An Exploratory Study of the Three Phases Analysis of Factor

Chinedu K. O[1]   George A. Osuji[2]   Chrisogonus K. Onyekwere[3]   Chukwuemeka B. Iwe[2]
1.Department of Statistics, Nnamdi Azikiwe University Awka, Nigeria
2.Meryland W.DC

Abstract
In this study, we examine factor analysis as a multivariate statistical tool, starting from the origin of factor analysis with regards to Spearman's approach of 1904 to the three phases of factor analysis. This is done with a view of determining the similarities and individual contributions of each of the three phases of factor analysis. This was achieved by examining the algorithms used in parameter estimations of the three phases of factor analysis. By inputting data into the algorithms and examining their outcomes and proffering recommendations based on the respective findings.

1 Introduction

Factor analysis is a statistical technique used to describe the inconsistency among observed, correlated variables in relation to lower number of unobserved variables called factors. It quests for combined variations in reaction to unobserved latent variables. Onyeagu (2003) describes factor analysis as a multivariate statistical technique which aims to describe, if possible, the covariance relationship among variables in terms of a few underlying but unobservable random quantities called factors. In other words it can be described as a multivariate statistical method which expresses "p" observed variables "z" in terms of "q" latent variables "f". Spearman (1904) published a capital idea. His celebrated article was titled "General Intelligence, Objectivity, Determined and Measured." This marked the beginning of the quantitative investigation of latent variables. He thought originally that all inter correlation among mental tests could be explained by assuming one general factor, along with a unique factor in each test. This he called the theory of two factors. Later this was modified to include group factors Spearman (1927). Spearman's theory was the conceptualization of the nature of a common factor – the element in common to two or more indicators (preferably three or more). Again, the theory highlighted the presence of two classes of factors; General (with one member) and Specific (with a potentially infinite number). It also stressed on the evaluation of empirical evidence on the tetrad difference criterion (i.e. on patterns in correlations among manifest variables) with no consideration of diagonal elements.

$$r_{ij.G} = \frac{r_{ij} - r_{iG}r_{jG}}{\sqrt{(1 - r_{iG}^2)(1 - r_{jG}^2)}} \tag{1}$$

where "G" is the general factor and "i" and "j" are the specific variables.

This theory was later developed extensively by Thurstone (1935, 1947). This is the most familiar multivariate procedure used in the behavioural sciences. This field of study was initially hampered by lack of adequate statistical Malenie (2004) defined factor analysis as a statistical model which expresses p observed variables z in terms of q unobservable latent variables f, where $q < p$. As time went on, another statistical model was developed for factor analysis. This time in the form of a linear model given in (Onyeagu 2003) as

$$X_i - \mu_i = L_{ij}F_j \dots + L_{ik}F_k + \in_i \tag{2}$$

where $i = 1, 2, \dots p$ and $j = 1, 2, \dots, k$ where $k < p$

where the p is a set observable random variables ,$X_1, X_2, \dots, X_p$ with means $\mu_1, \mu_2, \dots, \mu_p$, and some unknown constants L_{ij} with k unobserved random variables F_j where the $\in_i {}'s$ are independently distributed error terms with zero mean and infinite variance which may not be the same for all i,thus we let

$$var(\in_i) = \Psi, and\ E(\in_i) = 0 \tag{3}$$

In matrix terms, we have

$$X_{(px1)} - U = L_{(pxk)}F_{(kx1)} + \in_{(px1)} \tag{4}$$

If we have n observations, then we will have the dimensions $X_{p \times n}, L_{p \times k}$ and $F_{k \times n}$. Each column of X and F denote the values for one particular observations, and matrix L does not vary across observations. Also we note the following assumptions on F

    I.   F and $\in$ are independent
    II. $E(F) = 0$
    III. $cov(F) = E(FF') = I$ (to make sure that factors are uncorrelated)

Any solution of the chosen set of equations following the constraints for F is defined as the factor, and L as the loading matrix.

Factor analysis was later to be developed through the theory of latent structure analysis as proposed by Lazarfeld (1963) which gave the model as $p(t) = \int p(t|x)p(x)dx$. This was described by Bishop(2009) as one of the simplest forms of latent variable models and is based on the mapping $y(x; w)$, so that $t = Wx + \mu + u$, in which W and μ are adaptive parameters . The distribution p(x) is chosen to be a zero-mean unit covariance Gaussian distribution N(0,I), while the error model for u is also a zero mean Gaussian with a covariance matrix Ψ which is the diagonal element of the matrix. Silvia Bianconcini (2012) studied the nature of curiosity by analysing the agreements of junior high school students with large battery of statements ''I like to figure out how machinery works '',or ''I like to try new kinds of food'.' A factor analysis identified seven factors, three measuring enjoyment of problem –solving, learning and reading and computational technology. Rahn (2019) in his view stated that factor analysis is a useful tool for investigating variable relationships for complex concepts such as socio-economic status dietary patterns, or psychological scales. He enthused that it allows researchers to investigate concepts that are not easily measured directly by collapsing a large number of variables into a few interpretable underlying factors. Here the key concept of factor analysis is that multiple observed variables have similar patterns of responses because they are all associated with a latent variable.

In the study of factor analysis, it is a common knowledge that the correlation model has been of prominent use as well as the co-variance structure used in the linear model; all these were before the advent of the latent variable model discovered by Lazarsfeld (1968) which formed the crux of the General Linear Latent Variable Model (GLLVM) developed by Bartholomew and Knott (1999). All these make up the three phases of factor analysis and they all try to answer the questions posed by factor analysis which is data reduction, hypothesis testing, generating factor scores etc. The three models use different approaches to answer the basic problem posed by factor analysis.

However, there is a need to study these phases in order to ascertain the most suitable and convenient model fit for any type of data generated from sample survey used in factor analysis. Hence, the main purpose of the research is to determine the similarities and differences that exist in the three phases of factor analysis with a view to discovering a more comprehensive approach to the questions that factor analysis tends to answer. The research also examined the algorithms involved in the development of factor analysis as a multivariate statistical technique. This paper is organised as follows. Section 2 deals with the challenges of factor analysis, Section 3 discussed the methodology, section 4 deals with data analysis and results, and finally, section 5 contains summary and conclusions.

2 Challenges of Factor Analysis

The interpretation of factor analysis is based on empirical approach which is a solution that is convenient even if not completely true. More than one interpretation can be made of the same data factored the same way, and factor analysis cannot identify causality. Factor analysis is not without cost, however. It is mathematically complicated and entails diverse and numerous considerations in their formal training, and the sum is the major cost of factor analysis: most laymen, social scientists, and policy-makers find the nature and significance of the results incomprehensible for their work application. Its technical vocabulary includes strange terms such as eigenvalues, rotate, sample structure, orthogonal, loadings, and communality. Its results usually absorb a dozen or so pages in a given report, leaving little room for a methodological introduction or explanation of terms.

Philippe (2013) while working on (estimation of generalized linear latent models) stated that the general linear latent variable model (GLLVM) which is a statistical analysis presents a difficulty, since the latent variables are not observed , they must be integrated out from the likelihood function .and the calculations involved in this method is highly enormous. Add to this the fact that students do not ordinarily learn factor analysis as a course while in school.

3 The Three Phases of Factor Analysis
3.1 First phase of factor analysis
According to Yule (1927) the essence of what Spearman needed is contained in the formula for the partial correlation between two variables ,i and j say, given a third variable which Spearman calls G. This he gave as

$$r_{ij.G} = \frac{r_{ij} - r_{iG}r_{jG}}{\sqrt{(1-r^2_{iG})(1-r^2_{jG})}}, \; (i,j = 1,2,...p) \tag{5}$$

where $r_{ij.G}$ is the partial correlation between i and j

This is called the Spearman's approach. The study employed the principle of partial correlation given by Onyeagu (2003). Here partial correlation between $\underline{X}_1$ and $\underline{X}_2$ given by the matrix $\sum_{11} - \sum_{12}\sum_{22}^{-1}\sum_{21}$ which is also known as the matrix of partial variances and co variances. This can be demonstrated by letting $\delta_{ij.....q-1...p}$ denote the ij^{th} element in $\sum_{11.2}$, the matrix defined by the elements

$$\rho_{ij...q+1...p} = \frac{\delta_{ij...q+1...p}}{\sqrt{\delta_{ii...q+1...p}}\sqrt{\delta_{jj...q+1...p}}} \tag{6}$$

is the matrix of partial correlations where $\rho_{ij\ldots q+1.q+2\ldots p}$ is the partial correlation between X_i and X_j in the first set of X_1, holding variables in the second set $\underline{X_2}$ constant .The partial correlation coefficient allows us to measure the linear dependence of any two variables in the set by removing the linear association of the variables in the second set with the variables in the first set. This can be shown in the example below:

Suppose $X_1 = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $X_2 = X_3$ where $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$

The partial correlation between $\underline{X_1}$ and $\underline{X_2}$ can be written as

$$\rho_{12.3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} \tag{7}$$

Thus, working in a recursive manner any desired partial correlation can be obtained using Anderson (1958) formula

$$\rho_{ij.q+1,q+2\ldots p} = \frac{\rho_{ij.q+1\ldots p} - \rho_{ip.q+1\ldots p-1}\rho_{jp.q+1\ldots p-1}}{\sqrt{1-\rho_{ip.q+1\ldots p-1}^2}\sqrt{1-\rho_{jp.q+1\ldots p-1}^2}} \tag{8}$$

Consequently, we can write

$$r_{ij.G} = \frac{r_{ij} - r_{iG}r_{jG}}{\sqrt{(1-r_{iG}^2)(1-r_{jG}^2)}} \qquad (i,j = 1,2,\ldots p) \tag{9}$$

From this, we derive the correlation between r and j based on the assumption that their correlation can be totally explained by their common dependence on G implying that $r_{ij.G} = 0$, thus, the correlation matrix

$$R = \{r_{ij}\} = r_{iG}r_{jG} \tag{10}$$

If there are two or more (independent) underlying factors the correlation matrix would have the form

$$r_{ij} = \sum_{h=1}^{q} \lambda_{ih}\lambda_{jh} \quad (i \neq j) \tag{11}$$

The decision to use product moment correlation therefore, implies assumptions that item test of scores are linearly related to any underlying factor. Incidentally this became the central idea in the second phase of factor analysis. The major limitation of this phase was that, it's rudimentary in the sense that it was not capable of treating more than one factor at a time. Again it does not go beyond the level of partial correlation; so that all one needs to study was the pattern of the correlation matrix in order to get a picture of the nature of relationship between the variables involved. This method of dealing with factor analysis at the early stage of the development of factor analysis was hampered by the lack of adequate computational techniques and computer packages.

3.2 Second phase of factor analysis
The second phase of factor analysis which in modern notation according to Lawley and Maxwell (1963) supposed that

$$X_i = \lambda_{i1} y_1 + \lambda_{i2}y_2 + \cdots + \lambda_{pq}y_q + e_i; \quad \text{for } i = 1,2 \ldots p \tag{12}$$

which can be represented in terms of a probability distribution as in Bartholomew *et al* (1999) as $x_i|y_1 y_2 \ldots y_q \sim N(\mu_i + \sum_{i=1}^{q} \lambda_{ij}y_{i,}\Psi_i)$ given that $(i = 1,2 \ldots p)$, and $y_i \sim N(0,1)$. The covariance matrix $\sum$ has the form

$$\textstyle\sum = \lambda\lambda' + \Psi \tag{13}$$

Onyeagu (2003) describes the model this way

$$X_1 - \mu = L_{11}F_1 + L_{12}F_2 + \cdots . + L_{1m}F_m + \in_1$$
$$X_2 - \mu = L_{21}F_1 + L_{22}F_2 + \cdots . + L_{2m}F_m + \in_2$$
$$.. \qquad . .. \qquad \qquad ... \qquad \qquad \qquad \ldots$$
$$X_p - \mu = L_{p1}F_1 + L_{p2}F_2 + \cdots . + L_{pm}F_m + \in_p$$

In matrix form $\quad X_{(p\times 1)} - \mu = L_{(p\times m)}F_{(m\times 1)} + \in_{(p\times 1)}$

The coefficient L_{ij} is called the loading of the jth factor .The matrix L is the matrix of loadings, the ith specific factor $\in_i$ is associated only with the ith response x_i the p deviations $x_1 - \mu_1, x_2 - \mu_2$ are expressed in terms of $p + m$ random variables $F_1, F_2, \ldots, F_m, \in_1, \in_2, \ldots \in_p$ the assumptions are

a) $E(F) = 0_{(m\times 1)}$

b) $cov(F) = E(FF') = I_{(m\times m)}$, $E(\in) = 0_{(p\times 1)}$

c) $cov(\in) = (\in\in') = \Psi_{(p\times p)} = \begin{pmatrix} \Psi_1 & 0 \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \ldots & \Psi_p \end{pmatrix}$

And that F and $\in$ are independent so that $cov(\in, F) = E(\in F') = 0_{(p\times m)}$ this constitutes the orthogonal factor model which has the covariance as $\sum = cov(X) = LL' + \Psi$

The covariance structure is of the form

$$cov(X) = LL' + \Psi \text{ or } var(x) = L_{i1}^2 + \cdots L_{im}^2 + \Psi$$
$$cov(x_i, x_k) = L_{i1}L_{k1} + \cdots + L_{1m}L_{km}$$
$$cov(XF) = L \text{ or } cov(X_iF_j) = L_{ij}$$

and $x - \mu = LF + \in$ is linear in the common factors. Thus

$$\delta = L_{i1}^2 + L_{i2}^2 + \cdots L_{im}^2 + \Psi \tag{14}$$
$$var(X_i) = communuality + specific\ variance$$
$$\delta_{i1} = h_i^2 + \Psi_i \quad i = 1,2 \ldots p$$

3.2.1 Estimation of parameters

In the study the researcher examines ways to establish the covariance relationship given a dataset of say, $X_1, X_2, \ldots X_n$ observations on p correlated variables, which tends to address the problem posed by factor analysis which is to determine whether the factor model $X = U + LF + \in$, with its assumptions can adequately provide us with estimates of the data. This implies verifying the covariance relationship.

$$cov(X) = LL' + \Psi \text{ And } cov(X,F) = L$$

Here the population covariance matrix $\sum$ is estimated by the sample covariance matrix S or $\hat{\Sigma}$. The study used the principal component method (principal factor) and the maximum likelihood method to achieve this.

In the principal factor method, we factor out the covariance matrix using spectral decomposition theorem by Onyeagu et al (2003) which states that

$$\sum = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots + \lambda_p e_p e_p'$$

$$= [\sqrt{\lambda_1}e_1 \quad \sqrt{\lambda_2}e_2 \quad \ldots \sqrt{\lambda_p}e_p] \begin{vmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \sqrt{\lambda_p}e_p' \end{vmatrix} \tag{17}$$

where (λ_i, e_i) is the eigenvalue-eigenvector pair of $\sum$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

$$\sum_{p \times p} = L_{p \times p} L'_{p \times p} + 0_{p \times p} = LL'$$

The factor analysis representation is exact but not very useful since it does not allow for any variation in the specific factors $\in$, to overcome this the contribution of $\lambda_{m+1}e_{m+1}e_{m+1}' + \cdots \lambda_p e_p e_p'$ to $\sum$ in the spectral decomposition of $\sum$ giving rise to

$$\Sigma = (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2 \ldots \sqrt{\lambda_m}e_m) \begin{vmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \sqrt{\lambda_m}e_m' \end{vmatrix}$$

$$= L_{p \times m} L'_{m \times p}$$

the approximation becomes

$$\sum = LL' + \Psi \tag{18}$$

$$\sum = [\sqrt{\lambda_1}e_1 \quad \sqrt{\lambda_2}e_2 \quad \ldots \sqrt{\lambda_p}e_p] \begin{vmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \sqrt{\lambda_p}e_p' \end{vmatrix} + \begin{bmatrix} \Psi_1 & 0 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \cdots & \Psi_p \end{bmatrix} \tag{19}$$

where $\Psi = \delta - \sum_{j=1}^m l_{ij}^2, \ (i = 1,2..p)$

In maximum likelihood method, Bartholomew and Knott (1999) state that if $X \sim N_p(\mu, \sum)$, where the common factors F and specific factors $\in$ are assumed to be normally distributed. The maximum likelihood estimates of the factor loadings and specific variances can be obtained. The likelihood function for observations $X_1, \ldots, X_n$ may be written as

$$L(\Sigma) = constant + \frac{n}{2}\{ln|\Sigma^{-1}| - trace(\Sigma^{-1}S)\} \tag{20}$$

where S$= \sum_{i=1}^n (x_i - \mu)(x_i - \mu)'/n$.

This is first minimized with respect to μ. This is a standard problem and it is readily shown that $\hat{\mu} = \tilde{x}$. This is substituted into S, and then maximized with respect to $\wedge$ and Ψ, where

$$\Sigma = \wedge \wedge' + \Psi \tag{21}$$

The $\wedge$ is then properly defined by imposing the uniqueness condition $\wedge' \Psi^{-1} \wedge = \Delta$, a diagonal matrix. $\hat{L}$ and $\hat{\Psi}$ are then obtained by maximizing the likelihood function .

$$= (2\pi)^{\frac{-np}{2}}|\sum|^{-n/2}e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\Sigma_{j=i}^n (X_j - \overline{X})(X_j - \overline{X})' + n(\overline{X} - \mu)(\overline{X} - \mu)'\right)\right]}$$

$$= (2\pi)^{\frac{-(n-i)p}{2}}|\sum|^{-(n-1)/2}e^{-\frac{1}{2}tr\left[\Sigma^{-1}(\Sigma_{j=1}^n (X_j - \overline{X})(X_j - \overline{X})')\right]} \times (2\pi)^{-p/2}|\sum|^{-1/2}e^{-(n/2)(\overline{X} - \mu)\Sigma^{-1}(\overline{X} - \mu)'} \tag{22}$$

This is evaluated with the use of the computer program, thus the maximum likelihood estimators $\hat{L}, \hat{\Psi}, and\ \hat{\mu} = \overline{X}$ maximize the likelihood function $L(\mu, \sum)$ subject to $\hat{L}'\ \hat{\Psi}'\hat{L}$ being a diagonal. $\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots \hat{l}_{im}^2$ for $i =$

$1,2, \ldots, p$ which is termed the maximum likelihood estimates of the communalities for the covariance matrix.

$$\{proportion \ or \ total \ sample \ variance \ due \ to \ j^{th} factor\} = \frac{l_{1j} + l_{2j} + \cdots l_{pj}^2}{S_{11} + S_{22} + \cdots S_{pp}} \tag{23}$$

In the second phase of factor analysis the model provided us with rigorous and detailed answers to questions posed by factor analysis as well as makes available to us factor scores, which are used for further research work. One major limitation of this phase is that, it assumed that the specific variables were all continuous and thus ignoring the fact that data obtained from research work in most behavioural sciences are discrete, nominal or a mixture of the two in nature; in this way a lot of useful information is lost as a result of the assumptions.

3.3 Third phase of factor analysis
In the third phase of factor analysis emphasis is laid most, on latent variables. Latent variable models provide an important tool for the analysis of multivariate data. They offer a conceptual framework within which many disparate methods can be unified and a base from which new methods can be developed. A statistical model specifies the joint distribution of a set of random variables and it becomes a latent variable model when some of these variables –the latent variables are unobservable .The model known as general linear latent variable model (GLLVM) given in Bartholomew and Knott et al (2011), the model consist of two parts; first the prior distribution is given by

$$f(x) = \int h(y) \prod_{i=1}^{p} g_i (x_i/y) dy \tag{24}$$

It is represented by the density function $h(y)$ and the second element in the model is a set of conditional distribution of the set of the manifest variables (X_i) given the latent variable y. These are denoted by $g_i(X_i/y)(i = 1.2..p)$ where the subscript i and g reminds one that the form of distribution can vary with. A convenient family of distributions which turns out to have many useful properties in other branches of statistics is the one parameter exponential family. Suppose that latent variables combine to determine the value of the parameter, then one may have

$$g(X_i/\theta_i) = F_i(x_i)G_i(\theta_i)\exp(\theta_i \mu_i(x_i)) \quad \text{for } i = 1,2,\ldots p \tag{25}$$

where θ_i,is some function of y , the simplest assumption about the form of this function is to suppose that it is a linear function ,in which case we obtain,

$$\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \cdots \alpha_{iq}y_q \quad (i = 1,2,\ldots p).$$

This is the general linear latent variable model, the term linear refers to its linearity in the $\alpha_i s$. The difference between this model and the generalised linear model commonly used in statistics is that in the general linear latent variable model will have a set of $X's$ rather than a single dependent variable and here the $y's$ are unobservable. Thus we shall be predicting the $y's$ given the $'s$.

The exponential family of equations above includes the normal, gamma, and Bernoulli distributions as special cases, if $\bar{X}_i$ and θ are allowed to be vector valued it will also include the multinomial distribution.

3.3.1 Maximum likelihood estimation
The log-likelihood of the latent factor model for the mixture distribution is of the form

$$L(\pi_i, \mu_i, \textstyle\sum_i) = \sum_{n=1}^{N} ln\{\sum_{i=1}^{M} \pi_i f(y/i)\} \tag{26}$$

Maximization of this log likelihood is made complex by the presence of the summation inside the logarithm. A technique for performing the optimization is known as the expectation maximization (EM) algorithm can be employed according to Michael *et al* (1999).

$$R_{ni} \equiv f(i/y_n) = \frac{\pi_i f(y_n/i)}{\sum_j \pi_j f(y_n/j)} \tag{27}$$

An introductory account of this EM in the content of the factor model is given in the equation

$$f(y,x) = f(x)f(y/x) = f(x) \prod_{i=1}^{p} f(y_i/x) \tag{28}$$

The EM algorithm is based on the observation that if we have a set of indicator variables z_{ni} specifying which component i is responsible for generating in each data point y_n , then the log likelihood would take the form

$$L_{comp}(\{\pi_i, \mu_i, \textstyle\sum_i\}) = \sum_{n=1}^{N} \sum_{i=1}^{M} z_{ni} \ln\{\pi_i f(y/i)\} \tag{29}$$

and it's optimization would be straight forward, with the result that each component is filled independently to the corresponding group of data points. And the mixing coefficients are given by the fractions of the points in each group. The $\{z_{ni}\}$ are regarded as 'missing data' and the data set $\{y_n\}$ is said to be "incomplete" combining $\{y_n\}$ and $\{z_{ni}\}$ the corresponding "complete " data set is obtained with a log likelihood given in eq. (27). Though the values of $\{z_{ni}\}$ are unknown, but are determined by their posterior distribution given by R_{ni} .

As can be seen from the model, the third phase of factor analysis has laid to rest the issue of the nature of the specific variable, by using the conditional distribution of $f(y/x)$ and combining it with the mixing (or prior) distribution $f(y)$. In this way there is need to know the distribution of y in order to predict the various specific variables. Hence the need to search for a more advanced model which can tell us more about the distribution of the specific variables without making use of the prior distribution of y.

Taking a step further from the model for the third phase of factor analysis which is of the form

$$f(x) = \int_y^\infty \prod_{i=1}^p f(x_i \,/\, y) \, f(y) dy$$

where $f(y)$ is the prior (or mixing) distribution and $f(x_i/y)$ is the conditional distribution of x given y. And using the principles of Bayes theorem which states that $P(A/B) = \frac{P(A)P(B)}{P(B)}$ and applying same to the factor model gives

$$f(x) = \int_y^\infty \prod_{i=1}^p \frac{f(x_i)f(y)}{f(y)} \, f(y) dy \qquad (30)$$

Simplifying further gives

$$f(x) = \int_y^\infty \prod_{i=1}^p f(x_i) \, f(y) dy \qquad (31)$$

and integrating the mixing distribution over the range of y, we obtain $f(x) = \prod_{i=i}^p f(x_i) = f(X)$. In this way the parameter estimates are obtained by fitting distribution on x_i and obtaining the distribution of the specific variables, using the computer program R-console.

4 Results and Analysis
In this section the parameter estimates obtained for the three phases of factor analysis were evaluated using different sample sizes. The computational analysis of the simulated data for the parameters of the three phases of factor analysis were performed starting with the first phase.

4.1 The first phase of factor analysis
For this phase of factor analysis, with sample size n =379, and x =4, from a Binomial distribution, the partial correlation matrices can be obtained manually via eq. (8). The partial correlation estimates give us an insight of the contribution of the various variables to the general factor G in the above equation.

Table 4.1 shows the results obtained using SPSS version 20 for the first phase of factor analysis

Table 4.1: Correlation matrix

	Item 1	factor 2	factor 3	factor 4
factor 1	1.000	.655	.587	.520
factor 2	.655	1.000	.686	.609
factor 3	.587	.686	1.000	.707
factor 4	.520	.609	.707	1.000

Here it can be seen from the correlation matrix that items 3 and 4 are highly correlated followed by factors 2 and 3. Thus all the information about the model is all that can be found in the correlation matrix.

Table 4.2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.885	72.114	72.114	2.885	72.114	72.114
2	.516	12.904	85.019			
3	.328	8.196	93.215			
4	.271	6.785	100.000			

Extraction Method: Principal Component Analysis.

Table 4.2 shows the total variation explained by each of the components. Component one explained about 72.1% of the variance, followed by component two (12.9%) and lastly, component 4 which explained approximately 6.8% of the variance. This is also confirmed by the scree plot in fig 1. Only factor one appears significant
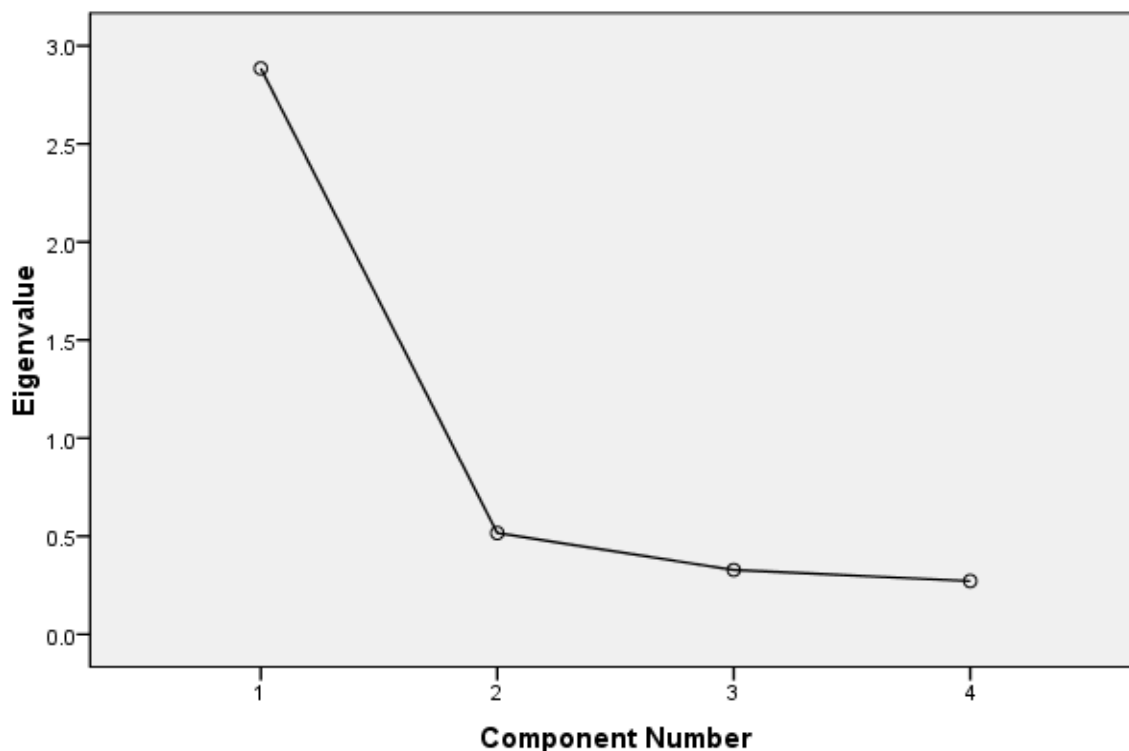
Scree Plot

Fig. 1: Scree Plot of the total variance explained

From the Scree plot diagram it is seen that only one variable is significant, which makes for the selection of one component only. The results obtained using the **third phase** is shown below

Table 4.3: Factor-Scores for observed response patterns

	Fact.1	Fact.2	Fact.3	Fact.4	Obs	Exp	Z 1	Se.Z1	Z 2	Se.Z2
1	0	0	0	0	1	0.810	-1.23	0.565	-.331	0.557
2	0	0	0	0	1	0.810	-.589	0.589	0.656	0.647
3	0	1	1	0	1	0.338	0.079	0.470	-1.35	0.567
4	1	0	1	1	1	0.451	0.736	0.625	1.121	0.510
5	1	1	0	1	1	0.401	1.053	0.453	-.357	0.468

Table 4.3 shows the individual distribution of factor scores, expected value of X, $E(X)$ and the standard error S.e of the posterior distribution of the data imputed. The analysis for the second data goes thus starting with the **first phase** of factor analysis which deals with the issue of partial correlation.

For the **first phase** of factor analysis these results were obtained.

Table 4.4 Correlation Matrix

		Fact.1	Fact. 2	Fact. 3	Fact. 4	Fact. 5	Fact. 6
Correlation	Fact.1	1.000	-.235	.072	-.086	.088	.082
	Fact. 2	-.235	1.000	.018	.101	.125	.020
	Fact. 3	.072	.018	1.000	.232	.348	.238
	Fact. 4	-.086	.101	.232	1.000	.218	.161
	Fact. 5	.088	.125	.348	.218	1.000	.264
	Fact. 6	.082	.020	.238	.161	.264	1.000

From the result obtained from the correlation matrix, it can be seen that the items are loosely correlated, the model does not provide us with much information concerning the number of factors in the model, since it is a one factor model thus it assumes that all of the variables (items) are grouped under one factor. But further analysis with the other models will prove otherwise.

Table 4.5: Communalities

	Initial	Extraction
Fact. 1	1.000	.737
Fact. 2	1.000	.568
Fact. 3	1.000	.314
Fact. 4	1.000	.392
Fact. 5	1.000	.238
Fact 6	1.000	.396

Table 4.5 shows the extracted communalities. From the table, it can be seen that the extracted communalities in the first and second factors indicate that most of the information contained in the model is contained in them. This is further explained in the table 4.6 below.

Table 4.6: Component analysis

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.767	29.453	29.453	1.767	29.453	29.453
2	1.278	21.306	50.759	1.278	21.306	50.759
3	.846	14.095	64.855			
4	.783	13.043	77.897			
5	.712	11.864	89.761			
6	.614	10.239	100.000			

Extraction Method: Principal Component Analysis.

From table 4.6, the first component has variance 29.45% while the second component has variance 21.31%. Fig. 2 shows the scree plot of the communalities.
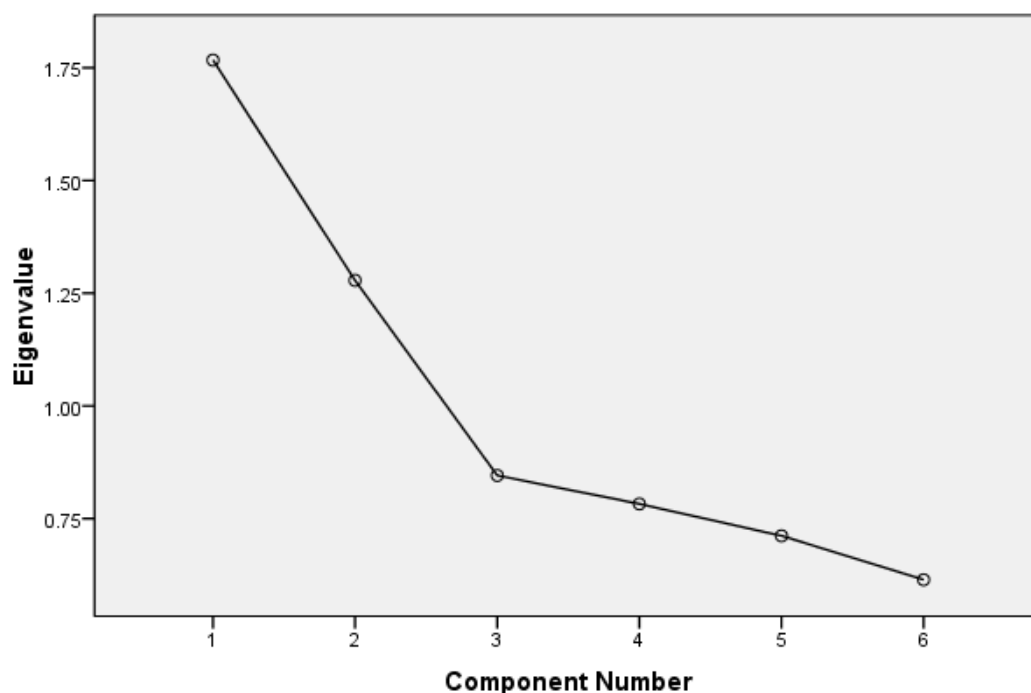

Fig 2: Scree plot of the Communalities extracted

Table   4.7:   Component Transformation Matrix

Component	1	2
1	.995	.101
2	.101	-.995

Extraction   Method: Principal Component Analysis.
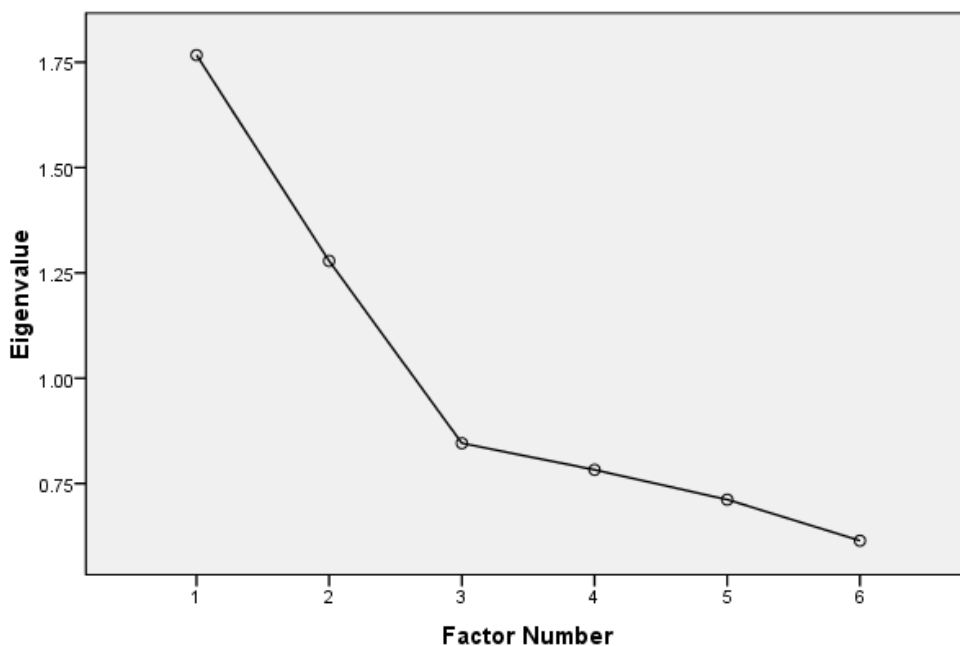
Fig 3:Scree Plot of the component transformation matrix

Table 4.8: Factor Matrix

	Factor	
	1	2
Item 1	.500	-.560
Item 2	-.075	.352
Item 3	.482	.287
Item 4	.228	.354
Item 5	.529	.330
Item 6	.382	.189

Extraction   Method: Maximum Likelihood.

In the factor matrix above we are able to obtain information about the two factors extracted from the model and their relationship with the other factors. With this, further analysis such as regression analysis can be carried out with the reduced data obtained from this factor analysis model.

Table 4.9.1: Goodness-of-fit Test

Chi-Square	df	Sig.
10.643	4	.031

Table   4.9.2:   Factor Transformation Matrix

Factor	1	2
1	.863	.506
2	.506	-.863

Table 4.9.3: summary $f(X)$

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1st Qu	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.3698	0.7962	0.2833	0.2525	0.3569	0.1461
3r Qu	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000
Max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Mean $f(X) = 0.1461233$

5 Summary of findings:

5.1 First phase of factor analysis

In the first phase, the model was able to derive the correlation matrices for the two data sets obtained by simulation using Minitab 16. Results as shown in table 4.1 and 4.2 led to the conclusion that the first phase of factor analysis is peripheral and lacks much content, because it starts and ends with the correlation matrix.

5.2 Second phase of factor analysis

This phase of factor analysis, which is a model based approach, goes beyond the provision of the correlation matrix to give a detailed explanation of the patterns in the matrix by providing an ANOVA table of the factor variances, table of communalities (this deals with the uniqueness of the error variances), and the SCREE plot diagram . Also the aspect of data reduction and provision of factor scores were adequately treated in this second phase of factor analysis. Thus this model approach may be considered most appropriate though it provides us with rigorous methods for answering the traditional questions addressed by factor analysis.

5.3 The third phase of factor analysis

The third phase of factor analysis was able to accommodate and treat adequately the various types of data (continuous, discrete and mixed) available. It also handled the issue of data reduction which is the main goal of latent variable analysis. In the third phase the posterior distribution is used to give a picture of the specific variables by providing estimates of the expectations, the standard error and the standardized values of the factor scores, using the logit /normal model.

In summary, the study explored the three phases of factor analysis by bringing to focus the synergy that exists among the three phases as can be seen from their individual correlation matrices. Furthermore, phases two and three of factor analysis share so much in common in their model attributes, such as, the uniqueness of the error variances, data reduction, and provision of factor scores. Though the factor scores of phase two and three are not exactly the same, but they nevertheless pursue the same objective, which is in providing data for further analysis. Both models also used eigenvalues to reduce the number of factors involved in the models. Again the research was able to show the different algorithms employed by the three phases and how they were put to use to ascertain there relevance in the study.

5.4 CONCLUSION

The three phases of factor analysis have one goal in common, that is to address the questions posed by factor analysis. This they all do, which still boils down to the issue of factor scores and its implementation.

However, based on the computational results from the three phases of factor analysis, it can be said that the rigorous approach of the second phase of factor analysis makes room for a better understanding of factor analysis and in making data available for further   analysis. Though it does not address adequately the issue of the nature of data involved, these it treats as continuous. But the third phase treats this issue with much caution. Hence the edge it has over the second phase of factor analysis model. Thus of the three phases it is the most advanced model as it is all encompassing especially in the proper manner it addresses each variable in the model.

REFERENCES
Anderson T.W.(1958) An Estimation of parameters in Latent Structure Analysis *Psychometrika* 19, 1-10.

Bartholomew David .J and Martin Knott (1999) latent models and factor Analysis 2ⁿᵈ edition, *London Arnold*.

Bartholomew David j . Martin Knott   and Irini Mousaki (2011), Latent models and Factor Analysis: A unified Approach *John Wiley and sons Ltd*.

Lazarsfeld . F. Paul (1950) *Measurement and Prediction*, *N.Y, Houghton – Mifflin*.

Lazarfeld.F. Paul and Henry N.W (1956) Latent Structure Analysis, *New York, Houghton-Mifflin*.

Melanie M. Wall,Yasua A. (2004) A Review Of Non-Linear Factor Analysis Methods And Applications. *IBM Thomas.J .Watson Research Centre.*

Michael E. Tipping, Christopher M. Bishop (1999) Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society: Series B, Vol 61(3), 611-622

Onyeagu S. I. (2003) Onyeagu .S.I, (2003), A First Course in Multivariate Statistical Analysis, Department of statistics Nnamdi Azikiwe University Awka. *Mega Concept,* 1ˢᵗ edition.

Philippe H., Elvezio R., Maria-pia V. F. (2013) Estimation of generalized linear latent variable models. *Journal of the Royal Statistical society*: Series B, 66(4), p893-908

Silvia B. and Silvia C. (2012) A General Multivariate Latent Growth Model with Applications to Student Achievement, *Journal of Educational and Behavioral Statistics.*

Spearman .C. (1904). General Intelligence Objectively Determined and Measured; American *Journal of Psychology* 15,201-393.

Spearman, c (1927) The Abilities of Man. *London Macmillan*

Thurstone .L.L (1931) Multiple Factor Analysis Psychological Review 38,406-427.

Thurstone .L.L (1947) *Mul*tiple Factor Analysis, *University of Chicago Press*.

Yule .G.U (1897) On the theory of Correlation*;* Journal of the Royal Statistical Society. 60,812-851.

.