International Journal of Statistical Distributions and Applications
2022; 8(4): 65-79
http://www.sciencepublishinggroup.com/j/ijsda
doi: 10.11648/j.ijsd.20220804.11
ISSN: 2472-3487 (Print); ISSN: 2472-3509 (Online)

Comparison Between Principal Components and Factor Analysis for Different Data

Ahmed Mohamed Mohamed Elsayed

Department of Basic Science, Al-Obour High Institute for Management, Computers and Informatics, Obour City, Egypt

Email address:

atabl18@yahoo.com

To cite this article

Ahmed Mohamed Mohamed Elsayed. Comparison between Principal Components and Factor Analysis for Different Data. *International Journal of Statistical Distributions and Applications.* Vol. 8, No. 4, 2022, pp. 65-79. doi: 10.11648/j.ijsd.20220804.11

Received: February 21, 2022; **Accepted**: March 21, 2022; **Published**: December 29, 2022

Abstract: Factor analysis (FA) is similar to the principal component analysis (PCA), but not the same. PCA can be considered as a basic of FA. PCA and FA aim to reduce dimension of a data, but the techniques are different. FA is clearly designed to identify the latent factors from the observed variables, PCA does not directly apply this aim. Eigenvalues of PCA are dispersed component loadings, with variance errors. FA assumes that the covariation of observed variables is due to the presence of latent variables. In contrast, PCA not depends on such causal relationship. If the factor model is incorrectly, then FA will give error results. PCA employs a transformation of the original data with no assumptions about the covariance matrix. PCA is used to determine linear combinations of the original variables and summarize the data set without losing information. For these reasons, we compared practically between FA and PCA using three different types of data. One of them is simulated data, and others are real data. R program is used for analysis the data, using suitable different packages and functions. Results are presented graphically and tabulated for the purposes of comparison. An obtained results interested for each data with three criteria: FA criterion is used to specify whereas a two factors are sufficient or not; the SS loadings specified the factor is worth keeping; the observed correlations between all original variables high or low; the Cattell's scree test, says to drop all further components after starting at the elbow. The PCA criterion, is used to determine the variable's importance, which have high Eigenvalue. The VRPC criterion is used to determine the variables tend to suitable factor.

Keywords: Principal Component Analysis, Factor Analysis, Rotation Process, Scree Test, Varimax Rotation, Eigenvalues

1. Introduction

FA is similar to PCA, in that FA also involves linear combinations of variables. Different from PCA, FA is a focused on the off diagonal correlations, while PCA focuses on explaining the terms on the diagonal. However, PCA also tends to fit relatively well the off diagonal correlations. Results of PCA and FA are too similar, but this is not always. FA is generally used when the research purpose is detecting data structure. If the factor model is incorrectly formulated, then FA will give erroneous results. FA takes into account the random error, whereas PCA fails to do that. Brown [4] explains this point. There are many articles discussed FA and PCA from many sides. Abdi and Williams [1] presented the definition and rules of PCA. Bandalos and Boehm-Kaufman [2] presented four common misconceptions in exploratory factor analysis (EFA). Emmanuel et al. [8] presented the robust PCA. Whereas Bouwmans and Zahzah [3] presented the robust PCA via Principal Component Pursuit. Chachlakiset al. [6] explained the L1- norm tucker tensor decomposition. Courtney [7] determined the number of factors to retain in EFA. Forkman et al. [10] tested the PCA when variables are standardized. Garrido et al. [11] presented a new look at Horn's parallel analysis with ordinal variables. Psychological Methods. Giorgia [12] presented PCA for stock portfolio management. Guan and Dy [13] devoted the sparse probabilistic PCA. Journee et al. [14] generalized the power method for sparse PCA. Larsen and Warne [16] estimated confidence intervals for Eigenvalues in EFA. Ledesma and Valero-Mora [17] determined the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. Liao et al. [18] presented a discriminant analysis of principal components as a new method for the analysis of genetically structured populations. Markopoulos et al. [19] presented the efficient L1-norm PCA via Bit Flipping. Miranda et al. [20] presented a new routes from minimal approximation error to principal components.

Moghaddam et al. [21], presented exact and greedy algorithms for the spectral bounds for sparse PCA. Ren et al. [22] presented a non-negative matrix factorization for the robust extraction of extended structures. Ritter [23] presented a comparison of distribution free and non-distribution free methods in factor analysis. Ruscio and Roche [24] determined the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Tran and Formann [26] explained the performance of parallel analysis in retrieving unidimensionality in the presence of binary data. Warne and Larsen [27] evaluated a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. Zou and Xue [28] presented a PCA.

This paper compared practically between FA and PCA using three different types of data, simulated, Hunua and Iris data. Software R program is used for analysis, using suitable different packages and functions.

This paper is ordered as: Section 2 present the definition and attributes of PCA. Section 3 presents the definition and attributes of FA. Section 4 presents the differences between PCA and FA. Section 5 presents the numerical examples. Section 6 presents the results and discussion. Finally, section 7 presents the conclusion.

2. PCA

The PCA of a collection of points in a real coordinate space are a sequencors. PCA creates variables that are linear combinations of the original variables. The new variables have the property that the variables are all orthogonal. PCA transformation can be helpful as a pre-processing step before clustering. PCA is seeking to reproduce the total variable variance. PCA is generally preferred for purposes of data reduction, but not when the goal is to detect the latent construct or factors. A best fitting minimizes the average squared distance from the points to the line. These directions constitute an orthonormal basis. PCA is the process of computing the principal components and using them to perform a change of basis data, sometimes using only the first few principal components and ignoring the rest. PCA is used for making predictive models. First principal component can be defined as a direction that maximizes the variance of the projected data. The principal components are Eigenvectors of the data's covariance matrix. Principal components are often computed by Eigen decomposition of the data covariance matrix. PCA is also related to canonical correlation analysis (CCA). The CCA defines coordinate systems that optimally describe the cross-covariance between two datasets. First principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. Second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted from the points. Singular values in Σ are the square roots of the Eigenvalues of the matrix X^TX. Each Eigenvalue is proportional to the portion of the variance associated with each Eigenvector. Sum of all the Eigenvalues is equal to the sum of the squared distances of the points from their multidimensional mean. PCA rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible into the first few dimensions. Remain values tend to be small and may be dropped with minimal loss of information. PCA is sensitive to the scaling of the variables. If we have just two variables and they have the same sample variance and are positively correlated, then PCA will rotatE 45° and the weights for the two variables with respect to the principal component will be equal. If we multiply all values of the first variable by 100°, then the first principal component will be almost the same with a small contribution from the other variable, whereas the second component will be almost aligned with the second original variable. This means that whenever the different variables have different units, PCA is a somewhat arbitrary method of analysis. Mean subtraction is necessary for performing PCA to ensure that the first principal component describes the direction of maximum variance. If mean subtraction is not performed, the first principal component might instead correspond more or less to the mean of the data. Mean of zero is needed for finding a basis that minimizes the mean square error of the approximation data. Mean centering is unnecessary if performing PCA on a correlation matrix, as the data are already centered after calculating correlations. Correlations are derived from the cross-product of two standard scores or statistical moments.

3. FA

The FA is used to describe variability among observed, correlated variables in terms of a potentially lower number of latent variables called factors. FA searches for such joint variations in response to latent variables. The observed variables are modelled as linear combinations of the potential factors plus error terms. FA is used in marketing, finance, and machine learning, and … etc.. It deal with data sets where there are large numbers of observed variables to reflect a smaller number of latent variables. FA is used to identify factors that explain a variety of results on different tests. FA in psychology is most often associated with intelligence research. However, it also has been used to find factors in a broad range of domains such as personality, attitudes, beliefs, etc. It is linked to psychometrics, as it can assess the validity of an instrument by finding if the instrument indeed measures the postulated factors. FA is a frequently used technique in cross cultural research. There are different types of factor analysis: Exploratory factor analysis (EFA): is used to identify complex interrelationships among items. Researcher makes no a priori assumptions about relationships among factors. Confirmatory factor analysis (CFA): is a more complex approach that tests the hypothesis that the items are associated with specific factors. CFA uses SEM to test a measurement model whereby loading on the factors allows for evaluation of relationships between observed variables and latent variables. SEM approaches can accommodate measurement error and are less restrictive than LSE. There are types of factor extraction: PCA is a method for factor extraction. The weights are computed to extract the maximum possible variance. The factor model must then be rotated for analysis. Canonical factor analysis (CAFA) is a different method of computing the same model as PCA, which uses the principal axis method. CAFA seeks factors have the highest canonical correlation with the observed variables. Common factor analysis (COFA) also principal axis factoring, it seeks the fewest factors which can account for the common variance (correlation) of a set of variables. Image factoring (IF) is based on the correlation matrix of predicted variables rather than actual variables, where each variable is predicted from the others using multiple regression. Alpha factoring (AF) is based on maximizing the reliability of factors assuming variables are randomly sampled from a universe of variables. Factor regression model (FRM) is a combinatorial model of factor model and regression model. It can be viewed as the hybrid factor model, whose factors are partially known.

There are some of terminology: Communality is the square of the standardized outer loading of an item. The squared factor loading is the percent of variance. To get the percent of variance in all the variables accounted for by each factor, add the sum of the squared factor loadings for that factor and divide by the number of variables. This is the same as dividing the factor's Eigenvalue by the number of variables. Sum of the squared factor loadings for all factors for a given variable is the variance in that variable accounted for by all the factors. Communality measures the percent of variance in a given variable explained by all the factors jointly. If the communality exceeds 1, there is a spurious solution, which may reflect too small sample or the choice to extract too many/few factors. Eigenvalues measure the amount of variation in the total sample accounted for by each factor. Ratio of Eigenvalues is the ratio of explanatory importance of the factors with respect to the variables. If a factor has a low Eigenvalue, then it is contributing little to the explanation of variances and may be ignored as less important than the factors with higher Eigenvalues. Component is retained if the associated Eigenvalue is bigger than the 95th percentile of the distribution of Eigenvalues derived from the random data. Kaiser criterion: Kaiser rule [15] is used to drop all components with Eigenvalues under 1, this being the Eigenvalue equal to the information accounted for by an average single item. Variation of this method has been created where a researcher calculates confidence intervals for each Eigenvalue and retains only components which have the entire confidence interval greater than 1. Cattell scree (test) plots the components as the X-axis and the corresponding Eigenvalues as the Y-axis. As one moves to the right, toward later components, Eigenvalues drop. When the drop ceases and the curve makes an elbow toward less steep decline, Cattell's scree test says to drop all further components after the one starting at the elbow [5]. Variance explained criterion: Some researchers simply use the rule of keeping enough factors to account for 90% of the variation. The purpose of a rotation is to produce factors with a mix of high, low and moderate loadings. From a mathematical viewpoint, there is no difference between a rotated and un-rotated matrix. Fitted model is the same, the uniquenesses are the same, and the proportion of variance explained is the same. Un-rotated output maximizes variance accounted for by the first and subsequent factors, and forces the factors to be orthogonal. Rotation serves to make the output more understandable: Pattern of loadings where each item loads strongly on only one of the factors, and much more weakly on the other factors. Rotations can be orthogonal or oblique. Varimax rotation is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor. Each factor will tend to have either large or small loadings of any particular variable. Varimax solution yields results which make it as easy as possible to identify each variable with a single factor. Orthogonality (independence) of factors is often an unrealistic assumption. Oblique rotations are inclusive of orthogonal rotation. It is allowing for factors that are correlated with one another. One may examine both a pattern matrix and a structure matrix. Structure matrix is simply the factor loading matrix as in orthogonal rotation. Pattern matrix contains coefficients which just represent unique contributions. More factors, Lower pattern coefficients. For oblique rotation, the researcher looks at both the structure and pattern coefficients when attributing a label to a factor. Quartimax rotation is an orthogonal alternative that minimizes the number of factors. This often generates a general factor on which most variables are loaded to a high or medium degree. Equimax rotation is a compromise between Varimax and Quartimax criteria. Promax rotation is an alternative non-orthogonal (oblique) rotation method. Higher-order factor analysis is a consisting of repeating steps factor analysis Oblique rotation factor analysis of rotated factors.

There are many advantages of FA: Reduction of number of variables, by combining two or more variables into a single factor. Identification of groups of inter related variables. Also, there are some disadvantages of FA: Each orientation is equally acceptable mathematically. Interpreting FA is based on using a heuristic, which is a solution that is convenient even if not true. More than one interpretation can be made of the same data factored the same way, and factor analysis cannot identify causality.

4. Differences Between PCA and FA

Fabrigar et al. [9] address a number of reasons used to suggest that PCA is not equivalent to FA: PCA is computationally quicker and requires fewer resources than FA. PCA and FA can produce similar results. Where data correspond to assumptions of the model, the results of PCA are inaccurate results. Researchers gain extra information from a PCA approach. Differences between PCA and FA are illustrated by Suhr [25]: PCA results in principal components that account for a maximal amount of variance for observed variables. FA accounts for common variance in the data. PCA inserts ones on the diagonals of the correlation matrix. FA adjusts the diagonals of the correlation matrix with the unique factors. PCA minimizes the sum of squared perpendicular distance to the component axis. FA estimates factors that influence responses on observed variables. The component scores in PCA represent a linear combination of the observed variables weighted by Eigenvectors. Observed variables in FA are linear combinations of the underlying and unique factors. Explained from PCA perspective, but it is not from FA perspective.

To compute the factor score for a given case for a given factor, one takes the case's standardized score on each variable, multiplies by the corresponding loadings of the variable for the given factor, and sums these products. Computing factor scores allows one to look for factor outliers. Also, factor scores may be used as variables in subsequent modeling. Researchers wish to avoid such subjective or arbitrary criteria for factor retention. A number of objective methods have been developed to solve this problem, allowing users to determine an appropriate range of solutions to investigate. The parallel analysis may suggest 5 factors while Velicer's MAP suggests 6 factors, so the researcher may request both 5 and 6 factor solutions and discuss each in terms of their relation to external data and theory.

5. Numerical Examples

FA is introduced by the 'factana' function of 'stats' package. The function performs maximum-likelihood factor analysis on a covariance matrix. The number of factors to be fitted is specified by the argument factors. Moreover, by the additional argument rotation the transformation of the factors may be specified by either 'Varimax', 'Oblique' or 'None' rotation. In addition to the data set the 'factanal' function requires an estimate of the number of factors. If we have a hypothesis about the latent variables, we may start with an informed guess. If we do not have any clue about the latent variables, we may start with an informed guess. If we do not have any clue about the number of factors and the number of variables in the data set is not too large, one may simply try out several values for initializing the model. Another, more sophisticated approach is to use PCA to get a good initial estimate of the number of factors. In the numerical examples, we just make a guess and set the number of factor to be. Further, we keep the defaults for the scores "none" and the rotation "Varimax". FA creates linear combinations of factors to abstract the variable's underlying communality. To the extent that the variables have underlying communality, fewer factors capture most of the variance in the data set. This allows us to aggregate a large number of observable variables in a model to represent an underlying concept, making it easier to understand the data. Variability in our data, X , is given by Σ and its estimate $\hat{\Sigma}$ is composed of the variability explained by the factors explained be a linear combination of the factors (communality) and of the variability, which can not be explained by a linear combination of the factors (uniqueness).

$$\hat{\Sigma} = \underbrace{\hat{\mathrm{A}}\hat{\mathrm{A}}^{T}}_{communality} + \underbrace{\hat{\Psi}}_{uniqueness} \tag{1}$$

The model output starts with the function call to remind us on the specifications of our function call. First chunk provides the 'uniquenesses', which range from 0 to 1. Uniqueness, sometimes referred to as 'noise' corresponds to the proportion of variability, which can not be explained by a linear combination of the factors. This is $\hat{\Psi}$ the in Equation (1) above. High uniqueness for a variable indicates that the factors do not account well for its variance. The 'loadings', which range from -1 to 1, $\hat{\mathrm{A}}$ in Equation (1) above. Loadings are the contribution of each original variable to the factor. Variables with a high loading are well explained by the factor. There is no entry for certain variables. That is because R does not print loadings less than 0.1. This is meant to help us spot groups of variables. By squaring the loading we compute the fraction of the variable's total variance explained by the factor. This proportion of the variability is denoted as communality. Another way to calculate the communality is to subtract the 'uniquenesses' from the 'observed correlations'. An appropriate factor model results in low values for uniqueness and high values for communality.

The loadings shows the proportion of variance explained by each factor. The row 'cumulative var' gives the cumulative proportion of variance explained. These numbers range from 0 to 1. The row 'proportion var' gives the proportion of variance explained by each factor, and the row 'SS loadings' gives the sum of squared loadings. This is sometimes used to determine the value of a particular factor. A factor is worth keeping if the 'SS loading' is greater than 1 (Kaiser's rule). The last section of the function output shows the results of a hypothesis test. Null hypothesis, H_0, is that the number of factors in the model is sufficient to capture the full dimensionality of the data set. Conventionally, we reject H_0 if the p-value is less than 0.05. Such a result indicates that the number of factors is too small. In contrast, we do not reject H_0 if the p-value exceeds 0.05. Such a result indicates that there are likely enough (or more than enough) factors capture the full dimensionality of the data set. High p-value in our example above leads us to not reject the H_0, and indicates that we fitted an appropriate model. This hypothesis test is available thanks to our method of estimation, maximum likelihood. Note that if you provide a covariance matrix to the 'factanal' function and not a data frame, the hypothesis test is not provided if we do not explicitly provide the number of observations as an additional argument to the function call. Using our model (1), we may calculate $\hat{\Sigma}$ and compare it to the observed correlation matrix (S) by simple matrix algebra. We now subtract the fitted correlation matrix ($\hat{\Sigma}$) from (S). The resulting matrix is called the residual matrix. Numbers close to 0 indicate that our factor model is a good representation of the underlying concept.

Let us apply a factor analysis on three types of different data. First one is generated or simulated data. Second is Hunua data. Last one is Iris data. We will discuss these data sets in details in the next subsections for two methods; present the FA and PCA for the types data as following:

5.1. Simulated Data

In this subsection, we generated a multivariate binary data for five independent variables using the marginals respectively (0.4, 0.6, 0.5, 0.3, 0.8) with n = 500 observations. for each binary variables. Then we get the correlation matrix for the binary variables to generate a multivariate normal data for five correlated variables, with $n = 500$ observations. The next results come from FA and PCA respectively as shown below:

Using R program, the function 'factanal', and the packages 'stats, nFactors', we have the next results:

5.1.1. Factor Analysis for Simulated Data

Table 1. *Maximum likelihood FA for simulated data.*

Variables	x1	x2	x3	x4	x5	SS loadings	Proportion var.	Cumulative var.
Uniqueness	**0.982**	**0.956**	**0.957**	**0.980**	**0.005**			
Communality	**0.018**	**0.044**	**0.043**	**0.020**	**0.995**			
Factor1	-	-	-	-	0.997	1.004	0.201	0.201
Factor2	0.112	0.200	0.204	0.140	-	0.115	0.023	0.024

Chi square statistic is 0.75 on 1 degree of freedom. P-value is 0.386. Two factors are sufficient for the simulated data. From Table 1 a high uniqueness for variables x1, x2, x3 and x4 indicate that the factors do not account well for its variance. But a low uniqueness for the variable x5 indicates that the factors do account well for its variance. Also, the variables x1, x2, x3 and x4 with a high loading are well explained by factor2. But the variable x5 explained by factor1. An appropriate factor model results in low values for uniqueness and high values for communality, a factor2 is good for the simulated data. Factor1 is worth keeping if the 'SS loading' is greater than 1.

Table 2. *Correlations (S) for simulated data.*

Variables	x1	x2	x3	x4	x5
x1	1.000000	0.016844	0.008003	0.044034	-0.069885
x2	0.016844	1.000000	0.049388	0.013203	0.064135
x3	0.008003	0.049388	1.000000	0.027447	-0.029514
x4	0.044034	0.013203	0.027447	1.000000	0.004473
x5	-0.069885	0.064135	-0.029514	0.004473	1.000000

The observed correlations between all original variables are too low.

Table 3. *Residual matrix ($S - \hat{\Sigma}$) for simulated data.*

Variables	x1	x2	x3	x4	x5
x1	0.000008	-0.001363	-0.017543	0.028352	-3.0e-06
x2	-0.001363	-0.000002	0.010597	-0.014813	1.0e-06
x3	-0.017543	0.010597	0.000003	-0.001081	-1.0e-05
x4	0.028352	-0.014813	-0.001081	-0.000002	1.5e-05
x5	-0.000003	0.000001	-0.000010	0.000015	0.0e+00

We subtracted the fitted correlation matrix $\hat{\Sigma}$ from our observed correlation matrix S. The resulting matrix is called the residual matrix. Numbers close to 0 indicate that our factor model is a good representation. Let us fit two factor models, one with 'no' rotation, one with 'Varimax' rotation, and one with 'Promax' rotation, and make a scatter plot of the first and second loadings.

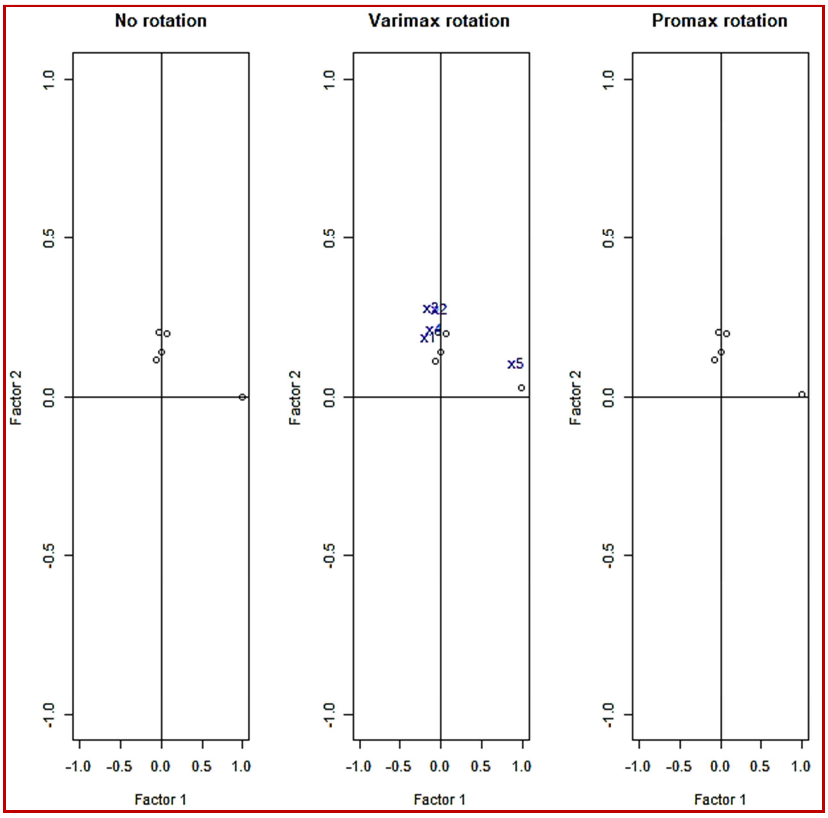

Figure 1. *Rotation processes for simulated data.*

It is clear that the variables x1, x2, x3, x4 tend to factor2, we can put them into factor2, whereas the variable x5 tend to factor1. This is too clear using 'Varimax' rotation.

Table 4. *Number of factors to extract for simulated data.*

Optimal coordinates	Acceleration factor	Parallel	Kaiser (Eigenvalues)
0	2	5	2

Number of factors in Acceleration factor is 2 factors and it suitable for simulated data with 5 variables.

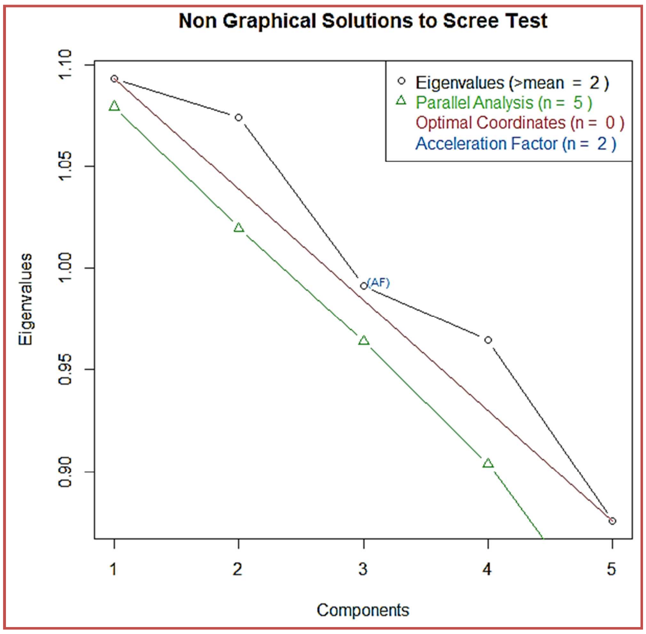

Figure 2. *Scree plot for simulated data.*

Cattell's scree test says to drop all further components (Fourth and Fifth) after (Third) starting at the elbow.

5.1.2. PCA for Simulated Data

Using R program, the function 'princomp', we have the next results:

Table 5. *Eigenvalues for simulated data.*

X1	X2	X3	X4	X5
1.0934456	1.0746981	0.9915222	0.9646287	0.8757054

The variables x1 and x2 are good and have high Eigenvalues.

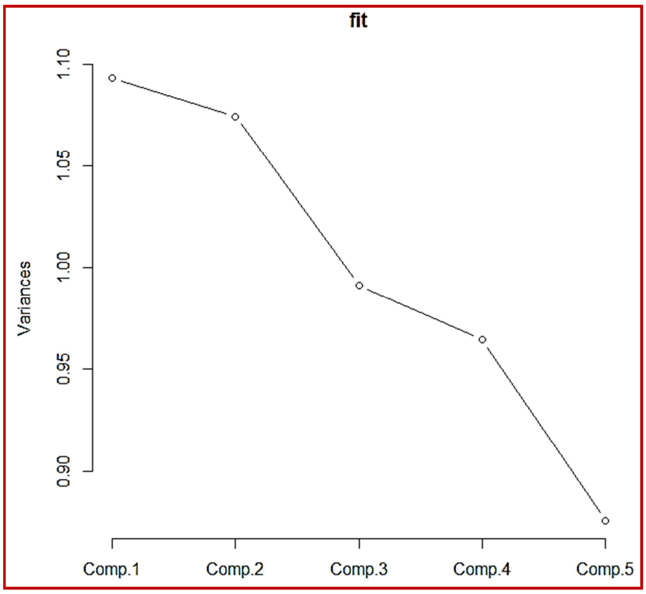

Figure 3. *Variances of components for simulated data.*

Table 6. *Importance of components for simulated data.*

Components	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.0456795	1.0366764	0.9957521	0.9821552	0.9357913
Proportion of Variance	0.2186891	0.2149396	0.1983044	0.1929257	0.1751411
Cumulative Proportion	0.2186891	0.4336287	0.6319332	0.8248589	1.0000000

Table 7. *Loadings for simulated data.*

Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
x1	0.631	-	0.288	0.548	0.461
x2	-0.116	0.709	-	0.482	-0.496
x3	0.289	0.468	-0.665	-0.357	0.358
x4	0.336	0.369	0.605	-0.583	-0.211
x5	-0.627	0.367	0.322	-	0.607

Components	Comp.5
Loadings	1.0
Proportion var.	0.2
Cumulative var.	1.0

5.1.3. VRPC for Simulated Data

Using R program, the function 'principal', and the package 'psych', we have the next results:

Table 8. *Standardized loadings for simulated data: Pattern matrix based upon correlation matrix.*

Components	RC1	RC2
SS loadings	1.09	1.08
Proportion var.	0.22	0.22
Cumulative var.	0.22	0.43
Proportion Explained	0.50	0.50
Cumulative Proportion	0.50	1.00

Mean item complexity = 1.4. Two components are not sufficient. The root mean square of the residuals (RMSR) is 0.22 with the empirical chi square 501.9 with P. value < 3.7e-111. Fit based upon off diagonal values = -31.29.

For factor analysis 'Varimax' rotation, the simulated data, the variables x1, x2, x3 and x4 have high loadings for the factor2, then they have something in common. Figure 4 appears that factor1 accounts x5. Whereas factor1 accounts for x1, x2, x3 and x4.

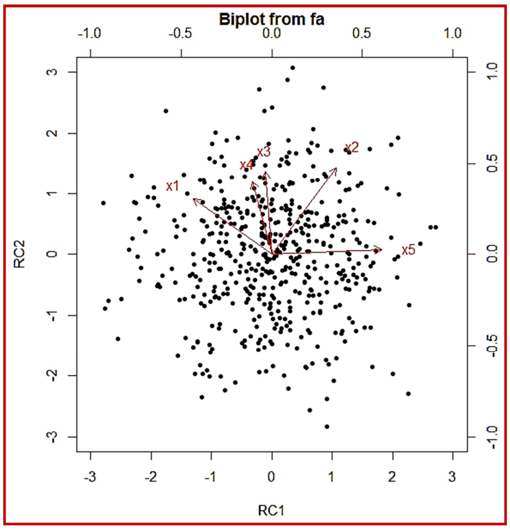

Figure 4. *Biplot of FA for simulated data.*

5.2. Hunua Data

These were collected from the Hunua Ranges, a small forest in southern Auckland, New Zealand. At 392 sites in the forest, the presence/absence of 17 plant species was recorded, as well as the altitude. Each site was of area size 200 m^2.

5.2.1. FA for Hunua Data

Using R program, the function 'factanal', and the packages 'stats, nFactors', we have the next results:

From Table 9, a high uniqueness for all variables except "daccup" indicate that the factors do not account well for its variance. Also, a " daccup " with a high loading is well explained by the factor2. Since an appropriate factor model results in low values for uniqueness and high values for communality, the factor2 is good for the variable ' daccup'. From "SS loadings", factor1 is worth keeping.

Table 9. *Maximum likelihood FA for hunua data.*

Variables	**agaaus**	**beitaw**	**corlae**	**cyadea**	**cyamed**	**daccup**	**dacdac**	**eladen**	**hedarb**
Uniqueness	**0.839**	**0.620**	**0.944**	**0.894**	**0.839**	**0.005**	**0.874**	**0.994**	**0.834**
Communality	**1.608442e-01**	**3.802777e-01**	**5.649836e-02**	**1.058853e-01**	**1.605772e-01**	**9.950003e-01**	**1.257536e-01**	**6.114275e-03**	**1.662007e-01**
Factor1	-0.399	0.532	0.230	0.307	0.394	-0.509	-0.109	-	0.398
Factor2	-	0.312	-	0.107	-	0.858	0.337	-	-

Variables	**hohpop**	**kniexc**	**kuneri**	**lepsco**	**metrob**	**neslan**	**rhosap**	**vitluc**
Uniqueness	**0.988**	**0.919**	**0.648**	**0.919**	**1.000**	**0.975**	**0.718**	**0.945**
Communality	**1.224671e-02**	**8.129281e-02**	**3.515201e-01**	**8.087977e-02**	**1.190216e-06**	**2.489148e-02**	**2.824767e-01**	**5.504454e-02**
Factor1	-	0.140	-0.539	-0.118	-	0.117	0.513	0.222
Factor2	-	0.248	-0.247	-0.259	-	0.106	0.140	-

Variables	**SS loadings**	**Proportion Var**	**Cumulative Var**
Factor1	1.834	0.108	0.108
Factor2	1.212	0.071	0.179

Chi square statistic is 302.78 on 103 degrees of freedom. P-value is 1.53e-21. Two factors are not sufficient for Hunua data.

Table 10. *S for hunua data.*

Variables	**hohpop**	**kniexc**	**kuneri**	**lepsco**	**metrob**	**neslan**	**rhosap**	**vitluc**
agaaus	-0.037114	0.143306	0.186334	-0.017856	0.008629	0.019321	-0.198111	-0.151687
beitaw	-0.059785	0.279432	-0.423300	-0.203572	0.077050	0.168472	0.201053	0.120120
corlae	-0.018288	-0.066591	-0.074363	-0.062272	0.012127	-0.052680	0.163486	0.369843
cyadea	-0.050153	0.069583	-0.077047	-0.076422	-0.008431	0.020159	0.284249	-0.028213
cyamed	-0.043033	0.058431	-0.231632	-0.071216	-0.014488	0.022051	0.303075	0.053919
daccup	-0.047565	0.141540	0.062414	-0.161963	0.000631	0.030829	-0.140517	-0.047588
dacdac	0.062603	-0.031807	-0.046198	-0.110715	-0.008399	0.009087	0.025297	0.113552
eladen	-0.014284	0.086389	-0.061576	-0.048639	0.111585	-0.041147	-0.100211	0.007058
hedarb	-0.029235	0.090943	-0.251613	-0.099549	0.046529	0.100006	0.158368	-0.082933
hohpop	1.000000	-0.012614	0.098250	-0.017462	-0.013263	-0.014772	-0.035977	-0.034257
kniexc	-0.012614	1.000000	-0.106616	-0.110477	0.068639	-0.010152	0.111684	0.015548
kuneri	0.098250	-0.106616	1.000000	0.008554	0.074519	-0.069097	-0.339457	-0.142166
lepsco	-0.017462	-0.110477	0.008554	1.000000	-0.045161	-0.050301	-0.122504	-0.088176
metrob	-0.013263	0.068639	0.074519	-0.045161	1.000000	0.033797	-0.093045	0.021192
neslan	-0.014772	-0.010152	-0.069097	-0.050301	0.033797	1.000000	0.089205	0.000676
rhosap	-0.035977	0.111684	-0.339457	-0.122504	-0.093045	0.089205	1.000000	0.217076
vitluc	-0.034257	0.015548	-0.142166	-0.088176	0.021192	0.000676	0.217076	1.000000

The observed correlations between all species of hunua data is low.

Table 10. *Continued.*

Variables	agaaus	beitaw	corlae	cyadea	cyamed	daccup	dacdac	eladen	hedarb
agaaus	1.000000	-0.191522	-0.106310	-0.203506	-0.198297	0.238344	-0.152379	-0.038280	-0.193740
beitaw	-0.191522	1.000000	0.132864	0.143659	0.144296	-0.003215	0.006641	0.130807	0.296369
corlae	-0.106310	0.132864	1.000000	0.047602	0.015248	-0.169624	-0.031151	-0.050940	-0.043441
cyadea	-0.203506	0.143659	0.047602	1.000000	0.255477	-0.064669	-0.058388	0.001805	0.179530
cyamed	-0.198297	0.144296	0.015248	0.25547	1.000000	-0.135862	-0.057952	0.000615	0.150970
daccup	0.238344	-0.003215	-0.169624	-0.064669	-0.135862	1.000000	0.345391	-0.045930	-0.128802
dacdac	-0.152379	0.006641	-0.031151	-0.058388	-0.057952	0.345391	1.000000	-0.055241	-0.107897
eladen	-0.038280	0.130807	-0.050940	0.001805	0.000615	-0.045930	-0.055241	1.000000	0.032573
hedarb	-0.193740	0.296369	-0.043441	0.179530	0.150970	-0.128802	-0.107897	0.032573	1.000000
hohpop	-0.037114	-0.059785	-0.018288	-0.050153	-0.043033	-0.047565	0.062603	-0.014284	-0.029235
kniexc	0.143306	0.279432	-0.066591	0.069583	0.058431	0.141540	-0.031807	0.086389	0.090943
kuneri	0.186334	-0.423300	-0.074363	-0.077047	-0.231632	0.062414	-0.046198	-0.061576	-0.251613
lepsco	-0.017856	-0.203572	-0.062272	-0.076422	-0.071216	-0.161963	-0.110715	-0.048639	-0.099549
metrob	0.008629	0.077050	0.012127	-0.008431	-0.014488	0.000631	-0.008399	0.111585	0.046529
neslan	0.019321	0.168472	-0.052680	0.020159	0.022051	0.030829	0.009087	-0.041147	0.100006
rhosap	-0.198111	0.201053	0.163486	0.284249	0.303075	-0.140517	0.025297	-0.100211	0.158368
vitluc	-0.151687	0.120120	0.369843	-0.028213	0.053919	-0.047588	0.113552	0.007058	-0.082933

The observed correlations between all species of hunua data is low.

Table 11. *Residual matrix ($S - \hat{\Sigma}$) for hunua data.*

Variables	hohpop	kniexc	kuneri	lepsco	metrob	neslan	rhosap	vitluc
agaaus	-0.057972	0.189017	-0.018559	-0.054250	0.008681	0.061749	0.000643	-0.066279
beitaw	0.001709	0.127419	-0.059542	-0.060124	0.076590	0.073174	-0.115394	-0.021676
corlae	-0.009758	-0.083508	0.034220	-0.051116	0.012138	-0.073105	0.054401	0.323567
cyadea	-0.021385	0.000036	0.114992	-0.012602	-0.008618	-0.027134	0.111687	-0.104556
cyamed	-0.011849	-0.015340	-0.000946	-0.005425	-0.014662	-0.032005	0.090762	-0.039145
daccup	-0.000189	-0.000159	-0.000027	0.000094	-0.000160	-0.000109	0.000026	0.000038
dacdac	0.086826	-0.100223	-0.021889	-0.036314	-0.008732	-0.013703	0.034086	0.112161
eladen	-0.010128	0.077238	-0.021354	-0.041297	0.111574	-0.049526	-0.139136	-0.009683
hedarb	0.003284	0.013691	-0.015547	-0.030264	0.046341	0.044193	-0.057991	-0.177923
hohpop	0.000000	0.018843	0.042292	-0.048510	-0.013150	0.002163	0.008555	-0.013564
kniexc	0.018843	0.000000	0.030242	-0.029694	0.068341	-0.052800	0.005011	-0.034455
kuneri	0.042292	0.030242	0.000000	-0.118896	0.074912	0.020165	-0.028496	-0.003747
lepsco	-0.048510	-0.029694	-0.118896	0.000001	-0.044857	-0.009154	-0.025774	-0.042315
metrob	-0.013150	0.068341	0.074912	-0.044857	0.000000	0.033656	-0.093318	0.021058
neslan	0.002163	-0.052800	0.020165	-0.009154	0.033656	0.000000	0.014300	-0.033377
rhosap	0.008555	0.005011	-0.028496	-0.025774	-0.093318	0.014300	0.000000	0.092631
vitluc	-0.013564	-0.034455	-0.003747	-0.042315	0.021058	-0.033377	0.092631	0.000000

Table 11. *Continued.*

Variables	agaaus	beitaw	corlae	cyadea	cyamed	dacdac	dacdac	eladen	hedarb
agaaus	0.000000	0.007870	-0.012183	-0.085224	-0.044327	0.000081	-0.209895	-0.006922	-0.038321
beitaw	0.007870	0.000000	0.029907	-0.053129	-0.088471	-0.000184	-0.040383	0.091600	0.057521
corlae	-0.012183	0.029907	0.000000	-0.016443	-0.070534	-0.000044	0.014710	-0.069256	-0.129629
cyadea	-0.085224	-0.053129	-0.016443	0.000000	0.126462	0.000377	-0.060684	-0.021381	0.047834
cyamed	-0.044327	-0.088471	-0.070534	0.126462	0.000000	0.000196	-0.040141	-0.029508	-0.012343
daccup	0.000081	-0.000184	-0.000044	0.000377	0.000196	0.000000	0.000315	-0.000213	-0.000006
dacdac	-0.209895	-0.040383	0.014710	-0.060684	-0.040141	0.000315	-0.000001	-0.044326	-0.093376
eladen	-0.006922	0.091600	-0.069256	-0.021381	-0.029508	-0.000213	-0.044326	0.000000	0.002157
hedarb	-0.038321	0.057521	-0.129629	0.047834	-0.012343	-0.000006	-0.093376	0.002157	0.000000
hohpop	-0.057972	0.001709	-0.009758	-0.021385	-0.011849	-0.000189	0.086826	-0.010128	0.003284
kniexc	0.189017	0.127419	-0.083508	0.000036	-0.015340	-0.000159	-0.100223	0.077238	0.013691
kuneri	-0.018559	-0.059542	0.034220	0.114992	-0.000946	-0.000027	-0.021889	-0.021354	-0.015547
lepsco	-0.054250	-0.060124	-0.051116	-0.012602	-0.005425	0.000094	-0.036314	-0.041297	-0.030264
metrob	0.008681	0.076590	0.012138	-0.008618	-0.014662	-0.000160	-0.008732	0.111574	0.046341
neslan	0.061749	0.073174	-0.073105	-0.027134	-0.032005	-0.000109	-0.013703	-0.049526	0.044193
rhosap	0.000643	-0.115394	0.054401	0.111687	0.090762	0.000026	0.034086	-0.139136	-0.057991
vitluc	-0.066279	-0.021676	0.323567	-0.104556	-0.039145	0.000038	0.112161	-0.009683	-0.177923

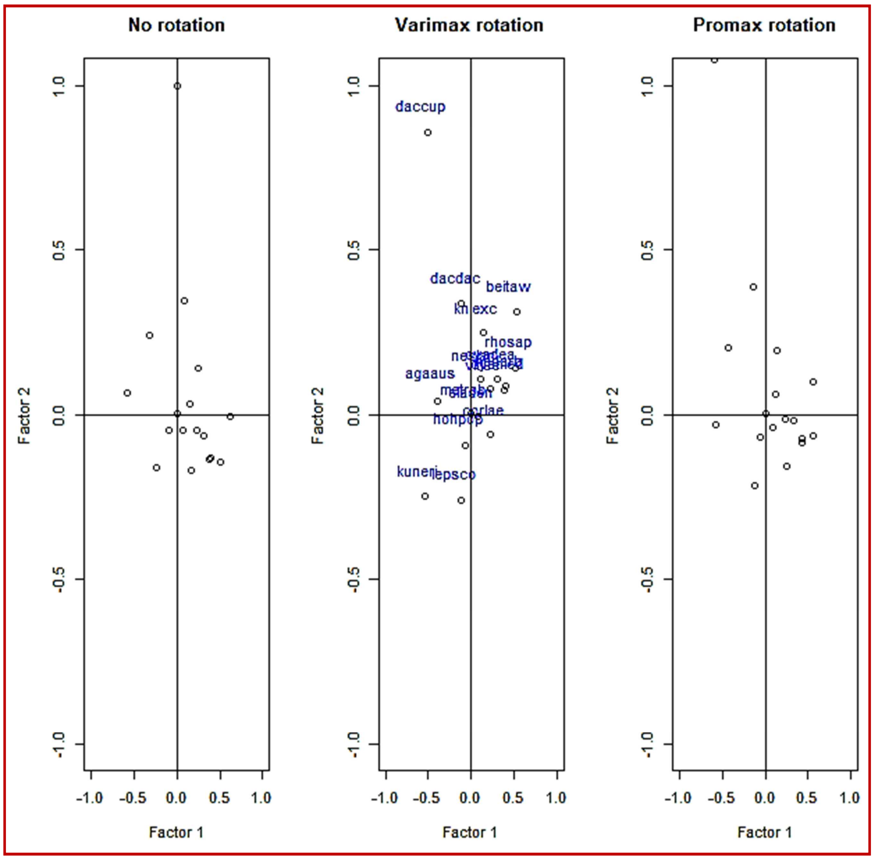

__Figure 5.__ Rotation processes for hunua data.

The variable 'daccup' tend to factor2. The rest variables tend to factor1.

__Table 12.__ Number of factors for to extract hunua data.

Optimal coordinates	Acceleration factor	Parallel	Kaiser (Eigenvalues)
5	1	7	7

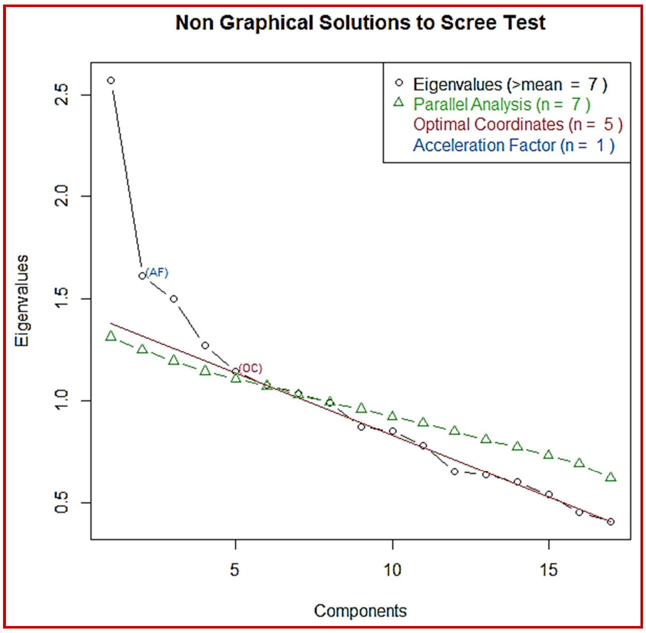

__Figure 6.__ Scree plot for hunua data.

Cattell's scree test says to drop all further components (Third to Seventeenth) after (Second) starting at the elbow.

5.2.2. PCA for Hunua Data

Using R program, the function 'princomp', we have the next results:

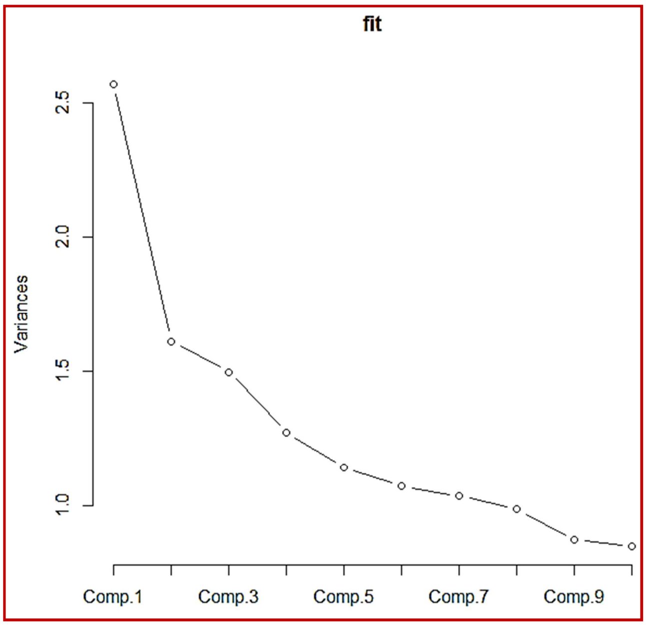

__Figure 7.__ Variances of components for hunua data.

Table 13. *Eigenvalues for hunua data.*

agaaus	beitaw	corlae	cyadea	cyamed	daccup'	dacdac	eladen	hedarb
2.5683761	1.6116994	1.4990880	1.2728054	1.1421284	1.0750427	1.0377820	0.9895705	0.8744913

Species 'agaaus, beitaw, corlae, cyadea, cyamed, daccup and dacdac' have high Eigenvalues.

Table 13. *Continued.*

agaaus	hohpop	kniexc	kuneri	lepsco	metrob	neslan	rhosap	vitluc
2.5683761	0.8529391	0.7791596	0.6546264	0.6399470	0.6012691	0.5394572	0.4520016	0.4096160

Species 'agaaus, beitaw, corlae, cyadea, cyamed, daccup and dacdac' have high Eigenvalues.

Table 14. *Importance of components for hunua data.*

Variables	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17
Standard deviation	0.92354700	0.88270019	0.80908988	0.79996687	0.77541545	0.73447752	0.67231066	0.64001253
Proportion of Variance	0.05017289	0.04583292	0.03850744	0.03764394	0.03536877	0.03173278	0.02658833	0.02409506
Cumulative Proportion	0.76023076	0.80606368	0.84457112	0.88221506	0.91758383	0.94931661	0.97590494	1.00000000

Table 14. *Continued.*

Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation	1.6026154	1.26952725	1.22437251	1.12818678	1.06870408	1.03684266	1.0187158	0.99477157	0.93514241
Proportion variance	0.1510809	0.09480585	0.08818165	0.07487091	0.06718402	0.06323781	0.0610460	0.05821003	0.05144067
Cumulative proportion	0.1510809	0.24588680	0.33406845	0.40893935	0.47612338	0.53936118	0.6004072	0.65861721	0.71005787

Table 15. *Loadings for hunua data.*

Variables	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17
agaaus	0.105	0.137	0.241	0.357	0.210	-	0.393	0.303
beitaw	-0.126	-0.137	0.284	-0.268	-0.256	-0.187	-	0.602
corlae	-0.170	-0.177	0.416	-	-	0.489	-0.111	-0.180
cyadea	-	-0.585	0.191	0.139	-	-0.235	0.287	-
cyamed	0.334	0.510	0.263	-0.389	0.248	0.145	-	-
daccup	-	-	0.260	-	0.338	-0.155	-0.562	-
dacdac	-	-	-	-	-	0.408	0.441	0.169
eladen	0.567	-0.154	-	0.210	0.143	0.198	-	-
hedarb	-0.450	-	-0.212	0.123	0.586	0.193	-	-
hohpop	-	-	0.198	0.107	0.116	-0.114	-	-
kniexc	-0.102	-0.208	-0.338	-0.302	-	0.222	-	-0.280
kuneri	-	-0.180	-0.301	-0.326	0.102	0.203	-0.205	0.487
lepsco	-	-0.353	0.113	-	0.261	-	-0.101	0.202
metrob	-0.202	0.214	-	0.239	-0.107	-	-	-
neslan	0.459	-0.199	-	-0.120	-	0.111	-	-0.170
rhosap	0.123	-	-0.288	0.528	-0.203	0.195	-0.385	0.259
vitluc	0.127	-	-0.346	-	0.420	-0.460	0.151	-

Table 15. *Continued.*

Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
agaaus	0.301	0.257	-	0.132	0.529	-	0.113	0.114	-
beitaw	-0.399	0.295	-	0.205	-	-0.196	-	0.117	-
corlae	-0.179	-0.266	-0.403	0.352	0.175	-	0.166	-	-0.139
cyadea	-0.282	-	0.163	-0.242	-	0.461	-	-0.261	-
cyamed	-0.324	-	0.126	-0.213	-	0.262	-	-0.114	0.228
daccup	0.162	0.538	-0.240	-0.226	-	-	-0.108	-0.135	-
dacdac	-	0.260	-0.476	-0.322	-0.384	-	-0.157	-0.103	0.130
eladen	-	0.134	0.172	0.446	-0.339	-	-0.357	-	-0.215
hedarb	-0.305	-	0.344	-	-0.175	-0.131	0.138	-	-0.237
hohpop	-	-	-	-	-0.393	0.225	0.300	0.751	0.194
kniexc	-0.134	0.428	0.114	0.178	0.286	0.221	-0.147	0.306	0.339
kuneri	0.395	-	-	-	-	0.391	0.304	-0.147	-
lepsco	0.144	-0.346	0.177	-	-	-0.366	-0.300	-	0.578
metrob	-	0.147	-	0.424	-0.287	0.166	0.284	-0.404	0.522
neslan	-0.101	0.173	-	-0.106	-0.238	-0.474	0.622	-0.119	-
rhosap	-0.405	-	-0.123	-0.214	0.238	0.103	-	-	0.149
vitluc	-0.198	-0.113	-0.532	0.284	-	-	-	-	0.104

Variables	Comp.17
SS loadings	1.00
Proportion var.	0.059
Cumulative var.	1.00

5.2.3. VRPC for Hunua Data

Using R program, the function 'principal', and the package 'psych', we have the next results:

Table 16. *Standardized loadings for hunua data.*

Components	RC1	RC2
SS loadings	2.5	1.68
Proportion var.	0.15	0.10
Cumulative var.	0.15	0.25
Proportion Explained	0.6	0.40
Cumulative Proportion	0.6	1.00

Mean item complexity = 1.3. Tow components are not sufficient. The root mean square of the residuals (RMSR) is 0.09. with the empirical chi square 868.16 with prob < 2.5e-121. Fit based upon off diagonal values = 0.51.

For factor analysis 'Varimax' rotation for hunua data, the variable 'daccup' has a high loadings for the factor2. Figure 8 appears that factor1 and factor2 account almost all variables.

5.3. Iris Data

The obtained results are:

5.3.1. FA for Iris Data

Using R program, the function 'factanal', and the packages 'stats, nFactors', we have the next results:

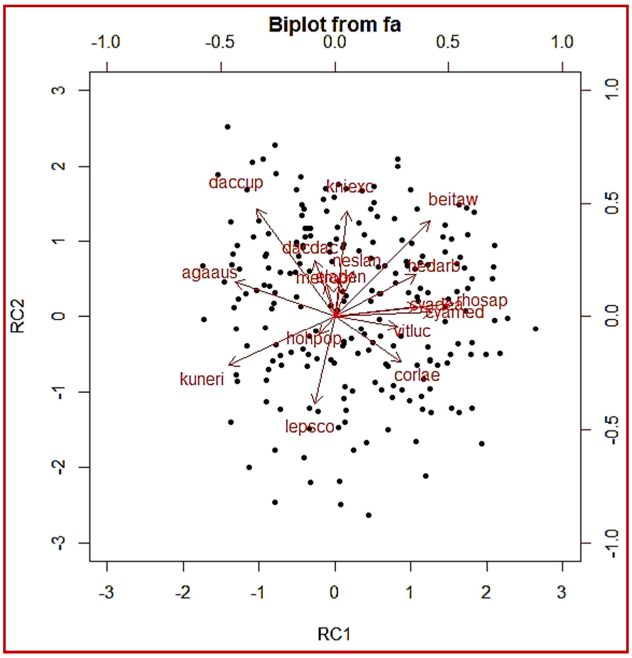

Figure 8. *Biplot of FA for hunua data.*

Table 17. *Maximum likelihood FA for iris data.*

Variables	Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species	SS loadings	Proportion	Cumulative
Uniqueness	0.005	0.614	0.022	0.043	0.054			
Communality	0.995	0.386	0.978	0.957	0.946			
Factor1	0.997	-0.108	0.865	0.810	0.774	3.010	0.602	0.602
Factor2	-	-0.612	0.480	0.548	0.590	1.252	0.250	0.852

Chi square statistic is 44.93 on 1 degree of freedom. P-value is 2.05e-11. Two factors are not sufficient for Iris data.

From Table 17 a high uniqueness for "Sepal. Width" indicate that the factors do not account well for its variance. The remain variables have low uniquenesses, this indicate that the factors do account well for its variance. Also, a "Sepal. Width" with a high loading is well explained by the factor2. An appropriate factor model results in low values for uniqueness and high values for communality, the factor2 is good for "Sepal. Width". From "SS loadings", a factor1 is worth keeping.

Table 18. *S for iris data.*

Variables	Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species
Sepal. Length	1.000000	-0.117570	0.871754	0.817941	0.782561
Sepal. Width	-0.117570	1.000000	-0.428440	-0.366126	-0.426658
Petal. Length	0.871754	-0.428440	1.000000	0.962865	0.949035
Petal. Width	0.817941	-0.366126	0.962865	1.000000	0.956547
Species	0.782561	-0.426658	0.949035	0.956547	1.000000

The observed correlations between original variables are good.

Table 19. *Residual matrix ($S - \hat{\Sigma}$) for iris data.*

Variables	Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species
Sepal. Length	-0.000021	0.001692	0.000249	-0.000289	-0.000007
Sepal. Width	0.001692	0.000022	-0.041454	0.056566	0.017665
Petal. Length	0.000249	-0.041454	0.000000	-0.000719	-0.003019

Variables	Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species
Petal. Width	-0.000289	0.056566	-0.000719	0.000000	0.006672
Species	-0.000007	0.017665	-0.003019	0.006672	0.000000

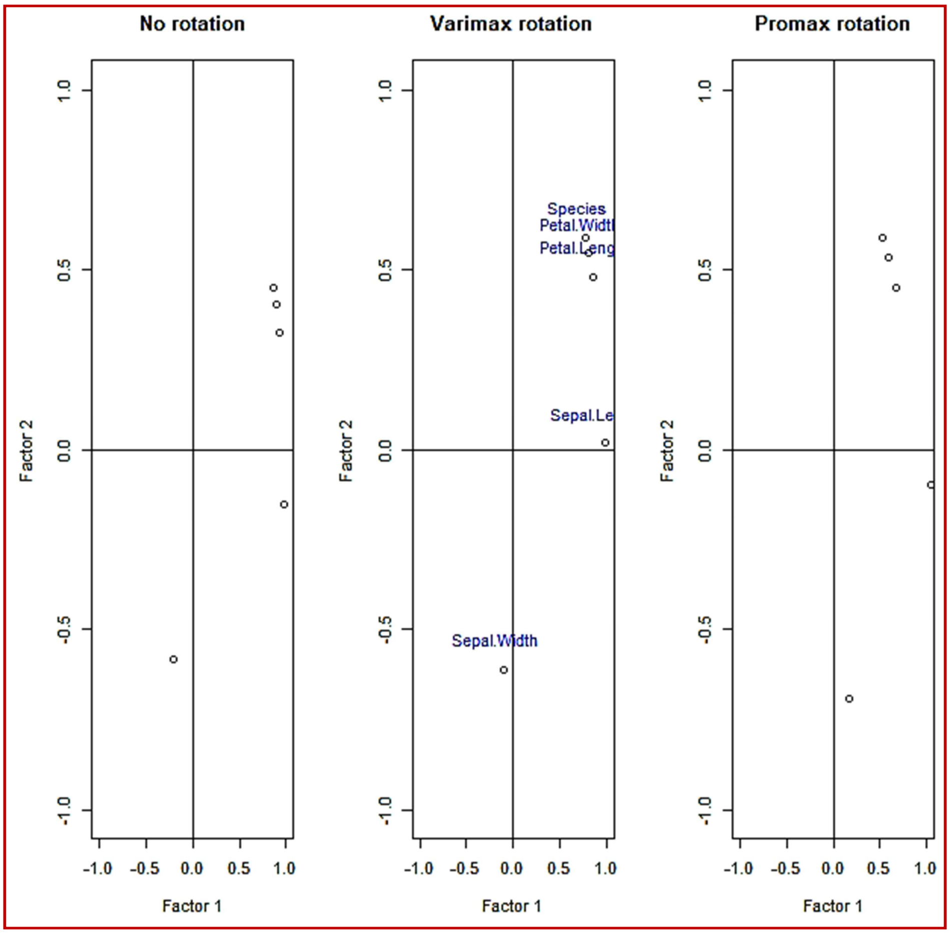

Figure 9. *Rotation processes for iris data.*

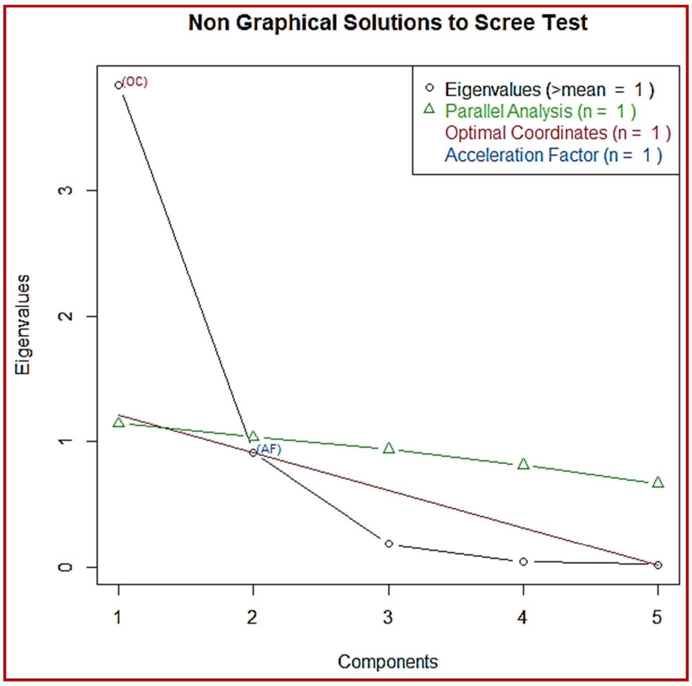

Figure 10. *Scree plot for iris data.*

Cattell's scree test says to drop all further components (Third to Fifth) after (Second) starting at the elbow.

Table 20. *Number of factors to extract for iris Data.*

Optimal coordinates	Acceleration f.	Parallel	Kaiser (Eigenvalues)
1	1	1	1

5.3.2. PCA for Iris Data

Using R program, the function 'princomp', we have the next results:

Table 21. *Eigenvalues for iris data.*

Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species
3.83701790	0.91413636	0.18622615	0.04208608	0.02053351

Variable 'Sepal. Length' has high Eigenvalue.

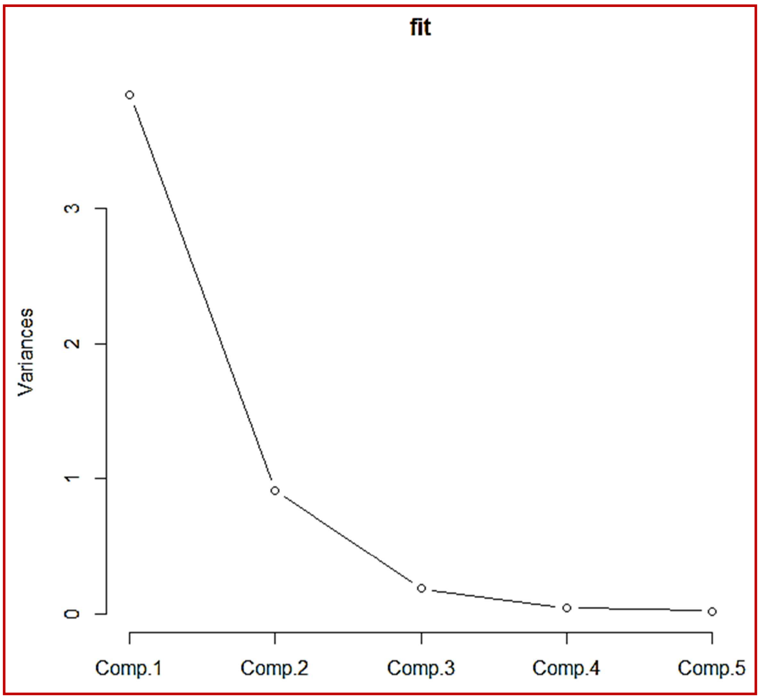

Figure 11. *Variances of components for iris data.*

Table 22. *Importance of components for iris data.*

Components	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.9588307	0.9561048	0.43153927	0.205148913	0.143295198
Proportion of Variance	0.7674036	0.1828273	0.03724523	0.008417215	0.004106703
Cumulative Proportion	0.7674036	0.9502309	0.98747608	0.995893297	1.000000000

Table 23. *Loadings for iris data.*

Variables	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Sepal. Length	0.445	0.382	0.751	0.141	0.270
Sepal. Width	-0.233	0.921	-0.287	-	-0.122
Petal. Length	0.506	-	-	-0.243	-0.827
Petal. Width	0.497	-	-0.385	-0.613	0.474
Species	0.495	-	-0.452	0.739	-

Components	Comp.5
Loadings	1.0
Proportion var.	0.2
Cumulative var.	1.0

5.3.3. VRPC for Iris Data

Using R program, the function 'principal', and the package 'psych', we have the next results:

Table 24. *Standardized loadings for iris data.*

Components	RC1	RC2
SS loadings	3.55	1.20
Proportion var.	0.71	0.24
Cumulative var.	0.71	0.95
Proportion Explained	0.75	0.25
Cumulative Proportion	0.75	1.00

Mean item complexity = 1.1. Two components are sufficient. The root mean square of the residuals (RMSR) is 0.03 with the empirical chi square 2.82 with prob < 0.093. Fit based upon off diagonal values = 1.

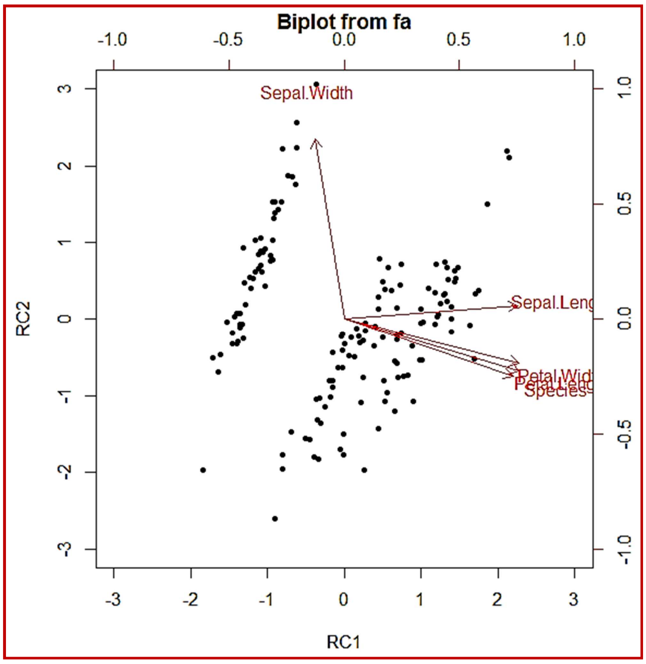

Figure 12. Biplot of FA for iris data.

For factor analysis 'Varimax' rotation for iris data, the variable 'Sepal. Width' has a high loadings for the factor2. Figure 12 appears that factor1 accounts all variables except the variable 'Sepal. Width'.

6. Results and Discussion

From the obtained results from tables 1 to 24, we have:

For simulated data: FA: Two factors are sufficient for the simulated data. From 'SS loading', Factor1 is worth keeping. The observed correlations between the original variables are too low. Number of factors in Acceleration factor is 2 factors of 5 variables. Cattell's scree test says to drop all further components (Fourth and Fifth) after (Third) starting at the elbow. PCA: The variables x1 and x2 are good and have high Eigenvalues. VRPC: Mean item complexity = 1.4. Two components are not sufficient. The root mean square of the residuals (RMSR) is 0.22 with the empirical Chi square 501.9 with P. value < 3.7e-111. Fit based upon off diagonal values = -31.29. For factor analysis 'Varimax' rotation, the variables x1, x2, x3, x4 tend to factor2, whereas the variable x5 tend to factor1.

For hunua data: FA: Two factors are not sufficient for Hunua data. From "SS loadings", factor1 is worth keeping. The observed correlations between all species of hunua data is low. Number of factors in Acceleration factor is 1 factor of 17 variables. Cattell's scree test says to drop all further components (Third to Seventeenth) after (Second) starting at the elbow. PCA: Species 'agaaus, beitaw, corlae, cyadea, cyamed, daccup and dacdac' have high Eigenvalues.

VRPC: Mean item complexity = 1.3. Tow components are not sufficient. The root mean square of the residuals (RMSR) is 0.09. with the empirical chi square 868.16 with P. value < 2.5e-121. Fit based upon off diagonal values = 0.51. For factor analysis 'Varimax' rotation, the variable 'daccup' tend to factor2. The rest 16 variables tend to factor1.

For iris data: FA: Two factors are not sufficient for iris data. From "SS loadings", a factor1 is worth keeping. The observed correlations between all original variables are good. Number of factors in Acceleration factor is 2 factors with 5 variables. Cattell's scree test says to drop all further components (Third to Fifth) after (Second) starting at the elbow. PCA: Variable 'Sepal. Length' has high Eigenvalue. VRPC: Mean item complexity = 1.1. Two components are sufficient. The root mean square of the residuals (RMSR) is 0.03 with the empirical chi square 2.82 with P. value < 0.093. Fit based upon off diagonal values = 1. It is clear that the variables 'Sepal. Length, Petal. length, Petal. Width and Species' tend to factor1, whereas the variable 'Sepal. Width' tend to factor2. This is too clear using 'Varimax' rotation.

7. Conclusion

In this paper, the characteristics of principal component analysis and the factor analysis techniques and their attributes are presented. Also, the similarity and difference between the two techniques are explained. The practical comparison between the two techniques are presented using three types of data. R program is used for analysis, using suitable different packages and functions. The results are presented graphically and tabulated form, for the process of comparison. The obtained results interested for each data with three criteria. FA criterion is used to specify whereas two factors are sufficient or not; from SS loadings, we specified the factor is worth keeping; The observed correlations between all original variables high or low; Cattell's scree test says to drop all further components, after starting at the elbow. PCA criterion is used to determine the variable's importance, which have high Eigenvalue. VRPC criterion is used to determine the variables tend to suitable factor.

References

[1] Abdi, H. and Williams, J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2 (4): pp. 433–459. doi: 10.1002/wics.101.

[2] Bandalos, L. and Boehm-Kaufman, R. (2008). Four common misconceptions in exploratory factor analysis. In Lance, Charles E.; Vandenberg, Robert J. (eds.). Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in the Organizational and Social Sciences. Taylor & Francis. pp. 61–87. ISBN 978-0-8058-6237-9.

[3] Bouwmans, T. and Zahzah, E. (2014). Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. Computer Vision and Image Understanding 122: pp. 22–34. doi: 10.1016/j.cviu.2013.11.009.

[4] Brown, D. (2009). Principal components analysis and exploratory factor analysis – Definitions, differences and choices. Shiken: JALT Testing & Evaluation SIG Newsletter.

[5] Cattell, R. B. (1978). Use of factor analysis in behavioral and life sciences. New York: Plenum.

[6] Chachlakis, G., Prater-Bennette, A. and Markopoulos, P. (2019). L1-norm Tucker Tensor Decomposition. IEEE Access. 7: pp. 178454–178465. doi: 10.1109/ACCESS.2019.2955134.

[7] Courtney, R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. Practical Assessment, Research and Evaluation, 18 (8). Available online: http://pareonline.net/getvn.asp?v=18&n=8

[8] Emmanuel, J., Xiaodong, L., Yi, M. and John, W. (2011). Robust principal component analysis". Journal of the ACM. 58 (3): doi: 10.1145/1970392.1970395.

[9] Fabrigar et al. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods.

[10] Forkman J., Josse, J. and Piepho, P. (2019). Hypothesis tests for principal component analysis when variables are standardized. Journal of Agricultural, Biological, and Environmental Statistics. 24 (2): pp. 289–308. doi: 10.1007/s13253-019-00355-5.

[11] Garrido, L., Abad, J. and Ponsoda, V. (2012). A new look at Horn's parallel analysis with ordinal variables. Psychological Methods. Advance. online publication. doi: 10.1037/a0030005.

[12] Giorgia, P. (2017). Principal component analysis for stock portfolio management. International Journal of Pure and Applied Mathematics 115 (1).

[13] Guan, Y. and Dy, J. (2009). Sparse probabilistic principal component analysis. Journal of Machine Learning Research Workshop and Conference Proceedings 5: p. 185.

[14] Journee, M., Nesterov, Y., Richtarik, P. and Sepulchre, R. (2010). Generalized power method for Sparse principal component analysis. Journal of Machine Learning Research. 11: pp. 517–553.

[15] Kaiser, F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement. 20 (1): pp. 141–151. doi: 10.1177/001316446002000116.

[16] Larsen, R. and Warne, T. (2010). Estimating confidence intervals for Eigenvalues in exploratory factor analysis. Behavior Research Methods. 42 (3): pp. 871–876. doi: 10.3758/BRM.42.3.871.

[17] Ledesma, R. and 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): pp. 1–11.

[18] Liao, T, Jombart, S., Devillard, F. and Balloux (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations" doi: 10.1186/1471-2156-11-94.

[19] Markopoulos, P., Kundu, S., Chamadia, S. and Pados, A. (2017). Efficient L1-Norm principal-component analysis via bit flipping". IEEE Transactions on Signal Processing. 65 (16): pp. 4252-4264. doi: 10.1109/TSP.2017.2708023.

[20] Miranda, A., Borgne, Y. and Bontempi, G. (2008). New routes from minimal approximation error to principal components, neural processing letters, Springer 27.

[21] Moghaddam, B., Weiss, Y. and Avidan, S. (2005). Spectral bounds for Sparse PCA: Exact and greedy algorithms. Advances in Neural Information Processing Systems 18. MIT Press.

[22] Ren, B., Pueyo, L., Zhu, B. and Duchêne, G. (2018). Non-negative matrix factorization: Robust extraction of extended structures". The Astrophysical Journal. 852 (2): p. 104. doi: 10.3847/1538-4357/aaa1f2.

[23] Ritter, N. (2012). A comparison of distribution-free and non-distribution free methods in factor analysis. Paper presented at Southwestern Educational Research Association (SERA) Conference 2012, New Orleans, LA (ED529153).

[24] Ruscio, J. and Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychological Assessment. 24 (2): pp. 282–292. doi: 10.1037/a0025697. PMID 21966933.

[25] Suhr, D. (2009). Principal component analysis vs. exploratory factor analysis. SUGI 30 Proceedings.

[26] Tran, S., and Formann, K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. Educational and Psychological Measurement, 69, pp. 50-61.

[27] Warne, T. and Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. Psychological Test and Assessment Modeling. 56: pp. 104–123.

[28] Zou, H. and Xue, L. (2018). A selective overview of Sparse principal component analysis: pp. 1311–1320. doi: 10.1109/JPROC.2018.2846588.