DINASTI PUBLISHER
DIJEFA:
Dinasti International Journal of Economics, Finance & Accounting
https://dinastipub.org/DIJEFA dinasti.info@gmail.com +62 811 7404 455
E-ISSN: 2721-303X
P-ISSN: 2721-3021

DOI: https://doi.org/10.38035/dijefa.v7i4

Received: 27 August 2026 | **Revised:** 16 September 2026 | **Accepted:** 23 September 2026

Determinants of Life Expectancy and Clustering: K-Means Panel for the Pandemic-Post-Covid-19 Pandemic Period in Indonesia

Sabrina Dwi Anindya[1*], Grace Natalia Marpaung[2]
[1]Universitas Negeri Semarang, Semarang, Indonesia, sabrinaanindya26@students.unnes.ac.id
[2]Universitas Negeri Semarang, Semarang, Indonesia, gracenatalia@mail.unnes.ac.id

*Corresponding Author: sabrinaanindya26@students.unnes.ac.id[1]

Abstract: Inequalities in life expectancy across Indonesia remain a challenge for health development, particularly during and after the COVID-19 pandemic. Previous studies have examined socioeconomic, educational, sanitation, and healthcare factors, but findings remain inconsistent, with limited integration of determinant analysis and provincial clustering. This study examines the determinants of life expectancy and classifies Indonesian provinces based on similar health characteristics during 2020–2025. A quantitative approach was employed using panel data from 34 provinces. Panel regression was estimated using Common Effect, Fixed Effect, and Random Effect models, with model selection based on the Chow, Hausman, and Lagrange Multiplier tests, followed by Feasible Generalized Least Squares (FGLS). K-Means clustering was then applied to group provinces based on socioeconomic and health characteristics. The results show that per capita expenditure, mean years of schooling, and proper sanitation positively affect life expectancy, while the number of health care facilities is insignificant, suggesting that service quality and equitable distribution may matter more than facility quantity. K-Means identified three provincial clusters, revealing persistent disparities, particularly between eastern and more developed regions. The study contributes an integrated regression-clustering framework for targeted, regionally differentiated public health policies.

Keywords: Life Expectancy, Health Care Facilities, Per Capita Expenditure, Mean Years of Schooling, Proper Sanitation.

INTRODUCTION

The quality of human resources is an important foundation for the progress of a region because sustainable development depends not only on physical capital and natural resources, but also on healthy, educated, and productive populations. From the perspective of the capability approach, development should extend the real freedom of individuals to live the lives they value so that human beings are placed as the main goal of development, not just an economic instrument (Mon, 1999). Health is one of the important prerequisites in the process because good health conditions allow people to obtain education, work productively, and participate in economic activities. In line with the social determinants of health approach, WHO emphasizes that health is not only influenced by medical services, but also by the social, economic, and environmental conditions in which a person lives (World Health Organization, 2008).

One of the indicators used to describe the achievements of health development is the Life Expectancy Rate (AHH), which is the estimated average length of a person's life from birth based on the pattern of mortality that prevails in a period (Indonesian Central Statistics Agency, 2025). AHH indirectly reflects various conditions related to health, such as nutrition, access to health services, sanitation, and community welfare. In the measurement of the Human Development Index (HDI), AHH is an indicator of the dimensions of longevity and healthy living along with the dimensions of education and decent living standards (Central Bureau of Statistics, 2025). Therefore, the increase in AHH not only indicates an improvement in health status, but is also an important part of efforts to improve the quality of human development and community welfare in a sustainable manner.

If examined further, the trajectory of AHH achievement in the Province of Indonesia during 2020–2025 shows a pattern that has consistently increased, although at a rate that is not always uniform between years. In 2020 to 2023, when the COVID-19 pandemic hit, there were several regions that had low life expectancy rates such as the provinces of Maluku, North Maluku, Papua, and West Papua which had life expectancy rates below 71 percent, compared to provinces with high life expectancy rates above 72 percent such as provinces in Java, Sumatra, and Bali. Entering the recovery period, this achievement continues to creep up to above 70 percent for the provinces of Maluku and North Maluku, the increase is an acceleration of growth compared to previous years (Central Bureau of Statistics, 2025). This positive trend will continue in 2025, this life expectancy gap is a challenge for Public Health policy.

A number of theoretical frameworks can explain the influence of AHH systematically. Health Production Function Grossman views health as health capital that can be improved through investment, including access to health facilities and workers; The more adequate health services are, the greater the chance of people maintaining health conditions and increasing life expectancy (Grossman, 1972). Capability Approach Sen sees poverty as a limitation of the ability to obtain proper nutrition, health services, and the environment, so that poverty is estimated to be negatively related to AHH (Mon, 1999). Meanwhile, Human Capital Theory explains that economic capabilities reflected through per capita expenditure allow households to meet their needs for better nutrition, health, and quality of life (Becker, 2020; Schultz, 1961). Furthermore, the perspective of Social Determinants of Health places sanitation and access to clean water as important environmental factors because poor sanitation conditions increase the risk of disease and can lower the degree of public health (World Health Organization, 2008). These four perspectives provide the basis that the difference in AHH between provinces is not only determined by health services, but also by socio-economic and environmental conditions.

Previous empirical findings support the relationship, but have not shown consistent results. Setyadi et al. (2023) found that health infrastructure, including doctors, health centers, and hospitals, had a significant effect on AHH, while Ardianta et al. (2025) found a negative relationship between health facilities and AHH in Central Kalimantan which showed that the availability of facilities did not necessarily reflect the effectiveness of services. In the economic aspect, (Valiant Kevin et al., 2022) shows the relationship between per capita expenditure and the achievement of provincial AHH in Indonesia, while Mahendra (2026) emphasized the importance of sanitation and clean water for AHH. Purwaningsih et al. (2021) It also found a link between sanitation and clean water access to human development, although the impact may emerge gradually. In addition, Mahendra (2026) shows that the average length of school has a positive and significant effect on AHH. The difference in findings indicates that the influence of the AHH factor may depend on regional characteristics, observation periods, and socioeconomic conditions, so further testing is needed using provincial panel data in Indonesia during the pandemic and post-COVID-19 pandemic period.

This study aims to identify the determinants of life expectancy and group provinces in Indonesia that have similar characteristics. This provincial group can form a cooperation group. This cooperation group must cooperate in Public Health policies related to life expectancy gaps. Therefore, the variation in direction and strength of the relationship between these five variables is an important gap to be studied further, especially in the context of the pandemic period and post-COVID-19 pandemic which has not been touched much by previous literature, so this study seeks to fill this gap through the analysis of panel data from 34 provinces in Indonesia during the post-COVID-19 pandemic period.

METHODS

This study uses a quantitative approach with **panel data of 34 provinces in Indonesia during the period 2020–2025**, which covers the pandemic and post-pandemic period of COVID-19. The data used is secondary data sourced from the Central Statistics Agency (BPS) and official government publications. The use of panel data allowed the study to capture variations between provinces as well as changes between times in explaining the Life Expectancy Rate (AHH) factor (Central Bureau of Statistics, 2025; Putri & Hidayat, 2025). The research model is formulated as:

$$AHH_{it} = \alpha + \beta_1 FK_{it} + \beta_2 PPK_{it} + \beta_3 RLS_{it} + \beta_4 SL_{it} + \mu_i + \lambda_t + \varepsilon_{it} \ ,$$

AHH As life expectancy, health facilities, per capita expenditure, average length of schooling, decent sanitation, indicate province and year. The specification refers to previous research that showed the association of health, socio-economic, and environmental factors with variations in life expectancy $FKPPKRLSSLit$(Gibson & Olivia, 2020; Kristanto et al., 2019). Estimation is carried out through Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM), then model selection is carried out using Chow, Hausman, and Lagrange Multiplier tests according to the characteristics of the data. The use of a panel approach allows interprovincial heterogeneity to be accounted for more systematically, while the choice between FEM and REM is concerned with assumptions regarding the relationship between individual effects and explanatory variables (Mutl & Pfaffermayr, 2011).

Testing of the assumptions and feasibility of the model is carried out to ensure that the estimate does not experience statistical problems that can interfere with the validity of the inference. Multicollinearity was examined using the Variance Inflation Factor (VIF), with a high VIF value indicating a strong linear correlation between independent variables. Heteroscedasticity was tested using the Breusch–Pagan/Cook–Weisberg test, while the autocorrelation in the panel data was tested using the Wooldridge test. In general, the Breusch–Pagan statistic can be expressed as LM = nR^2. If heteroscedasticity or dependency in error is found, inferential estimation is carried out using a robust or cluster-robust standard error. This approach is important because conventional error standards can result in inaccurate inferences when the error structure in the panel data does not meet the assumptions of homogeneity and independence (Stock & Watson, 2008). If there is an indication of interunit dependence, a robust covariance matrix approach to spatial and temporal dependencies can also be considered (Driscoll & Kraay, 1998).

To identify provincial groups with relatively similar AHH characteristics and their determining factors, this study applied K-Means clustering. Before clustering, all variables used in the grouping were standardized using z-score so that differences in units of measurement did not cause certain variables to dominate the formation of standardized clusters using

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (n_i - y_i)^2} \ ,$$

d(x, y) = distance of data x to data y, xi = value of the ith variable of data x, yi = value of the ith variable of data y, so that the difference in units does not dominate the distance. The number of clusters was determined by considering ***the Elbow Method***, while the interpretation of the results was directed at the characteristics of each group so as to result in more specific policy segmentation; *the K-Means* approach has been used to identify similarities in characteristics between Indonesian provinces in the context of COVID-19 and nonlinear panel research (Muayyad et al., 2022).

RESULTS AND DISCUSSION

Results

Based on Table 1, the descriptive statistical results show that during the study period with 204 observations, the average Life Expectancy Rate (AHH) was 73.01%, with a minimum value of 67.59% and a maximum of 76.27%, as well as a standard deviation of 1.93 which shows a relatively low variation in AHH between regions. The Health Facility (FK) variable has an average of 353.24 units, with a minimum number of 1,004 units and a maximum of 992 units, as well as a standard deviation of 259.72, which indicates that there is a considerable difference in the availability of health facilities between regions. Furthermore, Per Capita Expenditure (PPK) has an average of IDR 102.50, with a minimum value of 1 and a maximum of 204, and a standard deviation of 59.03. The Average School Length (RLS) variable has an average of 8.92 years, with a minimum value of 6.69 years and a maximum of 11.59 years, in the Proper Sanitation (SAN) control variable has an average of 82.70%, with a minimum value of 40.31% and a maximum of 98.20% and a standard deviation of 8.94. Overall, the results show that there are variations in socio-economic and health characteristics between regions, especially in the availability of health facilities and per capita expenditure, which are relevant for further analysis as factors related to AHH.

Table 1. Descriptive Statistics of Research Variables

Variable	N	Red	Std.Dev.	Min	Max
AHH (%)	204	73,01466	1,933561	67,59	76,27
FK(Unit)	204	353,2429	259,7194	1,004	992
PPK (Rp)	204	102,5	59,03389	1	204
RLS (Year)	204	8,915245	0,9152923	6,69	11,59
SAN (%)	204	82,69799	8,936167	40,31	98,2

Source: Output Stata, processed by the author (2026)

Table 2 presents the Pearson correlation matrix to provide a preliminary picture of the bivariant relationship between independent and dependent variables, as well as detect potential multicollinearity between independent variables in the model.

Table 2. Intervariable Correlation Matrix

	AHH	FK	PPK	RLS	SAN
AHH	1,0000				
FK	0,0607	1,0000			
PPK	0,3439	-0,0514	1,0000		
RLS	0,3804	-0,0147	0,3972	1,0000	
SAN	0,5634	-0,0307	0,1799	0,4816	1,0000

Source: Output Stata, processed by the author (2026)

Based on Table 2, it shows that AHH is positively correlated with all independent variables, with the greatest relationship strength being in SAN (r = 0.5634), followed by RLS (r = 0.3804) and PPK (r = 0.3439), while the correlation with FK is very weak (r = 0.0607). This pattern is consistent with*Social Determinants of Health*, which places sanitation as an important environmental factor for achievement (World Health Organization, 2008), and*Human Capital Theory*that emphasizes the role of education and economic ability in health (Becker, 2020; Schultz, 1961). The weak correlation of FK to AHH is in line with the finding that the availability of health facilities does not always reflect the effectiveness of health services in a region. On the side of independent variables, the highest correlation occurred between RLS and SAN (r = 0.4816) and PPK and RLS (r = 0.3972), but because all correlation values were still far below the threshold of 0.8–0.9 which indicates serious multicollinearity Ardianta et al., 2025) (Gujarati & Porter, 2012), these variables are suitable for use together in a single regression model, this result is also further confirmed by the VIF test in Table 3.

Table 3. Results of the Variance Inflation Factor (VIF) Test

Variable	VIVID	1/VIF
RLS	1,50	0,668119
SAN	1,30	0,767304
PPK	1,19	0,839914
FK	1,00	0,996548
BRIGHT RED	1,25	

Source: Output Stata, processed by the author (2026)

The results of the VIF test showed that all independent variables in the model had a VIF value below 10 with an average VIF of 1.25. According to (Gujarati & Porter, 2012), a VIF value that is below the critical limit of 10 indicates free multicollinearity. After the multicollinearity test, a panel estimation model that matches each other is selected.

Table 4. Summary of Model Selection Results

Test	Value	p-value	Remarks
Chow Test	F(33.166) = 40.11	0,0000	H0 is subtracted until the FEM model is selected.
Hausman Test *(sigmamore)*	Chi2(4) = 5.72	0,2209	H0 is accepted until the REM model is selected.
Breusch-Pagan LM Test	Chibar2(01) =350.04	0,0000	H0 is rejected until the REM model is selected.

Source: Output Stata, processed by the author (2026)

Based on table 4, first, the chow test obtained a value of F(33.166) = 40.11 with a p-value of 0.0000 which is smaller than alpha = 5%, so that H0 is rejected and FEM is more appropriate compared to CEM. Second, the Hausman Test is used as an option *Sigmamore* Profit outperforms chi-square value negative due to instability of the estimated variance on the standard test (Schreiber, 2008; StataCorp, 2024). The third Breusch-Pagan Langrange Multiplier test yielded a chibar value of 2(01) = 350.04 with a p-value of 0.0000 which is smaller than alpha = 5%, so H0 is rejected and REM is more appropriate. Once the REM model is selected, a classical asusmsi test is performed to ensure that the BLUE assumption is fulfilled *(Best Linear Unbiased Estimator).*

Table 5. Summary of Classical Assumption Test Results

Test	Value	p-value	Remarks
Heteroscedasticity Test	chi2(34) = 342038.04	0,0000	H0 is rejected so that heteroscedasticity exists
Autocorrelation Test	F(1, 33) = 3.317	0,0776	H0 is accepted so autocorrelation is free
Selected models	FGLS with panels(hetero) — corrects heteroscedasticity		

Source: Output Stata, processed by the author (2026)

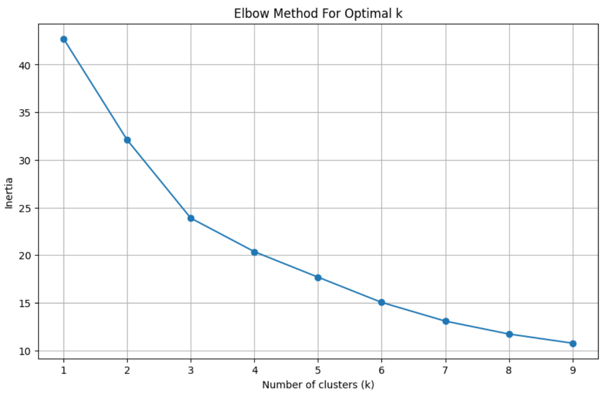
Based on the results of the classical assumption test in table 5, it shows the existence of heteroscedasticity in the REM model based on *Modified Wald Test* chi2(34) = 342038.04; p-value = 0.0000 < 0.05. The Wooldrige Test indicated a significant autocorrelation free of F(1, 33) = 3.317; p-value = 0.0776 > 0.05. Therefore, the estimation is continued using FGLS with panel specifications (hetero) to correct heteroscedasticity between panels, in line.

Table 6. Summary of FGLS Model Estimated Results

AHH	Coefficients	Std. Error	Z-Statistics	P-value
FK	0,0002763	0,0001966	1,41	0,160
PPK	0,0068564	0,0012045	5,69	0,000
RLS	0,1738649	0,0831856	2,09	0,037
SAN	0,0928702	0,0094616	9,82	0,000
Wald chi2(4) = 310.18		Prob > chi2 = 0.0000		

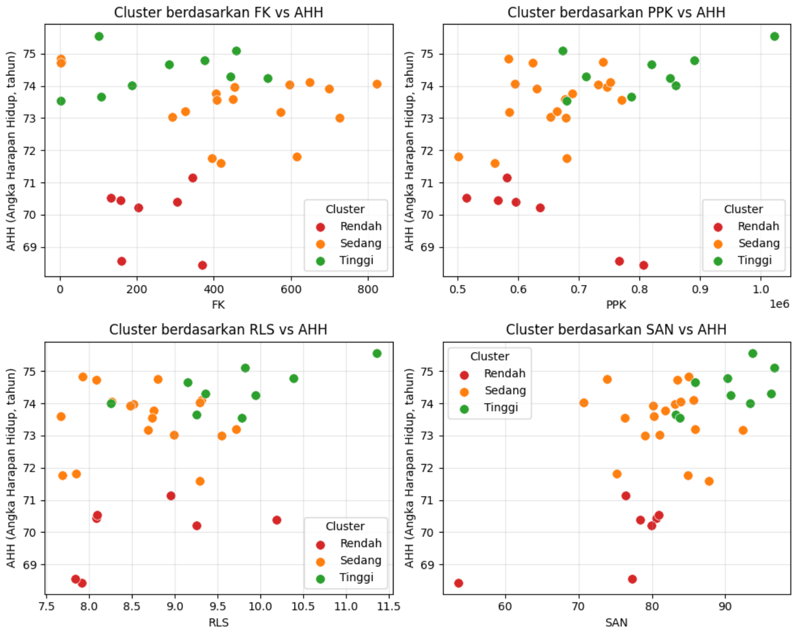
Source: Output Stata, processed by the author (2026)

Based on the results of the FGLS estimate in Table 6, it shows that the overall model is significant (Wald chi2(4) = 310.18; Prob > chi2 = 0.0000). The Per Capita Expenditure (PPK) variable has a positive and significant effect on the AHH (coefficient = 0.0068564; p = 0.000), which means that the increase in people's economic ability contributes to the increase in life expectancy, in line with*Human Capital Theory* (Becker, 2020; Grossman, 1972) and findings (Valiant Kevin et al., 2022) which shows the relationship between per capita expenditure and the achievement of provincial AHH in Indonesia. The Average Length of School Time (RLS) variable also had a positive and significant effect (coefficient = 0.1738649; p = 0.037), confirming the role of education as an investment in human capital that expands health literacy (Cutler & Lleras-Muney, 2006), while reinforcing the findings (Mahendra, 2026). The Decent Sanitation (SAN) variable had the strongest and most significant positive influence (coefficient = 0.0928702; p = 0.000), consistent with the perspective*Social Determinants of Health* (World Health Organization, 2008) as well as findings Mahendra (2026) and Purwaningsih et al. (2021) regarding the importance of sanitation for health achievements. In contrast, the Health Facility (FK) variable had no significant effect on AHH (coefficient = 0.0002763; p = 0.160), which is contrary to*Health Capital Theory*Grossman (1972) but in line with the findings Ardianta et al. (2025) indicates that the quantity of health facilities alone does not necessarily guarantee an improvement in the quality of services that have an impact on life expectancy, so that the aspect of equity and quality of services may be more decisive than the number of facility units alone.



Source: Python output, author-generated (2026)
Figure 1. Elbow Method For Optimal K

Based on the Elbow method in figure 1, the inertia value decreased significantly to k = 3, then sloped to k next. Therefore, the optimal number of clusters used in this study is k=3 which further results in three groups of Low, Medium, and High provinces as shown in the cluster profile (Figure 2 and Cluster Profile Table) which reflects the variations of PPK, RLS, SAN, and FK that have been tested for their effect on AHH in the previous regression model.



Source: Python output, author-generated (2026)
Figure 2. Clusters by Variables

Based on Figure 2, the scatter plot visualization shows that the SAN and PPK variables provide the clearest cluster separation, indicating the two variables as the main differentiators of characteristics between provincial groups. On the other hand, the FK variable showed an overlapping distribution pattern between clusters, confirming the findings in the regression analysis that health facilities did not have a significant effect on Life Expectancy, Table 7 presents the categories of each cluster

Table 7. Profile of Each Cluster

Categories	N_Provinsi	FK	PPK	RLS	SAN	AHH
Low	7	239,21	638133,55	8,61	75,3	69,96
Medium	18	435,47	659408,02	8,64	81,7	73,5
Height	9	277,48	810787,17	9,7	90,46	74,43

Source: Python output, author-generated (2026))

Table 7 of cluster profiles shows a consistent pattern of improvement in the PPK, RLS, SAN, and AHH variables as the cluster category rises from Low to High. The Low Cluster (7 provinces), namely Gorontalo, Maluku, North Maluku, Papua, West Papua, West Sulawesi, and Central Sulawesi, reflects the group of provinces with lagging health development, marked by the lowest AHH (69.96 years) supported by low access to sanitation (SAN=75.30%) and per capita expenditure (PPK=Rp638,133). On the other hand, the Tinggi cluster (9 provinces) including Bali, Banten, DI Yogyakarta, DKI Jakarta, East Kalimantan, North Kalimantan, Bangka Belitung Province, Riau Islands, North Sumatra, showed the most advanced health development conditions, with the highest AHH (74.43 years) supported by SAN (90.46%), PPK (Rp810,787), and RLS (9.70 years). Interestingly, the FK variable does not show a consistent pattern, the Medium cluster (18 provinces) including Aceh, Bengkulu, Jambi, West Java, Central Java, East Java, West Kalimantan, South Kalimantan, Central Kalimantan, Lampung, NTB, NTT, Riau, South Sulawesi, Southeast Sulawesi, North Sulawesi, West Sumatra, South Sumatra actually have the highest average FK (435.47) compared to the High cluster (277.48), indicating that the quantity of health facilities does not always reflect quality health services in a region.

Discussion

The results of the FGLS estimate show that Per Capita Expenditure (PPK), Average School Length (RLS), and Proper Sanitation (SAN) have a positive and significant effect on Life Expectancy (AHH), while Health Facilities (FK) are insignificant. This pattern is consistent with recent findings in various developing countries, which show that increasing life expectancy reflects progress in reducing preventable diseases, expanding access to health services, and strengthening the social determinants of health through economic growth, education, and health financing. Similar findings were also confirmed in cross-country studies showing that GDP per capita, literacy level, and per capita health expenditure have a positive effect on life expectancy, with GDP per capita having the strongest influence compared to other variables. On the other hand, the insignificance of FK shows that the relationship between the availability of facilities and health outcomes is not always linear, but is more determined by the equity and quality of access to services rather than just the number of units available (Li et al., 2025).

The positive influence of PPK on AHH in this study strengthens the findings Valiant Kevin et al. (2022) regarding the relationship between per capita expenditure and the achievement of provincial AHH in Indonesia, and in line with the Karunarathne et al. (2025) which shows that strengthening economic capacity encourages the acceleration of the recovery of life expectancy in the post-pandemic period. The findings are also consistent with recent cross-country evidence that any increase in GDP per capita of one percent can increase the total life expectancy of the population by about 0.2 percent, confirming that household economic ability remains one of the most consistent determinants of population health achievement in various developing country contexts (Yari et al., 2022).

The RLS variable was also shown to have a significant positive effect, reinforcing the findings Mahendra (2026) about the role of education for AHH in Indonesia. These results are in line with studies in OECD countries that affirm that each country seeks to improve life expectancy through interrelated economic, social, educational, and health policies, as well as a study in China that found that health expenditure, birth rate, and education have a positive effect on life expectancy, while inflation, population growth, and mortality rates have a negative effect. The consistency of the findings of education across country contexts strengthens the relevance of RLS as a structural determinant of AHH, although the magnitude of its influence in Indonesia is relatively more moderate than that of PPK and SAN, possibly because the impact of education on health is cumulative and only becomes significant in the medium to long term (Li et al., 2025).

Proper Sanitation (SAN) was recorded as the variable with the strongest influence on AHH, reinforcing the findings Mahendra (2026) on the importance of sanitation and clean water for health achievements, and in line with (Purwaningsih et al., 2021) regarding the relationship between sanitation and human development. The relevance of sanitation as a key environmental determinant was also confirmed in a study of seven developing countries (E-7) including Indonesia, which found that sanitation access, along with greenhouse gas emissions and water resource pressures, shape population well-being, where India and Indonesia lag behind China and Turkey due to structural inequality. The dominance of SAN in this study also emphasizes the urgency of access to sanitation and clean water in preventing disease transmission during and after the COVID-19 pandemic period (World Health Organization & United Nations Children's Fund, 2024).

On the other hand, the insignificance of FK contradicts the findings (Setyadi et al., 2023) that found that health infrastructure had a significant effect on AHH, but was consistent with

Ardianta et al. (2025) which found a negative relationship between health facilities and AHH in Central Kalimantan. The difference in direction of these findings reinforces the suspicion that the effectiveness of health facilities is highly dependent on the distribution and quality of services (Abbasi & Sohail, 2022; OECD & European Union, 2022), in line with a global study that found that life expectancy in publicly funded countries is significantly higher than in unsecured countries, confirming that access to and financing schemes of services is more decisive than the availability of facilities alone. This pattern is also reflected in the clustering results, where the Medium cluster actually has the highest average FK compared to the High cluster, reinforcing the argument that adding the quantity of facilities without equal distribution does not necessarily improve the AHH (Galvani-Townsend et al., 2022).

The results of the clustering of K-Means with k=3 optimally enriched the spatial understanding of the AHH determinant, where SAN and PPK were the main differentiators between provincial groups, while FK showed overlapping patterns. This approach is in line with the use of K-Means to identify similarities in characteristics between regions in the context of the pandemic and nonlinear panel data in Indonesia (Sadik, 2022), and strengthening Indonesia's panel data-based modeling that links the environment, education, government policies, and community economic conditions to life expectancy (Septianingsih et al., 2022). The lag of the Low cluster, which is dominated by the provinces of eastern Indonesia, is also confirmed by the latest official data showing the Papua region as the region with the lowest life expectancy among the 34 provinces in 2025, and consistent with the pattern of structural inequality that has been identified at the broader regional level (Ge & Wu, 2025). These findings confirm that the gap in health development between the eastern and western regions of Indonesia is persistent even though the national AHH trend continues to improve post-pandemic.

CONCLUSION AND RECOMMENDATIONS

Conclusion

The results of the FGLS estimate show that Per Capita Expenditure, Average School Length, and Decent Sanitation have a positive and significant effect on AHH, with Decent Sanitation as the strongest influencing variable, while Health Facilities are not proven to be significant, indicating that equity and quality of services are more decisive than just adding facility units. The results of the clustering of K-Means with optimal k=3 grouped the provinces into Low, Medium, and High categories, where Decent Sanitation and Per Capita Expenditure were the main differentiators between groups, while Health Facilities showed overlapping patterns that strengthened the regression findings. Provinces in the Low cluster, especially in the eastern region of Indonesia such as Maluku, North Maluku, Papua, and West Papua, reflect areas with backward health development, in contrast to the High cluster which is dominated by provinces with more established infrastructure and welfare.

Limitations

However, this study has a number of limitations that need to be considered in interpreting the results. First, the independent variables used are limited to four determinants (health facilities, per capita expenditure, average length of schooling, and decent sanitation) due to the availability of BPS secondary data, so that other potentially relevant factors such as the ratio of health workers, vaccination coverage, air quality, urbanization, as well as direct indicators of the impact of COVID-19 (number of cases or deaths) have not been modeled, and the variables of Health Facilities used are still in the form of unit quantities without taking into account the quality dimension or distribution of health workers in it. Second, the analysis is carried out at the provincial aggregate level, so that internal heterogeneity at the district/city level which may be quite large, especially in provinces with large areas such as Papua or

Kalimantan, cannot be captured in the model. Third, the FGLS approach used basically assumes a linear relationship between variables and does not accommodate the potential for endogeny or two-way causality relationships, for example the possibility that higher AHH contributes to increased economic productivity (expenditure per capita) rather than being a mere consequence. Fourth, the determination of the optimal number of clusters (k=3) through the Elbow Method is relatively exploratory and can produce different interpretations if validated with other methods such as the Silhouette Coefficient or the Davies Bouldin Index more strictly, while the K-Means algorithm itself is sensitive to centroid initialization and cluster shape assumptions that tend to be spherical.

Recommendations

Based on these limitations, further research is suggested to expand the scope of AHH determinant variables by including indicators of quality and equity of health services such as the ratio of doctors/health workers per population, accreditation of facilities, and geographical accessibility to test more sharply whether the main problem really lies in the distribution and quality, not the quantity of facilities. The use of data at the district/city level is also suggested to capture more detailed intra-provincial variations, while allowing the application of dynamic panel econometric approaches (e.g. System-GMM) to address the potential endogenicity and two-way causality relationships between AHH and socio-economic variables. In addition, given Indonesia's vast and island-wide geographical characteristics, future research may consider a spatial panel approach to capture the spillover effect between adjacent provinces. In the clustering aspect, the combination of K-Means with other validation methods such as hierarchical clustering or Density-Based Spatial Clustering (DBSCAN) can be used as a comparison to ensure the stability and robustness of the resulting provincial group segmentation.

Overall, despite these limitations, these findings still provide relevant policy implications that strengthening the Life Expectancy in Indonesia needs to be directed at increasing access to decent sanitation, strengthening people's purchasing power through economic empowerment programs, and expanding education as an investment in human capital, rather than simply increasing the number of health facility units without equal distribution of service quality. Provincial grouping based on cluster characteristics can be used as the basis for formulating more targeted and collaborative public health policies, especially for provinces in the Low Cluster in the eastern region of Indonesia that need priority interventions to narrow the gap in life expectancy between regions during the post-COVID-19 pandemic recovery.

REFERENCES

Abbasi, B. N., & Sohail, A. (2022). Ramification of healthcare expenditure on morbidity rates and life expectancy in the association of southeast asian nations countries: A dynamic panel threshold analysis. *The International Journal of Health Planning and Management*, *37*(6), 3251–3267. https://doi.org/10.1002/hpm.3551

Ardianta, B., Angin, P., Silaban, P. P., Ritonga, A. S., Dioni, N., Sari, Y., Saragih, A., Tampubolon, C. H., & Ompusunggu, P. (2025). Penerapan path analysis dalam menentukan faktor faktor mempengaruhi angka harapan hidup di Kalimantan Tengah. *Jurnal Penelitian Komunikasi Dan Informatika*, *4*(2).

Badan Pusat Statistik. (2025). *Indeks pembangunan manusia 2024*. https://www.bps.go.id/id/publication/2025/05/15/dd9a3ce7dfae1c733e46338f/human-development-index-2024.html

Badan Pusat Statistik Indonesia. (2025). *Dependency ratio hasil proyeksi penduduk tahun 2025* [Computer software]. https://www.bps.go.id/id/statistics-table/2/MTIwIzI=/dependency-ratio-hasil-proyeksi-penduduk.html

Becker, G. S. (2020). *Human capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research.

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: Evaluating theories and evidence*. National Bureau of Economic Research. http://www.dh.gov.uk/PolicyAndGuidance/HealthAndSocialCareTopics/HealthInequalities/fs/en

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, *80*(4), 549–560.

Galvani-Townsend, S., Martinez, I., & Pandey, A. (2022). Is life expectancy higher in countries and territories with publicly funded health care? Global analysis of health care access and the social determinants of health. *Journal of Global Health*, *12*. https://doi.org/10.7189/jogh.12.04091

Ge, X., & Wu, H. (2025). Public health policy pathways for balancing ecological challenges, healthcare systems, and social development in rapidly urbanizing economies. *Frontiers in Public Health*, *13*. https://doi.org/10.3389/fpubh.2025.1716258

Gibson, J., & Olivia, S. (2020). Direct and indirect effects of COVID-19 on life expectancy and poverty in Indonesia. *Bulletin of Indonesian Economic Studies*, *56*(3), 325–344.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255.

Gujarati, D. N., & Porter, D. C. (2012). *Dasar-dasar ekonometrika* (5th ed.). Erlangga.

Karunarathne, M., Buddhika, P., Priyamantha, A., Mayogya, P., Jayathilaka, R., & Dayapathirana, N. (2025). Restoring life expectancy in low-income countries: The combined impact of COVID-19, health expenditure, GDP, and child mortality. *BMC Public Health*, *25*(1). https://doi.org/10.1186/s12889-025-22109-4

Kristanto, E., Daerobi, A., & Samudro, B. R. (2019). Indonesian life expectancy: Role of health infrastructure and socio-economic status. *Signifikan: Jurnal Ilmu Ekonomi*, *8*(2), 241–254. https://doi.org/10.15408/sjie.v8i2.9579

Li, R., Zhan, Y., Tao, Y., Wang, R., Gu, X., He, Z., & Zhu, B. (2025). Economic and social determinants of life expectancy in China. *Frontiers in Public Health*, *13*. https://doi.org/10.3389/fpubh.2025.1554336

Mahendra, M. R. (2026). *Determinan angka harapan hidup di Indonesia (Analisis data panel tahun 2016–2024)*.

Mutl, J., & Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal*, *14*(1), 48–76.

OECD & European Union. (2022). *Health at a glance: Europe 2022: State of health in the EU cycle*. OECD Publishing. https://doi.org/10.1787/507433b0-en

Purwaningsih, F., Suharno, S., & Ahmad, A. A. (2021). Analisis pengaruh sanitasi dan akses air bersih terhadap Indeks Pembangunan Manusia di Jawa Tengah. *Jurnal Ilmiah Universitas Batanghari Jambi*, *21*(1), 56–62. https://doi.org/10.33087/jiubj.v21i1.1210

Putri, Y. M., & Hidayat, B. (2025). Apakah digitalisasi meningkatkan harapan hidup? Analisis panel data provinsi Indonesia 2022–2024. *Journal of Economics Research and Policy Studies*, *5*(1), 231–243. https://doi.org/10.53088/jerps.v5i1.1718

Sadik, K. (2022). Comparison study of distance measures in nonlinear panel data clustering with K-Means method. *International Journal of Sciences: Basic and Applied Research*, *62*(2), 106–117.

Schreiber, S. (2008). The Hausman test statistic can be negative even asymptotically. *Jahrbücher Für Nationalökonomie Und Statistik*, *228*(4), 394–405.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, *51*(1), 1–17.

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Septianingsih, A., Pertama, S. A., Kependudukan, D., Sipil, P., & Tangerang, K. (2022). Pemodelan data panel menggunakan random effect model untuk mengetahui faktor yang mempengaruhi umur harapan hidup di Indonesia. *Jurnal Lembaga Bahasa*, *3*(3), 112–120. https://doi.org/10.46306/lb.v3i3

Setyadi, S., Kustanto, A., & Widiastuti, A. (2023). Life expectancy in Indonesia: The role of health infrastructure, political, and socioeconomic status. *Iranian Economic Review*, *27*(3), 965–1005. https://doi.org/10.22059/ier.2021.85012

StataCorp. (2024). *Stata base reference manual*. Stata Press.

Stock, J. H., & Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, *76*(1), 155–174.

Valiant Kevin, A., Bhinadi, A., & Syari'udin, A. (2022). Pengaruh PDRB, angka harapan hidup, dan rata-rata lama sekolah terhadap kemiskinan di kabupaten/kota Provinsi Jawa Tengah tahun 2013–2021. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, *1*(12), 2959–2968. https://doi.org/10.54443/sibatik.v1i12.482

World Health Organization. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health*. World Health Organization.

World Health Organization & United Nations Children's Fund. (2024). *Progress on household drinking water, sanitation and hygiene 2000–2024: Special focus on inequalities*.

Yari, F., Agheli, L., Sadeghi, H., & Dizaji, S. F. (2022). The socioeconomic and political determinants of life expectancy in selected countries of the world. *Social Determinants of Health*, *8*(1). https://doi.org/10.22037/sdh.v8i1.37839