Science Journal of Applied Mathematics and Statistics
2023; 11(3): 48-55
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20231103.12
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

On Different Extraction Methods of Factor Analysis

Awogbemi Clement Adeyeye[1], Koyejo Samuel Olusegun[2], Olowu Abiodun Rafiu[3]

[1]Statistics Programme, National Mathematical Centre, Abuja, Nigeria
[2]Statistics Department, Federal University of Technology, Akure, Nigeria
[3]Mathematics Programme, National Mathematical Centre, Abuja, Nigeria

Email address:
awogbemiadeyeye@yahoo.com (Awogbemi Clement Adeyeye)

To cite this article:
Awogbemi Clement Adeyeye, Koyejo Samuel Olusegun, Olowu Abiodun Rafiu. On Different Extraction Methods of Factor Analysis. *Science Journal of Applied Mathematics and Statistics*. Vol. 11, No. 3, 2023, pp. 48-55.doi: 10.11648/j.sjams.20231103.12

Received: October 1, 2023; **Accepted**: November 8, 2023; **Published**: November 21, 2023

Abstract: This study aims at examining and comparing different methods of extracting factor analysis and applying such to real life scenario. Factor analysis simplifies complex and diverse relationships existing among a set of observed variables. This is carried out by unfolding common factor connecting unrelated variables that provide insight to the underlying data structure. Since common factors have unit variance, the variance of a given variable is partitioned into common variance and unique variance which were used to generate the total variance. The model assumptions for both random and non-random factor score analyses were examined to ascertain whether or not the model contains the model parameters to be estimated. Different methods of extracting factor analysis were examined and applied for possible comparison. The centroid method maximizes the sum of loadings without giving recourse to the signs; the principal factor method accounts for the maximum feasible amount of variance in the variables being factored and the maximum likelihood method maximizes the relationship between the sample of data and the population from which the sample is drawn. It was established that the principal component method is scale invariant while the maximum likelihood method of factor analysis provides the best estimate for the reproduced correlation matrix with convergence to the best value. It is therefore asserted that different extraction methods produce different solutions.

Keywords: Factors, Correlation Matrix, Eigen Value, Communality, Common Variance, Factor Scores

1. Introduction

Factor analysis is a data reduction multivariate method used for the extraction, rotation and naming of some underlying explanatory constructs that describe manifest variables that are innumerable and extracted from several interrelated manifest variables. The variables are technically rotated to explain the total variance in specific sets of manifest variables [9, 1]. It is in other way, a higher order reduction family of multivariate statistical techniques that best combines variables found to be naturally and practically measuring the same continuous latent construct, factor in either exploratory or confirmatory mode [10, 6].

Factor analysis seeks to resolve a large set of measured variables in terms of relatively few factors or categories which enables researchers to group variables into factors that may be treated as new variables by summing up the values of the original variables grouped into the factors. The identification of such new variable is subjectively determined by researchers considering the factors that are linear combinations of given data. Thus, the coordinates of each variable is measured to obtain the factor loadings which represents the correlation between the given variable and the factor placed in a matrix correlation between the variable and the factors.

Factor analysis is classified into two types, which are confirmatory factor analysis (has predetermined constraints on factor loadings) and explanatory factor analysis which has no constraints [12, 13].

There are different methods used for factor extraction, but all may not necessarily produce similar results. This implies that factor analysis is not just a single method, but a set of methods.

This study therefore examines some methods used in factor analysis with a view to establish some possible comparisons using real life applications.

1.1. Definitions of Terms

(a) Factor: A factor is an underlying dimension that accounts for several observed variables and also defines the way entities differ.

(b) Factor Loadings: These are values (factor variable correlations) that explain how closely variables are to each one of the factors discovered [4].

(c) Factor Scores: Factor scores indicate the degree to which each respondent earns high scores on the group of items that load high on each factor.

(d) Communality: A communality denoted by h^2 shows how much of each variable is accountedfor the underlying factor taken together. The communality of the i^{th} variable σ_{ii} is the portionof the variance of the i^{th} variable that is explained by the m common factors. It isdenoted by $\sigma_{ii} = h_i^2 + \Psi_i$, where σ_{ii} is the variance of X_i(sum of squared loading for X_i); $h_i^2 = (\lambda\lambda')_{ii} = \lambda_{i1}^2 + \lambda_{i2}^2 + \lambda_{i3}^2 + \ldots, + \lambda_{im}^2$ is the communality of X_i and Ψ_i is the specific variance of X_i [8]. Unique Variance: The unique variance of a variable, Ψ_i reflects the extent to which the common factors fail to account for the variance of the variable (portion left unexplained by the common factors).

(e) Manifest Variables: These are variables whose values are directly observed or measured.

(f) Eigen value: This refers to a mathematical index which depicts the magnitude of the variable among the several correlated variables that is accounted for by a single but much more illuminating underlying factor.

1.2. Preliminaries

1.2.1. Factor Analysis Model

Let N denote the number of variables and n denote the sample size. Then the observation y_{ij} for the ith variable and the jth sample is assumed to have the following decomposition:

$$y_{ij} = \sum_{k=1}^{r} l_{ik} f_{kj} + \sigma_i e_{ij,} \tag{1}$$

where $E = (e_{ij})N \times n$ is the noise matrix, $F_k = (f_{k1,}f_{k2,\ldots,}f_{kn,})^T$ denotes the k^{th} latent variable and $L = (l_{ik})N \times r$ is called the factor loading matrix. Let each observed variable be denoted by $Y_i = (y_{i1,}y_{i1,} \cdots, y_{in,})^T$ and the noise associated as $E_i = (e_{i1,}e_{i1,} \ldots e_{in,})^T$, then the vector form of (1) is

$$Y_i = \sum_{k=1}^{r} l_{ik} F_k + \sigma_i E_i \tag{2}$$

This exactly shows that all the observed variables can be explained by linear combinations of r common factors which is much smaller than N.

Let $Y = (y_{ij})N \times n$ be the data matrix and $F = (f_{kj})r \times n$ be the factor score matrix, then the matrix form of (1.1) is

$$Y = LF + \sum^{1/2} E \tag{3}$$

This has the interpretation that the data matrix can be expressed as a low-rank signal matrix $X = LF + E_i$ noise. Thus, factor analysis model can be used when a low-rank approximation of the data matrix is so desired.

The factor analytic model is expressed as

$$X = \lambda f + e, \tag{4}$$

where $X = m -$ dimensional vector of observed responses, $X' = (x_{1,}x_{2,\ldots,}x_m)$, $f =$ n-dimensional vector of unobservable variables called unique factors, $e' = (e_1, e_{2,\ldots,}e_n)$

$\lambda = m \times n$ matrix of unknown constants (factor loadings)so that

$$\lambda = \begin{pmatrix} \lambda_{11}\,\lambda_{12}\,\ldots\lambda_{1n} \\ \lambda_{21}\,\lambda_{22}\,\ldots\lambda_{2n} \\ \vdots \quad \vdots \quad \vdots \\ \lambda_{m1}\,\lambda_{m2}\ldots\,\lambda_{mn} \end{pmatrix}, ee' = \psi = \begin{pmatrix} \psi_1 & 0 \ldots & 0 \\ 0 & \psi_2 \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \ldots & \psi_m \end{pmatrix}, \text{Cov}(e, f') = 0$$

The model given by equation (1) with the imposed assumptions is an indication that the covariance matrix of the response vector X, denoted by $\sum_{XX}$ can be expressed as

$$\sum_{XX} = \lambda\phi\lambda' + \psi \tag{5}$$

where λ and ψ are as previously defined;

$$\phi = \begin{pmatrix} 1 & & & \\ \phi_{21} & 1 & & \\ \vdots & & & \\ \phi_{m1} & \phi_{m2} & \phi_{n,n-1} & 1 \end{pmatrix} \tag{6}$$

The factor analysis by equation (1) is otherwise expressed as a factor pattern:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \cdot \\ \cdot \\ X_m \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \ldots \lambda_{1n} \\ \lambda_{21} & \lambda_{22} \ldots \lambda_{2n} \\ & \vdots \\ \lambda_{m1}\,\lambda_{m2} & \ldots \lambda_{mn} \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \cdot \\ \cdot \\ f_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ \cdot \\ e_n \end{pmatrix} \Rightarrow X_i = \sum_{j=1}^{n} \lambda_{ij} f_j + e_i \tag{7}$$

Let the correlation matrix be R so that equation (5) becomes

$$R = \lambda\phi\lambda' + \psi, \tag{8}$$

where the $n \times n$ symmetric matrix ϕ contains the correlation matrix between the common factors and $\lambda\phi\lambda'$ is the common factor correlation matrix.

1.2.2. Model Assumptions for Random Factor Score Matrix

Let the columns of Y be randomly and independently drawn from the population, assuming that F is random to reduce the number of parameters to estimate [7]. Thus, the following assumptions hold:

1. The factor scores F and the noise E are random and the factor loading matrix L is non-random.
2. F and E are independent
3. For each latent variable k, ($f_{kj,j=1,2,\ldots n}$) are independently and identically distributed randomvariables with $(f_{kj}) = u_k$ and cov $(F_{.j})$= Λ_F, where $\Lambda_F \in \mathbb{R}^{r\times r}$is some positive definitesemi-definite matrix.
4. For eachvariable i, the noises ($e_{i1,}e_{i2,}\ldots e_{in,}$) are independently and identically distributed with $E(e_{ij})$= a_i, Var(e_{ij}) = 1 and $\sum = dia(\sigma^2_{1,}\sigma^2_{2,}, \ldots, \sigma^2_{N,})$
5. The entries of F and E follow Gaussian distribution with$E(f_{kj}) = 0$ and E (e_{ij})= 0

1.2.3. Model Assumptions for Non-Random Factor Score Analysis

We assume that a non-random variable F when the distributional assumption of F is very complex or estimating the low rank matrix $X = LF$ is more convenient to estimate than the factors. Comparing this with the random F assumptions, the model contains the model parameters to be estimated [11]. The assumptions for a non-random model are specified:

(a) The noises E are random, but the factor loading L and the factor score are non-random

(b) For each variable i, the noises $e_{i1,}\ e_{i2,}\ e_{i3,}, \ldots,\ e_{in,}$ are independently and identically distributed with $E(e_{ij}) = a_i$, Var(e_{ij}) =1 and $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$.

The non-random factor score model can be expressed as

$$Y = \alpha 1_n^T + LF + \Sigma^{\frac{1}{2}}\mathrm{E}, \tag{9}$$

where $E.I_n = 0$

1.2.4. Model Identification for Random Factor Score

When the number of factors r is not known, there exists an identification problem for r. Setting r = N and $\Sigma = 0$, the model becomes a correct model trivially. In order to do away with this scenario, r is defined as the minimum integer that the factor model exists. Thus, the uniqueness of r is automatically guaranteed since r = N gives a correct model. The normality for all the random variables is assumed and the identification of the factor analysis model parameters for both random and non-random factor score models are examined:

(a) Identification for Random Factor Score Model

For a random factor score model, certain constraints are required to identify each elements of the covariance matrix $L\Lambda_\mathrm{F}\mathrm{L}^\mathrm{T} + \Sigma$. With this, a sufficient condition for identification of $\Phi = L\Lambda_\mathrm{F}\mathrm{L}^\mathrm{T}$ and $\sum$ given r will be established.

Theorem 1

The sufficient condition for the identification of $\Sigma\ and\ \Phi = L\Lambda_\mathrm{F}\mathrm{L}^\mathrm{T} + \Sigma$ is that if any row of Φ is deleted, then there will be two disjoint subsets of rows of Φ of rank r.

The identification of L and Σ_F from Φ is examined under sparsity assumption. This is the same as the determination of U up to scaling and row or column permutation identity matrix for $\tilde{L} = LU$

Definition

A s-sparse family of L, where $s \geq r$ is defined as $\mathcal{L}(s) = \{L \in \mathbb{R}^{N\times r}\}$ such that L satisfies conditions (a) and (b) below:

(a) L is of rank r and each column of L contains at least s zeros

(b) For each column k, assume L_k is the matrix consisting of all rows of L which have zero in the k^{th} column. For any $k = 1,2,3, \ldots, r, L_k$is of the rank r-1.

Theorem 2

Considering Model Assumptions for Random Factor Score Matrix, the normality assumption and the identification conditions in Theorem 1, a necessary and sufficient condition for L in $\mathcal{L}$(s) identifiable up to scaling and permutation of row or column is that if a submatrix $L^* \in \mathbb{R}^{s\times r}$ of L is of the rank r-1, then it must be the sub-matrix off or some $m = 1,2, \ldots, r$.

Proof

For any $\tilde{L} = LU$, it will be shown that if $\tilde{L}$, L$\epsilon\mathcal{L}$, the condition in the theorem is a necessary and sufficient condition for U having exactly one non zero in each row and each column.

(i). Sufficient Condition

Since $\tilde{L}$ has rank r, U is a full rank and $L = \tilde{L}U^{-1}$. For any given m ϵ $\{1,2, \ldots, r\}$ as the rank of $\tilde{L}_m$ is r-1, there exists an s $\times$ r sub matrix$\tilde{L}^*$ of $\bar{L}_m$ that is of rank r-1, then $\tilde{L}^*U \in \mathbb{R}^{s\times r}$ also has rank r -1. Since $\tilde{L}^*U$ is a sub-matrix of L, then given the condition, $j_m(\tilde{L}^*U)$ are all equal to zero.Let $\tilde{L}^*{}_{(m)}$ be the sub-matrix of $\tilde{L}^*$ jettisoning the m^{th} column. Since $\tilde{L}^*{}_{(m)} \in \mathbb{R}^{s\times(r-1)}$ is of rank r-1, then the entries of $j_m\,th$ column except for mth row of U must be zero.

Since $m_1 \neq m_2$ which implies that $j_{m1} \neq j_{m2}$, then U has exactly one non-zero in every row and column.

(ii). Necessary Condition

If the condition in the given theorem is not satisfied, then there is a sub-matrix $\tilde{L}^* \in \mathbb{R}^{s\times r}$ of L that has rank r-1 with none of its column equal to zero. Thus, there exists $k\epsilon\mathbb{R}^r$ that has at least two non-zero entries and $L^*k = 0$ (assume that the first entry of k is non zero).Let $K_{-1} = (0I_{r-1})^T \in \mathbb{R}^{r\times(r-1)}$ and U =$kK_{-1} \in \mathbb{R}^{r\times r}$. Then U has rank r and its easy to find out that LU$\epsilon\mathcal{L}$. Therefore, L cannot be identified.

1.2.5. Model Identification for Non-Random Factor Score

For a non random factor score model, some constraints are required to identify the signal matrix X and noise covariance Σprovided r and some constraints are given to identify F and L inX =LF. In order to identify X and Σ of the model $Y = X + \Sigma^{\frac{1}{2}}E$, there is a need to be sure that if $Y = X' + \Sigma^{\frac{1}{2}}E'$, with X' is of rank r, Σ' diagonal E' a random matrix with identically independently distributed standard Gaussian entries, then $X = X'$ and $\Sigma' = \Sigma$. If $r = N$, then the model cannot be identified trivially. Thus, there is a need to have $r < N$ so that a necessary condition for identifying X and Σ is given.

Theorem 3

Given that $r < N$ with a known r, then a necessary condition for identifying X and Σis that is any row of x is removed, the remaining matrix remains that of rank r.

Proof

If there exists one row j (j^{th} row is removed), the remaining matrix $X_{(j)}$ has the rank $v < r.$ Let the remaining matrix of L after removing the j^{th} row be $L_{(j)}$ so that $L_{(j)} \in \mathbb{R}^{(p-1)\times r}$ has rank v. Thus, there exists a non zero vector $k \in \mathbb{R}^r$ that $L_{(j)} = 0$ since L is of full rank, Lk= 0. Thus implies that Lk has only one nonzero entry.

Let $X' = X + LvE_1^T$, where $E_1 \in \mathbb{R}^m$ is a random vector with independent and identically distributed standard Gaussian entries independent ofE. Then X' is of rank r and $\Sigma' + \Sigma - Lkk^T L^T$ is a diagonal, where X and Σ cannot be identified. This establishes the necessary condition.

2. Methods

The common variance of variable X_i is employed to generate the total variance. Different extraction methods of factor analysis are also examined in this section.

2.1. Generating the Total Variance from the Common Variance

Given the linear factor model in (7), each equation partitions the variable X_i into two uncorrelated parts

$$X_i = c_i + e_i, \tag{10}$$

where $c_i = \lambda_{i1}f_{1+}\lambda_{i2}f_{1+} \dots, +\lambda_{in}f_n$ is that part of each variable that is common to the other p-1 variables, and e_i is the part of each variable that is unique.

Since the common and unique parts of a variable are uncorrelated and common factors have unit variance, the variance X_i is partitioned to

$$var(X_i) = var(c_i) + var(e_i), \tag{11}$$

where $var(c_i)$ and $var(e_i)$ denote the common variance and the unique variance of X_i respectively.

Let the communality of i^{th} variable be h_i^2, then

$$var(X_i) = h_i^2 + \Psi_i, \tag{12}$$

where $var(X_i) = \Psi_i$. The sum of the squared elements in the ith row ofΛ is

$$var(c_i) = \sum_{j=1}^{n} \lambda_{ij}^2 = h_i^2 \tag{13}$$

The total contribution of factor f_j to the total variance of the entire set of variables is given by the eigen value of the factor f_j and computed as

$$Q_j = \sum_{j=1}^{n} \lambda_{ij}^2 = \lambda_j' \lambda_j, \tag{14}$$

where λ_j denotes the jth column of Λ. Equation (14) is the squared factor loadings, $\sum_{j=1}^{n} \lambda_{ij}^2$, for $j = 1,2, \dots, n$. The total contributions of all the common factors to the total variance among all the variables (total communality) is computed as

$$Q = \sum_{j=1}^{n} Q_j \tag{15}$$

The variance among all the variables that is accounted for by a factor f_j as a percentage of that accounted for by all the factors is given as

$$Q = \frac{Q_j}{Q} \tag{16}$$

Therefore, the total variance is written as

$$Tr(\Sigma_{XX}) = \sum_{j=1}^{n} Q_j + \sum_{i=1}^{m} \Psi_i = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij}^2 \sum_{i=1}^{m} \Psi_i \tag{17}$$

2.2. Extraction Procedures of Factor Analysis

2.2.1. Centroid Method

Centroid method maximizes the sum of loadings, disregarding signs. The method extracts the largest sum of absolute loadings for each factor in turn and it is defined by linear combinations in which all weights are either 1 or –1 [2]. The centroid method procedure is specified as follows:

(a) Compute a matrix of correlations, R with the product moment formula used for working out the correlation coefficients.

(b) If the correlation matrix is positive, the centroid method requires that the weights for all variables be +1.0. In case the correlation matrix is not a positive manifold, then reflections must be made before the first centroid factor is obtained.

(c) The first centroid factor is computed as follows:

1. The sum of the coefficients in each column of the correlation matrix is worked out.
2. Then the sum of these column sums (T) is obtained.
3. The sum of each column obtained as per (a) above is divided by the square root of T obtained in (b) above, resulting in what is called centroid loadings. The full set of loadings so obtained constitutes the first centroid factor (say C_1).
4. To obtain second centroid factor (say C_2), one must first obtain a matrix of residual coefficients. This is done for all possible pairs of variables and the resulting matrix of factor cross products may be named as Q_1.
5. For subsequent factors (C_3, C_4,…), the same process outlined is repeated.

2.2.2. Principal Factor Method

The principal factor method similar to principal component analysis, extracts factors such that each factor accounts for the maximum possible amount of the variance contained in the set of variables being factored. Principal component analysis finds the linear combinations of the observed variables to maximize the sample variance. Suppose $F = (Y - \breve{\varphi}1^D)(Y - \breve{\varphi}1^D)\frac{D}{m}$. Let the eigenvalue decomposition of F be $F = Q\Lambda Q^D$, where $\Lambda = diag(\lambda_1, \ldots, \lambda_M)$ is adiagonal matrix ($\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_M \geq 0$) and Q is an orthogonal $M \times M$ matrix. The eigenvectors, $(Q_{.1}, \ldots, Q_{.M})$ are called loadings and the rows of $Q^D(Y - \breve{\phi}1^D)$ are called principal components [5].

In factor analysis, the loadings of principal components are used to estimate the linear space of factors scores. Thus, principal components method aims at constructing out of a given set of variables X_j's (j =1,…, m) new variables, z_i known as the principal components that are linear combinations of Xs:

$$l_1 = b_{11}Y_1 + b_{12}X_2 + \cdots + b_{im}X_m$$

$$l_2 = b_{21}Y_1 + b_{22}X_2 + \cdots + b_{2m}X_m$$

$$\vdots$$

$$l_m = b_{m1}Y_1 + b_{m2}X_2 + \cdots + b_{mm}X_m, \qquad (18)$$

where b_{ij}'s are the loadings generated such that the principal components extracted satisfy these requirements:

(i) orthogonality (ii) l_1 has the maximum variance followed by l_2 and up tol_m.

The following steps in principal component approach:

a. Estimates of a_{ij} 's are obtained with X's and are transformed into orthogonal variables.

b. Regress Y on the principal components as $Y = \hat{y}_1 p_1 + \hat{y}_2 p_2 + \cdots, \hat{y}_n p_n (n < m)$.

c. Find b_{ij} of the initial model from $\hat{a}_{ij}$ and $\hat{y}_{ij}$ by transferring back from the p's into the standardized X's.

Alternatively, the correlation coefficients between pairs of K variables are obtained and arranged in form of a correlation matrix, R. Assuming the correlation matrix to be positive manifold, the sum of coefficients in each column is also obtained along with the vector of column sums (U_{a1}) which is also normalized (V_{a1}). This is carried out by squaring and summing the column sums in $U_{a1}$and dividing each element in $U_{a1}$by the square root of the sum of squares. The elements in V_{a1} are cumulatively multiplied by the first row of R to generate the first element in a new vector V_{a2}.In order to obtain the second elements U_{a2}, similar procedure is repeated and a new vector U_{a2} is produced which is subsequently normalized to obtain V_{a2}. Comparison is then made between V_{a1} and V_{a2}, and if they are identically near, convergence is said to have taken place. If not, the trial vectors will be used repeatedly until convergence occurs.

To generate the second factor, B, the solutions to V_b are obtained and the factor loadings for second component factor, B. The process is repeated until successive principal component factors are obtained.

2.2.3. Maximum Likelihood Method

Statistically, maximum likelihood method maximizes some relationship between the sample of data and the population from which the sample is drawn. The method obtains sets of factor loadings successively in a way that each explains as much as possible concerning the population correlation matrix as estimated from the sample correlation matrix.

Let R_s be the correlation matrix obtained from the data in a sample, then maximum likelihood method tends to extrapolate what is known from R_s. The loadings generated on the first factor are used to obtain a matrix of the residual coefficient. A significant test is then used to indicate whether it would be reasonable to extract a second factor. This is carried out repeatedly until factoring after the test of significance fails to reject the null hypothesis for residual matrix [2].

Let the fitting function that is maximized be the likelihood function given by

$$lnL = -\tfrac{1}{2}n[ln|\Sigma| + tr(\Sigma^{-1}S)]$$

$$= -\tfrac{1}{2}n[\ln(|\Lambda\phi\Lambda'| + \Psi) + tr\{(\Lambda\phi\Lambda' + \Psi)^{-1}S)\}], \quad (19)$$

where S is the sample covariance matrix calculated from a random sample of size n from a multivariate normal distribution. Equation (19) is to be differentiated with respect to Λ, $\phi$and Ψ and then solve the resulting system.

Maximizing ln L results to minimizing

$$\mathrm{F}(\Lambda, \phi, \Psi) = ln|\Sigma| + tr(\Sigma^{-1}S) - ln|S| - p, \qquad (20)$$

where n times the minimum value of F generates the likelihood ratio test statistic of goodness of fit.

In order to minimize F, its partial derivatives with respect to the elements Λ and the diagonal elements of Ψ is taken to have:

$$\frac{\partial F}{\partial \Lambda} = 2\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\Lambda \qquad (21)$$

$$\frac{\partial F}{\partial \Psi} = diag(\Sigma^{-1}(\Sigma - S)\Sigma^{-1}) \qquad (22)$$

where diag (.)is the diagonal matrix formed from (.) by replacing all non-diagonal elements of (.) with zeros. There are p equations, $\frac{1}{2}p(p+1)$ distinct elements in Σ, pq parameters in Λ, $\frac{1}{2}q(q+1)$ parameters in ϕ and p parameters in Ψ [14].

2.3. Statistical Test of Significance

The null hypothesis is that all the variance has been extracted by the hypothesized number of factors. The test of significance gives an asymptotic χ^2 statistic. Supposed at a specified probability level, the χ^2value is significant, then the residual matrix has significant variance in it and more factors are required to reproduce the correlations between the original variables. The test statistic is of the form

$$(n-1)ln\frac{|\Lambda\Lambda' + \Psi|}{|S|} + tr[(\Lambda\Lambda' + \Psi)^{-1}S] - p \qquad (23)$$

with degree of freedom given as $\frac{1}{2}[(p-q)^2 - p - q]$ [3].

3. Illustrative Examples

In order to apply different methods of factor analysis, illustrations are considered with their interpretations.

Application 3.1

The correlation matrix, R of the 8 variables in Table 1 is used to generate the variance accounted for common variance (eigen value), proportions of total variance and common variance related to centroid method and principal components method of factor analysis.

Table 1. *Correlation Matrix of Specified Variables.*

Variables ⇒ ⇓	A	B	C	D	E	F	G	H
A	1.0000	0.7090	0.2040	0.8100	0.6260	0.1130	0.1550	0.7740
B	0.7090	1.0000	0.0510	0.0890	0.5810	0.0980	0.0830	0.6520
C	0.2040	0.0510	1.0000	0.6710	0.1230	0.6890	0.5820	0.0720
D	0.0810	0.0890	0.6710	1.0000	0.0220	0.7980	0.6130	0.1110
E	0.6260	0.5810	0.1230	0.0220	1.0000	0.0470	0.2010	0.7240
F	0.1130	0.0980	0.6890	0.7980	0.4700	1.0000	0.8010	0.1200
G	0.1550	0.0830	0.5820	0.6130	0.2010	0.8010	1.0000	0.1520
H	0.7740	0.6520	0.0720	0.1110	0.7240	0.1200	0.1520	1.0000

Table 2. *Computation of Eigen values and Communalities using Centroid Method.*

Variables	First Factor Loadings	Second Factor Loadings	Communality (h^2)
A	0.9300	0.5630	0.7970
B	0.6180	0.5770	0.7150
C	0.6420	-0.5390	0.7030
D	0.6410	-0.6020	0.7730
E	0.6290	-0.5580	0.7070
F	0.6940	-0.6300	0.8790
G	0.6790	-0.5180	0.7290
H	0.6830	0.5930	0.8180
Eigen value	3.4900	2.6310	6.1210
Proportion of Total Variance	0.4400	0.3300	0.7700
Proportion of CommonVariance	0.5700	0.4300	1.0000

Table 3. *Computation of Eigen values and Communalities using Principal Components Method.*

Variables	First Principal Component	Second Principal Component	Communality (h^2)
A	0.6900	0.5700	0.8010
B	0.6200	0.5900	0.7330
C	0.6400	-0.5200	0.6800
D	0.6400	-0.5900	0.7580
E	0.6300	0.5700	0.7220
F	0.7000	-0.6100	0.8620
G	0.6800	-0.4900	0.7030
H	0.6800	-0.6100	0.8350
Eigen value	3.4900	2.6007	6.0921
Proportion of Total Variance	0.4360	0.3250	0.7610
Proportion of CommonVariance	0.5730	0.4270	1.0000

From Table 2, the first factor has a loading in excess of 0.33 on all variables. This is the general factor that represents what all the variables have in common. The second factor has all the loadings in excess with 50% negative signs. This represents a bipolar factor with a single dimension of poles.

Approximately, 77% of the total variance is common variance and the remaining 23% is made up of portions unique to individual variables. The common variance of approximately 57% is accounted for by the 1st factors and the remaining 43% by the second factor. Therefore, the two factors explain the common variance.

From Table 3, the eigen values total for the two principal components is 6.0921 with 44% of the total variance accounted for by the first principal component. A total variance of 33% is accounted for by the second principal component. Likewise, the common variance of approximately 57% is accounted for by the first principal component and approximately 43% of the common variance is accounted for by the second principal component. Thus, the two principal components altogether explain the common variance.

Application 3.2

The Illustrative examples are established by adapting a real life data from the characteristics of 14 selected countries' projects [14].

Table 4. *Data from Dimensionality of 14 Countrie' Projects.*

Characteristics	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
GNP per Capital($)	91	51	58	359	134	70	129	515	70	707	468	749	998	2334
Trade(millions of $)	2.729	407	349	1169	923	2689	1601	415	83	5395	1852	6530	18677	26836
Power (rank)[a]	7	4	11	3	5	10	8	2	1	6	9	13	12	14

Characteristics	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
Stability[b]	0	0	0	0	1	0	0	1	0	1	0	1	1	1
Freedom of group opposition[c]	2	1	0	1	1	2	1	2	1	2	0	0	2	2
Foreign conflict[d]	0	0	1	0	1	0	0	1	1	0	1	1	1	1
Agreement with US in UN[e]	69.1	-9.5	-41.7	64.3	-15.4	-28.6	-21.4	42.9	8.3	52.3	-41.7	-41.7	69	100
Defence budget (millions of $)	148	74	3054	53	158	410	267	33	29	468	220	34000	3934	40641
GNP for Defence (%)	2.8	6.9	8.7	2.4	6	1.9	6.7	2.7	25.7	6.1	1.5	20.4	7.8	12.2
Acceptance of International law[f]	0	0	0	0	1	1	0	1	0	1	0	0	0	1

[a]Inverse ranking based on population and energy production system; [b]0 = Unstable, [b]1= Stable; [c]0= Political opposition not permitted, [c]1 = Restricted opposition permitted, but no campaigns for control of government, [c]2 = Unrestricted; [d]1 = Intense foreign conflict; [d]0 = little, if any foreign conflict; [e]Percentage of votes in agreement – percentage in opposition; [f]0 = Not subscribed to statue of international court of justice, [f]1= subscribed with or without reservation; Cs= Countries

Table 5. *Correlation Matrix.*

Characteristics	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
GNP per capital	0.97									
Trade	0.93	0.97								
Power	0.55	0.66	0.89							
Stability	0.62	0.55	0.25	0.63						
Freedom of group opposition	0.31	0.40	-0.10	0.32	.91					
Foreign conflict	0.36	0.30	0.25	0.46	-0.32	0.61				
Agreement with US in UN	0.58	0.59	-0.07	0.36	.75	0.11	0.89			
Defence budget	0.79	0.71	0.66	0.49	-0.7	-0.38	-0.18	0.90		
Percentage of GNP for Defence	0.17	0.17	0.06	0.15	-0.28	-0.44	-0.11	0.47	0.73	
Acceptance of International Law	0.34	0.22	-0.2	0.56	0.57	-0.04	-0.24	0.14	-0.24	0.82

Table 6. *Dimensionality of Common Factor Solution for Principal Factor Method.*

Characteristics	f_1	f_2	f_3	f_4	h_i^2	ϕ_i^2
GNP per capital	0.95	-0.04	-0.07	0.05	0.92	0.08
Trade	0.94	-0.03	-0.26	0.01	0.95	0.05
Power	0.59	-0.47	-0.33	-0.53	0.96	0.04
Stability	0.71	0.07	0.44	0.02	0.70	0.30
Freedom of group opposition	0.39	0.81	-0.07	-0.01	0.82	0.18
Foreign conflict	0.37	-0.48	0.38	0.18	0.54	0.46
Agreement with US in UN	0.57	0.62	0.29	0.39	0.95	0.05
Defence budget	0.16	-0.43	-0.02	0.02	0.77	0.23
Percentage of GNP for Defence	0.20	0.50	0.15	0.41	0.49	0.51
Acceptance of International Law	0.42	0.51	0.56	-0.35	0.88	0.12
Total Variance	40.7	22.1	9.4	7.6	79.8	20.2
Common Variance	51	27.7	11.8	9.5		
Eigenvalue	4.07	2.21	0.94	0.76		

Table 7. *Dimensionality of Common Factor Solution for Maximum Likelihood Method.*

Characteristics	f_1	f_2	f_3	h_i^2	ϕ_i^2
GNP per capital	0.75	0.56	0.24	0.92	0.08
Trade	0.84	0.54	-0.01	0.95	0.01
Power	0.32	0.69	-0.06	0.96	0.40
Stability	0.52	0.20	0.58	0.65	0.35
Freedom of group opposition	0.83	-0.55	0.00	0.99	0.01
Foreign conflict	-0.01	0.57	0.36	0.45	0.55
Agreement with US in UN	0.80	-0.15	-0.09	0.66	0.34
Defence budget	0.39	0.71	0.27	0.73	0.27
Percentage of GNP for Defence	-0.06	0.42	0.07	0.18	0.82
Acceptance of International Law	0.47	0.31	0.67	0.77	0.23
Total Variance	33.5	25.5	10.6	69.5	30.5
Common Variance	48.2	36.6	15.2		
Eigenvalue	3.35	2.55	1.06		

Table 4 provides data on the 10 characteristics for 14 selected countries with the correlation matrix in Table 5, showing that 89% of the variation in power can be accounted for by the data on the remaining variables.

Table 6 provides four dimensional common factor solution for principal factor method. The table shows that 80% of the variance, among all the variables, is accounted for by the four common factors. The four dimensional common factor solution also accounts for a substantial amount of the variance in GNP per capital, trade, power, acceptance of International and freedom of opposition, but has not done well for the variables involving percentage of GNP for defence and foreign conflict.

Table 7 presents a three dimensional common factor solution for maximum likelihood method as a result of convergence failure test of a four dimensional common factor solution on Table 1. The factor loadings, communalities, and unique variances obtained are quite different from the results of principal factor method.

4. Conclusion

This study has shown that factor analysis serves as data reduction method used for investigating interdependences and distinguishing different types of variances. It has also been established that different extraction methods generate different factor solutions. The disparity between factor analysis solutions depends on data considerations such as availability of sample size, number of variables, magnitude of the communalities and variation among the variables in terms of their communalities. A comparison of the extraction methods indicated that the centroid method maximizes the loadings regardless of the signs, the principal factor method maximizes the variance accounted for and it is also scale invariant while the maximum likelihood method provides the best estimate of reproduced correlation matrix in a population with convergence to the best value.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] C. A. Mertler, R.V. Reinhart. Advanced and Multivariate Statistical Methods. Practical Application and Interpretation, New York: Routledge (2017).

[2] C.R. Kothari, G. Gaurav, Research Methodology, Methods and Techniques, 3rd ed., New Age International Publishers, New-Delhi (2013).

[3] F. Benaych-Georges, N. Raj-Rao, The Singular Values and Vectors of Low Rank Perturbations of Large Rectangular Random Matrices, Journal of Multivariate Analysis 111 (2012) 120–135.

[4] F. Walker, G. Welch, Demystifying Factor Analysis: How it Works and How to Use it, U.S.A., Xlibris Co-oporation (2010).

[5] I. Jolliffe, Principal Component Analysis, New York: Springer-Verlag (1986).

[6] L.S. Meyers, G. Gamst, A. J. Guarino, Applied Multivariate Research: Design and Interpretation, Los Angeles: SAGE. (2017).

[7] M.M. Wall, F. Wang, Generalized Common Spatial Factor Model, Biostatistics 4 (4) (2003) 569-582.

[8] N. Islam, M. Z. Mamun, Factors for Not Buying Life Insurance Policiesina Developing Country, A Case of Bangledesh, Journal of Business Administration, 1& 2: (2005) 31.

[9] P.J. Kpolorie, IBM SPSS Statistics Excellent Guide, U.S.A.: Amazon (2021). KDP.https://www.amazon.com/BM-SPSS-Statistics-Excellent-Guide-ebook/DP/B08JG4.

[10] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, New Jersey: Pearson Prentice Hall (2019).

[11] R. M. Warner, Applied Statistics: From Bi-variate through Multivariate Techniques, Los Angeles (2013).

[12] S. Noora, Factor Analysis as a Tool for Survey Analysis, American Journal of Applied Mathematics and Statistics (2021).

[13] T. M. Boron, Confirmatory Factor Analysis for Applied Research, New-York: Gullford Press (2015).

[14] R. D. William, G. Mathew, Multivariate Analysis, Methods and Applications, John Wiley &Sons Inc. (1984).