ITM Web of Conferences **58**, 04001 (2024) https://doi.org/10.1051/itmconf/20245804001
The 6th IICMA 2023

Variance The Estimation Eigen Value of Principal Component Analysis and Nonlinear Principal Component Analysis

Makkulau [1], *Andi* Tenri Ampa[1*], *Irma* Yahya[1], *Lilis* La Ome[1], and *La Ode* Saidi[2],

[1]Faculty of Mathematics and Natural Sciences, Universitas of Halu Oleo, Statistics Department, 2, Southeast Sulawesi, Indonesia
[2]Faculty of Mathematics and Natural Sciences, Universitas of Halu Oleo, Computer Science Department, Southeast Sulawesi, Indonesia

Abstract. Nonlinear Principal Component Analysis (PRINCALS) is an extension of Principal Component Analysis (Linear), which can reduce the variables of mixed scale multivariable data (nominal, ordinal, interval, and ratio) simultaneously. This study investigated variance the estimation eigen value of Principal Component Analysis Linear and Nonlinear. The result showed that variance the estimation eigen value of Principal Component Analysis is $\mathrm{Var}(\hat{\tilde{\lambda}}_S) = \mathbf{H}_S^{'}\mathbf{V}_S\mathbf{H}_S$ and variance the estimation eigen value of Nonlinear Principal Component Analysis is $\mathrm{Var}(\hat{\lambda}_R) = \mathbf{H}_R^{'}\mathbf{V}_R\mathbf{H}_R$. Variance the estimation eigen value of Nonlinear Principal Component Analysis better (efficient) than variance the estimation eigen value of Principal Component Analysis.

1 Introduction

Principal Component Analysis has been discussed, among others, by Joliffe, Bolton, and Vichi is an analysis to reduce variables, where the variables formed do not correlate with each other [1]. [2] discuss Principal Component Analysis in ordinal scale data, while Nonlinear Principal Component Analysis is a method for reducing variables from multivariable mixed scale data simultaneously.

Speaking of Principal Component Analysis, it will be faced with a determination of the eigen value ($\tilde{\lambda}$) which is the basis of determining the main components of the Principal Component Analysis. The eigen value obtained later is the eigen value of the samples because to study the population will be faced with limitations such as problems of time, cost, energy, etc.

Because the eigen value ($\tilde{\lambda}$) obtained is the eigen value of the samples that are expected to represent the population, then of course the eigen value is only an estimator of the true

*Corresponding author: anditenri.ampa@uho.ac.id

eigen value ($\tilde{\hat{\lambda}}$). An estimator is said to be good if it meets the criteria of being unbiased, efficient, and consistent.

This research is to determine variance the estimation eigen value (Var ($\tilde{\hat{\lambda}}$)) of Linear and Nonlinear Principal Component Analysis with Delta Method.

1.1 Principal Component Analysis

Principal Component Analysis (Linear Principal Component Analysis) is an analysis to reduce variables, where the variables formed do not correlate with each other ([1] and [3]). For more details the application of Principal Component Analysis can be seen in [4] and [5].

For example $\tilde{X} = [X_1, X_2, ..., X_j, ..., X_m]$ population random vector with mean vector $\boldsymbol{\mu}$, variance covariance matrix $\boldsymbol{\Sigma}$, and correlation matrix $\boldsymbol{\rho}$. Eigen value dan eigen vector of $\boldsymbol{\Sigma}$ are $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_j, e_j), \ldots, (\lambda_m, e_m)$, then principal component to j:

$$Y_j = \tilde{e}_j' \tilde{X} = e_{1j}X_1 + e_{2j}X_2 + \ldots + e_{jj}X_j + \ldots + e_{mj}X_m ; \quad j = 1, 2, \ldots, m \tag{1}$$

with $\text{Var}(Y_j) = \tilde{e}_j' \boldsymbol{\Sigma} \tilde{e}_j = \lambda_j$ and (2)

$$\text{Cov}(Y_j, Y_m) = \tilde{e}_j' \boldsymbol{\Sigma} \tilde{e}_j = 0 \tag{3}$$

For example $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_i, \ldots, \tilde{X}_n$ is a random sample of size n and dimension m. This sample has an average vector $\bar{\tilde{X}}$, variance-covariance matrix **S** and correlation matrix **R**. If **S** is variance-covariance matrix order of $m \times m$ with a pair of the estimation eigen value and eigen vector, namely:
$(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \ldots, \left(\hat{\lambda}_j, \hat{e}_j\right), \ldots, (\hat{\lambda}_m, \hat{e}_m)$, then the estimation principal component to j is:

$$\hat{Y}_j = \hat{\tilde{e}}_j' \tilde{X} = \hat{e}_{1j}X_1 + \hat{e}_{2j}X_2 + L + \hat{e}_{mj}X_m, \; j = 1, 2, \ldots, m \tag{4}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_j \geq \ldots \geq \hat{\lambda}_m \geq 0$, and $\hat{\lambda}_j$ is the estimation eigen value of matrix **S** and of being unbiased: $E(\hat{\lambda}_j) = \lambda_j$.

1.2 Nonlinear Principal Component Analysis *(PRINCALS)*

PRINCALS (***Prin****cipal* ***C****omponent Analysis by means of* ***A****lternating* ***L****east* ***S****quares*) or Principal Component Analysis using the alternative least squares approach commonly called Nonlinear Principal Component Analysis is a development of Principal Component Analysis [6].

PRINCALS is a method that analyzes mixed scale data simultaneously by grouping variables whose linear correlations are aligned into one principal [7].

If **R** is the matrix correlation sample order of $m \times m$ with the pair variance the estimation eigen value and eigen vector is $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \ldots, \left(\hat{\lambda}_j, \hat{e}_j\right), \ldots, (\hat{\lambda}_m, \hat{e}_m)$, and $\tilde{Z} = [Z_1, Z_2, \ldots, Z_m]$ observations are standardized with R correlation matrix, then the estimation principal component to-j is:

$$\hat{Y}_j = \hat{\tilde{e}}_j' \tilde{Z} = \hat{e}_{1j} Z_1 + \hat{e}_{2j} Z_2 + L + \hat{e}_{mj} Z_m \text{;} \quad j = 1, 2, \ldots, m \tag{5}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_j \geq \ldots \geq \hat{\lambda}_m \geq 0$, and $\hat{\lambda}_j$ is the estimation eigen value of R matrix and has an unbiased property:

$E(\hat{\lambda}_j) = \lambda_j$ is the estimation eigen value. (6)

PRINCALS here is based on the Meet Loss Theory, with the loss function homogeneity is:

$$\sigma_M(\mathbf{X}, \mathbf{Y}) = m^{-1} \sum_{j=1}^{m} (\mathbf{X} - \mathbf{G}_j \mathbf{Y}_j)'(\mathbf{X} - \mathbf{G}_j \mathbf{Y}_j) \tag{7}$$

If multivariable data without missing data, the estimation eigen value can be searched by PRINCALS from the formula [6]:

$$m^{-1}\mathbf{R}(\mathbf{Q}) \tag{8}$$

R(**Q**) = matrix the correlation matrix between the combined linear scores (The linear composite scores) of all sets of **Q** matrix on all dimensions.

1.3 Multivariate Normal Distribution

Several theorems on the Multivariate Normal distribution ([8] dan [9]).

Theorem 1:

If $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_i, \ldots, \tilde{X}_n \sim N_m(\tilde{\mu}, \mathbf{\Sigma})$, then

$$\bar{\tilde{X}} \sim N_m\left(\tilde{\mu}, \tfrac{1}{n}\mathbf{\Sigma}\right) \tag{9}$$

Theorema 2:

If $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_i, \ldots, \tilde{X}_n$ a random sample of population with finite mean vector $\tilde{\mu}$ variance-covariance matrix Σ finite, then:

$$\sqrt{n}\,(\bar{\tilde{X}} - \tilde{\mu}) \sim N_m(0, \mathbf{\Sigma}), \text{ for } n \to \infty \tag{10}$$

1.4 Delta Method

Delta method is a method for estimating the expected value and variance of random variables [10]. In this study, the Delta Method will be used variance the estimation eigen.

By looking at the equation:

$$E(y) = \beta_0 + \tilde{X}'\tilde{\beta} + \tilde{X}'\mathbf{B}\tilde{X}, \tag{11}$$

$$\tilde{X} = (X_1, X_2, \ldots, X_m), \quad \tilde{\beta} = (\beta_1, \beta_2, \ldots, \beta_m), 1$$

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1m} \\ \frac{1}{2}\beta_{12} & \beta_{22} & \cdots & \frac{1}{2}\beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}\beta_{1m} & \frac{1}{2}\beta_{2m} & \cdots & \beta_{mm} \end{bmatrix}$$

If **H** is a dimensionless matrix $[m(m+1)/2]$ x m, then:

$$\mathrm{H}' = \begin{bmatrix} Vec'((2\hat{d}_1\hat{d}_1') - diag(\hat{d}_1\hat{d}_1')) \\ Vec'((2\hat{d}_2\hat{d}_2') - diag(\hat{d}_2\hat{d}_2')) \\ \vdots \\ Vec'((2\hat{d}_\mathrm{m}\hat{d}_\mathrm{m}') - diag(\hat{d}_\mathrm{m}\hat{d}_\mathrm{m}')) \end{bmatrix} \quad (12)$$

where:

$\tilde{\hat{\mathrm{d}}}_j$, j = 1, 2, …, m is the eigen vector of matrix **B** whose eigen value is $\tilde{\lambda} = [\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m]$.

Estimator **B** is $\hat{\mathbf{B}}$ obtained from the Maximum Likelihood Estimation method and Vec($\hat{\mathbf{B}}$) with a Multivariate Normal distribution [11].

Estimator of $\tilde{\lambda}$ is $\tilde{\hat{\lambda}} = [\hat{\lambda}_1,\ \hat{\lambda}_2,\ \ldots,\ \hat{\lambda}_m]$, then $\sqrt{n}(\tilde{\hat{\lambda}} - \tilde{\lambda})$ with n to infinitely the asymptotic distribution of m-variate is normal with the mean vector zero and the variance-covariance matrix is $n\mathbf{H}'\mathbf{VH}$ where **H** as mentioned above and **V** is the matrix obtained from submatrix with dimension $[m(m+1)/2]$ x $[m(m+1)/2]$ at the bottom of the matrix $\mathbf{V}^*$ by multiplying the elements by the corresponding numbers 1/4 and 1/2, so that:

$$\mathrm{E}(\tilde{\hat{\lambda}}) = \tilde{\lambda} \text{ and } \mathrm{Var}(\tilde{\hat{\lambda}}) = \mathbf{H}'\mathbf{VH} \quad (13)$$

2 Research Methods

Broadly speaking, this study was conducted by determining of variance-covariance matrix **S** samples, the estimation eigen value and eigen vector from the **S**, matrix **H**, matrix $\mathbf{V}^*$, matrix s^2 and determine estimation confidence interval of the estimation eigen value $\tilde{\hat{\lambda}}$). From the results of the steps above, we will get a variety of feature root estimators (Var ($\tilde{\hat{\lambda}}$)) from Principal Component Analysis (Linear) and Nonlinear Principal Component Analysis.

3 Results and Discussion

Research in everyday life often cannot be done on populations for various reasons, such as time problems, cost issues, and others. This is where the foresight of researchers to determine the "good" parameter estimator that represents the population, namely the unbiased estimator, consistent, and efficient [9].

Principal Component Analysis (Linear) and Nonlinear Principal Component Analysis which is an analysis to reduce variables, the application can be seen in the research of [12]. This analysis can also be used for data derived from multivariate data (see [13]), especially those that do not have outliers. In the case of data in multivariate linear models [9], it is necessary to detect whether the model has an outlier or not as in [14], [15], [16], and in the research of [17].

For example, a population has eigen value λ. If a sample is taken from that population, then the estimation eigen value can be searched $\hat{\lambda}$. To check that it $\hat{\lambda}$ is a "good" estimator for λ. Next, the expectation, variance and distribution of the estimation eigen value will be searched.

3.1 Variance the Estimation Eigen Value of Principal Component Analysis (Linear)

A sample that comes from a multivariate normal distribution population with *n* objects and *m* variables, it can be written in the form of a matrix as follows:

Table 1. The Form of a Matrix.

Object	Variable					
	1	2	⋯	j	⋯	m
1	x_{11}	x_{12}	⋯	x_{1j}	⋯	x_{1m}
2	x_{21}	x_{22}	⋯	x_{2j}	⋯	x_{2m}
⋮	⋮	⋮		⋮		⋮
i	x_{i1}	x_{i2}	⋯	x_{ij}	⋯	x_{im}
⋮	⋮	⋮		⋮		⋮
n	x_{n1}	x_{n2}	⋯	x_{nj}	⋯	x_{nm}

If $\tilde{X} = [X_1 \quad X_2 \quad \dots \quad X_j \quad \dots \quad X_m]' \sim N_m(\tilde{\mu}, \Sigma)$

$\tilde{\mu} = [\ \mu_1\ \mu_2 \dots\ \mu_j\ \dots\ \mu_m\ ]'$

= vector average for the population

$\tilde{\bar{X}} = [\ \bar{X}_1\ \bar{X}_2 \dots\ \bar{X}_j\ \dots\ \bar{X}_m\ ]'$

= vector average for the sample

Σ = variance-covariance matrix for the population.

For example $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_i, \dots, \tilde{X}_n$, where $\tilde{X}_i \sim N_m(\tilde{\mu}, \Sigma)$, then vector average for the sample the *i* th from *n* objects and *m* buah variables:

$$\tilde{\bar{X}} = \tfrac{1}{n}[\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_i + \dots + \tilde{X}_n]$$

$$E(\tilde{\bar{X}}) = \tilde{\mu}, \quad Var(\tilde{\bar{X}}) = \frac{\Sigma}{n}, \text{ so to: } \tilde{\bar{X}} \sim N_m\left(\tilde{\mu}, \tfrac{\Sigma}{n}\right)$$

Estimator Σ is S_n or $\hat{\Sigma} = S_n$ = variance-covariance matrix of sample. Estimator Σ is S, if each element is multiplied by $\frac{1}{n-1}$.

For variance-covariance matrix S_n:

$E(S_n) = \frac{n-1}{n}\Sigma$, where S_n is a biased estimator for Σ.

In order not to be biased, a correction was made with factor C, so that:

$S = C\, S_n$

$E(S) = C\left(\frac{n-1}{n}\right)\Sigma$; where $C = \frac{n}{n-1}$, then:

$E(S) = \Sigma$

So the unbiased estimator of Σ is S, with elements:

$$s_{ij} = \tfrac{1}{n-1}\sum_{i=1}^{n}[(X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}'_j)]; \qquad i = 1, 2, \dots, i, \dots, n; \ \ j = 1, 2, \dots, j, \dots, m$$

Variance the estimation eigen value of Principal Component Analysis with the Delta Method are:

$$\text{Var[Vec(S)]} = \begin{bmatrix} Var(s_{11}) & Cov(s_{11}, s_{12}) & \dots & Cov(s_{11}, s_{1m}) \\ Cov(s_{11}, s_{12}) & Var(s_{22}) & \dots & Cov(s_{22}, s_{2m}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(s_{11}, s_{1m}) & Cov(s_{22}, s_{2m}) & \dots & Var(s_{mm}) \end{bmatrix}$$

$E(\mathbf{S}) = \mathbf{\Sigma}$ and $E[\text{Vec}(\mathbf{S})] = \text{Vec}[E(\mathbf{S})] = \text{Vec}(\mathbf{\Sigma})$

Because **S** is symmetrical ($s_{ij} = s_{ji}$; $i, j = 1, 2, \dots, m$), so:

$\text{Var}[\text{Vec}(\mathbf{S})] = \text{Var}[\text{Vec}(\mathbf{S}^*)] = \frac{2}{n-1}\text{M}_m(\mathbf{\Sigma}\otimes\mathbf{\Sigma}) = \mathbf{V}_S$

Therefore: $\mathbf{Vec(S)} \sim \text{N}_{m(m+1)/2}[\text{Vec}(\mathbf{\Sigma}), \mathbf{V}_S]$

Also found:

$$\mathbf{H}_S' = \begin{bmatrix} \text{Vec}'((2\tilde{\hat{d}}_{1S}\tilde{\hat{d}}_{1S}')\text{-diag}(\tilde{\hat{d}}_{1S}\tilde{\hat{d}}_{1S}')) \\ \text{Vec}'((2\tilde{\hat{d}}_{2S}\tilde{\hat{d}}_{2S}')\text{-diag}(\tilde{\hat{d}}_{2S}\tilde{\hat{d}}_{2S}')) \\ \text{M} \\ \text{Vec}'((2\tilde{\hat{d}}_{mS}\tilde{\hat{d}}_{mS}')\text{-diag}(\tilde{\hat{d}}_{mS}\tilde{\hat{d}}_{mS}')) \end{bmatrix}$$

and $\tilde{\hat{\lambda}}_S = [\hat{\lambda}_{(1)} \; \hat{\lambda}_{(2)} \; \dots \; \hat{\lambda}_{(m)}]'$ is the estimation eigen value of **S**.

From equation (13), it can be concluded that variance the estimation eigen value of Linear and Nonlinear Principal Component Analysis is:

$\text{Var}(\tilde{\hat{\lambda}}_S) = \mathbf{H}_S'\mathbf{V}_S\mathbf{H}_S$.

3.2 Variance the Estimation Eigen Value of Nonlinear Principal Component Analysis

Place the figure as close as possible after the point where it is first referenced in the text. If there is a large number of figures and tables it might be necessary to place some before their text citation.

A sample coming from a multivariate normal distribution population with *n* objects and *m* variable.

If $\tilde{X} = [X_1 \;\; X_2 \;\; \dots \;\; X_j \;\; \dots \;\; X_m]' \sim N_m(\tilde{\mu}, \Sigma)$

For example $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_i, \dots, \tilde{X}_n$ where:

$\tilde{X}_i \sim N_m(\tilde{\mu}, \Sigma)$,

then vector average for the sample the *i* th from *n* objects and *m* buah variables:

$$\bar{\tilde{X}} = \tfrac{1}{n}[\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_i + \dots + \tilde{X}_n]$$

$E(\bar{\tilde{X}}) = \tilde{\mu}$, $\text{Var}(\bar{\tilde{X}}) = \dfrac{\Sigma}{n}$, so that:

$$\bar{\tilde{X}} \sim N_m(\tilde{\mu}, \tfrac{\Sigma}{n})$$

Estimator $\mathbf{\Sigma}$ is S_n or:

$\hat{\boldsymbol{\Sigma}}$ = S_n = variance the estimation eigen value of sample and estimator $\boldsymbol{\Sigma}$ is S, if each element is multiplied by $\frac{1}{n-1}$.

For the variance matrix S_n, where $E(S_n) = \frac{n-1}{n}\Sigma$, S_n is a biased estimator for Σ. In order not to be biased, a correction is made with a factor of in order not to be biased, a correction is made with a factor of C, so that: S = C S_n and $E(S) = C\left(\frac{n-1}{n}\right)\Sigma$; where $C = \frac{n}{n-1}$, then E(S) = Σ, so the unbiased estimator of Σ is S, with elements:

$$s_{ij} = \frac{1}{n-1}\sum_{i=1}^{n}[(X_{ij} - \overline{X}_j)(X_{ij} - \overline{X}'_j)]; \quad i = 1, 2, \ldots, i, \ldots, n; \quad j = 1, 2, \ldots, j, \ldots, m$$

For standardized variables $\tilde{Z} = [Z_1 \ \ Z_2 \ \ldots \ Z_j \ \ldots \ Z_m]'$, write:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}; \ Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}; \ldots; \qquad Z_j = \frac{(X_j - \mu_j)}{\sqrt{\sigma_{jj}}}; \ldots; \ Z_m = \frac{(X_m - \mu_m)}{\sqrt{\sigma_{mm}}}$$

Matrix can be notated:

$$Z = D_\Sigma^{1/2}(\tilde{X} - \overline{\mu})$$

$$E(Z) = 0 \ \text{ and } \ Cov(Z) = D_\Sigma^{-1/2}\,\Sigma\, D_\Sigma^{-1/2} = \rho$$

So for the population	Medium for samples:
$\Sigma = D_\Sigma^{1/2}\ \rho\ D_\Sigma^{1/2}$;	$S = D_S^{1/2}\ R\ D_S^{1/2}$;
$\rho = D_\Sigma^{-1/2}\ \Sigma\ D_\Sigma^{-1/2}$,	$R = D_S^{-1/2}\ S\ D_S^{-1/2}$

When $\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_i, \ldots, \tilde{Z}_n$; $i = 1, 2, \ldots, n$ is a standardized observation:

$\tilde{Z} = [Z_1 \ \ Z_2 \ \ldots \ Z_j \ \ldots \ Z_m]'$; $j = 1, 2, \ldots, m$. $\tilde{Z}_i = D^{-1/2}(\tilde{X}_i - \tilde{\overline{X}})$, where $i = 1, 2, \ldots, n$.

3.3 Variance the estimation eigen value with the Delta Method of Nonlinear Principal Component Analysis

Variance the estimation eigen value with the delta method of nonlinear principal component analysis [13] are:

$$\mathbf{Var[Vec(}\frac{\mathbf{R}}{m}\mathbf{)] =} \ \frac{1}{m^2}\begin{bmatrix} Var(r_{11}) & Cov(r_{11}, r_{12}) & \cdots & Cov(r_{11}, r_{1m}) \\ Cov(r_{11}, r_{12}) & Var(r_{22}) & \cdots & Cov(r_{22}, r_{2m}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(r_{11}, r_{1m}) & Cov(r_{22}, r_{2m}) & \cdots & Var(r_{mm}) \end{bmatrix}$$

$E(Vec(\mathbf{R})) = Vec(E(\mathbf{R})) = Vec(\boldsymbol{\rho})$

$Var[Vec(\mathbf{R})] = Var[Vec(\mathbf{R}^{**})]$

$$= \frac{2}{n-1} M_m(\boldsymbol{\rho} \otimes \boldsymbol{\rho}) = \mathbf{V}_R$$

Therefore: $Vec\left(\frac{\mathbf{R}}{m}\right) \sim N_{m(m+1)/2}[Vec\left(\frac{\boldsymbol{\rho}}{m}\right), \mathbf{V}_R]$

Also found: $$\mathbf{H}_R^{'} = \begin{bmatrix} \mathrm{Vec}^{'}((2\hat{\tilde{\mathrm{d}}}_{1R}\hat{\tilde{\mathrm{d}}}_{1R}^{'}) - \mathrm{diag}(\hat{\tilde{\mathrm{d}}}_{1R}\hat{\tilde{\mathrm{d}}}_{1R}^{'})) \\ \mathrm{Vec}^{'}((2\hat{\tilde{\mathrm{d}}}_{2R}\hat{\tilde{\mathrm{d}}}_{2R}^{'}) - \mathrm{diag}(\hat{\tilde{\mathrm{d}}}_{2R}\hat{\tilde{\mathrm{d}}}_{2R}^{'})) \\ \vdots \\ \mathrm{Vec}^{'}((2\hat{\tilde{\mathrm{d}}}_{mR}\hat{\tilde{\mathrm{d}}}_{mR}^{'}) - \mathrm{diag}(\hat{\tilde{\mathrm{d}}}_{mR}\hat{\tilde{\mathrm{d}}}_{mR}^{'})) \end{bmatrix}$$

and:

$\hat{\lambda}_R = [\hat{\lambda}_{(1)} \ \hat{\lambda}_{(2)} \ \dots \ \hat{\lambda}_{(m)}]^{'}$ is a root estimator of $\frac{1}{m}\mathbf{R}$.

From equation (13) it can be concluded that variance the estimation eigen value of nonlinear principal component analysis $\mathrm{Var}(\hat{\lambda}_R) = \mathbf{H}_R^{'}\mathbf{V}_R\mathbf{H}_R$.

Next compare $\|\mathbf{T_S}\|$ with $\|\mathbf{T_R}\|$, where:

$$\mathrm{Var}(\hat{\lambda}_S) = \mathbf{T}_S = \mathbf{H}_S^{'}\mathbf{V}_S\mathbf{H}_S, \qquad \mathrm{Var}(\hat{\lambda}_R) = \mathbf{T}_R = \mathbf{H}_R^{'}\mathbf{V}_R\mathbf{H}_R$$

and obtained $\|\mathbf{T_S}\| \geq \|\mathbf{T_R}\|$.

Because $\|\mathbf{T_S}\| \geq \|\mathbf{T_R}\|$, then variance the estimation eigen value [var ($\hat{\lambda}_R$)] is better (efficient) than variance the estimation eigen value [var ($\hat{\lambda}_S$)].

This research is in line with [10, 13, and 18].

3.4 Conclusion and further research

With the Delta Method, variance the estimation eigen value of principal component analysis (Linear) are $\mathrm{Var}(\hat{\tilde{\lambda}}_S) = \mathbf{H}_S^{'}\mathbf{V}_S\mathbf{H}_S$ and variance the estimation eigen value of nonlinear principal component analysis are $\mathrm{Var}(\hat{\tilde{\lambda}}_R) = \mathbf{H}_R^{'}\mathbf{V}_R\mathbf{H}_R$. It was also found that variance the estimation eigen value of nonlinear principal component analysis [var ($\hat{\lambda}_R$)] is better (efficient) than variance the estimation eigen value of principal component analysis [var ($\hat{\lambda}_S$)].

Presumably this research can be continued by comparing variance the estimation eigen value of principal component analysis (Linear) and variance the estimation eigen value of nonlinear principal component analysis (Nonlinear) by applying them to real cases in daily life.

References

1. A.C. Rencher, Methods of Multivariate Analysis, 2nd edition. (John Wiley & Sons, New York, 2002)
2. P. Korhonen, and A. Siljamaki, Computational Statistics and Data Analysis, **26**, 411-424 (1998)
3. R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis, 6th edition. (Prentice Hall, New Jersey, 2007)
4. J. Shlens, A Tutorial on Principal Component Analysis (Center for Neural Science, New York University, 2009)

5. Makkulau and A.T. Ampa, *Determination of Factors Influencing Student Achievement with Principal Component Analysis*, Proceedings of the Indonesian Vocational Education National Seminar at UHO, 516 – 522, (2016)
6. A. Gifi, Nonlinear Multivariate Analysis, (Chichester, UK Wiley, 1990)
7. P.M. Kroonenberg, B.D. Harch, K.E. Basford, and A. Cruickshank, Journal of Agricultural, Biological, and Environmental Statistics, **2**, 3, 294 – 312 (1997)
8. J.A. Diaz-Garcia, G. Gonzalez-Farias, and V. Alvarado-Castro Applied Mathematical Sciences, **1**, 22, 1083-1100, (2007)
9. A.C. Rencher, and G.B, Schaalje, Linear Models in Statistics, 2nd edition. (John Wiley & Sons, New York, 2008)
10. S. Bisgaard, and B. Ankenman, Technometrics, **38**, 238-245, (1996)
11. J.R. Schott, Matrix Analysis for Statistics (John Wiley & Sons. Inc, New York, 1997)
12. Makkulau and L.O. Saidi, *Grouping of Production Results of Southeast Sulawesi Agricultural Commodities with Analysis of Main Components and Group Analysis,* Agriplus Unhalu Magazine (National Accredited), Kendari, (2006)
13. Makkulau, *Nonlinear Principal Component Analysis in Mixed Scale Multivariable Data*, ITS Thesis, Surabaya, (2002)
14. R.D. Cook, Technometrics, **42**, 1 (2000)
15. P. Filzmoser, of Statistics, **34**, 2, 127-138 (2005)
16. Makkulau, E. Cahyono, Mukhsar, A. Sani, L.O. Saidi, and A.T. Ampa, Global Journal of Pure and Applied Mathematics/GJPAM, **13**, 6, 2563-2578 (2017)
17. E. Cahyono, Makkulau, L.O., Saidi, and R.Raya, HIKARI Ltd, **4**, 7, 343-352 (2016)
18. Makkulau, G.N.A. Wibawa, A.T. Ampa, A.T.P. Makkulau, S. Harini, and A.D. Mulyanto, CAUCHY–Jurnal Matematika Murni dan Aplikasi, **8**,. 2, 7-15 (2023)
19. A.M. Tamonob, A. Saefuddin, and A.H. Wigena. Indonesian Journal of Statistics **20**, 2, 68-77 (2015)