KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
VOL. 22 (2) (2025), 83-89, p-ISSN: 693-7554, e-ISSN:2654-3990

Application of Principal Component Analysis on Factors Causing Inflation in West Kalimantan

Asri Rahmawati[1*] , Yuyun Eka Pratiwi [2] , Onelia Rochmah [3]

[1,3] Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University, West Kalimantan, 78124, Indonesia
[2] Statistics Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University, West Kalimantan, 78124, Indonesia

Abstract

Inflation is an important indicator in assessing the economic stability of a region. Inflation fluctuations in West Kalimantan are influenced by various economic and structural factors. This study aims to identify the main factors causing inflation in West Kalimantan using Principal Component Analysis (PCA). Secondary data for the 2024 timeframe was obtained from West Kalimantan's Central Statistics Agency (BPS). Economic variables that are suspected of influencing inflation are analyzed using PCA to be reduced to new dominant factors. The main components obtained are then interpreted economically to understand the structure of the causes of inflation. The results of the analysis show that the cumulative proportion of the two components reaches 90%, so the two main components are sufficient to represent the main structure of the data. This means that most of the information from the original variables can be effectively reduced into two main components. PCA successfully reduced the data dimension without losing significant information. The findings indicate that inflation in West Kalimantan is largely influenced by non-food groups, such as housing, education, and other services. These insights can serve as a basis for formulating inflation control policies that focus more on household non-food expenditure sectors.

Keywords: *Inflation; PCA; reduction; economic factors.*

1. Introduction

A major and ongoing increase in the cost of goods and services over a given time period is referred to as inflation. Since inflation is one of the macroeconomic indicators that shows the economic stability of a place, unchecked changes in the rate might lower purchasing power. In order to avoid causing macroeconomic diseases that may subsequently affect economic instability, the inflation growth rate is always tried to be moderate and stable [1]. Changes in the inflation rate can be caused by various underlying factors. In West Kalimantan Province, inflation is influenced by several factors, such as food prices, transportation, exchange rates, distribution of goods, and fiscal policy. In general, these factors are called the Consumer Price Index (CPI). The CPI is the main indicator in measuring inflation, and it consists of various expenditure groups such as food, housing, transportation, education, and other services. The large number of variables that make up the CPI often poses its own challenges in the analysis process, especially when these variables are correlated with each other. This can result in multicollinearity, which interferes with the accuracy of conventional statistical models' interpretations. Therefore, a statistical approach is needed that is able to

*****Corresponding author**. *E-mail address: asri.rahmawati@math.untan.ac.id*
Received: 1 July 2025, Accepted: 31 July 2025 and available online 31 July 2025
DOI: https://doi.org/10.33751/komputasi.v19i2.5260

reduce the dimensions of the data without eliminating important information and can identify hidden patterns from a complex set of variables.

A statistical approach is needed to describe the complexity problem between these variables to reduce the number of variables into simpler main factors but can still describe all existing factors. The Principal Component Analysis (PCA) method is one of the effective dimension reduction techniques for identifying hidden structures in multivariate data [2]. PCA can transform the original correlated variables into a set of uncorrelated principal components, which accounts for much of the variability in the data. As a result, it might be easier to classify and analyze the main causes of inflation.

The use of statistical and computational methods for data-driven decision-making has been widely applied in various fields [3]. Their study implements the Naive Bayes Classifier and Fuzzy Analytical Hierarchy Process (F-AHP) to classify and evaluate books eligible for publishing based on multiple criteria. Although the methods differ from Principal Component Analysis (PCA), the underlying principle of extracting key information from complex data to support structured decision-making is aligned. This illustrates how multivariate techniques, whether through classification models or dimensionality reduction like PCA, play a crucial role in simplifying data complexity for effective analysis and policy formulation.

Research that discusses the factors that affect inflation has been discussed in several articles. [4] discusses the factors that affect inflation with a panel data regression approach. [5] discusses the effect of inflation and economic growth on the unemployment rate in West Kalimantan. [6] discusses modeling the inflation rate using panel data regression. [7] discusses the dominant factors affecting inflation in Indonesia. [8] discusses the application of PCA to the factors that influence the length of completion of student theses. [9] discusses PCA in determining visitor satisfaction factors for digilib library services. [10] discusses the combination of PCA with the K-Means Algorithm for clustering Stunting Data.

The purpose of this research is to use the PCA technique to deconstruct the reasons of inflation in West Kalimantan into a collection of fundamental elements that more clearly and efficiently characterize every variable. Because PCA preserves the majority of the original data while reducing a large number of associated variables into a small number of main components, it is especially helpful. This simplification helps identify the key variables affecting economic circumstances, such inflation, and permits clearer analysis without sacrificing meaning. Based on important characteristics and pertinent data, the analysis's conclusions should help create more focused inflation control measures. Additionally, the goal of this research is to help local governments, academia, and economic stakeholders better and more precisely understand the underlying mechanisms of inflation.

2. Methods

In the early stages of this research, a literature review was conducted on the theories that will be used, namely principal component analysis. The factors influencing inflation in West Kalimantan will be reduced through the application of the Principal Component Analysis approach. Eleven elements influencing inflation in West Kalimantan in 2024 will be the data used in this study. The data was gathered from the 2024 West Kalimantan Province inflation rate and the consumer price index catalog. The details of the variables used in this study are in Table 1:

Table 1. Research Variables

No	Variables	Information	Data types
1.	X_1	Food, Beverages and Tobacco	Numeric
2.	X_2	Clothing and Footwear	Numeric
3.	X_3	Housing, Water, Electricity and Household Fuels	Numeric
4.	X_4	Household supplies, equipment, and routine maintenance	Numeric
5.	X_5	Health	Numeric
6.	X_6	Transportation	Numeric
7.	X_7	Information, Communication, and Financial Services	Numeric
8.	X_8	Recreation, Sports and Culture	Numeric
9.	X_9	Education	Numeric
10.	X_{10}	Provision of Food and Drinks/Restaurant	Numeric
11.	X_{11}	Personal Care and other services	Numeric

Data analysis was carried out using R-Studio and SPSS software. The following are the research stages carried out in this study [11].

a. Describing data characteristics using descriptive statistics.
b. Conducting a data suitability test using Bartlet and Kaiser-Mayers-Oklin (KMO).
c. Forming covariance and correlation matrices.
d. Calculating eigenvalues and eigenvectors.
e. Determine the number of principal components by using eigenvalues > 1 and the proportion of cumulative variance to the total or by using *a scree plot*.
f. Interpreting research results.

Principal component analysis was first proposed by *K Pearson* as a tool for fitting fields with orthogonal least squares, then developed by *T Hotelling* for the purpose of analyzing correlation structures [12]. The objectives of principal component analysis are as follows:

a. Reduce variables by preserving as much of the original data information as possible.
b. To obtain new variables that are orthogonal.
c. Transforming correlated independent variables into new independent variables so that there is no correlation between the independent variables.

Suppose a random vector is as follows:

$$x' = [x_1 \ x_2 \ \dots \ x_p]$$

With the covariance matrix Σ and having a pair of eigenvalues and eigenvectors $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. In the principal component analysis, pa linear combination is constructed from $x$namely

$$y_i = e_i' x = e_{i1}x_1 + e_{i2}x_2 + \cdots + e_{ip}x_p$$

$y_1, y_2, \dots, y_p$ are the main components, where they are not correlated with each other and have the greatest possible variance. So that

$$Var\ (y_i) = e_i' \textstyle\sum e_i, \quad \text{For} i = 1,2, \dots, p$$
$$cov(y_i, y_k) = e_i' \textstyle\sum e_k, \ \text{For} i, k = 1,2, \dots, p, i \neq k$$

The first main component:
A linear combination $e_1'x$ that maximizes var $(e_1'x)$ with $e_1'e_1 = 1$
The second main component:
The linear combination $e_2'x$ that maximizes var $(e_2'x)$ with $e_2'e_2 = 1$and cov $(e_1'x, e_2'x) = 0$. This means that the second principal component is uncorrelated with the first principal component.
The i-th principal component
The linear combination $e_i'x$ that maximizes var $(e_i'x)$ with $e_i'e_i = 1$and cov $(e_i'x, e_k'x) = 0$for $k < i$. This means that the -th principal component $i$is uncorrelated with the previous principal components.
There are several ways to determine $k$the main components, including [13]:

a. Using eigenvalues > 1
b. Using the cumulative proportion of variance to total variance. The cumulative proportion of the original independent variable variance explained by $k$the principal components is at least 80% and the proportion of the population variance is quite large.
c. Using a scree plot, namely a plot between $i$with λ_i, the selection of values $k$based on the scree plot is determined by looking at the occurrence of bends by deleting the main components that produce several small eigenvalues forming a straight-line pattern. If the curve is still steep, then additional components can still be added, conversely if the curve is already sloping, the determination of the number of components is complete even though the determination of the slope or steepness of the scree plot is subjective by each person.

3. Results and Discussion

This section will explain the stages of data analysis using the principal component analysis method. The explanation starts from descriptive statistics to the interpretation of the research results obtained.

3.1 Descriptive Statistics

This analysis makes use of monthly data on the West Kalimantan Province's consumer price index (CPI) evolution in 2024. Since the CPI is used to calculate inflation over time, the factors in the CPI have a significant impact on inflation. The CPI variables used in this study include Food, Beverages and Tobacco (X_1), Clothing and Footwear (X_2), Housing, Water, Electricity and Household Fuels (X_3), Household supplies, equipment and routine maintenance (X_4), Health (X_5), Transportation (X_6), Information, Communication and Financial Services (X_7), Recreation, Sports and Culture (X_8), Education (X_9), Food and Beverage Provision/Restaurants (X_{10}), Personal Care and other services (X_{11}). The CPI data unit is an index number, such as 2018=100, that indicates how the average level of a group of goods and services consumed by households changed during a given time period in comparison to a base year. This study uses the base year 2022=100 as its CPI. In January 2024, a value of 102.53 for the variable X_5 indicates a 2.53% improvement in health over the 2022 average. The data's descriptive statistics are displayed in Table 2.

Table 2. Descriptive Statistics of 11 Factor Variables Affecting Inflation in West Kalimantan 2024

Variables	Minimum	Maximum	To the track	Media	Variance
X_1	107.45	110.70	108.68	108.38	1.12
X_2	102.50	103.89	102.89	102.75	0.40
X_3	102.02	103.89	102.80	102.75	0.50
X_4	102.88	103.21	103.06	103.09	0.10
X_5	102.35	103.01	102.58	102.54	0.16
X_6	111.15	112.84	111.77	111.78	0.43
X_7	100.13	100.62	100.38	100.39	0.13
X_8	102.79	104.69	103.90	104.28	0.83
X_9	101.43	104.35	102.70	101.93	1.22
X_{10}	103.49	104.56	104.05	104.75	0.36
X_{11}	106.59	109.93	108.33	108.34	1.19

From Table 2, it can be seen that the variable X_1representing Food, Beverages and Tobacco has a minimum CPI value of 107.45 and a maximum of 110.70. The average CPI during the observation period $X_1$was recorded at 108.68 with a median value of 108.38. The variance of the variable $X_1$is 1.12 indicating fluctuations in this price group.

3.2 Data Feasibility Test

Before conducting analysis using the PCA method, a data feasibility test is needed to ensure that the variables in the data are correlated strongly enough. PCA aims to simplify the data by combining interrelated variables into several main components. If the variables are not correlated, PCA cannot produce the best final results [14].

a. Test *Bartlett's Test of Sphericity*

To determine if the correlation matrix is an identity matrix, the correlation between the variables in the sample is tested using Bartlett's test of sphericity. Table 3 displays the outcomes of the Bartlett's test of sphericity conducted with Rstudio.

Table 3. *Bartlett's Test of Sphericity* Results

Statistical Test	Mark
Chi-Square	159.31
Degrees of Freedom (df)	55
p *-value*	< 2.22e-16

Since the p-value is less than 0.05 according to the test results, it can be said that the correlation matrix differs from the identity matrix and that there is a strong enough correlation among the variables to move on to the PCA method stage.

b. KMO (*Kaiser-Mayer Olkin*) and MSA (*Measure of Sampling Adequacy*) tests

KMO and MSA tests were conducted to measure the sampling adequacy of each variable [3]. When the p-value of MSA of all variables is > 0.05, all variables can be included in the principal component analysis. The following KMO test results can be seen in Table 4.

Table 4. KMO Test Results

Statistical Test	Mark
p -*value*	0.58

Results from the test indicate that the p-value is 0.58, which is larger than 0.05, allowing the variables to be collected and processed further. Each variable is also assessed using the MSA test to determine which should be removed and which can be processed further. The MSA test results for each variable are shown in Table 5.

Table 5. MSA Test Results

Variables	*p-value*
X_1	0.24
X_2	0.75
X_3	0.71
X_4	0.47
X_5	0.48
X_6	0.23
X_7	0.82
X_8	0.51
X_9	0.66
X_{10}	0.72
X_{11}	0.63

Based on the test results, there are still several variables with values < 0.50, namely X_1, X_4, X_5, and X_6so so these four variables will not be included in the data analysis using PCA. The KMO and MSA tests will be repeated without including the variables X_1, X_4, X_5 and and X_6. Based on the retest, the KMO value was 0.7083 and the MSA test results for each variable were more than > 0.50. So the variables used for analysis using the PCA method are the variables $X_2, X_3, X_7, X_8, X_9, X_{10}$,and X_{11}.

3.3 Principal Component Analysis

Determining the covariance matrix that will be used to gauge the strength of the link between variables is the primary stage in the PCA approach [15]. The data's covariance matrix, which is displayed in Table 6.

Table 6. Covariance Matrix

Variables	X_2	X_3	X_7	X_8	X_9	X_{10}	X_{11}
X_2	0.1609	0.1900	-0.0438	0.2017	0.4103	0.1189	0.3919
X_3	0.1900	1.2596	-0.0456	0.1899	0.5085	0.1716	0.5539
X_7	-0.0438	-0.0456	0.0181	-0.0878	-0.1169	-0.0344	-0.1168
X_8	0.2017	0.1899	-0.0878	0.6927	0.5760	0.1506	0.5335
X_9	0.4103	0.5085	-0.1169	0.5760	1,5042	0.3577	0.2489
X_{10}	0.1189	0.1716	-0.0344	0.1506	0.3577	0.1336	0.4321
X_{11}	0.3919	0.5539	-0.1168	0.5335	1.2489	0.4321	0.4321

To make determining the primary components easier, the eigenvalue decomposition will be attempted once the covariance matrix has been obtained [12]. Table 7 displays the outcomes of the eigenvalue decomposition.

Table 7. Decomposition of Nilai Eigen

Components	Nilai Eigen	Proportion Variants (%)	Proportions Cumulative
1	5.5358	0.7908	0.7908
2	0.8182	0.1169	0.9077
3	0.2559	0.0365	0.9443
4	0.2238	0.0319	0.9762
5	0.1560	0.0222	0.9985
6	0.0076	0.0011	0.9996

7	0.0022	0.0003	1,0000

Table 7 show that one variable or component has a cumulative proportion of 79.08% and an eigenvalue greater than one. The eigenvalue for components 2–7, however, is less than 1. The next step is to look for the eigenvector, which is the primary component's coefficient. Table 8 displays the outcomes of the eigenvector value computation.

Table 8. Eigenvectors

Variables	Component						
	K_1	K_2	K_3	K_4	K_5	K_6	K_7
X_2	0.3992	0.0220	-0.6144	-0.2103	0.2261	-0.5954	-0.1124
X_3	0.3943	0.3321	-0.1980	-0.2332	0.3824	0.7022	0.0337
X_7	- 0.3652	0.4324	0.1379	0.3505	0.6975	-0.2219	-0.0437
X_8	0.2937	-0.7487	0.2745	0.1288	0.5098	0.0309	0.0165
X_9	0.3830	0.0846	-0.2146	0.8541	-0.2060	0.0729	0.1505
X_{10}	0.3952	0.2843	0.4849	-0.1782	-0.0237	-0.2937	0.6393
X_{11}	0.4031	0.2314	0.4556	0.0232	-0.1133	-0.1021	-0.7433

Component 1 absorbs the majority of the total variable, followed by component 2 absorbing the majority of the remaining variance, and so on, according to the eigenvector values in Table 8. Along with the eigenvalue and cumulative proportion, the scree plot can be used to determine the number of components that were produced. Figure 1 displays an image of the scree plot that was obtained.

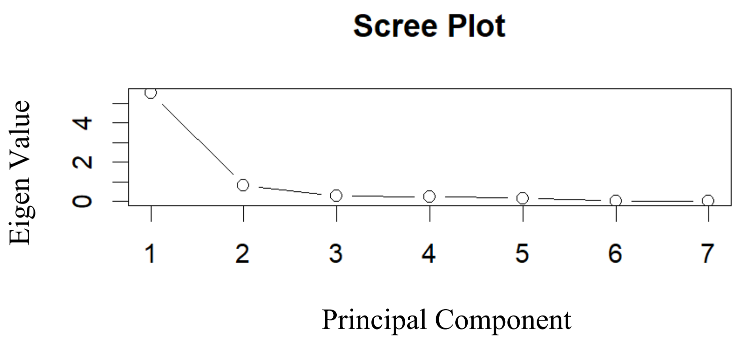

Figure 1. Scree Plot of Principal Component Analysis

Based on Figure 1, the connecting line between variables with other variables decreases sharply at a value of 1 on the y-axis. In addition, there are 2 components whose eigenvalues are greater than one. So it can be concluded that the number of components formed is 2. When viewed from the cumulative proportion, a minimum range of 70-90% can be taken so that the components formed can also be concluded as many as 2 components.

4 Conclusion

Based on the results of the principal component analysis of the seven selected CPI variables, it was found that two principal components were able to explain 90.77% the total variance of the data. The first component contributed the largest proportion of variance, 79.08%while the second component contributed 11.69%. The scree plot also showed a sharp decline after the second component, indicating that two PCA components were sufficient to represent information from the seven variables. PCA successfully reduced the data dimension without losing significant information. This shows that inflation in West Kalimantan is largely influenced by non-food groups, such as housing, education, and other services. This can also be used as a basis for formulating inflation control policies that are more focused on household non-food expenditure sectors.

References

[1] A. Salim, F. Fadilla, A. Purnamasari, “The Effect of Inflation on Indonesia’s Economic Growth,” *Economic Sharia: Journal of Islamic Economic Thought and Development*, Vol. *7* (1), pp. 17-28. 2021. [Online]. Available: https://ejournal.stebisigm.ac.id/index.php/esha/article/view/268#google_vignette

[2] I.T. Jolliffe, J. Cadima. “Principal component analysis: a review and recent developments.” *Philos Trans A Math Phys Eng Sci* . 2016. doi: 10.1098/rsta.2015.0202. [On line]. Available: https://pubmed.ncbi.nlm.nih.gov/26953178/

[3] M.D. Irwansyah, T.P. Negara, Erniyati, and P. Citra, Application of the Naïve Bayes Classifier Methode and Fuzzy Analitical Hierarchy Process in Determining Books Eligible for Publishing. *Komputasi: Jurnal ilmiah Ilmu Komputer dan Matematika*, vol. 21, no. 1 , pp. 55-67 , 2024. [Online]. Available : https://journal.unpak.ac.id/index.php/komputasi/article/view/6677

[4] B.E. Huruta and G. Sasongko, Faktor-faktor yang mempengaruhi inflasi daerah: Sebuah pendekatan regresi data panel. *Jurnal Ilmiah pendidikan Ekonomi*, Vol. 5, no. 2. 2020. [Online]. Available: https://journal.upgris.ac.id/index.php/equilibriapendidikan/article/download/7137/pdf

[5] A. Muharrahman, Pengaruh Inflasi dan pertumbuhan ekonomi terhadap tingkat pengangguran di kalimantan Barat. Jurnal pembangunan dan pemerataan, Vol. 6, no. 1. 2017. [Online]. Available: https://jurnal.untan.ac.id/index.php/jcc/article/view/17748

[6] D. Apriliawan, tarno, and H. Yasin, pemodelan laju inflasi di provinsi jawa tengah menggunakan regresi data panel. Vol. 2, No. 4, pp. 301-321. 2013 [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/article/download/3791/3676

[7] S. Djambak. Jurnal Ekonomi Pembangunan, Vol. 6, no. 1, pp. 23-38. 2008. [Online]. Available: https://jep.ejournal.unsri.ac.id/index.php/jep/article/download/4844/2591/10558

[8] M Wangge. Penerapan Metode principal Component Analysis (PCA) terhadap Faktor-faktor yang mempengaruhi lamanya penyelesaian skripsi mahasiswa program studi pendidikan matematika FKIP UNDANA. Vol. 5, No.2 , 2021. [Online]. Available: https://j-cup.org/index.php/cendekia/article/view/465

[9] S. manullang, D. Aryani, and H. Rusydah. Analisis Principal Component Analysis (PCA) dalam penentuan faktor kepuasan pengunjung terhadap layanan perpustakaan digilib. Vol. 7, no. 1, pp. 123-130. 2023. [Online]. Available: https://e-journal.hamzanwadi.ac.id/index.php/edumatic/article/view/14839/pdf

[10] G. F. Ibanez, G. W. Wiriasto, and Rosmaliati. Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunning. Vol. 5, no. 1, pp. 131-141. 2024. [Online]. Available: https://djournals.com/klik/article/view/1977/1139

[11] M. Noya van Delsen, A. Wattimena, and S. Saputri, “USE OF PRINCIPAL COMPONENT ANALYSIS METHOD TO REDUCE INFLATION FACTORS IN AMBON CITY”, *BAREKENG: J. Math. & App.* , vol. 11, no. 2, pp. 109-118, Dec. 2017. [Online]. Available: https://ojs3.unpatti.ac.id/index.php/barekeng/article/view/352

[12] R.A. Johnson, D.W. Wichern. 2002. Applied Multivariate Statistical Analysis, 5th edition. Pearson Education International.

[13] D.I. Purnama. “Principal Component Analysis on Sub-district Potential Data in Palu City Before the Earthquake and Tsunami Disaster of September 28, 2018”, *Journal of Mathematics, Statistics and Computation* , Vol *16* (1), pp. 25–32, 2019. [Online]. Available: http://journal.unhas.ac.id/index.php/jmsk/article/view/6329

[14] M. Irfan. (2024). “Principal Component Analysis on Diabetes Data”, *Sciencestatistics: Journal of Statistics, Probability, and Its Application* , Vol. *2* (2), pp. 48-58, 2024. [Online]. Available: https://scholar.ummetro.ac.id/index.php/sciencestatistics/article/view/5729

[15] Mariana. “Principal Component Analysis”, *Journal of Mathematics and Its Learning* , Vol 2(2), pp. 99-114, 2013. [Online]. Available: https://media.neliti.com/media/publications/293541- analis-komponen-utama-d5d8c65a.pdf