**Indonesian Journal of Statistics and Its Applications**
DOI: https://doi.org/10.29244/ijsa.v10i1p94-104

RESEARCH ARTICLE

# Segmentation of Health Status Using the FIMIX-PLS SEM Method: An Analysis Based on Socioeconomic Factors

Riwi Dyah Pangesti[1], Idhia Sriliana[1] , Susi Wijuniamurti[1], Alus Ahmad Suhaimi[1], Athaya Fairuzindah[1] and Anne Mudya Yolanda[2]

[1] Department of Statistics, The University of Bengkulu, Indonesia
[2] Department of Statistics, School of Mathematics, University of Leeds, United Kingdom

OPEN ACCESS

Check for updates

**Article History**
Received 10 May 2026
Revised 24 June 2026
Accepted 26 June 2026

**Keywords**
Health Status, Segmentation, FIMIX-PLS SEM, Sumbagsel, Socioeconomic

**Corresponding Author:**
Riwi Dyah Pangesti

**Email:**
rdyahpangesti@unib.ac.id

Department of Statistics, The University of Bengkulu, Bengkulu, Indonesia.

## ABSTRACT

The objective of this study is to model and segment the health status at the regency/city level in the Southern Sumatra region (Sumbagsel) by considering the complexity of relationships among the variables of Poverty, Economic Welfare, Environment, Utilization of Health Services, and Educational Attainment. The method used to form latent segments is Finite Mixture Partial Least Square Structural Equation Modeling (FIMIX-PLS SEM), which classifies latent heterogeneity based on finite mixture distributions into membership probabilities for each segment. The empirical findings of the FIMIX-PLS SEM segmentation suggest the presence of two optimal segments according to nearly all segment selection criteria. Segment 1 consists of 49 regencies/cities that are more strongly influenced by poverty and economic welfare. Segment 2 consists of 11 regencies/cities that are more strongly influenced by environmental factors, utilization of health services, and educational attainment. The formation of segments also increases the $R^2$ value, thereby increasing the variation in Health status that the exogenous latent factors can account for (Poverty, Economic Welfare, Environment, Utilization of Health Services, and Educational Attainment). FIMIX-PLS SEM improves the explanation of latent heterogeneity.

## Introduction

Public health status is a key indicator for assessing regional well-being. It is also a fundamental aspect that contributes across sectors to the achievement of all Sustainable Development Goals (SDGs), as it is closely related to social, economic, and environmental dimensions. Composite variables influencing the level of public health include education level, economic conditions, environmental quality, and access to healthcare services (Fitriasari et al., 2025; Rattermann et al., 2021). The complex relationships among these variables require an analytical approach capable of capturing latent structures and interactions simultaneously. However, these relationships are not always homogeneous, particularly in regions with diverse characteristics such as districts/cities in Southern Sumatra (Sumbagsel); therefore, an analytical approach that can accommodate such heterogeneity is needed.

A multivariate statistical method capable of simultaneously analyzing the relationships between latent variables and their corresponding indicators is a Structural Equation Modeling (SEM) (J. F. Hair et al., 2019; Kline, 2015). The covariance-based SEM (CB-SEM) approach has certain limitations, particularly regarding the requirement for big sample sizes. In addition, the indicators must satisfy the multivariate normality assumption (J. F. Hair et al., 2019; Kline, 2015). As an alternative, SEM

© 2026 Pangesti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY) license, allowing unrestricted use, distribution, and reproduction in any medium, provided proper credit is given to the original authors.
Print only if necessary. Let's protect our trees and the planet together.

based Partial Least Squares (PLS-SEM) gives greater adaptability in handling complex models, limited sample sizes, and do not satisfy the normally multivariate distributed indicator (J. F. Hair et al., 2019). PLS-SEM has been extensively applied across different research fields, including public health, to explore the relationships among factors influencing the level of public health (Otok et al., 2024; Sholiha & Salamah, 2015).

However, PLS-SEM alone may not adequately capture unobserved heterogeneity within a dataset. To address this limitation, the PLS-SEM using an approach Finite Mixture (FIMIX-PLS SEM) approach is applied to determine latent segments by assigning observations to classes based on posterior membership probabilities (Sarstedt & Ringle, 2010). This segmentation approach facilitates the examination of heterogeneous structural relationships across unobserved subpopulations that may exhibit different parameter patterns (Mukid et al., 2022). In this research, FIMIX-PLS is further utilized to investigate regional differences in public health determinants in South Sumatra (Sumbagsel). The model assumes that heterogeneity is present in the structural relationships, and that the latent segments follow a finite mixture distribution, typically approximated under a multivariate normal framework (Sarstedt & Ringle, 2010).

The urgency of this research lies in modeling public health status using PLS-SEM complemented by FIMIX-PLS SEM segmentation. This study aims to model and segment public health levels at the district/city level in the Southern Sumatra (Sumbagsel) region by considering the complexity of relationships among variables such as Poverty, Economic Well-being, Environment, Utilization of Healthcare Services, and Educational Attainment. This study's innovation is found in the use of FIMIX-PLS-SEM to classify latent segments within the population, thereby producing a more accurate model compared to a global PLS-SEM approach alone. By accounting for local differences in the relationships among variables, this study is expected to generate more targeted regional segmentation and serve as a foundation for formulating adaptive health policies in heterogeneous regions.

## Material and Methods

### *Data and Research Variables*

The information utilized in this research were obtained from publications of Statistics Indonesia (BPS) in 2024 (Badan Pusat Statistik, 2024a; 2024b; 2024c; 2024d) and the 2024 Indonesian Nutritional Status Survey (SSGI) (Ministry of Health of the Republic of Indonesia, 2024). This study's unit of analysis is made up 60 regencies/municipalities in the Southern Sumatra region (Sumbagsel). The sample size of 60 observations satisfies the minimum requirement for PLS-SEM based on the 10-times rule. In addition, FIMIX-PLS does not impose a specific optimal sample size requirement, as it is designed to identify unobserved heterogeneity in the data. Therefore, the sample used in this study is considered adequate for the application of both PLS-SEM and FIMIX-PLS. The endogenous latent variable examined is Health Status, which is reflected by Life Expectancy and Non-Prevalence of Stunting. The non-prevalence of stunting data were obtained by transforming the prevalence of stunting, calculated as the difference between 100 and the prevalence value, thereby representing the proportion of the population not affected by stunting.

$$\text{Non Prevalensi Stunting} = 100 - \text{Prevalensi Stunting} \tag{1}$$

**Table 1.** Indicator for Latent Variables

| Latent Variable | Symbol | Indicator |
|---|---|---|
| Education Attainment (X1) | X11 | Mean Years of Schooling |
| | X12 | Expected Years of Schooling |
| Poverty (X2) | X21 | Percentage of Poor Population |
| | X22 | Poverty Gap Index |
| | X23 | Poverty Severity Index |
| Economic Welfare (X3) | X31 | Per Capita Expenditure |
| | X32 | Gross Regional Domestic Product (GRDP) per Capita |
| Environment (X4) | X41 | Access to Safe Drinking Water |
| Utilization of Health Services (X5) | X52 | Complete Basic Immunization Coverage |
| Health Status (Y) | Y1 | Life Expectancy |
| | Y2 | Non-Prevalence of Stunting |

### Stages of Data Analysis

Data analysis and modeling in this study were performed as follows:

a. Specifying the PLS-SEM model

The PLS-SEM model have latent variable based on theories and previous studies about the determinants of public health status. There are two components to the research model: structural (inner) model and measurement (inner) model. The inner model explains the relationships between latent variables. Whereas, the outer model discusses the relationships between latent variables and their indicators. The inner model is defined in Equation 2 (Herrmann et al., 2005).

$$B\eta_i + \Gamma\xi_i = \zeta_i \tag{2}$$

In this equation, $\boldsymbol{\eta_i}$ is vector with a size of $(m \times 1)$ represents the $i$-th observation's endogenous latent variable vector, while $\boldsymbol{\xi_i}$ denotes the exogenous latent variables with dimention $(n \times 1)$. The coefficient matrix $\boldsymbol{B}$ of size $(m \times m)$ represents the associations between endogenous latent variables. In the meantime, how exogenous latent variables affect endogenous latent variables are captured by the matrix $\boldsymbol{\Gamma}$ with dimensions $(m \times n)$. Furthermore, $\boldsymbol{\zeta_i}$ denotes the vector of structural errors with dimensions of $(m \times 1)$.

Equations 3 and 4 define the measuring model with reflected indicators (Herrmann et al., 2005).

$$x = \Lambda_x\xi + \delta \tag{3}$$

$$y = \Lambda_y\eta + \varepsilon \tag{4}$$

where, $\boldsymbol{x}$ is the exogenous latent variable indicators vector with dimensions of $(q \times 1)$. $\boldsymbol{\Lambda_x}$ is the matrix containing the loading values of the indicators for the exogenous latent variables with dimensions of $(q \times n)$. $\boldsymbol{\delta}$ is the measurement error vector of exogenous indicators with dimensions of $(q \times 1)$. $\boldsymbol{y}$ is the vector of endogenous latent variable indicators with dimensions of $(p \times 1)$. $\boldsymbol{\Lambda_y}$ is the loading matrix of endogenous indicators with dimensions of $(p \times m)$. $\boldsymbol{\eta}$ is the endogenous latent variables vector with dimensions of $(m \times 1)$. $\boldsymbol{\varepsilon}$ is the measurement error vector of endogenous indicators with dimensions of $(p \times 1)$.

b. Global PLS-SEM Model Estimation

The next stage is to approximate the PLS-SEM model using all regency/city observations in the Southern Sumatra region without segmentation. The global PLS-SEM model estimation aims to obtain latent variable scores, which are subsequently used in the FIMIX-PLS analysis.

c. Outer Model Assessment

The objective of the outer model assessment is to determine the latent variables' convergence validity, discriminant validity and reliability. Average Variance Extracted (AVE) and loading value can be applied to assess convergent validity. Latent variables satisfy convergent validity standards if the AVE value is greater than 0.50, and indicators are considered acceptable when the loading value is greater than 0.70 or 0.6 if reliability latent variable has been achieved (Hair et al., 2017). The degree to which an indicator and its latent concept are related is reflected in outer loading. A higher loading value indicates that the indicator better represents the latent variable being measured. The AVE formula is presented in Equation 5 (J. F. Hair et al., 2019; Otok et al., 2024).

$$\text{AVE} = \frac{\sum_{i=1}^{M} \lambda_i^2}{\sum_{i=1}^{M} \lambda_i^2 + \sum_{i=1}^{M} Var(\varepsilon_i)} \tag{5}$$

where, $\lambda_i$ represents the loading value of the $i$-th indicator.

To assess Discriminant Validity can be used several measures among them, Heterotrait-Monotrait Ratio (HTMT), cross-loadings value, and the criteria of Fornell-Larcker. Good discriminant validity indicates that each latent variable is empirically distinct from the others. Therefore, the indicators should have stronger associations with their own latent variable than with other latent variables (J. Hair & Alamer, 2022).

Criteria of the Fornell-Larcker can be interpreted by comparing the AVE square roots of each latent variable with its correlations with other latent variables. Discriminant validity is considered satisfactory when the square root of the AVE for a latent variable is greater than its correlations with all other latent variables (J. Hair & Alamer, 2022; J. F. Hair et al., 2019; Härdle, 2011).

Cross-loading is interpreted by comparing the loading value of each indicator on its associated latent variable with its loading values on other latent variables. Discriminant validity is considered satisfactory when an indicator has a higher loading value on its own latent variables than on any other latent variables (J. Hair & Alamer, 2022; Härdle, 2011).

The HTMT is interpreted by examining the HTMT values between latent variables. Discriminant validity is considered satisfactory when the HTMT value is below the recommended threshold, commonly 0.85 or 0.90 (J. Hair & Alamer, 2022).

This journal article is © Pangesti et al. 2026

Latent variable reliability is evaluated using Composite Reliability (CR) or Cronbach's Alpha (CA). If the latent variable has CR or CA values no less than 0.70, then that latent variable own good consistency. The CR value can be calculated using the formula in Equation 6 (J. F. Hair et al., 2019).

$$\mathrm{CR} = \frac{(\sum_{i=1}^{M} \lambda_i)^2}{(\sum_{i=1}^{M} \lambda_i)^2 + \sum_{i=1}^{M} Var(\varepsilon_i)} \tag{6}$$

d. Inner Model Assessment

The assessment of the inner model is carried to measure the model's capability to describe the relationships among latent variables. Inner model assessment can be performed by examining the values of the Coefficient of Determination ($R^2$), Effect Size ($f^2$), dan Predictive Relevance ($Q^2$) (Dukalang et al., 2025; Hair et al., 2017; J. F. Hair et al., 2019; Herrmann et al., 2005).

The ability of exogenous variables to explain endogenous variables is gauged by the R^2 value. $R^2 = 0.75$, $R^2 = 0.50$ and $R^2 = 0.25$ respectively indicate substantial, moderate, and weak models (J. F. Hair et al., 2019; Henseler et al., 2009).

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \tag{7}$$

where, $R^2_{included}$ is the endogenous variable's $R^2$ value when a specific exogenous variable is incorporated into the model, and $R^2_{excluded}$ is the endogenous variable's $R^2$ value when a specific exogenous variable is removed from the model. The effect size ($f^2$) value is used to measure the contribution of each exogenous variable to the endogenous variable. According to (J. F. Hair et al., 2019; Henseler et al., 2009), $f^2 = 0.02, 0.15$, and $0.35$ respectively indicate small, medium, and large effects.

To measure the predictive capability of the model, can be used the predictive relevance ($Q^2$). A model is considered to have good predictive ability if $Q^2 > 0$ (Henseler et al., 2009; Sanderson, 2016).

$$Q^2 = 1 - \frac{SS_E}{SS_O} \tag{8}$$

where $SS_E$ is the sum of squared prediction errors, and $SS_O$ is the sum of squares of actual indicators.

e. Hypothesis Testing

Hypothesis testing was conducted using the bootstrapping procedure with 5,000 bootstrap subsamples at a 5% significance level (Hair et al., 2017). The hypotheses are:

$H_0$: $\gamma_i = 0$ (the path coefficient has no effect on the endogenous variable)

$H_1$: $\gamma_i \neq 0$ (the path coefficient has affects the exogenous variable).

The test statistics is:

$$t - val = \frac{\gamma_i}{SE(\gamma_i)} \tag{9}$$

where $\gamma_i$ is the path coefficient of the $i$-th exogenous variable, and $SE(\gamma_i)$ is the standard error of the $i$-th exogenous path coefficient obtained from the bootstrapping procedure.

The test statistic is presented in Equation 10. With a significance level of $\alpha = 5\%$, the rejection criterion is to reject $H_0$ if $p < \alpha$ or $t - val > t_{\alpha;(n-1)}$.

f. FIMIX-PLS Analysis

After the global model satisfies the validity and reliability criteria, FIMIX-PLS analysis is conducted to detect unobserved heterogeneity. The FIMIX-PLS analysis uses the latent variable scores obtained from the estimation of the global model as input for segmentation. Segmentation is performed by gradually estimating several possible numbers of segments, namely from two to five segments. The underlying premise of the FIMIX-PLS estimation is that the structural model exhibits heterogeneity, by assuming that $\eta_i$ (endogenous latent variables) follows a finite mixture distribution with a multivariate normal distribution function $f_{(i|k)}(\cdot)$ (Dukalang et al., 2025; Hair et al., 2017; Herrmann et al., 2005; Otok et al., 2024).

$$\boldsymbol{\eta_i} \sim \sum_{k=1}^{K} \rho_k \, f_{(i|k)}(\boldsymbol{\eta_i}|\boldsymbol{\xi_i}, \boldsymbol{B_k}, \boldsymbol{\Gamma_k}, \boldsymbol{\Psi_k}) \tag{10}$$

where, $\boldsymbol{\eta_i}$ is the vector of endogenous variables in the inner model (structural model) for $i = 1, 2, \ldots, I$. $\rho_k$ is the mixing proportion of the $k$-th latent segment, where $\rho_k > 0$ and $\sum_{k=1}^{K} \rho_k = 1$. $f_{(i|k)}(\cdot)$ represents the probability (odds) for the $i$-th case conditional on belonging to the $k$-th segment and the parameters $(\cdot)$. $\boldsymbol{B_k}$ is the matrix of path coefficients in the

inner model for the $k$-th latent segment, representing the relationships among endogenous latent variables, with dimensions of $(M \times M)$. $\boldsymbol{\Gamma_k}$ is the matrix of path coefficients in the inner model for the $k$-th latent segment, representing the relationships between exogenous and endogenous latent variables, with dimensions of $(M \times J)$. $\boldsymbol{\Psi_k}$ is an $(M \times M)$ matrix for the $k$-th latent segment containing the regression variances. $I$ is the number of observations, and $i$ denotes the $i$-th observation $i = 1,2, \dots, I$. $K$ is the total number of segments, and $k$ denotes the $k$-th segment $k = 1,2, \dots, K$.

g. Determination of the Optimal Number of Segments

According to (Sarstedt & Ringle, 2010), a number of information criteria, such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC), Minimum Description Length with Factor 5 (MDL5), Normed Entropy Statistic (EN), Non-Fuzzy Index (NFI), and Normalized Entropy Criterion (NEC), are used to determine the ideal number of segments. The best model is selected based on the smallest values of AIC, BIC, CAIC, MDL5, and NEC, as well as EN and NFI values that are close to one (Sarstedt & Ringle, 2010). In addition, segment size is also considered to avoid the formation of segments with excessively small numbers of observations.

h. Model Evaluation for Each Segment

After the optimal number of segments is obtained, each regency/city is classified into a segment based on the highest membership probability or maximum membership probability (Sarstedt & Ringle, 2010). The evaluation of the model for each segment is conducted by examining the $R^2$ values and the path coefficients of each segment.

## Result and Discussion

### *Results*

**Structural Equation Modeling based Partial Least Squares (PLS-SEM)**

The outer model and the inner model are the two models used in PLS-SEM. The relationships between latent variables and their indicators are represented by the outer model. The relationships between latent variables themselves are represented by the inner model. Convergent validity, discriminant validity, and reliability are used to assess the outer model for reflecting indicators. Meanwhile, the inner model is evaluated using $R^2$ and $f^2$.

Convergent validity is assessed based on the loading values or the Average Variance Extracted (AVE), which indicate that indicators within the same latent variable have a high correlation. Reliability is assessed based on the CR value. Table 2 shows the convergent validity (loading value and AVE) and reliability measures (CR).

**Table 2.** Measure of Convergent Validity and Reliability

| Latent Variable | Indicator | Loading values | AVE | CR |
|---|---|---|---|---|
| Education Attainment (X1) | Mean Years of Schooling (X11) | 0.970 | 0.629 | 0.765 |
| | Expected Years of Schooling (X12) | 0.942 | | |
| Poverty (X2) | Percentage of Poor Population (X21) | 0.938 | 0.917 | 0.971 |
| | Poverty Gap Index (X22) | 0.993 | | |
| | Poverty Severity Index (X23) | 0.940 | | |
| Economic Welfare | Per Capita Expenditure (X31) | 0.841 | 0.681 | 0.810 |
| | Gross Regional Domestic Product (GRDP) per Capita (X32) | 0.809 | | |
| Environment | Access to Safe Drinking Water (X41) | 1.000 | 1.000 | 1.000 |
| Health Services | Complete Basic Immunization Coverage (X52) | 1.000 | 1.000 | 1.000 |
| Health Status | Life Expectancy (Y1) | 0.932 | 0.913 | 0.955 |
| | Non-Prevalence of Stunting (Y2) | 0.624 | | |

According to Table 2, it can be viewed as loading values are no less than 0.60 and AVE values are no less than 0.5, suggesting that indicators satisfy good convergent validity. Good convergent validity shows that the indicators within the same latent variable own a higher correlation with each other than with indicators of other latent variables. According to the Table 3 also, can be viewed that all latent variables show CR values are above 0.7. This means that for all latent variable satisfy the good

reliability.

Table 3 shows the criterion of Fornell-Larcker, Table 4 presents the cross-loading values, and the Table 5 presents the HTMT values. These three criteria are used to assess whether the indicators employed also satisfy good discriminant validity.

**Table 3.** Fornell–Larcker Criterion

| | Health Status | Poverty | Economic Welfare | Environment | Health Services | Education Attainment |
|---|---|---|---|---|---|---|
| Health Status | **0.793** | | | | | |
| Poverty | -0.251 | **0.957** | | | | |
| Economic Welfare | 0.430 | -0.088 | **0.825** | | | |
| Environment | 0.593 | -0.155 | 0.440 | **1.000** | | |
| Health Services | 0.309 | 0.120 | 0.007 | 0.128 | **1.000** | |
| Education Attainment | 0.415 | -0.024 | 0.431 | 0.317 | -0.089 | **0.956** |

In Table 3, on the diagonal value show the roots of square of the AVE, while the values below the diagonal represent the correlations among latent variables. Based on Table 3, it is able to be seen that correlation among the latent variables less than the square root AVE values. Therefore, it could be interpreted that the all indicators satisfy discriminant validity. The indicators used for each latent variable can be clearly distinguished from the indicators of other latent variables.

**Table 4.** Cross-Loading Value

| Indicator | Education Attainment | Poverty | Economic Welfare | Environment | Health Services | Health Status |
|---|---|---|---|---|---|---|
| Mean Years of Schooling | **0.970** | -0.045 | 0.471 | 0.357 | -0.026 | 0.449 |
| Expected Years of Schooling | **0.942** | 0.008 | 0.331 | 0.230 | -0.168 | 0.327 |
| Percentage of Poor Population | -0.050 | **0.938** | -0.175 | -0.196 | 0.135 | -0.264 |
| Poverty Gap Index | -0.017 | **0.993** | -0.067 | -0.150 | 0.118 | -0.241 |
| Poverty Severity Index | 0.006 | **0.940** | 0.011 | -0.086 | 0.085 | -0.207 |
| Per Capita Expenditure | 0.516 | -0.269 | **0.841** | 0.402 | -0.117 | 0.369 |
| Gross Regional Domestic Product (GRDP) per Capita | 0.182 | 0.140 | **0.809** | 0.321 | 0.139 | 0.340 |
| Access to Safe Drinking Water | 0.317 | -0.155 | 0.440 | **1.000** | 0.128 | 0.593 |
| Complete Basic Immunization Coverage | -0.089 | 0.120 | 0.007 | 0.128 | **1.000** | 0.309 |
| Life Expectancy | 0.349 | -0.261 | 0.425 | 0.566 | 0.404 | **0.932** |
| Non-Prevalence of Stunting | 0.340 | -0.098 | 0.216 | 0.341 | -0.056 | **0.624** |

In Table 4, the values printed in bold are the outer loading values, which are compared with the cross-loading values in the same row as the corresponding loading value. If cross-loading less than loading value, hence the indicator is considered to own good discriminant validity. Based on Table 4, all indicators satisfy good discriminant validity.

**Table 5.** HTMT Values

| Latent Variables | Health Status | Poverty | Economic Welfare | Environment | Health Service |
|---|---|---|---|---|---|
| Poverty | 0.332 | | | | |
| Economic Welfare | 0.806 | 0.345 | | | |
| Environment | 0.831 | 0.154 | 0.601 | | |
| Health Service | 0.421 | 0.120 | 0.213 | 0.128 | |
| Education Attainment | 0.647 | 0.031 | 0.593 | 0.322 | 0.106 |

Table 5 shows that none HTMT value above 0.85. Therefore, it can be interpreted that all indicators have good discriminant validity. Indicator possess the good discriminant validity is mean the indicators of the corresponding latent variables have correlation that grater than correlation between indicator in other latent variable.

*This journal article is © Pangesti et al. 2026*

Based on the outer model assessment, it can be concluded that the indicators adequately reflect their respective latent variables. Therefore PLS-SEM modeling can be continued process.

The RMS theta value obtained from the analysis is 0.324. This value exceeds the threshold of 0.12, indicating that the reflective measurement model does not have good model fit. This suggests the presence of residual correlations among indicators as well as possible model specification errors. However, this condition may also reflect underlying data heterogeneity, thereby supporting the use of the FIMIX-PLS method for further segmentation analysis.

The inner model represents the relationships among latent variables in the structural model. Evaluation of the inner model can be conducted by examining the values of $R^2$ and $f^2$. According to the analysis results, the model obtained an $R^2 = 0{,}538$ value, which is categorized as moderate.

**Table 6.** Effect size values for the relationships between exogenous latent variables and health status. The table summarizes the contribution of poverty, economic welfare, environment, health services, and educational attainment to the endogenous latent variable.

| Direction of Relationship | $f^2$ Value | Description |
|---|---|---|
| Poverty -> Health Status | 0.092 | Small |
| Economic Welfare -> Health Status | 0.025 | Small |
| Environment -> Health Status | 0.236 | Medium |
| Health Service -> Health Status | 0.194 | Medium |
| Education Attainment -> Health Status | 0.115 | Small |

Based on the effect size ($f^2$) criteria proposed by (J. F. Hair et al., 2019), the results indicate that environmental factors and health service utilization have medium effect sizes on health status, while poverty, economic welfare, and educational attainment exhibit small effect sizes.

**Table 7.** Hypothesis Testing

| Latent Variables | Path Coefficient | Sample Mean | STDV | T value |
|---|---|---|---|---|
| Poverty | -0.211 | -0.207 | 0.092 | 2.280 |
| Economic Welfare | 0.129 | 0.161 | 0.114 | 1.127 |
| Environment | 0.382 | 0.368 | 0.109 | 3.493 |
| Health Service | 0.308 | 0.326 | 0.111 | 2.782 |
| Education Attainment | 0.261 | 0.239 | 0.139 | 1.877 |

Table 7 presents the significance of the influence between exogenous latent variables and endogenous latent variable (Health Status). Based on Table 7, the variables that possess a substantial effect on Health Status $T - value > 1.96$ are Poverty, Environment, and Utilization of Health Services. Meanwhile, the variables that do not possess a substantial effect on Health Status $T - value < 1.96$ are Economic Welfare and Educational Attainment.

Overall, the results indicate that increases in most exogenous variables tend to be associated with an improvement in health status, except for poverty, which shows a negative relationship. The positive path coefficients of Environment (0.382), Health Service (0.308), Education Attainment (0.261), and Economic Welfare (0.129) suggest that better environmental conditions, improved health services, higher educational attainment, and greater economic welfare are generally linked to higher health status. However, only Environment and Health Service are statistically significant, indicating that these two factors have a reliable and meaningful effect on improving health outcomes. Environment has a positive effect on Health Status ($\beta$ = 0.382), indicating that better environmental conditions are associated with higher health status. Specifically, a one standard deviation increase in Environment is associated with a 0.382 standard deviation increase in Health Status, assuming other variables remain constant. Similarly, Health Service has a positive effect on Health Status ($\beta$ = 0.308), indicating that a one standard deviation increase in Health Service is associated with a 0.308 standard deviation increase in Health Status.

In contrast, Poverty exhibits a significant negative effect (−0.211), implying that an increase in poverty leads to a decline in health status. This highlights poverty as a key barrier to health improvement. Meanwhile, Economic Welfare and Education Attainment, despite their positive directions, are not statistically significant in this model, suggesting that their effects on health status are not strong enough to be confirmed statistically. Overall, improvements in health status are primarily driven by better environmental conditions and health services, along with reductions in poverty.

*This journal article is © Pangesti et al. 2026*

**Approach Finite Mixture for PLS SEM (FIMIX-PLS SEM)**

The selection of the best segment is carried out by considering the criteria on Table 8.

**Table 8.** Criteria for Selecting the Optimal Segment

| Criterion | Two Segments | Three Segments | Four Segments | Five Segments |
|---|---|---|---|---|
| AIC | 106.573 | 98.059 | 89.571 | 75.509 |
| BIC | 133.800 | 139.946 | 146.118 | 146.717 |
| CAIC | 146.800 | 159.946 | 173.118 | 180.717 |
| HQ | 117.223 | 114.444 | 111.690 | 103.362 |
| MDL5 | 346.706 | 467.494 | 588.307 | 703.548 |
| EN | 0.901 | 0.886 | 0.802 | 0.854 |
| NFI | 0.914 | 0.877 | 0.778 | 0.822 |
| NEC | 5.911 | 6.869 | 11.901 | 8.742 |

The evaluation of segment solutions using multiple information criteria indicates that the two-segment model is the most optimal. This is supported by the lowest values of BIC, CAIC, MDL5, and NEC, as well as the highest entropy (EN) and NFI values compared to other segment alternatives. Although AIC and HQ continue to decrease with an increasing number of segments, the two-segment solution provides a more parsimonious and stable classification.

**Table 9.** Path Coefficients Based on the FIMIX-PLS SEM Results

| Latent Variables | Segment 1 | Segment 2 | Global Model |
|---|---|---|---|
| Poverty | -0.215 | -0.053 | -0.211 |
| Economic Welfare | 0.365 | 0.307 | 0.129 |
| Environment | 0.329 | 0.456 | 0.382 |
| Health Service | 0.276 | 0.438 | 0.308 |
| Education Attainment | -0.009 | 0.448 | 0.261 |
| $R^2$ | 0.521 | 1.000 | 0.538 |

Table 9 shows differences in the patterns of relationships among variables across segments. In Segment 1, the variables that have the greatest influence on Health Status are Economic Welfare (0.365) and Environment (0.329), while Poverty has a negative effect (-0.215) and Educational Attainment has almost no effect (-0.009). In contrast, in Segment 2, almost all variables have relatively strong positive effects, particularly Educational Attainment (0.448), Utilization of Health Services (0.438), and Environment (0.456), whereas the effect of Poverty is relatively small (-0.053). The $R^2$ values indicate that Segment 2 has a very high model explanatory power $R^2 = 1.000$ compared to Segment 1 $R^2 = 0.521$ and the global model $R^2 = 0.538$, indicating that the segmentation successfully revealed heterogeneity that was not visible in the global model.

**Table 10.** Regency/City Group Membership According to the FIMIX-PLS SEM Results

| Segment | Number of Members | Regency or City |
|---|---|---|
| 1 | 49 | Belitung, Bangka Barat, Bangka Tengah, Belitung Timur, Pangkal Pinang, Bengkulu Selatan, Rejang Lebong, Bengkulu Utara, Kaur, Seluma, Mukomuko, Kepahiang, Bengkulu Tengah, Kota Bengkulu, Kerinci, Merangin, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat, Bungo, Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim, Lahat, Musi Rawas, Musi Banyuasin, Banyuasin, Ogan Komering Ulu Timur, Ogan Ilir, Empat Lawang, Musi Rawas Utara, Palembang, Prabumulih, Pagar Alam, Lubuk Linggau, Lampung Barat, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Bandar Lampung, Metro. |
| 2 | 11 | Bangka, Bangka Selatan, Lebong, Tebo, Kota Jambi, Kota Sungai Penuh, Ogan Komering Ulu Selatan, Pali, Tanggamus, Tulang Bawang, Pesisir Barat. |

### Discussion

The results indicate that the PLS-SEM model has a moderate ability to explain the relationships among the latent variables. Furthermore, all indicators satisfied the required validity and reliability criteria. The global model reveals that the Environment and Health Service Utilization variables have substantial positive effects on improving Health Status in Southern Sumatra (Sumbagsel), whereas the Poverty variable has a substantial negative effect on Health Status. This finding is consistent with the World Health Organization (WHO, 2025), which highlights environmental conditions, access to healthcare services, and socioeconomic factors as important determinants of population health outcomes. In contrast, the Economic Welfare and Educational Attainment variables do not have significant effects on either improving or reducing Health Status in Sumbagsel. This condition may be attributed to the presence of unobserved latent heterogeneity among the research objects, which can lead to differences in structural relationships across groups that are not identified in the global model (Sarstedt & Ringle, 2010). Therefore, regional segmentation was conducted to identify distinct behavioral characteristics within each segment that cannot be directly observed.

The segmentation using FIMIX-PLS SEM successfully classified the regions into two optimal segments. This selection was based on several goodness-of-fit criteria, including AIC, BIC, CAIC, HQ, MDL5, EN, NFI, and NEC. Most of the optimal segment selection criteria indicated that the two-segment solution was the most optimal, based on the smallest AIC, BIC, CAIC, MDL5, and NEC values, as well as the largest EN and NFI values approaching 1. However, based on the HQ criterion, five segments were considered the most optimal. Based on these results, the researchers decided to select two segments as the optimal segmentation solution. This decision was also supported by the highest EN value (0.901), indicating a clear separation between segments. There are 49 regencies/cities in the Southern Sumatra region included in Segment 1 and 11 regencies/cities included in Segment 2.

Segment 1 is more strongly influenced by the Poverty and Economic Welfare variables. Therefore, to improve the Health Status in regions belonging to Segment 1, it is recommended to reduce poverty and improve economic welfare, particularly through increasing per capita expenditure. Segment 2 is more strongly influenced by the Environment, Utilization of Health Services, and Educational Attainment variables. Thus, to improve the Health Status in Segment 2 regions, the study suggests improving these three variables: improving environmental conditions by increasing access to safe drinking water, improving the Utilization of Health Services variable by increasing Complete Basic Immunization Coverage, and improving the Educational Attainment latent variable by increasing Expected Years of Schooling and Mean Years of Schooling.

The formation of these segments also increased the $R^2$ compared to the global model, meaning that a greater variation in Health Status could be explained by the model. Therefore, segmentation allows the model to better capture the heterogeneous relationships among regions by accounting for unobserved latent heterogeneity. Although the $R^2$ values increased after segmentation, Segment 2 consists of only 11 observations. Therefore, the results for this segment should be interpreted with caution, and future studies with larger samples are recommended to validate the stability of the identified relationships.

## Conclusion and Recommendations

The PLS-SEM model has a moderate ability to explain the relationships among latent variables, with indicators that have satisfied good validity and reliability criteria. The global model shows that the Environment and Utilization of Health Services variables have a significant effect on improving the Health Status of Sumbagsel, while Poverty has a significant negative effect on reducing it. The application of FIMIX-PLS SEM successfully formed two optimal regional segments, indicating differences in characteristics among regions, where Segment 1 is more influenced by Poverty and Economic Welfare, while Segment 2 is more influenced by Environment, Utilization of Health Services, and Educational Attainment. The formation of segments also increased the coefficient of determination $R^2$, demonstrating that segmentation is able to improve the accuracy of the analysis in explaining variations in Health Status by separating the behavioral characteristics of each segment that were previously not directly observable.

For upcoming studies, it is recommended to expand the scope of variables used in order to capture other factors influencing the Health Status of Sumbagsel, such as social, cultural, and basic infrastructure aspects. In addition, future studies may apply other segmentation or more complex modeling methods to compare results and improve the accuracy of the analysis of latent heterogeneity across regions. The use of panel data or time series data is also recommended to observe changes in regional health conditions over time. Furthermore, the segmentation results obtained can be utilized as a basis for formulating more specific and targeted policies according to the characteristics of each region.

## Conflicts of Interest

There are no conflicts of interest to declare.

This journal article is © Pangesti et al. 2026

## Acknowledgements

This research received funding support from the PNBP of FMIPA Universitas Bengkulu under the Development Research Scheme in 2025 with contract number: 4531/UN.12/HK/2025. The research team would like to express their sincere gratitude to FMIPA and LPPM Universitas Bengkulu for their support, which enabled this research to be carried out successfully.

## AI Writing Statement

During the preparation of this work, the author(s) used ChatGPT (OpenAI) to assist in language editing, translation, and grammar checking. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## References

Badan Pusat Statistik. (2024a). *Data dan informasi kemiskinan kabupaten/kota di Indonesia 2024* (Vol. 16). Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. (2024b). *Indeks pembangunan manusia provinsi di Sumatra Bagian Selatan 2024*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. (2024c). *Provinsi dalam angka di Sumatra Bagian Selatan 2024*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. (2024d). *Statistik kesejahteraan rakyat (Survei Sosial Ekonomi Nasional/Susenas) 2024*. Jakarta: Badan Pusat Statistik.

Dukalang, H. H., Otok, B. W., & Purhadi. (2025). Modified partial least square structural equation model with multivariate adaptive regression spline: Parameter estimation technique and applications. *MethodsX*, *14*: 103381. https://doi.org/10.1016/j.mex.2025.103381

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.). (2010). *Handbook of partial least squares: Concepts, methods and applications*. Berlin: Springer. https://doi.org/10.1007/978-3-540-32827-8

Fitriasari, E., & Umasugi, M. T. (2025). Determinan sosial dan intervensi dalam kesehatan: Scoping review tentang populasi rentan dan pendekatan edukasional. *Indonesian Journal of Nursing Research*, *8*(1): 42–52. https://doi.org/10.35473/ijnr.v8i1.3986

Hair, J., & Alamer, A. (2022). Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, *1*(3): 100027. https://doi.org/10.1016/j.rmal.2022.100027

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Andover: Cengage Learning.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation* modeling *(PLS-SEM)* (2nd ed.). Thousand Oaks, CA: Sage.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*: 277–319. https://doi.org/10.1108/S1474-7979(2009)0000020014

Herrmann, A., Hahn, C. H., Johnson, M. D., & Huber, F. (2005). Capturing customer heterogeneity using a finite mixture PLS approach. *SSRN Electronic Journal*: 243–269. https://doi.org/10.2139/ssrn.321004

This journal article is © Pangesti et al. 2026

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Mukid, M. A., Otok, B. W., & Suharsono, A. (2022). Segmentation in structural equation modeling using a combination of partial least squares and modified fuzzy clustering. *Symmetry*, *14*(11): 2431. https://doi.org/10.3390/sym14112431

Otok, B. W., Purhadi, Sriningsih, R., & Dila, D. S. (2024). Segmentation of toddler nutritional status using REBUS and FIMIX partial least square in Southeast Sulawesi. *MethodsX*, *12*: 102515. https://doi.org/10.1016/j.mex.2023.102515

Rattermann, M. J., Angelov, A., Reddicks, T., & Monk, J. (2021). Advancing health equity by addressing social determinants of health: Using health data to improve educational outcomes. *PLoS ONE*, *16*(3): e0247909. https://doi.org/10.1371/journal.pone.0247909

Sanderson, D. C. (2016). *The tenant as customer: Does good service enhance the financial performance of commercial real estate?* PhD thesis, University of Reading. https://centaur.reading.ac.uk/65939/

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modeling: A comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics*, *37*(8): 1299–1318. https://doi.org/10.1080/02664760903030213

Sholiha, E. U. N., & Salamah, M. (2015). Structural equation modeling-partial least square for modeling district/city health degrees in East Java. *Jurnal Sains dan Seni ITS*, *4*(2): 169–174. http://id.portalgaruda.org/index.php?ref=browse&mod=viewarticle&article=415116

*This journal article is © Pangesti et al. 2026*