International Journal of Applied Mathematics and Computing
Vol. 1, No.3 October 2024
e-ISSN: 3047-146X; Page 19-30
DOI ; https://doi.org/10.62951/ijamc.v1i3.26
Available online at: https://international.arimsi.or.id/index.php/IJAMC

Two Stage Lasso in Principal Component Analysis With an Application

Afraa A. Hamada
Department of Statistics, College of Administration and Economics,
University of Al- Qadisiyah, Iraq

Author Correspondence: afraa.hamada@qu.edu.iq

Abstract. *This paper will employ a novel approach that builds upon the lasso method, utilizing it in two stages. The first stage applies to the principal components to select the important principal component and exclude the unimportant ones. This technique is effective in identifying significant principal components while attempting to eliminate bias in selecting these components over others. Additionally, it removes the ranking in determining the principal components compared to classical methods. Moreover, the second stage involves determining the effective importance within each component by zeroing out the scores loading values within each component. To compare the performance of the proposed method in principal component analysis, a simulation approach can be used. Subsequently, the performance of the proposed method is tested using real data.*

Keywords: *Principal Component Analysis, Lasso, Two Stage Lasso, Factor Loading*

1. INTRODUCTION

A potent statistical method for dimensionality reduction, data compression, and exploratory data analysis is principal component analysis (PCA) (Wold, S., Esbensen, K., & Geladi, P. (1987)). We begin with vectors in (p)dimensions, (PCA) is mathematical approach is try to reducing the (p)dimensions into (q)dimensions subspace $(q \leq p)$ (Jolliffe, I. T. (2002)). It converts a large number of variables in a dataset into a smaller collection of variables (called principle components) but yet include the most crucial information. These new variables called principal components are produced by linearly combinations of the original variables in a dataset. It seeks to minimize the dimensionality of the data while capturing the greatest variance in it. Principal components reduce a large number of variables into a smaller set while preserving the important information, allowing for easier analysis and visualization. It assists in lowering the volume of data required for processing while preserving important data. It assists in lowering the volume of data required for processing while preserving important data. The eigenvalues greater than one are used to determine the major primary components. This strategy is regarded as classical and has a lot of disadvantages because it favours eigenvalues larger than 1(Zha, H., & Wang, H. (2002)). Principal components with eigenvalues somewhat less than one are disregarded. Its reliance on ranking is another drawback of this approach. In this paper, we will combine the Lasso approach (Tibshirani, R. (1996)) with principal component models to propose a novel and efficient way for choosing significant principal components. By eliminating the unnecessary principle components and

Received: July 18, 2024; Revised: August 19, 2024; Accepted: September 21, 2024; Online Available: September 25, 2024

keeping the relevant ones, this method allows us to automatically choose the important ones without the need for ranking, and so on. The second stage of using the Lasso technique is to apply this method for selecting important variables within each important component. In this paper, we propose a new approach via achieving components selection and variables selection by utilizing two stage lasso . The current paper is illustrated by using simulation scenario and a real data. This paper is organized as following: In Section two, we briefly review of principal component analyses, two stage lasso with principal component show in section three. In Section three, we use simulation experiments and real data to demonstrate the effectiveness of the proposed approach. Conclude and recommendation the current paper demonstrate in five Section.

2. LITERATURE REVIEW

We briefly review of principal component analyses

It is a statistical method that analyses the data using the primary dimensions. Through the identification of the primary factors (dimensions) that account for the majority of the variation in the data set, this technique seeks to simplify complex data(Jolliffe, I. T. (2002)). The principal component method focuses on reducing high dimensions as well as uncovering data patterns by analyzing the relationships between those dimensions, etc. Assume we have a dataset with (n)observations and(p)variables, which can be represented as a matrix (X)of dimensions($n * p$)Transforming this data to a standardized form can be achieved using the following formula: $z = \frac{x-\mu}{\sigma}$, In this case, the z is the standardized matrix ,the μ is the arithmetic mean and σ is the standard deviation. Before performing Principal Component Analysis (PCA), it is essential to standardize the range of the variable data using min-max scaling. This helps transform the data to a specific range, typically between 0 and 1, using the following formula: $\acute{x} = \frac{x-\min(x)}{\max(x)-\min(x)}$, where $\acute{x}$ is represented the scaled value , x is represented the scaled value, $\min(x)$ is represented the minimum value and $\max(x)$the maximum value. With these two transformations(Hastie, T., Tibshirani, R., & Friedman, J. (2009)), the variable data is now ready for analysis, as it eliminates different measurement units and addresses any outliers. Finding the standardized data covariance matrix might be the first step in performing Principal Component Analysis (PCA):

$$C = \frac{1}{n-1} Z\acute{Z} \qquad (1)$$

Where C is the covariance matrix, z is the standardized matrix. The eigenvalues (τ_i) and related eigenvectors (V) can be found from the covariance matrix (c) previously described.The following steps are used to do this:

$$CV_{ij} = \tau_i V_{ij} \quad (2)$$

The eigenvalues (τ_i) is computed through the characteristic equation $det(C - \tau_i I) = 0$, det is represented the determinant, I is the identity matrix of dimension equal to that of matrix C. V_{ij} is Eigenvectors: Reintroduce each eigenvalue (τ_i) into the equation: $(C - \tau_i I)V = 0$, The principal components' directions in the original feature space are represented by the eigenvectors (V) that result from this. In order to convert the original data into the principal component space and to comprehend the relevance of each principal component, these eigenvalues and eigenvectors are crucial(Zou, H., Hastie, T., & Tibshirani, R. (2006)). Thus, the mathematical model for Principal Component Analysis (PCA) can be represented by the following equation:

$$Z = XV_k \quad (3)$$

Where , Z is converted into the new principal component space from the data matrix , X is the original, standardized data matrix. $V_k$is represents the principal components as an eigenvector matrix ,which ,it corresponding K that has the largest eigenvalues. To determine the important and significant principal components, various methods can be used, including the Kaiser method, which focuses on principal components with eigenvalues greater than or equal to one. Another approach relies on the explained variance ratio, selecting components that account for more than 80% of the total variance, among other methods employed(Kaiser, H. F. (1960)).

Lasso in Principal Component Analysis

Lasso, which stands for Least Absolute Shrinkage and Selection Operator denoted by (Lasso), are especially helpful in statistical modeling . To improve the statistical model it generates' interpretability and prediction accuracy, it selects important principal and applies regularization(Bühlmann, P., & van de Geer, S. (2011)). From the equation (3), we can rewrite the this equation with more detail as following

$$Z_i = a_{ij} V_j \qquad [i = 1,2, \dots p\ ], [j = 1,2 \dots k] \quad (4) \quad ,$$

Principal Component Analysis can be re-modeled using the Lasso technique as a first stage to identify useful and important principal components, while excluding the unimportant components by setting their coefficients to zero((Zou, H., Hastie, T., & Tibshirani, R. (2006))). Therefore, we can rewrite the equation (4) with lasso technique

$$Z_i = a_{ij}V_j + \lambda \sum_{j=1}^{k} |a_{ij}| \qquad (5)$$

Z_i is the principal component,(a_{ij}) are the loadings for the corresponding eigenvectors term of V_j , λ is the regularization parameter which sets the penalties strength and the term$\sum_{j=1}^{k} \lambda |a_{ij}|$ encourages sparsity in the model by shrinking the coefficients towards zero (Tibshirani, R. (1996)).

Loading lasso in Principal Component Analysis

The coefficients that show the connection between the original variables and the principal components are referred to as loadings. They are essential for deciphering PCA results and figuring out the relative contributions of each original variable to each main component(Kaiser, H. F. (1960)) . The eigenvectors of the covariance matrix of the standardized data are where loadings are found. They show how much each initial variable affected the principal omponents (Jolliffe, I. T. (2002)). The loading of the Principal Component Analysis is computed by the formula:

$$L = V\sqrt{\omega} \qquad (6)$$

Where L is represented loadings matrix, V is represented matrix of eigenvectors. $\sqrt{\omega}$ is represented the square roots of the eigenvalues' diagonal matrix. According to the above equation, we observe that all original variables are related to the principal components, but with different loadings that determine the strength or weakness of each variable's contribution to the loading value. where Every loading value indicates how much of a matching original variable went into a certain principal component((Jolliffe, I. T. (2002))). While loadings near zero imply a moderate influence, high absolute values of loadings indicate that the variable has a large influence on that major component.in this paper we proposed a new technique for variable selection within principal component, via using the lasso technique for choosing important variable within one principal component and exclude unimportant variable when its loading closed from zero exactly as the following formula:

$$L = V\sqrt{\omega_i} + \lambda \sum_{j=1}^{k} |\omega_{ij}| \qquad (7)$$

L is stand for the principle component loadings, or loadings matrix((Toczydlowska, D et al 2017).

V is stand the component components matrix, or eigenvector matrix.

$\sqrt{\omega_i}$ is represents the eigenvalue's square root for the i-th principal component more detail ((Jolliffe, I et al 2003)) .

λ is represented the regularization parameter.

The term of $\lambda \sum_{j=1}^{k} |\omega_{ij}|$ is encourage sparsity, a penalty based on the loadings' absolute values is added.

From the equation (7) , the first part $V\sqrt{\omega_i}$ calculates the loadings using the square root of the appropriate eigenvalues of the eigenvectors. This establishes the relative contribution of each original variable to each primary component. but the second part $\lambda \sum_{j=1}^{k} |\omega_{ij}|$ by penalizing lesser values, the second term offers a regularization impact that helps to reduce over fitting and encourages the model to concentrate on the most significant loadings Bühlmann, P., & van de Geer, S. (2011).This concept is helpful in high-dimensional data analysis because it offers a means of controlling complexity through regularization and comprehending the contributions of original variables to the principal components(Zou, H., Hastie, T., & Tibshirani, R. (2006)).

3. RESULTS AND DISCUSSION

Simulation approach

In this section, simulated examples are used to evaluate the effectiveness of our proposed approach **(**Two stage Lasso in Principal Component Analysis)(TS LASSO PCA) . Through these simulations, we are able to assess the method's performance in a variety of scenarios and conditions. giving information about its resilience and efficacy for our proposed method . In order to assess the approach's advantages and disadvantages, we shall compare the outcomes with other methods, where appropriate, with current techniques. The our simulation in current study is implemented in two ways,

First way belong to choosing optimal principal component. in this study, we compered our proposed method with two other methods: First the (Kaiser, 1960) method is based on the eigenvalues greater than one to determine the important principal components. Second BARTLETT method, which it proposed by (BARTLETT 1951), this method is modified by(K.V.MARDIA 1979). For comparison, three criteria were used: Root Mean Square Error (RMSE), explained variance ratio(EVR), and the number of significant principal component(NSPC). The second way belong to selection important original variable within principal component . The matrix of original variables X are generate from multivariate normal

distribution $X \sim N_p(0, \Sigma_p)$. In our simulation three sample size has been used, ($n = 50,150 and\ 250$). There are three simulation examples

Simulation Example

In this simulation example, we simulated ten original variables some these variables have different correlation as following $\rho_1 = 0.45,\ \rho_2 = 0.65\ and\ \rho_3 = 0.99$. From the results show in the below table we see our proposed method have a good performance compared with other two methods. This is because the Root Mean Square Error (RMSE) calculated from our proposed method was significantly lower than that of the other methods. Therefore, our proposed approach demonstrates its superiority in identifying the important principal components. Also, We observe that the extracted variance according to our proposed method was higher than the extracted variance from the comparison methods. From these results, we conclude that our proposed approach has an advantage over the previous methods.

Table -1- The Root Mean Square Error (RMSE) ,explained variance ratio and number of important principal component in First Simulation Example

Sample size	Methods	$\rho_1 = 0.45$			$\rho_2 = 0.65$			$\rho_3 = 0.99$		
		RMSE	EVR	NSPC	RMSE	EVR	NSPC	RMSE	EVR	NSPC
n=50	Kaiser	1.862	61%	5	1.313	65%	5	1.562	64%	5
	BARTLETT	1.661	63%	4	1.223	67%	5	1.473	61%	5
	TSLASSO PCA	1.005	76%	4	0.822	74%	4	**0.683**	72%	5
n=150	Kaiser	1.313	58%	5	1.619	69%	5	1.114	64%	5
	BARTLETT	1.107	63%	6	1.323	64%	4	1.107	67%	5
	TSLASSO PCA	0.823	74%	5	**0.755**	77%	5	0.752	81%	6
n=250	Kaiser	1.764	62%	5	0.627	64%	4	1.473	59%	5
	BARTLETT	1.728	66%	5	0.582	68%	5	1.107	61%	6
	TSLASSO PCA	**0.531**	72%	7	**0.505**	76%	6	**0.626**	64%	5

The results are show in table -1- ,we can summarized via the below table

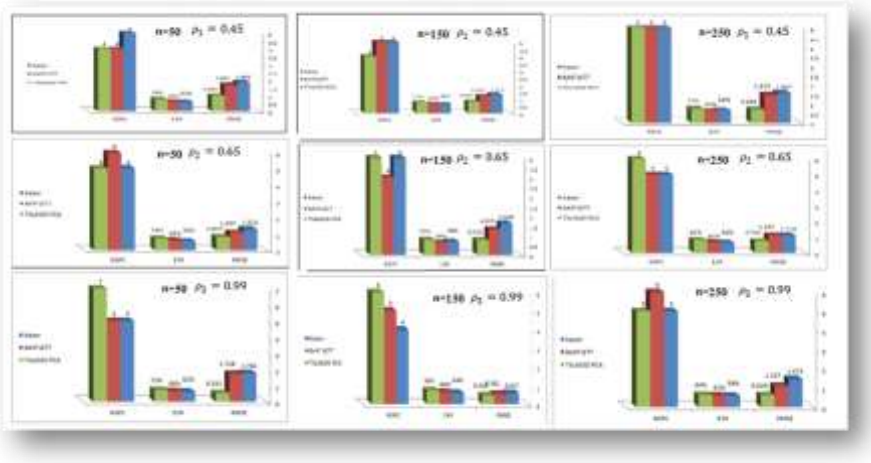

Figure -1- show the root mean square error (RMSE) ,explained variance ratio and number of important principal component

Simulation Example

In this simulation example, we simulated twenty original variables some these variables have different correlation as following $\rho_1 = 0.45,\ \rho_2 = 0.65\ and\ \rho_3 = 0.99$. From the results show in the below table we see our proposed method have a good performance compared with other two methods. This is because the Root Mean Square Error (RMSE) calculated from our proposed method was significantly lower than that of the other methods. Therefore, our proposed approach demonstrates its superiority in identifying the important principal components. Also, We observe that the extracted variance according to our proposed method was higher than the extracted variance from the comparison methods. From these results, we conclude that our proposed approach has an advantage over the previous methods.

Table -2- The Root Mean Square Error (RMSE) ,explained variance ratio and number of important principal component in secon Simulation Example

Sample size	Methods	$\rho_1 = 0.45$			$\rho_2 = 0.65$			$\rho_3 = 0.99$		
		RMSE	EVR	NSPC	RMSE	EVR	NSPC	RMSE	EVR	NSPC
n=50	Kaiser	1.562	65%	5	1.403	63%	6	1.562	57%	5
	BARTLETT	1.131	68%	5	1.234	62%	5	1.362	61%	5
	TSLASSO PCA	0.822	74%	5	**0.987**	71%	5	1.003	69%	4
n=150	Kaiser	1.651	61%	5	1.712	61%	6	1.496	64%	5
	BARTLETT	1.496	66%	4	**1.565**	67%	5	1.205	69%	5
	TSLASSO PCA	**0.964**	72%	4	**0.683**	75%	5	0.923	73%	4
n=250	Kaiser	1.563	64%	5	1.834	61%	7	1.373	71%	6
	BARTLETT	1.472	68%	5	1.607	67%	6	1.014	73%	5
	TSLASSO PCA	0.604	74%	5	0.892	76%	5	0.820	82%	6

The results are show in table -1- ,we can summarized via the below table

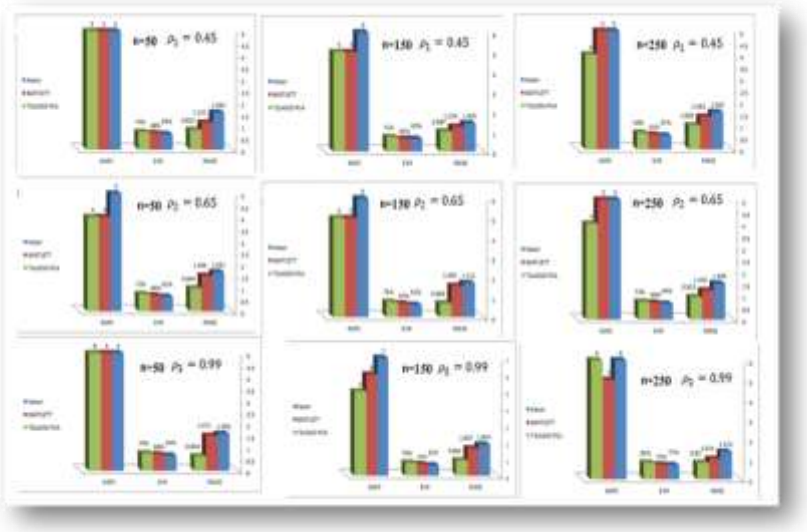

Figure -2- The Root Mean Square Error (RMSE) ,explained variance ratio and number of important principal component

The second way of our simulation

in first and second simulation examples are focus to testing best methods to selection important principal component. However, the selection of the original variable within the principal component is focus our simulation's second part .In this part, we used our proposed method for selection of the original variable within principal component, we choose some simulation scenario as following:

simulation at n=50 $\rho_1 = 0.45$ within first simulation example

First simulation example, there are (10) original variables, we used our proposed method **(**Two stage Lasso in Principal Component Analysis)(TS LASSO PCA) . in this simulation example the number of important principal component is (4), the total explained variance ratio of these important principal component is equal (76%). for selection of original variables as following table

Table-3- show loading scaling of important principal component

Original Variables	Comp1	Comp2	Comp3	Comp4	Comp5
X_1	0.8315	**0.5456**	**0.7654**		**0.5442**
X_2	**0.1744**	**0.8354**	**0.6554**		**0.6932**
X_3	**0.4007**	**0000**	**0000**		**0000**
X_4	**0000**	0.1684	0.1684		0.3421
X_5	**0000**	**-0.7865**	**-0.7865**		**0.2314**
X_6	**0.0828**	**-0.8516**	**-0.8516**		**0.5347**
X_7	**0000**	**0000**	**0000**		**0000**
X_8	**-0.0072**	**0.8028**	**0.8028**		**0.5427**
X_9	**0000**	**0.7743**	**0.7743**		**0.5633**
$X10$	**0.5812**	**0000**	**0000**		**0000**

From the results show in the above table, In first principal component there are 6 important original variables and four unimportant original variables. Also, in second principal component there are 7 important original variables and three unimportant original variables . In third principal component there are 7 important original variables and three unimportant original variables . In fourth principal component there are 7 important original variables and three unimportant original variables . Based on the above results, we find that our proposed method automatically eliminated the insignificant original variables by setting their loadings within the principal component to zero exactly , while retaining the important original variables.

Simulation at n=50 $\rho_1 = 0.45$ within second simulation example

Second simulation example, there are (20) original variables, we used our proposed method **(**Two stage Lasso in Principal Component Analysis)(TS LASSO PCA) . in this simulation example the number of important principal component is (5), the total explained variance ratio of these important principal component is equal (74%). for selection of original variables as following table.

Table-4- show loading scaling of important principal component

Original Variables	Comp1	Comp2	Comp3	Comp4	Comp5
X_1	0.5921	0.6853	0.5091	0.1183	0.3991
X_2	0.6725	0.3907	0.7820	0.4510	0.4872
X_3	**0.1674**	**-0.3419**	**-0.6773**	**0000**	**0.7211**
X_4	**0000**	**0000**	**0000**	**0000**	**0000**
X_5	**-0.3720**	**0.6772**	**0.9834**	**0.3294**	**0.8332**
X_6	0.4229	0000	0000	0000	0000
X_7	0000	0000	0.6733	0.5319	0000
X_8	0000	0000	0000	0000	0000
X_9	-0.6452	-0.3440	0.6745	0.3064	0.7701
$X10$	0.4907	0.7629	0.4542	0.1932	0.2241
X_{11}	**0.5782**	**-0.8546**	**-0.7662**	**-0.7042**	**0000**
X_{12}	**0000**	**0000**	**0000**	**0000**	**0000**
X_{13}	0.4051	0.8789	0.0940	0000	0000
X_{14}	0.5993	0.3923	0.6602	0.7719	0.8741
X_{15}	0000	0000	0000	0000	0000
X_{16}	**0.4637**	**0.6619**	**0.5624**	**0.3772**	**0.4562**
X_{17}	**0.5684**	**0.7638**	**0.3391**	**0.4932**	**0.6511**
X_{18}	**0.2832**	**0.6792**	**0.4997**	**0.4677**	**0.7724**
X_{19}	**0000**	**0000**	**0000**	**0000**	**0000**
X_{20}	**0.8734**	**0.7561**	**0.3307**	**0.9472**	**0.4519**

From the results show in the above table, In first principal component there are 15 important original variables and five unimportant original variables. Also, in second principal component there are 14 important original variables and six unimportant original variables .In third principal component there are 14 important original variables and six unimportant original variables . In fourth principal component there are 12 important original variables and eight unimportant original variables . In fifth principal component there are 11 important original variables and nine unimportant original variables . Based on the above results, we find that our proposed method automatically eliminated the insignificant original variables by setting their loadings within the principal component to zero exactly , while retaining the important original variables.

Real dataset

In this study, we will focus on medical dataset are consisting of 249 observations collected from Marjan Hospital in the city of Hilla. This study aims to identify the most significant variables affecting the increase of Interleukin-27 (IL-27) hormone. There are many variables that affect the increase of Interleukin-27 (IL-27) hormone, such asx_1: White Blood Cells denoted by (WBC), x_2: Monocyte denoted by (MON), x_3: Platelets denoted by (PLT), x_4: Red Blood Cells denoted by (RBC), x_5: Hemoglobin denoted by (HGB), x_6: Hematocrit denoted by (HCT), x_7: Mean Corpuscular Hemoglobin denoted by (MCH), x_8: Mean Corpuscular Hemoglobin denoted by (MCH), x_9: Interferon denoted by (IFN), x_{10}: AGE and x_{11}: sex. To test the effectiveness of our proposed method, it will be compared with the comparison methods using real data as follows:

Table -5- The Root Mean Square Error (RMSE) ,explained variance ratio and number of important principal component in real dataset

Methods	RMSE	EVR	NSPC
Kaiser	1.5621	68%	5
BARTLETT	1.0562	69%	5
TSLASSO PCA	0.9342	79%	4

From the results in the above table we see our proposed method have a good performance compared with other two methods. This is because the Root Mean Square Error (RMSE) calculated from our proposed method was significantly lower than that of the other methods. Therefore, our proposed approach demonstrates its superiority in identifying the important principal components. Also, We observe that the extracted variance according to our proposed method was higher than the extracted variance from the comparison methods. From these results, we conclude that our proposed approach has an advantage over the previous methods. Based on the simulation results and the outcomes from the real data, we find that our proposed method exhibits good performance. Therefore, we will employ the proposed method in data analysis, where we identify that the number of important components is 4 principal components out of a total of 11. These 4 principal components accounted for 79% of the total variance. The significance of the original variables can be illustrated as follows:

Table-6- show loading scaling of important principal component in real dataset

Original Variables	Name Variables	Comp1	Comp2	Comp3	Comp4
X_1	WBC	0.4134	0.5617	0.4625	0.6723
X_2	MON	**0000**	**0000**	**0000**	0000
X_3	PLT	-0.8983	-0.3895	**-0.4728**	**-0.2553**
X_4	RBC	**0000**	**0000**	0000	**0000**
X_5	HGB	0.5073	0.5804	**0.7393**	0.5612
X_6	HCT	-0.7244	-0.6723	**-0.3439**	**-0.4558**
X_7	MCH	**0000**	**0000**	0000	0000
X_8	MCH	**0.7312**	**0.5738**	0.7681	**0000**
X_9	IFN	**0.6513**	**0.4514**	**0.4574**	**0.5612**
$X10$	AGE	0000	0000	**0000**	0000
X_{11}	sex	**0.7553**	**0.553**	0.704	0.7124

The first component has seven important original variables and four unimportant original variables. Also. In the second component has seven important original variables and four unimportant original variables. Also. In the third component has seven important original variables and four unimportant original variables. But the forth component has six important original variables and five unimportant original variables. Based on the above results, we find that our proposed method automatically eliminated the insignificant original variables by setting their loadings within the principal component to zero exactly , while retaining the important original variables.

4. CONCLUSIONS AND RECOMMENDATION

Conclusions

Our proposed method balances bias and variance in selecting the important principal components, as well as in choosing the original variables within the same component. Based on the results presented in the simulation under various correlations. We observe that our proposed method demonstrated outstanding performance in automatically selecting the important principal components by setting the eigenvalues of the insignificant components to zero exactly. This method offers several advantages, including eliminating the bias in the selection of principal components and avoiding the ordering in the selection of principal component. Second stage of our proposed method . The process of selecting original variables focuses on employing the Lasso technique, which selects original variables in a manner that offers several advantages. One of the most significant benefits is that the selection is automatic, leading to time savings. Additionally, the exclusion of insignificant variables occurs by setting the loadings of the original variables within the component to zero exactly.

Recommendations

We recommend using our proposed method in analysis the dataset with high dimensional data. Because of this proposed method has many features for selection principal component and selection of original variables within the same component. We recommend expanding the current study to include a other legalization tools that possess better features than the Lasso technique and have high practical applicability. This would provide significant advantages in selecting the optimal principal component that is highly beneficial for analyzing the studied phenomenon.

REFERENCES

Bartlett, M. S. (1951). The effect of standardization on a chi-squared test. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *13*(2), 190–197.

Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Jolliffe, I. T., Trendafilov, N. T., & Uddin, M. (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, *12*(3), 531–547.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *58*(1), 267–288.

Toczydlowska, D., Peters, G. W., Fung, M. C., & Shevchenko, P. V. (2017). Stochastic period and cohort effect state-space mortality models incorporating demographic factors via probabilistic robust principal components. *Risks*, *5*(3), 42.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1–3), 37–52.

Zha, H., & Wang, H. (2002). Principal component analysis and its applications. In *Proceedings of the IEEE International Conference on Data Mining*.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286.