

# Application of Principal Component Regression in Analyzing Factors Affecting Human Development Index

Sumarni Susilawati<sup>1</sup>, Didiharyono<sup>2</sup>

<sup>1</sup>Universitas Muhammadiyah, Makassar, Indonesia

<sup>2</sup>IBK Nitro, Makassar, Indonesia

## Article Info

### Article history:

Received : 09-09-2022

Revised : 05-23-2023

Accepted : 05-25-2023

### Keywords:

Economic;  
Education;  
Human Development Index;  
Multicollinearity;  
Principal Component Regression.

## ABSTRACT

The human development index is an indicator to measure the quality of people's lives. If the human development index number increases, the better the quality of people's lives. There are many factors or variables that affect the level of the human development index, ranging from economic issues, education, health and other factors. However, not all factors have a positive and significant effect. Thus, this study aims to determine the factors that significantly affect the human development index in South Sulawesi. The method used in this study is principal component regression which involves many variables. The variables involved are expected length of schooling, average length of schooling, percentage of population with the highest Diploma, Bachelor and Masters education, school enrollment rate for people aged 7-24 years, percentage of poor people, spending per capita, and life expectancy. From the results of data processing using principal component analysis, 4 main components are obtained which represent the other components, for principal component regression, taking into account the cumulative proportion of > 80%. The results of this study indicate that the human development index in South Sulawesi is influenced by all the variables involved, which is equal to 95.7%. With the variable percentage of poverty being one of the variables that has a negative effect on HDI in South Sulawesi which shows that the higher the percentage of poverty, the lower the human development index. Thus, in order to increase the human development index in Indonesia, it is necessary to take strategic steps to improve people's welfare.



Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020

DOI: <https://doi.org/10.30812/varian.v6i2.2366>

### Corresponding Author:

Sumarni Susilawati,  
Universitas Muhammadiyah, Makassar, Indonesia.  
Email: [sumarnisusilawati8@gmail.com](mailto:sumarnisusilawati8@gmail.com)

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## A. INTRODUCTION

The very important role of human resources in the creation of a development which aims to create a healthy community environment and live a productive life. To achieve sustainable development, Human resources must be able to develop and optimize their capabilities. In a simple sense, development can be interpreted as an effort or process to carry out change for the better. In its implementation, development has various the complexity of the problem. The development process occurs in all aspects of people's lives, both economic, political, social and cultural aspects (Mahran, 2012).

In 1990 the United Nations Development Program (UNDP) for the first time introduced the concept of the Human Development Index (HDI) or the Development Index Humans, where this concept combines the life expectancy index, education index and purchasing power index.

The human development index is an indicator to measure the quality of life or community welfare. A nation and a country is said to be advanced when it has a human development index that continues to increase. Because the human development index (HDI) has an indicator of knowledge, and a decent standard of living. So to know or calculate the human development index can be seen

from three things, namely education, health and economy. For education, it can be measured from the school participation rate, the expected length of school and the average length of schooling as well as several other things that support education. Meanwhile, from the economic sector, it can be seen in terms of the work participation rate, poverty rate, unemployment rate, per capita expenditure and also the Gini ratio. While the health factor can be measured from the human life expectancy.

Based on data from the Central Statistics Agency for South Sulawesi, the HDI in South Sulawesi in 2021 increased by 0.31% from the previous year. Where it currently reaches 72.24%, which is accompanied by an increase in the school participation rate for children aged 7 years which reaches 13.52%. This figure is almost the same as the population aged 25 years and has completed education at the diploma and undergraduate level as much as 13.45% (Badan Pusat Statistika., 2022).

Other factors that can also affect the HDI in South Sulawesi are the per capita expenditure figures, the percentage of poverty and unemployment. The rate of per capita expenditure rose to 0.95%, and the percentage of poverty also increased by 0.11%. Meanwhile, the unemployment rate decreased by 0.59%.

Research related to the human development index was carried out by Winanda (2021) with the results of research that, the percentage of poor people and economic growth, has a significant negative effect on the development index in South Sulawesi. On the other study (Hasanah et al., 2021) conducted research on the human development index in South Sulawesi using a nonparametric spline regression, with the results of the research that the participation rate of work rate, student and school ratio, population density, health facilities and gross regional domestic product greatly influence the increase in the human development index in South Sulawesi. Furthermore, (Humaira and Nugraha, 2018) the influencing variables HDI are Life Expectancy (AHH), Adjusted Per Capita (Expenditure), School Average (RLS), School Expectation (HLS), and Gross Regional Domestic Product at Constant Price (GRDP).

Referring to the existing data, there are many factors related to the increase in HDI. However, it is necessary to know which factors are more dominant in supporting the improvement of the HDI in South Sulawesi, so that plans and strategies for increasing the HDI can be more well planned and on target. To find out, further analysis using statistical methods is needed. The method used in research involving many factors or multivariable is multiple regression and factor analysis which is divided into several methods. One of them is principal component analysis (PCA).

Several previous studies only involved two to six independent variables which were considered to have an effect on the human development index, so this study will involve more factors which are considered to have an effect on the development index in South Sulawesi, such as old school expectations, average length of schooling, percentage of population with highest education Diploma, undergraduate and postgraduate, school enrollment rate for population aged 7-24 years, gini ratio, percentage of poor people, expenditure per capita, labor force and life expectancy. So that principal component regression (PCR) will be used, which is a combination of principal component analysis and regression analysis. This method is used because it involves many factors or variables that allow multicollinearity between variables to occur. Or these variables have a very close correlation which will produce a large standard error that affects the results of the analysis. So several studies suggest using PCR to overcome multicollinearity problems in data.

## 1. Human Development Index

The Human Development Index (HDI) explains how the population can access development outcomes in terms of income, health, education and so on. HDI is formed on three basic dimensions, namely a long and healthy life, knowledge and a decent standard of living. Economic development is an important topic for developed countries and developing in search of growth in production and consumption. There are several benefits for countries that want to improve economic development through investment human development. The main benefit is to improve the welfare of citizens country (Ridha and Parwanto, 2020).

Local government responsible for the needs of the population working on development human resources by allocating funds to the education and health sectors. Cost training is used to build and maintain facilities and infrastructure for guide quality investment models (Muliza et al., 2017).

Government spending can be interpreted as a financial plan that describes policy choices for a certain period in the future will come. Consumption pattern describes budget allocation based on structure certain (Astri et al., 2013). Depending on the form of spending used, provincial government spending is called the selected budget. According to In the Indonesian budget method, there are two types of government spending: daily expenditures and development expenditures. Daily expenses are operational expenses government and payment of wages to employees and other expenses. Shopping Development is expenditure which is classified as government investment expenditure. Including investment in education and health (Astri et al., 2013).

HDI was introduced by the association of nations published by the Unit Nation Development Program (UNDP) in 1996. In the publication it was written that human development is a process to improve aspects of people's lives. The aspect of

life in question is an adequate level of education, healthy living and a decent life. HDI is one of the important indicators in seeing the increase in human resource development and measuring the government's performance in improving the quality of education, health, and other services. As a country that is committed to increasing HDI, Indonesia uses HDI data as a reference for determining general allocation funds in all sectors that affect human development.

The indicators used in measuring HDI are the expected length of school and the average length of school as an indicator of knowledge. For indicators of decent living standards, per capita GNI data is used. Meanwhile, for indicators of healthy life and longevity, data on life expectancy at birth are used.

(Meirianti, 2016) conducted an analysis using the fixed effect model of HDI data in 38 Regencies / Cities in East Java Province in 2010-2014. The results show that the level of poverty, health spending and economic spending have a positive and significant impact, education spending has a positive but not significant effect on the HDI.

## 2. System Equilibrium Point

Regression analysis is a statistical method used to identify variables that affect other variables. In regression analysis, the terms dependent variable and independent variable are known. Cases that have one dependent variable and one independent variable can be analysed using simple linear regression analysis. Meanwhile, cases that have two or more independent variables can be solved using multiple regression analysis.

The multiple linear regression equation model is

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k + \varepsilon \quad (1)$$

To calculate the regression coefficient, two methods can be used, namely the ordinary least square.

### Calculating $\beta_1$ and $\beta_2$ using ordinary least squares method

#### Calculating $\alpha$ , $\beta_1$ and $\beta_2$

$$\beta_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1)(\sum x_2) - (\sum x_1 x_2)^2} \quad (2)$$

$$\beta_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1)(\sum x_2) - (\sum x_1 x_2)^2} \quad (3)$$

$$\alpha = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 \quad (4)$$

### Multicollinearity

Multicollinearity is the occurrence of a linear relationship between independent variables in the regression equation. Multicollinearity causes the estimator to have a large variance, making it difficult to get an accurate estimate (Faizia et al., 2019). To find out what happens in the regression model, a multicollinearity test is performed to see if there is a high or perfect correlation between the independent variables. A good regression model should have no correlation between variables. If there is a high correlation between the independent variables, then the relationship between the independent variables and the dependent variable is disrupted.

Multicollinearity between independent variables can be detected by using the Variance Inflation Factors values with the following conditions:

1. If VIF value  $> 10$  and Tolerance value  $< 0.10$  then there is multicollinearity
2. If VIF value  $< 10$  and Tolerance value  $> 0.10$  then there is not multicollinearity.

The following is the equation for calculating the VIF value:

$$VIF = \frac{1}{1 - R_j^2} \quad (5)$$

Where  $R_j^2$  is the coefficient of determination between the independent variables on the regression equation model. If the independent variable  $X_j$  does not collaborate with other independent variables, then the value of  $R_j^2$  will be small and the VIF value will be close to 1 (Widiyawati, 2021).

### 3. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method used to reduce or simplify data to facilitate data interpretation (Maubanu and Kartiko, 2018). The advantage of this PCA method is that it can be used for all conditions of research data, and there is no need to reduce the number of original variables. However, please note that PCA is not the way to the end. PCA is often used as an input to other methods such as multiple regression and cluster analysis.

Linear algebraically Principal component is a special linear combination of  $p$  random variabel  $X_1, X_2, X_3, \dots, X_p$ . Geometrically this linear combination is a new coordinate system obtained from the rotation of the original system with  $X_1, X_2, X_3, \dots, X_p$  as the coordinate axes. The new axis is the direction with maximum variability and gives a simpler covariance.

Principal component depends entirely on the covariance matrix  $\sum$  or correlation matrix  $\rho$  (Tazliqoh et al., 2015). Suppose the random vector  $X' = X_1, X_2, X_3, \dots, X_p$  have a covarian matrix  $\sum$  with eigen value  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , then a linear combination is formed:

$$\begin{aligned} Y_1 &= a'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k \\ Y_2 &= a'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2k}X_k \\ &\vdots \\ Y_k &= a'_k X = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kk}X_k \end{aligned} \quad (6)$$

Principal component development does not require normal multivariate assumptions. But to get a normal multivariate can use interpretation.

To find variance  $Y_i$  and covariace  $(Y_j, Y_k)$  used equation:

$$Var(Y_j) = a'_j \sum a_j, \quad j = 1, 2, \dots, k \quad (7)$$

$$Cov(Y_j, Y_k) = a'_j \sum a_j, \quad j, k = 1, 2, \dots, k \quad (8)$$

Principal component is a linear combination  $Y_1, Y_2, \dots, Y_p$  which has no correlation. Next, if we use the covariance matrix  $\sum$  with paired eigenvectors  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_i, e_i)$  where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ . Until the Principal component to  $i$  is,  $Y_j = e'_j X = e_{j1}X_1 + e_{j2}X_2 + \dots + e_{jk}X_k, \quad j = 1, 2, \dots, k$  then

$$Var(Y_j) = e'_j \sum e_j, \quad j = 1, 2, \dots, k \quad (9)$$

$$Cov(Y_j, Y_k) = e'_j \sum e_j, \quad j, k = 1, 2, \dots, k \quad (10)$$

If several  $\lambda_i$  are equal, then the appropriate vector choice is  $e_i$  then  $Y_i$  not single. This formula is an uncorrelated principal component equation and has the same variance to the eigenvalues in the covariance matrix  $\sum$ .

$Y_1$  is the first component that satisfies the maximum value of  $e'_1 \sum e_1 = \lambda_1$ .  $Y_2$  is the second component that satisfies the remaining variance other than the first component by maximizing the value of  $e'_2 \sum e_2 = \lambda_2$ ,  $Y_p$  is the  $p$ -th component that satisfies the remainder of the variance other than  $Y_1, Y_2, \dots, Y_{k-1}$  by maximizing value  $e'_k \sum e_k = \lambda_k$ . Order  $Y_1, Y_2, \dots, Y_k$  must fulfill  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ .

Meanwhile, the total variance of the principal component population is

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{kk} = \sum_{i=1}^k Var(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_k = \sum_{i=1}^k Var(Y_i) \quad (11)$$

So the proportion of the total population variance described by the principal component is

$$\left( \begin{array}{l} \text{the proportion of the} \\ \text{total variance explained} \\ \text{by the principal component to } k \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_k}, \quad j = 1, 2, \dots, k \quad (12)$$

If the total variance of the population is greater than 80% for a large number of variables, it can be explained by two or three components, then these components can replace the original (original) variables, without losing a lot of information.

If  $Y_1 = e'_1 X, Y_2 = e'_2 X, \dots, Y_k = e'_k X$  is the Principal component that is formed from the covariance matrix, then

$$\rho_{Y_i, X, K} = \frac{e_{jk} \sqrt{\lambda_k}}{\sqrt{\sigma_{kk}}} \quad j, k = 1, 2, \dots, k \tag{13}$$

Is the correlation coefficient between components  $Y_i$  and variable  $X_k$ , where  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_j, e_j)$  eigen value and eigen vector pairs.

**4. Principal Component Regression**

Principal Component Regression (PCR) is a combination of regression analysis and PCA. The data were first analyzed using PCA. From the results of the analysis, new variables will be obtained which are the result of a linear combination of the previous variables. The resulting new variables are free from multicollinearity or not correlated with one another.

Principal component regression equation based on covariance matrix or correlation matrix, namely  $X_1, X_2, \dots, X_k$  replaced with a standardized variable, namely  $Z_1, Z_2, \dots, Z_k$ . Both equations will be in accordance with the observed variables.

If the number of principal components formed is given the notation  $K_1, K_2, \dots, K_n$ . Then the general form of the principal component equation is

$$Y = \alpha + \beta_1 K_1 + \beta_2 K_2 + \dots + \beta_k K_k \tag{14}$$

The principal component is a linear combination of the standard variables  $Z$ , then the form of the equation is :

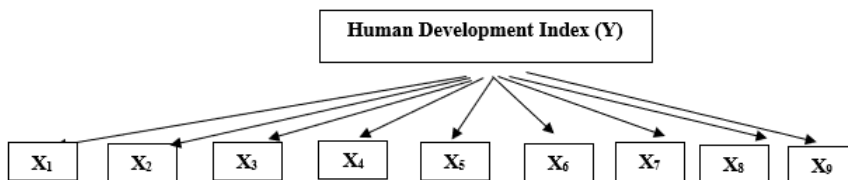
$$\begin{aligned} K_1 &= a_{11} Z_1 + a_{12} Z_2 + \dots + a_{1k} Z_k \\ K_2 &= a_{21} Z_1 + a_{22} Z_2 + \dots + a_{2k} Z_k \\ &\vdots \\ K_k &= a_{k1} Z_1 + a_{k2} Z_2 + \dots + a_{kk} Z_k \end{aligned} \tag{15}$$

Next, the 14th equation is substituted into the 13th equation. Then the equation is formed:

$$\begin{aligned} Y &= \beta_0 + \beta_1(a_{11} Z_1 + a_{12} Z_2 + \dots + a_{1k} Z_k) + \beta_2(a_{21} Z_1 + a_{22} Z_2 + \dots + a_{2k} Z_k) \\ &+ \dots + \beta_n(a_{n1} Z_1 + a_{n2} Z_2 + \dots + a_{nk} Z_k) \end{aligned} \tag{16}$$

**B. RESEARCH METHOD**

In this study, secondary data obtained from (Badan Pusat Statistika., 2022) the South Sulawesi Statistics Agency in 2021 were used. The variables in this study used one dependent variable and eight independent variables. The dependent variable used is the Human Development Index ( $Y$ ). Meanwhile, for the Independent variable, the old school hope ( $X_1$ ), Average length of school ( $X_2$ ), Percentage of Population with the highest education Diploma, graduate dan post graduate ( $X_3$ ), School participation rate of the population aged 7-24 years ( $X_4$ ), gini ratio ( $X_5$ ), percentage of poor people ( $X_6$ ), per capita expenditure ( $X_7$ ), Labor ( $X_8$ ) Life expectancy ( $X_9$ ).



The steps of data analysis carried out in the study:

1. Looking for data to be used in research
2. Determine the dependent and independent variables.

3. Processing data using Minitab with the following steps:

- (a) Perform multiple regression with Ordinary Least Square Method.
- (b) Checking Multicollinearity with Variance Inflation Factor.
- (c) Checking KMO and MSA assumptions.
- (d) Perform principal component analysis by calculating eigenvalues, eigenvectors and principal component scores.
- (e) Determine the number of main components to be used.
- (f) Regressing between the principal component scores and the dependent variable
- (g) Analyze the results of principal component regression.

## C. RESULTS AND DISCUSSION

### 1. Regression Analysis and Multicollinearity

#### 1.1. Estimation of regression parameters using the Ordinary Least Square method

By using the Ordinary Least Square method, the estimated parameters of the regression model are obtained South Sulawesi province's human development index is as follows:

$$Y = 17.6 + 0.622X_1 + 1.56X_2 - 0.0559X_3 + 0.0370X_4 - 8.52X_5 - 0.0575X_6 + 0.000849X_7 - 0.114X_8 + 0.359X_9 \quad (17)$$

Based on the results of the regression analysis which shows the number of determination coefficient is 99.7%. This shows that the independent variables can explain the dependent variable by 99.7%.

#### 1.2. Detect Multicollinearity

Multicollinearity that occurs in the independent variables can be detected by looking at each VIF value. If the VIF value  $> 10$ , it can be concluded that there is a case of multicollinearity between the independent variables. By using the Minitab application, VIF values are obtained in the following table:

**Table 1.** VIF Value

| Variable | VIF  |
|----------|------|
| $X_1$    | 10.7 |
| $X_2$    | 13.5 |
| $X_3$    | 7.5  |
| $X_4$    | 4.1  |
| $X_5$    | 1.9  |
| $X_6$    | 2.7  |
| $X_7$    | 3.3  |
| $X_8$    | 1.7  |
| $X_9$    | 3.1  |

Based on table 1, it can be seen that the VIF values in the variables  $X_1$  and  $X_2 > 10$ . So, it can be concluded that there is multicollinearity between the independent variables. For more details can be seen from the results of the following correlation analysis:

**Table 2.** Correlation Coefficient Between Independent Variables

|       | $X_1$  | $X_2$  | $X_3$  | $X_4$  | $X_5$  | $X_6$  | $X_7$  | $X_8$ |
|-------|--------|--------|--------|--------|--------|--------|--------|-------|
| $X_1$ | 0.909  | 1      |        |        |        |        |        |       |
| $X_2$ | 0.879  | 0.917  | 1      |        |        |        |        |       |
| $X_3$ | 0.587  | 0.48   | 0.398  | 1      |        |        |        |       |
| $X_4$ | 0.404  | 0.394  | 0.349  | 0.448  | 1      |        |        |       |
| $X_5$ | -0.433 | -0.384 | -0.413 | 0.057  | -0.365 | 1      |        |       |
| $X_6$ | 0.498  | 0.608  | 0.577  | -0.099 | 0.116  | -0.582 | 1      |       |
| $X_7$ | -0.227 | -0.168 | -0.156 | 0.063  | -0.06  | 0.327  | -0.431 | 1     |
| $X_8$ | 0.572  | 0.584  | 0.496  | 0.646  | 0.24   | -0.186 | 0.123  | 0.243 |

From the results of calculating the correlation between the independent variables, it can be seen that there are several variables that have very strong correlation coefficients. This shows a linear relationship between variables. Some variables that have a correlation value > 0.8, namely  $X_1$  and  $X_2$ ,  $X_1$  and  $X_3$ ,  $X_2$  and  $X_3$ . So that there is multicollinearity which can be overcome using principal component regression.

**2. Principal Component Analysis**

Principal Component Analysis basically aims to reduce data, without eliminating a lot of information from the Independent variables. The results of data reduction will form a new component that involves all variables. Prior to PCA analysis, data from the original variables were standardized using the Minitab application, to eliminate correlations between variables.

**Table 3.** Principal Component that is formed

| Variable | $K_1$  | $K_2$  | $K_3$  | $K_4$  | $K_5$  | $K_6$  | $K_7$  | $K_8$  | $K_9$  |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $Z_1$    | -0.452 | 0.041  | 0.082  | 0.158  | 0.059  | 0.358  | 0.184  | -0.564 | -0.528 |
| $Z_2$    | -0.452 | 0.012  | 0.196  | 0.057  | 0.247  | -0.03  | -0.171 | -0.336 | 0.742  |
| $Z_3$    | -0.434 | -0.028 | 0.22   | 0.016  | 0.344  | 0.315  | -0.178 | 0.7    | -0.15  |
| $Z_4$    | -0.272 | 0.48   | -0.213 | 0.37   | -0.225 | -0.111 | 0.592  | 0.258  | 0.178  |
| $Z_5$    | -0.252 | 0.115  | -0.794 | -0.193 | 0.346  | -0.253 | -0.233 | -0.04  | -0.128 |
| $Z_6$    | 0.265  | 0.376  | 0.261  | 0.527  | 0.455  | -0.373 | -0.223 | -0.071 | -0.204 |
| $Z_7$    | -0.296 | -0.461 | 0.238  | -0.117 | 0.071  | -0.695 | 0.327  | 0.048  | -0.181 |
| $Z_8$    | 0.114  | 0.493  | 0.25   | -0.684 | 0.326  | 0.008  | 0.323  | -0.057 | -0.01  |
| $Z_9$    | -0.307 | 0.395  | 0.208  | -0.196 | -0.573 | -0.272 | -0.489 | 0.01   | -0.162 |

Next, determine the main components selected for further analysis by looking at the eigenvalues and the proportion of variance, where the cumulative value of the proportion is > 80%. Eigenvalues and cumulative proportions are presented in the following table:

**Table 4.** Nilai Eigen dan Proporsi Kumulatif

| Variable           | $K_1$  | $K_2$  | $K_3$  | $K_4$  | $K_5$  | $K_6$  | $K_7$  | $K_8$ | $K_9$  |
|--------------------|--------|--------|--------|--------|--------|--------|--------|-------|--------|
| <b>Eigen Value</b> | 4.3597 | 1.8808 | 0.9255 | 0.7631 | 0.4955 | 0.2754 | 0.1729 | 0.081 | 0.0471 |
| <b>Proportion</b>  | 0.484  | 0.209  | 0.103  | 0.085  | 0.055  | 0.031  | 0.019  | 0.009 | 0.005  |
| <b>Cumulative</b>  | 0.484  | 0.693  | 0.796  | 0.881  | 0.936  | 0.967  | 0.977  | 0.995 | 1      |

Next, calculate the score of the selected main components. Based on table 4, the equation model is obtained

$$\begin{aligned}
 K_1 &= -0.452Z_1 - 0.452Z_2 - 0.434Z_3 - 0.272Z_4 - 0.252Z_5 + 0.265Z_6 - 0.296Z_7 + 0.114Z_8 - 0.307Z_9 \\
 K_2 &= 0.041Z_1 + 0.012Z_2 - 0.028Z_3 + 0.480Z_4 + 0.115Z_5 + 0.376Z_6 - 0.461Z_7 + 0.493Z_8 + 0.395Z_9 \\
 K_3 &= 0.082Z_1 + 0.196Z_2 + 0.220Z_3 - 0.213Z_4 - 0.794Z_5 + 0.261Z_6 + 0.238Z_7 + 0.250Z_8 + 0.208Z_9 \\
 K_4 &= 0.158Z_1 + 0.057Z_2 + 0.016Z_3 + 0.370Z_4 - 0.193Z_5 + 0.527Z_6 - 0.117Z_7 - 0.684Z_8 - 0.196Z_9
 \end{aligned}
 \tag{18}$$

From the equation above, the main component scores for each component are obtained in the following table:

**Table 5.** Principal Component Scores

| $Z_y$    | $K_1$    | $K_2$    | $K_3$    | $K_4$    |
|----------|----------|----------|----------|----------|
| -0.78044 | 1.316374 | 0.810733 | 0.622029 | 0.461409 |
| -0.31474 | 0.426952 | -0.71469 | -0.12596 | -0.22298 |
| -0.47248 | 2.383531 | -0.33824 | 2.041599 | -1.64236 |
| -1.58164 | 3.652625 | -0.37212 | 0.339045 | 0.358409 |
| -0.79045 | 1.965794 | -1.3846  | -0.43709 | -0.83694 |
| -0.14699 | 0.173062 | 0.554938 | 0.20377  | -0.482   |
| -0.78294 | 0.681373 | 0.357542 | -0.96219 | -0.78083 |
| -0.11695 | 0.389947 | -0.92388 | -0.13859 | 0.562476 |
| -0.4174  | 1.481484 | -0.4381  | 0.298424 | 1.057104 |
| 0.063324 | -0.25902 | -0.92141 | 0.035086 | 0.929522 |
| -1.12095 | 1.407382 | 0.124998 | -1.38534 | 0.233683 |
| -0.47248 | -0.48801 | 0.131383 | -2.64094 | 0.275803 |



|          |          |          |          |          |
|----------|----------|----------|----------|----------|
| -0.31474 | 0.895521 | -1.63946 | -0.23654 | -0.55492 |
| 0.165978 | 0.138362 | -1.79408 | -0.11895 | -0.36631 |
| 0.143444 | -0.37847 | -0.31653 | -0.66343 | 1.180797 |
| 0.508991 | -0.66276 | 1.935921 | 0.863218 | 0.36317  |
| -0.00678 | 0.299728 | 0.986802 | -0.42079 | 0.846569 |
| -0.34729 | -0.62305 | 3.939652 | 0.035559 | -0.36176 |
| -0.21459 | 1.882365 | -0.6693  | 1.343624 | 0.908622 |
| 0.616652 | -0.80018 | 0.388768 | -0.92412 | -1.77416 |
| -0.28219 | -0.47717 | 2.842838 | 0.21798  | 0.030137 |
| 2.950145 | -6.11279 | -1.26583 | -0.00645 | -0.14925 |
| 1.835977 | -3.49981 | -0.92288 | 0.734727 | -1.21205 |
| 1.878541 | -3.79324 | -0.37246 | 1.325314 | 1.175847 |

### 3. Principal Component Regression

Principal Component Regression is done by regressing the selected main component with the dependent variable. The results of the regression can be seen in the following table:

**Table 6.** Result Of Regression

| Predictor    | Coef     | SE Coef      | T      | P                 |
|--------------|----------|--------------|--------|-------------------|
| Constant     | 0.00000  | 0.04235      | 0.00   | 1.000             |
| K1           | -0.44882 | 0.02072      | -21.66 | 0.000             |
| K2           | -0.13075 | 0.03157      | -4.14  | 0.001             |
| K3           | 0.24051  | 0.04497      | 5.35   | 0.000             |
| K4           | -0.02954 | 0.05744      | -0.58  | 0.575             |
| S = 0.207455 |          | R-Sq = 96.4% |        | R-Sq(adj) = 95.7% |

Based on the results of data processing in table 6, the value of 0.0000 for constant coefficient  $\beta$ , -0.44882 for  $\beta_1$ , -0.13075 for  $\beta_2$ , 0.24051 for  $\beta_3$ , and -0.02954 for  $\beta_4$ . Thus, the regression equation obtained with the OLS method is as follows:

$$Y = 0.0000 - 0.449K_1 - 0.131K_2 + 0.241K_3 - 0.029K_4 \tag{19}$$

The coefficient of determination of the principal component regression equation model is 0.95. That means that the independent variable is able to explain the dependent variable by 95.7%, and 4.3% of it is explained by other variables not included in the study.

**Table 7.** Analysis Of Variance

| Source         | DF | SS      | MS         | F      | P     |
|----------------|----|---------|------------|--------|-------|
| Regression     | 4  | 22.1823 | 5.5456     | 128.85 | 0.000 |
| Residual Error | 19 | 0.88177 | 0.0430     |        |       |
| <b>Total</b>   |    |         | 23 23.0000 |        |       |

In table 8, it can be seen that all independent variables have a significant influence on the dependent variable. This can be seen from the significance figure which is smaller than the significance = 0.05.

If the 17th equation is substituted into the 18th equation, then the regression model is

$$Y = 0.213Z_1 + 0.246Z_2 + 0.251Z_3 + 0.045Z_4 - 0.122Z_5 - 0.011Z_6 + 0.254Z_7 - 0.035Z_8 + 0.142Z_9 \tag{20}$$

From the regression model, it can be seen that several independent variables have a positive effect on the dependent variable such as  $Z_1, Z_2, Z_3, Z_4, Z_7,$  and  $Z_9$  which means, when the value of the independent variable increases, the value of the dependent variable will also increase. While the variable  $Z_5, Z_6$  and  $Z_8$  has a negative effect on the dependent variable. If the value of the variable  $Z_5$  increases, the value of the dependent variable will decrease.

The results of this study indicate that the independent variables that have a significant positive effect on the human development index are expected length of schooling, average length of schooling, percentage of residents with the highest education Diploma, bachelor and postgraduate, school enrollment rate of residents aged 7-24 years, per capita spending, and life expectancy.



By looking at the variable coefficients, the most influential is the percentage of diploma, undergraduate and postgraduate education. So to improve the quality of life of the community, it must start from the development and development of education.

Several other studies have also shown that educational indicators have a very significant effect, such as research (Mahya and Widowati, 2021) related to the human development index in the province which involves the variables average length of schooling and expected length of schooling. Then, research (Nurkuntari et al., 2016) that educational indicators, namely the average length of schooling, and the expected length of schooling, are indicators that have a significant effect on increasing the human development index.

#### D. CONCLUSION AND SUGGESTION

There are 9 variables used in this study. but these variables can be replaced with four new main component variables, namely  $K_1$ ,  $K_2$ ,  $K_3$ , and  $K_4$ , with a variance of 88.1%. The newly formed variable has a value of 95.7% which indicates the magnitude of the influence of the independent variable on the dependent variable. Based on the principal component regression model formed, it is known that the average length of schooling, residents with Diploma, Bachelor, and Masters education levels, school enrollment rates for residents aged 7-24 years, expenditure per capita, and life expectancy have a positive effect on the increase in human development index. Meanwhile, the variable percentage of the poverty rate has a negative effect on the HDI in South Sulawesi. In line with previous research related to the human development index conducted by (Winanda, 2021) that the percentage of poor people has a significant negative effect on the development index in South Sulawesi. This means that, reducing poverty through improving education and the economy is needed to improve the quality of life of the people.

#### REFERENCES

- Astri, M., Nikensari, S. I., and Kuncara, H. (2013). Pengaruh pengeluaran pemerintah daerah pada sektor pendidikan dan kesehatan terhadap indeks pembangunan manusia di indonesia. *Jurnal Pendidikan Ekonomi dan Bisnis (JPED)*, 1(1):77–102.
- Badan Pusat Statistika. (2022). Statistik pendidikan provinsi sulawesi selatan 2021 (1st ed.).
- Faizia, T., Prahutama, A., and Yasin, H. (2019). Pemodelan indeks pembangunan manusia di jawa tengah dengan regresi komponen utama robust. *Jurnal Gaussian*, 8(2):253–271.
- Hasanah, N., Bahri, S., and Fitriyani, N. (2021). The effect of human development index on poverty model in indonesia using penalized basis spline nonparametric regression. In *IOP Conference Series: Materials Science and Engineering*, volume 1115, page 012055. IOP Publishing.
- Humaira, U. H. and Nugraha, J. (2018). Analysis of factors affecting the human development index in west kalimantan province using data panel data regression. *EKSAKTA: Journal of Sciences and Data Analysis*, pages 97–105.
- Mahrany, Y. (2012). Pengaruh indikator komposit indeks pembangunan manusia terhadap pertumbuhan ekonomi di sulawesi selatan. *Skripsi: Sarjana Fakultas Ekonomi dan Bisnis Universitas Hassanudin, Makassar*.
- Mahya, A. J. and Widowati, W. (2021). Analisis pengaruh angka harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran per kapita terhadap indeks pembangunan manusia. *Prismatika: Jurnal Pendidikan Dan Riset Matematika*, 3(2):126–139.
- Maubanu, E. and Kartiko, K. (2018). Analisis komponen utama untuk mengatasi multikolinearitas pada faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di provinsi nusa tenggara timur. *Jurnal Statistika Industri dan Komputasi*, 3(01):21–30.
- Meirianti, F. (2016). *Pengaruh Kemiskinan, Belanja Pemerintah Bidang Pendidikan, Kesehatan dan Ekonomi Terhadap Tingkat IPM di 38 kabupaten/kota Provinsi Jawa Timur, 2010-2014*. PhD thesis, Universitas Gadjah Mada.
- Muliza, M., Zulham, T., and Seftarita, C. (2017). Analisis pengaruh belanja pendidikan, belanja kesehatan, tingkat kemiskinan dan pdrb terhadap ipm di provinsi aceh. *Jurnal Perspektif Ekonomi Darussalam (Darussalam Journal of Economic Perspec)*, 3(1):51–69.
- Nurkuntari, Y., Fauzi, F., and Darsyah, M. Y. (2016). Analisis jalur terhadap faktor-faktor yang mempengaruhi indeks pembangunan manusia. *Value Added: majalah ekonomi dan bisnis*, 12(2).

- Ridha, M. R. and Parwanto, N. B. (2020). The effect of foreign direct investment, human development and macroeconomic condition on economic growth: Evidence from indonesia. *Journal of Indonesian Applied Economics*, 8(2):46–54.
- Tazliqoh, A. Z., Rahmawati, R., and Safitri, D. (2015). Perbandingan regresi komponen utama dengan regresi ridge pada analisis faktor-faktor pendapatan asli daerah (pad) provinsi jawa tengah. *Jurnal Gaussian*, 4(1):1–10.
- Widiyawati (2021). Pemilihan model regresi terbaik pada kasus multikolinearitas dengan regresi komponen utama dan stepwise regression. *Skripsi, Fakultas Mipa UNILA*.
- Winanda, E. (2021). *Faktor-faktor yang mempengaruhi indeks pembangunan manusia di Sulawesi selatan*. PhD thesis, Fakultas Ekonomi Universitas Muhammadiyah Makassar.