Volume 8 No 1 June (2025)

e-ISSN: 2615-3270
p-ISSN: 2615-3599

Eigen Mathematics Journal

Homepage Journal: https://eigen.unram.ac.id/index.php/eigen

Analysis of Factors that Influence Poverty in West Nusa Tenggara Using Principal Component Regression

Lisa Harsyiah[1*], Zulhan Widya Baskara[2], Dina Eka Putri[3], Rifdah Fadhilah[4]

[1, 2, 3] *Department of Statistics, Universitas Mataram, Indonesia*
[4] *Department of Mathematics, Universitas Mataram, Indonesia*
[*] *Corresponding author: lisa_harsyiah@unram.ac.id*

ABSTRACT

West Nusa Tenggara (NTB) is one of the provinces in Indonesia with a percentage of poor people according to the March-September period in 2019, namely $14.56\% - 13.88\%$, while in 2020 it was $13.97\% - 14.23\%$ and in 2021 the percentage was $14.14\% - 13.83\%$. The factors suspected of influencing poverty in each province have different conditions each year, so repeated observations are needed on poverty data and the factors that influence it. If the data contains multicollinearity, then one of the classic assumptions of multiple linear regression is not met so that the problem of multicollinearity needs to be addressed. The Principal Component Regression (PCR) method is the most consistent compared to the ridge and least square regression methods in solving multicollinearity problems. This study aims to analyze poverty in NTB using the PCR method. The data used in this study are the number of poor people and factors influencing poverty based on districts in NTB in 2020 – 2022. Based on the calculation results, it was obtained that Component 1 with an eigenvalue of 4.008 explained 57.2% of the variance, while Component 2 with an eigenvalue of 1.740 explained 82.1% of the variance. Both components significantly affect poverty according to the results of simultaneous and partial tests. This model has an R^2 value of 0.302 or 30.2% and the remaining 69.8% is influenced by external factors (error). The R^2 value is classified as a weak category and it is recommended to add other factors that affect poverty including access to electricity, access to sanitation, access to clean drinking water, and government spending.

Keywords: Poverty, Multicollinearity, Principal Component Regression

Received	:	01-07-2024;
Revised	:	08-04-2025;
Accepted	:	08-05-2025;
Published	:	11-06-2025;

DOI: https://doi.org/10.29303/emj.v8i1.229

1. Introduction

Poverty is a global problem that can be found in almost all countries in the world, both developed and developing countries. Indonesia is one of the developing countries with a fairly high poverty rate. Based on data released by the BPS–Statistics Indonesia in 2021, the number of poor people in Indonesia reached 26.50 million people. Even though this number decreased by 0.01 million people from the number of poor people in 2020, the number of poor people in Indonesia is still relatively high.

West Nusa Tenggara (NTB) is one of the provinces in Indonesia which has a percentage of poor people according to the March-September period in 2019, namely $14.56\% - 13.88\%$, while in 2020 it is $13.97\% - 14.23\%$

and in 2021 percentage of 14.14% – 13.83% (BPS, 2022). Based on the percentage of poor people since 2019 – 2021, the poverty level is still very high, so both central and regional governments need to eradicate poverty in NTB, so it is necessary to conduct research on the factors that significantly influence poverty. Factors thought to influence poverty include gross regional domestic product (GRDP), literacy rates, unemployment, government spending: health and education [1]. The factors that are thought to influence poverty in each province have different conditions each year, so repeated observations of poverty data and the factors that influence it are needed. Furthermore, to determine the relationship pattern or influence of predictor factors on poverty, a statistical method is needed, namely regression analysis [2].

Regression analysis is a statistical analysis used to determine the pattern of relationships between variables, namely predictor variables (X) and response variables (Y). Data analysis that can be used on response variables is quantitative data type, namely linear regression. There are two types of linear regression, namely simple linear regression and multiple linear regression. Simple linear regression is used to determine the relationship or influence of a predictor variable on the response variable. Meanwhile, multiple linear regression is used to determine the relationship or influence that more than one predictor variable has on the response variable [3]. The regression analysis used to determine the factors that significantly influence poverty uses multiple linear regression analysis because it has more than one predictor variable. If the data contains multicollinearity, then one of the classic assumptions of multiple linear regression is not fulfilled so the multicollinearity problem needs to be overcome [4]. There are several methods for resolving multicollinearity, including least squares, Principal Component Regression (PCR), ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), and latent roots [5]. However, in applying these methods it was found that Ridge Regression and PCR were the best methods for resolving multicollinearity because they had the smallest Average Mean Square Error (AMSE) values compared to other methods [6].

The PCR method is a combination of Principal Component Analysis (PCA) and regression analysis. PCR works by reducing the initial independent variables into new variables called principal components which are mutually independent, then regressing these principal components to obtain a regression model. Ridge regression is a modification of the least squares method that works by adding a bias constant to the main diagonal of the variance-covariance matrix. Ridge regression and PCR both obtain analysis results that have smaller variances [7]. Based on research conducted by [8] entitled "Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis" found that the Principal Component Regression (PCR) method was the most consistent compared to the ridge and least square regression methods in solving multicollinearity problems. This is because all the independent variables studied are significant to the dependent variable.

Given the many factors that influence poverty, the problem of multicollinearity often occurs when it involves many variables. For example, research by [9] explains that there is a problem of multicollinearity in the variables of electricity access, sanitation access, access to clean drinking water, GRDP, and government spending. The five variables used are factors that influence the percentage of poor people in Java-Bali and West Nusa Tenggara. Although research related to poverty has been conducted, the use of Principal Component Regression (PCR) to identify the factors that most influence poverty in NTB is a more modern and appropriate approach. This method not only overcomes the problem of multicollinearity, it will also provide clearer results about which variables are most significant in explaining variations in poverty. Therefore, this study aims to analyze poverty in NTB using the PCR method with the title "Analysis of Factors Affecting Poverty in NTB Using Principal Component Regression".

2. Research Methods

2.1. Data Source

The data used in this research is the number of poor people and factors that influence poverty according to districts/cities in NTB in 2020 – 2022. This data is secondary data sourced from the Central Statistics Agency (BPS) of NTB Province.

The research data is part of the dependent variable and several independent variables. These variables are presented in Table 1 as follows.

Table 1. Research Variable

Symbol	Variable	Unit
Y	Percentage of Poor People	Percentage (%)
X_1	Population Density	Souls/Km^2
X_2	Human Development Index	Percentage (%)
X_3	Open Unemployment Rate	Percentage (%)
X_4	Expenditure Per Capita	Thousand Rupiah/Person
X_5	GRDP Per Capita at Current Prices	Rupiah/Person
X_6	GRDP Per Capita at Constant Prices	Rupiah/Person
X_7	Mean Years of Schooling	Year

2.2. Research Procedures

This study uses the Principal Component Regression (PCR) method. The analysis steps are as follows:

a. Determine the data description, namely the average and standard deviation of poverty data and the factors that influence it.

b. Standardizing research data using centralization and scaling methods. Standardization is a method that transforms research variables into the form [10]:

$$y_i^* = \frac{y_i - \bar{Y}}{S_Y}, \quad i = 1, 2, \ldots n. \tag{1}$$

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}, \quad i = 1, 2, \ldots n, \quad j = 1, 2, \ldots k. \tag{2}$$

with

$$S_Y = \sqrt{\frac{\sum_i (Y_i - \bar{Y})^2}{n-1}} \quad \text{and} \quad S_j = \sqrt{\frac{\sum_i (X_{ij} - \bar{X}_j)^2}{n-1}}$$

Based on the variables that have been standardized, the general form of the multiple linear regression equation model is obtained as follows [7]:

$$y_i^* = b_1 Z_{i1} + b_2 Z_{i2} + \cdots + b_k Z_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{3}$$

c. Carry out parameter estimation using the Ordinary Least Square (OLS) method through the following equation [7]:

$$\hat{\beta} = (X^t X)^{-1} X^t Y \tag{4}$$

where $\hat{\beta}$ estimated parameter β_j using the OLS method.

d. Detecting multicollinearity, if multicollinearity is idetected then it is resolved using the PCR method. Testing is detected through the Variance Inflation Factor (VIF) [11]:

$$VIF = \frac{1}{1 - R^2} \tag{5}$$

where R^2 is the coefficient of determination. If the $VIF \geq 10$, then multicollinearity occurs between the independent variables.

e. Determine the eigenvalues (λ_j) of the correlation matrix (ρ) using the following equation [12]:

$$|\rho - \lambda I = 0| \tag{6}$$

f. Determine the principal component score (K) formed using the following equation [12]:

$$(\rho - \lambda I) a_j = 0 \tag{7}$$

g. Conducting multiple linear regression of the dependent variable against the principal components. Thus, the principal component regression model is obtained as follows [13]:

$$Y = w_0 + w_1 K_1 + w_2 K_2 + \cdots + w_m K_m \tag{8}$$

The selected principal components $K_1, K_2, \ldots, K_m$ are linear combinations of the standardized variables as follows.

$$\begin{aligned} K_1 &= a_1^t Z = a_{11} Z_1 + a_{12} Z_2 + \cdots + a_{1k} Z_k \\ K_2 &= a_2^t Z = a_{21} Z_1 + a_{22} Z_2 + \cdots + a_{2k} Z_k \\ &\vdots \\ K_m &= a_m^t Z = a_{m1} Z_1 + a_{m2} Z_2 + \cdots + a_{mk} Z_k \end{aligned} \tag{9}$$

h. Testing the significance of PCR model parameters simultaneously and partially. The explanation is as follows [7]:

- Simultaneous Test
  The following is the hypothesis of the simultaneous test for the F test.

  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$
  $H_1 : \exists \beta_i \neq 0, (i = 1, 2, 3, \ldots, k)$

  Simultaneous test statistics using the F test are as follows:

  $$F_{\text{count}} = \frac{\frac{SSR}{k}}{\frac{SSE}{(n-k-1)}} = \frac{MSR}{MSE} \tag{10}$$

  Decision-making criteria, namely if $F_{\text{count}} > F_{(\alpha;df1;df2)}$ or p value $< \alpha$, then reject H_0, which means that the independent variable simultaneously influences the dependent variable.
- Partial Test
  The following is the hypothesis of the partial test for the t test.

  $H_0 : \beta_j = 0, \quad j = 1, 2, 3, \ldots, k \quad$ (j parameter is not significant)
  $H_1 : \beta_j \neq 0, \quad j = 1, 2, 3, \ldots, k \quad$ (j parameter is significant)

  Partial test statistics using the t test are as follows:

  $$t_{\text{count}} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{11}$$

  Decision-making criteria, namely if $|t_{\text{count}}| > t_{(\alpha;df)}$ or p value $< \alpha$, then reject H_0, which means that the independent variable partially influences the dependent variable.

i. Changing the multiple linear regression equation into the standard variable form, the following model is obtained [13]:

$$Y = w_0 + w_1 K_1 + w_2 K_2 + \cdots + w_m K_m + \varepsilon \tag{12}$$

j. Changing the multiple linear regression equation back into its original variable form, the following model is obtained [13]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{13}$$

k. Conducting classical assumption tests of multiple linear regression including normality tests, heteroscedasticity tests, and autocorrelation tests. The explanation is as follows:

- Normality Test
  The Kolmogorov-Smirnov test hypothesis testing is as follows.

  H_0 : Error is normally distributed ($\mu = 0$)
  H_1 : Error is not normally distributed ($\mu \neq 0$)

  The test statistics for the Kolmogorov-Smirnov test are as follows:

$$D = \text{Max}|F_0(\varepsilon_i) - S_N(\varepsilon_i)| \tag{14}$$

  Decision-making criteria, namely if the value of $D \leq D_{table}$ or p value $> \alpha$ then it fails to reject H_0, meaning the error is normally distributed [14].
- Heteroscedasticity Test
  The Glejser test hypothesis testing is as follows.

  H_0 : No heteroscedasticity occurs
  H_1 : Heteroscedasticity occurs

  The test statistics for the Glejser test are as follows:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{15}$$

  Decision-making criteria, namely if the value of $|t_{\text{count}}| < t_{(\alpha;df)}$ or p value $> \alpha$ then fail to reject H_0 [14].
- Autocorrelation Test
  The Durbin-Watson test hypothesis testing is as follows.

  $H_0 : \rho_1 = \rho_2 = \cdots = \rho_n = 0 \quad$ (There is no autocorrelation)
  $H_1 : \rho_i \neq 0, \quad i = 1, 2, 3, \ldots, n \quad$ (There is autocorrelation)

  The Durbin Watson method test statistics are as follows:

$$d = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} \tag{16}$$

  Decision making criteria, namely if $dU < d < 4 - dU$ then fail to reject fail to reject H_0 meaning there is no autocorrelation [11].

l. Conduct a goodness of fit test of the regression model formed from the PCR method by looking at the value of the determination coefficient obtained. The test uses the following equation [15]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{17}$$

where the R^2 value has a range between 0 and 1. A regression model is said to be appropriate if the R^2 value approaches 1, and conversely, a regression model is increasingly inappropriate when the R^2 value approaches 0.

3. Result and Discussion

3.1. Descriptive Statistics

A complete explanation of the data description is presented in Table 2 below

Table 2. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Y	30	14.66	190.84	73.0823	51.82699
X_1	30	77	7,189	1,096.03	2,055.518
X_2	30	65.80	80.67	71.5473	4.16375
X_3	30	0.38	6.83	3.5053	1.44703
X_4	30	18,566	32,839	23,205.73	4,098.946
X_5	30	15,152	207,626	40,495.40	46,949.075
X_6	30	10,577	134,274	27,550.10	31,869.650
X_7	30	5.91	10.94	7.9360	1.46796

Table 2 provides information regarding the amount of data for each variable, totaling 30 data, with minimum, maximum, average and data distribution values. The standard deviation value is used as a measure of whether the data is distributed homogeneously or not. Variables with high standard deviations are variables X_1, X_4, X_5, and X_6 indicating high data variation, which means non-homogeneous data distribution. In contrast, variables with small standard deviations are variables X_2, X_3, and X_7 have data that is more centered around the mean, indicating higher stability or homogeneity in the variables.

3.2. Data Standardization

The data used has different units, so the first step is to standardize the data. The process is carried out by transforming the initial variables into standard variables using the average value and standard deviation. The standard variables are symbolized $Y^*, Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7$.

3.3. Parameter Estimation

Estimating the parameters of the standardized variables using the Ordinary Least Square (OLS) method. The results of the parameter estimation calculations are presented in Table 3 below.

Table 3. OLS Method Parameter Estimation

Variable	β_j Estimate
Z_1	-0.189
Z_2	2.507
Z_3	0.203
Z_4	-1.420
Z_5	1.321
Z_6	-1.412
Z_7	-2.280

Table 3 explains the role of coefficients in the OLS method which shows the direct influence of each independent variable on the dependent variable. The variables that have a positive influence on the Y variable are standard variables Z_2, Z_3, and Z_5. While the variables that have a negative influence on Y are standard variables Z_1, Z_4, Z_6, and Z_7.

3.4. Multicollinearity Test

Multicollinearity can be detected using the correlation coefficient or VIF value. If a VIF ≥ 10 is obtained, it is free from multicollinearity problems. The results of the VIF calculation are presented in Table 4 below.

Table 4. Multicollinearity Test Results

Variable	VIF
Z_1	4.542
Z_2	45.233
Z_3	3.691
Z_4	16.095
Z_5	331.065
Z_6	336.343
Z_7	19.377

Table 4 explains that the variables human development index (Z_2), expenditure per capita (Z_4), GRDP per capita at current prices (Z_5), GRDP per capita at constant prices (Z_6), and mean years of schooling (Z_7) have values $VIF > 10$, then the regression model experiences multicollinearity problems. Therefore, to overcome this problem, the Principal Component Regression (PCR) method was used.

3.5. Correlation Matrix

The correlation matrix shows the level of correlation or relationship between independent variables. An explanation of the correlation matrix results can be seen as follows.

$$\rho = \begin{bmatrix} 1.000 & 0.715 & 0.583 & 0.765 & -0.028 & -0.028 & 0.320 \\ 0.715 & 1.000 & 0.723 & 0.783 & 0.323 & 0.321 & 0.819 \\ 0.583 & 0.723 & 1.000 & 0.785 & 0.471 & 0.482 & 0.463 \\ 0.765 & 0.783 & 0.785 & 1.000 & 0.379 & 0.378 & 0.336 \\ -0.028 & 0.323 & 0.471 & 0.379 & 1.000 & 0.998 & 0.318 \\ -0.028 & 0.321 & 0.482 & 0.378 & 0.998 & 1.000 & 0.317 \\ 0.320 & 0.819 & 0.463 & 0.336 & 0.318 & 0.317 & 1.000 \end{bmatrix}$$

The correlation formed has a weak to very strong level, where there is also a positive or negative relationship. Then to determine the eigenvalue, you can use the correlation matrix. The results of the calculation of Equation (6) will obtain the eigenvalue.

3.6. Determine Principal Component

Determining the number of principal components formed is to look at eigenvalues that are greater than one. Based on the eigenvalues, the total diversity value for each component is also obtained. The eigenvalues and total diversity are presented in Table 5 below.

Table 5. Eigenvalues and Total Diversity

Component	Eigenvalues	Total Diversity
K_1	4.008	0.572
K_2	1.740	0.821
K_3	0.834	0.940
K_4	0.267	0.978
K_5	0.137	0.997
K_6	0.013	0.999
K_7	0.001	1.000

Table 5 shows the number of principal components that can be formed. The results obtained were 2 components that could be formed from the seven variables that had been analyzed, namely 4.008 and 1.740. The first eigenvalue of $4.008 > 1$ then becomes Component 1 and explained 57.2% of the variation. Meanwhile, the second eigenvalue is $1.740 > 1$, so it becomes Component 2 and explained 82.1% of the variation. Apart from that, a visualization of the main components formed can be seen in Figure 1 below.

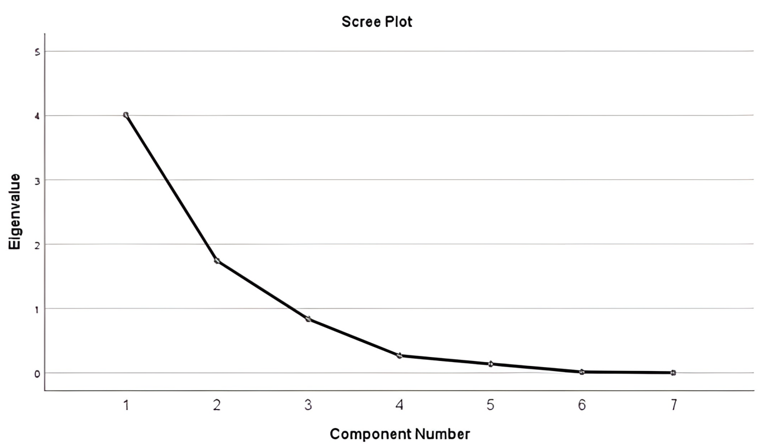

Figure 1. Scree Plot Eigenvalues

The scree plot image can show the many components that are formed. The method is to look at the component values that have eigenvalues > 1. The results in Figure 1 show two component points that have eigenvalues > 1. Therefore, it can be interpreted that there are two main components formed.

3.7. Determine Principal Component Scores

The principal component scores will be used to form the principal component regression equation through the calculation of eigenvectors. The eigenvectors are obtained using Equation (7) by substituting the eigenvalues of the two principal components. Then the results of the eigenvector calculations are displayed in Table 6 below.

Table 6. Principal Component Scores

Variable	K_1	K_2
Z_1	0.667	-0.630
Z_2	0.911	-0.282
Z_3	0.874	-0.048
Z_4	0.866	-0.231
Z_5	0.614	0.776
Z_6	0.616	0.777
Z_7	0.681	-0.053

Table 6 explains that the main component scores obtained are the relationship between the standard variable Z_j and the principal components. This relationship can be written into the following equation.

$$K_1 = 0.667Z_1 + 0.911Z_2 + 0.874Z_3 + 0.866Z_4 + 0.614Z_5 + 0.616Z_6 + 0.681Z_7 \quad (18)$$

$$K_2 = -0.630Z_1 - 0.282Z_2 - 0.048Z_3 - 0.231Z_4 + 0.776Z_5 + 0.777Z_6 - 0.053Z_7 \quad (19)$$

3.8. Regressing Principal Component

After obtaining the main component scores, the new variables K_1 and K_2 will be used for regression analysis to obtain the parameter estimates in Table 7 below.

Table 7. Parameter Estimation

Model	Unstandardized Coefficients	
	B	Std. Error
Constant	$-6.600e^{-16}$	0.158
K_1	-0.331	0.161
K_2	-0.439	0.161

Based on the parameter estimation results, the Principal Component Regression (PCR) model is obtained as follows:

$$Y = -6.600e^{-16} - 0.331K_1 - 0.439K_2 \quad (20)$$

3.9. Parameter Significance Test

The parameter significance test results consist of two tests, namely simultaneous and partial tests.

a. Simultaneous Test

The simultaneous test uses the F test with the following hypothesis:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1 : $\exists\beta_i \neq 0, \quad (i = 1, 2, 3, \ldots, k)$

The results of simultaneous test calculations are presented in Table **??** below.

Table 8. Simultaneous Test Results

	df	F_{count}	p value
Regression	2	5.852	0.008
Residual	27		
Total	29		

Table 8 shows that the $F_{(0.05;2;27)}$ value is 3.354 with $\alpha = 5\%$. Comparing the value of $F_{\text{count}} > F_{(0.05;2;27)}$, namely $5.852 > 3.354$, the decision is to reject H_0. Therefore, the conclusion shows that simultaneously the independent variables have a significant influence on the dependent variable.

b. Partial Test

The partial test uses the t test with the following hypothesis:

$H_0 : \beta_j = 0, \quad j = 1, 2, \cdots, k \quad$ (j parameter is not significant)
$H_1 : \beta_j \neq 0, \quad j = 1, 2, \ldots, k \quad$ (j parameter is significant)

The results of the partial test calculations are presented in Table 9 below.

Table 9. Partial Test Results

Variable	t_{count}	$\lvert t_{\text{count}}\rvert$	p value	VIF
Constant	0.000	0.000	1.000	
K_1	-2.062	2.062	0.049	1.000
K_2	-2.730	2.730	0.011	1.000

Table 9 shows a $t_{(0.05;27)}$ value of 2.052 with $\alpha = 5\%$. Comparison of the values of $|t_{\text{count}}| > t_{(0.05;27)}$, namely $2.062 > 2.052$ and $2.730 > 2.052$, then the decision is to reject H_0. Therefore, the conclusion shows that partially the independent variable has a significant influence on the dependent variable. Apart from that, the VIF value is also displayed where there is no multicollinearity so that the PCR method can eliminate multicollinearity between independent variables by producing principal components that are independent of each other.

3.10. Regression Equation in Standard Variables

Returns the principal component regression model to an equation containing the following standard variables:

$$Y = -6.600e^{-16} + 0.0558Z_1 - 0.1777Z_2 - 0.2682Z_3 - 0.1852Z_4 - 0.5439Z_5 - 0.5450Z_6 - 0.2022Z_7 \tag{21}$$

Equation (21) is used in the initial stage of PCR, where regression is performed on the principal components (Z_i). Furthermore, the standard variables Z_i are returned to the initial variables X_i using a linear combination to make them easier to interpret as in equation (22).

3.11. Regression Equation in Initial Variables

Returning the equation into the initial variable, namely X_i, as follows:

$$Y = -6.600e^{-16} + 2.7146X_1 - 0.0425X_2 - 0.1853X_3 - 4.5182X_4 - 1.1585X_5 - 1.7101X_6 - 0.1377X_7 \tag{22}$$

Equation (22) explains that the population density variable (X_1) increases the amount of poverty in West Nusa Tenggara. Therefore, emphasis is needed to reduce the amount of poverty in West Nusa Tenggara. While other variables can reduce the amount of poverty in West Nusa Tenggara, it is necessary to increase the variables human development index (X_2), open unemployment rate (X_3), expenditure per capita (X_4), GRDP per capita at current prices (X_5), GRDP per capita at constant prices (X_6), and average years of schooling (X_7).

3.12. Classical Assumption Test

The results of testing errors through several classic assumption tests are as follows.

a. Normality Test

Normality testing uses the Kolmogorov Smirnov test with the following hypothesis:

H_0 : Error is normally distributed ($\mu = 0$)
H_1 : Error is not normally distributed ($\mu \neq 0$)

The results of the normality test are shown in Table 10 below.

Table 10. Normality Test Results

Test Statistical Value	p value
0.138	0.153

Table 10 shows a value of $D_{\text{table}} = 0.242$ with $\alpha = 0.05$. If the value of $D < D_{\text{table}}$ is compared, namely $0.138 < 0.242$, then the decision fails to reject H_0. Apart from that, the p-value shows a result of $0.153 > 0.05$, so the decision fails to reject H_0. Thus, the conclusion is that the error is normally distributed.

b. Heteroscedasticity Test

Heteroscedasticity testing uses the Glejser test with the following hypothesis:

H_0 : No heteroscedasticity occurs
H_1 : Heteroscedasticity occurs

The results of the heteroscedasticity test are shown in Table 11 below.

Table 11. Heteroscedasticity Test Results

Variable	t_{count}	p value
K_1	0.324	0.749
K_2	-1.241	0.225

Table 11 shows the value $t_{(0.05;27)} = 2.052$ with $\alpha = 0.05$. Comparing the value of $t_{\text{count}} < t_{(0.05;27)}$, namely $0.324 < 2.052$ and $-1.241 < 2.052$, then the decision fails to reject H_0. Apart from that, the p value of 0.749 and 0.225 indicates that if it is greater than 0.05, the decision fails to reject H_0. Thus, the conclusion is that heteroscedasticity does not occur or the error variance of the method used is consistent.

c. Autocorrelation Test

Autocorrelation testing uses the Durbin-Watson test with the following hypothesis:

$H_0 : \rho_1 = \rho_2 = \cdots = \rho_n = 0$ (There is no autocorrelation)
$H_1 : \rho_i \neq 0, \quad i = 1, 2, 3, \ldots, n$ (There is autocorrelation)

The results of the autocorrelation test are shown in Table 12 below.

Table 12. Autocorrelation Test Results

d Value	Durbin Watson Table Values	
	dL	dU
1.690	0.748	1.814

Table 12 explains that the Durbin-Watson test from the principal component regression method obtained a value $d_U < d < 4 - d_U$, namely $0.748 < 1.690 < 2.186$ so that the decision failed to reject H_0, which gave the conclusion that there was no autocorrelation.

3.13. Goodness of Fit Test

Test the goodness of the regression model by looking at the coefficient of determination (R^2) which is shown in Table 13 below.

Table 13. Model Goodness Results

R	R^2	R^2_{adj}	SSE
0.550	0.302	0.251	0.86560713

Table 13 explains the R^2 value of 0.302 or 30.2% and the remaining 69.8% is external factors (error). The Adjusted R Square (R^2_{adj}) value of 0.251 indicates that the variation in the independent variable is able to explain the variation in the dependent variable by 25.1%, while the remaining 74.9% explains other factors that were not studied. Std. The Error of the Estimate (SEE) is 0.86560713, where the smaller the SEE value, the more accurate the regression model makes predictions.

According to [16], the R^2 value is categorized as weak if it is more than 0.19 but lower than 0.33. This study obtained the R^2 value using PCR on poverty data in NTB which is categorized as weak, meaning that the independent variable is not dominant in influencing the dependent variable. So, there are other variables that contribute to the dependent variable. Like the research conducted by [9] explains that the factors that influence the percentage of poor people in Java-Bali and NTB are access to electricity, access to sanitation, access to clean drinking water, GRDP, and government spending. Therefore, new variables can be added besides GRDP in further research.

4. Conclusions

Based on the results of the analysis conducted, several conclusions were obtained as follows:

a. Two main components were formed from the 7 initial variables representing the diversity of data, namely Component 1 with an eigenvalue of 4.008 explaining 57.2% of the variance, while Component 2 with an eigenvalue of 1.740 explaining 82.1% of the variance. These two components significantly influence poverty according to the results of simultaneous and partial tests.

b. Regression model using PCR on poverty data in NTB

$$Y = -6.600e^{-16} + 2.7146X_1 - 0.0425X_2 - 0.1853X_3 - 4.5182X_4 - 1.1585X_5 - 1.7101X_6 - 0.1377X_7$$

c. The accuracy of the model obtained an R^2 value of 0.302 or 30.2% and the remaining 69.8% was influenced by external factors (error). The R^2 value is classified as a weak category and it is recommended to add other factors that affect poverty including access to electricity, access to sanitation, access to clean drinking water, and government spending.

Acknowledgement

Thank you to the Faculty of Mathematics and Natural Sciences, University of Mataram for facilitating this research to be completed.

References

[1] R. Rahmawati, L. Harsyiah, and Z. Baskara, “Comparison of ridge regression and principal component regression in overcoming multicollinearity of factors affecting poverty in indonesia,” in *International Conference on Mathematics: Pure, Applied and Computation*, pp. 41–51, Springer, 2023. https://doi.org/10.1007/978-981-97-2136-8_4.

[2] D. N. Gujarati, *Basic Econometrics*. Boston: McGraw Hill, 4th ed ed., 2003.

[3] A. Astuti, *Partial Least Square (PLS) and Principal Component Regression (PCR) for Linear Regression with Multicollinearity in the Case of the Human Development Index in Gunung Kidul Regency*. PhD thesis, Universitas Negeri Yogyakarta, Yogyakarta, 2014.

[4] S. Pujilestari, N. Dwidayati, and S. , “Pemilihan Model Regresi Linier Berganda Terbaik pada Kasus Multikolinieritas berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise,” *UNNES Journal of Mathematics*, vol. 6, no. 1, pp. 61–71, 2017. https://journal.unnes.ac.id/sju/ujm/article/view/11719.

[5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY: Springer New York, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[6] N. Herawati, K. Nisa, E. Setiawan, N. , and T. , “Multicollinearity, LASSO, Ridge Regression, Principal Component Regression,” *International Journal of Statistics and Applications*, vol. 8, pp. 167–172, 2018. http://article.sapub.org/10.5923.j.statistics.20180804.02.html.

[7] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021. https://www.kwcsangli.in/uploads/3--Introduction_to_Linear_Regression_Analysis_-_5th_ed._Douglas_C._Montgomery_-_Elizabeth_A._Peck_-_and_G._.pdf.

[8] S. Çankaya, S. Eker, and S. H. Abacı, “Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis,” *Turkish Journal of Agriculture - Food Science and Technology*, vol. 7, no. 8, pp. 1166–1172, 2019. https://doi.org/10.24925/turjaf.v7i8.1166-1172.2515.

[9] S. N. B. Sitepu, T. Haryanto, and N. M. Sukartini, “Faktor-Faktor Yang Mempengaruhi Persentase Penduduk Miskin Pulau Jawa-Bali dan Nusa Tenggara Barat,” *PROSENAMA*, vol. 2, no. 1, pp. 103–113, 2022. https://prosenama.upnjatim.ac.id/index.php/prosenama/article/view/32.

[10] M. H. Kutner, C. Nachtsheim, J. Neter, and W. Li, eds., *Applied Linear Statistical Models*. The McGraw-Hill/Irwin Series Operations and Decision Sciences, Boston: McGraw-Hill Irwin, 5th ed ed., 2005.
[11] N. R. Draper and H. Smith, *Applied Regression Analysis*. Wiley Series in Probability and Statistics, Wiley, 3 ed., 1998. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118625590.
[12] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ: Pearson/Prentice Hall, 6. ed ed., 2007.
[13] A. Mattjik and I. Sumertajaya, *Sidik Peubah Ganda Dengan Menggunakan SAS*. IPB PRESS, 2011.
[14] I. Ghozali, *Aplikasi Analisis Multivariete SPSS 23*. Semarang: Badan Penerbit Universitas Diponegoro, 8 ed., 2016.
[15] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics, New York, NY: Springer US, 2021. https://link.springer.com/book/10.1007/978-1-4614-7138-7.
[16] W. W. Chin, “The Partial Least Squares Approach to Structural Equation Modeling,” in *Modern Methods for Business Research*, Methodology for Business and Management, pp. 295–336, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 1998.