Interrelations Between Population Dynamics and Service Accessibility in Indonesian Provinces: A 2023 Canonical Correlation Study

Annisa Indriani  
*Bina Nusantara University*  
Computer Science and Statistics DepartmentJakarta, Indonesia 11530  
[annisa.indriani@binus.ac.id](mailto:annisa.indriani@binus.ac.id)

Elvano Jethro Mogi Pardede  
*Bina Nusantara University*  
Computer Science and Statistics DepartmentJakarta, Indonesia 11530  
[elvano.pardede@binus.ac.id](mailto:elvano.pardede@binus.ac.id)

Rasyad Muhammad Ramdhanazuri  
*Bina Nusantara University*  
Computer Science and Statistics DepartmentJakarta, Indonesia 11530  
[rasyad.ramdhanazuri@binus.ac.id](mailto:rasyad.ramdhanazuri@binus.ac.id)

Dira Abiyyu Bagaspati  
*Bina Nusantara University*  
Computer Science and Statistics  
Department  
Jakarta, Indonesia 11530  
[email@binus.ac.id](mailto:)

Margaretha Ohyver  
*Bina Nusantara University*  
Computer Science and Statistics  
DepartmentJakarta, Indonesia 11530  
[mohyver@binus.edu](mailto:mohyver@binus.edu)

*Abstract*—This study examines the relationship between demographic factors and access to basic services across Indonesian provinces using Canonical Correlation Analysis (CCA). The analysis focuses on six variables: population growth rate, population percentage, and population density as demographic factors; and access to drinking water, sanitation, and healthcare as basic services. The study aims to uncover the multidimensional relationships between these variables to inform more effective and equitable policy interventions. The results reveal a significant canonical correlation between demographic changes and service availability, highlighting the inverse relationship between population density and access to basic services. These findings offer valuable insights for addressing regional disparities and improving resource allocation in Indonesia.

Keywords—canonical correlation, health, demographics, multivariate statistics

# Introduction

The United Nations Sustainable Development Goals (SDGs) provide a universal framework for addressing critical global challenges. The 3rd goal of the SDGs emphasizes ensuring good health and well-being for all [1], aiming to reduce mortality rates, improve access to healthcare, and enhance overall life expectancy. Simultaneously, the 6th goal focuses on ensuring availability and sustainable management of water and sanitation for all [1], highlighting the necessity of clean water and adequate sanitation facilities as fundamental human rights.

Analyzing the relationship between demographic factors and access to basic services is crucial for several reasons. Demographic factors such as population growth rate, population percentage, and population density provide essential insights into the distribution and dynamics of the population across provinces. These factors can significantly influence the demand for resources and services. Basic services, including access to drinking water, sanitation, and healthcare, are critical determinants of public health and quality of life. Understanding how these services are distributed in terms of demographic factors can reveal inequalities and inform policy decisions to address inequalities.

Canonical correlation analysis (CCA) is a powerful statistical method used to explore the relationship between two sets of variables. In this context, CCA can help explain the complex interdependencies between demographic factors and access to basic services. By identifying the underlying correlations, policymakers can gain a better understanding of how demographic changes impact service availability and vice versa. This analysis is especially important in a diverse and densely populated country like Indonesia, where regional inequalities can significantly affect development outcomes.

Canonical correlation analysis (CCA) is a powerful statistical method used to explore the relationship between two sets of variables. In this context, CCA can help explain the complex interdependencies between demographic factors and access to basic services. By identifying the underlying correlations, policymakers can gain a better understanding of how demographic changes impact service availability and vice versa. This analysis is especially important in a diverse and densely populated country like Indonesia, where regional inequalities can significantly affect development outcomes.

[2] stated that one of the impacts of high population density is the difficulty in meeting basic needs such as clean water, sanitation, healthcare services, and education, as well as increased unemployment, poverty, and access to adequate infrastructure. This aligns with [3], who observed that densely populated slum areas are often characterized by poor housing conditions and a lack of access to essential services such as clean water, sanitation, electricity, and healthcare facilities .

Several studies have addressed these issues using canonical correlation analysis. [4]measured the relationship between infrastructure indicators and the Human Development Index (HDI) indicators using canonical correlation analysis, finding that infrastructure indicators are related to HDI indicators, but HDI indicators have a weak relationship with infrastructure indicators. [5] employed canonical discriminant analysis to examine the relationship between socioeconomic factors and HDI, concluding that variations in city/regency welfare based on HDI values can be explained by these socioeconomic factors.

By exploring the canonical correlation between demographic factors and access to basic services, this study aims to uncover the multidimensional relationships that can inform more effective and equitable policy interventions. The findings of this research will provide valuable insights for addressing regional disparities and improving the allocation of resources across Indonesian provinces.

# Methodology

## Dataset

The dataset consists of six variables grouped into two categories, each indicating various aspects of provincial development and living conditions. All data were collected throughout 2023 and each row in the dataset corresponds to one of Indonesia's provinces.[6], [7]

### Group 1: Demographic Factors[6]

* Population Growth Rate per Year: This variable measures the population growth rate for the year 2023, reflecting the changes in the population size during this specific year in each province.
* Population Percentage: This variable represents the proportion of the population within a specific province relative to the total population of Indonesia in 2023.
* Population Density per square km, 2023: This metric indicates how densely populated an area is in 2023, calculated as the number of people living per square kilometer in each province.

### Group 2: Access to Basic Services[7]

* Proportion of Households with Access to Drinking Water: This variable measures the percentage of households within a province that have access to clean and safe drinking water in 2023.
* Proportion of Households with Access to Basic Sanitation: This variable indicates the percentage of households with access to essential sanitation facilities in 2023 for each province.
* Proportion of Households with Access to Basic Health Services, 2023: This metric shows the percentage of households that have access to primary healthcare services in 2023 for each province.

## Variable Selection

Demographic factors can significantly affect the demand for basic services. For example, a high population growth rate in a single year in a province can lead to increased pressure on existing infrastructure and resources, requiring rapid adjustments in services supply [8]. Similarly, understanding the proportion of population in each province is critical for equal distribution of resources. Larger populations may require more extensive infrastructure and services, while smaller populations may face challenges due to less viable resources. Areas with high population density may experience difficulties with overcrowded facilities and overloaded services, while areas with low population density may experience problems of insufficient service coverage and accessibility issues. [9]

Population Growth Rate per Year is selected as it reflects the annual changes in the population size during 2023, providing insights into how rapidly populations are increasing within that specific year in each province. Rapid population growth can significantly impact infrastructure development and resource allocation, necessitating swift policy responses to accommodate the growing population and maintain service quality [10].

Population Percentage helps identify the relative size of the population in each province for the year 2023. This information is vital for understanding regional disparities and ensuring that resources and services are distributed equitably. Provinces with larger population percentages might require more extensive infrastructure and services to meet the needs of their residents [9].

Population Density per square km, is a key indicator of how population distribution affects living conditions and resource availability. High population density can strain infrastructure and resources, while low density might imply underutilization of services and facilities. Analyzing population density helps in planning and optimizing service delivery to match the specific needs of each province [9], [11].

Proportion of Households with Access to Drinking Water is a fundamental measure of living standards and public health. Access to clean and safe drinking water is crucial for preventing waterborne diseases and ensuring the well-being of the population. This variable measures the percentage of households within each province that have access to drinking water in 2023, highlighting regional disparities in basic service provision [12].

Proportion of Households with Access to Basic Sanitation directly impacts public health outcomes and the quality of life within each province. Proper sanitation facilities are essential for preventing diseases and maintaining a healthy environment. This variable indicates the percentage of households with access to essential sanitation facilities in each province for 2023, reflecting the effectiveness of regional sanitation policies [13].

Proportion of Households with Access to Basic Health Services is a critical indicator of the health system's reach and effectiveness in providing necessary care to the population. Access to primary healthcare services is vital for addressing basic health needs, preventing diseases, and ensuring overall health security. This metric shows the percentage of households that have access to primary healthcare services in each province for 2023, underscoring the importance of equitable healthcare access [14], [15].

## Canonical Correlation

Canonical Correlation Analysis (CCA) is a multivariate statistical method to analyze paired sets of variables. The observation variables can be partitioned into two sets which can be considered as two views of the data [16]. This method is used to understand the relationship between two sets of variables that contain at least two variables in each set [17]. CCA is useful when dealing with multiple correlated variables from different groups or to explore the relationship between two modalities of data.

### Variance Covariance Matrix

The variance covariance matrix is a matrix that contains a measure of the linear relationship between two variables [18]. The following is an example of sample variable covariance matrix for variables:



From (1), The diagonal matrix contains the sample variance of p variables. Meanwhile, the values outside the diagonal matrix contain the sample covariance of a pair p variables [19].

### Canonical Weights

Canonical weights refer to the coefficients or weights assigned to the original variables in canonical correlation analysis (CCA). In CCA, the goal is to find linear combinations of variables from two sets that maximize the correlation between these combinations, known as canonical variates. Use the equation below to calculate canonical weights [20]:

2

### Canonical Variates

Canonical variates as known as canonical functions are linear combinations of the original variables to maximize the correlation between the two sets of the two linear functions [21]. They indicate the strength and direction of the relationship between each original variable and the canonical variates, helping to interpret the results of canonical correlation analysis. To define canonical variates, use equation below [20]:

3

Equations (3) are to find the k-th pair of canonical variates.

### Canonical Loadings

Correlations between the original variables and the canonical variates, used to interpret the importance of each original variable in the canonical variates. To calculate the canonical loadings, use the equation below:

4

or

5

Equation (4) is to calculate the canonical loading of on and (5) is to calculate the canonical loading of on .

### Canonical Significant Test

There are two hypotheses that will be tested in the canonical correlation analysis, namely the joinly significance test and the partial significance test.

#### Joinly Significance Test

Wilks' Lambda is a test of how well a model can explain the variance between and within each set of variables in the joint covariance matrix between two sets of variables. Below are hypotheses of Wilks’ Lambda for canonical correlation analysis:

The significance of canonical correlations can be tested by [22]:

6

From the results obtained from the calculation of the Wilks' lambda test, h0 will be rejected if . Note that (6) can be expressed in the form of squared canonical correlation [22]:

7

Note that are calculated from where the size of r is .

#### Partial Significance Test

For the partial significance test, we can modify Wilks' lambda from the squared canonical correlation form [22]:

8

Canonical correlation analysis helps researchers identify patterns or structures that may exist between two sets of variables. It is commonly used in fields such as psychology, sociology, and economics to explore complex relationships between several types of data. By understanding these relationships, researchers can gain insights into underlying processes and make informed decisions.

## Assumption

### Linearity

The relationship between variables in each set should be linear. This means that changes in one set of variables should correspond to proportional changes in the other set. To test linearity, we can use Pearson’s correlation coefficient and then compute them using Fisher’s transformation. First, calculate the Pearson’s correlation coefficient [23]:

9

After calculating the Pearson’s correlation coefficient, now compute them using Fisher’s transformation [24]:

10

The p-value associated with the test statistic () is calculated based on the t-distribution with degrees of freedom. The hypothesis of linearity assumption as follow:

Then compate the p-value with alpha, if p-value then is rejected.

### Multivariate Normality

This assumption states that the variables in each set should follow a multivariate normal distribution. In practice, this can be assessed by testing the significance of each canonical function or by examining the normality of individual variables. To test multivariate normality, we can use Generalized Shapiro-Wilk (GWS) test, below is the formula for GSW [25]:

11

Note that is eigenvalues of covariance matrix. After calculating the test statistic (), obtain the p-value from the distribution of . The hypothesis of GWS as follows:

Then compate the p-value with alpha, if p-value then is rejected.

### Multicollinearity

Multicollinearity occurs when there are high correlations between variables within the same set. This can make it difficult to estimate the relationships between variables accurately. To test multicollinearity, we can use VIF. If the VIF value is greater than 10, then the variable is multicollinear. Below is the formula for VIF [26]:

12

## Yeo Johnson Transformation

The Yeo-Johnson transformation is a statistical method developed by Yeo and Johnson (2000) to modify data distributions in a versatile manner [27]. Unlike the Box-Cox transformation, which is limited to positive values and specific assumptions like normality, the Yeo-Johnson transformation accommodates datasets containing positive, negative, and zero values. This flexibility addresses the shortcomings of the Box-Cox method by introducing a parameter λ that allows for a wider range of data distributions to be normalized or adjusted towards normality. Here is the equation of the Yeo-Johnson transformation:

(13)

where denotes the original data value, symbolizes the transformation parameter that can cover any real number, including zero, and represents the data point after undergoing the transformation process.

# Result and Discussion

## Canonical Correlation Analysis

We conducted a canonical correlation analysis to explore the relationship between two groups of variables: demographic factors and access to basic services across Indonesian provinces. Group 1 includes Population Growth Rate per Year, Population Percentage, and Population Density per square km for 2023. Group 2 consists of Proportion of Households with Access to Drinking Water, Proportion of Households with Access to Basic Sanitation, and Proportion of Households with Access to Basic Health Services for 2023.

## Assumptions for Canonical Correlation

Before performing the canonical correlation analysis, it was essential to ensure that the data met the underlying assumptions of linearity, normality, and multicollinearity. The linearity assumption checks whether the relationships between the variables are linear, the normality assumption verifies that the data distribution follows a normal curve, and the multicollinearity assumption ensures that the variables are not too highly correlated with each other. First, we need to calculate the VIF values.

1. Multicollinearity Assumption (Pre-Transformation)

| Variable | VIF |
| --- | --- |
| *X*1 | 2.345280 |
| *X*2 | 1.301028 |
| *X*3 | 1.989194 |
|  | 1.310361 |
|  | 2.488587 |
|  | 2.407745 |

Table I shows that all VIF values of each variable are smaller than 10. This indicates that there are no multicollinear variables. Next, test the linearity assumption on each variable. The following are figures regarding the linearity assumption test:



1. Linearity Correlation Plot (Pre-Transformation)

Using = 0.05, Figure 1 shows that many pairs of variables do not meet the assumption of linearity or there is no linear correlation. This is indicated by the absence of the “\*\*\*” symbol which indicates that the variable pair has a linear correlation. Which means, there are only two pairs of variables that meet the assumption of linearity. Since there are only two pairs, we assume the linearity test is not satisfied. Next do the normal multivariate distribution test, the following are the results of the Shapiro-Wilk test:

1. Shapiro-Wilk test (Pre-Transformation)

| Variable | p-value |
| --- | --- |
| Group 1/ | < 2.2e-16 |
| Group 2/ | 0.000109 |

Using = 0.05, Table II shows that both groups do not meet the multivariate normal assumption test. This is because the p-value is smaller than alpha. Thus, the null hypothesis is rejected which proves that the groups are not multivariate normally distributed. Upon conducting preliminary analyses, we found that the assumptions of linearity and normality were not satisfied. Specifically, the linearity assumption was violated as indicated by the scatterplot analysis, and the normality assumption was violated as revealed by the Shapiro-Wilk test.

## Data Transformation

1. Shapiro-Wilk test (Post-Transformation)

| Transformation Type | p-value | |
| --- | --- | --- |
| Group 1/X | Group 2/Y |
| Logaritmic | 0.4374 | 3.822e-08 |
| Square Root | 1.627e-13 | 2.559e-06 |
| Box-Cox | 0.7787 | 0.0215 |
| Yeo-Johnson | 0.3552 | 0.881 |

To address the violations of assumptions, we applied various data transformation techniques. After experimenting with several transformations, including logarithmic, square root, and Box-Cox transformations, we found that with = 0.05, the Yeo-Johnson transformation was the most effective in normalizing the data. Table X. shows that The Yeo-Johnson transformation is the only transformation that passed the normality test. The Shapiro-Wilk test results indicated that both groups were multivariate normally distributed.

## Post-Transformation Analysis

After obtaining a method that can make the normality assumption of this data satisfied, by using the Yeo Johnson Transformation, we proceed to test the linearity and multicollinearity assumptions.

1. Multicollinearity Assumption (Post-Transformation)

| Variable | VIF |
| --- | --- |
| *X*1 | 1.810824 |
| *X*2 | 2.552850 |
| *X*3 | 3.469298 |
|  | 1.264566 |
|  | 1.954197 |
|  | 1.775094 |

Table IV shows that all VIF values of each variable are smaller than 10. This indicates that there are no multicollinear variables. Next, test the linearity assumption on each variable. The following are figures regarding the linearity assumption test:



1. Linearity Correlation Plot (Post-Transformation)

Using = 0.05, Figure 2 shows that many pairs of variables do not meet the linearity assumption or there is no linear correlation. However, compared to the previous linearity test, this result shows that more pairs of variables have a linear correlation. That is, there are 10 pairs of variables that correlated linearly. Thus, we assume that the linearity test is satisfied.

## Canonical Correlation Results

The canonical correlation analysis revealed a significant relationship between the two groups of variables. The first canonical correlation coefficient was , indicating a strong relationship between the demographic factors and access to basic services. The results suggest that as population dynamics change, there are corresponding variations in the accessibility of essential services.

Further interpretation will be conducted on the first canonical function as it is jointly and partially significant. The table below presents coefficients of the canonical weights of independent variables in the first function along with the linear combination equation of these canonical weights.

1. Canonical Weights For Independent Variables

| Variable | Canonical Weights |
| --- | --- |
| *X*1 | 0.06467549 |
| *X*2 | 0.12229054 |
| *X*3 | -0.17921883 |

(14)

Based on the canonical weight coefficients in Table V, it can be concluded that the order of independent variables contributing most to least to their canonical variable is , , and . This implies that within the demographic factors group, the population density per square kilometer (km²) is the variable most contributing to the relationship between demographic conditions and access to basic services, with a coefficient of -0.17921883. The negative value in the weight indicates an inverse relationship with the generated canonical component, suggesting that as population density increases, access to basic services may tend to decrease. Furthermore, the demographic factors variables contributing to the relationship are population percentage and population growth rate per year.

Next, the table below presents coefficients of the canonical weights of dependent variables in the first function along with the linear combination equation of these canonical weights.

1. Canonical Weights For Dependent Variables

| Variable | Canonical Weights |
| --- | --- |
| *Y*1 | -0.02378430 |
| *Y*2 | -0.11041849 |
| *Y*3 | -0.08040095 |

(x)

Based on the canonical weight coefficients in Table VI, it can be concluded that the order of dependent variables contributing most to least to their canonical variable is , , and . This means that within the access to basic services group, the proportion of households with access to basic sanitation contributes most significantly to the relationship between demographic conditions and access to basic services, with a coefficient of -0.11041849. The negative value in the weight indicates an inverse relationship with the generated canonical component, suggesting that as access to basic sanitation increases, there is a tendency for the canonical component related to access to basic services to decrease. Furthermore, the access to basic services variables contributing to the relationship are the proportion of households with access to basic health services and the proportion of households with access to clean drinking water.

Further discussion will focus on the canonical loadings of the first canonical function. The table below presents coefficients of the canonical loadings of dependent variables in the first function.

1. Canonical Loadings For Independent Variables

| Variable | Canonical Loadings |
| --- | --- |
| *X*1 | -0.7006847 |
| *X*2 | 0.1297983 |
| *X*3 | 0.8070225 |

From the coefficients of the canonical loadings in Table VII, it can be seen that for independent variables, , which is population density per square kilometer (km²), has the highest loading of 0.8070225. This indicates that is the most influential variable and has a strong relationship in forming . Additionally, the population growth rate per year shows a fairly strong negative correlation of -0.7006847, meaning that as the population growth rate increases, the value of tends to decrease. Meanwhile, the population percentage shows a weak positive correlation of 0.1297983.

Next, the table below presents coefficients of the canonical loadings of dependent variables in the first function.

1. Canonical Loadings For Dependent Variables

| Variable | Canonical Loadings |
| --- | --- |
| *Y*1 | 0.2404717 |
| *Y*2 | 0.9086484 |
| *Y*3 | 0.8460922 |

From the coefficients of the canonical loadings in Table VII, it can be seen that for dependent variables, , which is the proportion of households with access to basic sanitation, has the highest loading of 0.9086484. This indicates that is the most influential variable and has a strong relationship in forming . Additionally, the proportion of households with access to basic health services shows a strong positive correlation of 0.8460922, meaning that as access to basic health increases, the value of tends to increase. Meanwhile, the proportion of households with access to clean drinking water shows a weak positive correlation of 0.2404717.

In addition to canonical loadings, there are also cross-canonical loadings. The table below presents coefficients of the cross-canonical loadings of independent variables in the first function.

1. Canonical Cross Loadings For Independent Variables

| Variable | Canonical Cross Loadings |
| --- | --- |
| *X*1 | -0.50977552 |
| *X*2 | 0.09443332 |
| *X*3 | 0.58714038 |

From the coefficients of the cross-canonical loadings in Table IX, for independent variables, , which is population density per square kilometer (km²), has the highest cross-loading of 0.58714038. This indicates that is the most influential variable and has a fairly strong relationship with the first canonical component of the dependent variable (). This means that as population density increases, the value of the first canonical function of the dependent variable also tends to increase. Additionally, the population growth rate per year shows a fairly strong negative correlation of -0.50977552, meaning that as the population growth rate increases, the value of tends to decrease. Meanwhile, the population percentage shows a very weak positive correlation of 0.09443332.

Next, the table below presents coefficients of the cross-canonical loadings of dependent variables in the first function.

1. Canonical Cross Loadings For dependent Variables

| Variable | Canonical Cross Loadings |
| --- | --- |
| *Y*1 | 0.1749525 |
| *Y*2 | 0.6610772 |
| *Y*3 | 0.6155651 |

From the coefficients of the cross-canonical loadings in Table X, it can be seen that for dependent variables, , which is the proportion of households with access to basic sanitation, has the highest cross-loading of 0.6610772. This indicates that is the most influential variable and has a fairly strong relationship with the first canonical component of the independent variable (). This means that as access to basic sanitation increases, the value of the first canonical function of the independent variable also tends to increase. Additionally, the proportion of households with access to basic health services shows a fairly strong positive correlation of 0.6155651, meaning that as access to basic health services increases, the value of tends to increase. Meanwhile, the proportion of households with access to clean drinking water shows a weak positive correlation of 0.1749525.

# Conclusion

In this study, canonical correlation analysis was used to examine the relationship between the group of demographic factor variables and the group of basic services access variables. Before testing, the data was first transformed using the Yeo-Johnson transformation because the data did not meet the assumption of multivariate normal distribution. After the transformation, it was ensured that there was no multicollinearity in the data and that many pairs of variables were linearly correlated. Once all assumptions were confirmed, the testing proceeded, resulting in three canonical functions. However, significance testing, both jointly and partially, revealed that only the first canonical function was significant, with a canonical correlation value of 0.72753905. This value indicates a strong relationship between the group of demographic factor variables and the group of basic services access variables.

Based on the analysis of canonical weights, loadings, and cross-loadings, it can be concluded that for the independent variables in the first canonical function, the variable population density per square kilometer (km²) has the greatest and most significant contribution in linking demographic conditions with access to basic services. The negative weight on this variable indicates an inverse relationship, where an increase in population density tends to reduce access to basic services. This is supported by the high canonical loading (0.8070225) and cross-loading (0.58714038), indicating that this variable is highly influential in forming the first canonical component for both the independent and dependent variables.

For the dependent variables in the first canonical function, it can be concluded that the variable proportion of households with access to basic sanitation has the greatest and most significant contribution in linking household access to basic facilities with demographic conditions. The negative weight on this variable indicates an inverse relationship, where an increase in access to basic sanitation is associated with a decline in the measured demographic conditions. This is reinforced by the high canonical loading (0.9086484) and cross-loading (0.6610772), indicating that this variable is highly influential in forming the first canonical component for both the independent and dependent variables.

##### References

[1] UN, “The 17 Sustainable Development Goals,” United Nation. Accessed: Jun. 25, 2024. [Online]. Available: https://sdgs.un.org/goals

[2] M. W. Massuanna, F. Malinda, R. A. Syafaririn, and W. M. Alam, “KEPADATAN PENDUDUK DI PERKOTAAN: DAMPAK TERHADAP KUALITAS LINGKUNGAN MASYARAKAT,” *Jurnal Ilmiah Multidisiplin Terpadu*, vol. 8, no. 5, 2024.

[3] U. Handoyo, “Pengaruh Kepadatan Penduduk Terhadap Indeks Pembangunan Manusia Di Indonesia,” *Journal of Nusantara Economic Science*, vol. 1, no. 1, pp. 19–23, 2023.

[4] R. R. Mahani, “ANALISA KORELASI KANONIK UNIT INFRASTRUKTUR DENGAN KUALITAS SUMBER DAYA MANUSIA (SDM) DI INDONESIA PERIODE TAHUN 2017-2021.” 2021.

[5] A. Sofia, “Pengelompokkan Kesejahteraan Berdasarkan Faktor-Faktor Sosial Ekonomi Terhadap Klasifikasi Indeks Pembangunan Masyarakat di Provinsi Aceh Menggunakan Analisis Diskriminan,” *Indonesian Journal of Applied Mathematics*, vol. 4, no. 1, pp. 11–18, 2024.

[6] BPS, “Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk, Kepadatan Penduduk, Rasiu Jenis Kelamin Penduduk menurut Provinsi.” Jul. 2024. [Online]. Available: https://www.bps.go.id/id/statistics-table/3/V1ZSbFRUY3lTbFpEYTNsVWNGcDZjek53YkhsNFFUMDkjMw==/penduduk–laju-pertumbuhan-penduduk–distribusi-persentase-penduduk–kepadatan-penduduk–rasio-jenis-kelamin-penduduk-menurut-provinsi.html?year=2023

[7] BPS, “Proporsi Rumah Tangga dengan Akses terhadap Pelayanan Dasar menurut Provinsi.” Jul. 2024. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MjAxNiMy/proporsi-rumah-tangga-dengan-akses-terhadap-pelayanan-dasar-menurut-provinsi–persen-.html

[8] E. Penno, R. Gauld, and R. Audas, “How are population-based funding formulae for healthcare composed? A comparative analysis of seven models,” *BMC Health Serv Res*, vol. 13, no. 1, p. 470, Dec. 2013, doi: 10.1186/1472-6963-13-470.

[9] J. C. Ohls and T. J. Wales, “Supply and Demand for State and Local Services,” *Rev Econ Stat*, vol. 54, no. 4, p. 424, Nov. 1972, doi: 10.2307/1924569.

[10] M. Macalou, S. I. Keita, A. B. Coulibaly, and A. K. Diamoutene, “Urbanization and food security: evidence from Mali,” *Front Sustain Food Syst*, vol. 7, Jul. 2023, doi: 10.3389/fsufs.2023.1168181.

[11] M. Sang, J. Jiang, X. Huang, F. Zhu, and Q. Wang, “Spatial and temporal changes in population distribution and population projection at county level in China,” *Humanit Soc Sci Commun*, vol. 11, no. 1, p. 288, Feb. 2024, doi: 10.1057/s41599-024-02784-1.

[12] A. Al-kassab-Córdova, C. Silva-Perez, P. Robles-Valcarcel, G. Bendezu-Quispe, A. I. Ortiz, and V. A. Benites-Zapata, “Spatial and socioeconomic inequalities in the access to safe drinking water in Peruvian households,” *J Water Health*, vol. 21, no. 4, pp. 525–535, Apr. 2023, doi: 10.2166/wh.2023.316.

[13] S. Kant, R. Kaur, A. Lohiya, F. Ahamed, S. Malhotra, and P. Haldar, “Access and utilization of sanitation facilities in a Rural Area of Haryana, North India,” *Indian J Public Health*, vol. 64, no. 4, p. 357, 2020, doi: 10.4103/ijph.IJPH\_416\_19.

[14] R. E. Narita, “Consumption of Healthcare Services in the United States: The Impact of Health Insurance,” *Journal of Risk and Financial Management*, vol. 16, no. 5, p. 277, May 2023, doi: 10.3390/jrfm16050277.

[15] P. O. Otieno *et al.*, “Access to primary healthcare services and associated factors in urban slums in Nairobi-Kenya,” *BMC Public Health*, vol. 20, no. 1, p. 981, Dec. 2020, doi: 10.1186/s12889-020-09106-5.

[16] V. Uurtio, J. M. Monteiro, J. Kandola, J. Shawe-Taylor, D. Fernandez-Reyes, and J. Rousu, “A Tutorial on Canonical Correlation Methods,” Nov. 2017, doi: 10.1145/3136624.

[17] B. Thompson, *Canonical correlation analysis: Uses and interpretation*, vol. 47. Sage, 1984.

[18] D. Weenink, “Canonical correlation analysis,” in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Amsterdam: University of Amsterdam, Aug. 2003, pp. 81–99.

[19] M. F. Asbah, S. Sudarno, and Safitri D., “PENENTUAN KOEFISIEN KORELASI KANONIK DAN INTERPRETASI FUNGSI KANONIK MULTIVARIAT,” *Jurnal Gaussian*, vol. 2, no. 2, pp. 119–128, Jun. 2013.

[20] R. A. Johnson and D. W. Winchern, *Applied multivariate statistical analysis.*, 6th ed. 2002.

[21] A. C. Rencher, “Interpretation of Canonical Discriminant Functions, Canonical Variates, and Principal Components,” *Am Stat*, vol. 46, no. 3, pp. 217–225, Aug. 1992, doi: 10.1080/00031305.1992.10475889.

[22] A. C. Rencher and M. G. Schimek, *Methods of multivariate analysis*, 3rd ed., vol. 12. John Wiley & Sons, Inc: John Wiley & Sons, Inc, 1997.

[23] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, “Effective use of Pearson’s product–moment correlation coefficient,” *Anim Behav*, vol. 93, pp. 183–189, Jul. 2014, doi: 10.1016/j.anbehav.2014.05.003.

[24] R. F. Bartlett, “Linear Modelling of Pearson’s Product Moment Correlation Coefficient: An Application of Fisher’s z-Transformation,” *The Statistician*, vol. 42, no. 1, p. 45, 1993, doi: 10.2307/2348110.

[25] E. González-Estrada and W. Cosmes, “Shapiro–Wilk test for skew normal distributions based on data transformations,” *J Stat Comput Simul*, vol. 89, no. 17, pp. 3258–3272, Nov. 2019, doi: 10.1080/00949655.2019.1658763.

[26] A. Alin, “Multicollinearity,” *WIREs Computational Statistics*, vol. 2, no. 3, pp. 370–374, May 2010, doi: 10.1002/wics.84.

[27] L. Yang, T. Nguyen-Thoi, and T. T. Tran, “Predicting the friction angle of clays using a multi-layer perceptron neural network enhanced by yeo-johnson transformation and coral reefs optimization,” *Journal of Rock Mechanics and Geotechnical Engineering*, 2024.