**Research Article***Copyright © All rights are reserved by Onu Obineke Henry*

# Factor and Cluster Analysis on the Evaluation of the Knowledge of use of Contraceptive Methods by Women in Nigeria.

**Chris Chinedu, Joy Nonso, Ijomah Maxwell Azubuike and Onu Obineke Henry\****Mathematics/Statistics, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State, Nigeria*

**\*Corresponding author:** Onu Obineke Henry, Mathematics/Statistics, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State, Nigeria.

**Received Date:** September 15, 2022

**Published Date:** September 29, 2022

**Abstract**

This research aims to identify the most contributory factors of married women currently using contraceptives as family planning in Nigeria by factor analysis and cluster analysis techniques and then compare the results to get the most important variables. Data was a secondary data extracted from National Population Commission. In the practical section, principal Component method in factor analysis and Ward's Hierarchical method in cluster analysis were used to determine the most influential variables in the use of modern contraceptives among married women migration. From this study, it was reached that there are similarities of the variables extracted in factor analysis and cluster analysis, that is, the number of cluster equals the number of factors, as well as the sequence of variables within the factors are the same sequence of variables in the clusters except the factor and cluster (four and five) there are an exchange in place of variables, and it was the most important influencing variables are any method, any modern method, withdrawal, periodic abstinence, male condom, injectable and IUD.

**Keywords:** Factor analysis; Principal component analysis; Cluster analysis; Modern contraceptives

**Introduction**

Generally, the trend in this research is to use the quantitative measurement method and statistical curriculum and as well classify the scientific phenomena and analyse the reciprocal relation between phenomena by the basics objective, and the complications in most phenomena require the researcher to collect data for many variables, hence the collection of data is important in multivariate analysis methods, and in this method you need to understand and explain the interrelationships between the variables affecting the phenomenon.

Comparing the factor analysis with cluster analysis entails approaching a data collection from two opposing viewpoints. The

basic rationale of both approaches is categorization. Classification in either approach is based on homogeneity. Classification in either technique is based on uniformity. Homogeneity in cluster analysis indicates that research units, which are positioned in the rows of the data matrix and might be individuals or groups of individuals, are categorised into clusters based on their similarity on variables. Clusters are best defined by object homogeneity within the cluster and object heterogeneity between clusters. Factor analysis, on the other hand, focuses on the homogeneity of variables that results from the similarity of values supplied to variables by respondents. Variables are located in the matrix's columns and classified as factors or dimensions from the perspective of the data matrix [10].

## Materials and Methods

### Factor analysis (FA)

Factor analysis (FA) is a type of multivariate statistical method that focuses on data reduction and summarization. It deals with the problem of analysing the interdependence of a large number of variables and then explaining these variables in terms of their common, underlying factors.

The results and interpretations of methods are similar, but the mathematical models are different, so this method is a summary of the principal component (PCA) method [13]. The FA method is concerned with the correlations between a large number of quantitative variables. It reduces the number of primary variables by calculating a smaller number of new variables known as factors. This reduction is achieved by grouping variables into factors, which means that each variable within each factor is closely correlated, whereas variables belonging to different factors are less correlated [2, 13, 10, Sharma, 1996]

### Factor model

Explains the factor model of observed variables ( $k$ ) in sample size ( $n$ ) by a linear function ( $m$ ) of the common factors, where ( $m < k$ ) and  $k$  is a unique factor for each variable namely that:

$$\bar{X}_{k \times 1} = \bar{\mu}_{k \times 1} + A_{k \times m} \bar{F}_{m \times 1} + \bar{U}_{k \times 1} \quad (1)$$

Where:

X: random vector of variables.

A: factor loadings matrix of any constant matrix.

F: random vector of common factors.

U: random vector of a unique factor.

$\mu$ : average vector

Basic assumption of factor analysis is the correlation between a variety of variables known as intra-correlation that affect by the existence of common factors and return to the reality of these factors, i.e.,  $Cov(X_i, X_j) \neq 0$ , where the factor analysis explains these correlations that have minimum number of independent factors among them. These assumptions take the standard value of the variables, which are distributed normal distribution with mean (0) and variance (1), as well as to get rid of different variables and measurement units. Since the averages vector of both the common factor and unique factor are zero vectors depending on the assumption that the zero vector average variables:

$$E(X) = \mu = 0$$

So, factor model is as follows:

$$\bar{X}_{k \times 1} = A_{k \times m} \bar{F}_{m \times 1} + \bar{U}_{k \times 1} \quad (2)$$

The covariance matrix for each vector of the common factors (F) and the unique factors (U) (the assumption being independent) are as follows:

$$E \begin{bmatrix} \bar{F} \\ \bar{U} \end{bmatrix} \begin{bmatrix} \bar{F}' \\ \bar{U}' \end{bmatrix} = \begin{bmatrix} E(\bar{F}\bar{F}') & E(\bar{F}\bar{U}') \\ E(\bar{U}\bar{F}') & E(\bar{U}\bar{U}') \end{bmatrix} = \begin{bmatrix} \Phi_{m \times m} & 0_{m \times k} \\ 0_{k \times m} & \Psi_{k \times k} \end{bmatrix} \quad (3)$$

Under this assumption, we have three types of variation namely:

1. Common Variance is a part of the total variation and is correlated the rest of variables.
2. Specific Variance is a part of the total variation which in not correlated with the rest of variables but correlated with the same variable.
3. Error Variance is a part of exist variation during errors in the drawing of the sample or measured, or any other changes lead to volatility in the data.

Share all of the common's variation, specific variance, and error variance to create in the total variation which represents the equation.

$$1 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + S_j^2 + \ell_j^2 \quad (4)$$

The square roots of common ratios of  $a_{j1}$  to  $a_{jm}$  to return loading factors and represents the amount of the correlate of variable  $j$  for each factor [2, 14].

### Communalities

The communalities of the variable  $j$  (standard value) are the sum square of factors loading analysis of variable and represents the variation ratio which interprets summary factors from the correlation matrix analyzed for these variables, and symbolized by the symbol and the following formula:  $h_j^2$

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 = \sum_{p=1}^m a_{jp}^2 \quad (5)$$

Where:

$a_{jp}$  represents a weight of factor  $p$  for variable  $j$  which a matrix of factors known in factors loading. Therefore, the total variation of the variable  $j$  represents by equation.

$$h_j^2 + s_j^2 + \ell_j^2 = 1 \quad (6)$$

Where:

$s_j$  represents the total variation that is associated to each variable. One characteristic of  $h_j^2$  is positive that is between zero and one ( $h_j^2 + s_j^2 + \ell_j^2 = 1$ ).

The extent  $h_j^2$  of overlap between the variables and summary factors: if  $h_j^2$  is large and close to one, this means that the variable interferes with summary factor. If  $h_j^2$  is equal to zero, it means that the summary factor is not able to explain any part of the variation. If  $h_j^2$  is between zero and one that indicate the partial overlap between the variables and factors. [13, Sharma, 1996].

### Principal Components Analysis

Principal components analysis is a technique for transforming variables in a multivariate data set into new variables that are uncorrelated with each other and account for reducing proportions of the original variables' total variance (Sharma, 1996). Each new

variable is defined as a linear combination of the old ones. The method is described in detail in [5] but in summary. The first principal component,  $\mathbf{e}_1$ , is defined as the linear combination of the original variables,  $x_1, x_2, \dots, x_k$ , that accounts for the maximal amount of the variance of the  $x$  variables amongst all such linear combinations [2, 1].

The second principal component,  $\mathbf{e}_2$ , is defined as the linear combination of the original variables,  $x_1, x_2, \dots, x_k$  that accounts for a maximal amount of the remaining variance subject to being uncorrelated with  $\mathbf{e}_1$ . Subsequent components are defined similarly.

The matrix of data  $x_{n \times m}$  contains (m) columns of variables and (n) rows of observations. Matrix of data calculated by correlation matrix and the application of PCA by the correlations matrix to get m of characteristic roots that symbolizes by  $\lambda_i$  in decreasing order  $\lambda_1 > \lambda_2 > \dots > \lambda_m$  which represent variations of summary factors. [2, 1].

Explain of characteristic vector placed as factors in the linear combination of the original variables to give ( $\mathbf{e}_j$ ) of the value  $i$  of the principal component  $j$  -

$$\mathbf{e}_j = a_{1j}X_{i1} + a_{2j}X_{i2} + \dots + a_{kj}X_{ik} = \sum_{l=1}^k a_{lj}X_{il} \quad (7)$$

Where:

$\mathbf{e}_j$  represents the principal component  $j$ ,  $a_j$  represents the coefficient of variable (i) of the component (j) which are values of Eigen Vectors accompanying of Eigen roots.

In the correlation matrix user, the variation  $j$  equal to  $\text{Var}(\mathbf{e}_j) = \lambda_i$  and  $\lambda_i$  represents a characteristic root of the principal component  $j$ . It can be obtained by summary factors to get factors  $F_1, F_2, \dots, F_m$  by dividing each principal component of the standard deviation as follows:

$$F_i = \frac{\mathbf{e}_j}{\sqrt{\text{Var}(\mathbf{e}_j)}} = \frac{\mathbf{e}_j}{\sqrt{\lambda_i}} \quad (8)$$

Where:

$F_i$  is a factor (i) and it can be represented as ( $\mathbf{e}_j$ ) as follows:

$$\mathbf{e}_i = F_i \sqrt{\lambda_i} \quad (9)$$

Transfer Principle Component Model to the Factor Model

$$X_i = \sum_{j=1}^m a_j \mathbf{e}_j$$

$$i = 1, 2, \dots, k; j = 1, 2, \dots, m$$

$$= \sum_{j=1}^m a_j F_j \sqrt{\lambda_j}$$

$$\therefore L_j = a_j \sqrt{\lambda_j}$$

$$= \sum_{j=1}^m L_j F_j \quad (10)$$

## Rotation of Factors

After obtaining the initial factor solutions, one is interested to rotate the loadings. The goal of factor rotation is to find simpler factor structure that can use to make interpretation of the resulting factors easily and meaningfully and to determine the appropriate number of factors. It is a way of maximizing high loadings and minimizing low loadings so that the simplest possible structure is achieved. To accomplish this and for the simplicity of this study, orthogonal rotations are done using the varimax procedure. Here we only focus on the orthogonally rotated solutions as they can produce more simplified factor structures from a large amount of data [13, 5; Sharma, 1996; 3].

In this research we used loading method by orthogonal method (Kaiser - Varimax) that applies the definition of the variation by factor loading (i).

$$\text{Var}(F_k) = \frac{1}{P} \left[ \sum_{j=1}^p (a_{jk}^2)^2 - \frac{1}{P} \left( \sum_{j=1}^p a_{jk}^2 \right)^2 \right],$$

$$k = 1, 2, \dots, m \quad (11)$$

Where:

$\text{Var}(F_k)$  represents a variation factor (k),

$a_{jk}$  represents the value of the saturation variable (j) factor (k).

$p$  represents the number of variables.

## Choosing the number of factors

It is crucial to pick the right number of common factors. We make a graph pair  $(j, \lambda_j)$ , (the "scree plot") and watch for when it starts to become "flat" (Cattell, 1966). The Kaiser criterion is another criterion for dealing with the number of factors issue (Kaiser 1960). When the eigenvalues are greater than one ( $\lambda > 1$ ), a factor  $j$  is important. If the number of factors discovered by the Kaiser Test is equal to the number of factors discovered by the "scree plot," we can proceed with the other procedures; otherwise, we must choose one of the previously obtained results. If one of the produced results from the "scree plot" graphic is chosen, the preceding methods must be performed in order to arrange the best number of elements. As the number of factors increases, so do the results. Optional methods include Kaiser's (Kaiser 1956) "eigenvalues greater than one" rule, the scree plot, a priori theory, and preserving the number of components that accounts for the greatest proportion of variance or provides the best interpretable solution [13, 12].

## Variance extracted

Another strategy based on a similar conceptual structure is to keep the number of components that account for a particular percentage of the retrieved variance. The research differs on how much variance must be explained before the number of components is considered sufficient. The majority believes that 75-90% of the variance should be accounted for; however, some believe that as

little as 50% of the variance explained is acceptable. This seemingly broad norm must be examined in connection to the fundamental distinctions across extraction methods, as with any criterion method that is primarily based on variance. When evaluating the value of percent of variance extracted, the amount of variation that was included for extraction must be considered [3]. Component analysis contains more variance to be described, implying that higher percentages of explained variance are expected than would be required if only common variance is included [3].

## Cluster Analysis

Cluster analysis is a class of multivariate techniques that aims at grouping objects or respondents based on the characteristics they have (Malhotra & Birks, 2007; Hair et al. 2009). The clusters formed based on the analysis should demonstrate high homogeneity within clusters and high heterogeneity between clusters. Cluster analysis differs from factor analysis in the sense that the former is concerned with reducing the number of objects by grouping them into clusters whereas the latter aims at reducing the number of variables by grouping them to form new constructs or factors [5, Sharma, 1996].

The first task in cluster analysis is to select the variables that are used for forming clusters. The selected variables should be meaningful in describing the similarities between objects in consideration of the research questions.

There are two key assumptions with regard to cluster analysis. The most important one states that the basic measure of similarity, which is the basis of clustering, must be a valid measure of the similarity between objects. A second major assumption is that there is theoretical justification for structuring the objects into clusters. These two assumptions are both valid for my data set and will be used in the following analysis [5, Sharma, 1996].

## Basic concepts in the cluster analysis

### Element

Element ( $X_i$ ) is a vector distance for measurement for the  $n$ -dimensional  $X_i = (X_{i1}, X_{i2}, \dots, X_{in})$  and elements are numerical amounts of possible measurement (Yan, 2005).

#### A. Element selection

The goal of cluster analysis is to investigate the cluster structure in data without knowing the true categorization. The elements chosen for research will, in particular, define the data structure. The underlying clusters or populations should be represented by an ideal sample. Outliers (datapoints that fall outside the general region of any cluster) should be avoided as much as possible in order to ease the recovery of distinct and trustworthy cluster structures, even if they do occasionally form a single cluster (Yan, 2005).

#### B. Variable selection

Data investigated in cluster analysis are often recorded on multiple variables. The importance of choosing an appropriate set of variables for clustering is obvious (Yan, 2005).

## Similarity Measures

An essential step in the cluster analysis procedure is to obtain

a measure of the similarity or "proximity" between each pair of objects under study. The Minkowski metric or  $L_q$  norm calculates the distance  $d$  between the two objects  $x$  and  $y$  by comparing the values of their  $n$  features. The Minkowski metric can be applied to frequency, probability and binary values [13, 9].

$$d(x, y) = L_q(x, y) = \sqrt[q]{\sum_{i=1}^p |x_i - y_i|^q} \quad (12)$$

Where:

$d(x, y)$  is the distance between  $x$  and  $y$  and

$p$  is the number of variables

Two important special cases of the Minkowski metric are  $q=1$  and  $q=2$ . The most commonly used measure of similarity is Euclidean distance or  $L_2$  norm, which can be expressed as follows:

$$d(x, y) = L_2 = \sqrt{\sum_{i=1}^p |x_i - y_i|^2} \quad (13)$$

Manhattan distance or City block distance or  $L_1$  norm as follows:

$$d(x, y) = L_1 = \sum_{i=1}^p |x_i - y_i| \quad (14)$$

## Cluster

Is a collection of adjacent elements in the statistical community and homogenous to some extent, also known as a group of similar things, that are different from the other clusters [9].

## Dendrogram

To visualize the clustering results various versions of the dendrogram have been proposed and widely used. The main objective of this visualization is a tree-based structure yielded by a hierarchical algorithm which is easy to understand the similarities/dissimilarities between elements. It demonstrates both the cluster and the sub-cluster relationships and also the order in which the clusters are merged for agglomerative algorithm or split for divisive case. Usually, a dendrogram is read from left to right. In a dendrogram, the data points are represented along the bottom and referred to as external/terminal nodes, and clusters to which the data belong are formed by joining with the internal nodes. The name of objects added to the terminal nodes are known as labels. The height of the dendrogram shows the distance between objects or clusters. Therefore, the small values of the height indicate a high similarity between the objects and the large value of the height indicates more distance. This can be shown in the figure below. The coalition between each internal node and the height can be explained mathematically by the following condition: [13, 9].

$$h(Q) \leq h(R) \Leftrightarrow Q \subseteq R$$

For every subset of objects  $Q$  and  $R$  and  $Q \cap R \neq \Phi$

Here,  $h(Q)$  and  $h(R)$  represent the highest of  $Q$  and  $R$  respectively.

## Techniques of Clustering

The focus was on hierarchical cluster which was summarized by [8] and are employed in this study:

### Hierarchical Cluster Analysis

Cluster analysis is a technique for identifying trends in a data collection by grouping (multivariate) observations into clusters. Hierarchical clustering is one of the cluster analyses approaches because it involves grouping and categorizing data into clusters, and these clusters are very similar to each other in terms of similarity within clusters, but they differ or are dissimilar between clusters. Hierarchical cluster analysis is regarded as a good, active method and tool for analysing data in various ways. This method (Hierarchical cluster) is used to analyse a data matrix that consists of (n) elements, each of which has a number of variables (p). Hierarchical approaches are classified into two types: agglomerative and divisive [13, Shen, 2007].

### Agglomerative Methods

Agglomerative hierarchical clustering begins with n clusters, each containing a single object in the data. The two clusters with the shortest between-cluster distances are fused in the second phase, and they will be considered as a single cluster in the following step. As the operation progresses, a single cluster holding all n objects is formed.

**Divisive methods:** The divisive approach goes in the opposite direction. That is, it starts with a single cluster with all the data points in it, and iteratively divides it into two clusters that are furthest apart from one another.

The structure in (Figure B) resembles an evolutionary tree, and it is in biological applications that hierarchical classifications are perhaps most relevant. According to Rohlf (1970), a biologist, „all other things being equal“, aims for a system of nested clusters.

### Steps of Clustering [10]

Agglomerative hierarchical clustering algorithm steps in below

- compute the distance matrix.
- Repeat
- Merge the closest two clusters.
- Update the distance matrix to reflect the proximity between the new cluster and the original clusters.
- Until only one cluster remains.

### Clustering Methods

There are two types of clustering methods such as Nearest Neighbor (Single Linkage) and Furthest Neighbor (Complete Linkage):

#### Single Linkage (Nearest Neighbor Cluster Distance)

The distance  $d$  between two clusters  $C_i$  and  $C_j$  is defined as the minimum distance between the cluster objects. The cluster distance measure is also referred to as single linkage. Typically, it causes a chaining effect concerning the shape of the clusters, i.e.,

whenever two clusters come close to each other they stick together even though some members might be far from each other [13].

$$d(C_i, C_j) = d_{\min}(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \quad (15)$$

#### Complete Linkage (Furthest Neighbor Cluster Distance)

The maximum distance between cluster objects is defined as the distance  $d$  between two clusters  $C_i$  and  $C_j$ . Complete-linkage distance is another name for cluster distance. Because every object within a cluster is expected to be adjacent to every other object within the cluster, and outlying objects are not incorporated, it typically creates compact clusters with tiny widths (Yim & Ramdeen, 2015).

$$d(C_i, C_j) = d_{\max}(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y) \quad (16)$$

### Wald Method

The loss of information (or increase in inaccuracy) while merging two clusters is defined as the distance  $d$  between them (Ward, 1963). The sum of distances between the cluster's objects and the cluster centroid  $cen_C$  is used to calculate the error of a cluster  $C$ . When two clusters are combined, the combined cluster's error is greater than the sum of the errors of the two individual clusters, resulting in information loss. However, the merging is done on the most homogenous clusters in order to unite them and reduce the amount of variation within the merged clusters. Ward's approach is prone to producing small, compact clusters. It is a least squares method, so implicitly assumes a Gaussian model (Sharma, 1996).

$$d(C_i, C_j) = d_{ward}(C_i, C_j) = \sum_{x \in (C_i \cup C_j)} d(x, cen_C) - \left[ \sum_{x \in C_i} d(x, cen_{C_i}) + \sum_{x \in C_j} d(x, cen_{C_j}) \right] \quad (17)$$

## Results and Discussion

This part includes the application of two types of multivariate analysis (factor analysis and cluster analysis) to identify the most important method of modern contraceptives among married women and the comparison between the results of the analyzes in order to know the extent of conforming between them. Results were extracted using the statistical program (SPSS V.22).

### The Data Collection

The data used for this research was a secondary data extracted from the National Population Commission (NPC 2020). The information on the knowledge and use of family planning methods was obtained from female respondents by asking them to mention any ways or methods by which couple delay or avoid pregnancy. However, the actual sample size used in analysis is 7,621.

### Data analysis

#### Results of factor analysis

The contraceptives indices used in this study were numbered as follows:

1. No method,
2. Female sterilization,
3. Male sterilization,
4. Any method,
5. Any traditional methods,
6. IUD,
7. No of women age 15-49 years currently married or in union,
8. Any modern method,
9. LAM,
10. Others,
11. Withdrawals,
12. Periodic abstinence,
13. Injectables,
14. Diaphragm,
15. Implant,
16. Female condom,
17. Male condom,
18. Pills.

(Table 1) shows the correlation matrix which is typically used as an eyeball test and gets a feeling for which variable is strongly associated with which variable.

**Table 1:** The correlation matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1																	
2	-0.34	1																
3	-0.13	0.2	1															
4	-0.74	0.4	0.13	1														
5	-0.63	0.4	0.08	0.7	1													
6	-0.63	0.1	0.25	0.7	0.25	1												
7	0.37	0.2	-0.23	0	-0.4	-0.22	1											
8	-0.95	0.3	0.12	1	0.51	0.71	-0.29	1										
9	-0.49	0.1	-0.12	0.5	0.44	0.16	-0.19	0.43	1									
10	-0.1	0.5	0.07	0.1	0.1	0.06	0.12	0.08	0.11	1								
11	-0.73	0.2	0.08	0.7	0.87	0.41	-0.31	0.55	0.38	-0.13	1							
12	-0.58	0.3	0.07	0.6	0.86	0.09	-0.47	0.36	0.36	-0.15	0.64	1						
13	-0.65	0.1	0.08	0.7	0.07	0.59	-0.15	0.81	0.08	0.003	0.14	0.001	1					
14	-0.28	0.1	0.61	0.3	0.25	0.18	-0.08	0.2	0.16	-0.06	0.43	0.29	0.15	1				
15	-0.54	0.3	0.19	0.5	0.26	0.19	-0.13	0.57	0.08	0.11	0.19	0.22	0.43	0	1			
16	-0.25	0.5	0.16	0.2	0.19	0.19	-0.06	0.24	-0.002	0.26	0.14	0.05	0.07	-0.23	0.12	1		
17	-0.85	0.5	0.23	0.8	0.73	0.59	-0.43	0.77	0.48	0.07	0.74	0.57	0.36	0.28	0.32	0.3	1	
18	-0.53	-0.1	-0.18	0.5	0.07	0.45	0.04	0.65	0.19	0.04	0.12	-0.19	0.62	-0.25	0.31	0.3	0.21	1

(Table 2) provides the Kaiser-Meyer- Olkin (KMO) and Bartlett's test of sphericity. The KMO criterion can have values between (0,1) where the usual interpretation is that 0.8 indicates a good adequacy to use the data in a factor analysis. If the KMO criterion is less than 0.5 we cannot extract in some meaningful way. However, the overall Measure of Sampling Adequacy (MSA) for the

set of variables included in the analysis was 0.565, which exceeds the minimum requirement of 0.50 for overall MSA and also means this test does not find the level of correlation to be too low for factor analysis. The variables remaining in the analysis satisfy the criteria for appropriateness of factor analysis.

**Table 2:** KMO and Bartlett's Test.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.565	
Bartlett's Test of Sphericity	Approx. Chi-Square	946.324
	Df	153
	Sig.	0

(Table 3) which is the next table is the Anti-image Matrices. The image theory splits into image and anti-image. The rule of thumb is

that in the anti-image covariance matrix only a maximum of 25% of all non-diagonal cells should be greater than (0.09).

Table 3: Anti-image Matrices.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.002	-0.001	0.008	0.002	0.02	-0.003	0.005	0.005	0.004	0.003	0.002	0.002	0.001	8.00E-04	0.003	0.007	0.001	0.002
2	-0.001	0.237	-0.052	-0.001	0.002	0.103	-0.012	0.001	0.084	-0.04	0.011	-0.01	0.026	-0.001	0.068	-0.147	0.023	0.062
3	0.008	-0.052	0.54	-0.001	0.003	0.063	0.002	0.002	0.045	0.035	0.022	-0.01	0.01	-0.005	-0.006	0.125	0.007	0.022
4	0	-0.001	-0.001	6.00E-06	0	-0.001	-8.00E-06	-8.00E-06	7.00E-06	0	1.00E-06	0.022	0	0.001	7.00E-06	0.002	0	-3.00E-06
5	0	0.002	0.003	0	0.001	0.002	-0.003	0	0	-0.002	0	0	0	-0.003	0	-0.006	0.001	0
6	-0.003	0.103	-0.064	-0.001	0.002	0.148	0.021	0	0.072	-0.028	-0.002	-0.001	0.025	0.049	0.078	-0.073	0.029	0.049
7	0.005	-0.012	0.063	0.001	-0.003	0.021	0.564	-0.002	0.036	-0.043	-0.015	-0.008	0.014	-0.031	0.023	0.059	0.019	0.003
8	0.005	0.001	0.002	-8.00E-06	0	0	-0.002	0	0	0	-5.00E-06	-0.001	-0.001	-0.002	-0.001	-0.003	0	0
9	0.004	0.084	0.045	7.00E-06	0	0.072	0.036	0	0.131	0.028	0.027	0	-1.00E-06	0.017	0.099	0.008	0.043	0.069
10	0.003	-0.04	0.035	0	-0.002	-0.028	-0.043	0	0.028	0.138	0.044	0.022	0.04	-0.027	0.029	0.044	0.019	0.023
11	0.002	0.011	0.022	1.00E-06	0	-0.002	-0.015	-5.00E-06	0.029	0.049	0.026	0.027	0.012	-0.2	0.025	0.012	0.011	0.018
12	0.002	-0.01	0.022	0	-0.001	-0.008	-0.001	0	0.022	0.065	0.027	0.041	0.01	-0.01	0.02	0.034	0.01	0.015
13	0.001	0.026	0.01	0	0	0.025	0.014	-1.00E-06	0.04	0.016	0.012	0.01	0.017	-0.001	0.037	-0.004	0.021	0.027
14	8.00E-06	-0.001	-0.005	0.001	-0.003	0.049	-0.031	-0.002	0.017	-0.027	-0.02	-0.01	-0.001	0.055	0.047	0.075	0.012	0.052
15	0.003	0.068	-0.006	7.00E-06	0	0.078	0.023	-0.001	0.099	0.029	0.025	0.02	0.037	0.045	0.045	-0.009	0.049	0.072
16	0.007	-0.147	0.125	0.002	-0.006	-0.073	0.059	-0.003	0.008	0.044	0.012	0.034	-0.004	0.075	0.119	0.363	-0.02	-0.025
17	0.001	0.023	0.007	0	0.001	0.029	0.019	0	0.043	0.019	0.011	0.01	0.021	0.021	0.049	-0.019	0.037	0.039
18	0.002	0.062	0.022	0.0003	0	0.049	0.003	0	0.069	0.023	0.018	0.015	0.027	0.055	0.072	-0.025	0.039	0.07

Anti-image Covarian

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.864	-0.032	0.274	0.39	-0.237	-0.191	0.161	-0.243	0.308	0.208	0.295	0.256	0.21	0.003	0.246	0.313	0.143	0.215
2	-0.032	0.469	-0.144	-0.191	0.207	0.552	-0.032	0.113	0.479	0.222	0.144	-0.1	0.403	-0.002	0.404	-0.503	0.243	0.481
3	0.274	-0.144	0.47	-0.144	0.177	-0.277	0.114	0.201	0.169	0.128	0.18	0.149	0.102	-0.011	-0.025	0.281	0.053	0.115
4	0.39	-0.191	-0.144	0.71	-0.968	-0.231	0.201	-0.973	0.003	0.055	0.001	0.133	-0.112	0.192	0.027	0.435	-0.22	-0.018
5	-0.237	0.207	0.71	-0.968	0.627	0.241	-0.14	0.972	0.012	-0.194	-0.124	-0.229	0.105	-0.178	-0.021	-0.432	0.22	0.019
6	-0.191	0.525	-0.968	-0.231	0.241	0.581	0.071	0.1	0.52	-0.198	-0.029	-0.099	0.486	0.189	0.587	-0.315	0.393	0.484
7	0.161	-0.032	-0.231	0.201	-0.14	0.071	0.827	-0.201	0.131	-0.155	-0.126	-0.009	0.143	-0.061	0.091	0.131	0.131	0.015
8	-0.243	0.113	0.201	-0.973	0.071	0.827	-0.201	0.702	-0.09	-0.066	-0.03	-0.151	-0.011	-0.227	-0.14	-0.382	0.111	-0.096
9	0.308	0.479	0.169	0.003	0.1	-0.201	0.131	-0.094	0.319	0.207	0.489	0.306	0.864	0.069	0.799	0.037	0.627	0.722
10	0.208	-0.222	0.128	0.055	-0.194	0.131	-0.155	-0.066	0.207	0.168	0.819	0.861	0.325	-0.109	0.225	0.196	0.265	0.234
11	0.295	0.144	0.18	0.001	-0.124	-0.029	-0.126	-0.03	0.489	0.819	0.591	0.832	0.545	-0.187	0.441	0.127	0.363	0.428
12	0.246	-0.1	0.149	0.133	-0.279	-0.099	-0.099	-0.151	0.306	0.861	0.832	0.567	0.366	-0.071	0.288	0.278	0.258	0.289
13	0.21	0.405	0.102	-0.12	0.105	0.426	0.143	-0.011	0.846	0.325	0.545	0.366	0.417	-0.012	0.824	-0.049	0.833	0.785
14	0.003	-0.002	-0.011	0.192	-0.178	0.189	-0.061	-0.227	0.069	-0.109	-0.187	-0.071	-0.012	0.685	0.195	0.184	0.092	0.31
15	0.246	0.404	-0.025	0.027	-0.021	0.587	0.091	-0.14	0.799	0.225	0.411	0.288	0.824	0.195	0.319	-0.403	0.74	0.789
16	0.313	-0.503	0.281	0.453	-0.432	-0.315	0.131	-0.382	0.037	0.196	0.127	0.278	-0.049	0.181	-0.043	0.39	-0.16	-0.158
17	0.143	0.243	0.053	-0.223	0.22	0.393	0.131	0.111	0.627	0.265	0.363	0.258	0.833	0.092	0.74	-0.163	0.628	0.779
18	0.215	0.481	0.115	-0.018	0.019	0.484	0.015	-0.096	0.722	0.234	0.428	0.289	0.785	0.36	0.789	-0.158	0.779	0.37

Anti-image Correlation



The second part of the table is the anti-image correlations where the diagonal elements of that matrix are the MSA values. A figure of 0.8 indicates good accuracy; if MSA is less than 0.5, we exclude the variable from the analysis. We find that those who took pills has a MSA value of 0.370 and might be a candidate for exclusion, we can proceed with our factor analysis.

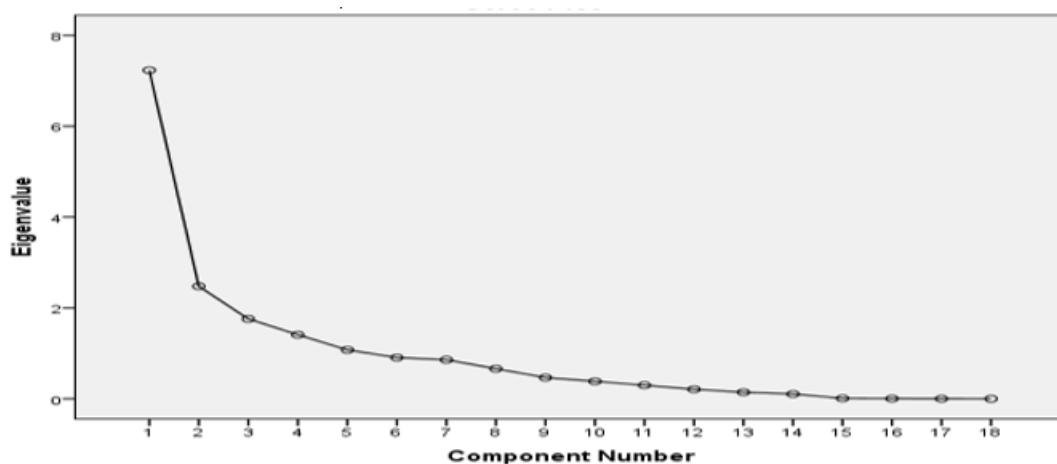
The next which is Table 4 shows the communality. The communality of a variable is the variance of the variable that is explained by all factors together. Mathematically it is sum of the squared factor loadings for each variable. A rule of thumb is for all communalities to be greater than 0.5. This holds true in this example, for all the communalities are greater than 0.5 as can be seen in the above table.

**Table 4:** The communalities.

	Initial	Extraction
No method	1	0.992
Female sterilization	1	0.803
Male sterilization	1	0.768
Any method	1	0.994
Any traditional method	1	0.916
IUD	1	0.618
No of women aged 15-49 years currently married or in union	1	0.622
Any modern method	1	0.989
LAM	1	0.524
Others	1	0.737
Withdrawal	1	0.789
Periodic abstinence	1	0.801
Injectable	1	0.845
Diaphragm/foam/jelly	1	0.751
Implant	1	0.39
Female condom	1	0.738
Male condom	1	0.833
Pills	1	0.841

(Table 5) shows the proportion of variability which is firstly explained with all factors together and then only with the factors before rotation. The table also includes the eigenvalues of each factor. The eigenvalue is the sum of squared factor loadings for each factor. SPSS extracts all factors that have an eigenvalue greater than

0.1. The result shows five factors. The table also shows us the total explained variance before and after rotation. The rule of thumb is that the model should explain more than 70% of the variance. In this study, the model explains 77% (Figure 1).



**Figure 1:** The scree plot.

**Table 5:** Total Variance Explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.232	40.176	40.176	7.232	40.176	40.176	4.759	26.438	26.438
2	2.475	13.752	53.928	2.475	13.752	53.928	4.621	25.674	52.112
3	1.758	9.768	63.696	1.758	9.768	63.696	1.815	10.082	62.194
4	1.409	7.830	71.526	1.409	7.830	71.526	1.413	7.849	70.043
5	1.076	5.980	77.505	1.076	5.980	77.505	1.343	7.462	77.505
6	.907	5.040	82.545						
7	.858	4.768	87.314						
8	.662	3.677	90.990						
9	.467	2.593	93.584						
10	.384	2.133	95.716						
11	.299	1.659	97.375						
12	.208	1.158	98.532						
13	.144	.801	99.333						
14	.105	.585	99.918						
15	.009	.052	99.970						
16	.004	.023	99.992						
17	.001	.007	100.000						
18	3.886E-5	.000	100.000						

Extraction method: Principal Component Analysis

The scree plot graphs the eigenvalue against the factor number. We can see the value of the first and second column of the table immediately above. From the second factor on, we can see that the line is almost flat, meaning that each successive factor is accounting for smaller and smaller amounts of the total variance.

(Table 6, 7) which displays the factor matrix, and the rotated factor matrix are the keys in interpreting our factor. The factor loadings that are shown in the tables are the correlation coefficient between the variable and the factor. The factor loadings should be greater than 0.4 and the structure should be easy to interpret.

**Table 6:** The Factor Matrix.

	Factor				
	1	2	3	4	5
No method	-0.99				
Female sterilization	0.41		0.769	0.163	
Male sterilization	0.174		0.12	0.84	
Any method	0.989				
Any traditional method	0.784	-0.516		-0.161	
IUD	0.641	0.377	-0.139	0.205	
No of women age 15-49 years currently married or in union	-0.446	0.309		-0.23	0.523
Any modern method	0.922	0.349	-0.107		
LAM	0.496	-0.183	-0.118	-0.423	0.228
Others		0.181	0.708		0.44
Withdrawal	0.763	-0.411	-0.149	-0.126	
Periodic abstinence	0.629	-0.591		-0.117	-0.206
Injectable	0.584	0.618	-0.251	0.214	0.114
Diaphragm/foam/jelly	0.302	-0.437	-0.244	0.194	0.61
Implant	0.506	0.277		0.216	

Female condom	0.279	0.236	0.632	-0.343	-0.296
Male condom	0.892	-0.162			
Pills	0.453	0.706	-0.168	-0.311	-0.11
Extraction Method: Principal Component Analysis.					

**Table 7:** Rotated Factor Matrix.

	Factor				
	1	2	3	4	5
No method	-0.644	-0.741	-0.157		
Female sterilization	0.302		0.788	0.264	-0.123
Male sterilization		0.15	0.106	0.839	0.172
Any method	0.633	0.753	0.151		
Any traditional method	0.927	0.137	0.175		
IUD	0.179	0.75		0.15	
No of women age 15-49 years currently married or in union	-0.568	-0.101	0.179	-0.451	0.234
Any modern method	0.408	0.898	0.118		
LAM	0.515	0.19		-0.434	0.175
Others	-0.133		0.841		
Withdrawal	0.831	0.254			0.178
Periodic abstinence	0.891				
Injectable		0.912			
Diaphragm/foam/jelly	0.335				0.796
Implant	0.137	0.543	0.192	0.192	
Female condom	0.176	0.143	0.499	-0.147	-0.645
Male condom	0.733	0.481	0.223	0.111	
Pills		0.771		-0.367	-0.322

For analysis and interpretation factor analysis result in (Table 6, 7) we are only concerned with rotation sums of squared loadings. Here note that the first factors explain relatively large amounts of variance for 36.7% of the total variance and included seven variables (Any method, Any traditional method, Any modern method, LAM, Withdrawal, Periodic abstinence, male condom), the second factor explained 36.7% of the total variance and contain seven variables (Any method, IUD, Any modern method, Injectables, Implant, Male condom and Pills), the third factor has been interpreted 16.0% of the total variance with variables (Female sterilization, others and Female condom), the fourth factor explains 5.3% of the variance and included one variable (Male sterilization) and factor five explains 5.3% of the total variance and contain variable (Diaphragm). All the

remaining factors are not significant.

### The Results of cluster analysis

Cluster analysis was used to determine the relationship between the variables studied in terms of the degree of similarity or difference and classify of the variables into homogeneous groups among them. Cluster analysis has been used to classify variables that evaluates the use of contraceptives by women into five clusters separately which is the strong degree of homogeneity among the members of one cluster and weak between members of different clusters. We have used Ward's Method among the methods of cluster analysis and have used Euclidean Distance measure to create the matrix distance. The results were as shown in the (Figures 2, Table 9).

**Table 8:** Rotated factor matrix and communalities.

Variables	1	2	3	4	5	Communalities Extraction
Female sterilization			0.788			0.803
Male sterilization				0.839		0.768
Any method	0.633	0.753				0.994
Any traditional method	0.927					0.916

IUD		0.75				0.618
Any modern method	0.408	0.898				0.989
LAM	0.515					0.524
Others			0.841			0.737
Withdrawal	0.831					0.789
Periodic Abstinence	0.891					0.801
Injectables		0.912				0.845
Diaphragm				0.796		0.751
Implant		0.543				0.39
Female condom			0.499			0.738
Male condom	0.892	0.481				0.833
Pills		0.771				0.841

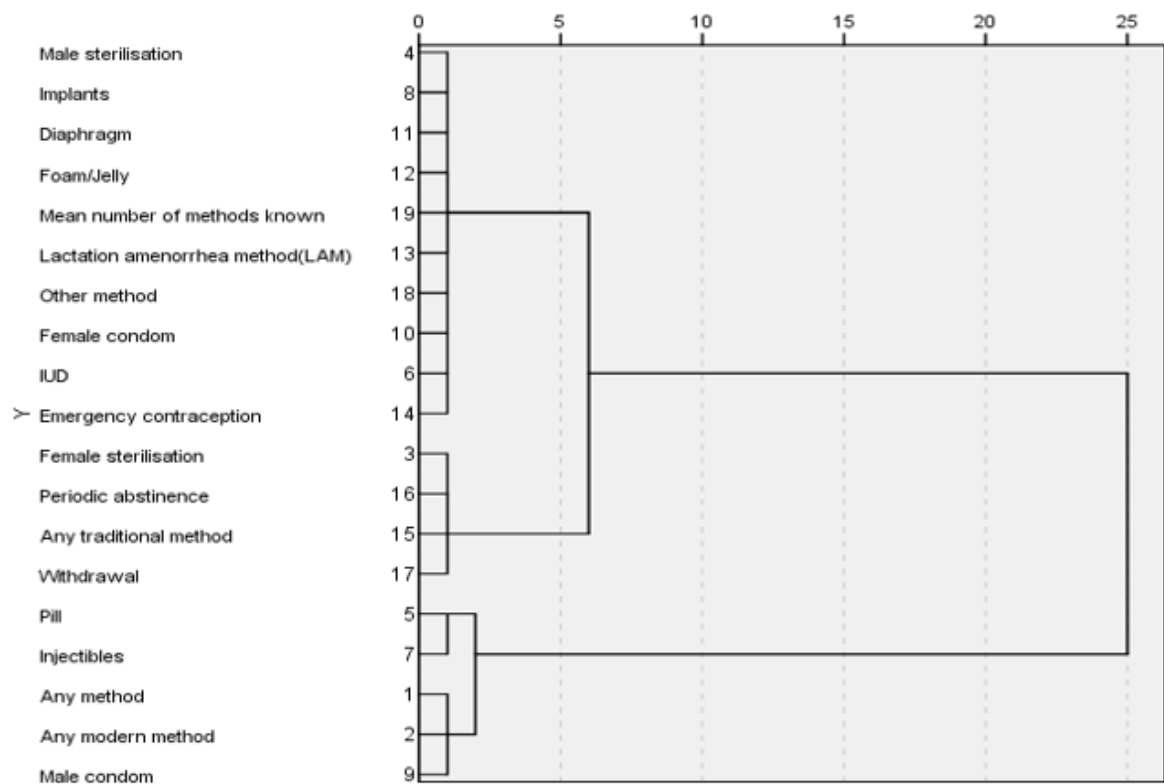


Figure 2: The Dendrogram using Wald method.

(Figure 5) shows five number of clusters that used in this analysis as well as variables belong to each cluster. Vertical axis (Y) in the structure represents the distance between merge variables

and the horizontal axis (X) represents the merge variables. (Table 9) indicates the number of clusters and the proportion of variables within each cluster as well as variables within each cluster.

**Table 9:** Cluster analysis of data.

Cluster	No of member	Variables	Proportion of variables in cluster
1	7	Any method, Female condom Any traditional method, Any modern method, LAM, Withdrawals, Periodic Abstinence.	36.70%
2	7	IUD, Female condom, Any method, Any modern method, Periodic Abstinence, Diaphragm, male condom.	36.70%
3	3	Others, Female sterilization, Implant	16%
4	1	Injectables	5.30%
5	1	Male sterilization	5.30%
Total	19		100%

### A comparison of the results of factor analysis and cluster analysis

**Table 10:** Comparison between the results of factor analysis and cluster analysis.

Factor	Variables in factor	Cluster	Variables in cluster
1	Any method, Any traditional method, Any modern method, LAM, Withdrawals, Periodic Abstinence, Female condom.	1	Any method, Female condom Any traditional method, Any modern method, LAM, Withdrawals, Periodic Abstinence.
2	Any method, IUD, Any modern method, Periodic Abstinence, Diaphragm, Female condom, Male condom.	2	IUD, Female condom, Any method, Any modern method, Periodic Abstinence, Diaphragm, male condom.
3	Female sterilization, Others, Implant.	3	Others, Female sterilization, Implant
4	Male sterilization	4	Injectables
5	Injectables	5	Male sterilization

(Table 10) contains a comparison between factor and cluster analysis of the data in order to determine the compatibility of the results of analyses, which illustrate the number of variables in each factor as same as number of variables in each cluster and both analyses gave the closest results.

Cluster analysis showed the lowest distance between Any method and any modern method is 4.825 while the highest distance between Withdrawal method and Periodic Abstinence is 9,568.173 as well as factor analysis in table 1 shown, the highest correlation between Any method and any modern methods is 0.957 while the lowest correlation between withdrawal method and Periodic Abstinence is 0.638.

## Conclusions and recommendations

### Conclusion

The most important conclusions were summarized through practical part as follows:

1. The analysis shows that there are quite similarities between the results of factor analysis and cluster analysis which the number of factors equal to the number of clusters, as well as in terms of inclusion, the factor analysis and cluster analysis have the same variables in each cluster and factor with the same sequence except the factor and cluster 4, 5 with difference place of variables.
2. Both results show that there are five significant factors and five important clusters from which to determine the variables in

evaluating the use of contraceptives the ratio explain is 77.5% of the total variance.

3. The lowest distance is between Any method and any modern method while the highest distance is between Withdrawal method and Periodic Abstinence as well as the lowest correlation is between Any method and any modern method while the highest correlation is between Withdrawal method and periodic abstinence.
4. In this research reveals that
  1. The contraceptives prevalence rate is 15% among currently married women using Any method of contraception.
  2. Any method, any modern method, Withdrawal, Periodic Abstinence were mostly used in Nigeria.
  3. Family planning with the use of traditional method of contraceptive in Nigeria is very low.

### Recommendations

1. We recommend the possibility of using factor analysis in the classification gives the same results as cluster analysis, especially when the application analyzes.
2. Taking into account the findings of a study of use of contraceptives in family planning and try to be educated, campaigned and encourage couples to limit or space the number of children they want to have through the help of government via Health sectors.

3. Comparisons to be made between other methods of cluster analysis and the other ways the factor analysis.

### Acknowledgement

None.

### Conflict of Interest

No conflict of interest.

### References

1. Johnson R A, Wichern D W (2013) Applied Multivariate Statistical Analysis, (6<sup>th</sup> Edn.), Pearson Prentice Hall.
2. Rencher A C (2002) Methods of Multivariate Analysis, Second edition, Brigham Young University, John Wiley & Sons, Inc, USA.
3. Anderson T W (1984) An Introduction to Multivariate Statistical Analysis, (2<sup>nd</sup> Edn.), John Wiley and Sons, New York-U.S.A.
4. Seber G A F (1984) Multivariate Observations. John Wiley and Sons, New York - USA.
5. Everitt B S (2005) An R and S-PLUS Companion to Multivariate Analysis, Springer-Verlag London.
6. Afifi A, Clark V (1984) Computer Aided Multivariate Analysis, Lifetime Learning Publications, USA.
7. Beavers A S (2013) Practical Considerations for Using Exploratory Factor Analysis in Educational Research 18: 6.
8. Muca M, Puka L, Bani K (2013) Principal Components and the maximum likelihood method as tools to analyze large data with a psychological testing example, European Scientific Journal 9(20): 176-184.
9. Hair JR (2010) Multivariate Data Analysis, Seventh Edition, Pearson Prentice Hall.
10. Han J, Kamber M, Pei J (2012) Data Mining "Concepts and Techniques", (3<sup>rd</sup> Edn.), Morgan Kaufmann, USA.
11. Clatworthy J, Deanna Buick, Matthew Hankins, John Weinman, Robert Horne, et al. (2005) The use and reporting of cluster analysis in health psychology: A review. British Journal of Health Psychology, 10(3): 329-358.
12. Everitt B S (2011) Cluster Analysis. Chichester, UK: John Wiley & Sons, Ltd.
13. Finch H (2005) Comparison of distance measures in cluster analysis with dichotomous data. Journal of Data Science 3(1): 85-100.
14. Kaufman L, Rousseeuw P J (1990) Finding Groups in Data "An Introduction to Cluster Analysis", John Wiley and Sons.