Black Sea Journal of Health Science

CONFIGURAL FREQUENCY ANALYSIS FOR MEDICAL SCIENCES WITH SPSS

Fatih ÜÇKARDEŞ[1]*

[1]*Department of Biostatistics and Medical Informatics, Adıyaman University, Research and Education Hospital, 02040, Adıyaman, Türkiye*

Abstract: The aim of this study was to analyze the Configural Frequency Analysis (CFA) with the help of the CFA module prepared for SPSS and to interpret the results. The data set used in this study was taken from Von EYE and Nesselroade (1992). This dataset included 698 patients with four different behavioral disorders (L = logorrhea that is, uncontrolled, uninterrupted speaking, R = running away from home, F = fear of dark, and B = bedwetting), two different (1 = present, 2 = absent) status. The multiway contingency form is 2x2x2x2 and includes 16 configurations in total. The CFA module for SPSS uses the first degree independence model. It uses Z test and Asymptotic Binomial test to determine Type and Antitype configurations. Holm's method was used for control Family wise error rate. The Pearson Chi-square test was found significantly under assumption of total independence of all four variables (P<0.01) Subsequently, CFA analysis was performed. As a result, CFA analysis, Types 4 and antitypes 3 of the 16 configurations were significantly found. The rest of that were non-significance. It has been shown that the CFA method can be done with the help of the SPSS program thanks to the CFA module prepared for SPSS, without the need to learn any programming language. This study will be an important work for researchers using multiway contingency tables.

Keywords: Multiway frequency table, Contingency table, Configural Frequency Analysis, SPSS

*Corresponding author: Department of Biostatistics and Medical Informatics, Adıyaman University, Research and Education Hospital, 02040, Adıyaman, Türkiye
E mail: fatihuckardes@gmail.com (F. ÜÇKARDEŞ)
Fatih ÜÇKARDEŞ   https://orcid.org/0000-0003-0677-7606

1. Introduction

Configurational frequency analysis (CFA) was first proposed by Gustav A. Lienert in 1968. CFA is a non-parametric statistical method recommended for the analysis of multiway contingency tables (Lienert, 1968; Von Eye, 1990). This method is used for cell-based statistical checks of rxc, rxcxm or more arranged contingency tables. In this respect, it differs from tests such as Pearson Chi-square, Likelihood Ratio and Fisher Exact test used in the analysis of contingency tables (Öğüş and Yazıcı, 2011). These methods generally test the presence or absence of relationships between categorical variables. CFA analysis, on the other hand, statistically controls the observed frequency and expected frequency of categorical variables on a cellular basis separately. Therefore, it differs from the methods mentioned above. The variable-based logarithmic linear (LDM) method is also used for the analysis of multiway contingency tables. Unlike other methods, the LDM method is used to statistically control the main effects of the variables and the interaction effects with the help of a model (Bek, 2017). Barış (2020) reported that CFA, unlike LDM, is cell-based but is not variable-based, and emphasized that this method recognizes each cell of the multiway contingency tables separately.

Von Eye (2000) developed a program for CFA analysis, written in Fortran's language and running in MS-DOS. Similarly, it was developed modules written in the R programming language (Funke et al., 2007). However, since The MS-DOS operating system is not used much nowadays. It is necessary to learn this software language to use the R program. For such reasons, its use in the literature is quite limited, except for research to develop the CFA method.

SPSS is highly preferred by researchers because it has an interface and does not require knowing a programming language. There is no CFA module available in the SPSS program. There is a CFA module written for this program. CFA analysis can be run via this CFA module.

The aim of this study is to show step by step how to integrate the CFA module into the SPSS program and interpret the analyzes and results. In this respect, it is the first publication made with SPSS. In this study it is also aimed to provide more opportunities for using CFA in the analysis of contingency tables. In this respect, this study is thought to be an important resource for researchers.

2. Material and Methods

The sample dataset used in this study was taken from Von Eye and Nesselroade (1992) (Figure 1). Six hundred and ninety-eight children were scored (1 = present, 2 = absent) on each of the following behavior disorders: L =

logorrhea that is, uncontrolled, uninterrupted speaking, R = running away from home, F = fear of dark, and B = bedwetting the multiway contingency form is 2x2x2x2 and includes 16 configurations in total.

Figure 1. First step Categorical variables should be defined as nominal scales.

For your CFA analysis, the Chi-square value of the general (global) relationship should be significant. Next, cellular CFA analysis is started. The definition of "type" and "antitype" cells is used in the CFA method. These values take values according to the observed (Gij) and expected (Bij) frequencies. Gij>Bij and Gij<Bij are defined as "type" and "antitype" cells, respectively (von Eye, 1990).

Test approaches suggested by von EYE (1990) and Krauth et al. (1973) were taken into account in the CFA module used in this study. Both methods use the Asymptotic Binomial test approach. Moreover, the program gives the results of the Z test statistics as standard.

The CFA method defines each of the category components of the cells in the contingency tables as a configuration (von Eye et al., 2010). These configurations are shown like "111, 112,..., 222" (Lienert and Krauth, 1975).

Von Eye et al (2010) reported that if the number of configurations increases in CFA analysis, this leads to an increase in Family wise error rate (FWER). Therefore, The Holm's correction method was used to control FWER (Holm, 1979).

The Holm's correction is given in Equation 1;

$$\alpha_i^* = \alpha/(k-i+1), \; i=1,2,...,k \tag{1}$$

calculated by the formula given above. Here k is the total number of configuration and i is the number of step. The adjusted alpha value of first configuration calculates as

$\alpha_1^* = \alpha/(k-1+1)$, for the second $\alpha_2^* = \alpha/(k-1)$,..., similarly, lastest test $\alpha_k^* = \alpha$, respectively (Bek, 2017).

The Holm method involves a step-down method (Vickerstaff et al., 2019). The smallest probability value (P) is compared with the smallest adjusted alpha value and these processes are continued step by step. When it becomes non-significance for the first time, the process is stopped. Probability values after that are also accepted non-significance.

2.1. SPSS Installation Instructions of CFA Extension

1. Download CFA Extension (Grüner, 2021)
2. Open IBM SPSS Statistics
3. SPSS→ Extensions → Install Local Extension Bundle… → Click
4. Locate and double-click the CFA.spe file. (It will usually be in your Downloads folder.)

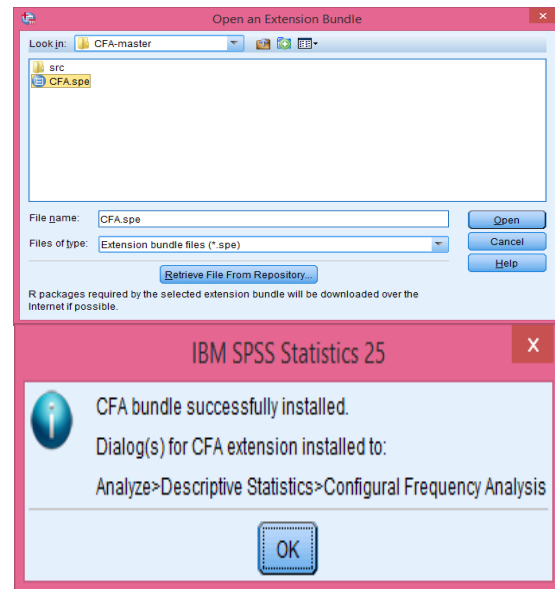SPSS →Analyze →Descriptive Statistics → Configural Frequency Analysis will be installed (Figure 2).

Figure 2. The exhibition of the installation of the CFA.spe extension in SPSS.

3. Results

The Pearson Chi-square statistics must be significance to perform the CFA analysis. First step, the Pearson Chi-square test of multiway contingency tables performs as below:

Step 1. SPSS → Data→Weight cases → The observed data assign as frequency variable and OK (Figure 3).
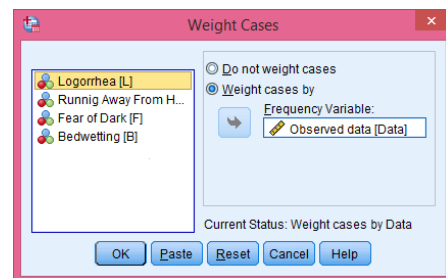

Figure 3. The observed data assign as frequency variable.

Step 2. SPSS → Analyze → Descriptive Statistics → Crosstabs (Figure 4).

Step 3. Crosstabs: Layer 1: Running Away Home From, Layer 2: Logorrhea and Statistics →Chi-square, Click→ Continue →OK (Table 1).

The Pearson Chi-square statistics were found to be significant in Table 1 (P=0.0016). It was determined that there was a significant relationship between categorical variables under the assumption independent. Subsequently, Second Step: the CFA analysis can be continued.
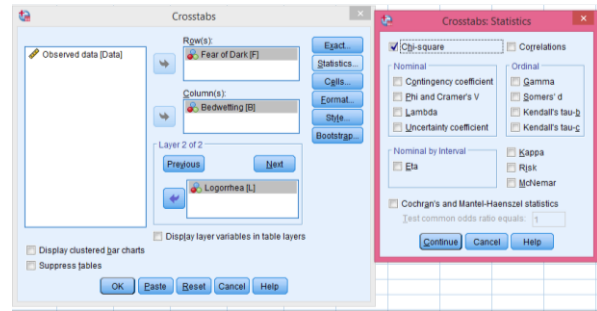

Figure 4. The Pearson Chi-square test show how to make in the Multiway contenjans analysis.

Table 1. Frequency table and significance level of Pearson Chi-square statistics

Logorrhea	Running Away From Home			Bedwetting		Total
				1	2	
1	1	Fear of Dark	1	23	11	34
			2	18	15	33
		Total		41	26	67
	2	Fear of Dark	1	7	20	27
			2	12	16	28
		Total		19	36	55
	Total	Fear of Dark	1	30	31	61
			2	30	31	61
		Total		60	62	122
2	1	Fear of Dark	1	18	43	61
			2	13	43	56
		Total		31	86	117
	2	Fear of Dark	1	38	122	160
			2	50	249	299
		Total		88	371	459
	Total	Fear of Dark	1	56	165	221
			2	63	292	355
		Total		119	457	576
Total	1	Fear of Dark	1	41	54	95
			2	31	58	89
		Total		72	112	184
	2	Fear of Dark	1	45	142	187
			2	62	265	327
		Total		107	407	514
	Total	Fear of Dark	1	86	196	282
			2	93	323	416
		Total		179	519	698

Pearson Chi-square= 5.841; P=0.0016; 1= present; 2= absent.

SPSS → Analyze → Descriptive Statistics → Configural Frequency Analysis →Click (Figure 5).

The analysis results of Figure 5 are given in Table 2. The Configurations column has been ordered as Logorrhea, Running Away from Home, Fear of Dark and Bedwetting, respectively. This order that is like reported by Von EYE and Nesselroade (1992). The variables assigned to the row and column boxes are in the last two orders, respectively. The variable that is desired in the first order of the configuration is started in the last order in the Control box.
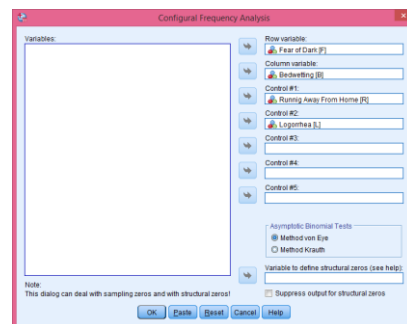

Figure 5. Configural Frequency Analysis.

If CFA analysis of two categorical variables is calculated, the only row and column boxes are used. As shown in Figure 5, CFA analysis can be analyzed seven categorical variables in totally.

The interpretation of the results of the CFA analysis is carried out according to the configurations. That is, it concerns a single cell or a group of cells. CFA analysis results are given in Table 2. According to results, case increase in 1111, 1121 and 2222 type configurations was statistically significantly higher than expected. While 3.33 cases were expected in the 1111 configuration where all symptoms were seen together, 23 cases were observed. While 4.91 cases were expected in 1121 configuration, which is another configuration in which Fear of Dark symptoms were not seen, 18 cases were observed. In the 2222 configuration, which showed no symptoms, 187.97 cases were expected, while 249 cases were observed. When the standardized residual values of these configurations (1st, 3rd and 16th row numbers) were examined, it was seen that the number of cases increased (Figure 6). All these increase in cases were found to be statistically significant.

Table 2. Test statistic for the asymptotic binomial test: after von Eye

	Configurations				Empirical cell counts	Expected cell counts	Standardized residuals	P-values for the standardized residuals	Test statistic for the asymptotic binomial test: after von Eye	P-values for the asymptotic binomial test	Alpha level for the Holm correction	Types *
	L	R	F	B								
1	1	1	1	1	23	3.332	10.775	0.0000	10.8	0.0000	0.0031	Type
2	1	1	1	2	11	9.661	0.431	0.3333	0.272	0.3929	0.025	
3	1	1	2	1	18	4.915	5.902	0.0000	5.923	0.0000	0.0033	Type
4	1	1	2	2	15	14.252	0.198	0.4215	0.2	0.4207	0.05	
5	1	2	1	1	7	9.308	-0.757	0.2247	-0.927	0.1771	0.0071	
6	1	2	1	2	20	26.988	-1.345	0.0893	-1.372	0.085	0.0056	
7	1	2	2	1	12	13.731	-0.467	0.3202	-0.472	0.3185	0.0125	
8	1	2	2	2	16	39.812	-3.774	0.0001	-3.886	0.0001	0.0038	Antitype
9	2	1	1	1	18	15.732	0.572	0.2837	0.578	0.2815	0.0083	
10	2	1	1	2	43	45.613	-0.387	0.3494	-0.4	0.3445	0.0167	
11	2	1	2	1	13	23.207	-2.119	0.0171	-2.155	0.0156	0.0045	
12	2	1	2	2	43	67.288	-2.961	0.0015	-3.115	0.0009	0.0042	Antitype
13	2	2	1	1	38	43.946	-0.897	0.1849	-0.927	0.1771	0.0063	
14	2	2	1	2	122	127.419	-0.48	0.3156	-0.531	0.2977	0.01	
15	2	2	2	1	50	64.828	-1.842	0.0328	-1.934	0.0266	0.005	
16	2	2	2	2	249	187.966	4.452	0.0000	5.208	0.0000	0.0036	Type

* Following Krauth= Result for the binomial test using Holm's correction; 1=present; 2= absent.
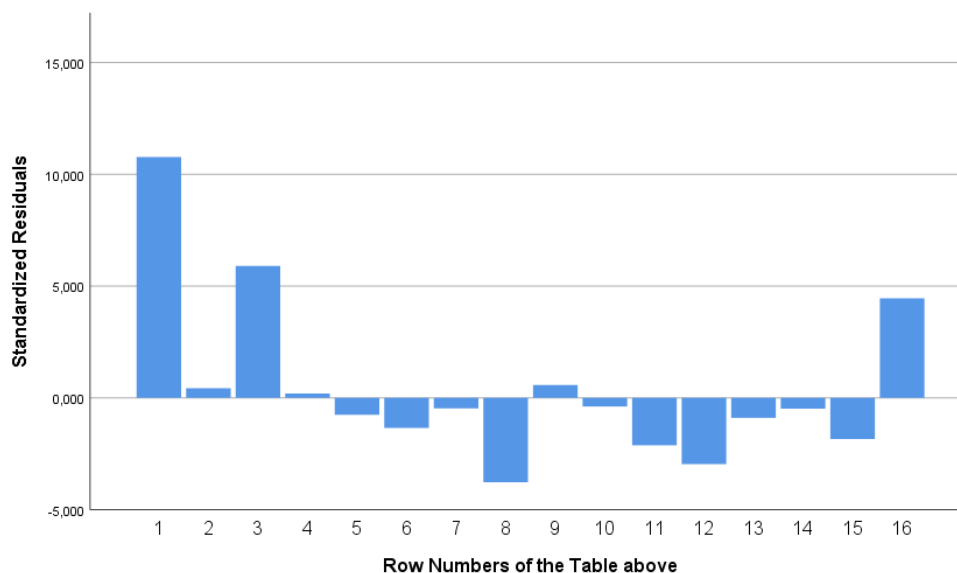

Figure 6. The plot of standardized residual values according to the row number of configurations.

1222 and 2122 are antitype configurations. These configurations (syndromes) had significantly fewer cases than expected. In the 1222 configuration where only Logerra syndrome was seen, 39.8 cases were expected, while 16 cases were observed. While 67.29 cases were expected in the 2122 configuration, where only Running away from syndrome was seen, 43 cases were observed. When the standardized residual values of these configurations (8th and 12th row numbers) were examined, it was seen that the number of cases decreased. These case decreases were statistically significant (Figure 6). Other configurations were found to be non-significance.

4. Discussion

In this study, the data of Von EYE and Nesselroade (1992) were reanalyzed to compare the results of the CFA module. The results were the same as those reported by Von EYE and Nesselroade (1992). There are various models of independence in CFA analysis. Bek (2017) mentioned that CFA independent models use loglinear, graphical or bayesian approaches. The CFA module, on the other hand, uses the first order degree general CFA independence model with log linear model including the main effects. Bek (2017) and Barış (2020) reported general detailed information on independence models.

In this study, four categorical variables were analyzed together. Using by the CFA module, statistical analyzes of seven categorical variables can be performed. Multidimensional scaling method (MDS) is frequently preferred in the interpretation of multinomial (such as 4x5) categorical variables (Öğüş and Yazıcı, 2011). By the CFA module, it can be easily interpreted these variables without using MDS.

Various test techniques are used to identify type and antitype cells in CFA. In the CFA module, it uses the Z test and the Asymptotic Binomial test approach suggested by Von Eye (1990) and Krauth and Lienert (1973). Von EYE (1990) emphasized that the Asymptotic Binomial test approach is a nonparametric test and does not require an assumption about distribution. Von Eye (2002) conducted a simulation study for the tests used to determine the type and antitype in CFA. As a result of this study, He was reported that there was no significant deviation between these tests with an increase in sample size.

Von Eye et al. (2010) mentioned that as the number of configurations increases in CFA analysis. This leads to an increase in Family wise error rate (FWER). Vickerstaff et al. (2019) appointed that there are many methods such as Bonferroni and Holm to control FWER. The Bonferroni correction uses a fixed adjusted alpha value, while the Holm's correction uses a Step-down adjusted alpha values. Vickerstaff et al. (2019) emphasized that The Bonferroni and Holm methods provide the same disjunctive power. However, von Eye et al. (2010) stated that Holm is less conservative than Bonferroni, but it is effective only if the tail probabilities are ordered in an ascending sequence.

There are two statistical test options belong to Von Eye and Krauth methods in the CFA module prepared for SPSS, and Z test is standard. Bek (2017) reported that seven statistical tests are calculate in the CFA program written by Von Eye (2000). Four statistical tests are available in the R package prepared by Funke et al. (2007).

As a result of this study, it was shown that the CFA method can be performed with SPSS. In addition, it is thought that those who need multiway contingency tables can perform these analyzes without the need to learn the R programming language or MS-DOS operating systems.

References

Barış M. 2020. Konfigüral frekans analizinde kullanılan test yöntemlerinin karşılaştırılması. MSc Thesis, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Samsun, Türkiye, pp. 66.

Bek Mİ. 2017. Örgüsel frekans analizi ve hayvancilik araştırmalarında kullanılması. PhD Thesis, Kahramanmaraş Üniversitesi Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Kahramanmaraş, Türkiye, pp. 53.

Funke S, Mair P, von Eye A. 2007. Analysis of configural frequencies. Program module in R. s.l. URL: http://cran.r-project.org/ (access date: February 12, 2022).

Gruner H. 2021. Informationen zu SPSS. URL: https://gruener.userpage.fu-berlin.de/dialogs/CFA.zip (access date: May 23, 2022).

Holm S. 1979. A simple sequentially rejective multiple test procedure. Scand J Stat, 6: 65-70.

Krauth J, Lienert GA. 1973. KFA. Die Konjiguratiom frequenz analyse und ihre An wendung in Psychologie und Medizin. Freiburg: Alber.

Lienert GA. 1968. Die "Konfigurationsfrequenzanalyse" als Klassifikationsmethode in der klinischen Psychologie. Bericht über den, 26: 244-253.

Öğüş E, Yazıcı CA. 2011. Comparision of log-linear analysis and correspondence analysis in two-way contingency tables: A medical application. Balkan Med J, 28:143-147.

Vickerstaff V, Omar ZR, Ambler G. 2019. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomized controlled trials with multiple primary outcomes. BMC Med Res Methodol, 19: 129.

Von Eye A, Mair P, Mun EY. 2010. Advances in configural frequency analysis. Guilford Press. New York. US, pp. 306.

Von Eye A, Nesselroade JR. 1992. Types of change: Application of configural frequency analysis in repeated measurement designs. Exp Aging Res, 18(3-4): 169-183.

Von Eye A. 1990. Introduction to configural frequency analysis: The search for types and antitypes in cross-classification. Cambridge University Press, Cambridge, UK, pp. 284.

Von Eye A. 2000. Configural Frequency Analysis. A program for 32 bit Windows operating systems. Manual for program Version 2000. (Version 2), East Lansing, MI, US.

Von Eye A. 2002. Configural frequency analysis: Methods, models and applications. Lawrence Erlbaum Associates, Inc. New York. US, pp. 445.