Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

UNIVERSAL WISER PUBLISHER

Research Article

Robust Dimensionality Reduction: A Bootstrap-Based Evaluation of PCA with Applications in Nutritional and Environmental Sciences

Zakiah I. Kalantan[1*], **Lujain S. Alharbi**[1], **Maryam H. Al-Zahrani**[2], **Sulafah M. Saleh Binhimd**[1]

[1] Department of Statistics, Faculty of Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[2] Department of Biochemistry, Faculty of Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
E-mail: zkalanten@kau.edu.sa

Received: 4 November 2024; **Revised:** 2 January 2025; **Accepted:** 15 January 2025

Abstract: The complex structures of vast amounts of data provide a considerable challenge to researchers. Dimensional reduction methods are reliable and transform high-dimensional data into lower-dimensional representations while preserving most original information. Principal Component Analysis (PCA) is a commonly used approach for dimensionality reduction that transforms data into a lower-dimensional space while preserving important information. The variability within the sample may affect the stability and reliability of PCA results.The constraint of existing approaches compromises the accuracy of PCA stability assessments in practical data scenarios. These methodologies frequently depend on linear assumptions and encounter difficulties when addressing high-dimensional data. This study used the bootstrap method to assess the stability of PCA by assessing the variability of eigenvalues and principal components over several bootstrap iterations. We evaluate how stability metrics, particularly confidence intervals for eigenvalues and the proportion of variance clarified, can assist in determining the optimal number of principal components. The results indicate that the bootstrap provides a helpful framework for evaluating the robustness of PCA and guiding informed decisions on dimensionality reduction in many applications, including data compression, visualization, and classification. Moreover, the results illustrate the efficacy of this method in enhancing the reliability and interpretability of PCA findings among distinct data-driven research endeavors. This study enhances understanding of how principal component analysis (PCA) tackles data unpredictability while delivering valuable insights for professionals in several disciplines.

Keywords: big data, data visualization, principal component analysis, bootstrap, stability

MSC: 62H25, 62P10, 62F40

1. Introduction

In contrast to data visualization's ability to synthesize data for increased understanding, big data analysis is troubled by the "curse of dimensionality", the phenomenon in which determining complexity and data sparsity grow exponentially with an increasing number of characteristics [1], causing significant computing challenges, increased possibility of overfitting, and reduced advantages when including additional features. This sparsity requires exponentially more data to sustain sufficient data density, resulting in heightened computing expenses and the potential for models to overfit noise and include extraneous information. Furthermore, the immediate improvement of model performance may not be achieved

DOI: https://doi.org/10.37256/cm.6120256016

by incorporating supplementary characteristics; in reality, it may be reduced by the introduction of redundancy and noise. Dimensionality reduction techniques transform high-dimensional data into lower-dimensional subspaces, preserving important information and alleviating the challenges posed by the curse of dimensionality. These methodologies diminish computing complexity, augment generalization, and boost interpretability by emphasizing the most pertinent characteristics and mitigating the influence of noise. Researchers use several machine learning approaches to identify patterns and underlying dimensions within large datasets, hence facilitating the implementation of dimensionality reduction strategies. These procedures may be classified as supervised and unsupervised, with unsupervised methods concentrating on recognizing similarities and disparities within the dataset.

Principal Component Analysis (PCA) is among the most often used approaches for dimension reduction [2]. This is an unsupervised method aimed at reducing the dimensionality of a dataset by identifying the main components that indicate the directions of maximum variation in the data. PCA preserves the most variance in the original dataset while minimizing the number of variables. Fisher and Mackenzie [3] first presented it in the context of modeling response data. They proposed that PCA was more appropriate for this purpose than analysis of variance. The authors also delineated the Non-linear Iterative Partial Least Squares (NIPALS) algorithm, a technique for executing PCA. Hotelling [4] advanced PCA to its present iteration. PCA has been reintroduced and used throughout several scientific domains, including facial recognition, handprint identification, and mobile robots. Rao [5] significantly advanced the subject of PCA by presenting novel concepts for its use, interpretation, and extension.

PCA has different drawbacks that need other approaches. This involves dependence on algorithms, and which can produce biases and lead to inaccurate results. Various modifications of traditional PCA have been developed to overcome its shortcomings [6]. For instance, probabilistic PCA (PPCA), presented by Tipping and Bishop in 1999, and generalized PCA (GPCA), defined by Vidal et al. [7]. This work utilizes a Gaussian mixture model (GMM) to estimate the probability density function (PDF) of process data under typical operating circumstances. Training a Gaussian Mixture Model (GMM) using the expectation-maximization (EM) approach on process data might be arduous or infeasible for high-dimensional and collinear process variables. Xu et al. [8] proposed an innovative multimode process monitoring method with a PCA mixture model to resolve this problem. Current methodologies exhibit sensitivity to noise and outliers, which might affect the stability and selection of appropriate principal components. This sensitivity constrains their effectiveness in practical situations characterized by poor data quality. Certain current methodologies emphasize evaluating the statistical significance of principal components (PCs) while neglecting their interpretability, therefore obstructing comprehension of the underlying data structures and diminishing the practical applicability of principal component analysis (PCA).

Assessing the statistical relevance of PCs and offering meaningful interpretations are essential for researchers. computer complexity is a substantial obstacle to real-time data processing and applications constrained by restricted computer resources. Researchers want precise and effective techniques to enhance the accessibility of PCA for extensive data analysis. In order to prevent erroneous interpretations and draw precise conclusions, it is imperative that personal computers are stable. The optimal quantity of principal components effectively encapsulates significant data variance while reducing noise and preventing overfitting; inadequate principal components lead to the omission of essential information, whilst an abundance adds noise and complicates interpretation. Improved methods for assessing PCA stability and determining suitable principal components are necessary to improve the reliability, interpretability, and efficacy of PCA among diverse practical applications. Researchers can improve the efficacy of PCA for data analysis and predictive modeling by addressing the drawbacks of existing methodologies and using more robust and efficient techniques [6, 8].

According to Efron [9, 10], bootstrapping is a resampling technique that entails the repetitive extraction of samples from a dataset, with replacement, in order to estimate the distribution of a statistic. This method facilitates the calculation of confidence intervals, standard errors, and bias correction independent of parametric assumptions on the underlying population. Numerous statistical environments have made use of the bootstrap method, such as testing of hypotheses, regression analysis, and confidence interval construction. It additionally has been used as a tool for data analysis using clustering and integrating models [11, 12]. Binhimd and Coolen [13] have developed a novel bootstrap technique known as NPI-B.

They contrasted this approach with Efron's traditional bootstrap method (Ef-B) and found that NPI-B exhibited more reproducibility. Binhimd and Almalki [14] have examined the comparison of Ef-B and NPI-B using various methodologies. In PCA, bootstrapping may provide confidence intervals for the loadings of the principal components, enabling researchers to evaluate the reliability of the associations between original variables and the principal components. Bootstrap methodologies may assist in assessing the stability of the components. Through repeated resampling and the execution of PCA, one may ascertain the number of components that consistently account for the variation across several samples. The bootstrap distribution-free resampling technique [9] is often used to assess the variance of estimators or to establish tolerance zones in graphical representations derived from principal axis methods such as principal component analysis (PCA) [9]. Daudin et al. [15] tackled the issue of identifying the ideal number of components and proposed confidence intervals for locations inside the subspace defined by the primary axes. This research examined diverse methodologies, including those potentially constrained by the use of duplicated samples in the resampling procedure [15]. Linting et al. [16] used the nonparametric bootstrap method to evaluate the stability of nonlinear PCA outcomes. They used confidence ellipses for eigenvalues, component loadings, and individual scores, as well as confidence intervals for variable transformations. Also examined techniques to enhance stability, including Procrustes rotation, bias estimation, and the balanced bootstrap.

To evaluate the stability of their nonlinear PCA method against that of linear PCA, they used an identical bootstrap procedure on linear PCA using the same dataset. Enhancing the accuracy of the bootstrap approach may be achieved by augmenting the quantity of bootstrapped samples. Nonetheless, this will also elevate the computational expense. In practical PCA applications, determining the number of components to retain is a critical issue. This entails evaluating the stability of the components or variables. In statistical modeling, stability is the characteristic whereby a dataset originating from the same underlying distribution yields relatively consistent parameters. This indicates that a significant and legitimate parameter (component) should not readily vanish when the dataset experiences superfluous alterations [17]. Certain stability criteria use computationally intensive methods, such as cross-validation, bootstrap, or jackknife procedures. Bootstrapping has been used in the literature to estimate the confidence area for a differentiable function of the covariance matrix of the original data after the application of PCA to the examined dataset. The study findings include considerable practical consequences, as they provide a dependable approach for determining the ideal amount of components to maintain in PCA, thereby improving the dependability of the results.

This work presents new stability criteria for Principal Component Analysis (PCA) based on a risk function that measures the distance between orthogonal projectors. This invention considerably improves the field by enabling a more comprehensive evaluation of PCA stability, therefore augmenting the efficiency and effectiveness of this prevalent approach. The efficacy of our recommended techniques is meticulously assessed via the analysis of real-world datasets. A fundamental component of this approach, the bootstrap method substantially improves the strength of PCA findings. Bootstrapping reduces the effect of small fluctuations and noise by constantly resampling the data and averaging the results. Hence, it is especially helpful for datasets that are either restricted or erratic. This method quickly solves the need for exact and consistent PCA findings, which are necessary for effective data interpretation and analysis.

2. Materials and methods

This section provides an overview of the techniques used in this study and a theoretical introduction to Principal Component Analysis (PCA), a widely used technique for dimensionality reduction. The bootstrap technique is a resampling procedure that is used to evaluate the stability and variability of statistical estimates, particularly those derived using principal component analysis (PCA). These methods are discussed in Sections 2.1 and 2.2. Section 2.3 delineates the bootstrap method approaches, focusing on the critical stages of the bootstrap process. It employs confidence intervals and examines the differences between bootstrap trials to determine stability. Section 2.4 lays out the steps for using a bootstrap stability assessment to perform PCA. These steps include preparing the data and running PCA.

2.1 *Principal component analysis*

PCA is a dimension reduction method that projects data with high-dimensional space into a lower-dimensional subspace. The new data can then be more easily visualized and analyzed while retaining as much of the original data's variation as possible. This has applications in various fields, such as face recognition and image compression, and it is a common technique for finding patterns in high-dimensional data.

Hence, let $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ be a random vector that has a probability density function, $f(x)$, with the mean and variance denoted as μ and Σ, respectively. Assume that a sample of size N is drawn from the random vector, $\boldsymbol{X}$, yielding the data $\boldsymbol{Z} \in R^{N \times D}$, where $x_i = (x_{i1}, \ldots, x_{iD})$ for $i = 1, \ldots, N$. The data matrix, $\boldsymbol{Z}$, has the following structure:

$$\boldsymbol{Z} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1j} & \cdots & x_{1D} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2j} & \cdots & x_{2D} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3j} & \cdots & x_{3D} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{Nj} & \cdots & x_{ND} \end{bmatrix} \tag{1}$$

The main idea of PCA is to illustrate the variation that appears in a dataset that has correlated variables, $\boldsymbol{X^T} = (X_1, \ldots, X_D)$, and to define a new dataset of uncorrelated variables, $\boldsymbol{Y^T} = (Y_1, \ldots, Y_D)$, where each Y_i, $i = 1, \ldots, D$, is a linear combination of X. The variables in the new data are the PCs, and the first PC Y_1, is the following linear combination:

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1D}X_D. \tag{2}$$

Among all possible linear combinations of the variables, Y_1 has the maximum variance, which could increase by increasing the coefficients $\boldsymbol{a_1} = (a_{11}, a_{12}, \cdots, a_{1D})$, this yields to build a constraint on these coefficients. A reasonable constraint is to require that $a_1^\top a_1 = 1$. The variance of Y_1 is a linear function of X variables, which is given as $a_1^\top S a_1$, and, $\boldsymbol{S}$ is the $D \times D$ covariance matrix of $\boldsymbol{X}$. The Lagrange multiplier is used to maximize a function of multiple variables subject to one or more constraints. Without loss of generality, it is assumed that the mean of the data is zero, i.e., $\bar{x}_i = \frac{\sum x_{ij}}{N}$(for $j = 1, \ldots, D$), where $\bar{x}_i$ represent the mean of the i-th variable.

The second PC, Y_2, is defined as the following linear combination:

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2D}X_D \tag{3}$$

i.e., $Y_2 = a_2X$, where $\boldsymbol{a}_2 = (a_{21}, a_{22}, \cdots, a_{2D})$ and $\mathrm{X} = (X_1, X_2, \cdots, X_D)$. This achieves the highest variance while adhering to the specified limitations:

$$a_2^\top a_2 = 1,\ a_2^\top a_1 = 0 \tag{4}$$

and the total variance of the D principal components will equal the total variance of the original variables.

In PCA, squared factor loadings indicate the amount of variation in each original variable that a given principal component (PC) may elucidate. A greater squared loading signifies a more robust association between the variable and the principal component, suggesting that the principal component explains a bigger share of the variable's variation. Despite eigenvalues correlating to the variance that comprises each principal component, they do not directly indicate the proportion of the total variance in the dataset that a single principal component accounts for. To determine the proportion of total variance explained by a principal component (PC), divide the eigenvalue of the PC by the sum of all eigenvalues.

According to [18], the sum of squared loadings for a single variable across all PCs should be 1, which implies that the sum of all PCs accounts for all of the variable's variance. Although the research did not explicitly evaluate the assumption of multivariate normality for the data, it is crucial to recognize that the principal components acquired through PCA are statistically independent and uncorrelated under this assumption. This implies that the principal components are discrete and independent data variation sources. It is essential to recognize that while the principal components remain orthogonal (uncorrelated), they may not exhibit statistical independence without multivariate normality.

2.2 *Bootstrapping*

A bootstrap's primary tenet is that, in some circumstances, it is preferable to estimate a population parameter only from the available data without making any assumptions about underlying distributions. Compared to traditional inferences, bootstrap approaches can often be considerably more accurate. This method, suitable for a range of statistical inferences, is a valuable tool for small sample sizes, allowing for the estimation of sampling distributions and the application of various statistical techniques [10]. The issue with this method is that different findings can be received each time a sample is gathered. Theoretically, the standard deviation of a point estimate for repeated population samplings may be high, which could bias the estimate [19].

2.3 *Approaches using bootstrapping*

Bootstrapping is a resampling method that is employed to approximate the sampling distribution of a statistic. It comprises the repeated sampling of data with replacement from the original dataset. The bootstrap technique produces several samples and calculates the statistic of interest for each, yielding a subjective evaluation of the sampling distribution for that statistic. This reduces the need for assumptions regarding the underlying population distribution and facilitates the calculation of standard errors and the construction of confidence intervals. The theory of central limits is essential in several statistical inference methods; nonetheless, it is critical to recognize that the bootstrap method's validity is not directly dependent on it. The bootstrap offers a framework for estimating the sampling distribution of a statistic independent of the underlying population distribution [20–22].

The process of constructing a bootstrapped confidence interval is demonstrated below:

- Generate B bootstrap samples by sampling with replacement from the original dataset.
- Calculate the covariance matrix for each bootstrap sample.
- Calculate the maximum eigenvalue λ_i^* for each bootstrap covariance matrix. This yields a collection of B bootstrap maximum eigenvalues.
- Determine the difference (d_i) between the maximum eigenvalue of the original data λ_{true} and each of the B bootstrap maximum eigenvalues. $d_i = \lambda_{true} - \lambda_i^*$.
- Determine the 2.5 th percentile (a) and the 97.5 th percentile (b) of the distribution of the differences (d_i).
- The 95% confidence interval for the true maximum eigenvalue (λ_{true}) is expressed as:

$$(\lambda_{mean}^* - b,\ \lambda_{mean}^* + a)$$

where λ_{mean}^* is the mean of the B bootstrap maximum eigenvalues.

An abundance of stability metrics for PCA have been proposed in the literature. This study used the bootstrap method, a popular resampling technique, to evaluate the durability of PCA results. Other notable strategies include:

- Resampling techniques:
  - ○ The jackknife method, similar to the bootstrap procedure, involves the methodical exclusion of a single observation at a time, followed by the recalibration of the PCA.

○ The significance of patterns that appear in the PCA results may be evaluated using permutation tests, which involve the random rearrangement of data within groups or variables.

• Cross-validation methods:

○ K-fold cross-validation entails partitioning the dataset into k segments, doing PCA on k-1 segments, and assessing outcomes on the remaining segment. The procedure is executed k times, and the outcomes are averaged to evaluate the stability and generalization efficacy of the PCA model.

• Stability selection:

○ This methodology integrates resampling techniques, such as subsampling, with variable selection methods to discern stable features across various subsamples. This study examines the traditional bootstrap method's simplicity, broad applicability, and established utilization in diverse statistical contexts. Additionally, we examined the impact of the quantity of bootstrapped samples on the stability of PCA.

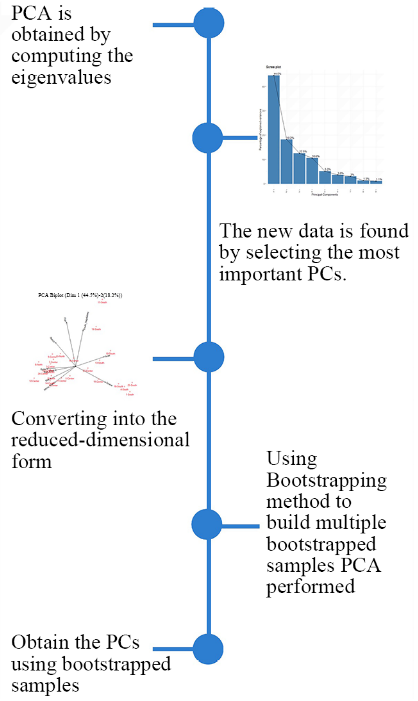
PCA is obtained by computing the eigenvalues
The new data is found by selecting the most important PCs.
Converting into the reduced-dimensional form
Using Bootstrapping method to build multiple bootstrapped samples PCA performed
Obtain the PCs using bootstrapped samples

Figure 1. The step of implementing PCA with bootstrapping

2.4 *Algorithm of PCA with bootstrapping*

This section illustrates the steps of the proposed method, as shown in Figure 1:

(1) Use PCA to lower the dimensionality of a p-dimensional dataset, starting with calculating the eigenvalues and eigenvectors from the covariance matrix to pinpoint the primary components.

(2) Select the m eigenvectors that fit the m largest eigenvalues, where m is the new feature subspace's $(m \leq p)$ number of dimensions; in this work, we used the scree plot to determine m.

(3) Transform the initial data set into a lower-dimensional representation.

(4) Use the bootstrap method to create several bootstrapped samples for PCA computation.

(5) Acquire the main components (eigenvectors) by employing bootstrapped samples.

(6) Determine the confidence interval for the principal components.

In evaluating the stability of PCA findings, the intervals used must demonstrate the following attributes:

1. Narrow intervals: The calculated eigenvalues, eigenvectors, and loadings suggest more confidence. This means that small changes have less of an effect on the PCA answer in the data.

2. Uniform Intervals: Consistency in the intervals between various bootstrap repetitions or resampling techniques is advantageous. The variability observed in these intervals indicates that the PCA solution is quite sensitive to slight changes in the data, highlighting its instability.

3. The intervals must provide enough coverage. Confidence intervals at the 95% level are anticipated to include the genuine population parameters about 95% of the time. This verifies that the intervals precisely represent the uncertainty linked to the PCA findings.

The stability of PCA results must be evaluated by maintaining narrow intervals, which suggest a high level of confidence in the estimated eigenvalues, eigenvectors, and loadings. A stable PCA solution is shown by bootstrap repeats with consistent intervals, whereas inconsistent intervals imply susceptibility to data fluctuations. The bootstrap method is beneficial for assessing the stability of PCA and determining the optimal number of components; however, it does not by nature produce “more accurate” results. It provides evaluations of variability and uncertainty. Cronbach’s alpha and McDonald’s omega are measures used to assess internal consistency reliability, measuring the extent to which items on a scale represent the same underlying construct.

Projecting the full data matrix onto the bootstrapped principal components produces row scores. Improving the bootstrap technique involves strategies like increasing the number of bootstrap samples, applying diverse resampling schemes, using various measures of variability, and performing statistical tests. Techniques include the application of eigenvalue shrinkage estimators and alternative principal component selection methods. The effectiveness of the bootstrap method is contingent upon the specific application and parameters, highlighting the necessity for a customized approach [23, 24].

This study utilized the traditional bootstrap method for stability assessment. This method entails repeated sampling with replacement from the original dataset to produce multiple bootstrap replicates. Observing the changes in eigenvalues, eigenvectors, and the explained variance ratio over repetitions is one approach to assessing the consistency of PCA results. Because of its simplicity, wide adaptability, and well-known use in several statistical contexts, the conventional bootstrap method was chosen for this investigation. However, future studies may consider other resampling techniques, such as the Nonparametric Predictive Inference Bootstrap (NPI-B) or the Jackknife. Future research may explore the advantages of alternative resampling methods, including NPI-B, in evaluating the stability of PCA outcomes. An analysis of the performance of various resampling methods regarding accuracy, computational efficiency, and robustness to diverse data characteristics would yield important insights into the most appropriate approach for particular applications.

3. Experiments results

This section explains the implementation of the bootstrap method in PCA on the selected data, protein, chemical and metabolic syndrome datasets. It also offers a discussion across the subsequent subsections: Section 3.1 presents the results for the protein data; Section 3.2 presents the results for the chemical data; Section 3.3 presents the results for the metabolic syndrome data.

3.1 *Protein data*

The protein consumption data consisted of 25 rows of observations and 11 variables that were collected from 25 European countries for 9 kinds (groups) of food Gabriel [25], the implementation relied solely on quantitative data, consisting of 9 variables. PCA was first implemented on these data to give an insight into their structures and calculate their variability; results are shown in Figure 2, which shows that 91% of the total variance was explained by five components, with PC1 explaining 44.5% and PC2 explaining 18.2%. PC3 further explained 12.5% of the total variance, with PC4 and PC5 explaining 10.6% and 5.2%, respectively. This offered a perspective of these components’ comparative relationships to the dataset, with just five PCs explaining 91.43% of the total variance. Very briefly, the most variance was explained by the first PC, which was followed by the second and so on. We struck a balance between maintaining the information in the data and lowering the dimensionality by selecting a suitable number of major components. For evaluation of the relationships between the multiple variables and the relevant PCs, Table 1 illustrates the squared factor loading for each variable; the total percentage of the variance that the two components could explain was around 62.7%. Consider the variables Red_Meat, White_Meat, Eggs, Milk, Fish, Cereal, Starch, Nuts and Fruits_Vegetables; we represent them with X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, respectively. As illustrated in Eq. (5) and Eq. (6), one can infer that the highest contributing variables in PC1 were Cereal and Nuts while Eggs provided the highest negative level of contribution. In PC2, the highest contributing variables were Fish and Fruits_Vegetables, while Eggs provided a low negative contribution level.

$$\mathrm{PC1} = -0.303\mathrm{X}_1 - 0.311\mathrm{X}_2 - 0.427\mathrm{X}_3 - 0.378\mathrm{X}_4 - 0.136\mathrm{X}_5 + 0.438\mathrm{X}_6 - 0.297\mathrm{X}_7 + 0.420\mathrm{X}_8 + 0.110\mathrm{X}_9 \tag{5}$$

$$\mathrm{PC2} = -0.056\mathrm{X}_1 - 0.237\mathrm{X}_2 - 0.035\mathrm{X}_3 - 0.185\mathrm{X}_4 + 0.647\mathrm{X}_5 - 0.233\mathrm{X}_6 + 0.353\mathrm{X}_7 + 0.143\mathrm{X}_8 + 0.536\mathrm{X}_9 \tag{6}$$

$$\mathrm{PC3} = -0.298\mathrm{X}_1 + 0.624\mathrm{X}_2 + 0.182\mathrm{X}_3 - 0.386\mathrm{X}_4 - 0.321\mathrm{X}_5 + 0.096\mathrm{X}_6 + 0.243\mathrm{X}_7 - 0.054\mathrm{X}_8 + 0.407\mathrm{X}_9 \tag{7}$$

$$\mathrm{PC4} = -0.646\mathrm{X}_1 + 0.037\mathrm{X}_2 - 0.313\mathrm{X}_3 + 0.003\mathrm{X}_4 + 0.216\mathrm{X}_5 + 0.006\mathrm{X}_6 + 0.337\mathrm{X}_7 - 0.330\mathrm{X}_8 - 0.462\mathrm{X}_9 \tag{8}$$

$$\mathrm{PC5} = 0.322\mathrm{X}_1 - 0.300\mathrm{X}_2 + 0.079\mathrm{X}_3 - 0.200\mathrm{X}_4 - 0.290\mathrm{X}_5 + 0.238\mathrm{X}_6 + 0.736\mathrm{X}_7 + 0.151\mathrm{X}_8 - 0.234\mathrm{X}_9 \tag{9}$$

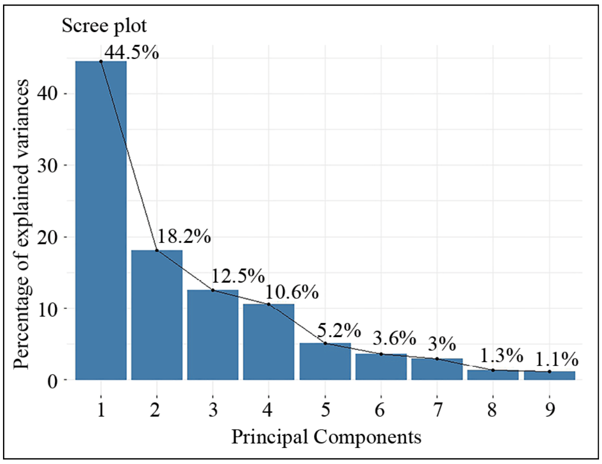

Figure 2. Protein data: PCA scree plot

Table 1. Eigenvector coefficients of protein data

Variable (food)	PC1	PC2	PC3	PC4	PC5
Red_meat	0.303	0.056	0.298	0.646	0.322
White_meat	0.311	0.237	0.624	0.037	0.300
Eggs	0.427	0.035	0.182	0.313	0.079
Milk	0.378	0.185	0.386	0.003	0.200
Fish	0.136	0.647	0.321	0.216	0.290
Cereal	0.438	0.233	0.096	0.006	0.238
Starch	0.297	0.353	0.243	0.337	0.736
Nuts	0.420	0.143	0.054	0.330	0.151
Fruits_vegetables	0.110	0.536	0.407	0.462	0.234

A bootstrap method was used to assess the stability of previous results from 1,000 samples. The initially generated data matrix corresponded with the bootstrapped principal components, as seen in Figures 3 and 4. The results emphasized the reliability of variable representation and their interconnections. The picture illustrates the observations as points on a plane defined by two principal components (synthetic variables), while the original variables are shown as vectors. These graphics illustrate the interactions among the variables and emphasize the uniformity of variable representation as well as the correlation between variables and measures. Certain components have a positive connection, including Cereal, Nuts, and Fruits_Vegetables, with a strong association between Fish, Eggs, and Cereal, as seen in Figure 5.

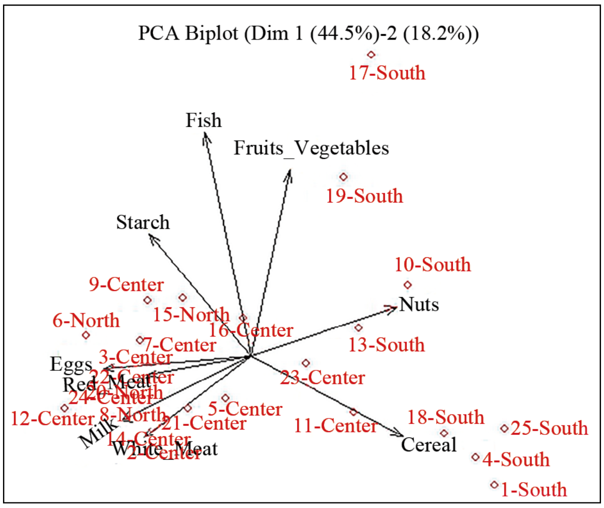

Figure 3. Protein data: PCA biplot

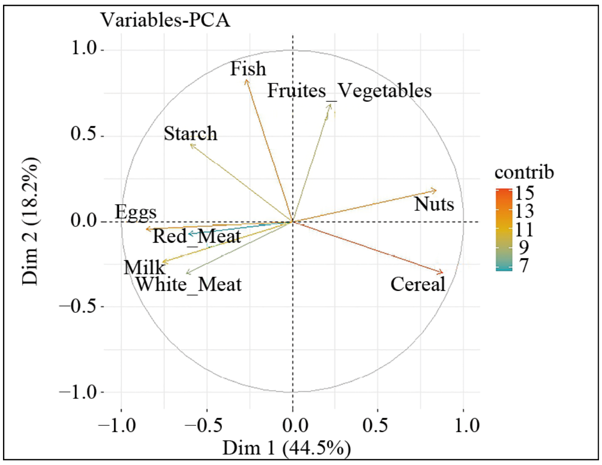

Figure 4. Protein data: variable factor map

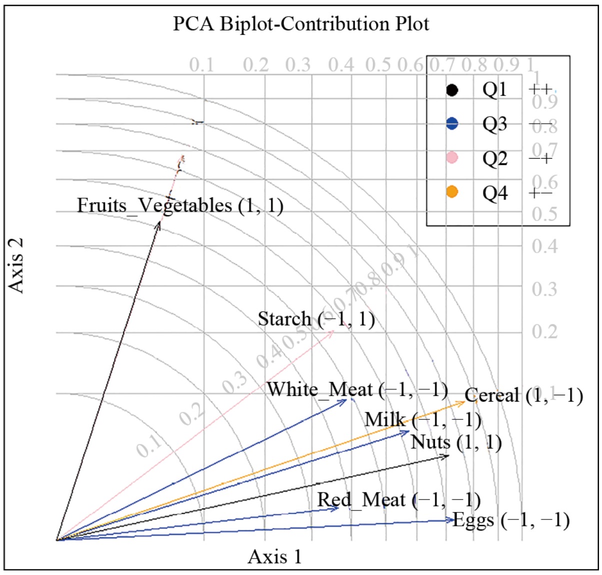

Figure 5. Protein data: PCA biplot contribution plot

Additionally, Table 2 presents a matrix illustrating the variance proportions (columns) for each bootstrapped sample (rows). The bootstrapped mean for PC1 in our analyzed samples was 46.95. The 95% confidence range for the variance difference between the test and control populations was [38.577, 56.741], while the 95% confidence interval for the mean difference between the test and control populations was [38.041, 55.867]. For PC2, the bootstrapped mean was 19.77, with 95% confidence intervals for the mean and variance differences being [13.628, 25.913] and [14.508, 26.407], respectively.

An analysis of consumption of proteins data demonstrated the effectiveness of Principal Component Analysis (PCA) in discerning significant patterns and reducing dimensionality. The data's properties, particularly the presence of several connected variables about food consumption, provide an appropriate scenario for the use of PCA. PCA successfully delineated the data's underlying structure by finding major components that accounted for a significant portion of the overall variance.

The bootstrap analysis significantly demonstrated the stability of the PCA results. The determined confidence intervals for the mean and variance of each major component enabled an assessment of the robustness of the identified components. Smaller confidence intervals indicate improved stability and more certainty in the identified principal components. This technique enhanced our understanding and proficiency in interpreting PCA findings by providing a more reliable explanation of the underlying data structure and the relationships among variables.

Table 2. Protein data: accounted variance

	Initial	Bootstrapped mean	CI: P2.5	CI: P97.5	CI: MEI	CI: MES
Dim 1	44.516	46.954	38.577	56.741	38.041	55.867
Dim 2	18.167	19.771	14.508	26.407	13.628	25.913
Dim 3	12.532	13.015	9.343	16.828	9.219	16.811
Dim 4	10.607	8.861	5.597	12.260	5.457	12.265
Dim 5	5.154	5.028	3.119	7.589	2.750	7.306
Dim 6	3.613	3.120	1.886	4.725	1.684	4.555
Dim 7	3.018	1.872	0.946	3.039	0.778	2.967
Dim 8	1.292	0.938	0.370	1.524	0.349	1.527
Dim 9	1.101	0.441	0.080	0.953	0.001	0.884

3.2 *Chemical data*

This chemical dataset was obtained by the Department of Ecology at the University of León in Spain and included 16 variables and 324 observations. The data can potentially be accessed with the R program [25]. Principal Component Analysis (PCA) was performed on 11 quantitative variables, revealing that five components explained 92.7% of the total variance, with PC1 contributing 51.1% and PC2 providing 18.4%. This analysis provided a perspective on the relationships between variables, as shown in Figure 6. For evaluation of the relationship between the multiple variables and the relevant PCs, Table 3 illustrates the squared factor loading for each variable; the total percentage of the variance that these two components could explain was around 69.5%. Consider the variables pH, ALKALINITYmeql, CO2free, NNH4mgl, NNO3mgl, SRPmglP, TPmgl, TSSmgl, CONDUCTIVITYmScm, TSPmglP and Chlorophyllamgl, represented by X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, respectively. As illustrated in Equations (10) and (11), one can infer that the highest contributing variables in PC1 were SRPmglP and TSPmglP while TSSmgl provided the highest negative contribution level. In PC2, the highest contributing variables were pH and Chlorophyllamgl, while SRPmglP provided a low negative contribution level.

$$\begin{aligned} \mathrm{PC1} = {} & 0.094\mathrm{X}_1 + 0.395\mathrm{X}_2 - 0.085\mathrm{X}_3 + 0.314\mathrm{X}_4 + 0.276\mathrm{X}_5 + 0.403\mathrm{X}_6 \\ & + 0.378\mathrm{X}_7 - 0.106\mathrm{X}_8 + 0.402\mathrm{X}_9 + 0.403\mathrm{X}_{10} + 0.114\mathrm{X}_{11}, \end{aligned} \tag{10}$$

$$\begin{aligned} \mathrm{PC2} = {} & 0.573\mathrm{X}_1 - 0.044\mathrm{X}_2 - 0.496\mathrm{X}_3 - 0.202\mathrm{X}_4 - 0.110\mathrm{X}_5 - 0.029\mathrm{X}_6 \\ & + 0.112\mathrm{X}_7 + 0.351\mathrm{X}_8 - 0.086\mathrm{X}_9 + 0.005\mathrm{X}_{10} + 0.476\mathrm{X}_{11}, \end{aligned} \tag{11}$$

$$\begin{aligned}\mathrm{PC3} = {} & 0.365\mathrm{X}_1 - 0.077\mathrm{X}_2 - 0.486\mathrm{X}_3 - 0.012\mathrm{X}_4 + 0.258\mathrm{X}_5 - 0.052\mathrm{X}_6 - 0.205\mathrm{X}_7 - 0.566\mathrm{X}_8 \\ & + 0.016\mathrm{X}_9 - 0.077\mathrm{X}_{10} - 0.432\mathrm{X}_{11}, \end{aligned} \tag{12}$$

$$\begin{aligned}\mathrm{PC4} = {} & 0.082\mathrm{X}_1 - 0.199\mathrm{X}_2 + 0.216\mathrm{X}_3 - 0.475\mathrm{X}_4 + 0.726\mathrm{X}_5 \\ & + 0.062\mathrm{X}_6 + 0.050\mathrm{X}_7 + 0.334\mathrm{X}_8 + 0.069\mathrm{X}_9 + 0.053\mathrm{X}_{10} - 0.167\mathrm{X}_{11} \end{aligned} \tag{13}$$

$$\begin{aligned}\mathrm{PC5} = {} & 0.024\mathrm{X}_1 + 0.127\mathrm{X}_2 + 0.013\mathrm{X}_3 + 0.504\mathrm{X}_4 + 0.414\mathrm{X}_5 - 0.363\mathrm{X}_6 - 0.364\mathrm{X}_7 \\ & + 0.160\mathrm{X}_8 + 0.222\mathrm{X}_9 - 0.362\mathrm{X}_{10} + 0.295\mathrm{X}_{11}. \end{aligned} \tag{14}$$

This analysis provided a perspective on the relationships between variables. A bootstrap analysis was performed to assess the stability of these results. The original data matrix was projected onto the bootstrapped PCs, and the results were visualized in Figures 7 and 8. These figures highlighted the consistency of variable representation and the associations between variables. Some factors were positively connected, such as NNH4mgl, NNO3mgl, SRPmglP, TPmgl, CONDUCTIVITYmScm and TSPmglP, with high correlations in between, as represented in Figure 9.

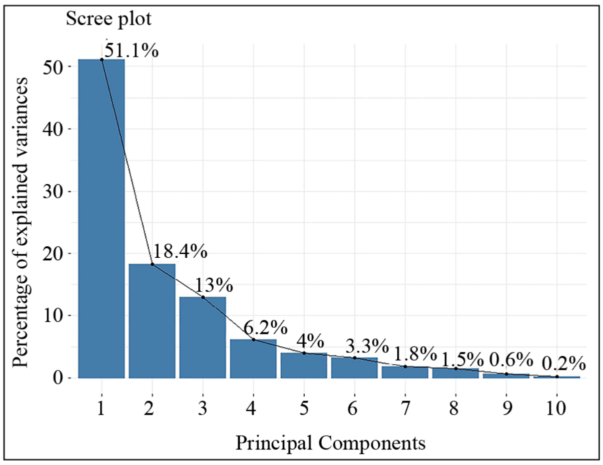

Figure 6. Chemical data: PCA scree plot

Table 3. Eigenvector coefficients of chemical data

Variable (food)	PC1	PC2	PC3	PC4	PC5
pH	0.094	0.573	0.365	0.082	0.024
ALKALINITYmeql	0.395	0.044	0.077	0.199	0.127
CO2free	0.085	0.496	0.486	0.216	0.013
NNH4mgl	0.314	0.202	0.012	0.475	0.504
NNO3mgl	0.276	0.110	0.258	0.726	0.414
SRPmglP	0.403	0.029	0.052	0.062	0.363
TPmgl	0.378	0.112	0.205	0.050	0.364
TSSmgl	0.106	0.351	0.566	0.334	0.160
CONDUCTIVITYmScm	0.402	0.086	0.016	0.069	0.222
TSPmglP	0.403	0.005	0.077	0.053	0.362
Chlorophyllamgl	0.114	0.476	0.432	0.167	0.295

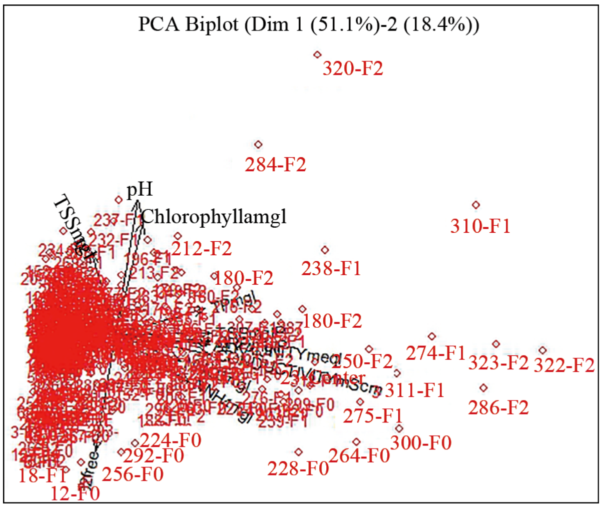

Figure 7. Chemical data: PCA biplot

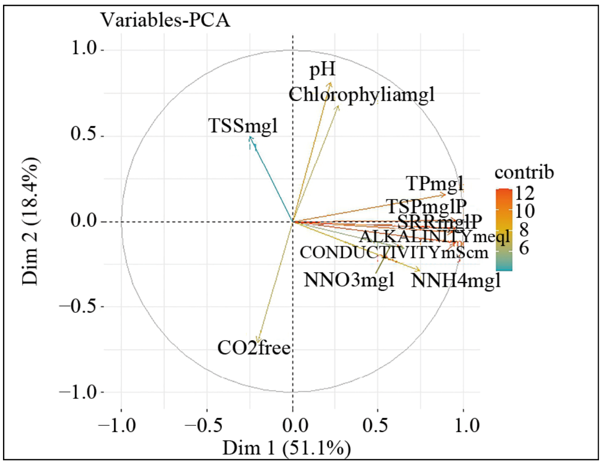

Figure 8. Chemical data: variable factor map

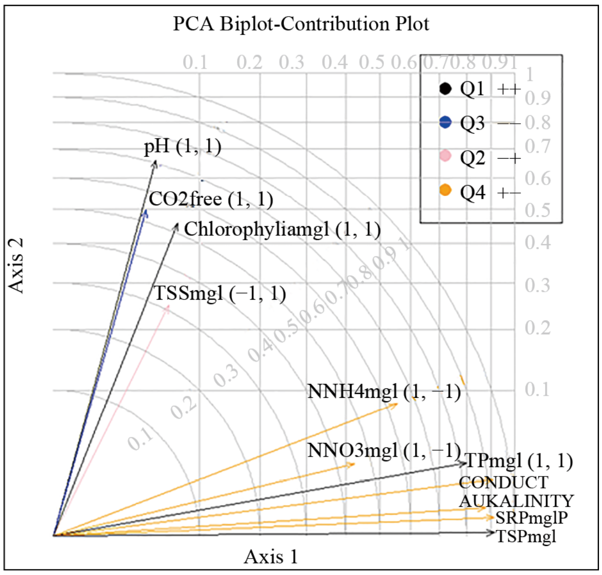

Figure 9. Chemical data: PCA biplot contribution plot

Additionally, Table 4 presents the variance proportions of each bootstrapped sample (columns). The bootstrapped mean of PC1 in our observed samples was 51.32. The 95% confidence interval for the variance difference estimate between the test and control groups was [49.330, 53.395]. The lower and upper bound of 95% confidence interval for the mean difference equals [49.308, 53.330]. For PC2, the bootstrapped mean was 18.465, with 95% confidence intervals for the mean and variance differences of [16.997, 19.932] and [17.150, 20.068], respectively. One can observe that five principal components explained 92.7% of the total variance. Bootstrap analysis validates the stability of these components. Assessing the confidence intervals for the mean and variance of each principal component allows us to ascertain the robustness of the PCA solution. The smaller confidence intervals signify enhanced stability and greater certainty in the selected major components. A comprehensive understanding of the foundational structure of chemical data and the interrelationships among variables enhances the dependability of PCA.

Table 4. Chemical data; accounted variance

	Initial	Bootstrapped mean	CI: P2.5	CI: P97.5	CI: MEI	CI: MES
Dim 1	51.067	51.319	49.330	53.395	49.308	53.330
Dim 2	18.355	18.465	17.150	20.068	16.997	19.932
Dim 3	12.961	12.879	11.529	14.170	11.540	14.217
Dim 4	6.191	6.284	5.262	7.424	5.170	7.399
Dim 5	4.014	4.031	3.289	4.844	3.240	4.822
Dim 6	3.296	3.118	2.516	3.700	2.519	3.716
Dim 7	1.767	1.766	1.465	2.185	1.419	2.112
Dim 8	1.476	1.382	1.033	1.693	1.049	1.715
Dim 9	0.640	0.539	0.248	0.793	0.225	0.853
Dim 10	0.197	0.184	0.108	0.295	0.092	0.277
Dim 11	0.036	0.034	0.019	0.054	0.017	0.051

3.3 *Metabolic syndrome data*

This cross-sectional study, which included 172 students, was conducted at KAU between December 2017 and April 2018 [26]. Female science students between 18 and 25 met the inclusion requirements. Pregnancy and lactation were the conditions for disqualification. This study was approved by the Faculty of Medicine, KAU's biomedical ethics section, and the General Directorate of Health Affairs of the College of Medicine. The data variables are defined as follows;

Body mass index (BMI), weight (Wt), waist circumference (Waist), hip circumference (Hips), waist-to-hip ratio (Whr), cholesterol (Chol), triglycerides (TGL), HDL-cholesterol (HDL), and LDL-cholesterol (LDL).

Hence, the PCA is implemented on data variables and the results are shown in Figure 10, which shows that 89.3% of the total variance was explained by five components, with PC1 explaining 37.8% and PC2 explaining 17.1%. PC3 further explained 13% of the total variance, with PC4 and PC5 explaining 11.2% and 10.2%, respectively. This offered a perspective on these components' comparative relationships to the dataset, with just five PCs explaining 89.3% of the total variance. For evaluation of the relationship between the multiple variables and the relevant PCs, Table 5 illustrates the squared factor loading for each variable; the total percentage of the variance that the two components could explain was around 54.9%. Consider the variables BMI, Wt, Waist, Hips, Whr, Chol, TGL, HDL and LDL, represented by X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, respectively. As illustrated in Equations (15) and (16), one can infer that the highest contributing variables in PC1 were BMI and Wt, with no negative levels of contribution. In PC2, the highest contributing variables were Waist and Whr while BMI provided a low negative level of contribution.

$$\mathrm{PC1} = 0.517\mathrm{X}_1 + 0.515\mathrm{X}_2 + 0.475\mathrm{X}_3 + 0.408\mathrm{X}_4 + 0.122\mathrm{X}_5 + 0.130\mathrm{X}_6 + 0.177\mathrm{X}_7 + 0.051\mathrm{X}_8 + 0.083\mathrm{X}_9 \tag{15}$$

$$\mathrm{PC2} = -0.091\mathrm{X}_1 - 0.080\mathrm{X}_2 + 0.290\mathrm{X}_3 - 0.471\mathrm{X}_4 + 0.741\mathrm{X}_5 + 0.281\mathrm{X}_6 + 0.083\mathrm{X}_7 + 0.180\mathrm{X}_8 - 0.102\mathrm{X}_9 \tag{16}$$

$$\mathrm{PC3} = 0.066\mathrm{X}_1 + 0.085\mathrm{X}_2 + 0.079\mathrm{X}_3 + 0.106\mathrm{X}_4 - 0.038\mathrm{X}_5 - 0.040\mathrm{X}_6 - 0.615\mathrm{X}_7 + 0.339\mathrm{X}_8 - 0.688\mathrm{X}_9 \tag{17}$$

$$\mathrm{PC4} = -0.065\mathrm{X}_1 - 0.025\mathrm{X}_2 - 0.117\mathrm{X}_3 + 0.050\mathrm{X}_4 - 0.224\mathrm{X}_5 + 0.608\mathrm{X}_6 - 0.197\mathrm{X}_7 + 0.581\mathrm{X}_8 + 0.424\mathrm{X}_9 \tag{18}$$

$$\mathrm{PC5} = 0.053\mathrm{X}_1 + 0.028\mathrm{X}_2 - 0.034\mathrm{X}_3 + 0.001\mathrm{X}_4 + 0.002\mathrm{X}_5 + 0.600\mathrm{X}_6 - 0.360\mathrm{X}_7 - 0.708\mathrm{X}_8 - 0.057\mathrm{X}_9. \tag{19}$$

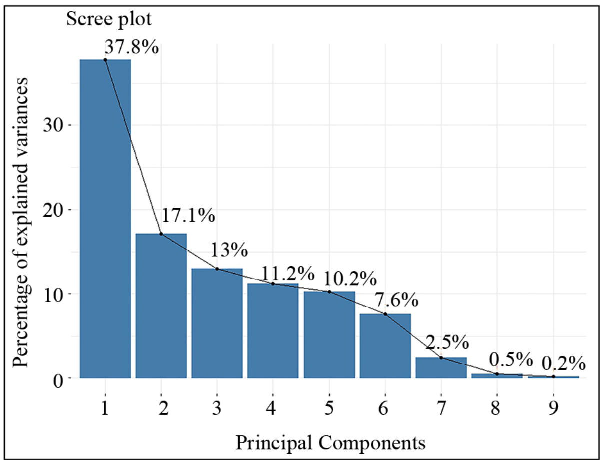

Figure 10. Metabolic syndrome data: PCA scree plot

Table 5. Eigenvector coefficients of metabolic syndrome data

Variable (food)	PC1	PC2	PC3	PC4	PC5
BMI	0.517	0.091	0.066	0.065	0.053
Wt	0.515	0.080	0.085	0.025	0.028
Waist	0.475	0.290	0.079	0.117	0.034
Hips	0.408	0.471	0.106	0.050	0.001
Whr	0.122	0.741	0.038	0.224	0.002
Chol	0.130	0.281	0.040	0.608	0.600
TGL	0.177	0.083	0.615	0.197	0.360
HDL	0.051	0.180	0.339	0.581	0.708
LDL	0.083	0.102	0.688	0.424	0.057

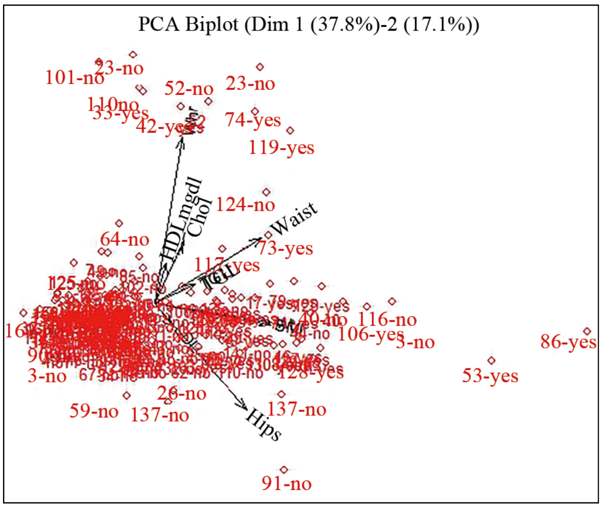

Figure 11. Metabolic syndrome data: PCA biplot

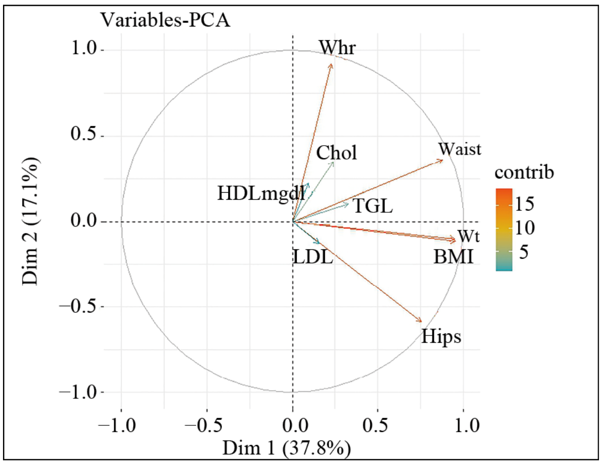

Figure 12. Metabolic syndrome data: variable factor map

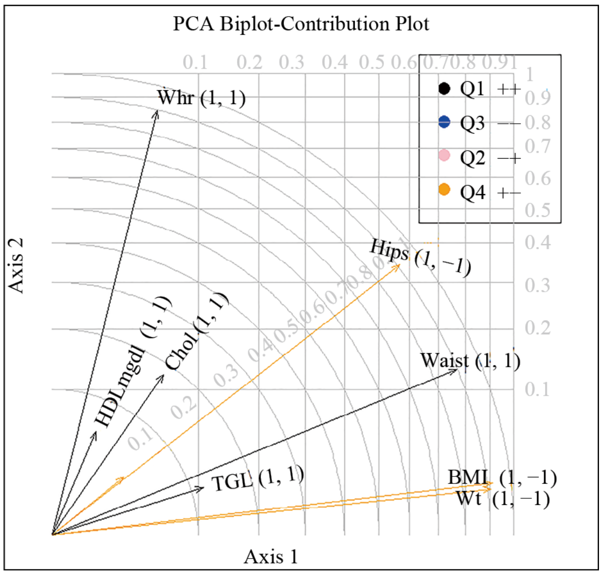

Figure 13. Metabolic syndrome data: PCA biplot contribution plot

The results of projected data onto the bootstrap PC are displayed in Figures 11 and 12. The observations are shown as dots on a two-dimensional plane formed by the first two main components, with the original variables shown as vectors. The figures show the interactions between the variables, which also demonstrate the consistency of variable representation and the relationship between measurements and variables. Numerous indicators exhibited positive correlations, particularly WHR, cholesterol, and HDL, with a strong association seen among WHR, waist circumference, hip measurements, weight, and BMI, as shown in Figure 13. Table 6 displays the outcomes of the bootstrap analysis. The bootstrapped mean for PC1 was 38.16, accompanied by a 95% confidence interval of [34.668, 41.645]. The 95% confidence interval for the variance of PC1 is [34.795, 41.608]. The bootstrapped mean for PC2 was 17.095, with 95% confidence intervals of [14.860, 20.175] for the mean and [14.995, 20.240] for the variance. Narrower confidence intervals for each principal component's mean and variance indicate enhanced stability. Consistent intervals among bootstrap replicates would enhance the stability of the PCA solution. The findings offer insights into the robustness of PCA results and the reliability of identified principal components in elucidating the relationships between anthropometric and lipid profile variables in this student population. Five principal components explain the 89.3% of the total variance; this indicates that PCA successfully reduced the dimensionality of the student health data.

The analysis demonstrated significant correlations between BMI, weight, and the first principal component, whereas waist circumference and waist-to-hip ratio prominently affected the second principal component. The bootstrap study demonstrated the reliability of PCA results, with restricted confidence ranges for each principal component. This improves confidence in the correlations between anthropometric and lipid profile variables, yielding a more reliable understanding of the data's framework and consequences for student health.

Table 6. Metabolic syndrome data; accounted variance

	Initial	Bootstrapped mean	CI: P2.5	CI: P97.5	CI: MEI	CI: MES
Dim 1	37.754	38.156	34.795	41.608	34.668	41.645
Dim 2	17.095	17.518	14.995	20.240	14.860	20.175
Dim 3	13.008	13.526	11.926	15.456	11.743	15.309
Dim 4	11.208	11.251	9.979	12.765	9.865	12.636
Dim 5	10.213	9.397	7.931	10.575	8.020	10.774
Dim 6	7.566	7.143	5.637	8.688	5.650	8.635
Dim 7	2.478	2.356	1.573	3.185	1.559	3.153
Dim 8	0.508	0.506	0.330	0.749	0.297	0.715
Dim 9	0.169	0.148	0.087	0.210	0.085	0.210

4. Discussion

In conclusion, this paper explores the critical importance of assessing the stability of Principal Components Analysis (PCA) and illustrates the effectiveness of PCA in a variety of disciplines. The effectiveness of PCA in identifying significant patterns and relationships within complex data was demonstrated by its application to three distinct datasets: protein consumption, the chemical composition of water bodies, and anthropometric data of female students. PCA highlighted unique dietary trends in protein consumption data throughout European nations, emphasizing the impact of cereal and nut intake on one main component and the opposing effect of eggs on another. PCA efficiently identified critical water quality factors in the chemical data, including suspended particles and nutrients, which significantly impacted the first main component, although pH and chlorophyll levels were notable in the second component. PCA analysis of student health data indicated robust correlations between BMI, weight, and the first main component, although waist circumference and waist-to-hip ratio significantly impacted the second principal component.

Moreover, the bootstrap approach substantially improved the reliability and interpretability of the PCA outcomes. By evaluating the stability of eigenvalues, eigenvectors, and loadings via confidence intervals, we attained a more profound comprehension of the robustness of the discovered major components. Narrower confidence ranges, seen in all datasets, provide compelling evidence for the stability of the PCA solutions, suggesting that the revealed patterns were not just products of sampling variability. The significant correlations between BMI and waist circumference, together with the narrow confidence intervals for the principal components in the student health data, suggest that they are likely relevant to other comparable populations.

This study emphasizes the need for a thorough stability evaluation method that ensures the accuracy and practicability of PCA results. The results have substantial ramifications for several domains, including public health, environmental science, and nutrition research, where informed decision-making and effective therapies rely on the reliability and comprehensiveness of data analysis. Future research will focus on the efficacy of the bootstrap method compared to alternative stability assessment methods for PCA, such as jackknife resampling and cross-validation approaches.

Data availability statement

The "Metabolic Syndrome Data" used in this study was provided by the third author. Data availability can be requested from Maryam H. Al-Zahrani upon reasonable request and with the third author's permission.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Einbeck J, Kalantan Z, Kruger U. Practical considerations on nonparametric methods for estimating intrinsic dimensions of nonlinear data structures. *International Journal of Pattern Recognition and Artificial Intelligence*. 2020; 34(09): 2058010.

[2] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1901; 2(11): 559-572.

[3] Fisher RA, Mackenzie WA. Studies in crop variation. II. The manurial response of different potato varieties. *The Journal of Agricultural Science*. 1923; 13(3): 311-320.

[4] Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933; 24(6): 417.

[5] Rao CR. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics*. 1964; 26(4): 329-358.

[6] Alqahtani NA, Kalantan ZI. Gaussian mixture models based on principal components and applications. *Mathematical Problems in Engineering*. 2020; 2020(1): 1202307.

[7] Vidal R, Ma Y, Sastry S. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005; 27(12): 1945-1959.

[8] Xu X, Xie L, Wang S. Multimode process monitoring with PCA mixture model. *Computers and Electrical Engineering*. 2014; 40(7): 2101-2112.

[9] Efron B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*. 1979; 7(1): 1-26.

[10] Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. 1st ed. New York: Chapman and Hall/CRC; 1994.

[11] Fang Y, Wang J. Selection of the number of clusters via the bootstrap method. *Computational Statistics and Data Analysis*. 2012; 56(3): 468-477.

[12] Jaki T, Su TL, Kim M, Van Horn ML. An evaluation of the bootstrap for model validation in mixture models. *Communications in Statistics-Simulation and Computation*. 2018; 47(4): 1028-1038.

[13] Binhimd S, Coolen F. Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov-Smirnov test. *Journal of Statistical Theory and Practice*. 2020; 14(2): 1-13.

[14] Binhimd S, Almalki B. Bootstrap methods and reproducibility probability. *American Scientific Research Journal for Engineering, Technology, and Sciences*. 2019; 59(1): 76-80.

[15] Daudin JJ, Duby C, Trecourt P. Stability of principal component analysis studied by the bootstrap method. *Statistics: A Journal of Theoretical and Applied Statistics*. 1988; 19(2): 241-258.

[16] Linting M, Meulman JJ, Groenen PJ, Van der Kooij AJ. Stability of nonlinear principal components analysis: An empirical study using the balanced bootstrap. *Psychological Methods*. 2007; 12(3): 359.

[17] Hennig C. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*. 2007; 52(1): 258-271.

[18] Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. USA: Pearson; 2002.

[19] Hinkley DV. Bootstrap methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1988; 50(3): 321-337.

[20] Grzegorzewski P, Romaniuk M. Bootstrap methods for epistemic fuzzy data. *International Journal of Applied Mathematics and Computer Science*. 2022; 32(2): 285-297.

[21] Hutson AD. A composite quantile function estimator with applications in bootstrapping. *Journal of Applied Statistics*. 2000; 27(5): 567-577.

[22] Johnson RW. An introduction to the bootstrap. *Teaching Statistics*. 2001; 23(2): 49-54.

[23] Chateau F, Lebart L. Assessing sample variability in the visualization techniques related to principal component analysis: Bootstrap and alternative simulation methods. In: *COMPSTAT: Proceedings in Computational Statistics 12th Symposium held in Barcelona, Spain, 1996*. Heidelberg: COMPSTAT, Physica Verlag; 1996. p.205-210.

[24] Jaadi Z. Principal Component Analysis (PCA): A step-by-step explanation. *Ανάκτηση*. 2024; 9(06): 2024.

[25] The R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.

[26] Balgoon MJ, Al-Zahrani MH, Alkhattabi NA, Alzahrani NA. The correlation between obesity and metabolic syndrome in young female university students in the Kingdom of Saudi Arabia. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*. 2019; 13(4): 2399-2402.