OPTIMIZING CORRESPONDENCE ANALYSIS WITH SINGULAR VALUE DECOMPOSITION: A STUNTING DATA CASE STUDY

U. S. PASARIBU[1*], N. A. SARI[3†], N. F. SA'IDAH[3], S. B. HARMADI[4]

[1]Departement of Mathematics, Institut Teknologi Bandung, Bandung, Indonesia
[2]Bachelor of Mathematics Study Program, Institut Teknologi Bandung, Bandung, Indonesia
[3]Departement of Industrial Engineering, Institut Teknologi Bandung, Cirebon, Indonesia
[4]Departement of Development Study, Institut Teknologi Sepuluh November, Surabaya, Indonesia
[1]udjianna@itb.ac.id; 210120076@mahasiswa.itb.ac.id;
[3]nurfaizatus@itb.ac.id;[4]sonny.harmadi@its.ac.id;

ABSTRACT

Correspondence Analysis (CA) is a statistical technique used to map relationships between qualitative variables. It visualizes data in a low-dimensional space, enabling the interpretation of complex relationships. This study addresses the challenge of visualizing contingency tables with more than three categories using Singular Value Decomposition (SVD) for dimensionality reduction. We apply this approach to stunting data collected by the Indonesian Population Coalition in 2023, focusing on variables such as the district of residence, fever management methods, educational level of caregivers, and sources of information on stunting. The analysis reveals significant associations among these variables, providing insights that could inform public health strategies. This work underscores the utility of CA and SVD in handling high-dimensional qualitative data, particularly in health-related studies.

Keywords: Correspondence Analysis, Singular Value Decomposition, Stunting, Dimensionality Reduction, Public Health

1 INTRODUCTION

Correspondence Analysis (CA) is a multivariate analysis method that studies the relationships between two or more qualitative random variables. CA is used to convert categorical data into principal coordinates, which are then used to construct low-dimensional maps (Tuffery, 2018). The input for CA is a contingency table, also known as a cross-tabulation table. A contingency table illustrates the frequency distribution of two or more categorical variables. Each cell in the contingency table results from counting the number of combinations of row variables (v.b) and column variables (v.k) observed. The matrix where each element represents the frequency from the contingency table is called the cross-tabulation matrix, usually denoted by matrix $\mathbf{N} = \left(n_{ij}\right)$ of size $I \times J$ with $I, J \geq 2$. CA calculations often use the correspondence matrix, where each element is the frequency relative to the total number of observed

---

* Corresponding Author

† Corresponding Author

objects/individuals. Generally, the correspondence matrix is denoted by matrix $\mathbf{P} = (p_{ij}) = \left(\frac{n_{ij}}{n_{..}}\right)$ of size $I \times J$.

A fundamental concept of CA is the row profile and column profile of the cross-tabulation matrix $\mathbf{N}$, Tuffery [18]. Let the row profile and column profile matrices of size $I \times J$ be denoted as $\boldsymbol{\mathcal{P}}_{\mathbf{row}}$ and $\boldsymbol{\mathcal{P}}_{\mathbf{col}}$, respectively. The elements of the row profile and column profile are the relative frequencies with respect to the number of objects/individuals. The columns of the row profile and column profile represent the position vectors of the observed categories, which can be visualized in a plot. However, this is only applicable when the contingency table has no more than three categories. If there are more than three variables, the row and column profile vectors cannot be directly visualized, necessitating a dimensionality reduction process such as Singular Value Decomposition (SVD). This decomposition is performed on the standardized residual matrix $\mathbf{Z}$, which represents the association between row variables and column variables.

The case study uses stunting data from the Indonesian Population Coalition in 2023. Stunting is a developmental disorder in children caused by a lack of nutrition. In 2022, the stunting rate in Indonesia reached 21.6%, exceeding the World Health Organization's standard threshold, making stunting a serious public health issue, Kementerian Kesehatan Republik Indonesia [14]. Factors contributing to stunting involve a complexity of aspects such as inadequate nutrition, recurrent infections, poor sanitation, inappropriate disease management in infants, and limited healthcare services. The characteristic of stunting can be seen in a child's height being shorter than the age standard, World Health Organization [20]. The selection of stunting data as the focus of the research is based on the urgency of this public health issue in Indonesia. The case data to be analysed includes aspects of education, social, and environmental factors, such as: the district where the child's guardian resides and the management of fever in infants for CA $2 \times J$, as well as the highest level of education of the guardian and sources of stunting information for CA $3 \times J$.

CA is also associated with the chi-square test χ^2, allowing the hypothesis of dependence to be tested using Pearson's ratio matrix $\boldsymbol{\phi}$, Beh and Lombardo [3]. The reduction of Pearson's ratio matrix $\boldsymbol{\phi}$ results in a standard residual matrix $\mathbf{Z}$, which serves as the basis for SVD to obtain orthonormal left and right singular matrices. Since the left and right singular matrices are orthonormal, their vector lengths are 1 and the principal coordinates will range between -1 and 1. Principal coordinates are obtained from the SVD result matrix, allowing the creation of an AK map with reduced dimensions, Beh and Lombardo [3]. This methodological approach underscores the utility of CA and SVD in public health research, particularly in addressing complex issues like stunting.

2 THEORY

2.1 Correspondence Analysis

Suppose there are two categorical random variables $X = \{1,2,\dots,I\}$ and $Y = \{1,2,\dots,J\}$, with $I,J \geq 2$. Each of X and Y are random variables for rows and columns. Thus, a two-way contingency table can be constructed where n_{ij} is the frequency in row i and column j, $n_{i.}$ is the marginal frequency for row i, $n_{.j}$ is the marginal frequency for column j, and $n_{..}$ is the total frequency. From the contingency table, the cross-tabulation matrix $\mathbf{N} = (n_{ij})$, row profile matrix $\mathcal{P}_{\mathbf{row}} = \left(\frac{n_{ij}}{n_{i.}}\right)$ and column profile matrix $\mathcal{P}_{\mathbf{col}} = \left(\frac{n_{ij}}{n_{.j}}\right)$ can be constructed. Furthermore, the correspondence matrix from the relative frequencies with respect to the total observations is $\mathbf{P} = (p_{ij}) = \left(\frac{n_{ij}}{n_{..}}\right)$. Let $\vec{\boldsymbol{r}} = (\boldsymbol{r_1} \quad \boldsymbol{r_2} \quad \cdots \quad \boldsymbol{r_I})^T$ where $\boldsymbol{r_i} = \sum_{i=1}^{I} \frac{\boldsymbol{n_{ij}}}{\boldsymbol{n_{..}}}$ and $\vec{\boldsymbol{c}} = (\boldsymbol{c_1} \quad \boldsymbol{c_2} \quad \cdots \quad \boldsymbol{c_J})^T$ where $\boldsymbol{c_j} = \sum_{j=1}^{J} \frac{\boldsymbol{n_{ij}}}{\boldsymbol{n_{..}}}$. A matrix $\mathbf{Z}$ can be constructed involving the marginal probability mass functions of the row variables $\vec{\boldsymbol{r}}$ and the column variables $\vec{\boldsymbol{c}}$, Beh and Lombardo [3].

$$\mathbf{Z} = diag_{\frac{1}{\sqrt{r}}}(\mathbf{P} - \mathbf{rc^T})diag_{\frac{1}{\sqrt{c}}} = \left(\frac{n_{ij} - \frac{n_{i.}n_{.j}}{n_{..}}}{\sqrt{n_{i.}n_{.j}}}\right) \quad \textbf{\textit{(1)}}$$

Table 1 Contingency Table $I \times J$

Y / X	**1**	**2**	**...**	**J**	**Total Baris**
1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I,}$
Total Kolom	$n_{.1}$	$n_{.2}$	...	$n_{.J}$	$n_{..}$

The Pearson Chi-Square test is used to determine whether there is a dependence between two categorical random variables. Let $P(X = i, Y = j) \approx p_{ij}$, $P(X = i) \approx p_{i.}$, and $P(Y = j) \approx p_{.j}$. The hypotheses are formulated as follows:

$$H_0: p_{ij} = p_{i.}p_{.j};\ H_1: p_{ij} \neq p_{i.}p_{.j} \quad (2)$$

The χ^2 statistic for H_0 is given by:

$$\chi^2 = n_{..} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n_{..}}\right)^2}{\left(n_{i.}\frac{n_{.j}}{n_{..}}\right)} \quad (3)$$

with degree of freedom $v = (I - 1)(J - 1)$. H_0 is rejected when p-value is smaller than the significance level α, Walpole et al. [19].

2.2 Singular Value Decomposition

Let $K = \min\{I, J\}$, SVD decomposes the standard residual matrix $\boldsymbol{Z}$ into the product of three matrices:

$$\boldsymbol{Z} \quad = \quad \widetilde{\boldsymbol{A}} \qquad \Delta \qquad \widetilde{\boldsymbol{B}}^T$$

$(I \times J) \qquad (I \times K) \;\; (K \times K) \;\; (K \times J)$ ***(4)***

Then, matrix $\mathbf{M}$ can be constructed as follows:

$$\boldsymbol{M} = \begin{cases} \boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}} = \widetilde{\boldsymbol{A}}\Delta\widetilde{\boldsymbol{B}}^{\boldsymbol{T}}\widetilde{\boldsymbol{B}}\Delta\widetilde{\boldsymbol{A}}^{\boldsymbol{T}} = \widetilde{\boldsymbol{A}}\Delta^2\widetilde{\boldsymbol{A}}^{\boldsymbol{T}}, \text{jika } \boldsymbol{I} \leq \boldsymbol{J} \\ \boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z} = \widetilde{\boldsymbol{B}}\Delta\widetilde{\boldsymbol{A}}^{\boldsymbol{T}}\widetilde{\boldsymbol{A}}\Delta\widetilde{\boldsymbol{B}}^{\boldsymbol{T}} = \widetilde{\boldsymbol{B}}\Delta^2\widetilde{\boldsymbol{B}}^{\boldsymbol{T}}, \text{jika } \boldsymbol{I} > \boldsymbol{J} \end{cases} \quad (5)$$

Columns $I \times K$ of matrix $\widetilde{\boldsymbol{A}}$ are the orthonormal eigenvectors of $\boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}}$, while columns $J \times K$ of matrix $\widetilde{\boldsymbol{B}}$ are the orthonormal eigenvectors of $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z}$. Since the columns of $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{B}}$ are orthonormal. $\widetilde{\boldsymbol{A}}^{\boldsymbol{T}}\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{B}}^{\boldsymbol{T}}\widetilde{\boldsymbol{B}} = \boldsymbol{I}$, Beh and Lombardo [3]. The orthonormal eigenvectors in the columns of $\tilde{A}$ correspond to the eigenvalues of$\boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}}$, similarly, the orthonormal eigenvectors in the columns of $\widetilde{\boldsymbol{B}}$ correspond to the eigenvalues of $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z}$. The values $\lambda_1, \lambda_2, \ldots, \lambda_K$ in the matrix $\Delta = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_K})$ are the eigenvalues of $\boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}}$ or $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z}$. The values $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_K}$ in matrix Δ are called the singular values of $\mathbf{Z}$. Subsequently, row and column coordinate matrices representing the positions of the row and column variables can be obtained, Beh and Lombardo [3].

$$\mathbf{F} = diag\frac{1}{\sqrt{r}}\widetilde{\boldsymbol{A}}\Delta \quad (6)$$

$$\mathbf{G} = diag\frac{1}{\sqrt{c}}\widetilde{\boldsymbol{B}}^{\boldsymbol{T}}\Delta \quad (7)$$

The proportion of data variance explained by the CA map of the principal coordinates is indicated by the inertia value. The total percentage of inertia across all coordinate axes on the CA map is 100%. The total inertia on the CA map is given by $\sum_{i=1}^{K}\lambda_i$. Based on the total inertia, the contribution of each dimension can be determined by the percentage of inertia.

$$I_i = \frac{\lambda_i}{\sum_{i=1}^{K}\lambda_i} \times 100\%, \text{dengan } i = 1,2, \ldots, K \quad (8)$$

2.3 Simplification of CA

2.3.1 AK $2 \times \boldsymbol{J}$

Let matrix $\mathbf{M} = \mathbf{Z}\mathbf{Z}^{\mathbf{T}}$, to minimize the computational process of the matrix, mathematical calculations are presented in Theorem 1, Lemma 1, Lemma 2, Lemma 3, Lemma 4, Lemma 5, and Theorem 2 by Ginanjar [9].

Teorema 1. If the size of matrix $\mathbf{N}$ is $I \times J$ and $\mathbf{Z}\mathbf{Z}^{\mathbf{T}} = \mathbf{M}$, then 0 is an eigenvalue of the matrix $\mathbf{M}$.

Lemma 1. If the size of the matrix $\mathbf{N}$ is $I \times J$ and $\mathbf{Z}\mathbf{Z}^{\mathbf{T}} = \mathbf{M}$, then $m_{ik} = \frac{1}{\sqrt{n_{i\cdot}n_{k\cdot}}}\left(\sum_{j=1}^{J}\frac{n_{ij}n_{kj}}{n_{\cdot j}} - \frac{n_{i\cdot}n_{k\cdot}}{n_{\cdot\cdot}}\right), \quad i = k = 1,2, \ldots, I.$

Lemma 2. If the size of the matrix $\mathbf{N}$ is $2 \times J$ and $\mathbf{Z}\mathbf{Z}^{\mathbf{T}} = \mathbf{M}$, then the elements of $\mathbf{M} = (m_{ik})$ with $i = k = 1,2$ are obtained as $m_{11} = \frac{n_{2\cdot}}{n_{1\cdot}}m_{22} = -\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}m_{12} = -\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}m_{21}$.

Lemma 3. If the size of the matrix $\mathbf{N}$ is $2 \times J$ and $\mathbf{Z}\mathbf{Z}^{\mathbf{T}} = \mathbf{M}$, then the first eigenvalue (λ_1) corresponding to $\mathbf{M}$ is $\frac{n_{\cdot\cdot}}{n_{1\cdot}n_{2\cdot}}\left(\sum_{j=1}^{J}\frac{n_{1j}^2}{n_{\cdot j}} - \frac{n_{1\cdot}{}^2}{n_{\cdot\cdot}}\right)$, and the second eigenvalue(λ_2) corresponding to $\mathbf{M}$ adalah 0.

Lemma 4. If the size of the matrix $\mathbf{N}$ is $2 \times J$ and $\mathbf{ZZ^T} = \mathbf{M}$, then the orthonormal eigenvectors of $\mathbf{M}$ are $\widetilde{\boldsymbol{A}} = (\vec{a}_1 \quad \vec{a}_2) = \frac{1}{\sqrt{n}}\begin{pmatrix} -\sqrt{n_{2\cdot}} & \sqrt{n_{1\cdot}} \\ \sqrt{n_{1\cdot}} & \sqrt{n_{2\cdot}} \end{pmatrix}$.

Lemma 5. If the size of the matrix $\mathbf{N}$ is $2 \times J$ and $\mathbf{ZZ^T} = \mathbf{M}$, then the elements of the orthonormal eigenvector corresponding to the first eigenvalue (λ_1) of $\mathbf{M}$ are $b_{j1} = \frac{n_{2j}n_{1\cdot} - n_{1j}n_{2\cdot}}{\sqrt{\lambda n_{1\cdot} n_{2\cdot} n_{\cdot j} n_{\cdot\cdot}}}$; $j = 1,2,\dots,J$.

Teorema 2. If the size of the matrix $\mathbf{N}$ is $2 \times J$, t hen the coordinates for the row variable $\boldsymbol{F}$ are given by $\boldsymbol{F} = \sqrt{\frac{n_{\cdot\cdot}}{n_{1\cdot}^2}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{\cdot j}} - 1}\begin{pmatrix} -1 & 0 & \dots & 0 \\ \frac{n_{1\cdot}}{n_{2\cdot}} & 0 & \dots & 0 \end{pmatrix}$ and the coordinates for the column variable are $\boldsymbol{G} = \begin{pmatrix} g_1 & 0 \\ g_2 & 0 \\ \vdots & \vdots \\ g_J & 0 \end{pmatrix}$ with $g_j = \frac{n_{2j}n_{1\cdot} - n_{1j}n_{2\cdot}}{n_{\cdot j}\sqrt{n_{1\cdot}n_{2\cdot}}}$.

2.3.2 AK $3 \times J$

Given a matrix $\mathbf{M} = \mathbf{ZZ^T}$, to minimize the computational process, a mathematical analysis is presented in Lemma 6, Lemma 7, Lemma 8, and Theorem 3 by Ginanjar [9].

Lemma 6. If the size of the matrix $\mathbf{N}$ is $3 \times J$ with $J \geq 3$, and $\mathbf{M} = (m_{ik})$ is symmetric and has a non-trivial solution, let $f = m_{11} + m_{22} + m_{33}, g = m_{11}m_{22} + m_{11}m_{33} + m_{22}m_{33}$, and $h = m_{12}^2 + m_{13}^2 + m_{23}^2$. Then the eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ of $\mathbf{M}$ are $\lambda_1 = \frac{f + \sqrt{f^2 - 4(g-h)}}{2}, \lambda_2 = f - \lambda_1, \lambda_3 = 0$.

Lemma 7. If the size of the matrix $\mathbf{N}$ is $3 \times J$ with $J \geq 3$, and $\mathbf{M} = (m_{ik})$ is a 3×3 symmetric matrix with a non-trivial solution, and λ_k is the k-th eigenvalue of the matrix $\mathbf{M}$, let $\acute{a}_{1k} = m_{12}m_{23} - m_{13}m_{22} + m_{13}\lambda_k, \acute{a}_{2k} = m_{12}m_{13} - m_{23}m_{11} + m_{23}\lambda_k$, and $\acute{a}_{3k} = (m_{11} - \lambda_k)(m_{22} - \lambda_k) - m_{12}^2$, Then the matrix $\widetilde{\boldsymbol{A}}$, where each column is an orthonormal eigenvector corresponding to λ_k is $\widetilde{\boldsymbol{A}} = (a_{ik}), a_{ik} = \frac{\acute{a}_{ik}}{\sqrt{\sum_{i=1}^{3}\acute{a}_{1k}^2}}, i = k = 1,2,3$.

Lemma 8. If the size of the matrix $\mathbf{N}$ is $3 \times J$ with $J \geq 3$, and λ_k is the k-th eigenvalue of the matrix $\mathbf{M}$, and a_{ik} is the i-th element of the left singular vector corresponding to λ_k then b_{jk} the j-th element of the right singular vector corresponding to λ_k is given by $b_{jk} = \frac{1}{n_{\cdot\cdot}\sqrt{n_{\cdot j}\lambda_k}}\left(\sum_{i=1}^{3}\frac{a_{ik}}{\sqrt{n_{i\cdot}}}(n_{ij}n_{\cdot\cdot} - n_{i\cdot}n_{\cdot j})\right)$.

Teorema 3. If $\mathbf{N} = (n_{ij})$ is a contingency table matrix of size $3 \times J$ with $J \geq 3$ and $\mathbf{M} = (m_{ik})$ is symmetric and has a non-trivial solution for $i, k = 1,2,3$ then the principal coordinates for rows and columns are given by $\mathbf{F} = (f_{ik})$, with $f_{ik} = \sqrt{\frac{n_{\cdot\cdot}\lambda_k}{n_{i\cdot}}}\frac{\acute{a}_{ik}}{\sqrt{\sum_{i=1}^{3}\acute{a}_{1k}^2}}$ and $\mathbf{G} = (g_{jk})$, with

$$g_{jk} = \frac{1}{n_{\cdot j}\sqrt{n_{\cdot\cdot}}}\left(\sum_{i=1}^{3}\frac{\acute{a}_{ik}}{\sqrt{n_{i\cdot}\sum_{i=1}^{3}\acute{a}_{1k}^2}}(n_{ij}n_{\cdot\cdot} - n_{i\cdot}n_{\cdot j})\right).$$

3 MATHEMATICAL RESULTS

3.1 Case Study: Districts and Methods for handling fever in toddlers

The data focuses on the province of Banten, specifically the regencies of Lebak and Pandeglang, with a total of 392 respondents. The respondents are relatives or guardians of children under five years old. The data is presented in a contingency table as shown in Table 2. From Table 2, column profiles and row profiles can be obtained which can then be depicted in two-dimensional and three-dimensional graphs, respectively. Column and row profile graphs can be reduced to one dimension by performing projections.

Table 2 Case Study Districts and Methods for handling fever in toddlers

Districts	Methods for handling fever in toddlers			Row totals
	Visiting a midwife	Visiting a doctor/community health center/hospital	Others	
Lebak	72	113	29	214
Pandeglang	73	85	20	178
Column totals	145	198	49	392

$$\mathcal{P}_{\text{col}} = \begin{matrix} \\ \text{Lebak} \\ \text{Pandeglang} \end{matrix} \begin{matrix} \text{Visiting a midwife} & \text{Visiting a doctor/ ...} & \text{Others} \\ \begin{pmatrix} \frac{72}{145} & \frac{113}{198} & \frac{29}{49} \\ 1-\frac{72}{145} & 1-\frac{113}{198} & 1-\frac{20}{49} \end{pmatrix} \end{matrix}$$

$$\mathcal{P}_{\text{row}} = \begin{matrix} \text{Visiting a midwife} \\ \text{Visiting a doctor/...} \\ \text{Others} \end{matrix} \begin{matrix} \text{Lebak} \quad \text{Pandeglang} \\ \begin{pmatrix} \frac{72}{214} & \frac{73}{178} \\ \frac{113}{214} & \frac{85}{178} \\ 1-\frac{185}{214} & 1-\frac{158}{178} \end{pmatrix} \end{matrix}$$

The column and row profile plots are shown in Figures 1 and 2. Subsequently, using the chi-square test χ^2, the independence of the two variables in Table 2 was tested, yielding a chi-square statistic of $\chi^2 = 0{,}23331$ with 2 degrees of freedom. The calculated P-value was 0.3114, leading to the acceptance of the null hypothesis H_0. Therefore, there is not enough evidence to suggest a significant dependence between the two variables in Table 2 at the 0.05 significance level.

Using the simplification of CA by Ginanjar [9], the eigenvalues obtained are $\lambda_1 = \frac{392}{214\times178}\left(\frac{72^2}{145}+\frac{113^2}{198}+\frac{29^2}{49}-\frac{214^2}{392}\right) = \frac{7604}{1277595}$ and $\lambda_2 = 0$. The obtained eigenvalues can be used to determine the percentage of inertia of the dimension on the CA map. The percentage of the first dimension is $I_1 = 100\%$. The coordinates for the row variables and column variables, represented by matrices FFF and GGG, can be directly computed using information from the contingency table. This allows for the construction of a correspondence analysis map, as shown in Figure 3.

$$\mathbf{F} = \sqrt{\frac{392}{214^2}\cdot\left(\frac{72^2}{145}+\frac{113^2}{198}+\frac{29^2}{49}-1\right)}\begin{pmatrix} -1 & 0 \\ \frac{214}{178} & 0 \end{pmatrix} = \begin{pmatrix} -0{,}0704 & 0 \\ 0{,}0846 & 0 \end{pmatrix}$$

$$\mathbf{G} = \begin{pmatrix} \frac{73\cdot214-72\cdot178}{145\sqrt{214\cdot178}} & 0 \\ \frac{85\cdot214-113\cdot178}{198\sqrt{214\cdot178}} & 0 \\ \frac{20\cdot214-29\cdot178}{49\sqrt{214\cdot178}} & 0 \end{pmatrix} = \begin{pmatrix} 0{,}0992 & 0 \\ -0{,}0498 & 0 \\ -0{,}0922 & 0 \end{pmatrix}$$

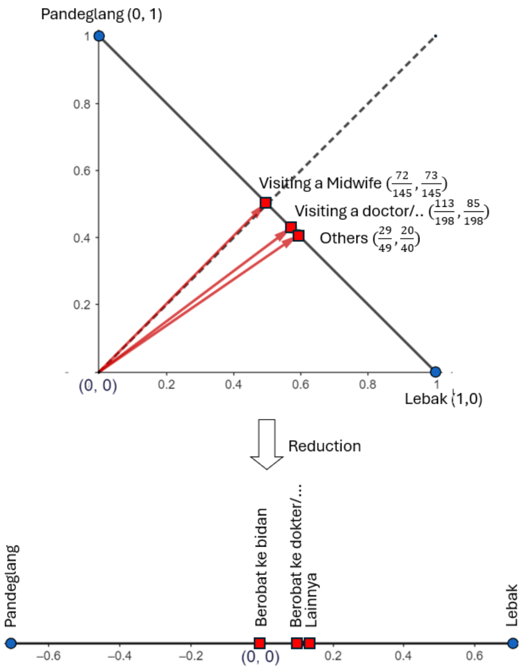

Figure 1 Reduction of Column Profiles (Districts and Methods for handling fever in toddlers)

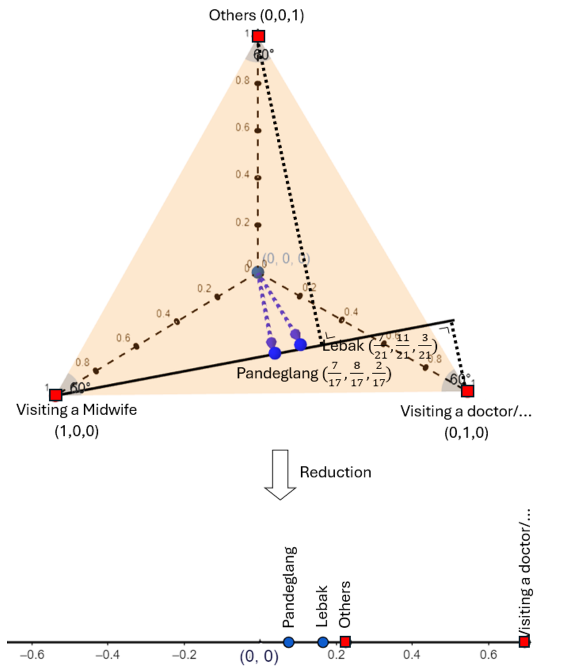

Figure 2 Reduction of Row Profiles (Districts and Methods for handling fever in toddlers)

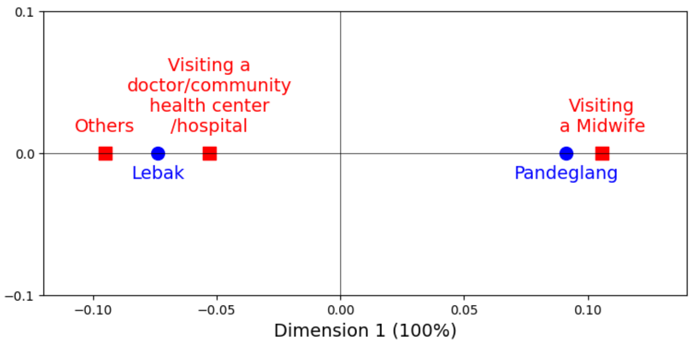

Figure 3 CA Map (Districts and Methods for handling fever in toddlers)

3.2 Case Study: Education level and Sources of stunting information

A chi-square χ^2 test was conducted to test the independence between two variables in Table 3. The chi-square test statistic was found to be $\chi^2 = 45{,}603$ with 10 degrees of freedom. The p-value calculated was $1{,}692 \times 10^{-6}$ leading to the rejection of the null hypothesis H_0.

Table 3 Case Study: Respondents' Last Education Level and Sources of Information on Stunting

Education level	Sources of stunting information						Row totals
	Community Health Center	Integrated Health Post	Social Media	Electronic Media	Midwife	Others	
$\leq$ Jr High	107	8	10	3	37	2	167
Sr High	98	5	36	7	22	1	169
University	19	4	17	5	8	3	56
Column totals	224	17	63	15	67	6	392

Subsequently, the column profiles and row profiles can be obtained as follows:

$$\mathcal{P}_{\textbf{kolom}} = \begin{array}{r} \\ \leq \text{ Jr High} \\ \text{Sr High} \\ \text{University} \end{array} \begin{array}{cccccc} \text{Community Health Center} & \text{Integrated Health Post} & \text{Social Media} & \text{Electronic Media} & \text{Midwife} & \text{Others} \end{array}$$

$$\begin{pmatrix} \frac{107}{224} & \frac{8}{17} & \frac{10}{63} & \frac{3}{15} & \frac{37}{67} & \frac{2}{6} \\ \frac{98}{224} & \frac{5}{17} & \frac{36}{63} & \frac{7}{15} & \frac{22}{67} & \frac{1}{6} \\ 1-\frac{205}{224} & 1-\frac{13}{17} & 1-\frac{46}{63} & 1-\frac{10}{15} & 1-\frac{59}{67} & 1-\frac{3}{6} \end{pmatrix}$$

Using the simplification of CA by Ginanjar [9], the eigenvalues were obtained as $\lambda_1 = 0{,}094$, $\lambda_2 = 0{,}023$, $\lambda_3 = 0$. This resulted in the first dimension is $I_1 = 80{,}45\%$ and the second dimension is $I_2 = 19{,}55\%$., respectively. The coordinates for the row and column variables, represented by matrices $\mathbf{F}$ and $\mathbf{G}$, can be calculated using Theorem 3, leading to the CA map shown in Figure 5.

$$\mathbf{F} = \begin{pmatrix} -0{,}300 & -0{,}094 \\ 0{,}098 & 0{,}166 \\ 0{,}599 & -0{,}222 \end{pmatrix} \text{ and } \mathbf{G} = \begin{pmatrix} -0{,}162 & 0{,}060 \\ 0{,}093 & -0{,}315 \\ 0{,}555 & 0{,}134 \\ 0{,}606 & -0{,}101 \\ -0{,}202 & -0{,}157 \\ 0{,}705 & -0{,}760 \end{pmatrix}$$

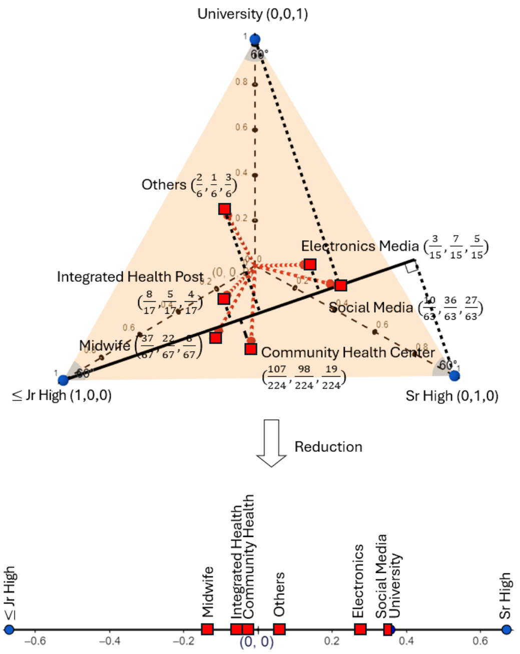

Figure 4 Reduction of Column Profiles (Education level and Sources of stunting information)

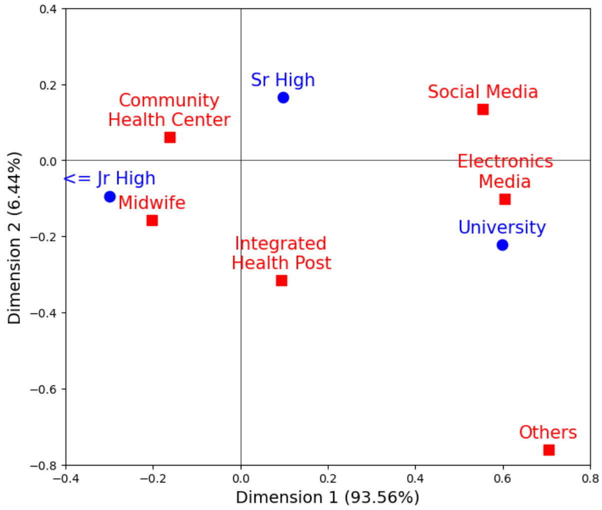

Figure 5 CA Map (Education level and Sources of stunting information)

4 CONCLUSION

The results for the left singular matrix, right singular matrix, eigenvalues, and principal coordinates from CA are more accurate when calculated directly from the elements of matrix $\mathbf{N}$. The number of non-zero singular values $\sqrt{n}$ in CA $2 \times J$ with $I \leq J$ is $\min(2, J) - 1 = 1$. Meanwhile, the number of non-zero singular values $\sqrt{n}$ in CA $3 \times J$ with $I \leq J$ is $\min(3, J) - 1 = 3$. The CA map generated from the $2 \times J$ matrix $\mathbf{N}$ is one-dimensional, with the percentage of inertia explained by the first dimension being 100%. The CA map generated from the $2 \times J$ matrix $\mathbf{N}$ is two-dimensional, with the sum of the percentage of inertia explained by the first and second dimension being 100%.

The Correspondence Analysis (CA) mapping provides insightful visualizations of associations between variables. In our study, the proximity of "Seeking treatment at a doctor/community health centre/hospital" to "Others" in the column profile reduction graph suggests a similar distribution pattern. A one-dimensional reduction places "Lebak" near "Others," indicating a notable association. Additionally, the close positioning of "University" with "Social Media" and "Electronic Media" implies a relationship that warrants further analysis using the CA map.

The CA map reveals meaningful associations between different variables: For example, respondents in Pandeglang are inclined to treat children's fever by visiting a midwife, whereas those in Lebak prefer doctors or community health centres. In terms of education level and stunting information sources, individuals with lower educational attainment often rely on midwives and community health centres, while those with higher education levels tend to access information through social and electronic media.

These findings demonstrate the practical application of CA in uncovering relationships within complex datasets, relevant to the conference's focus on advanced methodologies in data analysis and their implications in public health and social sciences. The ability to discern these associations can inform targeted public health strategies and interventions, highlighting the importance of sophisticated analytical techniques in addressing critical issues like stunting in Indonesia.

5 REFERENCES

[1] Afrianti, Y.S., Pasaribu, U.S., Ardy, Husaini., and Lestari, K.E. 2024. Elliptic Confidence on Correspondence Analysis Stainless Steels Corrosion Problem. Engineering Letters, 32(7).

[2] Banerjee, S., and Roy, A. 2014. Linear Algebra and Matrix Analysis for Statistics, Taylor & Francis Group, LLC, Boca Raton.

[3] Beh, E.J., and Lombardo, R. 2014. Correspondence Analysis Theory, Practice and New Strategies, John Wiley & Son, Ltd, West Sussex.

[4] Brzezińska, J. 2018. Singular Value Decomposition Approaches in A Correspondence Analysis with The Use of R. Folia Oeconomica Stetinensia, 18(2), pp 178-189. doi:10.2478/foli-2018-0026.

[5] Dewan Perwakilan Rakyat Republik Indonesia. 2023. Info Singkat: Terobosan dalam Mengatasi Stunting di Indonesia. Available online: https://berkas.dpr.go.id/pusaka/files/info_singkat/Info%20Singkat-XV-14-II-P3DI-Juli-2023-196.pdf.

[6] Dinas Kesehatan Provinsi Papua. 2024. Menkes Budi soroti lambatnya penurunan angka stunting di Indonesia. Available online: https://dinkes.papua.go.id/menkes-budi-soroti-lambatnya-penurunan-angka-stunting-di-indonesia.

[7] Dwi Nur'aini, K., Suryani, D.R., Riyana, M., Sianturi, M., Mukhaiyar, U., and Nurvitasari, E. 2018. Correspondence Analysis to Identify Alumni Profiles: A Case Study on ITB Alumni Class of 2008.

[8] Ginanjar, I., Pasaribu, U.S., and Barra, A. 2016. Simplification of correspondence analysis for more precise calculation which one qualitative variables is two categorical data. 11(3). Available online: www.arpnjournals.com.

[9] Ginanjar, I., Pasaribu, U.S., and Barra, A. 2019. Penyederhanaan Analisis-Korespondensi Untuk Meningkatkan Akurasi Koordinat Utama. Dissertation. Faculty of Mathematics and Natural Sciences, Mathematics, Institut Teknologi Bandung, Bandung.

[10] Greenacre, M.J. 1984. Theory and applications of correspondence analysis. Chapman & Hall/CRC.

[11] Greenacre, M.J. 2017. Correspondence Analysis in Practice. Chapman & Hall/CRC.

[12] Hardle, W.K., and Simar, L. 2015. Applied Multivariate Statistical Analysis Fourth Edition, Springer, Berlin.

[13] Kementerian Kesehatan Republik Indonesia. n.d. 1 dari 3 balita Indonesia derita stunting. P2PTM Kemenkes. Available online: https://p2ptm.kemkes.go.id/tag/1-dari-3-balita-indonesia-derita-stunting.

[14] Kementerian Kesehatan Republik Indonesia. 2023. Mengenal lebih jauh tentang stunting. Available online: https://yankes.kemkes.go.id/view_artikel/2657/mengenal-lebih-jauh-tentang-stunting.

[15] Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan. 2023. Perlu terobosan dan intervensi tepat sasaran lintas sektor untuk atasi stunting. Kemenko PMK.

Available online: https://www.kemenkopmk.go.id/perlu-terobosan-dan-intervensi-tepat-sasaran-lintas-sektor-untuk-atasi-stunting.
[16] Lestari, K.E., Utami, M.R., and Yudhanegara, M.R. 2022. Simple Algorithm to Construct Circular Confidence Regions in Correspondence Analysis Using R. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 16(1), pp 065-074. doi:10.30598/barekengvol16iss1pp065-074.
[17] Mair, P. 2018. Modern psychometrics with R. Springer International Publishing. doi:10.1007/978-3-319-93177-7.
[18] Tuffery, S., and Riesco, R. 2005. Data mining and statistics for decision making. John Wiley & Sons, Ltd, Chichester.
[19] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability & Statistics for Engineers & Scientists* (9th ed.). Pearson Education, Boston.
[20] World Health Organization. 2024. Malnutrition in children: Stunting, wasting, overweight, and underweight. Available online: https://apps.who.int/nutrition/landscape/help.aspx?menu=0&helpid=391&lang=EN.
[21] World Health Organization. 2015. Stunting in a nutshell. Available online: https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell.