ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2018-61399

A Robust Approach to the Identification of Latent Factor Structures

Latent Faktör Yapılarının Belirlenmesinde Sağlam Bir Yaklaşım

Özlem ALPU,[a]
Hatice ŞAMKAR[a]

[a] Department of Statistics,
Eskişehir University
Faculty of Arts and Sciences,
Eskişehir

Received: 10.05.2018
Received in revised form: 28.06.2018
Accepted: 29.06.2018
Available Online: 07.09.2018

Correspondence:
Hatice ŞAMKAR
Eskişehir University
Faculty of Arts and Sciences,
Department of Statistics, Eskişehir,
TURKEY/TÜRKİYE
hfidan@ogu.edu.tr

ABSTRACT Objective: Factor analysis is a method for modeling observed variables using a fewer number of latent factors not directly observable. The analysis is based on the estimation of correlation or covariance matrix. However, outliers can severely affect the covariance matrix (or correlation matrix) and hence the parameter estimation and latent factor structures. The influence of outliers is curbed using robust scatter matrix instead of classical one. Usage of the robust scatter matrix within factor analysis leads to robust factor analysis. This study aims to compare classical factor analysis with a robust counterpart that is resistant to the effect of outliers in a multivariate data set. **Material and Methods:** The data regarding environmental sensitivity of university students who were members of international student organizations were used. An e-mail survey was administered to a sample of university students who are members of AISEC and AEGEE student platforms in order to obtain the data regarding environmental sensitivity. **Results:** In classical factor analysis, the six factors explained 64.6% of the total variance, while it explained 68.3% of the total variance in robust factor analysis. **Conclusion:** The analysis results revealed that robust factor analysis decreased the effect of outliers and provided estimations which fit most of the data, which had a higher variance explained, and which included factors with more conceptually meaningful variables.

Keywords: Robust estimators; robust factor analysis;
environmental attitudes and behaviors

ÖZET Amaç: Faktör analizi çok sayıdaki gözlenen değişkenleri az sayıda doğrudan gözlenemeyen gizli faktörleri kullanarak modellemeye çalışan bir tekniktir. Analiz korelasyon ve kovaryans matrisinin kestirimine dayanır. Bununla birlikte, her iki matris ve dolayısıyla da gizli faktör yapıları ile parametre kestirimleri aykırı değerlerden ciddi şekilde etkilenir. Klasik tekniklerle hesaplanan matrisler yerine sağlam tekniklerle hesaplanan matrisleri kullanarak aykırı değerlerin etkisini gidermek mümkündür. Faktör analizinde kullanılan korelasyon veya kovaryans matrisi yerine sağlam korelasyon veya kovaryans matrisinin kullanımı araştırmacıları sağlam faktör analizi kullanımına götürür. Bu çalışma çok değişkenli bir veri setinde aykırı değerlerin etkisine dirençli olan sağlam faktör analizi ile klasik faktör analizini karşılaştırmayı amaçlar. **Gereç ve Yöntemler:** Çalışmada uluslararası öğrenci organizasyonlarına üye olan üniversite öğrencilerinin çevre duyarlılıklarına ilişkin veriler ele alınmıştır. Bu amaçla AISEC ve AEGEE öğrenci platformlarına üye olan 505 üniversite öğrencisine e-posta aracılığıyla anket uygulanmıştır. **Bulgular:** Klasik faktör analizinde toplam varyansın %64.6'sı altı faktör tarafından açıklanırken, sağlam faktör analizinde bu oran %68.3 olarak bulunmuştur. **Sonuç:** Analiz sonuçları sağlam faktör analizinin aykırı değerlerin etkisini azaltarak daha yüksek varyans açıklanma oranının ve değişkenlerin kavramsal olarak daha anlamlı faktörlerde toplanması sonucu verilere daha iyi uyum sağlayan kestirimlerin elde edildiğini göstermiştir.

Anahtar Kelimeler: Sağlam kestiriciler; sağlam faktör analizi;
çevresel tutum ve davranışlar

For the measurement of individuals' sensitivities, behaviors, opinions and attitudes towards any subject matter within behavioral sciences, there are indirect measurement criteria. For the purpose of measuring such concepts, scales which mostly involve combined use of variables considered to measure the same concept have been developed. In scales involving Likert-type items, factor analysis is commonly used to see which item forms a group with other items that serves a similar goal and to reveal the effect of these items within the group.[1-7]

Generally, data gathered from research in social or behavioral sciences have heavier tails when compared to normal distributed data.[8] In addition, presence of outliers in data sets results in violation of the normality assumption and causes the variables to fail to gather under appropriate factors. However, it is important to remember that one goal of factor analysis is to reveal the real causes underlying a number of measureable and observable variables or to identify the latent dimensions that cannot be measured or observed.[9] In this respect, it is a well-known fact that outliers affect these structures and have bad influence on the performance of the analysis.[10] Several studies in the literature note that factor analysis is commonly used without controlling whether data fulfill its requirements.[11-14]

In this study, in order to examine the influence of outliers on latent factor structures, classical factor analysis is compared with a robust counterpart based on fast minimum covariance determinant (MCD) estimators via a real data set on environmental sensitivity. The second part of the study focuses on classical and robust factor analysis. The third part of the study aims at determining the factors influential on the environmental sensitivity behaviors of university students who were members of AISEC (Association Internationale des Étudiants en Sciences Économiques et Commerciales) and AEGEE (Association des Etats Généraux des Etudiants de l'Europe) student platforms. In the last part, the results of robust factor analysis are interpreted in comparison with the results of classical factor analysis.

MATERIAL AND METHODS

A BRIEF INTRODUCTION TO FACTOR ANALYSIS

Factor analysis is regarded as a method for interpreting multivariate data structures and dimension reduction.

The mathematical model of classical factor analysis is given below:

$$y_{1i} = \nu_1 + \lambda_{11}\eta_{1i} + \lambda_{12}\eta_{2i} + \cdots + \lambda_{1k}\eta_{ki} + \varepsilon_{1i}$$

$$\ldots$$

$$y_{ji} = \nu_j + \lambda_{j1}\eta_{1i} + \lambda_{j2}\eta_{2i} + \cdots + \lambda_{jk}\eta_{ki} + \varepsilon_{ji} \tag{1}$$

$$\ldots$$

$$y_{pi} = \nu_p + \lambda_{p1}\eta_{1i} + \lambda_{p2}\eta_{2i} + \cdots + \lambda_{pk}\eta_{ki} + \varepsilon_{pi.}$$

where p denotes the number of variables and k denotes the number of factors; y_j are the variables represented in latent factors; ν_j are the means; λ_{jk} are factor loadings; η_{ki} are factor values; and ε_{ji} are error terms with zero means and correlations of zero with the factors (i=1,2,…,n, and j=1,2,…,p). Therefore, according to this model, it is assumed that there are k latent factors and that each observed variable is a linear function of these factors.

This model can generally be reduced to matrix notation as shown below:

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon}. \quad (2)$$

where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_p)'$ is the mean vector of the observed variables, $\boldsymbol{\Lambda} \in \mathbb{R}^{\boldsymbol{p}\times\boldsymbol{\kappa}}$ is the matrix of factor loadings reflecting the relations between the observed variables and common factors, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)'$ is a vector of common factors, $k<p$, and the error term is $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)'$, $E(\boldsymbol{\eta}) = \boldsymbol{0}, Cov(\boldsymbol{\eta}) = \boldsymbol{I}$, and $Cov(\boldsymbol{\eta}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$. $Cov(\boldsymbol{\varepsilon}) = diag(\boldsymbol{\Psi})$, and $\boldsymbol{\Psi} = (\psi_1, \dots, \psi_p)$ refers to the diagonal matrix.[15] $\boldsymbol{\Psi}$ is the matrix of error term variances / covariances with the population covariance matrix $\boldsymbol{\Sigma}$ of observed variables. The estimation of the factor model in Eq.(2) is based on the covariance structure below:[16]

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Lambda}' + diag(\boldsymbol{\Psi}). \quad (3)$$

Eq.(3) is called common factor decomposition. In factor analysis, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ for a given $\boldsymbol{\Sigma}$ are estimated by using this decomposition.[17]

INFLUENCE OF OUTLIERS

Factor analysis utilizes the correlations between a large number of variables to determine the latent factor constructs in multivariate data. For this purpose, first, the correlation or covariance matrix is calculated. The use of covariance matrix is usually preferred in practice when the observable variables are measured on the same scales. The estimations of Λ and Ψ matrices obtained by using the decomposition in Eq. (3) are quite good if the dataset fulfills the requirements of factor analysis. Factor analysis is based on the assumption of normality and uses the location and scale parameter estimations of the distribution. The presence of outliers in dataset can result in misspecification of the model as well as in biased parameter estimations and cause the analysis results to be found meaningless.[18]

The covariance matrix of observed variables, $\boldsymbol{\Sigma}$, is quite sensitive to outliers. As the sample covariance matrix has 0% breakdown point, even the presence of any bad observation could cause distortion of the estimation.[19] In case of presence of an outlier in the dataset, the performance of classical factor analysis based on $\boldsymbol{\Sigma}$ matrix will deteriorate depending on the outliers.[10] In such a situation, estimating the covariance matrix of variables with the help of a robust method instead of estimating it with classical methods allows obtaining robust parameter estimations. In other words, when unknown parameters of Λ and Ψ are estimated via decomposition of robust covariance matrix, the structure resistant to uncontrolled influence of outliers can be obtained.

PLUGGING IN ROBUST SCATTER MATRIX

In this study, as the robust estimator of $\boldsymbol{\Sigma}$ matrix, the location and scatter estimators of the MCD method were used.

In the literature, there are robust estimators used as an alternative to MCD. Some of these are the minimum volume ellipsoid (MVE) estimators of Rousseeuw, the translated-biweights (TBS) estimator derived by Rocke (1996), the orthogonal Gnanadesikan–Kettenring (OGK) estimator suggested by Maronna and Zamar (2002), the median ball algorithm (MBA) proposed by Olive (2004), the mimimum generalized variance (MGV) and projection methods.[20-25] Wilcox (2008) compared some of the methods (OGK, TBS, MGV and MBA), and stated that no single method was always best.[26] He found that the MGV and projection methods performed relatively well when the number of variables is not too large (meaning that $p \leq 9$) and that these two methods break down with $p > 9$ variables. Zygmont and Smith (2014) specified that MVE method works well if the number of variables is less than 10.[14] Because there are many robust estimators and none of these methods is always best, researchers are forced to choose among these methods. In view of size and structure of the data, the MCD method among robust estimators has been preferred in this study.

The purpose of the MCD method is to obtain the subset including h observations (out of n) which will minimize the covariance matrix's determinant. The mean calculated via h observations will form the location estimation of MCD, and the covariance matrix obtained via the same observations will form the scatter estimation of MCD. However, to run this procedure, how to choose the initial subset C_1 including h observation has to be decided.

CHOOSING THE INITIAL SUBSET FOR THE MCD

In order to determine the initial subset C_1, Rousseeuw and Van Driessen (1999) considered the two situations below:[27]

1- Drawing a random h-subset C_1.

2- Drawing a random (p+1)-subset C_0. Since the lowest number of elements that a non-singular covariance matrix should have is p+1, the researchers suggested starting with a random subset of this size and adding elements until obtaining h elements.

In MCD method, after choosing the initial subset, it is necessary to determine the subset that will minimize the determinant of covariance matrix. On the other hand, it is quite difficult and almost impossible to find the subset including h observations that will minimize the covariance matrix's determinant especially in cases of a large number of n and p. In order to find a solution to this problem, various alternative algorithms were suggested. One algorithm developed by Hawkins (1994) suggests a steepest descent method to find a h-subset.[28] However, in order to speed up this method, the fast MCD method based on exchange of more than one observation in each step instead of exchange of an observation pair was developed later by Rousseeuw and Van Driessen (1999).[27]

ALGORITM OF FAST MCD

The basis of the fast MCD depends on the C-Step. Let $\{X_1, X_2, \dots, X_n\}$ be a random data set of p-variate observations. The method starts with the selection of $|C_1| = h$ from $C_1 \subset \{1,2,\dots,n\}$. The h value is considered to be made up of a minimum number of observations which do not include any outliers. The mean ($\bar{\boldsymbol{x}}_{C_1} := (1/h)\sum_{i\in C_1} \boldsymbol{x}_i$) and the covariance matrix ($\boldsymbol{S}_{C_1} := (1/h)\sum_{i\in C_1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{C_1})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{C_1})'$) are calculated based on C_1 subset. In addition, the distances between the observation values for each i element and the mean C_1 subset ($D_{C_1}(\boldsymbol{x_i}, \bar{\boldsymbol{x}}_{\boldsymbol{C_1}}) = D_{C_1}(i) := \sqrt{\left((\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{C_1})' \boldsymbol{S}_{C_1}^{-1}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_{C_1})\right)}$ for $i = 1, \dots, n$) are calculated. Following this, these distances are ordered ($(D_{C_1})_{1:n} \le (D_{C_1})_{2:n} \le \dots \le (D_{C_1})_{n:n}$). The observations that the lowest h number of these ordered distances ($\{D_{C_1}(i); i \in C_2\} := \{(D_{C_1})_{1:n} \le \dots \le (D_{C_1})_{h:n}\}$) belong to are assigned to the C_2 subset. Based on the new C_2 subset, the mean, covariance matrix and distances are calculated. Then det$(\boldsymbol{S}_{C_2})$ is less than or equal to det$(\boldsymbol{S}_{C_1})$ if and only if $\bar{\boldsymbol{x}}_{C_2} = \bar{\boldsymbol{x}}_{C_1}$ and $\boldsymbol{S}_{C_2} = \boldsymbol{S}_{C_1}$.[27]

Rousseeuw and Driessen (1999) refer to the process above as a C-step and define that "C stands for 'concentration' because the method concentrates on the h observations with smallest distances, and $(\boldsymbol{S}_{C_2})$ is more concentrated than $(\boldsymbol{S}_{C_1})$.[27] Repeating C-steps yields an iteration process. If det $(\boldsymbol{S}_{C_2}) = 0$ or det$(\boldsymbol{S}_{C_2}) =$ det$(\boldsymbol{S}_{C_1})$, the process is stopped; otherwise, another C-step yielding det $(\boldsymbol{S}_{C_3})$ continues". In other words, the process above is repeated until h observation that makes the distances smallest remains unchanged. In this way, the location and scale estimations calculated via the subset provide fast MCD estimations.

The fast MCD estimations obtained with this method are not much likely to be influenced by outliers. Therefore, distances based on MCD that is a robust estimator are highly likely to determine outliers. For outliers, these distance values will be bigger than other observations.[27]

Robust factor analysis conducted with the use of covariance matrix, which is less sensitive to outliers and which is based on fast MCD method, provides estimations fitting most of the data with a higher variance explained. Furthermore, the analysis gives factors including more conceptually meaningful variables.

DATA COLLECTION AND PARTICIPANTS

An e-mail survey was administered to a sample of university students who are members of AISEC and AEGEE student platforms in order to obtain the data regarding environmental sensitivity. AIESEC, which has 70,000 active members and spans 126 countries, is a global platform for young people and a non-political, independent and not-for-profit organization. AEGEE, which brings 13000 students from 40 different countries, is one of Europe's biggest interdisciplinary student organizations. Like AISEC, AEGEE is another non-governmental, politically independent non-profit organization. For a population of 83,000 students in these organizations, considering the probable number of questionnaire forms which might not have been returned, the questionnaire forms were sent to 650 randomly selected international university students from two global platforms (AISEC and AEGEE) through e-mail. A total of 541 students responded to our invitations and filled in the form. However, after the questionnaire forms which were not completed or filled out correctly were excluded, the analyses were conducted on the data collected from 505 participants.

In the study, the purpose was to compare the results of robust factor analysis and classical factor analysis via a real dataset. For this purpose, the questionnaire form was made up of 5-point Likert-type 24 items prepared in line with the scales reported in studies conducted by Cabuk and Karacaoglu (2003) and Yesilyurt et al. (2013).[6,29] The items included in the questionnaire scored as 1–20 (Strongly disagree), 21–40 (Disagree), 41–60 (Neutral), 61–80 (Agree), and 81–100 (Strongly agree). The items in questionnaire form were presented in Appendix.

EVALUATION OF THE APPROPRIATENESS OF THE DATA TO FACTOR ANALYSIS

A sample size for factor analysis should be at least 5:1 case-to-variable ratio, but the more admissible ratio is 10:1. Some researchers state that the sample size can be at least 20 cases for each variable.[30] Sample size in this study satisfied 20:1 case-to-variable ratio. In addition, several calculations were carried out to determine the appropriateness of the data set to factor analysis. First, according to Barlett's Test of Sphericity, the results indicate that the cross-correlations among the variables was found statistically significant to conduct factor analysis ($p=0.0001<\alpha=0.05$). In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy was examined and obtained to be 0.833. Depending on the results of both tests, the dataset was found appropriate to factor analysis.[31-33] Moreover, the normality of the data set was examined, and they were found to be distributed non-normally (Generalized Shapiro–Wilk's $W = 0.97845$, p-value $= 2.2\times10^{-16}$).

EXAMINING THE OUTLIERS IN THE DATASET

One of the necessities of robust analysis is to identify the outliers. For this purpose, the plot of the robust distances (RD) against the Mahalanobis distances (MD) is drawn.

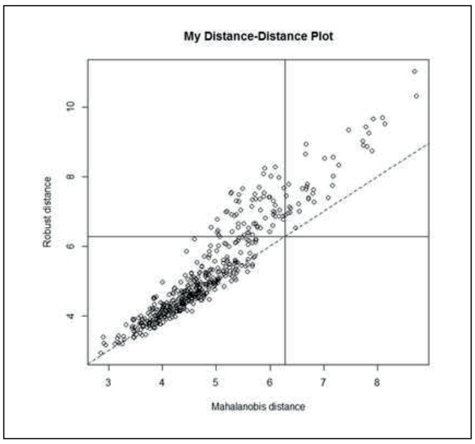

FIGURE 1: Distance of outliers.

It is possible to determine the outliers via the plot presented in Figure 1. If the data set is not contaminated, both robust and Mahalanobis distances give the same results, and all the points are found near the dashed line. The vertical and horizontal lines cut by the threshold value $\sqrt{\chi^2_{24;0.975}} \approx 6.27408$ divides the plot into four parts. The first part ($MD(x_i) \leq 6.27408$ and $RD(x_i) \leq 6.27408$) includes good observations that are not marked as outliers by both distances in the dataset. The second part ($MD(x_i) > 6.27408$ and $RD(x_i) \leq 6.27408$ shows the observations determined mistakenly as outliers by the classical method; however, in the plot, there is no observation in this part. The third part ($MD(x_i) > 6.27408$ and $RD(x_i) > 6.27408$) includes observations identified as outliers. The fourth part ($MD(x_i) \leq 6.27408$ and $RD(x_i) > 6.27408$ includes the masked outliers; in other words, this part includes the observations which are not determined as outliers by Mahalanobis distance but which are identified as outliers by robust distance.[18] According to the plot, it is seen that there are a number of outliers in the dataset.

The two most common methods used to determine the number of factors are the Scree Test and the Kaiser's little jiffy, in which the eigenvalues higher than 1 are selected. However, these two methods are not considered to be reliable since they overestimate or underestimate the number of factors.[34,35] Furthermore, Cattel (1966) regards scree test as a subjective method and states that its reliability is low.[36,37]

As an alternative to these two methods to determine the number of factors, Horn (1965) suggested parallel analysis.[38] In this study, the parallel analysis was used to determine the number of factors and six factors were obtained.

Considering the method of parallel analysis used to determine the number of factors, the robust factor analysis results were obtained by using the fast MCD estimators after the presence of outliers was determined. In addition, in both classical and robust factor analyses, the principal component analysis was used as the factor extraction method. Table 1 presents the factor loadings with varimax rotation, the communality and the total variance explained with the contribution of each factor for both analyses.

As the results of classical factor analysis were influenced considerably by outliers, related interpretations will be based only on robust factor analysis.

RESULT AND DISCUSSION

According to Table 1, the first six factors explained 64.6% of the total variance in classical factor analysis, while it explained 68.3% of the total variance in robust factor analysis.

In robust factor analysis, the variables contributing to the explanation of the first factor were X1-X5. This factor was called "environmental conservation". The variables contributing to the explanation of the second factor were X6-X9, and the factor was called "environmental product use". The third factor included the variables of X10-X12, and this factor was called "recycling". The name of the fourth factor that the variables of X13-X18 contributed to most was determined as "reduce/reuse". The variables contributing to the explanation of the fifth factor were X19-X21, and this factor was called "environmental response". As for the last factor, it was made up of the variables of X22-X24, and this factor was named "environmental education". When the gathering of the variables in the factors was examined, it was seen that the variables in robust factor analysis were gathered in a more conceptually meaningful manner when compared to the classical factor analysis.

CONCLUSION

Classical factor analysis is quite sensitive to outliers. This problem can be overcome by using robust estimators. The present study compared the results of classical and robust factor analyses to reveal the effects of outliers on classical factor analysis via a real dataset. Robust factor analysis based on fast MCD estima-

TABLE 1: Results of classical and robust factor analyses

	Classical factor analysis							Robust factor analysis						
Variables	Factor loadings						Commu-nalities	Factor loadings						Commu-nalities
	F1	F2	F3	F4	F5	F6	h^2	F1	F2	F3	F4	F5	F6	h^2
X1	0.865						0.689	0.867						0.638
X2	0.828						0.670	0.862						0.628
X3	0.852						0.904	0.817						0.919
X4	0.621						0.908	0.696						0.912
X5	0.477						0.429	0.636						0.525
X6		0.914					0.935		0.908					0.922
X7		0.906					0.686		0.903					0.731
X8		0.894					0.530		0.899					0.575
X9				0.487			0.939		0.526					0.941
X10			0.937				0.910			0.945				0.904
X11			0.916				0.418			0.926				0.421
X12			0.918				0.905			0.924				0.928
X13				0.559			0.705				0.661			0.767
X14					0.405		0.638				0.619			0.596
X15				0.612			0.460				0.604			0.449
X16				0.625			0.344				0.550			0.451
X17				0.626			0.350				0.548			0.500
X18				0.523			0.500				0.496			0.576
X19					0.817		0.486					0.858		0.394
X20					0.749		0.356					0.725		0.668
X21					0.750		0.782					0.663		0.825
X22				0.496			0.746						0.712	0.792
X23						0.733	0.757						0.692	0.737
X24						0.743	0.429						0.650	0.502
	F1	F2	F3	F4	F5	F6		F1	F2	F3	F4	F5	F6	
Eigenvalue	3.067	3.042	2.828	2.735	2.281	1.524		3.495	3.353	3.031	2.715	1.938	1.844	
Prop. Var.	0.128	0.127	0.118	0.114	0.095	0.064		0.146	0.140	0.126	0.113	0.081	0.077	
Cumul.Var.	0.128	0.255	0.373	0.487	0.582	0.646		0.146	0.286	0.412	0.525	0.606	0.683	

Prop.Var.: Proportional Variance; Cumul.Var.:Cumulative Variance

tors and classical factor analysis based on sample mean and covariance matrix were applied to the data regarding the environmental sensitivity of university students who were members of AISEC and AEGEE student platforms. When the results were compared with respect to each factor's proportion of explaining the total variance, it was seen that the proportion of variance explained by robust factor analysis was higher than classical factor analysis with 3.7%.

In addition, the presence of outliers in the study changed the relative importance of the factors. Moreover, the variables contributing to the explanation of some factors in robust factor analysis changed in classical factor analysis; in other words, variables X9, X14 and X22 had a different position in classical factor analysis.

In factor analysis, a factor including fewer than three variables is not desirable.[39-41] In classical factor analysis, the factor F6 consisted of two variables. On the other hand, all the factors in robust factor analysis were made up of at least three variables.

Consequently, the robust factor analysis aims at determining a factor structure which will not bias the parameter estimations and which fits most of the data by reducing the effect of outliers. In this study, use of robust factor analysis was suggested for the related dataset because of the variables gathered in a more conceptually meaningful manner and a higher rate of variance explained.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Ozlem Alpu, Hatice Samkar;* ***Design:*** *Ozlem Alpu, Hatice Samkar;* ***Control/Supervision:*** *Ozlem Alpu, Hatice Samkar;* ***Data Collection and/or Processing:*** *Ozlem Alpu, Hatice Samkar;* ***Analysis and/or Interpretation:*** *Ozlem Alpu, Hatice Samkar;* ***Literature Review:*** *Ozlem Alpu, Hatice Samkar;* ***Writing The Article:*** *Ozlem Alpu, Hatice Samkar;* ***Critical Review:*** *Ozlem Alpu, Hatice Samkar;* ***References and Fundings:*** *Ozlem Alpu, Hatice Samkar.*

REFERENCES

1. Leeming FC, Dwyer WO, Bracken BA. Children's environmental attitude and knowledge scale: construction and validation. The Journal of Environmental Education 1995;26(3):22-31.
2. La Trobe HL, Acott TG. A modified NEP/DSP environmental attitudes scale. The Journal of Environmental Education 2000;32(1):12-20.
3. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. The Counseling Psychologist 2006;34:806-38.
4. Sasmaz Oren F, Kiyici G, Erdogmus E, Sevinc OS. [The scale on environmentally aware teachers' qualities: a validity and reliability study]. Inonu University Journal of the Faculty of Education 2010;11(1):133-52.
5. Markle GL. Pro-environmental behavior: does it matter how it's measured? Development and validation of the pro-environmental behavior scale (PEBS). Human Ecology 2013;41(6):905-14.
6. Yesilyurt S, Gul S, Demir Y. [Biology teacher candidates' environmental awareness and environmental sensitivity: a scale development study]. Mehmet Akif Ersoy University Journal of Education Faculty 2013;25:38-54.
7. Dikmenli Y, Konca AS. [Validity and reliability study of environmental consciousness scale]. Journal of Theory and Practice in Education 2016;12(6):1273-89.
8. Hayashi K, Yuan KH. Robust bayesian factor analysis. Structural Equation Modeling 2003;10(4):525-33.
9. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis. 5th ed. New Jersey: Prentice Hall; 1998. p.730.
10. Pison G, Rousseeuw PJ, Filzmoser P, Croux C. Robust factor analysis. Journal of Multivariate Analysis 2003;84(7):145-72.
11. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 1999;4(3):272-99.
12. Russell DW. In search of underlying dimensions: the use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. Personality and Social Psychology Bulletin 2002;28(12):1629-46.

13. Treiblmaier H, Filzmoser P. Exploratory factor analysis revisited: how robust methods support the detection of hidden multivariate data structures in IS research. Information and Management 2010;47(4):197-207.
14. Zygmont C, Smith MR. Robust factor analysis in the presence of normality violations, missing data, and outliers: empirical questions and possible solutions. The Quantitative Methods for Psychology 2014;10(1):40-55.
15. Kosfeld R. Robust exploratory factor analysis. Statistical Papers 1996;37(2):105-22.
16. Basilevsky A. Statistical Factor Analysis and Related Methods: Theory and Application. 1st ed. New York: John Wiley & Sons; 1994. p.768.
17. Tanaka Y, Odaka Y. Influential observations in principal factor analysis. Psychometrika 1989;54(3):475-85.
18. Marcoulides GA, Moustaki I. Latent Variable and Latent Structure Models. 1st ed. London: Lawrence Erlbaum Associates Publishers; 2002. p.279.
19. Yuan KH, Bentler PM. Robust mean and covariance structure analysis. Br J Math and Stat Psychol 1998;51(Pt 1):63-88.
20. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust Statistics: The Approach Based on Influence Function. 1st ed. New York: John Wiley & Sons; 1986. p.502.
21. Rousseeuw PJ. Multivariate estimation with high breakdown point. In: Grossmann W, Pflug G, Vincze I, Wertz W, eds. Mathematical Statistics and Applications. Vol B. 1st ed. Dordrecht: Reidel Publishing Company; 1985. p.283-97.
22. Rousseeuw PJ, Leroy AM. Robust Regression and Outlier Detection. 1st ed. New York: John Wiley & Sons; 1987. p.329.
23. Rocke DM. Robustness properties of S-estimators of multivariate location and shape in high dimension. Annals of Statistics 1996;24(3):1327-45.
24. Maronna RA, Zamar RH. Robust estimates of location and dispersion for high-dimensional datasets. Technometrics 2002;44(4):307-17.
25. Olive DJ. A resistant estimator of multivariate location and dispersion. Computational Statistics and Data Analysis 2004;46:93-102.
26. Wilcox RR. Some small-sample properties of some recently proposed multivariate outlier detection techniques. J Stat Comput Simul 2008;78:701-12.
27. Rousseeuw PJ, Driessen KV. A fast algorithm for the minimum covariance determinant estimator. Technometrics 1999;41(3):212-23.
28. Hawkins DM. The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data. Computational Statistics and Data Analysis 1994;17(2):197-210.
29. Cabuk B, Karacaoglu OC. [Examining the environmental sensitivities of university students]. Ankara University Journal of Faculty of Educational Sciences 2003;36(1-2):189-98.
30. Hair Jr JF, Black WC, Babin BJ, Anderson R, Tatham RE. Multivariate Data Analysis. 7th ed. Upper Saddle River: Prentice Hall; 2009. p.800.
31. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. New Jersey: Pearson Education, Inc; 2013. p.1056.
32. Meyers LS, Gamst G, Guarino AJ. Applied Multivariate Research: Design and Interpretation. 2nd ed. Thousand Oaks: Sage Publications; 2006. p.722.
33. Hutcheson GD, Sofroniou N. The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models. 1st ed. Thousand Oaks: Sage Publications; 1999. p.288.
34. Thompson B. Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. 1st ed. Washington DC: American Psychological Association; 2004. p.195.
35. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment 1995;7(3):286-99.
36. Cattell RB. The scree test for the number of factors. Multivariate Behav Res 1966;1(2):245-76.
37. Streiner DL. Factors affecting reliability of interpretations of scree plots. Psychological Reports 1998;83(2):687-94.
38. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965;30(2):179-85.
39. O'Rourke N, Hatcher L. A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling. 2nd ed. North Carolina: SAS Institute, Inc; 2013. p.444.
40. Osborne JW, Costello AB. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research and Evaluation 2005;10(7):1-9.
41. Velicer WF, Fava JL. Effects of variable and subject sampling on factor pattern recovery. Psychological Methods 1998;3(2):231-51.