Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.15, No.1, 2025

Some Conditions That Affect Practical Application of Factor Analysis

Kwashiga S. Adjorlolo [1*]

1. Department of Mathematics, Presbyterian Senior High School, Bechem, Ghana.

* E-mail of the corresponding author: askwashiga@gmail.com

Abstract

The study examined some conditions that must be satisfied in order to perform Factor Analysis. The objective was to determine whether or not the accepted pre-requisite tests always prove that the dataset will produce practical factor solution. Some of the conditions examined are the Kaiser-Meyer-Olkin test of sampling adequacy and the Bartlett's test of sphericity. Two datasets were used in this study namely Sales Performance and Personality Types. Both datasets were subjected to the pre-requisite tests and the extraction of various factor solutions. Both datasets passed the pre-requisite tests. One of the datasets was found not to produce significant factor solution; the other produced a practical factor solution.

Keywords: factor analysis, Kaiser-Meyer-Olkin test, Bartlett's test

1. Introduction

Research into the area of dimension reduction techniques is important because most researchers and students use these techniques daily. Data reduction techniques are applied where the goal is to aggregate or amalgamate the information contained in large datasets into manageable smaller information. Data reduction techniques can include simple tabulation, aggregation (computing descriptive statistics) or more sophisticated techniques like Principal Component Analysis and Factor Analysis. Factor analysis is a statistical method to explain a large number of interrelated variables in terms of a potentially low number of unobserved variables. Factor analysis reduces the complexity and reveals the underlining structure of the data set. Factor analysis is over a century old. In psychology, the factor model dates back at least to Spearman (1904), who is sometimes credited with the invention of factor analysis. The technique is later also applied to social science, economics, finance and marketing, signal processing, bioinformatics etc. The latent factors discovered by factor analysis make the observed variables more understandable. Factor analysis has been used for several years for finding underlying factors within a large set of variables. There exist pre-determined tests that are run on the dataset to determine whether factor analysis will be possible or not. These tests or conditions include Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The main justification for conducting this study is to examine if datasets that pass the pre-requisite conditions can produce practical factor solutions.

2. Literature Review

2.1 Factor Analysis

Factor analysis is a collection of methods used to examine how underlying constructs influence the responses on the number of measured variables. The key concept of factor analysis is that multiple indicator variables have similar patterns of responses as they are all associated with a latent (i.e., not directly measured) variable (Nkansah, 2018). The essential purpose of FA is to describe, if possible, the covariance relationships amongst many variables in terms of few underlying, but unobservable random quantities called factors.

2.2. Assessing the Factor Suitability of the Data

2.2.1. Sample Size and Sample to Variable Ratio

In reviewing literature, a general opinion has emerged, suggesting that ratio criteria do not provide an accurate guide (Guadagnoli & Velicer, 1988; Hogarty et al., 2005; Osborne & Costello, 2004). Fabrigar et al. (1999) and MacCallum et al. (2001), further support that stable solutions can be reached with samples as low as 100 when three to four strong items (loadings of 0.70 or greater) make up a factor, suggesting that weaker relationships need a larger sample size. A strong solution, made up of stable factors, reduces the influence of the sample size; however,

a larger sample size decreases sampling error resulting in more stable solutions (Hogarty et al., 2005). Determination of sample size sufficiency is dependent upon the stability of the solution; therefore, the adequacy of a sample cannot be fully determined until the analysis has been conducted.

2.2.2. Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity

Burton and Mazerolle **(**2011) states that prior to the extraction of the factors, there are some tests which must be conducted to examine the adequacy of the sample and the suitability of data for factor analysis. Measures of sampling adequacy evaluate how strongly an item is correlated with other items in the correlation matrix (Burton and Mazerolle, 2011). The sampling adequacy can be assessed by examining the Kaiser-Meyer-Olkin (KMO) (Kaiser**,** 1970). The Kaiser-Meyer-Olkin Test of Sampling Adequacy (KMO) is a measure of the shared variance in the items. The following guideline is suggested for assessing the measure.

Table 1.1 **Interpretation Guidelines for the Kaiser-Meyer-Olkin MSA**

KMO Value	Degree of Common Variance
0.90 to 1.00	Marvelous
0.80 to 0.89	Meritorious
0.70 to 0.79	Middling
0.60 to 0.69	Mediocre
0.50 to 0.59	Miserable
0.00 to 0.49	Don't Factor

By the guideline in Table 1.1, it is generally expected that to have satisfactory results, the overall KMO measure should be 0.8 or higher (Nkansah, 2018). This rule of thumb appears to have been accepted widely, although a measure of above 0.6 is acceptable (Rencher, 2002). Bartlett's test of Sphericity (Bartlett 1950) provides a chi-square output that must be significant. The null hypothesis of Bartlett's test states that the observed correlation matrix is equal to the identity matrix, suggesting that the observed matrix is not factorable (Pett et al., 2003). It indicates that the matrix is not an identity matrix and accordingly it should be significant ($p<0.05$) for factor analysis to be suitable (Hair, Anderson et al. 1995; Tabachnick and Fidell 2001). In summary, if the KMO indicates sample adequacy and Bartlett's test of sphericity indicates that the item correlation matrix is not an identity matrix, then researchers can move forward with the factor analysis (Netemeyer, Bearden et al. 2003).

2.3. The Correlation Matrix

A correlation matrix should be used in the factor analysis process displaying the relationships between individual variables. Henson and Roberts (2006) pointed out that a correlation matrix is most popular among investigators. Tabachnick and Fidell (2007) recommended inspecting the correlation matrix (often termed Factorability of R) for correlation coefficients over 0.30. Hair et al. (1995) categorized these loadings using another rule of thumb as ±0.30=minimal, ±0.40=important, and ±.50=practically significant. If no correlations go beyond 0.30, then the researcher should reconsider whether factor analysis is the appropriate statistical method to utilize. In other words, a factorability of 0.3 indicates that the factors account for approximately 30% relationship within the data, or in a practical sense, it would indicate that a third of the variables share too much variance, and hence becomes impractical to determine if the variables are correlated with each other or the dependent variable (multicollinearity) (Williams, Brown et al. 2010)**.** If the correlation matrix is an identity matrix (there is no relationship among the items), factor analysis should not be applied.

2.4 Method of Factor Extraction

There are several ways to extract factors: principal components analysis (PCA), principal axis factoring (PAF), image factoring, maximum likelihood, alpha factoring, unweighted least squares, generalized least squares and canonical (Tabachnick and Fidell 2001; Thompson 2004; Costello and Osborne 2005). However, principal components analysis and principal axis factoring are used most commonly in studies (Tabachnick and Fidell 2001; Thompson 2004; Henson and Roberts 2006). The decision whether to use PCA and PAF is fiercely debated among analysts (Henson and Roberts 2006), although the practical differences between the two are often insignificant (Thompson 2004) and according to (Gorsuch 1983), when factors have high reliability or there are thirty or more factors, the is not significant differences. Thompson (2004) stated that the reason why PCA is mostly used is that

it is the default method in many statistical software. PCA is suggested to be used when no prior theoretical basis or model exists (Gorsuch 1983). Moreover, (Pett, Lackey et al. 2003) recommended using PCA in establishing preliminary solutions in EFA. According to (Costello and Osborne 2005), factor analysis is preferable to principal components analysis which is only a data reduction approach. If researcher have initially developed an instrument with several items and is interested in reducing the number of items, then the PCA is useful (Netemeyer, Bearden et al. 2003). It is computed without regard to any underlying structure caused by latent variables; components are calculated using all of the variance of the manifest variables, and all of that variance appears in the solution (Ford, MacCallum et al. 1986). When the factors are uncorrelated and communalities are moderate it can produce inflated values of variance accounted for by the components (Gorsuch 1997). On the other hand, (Fabrigar, Wegener et al. 1999) stated that if data are relatively normally distributed, maximum likelihood (ML) is the best choice because "it allows for the computation of a wide range of indexes of the goodness of fit of the model and permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals." Overall, according to (Costello and Osborne 2005), maximum likelihood or principal axis factoring will give researcher the best results, depending on if data are generally normally-distributed or significantly non-normal, respectively.

2.5. Factor Loadings

Simple structure is achieved when each factor is represented by several items that each load strongly on that factor only (Pett et al., 2003; Tabachnick and Fidell, 2001). Practically, "several items" is generally considered to be at least three to five items with strong loadings (Guadagnoli & Velicer, 1988). An item is considered to be a good identifier of the factor if the loading is 0.70 or higher and does not significantly cross load on another factor greater than 0.40. These guidelines vary slightly within literature. Tabachnick and Fidell (2001) suggest that the secondary loading (or cross-loading) should be not greater than **0**.32. Costello and Osborne (2005) suggest that a loading of 0.50 is enough to be considered "strong," while Guadagnoli and Velicer (1988) state that the loading should be 0.60 or greater. Generally, a communality (loading) of 0.70 or greater is ideal because that suggests that approximately 50% of the variance of that item is accounted for by the factor. . If an item is not significantly correlated to any of the factors (generally considered to be less than 0.30) and does not provide a conceptually vital dimension to the measure, the item should be removed. Additionally, a complex variable, or a variable that loads on more than one factor, should be removed if the cross-loading is greater than 0.40 (Schonrock-Adema et al., 2009). Once the weak items have been removed, the data should be factored again without the presence of that item for a more refined solution (Pett et al., 2003). Interpretation of the factor also requires that each factor be sufficiently identified. This means that a factor contains at least three to five items with significant loadings in order to be considered a stable and solid factor (Costello & Osborne, 2005). More importantly, the items and the factors should make sense conceptually.

3. Materials and Methods

The data and statistical techniques utilized in this study to meet the stated goals are the main topics of this section. In this section, some tests will be described into detail.

3.1. Data Description

Two datasets have been used to carry out the study. The following provides
descriptions of these datasets.

Dataset 1 (Performance of Sales Personnel): The data covers assessment of performance of sales personnel employees of a marketing company (Johnson & Wichern, 2007). The firm attempts to evaluate the quality of its sales staff and tries to find an examination, or series of tests, that may reveal the potential for good performance in sales. It has selected a random sample of 50 salespeople and has evaluated each on three measures of performance: growth of sales, profitability of sales, and new account sales. These measures have been converted to a scale, on which 100 indicates "average" performance. Each of the 50 individuals would take each of four tests, which purportedly measures creativity, mechanical reasoning, abstract reasoning, and mathematical ability, respectively.

Dataset 2 (Difference Personality types): The data covers a set of responses from a personality questionnaire obtained from 400 participants. There were 13 variables personality characteristics. These characteristics include talkative, finds fault, does a thorough job, depressed, original, reserved, helpful, careless, relaxed, starts quarrels, reliable, tense and ingenious.

3.2. Some Conditions Required Before Factor Analysis

Before factor analysis is done, some tests are run on the data to show that a practical factor solution can be achieved. Some of these conditions include Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

3.2.1. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy

The Kaiser-Meyer-Olkin measure of sampling adequacy is used to measure the appropriateness of factor analysis. The Kaiser-Meyer-Olkin Measure is a statistic that indicates the proportion of variance in the variables that might be caused by underlying factors. High values (close to 1.0) generally indicate that a factor analysis may be useful with your data. If the value is less than 0.50, the results of the factor analysis probably won't be very useful. The Kaiser-Meyer-Olkin measure of sampling adequacy is used to determine the suitability of factor analysis and whether the partial correlations among the different variables are small (Bisschoff & Kade, 2010). The Kaiser-Meyer- Olkin measure of sampling adequacy presents an index ranging from 0 to 1 of the amount of variance among the different variables where a value of 0 indicates that factor analysis is not suitable and a value of 1 indicates that factor analysis is suitable for the study (Field, 2000). KMO values smaller than 0.5 indicates that factor analysis is not suitable and a KMO value of 0.6 should be present before factor analysis can be considered. Values between 0.5 and 0.7 are considered average, values between 0.7 and 0.8 are good and values between 0.9 and 1.0 are excellent.

The specific form of the KMO measure is given by

$$KMO_J = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}}$$

where

- R= $[r_{ij}]$ is the correlation matrix
- U=$[u_{ij}]$ the partial covariance matrix

3.2.2. Bartlett's Test of Sphericity

The Bartlett's test of sphericity is used to test if the original variables are independent and have a common variance. **Bartlett's test for sphericity** compares the correlation matrix (a matrix of Pearson correlations) of the data to the identity matrix. In other words, it checks if there is a redundancy between variables that can be summarized with some factors. **Bartlett's test of sphericity** tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection (Field, 2000). Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with the data.

The formula for the chi-square value is:

$$\chi^2 = -\left((n-1) - \frac{2*p-5}{6}\right) * \log(|R|)$$

where

- n is the number of observations,
- p is the number of variables,
- R is the correlation matrix.

The chi square test is then performed on $\frac{p^2-p}{2}$ degrees of freedom.

In summary, the test statistic calculates the determinate of the matrix of the sums of products and cross-products (S) from which the intercorrelation matrix is derived. The determinant of the matrix S is converted to a chi-square statistic and tested for significance. The null hypothesis is that the intercorrelation matrix comes from a population in which the variables are noncollinear (i.e. an identity matrix) and that the non-zero correlations in the sample matrix are due to sampling error.

3.3. Factor Retention Methods

3.3.1. Eigenvalues

Let A be k*k square matrix and I be k*k identity matrix. The scalar $\lambda_1, \lambda_2, \dots, \lambda_k$ satisfying the polynomial equation $|A - \lambda I| = 0$ are called the eigenvalues of the matrix A. $|A - \lambda I| = 0$ is known as the characteristic equation. Also, if $AX = \lambda X$ where X is a k*1 vector, then X is known as the eigenvector of the matrix A. The most popular method for deciding on the retention of factors is Kaiser's eigenvalue greater than one criterion (Fabrigar et al., 1999). This method specifies all factors with eigenvalues greater than one are retained for interpretation. Some researchers argue this method oversimplifies the situation and also has a tendency to overestimate the number of factors to retain (Zwick and Velicer, 1986). According to Ledesma and Pedro, 2007, this method may lead to arbitrary decisions, for example, it does not make sense to retain a factor with an eigenvalue of 1.01 and discard a factor with an eigenvalue of 0.99.

3.3.2. Cattell's Scree Plot

A technique which overcomes some of the deficiencies inherent in Kaiser's approach is the Catell's scree test (Cattell and Vogelmann, 1977). This is a plot of the eigenvalues associated with each of the factors extracted, against each factor. The scree plot is used to determine the number of factors to retain in factor analysis. In a scree plot, it is desirable to find a sharp reduction in the size of the eigenvalues. When the eigenvalues drop dramatically in size, an additional factor would add relatively little to the information already extracted.

3.4. Factor Analysis

The Factor Analysis Model

The "traditional" or "classical" factor analysis model is defined by the equation

$$\boldsymbol{Z_J = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + \varepsilon_j} \qquad \boldsymbol{(j = 1, 2, \dots, n)}$$

In this model, a variable $\mathbf{Z_j}$ is described by a linear combination of common factors $\mathbf{(F_1, F_2, \dots, F_m)}$ and a unique factor ε_j. Coefficients or loadings for the common factors are represented with $a_{j1}, a_{j2}, \dots, a_{jm}$; the number of common factors (**m**) is normally smaller than the number of observed variables, **n** (Harman, 1976). When considering the value of a specific variable, $\boldsymbol{x_j}$, for a given individual, **i**, the factor model can be written as:

$$Z_{ji} = \sum_{p=1}^{m} a_{jp} F_{pi}$$

where

- f_{pi} is the common factor **p** for individual **i**;
- $a_{jp}F_p$ represents the contribution of the factor on the linear composite.
- The residual error is given by ε_j (Harman, 1976).

The factor analytic model provides estimates for the values of loadings on common factors (Harman, 1976).

3.5. Factor Extraction Methods

According to Merrifield (1974), dimensional options include the methods that social science researchers employ to extract factors from "person by task matrices" (Merrifield, 1974). The most commonly used methods include maximum likelihood, principal axis factors with prior estimates of communalities, and "iterative principal factors" (Fabrigar et al., 1999). The relative utility of each method is dependent on the researchers' intentions and the distributions of observed data (Fabrigar et al., 1999).

3.5.1. Maximum Likelihood Estimation

Assumptions

- Data are independently sampled from a multivariate normal distribution
- Mean = 0 and variance = 1

Suppose $X_i \sim iid\ N(\mu,\ LL^1 + \Psi)$ is a multivariate normal vector. The log-likelihood function for a sample of n observations has the form

$$LL(\mu, L, \Psi) = -\frac{nplog(2\Pi)}{2} + \frac{nlog(|\sum^{-1}|)}{2} - \frac{\sum_{i=1}^{n}(X_i - \mu)'\sum^{-1}(X_i - \mu)}{2}$$

where $\sum = LL' + \Psi$

Maximum likelihood factoring allows the researcher to test for statistical significance in terms of correlation among factors and the factor loadings, but this method for estimating factor models can yield distorted results when observed data are not multivariate normal (Costello & Osborne, 2005; Fabrigar et al., 1999). Based on the assumption that a specified number of factors exists in a population, maximum likelihood factor analysis yields estimate of factor loadings for a given sample size and number of observed variables (Harman, 1976). When the observed variables exhibit multivariate normality and the sample size is large, maximum likelihood strategies facilitate the calculation of confidence intervals for the estimated loadings (Chen, 2003).

3.5.2. Principal Component Method

Perhaps the most widely used method for determining a first set of loadings is the **principal component method**. This method seeks values of the loadings that bring the estimate of the total communality as close as possible to the total of the observed variances.

Let X_i be a vector of observations for the i^{th} subject.

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ . \\ . \\ . \\ x_{ip} \end{pmatrix}$$

S denotes the sample variance-covariance matrix and it is expressed as

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$$

There are *p* eigenvalues for this variance-covariance matrix as well as corresponding eigenvectors for this matrix.

Eigenvalues of S= $\widehat{\lambda_1}, \widehat{\lambda_2}, \dots, \widehat{\lambda_p}$

Eigenvectors of S= $\widehat{e_1}, \widehat{e_2}, \dots, \widehat{e_p}$

The variance-covariance matrix expressed in the form of the eigenvalues and eigenvectors

$$\sum = \sum_{i=1}^{p}\lambda_i e_i e_i' \cong \sum_{i=1}^{m}\lambda_i e_i e_i' = \left(\sqrt{\lambda_1 e_1}\ \sqrt{\lambda_2 e_2}\ \dots\ \sqrt{\lambda_m e_m}\right)\begin{pmatrix} \sqrt{\lambda_1 e_1'} \\ \sqrt{\lambda_2 e_2'} \\ \vdots \\ \sqrt{\lambda_m e_m'} \end{pmatrix} = LL'$$

The idea behind the principal component method is to approximate the expression above. Instead of summing from 1 to *p*, the summing is now done from 1 to *m*, ignoring the last *p-m* terms.

The factor loadings are estimated with $\widehat{\iota_{ij}} = \widehat{e_{ji}}\sqrt{\widehat{\lambda_j}}$. This forms the matrix L of factor loadings in the factor analysis. This is followed by the transpose of L.

Recall: $\sum = LL' + \Psi$

$\Psi = \sum - LL'$

Therefore, the specific variance, the diagonal elements of Ψ, are estimated with the expression below.

$$\widehat{\Psi_i} = S_i^2 - \sum_{i=1}^{m}\lambda_j\widehat{e_{ji}^2}$$

where

- $\widehat{\Psi_i}$ is the communality or unique variance
- S_i^2 is the sample variance
- $\sum_{i=1}^{m}\lambda_j\widehat{e_{ji}^2}$ is the sum of the squared factor loadings

The approach of the principal component method is to calculate the sample covariance matrix S from a sample of data and then find an estimator that can be used to factor s. The principal component method seeks values of the

loadings that bring the estimate of the total communality as close possible to the total of the observed variances.

4. Results and Discussion

4.1. Factor Analysis

The aims of the research include **examining some conditions for practical factor solution,** determining whether sample size affects the factor analysis process and to compare the values of the pre-requisite tests and their suitability in confirming practical factor solution.

4.1.1. Correlation Matrix

Table 3.1 Correlation Matrix on Sales Performance

	Sales growth	Sales profitability	New account sales	Creativity	Mechanical reasoning	Abstract reasoning
Sales profitability	0.926					
New account sales	0.884	0.843				
Creativity	0.572	0.542	0.700			
Mechanical reasoning	0.708	0.746	0.637	0.591		
Abstract reasoning	0.674	0.465	0.641	0.147	0.386	
Mathematics	0.927	0.944	0.853	0.413	0.575	0.566

From Table 3.1, ‘Mathematics’, ‘Sales growth’ and ‘Sales profitability’ are likely to share the same factor since they are highly correlated. Also, ‘New account sales’ has strong relationships with ‘Mathematics’, ‘Sales growth’ and ‘Sales profitability’ and positive moderate relationships with ‘Creativity’, ‘Mechanical reasoning’ and ‘Abstract reasoning’. The correlation matrix shows the strength of association between the variables to be analyzed.

4.1.2. KMO and Bartlett's Test

Table 3.2 KMO and Bartlett's Test for Sales Performance

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.616
Bartlett's Test of Sphericity	Approx. Chi-Square	499.661
	Df	21
	Sig.	0.000

Table 3.3 KMO and Bartlett’s Test for Personality

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.712
Bartlett's Test of Sphericity	Approx. Chi-Square	1132.561
	Df	78
	Sig.	0.000

Tables 3.2 and 3.3 show the Kaiser-Meyer-Olkin (KMO) measure and the Bartlett’s test of sphericity which are used to check the adequacy of the sample size and whether or not the correlation matrix is suitable for factor analysis. Both KMO values of 0.616 (for the sales performance data) and 0.710 (for the personality data) are greater than the minimum threshold of 0.50 required for factor analysis. The Bartlett’s test *p*-value of 0.000 for

both datasets that at least some of the variables are inter-correlated and therefore the data is suitable for factor analysis.

4.1.3. Extracted Communalities

Tables 3.4 shows the communalities of variables in the sales performance and personality studies.

Table 3.4 Communalities for Personality and Sales Performance

Indicators	Extraction	Indicators	Extraction
Talkative	0.764	Sales growth	0.995
finds fault	0.769	Sales profitability	0.987
does a thorough job	0.698	New account sales	0.998
Depressed	0.607	Creativity	0.999
Original	0.629	Mechanical reasoning	0.999
Reserved	0.706	Abstract reasoning	0.998
Helpful	0.562	Mathematics	0.993
Careless	0.518		
Relaxed	0.643		
starts quarrels	0.576		
Reliable	0.659		
Tense	0.681		
Ingenious	0.732		

In principal component method, all variables are assigned an initial variance (total communality) of 100%. The extracted communalities are at least 0.50 and above. This means that, at least 50% of the initial communality of each variable accounts for factors in the final factor solution.

4.1.4. Total Variance Explained

Table 3.5 displays the number of factors that can be derived from the sales performance variables.

Table 3.5 Total Variance explained by Extracted factors for Sales Performance

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.035	71.923	71.923	5.035	71.923	71.923	2.877	41.100	41.100
2	0.934	13.336	85.259	0.934	13.336	85.259	1.446	20.662	61.761
3	0.498	7.113	92.372	0.498	7.113	92.372	1.396	19.949	81.711
4	0.421	6.018	98.390	0.421	6.018	98.390	1.153	16.469	98.180
5	0.081	1.158	99.547	0.081	1.158	99.547	.096	1.368	99.547
6	0.020	0.291	99.838						
7	0.011	0.162	100.000						

When two factors are extracted, it accounts for 85.259% of the total variance of the factor analysis model. Three, four and five factors explain 92.372%, 98.390% and 99.547% respectively.

Table 3.6 displays the number of factors that can be derived from the variables in the personality dataset.

Table 3.6 Total Variance explained by Extracted Factors for Personality

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.09	23.759	23.759	3.089	23.759	23.759	2.158	16.604	16.604
2	1.77	13.641	37.400	1.773	13.641	37.400	1.962	15.093	31.696
3	1.37	10.563	47.963	1.373	10.563	47.963	1.675	12.884	44.580
4	1.25	9.606	57.569	1.249	9.606	57.569	1.423	10.943	55.523
5	1.06	8.154	65.723	1.060	8.154	65.723	1.326	10.199	65.723
6	0.747	5.748	71.471						
7	0.713	5.484	76.955						
8	0.626	4.818	81.773						
9	0.579	4.453	86.226						
10	0.519	3.991	90.217						
11	0.486	3.740	93.957						
12	0.401	3.083	97.040						
13	0.385	2.960	100.000						

4.2 Cattell's Scree Plot

The scree plots in Figures 3.1 and 3.2 show the factors or components that contribute to the total variance of the factor model.

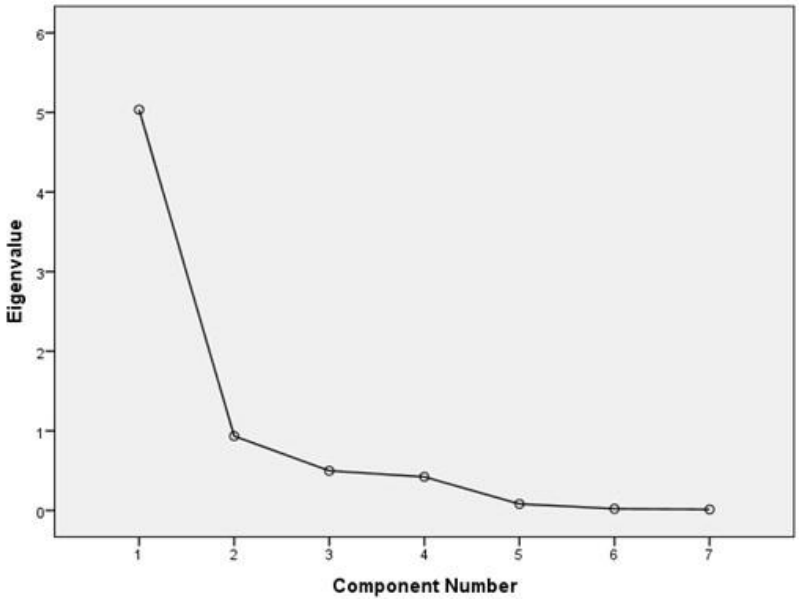

Fig 3.1 Scree plot of the sales performance dataset

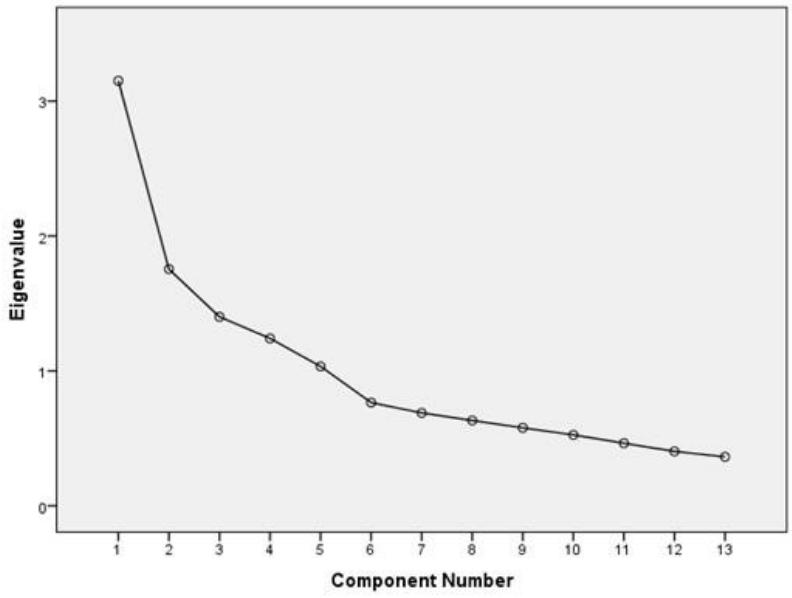

Figure 3.2 Scree plot of the personality dataset

4.3. Extraction and Rotation of Factors for the Sales Performance Data

Table 3.7 One-Factor Solution for Two Different Methods

	PC	ML
	1	1
Sales growth	0.973	0.975
Sales profitability	0.943	0.959
New account sales	0.945	0.902
Creativity	0.660	0.567
Mechanical reasoning	0.783	0.712
Abstract reasoning	0.649	0.615
Mathematics	0.914	0.953

From Table 3.7, all the variables loaded significantly on the Factor under the two different methods of extraction.

Table 3.8 Two-Factor Solution for Two Different Methods

	PC		ML	
	1	2	1	2
Sales growth	0.785	0.585	0.852	0.452
Sales profitability	0.670	0.664	0.868	0.419
New account sales	0.685	0.651	0.717	0.602
Creativity	0.043	0.923	0.147	0.989
Mechanical reasoning	0.379	0.743	0.502	0.523
Abstract reasoning	0.898	-0.012	0.620	0.057
Mathematics	0.801	0.482	0.946	0.277

From Table 3.8, three variables, namely “Sales growth”, “New account sales” and “Sales profitability” load significantly on both Factor 1 and Factor 2 under the Principal Component method. That is, they load above 0.50 on both factors.

Also, with the Maximum Likelihood method, there are significant loadings on both factors by the variables “New account sales” and “Mechanical reasoning”.

Table 3.9 Three-Factor Solution using Principal Component Method

	Component		
	1	2	3
Sales growth	0.779	0.387	0.452
Sales profitability	0.908	0.356	0.189
New account sales	0.616	0.548	0.484
Creativity	0.213	0.952	0.047
Mechanical reasoning	0.552	0.607	0.146
Abstract reasoning	0.286	0.060	0.950
Mathematics	0.909	0.181	0.328

From Table 3.9, "New account sales" and "Mechanical reasoning" load significantly on Component 1 and Component 2. Also, Component 3 is a single indicator component which defeats the idea of factor analysis.

Table 3.10 Three-Factor Solution using Maximum Likelihood Estimation

	Factor		
	1	2	3
Sales growth	0.794	0.374	0.437
Sales profitability	0.912	0.316	0.184
New account sales	0.652	0.544	0.437
Creativity	0.255	0.966	0.019
Mechanical reasoning	0.541	0.464	0.208
Abstract reasoning	0.300	0.054	0.952
Mathematics	0.918	0.179	0.296

In Table 3.10, "New account sales" loads significantly on Factor 1 and on Factor 2.

Table 3.11 Four-Factor Solution using Principal Component Method

	Component			
	1	2	3	4
Sales growth	0.765	0.316	0.417	0.322
Sales profitability	0.863	0.252	0.155	0.394
New account sales	0.660	0.548	0.437	0.175
Creativity	0.233	0.934	0.013	0.257
Mechanical reasoning	0.353	0.308	0.169	0.865
Abstract reasoning	0.297	0.030	0.944	0.132
Mathematics	0.927	0.159	0.280	0.172

From Table 3.11, "New account sales" loads significantly on Components 1 and 2. Factors 3 and 4 are also single indicator factors.

Attempt to determine the significance of the Four-factor solution does not yield convergent result. The error message indicates that it was not possible to extract four factors from the seven-item sales performance correlation matrix using maximum likelihood methods.

Table 3.12 provides the Chi-square estimates, the degrees of freedom and the *p*-values for the various factor solutions.

Table 3.12 Goodness-of-fit for various Factor Solutions for the Sales Performance Data

Model	Chi-Square	Df	*p*-value
1	162.715	14	0.000
2	117.114	8	0.000
3	61.651	3	0.000
4	-	-	-

These *p*-values are all significant ($p<0.05$). Thus, the one factor solution, two factor solution and three factor solution are all significant. The chi-square goodness of fit test **tests** the null hypothesis, which states that, compared with a one-factor model (i.e., all of the items load on a single factor), the fit of the data with the number of factors chosen (**m**) is adequate. If the test statistic is significant, that means m factors are not enough and thus should try m+1 extraction. Thus, in this test, we seek the value of m for which the null hypothesis is not rejected.

Interpretation of Factors Extracted from Sales Performance Data

In an attempt to extract various factor solutions, several indicators cross-load significantly on two factors. This results in solutions that are not interpretable. Hence, it is difficult to provide labels for the factors.

4.4. Extraction and Rotation of Factors for the Personality Data

Table 3.13 Two-Factor Solution for Principal Component Method

	Component	
	1	2
talkative	0.404	-0.332
finds fault	-0.129	0.315
does a thorough job	0.767	0.084
depressed	-0.190	0.706
original	0.512	-0.305
reserved	-0.058	0.483
helpful	0.651	-0.168
careless	-0.532	0.195
relaxed	0.012	-0.720
starts quarrels	-0.356	0.145
reliable	0.794	-0.059
Tense	0.049	0.786
ingenious	0.294	0.116

From Table 3.13, five variables, namely "does a thorough job", "original", "helpful", "careless" and "reliable" load significantly on Factor 1. Also, "tense", "relaxed", load significantly on Factor 2.

Table 3.14 Three-Factor Solution for Principal Component Method

	Component		
	1	2	3
talkative	0.267	-0.015	0.799
finds fault	-0.199	0.509	0.322
does a thorough job	0.769	0.068	0.052
depressed	-0.144	0.630	-0.342
original	0.454	-0.185	0.391
reserved	0.071	0.185	-0.745
helpful	0.637	-0.155	0.154
careless	-0.551	0.262	0.026
relaxed	0.010	-0.752	0.093
starts quarrels	-0.405	0.282	0.210
reliable	0.782	-0.048	0.145
tense	0.063	0.788	-0.162
ingenious	0.298	0.106	-0.007

From Table 3.14, the variables “talkative” and “reserved” load significantly on Factor 3. On Factor 2, “depressed” and “relaxed” have significant loadings.

Table 3.15 Four-Factor Solution for Principal Component Method

	Component			
	1	2	3	4
talkative	0.294	-0.012	0.829	-0.001
finds fault	-0.259	0.424	0.281	0.327
does a thorough job	0.748	0.089	0.016	0.184
depressed	-0.190	0.634	-0.310	0.026
original	0.409	-0.277	0.245	0.527
reserved	0.035	0.202	-0.767	0.041
helpful	0.654	-0.110	0.154	-0.002
careless	-0.568	0.222	0.042	0.005
relaxed	0.034	-0.796	-0.002	0.127
starts quarrels	-0.487	0.131	0.083	0.543
reliable	0.790	0.003	0.145	0.032
tense	0.016	0.808	-0.108	0.023
ingenious	0.198	-0.031	-0.205	0.743

In Table 3.15, “reliable”, “helpful”, “careless” and “does a thorough job” are significant on Factor 1. Also, the variables “depressed” and “relaxed” are significant on Factor 2. On Factor 3, “talkative” and “reserved” are significant. “Ingenious”, “original” and “starts quarrel” also load significantly on Factor 4.

Table 3.16 Five-Factor Solution for Principal Component Method

	Component				
	1	2	3	4	5
talkative	0.270	-0.049	0.833	0.067	-0.019
finds fault	0.086	0.189	0.077	0.851	-0.081
does a thorough job	0.805	0.037	0.014	0.010	0.170
depressed	-0.239	0.689	-0.257	0.001	0.069
original	0.298	-0.243	0.302	-0.035	0.605
reserved	0.158	0.169	-0.810	0.068	-0.021
helpful	0.416	0.021	0.309	-0.504	0.267
careless	-0.663	0.290	0.082	0.023	0.044
relaxed	0.010	-0.790	-0.019	-0.105	0.155
starts quarrels	-0.355	0.037	-0.017	0.578	0.322
reliable	0.751	0.006	0.196	-0.195	0.118
tense	0.019	0.817	-0.075	0.100	0.015
ingenious	0.061	0.045	-0.124	0.003	0.842

In the Five-factor solution, “does a thorough job”, “careless” and “reliable” load significantly on Factor 1. “Depressed” and “relaxed” also load significantly on Factor 2. Factor 3 is indicated by “talkative” and “reserved”. On Factor 5, the indicators are “original” and “ingenious”.

Table 3.17 Six-Factor Solution for Principal Component Method

	Component					
	1	2	3	4	5	6
talkative	0.248	-0.049	0.833	-0.004	0.108	-0.089
finds fault	0.031	0.176	0.085	0.015	0.872	0.168
does a thorough job	0.795	0.037	0.013	0.170	0.059	-0.117
depressed	-0.174	0.703	-0.259	-0.009	-0.118	0.232
original	0.355	-0.240	0.311	0.530	-0.138	0.173
reserved	0.134	0.162	-0.810	0.021	0.110	-0.050
helpful	0.484	0.037	0.305	0.150	-0.561	-0.042
careless	-0.732	0.265	0.086	0.168	0.094	-0.105
relaxed	0.024	-0.792	-0.014	0.131	-0.125	0.019
starts quarrels	-0.160	0.079	-0.007	0.069	0.182	0.899
reliable	0.704	-0.002	0.193	0.165	-0.069	-0.315
Tense	-0.005	0.811	-0.077	0.061	0.130	-0.036
ingenious	0.033	0.015	-0.108	0.917	0.009	-0.002

From Table 3.17, the 6th Factor is a One-Indicator Factor. Two variables namely, "does a thorough job" and "reliable" loads significantly on Factor 1. "Finds fault" and "helpful" also are significant on Factor 5.

Table 3.18 Goodness-of-fit for various Factor Solutions for the Personality Data

Model	Chi-Square	df	*p*-value
1	532.314	65	0.000
2	285.119	53	0.000
3	144.205	42	0.000
4	88.430	32	0.000
5	38.671	23	0.022
6	22.540	15	0.094

The *p*-values are all significant ($p<0.05$) for One-Factor solution through to Five-Factor solution. The *p*-value for the Six-Factor solution is not significant ($p>0.05$) and as such the null hypothesis is not rejected. This means that a Six-Factor solution is theoretically suitable for the data.

4.5. Interpretation of Factors Extracted from Personality Data

The higher the absolute value of a loading of an indicator on a factor, the more influential the variable is on the factor. However, a cut-off value of 0.50 and above is used to ensure that only variables of practical significance are included in the final factor solution. The factors are therefore labelled based on the loadings of the variables shown in the above Table 3.17 so that the higher the absolute value of a variable's loading on a factor, the more influential the variable is in naming the factor. The factors are labelled as follows:

- Factor 1: Conscientiousness
- Factor 2: Neuroticism
- Factor 3: Extraversion
- Factor 4: Openness to experience
- Factor 5: Agreeableness

5. Conclusion

This study examined some of the conditions that are required for practical factor solution in factor analysis. Some of the conditions studied include Bartlett's test of sphericity and Kaiser-Meyer-Olkin measure of sampling adequacy. Two different datasets were used which are referred to in this study as the Sales Performance dataset and personality dataset. The Sales Performance dataset contained 7 variables and was collected from 50 respondents. The Personality dataset measured 13 variables and was collected from 400 participants. Using the concept of exploratory factor analysis, the study variables in both datasets were subjected to statistical testing. The values for the Kaiser-Meyer-Olkin measure for both datasets were appropriate for factor analysis based on reviewed literature. Also, the Bartlett's test for sphericity was also significant for both datasets.

Although both pre-tests for the sales performance data were appropriate, practical factor solution could not be achieved after the extraction and rotation of two, three and four factors.

The personality dataset also passed both the Kaiser-Meyer-Olkin measure and the Bartlett's test of sphericity. Two, three, four, five and six factors were extracted and rotated. This six-factor solution was theoretically and practically correct. The study shows that the KMO and the Bartlett's test of sphericity may not be golden rules for determining suitability of data for factor analysis. The significance of larger sample size for practical factor solution is consistent with findings in the literature. Such relevance can also be attributed to the number of variables in the data.

Following the outcome of the study, the following recommendations are made for consideration to add to already literature in an attempt to reduce the controversy surrounding when factor analysis must be used on a multivariate data.

1. The necessary size may depend on the complexity of the model e.g. number of factors. In any case and as has been established in this study, a large sample size is preferable to a small sample size.

2. Further research should be conducted using more datasets with varying sample sizes and number of variables to address the different ideologies concerning when factor analysis must be conducted and the proper preliminary conditions that must be satisfied before data can be declared fit for factor.

References

Bartlett, M. S. (1950). "Tests of significance in factor analysis. ." British Journal of Psychology 3(2): 77-85.

Burton L. J., & Mazerolle S. M. (2011). "Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research." *Athletic Training Education Journal*, 6(1).

Cattell R. B., & Vogelmann S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. Multivariate Behavioral Research, 12(3).

Chen, R. (2003). A SAS/IML procedure for maximum likelihood factor analysis. Behavioral Research Methods, 35(3).

Comrey, A. L. & Lee, H. B. (1992). A first course in factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Costello, A. B., & Osborne, J. W. (2005). Recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation, 10 (7).

Fabrigar, L. R., MacCallum, R. C., Wegener, D. T., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods (3). 272-299.

Field, A. (2000). Discovering Statistics using SPSS for Windows. London – Thousand Oaks – New Delhi: Sage publications.

Ford, J. K., R. C. MacCallum, et al. (1986). "The Application of Exploratory Factor Analysis in Applied-Psychology - a Critical Review and Analysis." *Personnel Psychology*.

Gorsuch, R. (1983). *Factor Analysis*., Hillsdale, NJ: Erlbaum.

Guadagnoli, E. & Velicer, W. F. (1988). *Relation of sample size to the stability of component patterns*. Psychological Bulletin 103(2).

Harman, H. H. (1976). Modern factor analysis (3rd ed.). Chicago: University of Chicago Press.

Hair J., Anderson R. E., Tatham R. L., Black W. C. *Multivariate data analysis*. 4th ed. New Jersey: Prentice-Hall Inc; 1995.

Hayton, J. C., D. G. Allen, et al. (2004). *Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis.* Organizational Research Methods 7:191-205.

Henson, R. K. & Roberts J. K. (2006). *Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice.* Educational and Psychological Measurement 66(3).

Hogarty K, Hines C, Kromrey J, Ferron J, Mumford K. (2005). *The Quality of Factor Solutions in Exploratory Factor Analysis: The Influence of Sample Size, Communality, and Overdetermination*. Educational and Psychological Measurement. 65(2).

Hurley, A. E., T. A. Scandura, et al. (2004). *Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives.* Journal of Organizational Behavior 18: 667-683.

Johnson, R.A., & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). NJ: Prentice-Hall Int., Inc.

Kaiser H. F. (1970). A Second-Generation Little Jiffy. Psychometrika 35(4):401-15.

Ledesma, R. D., & Valero-Mora P. (2007). *Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis.* Practical Assessment, Research & Evaluation 12(2).

MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychological Methods*. 1999;4(1):84-99.

Merrifield, P. R. (1974). Factor analysis in educational research. Review of research in education, 2, 393-434.

Netemeyer, R. G., W. O. Bearden, et al. (2003). *Scaling Procedures: Issues and Applications.* London, Sage.

Nkansah, B.K. (2018). On the Kaiser-Meier-Olkin's Measure of Sampling Adequacy: *Mathematical Theory and Modelling*, 8(7), ISSN 2225-0522.

Pett M. A., Lackey N. R., Sullivan J. J. (2003). *Making Sense of Factor Analysis: The use of factor analysis for instrument development in health care research.* California: Sage Publications Inc.

Rencher, A.C. (2002). Methods of Multivariate Analysis: (2nd Ed.). New Jersey: John Wiley & Sons.

Sapnas, K. G. & Zeller, R. A. (2002). *Minimizing sample size when using exploratory factor analysis for measurement.* Journal of Nursing Measurement. 10(2): 135-153.

Schonrock-Adema, J., Heijne-Penninga, M., van Hell, E. A., & Cohen-Schotanus, J. (2009). *Necessary steps in factor analysis: Enhancing validation studies of educational instruments.* 31(6).

https://doi.org/10.1080/01421590802516756

Spearman, C. (1904). *General intelligence, objectively determined and measured*. American Journal of Psychology 15: 201-293.

Tabachnick, B. G. & L. S. Fidell (2001). *Using multivariate statistics* (3rd ed.). Needham Heights, MA, Allyn & Bacon.

Tabachnick, B.G., & Fidell, L.S. (2007*).* Using multivariate statistics (5th ed.). Boston, Pearson.

Thompson B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.

Velicer, W. F., & Jackson, D. N. (1990). *Component Analysis Versus Common Factor Analysis - Some Further Observations.* Multivariate Behavioral Research, 25(1): 97-114.

Williams, B., Brown,T., et al. (2010). *Exploratory factor analysis: A five-step guide for novices.* Australasian Journal of Paramedicine 8(3).

Zwick, W. R. and W. F. Velicer (1986). *Comparison of five rules for determining the number of components to retain.* Psychological Bulletin 99: 432-442.