Investigation of the Methods for the Comparison of Factor Patterns of Independent Two Groups: A Simulation Study

İki Bağımsız Grubun Faktör Yapılarının Karşılaştırılmasında Kullanılan Yöntemler: Bir Simülasyon Çalışması

Bahar TAŞDELEN,[a]
Semra ERDOĞAN,[a]
Gülhan OREKICI TEMEL[a]

[a]Department of Biostatistics & Bioinformatics, Mersin University Faculty of Medicine, Mersin

Geliş Tarihi/*Received:* 25.01.2012
Kabul Tarihi/*Accepted:* 25.04.2012

Yazışma Adresi/*Correspondence:*
Semra ERDOĞAN
Mersin University Faculty of Medicine, Department of Biostatistics & Bioinformatics, Mersin, TÜRKİYE/TURKEY
semraerdogann@gmail.com

ABSTRACT Objective: In this study, when a scale being applied to two different groups, methods used to compare factor structures were discussed and the dependencies of these methods on the sample size, number of variables, and number of factors were investigated with simulation. **Material and Methods:** In this research, different number of variables (10, 20, and 40) and number of factors extracted (2, 5, 10 and 20) were treated for different sample sizes (50, 100, 200 and 500) with a 1000 replicated simulation study. In order to be able to distinguish factors good, the correlations between variables in the same factor were high (r=0.80), the correlations between variables in different factors were kept as low for each group. **Results:** Root Mean Square Coefficient (RMS) increases as the number of variables increase, when the number of factors is fixed. We can also see decreasing in the values of RMS and increasing in the Pearson correlation coefficients with sample size. Since it is difficult to obtain similar factor structures with excessive factors, correlation coefficient decreases as number of factors increase, when the number of variables is fixed. **Conclusion:** It was agreed that the sample size should be at least 10 times larger than the number of variables ($n \geq 10p$). Furthermore, the number of variables should be two times larger than the number of factors ($p > 2f$; p: number of variables, f: number of factors).

Key Words: Coefficient of congruence; factor analysis; structural equation modeling, root mean square coefficient

ÖZET Amaç: Bu çalışmada, bir ölçek iki farklı grupta uygulanırken faktör yapılarının karşılaştırılması amacıyla kullanılan yöntemler ele alınmış ve bu yöntemlerin örnek genişliği, değişken sayısı ve faktör sayısına ne derece bağımlı olduğu bir simülasyon çalışmasıyla incelenmiştir. **Gereç ve Yöntemler:** Çalışmada, farklı değişken sayıları (10, 20 ve 40) ve faktör sayıları (2, 5, 10 ve 20) alınarak farklı örneklem büyüklükleri (50, 100, 200 ve 500) için veriler üretilmiş ve simülasyonlar 1000 kez tekrar edilmiştir. Her grupta faktörleri iyi ayırt edebilmek amacıyla, aynı faktördeki değişkenler arası korelasyonlar yüksek (r=0,80), farklı faktörlerdeki değişkenler arası korelasyonlar ise düşük tutulmuştur. **Bulgular:** Faktör sayıları sabit iken değişken sayısı arttıkça kareler ortalamasının karekökü (RMS) değeri artmaktadır. Ayrıca, örneklem büyüklüğüne bağlı olarak Pearson korelasyon değerinin arttığı, RMS değerinin de azaldığı gözlenmektedir. Faktör sayısının artışı ile benzer faktör yapılarını elde etmek zor olduğundan, değişken sayısı sabit iken faktör sayısı arttığında korelasyon katsayısı azalmaktadır. **Sonuç:** Gruplar arası faktör yapısının karşılaştırıldığı araştırmalarda örnek genişliğinin, değişken sayısının en az 10 katı ($n \geq 10p$) olması ve değişken sayısının ortak faktör sayısının 2 katından daha fazla ($p > 2f$; p=değişken sayısı, f= faktör sayısı) olması önerilebilir.

Anahtar Kelimeler: Congruence katsayısı; Faktör analizi; Yapısal eşitlik modellemesi, kareler ortalamasının karekökü

Turkiye Klinikleri J Biostat 2012;4(2):60-9

Most of the epidemiological studies which have descriptive or analytical require measurement of qualitative variables using specific scale. In behavioral sciences, many variables are not numerical. The satisfaction or attitude scales, intelligence and personality tests are measuring tools which are used to transform perceptual variables to numerical values.

Generally, researchers apply a scale to two or more populations (cultures) or different time points and want to evaluate similarities of factor structures (factorial invariance) by means of factor analysis.[1] In addition to validity and reliability analysis, explanatory factor analysis is used to detect optimal factor structure which explains variation between the scale points in the first population. Afterwards, factor structure can be tested whether it is valid for other populations. In this situation, confirmatory factor analysis is used to verify that a scale developed for a population can be used for other populations. There are so many methods which have been suggested to compare factor structures of different samples. In addition to traditional methods, structural equation modeling is popular.[2,3] In confirmatory factor analysis, Cattel's similarity index, Pearson correlation coefficient, Congruence coefficient (CC) and root mean square of factor loadings are commonly used to compare factor structures. Nesselroade and Baltes (1970-1971), Korth and Tucker (1975-1979) have examined the sampling distribution of congruence coefficients for different numbers of variables and sample sizes. In 1970, Nesselroad and Baltes carried out a simulation study reference to the number of variables and the number of factors to compare factor structures of two populations and they used oblique rotation technique to investigate the distribution of congruence coefficient. Later, Nesselroade, Baltes, and Labouvie reanalyzed congruence coefficients for various sample sizes (50, 100, 200), the number of variables (15, 30, 45), and the number of factors extracted (5, 10). The conclusions showed that the number of variables and the number of factors were the significant main effects.[4] While some researchers have found the Cattel's similarity index to be more reliable than CC, the others have declared that the Cattel's index were not common in the literature.[5]

In this study, to evaluate similarities of factor structures when a scale applied to two independent populations, structural equation models were examined as well as traditional methods with simulations. The performances of these methods were investigated for various sample sizes, the number of factors, and the number of variables on. The first purpose of the study is to infer which comparison methods must be preferred and the second is to investigate the dependencies of these methods on the sample size, the number of variables, and the number of factors.

MATERIAL AND METHODS

There are so many methods which have been suggested to compare two groups' factor structures. These methods are called as Pearson correlation coefficient, Root Mean Square (RMS), Coefficient of congruence (CC) and Structural Equation Models (SEM).

PEARSON CORRELATION COEFFICIENT

As comparing factor structures of two groups, the first and the simple method is Pearson correlation coefficient. The correlation between factor loadings of each variable in two groups is computed. Pearson r correlation value not only determines the difference in factor loads of two factors but also the difference reveals the relative importance of factor loads. It is important to have numerous factors which have factor load with small values. These loads with small values would cause the value of correlation coefficient to increase, in other words would cause factors to be related with each other.[6] Pearson correlation coefficient can be formulated as Equation 1. Where, x and y are the factor loadings of each group.

$$r = \frac{\sum_{i,j} x_{i,j} y_{i,j}}{\sqrt{\left(\sum_{i,j} x_{i,j}^2\right)\left(\sum_{i,j} y_{i,j}^2\right)}} \quad (1)$$

ROOT MEAN SQUARE COEFFICIENT (RMS)

The differences between the factor loading in one group and the other group for each variable are determined and then the mean of these differences is calculated. The last, the square root of the mean is found. This method can be formulated as Equation 2.[7]

$$RMS = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{i1}-Y_{i2}\right)^2}{n}} \tag{2}$$

Where, X and Y are the factor loadings of each group. The value of RMS is between intervals 0-2. If there is a perfect direction and magnitude match between two groups, RMS is zero. As the coefficient departs from zero, the factors of two groups are less alike.[5] However, since the value of RMS for an acceptable agreement between factors was not known, it was said that the other similarity indexes as well as RMS should be used to make decision on factor similarity.

CONGRUENCE OF COEFFICIENT (CC)

Congruence coefficient produced by Burt in 1948 is defined as the best measure of the similarity between two configurations. Let's take two matrices; X and Y. When the number of rows and columns are equal to each other, congruence coefficient is defined. These two matrices are represented as factor loads or factor indicator. This coefficient measures the similarity of two matrices. They are indicated as φ or r_c and formulated in three different ways given below.[8,9]

$$\varphi = r_c = \frac{\sum_{i,j} x_{i,j} y_{i,j}}{\sqrt{\left(\sum_{i,j} x_{i,j}^2\right)\left(\sum_{i,j} y_{i,j}^2\right)}} \tag{3}$$

$$r_c = \frac{vec\{X\}^T vec\{Y\}}{\sqrt{\left(vec\{X\}^T vec\{X\}\right)\left(vec\{Y\}^T vec\{Y\}\right)}} \tag{4}$$

$$r_c = \frac{trace\{XY^T\}}{\sqrt{trace\{XX^T\}\, trace\{YY^T\}}} \tag{5}$$

In determination of confidence interval, Fisher's Z transform is used for congruence correlation coefficient and can be formulated as Equation 6 and 7. Here; $\log_e$ indicates natural logarithm; r_c indicates congruence coefficient.[4,9]

$$Z_{r_c} = \log_e \sqrt{\frac{(1+r_c)}{(1-r_c)}} \tag{6}$$

$$Z_{r_c} = 0{,}5 * \ln \frac{(1+r_c)}{(1-r_c)} \tag{7}$$

The values of congruence coefficient vary between -1 (perfect negative similarity) through +1 (perfect positive similarity). A zero value of the congruence coefficient shows that there is no similarity between two configurations.[4,5,8,10]Although the formulation of congruence coefficient is similar to the formulation of correlation coefficient, the usage areas and theorical foundations of two coefficients differ. When the population congruence coefficient was equal to zero, the sampling distribution of the congruence coefficient was similar to the sampling distribution of the correlation coefficient.[4]

There are different criteria to comment congruence coefficient. Some researchers decided that congruence coefficient must be 0.85 or above for the virtually equal factor configurations. The others defined a 0.80 coefficient of congruence as robust. Another researcher described congruence coefficient must be at least 0.70 for agreement.[5,11-13] On the other hand, Cheung et al. (2003) interpreted congruence coefficient as satisfactory if it was greater than or equal to 0.90.[2]

STRUCTURAL EQUATION MODELS (SEM)

In addition to traditional methods, structural equation models (SEM) can be used as a confirmatory factor analysis approach. SEM is used to explain causal relationships between observed and latent variables (factors). The best known model of structural equation models, linear structural relationships (LISREL), was introduced by Jöreskog and Sörbom.[14-17] Using SEM, the factor loadings of one group are constrained to be equal to the other groups and the model fit can be tested by using some fit indexes such as goodness of fit index (GFI). GFI should be greater than or equal to 0.90 to accept the model.[18]

SIMULATION STUDY

In this study, that a scale was applied to two different groups and their factor structures were similar, was assumed. In other words, the factor structures of two groups were common. The main aim of this study is to evaluate the performances of structural equation models as well as traditional methods for various sample sizes, the number of factors, and the number of variables. Hence, a 1000 replicated simulation study was organized for different number of variables (10, 20, and 40), for different number of factors extracted (2, 5, 10 and 20) and for different sample sizes (50, 100, 200 and 500).

Step 1. Firstly, the multivariate normal distributed data were generated by assuming high correlation (r=0.80) between variables in the same factor and low correlations among variables in different factors. The correlations were fixed as r=0.20 among item responses between factors on each group. A multivariate normal random data generator MVN, a SAS Macro downloaded from website (www.support.sas.com), was used to simulate data assuming common mean and variance ($\mu=0; \sigma^2=1$) for all variables. The variance-covariance matrix was assumed equal to correlation matrix, because of the standard normal distribution and given in the following way for 10-variable and 2-factor model. The simulation program was given in Appendix-I.

$$\begin{bmatrix} 1 & & & & & & & & & \\ 0.80 & 1 & & & & & & & & \\ 0.80 & 0.80 & 1 & & & & & & & \\ 0.80 & 0.80 & 0.80 & 1 & & & & & & \\ 0.80 & 0.80 & 0.80 & 0.80 & 1 & & & & & \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.80 & 1 & & & & \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.80 & 0.80 & 1 & & & \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.80 & 0.80 & 0.80 & 1 & & \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.80 & 0.80 & 0.80 & 0.80 & 1 & \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.80 & 0.80 & 0.80 & 0.80 & 0.80 & 1 \end{bmatrix}$$

This correlation matrix was also extended to all number of variables and factors.

Step2. In each group, the factor loadings were found using principal axis factoring to avoid inflated factor loadings.[19] Varimax rotation technique was applied to achieve a simpler factor structure that can be meaningfully interpreted and rotated factor loadings were estimated and saved. The cumulative proportions of variance explained were greater than 60%. For example, the factor loadings for 10-variable and 2-factor model were found as in the following way.

Group1	Factor 1	0.86	0.86	0.85	0.79	0.78	0.10	0.19	0.28	0.16	0.31
	Factor 2	0.13	0.25	0.14	0.25	0.31	0.86	0.80	0.78	0.78	0.76
Group2	Factor 1	0.88	0.83	0.81	0.77	0.70	0.25	0.38	0.41	-0.18	0.35
	Factor 2	0.18	0.22	0.27	0.14	0.14	0.82	0.74	0.73	0.73	0.69

Step3. RMS, Pearson correlation and congruence coefficients were calculated for each replication and then average values were obtained.

Step4. GFI values for 1000 replications were estimated from confirmatory factor analysis and the averaged value were calculated. The confirmatory factor analyses for the correlation matrix of the second group were done using factor loadings of the first group. CALIS procedure of SAS 9.1.3 software was used for this process.

RESULTS

To compare factor structures, the averaged values of the Pearson correlation coefficient, RMS, CC and GFI for various combinations of the number of variables, factors extracted and sample size are presented in Tables 1, 2 and 3.

The mean change of RMS coefficients for different sample sizes (50, 100, 200, 500) is illustrated for 10 variables with 2 and 5 factors, 20 variables with 2, 5 and 10 factors, 40 variables with 2, 5, 10 and 20 factors separately (Figure 1). Similarly, the mean change of correlation coefficients for different sample sizes (50, 100, 200, 500) is illustrated for 10 variables with 2 and 5 factors, 20 variables with 2, 5 and 10 factors, 40 variables with 2, 5, 10 and 20 factors separately (Figure 2).

The results found in this study have indicated that RMS increases as the number of variables increase, when the number of factors is fixed. We can also see decreasing in the values of RMS and increasing in the Pearson correlation coefficients with sample size (Figure 1 and 2). Since it is difficult to obtain similar factor structures with excessive factors, correlation coefficient decreases as

TABLE 1: The averaged values of RMS, Pearson correlation coefficient, CC and GFI for various combinations of the number of factors and sample size for 10 variables.

Number of Factors		Sample Size	RMS	r	CC	GFI
2	1	50	0.0354	0.9713	0.9902	0.8424
	2		0.0381	0.9581	0.9871	
	1	100	0.0217	0.9792	0.9926	0.8826
	2		0.0193	0.9828	0.9939	
	1	200	0.0079	0.9949	0.9981	0.9522
	2		0.0087	0.9945	0.9975	
	1	500	0.0033	0.9974	0.9991	0.9767
	2		0.0041	0.9955	0.9985	
5	1	50	0.0548	0.9232	0.9557	0.8872
	2		0.0435	0.9482	0.9693	
	3		0.0435	0.9496	0.9702	
	4		0.0644	0.9227	0.9526	
	5		0.0532	0.9374	0.9444	
	1	100	0.0242	0.9679	0.9467	0.9313
	2		0.0229	0.9680	0.9811	
	3		0.0182	0.9812	0.9400	
	4		0.0190	0.9792	0.9919	
	5		0.0243	0.9683	0.9909	
	1	200	0.0180	0.9626	0.9810	0.9636
	2		0.0105	0.9879	0.9808	
	3		0.0182	0.9526	0.9879	
	4		0.0106	0.9843	0.9875	
	5		0.0130	0.9836	0.9803	
	1	500	0.0051	0.9947	0.9961	0.9801
	2		0.0044	0.9961	0.9969	
	3		0.0045	0.9960	0.9968	
	4		0.0045	0.9958	0.9965	
	5		0.0052	0.9948	0.9958	

number of factors increase, when the number of variables is fixed (Figure 2). The results also show that the changes of RMS and Pearson correlation coefficient are more stable for the small numbers of factors.

As comparing factor structures when the number of variables is 10, minimum correlation coefficient is obtained with 5 factors and 50 subjects (Table 1). A similar situation is observed with 10 factors and 100 subjects and 20 factors and 200 subjects for 20 and 40 variables (Table 2 and Table 3). These results show that the suggestions as explained above about the sample size, the number of variables and factors are important. Otherwise, the comparison of factor structures may give some misleading results.

The mean change of congruence coefficients for different sample sizes (50, 100, 200, 500) is illustrated for 10 variables with 2 and 5 factors, 20 variables with 2, 5 and 10 factors, 40 variables with 2, 5, 10 and 20 factors separately (Figures 3). When the number of variables is 10, CC takes too close values to 1 and changes from 0.95 to 1 as sample size increases for 2 and 5 factor situations separately. For 20 variables, CC is over 0.84 with 2 and 5 factors, it decreases until 0.56 with 10 factors for small sample sizes ($n \leq 200$). For 40 variables, CC is over 0.75 with 2, 5 and 10 factors, it decreases until 0.69 with 20 factors for 200 subjects.

The values of GFI are also related to sample size (Figure 4). When the numbers of factors are 2

TABLE 2: The averaged values of RMS, Pearson correlation coefficient, CC and GFI for various combinations of the number of factors and sample size for 20 variables.

Number of Factors		Sample Size	RMS	r	CC	GFI
2	1	50	0.0462	0.9783	0.9915	0.7288
	2		0.0533	0.9664	0.9872	
	1	100	0.0263	0.9851	0.9944	0.8617
	2		0.0266	0.9834	0.9940	
	1	200	0.0122	0.9931	0.9975	0.9259
	2		0.0114	0.9938	0.9979	
	1	500	0.0048	0.9975	0.9991	0.9611
	2		0.0055	0.9962	0.9987	
5	1	50	0.0499	0.9285	0.8562	0.7002
	2		0.0448	0.9322	0.8389	
	3		0.0422	0.9408	0.9255	
	4		0.0414	0.9328	0.9348	
	5		0.0461	0.9090	0.9006	
	1	100	0.0207	0.9699	0.9570	0.8615
	2		0.0199	0.9663	0.8872	
	3		0.0216	0.9623	0.9304	
	4		0.0212	0.9701	0.9289	
	5		0.0213	0.9633	0.9071	
	1	200	0.0155	0.9457	0.9538	0.8954
	2		0.0106	0.9854	0.9759	
	3		0.0122	0.9376	0.9104	
	4		0.0183	0.9488	0.9032	
	5		0.0163	0.9609	0.9476	
	1	500	0.0040	0.9933	0.9887	0.9604
	2		0.0037	0.9942	0.9842	
	3		0.0045	0.9924	0.9838	
	4		0.0036	0.9946	0.9932	
	5		0.0043	0.9864	0.9872	
10	1	100	0.0319	0.8568	0.8951	0.7957
	2		0.0389	0.8625	0.8631	
	3		0.0328	0.8394	0.7716	
	4		0.0275	0.8782	0.8605	
	5		0.0368	0.8645	0.8454	
	6		0.0304	0.8624	0.8536	
	7		0.0407	0.7499	0.7518	
	8		0.0423	0.7542	0.8029	
	9		0.0350	0.7424	0.7926	
	10		0.0473	0.6425	0.5574	
	1	200	0.0264	0.8713	0.8633	0.7960
	2		0.0250	0.8630	0.8608	
	3		0.0232	0.8402	0.7868	
	4		0.0195	0.8957	0.8549	
	5		0.0227	0.8603	0.8587	
	6		0.0243	0.8399	0.8596	
	7		0.0249	0.7780	0.7557	
	8		0.0293	0.7626	0.8015	
	9		0.0218	0.7742	0.7860	
	10		0.0297	0.7742	0.5573	
	1	500	0.0040	0.9935	0.9628	0.9479
	2		0.0046	0.9920	0.9511	
	3		0.0037	0.9932	0.9813	
	4		0.0042	0.9941	0.9532	
	5		0.0044	0.9926	0.9471	
	6		0.0042	0.9935	0.9946	
	7		0.0042	0.9937	0.9725	
	8		0.0046	0.9924	0.9720	
	9		0.0043	0.9938	0.9677	
	10		0.0046	0.9923	0.9604	

TABLE 3: The averaged values of RMS, Pearson correlation coefficient, CC and GFI for various combinations of the number of factors and sample size for 40 variables.

Number of Factors		Sample Size	RMS	r	CC	GFI
2	1	200	0.0148	0.9837	0.9252	0.8300
	2		0.0138	0.9851	0.9484	
	1	500	0.0095	0.9955	0.9982	0.9314
	2		0.0103	0.9941	0.9978	
5	1	200	0.0161	0.9627	0.8628	0.5879
	2		0.0164	0.9642	0.8697	
	3		0.0145	0.9654	0.9063	
	4		0.0160	0.9656	0.9191	
	5		0.0136	0.9619	0.9221	
	1	500	0.0051	0.9879	0.9658	0.9243
	2		0.0061	0.9871	0.9552	
	3		0.0061	0.9844	0.9549	
	4		0.0056	0.9861	0.9468	
	5		0.0058	0.9869	0.9544	
10	1	200	0.0172	0.9316	0.7556	0.8253
	2		0.0154	0.9381	0.8312	
	3		0.0159	0.9313	0.8463	
	4		0.0143	0.9376	0.8780	
	5		0.0178	0.9277	0.7976	
	6		0.0169	0.9295	0.8268	
	7		0.0172	0.9235	0.7702	
	8		0.0169	0.9327	0.8230	
	9		0.0155	0.9326	0.8231	
	10		0.0175	0.9236	0.8512	
	1	500	0.0065	0.9751	0.9564	0.9171
	2		0.0070	0.9698	0.9057	
	3		0.0054	0.9729	0.9337	
	4		0.0053	0.9709	0.9517	
	5		0.0066	0.9754	0.8979	
	6		0.0065	0.9712	0.9256	
	7		0.0067	0.9729	0.9188	
	8		0.0067	0.9732	0.9029	
	9		0.0069	0.9717	0.9514	
	10		0.0064	0.9718	0.9394	
20	1	200	0.0189	0.8699	0.7505	0.8260
	2		0.0179	0.8873	0.7333	
	3		0.0134	0.8866	0.8633	
	4		0.0163	0.8698	0.7776	
	5		0.0177	0.8710	0.7392	
	6		0.0162	0.8781	0.7237	
	7		0.0184	0.8660	0.7701	
	8		0.0151	0.8872	0.7826	
	9		0.0166	0.8706	0.7565	
	10		0.0172	0.8637	0.6925	
	11		0.0189	0.8520	0.7363	
	12		0.0151	0.8898	0.7094	
	13		0.0217	0.8557	0.8530	
	14		0.0143	0.8621	0.7562	
	15		0.0168	0.8759	0.8486	
	16		0.0170	0.8628	0.7794	
	17		0.0161	0.8710	0.7132	
	18		0.0195	0.8657	0.7377	
	19		0.0155	0.8523	0.8018	
	20		0.0215	0.8583	0.7013	

continued →

TABLE 3: *continued.*

Number of Factors		Sample Size	RMS	r	CC	GFI
	1		0.0096	0.9538	0.8479	
	2		0.0064	0.9442	0.9186	
	3		0.0064	0.9434	0.9153	
	4		0.0066	0.9478	0.8512	
	5		0.0061	0.9498	0.8684	
	6		0.0057	0.9540	0.9172	
	7		0.0063	0.9451	0.9127	
	8		0.0068	0.9459	0.8619	
	9		0.0063	0.9420	0.8747	
20	10	500	0.0076	0.9486	0.8571	0.9152
	11		0.0074	0.9397	0.8318	
	12		0.0069	0.9428	0.8768	
	13		0.0063	0.9551	0.8623	
	14		0.0061	0.9445	0.8939	
	15		0.0073	0.9419	0.8745	
	16		0.0073	0.9417	0.8079	
	17		0.0063	0.8572	0.8835	
	18		0.0062	0.9450	0.8847	
	19		0.0072	0.9381	0.8647	
	20		0.0064	0.9466	0.8793	

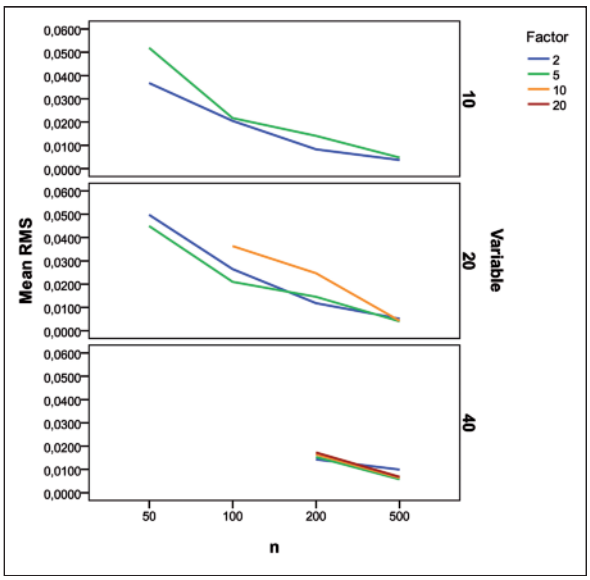

FIGURE 1: The mean change of RMS coefficients for the numbers of variables are 10, 20 and 40, the numbers of factors are 2, 5, 10 and 20, sample sizes are 50,100, 200 and 500.

(See for colored form http://biyoistatistik.turkiyeklinikleri.com/)

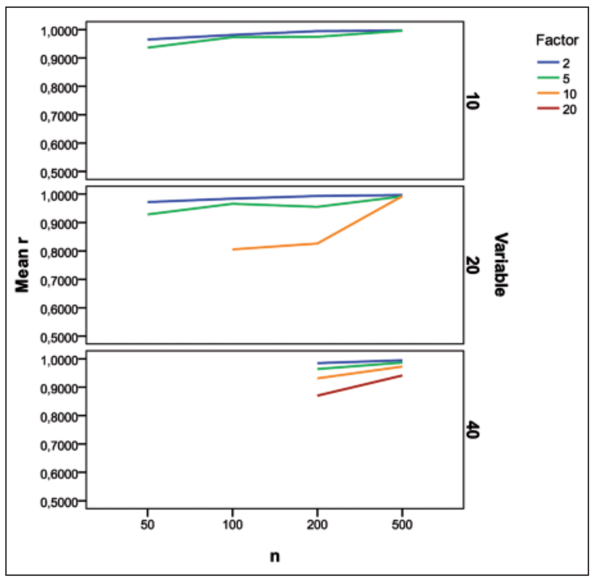

FIGURE 2: The mean change of correlation coefficients for the numbers of variables are 10, 20 and 40, the numbers of factors are 2, 5, 10 and 20, sample sizes are 50,100, 200 and 500.

(See for colored form http://biyoistatistik.turkiyeklinikleri.com/)

and 5, GFI ranges from 0.84 to 0.98 as sample size increases for 10 variables. On the other hand, it takes values in the interval 0.70-0.96 and 0.59-0.93 for 20 and 40 variables separately.

CONCLUSION

In addition to finding common structure of data set, factor analysis is used as a data reduction tech-

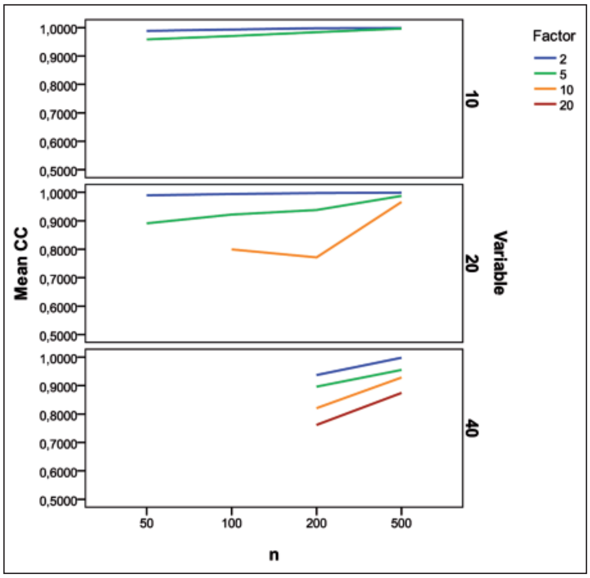

FIGURE 3: The mean change of congruence coefficients for the numbers of variables are 10, 20 and 40, the numbers of factors are 2, 5, 10 and 20, sample sizes are 50,100, 200 and 500.
(See for colored form http://biyoistatistik.turkiyeklinikleri.com/)

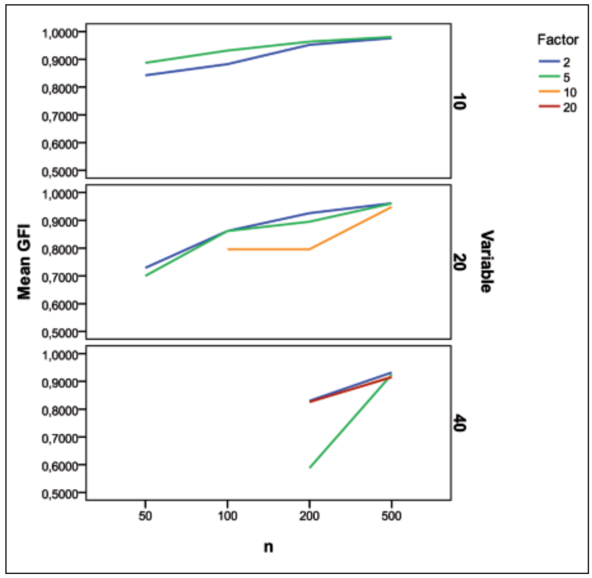

FIGURE 4: The mean changes of GFI for the numbers of variables are 10, 20 and 40, the numbers of factors are 2, 5, 10 and 20, sample sizes are 50,100, 200 and 500.
(See for colored form http://biyoistatistik.turkiyeklinikleri.com/)

nique to explain variability among observed variables in terms of fewer unobserved variables (factors). Increase in the number of factors along variables is not expected or wanted. When the factor analysis is done the relationship between the number of variables and sample size should be noticed. There are different approaches in the literature about the minimum sample size needed to conduct factor analysis. Some have suggested the ratio of sample size to number of variables was no lower than 5.[20] Some have recommended at least 150 - 300 cases. Froman (2001) recommends at least 300 cases.[21] Another important task for factor analysis is to determine the number of factors. Although there are also different approaches in the literature, the number of factors should not exceed half the number of variables.[22] Thurstone recommended at least three variables per factor for doing explanatory factory analysis.[23]

Since the techniques used for the comparison of factor structures were investigated in this study, we could not ignored the relationships between the coefficients (RMS, CC, r, GFI) and sample size, the number of variables and factors. Some researchers have investigated the effect of the sample size, the

```
data variance_covariance;
input m1-m10;
cards;
1.0 0.8 0.8 0.8 0.8 0.2 0.2 0.2 0.2 0.2
0.8 1.0 0.8 0.8 0.8 0.2 0.2 0.2 0.2 0.2
0.8 0.8 1.0 0.8 0.8 0.2 0.2 0.2 0.2 0.2
0.8 0.8 0.8 1.0 0.8 0.2 0.2 0.2 0.2 0.2
0.8 0.8 0.8 0.8 1.0 0.2 0.2 0.2 0.2 0.2
0.2 0.2 0.2 0.2 0.2 1.0 0.8 0.8 0.8 0.8
0.2 0.2 0.2 0.2 0.2 0.8 1.0 0.8 0.8 0.8
0.2 0.2 0.2 0.2 0.2 0.8 0.8 1.0 0.8 0.8
0.2 0.2 0.2 0.2 0.2 0.8 0.8 0.8 1.0 0.8
0.2 0.2 0.2 0.2 0.2 0.8 0.8 0.8 0.8 1.0
;

data means;
input m1;
cards;
0
0
0
0
0
0
0
0
0
0
;
%inc 'c:\Documents and Settings\XP\Belgelerim\My SAS Files\9.1\mvn.sas';
%mvn(varcov=variance_covariance, means=means,
n=50,sample=factorn50v10p2)
```

APPENDIX-I: A multivariate normal random data generation for 10-variable and 2-factor model assuming standart normal distribution for all variables.

number of variables, and the number of factors on the sampling distribution of congruence coefficients.[4,22] However, the superiorities of the congruence coefficient, Pearson correlation coefficient, RMS and SEM statistics have not been investigated together yet. Hence, the question about which statistics are more affected from the sample size, the number of variables, and the number of factors has been examined in this study.

If we compare all the results of RMS, r, CC and GFI, we can recommend that the ratio of sample size to number of variables is no lower than 10 and the ratio of the number of variables to number of factors extracted is greater than two ($p > 2f$; p: number of variables, f: number of factors). Furthermore, whatever the number of variables and the number of factors are, all index and coefficients take stable and optimum values with 500 subjects. The results found in this study have shown that all indexes and coefficients are concordant themselves.

This study may be explanatory for the researchers who want to know which criteria is more reliable for the comparison of similar factor structures in the cross-cultural studies and which combinations of the sample size, the number of variables, and the number of factors must be used to obtain acceptable and reliable results.

In this study, the sampling distributions of these coefficients have been evaluated assuming high similarity between two factor structures. Further investigation may be carried out under poor similarity between two factor structures.

REFERENCES

1. Hoyle RH, Smith GT. Formulating clinical research hypotheses as structural equation models: A conceptual overview. J Consult Clin Psych 1994;62(3):429-40.
2. Cheung MWL, Leung K, Au K. Evaluating multilevel models in cross-cultural research: an illustration with social axioms. J Cross Cult Psychol 2006;37(5):522-41.
3. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. Counseling Psychologist 2006;34(6):806-38.
4. Broadbooks WJ, Elmore PB. A Monte Carlo study of the sampling distribution of the congruence coefficient. Educ Psychol Meas 1987;47(1):1-11.
5. Brauchle PE, Azam MS. Factorial invariance of the Occupational Work Ethic Inventory (OWEI). J Vocat Educ Res 2004;29(2):121-31.
6. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 3rd ed. New York: HarperCollins; 1996. p.880.
7. Rummel RJ. Chapter 20: Factor Comparison. Applied Factor Analysis. 1st ed. United States of America: Northwestern University Press; 1970. p.461.
8. Abdi H. RV coefficient and congruence coefficient. In: Salkind NJ, ed. Encyclopedia of Measurement and Statistics. 1sted. Thousand Oaks (CA): Sage; 2007. p. 849-53.
9. Nimon K, Reio TG. Measurement ınvariance: a foundational principle for quantative theory building. Human Resource Development Review 2011;10(2):198-214.
10. Dreger RM. Sets of basic programs for factor matrix comparisons: ahmavaara transformation, congruence coefficient, salient similarity index. Educ Psychol Meas 1985;45(1):167-72.
11. Koschat MA, Swayne DF. A weighted procrustes criteria. Psychometrica 1991;56(2):229-39.
12. Ommundsen R, Hak T, Mörch S, Larsen KS, Van der Veer K. Attitudes toward illegal immigration: a cross-national methodological comparison. J Psychol 2002;136(1):103-10.
13. Sakamoto S, Kijima N, Tomoda A, Kambara M. Factor structures of the Zung Self-Rating Depression Scale (SDS) for undergraduates. J Clin Psychol 1998;54(4):477-87.
14. Bollen KA. Structural equation models. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics. 2nd ed. England: John Wiley; 2005. p.5269-78.
15. Loehlin JC. Chapter 3: Fitting path and structural models to data from a single group on a single occasion; Chapter 4: Fitting models involving repeated measures or multiple groups. Latent Variable Models: An Introduction to Factor, path and Structural Equation Analysis. 4th ed. London: Lawrence Erlbaum Associates; 2004. p.87-117.
16. Krzanowski WJ, Marriott FHC. Multivariate analysis. Part 2: Classification, covariance structures and repeated measurements. J Classif 1995;15(2):294-7.
17. Sharma, S. Applied multivariate techniques. 1st ed. New York: John Wiley&Sons Inc; 1996. p.90-185.
18. Hu L, Bentler PM. Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55.
19. Knäuper B, Rabiau M, Cohen O, Patriciu N. Compensatory health beliefs: scale development and psychometric properties. Psychology & Health 2004;19(5):607-24.
20. Bryant FB, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR, eds. Reading and Understanding Multivariate Analysis. 1st ed. Washington, DC: American Psychological Association Books; 1995. p.99-136.
21. Froman RD. Elements to consider in planning the use of factor analysis. Southern Online Journal of Nursing Research 2001;5(2):1-22.
22. Korth B, Tucker LR. The distribution of chance congruence coefficients from simulated data. Psychometrika 1975;40(3):361-72.
23. Kim JO, Mueller CW. Confirmatory factor analysis. Factor Analysis: Statistical Methods and Practical Issues. In: Uslaner EM, ed. Quantitative Applications in the Social Sciences Series. Thousand Oaks, CA: Sage Publications; 1978. p.55-9.