CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 904-915
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Integrating Path Analysis and Kendall's Tau-based Principal Component Analysis to Identify Determinants of Child Health

Viky Iqbal Azizul ALim*, Atiek Iriany, Adji Achmad Rinaldo Fernandes, Solimun, and Candra RWS Weni Utomo

Department of Statistics, Brawijaya University, Indonesia

Abstract

This study develops a latent variable path analysis model using a Mixed-Scale Principal Component Analysis (PCA) approach based on Kendall's Tau correlation to identify key determinants of child health in Batu City, Indonesia. Primary data were collected from 100 mothers with children under five years old through questionnaires. The variables examined include Family Demographics, Nutritional Consumption, and Child Health Condition, each measured using mixed-scale indicators (ordinal and numerical). Kendall's Tau-based PCA was applied to reduce data dimensionality and construct latent variables, which were then integrated into a path analysis model. The results show that maternal age is the most dominant indicator in shaping the Family Demographics construct, while balanced nutritional food is the strongest indicator forming the Nutritional Consumption construct. Path analysis further reveals that Family Demographics significantly affect Child Health Condition both directly and indirectly through Nutritional Consumption, with a coefficient of determination of 77.62%. These findings underscore the critical role of demographic and nutritional factors in determining child health outcomes and highlight the methodological advantage of Kendall's Tau-based mixed-scale PCA for analyzing heterogeneous indicator data within a structural path framework.

Keywords: Child Health; Kendall's Tau Correlation; Latent Variable Modeling; Mixed-Scale Principal Component Analysis; Path Analysis

1 Introduction

Path analysis is an extension of regression analysis designed to accommodate multiple structural equations simultaneously [1]. A defining feature of such models is the presence of at least one exogenous variable, one intervening endogenous variable, and one pure endogenous variable. Exogenous variables influence other variables but are not influenced themselves, while pure endogenous variables are only affected by others. Intervening endogenous variables, in contrast, serve both as predictors and outcomes within the model structure.

Traditionally, path analysis has been applied to manifest variables—those directly observable and measurable. When latent variables are involved, however, auxiliary techniques are required. Latent variables represent theoretical constructs that cannot be measured directly but must instead be inferred from multiple indicators. Principal Component Analysis (PCA) is one of the

*Corresponding author. E-mail: vikyiqbal30@gmail.com

Submitted: December 26, 2024 Reviewed: June 4, 2025 Accepted: August 27, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.31156

most widely used techniques for this purpose. PCA reduces data dimensionality by transforming correlated variables into a smaller set of uncorrelated components that retain most of the original variance [2]. Nevertheless, conventional PCA assumes linearity and is limited to metric-scale data, which restricts its application in many real-world settings.

In practice, particularly in social and health research, indicators are rarely measured on a single scale. Instead, they often include a mixture of numerical, ordinal, and categorical data. To address this, Nonlinear Principal Component Analysis (NLPCA) has been developed, using optimal scaling to transform non-metric variables into quantitative values [3], [4]. More recently, Kendall's Tau-based PCA has emerged as a powerful alternative for handling ordinal and mixed-scale data. Studies such as [5] and [6] demonstrate that Kendall's Tau correlation is particularly effective when variables are ordinal or non-normally distributed, providing greater flexibility than conventional PCA.

One domain where such methodological advances are highly relevant is the study of child stunting, a major public health challenge. Stunting is characterized by impaired growth and development resulting from chronic malnutrition, especially during the first 1,000 days of life [7]. It is influenced by multiple factors, including environmental quality, household economic status, nutritional intake, and child health conditions. For instance, poor access to clean water and sanitation can increase infection risks and impair nutrient absorption [8], while limited household income may reduce access to nutritious foods and healthcare services [9]. Similarly, nutritional deficiencies during critical developmental stages substantially elevate stunting risk [10].

This study introduces a methodological integration that combines path analysis with Kendall's Tau-based Mixed-Scale PCA to construct latent variable models capable of handling mixed-scale survey data. Unlike previous studies that rely exclusively on metric-scale PCA or standard path models, this approach allows for the analysis of complex datasets where variables are heterogeneous in scale and distribution. The contribution of this work lies in improving both the accuracy and interpretability of structural models in the context of child health research. Specifically, it targets the determinants of stunting using primary data collected from Batu City, Indonesia. By addressing the limitations of traditional methods, the proposed framework offers a more robust analytical tool for advancing public health research and informing policy interventions.

The remainder of this paper is organized as follows. Section 2 details the study design, including data collection, variable definitions, construction of latent variables via Kendall's Tau-based Mixed-Scale PCA, the path analysis specification, and model validity metrics. Section 3 reports the empirical findings, covering measurement results, parameter estimation, assumption testing, direct and indirect effects with bootstrap and Sobel mediation tests, and the final path diagram. Section 4 synthesizes the main findings, policy implications, limitations, and directions for future research.

2 Methods

To investigate the complex interplay among the determinants of child health, particularly in the context of stunting in Batu City, this study adopts a robust and integrated methodological framework. Specifically, a combination of Kendall's Tau-based Principal Component Analysis (PCA) and path analysis is employed to construct latent variables from mixed-scale data and to examine their structural relationships. The following subsections describe the methodological procedures in detail, beginning with data collection, followed by the definition of research variables, dimensionality reduction techniques, and the specification of the structural model.

Accordingly, the first step in the methodological process was to gather reliable primary data from the target population, as outlined in the following subsection.

2.1 Data Collection

Primary data were collected on three main constructs: Family Demographics (X_1), Nutritional Consumption (Y_1), and Child Health Conditions (Y_2). The study population consisted of households in Batu City with children under the age of five, and the respondents were mothers of toddlers. A two-stage sampling technique was employed to ensure representativeness of the sample. In the first stage, cluster sampling was applied, while in the second stage purposive sampling was used to refine the selection of respondents.

The required sample size was calculated using the Slovin formula:

$$n = \frac{N}{1 + Ne^2} = \frac{135}{1 + 135(0.05)^2} = 100,$$

where N is the population size (135 households) and e is the margin of error (0.05). Based on this calculation, the final sample consisted of 100 mothers with toddlers, which was considered sufficient to support the subsequent statistical analyses.

2.2 Research Variables

The research variables were grouped into latent constructs supported by multiple indicators, based on both theoretical considerations and empirical findings on child health. Three main constructs were identified: Family Demographics (X_1), Nutritional Consumption (Y_1), and Child Health Conditions (Y_2). These constructs represent key dimensions of child well-being, and their hypothesized relationships are summarized in Fig. 1.

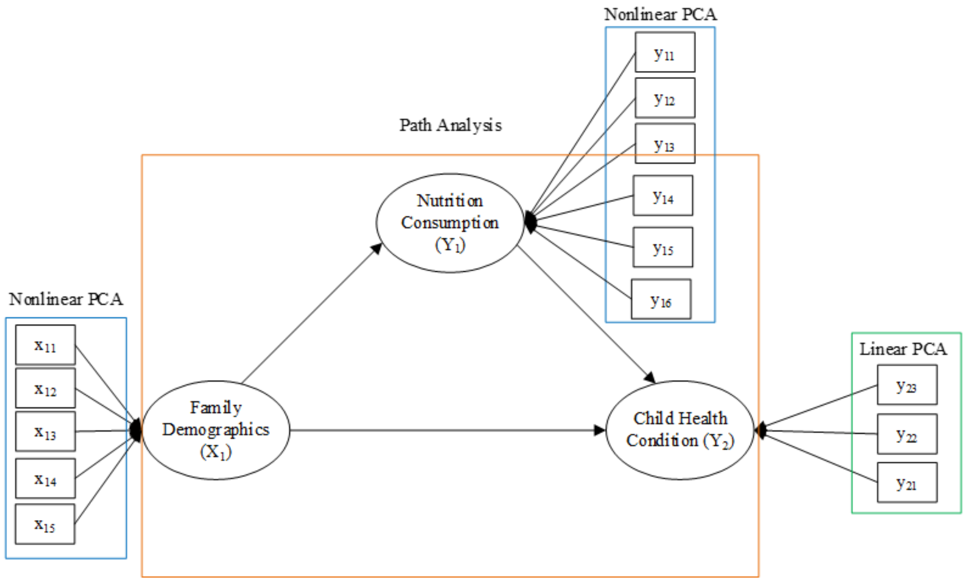

Figure 1: Research framework of the study

Family Demographics (X_1) includes indicators such as gender of the first child, parents' age at marriage, maternal education, and region of residence. These variables reflect the background context in which children grow. Nutritional Consumption (Y_1) is measured through indicators related to dietary adequacy and diversity, such as balanced nutrition, frequency of eating, and consumption of fruits, vegetables, and proteins. Finally, Child Health Conditions (Y_2) is the endogenous construct, consisting of physical, psychological, and social dimensions of health. A summary of the variables, indicators, and their measurement scales is presented in Table 1.

Table 1 also shows that the indicators are measured on mixed scales—some ratio, some ordinal, and some categorical. To address this diversity, Principal Component Analysis (PCA) based on Kendall's Tau correlation was employed. This method is appropriate because it accommodates ordinal and categorical data when forming latent variables. The details of the coding scheme used for categorical and ordinal indicators are provided in Table 2.

Table 1: Research Variables

Variable	**Indicator**	**Scale**
Family Demographics (X_1)	Gender of first child (X_{11})	Ordinal
	Father's age at marriage (X_{12})	
	Mother's age at marriage (X_{13})	Ratio
	Last education (X_{14})	
	Region of Residence (X_{15})	Ordinal
Nutrition Consumption (Y_1)	Balanced Nutrition Food (Y_{11})	
	Carbohydrate and Protein Consumption (Y_{12})	
	Fruit and Vegetable Consumption (Y_{13})	Ordinal
	Frequency of Eating (Y_{14})	
	Animal and Vegetable Side Dishes (Y_{15})	
	Food spent (Y_{16})	
Child Health Condition (Y_2)	Physical (Y_{21})	
	Psychic (Y_{22})	Likert
	Social (Y_{23})	

Table 2: Observation Data on Mixed-Scale Variables

Variable	**Indicator**	**Observation Data**	
		Category	Code
Family demographics (X_1)	Gender of first child (X_{11})	Male	0
		Female	1
	Father's age at marriage (X_{12})	Year	–
	Mother's age at marriage (X_{13})	Year	–
	Last education (X_{14})	Year	–
	Region of Residence (X_{15})	Village	0
		City	1
Nutrition Consumption (Y_1)	Balanced Nutrition Food (Y_{11})	Low	1
		Medium	2
		High	3
	Carbohydrate and Protein Cons. (Y_{12})	Low	1
		Medium	2
		High	3
	Fruit and Vegetable Cons. (Y_{13})	Low	1
		Medium	2
		High	3
	Frequency of Eating (Y_{14})	Low	1
		Medium	2
		High	3
	Animal and Veg. Side Dishes (Y_{15})	Low	1
		Medium	2
		High	3
	Food spent (Y_{16})	Low	1
		Medium	2
		High	3

2.3 Principal Component Analysis Based on Kendall's Tau Correlation

Kendall's tau correlation coefficient is a statistic that measures the strength and direction of the association between two ordinal variables. Kendall's tau, often denoted by τ, evaluates how well the relationship between two variables can be described using their ranked values. In Principal Component Analysis (PCA), a Kendall's tau correlation matrix can be used as the input instead of the conventional Pearson correlation matrix [11]. This approach enables the extraction of principal components or latent variables from ordinal data, producing component scores that are suitable for further analysis. Thus, PCA based on Kendall's tau correlation provides a useful method for reducing the dimensionality of ordinal data while retaining essential information.

In the PCA measurement model, suppose we have indicator data $\{x_1, x_2, \ldots, x_n\}$ that constitute variable X. The equation for the j-th principal component based on Kendall's tau correlation is given in Eq. 1:

$$PC_j = b_{1j}X_1 + b_{2j}X_2 + b_{3j}X_3 + \cdots + b_{pj}X_p, \tag{1}$$

where b_{ij} denotes the weight (loading) of the i-th variable on the j-th component, and j is the index of the component, $j = 1, 2, \ldots, p$.

Eq. 1 is derived through several stages. The first step in PCA based on Kendall's tau correlation is to rank the indicator values $\{x_1, x_2, \ldots, x_n\}$ of variable X from 1 to n. After ranking, each pair of observations is compared sequentially. If the order between two observations is consistent (concordant, C), a value of $+1$ is assigned; if it is inconsistent (discordant, D), a value of -1 is assigned. Kendall's tau correlation is then calculated using Eq. 2:

$$r_\tau = \frac{2(C - D)}{\sqrt{N(N-1) - T_{x_i}}\,\sqrt{N(N-1) - T_{x_{i+1}}}}, \tag{2}$$

where

$$T_{x_i} = \sum_{j=1}^{s} \left(t^2_{j(x_i)} - t_{j(x_i)}\right) \quad \text{is the tie-correction factor for variable } x_i,$$

and

$$T_{x_{i+1}} = \sum_{j=1}^{s} \left(t^2_{j(x_{i+1})} - t_{j(x_{i+1})}\right) \quad \text{is the tie-correction factor for variable } x_{i+1}.$$

The correlation matrix of Kendall's tau is then given by Eq. 3:

$$\mathbf{r} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}. \tag{3}$$

r_{ij} denotes the sample Kendall's tau correlation between variables i and j; $\mathbf{r}$ is the sample correlation matrix; r_s denotes correlations between indicators within the same construct (if applicable). After forming the correlation matrix, the eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$$

are obtained by solving Eq. 4:

$$\det(\mathbf{r} - \lambda I) = 0, \tag{4}$$

where I is the identity matrix. The corresponding eigenvectors, which provide the component loadings (weights), satisfy Eq. 5:

$$(\mathbf{r} - \lambda_i I)\,\mathbf{b}_i = \mathbf{0}. \tag{5}$$

2.4 Path Analysis

Path analysis is an extension of regression analysis. Its development stems from the limitations of regression analysis, which cannot accommodate more complex data structures, particularly those involving multiple response variables. Path analysis was therefore introduced to overcome these shortcomings [12]. It is an analytical technique used to examine the causal relationships between predictor and response variables [13]. The purpose of path analysis is to identify both the direct and indirect effects of a set of exogenous variables on endogenous variables. The sum of these direct and indirect effects yields the total effect [14].

Unlike regression analysis, path analysis requires the use of standardized variables. Standardization is performed using a standard normal transformation with a mean of 0 and a variance of 1 [15], as shown in the following formula:

$$Z_{X_{ji}} = \frac{X_{ji} - \bar{X}_j}{s_{X_j}}, \tag{6}$$

where $Z_{X_{ji}}$ denotes the standardized value of the j-th exogenous variable for the i-th observation, X_{ji} is its observed value, $\bar{X}_j$ is the mean of the corresponding variable, and s_{X_j} is its standard deviation.

The formula for the standard deviation of exogenous and endogenous variables is given in Eq. 7:

$$s_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}(X_{ji} - \bar{X}_j)^2}{n-1}}. \tag{7}$$

Based on Fig. 1, the standardized equation models can be expressed as Eq. 8 and Eq. 9:

$$Z_{Y_{1i}} = \beta_{11} Z_{X_{1i}} + \varepsilon_{Y_{1i}}, \tag{8}$$

$$Z_{Y_{2i}} = \beta_{12} Z_{X_{1i}} + \beta_{22} Z_{Y_{1i}} + \varepsilon_{Y_{2i}}. \tag{9}$$

After the latent constructs were formed through Kendall's Tau-based PCA, the next step was to analyze the structural relationships among these constructs using path analysis.

2.5 Model Validity

The validity of a path analysis model can be evaluated through the coefficient of total determination [14]. This coefficient summarizes the proportion of variance in the observed data that is explained collectively by the structural model, thereby serving as an overall measure of model fit. It is defined as:

$$R^2_{T,adj} = 1 - (1 - R^2_{1,adj})(1 - R^2_{2,adj}) \cdots (1 - R^2_{k,adj}), \tag{10}$$

where $R^2_{k,adj}$ represents the adjusted coefficient of determination for the k-th inner model equation. This value can be computed using Eq. 11:

$$R^2_{k,adj} = 1 - \frac{\sum_{i=1}^{n}(y_{ki} - \hat{y}_{ki})^2/(n-p-1)}{\sum_{i=1}^{n}(y_{ki} - \bar{y}_k)^2/(n-1)}. \tag{11}$$

In this formulation, R^2_T denotes the total coefficient of determination, which reflects the explanatory power of the full model; R^2_k is the coefficient of determination for the k-th inner model equation; $\hat{y}_{ki}$ represents the predicted value of the i-th endogenous variable; and $\bar{y}_k$ is the mean of the corresponding endogenous variable. Taken together, these measures provide a comprehensive indication of how well the proposed path analysis model accounts for the variability in the data. A higher $R^2_{T,adj}$ value suggests a stronger model fit, thereby confirming the robustness of the structural relationships under study.

3 Results and Discussion

This section presents the findings derived from the statistical analysis conducted using the integrated approach of Kendall's Tau-based PCA and path analysis. The discussion is structured to first describe the construction and evaluation of latent variables, followed by an analysis of the structural model and interpretation of the relationships among key constructs. Each subsection highlights the empirical evidence supporting the theoretical framework of this study.

3.1 Latent Variable Measurement Method

In this study, the measurement models for family demographic variables (X_1) and nutritional consumption (Y_1) were developed using Principal Component Analysis based on Kendall's Tau correlation. In contrast, the child health condition variable (Y_2) was analyzed using conventional principal component analysis.

The measurement models were then evaluated to determine which indicators significantly contributed to the formation of each latent variable. This evaluation was carried out by assessing the extent to which the predetermined indicators were able to represent their respective constructs. The resulting component weights for each variable are presented in Table 3.

Table 3: Research Variables

Variable	Indicator	Weight Est	P-Value	Result
X_1	X_{11}	0.172	0.391	not sig
	X_{12}	0.538	0.007	sig
	X_{13}	0.671	0.001	sig
	X_{14}	0.851	0.000	sig
	X_{15}	0.371	0.063	not sig
Y_1	Y_{11}	0.828	0.000	sig
	Y_{12}	0.451	0.024	sig
	Y_{13}	0.820	0.000	sig
	Y_{14}	0.716	0.000	sig
	Y_{15}	0.377	0.059	not sig
	Y_{16}	0.719	0.000	sig
Y_2	Y_{21}	0.688	0.006	sig
	Y_{22}	0.552	0.001	sig
	Y_{23}	0.555	0.006	sig

It can be seen from Table 3 that indicators with higher component weights are considered more important and exert a stronger influence in forming the latent variables. For the demographic construct (X_1), the strongest indicator is last education ($X_{1.4}$) with a component weight of 0.851. For the nutritional consumption construct (Y_1), the most influential indicator is balanced nutritional food ($Y_{1.1}$) with a component weight of 0.828. For the child health condition construct (Y_2), the strongest indicator is the psychological factor ($Y_{2.1}$) with a component weight of 0.688.

Based on the obtained component weights, the measurement models of the latent variables can be expressed as follows:

1. Measurement model of family demographic variables:

$$X_1 = 0.172X_{1.1} + 0.538X_{1.2} + 0.671X_{1.3} + 0.851X_{1.4} + 0.371X_{1.5}$$

2. Measurement model of nutrition consumption variables:

$$Y_1 = 0.828Y_{1.1} + 0.451Y_{1.2} + 0.820Y_{1.3} + 0.716Y_{1.4} + 0.377Y_{1.5} + 0.719Y_{1.6}$$

3. Measurement model of child health condition variables:

$$Y_2 = 0.688Y_{2.1} + 0.552Y_{2.2} + 0.555Y_{2.3}$$

3.2 Parameter Estimation

Parameter estimation in path analysis was carried out using Ordinary Least Squares (OLS). The results of the parameter estimation are presented in Table 4.

Table 4: Path Analysis Coefficients

Parameter	Path Coefficient
$\beta_{x_1y_1}$	0.623
$\beta_{x_1y_2}$	0.217
$\beta_{y_1y_2}$	0.793

Table 4 shows both direct and indirect effects. Based on the estimated path coefficients, the structural equations can be written as:

$$Z_{Y_{1i}} = 0.623 Z_{X_{1i}} + \varepsilon_{Y_{1i}}, \tag{12}$$

$$Z_{Y_{2i}} = 0.217 Z_{X_{1i}} + 0.793 Z_{Y_{1i}} + \varepsilon_{Y_{2i}}. \tag{13}$$

3.3 Assumption Testing

The validity of the path analysis model was examined through assumption testing, including linearity, additivity, and normality of residuals.

1. **Linearity and Additivity Assumptions.** Linearity was tested using the Ramsey `RESET` method. The results are shown in Table 5. As indicated, all relationships between exogenous and endogenous variables produced p-values greater than $\alpha = 0.05$, leading to the acceptance of H_0 and confirming that the relationships are linear. Additivity was assessed by examining the structural model in Eq. 12 and Eq. 13. Since the variables do not interact with one another, the model is additive. Thus, both linearity and additivity assumptions were satisfied.
2. **Normality of Residuals.** The normality of residuals was tested using the Jarque–Bera method, with results presented in Table 6.

Table 5: Linearity Assumption Test Results

Variable	P-Value	Relationship
Family Demographics (X_1) → Nutrition Consumption (Y_1)	0.607	Linear
Family Demographics (X_1) → Child Health Condition (Y_2)	0.906	Linear
Nutrition Consumption (Y_1) → Child Health Condition (Y_2)	0.554	Linear

Table 6: Residual Normality Assumption Test Results

Endogenous Variable	P-Value	Conclusion
Nutrition Consumption (Y_1)	< 0.001	Not normal
Child Health Condition (Y_2)	< 0.001	Not normal

As shown in Table 6, both endogenous variables violate the normality assumption. To address this, hypothesis testing in this study was conducted using a bootstrap approach.

3.4 Direct Effect Test

To further evaluate the relationships specified in the structural model, hypothesis testing was carried out for both direct and indirect effects. The following subsections present the results of these tests in detail.

1. **Direct Effect Test.** Hypothesis testing was conducted using a bootstrap approach to determine the significance of the estimated model. The results of this test are presented in Table 7. Based on the table, the path coefficient values obtained from the OLS and bootstrap methods are consistent, indicating the robustness of the estimation. Specifically:
   - The direct effect of Family Demographics (X_1) on Nutrition Consumption (Y_1) is significant ($p < 0.001$), suggesting that better demographic conditions—such as maternal age, education, and household composition—positively influence nutritional intake in toddlers.
   - The direct effect of Nutrition Consumption (Y_1) on Child Health Condition (Y_2) is also highly significant ($p < 0.001$), confirming that adequate and balanced nutrition is crucial for improving child health outcomes. This finding is consistent with previous studies that identify nutrition as a key determinant of child health and stunting risk [7], [10].
   - In contrast, the direct effect of Family Demographics (X_1) on Child Health Condition (Y_2) is not statistically significant ($p = 0.096$), implying that demographic factors may influence child health indirectly, most likely through nutritional behavior.

   These results are consistent with the findings of [8], who emphasized the indirect role of socioeconomic and demographic factors operating through mediators such as nutritional practices and environmental access.
2. **Indirect Effect Test.** The indirect effect was tested to evaluate whether a mediating variable significantly contributes to the model. The results of the Sobel Test are presented in Table 8.

Table 7: Hypothesis Testing of Direct Influence

Variable	OLS Path Coefficient	Bootstrap Path Coefficient	P Value
Family Demographics (X1) → Nutrition Consumption (Y1)	0.623	0.621	< 0.001
Family Demographics (X1) → Child Health Condition (Y2)	0.217	0.219	0.012
Nutrition Consumption (Y1) → Child Health Condition (Y2)	0.793	0.790	< 0.001

Table 8: Hypothesis Testing of Indirect Influence Variable P-value Conclusion

Variable	P-Value	Conclusion
Family Demographics ($X1$) → Child Health Condition (Y_2)	0.096	Mediation
Family Demographics ($X1$) → Nutrition Consumption (Y_1) → Child Health Condition (Y_2)	0.014	Perfect

Based on Table 8, the Sobel test results indicate that the indirect effect of X_1 on Y_2 through Y_1 is statistically significant ($p = 0.014$), whereas the direct effect of X_1 on Y_2 remains non-significant ($p = 0.096$). This pattern satisfies the conditions for perfect mediation, where the influence of the independent variable on the dependent variable occurs entirely through the mediating variable.

In this context, the findings suggest that family demographic factors affect child health conditions only when mediated by nutritional consumption. For instance, a mother's age or education level may not directly improve a child's health unless they lead to better feeding practices or dietary quality. This is consistent with the conceptual model of stunting presented by Suherman and Nurhaidah [10], which emphasizes the role of nutritional adequacy as a mediating pathway linking structural household characteristics to child development outcomes.

These results highlight the critical role of nutrition as a mediator in health intervention models. They suggest that policy efforts aimed at improving child health and reducing stunting should not only address structural demographic improvements (e.g., education or economic support) but also ensure that such improvements are translated into enhanced nutrition-related behaviors and practices.

Furthermore, the use of Kendall's Tau-based Mixed-Scale PCA in this study adds methodological strength, as it enables robust modeling of latent constructs derived from mixed-scale indicators—an issue frequently encountered in field survey data.

3.5 Research Result Model

Based on the path significance test using the bootstrap approach, all variables exert significant influences, either directly or indirectly. The resulting research model is illustrated in Fig. 2.

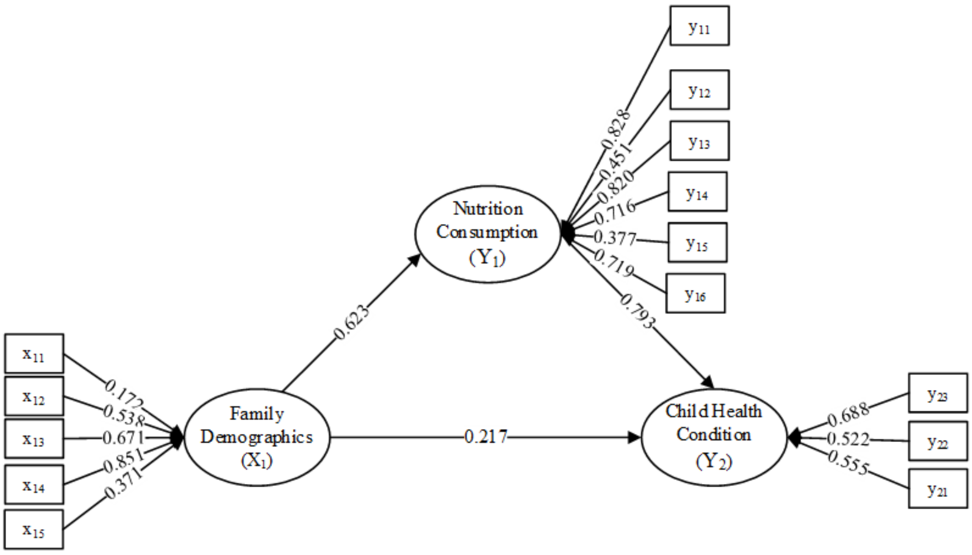

Figure 2: Path diagram of research results

3.6 Model Validity

The validity of the path model was assessed using the adjusted R^2 value. The adjusted R^2 was found to be 0.7762, indicating that 77.62% of the variance in the data can be explained by the model, while the remaining 22.38% is accounted for by other variables outside the model.

4 Conclusion

This study demonstrates that family demographic factors significantly influence nutritional consumption, which in turn has a strong and direct effect on child health conditions. The final path model reveals that Nutritional Consumption has the largest effect on Child Health Condition, with a path coefficient of 0.793, highlighting the central role of nutrition in early childhood health. Furthermore, mediation analysis using the Sobel test confirms a perfect mediation effect, indicating that Family Demographics influence Child Health Condition indirectly through Nutritional Consumption, while the direct effect is statistically insignificant.

The model exhibits strong explanatory power, with a coefficient of determination (R^2) of 77.62%, suggesting that the proposed model effectively captures the variance in child health outcomes based on the studied variables. These findings emphasize the importance of translating improvements in family demographics—such as maternal education, age, and household characteristics—into effective nutritional practices. Policymakers and health practitioners should therefore prioritize nutrition education and targeted family support programs, particularly for mothers with toddlers, to ensure that structural demographic improvements lead to better child health outcomes. Interventions focusing on nutrition can serve as critical leverage points for reducing stunting and promoting healthy early development.

This study has several limitations. First, the sample size was relatively small ($n = 100$) and limited to Batu City, which may reduce the generalizability of the findings. Second, the data relied on self-reported questionnaires, which are subject to recall and response bias. Third, although the model incorporated key variables, other important factors—such as environmental hygiene, access to healthcare services, and parenting practices—were not included and could also influence child health outcomes.

Future studies could expand the scope by employing larger and more diverse samples across different regions. Researchers are also encouraged to incorporate additional latent constructs, such as parenting quality, healthcare accessibility, and environmental sanitation, to enrich the model. Moreover, longitudinal studies are needed to capture causal dynamics over time, especially during the first 1,000 days of life, which represents a critical window for child development.

CRediT Authorship Contribution Statement

Viky Iqbal Azizul Alim: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing–Original Draft, Writing–Review & Editing, Project Administration. **Atiek Iriany:** Methodology, Validation, Resources, Writing–Review & Editing, Project Administration. **Adji Achmad Rinaldo Fernandes:** Conceptualization, Validation, Writing–Review & Editing, Supervision. **Solimun:** Methodology, Validation, Writing–Review & Editing, Supervision. **Candra RWS Weni Utomo:** Resources, Data Curation, Visualization, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted technologies

Generative AI or AI-assisted technologies were used solely for language support during manuscript preparation. Specifically, Grammarly and ChatGPT (version 4) were employed for translation from Indonesian to English and for proofreading/clarity edits. No AI tools were used for data analysis, statistical modeling, or generation of results and figures.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding. The authors thank the Department of Statistics, Brawijaya University, and the participating mothers in Batu City for their valuable time and responses that made this study possible.

Data and Code Availability

The data and code supporting the findings of this study are available from the corresponding author upon reasonable request and subject to confidentiality considerations.

References

[1] F. Murtagh and A. Heck, *Multivariate data analysis*. Springer Science & Business Media, 2012, vol. 131. DOI: 10.1007/978-94-009-3789-5.

[2] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th. Upper Saddle River, NJ: Prentice Hall, 2007. Available online.

[3] U. Kruger, J. Zhang, and L. Xie, "Developments and applications of nonlinear principal component analysis–a review," *Principal manifolds for data visualization and dimension reduction*, pp. 1–43, 2008. DOI: 10.1007/978-3-540-73750-6_1.

[4] A. Gifi, *Nonlinear multivariate analysis*. Wiley Chichester, 1990, vol. 1. Available online.

[5] H. Yu, F. Khan, and V. Garaniya, "An Alternative Formulation of PCA for Process Monitoring Using Distance Correlation," en, *Industrial & Engineering Chemistry Research*, vol. 55, no. 3, pp. 656–669, Jan. 2016. DOI: 10.1021/acs.iecr.5b03397.

[6] C. J. Guembou Shouop, M. Ndontchueng Moyo, E. J. Nguelem Mekongtso, O. Motapon, and D. Strivay, "Application of energy dispersive X-ray fluorescence, -ray spectrometry and multivariate statistical approach for the classification of soil/sand from Douala – Cameroon," en, *Radiation Physics and Chemistry*, vol. 188, p. 109 589, Nov. 2021. DOI: 10.1016/j.radphyschem.2021.109589.

[7] R. Hizriyani, "Pemberian Asi Ekslusif Sebagai Pencegahan Stunting," *J. Jendela Bunda Program Studi PG-PAUD Universitas Muhammadiyah Cirebon*, vol. 8, no. 2, pp. 55–62, 2021. DOI: 10.32534/jjb.v8i2.1722. Available online.

[8] A. Arifuddin et al., "Epidemiological model of stunting determinants in indonesia," *Healthy Tadulako Journal (Jurnal Kesehatan Tadulako)*, vol. 9, no. 2, pp. 224–234, 2023. DOI: 10.22487/htj.v9i2.928.

[9] N. B. Brahmana, V. S. Manalu, D. Nababan, T. R. Sinaga, and F. L. Tarigan, "Faktor-faktor yang berhubungan dengan kejadian stunting pada balita di desa marbun tonga marbun dolok kecamatan baktiraja kabupaten humbang hasundutan tahun 2021," *Journal of Healthcare Technology and Medicine*, vol. 7, no. 2, pp. 1674–1690, 2022. Available online.

[10] R. Suherman and N. Nurhaidah, "Analisis Faktor Determinan Stunting di Desa Pesa Kecamatan Wawo Kabupaten Bima," *Jurnal Manajemen Kesehatan Indonesia*, vol. 8, no. 2, pp. 120–126, Aug. 2020. DOI: 10.14710/jmki.8.2.2020.120-126.

[11] D. Zebrowska, "Elimination of characteristics concerning the performance of open-ended equity funds using PCA," *Stat. Transit*, vol. 24, no. 4, pp. 153–168, 2023. DOI: 10.59170/stattrans-2023-057.

[12] E. Arisoesilaningsih, I. Serafinah, R. Rurini, and R. Adji Achmad, *Pemodelan Pertumbuhan Umbi Amorphophallus onchophyllus pada Beberapa Agroforestri di Jawa Timur Menggunakan Program Smart PLS*. Malang: Universitas Brawijaya, 2010. Available online.

[13] B. M. Byrne, "Structural equation modeling with amos, eqs, and lisrel: Comparative approaches to testing for the factorial validity of a measuring instrument," *International journal of testing*, vol. 1, no. 1, pp. 55–86, 2001. DOI: 10.1207/S15327574IJT0101_4.

[14] S. Astutik et al., *Analisis Multivariat: Teori dan Aplikasinya dengan SAS*. Universitas Brawijaya Press, 2018. Available online.

[15] C. C. Li, *Path Analysis-a primer*. The Boxwood Press, 1975. Available online.