hypothesis for type 111 can be rejected at the multiple level of $\alpha = 0.001$, but the set of null hypotheses (111, 122, 212, 221) can be rejected at the multiple level $\alpha = 0.02$ only. Kieser and Victor (1999) had found here the type 111, too, but no further types. Otherwise, they could identify the cell 222 as an anti-type. This result of Kieser and Victor is of bribing beauty, although found in the sense of Bayes. We cite Kieser and Victor: „According to a priori postulation of the Psychotoxic Base Syndrome configuration (+++) by Leuner (1962), the alternative type model with T = {(+++), (---)} is fitted to the data (assuming the cell (+++) to be overfrequented and, as a consequence, the cell (---) to be underfrequented)." The use of Leuner's postulate is one reason, the different mathematical models may be another reason of the discrepancy between the results. Our new procedure does not select explicitly any type set during χ^2-decreasing, while Kieser and Victor are using the notion „a priori postulated type model". But nevertheless the LSD data will remain an interesting object also in future, and perhaps the next version of the new procedure will fit the beauty result of Kieser and Victor without any additional information in the sense of Bayes.

A further example of original data, which was used by Kieser and Victor, too, are the cancer data of Havemann et al. (1987). The efficacy of different types of chemotherapy was evaluated in four clinical trials involving a total number of 1127 patients which suffered from small-cell lung cancer. The proposed approach was used to investigate whether the early formation of distant metastases in liver (syndrome M1), skeleton (M2), brain (M3) and bone marrow (M4), which is encountered frequently in this type of carcinoma, occurs preferably in certain location patterns. Metastases are in table 5 coded either as absent ("1") or as present ("2").

Table 5:
Results to the cancer data of Havemann et al.

```
Dimension............         4
Degrees of freedom...        11
Maximal type strength      2.93


alfa=  0.010    beta=   21.4%   Korrektur=  -0.96
alfa=  0.050    beta=   15.8%   Korrektur=  -0.88
alfa=  0.250    beta=    8.2%   Korrektur=  -0.70
```

i j k l m	Nijk	Eijk	Vijk	Kor	Tw	KIW	Typ/Antityp
1 1 1 1 .	2	0.3	2.7	-0.72	0.194	1.000	0
1 1 1 2 .	8	3.2	11.7	-0.72	0.601	1.000	0
1 1 2 1 .	31	3.3	11.5	-0.92	2.880	0.020	1
1 1 2 2 .	53	32.3	50.4	-0.72	0.211	1.000	0
1 2 1 1 .	2	1.7	5.4	-0.72	0.779	1.000	0
1 2 1 2 .	15	16.1	23.8	-0.72	1.028	1.000	0
1 2 2 1 .	20	16.6	23.4	-0.72	0.397	1.000	0
1 2 2 2 .	104	161.5	102.7	-0.72	0.075	1.000	0
2 1 1 1 .	3	1.3	4.1	-0.72	0.272	1.000	0
2 1 1 2 .	8	12.2	17.9	-0.72	1.321	1.000	0
2 1 2 1 .	16	12.6	17.5	-0.72	0.208	1.000	0
2 1 2 2 .	67	122.7	77.1	-0.72	0.664	1.000	0
2 2 1 1 .	0	6.3	8.3	-0.72	1.578	1.000	0
2 2 1 2 .	64	61.0	36.4	-0.79	2.523	0.100	1
2 2 2 1 .	31	63.0	35.7	-0.72	0.455	1.000	0
2 2 2 2 .	703	613.0	157.0	-1.15	20.159	0.001	1

The new procedure selects the type 2222 at the multiple level of $\alpha = 0.001$, and at the multiple level of $\alpha = 0.02$ the type set (2222, 1121). Another type 2212 is detected, but not ensured because of its low significance (multiple level of $\alpha = 0.1$). Victor and Kieser had detected exactly the same types, but with other levels of significance: 2222 with $\alpha < 0.0001$, 2212 with $\alpha = 0.0016$ and 1121 with $\alpha = 0.022$. Despite these differences one can see that the two different approaches are leading to nearly identical results.

5. Résumé and conclusions

Regression analysis minimizes the sum of least squares between observed and estimated values of a dependent variable. The log-linear model is working similarly, but with the difference that here the sum of least squares of the logarithms is minimized between the observed and

estimated frequencies of a contingency table. In 1999 Victor and Kieser published in their paper "New Approach" a series of test statistics and methods, which one can use to detect and to remove type cells (or anti-type cells) in form of outliers. Now we can consider the remaining contingency table to be independent within the real meaning of the word.

It was our aim to write a comparable procedure, but in the context of the CFA and not in the context of the log-linear model. That means that we are working further with cell frequencies and their expectation values. What we had found is a kind of gradient method which decreases steadily the total χ^2 of the table and computes at the same time the so called Victor-expectation values, all without the knowledge of the type set and without use of the Deming-Stephan algorithm. The old procedure (cf. Weber 2000) computes the Victor-expectation values in cases only, where it has a type set selected, and uses for this numerical task the Deming-Stephan algorithm.

Simulations with constructed data showed that the new procedure has reached the goal. The originally put frequencies, i.e. that frequencies before we saddled a type or anti-type, have been estimated precisely by the new procedure. In all cases of original field data nobody knows, whether a type does really exist. But the new procedure finds in a large measure the same types, as the "new approach" of Kieser and Victor. Because these results are yielded by mathematically and algorithmically different models, the probability rises that real existing types have been detected.

Although the new procedure has got a very good rank in a greater comparing study, presented at the Lienert-meeting May 2002 in Kassel, Germany (von Weber, Lautsch and von Eye 2003), the authors see the discrepancies appearing with Lienert's LSD data, and will think about the future progress. In 1989 N. Victor made the proposal to prove all combinatorial type sets and to detect so in an explorative manner the type set with the highest probability. Then this type set is analysed confirmatorily. Meanwhile the speed of the computers is high enough, so that we could use this method. The limitation of the number of types to some few types per table, discussed by Kieser and Victor in their "new approach", is a good argument for the combinatorial search. Here the Deming-Stephan algorithm, perhaps slightly modified, would be used again.

References

1. Bergman, L. R. and von Eye, A.: A note on numerical approximation of the binomial test in configural frequency analysis, EDV in Medizin und Biologie, 17. 1969.
2. Cierzynski, T. C. and von Weber, St.: Simulation Experiments with CWA, a New Stable Regression Algorithm, and Comparisons with Two Other Stable Regression Algorithms, Statistical Software Newsletter 1992 (13) 4, 1992, 485-491.
3. Dunkl, E. and von Eye, A.: Kleingruppentests gegen Victor-Typen und -Syndrome, Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie 38. 1990.
4. Goldstein, M. and Dillon, W. (1984): Multivariate analysis methods and applications, New York: Wiley, 1984.
5. Habermann, S. J.: Analysis of qualitative data, Vol. 1., New York: 1978.
6. Havemann, K., Wolf, M., Holle, R., Gropp, C., Manke, H.-G., Hans, K., Schroeder, M., Heim, M., Victor, N., Georgii, A., Thomas, C., Pflüger, K.-H. and Bepler, G.: Alternating versus sequential chemotherapy in small cell lung cancer. A randomized German multicenter trial. Cancer 59, 1987, 1072-1082.
7. Holm, S.: A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics 6, 1979, 65-70.

8.   Kieser, M. and Victor, N.: A test procedure for an alternative approach to configural frequency analysis, Methodika 5, 1991, 87-97.
9.   Kieser, M. and Victor, N.: Configural Frequency Analysis (CFA) Revisited – a New Look at an Old Approach, Biometrical Journal 41, 1999, 967-983.
10. Krauth, J. und Lienert, G. A.: KFA - Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin, Alber, Freiburg/München, 1973.
11. Küchenhoff, H.: A note on a continuity correction for testing in three-dimensional configural frequency analysis, Biometrical Journal 28, 1986, 465-468.
12. Lautsch, E. und von Weber, St.: Die Konfigurationsfrequenzanalyse - Methoden und Anwendungen, Berlin: Volk und Wissen, 1990.
13. Lautsch, E. und von Weber, St.: Methoden und Anwendungen der Konfigurationsfrequenzanalyse, Weinheim: Beltz-Verlag, 1995.
14. Lehmacher, W.: A more powerful simultaneous test procedure in configural frequency analysis, Biometrical Journal, 23, 1981, 429-436.
15. Lienert, G. A.: Die Konfigurationsfrequenzanalyse als Klassifikationsmittel in der klinischen Psychologie, In Irle M.: Bericht über den 26. Kongress der Deutschen Gesellschaft für Psychologie in Tübingen 1968, Hogrefe, Göttingen: Hogrefe, 1969.
16. Lienert, G. A.: Konfigurationsfrequenzanalyse einiger Lysergsäurediäthylamid-Wirkungen, Arzneimittelforschung, 20, 912 (1970).
17. Lindner, K.: Eine exakte Auswertungsmethode zur Konfigurationsfrequenzanalyse. Psychologische Beiträge, 1984, Bd. 26.
18. Perli, H.-G., Hommel, G., and Lehmacher, W.: Sequentially rejective test procedures for detecting outlying cells in one- and two-sample multinomial experiments, Biometrical Journal 27, 1985, 885-893.
19. Victor, N.: An alternative approach to configural frequency analysis, Methodika 3, 1989, 61-73.
20. von Weber, St.: Ein Vergleich in der KFA verwendeter Tests mittels Simulationsrechnungen, Psychologische Beiträge, Bd. 42, Heft 3, 260-272, Pabst Science Publishers, 2000.
21. von Weber, St., Lautsch, E. and von Eye, A.: Tables-specific continuity corrections for configural frequency analysis, 2003, in this volume.
22. Weisberg, S.:Applied linear regression, Wiley: New York 1986.