Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 2 No. 3 2022, 17-23
DOI: *https://doi.org/10.22487/27765660.2022.v2.i3.16203*
eISSN: 2776-5660

CLASSIFIYING THE FACTORS INFLUENCING THE HUMAN DEVELOPMENT INDEX IN RIAU PROVINCE USING PRINCIPAL COMPONENT ANALYSIS

Gustriza Erda[1*], Sartika Mega Aulia[1], Zulya Erda[2]
[1]Universitas Riau
[2]Poltekkes Kemenkes Tanjung Pinang

e-mail: *gustrizaerda@lecturer.unri.ac.id*

ABSTRACT

The Human Development Index is a critical indicator of economic growth. Several factors, including average length of schooling (X_1), expected length of schooling (X_2), life expectancy at birth (X_3), number of health workers (X_4), number of health facilities (X_5), spending per capita (X_6), open unemployment rate (X_7), number of poor people (X_8), percentage of households with proper drinking water sources (X_9), and GRDP growth rate (X_{10}), can influence the Human Development Index. The purpose of this research was to simplify the factors that influence the human development index in Riau Province in 2021. Data analysis used R-Studio software by applying descriptive statistical analysis, Principal Component analysis, and Biplot analysis. The analysis revealed that the ten variables that influence human development index in Riau in 2021 can be divided into three categories: community service quality, health facilities, access, and economic conditions. These three factors can describe up to 80% of the diversity of the data.

Keywords: *Biplot, Descriptive Statistics, Human Development Index, Principal Component Analysis.*

Cite: Erda, G., Mega Aulia, S., & Erda, Z. (2023). Classifiying The Factors Influencing The Human Development Index in Riau Province using Principal Component Analysis. *Parameter: Journal of Statistics*, *2*(3), 17-23. https://doi.org/10.22487/27765660.2022.v2.i3.16203

[*]Received: Dec 2022; Reviewed: Jan 2023; Published: Jan 2023

INTRODUCTION

Human development is an important factor in economic growth because quality human resources are considered to have high productivity so as to increase the efficiency of economic activity which is ultimately believed to influence economic growth (Tjiptoherijanto, 1996). The economic development of a region is influenced by many factors, one of which is the human development index. The Human Development Index (HDI) is an illustration that can determine the ability of the population to absorb and manage sources of economic growth (Dewi & Sutrisna, 2014). A high level of human development in an area indicates that the area is capable of absorbing and managing sources of economic growth. The Human Development Index (HDI) is compiled based on indicators of achievement of a number of basic quality of life components (BPS Riau, 2020). These indicators include health, education, and the economy. This measurement uses three basic dimensions, namely longevity and healthy living, knowledge, and a decent standard of living.

The Central Bureau of Statistics (BPS) has been calculating HDI since 2014. It is recorded that human development in Riau Province has shown positive developments from year to year. In 2021, the HDI in Riau Province had reached 72.94 percent, an increase of 0.23 percent compared to 2020. The increase in HDI in 2021 included all of its constituent components. In terms of a decent standard of living in 2021, per capita expenditure had risen 0.57 percent compared to 2020, allegedly because the conditions of the Covid-19 pandemic are getting better. In terms of knowledge, it increased 0.08 percent from 2020 with the attainment of children aged 7 years having a life expectancy of 13.28 years or equivalent to the length of time they have completed education up to diploma 1. In addition, the average length of schooling for the population was 25 years and over also increased 0.05 percent with an achievement of 9.19 years. Meanwhile, from a health standpoint, babies born in 2021 had the hope of living up to 71.67 years, 0.07 years longer than in 2020. This shows that human development in Riau Province is getting better ([BPS] Badan Pusat Statistik, 2021).

The increase in the human development index is influenced by several factors, starting from education, health, and economic factors. According to (Herdiansyah & Kurniati, 2020) the implementation of good education can produce quality human development and support educational facilities. Education can be measured through the average length of schooling and expected length of schooling. In addition, human development is also influenced by life expectancy at birth, the number of health workers and the availability of good quality clean water can increase the level of health. The better the availability of clean water, the higher the chance to be healthy so that the human development index will increase (Rambe & Pulungan ,2022). The number of unemployed also affects economic growth in an area. According to Ezkirianto & Findi (2013), a large number of unemployment rates will have an impact on the low economy and will increase the number of poor people which can affect human development.

Research related to the Mapping of the Characteristics of the Human Development Index of Districts/Cities in East Java Province in 2016 using biplot analysis had been carried out by Ayu Utami et al., (2018) Based on this research, it is found that there are two main factors were formed, namely the economic education (infant mortality rate, the percentage of poor people, expenditure per capita per year, the expected length of schooling, the average length of schooling) and social health factors (percentage of the population with proper drinking water sources, the percentage of households with a clean and healthy lifestyle). In addition, Rifa'i & Hartono (2017) conducted a research about the Human Development Index and the influencing Factors using multiple regression analysis in Lampung province. It is found that life expectancy had a significant effect while the average length of schooling, number of health workers, percentage of poor population, per capita income, and economic growth did not have a significant effect.Based on this description, this study aims to simplify the factors that influence the human development index in Riau Province in 2021 using the Principal Component Analysis method so that the group that has the most influence on the human development index in Riau Province in 2021 is obtained.

MATERIALS AND METHODS

The data used in this study is secondary data published by the Central Bureau of Statistics for Riau Province in 2021. There are 10 variables used, namely; average length of schooling (X_1), expected length of schooling (X_2), life expectancy at birth (X_3), number of health workers (X_4), number of health facilities (X_5), spending per capita (X_6), open unemployment rate (X_7), number of poor people (X_8), percentage of households with proper drinking water sources (X_9), and GRDP growth rate (X_{10}). Data analysis was performed using R-Studio software. The data analysis steps carried out are:

1. Conducting descriptive statistical analysis to get an overview of the data used;
2. Conducting the Barlett test and Kaiser Mayer Olkin test;
3. Performing principal Component Analysis and classifying he factors of each principal component formed;
4. Visualizing the results of the Principal Component Analysis with Biplots to provide additional information regarding the relationship between variables and observations;
5. Interpreting the results and conclusions.

RESULTS AND DISCUSSION

Descriptive Statistical Analysis

An overview of the data affecting the human development index in Riau Province in 2021 is explained in Table 1. Based on Table 1, it can be seen clearly that the variables Life Expectancy at Birth (X_3), Percentage of Households with Adequate Drinking Water Sources (X_9), and Growth Rate of GRDP over Constant Price Base (X_{10}) has a higher median than the mean while other variables have a mean value that is above the median value. In addition, it can also be seen that the highest value is found in the variable Number of Health Workers (X_4), which is equal to 8074 people, while the lowest value is found in GRDP Growth Rate on the Basis of Constant Prices (X_{10}), which is equal to 0.51.

Table 1. Descriptive Statistics of factors influencing the Human Development Index in Riau Province in 2021

Variable	Mean	Median	Standard Deviation	Min	Max
Average length of schooling (X_1)	9.05	8.73	1.24	4.68	7.24
Expected length of schooling (X_2)	13.06	12.84	0.90	3.62	11.9
Life Expectancy at Birth (X_3)	70.25	70.61	1.42	4.63	67.7
Number of Health Workers (X_4)	2190.5	1575	1904.59	762	8074
Number of Health Facilities (X_5)	548.33	523	192.89	650	240
Expenditures Per Capita (X_6)	10753.92	10583.50	1653.66	6580	7780
Open Unemployment Rate (X_7)	4.18	3.0	1.98	6.23	2.06
Number of Poor Population (X_8)	41.74	41.14	18.41	64.1	10.5
Percentage of households with proper drinking water sources (X_9)	88.39	90.01	8.41	33.9	64.8
GDP Growth Rate at Constant Prices (X_{10})	3.61	3.81	1.60	0,51	5.47

Principal Component Analysis

The first step in carrying out principal component analysis is to carry out the Barlett test to test whether the correlation matrix between the variables formed is an identity matrix or not. If an identity matrix is formed, it means that there is no relationship between variables so that variables cannot be simplified using Principal Component Analysis (Nugroho, 2008). Barlett test results are shown in Table 2.

Table 2. Barlett Test Results

Chi-squared	Df	p-value	Conclusion
84,73	45	0,0003121	Correlation matrix ≠ identity matrix

Based on Table 2, it was found that the calculated chi-square value was 84.73 and this value was greater than the chi-square table (0.05; 6) of 61.65. This indicates that with a 95% confidence level, there is sufficient evidence to state that the correlation matrix formed is not an identity matrix. This indicates that there are t variables that are interconnected with each other so that they can be investigated further using principal component analysis.

Furthermore, the Kaiser Meyer Olkin Test (KMO) was carried out to find out whether all the data collected was sufficient to be factored and feasible to proceed with Principal Component análisis (Rahardjo, 2014). The KMO test scale ranges from 0 to 1. If the calculated KMO value is lower than 0.5, then the analysis is not feasible. Meanwhile, if the calculated KMO value is greater than 0.5, then the analysis is feasible. Based on the analysis, it was obtained that the KMO value was 0.55. This value is greater than 0.5 so that it can be used in Principal Component analysis.

After the data has been tested for feasibility, dimension reduction (simplification) is then carried out based on the criteria for the percentage of diversity in the data explained by several principal components. Good data diversity is with eigenvalues above 1. The diversity of the principal component analysis data is shown in Table 3.

Table 3. Variancies of Principal Component Analysis

Principal Component	Eigen Value	Proportion of Total Variance	Cumulative Proportion of Total Variance
1	5,20	0,5199	0,5199
2	1,74	0,1737	0,6935
3	1,14	0,1135	0,8070
4	0,80	0,0799	0,8869
5	0,50	0,0501	0,9370
6	0,33	0,0327	0,9698
7	0,20	0,0200	0,9898
8	0,06	0,0057	0,9955
9	0,04	0,0037	0,9992
10	0,01	0,0007	1,000

Based on Table 3, it can be seen that the eigenvalues for the principal components 1 to 3 are worth greater than 1 which indicated that there are three principal components formed for grouping factors that affect the human development index in Riau province in 2021, namely Principal component 1 (PC_1), Principal Component 2 (PC_2), and Principal Component 3 (PC_3). The diversity of data is also visualized using a Scree plot as shown in Figure 1. The scree plot shows that Principal Components 1, 2, and 3 have accumulated 80% of the total variance proportion.

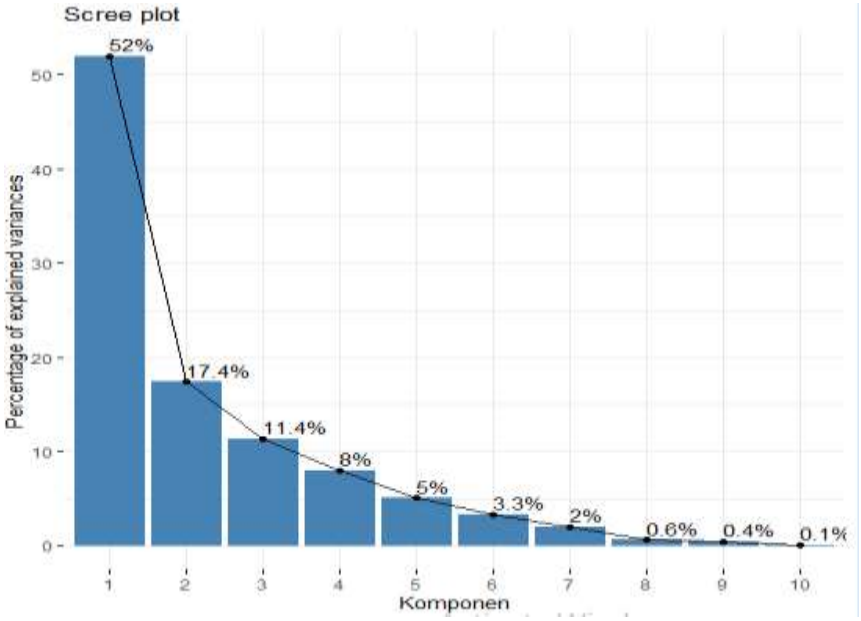

Figure 1. Scree plot of principal component variance

Furthermore, by taking into account the value of each of the principal components formed as described in Table 4, it is possible to determine the variables included in the Principal Component analysis.

Table 4. Results of the three Principal Components Formed

Variable	PC_1	PC_2	PC_3
Average length of schooling (X_1)	0,420	-0,141	0,070
Expected length of schooling (X_2)	0,387	0,065	-0,052
Life Expectancy at Birth (X_3)	0,341	-0,030	0,008
Number of Health Workers (X_4)	0,385	0,189	-0,145
Number of Health Facilities (X_5)	0,215	0,596	-0,173
Expenditures Per Capita (X_6)	0,393	-0,138	-0,126
Open Unemployment Rate (X_7)	0,359	-0,176	0,303
Number of Poor Population (X_8)	-0,118	0,687	-0,008
Percentage of households with proper drinking water sources (X_9)	0,237	0,222	0,429
GDP Growth Rate at Constant Prices (X_{10})	0,102	-0,122	-0,806

The equations obtained from the three principal components are formed as follows:

- Principal Component 1 (PC_1)
  $PC_1 = 0{,}420X_1 + 0{,}387X_2 + 0{,}341X_3 + 0{,}385X_4 + 0{,}215X_5 + 0{,}393X_6 + 0{,}359\ X_7 - 0{,}118X_8 + 0{,}237X_9 + 0{,}102X_{10}$ (1)
- Principal Component 2 (PC_2)
  $PC_2 = -0{,}141X_1 + 0{,}065X_2 - 0{,}030X_3 + 0{,}189X_4 + 0{,}596X_5 - 0{,}138X_6 - 0{,}176X_7 + 0{,}687X_8 + 0{,}222X_9 - 0{,}112X_{10}$ (2)
- Principal Component 3 (PC_3)
  $PC_3 = 0{,}070X_1 - 0{,}052\ X_2 + 0{,}008\ X_3 - 0{,}145X_4 - 0{,}173X_5 - 0{,}126X_6 + 0{,}303\ X_7 - 0{,}008X8 + 0{,}429X9 - 0{,}806X10$ (3)

In the formation of the Principal Components, the variables included in the Principal Components are obtained as follows:

Table 5. Principal Components Produced

Principal Component (PC)	Variables	New Component Name
PC_1	$X_1, X_2, X_3, X_4, X_6, X_7$	Community Service Quality
PC_2	X_5, X_8	Medical facility
PC_3	X_9, X_{10}	Access and Economic Conditions

Biplot Analysis

In order to find out additional information regarding the relationship between variables and observations, a plot presentation of observations and variables is performed simultaneously (Bartholomew, 2010). Based on the biplot analysis, it can be seen that the closeness between objects, the diversity of variables, the correlation between variables, and the relationship between variables in objects (Mattjik & Sumertajaya, 2011). The biplot results shown in Figure 2 explain the results as follows:

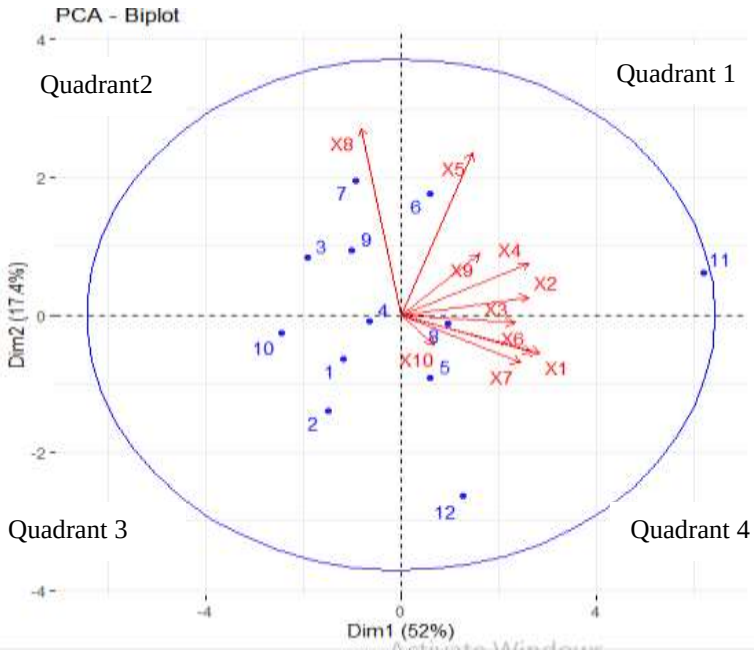

Figure 2: Principal Component Analysis Visualization Image – Biplot

Description:

1. Kuantan Singingi
2. Indragiri Hulu
3. Indragiri Hilir
4. Pelalawan
5. Siak
6. Kampar
7. Rokan Hulu
8. Bengkalis
9. Rokan Hilir
10. Kepulauan Meranti
11. Pekanbaru
12. Dumai

Proximity Between Objects

It can be seen that the object points are spread in all quadrants in the biplot. Proximity between objects is shown in quadrant 1, the position of objects that are far apart can be seen. While the quadrants that have the position of closeness between objects are in quadrant 2 (object 3 (Indragiri Hilir Regency), object 7 (Rokan Hulu Regency), and object 9 (Rokan Hilir Regency)), quadrant 3 (object -1 (Kuantan Singingi Regency), 2nd object (Indragiri Hulu Regency), 4th object (Pelawan Regency), and 10th object (Meranti Islands Regency) and quadrant 4 (5th object (Siak Regency) and object 8 (Bengkalis Regency)) This shows that the closeness between objects has relatively the same characteristics while the positions of objects that are far apart have different characteristics.

Variable Diversity

The diversity in the variables can also be explained by the 1st dimension of 52% and 17.4% of the 2nd dimension, so that overall the diversity that can be explained by the two dimensions is 69.4% of the total actual data diversity. The greatest variation occurs in the variable X_8 (number of poor people) because this variable has a long vector. While the small diversity occurs in the variable X_{10} (GRDP growth rate at constant prices) because it has a short vector.

Correlation Between Variables

The positive correlation between variables occurs in X_1 (average length of school) and X_6 (per capita expenditure) because these two variables have the same direction. This means that the variable increases the average length of school then increases. Meanwhile, the negative correlation occurs in the variables X_8 (number of poor people) and X_{10} (GRDP growth rate at constant base prices) because the two variables are in opposite directions. This means that the greater the number of poor people, the slower the GRDP growth rate at constant basic prices.

Variable Relationships on Objects

It can be seen that the 11th object (Pekanbaru City) is in the direction of the X_2 vector (Expectation of long school years). In accordance with the data where the expected length of schooling in the Regency is 15.5% above the overall average of 13.06%. In variable X_9 (Percentage of Households with Adequate Drinking Water Sources) and X_2 (Expectation of length of schooling) the opposite direction is object 1 (Kuantan Singingi Regency) and object 10 (Regency of Meranti Islands) which means the percentage of households with adequate drinking water and the expected length of schooling is below the average for all districts/cities.

Therefore, In biplots, the data information presented is determined based on the value of ρ^2 where if the value of $\rho 2$ gets closer to the value of one, it means that the biplot provides a better presentation of data (Leleury & Wokanubun, 2015). If ρ^2 is close to one value ≥70%, then the biplot provides a better presentation of the actual data information (Mattjik & Sumertajaya, 2011). Based on the calculation results, the arena value is $\rho^2 = 0.70$. This value is close to 1, which means that the biplot produced is very good and is able to explain 70% of the overall information contained in the data.

Discussion

The analysis used in this study is relatively simple, therefore further research can be developed using other advanced method such as Density-based spatial clustering of applications with noise (DBSCAN).

CONCLUSION

Based on the results, it is known that data on 10 factors that influence the human development index in Riau Province can be simplified into 3 factors which can still describe as much as 80% of the diversity of the data. The three new factors are:

1. The Community Service Quality Factor consists of the variables Average Length of Schooling, Years Expected to School, Life Expectancy at Birth, Number of Health Workers, Per Capita Spending, Open Unemployment Rate.
2. The health facility factor consists of the variable number of health facilities and the number of poor people.
3. Access Factor and Economic Condition, consisting of the variable Percentage of Household Sources of Adequate Drinking Water, GRDP Growth Rate at Constant Prices.

REFERENCES

Ayu Utami, D., Pramesti, W., Indrasetianingsih, A., & PGRI Adi Buana Surabaya Jl Dukuh Menanggal, U. (2018). Analisis Biplot pada Pemetaan Karakteristik Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Provinsi Jawa Timur tahun 2016. *J Statistika*, *11*(2), 17–22. www.unipasby.ac.idXIITelp./Fax.

[BPS] Badan Pusat Statistik. (2021). Indeks Pembangunan Manusia (IPM) Provinsi Riau Tahun 2021. *Berita Resmi Statistiik BPS Provinsi Riau*, 1–8.

Dewi, N. L. S., & Sutrisna, K. I. (2014). Pengaruh Komponen Indeks Pembangunan Manusia Terhadap Pertumubuhan Ekonomi Provinsi Bali. *E-Jurnal EP Unud*, *3*(3), 106–114.

Ezkirianto, R., & Findi, M. (2013). Analisis Keterkaitan Antara Indeks Pembangunan Manusia dan PDRB per Kapita di Indonesia. *Jurnal Ekonomi Dan Kebijakan Pembangunan*, *2*(1), 14–29.

Herdiansyah, D., & Kurniati, P. S. (2020). Pembangunan Sektor Pendidikan sebagai Penunjang Indeks Pembangunan Manusia di Kota Bandung. *Jurnal Agregasi : Aksi Reformasi Government Dalam Demokrasi*, *8*(1), 43–50. https://doi.org/10.34010/agregasi.v8i1.2765

Mattjik, A. A., & Sumertajaya. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB Press.

Nugroho, S. (2008). *Statistika Multivariat Terapan*. UNIB Press.

Rahardjo, B. (2014). Analisis Faktor untuk Mengetahui Pengaruh Persoal Selling dan Word of Mouth terhadap Keputusan Pembelian Suatu Studi Kasus Pada PT. Starmas Inti Alumunium Industry. *Jurnal Fakultas Ekonomi Universitas Budi Luhur Jakarta*, *21*(1), 1–19.

Rambe, I. H., & Pulungan, H. R. (2022). DESA LAYAK AIR BERSIH DAN SANITASI DI DESA LABUAN RASOKI. Marpokat: Jurnal Pengabdian Kepada Masyarakat, 1(1), 1-11.

Rifa'i, A., & Hartono. (2017). Indek Pembangunan Manusia dan Faktor Yang Mempengaruhinya di Daerah Perkotaan Provinsi Lampung. *Prosiding Seminar Nasional Metode Kuantitatif*, 195–213.

Tjiptoherijanto, P. (1996). *Sumber Daya Manusia dalam Pembangunan Nasional*. Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.