E-ISSN : 2686-6331
P-ISSN : 2686-6358
DINASTI INTERNATIONAL JOURNAL OF EDUCATION MANAGEMENT AND SOCIAL SCIENCE (DIJEMSS)
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455

DOI: https://doi.org/10.38035/dijemss.v5i5
Received: 15 June 2024, **Revised:** 26 June 2024, **Publish:** 07 July 2024

Distribution of Endemic Diseases Using Correspondence Analysis: Case Study of Endemic Diseases in Bali Province

Asa Vira Safitri[1*], I Komang Gde Sukarsa[2], Eka N Kencana[3]
[1]Udayana University, Bali, Indonesia, email: asavirasftr@gmail.com
[2]Udayana University, Bali, Indonesia, email: gedesukarsa@unud.ac.id
[3]Udayana University, Bali, Indonesia, email: i.putu.enk@unud.ac.id
[2,3]Social Statistics and Tourism Research Group, Udayana University

Corresponding Author: i.putu.enk@unud.ac.id

Abstract: One of the health problems faced by the Indonesian people and related stakeholders is the potential for extraordinary events (KLB) caused by the spread of endemic diseases, including pulmonary tuberculosis (TB), diarrhea, dengue hemorrhagic fever (DHF), and malaria. As infectious diseases, these four diseases are classified as endemic considering that their potential for emergence is closely related to the topography of an area. This research aims to elaborate and study the relationship between tuberculosis, diarrhea, dengue hemorrhagic fever, and malaria with the topography of 57 sub-districts in Bali Province. Using secondary data from 9 city district health offices in Bali regarding the number of each disease in 2022, correspondence analysis was applied. The topography of the 57 sub-districts in Bali Province is classified into 4 altitude groups, namely lowland, low-medium land, medium-highland, and low-highland. The contingency table as the main data for analysis is a matrix of size 4×4. The results of the analysis show that TB has the potential to occur in sub-districts regardless of altitude, diarrhea is observed in all four topographies, DHF tends to occur in the lowlands, and malaria has no tendency towards topography in Bali Province.

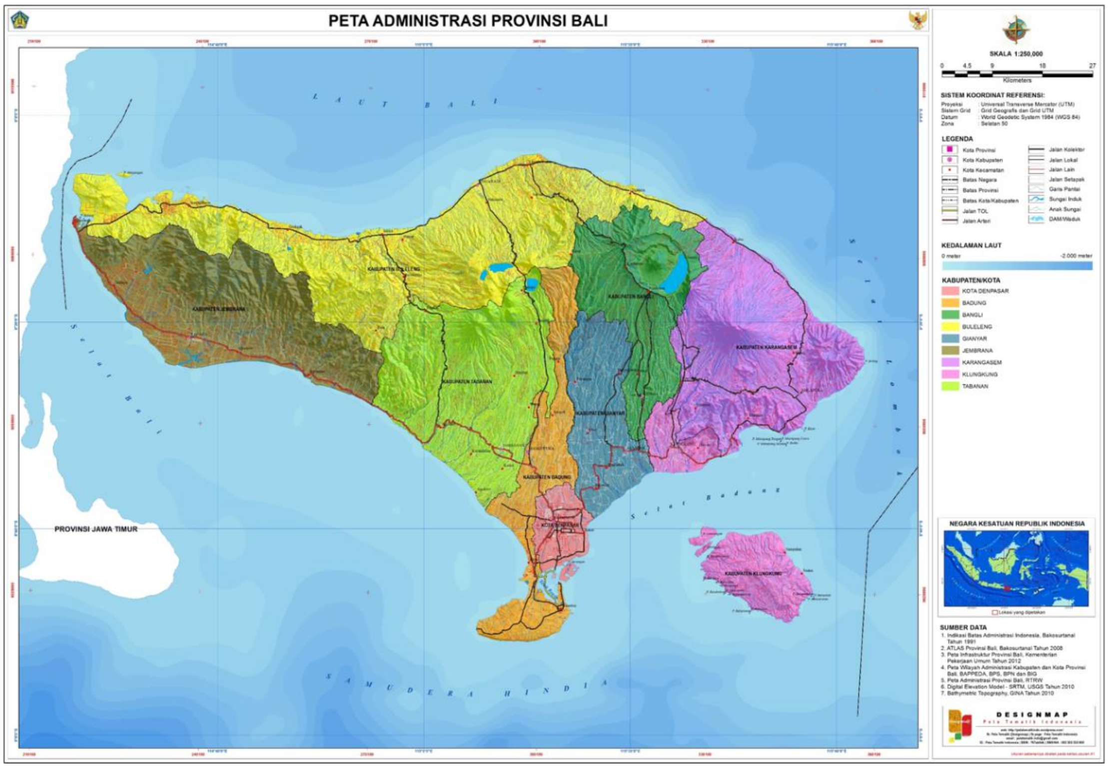
Keyword: Bali, Correspondence Analysis, Endemic Disease, Sub-District, Topography

INTRODUCTION

As the part of the multivariate analysis technique, correspondence analysis is a method of analyzing independence in the classification of statistical analysis methods (Hair et al., 2019). According to Greenacre (2017), correspondence analysis is a technique aimed at representing the rows and columns of a two-way contingency table as a set of points in a low-dimensional vector space. Utilizing visualization in $\mathfrak{R}^2$ with each axis representing dimensions explaining the two largest principal components formed from the categories of observed variables, each category is plotted with the abscissa and ordinate values representing the values on dimension I and dimension II (Johnson & Wichern, 2014). As a multivariate analysis technique, correspondence analysis has been applied in various research domains such as in the field of health (Greenacre, 1992), urban planning (Ali et al., 2018), and tourism (Calantone et al., 1989; Tsujioka et al., 2020). This research aims to elaborate on and study the correspondence patterns between endemic diseases, infectious diseases found in specific populations or regions whose distribution is related to geographical and topographical factors (Kalra et al., 2015), where the topography of the region is one of the environmental geographical factors that describes the elevation of certain areas (Lu'lu'a et al., 2023).

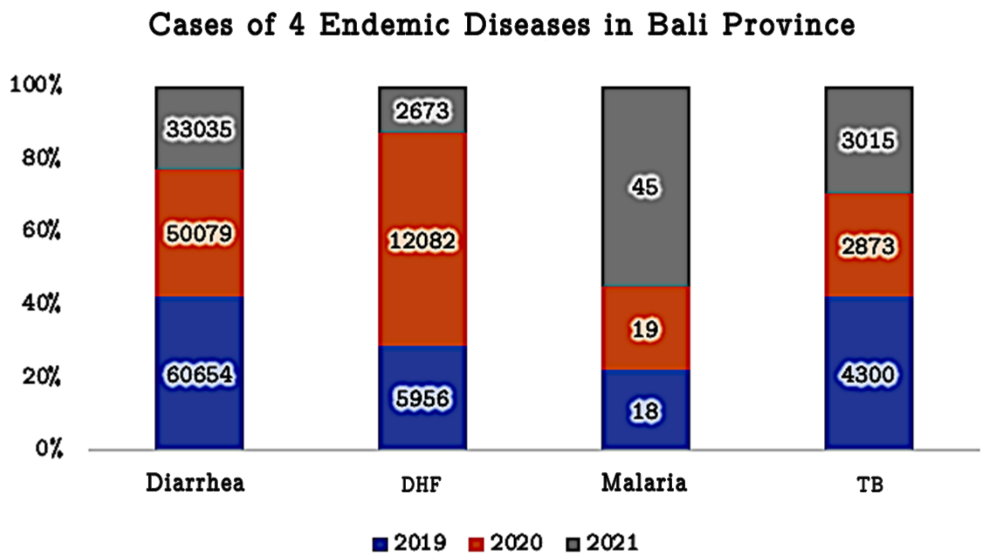
The Province of Bali, as 1 of the 38 provinces in Indonesia, is recorded to have an area of 5,590.15 km^2 with a total of 34 registered islands. The administrative region of the Province of Bali consists of 9 regencies and municipalities divided into 57 sub-districts. Tabanan Regency is noted to have 10 sub-districts and is the regency with the largest number of sub-districts in Bali, followed by Buleleng Regency with 9 sub-districts. The Central Statistics Agency (BPS) distinguishes the elevation of a region into 3 types of terrains: (a) lowland with an elevation of 0-400 m; (b) midland with an elevation of 401-700 m; and (c) highland with an elevation of more than 700 m above sea level (asl). The topography of sub-districts in the Province of Bali is highly variative (Figure 1), ranging from low-lying areas such as the 4 sub-districts in Denpasar City to stretching from lowlands to highlands such as the Kubu Sub-district in Karangasem Regency. Observing Figure 1, sub-districts in Bali can be grouped into 4 types of terrains: (a) lowland; (b) low-medium land; (c) medium-highland; and (d) low-highland.

Regarding with the endemic diseases in this province, pulmonary tuberculosis (TB), diarrhea, dengue hemorrhagic fever (DHF), and malaria are four endemic diseases that often became health problems for the population. Data from the Health Office of Bali Province shows that the number of cases of these four endemic diseases fluctuated during the period from 2019 to 2021 (Figure 2). In 2019 to 2021, diarrhea was the endemic disease with the highest number of cases, followed by DHF and TB. Malaria was also found, albeit with a relatively very small number of cases compared to the other three types of endemic diseases.



Source: analyzed data (2024)

Figure 1. Administrative Map of the 9 Regencies and Municipalities in the Province of Bali



Source: analyzed data (2024)

Figure 2. Number of Cases of Four Endemic Diseases in the Years 2019-2021

Considering the importance of public health development and Bali's position as a favorite tourist destination for both domestic and international tourists, which is highly vulnerable to various issues potentially disrupting the safety and comfort of tourism, one of the issues is the spread of endemic diseases in Bali. This research aims to examine the correlation between the region's topography and the number of cases of endemic diseases in Bali.

METHOD

Type and Data Source

This research utilizes data sourced from the health department of regencies and cities in Bali Province regarding the number of cases of TB, diarrhea, DHF, and malaria in each sub-district of every regency and city in 2022.

Research Variables

There are 2 variables involved in this research. Firstly, the type of endemic disease is a nominal scale variable with possible values of TB, diarrhea, DHF, and malaria. Secondly, the topography of sub-district areas is an ordinal scale variable with possible values of lowland (DR), low-medium land (DRS), medium-highland (DST), and low-highland (DRT).

Data Analysis Steps

The examination of the association between endemic disease types and sub-district topography using correspondence analysis is conducted following these steps:

1. Building a contingency table, cross-tabulation of two or more categorical variables with table elements being observed frequencies (Johnson & Wichern, 2014);
2. Examining the formed association between variables using the χ^2 test. The pair of hypotheses tested in this test is to determine whether ρ, the coefficient of association between the first variable and the second variable, is significant ($\rho \neq 0$) or not:

   H_0 : $\rho = 0$

   H_1 : $\rho \neq 0$

   H_0 cannot be accepted if the value $\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-e_{ij})^2}{e_{ij}^2}$ is smaller than $\chi^2_{\frac{\alpha}{2}(I-1)(J-1)}$ at the chosen significance level of test (α) with I, J representing the number of categories of the first and second variables, respectively, and n_{ij}, as well as $e_{ij} = \frac{n_i n_j}{n}$, representing the observed frequency and expected value of cell – i, j. Correspondence analysis can be conducted if H_0 is rejected, indicating an association between categories in both variables;

3. Constructing the correspondence matrix $\mathbf{P} = \frac{1}{n}\mathbf{K}$ with $\mathbf{K}$ representing the contingency table from step (1);
4. Calculating the row profile matrix R and column profile matrix C using equations (1) and (2):

$$\mathbf{R} = D_r^{-1} P \quad (1)$$
$$\mathbf{C} = D_c^{-1} P \quad (2)$$

In equations (1) and (2), D_r and D_c represent square matrices of dimensions I and J, respectively, with matrix elements on the off-diagonal being 0, as follows:

$$D_r = \begin{pmatrix} p_{1.} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & p_{i.} \end{pmatrix} \text{ and } D_c = \begin{pmatrix} p_{.1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & p_{.j} \end{pmatrix}$$

In D_r and D_c, $p_{i.}$ represents the total of row i from **P**, and $p_{.j}$ represents the total of column j from **P**, with $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$. The diagonal elements of both matrices are usually expressed as $\boldsymbol{r} = \mathbf{P}\underline{\mathbf{1}} = (p_{1.}, \ldots, p_{I.})$ and $\boldsymbol{c} = \mathbf{P}^{\mathbf{T}}\underline{\mathbf{1}} = (p_{.1}, \ldots, p_{.J})$;

5. Calculating the row profile matrix R and column profile matrix C using equations (1) and (2):
6. Determining the row and column profile coordinates is done through singular value decomposition (Greenacre, 2017; Johnson & Wichern, 2014). The coordinates of both profiles are calculated so that the analysis results can be visualized in $\mathfrak{R}^2$ or $\mathfrak{R}^3$.
7. Interpreting the obtained plot.

RESULTS AND DISCUSSION

The data on endemic disease cases of TB, diarrhea, DHF, and malaria in 2022 by sub-district in Bali Province ($\mathbf{X}_{57\times5}$) are summarized in a contingency table sized 4 × 4, as shown in Table 1.

Table 1. Number of Endemic Disease Cases versus Sub-district Topography in Bali Province

Sub-district Topography		**Disease Type**				**Total**
		TB	Diarrhea	DHF	Malaria	
Lowland	(DR)	1 256	19 681	2 947	19	23 903
Low-Medium Land	(DRS)	647	9 806	1 405	4	11 862
Medium-Highland	(DST)	108	3 716	107	0	3 931
Low-Highland	(DRT)	1 122	11 180	1 367	9	13 678
Total		3 133	44 383	5 826	32	53 374

Source: data analyzed (2024)

Table 1 shows the sub-districts in Bali Province with lowland topography (16 sub-districts) as areas with the highest number of disease cases, accounting for 45 percent of the total endemic disease occurrences in 2022, recorded at 53,374 cases. This is followed by sub-districts with low-highland topography (21 sub-districts) at 26 percent. The lowest number of disease cases was found in sub-districts with medium-highland topography (6 sub-districts) at 7 percent. In terms of disease types, diarrhea and DHF rank first and second, respectively, accounting for 83 percent and 11 percent.

A test for the significance of the association between endemic disease types and sub-district topography was conducted using the χ^2 test with the test result indicating $\chi^2_{\text{Hitung}} = 588.64$; exceeding the critical value $\chi^2_{(0.05)(9)} = 16.92$. Thus, H_0 stating no significant association between endemic disease types and topography is rejected at the 5 percent significance level. Considering the conclusion from this independence analysis, contingency analysis can be continued. The analysis was performed using **R 4.2.2** software (R Core Team, 2007) with the *package* `ca` (Nenadic & Greenacre, 2007) and a biplot of row profile coordinates with column profiles was formed using the `fvis_biplot()` function in the *package* `factoextra` (Kassambara & Mundt, 2020).

Using the contingency table in Table 1, a correspondence matrix is formed between sub-district topography and disease types $\mathbf{P} = \frac{1}{n}\mathbf{K}$ as follows:

$$\mathbf{P} = \frac{1}{53347}\begin{pmatrix} 1256 & 19681 & 2947 & 19 \\ 647 & 9806 & 1405 & 4 \\ 108 & 3716 & 107 & 0 \\ 1122 & 11180 & 1367 & 9 \end{pmatrix} = \begin{pmatrix} 0.023 & 0.369 & 0.055 & 0.0004 \\ 0.012 & 0.184 & 0.026 & 0.0001 \\ 0.002 & 0.070 & 0.002 & 0.0000 \\ 0.021 & 0.209 & 0.026 & 0.0002 \end{pmatrix}$$

Using equations (1) and (2), the row profiles and column profiles of **P** are stated in Tables 2 and 3.

Table 2. Row Profile of Endemic Disease in Bali Province

Topography	Type of Disease				Mass	Inertia
	TB	Diarrhea	DHF	Malaria		
DR	0.053	0.823	0.123	0.001	0.448	0.001
DRS	0.055	0.827	0.118	0.000	0.222	0.000
DST	0.027	0.945	0.027	0.000	0.074	0.007
DRT	0.082	0.817	0.100	0.001	0.256	0.003

Source: data analyzed (2024)

Table 3. Column Profile of Endemic Diseases in Bali Province

Topography	Type of Disease			
	TB	Diarrhea	DHF	Malaria
DR	0.401	0.207	0.034	0.358
DRS	0.443	0.221	0.084	0.252
DST	0.506	0.241	0.018	0.235
DRT	0.594	0.125	0.000	0.281
Mass	0.059	0.832	0.109	0.000
Inertia	0.004	0.001	0.006	0.000

Source: data analyzed (2024)

The row profile in Table 2 indicates that sub-districts with lowland topography (DR) have the highest endemic disease mass index compared to the other three topography categories. Sub-districts with low-highland topography (DRT) rank second. This is presumably due to the number of endemic disease cases occurring in villages within their administrative areas classified as lowland topography. The column profile in Table 3 shows that diarrhea is the endemic disease with the highest mass index, approximately 8 times higher than DHF which ranks second. Table 3 justifies that the types of endemic diseases with the highest number of cases in the 57 sub-districts in Bali Province, regardless of the topography of each sub-district, are diarrhea and DHF.

The final step before interpreting the results is to determine the coordinates of each disease category and topography. Visualization is done by creating a biplot in $\mathfrak{R}^2$ with the abscissa and ordinate representing dimension I and dimension II respectively. These are the two dimensions with the largest proportion of explained variance from the three possible eigenvalues formed by the minimum values of (r-1) and (c-1), where r and c represent the number of row and column categories. The results of the singular value decomposition are shown in Table 4.

Table 4. Eigenvalues and Proportion of Explained Variance

Dimension	Eigenvalue	Proportion of Variance (%)	Cumulative Proportion (%)
I	7.515×10^{-3}	68.14	68.14
II	3.467×10^{-3}	31.43	99.57
III	4.713×10^{-5}	0.43	100.00
Total	10.987×10^{-3}	100.00	

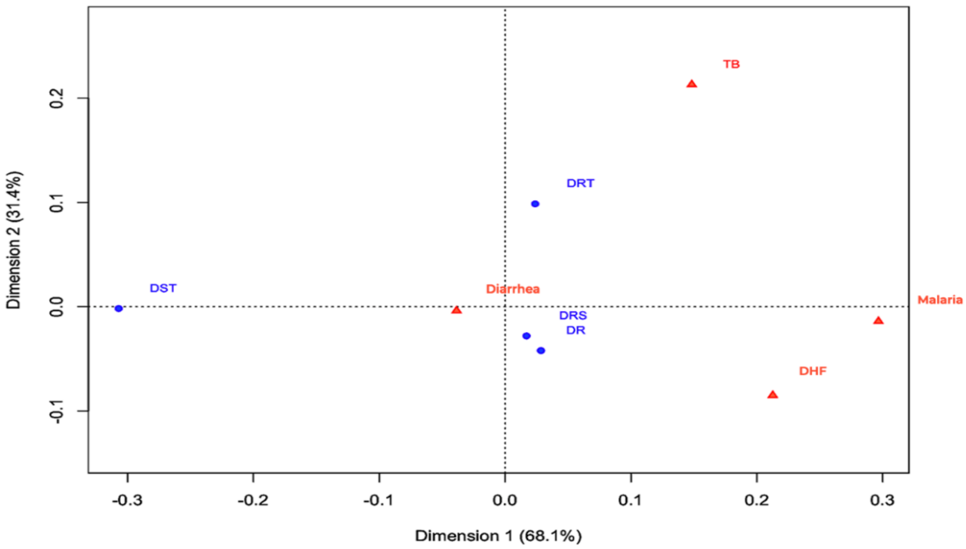
Source: data analyzed (2024)

In Table 4, the first two dimensions are able to explain the data variance almost perfectly at 99.6 percent. Thus, the visualization of rows (sub-district topography) versus endemic disease types using dimension I and dimension II is highly appropriate, with a 'loss' of information from data variance of only 0.4 percent. Table 5 shows the coordinates of each disease and topography category. The biplot is shown in Figure 3.

Table 5. Disease and Topography Category Coordinate

Variable	Category	Dimension I	Dimension II
Sub-district Topography	DR	0.028	-0.042
	DRS	0.017	-0.028
	DST	-0.307	-0.002
	DRT	0.024	0.099
Endemic Disease Type	TB	0.148	0.213
	Diarrhea	-0.039	-0.004
	DHF	0.213	-0.085
	Malaria	0.297	-0.014

Source: data analyzed (2024)



Source: data analyzed (2024)

Figure 3. Biplot of Topography versus Type of Disease

Figure 3 plots shows that TB is an endemic disease with the potential to occur in sub-districts regardless of altitude, from the lowlands to highlands. A study conducted by Kristini

& Hamidah (2020) on the potential transmission of pulmonary tuberculosis (TB) among family members of patients found that intense social interaction between patients and their family members (exceeding 8 hours/day) is one of the determinants of someone infected. Besides social interaction, another cause of TB is the type of occupation of the patient. In one of Indonesia's provinces, where the population's livelihood is in the agricultural sector, non-organic farming activities using various chemical elements have also observed TB cases in sub-districts where the residents work as farmers. This is consistent with a study by Diana (2019), which mentioned that 1 in 5 TB patients in Sekayam Sub-district, Senggau Regency, West Kalimantan Province, was exposed to *Mycobacterium tuberculosis* bacteria as an indirect result of plant spraying activities using chemical substances, this led to decreased body immunity and increased risk of respiratory infectious diseases.

In Figure 3, it is also evident that diarrhea is an endemic disease with the shortest distance to all four topographies of sub-districts in Bali. Diarrhea is recorded as the endemic disease with the highest number of cases in each sub-district's respective topography. The disease with the next highest number of cases after diarrhea is DHF. Paomey *et al*. (2019), in their research in the Malalayang Sub-district, Manado City, concluded that there is a correlation between DHF cases and the region's topography. Sub-districts with higher elevation tend to have lower DHF cases compared to other sub-districts. The last endemic disease examined in this research is malaria. Malaria is an endemic disease with the smallest number of cases in the four topographies of sub-districts in Bali. In fact, in sub-districts classified as medium-highland topography (DST), there were no recorded malaria cases in 2022.

CONCLUSION

Conclusion

The study on the correlation or association between regional topography and the number of endemic disease cases in Bali including TB, diarrhea, DHF, and malaria concluded: (a) the two extracted dimensions can explain 99.6 percent of the data variance; (b) TB and diarrhea cases were observed in all four topographical sub-districts in the Province of Bali; and (c) DHF cases tend to occur in sub-districts with lowland topography.

Suggestion

As a preventive measure against the increasing cases of the 4 endemic diseases in Bali, it is recommended for relevant parties (health departments, community health centers, and sub-health centers) in each regency and municipality to enhance the dissemination of clean and healthy living behaviors, with a focus on emphasizing the types of dominant endemic diseases according to the region's topography. In the realm of correspondence analysis application, it is advised for further research to include other variables such as population density and level of education with the applied analysis being multiple correspondence analysis.

REFERENCES

Ali, F., Dissanayake, D., Bell, M., & Farrow, M. (2018). Investigating car users' attitudes to climate change using multiple correspondence analysis. *Journal of Transport Geography*, *72*(February 2017), 237–247. https://doi.org/10.1016/j.jtrangeo.2018.09.007

Calantone, R. J., Di Benedetto, C. A., Hakam, A., & Bojanic, D. C. (1989). Multiple multinational tourism positioning using correspondence analysis. *Journal of Travel Research*, *28*(2), 25–32. https://doi.org/10.1177/004728758902800207

Diana, E. (2019). *Gambaran Kasus TB Pada Petani Di Wilayah Kecamatan Sekayam*. Universitas Muhammadiyah Pontianak.

Greenacre, M. (1992). Correspondence Analysis in Medical Research. *Statistical Methods in Medical Research*, *1*(1), 97–117. https://doi.org/10.1177/096228029200100106

Greenacre, M. (2017). *Correspondence Analysis in Practices* (3rd ed.). CRC Press.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (Eighth ed). Cengage Learning.
Johnson, R. A., & Wichern, D. W. and others. (2014). Applied Multivariate Statistical Analysis. In *British Library Cataloguing-in-Publication Data* (Sixth). Pearson.
Kalra, S., Kumar, A., Jarhyan, P., & Unnikrishnan, A. G. (2015). Endemic or epidemic? Measuring the endemicity index of diabetes. *Indian Journal of Endocrinology and Metabolism*, *19*(1), 5–7. https://doi.org/10.4103/2230-8210.144633
Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (R package version 1.0.7).
Kristini, T., & Hamidah, R. (2020). Potensi Penularan Tuberculosis Paru pada Anggota Keluarga Penderita. *Jurnal Kesehatan Masyarakat Indonesia*, *15*(1), 24. https://doi.org/10.26714/jkmi.15.1.2020.24-28
Lu'lu'a, N., Amalia, I. N. I., Nayumi, A. P., Jumlat, M. W., Amalia, D., Mardianto, M. F. F., & Ana, E. (2023). Pola Kecenderungan Penyakit Menular Terhadap Topografi Kabupaten/Kota di Jawa Timur Menggunakan Analisis Korespondensi. *Jurnal Sains Matematika Dan Statistika*, *9*(1), 91. https://doi.org/10.24014/jsms.v9i1.20853
Nenadic, O., & Greenacre, M. (2007). Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software*, *20*(3), 1–13.
Paomey, V. C., Nelwan, J. E., & Kaunang, W. P. J. (2019). Sebaran Penyakit Demam Berdarah Dengue Berdasarkan Ketinggian dan Kepadatan Penduduk di Kecamatan Malalayang Kota Manado Tahun 2019. *Kesmas*, *8*(6), 521–527.
R Core Team. (2007). *R: A Language and Environment for Statistical Computing* (4.0.3). R Foundation for Statistical Computing.
Tsujioka, S., Watanabe, K., & Tsukamoto, A. (2020). Tourism Analysis Using User-Generated Content: A Case Study of Foreign Tourists Visiting Japan on TripAdvisor. *Tourism and Sustainable Development Review*, *1*(1), 57–64. https://doi.org/10.31098/ tsdr.v1i1.9