**Scholars Journal of Agriculture and Veterinary Sciences (SJAVS)** **e-ISSN 2348–1854**
***Abbreviated Key Title: Sch. J. Agric. Vet. Sci.*** **p-ISSN 2348–8883**

A Unit of Scholars Academic and Scientific Society, India
www.saspublishers.com

# Investigation of Local Association in Animal Research for Multiway Cross Tabulated Count Data

**Mehmet Ilker BEK[1*], Ercan EFE[2]**
[1]Kahramanmaraş Sutcu Imam University, Department of informatics, Kahramanmaras, Turkey
[2]Kahramanmaras Sutcu Imam University, Faculty of Agriculture, Animal Science Dept, Division of Biometry and Genetics, Kahramanmaras, Turkey

**Original Research Article**

***Corresponding author**
*Mehmet Ilker BEK*

**Article History**
*Received: 11.06.2018*
*Accepted: 27.06.2018*
*Published: 30.06.2018*

**DOI:**
10.36347/sjavs.2018.v05i06.010

**Abstract:** In many animal researches mostly two way cross-tabulated data were evaluated by calculating a simple chi-squared test to see if the whole table deviated from the expected pattern. Such approaches to multiway frequency tables are mostly dissatisfying. Configural frequency analysis (CFA) is a general multipurpose tool and revolutionizes how we examine the cross-tabulation of two or more count variables [1]. CFA is a useful inferential tool used to evaluate the expected configural patterns in two-way to multiway cross tabulations of frequencies [2]. The results are possible types/antitypes depending on whether the observed cell frequencies are significantly lower/higher with respect to the base model. One of the goals of this study is to put the application of CFA technique into practice to investigate local associations for animal research in count data which was exemplified by simulated data for cause of mortality of calves. Also we compared the type and antitype cell numbers for different CFA approaches. The mortality data were not very well recorded in in Turkish cattle farming systems. Because of the difficulties of obtaining such risk data for calves' mortality [3], we used simulated data to evaluate the implementation of configural frequency analysis in calves' mortality. We have seen that CFA is a very powerful technique to investigate local configural associations, and CFA approaches must be evaluated for animal researches in the future to evaluate local relations.
**Keywords**: Local associations, Configural frequency analysis, Count Data analysis, Animal data analysis.

## INTRODUCTION

The standard analysis approaches of contingency tables which are used by animal science researchers to find out the relationships among categorical variables can be defined as Chi-Square decomposition techniques and log-linear modelling. Unfortunately such approaches to frequency tables are mostly dissatisfying.

CFA approaches can be applied in both exploratory and confirmatory research. CFA models are general multipurpose tools for analyzing categorical data. Configural frequency analysis and log-linear modeling are presented as cell-centered analytic approaches for the analysis of categorical or categorized data in multi-way contingency tables.

The development of CFA is proceeding at a rapid pace. The aim of this work is to introduce previously CFA methods to the animal science researchers who are not familiar with its benefits. In addition, recent developments in CFA method are presented along with an application and comparison between different approaches of CFA methods are carried out.

## MATERIALS AND METHODS

Configural Frequency Analysis (CFA) is a multivariate method for researches involving categorical variables [4]. CFA allows the researchers to focus on individual cells of a cross-classification instead of the variables that constitute this cross-classification. Results of standard methods of categorical data analysis such as log-linear modeling or logistic regression are expressed in terms of relationships among variables. In contrast, results from CFA are expressed in terms of cells of a table that are observed at different rates than expected under some base model. Therefore we need to find out cell based local relations. The patterns of categories that define a cell, that is, the cell indices, are called configurations [5]. If a cell contains significantly more cases than expected, it is said to constitute a CFA type. If a cell contains significantly fewer cases than expected, it is said to constitute a CFA antitype.

The classical approach specifies in the first step a base model and, then, examines either all cells or a selection of cells with the goal of finding those that contradict the base model. The Base Model, which can often be expressed in terms of a log-linear model [Log $(m_i) = X_i\lambda_i)$ ] which involves all variable relationships that are not of interest for the hypotheses under study, where $m_i$ is the array of expected frequencies in the cross-tabulation for configuration i, $X_i$ is the indicator matrix that contains all vectors needed for the intercept and all main effects, and $\lambda_i$ is the parameter vector. $\hat{m}_i$ is the estimated expected frequency for cell i, where i goes over all cells. Then, a general null hypothesis for CFA is: $H_0: E[n_i] = m_i$, where $n_i$ is observed frequency, $m_i$ is expected frequency estimated under some base model for configuration i. Exploratory CFA asks, under this null hypothesis, for each cell, whether the differences were statistically significant, and $E(n_i) > m_i$, that is said to constitute a CFA type, if $E(n_i) < m_i$, this cell i is said to constitute a CFA antitype. If in cell i, $E(n_i) = m_i$ this cell is called neutral cell, neither a type nor an antitype [6].

There are hierarchical and non-hierarchical versions of CFA[7]. In the classical non-hierarchical approaches the base models include two kinds of models, one is global models where all variables have the same status and the other is regional models where variables are grouped. The other type of CFA solution uses hierarchical models (HCFA). These models include two types of functional CFA solutions, fCFA and kv-CFA [8]. Each of the two kind of HCFA requires multiple CFA runs, because these approaches use a step down elimination technique in which cells were selected out that constitute types and antitypes. In contrast standard Configural Frequency Analysis (CFA) is a one-step procedure that determines which cells of a cross-classification contradict a base model. HCFA uses an iterative procedure that blanks out individual cells one at a time, until the base model fits or until there are no more cells that can be blanked out. A fitted final model describes the variable relationships within an incomplete table, that is, a table without the type and antitype cells. The base model for HCFA, thus changes to $Log(m) = X_s\lambda_s + X_f\lambda_f$ , the first part of this model is identical to the standard CFA base model, the second part of the model is functional part of the model. The functional part of the model is created in an iterative process. kv-CFA proposed by Kieser and Victor [9]. is the other hierarchical approach other than fCFA. In these models, fCFA approach involves a stepwise selection procedure; while kv-CFA approach involves forward inclusion routine. kv-CFA uses the overall goodness-of-fit LR-criterion; whereas fCFA blanks those cells out that are extreme based on the magnitude of residual scores.

In any CFA model to make a decision as to whether a cell constitute a CFA type or antitype, a number of statistical test has been proposed which were protected tests for test-wise $\alpha$. The choice of one of protected tests also affects the number of extreme cells, hence the final number of types and antitypes obtained from the same data set. Some other measurements such as RR (relative risk ratio) and Log (P) indicate individual characteristics of cells that constitute types and antitypes. These two coefficients are interpreted after the parsimonious solution obtained. The RR and Log (P) can be defined as.

$RR_i = n_i/m_i$ , where i indexes the cells in a cross-tabulation, it indicates relative frequency of the occurence of a configuration, given the expectation from the base model. Log(P) is defined, $Log(P_i) = -Log_{10}\big(Pr(X \geq n_i)\big)$ where $X \sim Poisson(m_i)$. Log(P) can be interpreted as the probability that the observed cell frequency is smaller than the expected cell frequency. The concordance of rank orders of these two statistics give some hints about the distribution of cell frequencies.

In this article the global non-hierarchical and hierarchical CFA methods and some statistical tests for the identification of types and antitypes, methods for protection of the family-wise $\alpha$ adjustment methods are illustrated and compared by using simulated artificial data for four categorical variables related with calf death. The rank orders of RR and Log(P) statistics are also interpreted for this data set.

The simulated experiment by employing the multinomial sampling schemes [10] on the calves' death includes four different categorical variables. The probability of observing the contingency table with cell frequencies $n_{1111}, n_{1112}, n_{1121}, n_{1122}, \ldots, n_{abcd}$, is given as the product of probabilities of observing each of (a, b, c, d) independent vectors with defined probabilities. A number of 900 calves obtained with different death probabilities under various conditions. The total number of observations calculated by the rule N=25xcell numbers [11]. The Table 1 is an asymmetrical table, because the marginal totals differ from each other.

The calves' death data obtained by simulation that are used to evaluate the classical and hierarchical model approaches and compared the number of types and antitypes that obtained for different statistical test with Bonferroni protection of test-wise $\alpha$ to define types and antitypes. Specifically, the choice of base model and critics of different model approaches of modeling in the CFA context of von Eye program and R program solution have been discussed.

**Table-1: The death frequencies data describe 900 calves that were breed in different barn management systems**

| | V(vaccination) | 1 | | 2 | | Total |
|---|---|---|---|---|---|---|
| | G(gender) | 1 | 2 | 1 | 2 | |
| B (barn system) | D (disease type) | | | | | |
| 1 | 1 | 44 | 68 | 39 | 63 | |
| | 2 | 51 | 83 | 15 | 23 | |
| | 3 | 6 | 8 | 3 | 3 | |
| 2 | 1 | 12 | 18 | 36 | 58 | |
| | 2 | 12 | 18 | 42 | 68 | |
| | 3 | 6 | 8 | 0 | 2 | |
| 3 | 1 | 24 | 38 | 33 | 53 | |
| | 2 | 3 | 3 | 3 | 3 | |
| | 3 | 15 | 25 | 6 | 8 | |
| Total | | | | | | N=900 |

B: Barn systems (# of categories=3; 1- The management system in which the mother and the other cattle and calves are kept together in the same barn, 2-: the barn system in which the cattle and calves kept in two group housing system, 3- the barns with individual calf pens keep calves separated from cattle) , D-Disease types (# of categories=3;: 1- respiratory system diseases, 2- digestive system diseases, 3- other trauma conditions), V- vaccination condition (# of categories=2;1- vaccinated, 2- non-vaccinated), G-Gender (# of categories=2; 1- Female, 2-Male)

The analyzes of the 3x3x2x2 cross tabulation of B (barn systems), D (disease types), V (vaccination) and G (gender) variables have been performed. The first categorical variable affecting calves' death is the conditions of housing barns for cattle to give birth and to raise their calf, shortly we call it barn systems with three levels (b1: The management system in which the mother and the other cattle and calves are kept together in the same barn with the death probability 45%, b2: the barn system in which the cattle and calves kept in two group housing system with the death probability 30%, and b3: the barns with individual calf pens keep calves separated from cattle with the death probability 25%). The second variable is the disease conditions caused to death with three levels (d1: respiratory system diseases with the death probability 54%, d2: digestive system diseases with the death probability 36%, and d3: other trauma conditions with the death probability 10%). The third variable is the vaccination condition with two levels (v1: vaccinated with death probability 49%, v2: not vaccinated with death probability 51 %,), and the last variable is the gender of calf with two levels (c1: female with death probability 39%, c2: male with death probability 61%). We illustrated the cross-tabulation of simulated artificial data in Table 1.

The data illustrated in Table 1 were arranged in a different way to use log linear models and CFA analysis. The data file must be restructured to be readable for the von Eye CFA program and also R program.

**RESULTS**

Base Model Definitions; Log-linear models are typically applied to find out the relationships among variables of two-way or multiway cross-classifications of categorical variables [12,13]. The Base Model, which can often be expressed in terms of a log-linear model that involves all variable relationships that are not of interest for the hypotheses under study [14].

The base models can include either global models where all variables have the same status or regional models where variables are grouped. The grouped variables can have different status. The Table 3 illustrates the available log-linear models which consider four variables (B, D, V and G). The base models of zero order CFA considers only chance effect, no variable effect. In this case all the expected values of cells are equal each other. If we consider four variables B, D, V and G in bracket notation the base model for first order CFA can be expressed as $[B],[D],[V],[G]$. The base model for second order CFA of these four variables can be defined as $[BD],[BV],[BG],[DV],[DG],[VG]$; this is hierarchical log-linear model, that is, when higher order effects are taken into account, all lower order effects of the variables included in the higher order effect terms are implied.

Equivalently we can illustrate this model by a long definition;

$$Ln\ (m_{ijk}) = \lambda_0 + \lambda_i^B + \lambda_j^D + \lambda_k^V + \lambda_l^G + \lambda_{ij}^{BD} + \lambda_{ik}^{BV} + \lambda_{il}^{BG} + \lambda_{jk}^{DV} + \lambda_{jl}^{DG} + \lambda_{kl}^{VG} \ .$$

**Table-2: Log-linear model definitions in global CFA for four variable (B,D,V and G) existence situation**

| Log-frequency model specification * | Definition |
|---|---|
| $Ln\ (m_{ijk}) = \lambda_0$ | Zero order base models. This model includes no variable effects, only chance effect exists. |
| $Ln\ (m_{ijk}) = \lambda_0 + \lambda_i^B + \lambda_j^D + \lambda_k^V + \lambda_l^G$ | First order base model. This model includes only main effects. Model of variable independence. |
| $Ln\ (m_{ijk}) = \lambda_0 + \lambda_i^B + \lambda_j^D + \lambda_k^V + \lambda_l^G + \lambda_{ij}^{BD} + \lambda_{ik}^{BV} + \lambda_{il}^{BG} + \lambda_{jk}^{DV} + \lambda_{jl}^{DG} + \lambda_{kl}^{VG}$ | Second order base model. This model includes main effects, first and second order associations. |
| $Ln\ (m_{ijk}) = \lambda_0 + \lambda_i^B + \lambda_j^D + \lambda_k^V + \lambda_l^G + \lambda_{ij}^{BD} + \lambda_{ik}^{BV} + \lambda_{il}^{BG} + \lambda_{jk}^{DV} + \lambda_{jl}^{DG} + \lambda_{kl}^{VG} + \lambda_{ijk}^{BDV} + \lambda_{ijl}^{BDG} + \lambda_{jkl}^{BVG} + \lambda_{jkl}^{DVG}$ | Third order base model. This model includes main effects, first, second and third order associations. |
| $Ln\ (m_{ijk}) = \lambda_0 + \lambda_i^B + \lambda_j^D + \lambda_k^V + \lambda_l^G + \lambda_{ij}^{BD} + \lambda_{ik}^{BV} + \lambda_{il}^{BG} + \lambda_{jk}^{DV} + \lambda_{jl}^{DG} + \lambda_{kl}^{VG} + \lambda_{ijk}^{BDV} + \lambda_{ijl}^{BDG} + \lambda_{jkl}^{BVG} + \lambda_{jkl}^{DVG} + \lambda_{ijkl}^{BDVG}$ | Forth order base model. This is saturated model that includes main effects and all possible interaction effects. |
| *: The subscripts (*i, j, k, l*) are index the estimated parameters, and the superscripts (*B, D, V, and G*) index the variables, $\lambda_0$ is the intercept. | |

The parameters not estimated are set equal to zero for CFA base model. The global CFA base model assigns all four variables the same status. Every log-linear model can be considered as a CFA base model. Four main effects (B, D, V and G) are part of the base model and cannot be reason for the emergence of types and antitypes, because this model is variable independence model. The standard base model is used when estimating expected values of cross tables for CFA approach, this base model is a model that types and antitypes will emerge. The different base models can be tried for further analysis to see if the higher order interaction effects release types and antitypes. The third order interaction effects model was fitted to see if more type and antitypes appeared. The classical log-linear model analysis can help to decide the base model.

**Table-3: The best model obtained with the classical Log-linear model solution of application data**

| Step No | Best No | F | Chi-Square | Prob. Level | Term Deleted | F | Chi-Square | Prob. Level | Hierarchical Model |
|---|---|---|---|---|---|---|---|---|---|
| 20 | 20 | 21 | 29,3 | 0,1063 | BDV | 4 | 26,3 | 0,0000 | G, DV,BV,BD |

The classical log-linear solution illustrated in Table 2. We can see that some of the interaction effects between two variables only emerge in the best fit hierarchical model. Two variables DV, BV and BD were associated with each other. But it is not obvious which combinations of categories of these variables are statistically significant, in other words, local associations are not obviously seen from this solution. This log linear modelling focuses on variables, that is, the result represents the relationships among variables, not local cells. In contrast CFA focuses on the discrepancies between some base model and data. These discrepancies appear in the form of CFA types and CFA antitypes.

The configurations emerging as types and antitypes are the cells contradicting the chosen base model fit. The existing types and antitypes represent the associations at the level of configuration rather than variables. Although the log-linear modeling and CFA approaches are different, they both use the same method to estimate the expected cell frequencies

**Table-4: The model fit results of the standard global CFA solution, and CFA types, antitypes for first order log-linear base model by using Z-test with Bonferroni correction**

| Pearson's $chi^2$ test | | | | Bonferroni-adjusted alpha = 0.0013889 | | |
|---|---|---|---|---|---|---|
| $Chi^2$ for CFA model = 353.0205 | | | | df = 29; p = 0.00000000 | | |
| LR-$Chi^2$ for CFA model = 344.7522 | | | | df = 29; p = 0.00000000 | | |
| | Configuration (cell indices) | $n_i$ (observed) | $m_i$ (Expected) | Z-test | p | Results* |
| 5 | 1211 | 51. | 27.91 | 4.36 | .0000062 | Type |
| 6 | 1212 | 83. | 43.86 | 5.90 | .0000000 | Type |
| 8 | 1222 | 23. | 45.45 | -3.33 | .0004335 | Antitype |
| 13 | 2111 | 12. | 28.87 | -3.14 | .0008428 | Antitype |
| 14 | 2112 | 18. | 45.37 | -4.06 | .0000241 | Antitype |
| 19 | 2221 | 42. | 19.94 | 4.93 | .0000004 | Type |
| 20 | 2222 | 68. | 31.34 | 6.54 | .0000000 | Type |
| 29 | 3211 | 3. | 14.71 | -3.05 | .0011300 | Antitype |
| 30 | 3212 | 3. | 23.12 | -4.18 | .0000142 | Antitype |
| 31 | 3221 | 3. | 15.24 | -3.13 | .0008554 | Antitype |
| 32 | 3222 | 3. | 23.95 | -4.28 | .0000092 | Antitype |
| 33 | 3311 | 15. | 4.08 | 5.39 | .0000000 | Type |
| 34 | 3312 | 25. | 6.42 | 7.33 | .0000000 | Type |
| *: The result is von Eye CFA program solution (von Eye, 2001). | | | | | | |

The results in Table 4 indicate the existence of 6 types and 7 antitypes suggest that the four conditions of breeding calves associates in 13 configurations. The death of calves occur more often than expected under independence model in the configurations indicating 6 CFA types, and that less often than expected independence model in the configuration indicating 7 CFA antitypes. The results in 4 is complemented by 6 types which are constituted by configurations 1211, 1212, 2221, 2222, 3311 and 3312 . The first type, constituted by configuration 1211, suggests that more death of calves occurred than expected from the base model.

The Table 5 presents a summary of von Eye program [15] output results for global CFA approaches and R program [16] output results for global and hierarchical models. Many statistical significance tests can be performed to decide whether a configuration constitutes a CFA type or CFA antitype [17,18]. In CFA hypothesis testing the null hypothesis is formulated at a level of single configurations, therefore the total number of the tests performed was changed; hence we need to protect the nominal significance threshold α against possible test-wise errors **[**19]. The Table 5 presents a summary of results for different CFA statistical tests and different model approaches. The number of types and antitypes in Table 5 indicate that the 36 configurations do not appear at equal rates. The number of types and antitypes varies with these 7 different statistical tests, and also global or hierarchical model preferences. To make results comparable, the Bonferroni-adjusted alphas were used to obtain CFA types and antitypes in Table 5. The revealed types and antitypes for Lehmacher test 7T/7A, Lehmacher with Küchenoff continuity correction test 7T/7A, Binomial test 6T/7A, Binomial test with Normal approximation 6T/7A, Anscombe's test 6T/7A, Z-test 6T/7A, Person $\chi^2$ test 6T/4A for non-hierarchical standard CFA solutions, and Person $\chi^2$ test 6T/1A Z-test 6T/6A with R program standard global CFA solution. In step down hierarchical CFA approaches fCFA 6T/5A, kvCFA 2T/10A. When sampling is multinomial and a base model for higher order CFA was specified, one of the binomial or Z-tests can be selected for simplicity. In spite of the large sample size, the resulting pattern of types and antitypes is not the same.

Some configurations occur type or antitype in all solutions, while some occur only certain method of CFA approaches. However, the number of CFA types and CFA antitypes do not make a significant change for the statistical test methods of identifying type and antitype in non-hierarchical CFA von Eye program approaches, because of large sample size. However, Anscombe's test released more antitypes than the other statistical tests, and Person $\chi^2$ test released less antitype than other statistical tests.

In R solution Person $\chi^2$ test released only one antitype. Configuration 2111 no longer constitutes antitype and neither do configurations 3211, 3221. Person $\chi^2$ test is more conservative than other tests especially for antitypes. In hierarchical R CFA solution, fCFA approaches more similar results. Configurations 1211, 1212, 2221, 2222, constitute CFA type for all solutions except kvCFA solution, configurations 3311, and 3312 constitute CFA type for kvCFA solution too, that is said to constitute more deaths of calves than expected death ($n_i > m_i$). The configurations 1221, 2211, 2212, constitute antitype only in kvCFA method. Configurations 1222, 2111, 2112, 3211, 3212, 3221, and 3222 constitute CFA antitype for most of the solutions, which is said to constitute fewer deaths of calves than expected

death($n_i < m_i$). There exist 6 CFA types and 7 CFA antitypes for most statistical test and different solutions, these configurations can be interpreted as following.

**Table-5: The number of types and antitypes for seven statistical tests with global and hierarchical CFA models**

| | **von Eye-2000 solutions** | | | | | | | **R solutions** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Methods for identification types and antitypes in Non-hierarchical CFA | | | | | | | Methods for Non-hierarchical CFA | | Methods for Hierarchical CFA | |
| 1211 | T | T | T | T | T | T | T | T | T | T | |
| 1212 | T | T | T | T | T | T | T | T | T | T | |
| 1221 | | | | | | | | | | | A |
| 1222 | A | A | A | A | A | A | A | | A | | A |
| 1322 | | | A | | A | | | | | | |
| 2111 | A | A | A | A | A | A | | | | A | A |
| 2112 | A | A | A | A | A | A | A | | A | A | A |
| 2211 | | | | | | | | | | | A |
| 2212 | | | | | | | | | | | A |
| 2221 | T | T | T | T | T | T | T | T | T | T | |
| 2222 | T | T | T | T | T | T | T | T | T | T | |
| 2321 | | | | | A | | | | | A | |
| 3122 | T | T | | | | | | | | | |
| 3211 | A | A | A | A | A | A | | | A | | A |
| 3212 | A | A | A | A | A | A | A | | A | A | A |
| 3221 | A | A | A | A | A | A | | | A | | A |
| 3222 | A | A | A | A | A | A | A | A | A | A | A |
| 3311 | T | T | T | T | T | T | T | T | T | T | T |
| 3312 | T | T | T | T | T | T | T | T | T | T | T |
| T(ype) and A(ntitype) ratio | 7T/ 7A | 7T/ 7A | 6T/ 8A | 6T/ 7A | 6T/ 9A | 6T/ 7A | 6T/ 4A | 6T/ 1A | 6T/ 6A | 6T/ 5A | 2T/ 10A |
| **T** abbreviation indicates CFA type, and **A** indicates CFA Antitype which exits if the related configuration contradicts the hypothesis. | | | | | | | | | | | |

The Table 6 displays the descriptive statistics RR (relative risk ratio), Log(P) and Z-statistics for the results of standard CFA, and their rank scores.

The first descriptive statistics for CFA is RR (relative risk ratio), defined as $RR_i = n_i/m_i$, where i goes over all cells in a contingency table. If $RR_i = 1$, the observed numbers for cell (i) is the same as the expected number. This means the base model describe data adequately. The RR statistics is descriptive, hence if $RR_i$ are the same for different configurations for the $m_i = 10$, and $m_i = 0.0010$, they do not interpret same situations [20].

The other descriptive statistics for the use in global CFA is Log(P), defined as
$LogP = -log_{10}(Pr[X \geq n_i])$, where $X \sim Poisson(m_i)$.

The configurations identified as types and antitypes are among the most extreme ones in the rank order of Log(P) values. This statistics are used in the context of descriptive manner, therefore rather than printing a probability for the RR score, the scores are ranked and so are the Log(P) and Z-test statistics and the ranks are printed.

The configurations identified as types and antitypes are among the most extreme ones in the rank order of Log(P) values. However, that the most extreme RR (1321, 1322), the most extreme Log(P) (3122) constitutes neither a type nor an antitype. The rank orders of measures differ. The spearman correlation between the ranks of Log (P) and Z-statistics is 0.914, that is higher than the correlation between RR and Z-statistics, R=0.160. These correlations show that information about relative risk RR does not carry much information about magnitude of Z, on the left hand side of distribution there must be disagreement. The reason for this disagreement is existing antitype configurations. Interpretation of these correlations is very important for evaluating spars contingency tables. In general RR statistics do not release much information about Z-statistics. The agreement of the configuration defined as CFA type on the positive side of distribution is much higher. RR and Log (P) descriptive measures describe the different characteristics of data

distribution. RR indicates antitypes by values that approaches zero. If $m_i$ is large, Log (P) and Z statistics is expected to behave similarly. These three measures correlate strongly when m is very small. Although the shape of the relationships is not linear, these three measures point to the same configurations as types when $m_i$ is very small. When $m_i$ increases and approximates $0.5n_i$, the three measures increasingly reflect different characteristics of the data.

**Table-1.6: The rank scores of RR, Log(P) and Z-statistical test for standard Local CFA solution results**

| Cell | Type/Antitype | | RR | Rank | Log(P) | Rank | Z test | Rank |
|---|---|---|---|---|---|---|---|---|
| -------- | Specification | | -------- | -------- | -------- | -------- | -------- | -------- |
| 1111 | | | 1.051 | 16 | .697 | 27 | 0,32 | 33 |
| 1112 | | | 1.033 | 17 | .766 | 23 | 0,27 | 35 |
| 1121 | | | .899 | 20 | .650 | 28 | -0,66 | 27 |
| 1122 | | | .924 | 19 | .750 | 24 | -0,62 | 29 |
| 1211 | T | | 1.827 | 6 | 4.283 | 9 | 4,36 | 6 |
| 1212 | T | | 1.892 | 5 | 7.065 | 3 | 5,90 | 3 |
| 1221 | | | .519 | 25 | 1.725 | 15 | -2,58 | 16 |
| 1222 | A | | .506 | 26 | 2.937 | 10 | -3,33 | 10 |
| 1311 | | | .774 | 21 | .274 | 36 | -0,63 | 28 |
| 1312 | | | .657 | 22 | .445 | 33 | -1,19 | 24 |
| 1321 | | | .373 | 29 | .529 | 31 | -1,77 | 21 |
| 1322 | | | .238 | 30 | 1.497 | 17 | -2,70 | 15 |
| 2111 | A | | .416 | 27 | 2.474 | 11 | -3,14 | 11 |
| 2112 | A | | .397 | 28 | 4.379 | 8 | -4,06 | 9 |
| 2121 | | | 1.203 | 11 | .962 | 20 | 1,11 | 25 |
| 2122 | | | 1.233 | 10 | 1.336 | 19 | 1,60 | 23 |
| 2211 | | | .623 | 23 | .774 | 22 | -1,65 | 22 |
| 2212 | | | .595 | 24 | 1.363 | 18 | -2,22 | 19 |
| 2221 | T | | 2.105 | 4 | 4.968 | 4 | 4,93 | 5 |
| 2222 | T | | 2.169 | 3 | 8.024 | 1 | 6,54 | 2 |
| 2311 | | | 1.122 | 13 | .400 | 34 | 0,28 | 34 |
| 2312 | | | .952 | 18 | .339 | 35 | -0,13 | 36 |
| 2321 | | | .000 | 36 | .705 | 26 | -2,35 | 17 |
| 2322 | | | .230 | 31 | .876 | 21 | -2,27 | 18 |
| 3111 | | | 1.087 | 15 | .615 | 30 | 0,41 | 32 |
| 3112 | | | 1.096 | 14 | .737 | 25 | 0,56 | 30 |
| 3121 | | | 1.443 | 8 | 1.637 | 16 | 2,11 | 20 |
| 3122 | | | 1.475 | 7 | 2.418 | 12 | 2,84 | 14 |
| 3211 | A | | .204 | 32 | 2.041 | 14 | -3,05 | 13 |
| 3212 | A | | .130 | 34 | 4.592 | 7 | -4,18 | 8 |
| 3221 | A | | .197 | 33 | 2.188 | 13 | -3,13 | 12 |
| 3222 | A | | .125 | 35 | 4.868 | 5 | -4,28 | 7 |
| 3311 | T | | 3.670 | 2 | 4.596 | 6 | 5,39 | 4 |
| 3312 | T | | 3.892 | 1 | 7.667 | 2 | 7,33 | 1 |
| 3321 | | | 1.417 | 9 | .618 | 29 | 0,85 | 26 |
| 3322 | | | 1.202 | 12 | .508 | 32 | 0,52 | 31 |

**DISCUSSION**

The selection of the statistical test for global base model did not affect the number of CFA types and antitypes too much. Binomial test and Anscombe's test tend to have more antitype.

Nevertheless, the results obtained in most of the CFA solutions, there are 6 CFA types and 7 CFA antitypes, these configurations are very important for evaluation of local associations. The first two CFA types constitute by configurations 1211, 1212, suggest that more death occurs from digestive diseases for not vaccinated calves on both gender appear in the barn systems in which mother and other calves breed together. The next two CFA types constitute by configurations 2221, 2222 consisting of digestive system diseases, and the barn system in which the cattle and calves kept in two separate group housing system, non-vaccinated calves.

The biggest discrepancies in Table 1.5 are those Pearson $\chi^2$ tests and the other tests for global CFA model. The proportions of T/A for the binomial test and its approximations are very similar with z-test. When the sample size is

relatively large, z-test can be trusted. The chi-square component test is biased against antitypes; it prevents researchers to identify antitypes.

The configurations 1211 and 1212 are CFA type, these indicates that the more death occurred in the management system in which the mother and the other cattle and calves are kept together in the same barn; for digestive system diseases; on vaccinated; male and female calves. That means in the mixed barn management system for vaccinated female and male calves' death occurred more than expected because of digestive diseases.

The configurations 2221 and 2222 are CFA types; indicate that in the management system in which the calves kept together, but separated from mothers' cows, the calves' death also occurred more than expected because of digestive diseases.

The configurations 3311 and 3312 are CFA types, which means, the calves' death occurred more than expected because of the other trauma condition especially with vaccinated calves in both gender in the management system which the barns with individual calf pens keep calves separated from cattle.

The descriptive measures RR and Log (P) are sensitive to different distributional characteristics of data than the residual based statistics. The first 6 ranks of types according to z-test are similar to the ranks of RR statistics, so there is a good harmony among ranks of types. The Pearson $\chi^2$-test is less powerful in particular when $n_i < m_i$, that is, when antitypes could be detected. The Pearson $\chi^2$-test also yields an inconsistent pattern of higher and lower tail probabilities in R global CFA solution. We need guidance concerning the selection of tests. When sampling is multinomial and a base model for higher order CFA was specified, one of the binomial or Z-tests can be selected for simplicity.

Vaccination protects both genders from deaths of respiratory system disease in the second type barn system. Keeping calves individual separate pens decreases the number of calves' death caused by digestive system diseases.

These results are interesting because the calves death appearing in different management systems by different cause of diseases and vaccination status.

**REFERENCES**

1. Von Eye A, Gutiérrez Peña E. Configural frequency analysis: The search for extreme cells. Journal of Applied Statistics. 2004 Oct 1;31(8):981-97.
2. Stemmler M, Lösel F. The stability of externalizing behavior in boys from preschool age to adolescence: A person-oriented analysis. Psychological Test and Assessment Modeling. 2012;54(2):195.
3. Wymann MN. *Calf mortality and parasitism in periurban livestock production in Mali* (Doctoral dissertation, University_of_Basel).
4. Lienert GA, Krauth J. Configural frequency analysis as a statistical tool for defining types. Educational and Psychological Measurement. 1975 Jul;35(2):231-8.
5. Doğan İ, Doğan N. Configural Frequency Analysis as A Statistical Tool. Turkiye Klinikleri Journal of Biostatistics. 2016 Jul 1;8(2).
6. Stemmler M. Person-Centered Methods: Configural Frequency Analysis (CFA) and Other Methods for the Analysis of Contingency Tables. Springer; 2014 May 24.
7. Von Eye A, Mair P. A functional approach to configural frequency analysis.2017.
8. R Development Core Team: R. A Language and Environment for Statistical Computing. Vienna, Austria. [Retrieved from http:// www.R-project.org/], 2015.
9. Kieser M, Victor N. Configural Frequency Analysis (CFA) revisited—a new look at an old approach. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 1999 Dec;41(8):967-83.
10. Wise ME. Multinomial probabilities and the chi$^2$ and the chi$^2$ distributions. Biometrika. 1963; 50:145-154.
11. Von Eye A. Configural frequency analysis: Mehods, models and applications. Mahwah, New Jersey : Lawrance Erlbaum Associates, Publishers, 2002.
12. Agresti A. An introduction to cathegorical data analysis. New York : *Wiley*. 1996.
13. Von Eye A, Mun EY. Log-linear modelling: Concepts, interpretation, and application. Hoboken, NJ: John Wiley and Sons, *Inc*. 2013.
14. Von Eye A. Base models for configural frequency analysis. Psychology Science. 2004; 46(1):150-170.
15. Von Eye A. Configural frequency analysis-Version 2000. A Program for 32 bit Windows Operating Systems. [ed.] Alexander Von Eye. Methods of Psychological Research Online. [http://www.mpr-online.de]. 2001; 6(2), 130-139.
16. Funke S, Mair P, von Eye A. Analysis of configural frequencies. Program module in R., [http://cran.r-project.org/], 2007.

17. Von Eye A. The odds favor antitypes- a comparison of tests for the identification of configural types and antitypes. Methods of Psychological Research. 2002; 7(3):1-29.
18. Von Eye A, Mair P, Eun-Young M. Advances in configural frequency analysis. New York: The Guilford Press. 2010.
19. Victor N, Kieser M. Configural frequency analysis and association analyses in contingency tables. Computational Statistics & Data Analysis. 2003; 44(1-2):419-29.
20. Von Eye A, Pena EG. A comparison of measures used for Configural Frequency Analysis of large,sparse cross-classifications. Psychology Science. 2005; 47(3-4):377-390.