Jurnal Matematika Murni dan Terapan "εpsilon"
http://ppjp.ulm.ac.id/journals/index.php/epsilon
p-ISSN: 1978-4422 | e-ISSN 2656-7660
Vol.15 No.1 (Juni 2021) Hal. 58-65

BIPLOT ANALYSIS ON PRINCIPAL COMPONENTS OF HUMAN DEVELOPMENT IN ASEAN COUNTRIES

Deva Apriani Nurul Huda[1], Pardomuan Robinson Sihombing[2]

[1]*Badan Pusat Statistik (BPS) Sumatera Utara, Jl. Asrama No. 179 Medan*
[2]*Badan Pusat Statistik (BPS) RI, Jl.dr Sutomo No 6-8 Jakarta*
email*: robinson@bps.go.id*

ABSTRACT

The Human Development Index (HDI) has been the key indicator for assessing the development of a country throughout the years. It is conducted from four indicators that represent the health dimension, the education dimension, and the standard of living dimension. In ASEAN countries, the HDI tends to rise from year to year, with some countries can achieve the very high and high level of human development, while the others are still in the medium level. The aim of this study is to find the information about relative positions, characteristic similarities between ASEAN countries and the diversity of the components that construct the human development index. The Principal Component Analysis Biplot used divides the ten countries into four groups. Group 1 are the countries with the high scores especially in GNI per capita, group 2 are the ones with high scores especially in the mean years of schooling, group 3 have low scores especially in GNI per capita, and group 4 have low scores especially in the mean years of schooling

Keywords : ASEAN, Biplot, HDI, PCA

ABSTRAK

Indeks Pembangunan Manusia (IPM) adalah indikator penting untuk menilai pembangunan sebuah negara selama bertahun-tahun. IPM disusun dari empat indikator yang mewakili dimensi kesehatan, pendidikan dan standar kehidupan. Di negara-negara ASEAN, IPM cenderung meningkat dari tahun ke tahun, beberapa negara mencapai level IPM sangat tinggi dan tinggi, sementara lainnya masih di level medium. Tujuan penelitian ini adalah untuk mendapatkan informasi mengenai posisi relatif, kesamaan karakteristik di antara negara-negara ASEAN dan keragaman komponen-komponen yang menyusun IPM. Komponen Utama dalam Analisis Biplot adalah membagi 10 negara kedalam empat kelompok. Kelompok 1 terdiri dari negara-negara dengan skor tinggi terutama pada PNB perkapita, kelompok 2 terdiri dari negara-negara dengan skor tinggi terutama pada rata-rata lama sekolah, kelompok 3 memiliki skor rendah terutama pada PNB perkapita, dan kelompok 4 memiliki skor rendah terutama pada rata-rata lama sekolah

Kata kunci: *ASEAN, Biplot, IPM, PCA*

1. INTRODUCTION

The Human Development Index (HDI) is a composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living. It is a statistical tool used to measure a country's overall achievement in its social and economic dimensions. The social and economic dimensions of a country are based on the health of people, their level of education attainment and their standard of living (UNDP, 2016). Pakistani economist Mahbub ul Haq created HDI in 1990 which was further used to measure the country's development by the United Nations Development Program (UNDP). Calculation of the index combines four major indicators: life expectancy for health, expected years of schooling and mean of years of schooling for education and Gross National Income per capita for standard of living. Every year UNDP ranks countries based on the HDI report released in their annual report. HDI is one of the best tools to keep track of the level of development of a country, as it combines all major social and economic indicators that are responsible for economic development (UNDP, 2010)

The Association of Southeast Asian Nations (ASEAN) is a regional intergovernmental organization comprising ten southeast Asian states. ASEAN promotes Pan-Asianism and intergovernmental cooperation and facilitates economic, political, security, military, educational, and socio-cultural integration amongst its members and Asian nations. Table 1 shows the HDI in ten member countries for the past few years. In 2018, Singapore, Brunei Darussalam and Malaysia are among the countries with very high human development, while Thailand, Philippines and Indonesia is among those with high human development, and the rest are with medium human development.

Table 1. Human Development Index Trends in ASEAN Countries, 2014-2018

Country	2014	2015	2016	2017	2018	World Rank (2018)	HDI (2018)
Singapore	0.924	0.925	0.930	0.932	0.935	9	very high
Brunei	0.864	0.865	0.852	0.853	0.845	43	very high
Malaysia	0.787	0.789	0.799	0.802	0.804	61	very high
Thailand	0.738	0.740	0.748	0.755	0.765	77	high
Philippines	0.679	0.682	0.696	0.699	0.712	106	high
Indonesia	0.686	0.689	0.691	0.694	0.707	111	high
Vietnam	0.678	0.683	0.689	0.694	0.693	118	medium
Laos	0.582	0.586	0.598	0.601	0.604	140	medium
Myanmar	0.552	0.556	0.574	0.578	0.584	145	medium
Cambodia	0.558	0.563	0.576	0.582	0.581	146	medium

Source: UNDP (2019).

Ranking and grouping of countries based on their HDI have been conducted several times. Biswas and Caliendo (2002) used Principal Component Analysis (PCA) to rank countries based on their life expectancy, education and GDP indexes; and compared the result to the HDI ranking by UNDP (2019). Mylevaganam (2017) categorized and ranked the member states of the UN based on PCA and K-means clustering algorithm using 2015 HDI data.

In statistical sciences, there are many methods that can be used to describe the relative positions of objects with several variables simultaneously. One of them is the Principal Component Analysis Biplot. PCA Biplot is a two-dimensional chart that represents the relationship between the rows and columns of a table. This study aims to find the information about relative positions, characteristic similarities between objects and the diversity of variables in ASEAN countries by the components that construct the human development index.

2. METHOD

The data used in this study is retrieved from the UNDP, both from their annual publication, The Human Development Reports, also the data available in the UNDP website. The variables are the ones that construct the Human Development Index (HDI) as follow (UNDP, 2018),

Life expectancy at birth (X_1), is the number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life.

Expected years of schooling (X_2), is the number of years of schooling that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout the child's life.

Mean years of schooling (X_3), is the average number of years of education received by people ages 25 and older, converted from education attainment levels using official durations of each level.

Gross national income (GNI) per capita (X_4), is the aggregate income of an economy generated by its production and its ownership of factors of production, less the incomes paid for the use of factors of production owned by the rest of the world, converted to international dollars using PPP rates, divided by midyear population.

a. Principal Component Analysis

Principal component analysis (PCA) was first introduced by Karl Pearson in the early 1900's. In PCA a set of p correlated variables is transformed to a smaller set of uncorrelated hypothetical constructs called principal components (PCs). The PCs are used to discover and interpret the dependences that exist among the variables, and to examine relationships that may exist among individuals. The PCs may be used to stabilize estimates, evaluate multivariate normality, and to detect outliers (Jolliffe, 2002; Timm, 2002).

PCA deals with a single sample of n observation vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ that form a swarm of points in a p-dimensional space. If the variables $y_1, y_2, \ldots, y_p$ in $\boldsymbol{y}$ are correlated, the ellipsoidal swarm of points is not oriented parallel to any of the axes represented by $y_1, y_2, \ldots, y_p$. We wish to find the natural axes of the swarm of points with origin at $\bar{y}$, the mean vector of $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ (Rencher & Christensen, 2012). The principal component can be written as:

$$\boldsymbol{z}_i = \boldsymbol{A}\boldsymbol{y}_i. \tag{1}$$

Since $\boldsymbol{A}$ is orthogonal, $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$, and the distance to the origin is unchanged:

$$\boldsymbol{z}_i'\boldsymbol{z}_i = (\boldsymbol{A}\boldsymbol{y}_i)'(\boldsymbol{A}\boldsymbol{y}_i) = \boldsymbol{y}_i'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{y}_i = \boldsymbol{y}_i'\boldsymbol{y}_i.$$

The sample covariance matrix of $\boldsymbol{z}$, $\boldsymbol{S}_{\boldsymbol{z}} = \boldsymbol{ASA}'$ is diagonal,

$$\boldsymbol{S}_{\boldsymbol{z}} = \boldsymbol{ASA}' = \begin{bmatrix} s_{z_1}^2 & 0 & \cdots & 0 \\ 0 & s_{z_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{z_p}^2 \end{bmatrix}, \tag{2}$$

where $\boldsymbol{S}$ is the sample covariance matrix of $\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n$. Since $\boldsymbol{C}'\boldsymbol{S}\boldsymbol{C} = \boldsymbol{D} = diag(\lambda_1, \lambda_2, \dots, \lambda_p)$ where λ_i's are eigenvalues of $\boldsymbol{S}$ and $\boldsymbol{C}$ is an orthogonal matrix whose columns are normalized eigenvectors of $\boldsymbol{S}$. Thus the orthogonal matrix $\boldsymbol{A}$ that diagonalizes $\boldsymbol{S}$ is the transpose of the matrix $\boldsymbol{C}$:

$$\boldsymbol{A} = \boldsymbol{C}' = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix}, \tag{3}$$

where $\boldsymbol{a}_i'$ is the ith normalized ($\boldsymbol{a}_i'\boldsymbol{a} = 1$) eigenvectors of $\boldsymbol{S}$. The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ of $\boldsymbol{S}$ are the variances of the principal components:

$$s_{z_p}^2 = \lambda_i. \tag{4}$$

The proportions of variance explained by the first k components are:

$$\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_p} = \frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{\sum_{j=1}^{p} s_{jj}}. \tag{5}$$

Principal components also have the property of being uncorrelated in the sample; that is, the covariance of z_i and z_j is zero:

$$s_{z_i z_j} = \boldsymbol{a}_i'\boldsymbol{S}\boldsymbol{a}_j = 0 \quad \text{for} \quad i \neq j. \tag{6}$$

Since $\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}$ has no maximum if $\boldsymbol{a}$ is unrestricted, we seek the maximum of

$$\lambda = \frac{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}. \tag{7}$$

The maximum value of λ is given by the largest eigenvalue in the expression

$$(\boldsymbol{S} - \lambda\boldsymbol{I})\boldsymbol{a} = 0. \tag{8}$$

If we are retaining k components, we calculate

$$\begin{aligned} z_{1i} &= \boldsymbol{a}_1'\boldsymbol{y}_i \\ z_{2i} &= \boldsymbol{a}_2'\boldsymbol{y}_i \\ &\vdots \\ z_{ki} &= \boldsymbol{a}_k'\boldsymbol{y}_i \end{aligned} \tag{9}$$

for $i = 1, 2, \dots, n$. These are sometimes referred to as *component scores*. In vector form (9) can be written as

$$\boldsymbol{z}_i = \boldsymbol{A}_k\boldsymbol{y}_i, \tag{10}$$

where

$$\boldsymbol{z}_i = \begin{pmatrix} z_{1i} \\ z_{2i} \\ \vdots \\ z_{ki} \end{pmatrix} \quad \text{and} \quad \boldsymbol{A}_k = \begin{pmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_k' \end{pmatrix}.$$

b. PCA Biplots

PCA is inextricably linked to the singular value decomposition (SVD). This powerful results in matrix theory provides the solution to the classical PCA problem and the solution is in a format leading directly to the biplot display (Greenacre, 2010).

The (n x p) matrix $\boldsymbol{X}$ of n observations on p variables measured about their sample means can be written:

$$\boldsymbol{X} = \boldsymbol{ULA}', \tag{11}$$

where $\boldsymbol{U}$, $\boldsymbol{A}$, are (n x r), (p x r) matrices respectively, each with orthonormal columns, $\boldsymbol{L}$ is an (r x r) diagonal matrix with elements $l_1^{1/2} \geq l_2^{1/2} \geq \cdots \geq l_r^{1/2}$ and r is the rank of $\boldsymbol{X}$. Let $\boldsymbol{G} = \boldsymbol{UL}^{\alpha}$, $\boldsymbol{H}' = \boldsymbol{L}^{1-\alpha}\boldsymbol{A}'$. The (i,j)th element of $\boldsymbol{X}$ can be written

$$x_{ij} = \boldsymbol{g}_i'\boldsymbol{h}_j, \tag{12}$$

where $\boldsymbol{g}_i'$, $i = 1,2,\dots,n$ and $\boldsymbol{h}_j$, $j = 1,2,\dots,p$ are the rows of $\boldsymbol{G}$ and $\boldsymbol{H}$, respectively. If $\boldsymbol{X}$ has rank 2, all could be plotted as points in two-dimensional space. So (11) can be written

$$x_{ij} = \sum_{k=1}^{r} u_{ik} l_k^{1/2} a_{jk}, \tag{13}$$

which is often well approximated by

$${}_m\tilde{x}_{ij} = \sum_{k=1}^{m} u_{ik} l_k^{1/2} a_{jk}, \text{ with } \quad m < r. \tag{14}$$

The Mahalanobis distance between two observations $\boldsymbol{x}_h$, $\boldsymbol{x}_i$, assuming that $\boldsymbol{X}$ has a rank p so that $\boldsymbol{S}^{-1}$ exists, is defined as

$$\delta_{hi}^2 = (\boldsymbol{x}_h - \boldsymbol{x}_i)'\boldsymbol{S}^{-1}(\boldsymbol{x}_h - \boldsymbol{x}_i), \tag{15}$$

or

$$\begin{aligned} \delta_{hi}^2 &= (\boldsymbol{g}_h - \boldsymbol{g}_i)'\boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{H}(\boldsymbol{g}_h - \boldsymbol{g}_i) \\ &= (n-1)(\boldsymbol{g}_h - \boldsymbol{g}_i)'\boldsymbol{LA}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{AL}(\boldsymbol{g}_h - \boldsymbol{g}_i). \end{aligned} \tag{16}$$

.

3. RESULT AND DISCUSSION

Among ten ASEAN countries, life expectancy at birth varies from 66.9 (Myanmar) to 83.5 years (Singapore) with mean of 73.41 years and standard deviation of 4.79 years. Expected years of schooling varies from 10.3 (Myanmar) to 16.3 years (Singapore) with mean of 12.99 years and standard deviation of 1.73 years. Mean years of schooling varies from 4.8 (Cambodia) to 11.5 years (Singapore) with mean of 7.91 years and standard deviation of 2.18 years. GNI Per Capita varies from US$3,597 (Cambodia) to US$83,793 (Singapore) with mean of US$24,623 and standard deviation of US$28,526.21. It is concluded that Singapore is the top in all four components; Myanmar is the lowest in life expectancy at birth and expected years of schooling; and Cambodia is the lowest in mean years of schooling and GNI per capita. GNI per capita has the highest standard deviation among four variables (Table 3).

Table 2. Summary of the Variables

Variable	Min	Median	Mean	Max	Standard Deviation
Life expectancy at birth (X_1) (years)	66.90	73.40	73.41	83.50	4.79
Expected years of schooling (X_2) (years)	10.30	12.80	12.99	16.30	1.73
Mean years of schooling (X_3) (years)	4.80	8.10	7.91	11.50	2.18
Gross national income (GNI) per capita (X_4)	3,597	10,398	24,623	83,793	28,526

Using two principal components results in 94.44% (Table 3) proportion explained. From the model we can see that the first principal component is the mixture of all four variables, while the second principal component is a trade-off between X_3 and X_4. The two principal components can be written in the models (Table 4):

$$PC_1 = \quad -0.5179X_1 \ -0.5237X_2 \ -0.4910X_3 \ -0.4653X_4$$
$$PC_2 = \quad 0.1984X_1 \ +0.1209X_2 \ +0.4691X_3 \ -0.8527X_4$$

Table 3. Model Quality

Principal Component	Eigenvalue	Proportion of Variance Explained	Cumulative Proportion of Variance Explained
PC1	3,4388	85,97%	85,97%
PC2	0,3387	8,47%	94,44%
PC3	0,1792	4,48%	98,92%
PC4	0,0432	1,08%	100,00%

Table 4. Linear Combinations Between Variables and Principal Components, Obtained from the Eigenvectors of the Correlation Matrix

	PC1	PC2	PC3	PC4
X_1	-0.5179	0.1984	-0.5030	-0.6629
X_2	-0.5237	0.1209	-0.3950	0.7451
X_3	-0.4910	0.4691	0.7333	-0.0324
X_4	-0.4653	-0.8520	0.2306	-0.0665

Angles formed between variables in the biplot (Figure 1) represent the relationships between the variables. We can see that the closest relationship with positive correlation is between life expectancy at birth (X_1) and expected years of schooling (X_2), since the two variables form the smallest angle. While mean years of schooling (X_3) and GNI per capita barely have close relationship and positive correlation, since the angle they form is about 60 degrees.

The lengths of the variable vectors in the biplot represent variable diversities. All four variables have relatively similar length, meaning that the diversities between variables doesn't really vary or they have relatively similar diversities.

The group forming based on the quadrants represents the proximity between objects. It can be said that countries in each group resemble each other's characteristics. The biplot forms four groups. Group 1 includes Singapore and Brunei Darussalam. Group 2 includes Malaysia and Thailand. Group 3 includes

Philippines, Indonesia and Vietnam. Group 4 includes Laos, Myanmar and Cambodia.

Group 1 (lower left quadrant) are the countries with the high scores in human development components, especially in GNI per capita (X4). Group 2 (upper left quadrant) are the ones with high scores, especially in the other three components (X1, X2 and X3). Group 3 (upper right quadrant) have low scores, especially in GNI per capita (X4). Group 4 (lower right quadrant) have low scores, especially in the other three components (X1, X2 and X3).

The overall grouping result is similar to the UNDP ranking. Countries with highest HDI are in Group 1, followed by Group 2 and Group 3, while Group 4 consists of countries with the lowest HDI. Furthermore, it is obvious that Group 1 is distinguishable by GNI per capita from other groups, since Singapore and Brunei have much higher GNI per capita compared to the other eight countries.

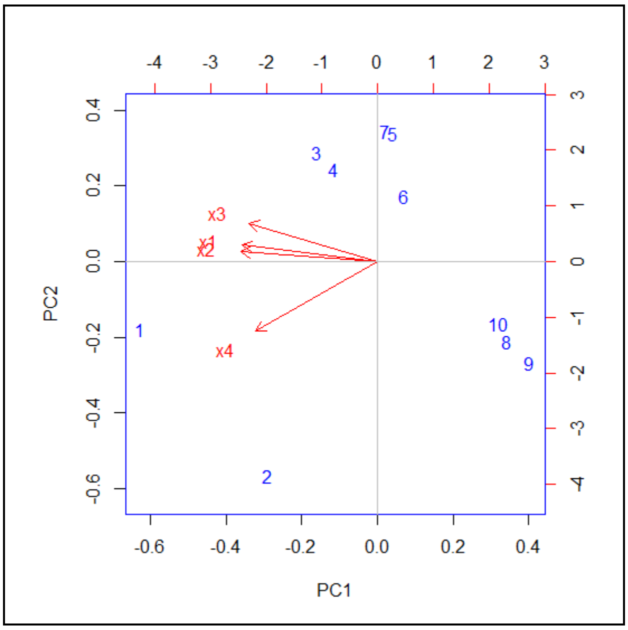

Figure 1. Biplot of the (Scaled) First Two Principal Components

4. CONCLUSION

From the analysis conducted with Principal Component Analysis Biplot in four variables assessing the human development in ASEAN countries, the conclusions derived are as follows,

a. Singapore and Brunei have similar characteristics in the components that conduct the human development. They have high scores in all four components, especially in GNI per capita.

b. Malaysia and Thailand have similar characteristics in the components that conduct the human development. They have high scores in all four components, especially in life expectancy at birth, expected years of schooling and mean years of schooling.

c. Philippines, Indonesia and Vietnam have similar characteristics in the components that conduct the human development. They have low scores in all four components, especially in GNI per capita.
d. Laos, Cambodia and Myanmar have similar characteristics in the components that conduct the human development. They have low scores in all four components, especially in life expectancy at birth, expected years of schooling and mean years of schooling.
e. It is expected for other researchers to explore the analysis on the human development components specially in ASEAN with further and more various methods. Other researches using the principal component analysis methods are also very encouraged in hope to compare the results with this study.

REFERENCES

[1]. Biswas, B., & Caliendo, F. (2002). A Multivariate Analysis of The Human Development Index. *Indian Econ. J*, *49*(4), 96–100,.

[2]. Greenacre, M. (2010). *Biplots in Practice.* Bilbao: Fundación BBVA.

[3]. Jolliffe, I. T. (2002). *Principal Component Analysis, Second.* New York: Springer-Verlag, Inc.

[4]. Mylevaganam. (2017). The Analysis of Human Development Index (HDI) for Categorizing the Member States of the United Nations (UNS) . *Open J. Appl. Sc*, *7*(12), 661–690.

[5]. Rencher, A. C., & Christensen, W. F. (2012). *Methods of Multivariate Analysis, Third* . Hoboken: John Wiley & Sons, Inc.

[6]. Timm, N. H. (2002). *Applied Multivariate Analysis.* New York: Springer-Verlag, Inc.

[7]. UNDP. (2010). *About Human Development.* New York: The United Nation Development Programme.

[8]. UNDP. (2016). *Human Development Report 2016: Human Development for Everyone.* New York: The United Nation Development Programme.

[9]. UNDP. (2018). *Human Development Indices and Indicators: 2018 Statistical Update.* New York: The United Nation Development Programme.

[10]. UNDP. (2019). *Human Development Report 2019: Beyond Income, Beyond Averages, Beyond Today.* New York:The United Nation Development Programme.