Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:132
https://doi.org/10.28919/cmbn/8915
ISSN: 2052-2541

# REVEALING THE HIDDEN PATTERN OF UNDER-FIVE MALNUTRITION PREVALENCE DISTRIBUTION IN WEST JAVA-INDONESIA FROM CANONICAL CORRESPONDENCE ANALYSIS AND PREDICTIVE CLUSTERING PERSPECTIVE

KARUNIA EKA LESTARI[1,*], ATTIN WARMI[1], SRI WINARNI[2], SISILIA SYLVIANI[3], RISNAWITA[4], EDWIN SETIAWAN NUGRAHA[5], MOKHAMMAD RIDWAN YUDHANEGARA[1]

[1]Mathematics Education Department, Universitas Singaperbangsa Karawang, Karawang 41361, Indonesia

[2]Statistics Department, Universitas Padjajaran, Sumedang 45363, Indonesia

[3]Mathematics Department, Universitas Padjajaran, Sumedang 45363, Indonesia

[4]Mathematics Education Department, UIN Sjeh M. Djamil Djambek, Bukittinggi 26181, Indonesia

[5]Study Program of Actuarial Science, President University, Bekasi 17530, Indonesia



**Abstract:** The global burden of under-five malnutrition has become a critical and urgent health and welfare problem to the present. Since under-five malnutrition has varying determinant factors, this study offers another perspective to reveal the hidden patterns of the malnutrition prevalence distribution among under-fives by exploring their associations. Canonical correspondence analysis and predictive clustering were employed to (1) reveal the association between the under-five malnutrition prevalence with nutritional intake, environmental health, spatial, and socio-economic factors and (2) predict the clustering of under-five malnutrition emergency status for a certain region. The

[*]Corresponding author
E-mail address: karunia@staff.unsika.ac.id
Received September 21, 2024

results suggest that (1) there exists a statistically significant association between the prevalence of under-five malnutrition and its determinant factors, and (2) three clusters have been identified to represent the under-five malnutrition emergency status (critical, serious, warning) and a model has been constructed with an accuracy rate of 83.3% in predicting a certain region status with given attribute variable values. The designation of under-five malnutrition emergency status is desired to provide direction regarding the priority level in handling malnutrition for certain regions.

**Keywords:** canonical correspondence analysis; predictive clustering; ordination techniques; under-five malnutrition.

**2020 AMS Subject Classification:** 62H25, 62H30.

## 1. INTRODUCTION

The global burden of malnutrition has been an extremely critical and urgent issue. Malnutrition refers to either a nutrient intake deficiency or excess, an imbalance of essential nutrients, or disturbances in nutrient utilization [1]. Undernutrition is characterized by a nutritional disorder resulting from insufficient food intake, infectious/parasitic diseases, and poor absorption alone or in combination [2]. Therefore, in some literature, malnutrition is considered synonymous with undernutrition, defined as the intake of energy and nutrients that are insufficient to adequately supply the individual's needs for proper health. [3]. The World Health Organization (WHO) estimates that globally, by 2022, around 230 million under-fives suffer from malnutrition and slightly more than half of under-five deaths are related to malnutrition, especially in low-middle income countries [4]. Taylor [5] reported that more than 30 million under-five in the 15 countries worst affected by the recent famine were suffering from acute malnutrition, while 8 million suffered from wasting, which is the severest and most lethal malnutrition manifestation.

Despite a decline in malnutrition rates over the past decade, Indonesia has one of the highest rates of maternal and under-five malnutrition in the world. Indonesia also struggles with the triple burden of malnutrition: stunting, underweight, and wasting [6]. The UNICEF Indonesia [7] reports that by 2022, 1) one in five under-five are stunted; 2) one in twelve under-five are wasted; 3) 1.5 million infants are born with low birth weight, which indicates maternal and child malnutrition;

and 4) only one in two infants aged 0-5 months are exclusively breastfed. In addition, children who are moderately and severely wasted are three times more likely to be stunted, and those who are severely wasted risk almost 12 times more dying than healthy children [8]. As a province with the largest population, West Java, which reached 49,405 people in 2022, has a 20.2% malnutrition prevalence [9]. The fact suggests that Indonesians under-five are facing a massive vulnerability to malnutrition with a wide prevalence of stunting, underweight, and wasting.

*Stunting* is defined as low length/height for age (LfA or HfA) and is identified by measuring an under-five length at lying down for under-five under 2 years of age and height at standing for under-five 2 years of age or older [4], [10]. The under-five is categorized as stunted if the LfA or HfA score is below -2 standard deviations (SD) from the median of the WHO child growth standards referred to as the LfA-z-score (LAZ) or HfA-z-score (HAZ), which means that the under-five length/height is too low for age and sex [11]. Stunting is the devastating impact of poor nutrition in utero and during childhood. Stunted children can have irreversible physical and cognitive disorders that inherent linear growth retardation. Children who suffer from stunting will never reach their optimal height, and their brains will never develop to their ideal potential. These children start life at a disadvantage with consequences that continue into adulthood, such as facing learning difficulties at school, earning less as adults, and facing social participation barriers [4]. Thus, the harmful effects of stunting occur over a lifetime and even affect the next generation.

*Underweight* is defined as low weight-for-age (WfA) and is identified by measuring the weight of under-five [12]. An under-five is designated as underweight if the WfA score is below -2SD from the median of the WHO child growth standards referred to as the WfA-z-score (WAZ), which means the under-five weight is too low for age and sex [11]. Being underweight occurs due to rapid weight loss or failure to gain weight, commonly associated with nutritional deficiencies. Fishman *et al*. [13] reported that underweight in under-five is a risk factor for death from diarrhea, pneumonia, measles, and malaria.

*Wasting* is determined based on the weight-for-height index (WfA/LfA or WfA/HfA). An under-five is recognized as wasted if the WfA/LfA or WfA/HfA score is below -2SD from the

WHO median child growth standard referred to as the WfA/LfA-z-score (WAZ/LAZ) or WfA/HfA-z-score (WAZ/HAZ). Childhood wasting is a life-threatening consequence of poor nutrition or recurrent illness. Under-five suffering from wasting possess weakened immunity, are susceptible to long-term developmental delays, and face an increased risk of death, especially when it is chronic wasting. Under-five with chronic wasting require early detection and timely treatment to survive [4].

The Food and Agriculture Organization (FAO) and four other United Nations agencies, including the United Nations Children's Fund (UNICEF), the World Food Programme (WFP), the UN Refugee Agency (UNHCR), and WHO are building an alliance to appeal rapid action for preventing, detecting and treating acute under-five malnutrition in countries most severely affected by famine. The agencies called for more investment in a global action plan to reduce maternal and under-five malnutrition prevalence by addressing food, health, water and sanitation, and social protection systems [12]. Likewise, the Indonesian government has committed to reduce the prevalence of stunting, underweight, and wasting [8]. Joint action to accelerate the realization of these targets is becoming urgent today. The broadened interventions and roadmap for global action plans to reduce under-five mortality and morbidity, along with improving the overall well-being of society, should be carefully planned. Investigating the pattern of malnutrition prevalence distribution could provide a foundation for designing such interventions. Therefore, exploring determinant factors that might be associated with the under-five malnutrition prevalence is a key starting point.

Malnutrition has varying determinant factors. Numerous studies have been devoted to socio-economic aspects as determinant factors of malnutrition among under-five by employing various multivariate statistical methods, especially in low-middle income countries. Some of them are Daniel *et al*. [14], who applied a multivariate regression model to describe the association between maternal chronic energy deficiency (CED) in the last trimester of pregnancy and the prevalence of stunting in infants during the first 3.5 months of life. The study found that Cambodian infants born to CED mothers faced a 1.6% higher probability of being stunted compared with infants born to non-CED mothers. Uthman & Aremu [15] employed meta-regression analysis to investigate the association between malnutrition prevalence and residence location by comparing rural and urban populations in sub-Saharan Africa. The study mentioned that malnourished rural pregnant women are more likely to deliver low birth weight babies who face an increased risk of mortality in infancy and are more likely to suffer from stunting during childhood than urban mothers.

Kumera *et al*. [16] performed logistic regression to identify the malnutrition status of pregnant women attending antenatal care at the University of Gondar Hospital, Ethiopia, considering a set of factors associated with malnutrition during pregnancy (including anemia and socio-demographic aspects) to reduce maternal-child morbidity and mortality. Multivariable logistic regression analysis revealed that anemia was one of the significant predictors of malnutrition. Bühler *et al*. [17] conducted conditional multivariate regression methods to analyze the associations between malnutrition causes and different food security dimensions in Cambodian and Lao PDR children, as one of their findings suggests that individual-level malnutrition was much more prominent than household-level food insecurity.

Yanuar *et al*. [18] constructed a stunting rate model using Tobit quantile regression with exclusive breastfeeding as one of the 10 predictors in such a model. The study found that exclusive breastfeeding provides a substantial impact on stunting rate at middle quantiles. Fitri and Wijayanto [19] compared four linear models from the generalized linear and autoregressive family to model stunting cases in Java, Indonesia. The models involved predictor variables such as exclusive breastfeeding, adequate sanitation, and poverty. The comparative study revealed that stunting cases in Java Island could be mitigated by estimating the relative risk levels of these variables using a Conditional Autoregressive Bayesian Model (CAR BYM). Djara *et al*. [20] considered a spatial autoregressive model to explain the prevalence of stunting and underweight among under-fives in Java, Indonesia, by engaging seven explanatory variables, including human development index and proper drinking water.

The aforementioned and other related studies primarily rely on regression-based algorithms to analyze the association among under-five malnutrition determinants. The remaining studies employed descriptive statistics and chi-squared tests to investigate these associations. Indeed, Braghetta [21] has undertaken a systematic review using simple descriptive statistics to describe the connection between under-five malnutrition and the scarcity of safe drinking water in Guatemala. Hasanain *et al*. [22] relied on the chi-squared test to conclude that safe drinking water in Iraq could lead to increased in diarrhea cases and ultimately to under-five malnutrition. However, it is challenging to find literature that examines the association between under-five malnutrition prevalence and its determinants by performing canonical correspondence analysis (CCA) and predictive clustering, as illustrated in the folowing bibliometric analysis.

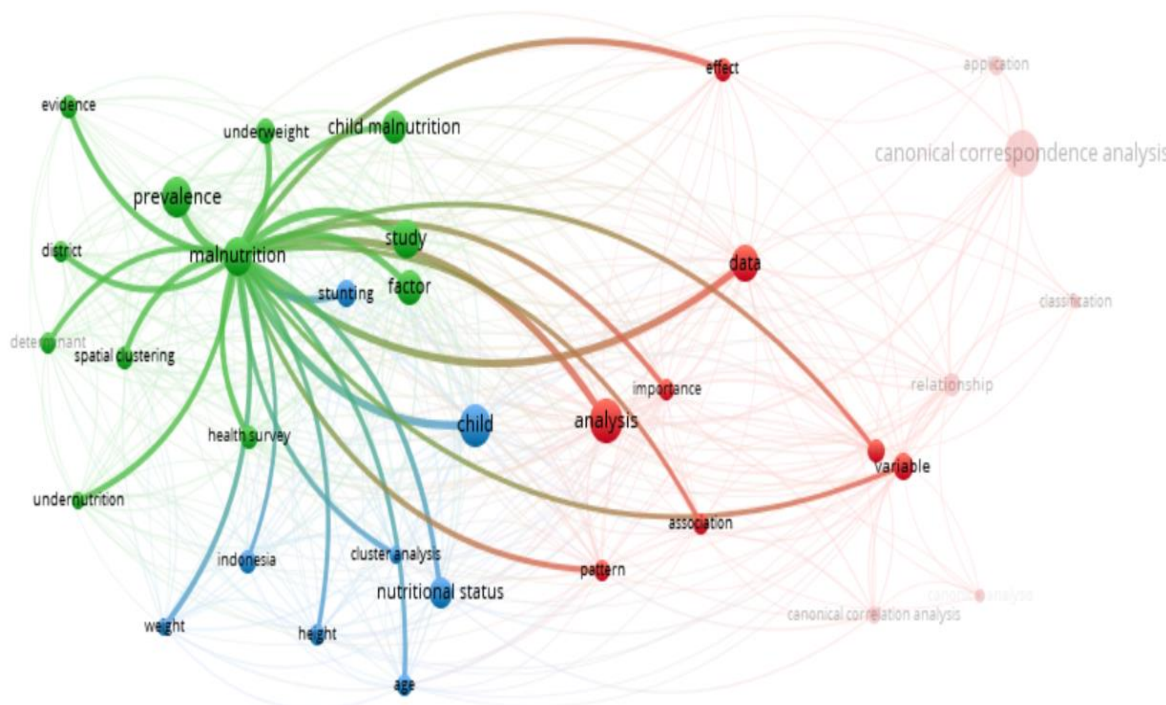



**Figure 1**. Visualization of bibliometric analysis to portray research gaps and potential novelty

This study offers another perspective to reveal the hidden patterns of the malnutrition prevalence distribution among under-fives by exploring its associations with various determinants, including nutritional intake, environmental health, and socio-economic factors. The under-five malnutrition prevalence was explored in a spatial manner by considering districts and cities in West Java, Indonesia, as a case study. Nutritional intake factors are represented by the ratio of pregnant women with CED and anemia for each region, as well as the percentage of exclusive breastfeeding of infants under six months of age. Environmental health factors are considered in terms of access to safe drinking water for pregnant women and access to proper sanitation for households. Socio-economic factors are measured using three indicators: poverty line, human development index (HDI), and food security index (FSI). The two main concerns in this study are (1) revealing the association between the under-five malnutrition prevalence with nutritional intake, environmental health, socio-economic and spatial factors and (2) predicting the clustering of under-five malnutrition emergency status for a region with given attributes.

CCA proposed by Ter Braak [23] as a multivariate direct gradient analysis incorporating ordination techniques, was adopted to detect hidden patterns of variation in the composition of spatial aspects representing the under-five malnutrition prevalence for each district and cities in West Java explained by a set of determinant variables. The association between these variables is visualized graphically by CCA plots. Afterwards, predictive clustering that combines k-means clustering and linear discriminant analysis (LDA) is employed to determine the cluster of regions in West Java according to their malnutrition status and predict the malnutrition status of a region if information on nutritional intake, environmental health, and socio-economic factors are known.

## 2. MATERIALS AND METHODS

This section briefly provides conceptual theory and methods for data analysis, including canonical correspondence analysis, predictive clustering, data sources, and variables.

### 2.1. Canonical Correspondence Analysis

The CCA input takes from a two-way contingency table represented by a bivariate frequency matrix $\mathbf{N} = \left(n_{ij}\right)$ consisting of $\mathscr{i}$ rows $(i = 1, \cdots, \mathscr{i})$ and $\mathscr{j}$ columns $(j = 1, \cdots, \mathscr{j})$, such that the elements of $n_{ij}$ are non-negative integers [24], [25]. For simplicity, consider $\mathscr{i} \geq \mathscr{j}$. As a result, the row marginal frequencies $n_{i\bullet}$ are collected in a diagonal matrix $\mathbf{D}_{\mathscr{i}}$ of size $\mathscr{i} \times \mathscr{i}$ and the column marginal frequencies $n_{\bullet j}$ as the diagonal elements of matriks $\mathbf{D}_{\mathscr{j}}$ of size $\mathscr{j} \times \mathscr{j}$. Let

$\mathbf{w}_{\imath}$ and $\mathbf{w}_{\jmath}$ be vectors of length $\imath$ and $\jmath$, respectively, whose elements are all one. Thus, the total frequency can be written as $n = \mathbf{w}_{\imath}^{t}\, \mathbf{N}\, \mathbf{w}_{\jmath}$.

Suppose one seeks row score $\mathbf{X}$ and column score $\mathbf{Y}$ such that the correlation in the bivariate table $\mathbf{N}$ is as large as possible. It implies maximizing $\lambda(x, y) = n^{-1}\mathbf{x}^{t}\mathbf{N}\mathbf{y}$ over the vector of row scores $\mathbf{x}$ and the vector of column scores $\mathbf{y}$. These vectors are centered by

$$\mathbf{w}_{\imath}^{t}\mathbf{D}_{\imath}\mathbf{x} = \mathbf{0} \text{ and } \mathbf{w}_{\jmath}^{t}\mathbf{D}_{\jmath}\mathbf{y} = \mathbf{0}, \tag{1}$$

and normalized by

$$\mathbf{x}^{t}\mathbf{D}_{\imath}\mathbf{x} = N \text{ and } \mathbf{y}^{t}\mathbf{D}_{\jmath}\mathbf{y} = N. \tag{2}$$

The vectors are centered and normalized, then called standardized. In turn, the Lagrange multiplier $(\xi_x, \xi_y, \mu_x, \mu_y)$ is formed to solve the optimized constraint $\lambda(x, y)$ on the vector $(x, y)$ that satisfies the such conditions by setting its derivative equal to zero. Thus, the optimal $\mathbf{x}$ and $\mathbf{y}$ should meet the centering and normalization conditions in Eqs. (1) and (2), along with the stationary equations as follows

$$\mathbf{N}\mathbf{y} = \xi_x\mathbf{D}_{\imath}\mathbf{x} + \mu_x\mathbf{D}_{\imath}\mathbf{w}_{\imath} \text{ and } \mathbf{N}^{t}\mathbf{x} = \xi_y\mathbf{D}_{\jmath}\mathbf{y} + \mu_y\mathbf{D}_{\jmath}\mathbf{w}_{\jmath}. \tag{3}$$

By taking the constraints in Eqs. (1) and (2), the Lagrange multiplier fulfills $\mu_x = \mu_y = 0$ and $\xi_x, = \xi_y = \sigma(x, y)$. This leads to a simpler system that satisfies these conditions, such as

$$\mathbf{N}\mathbf{y} = \sigma\mathbf{D}_{\imath}\mathbf{x} \text{ and } \mathbf{N}^{t}\mathbf{x} = \sigma\mathbf{D}_{\jmath}\mathbf{y}. \tag{4}$$

The system in Eq. (4) is a singular value problem. De Leeuw and Mair [26] found the stationary value of $\sigma$ to be the singular value of the $\mathbf{S}$, that is

$$\mathbf{S} = \mathbf{D}_{\imath}^{-\frac{1}{2}}\mathbf{N}\mathbf{D}_{\jmath}^{-\frac{1}{2}}. \tag{5}$$

Since $\imath \geq \jmath$, the following singular value decomposition (SVD) is obtained.

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{t}. \tag{6}$$

Matrix $\boldsymbol{\Sigma}$ is a diagonal matrix containing $\min(\imath, \jmath) = \jmath$ singular values in descending order. The $\imath \times \imath$ matrix $\mathbf{U}$ consists of left singular vectors, while the $\jmath \times \jmath$ matrix $\mathbf{V}$ consists of right singular vectors. Matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal, which means $\mathbf{U}^{t}\mathbf{U} = \mathbf{I} = \mathbf{V}^{t}\mathbf{V}$.

The $j$ solution of the stationary Eq. (4) yields the following row scores and column scores.

$$\mathbf{X} = \sqrt{n}\mathbf{D}_i^{-\frac{1}{2}}\mathbf{U} \quad \text{and} \quad \mathbf{Y} = \sqrt{n}\mathbf{D}_j^{-\frac{1}{2}}\mathbf{V}. \tag{7}$$

Here, $\mathbf{X}$ is a row score matrix of size $i \times j$, and $\mathbf{Y}$ is a column score matrix of size $j \times j$ (excluding repeated singular values, since its solution is uniquely determined). Later, de Leeuw and Mair [26] introduced CCA from a general point of view that has a row covariate $\mathbf{A}$ on the row marginal frequency $n_{i\bullet}$ and/or a covariate column $\mathbf{B}$ on the column marginal frequency $n_{\bullet j}$. Therefore, CCA can be derived by taking a linear regression of $\mathbf{A}$ and $\mathbf{B}$ on the row score $\mathbf{X}$ and column score $\mathbf{Y}$, respectively;

$$\mathbf{X} = \mathbf{AP} \quad \text{and} \quad \mathbf{Y} = \mathbf{BQ}, \tag{8}$$

with $\mathbf{A}$ and $\mathbf{B}$ being matrices of size $i \times a$ and $j \times b$, and weighted matrices $\mathbf{P}$ and $\mathbf{Q}$. Nonetheless, for the sake of generality, assume that $\mathbf{A}$ and $\mathbf{B}$ have full column rank.

Suppose that $\mathbf{w}_i$ and $\mathbf{w}_j$ are the column spaces of $\mathbf{A}$ and $\mathbf{B}$, respectively, such as to get

$$(\mathbf{A}^t\mathbf{NB})\mathbf{Q} = (\mathbf{A}^t\mathbf{D}_i\mathbf{A})\mathbf{P\Sigma} \ \text{ and } \ (\mathbf{B}^t\mathbf{N}^t\mathbf{A})\mathbf{P} = \left(\mathbf{B}^t\mathbf{D}_j\mathbf{B}\right)\mathbf{Q\Sigma}. \tag{9}$$

The row scores $\mathbf{X}$ and column scores $\mathbf{Y}$ in Eq. (8) comply with the following standardization conditions.

$$\mathbf{P}^t(\mathbf{A}^t\mathbf{D}_i\mathbf{A})\mathbf{P} = n\mathbf{I} = \mathbf{Q}^t\left(\mathbf{B}^t\mathbf{D}_j\mathbf{B}\right)\mathbf{Q}. \tag{10}$$

If $\mathbf{w}_i = \mathbf{Ag}$ and $\mathbf{w}_j = \mathbf{Bh}$, then $(g, \text{h})$ is defined as the solution of Eq. (9) with $\sigma = 1$. Thus, there is another dominant trivial solution that ensures all pairs of singular values are at the center.

The remaining problem to be solved in CCA is the SVD of $\mathbf{S}$, as follows

$$\mathbf{S} = (\mathbf{A}^t\mathbf{D}_i\mathbf{A})^{-\frac{1}{2}}\mathbf{A}^t\mathbf{NB}\left(\mathbf{B}^t\mathbf{D}_j\mathbf{B}\right)^{-\frac{1}{2}}. \tag{11}$$

The matrix $\mathbf{S}$ in Eq. (11) is referred to as the canonical standard residual matrix. Taking the inverse of the symmetric square root of $\mathbf{A^tD}_i\mathbf{A}$ and $\mathbf{B^tD}_j\mathbf{B}$, suppose that Eq. (6) is the SVD of $\mathbf{S}$, according to Eq. (11) the following expression is derived

$$(\mathbf{A}^t\mathbf{D}_i\mathbf{A})^{-\frac{1}{2}}\mathbf{A}^t\mathbf{NB}\left(\mathbf{B}^t\mathbf{D}_j\mathbf{B}\right)^{-\frac{1}{2}} = \mathbf{U\Sigma V}^t, \tag{12}$$

such that

$$\mathbf{P} = (\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}\mathbf{U} \text{ and } \mathbf{Q} = \left(\mathbf{B}^t\mathbf{D}_{j}\mathbf{B}\right)^{-\frac{1}{2}}\mathbf{V}, \tag{13}$$

which is the optimal solution for the weights in the maximum correlation problem, and the corresponding scores, respectively;

$$\mathbf{X} = \mathbf{A}(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}\mathbf{U} \text{ and } \mathbf{Y} = \mathbf{B}\left(\mathbf{B}^t\mathbf{D}_{j}\mathbf{B}\right)^{-\frac{1}{2}}\mathbf{V}. \tag{14}$$

The mapping of $\mathbf{X}$ and $\mathbf{Y}$ column vectors yields standardized coordinates on the CCA plot [27]. Both $\mathbf{X}$ and $\mathbf{Y}$ are normalized and orthogonal, except for the centrally dominant solution. If $\mathbf{w}_{i}$ and $\mathbf{w}_{j}$ are the first columns of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively, then the elements of the first row and column of $\mathbf{S}$ are equal to zero, except for the element of $s_{11}$ that takes the value one. Under the independence assumption, the rest of the elements of $\mathbf{S}$ are asymptotically independent and standard normally distributed, $S\sim\mathcal{N}(0,1)$. In this way, the square of the singular values $\sigma^2$ corresponds to the eigenvalues $\lambda$ of $\mathbf{S}^t\mathbf{S}$ and $\mathbf{S}\mathbf{S}^t$ reflects the principal inertia [28], [29]. The total inertia in CCA is determined by

$$n\sum_{m=1}^{\mathcal{M}}\sigma_m^2 = n\sum_{m=1}^{\mathcal{M}}\lambda_m = n(tr\ \mathbf{S}^t\mathbf{S} - \mathbf{1}) = \chi^2\left(\mathbf{N}_{A,B}\right), \tag{15}$$

which is asymptotically chi-square distributed with degrees of freedom $\nu = (a-1)(b-1)$. It indicates that CCA computes the canonical partition of the chi-squared components corresponding to the orthogonal contrasts (covariates) $\mathbf{A}$ and $\mathbf{B}$. Intra-variable distances between the $i$-th and $i'$-th rows are measured by the following Benzécri distance [26].

$$\begin{aligned}
\delta_{ii'}^2\left(\mathbf{N}_{A,B}\right) &= (\mathbf{c}_i - \mathbf{c}_{i'})^t(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^t\mathbf{N}\mathbf{B}\left(\mathbf{B}^t\mathbf{D}_{j}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^t\mathbf{N}^t\mathbf{A}(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\mathbf{1}}(\mathbf{c}_i - \mathbf{c}_{i'}),\\
&= (\mathbf{c}_i - \mathbf{c}_{i'})^t(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}\mathbf{S}\mathbf{S}^t(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}(\mathbf{c}_i - \mathbf{c}_{i'}),\\
&= (\mathbf{c}_i - \mathbf{c}_{i'})^t(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}\mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^t(\mathbf{A}^t\mathbf{D}_{i}\mathbf{A})^{-\frac{1}{2}}(\mathbf{c}_i - \mathbf{c}_{i'}),\\
&= (\mathbf{c}_i - \mathbf{c}_{i'})^t\mathbf{X}\mathbf{\Sigma}^2\mathbf{X}^t(\mathbf{c}_i - \mathbf{c}_{i'}),\\
&= (\mathbf{c}_i - \mathbf{c}_{i'})^t\mathbf{A}\mathbf{P}\mathbf{\Sigma}^2\mathbf{P}^t\mathbf{A}^t(\mathbf{c}_i - \mathbf{c}_{i'}).
\end{aligned} \tag{16}$$

Here, $\mathbf{c}_i$ and $\mathbf{c}_{i'}$ are unit vectors of length $i$. Similarly, the Benzécri distance between the $j$-th and $j'$ −th columns is defined as

$$\delta_{jj'}^2(\mathbf{N}_{A,B}) = \left(\mathbf{d}_j - \mathbf{d}_{j'}\right)^t \mathbf{BQ\Sigma^2Q}^t\mathbf{B}^t\left(\mathbf{d}_j - \mathbf{d}_{j'}\right), \tag{17}$$

where $\mathbf{d}_j$ and $\mathbf{d}_{j'}$ are unit vectors of length $j$.

The last two equations imply that 1) the Benzécri distance between rows of $\mathbf{N}_{A,B}$ is equivalent to the Euclid distance between rows of $\mathbf{X\Sigma}$; and 2) the Benzécri distance between columns of $\mathbf{N}_{A,B}$ is equivalent to the Euclid distance between rows of $\mathbf{Y\Sigma}$. Therefore, the row principal coordinate $\mathbf{X}^*$ and column principal coordinate $\mathbf{Y}^*$ on the CCA plot are determined by

$$\mathbf{X}^* = \mathbf{AP\Sigma} \text{ and } \mathbf{Y}^* = \mathbf{BQ\Sigma}. \tag{18}$$

Substituting Eqs. (8) and (9) into the last equation yields

$$\mathbf{X}^* = \mathbf{D}_i^{-1}\mathbf{NY} \text{ and } \mathbf{Y}^* = \mathbf{D}_j^{-1}\mathbf{N}^t\mathbf{X}. \tag{19}$$

Let $\mathbf{Z}_{A} = (\mathbf{A}^t\mathbf{D}_i\mathbf{A})^{-1}\mathbf{A}^t\mathbf{D}_i\mathbf{X}^*$ and $\mathbf{Z}_{B} = \left(\mathbf{B}^t\mathbf{D}_j\mathbf{B}\right)^{-1}\mathbf{B}^t\mathbf{D}_j\mathbf{Y}^*$, in algebraic manipulation manner of the last two equations leads

$$\mathbf{Z}_{A} = \mathbf{P\Sigma} \text{ and } \mathbf{Z}_{B} = \mathbf{Q\Sigma}, \tag{20}$$

which are the canonical coefficients for row covariate $\mathbf{A}$ and column covariate $\mathbf{B}$, respectively. The $m$-dimensional CCA plot is constructed by depicting the columns of the row principal coordinate matrix and the columns of the canonical coefficient matrix for the row $\mathbf{A}$ and/or column $\mathbf{B}$ covariates. In particular, plotting the first two column vectors of $\mathbf{X}^*, \mathbf{Y}^*$, $\mathbf{A}$ and/or $\mathbf{B}$ produces a two-dimensional CCA plot [30]. Broadly defined, there are nine steps required for constructing a CCA plot, as condensed in Algorithm 1.

**Algorithm 1**. The two-dimensional CCA plot construction

Step 1: Input an $i \times j$ bivariate frequency matrix $\mathbf{N}$ and an $i \times a$ row covariate matrix $\mathbf{A}$ and/or a $j \times b$ column covariate matrix $\mathbf{B}$.

Step 2: Determine the diagonal matrix of row and column marginal frequencies, $\mathbf{D}_i$ and $\mathbf{D}_j$.

Step 3: Calculate the canonical standardized residual matrix $\mathbf{S}$, Eq. (11).
Step 4: Find the SVD of $\mathbf{S}$, Eq. (12).
Step 5: Calculate the weighted matrix $\mathbf{P}$ and $\mathbf{Q}$, Eq. (13).
Step 6: Calculate standard coordinates matrices $\mathbf{X}$ and $\mathbf{Y}$, Eq. (14).
Step 7: Calculate principal coordinates matrices $\mathbf{X}^*$ and $\mathbf{Y}^*$, Eq. (18).
Step 8: Find canonical coefficients covariates $\mathbf{Z}_{A}$ and $\mathbf{Z}_{B}$, Eq. (20).
Step 9: Plotting the first two columns of matrices $\mathbf{X}^*, \mathbf{Y}^*, \mathbf{A}$ and/or $\mathbf{B}$.

## 2.2. Predictive Clustering

Predictive clustering is initiated accomplished by creating $k$ clusters that partition the dataset based on the given attributes using the k-means clustering method. The clusters produced are converted into a prediction set. This set is then used to predict the probability of a cluster's membership based on the specified attributes. The linear discriminant analysis algorithm is applied at this stage to find the directions that maximize the separation between clusters and then employ these directions to predict the clusters of a region. These directions are referred to as linear discriminant functions, which are linear combinations of variable attributes (predictors).

**K-means Clustering**

K-means clustering is widely used as an unsupervised machine learning algorithm to partition a given dataset into a set of k groups (clusters), which $k$ represents the number of groups predefined by the observer/researcher [31]. The underlying idea behind k-means clustering is to define clusters that minimize the total intra-cluster variation, also known as the total within-cluster variation.

$$\mathcal{E}(C_{\Bbbk}) = \sum_{x_i \in C_{\Bbbk}} (x_i - \mu_{\Bbbk})^2, \tag{21}$$

with $x_i$ is the data points belonging to cluster $C_{\Bbbk}$ and $\mu_{\Bbbk}$ is the average (centroid) of the points assigned to cluster $C_{\Bbbk}$ for $\Bbbk = 1, \cdots, k$. Each observation $(x_i)$ is assigned to a particular cluster in such a way that it minimizes the sum of squared distances of the observations to the centers of the assigned clusters [32]. Total within-cluster $(SS_{WC})$ variation is defined as

$$SS_{WC} = \sum_{\Bbbk=1}^{k} \mathcal{E}(C_{\Bbbk}) = \sum_{\Bbbk=\mathbf{1}}^{k} \sum_{x_i \in C_{\Bbbk}} (x_i - \mu_{\Bbbk})^2. \tag{22}$$

**Algorithm 2**. Unsupervised clustering using k-means.

Step 1: Specify the number of $k$ clusters to be established by the observer/researcher.
Step 2: Choose $k$ objects randomly from the dataset as the initial cluster center (centroid).
Step 3: Assign each observation to the nearest center, corresponding to the Euclidean distance between the object and its centroid.
Step 4: Update the cluster centroid by calculating the new average value of all data points in the cluster, for each of the $k$ clusters.
Step 5: Minimize $SS_{WC}$ iteratively by alternating steps 3 and 4 until the cluster assignment stops changing or it reaches the maximum number of iterations.

The algorithm classifies objects in several groups (clusters), ensuring that objects in the same cluster are as similar as possible (high intra-class similarity). In contrast,objects from different clusters are set as dissimilar as possible (low inter-class similarity). Each cluster is represented by its center (centroid), which reflects the average of the points assigned to corresponding the cluster [33]. The centroid of the $k$-th cluster is a vector of length $\ell$ containing the average of all variables for observations in the $k$-th cluster with $\ell$ denoting the number of variables.

**Linear Discriminant Analysis**

LDA is a technique in supervised machine learning for solving multi-class classification problems, including prediction or assignment of observations into groups/clusters using a linear discriminant function. This function assigns an individual sampling unit to the one cluster that is best possible for that individual together with separating multiple clusters with multiple features by reducing the data dimensionality [34]–[36]. Here is the general equation for modeling the linear discriminant function ($LD$).

$$LD = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p, \tag{22}$$

with discriminant coefficient $\beta_j$ and attribute variable $X_j$ as a predictor for $j = 1, 2, \cdots, p$.

**Algorithm 3**. Predictive clustering based on linear discriminant function.

Step 1: Input data matrix $\mathbf{X}$ of size $n \times p$ with $n$ observations and $p$ attribute or predictor variables.
Step 2: Assign a single categorical target variable consisting of $k$ categories which represent the $k$ clusters acquired from the k-means clustering algorithm.
Step 3: Split the data matrix into training data and test data.
Step 4: Establish the linear discriminant model ($LD$), Eq. (22).
Step 5: Predict the model from the test data.
Step 6: Draw the linear discriminant function on a two-dimensional plot.

LDA reduces dimensionality using the X-Y axis (two-dimensional plot) to create a new axis, which straight-line separates the different classes and projects the data onto such a new axis. The new axis is constructed in a manner to maximize the distance between the averages of the two groups/clusters and minimize the variance within each class.

## 2.3. Data Sources and Variables

The data employed are secondary data from government sources that can be accessed online at the Satu Data Indonesia (SDI) portal by visiting www.katalog.data.go.id. SDI is Indonesia's official open data portal managed by the One Data Indonesia Secretariat at the central level, Ministry of National Development Planning (Bappenas). Nine datasets collected from the SDI portal are detailed in Table 1 as variables to be analyzed in this study.

**Table 1**. Variables description

| Variabel | Description |
|---|---|
| Region ($X$) | A categorical variable expressing 27 different regions in West Java, Indonesia, divided into 18 districts; Bogor Dist. ($X_1$), Sukabumi Dist. ($X_2$), Cianjur Dist. ($X_3$), Bandung Dist. ($X_4$), Garut Dist. ($X_5$), Tasikmalaya Dist. ($X_6$), Ciamis Dist. ($X_7$), Kuningan Dist. ($X_8$), Cirebon Dist. ($X_9$), Majalengka Dist. ($X_{10}$), Sumedang Dist. ($X_{11}$), Indramayu Dist. ($X_{12}$), Subang Dist. ($X_{13}$), Purwakarta Dist. ($X_{14}$), West Bandung Dist. ($X_{15}$), Pangandaran Dist. ($X_{16}$), Karawang Dist. ($X_{17}$), Bekasi Dist. ($X_{18}$), and 9 cities; Bogor City ($X_{19}$), Sukabumi City ($X_{20}$), Bandung City ($X_{21}$), Cirebon City ($X_{22}$), Bekasi City ($X_{23}$), Depok City ($X_{24}$), Cimahi City ($X_{25}$), Tasikmalaya City ($X_{26}$). Banjar City ($X_{27}$) |
| Malnutrition ($Y$): | A categorical variable consists of three categories representing different manifestations of under-five malnutrition prevalence. |
| Stunting ($Y_1$) | The number of cases (prevalence) of under-fives who are stunting (count) in West Java, Indonesia 2023. |
| Underweight ($Y_2$) | The number of cases (prevalence) of under-fives who are underweight (count) in West Java, Indonesia 2023. |
| Wasting ($Y_3$) | The number of cases (prevalence) of under-fives who are wasting (count) in West Java, Indonesia 2023. |
| Determinants factors ($A$): | A quantitative variable as a row covariate representing determinant factors that might be associated with the under-five malnutrition prevalence such as nutritional intake, environmental health, and socio-economic factors. |
| CED ($A_1$) | The ratio of the number of pregnant women suffering from CED to the total population of pregnant women for each region in West Java 2023 (in percent). Pregnant women who are at risk of CED have an upper arm circumference (ALC) measurement below 23.5 cm or a body mass index (BMI) at pre-pregnancy or fisrt trimester below 18.5 kg/m$^2$ [37]. |
| Anemia ($A_2$) | The ratio of the number of anemic pregnant women to the total population of pregnant women for each region in West Java 2023 (in percent). WHO standard for anemia in pregnancy is diagnosed when the hemoglobin (Hb) level less than [38]. |
| Breastfeeding ($A_3$) | Percentage of exclusive breastfeeding among infants under six months in West Java 2023. |
| Drinking water ($A_4$) | Percentage of pregnant women who own access to safe drinking water to the total population of pregnant women for each region in West Java 2023. |
| Sanitation ($A_5$) | Percentage of families with access to proper sanitation in West Java 2023. |
| Poverty ($A_6$) | Poverty line in West Java 2023. The poverty line (as score) is the minimum expenditure required for a person consumption (food and non-food ) which is assumed not living in poverty [39]. |
| HDI ($A_7$) | Human Development Index (HDI) in West Java 2023. The index (as score) is calculated by taking four components, including life expectancy at birth, expected years of schooling, mean years of schooling and expenditure per capita [40]. |
| FSI ($A_8$) | Food Security Indeks (FSI) in West Java 2023. The index (as score) is calculated according to three dimensions covering aspects of food availability, affordability and utilization [41]. |

## 3. MAIN RESULTS

Based on data acquired from the SDI portal, the total prevalence of malnutrition in West Java, Indonesia, was highest during the COVID-19 pandemic, which took place around 2019-2020. The figure decreased steadily from 2021 to 2023, indicating that malnutrition was tackled in the framework of post-pandemic socio-economic and health recovery from the West Java regional government during this period.

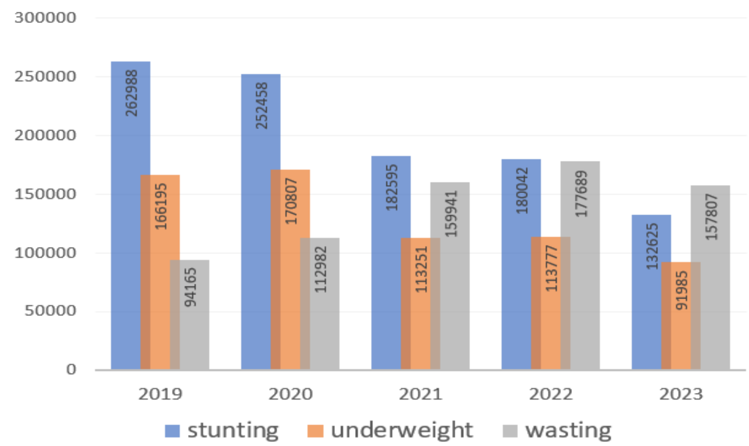



**Figure 2**. Distribution of total under-five malnutrition prevalence in West Java in the last five years

The "Jabar Zero New Stunting" program launched in 2022 by the West Java regional government successfully reduced the stunting prevalence from 262,988 cases in 2019 to 132,625 in 2023. The other two manifestations of malnutrition have also decreased, with a total prevalence of underweight and wasting in 2023 of 91,985 and 157,807 cases, respectively. However, these figures should not be neglected. Accelerated efforts to reduce malnutrition among under-five need to be pursued in a convergent, holistic, integrative, and quality manner through multi-sectoral cooperation from the center, regions, and urban-rural areas. For this reason, a spatial analysis of the malnutrition prevalence distribution in each district and city in West Java is needed.

### 3.1. Hidden Pattern of Under-five Malnutrition Prevalence Distribution

Spatial analysis was conducted descriptively to reveal and describe the hidden patterns of under-five malnutrition distribution in West Java, Indonesia, in 2023. Each region on the West Java map is assigned a color that represents its prevalence level, starting from the green zone for lower prevalence to the red zone for highest prevalence.

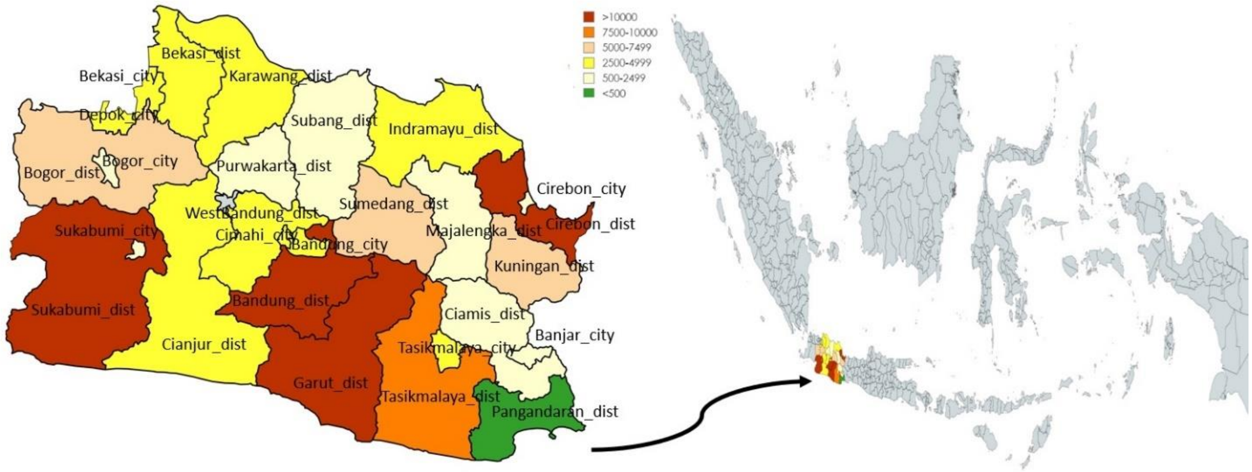



**Figure 3**. Stunting prevalence distribution among under-five children in West Java 2023

In Figure 3, four regions are in the red zone, including Sukabumi, Bandung, Garut, and Cirebon Districts, with stunting prevalence exceeding 10,000 cases. Pangandaran District has only one region in the green zone with stunting prevalence below 500 cases. In a nutshell, the figure shows the pattern of stunting distribution where neighboring regions tend to be in the same color zone. In contrast, the urban regions tend to be in the color zone with fewer stunting prevalence cases than the rural regions. For example, Bandung City lies in the yellow zone with a stunting prevalence of around 2,500-4,999 cases, less than Bandung District which falls in the red zone with a prevalence of over 10,000 cases.

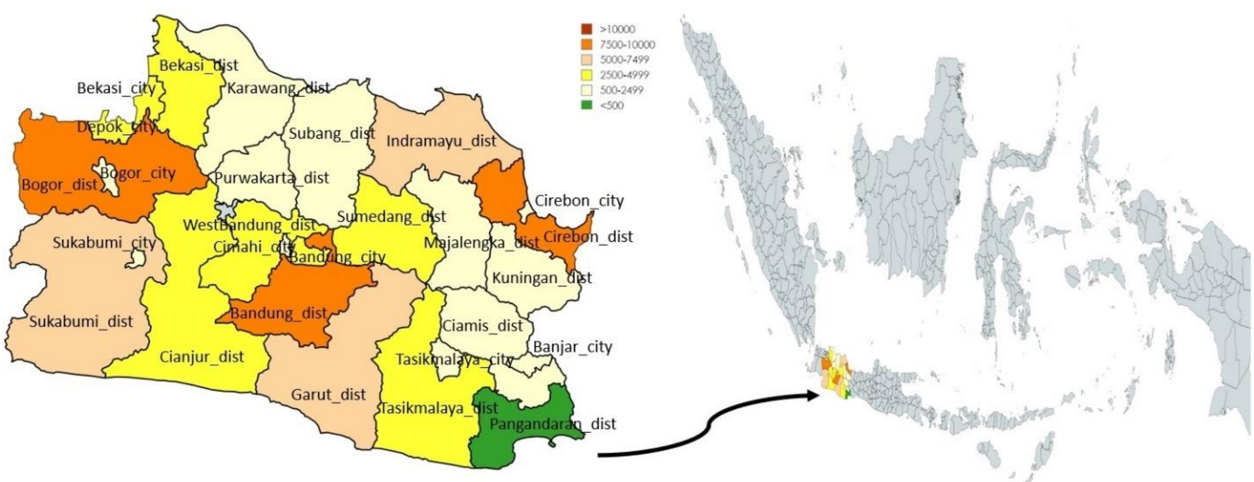



**Figure 4**. Underweight prevalence distribution among under-five children in West Java 2023

The fact that none of the regions are in the red zone indicates that the distribution of underweight prevalence in the districts in West Java is less than 10,000 cases (in Figure 4). Pangandaran District stands as the sole region in the green zone whose underweight prevalence cases are less than 500. Underweight has a similar distribution pattern to the stunting prevalence distribution, where nearby regions tend to be under the same color zone. However, the urban regions tended to be in the color zone with slightly fewer underweight cases compared to the rural regions. For example, Cirebon City is located in the light-yellow zone with its underweight prevalence around 500-2,499 cases, less than Cirebon District which belongs to the orange zone with its prevalence around 7,500-10,000 cases.

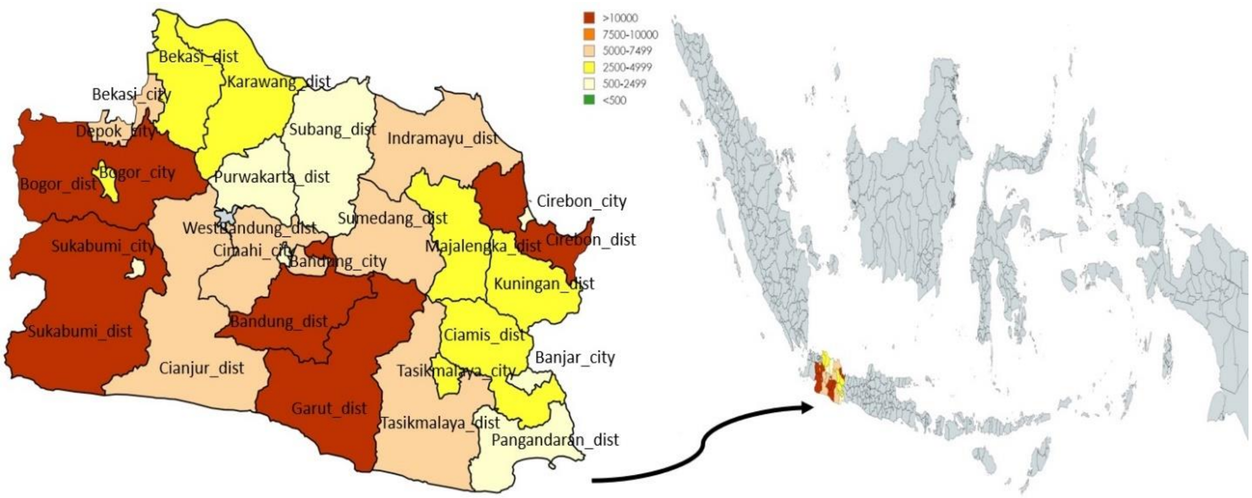
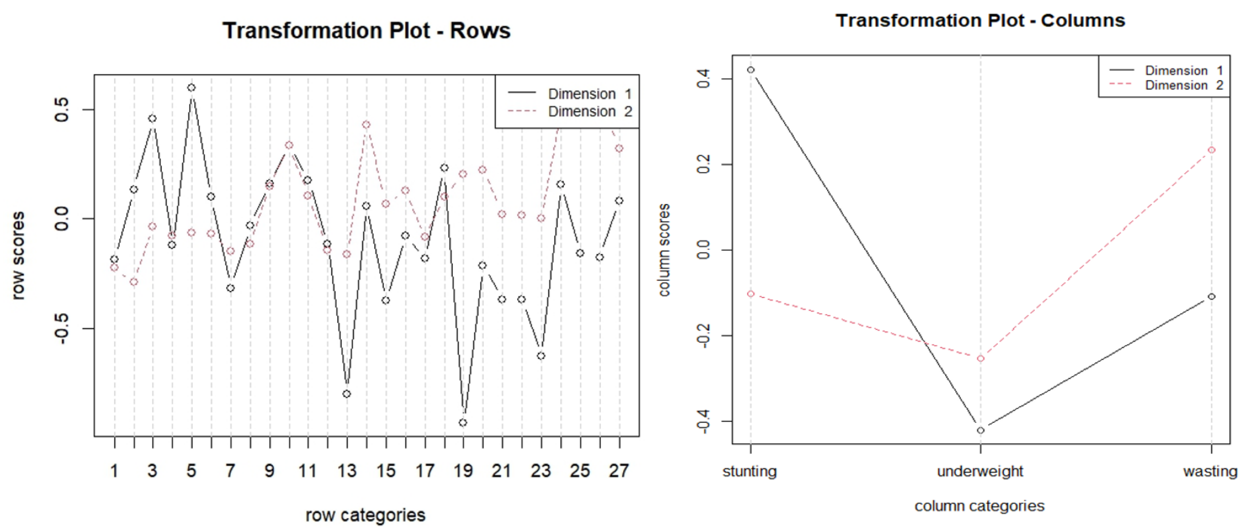



**Figure 5**. Wasting prevalence distribution among under-five children in West Java 2023

Based on Figure 5, the distribution of wasting prevalence has more red zones than the stunting prevalence distribution. Five districts are within the wasting red zone. The four districts in the stunting red zone are also within the wasting red zone and adding Bogor District, which borders Sukabumi District in the red zone. Pangandaran district is no longer in the green zone, indicating that the wasting prevalence in this region is greater than the stunting and underweight prevalence. Since none of the regions are in the green zone, the wasting prevalence in West Java is relatively greater than the other two manifestations of malnutrition. Generally, the bordering regions tend to be in the same color zone.

### 3.2. Under-five Malnutrition Prevalence from CCA's Perspective

Considering Table 4 (available in supplementary data), "Region" is established as the row categorical variable $(\mathbf{X})$ and "Malnutrition" as the column categorical variable $(\mathbf{Y})$, leading to the $27 \times 3$ bivariate frequency matrix $\mathbf{N}$ and the $27 \times 8$ row covariate matrix $(\mathbf{A})$ representing the determinant factors consisting of nutritional intake factors (CED, Anemia, Breastfeeding), environmental health factors (Sanitation, Drinking water), and socio-economic factors (Poverty, HDI, FSI). The chi-square test concluded that regions are statistically significantly associated with malnutrition prevalence and its determinants at a 5% significance level, since $\chi^2_{stat} = 24{,}282$ with 260 degrees of freedom and $p - value = 2.2 \times 10^{-16}$.




(a) (b)

**Figure 6**. Transformation plot from row categories to row scores (a) and from column categories to column scores (b)

The transformation plot in Figure 6 plots the original category values against the corresponding quantification (score) to reveal trends that might not be captured in the CCA plot. This plot displays the row (or column) category values on the $x$-axis into row (or column) scores in the dimensions of the CCA plot on the $y$-axis [42]. As an illustration, the stunting category in right-hand plot (b) has a column score value around 0.4 in the first dimension and about -0.1 in the second dimension. It implies the stunting category will be drawn at the Quadrant IV on a CCA plot, since it has a positive $x$-coordinate and negative $y$-coordinate.

The row transformation plot (a) has interesting patterns. Row categories that have similar row score patterns in Dim1-Dim2 will be plotted close together on the CCA plot, such as the pair $X_{21}$ (Bandung City) and $X_{22}$ (Cirebon City), $X_9$ (Cirebon District) and $X_{11}$ (Sumedang District). On the other side, $X_{21}$ (Bandung City), $X_{22}$ (Cirebon City), and $X_{23}$ (Bekasi City), whose row scores approximately zero on Dim2, will be plotted close to the $x$-axis indicating that they contribute mainly to the first axis. Instead, $X_{14}$ (Purwakarta District) with a row score close to zero on Dim1 will be drawn closely to the $y$-axis, indicating such a category contributes more strongly to the second axis [43]. If a category has a row score of approximately zero in both dimensions, it will be plotted in proximity to the origin $(0,0)$. It implies that the contribution of the corresponding category to the association structure between variables can be neglected [44].

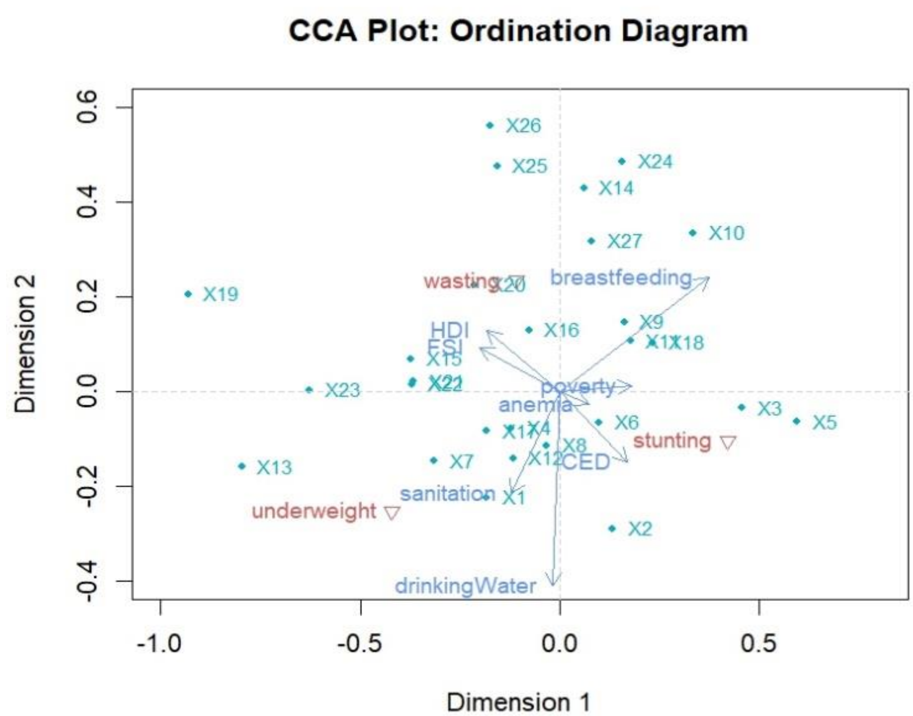



**Figure 7**. The CCA plot captured hidden patterns of association between the variables of interest

The plot in Figure 7 is obtained by applying Algorithm 1 to the malnutrition dataset (Table 4). This figure reveals that stunting prevalence is more closely associated with nutrition intake factors, especially CED and anemia in pregnant women, than other determinants. Pregnant women who fail to meet their nutritional needs could be exposed to CED, which harms maternal health and increases the risk of anemia [37]. In another case, CED might cause disrupted fetal growth, increasing the maternity risk, miscarriage, premature birth, postpartum hemorrhage, congenital anomalies, and babies born with low birth weight, which all lead to an increase in the risk of stunting [45]. Normal pregnant women require an additional 180-300 kcal of energy and 30 grams of protein per day to gain 0.5 kg/week [38]. CED pregnant women, need an additional energy intake of 500 kcal/day from their daily energy intake, where less than 25% of the energy in supplementary food comes from protein [46].

Sukabumi District $(X_2)$, Cianjur District $(X_3)$, Garut District $(X_5)$, and Tasikmalaya District $(X_6)$ are the regions associated with these variables. Another interesting point is the fact that Sukabumi District geographically borders Cianjur District. In contrast, Garut District abuts Tasikmalaya District, implying there exists a demographic similarity among the regions in terms of the association between stunting prevalence and nutrition intake and spatial factors. However, the total prevalence of stunting, percentage of pregnant women with CED and anemia in respective districts merits more attention from the West Java regional government by provision of blood folate and folic acid tablets to pregnant women.

In the same way, it was revealed that “underweight” tends to be closer to healthy environmental factor such as “sanitation” and “drinking water”. Lack of access to safe drinking water and poor sanitation pose health risks, including parasitic diseases such as diarrhea or intestinal infections, which are the leading causes of being underweight [47], [48]. Regions closely associated with these variables would be Bogor District $(X_1)$, Bandung District $(X_4)$, Ciamis District $(X_7)$, Kuningan District $(X_8)$, Sumedang District $(X_{11})$, and Indramayu District $(X_{12})$. According to Figure 4, Bogor District and Bandung District are in the underweight orange zone with the highest frequency of underweight prevalence. Geographically, Sumedang District borders Indramayu District and Ciamis District adjacent Kuningan District. It indicates that there might be potential demographic similarity among the regions in terms of the association between wasting prevalence, healthy environmental, and spatial factors. Thus, local governments could design some mechanisms to reduce underweight prevalence through improved access to safe drinking water and proper sanitation, reduced exposure to waterborne pathogens, along enhance health and hygiene environmental holistically.

The other patterns reveal that the wasting is closely associated with socio-economic factors such as HDI and FSI. In spatial terms, the regions associated with these variables include West Bandung District $(X_{15})$, Pangandaran District $(X_{16})$, Sukabumi City $(X_{20})$, Bandung City $(X_{21})$, and Cirebon City $(X_{22})$. As exhibited in Figure 5, West Bandung District ia a neighbor to Bandung City, but the other regions are neither close together nor fall within the same wasting color zone. It indicates that wasting prevalence is not associated with spatial factors. Nonetheless, an interesting pattern is revealed by the fact that these regions are the leading domestic tourist destinations in West Java. It suggests that high regional income does not imply adequate socio-economic welfare, as reflected in the HDI and FSI scores. Thus, the West Java regional government should conduct evaluations and audits related to equalizing the community's social welfare.

Taken together, the vectors representing “stunting”, “underweight”, and “wasting” heads in opposite directions, indicating that these different manifestations of malnutrition are mutually independent. It affirms that there is no under-five is identified in more than one case. In particular, the “poverty” vector narrowly coinciding with the $x$-axis indicates that the poverty line contributes to malnutrition prevalence and other determinant factors in almost equal proportions. Therefore, poverty is solid major determinant of maternal malnutrition, poor childcare, household food insecurity, and unhealthy environment [13]. For the rest, “breastfeeding” has a separate pattern from the other variables, indicating that the exclusive breastfeeding factor should be analyzed differently.

### 3.3. Under-five Malnutrition from Predictive Clustering’s Perspective

As a first step in predictive clustering, the optimal number of clusters is determined by the WSS (within the sum of squares) method. The plot in Figure 8 exhibits the within-group variance, which decreases as $k$ increases. However, there is an elbow at $k = 3$. This elbow indicates that the additional clusters beyond the third cluster have small values, making $k = 3$ the optimal cluster for partitioning the given dataset.

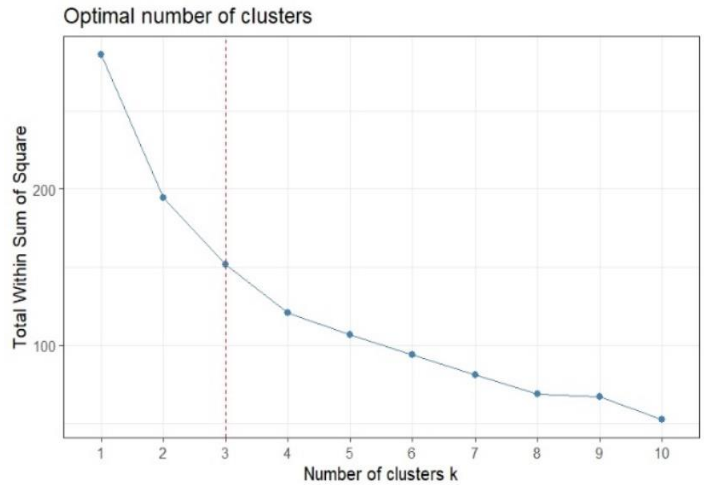


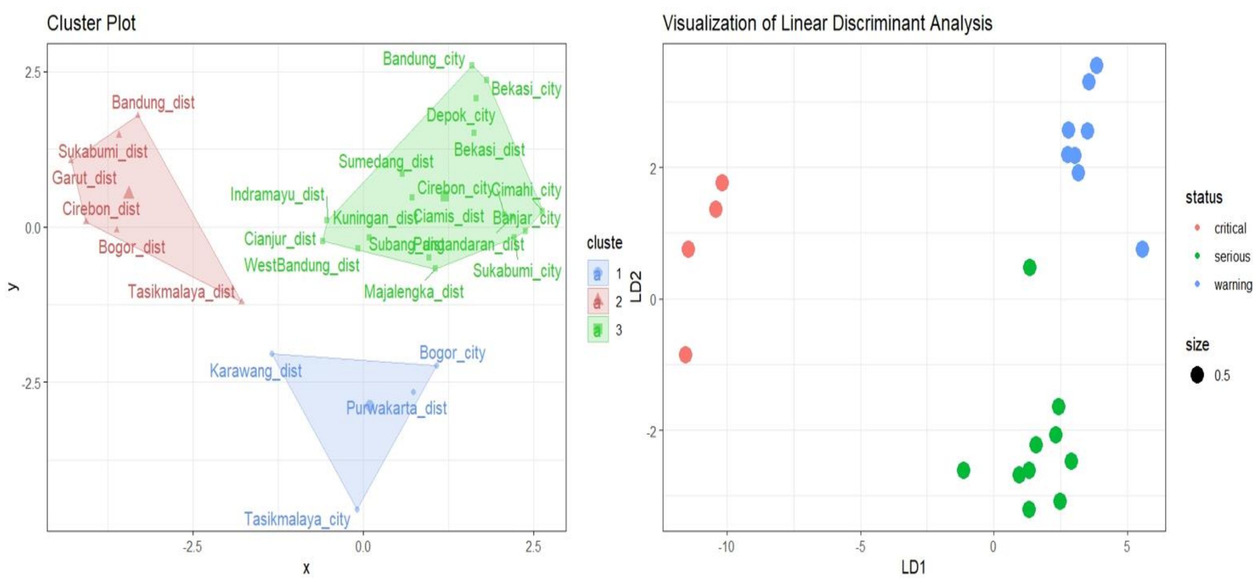

**Figure 8**. Evaluate the quality of clusters by the WSS method

The centroid of each cluster is summarized in Table 2. Each centroid is arranged from the average value of each attribute variable.

**Table 2**. Cluster centroid for malnutrition dataset (in standardized values)

| Cluster | Y1 | Y2 | Y3 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.59 | -0.67 | -0.60 | -0.30 | 0.42 | -0.55 | -1.75 | -1.54 | 0.17 | 0.20 | -0.15 |
| 2 | 1.59 | 1.51 | 1.58 | 1.55 | 1.02 | -0.05 | 0.01 | 0.07 | 0.18 | -0.66 | -0.81 |
| 3 | -0.42 | -0.37 | -0.42 | -0.48 | -0.46 | 0.15 | 0.41 | 0.34 | -0.10 | 0.19 | 0.32 |

The clusters resulting from the k-means clustering procedure (Algorithm 2) are exhibited in Figure 9 (a). The clusters that have been generated are interpreted by comparing their centroids.




(a) (b)

**Figure 9**. The predictive clustering result by k-means clustering (a) and LDA (b)

By considering the centroid value of each status, we define the three malnutrition emergency statuses as 1= “warning”, 2= “critical”, and 3= “serious”. This definition is derived from interpreting of each variable attribute distribution in each cluster, as shown in Figure 10.

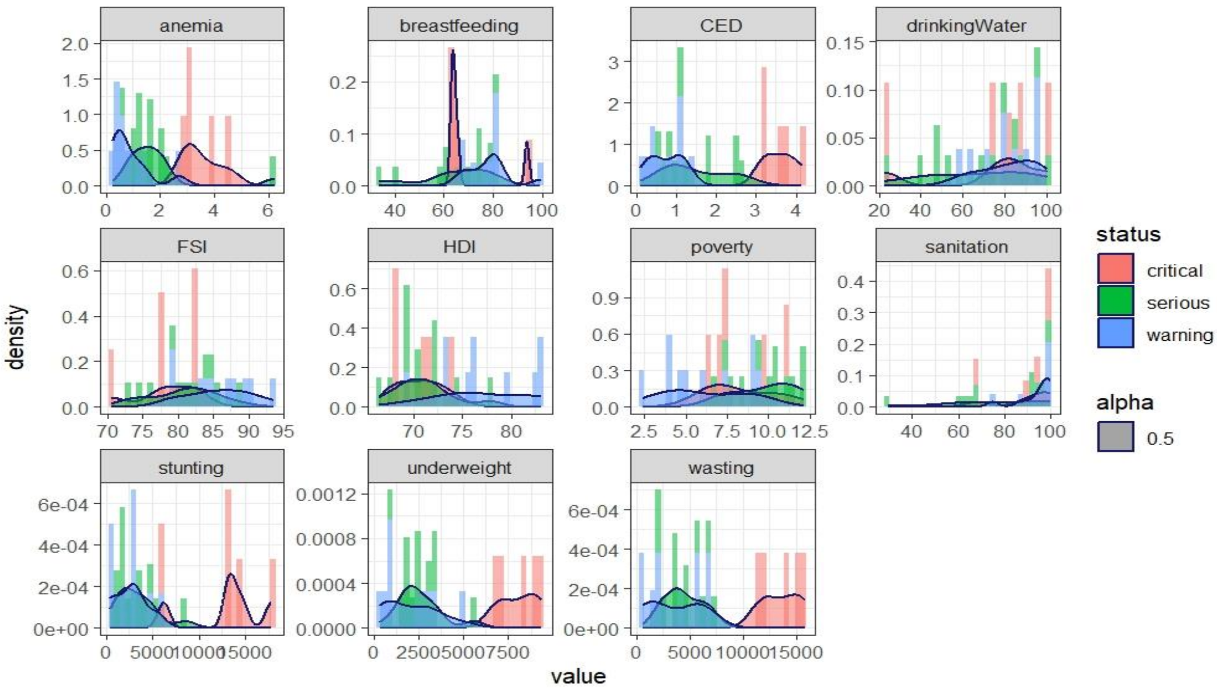



**Figure 10**. Attribute variables distribution by under-five malnutrition emergency status

Based on Figure 10, Cluster 2 is defined as “critical” in terms of under-five malnutrition emergency status by the fact that it is characterized by a high prevalence of malnutrition due to stunting, underweight, and wasting. Other characteristics include a high percentage of anemic and CED pregnant women, a relatively lower percentage of exclusive breastfeeding, low FSI and HDI scores, and a high poverty line. Cluster 1 is defined as “serious” as it is characterized by a moderate prevalence of malnutrition and other attribute variables that also tend to be relatively moderate. Cluster 3 is designated as a “warning” due to its characteristics of a slightly lower malnutrition prevalence, a low percentage of anemic pregnant women and CED, a medium-high percentage of exclusive breastfeeding, high FSI and HDI scores, and a low poverty line. The distribution of safe drinking water and proper sanitation in the three clusters tends to take a similar shape.

In the remaining steps, a linear discriminant function is sought to predict the under-five malnutrition emergency status of a certain region based on the features of the given attribute variables. Based on Eq. (22) and by employing Algorithm 3, the following first two linear discriminant functions are obtained

$$LD_1 = -2.19Y_1 - 2.18Y_2 + 1.91Y_3 - 0.76A_1 - 1.18A_2 + 0.12A_3 - 1.15A_4 + 2.14A_5 + 1.94A_6 + 1.48A_7 + 0.99A_8 \ and$$

$$LD_2 = 3.55Y_1 + 0.94Y_2 - 1.87Y_3 - 0.54A_1 - 2.018A_2 + 0.61A_3 + 1.43A_4 - 0.37A_5 - 0.74A_6 + 0.76A_7 - 0.31A_8.$$

The percentage of separation archived by each discriminant function is 85.39% and 14.61%, respectively. Furthermore, Figure 9 (b) illustrates that $LD_1$ on the $x$-axis and $LD_2$ on the $y$-axis fairly well separate the three malnutrition emergencies since no overlapping points exist. Evaluating the test data confirmed that 83.3% of the discriminant models $LD_1$ and $LD_2$ are accurate in predicting the emergency status of malnutrition in a region. Later, the model will be used to predict the under-five malnutrition emergency status of several regions around West Java with the given attribute variable values as in Table 3. Prediction outcomes are tabled in the last column.

**Table 3**. Cluster prediction for malnutrition emergency status by using the LDA model

| District/ City | Y1 | Y2 | Y3 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | Cluster Prediction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Jakarta District | 2837 | 937 | 565 | 2.05 | 6.81 | 75.15 | 92.35 | 98.35 | 4.90 | 82.9 | 83.56 | Warning |
| Thousand Islands District | 320 | 243 | 68 | 0.91 | 1.60 | 72.92 | 85.66 | 99.52 | 14.11 | 75.34 | 80.27 | Warning |
| Surakarta City | 950 | 1639 | 1017 | 3.61 | 9.78 | 83.96 | 94.57 | 97.2 | 9.03 | 82.21 | 80.27 | Serious |

Defining malnutrition emergency status is intended to provide direction regarding the priority level in handling under-five malnutrition in certain regions. Crisis status means that urgent, unexpected, and possibly dangerous malnutrition manifestations pose a critical risk to health and life, requiring immediate actions. Serious status is recognized as a condition where persistent malnutrition manifestation continues and has the potential to disrupt life and health, hence requiring serious action. Warning status is considered a condition of malnutrition manifestation that signals a warning of an increase in the prevalence frequency along with attribute variable values that start to deviate from expected conditions, thus requiring preventive and precautionary actions.

## 4. Conclusion

This study highlights two conclusions (1) The canonical correspondence analysis perspective reveals that there exists a statistically significant association between the prevalence of malnutrition and several determinant factors including nutritional intake, environmental health, socio-economic, and spatial factors. The stunting prevalence is more closely associated with nutritional intakes, such as CED and anemia; underweight tends to be closer to environmental health factors, such as sanitation and drinking water; and the wasting prevalence is relatively closely associated with socio-economic factors such as HDI and FSI; (2) from the perspective of predictive clustering, three clusters are identified to represent the under-five malnutrition emergency status (critical, serious, warning) and a model has been constructed with an accuracy rate of 83.3% in predicting a certain region status with given attribute variable values. The designation of under-five malnutrition emergency status is desired to provide direction for the regional government concerned regarding the priority level in handling malnutrition in certain regions.

## Acknowledgment

The authors would like to thank LPPM Universitas Singaperbangsa Karawang for the research funding through the “Competitive Grant-Fundamental Research 2024” scheme. We also thanks Dian Pebriana, Doni Ramdan, and Julia Permata as the research assistant team for their technical contributions.

## Conflict of Interests

The authors declare that there is no conflict of interests.

## SUPPLEMENTARY DATA

**Table 4.** Malnutrition datset of West Java-Indonesia 2023

| X | Y1 | Y2 | Y3 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bogor_dist | 6231 | 9332 | 12066 | 3,80 | 3,86 | 63,76 | 68,41 | 87,5 | 7,27 | 71,78 | 70,57 |
| Sukabumi_dist | 14671 | 6871 | 11690 | 4,15 | 2,99 | 63,58 | 100 | 100 | 7,01 | 68,49 | 77,65 |
| Cianjur_dist | 4893 | 3300 | 6268 | 2,57 | 1,01 | 81,65 | 68,71 | 100 | 10,22 | 66,55 | 78,99 |
| Bandung_dist | 13321 | 8893 | 15149 | 3,21 | 2,96 | 63,61 | 99,74 | 74,6 | 6,4 | 73,74 | 82,48 |
| Garut_dist | 17805 | 7147 | 13872 | 3,21 | 4,58 | 93,61 | 88,94 | 81,5 | 9,77 | 68,11 | 78,04 |
| Tasikmalaya_dist | 8584 | 3527 | 7333 | 2,61 | 2,08 | 69,5 | 64,47 | 84,6 | 10,28 | 67,76 | 75,68 |
| Ciamis_dist | 2020 | 2498 | 3747 | 1,80 | 1,71 | 58,37 | 92,94 | 95,4 | 7,42 | 72,05 | 82,36 |
| Kuningan_dist | 5237 | 2299 | 4482 | 1,28 | 2,23 | 73,86 | 94,62 | 80,3 | 12,12 | 70,72 | 80,95 |
| Cirebon_dist | 13353 | 8358 | 15799 | 3,65 | 3,19 | 65,57 | 95,59 | 23,9 | 11,2 | 70,95 | 82,03 |
| Majalengka_dist | 1924 | 1930 | 3702 | 0,69 | 0,70 | 79,15 | 98,72 | 53,7 | 11,21 | 69,13 | 81,39 |
| Sumedang_dist | 5791 | 2665 | 6748 | 1,35 | 0,82 | 81,27 | 98,62 | 79 | 9,36 | 73,18 | 84,88 |
| Indramayu_dist | 3440 | 5564 | 6597 | 1,87 | 1,68 | 60,88 | 96 | 74,3 | 12,13 | 69,25 | 84,52 |
| Subang_dist | 1667 | 1894 | 2343 | 1,10 | 2,01 | 33,6 | 100 | 68,4 | 9,52 | 70,7 | 85,31 |
| Purwakarta_dist | 984 | 854 | 2066 | 1,05 | 1,28 | 74,87 | 63,14 | 48,6 | 8,46 | 72,09 | 80,42 |
| Karawang_dist | 2696 | 2373 | 4450 | 2,56 | 6,23 | 61,4 | 59,81 | 39,1 | 7,87 | 72,35 | 88,48 |
| Bekasi_dist | 2997 | 3148 | 4387 | 1,06 | 0,65 | 75,57 | 92,28 | 95,5 | 4,93 | 75,76 | 88,05 |
| WestBandung_dist | 4323 | 3246 | 5761 | 0,93 | 1,59 | 64,46 | 92,82 | 88,2 | 10,52 | 69,61 | 74,19 |
| Pangandaran_dist | 425 | 340 | 638 | 0,41 | 0,60 | 80,51 | 100 | 96,6 | 8,98 | 69,38 | 83,57 |
| Bogor_city | 1369 | 1581 | 2622 | 0,81 | 1,20 | 41,59 | 67,45 | 47,4 | 6,67 | 77,85 | 83,78 |
| Sukabumi_city | 614 | 961 | 1431 | 0,46 | 0,32 | 71,35 | 95,53 | 72,2 | 7,5 | 76,32 | 79,08 |
| Bandung_city | 4706 | 3390 | 6794 | 1,12 | 1,39 | 79,95 | 100 | 96,4 | 3,96 | 83,04 | 90,46 |
| Cirebon_city | 2043 | 1074 | 1990 | 0,58 | 0,40 | 66,44 | 99,18 | 80,5 | 9,16 | 76,46 | 83,77 |
| Bekasi_city | 2987 | 4859 | 5490 | 0,89 | 1,25 | 68,03 | 96,29 | 90,7 | 4,1 | 83,03 | 93,55 |
| Depok_city | 3283 | 2543 | 5494 | 1,10 | 2,72 | 98,91 | 99,09 | 59,31 | 2,38 | 82,38 | 89,47 |
| Cimahi_city | 2890 | 869 | 2278 | 0,30 | 0,26 | 79,86 | 74,96 | 63,6 | 4,66 | 79,46 | 87,17 |
| Tasikmalaya_city | 3595 | 1786 | 3903 | 0,55 | 0,98 | 74,75 | 29,66 | 22,9 | 11,53 | 74,47 | 72,71 |
| Banjar_city | 776 | 683 | 707 | 0,10 | 0,45 | 80,93 | 86,95 | 86,4 | 6,14 | 73,08 | 79,55 |

## REFERENCES

[1] World Health Organization, Malnutrition, https://www.who.int/health-topics/malnutrition.

[2] J.E. Morley, W.A. Van Staveren, Undernutrition: diagnosis, causes, consequences and treatment, in: Food for the Ageing Population, Elsevier, 2009: pp. 153–168. https://doi.org/10.1533/9781845695484.2.153.

[3] J.M. Lecerf, F. Andreelli, La dénutrition, Méd. Mal. Métab. 16 (2022), 398–399. https://doi.org/10.1016/j.mmm.2022.05.003.

[4] S. Das, Z. Hossain, M.K. Nesa, Levels and trends in child malnutrition in Bangladesh, Asia-Pac. Popul. J. 24 (2009), 51–78. https://doi.org/10.18356/6ef1e09a-en.

[5] L. Taylor, Urgent action is needed to tackle malnutrition crisis in 30 million children, says WHO, BMJ. 380 (2023), p99. https://doi.org/10.1136/bmj.p99.

[6] H. Andriani, E. Friska, M. Arsyi, et al. A multilevel analysis of the triple burden of malnutrition in Indonesia: trends and determinants from repeated cross-sectional surveys, BMC Public Health 23 (2023), 1836. https://doi.org/10.1186/s12889-023-16728-y.

[7] UNICEF, Kementerian Kesehatan Republik Indonesia, Menuju masa depan indonesia bebas masalah kekurangan gizi: Mengatasi wasting dan menurunkan prevalensi stunting, pp. 1–12, 2023. www.unicef.or.id.

[8] UNICEF, MOH, Mid-upper arm circumference ( MUAC ) tapes: A simple tool to detect child wasting and save lives in children aged between 6 months and 5 years old, pp. 1–12, 2023. www.unicef.or.id.

[9] BPS Statistics Indonesia, Indonesia stunting special index report 2021-2022, 2023.

[10] A.P. Campos, M. Vilar-Compte, S.S. Hawkins, Association between breastfeeding and child overweight in Mexico, Food Nutrition Bull. 42 (2021), 414–426. https://doi.org/10.1177/03795721211014778.

[11] UNICEF, WHO, World Bank Group, Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-world bank joint malnutrition estimates, 2021.

[12] Global Action Plan on Child Wasting, Global action plan on child wasting, https://www.childwasting.org.

[13] WHO, Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors, 2004. https://www.who.int/publications/i/item/9241580313.

[14] D.E. Kpewou, E. Poirot, J. Berger, et al. Maternal mid-upper arm circumference during pregnancy and linear growth among Cambodian infants during the first months of life, Maternal Child Nutr. 16 (2020), e12951. https://doi.org/10.1111/mcn.12951.

[15] O. A. Uthman, O. Aremu, Malnutrition among women in sub-Saharan Africa: Rural-urban disparity, Rural Remote Health, 8 (2008), 1–11.

[16] G. Kumera, D. Gedle, A. Alebel, F. Feyera, S. Eshetie, Undernutrition and its association with socio-demographic, anemia and intestinal parasitic infection among pregnant women attending antenatal care at the University of Gondar Hospital, Northwest Ethiopia, Maternal Health Neonatol. Perinatol. 4 (2018), 18. https://doi.org/10.1186/s40748-018-0087-z.

[17] D. Bühler, R. Hartje, U. Grote, Matching food security and malnutrition indicators: evidence from Southeast Asia, Agric. Econ. 49 (2018), 481–495. https://doi.org/10.1111/agec.12430.

[18] F. Yanuar, A.S. Deva, Rudiyanto, et al. Bayesian regularized tobit quantile to construct stunting rate model, Commun. Math. Biol. Neurosci. 2023 (2023), 108. https://doi.org/10.28919/cmbn/7976.

[19] D.J. Fitri, A. Djuraidah, H. Wijayanto, Bayesian conditional negative binomial autoregressive model: a case study of stunting on Java Island in 2021, Commun. Math. Biol. Neurosci. 2024 (2024), 18. https://doi.org/10.28919/cmbn/8281.

[20] V.A.D. Djara, Y. Andriyana, L. Noviyanti, Modelling the prevalence of stunting toddlers using spatial autoregressive with instrument variable and S-estimator, Commun. Math. Biol. Neurosci. 2022 (2022), 29. https://doi.org/10.28919/cmbn/7234..

[21] A. Braghetta, Drawing the connection between Malnutrition and lack of safe drinking water in Guatemala, J. AWWA 98 (2006), 97–106. https://doi.org/10.1002/j.1551-8833.2006.tb07663.x.

[22] F.G. Hasanain, M. Jamsiah, M.I. Zaleha, et al. Association between drinking water sources and diarrhea with malnutrition among kindergarten's children in Baghdad city, Iraq, Malays. J. Public Health Med. 12(2012), 45-48.

[23] C.J.F. Ter Braak, Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis, Ecology 67 (1986), 1167–1179. https://doi.org/10.2307/1938672.

[24] K.E. Lestari, U.S. Pasaribu, S.W. Indratno, H. Garminia, The comparative analysis of dependence for three-way contingency table using Burt matrix and Tucker3 in correspondence analysis, J. Phys.: Conf. Ser. 1245 (2019), 012056. https://doi.org/10.1088/1742-6596/1245/1/012056.

[25] K.E. Lestari, U.S. Pasaribu, S.W. Indratno, H. Garminia, Generating roots of cubic polynomials by Cardano's approach on correspondence analysis, Heliyon 6 (2020), e03998. https://doi.org/10.1016/j.heliyon.2020.e03998.

[26] J. de Leeuw, P. Mair, Simple and canonical correspondence analysis using the R package anacor, J. Stat. Softw. 31 (2009), 1–18.

[27] L. Doey, J. Kurta, Correspondence analysis applied to psychological research, Tutor. Quant. Methods Psychol. 7 (2011), 5–14. https://doi.org/10.20982/tqmp.07.1.p005.

[28] K.E. Lestari, M.R. Utami, M.R. Yudhanegara, Exploratory analysis on adaptive reasoning of undergraduate student in statistical inference, Int. J. Instruction 15 (2022), 535-554.

[29] K.E. Lestari, M.R. Utami, M.R. Yudhanegara, Sequential exploratory design by performing correspondence analysis to investigate procedural fluency of undergraduate student, AIP Conf. Proc. 2588 (2023), 050004. https://doi.org/10.1063/5.0111974.

[30] K.E. Lestari, U.S. Pasaribu, S.W. Indratno, H. Garminia, The reliability of crash car protection level based on the circle confidence region on the correspondence plot, IOP Conf. Ser.: Mater. Sci. Eng. 598 (2019), 012061. https://doi.org/10.1088/1757-899X/598/1/012061.

[31] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297, Oakland, 1967.

[32] M.R. Yudhanegara, S.W. Indratno, R.K.N. Sari, Clustering for item delivery using rule-k-means, J. Indones. Math. Soc. 26 (2020), 185–191. https://doi.org/10.22342/jims.26.2.871.185-191.

[33] K.E. Lestari, M.R. Yudhanegara, Empirical study of mathematical investigation skill on graph theory, Math. Teach. Res. J. 16 (2024), 4–27.

[34] G.J. McLachlan, Discriminant analysis and statistical pattern recognition, Wiley, Hoboken, 2004.

[35] A.C. Rencher, Methods of multivariate analysis, Wiley, 2002. https://doi.org/10.1002/0471271357.

[36] Z. Wang, Y.-H. Shao, L. Bai, et al. MBLDA: A novel multiple between-class linear discriminant analysis, Inf. Sci. 369 (2016), 199–220. https://doi.org/10.1016/j.ins.2016.05.031.

[37] E.K. Wati, R. Murwani, M.I. Kartasurya, S. Sulistiyani, Determinants of chronic energy deficiency (CED) incidence in pregnant women: A cross-sectional study in Banyumas, Indonesia, Narra J. 4 (2024), e742. https://doi.org/10.52225/narra.v4i1.742.

[38] G. Stephen, M. Mgongo, T.H. Hashim, et al. Anaemia in Pregnancy: Prevalence, Risk Factors, and Adverse Perinatal Outcomes in Northern Tanzania, Anemia 2018 (2018), 1846280. https://doi.org/10.1155/2018/1846280.

[39] BPS-Statistics Indonesia, Indonesia poverty profile in March 2023, 2023. https://www.bps.go.id.

[40] V. Verawaty, M. Gunarto, R. Wahasusmiah, C.I. Merina, Determinants of human development index in Indonesia, in: Proceedings of the International Conference on Industrial Engineering and Operations Management, IEOM Society International, Singapore, Singapore, 2021. https://doi.org/10.46254/AN11.20210750.

[41] R. Virtriana, A. Riqqi, T.S. Anggraini, et al. Development of spatial model for food security prediction using remote sensing data in West Java, Indonesia, ISPRS Int. J. Geo-Inf. 11 (2022), 284. https://doi.org/10.3390/ijgi11050284..

[42] K.E. Lestari, U.S. Pasaribu, S.W. Indratno, Graphical depiction of three-way association in contingency table using higher-order singular value decomposition Tucker3, J. Phys.: Conf. Ser. 1280 (2019), 022035. https://doi.org/10.1088/1742-6596/1280/2/022035.

[43] K.E. Lestari, M.R. Utami, M.R. Yudhanegara, The contribution biplot on correspondence analysis to investigate the floricultural crops production in West Java, AIP Conf. Proc. 2867 (2024), 020007. https://doi.org/10.1063/5.0224521.

[44] Y.S. Afrianti, U.S. Pasaribu, H. Ardy, et al. Elliptic confidence on correspondence analysis stainless steels corrosion problem, Eng. Lett. 32 (2024), 1332–1339.

[45] P. Pandey, P. Bajpai, S. Jain, et al. Maternal empowerment holds the key to reducing stunting during first 1000 days of life: Evidence from a case-controlled study, Ann. Trop. Med. Public Health, 10 (2017), 667–677.

[46] M.A. Kominiarek, P. Rajan, Nutrition recommendations in pregnancy and lactation, Med. Clin. North Amer. 100 (2016), 1199–1215. https://doi.org/10.1016/j.mcna.2016.06.004.

[47] F. Murtaza, M. Muzaffar, T. Mustafa, et al. Water and sanitation risk exposure in children under-five in Pakistan, J. Family Community Med. 28 (2021), 103. https://doi.org/10.4103/jfcm.jfcm_149_21.

[48] M. Nadeem, M. Anwar, S. Adil, et al. The association between water, sanitation, hygiene, and child underweight in Punjab, Pakistan: An application of population attributable fraction, J. Multidiscip. Healthc. 17 (2024), 2475–2487. https://doi.org/10.2147/JMDH.S461986.