Multiple Correspondence Analysis for Identifying the Contribution of Infant Mortality Indicator Categories

Titi Purwandari¹, Irlandia Ginanjar¹, Dhita Diana Dewi¹

¹Department of Statistics, Universitas Padjadjaran, Indonesia.

Corresponding Author : titipurwandari@yahoo.com

Abstract. Infant mortality rate is one of several other indicators that are very susceptible in order to illustrate the mark of public health problems. This indicator is substantial to know the cause of the deaths as well as how success the maternal and child health program. This study aims to determine the characteristics of the causes of infant deaths based on age groups and sub-districts in the city of Bandung using multiple correspondence analysis. The results of the study concluded that asphyxia, early neonatal age groups provide significant positive dependence on the sub-district of Bojongloa Kaler, Bojongloa Kidul, Cicendo, Cidadap, Lengkong. Asphyxia, infant age groups provide significant positive dependence on sub-district of Cibeunying Kidul, Cinambo, Kiara Condong. Low birth weight, early neonatal age groups provide significant positive dependence on sub-district of Andir, Rancasari. Congenital abnormalities, early neonatal age groups provide significant positive dependence on sub-district of Babakan Ciparay, Batununggal, Sukajadi. Congenital abnormalities, advanced neonatal age groups provide significant positive dependence on sub-district of Bandung Kidul. Congenital abnormalities and infant age groups provide a significant positive dependence on sub-district of Coblong. Pneumonia, the early neonatal age group provide a significant positive dependence on the sub-district of Bandung Kulon. Sepsis, infant age groups provide significant positive dependence on sub-district of Bandung Wetan.

Keywords: multiple correspondence analysis, dependency, infant mortality

1. Introduction

Generally, multivariate analysis is all statistical methods that analyze multiple measurements simultaneously on each object or individual being observed. Multivariate analysis techniques are classified into two, namely dependency techniques, where variables are categorized into independent variables that affect and independent variables that are affected, and interdependence techniques, where variables are not differentiated into independent variables and independent variables, but each variable has the same level [1].

Karl Pearson and R.A. Fisher has developed ways to analyze categorical data, including graphical representations, so that many statistical methods can be used to measure, model, visualize, and examine how variables are related to one another. The techniques from Pearson and Fisher became the basis for the construction of correspondence analysis. Correspondence analysis has a broad use in the social and scientific spheres as a method for describing patterns of
association in a frequency table. Multiple Correspondence Analysis (MCA) is a development of correspondence analysis used to determine the relationship of the categories between one and another. The variables used in this analysis are more than two categorical variables.

One of the studies using MCA was held by Pangastuti (2013) [2] who examined the positioning of a laptop brand to other brands on the market and the proximity between brands that is influenced by processor factors, designating laptops for consumers, feature sets, support and warranty, laptop resistance to damage and laptop lifespan when experiencing damage to the hardware the first time. In addition, Nurfitasari (2017) [3] examined criminal acts that occurred in the City and Regency of Madiun during January to December 2016. The observed variables include crime scene, time and area of crime. While Ariyanti (2017) [4] examined the tendency of Land Value Zones (ZNT) in Surabaya with number of public facilities and the territorial division as variables.

Determining the p-value is very important in order to see the level of significance of contributions between categories in multiple data, considering that in every study not all categories have a significant influence on the others. Thus, this study will elaborate the significance of the contribution of categories to qualitative variables in MCA which is the development of p-values based on confidence ellipse for two categories.

Some studies that use the p-value confidence ellipse approach to see the level of significance of contributions between categories in the data include Beh [5] that used elliptical confidence regions in simple correspondence analysis to analyze the data classification of maternal and infant attachments, the variables involved are maternal attachment classification and baby's response.

The case study in this research is about infant mortality in city of Bandung. Infant mortality in Bandung has decreased from 2016 to 2018. Despite of the decline, Infant Mortality Rate (IMR) in Bandung is still considered a problem that must be resolved, because the spread of infant mortality in 2018 has expanded to four villages, namely in 29 sub-districts out of 30 sub-districts compared to 2017 which is spread in 25 sub-districts. In addition, the treatment carried out by the Health Office of Bandung City in dealing with cases of infant death is same for each sub-district, although each sub-district has different characteristics in cases of infant death. Infant mortality is one of essential indicators to determine the degree of public health. Therefore, a preventive effort is needed to reduce IMR in Bandung, so a statistical method is needed to identify the characteristic patterns of causes of infant mortality in the sub-districts and infant mortality indicators as well as to find out information about categories that significantly contribute to the dependency structure.

Some studies conducted on infant mortality cases including Oktari et al. [6] who did modeling of the number of infant deaths in Padang in 2013 and 2014 with a negative binomial regression approach, the variables involved were LBW percentage, percentage of babies given exclusive breastfeeding, and others. Hajarisman et al. [7] estimated the IMR through a two-level hierarchical Poisson Bayes regression model, the variables involved were the number of live births, the number of infant deaths, and others. In addition, Elyana and Srinandi [8] carried out modeling of Infant Mortality Rates using the GWPR approach in Bali Province, the variables involved were the number of health facilities, the average length of breastfeeding, and others.

In order to identify the characteristics of causes of infant death and their predisposition patterns, multiple correspondence analysis can be used. Various methods had been developed to determine the statistical significance of a category on dependency structures, including building the confidence circles. However, the weakness of confidence circles is that they do not consider unbalanced weights in the correspondence plot axis. Therefore, this study uses confidence ellipses to reflect unbalanced weights on the axis of the correspondence plot. Using confidence ellipses is also better because it includes information contained in a higher dimension, so this plot provides a more comprehensive description of the dependencies of two qualitative variables. Approximate p-values confidence ellipses are designed to determine the statistical significance of a category on the dependency structure between two qualitative variables. This study aims to introduce the approximate p-values of confidence ellipses to multiple correspondence analysis using data on infant mortality in Bandung.
2. Methods

The data used in this study are data on infant mortality indicators in the city of Bandung in January 2018-June 2019. Data were obtained from Puskesmas reports of each sub-district in Bandung that reported to the Bandung City Health Office. Result of the critical study found eight qualitative variables that were indicators of infant mortality. Chi-square test results indicated that there were dependencies between causes of death, age groups, and sub-districts. In this study, there are 15 categories of causes of infant death, namely severe anemia [C1], asphyxia [C2], aspiration [C3], LBW [C4], DHF [C5], dehydration [C6], diabetes mellitus [C7], diarrhea [C8], immature [C9], viral infection [C10], congenital abnormalities [C11], pneumonia [C12], premature [C13], sepsis [C14], and others [C15]. There were 29 sub-districts in the city of Bandung which had cases of infant deaths including Andir [D1], Antapani [D2], Arcamanik [D3], Astana Anyar [D4], Babakan Ciparay [D5], Bandung Kidul [D6], Bandung Kulon [D7], Bandung Wetan [D8], Batununggal [D9], Bojongloa Kaler [D10], Bojongloa Kidul [D11], Buahbatu [D12], Cibeunying Kaler [D13], Cibeunying Kidul [D14], Cibiru [D15], Cicendo [D16], Cidadap [D17], Cinambo [D18], Coblong [D19], Gedeage [D20], Kiaracorong [D21], Lengkong [D22], Mandalajati [D23], Rancasari [D24], Regol [D25], Sukajadi [D26], Sukasari [D27], Sumur Bandung [D28], and Ujung Berung [D29]. In this study the age category is categorized into three groups namely early neonates (0-7 days) [A1], advanced neonates (8-28 days) [A2], and infants (29 days - <1 year) [A3].

2.1. Multiple Correspondence Analysis (MCA)

The data used in the MCA are data from the indicator matrix. Indicator matrix is a matrix where rows represent objects and columns are dummy variables that represent characteristics. The element of the indicator matrix is only 0 or 1. For example, \( Y = (y_{ik}) \) is a data matrix of size \( n \times p \) where \( n \) represents the quantity of objects, \( p \) represents the number of category variables, with \( i = 1, 2, \ldots, n \) and \( k = 1, 2, \ldots, p \). If \( q_k \) represents the number of categories of the \( k \)-th variable and \( Z_k = (z_{ijk}) \) is an indicator matrix for the \( k \)-th variable which element size is \( n \times q_k \) so with \( z_{ijk} \) is the \((i,j)\)-th element of \( Z_k \) with \( j = 1, 2, \ldots, q_k \), the indicator matrix is denoted as follows:

\[
Z = \left( Z_1 \left| Z_2 \right| \ldots \left| Z_p \right) \right.
\]

in the size of \( n \times Q \) with \( Q = \sum_{k=1}^{p} q_k \).

The symmetric matrix resulted by the matrix multiplication called the Burt matrix (Greenacre, 2007) has the following notation:

\[
B = Z^T Z = \begin{bmatrix}
Z_1^T Z_1 & Z_1^T Z_2 & \cdots & Z_1^T Z_p \\
Z_2^T Z_1 & Z_2^T Z_2 & \cdots & Z_2^T Z_p \\
\vdots & \vdots & \ddots & \vdots \\
Z_p^T Z_1 & Z_p^T Z_2 & \cdots & Z_p^T Z_p
\end{bmatrix} = \begin{bmatrix}
D_1 & N_{12} & \cdots & N_{1p} \\
N_{12}^T & D_2 & \cdots & N_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
N_{1p}^T & N_{2p}^T & \cdots & D_p
\end{bmatrix} = (b_{\ell\ell})
\]

\( D_k \) is a diagonal matrix of the marginal frequency of the \( j \)-th category of the \( k \)-th variable for \( k = 1,2,\ldots, p \), likewise \( N_{kk'} \) is a two-way contingency table of the variables of the \( k \)-th and \( k' \)-th, \( k \neq k' \) and \( \ell = 1,2,\ldots, Q \).

Burt correspondence matrix is a matrix where each element is a frequency matrix relative to the grand total \((b)\) Burt matrix \((B)\), with the following notation:

\[
P = \frac{1}{b} B = (p_{\ell\ell})
\]

The proportion of the Burt column shows the proportion of one category to all, where the proportion of the Burt column \((c)\) and the mass of the Burt matrix row are the same, with notation:

\[
c = \frac{1}{b} B 1
\]

where \( 1 \) is a vector sized \( q \times 1 \) and each element has a value of 1.

The main coordinates are obtained from a standard residual matrix that represents associations between categories, (Greenacre, 2007) [9] showing the following calculation:
\[ S = D_r^{-\frac{1}{2}}(P - P11^TP)D_c^{-\frac{1}{2}} \]

where \( P \) is the correspondence matrix, \( D_r \) is the row diagonal matrix, and \( D_c \) is the column diagonal matrix. Because in the Burt matrix the proportions of rows and columns are the same, so \( D_r = D_c \) and \( P1 = c \) and \( 1^TP = c^T \), thus:

\[ S = D_c^{-\frac{1}{2}}(P - cc^TP)D_c^{-\frac{1}{2}} \]

Because the residual matrix is searched based on the Burt matrix, the diagonal of the standard Burt residual matrix is decomposed to get the mutually orthonormal main coordinates, besides the decomposition results can also calculate the inertia which represents the variance of data included in each major coordinate.

This study used EVD (Eigen Value Decomposition) because the results obtained from EVD based on Burt matrix are more refined in getting the main coordinates as the basis for ellipse formation.

Column \( V \) is a mutually orthogonal eigen vector, \( V \) is an orthogonal matrix \( V^{-1} = V^T \) then \( VV^T = V^TV = I \) or in another matrix notation can be written as:

\[ S = VAV^T \]

where, \( \Lambda \) is a diagonal matrix of eigen value \( \lambda_m \), for \( m = 1,2,\ldots,M \), where \( M = Q - p \) thus \( \Lambda = \text{diag}(\vec{\lambda}) \), \( \vec{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m, \ldots, \lambda_M)^T \) and \( \lambda_1 > \lambda_2 > \ldots > \lambda_m = \ldots > \lambda_M \). In addition, \( V \) is the dimension of \( Q \times M \), a matrix containing an eigen vector of \( S \) which columns correspond with \( \lambda_m \).

The linear combination of eigen vectors of value of dependence between row categories and column categories or information, total inertia will be able to show the percent of the missing categories or information, total inertia has the following notation:

\[ \text{trace} (F^TF) = \text{trace} (\Lambda) \]

\[ \tau_d = \left( \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_d}{\sum_{m=1}^{M} \lambda_m} \right) \tag{1} \]

and the scope of variances for each dimension is as follows:

\[ \phi_d = \left( \frac{\lambda_d}{\sum_{m=1}^{M} \lambda_m} \right) \]

where \( \tau_d \) is the variances which are covered, \( \phi_d \) is the scope of the variances of each dimension, \( \lambda_d \) is the \( d^{th} \) eigen and \( d = (1,2,\ldots,m) \).

In the main inertia, the axis of the correspondence plot where the first inertia is greater than the second inertia, so that inertia can be used to build the confidence area, taking into account of the area of ellipse.

2.2. Elliptical Confidence Regions

The two-dimensional ellipse equation \( m = M(=2) \) centered on \((f_{\ell1}, f_{\ell2})\) has the corresponding plot of the axis length as follows:

\[ x_{\ell m(a)} = \lambda_m \left( \frac{X_{(\text{table})a}}{X_{(\text{count})}} \right) \left( \frac{1}{p_{\ell jk}} - \sum_{m=3}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 \right) \]

where \( x_{\ell m(a)} \) is the two-dimensional ellipse axis, \( \lambda_m \) is the eigen value, \( X_{(\text{table})a} \) is the chi-square table, \( X_{(\text{count})} \) is the calculated chi-square, \( p_{\ell jk} = \text{diag} (B)/n \), where \( x_{\ell1(a)} \) is the semi-major axis length of the confidence ellipse, and \( x_{\ell2(a)} \) is the semi-minor axis length of the ellipse with free
degrees ($\ell-1)(m-1)$. While the confidence ellipse for $m$ dimensions, $m = M (>2)$, then the $\ell^n$ row category, can be built with the length of the semi-axis along the $m^n$ main axis with the following equation:

$$x_{\ell m} = \frac{\chi^2_{(\text{table}) \alpha} \lambda_m}{\chi^2_{(\text{count})} \left( \frac{1}{p_{jk}} - \sum_{m=1}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 \right)}$$

where $x_{\ell m}$ is the dimension ellipse $m$ axis, $\lambda_m$ is the eigenvalue $\chi^2_{(\text{table}) \alpha}$ is the chi-square table, $\chi^2_{(\text{count})}$ is the chi-square count, $p_{jk} = \text{diag} (B) / n$.

After forming the ellipse and knowing the length of the axis, the next step is to find out the significance of the existing categories, by comparing Pearson's chi-square or p-value with a significance level ($\alpha$). The p-value calculation introduced by Beh and Lambardo is for two qualitative variables, while this study will be developed by using more than two qualitative variables.

2.3. Test the Significance of Categories of Qualitative Variables

This study examined the significance of categories through p-values between categories with more than three qualitative variables based on the Burt matrix, so that the following equation obtained was:

$$p - \text{value}_{i,M} \approx P \left\{ \chi^2_{(\text{table}) \alpha} > \chi^2_{(\text{count})} \left( \frac{1}{p_{jk}} - \sum_{m=M+1}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 \right)^{-1} \sum_{m=1}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 \right\} \quad (2)$$

The value of $\sum_{m=M+1}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 = 0$ in higher dimensions so that Equation (2) is clearly close to the origin of $M$, then in $M$ the p-value dimension can be denoted as:

$$p - \text{value}_{i,M} \approx P \left\{ \chi^2_{(\text{table}) \alpha} > \chi^2_{(\text{count})} p_{jk} \sum_{m=M+1}^{M} \left( \frac{f_{\ell m}}{\lambda_m} \right)^2 \right\}$$

The equation above reflects the relative weighting of each axis in the M dimensional correspondence plot, taking into account the inequality of the main inertia values.

2.4. Identify Distances Between Categories

In the MCA, information is obtained from the results of the distance between categories. In this paper the distance between categories for M dimensions was calculated using the Euclidean distance with the following equation:

$$d^2 = (x_j - x_j')^T(x_j - x_j')$$

where

- $d$ : Euclidean distance
- $x_j$ : vector for the $j$th category
- $x_{j'}$ : vector for $j'$th category

The smaller the value of $d$, the more similar the two vectors that are matched/compared. Otherwise, the greater the value of $d$, the more different the two vectors that are matched/compared (Budi Santosa, 2007) [11]. This study on decision making will be based on the Euclidean distance matrix between categories.

3. Result and Discussion

Before conducting an MCA, the relationships between variables are tested using the chi-square test with the following hypotheses:

$H_0$: Causes of death, age group, and sub-district are independent
H₁: Causes of death, age group, and sub-districts are dependent

With a significant level of α = 5%, and based on a significant value, if \( \text{sign} > 0.05 \), then the one accepted is \( H₀ \) that the conclusion is the causes of death, age group, and sub-district are independent. Otherwise, if the sign < 0.05, then it will reject \( H₀ \) that the conclusion is the cause of death, age groups, and sub-districts are dependent. By using software R the following results are obtained:

\[ \chi^2 = 2828.86, \text{df} = 1780, p - \text{value} = 0.022471004 \]

P-value <0.05, then \( H₀ \) is rejected, so the cause of death, age group, and sub-districts are dependent.

3.1. Multiple Correspondence Analysis (MCA)

MCA is used to determine the relationship between one category with another categories of three or more variables simultaneously. In this analysis, the dimensions, main inertia and coordinate values will be determined as follows:

| \( D \) | \( \lambda_m \) | \( \phi_d \) | \( \tau_d \) |
|--------|----------------|--------|--------|
| 1      | 0.483903       | 8.426483 | 8.426483 |
| 2      | 0.356877       | 6.214502 | 14.64099 |
| 3      | 0.305879       | 5.326454 | 19.96744 |
| 4      | 0.292262       | 5.089324 | 25.05676 |
| 5      | 0.274303       | 4.776602 | 29.83337 |
| ...    | ...            | ...    | ...    |
| 40     | 0.021198       | 0.36912 | 99.14021 |

Based on the analysis results, an eigenvalues, inertia, and number of dimensions are presented in Table 1. Based on these results the best results are shown by 40 dimensions. In this case the researcher sets 2 dimensions to explain the results using MCA, because if the dimensions used are large it will be difficult to interpret the results of the analysis. The results of the analysis using MCA for 2 dimensions can explain as much as 14.64% of the diversity of data. After determining the number of dimensions to be used in the analysis, it is necessary to know the main coordinates. This can be seen from the results of the main coordinate columns for each level of categorical variables in Table 2.

| Category | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | ...  | Dim 40 |
|----------|-------|-------|-------|-------|-------|------|-------|
| D1       | -0.3135 | 0.379148 | -0.06902 | 0.104771 | 0.205986 | ...  | 0.130767 |
| D2       | 1.234489 | -0.88488 | -2.40728 | 2.396303 | 0.234229 | ...  | 0.05686 |
| D3       | 1.057371 | 0.306973 | 0.348325 | -0.25668 | 0.049696 | ...  | 0.350563 |
| D4       | -0.49006 | -0.88412 | 0.068159 | -1.12687 | 0.54267 | ...  | -0.42711 |
| D5       | -0.3373 | 0.08431 | 0.345109 | 0.185407 | 0.370072 | ...  | 0.135957 |
| D6       | 0.13611 | 0.872286 | 1.032018 | 0.487229 | 0.973331 | ...  | -0.30188 |
| D7       | 0.301491 | 0.674743 | -0.28285 | -0.57232 | -1.16785 | ...  | 0.076495 |
| D8       | 1.108922 | 0.362017 | 2.110914 | 1.162936 | 1.005754 | ...  | -0.05994 |
| D9       | -0.08295 | 0.226089 | -0.2331 | -0.11212 | 0.316202 | ...  | -0.03743 |
| ...      | ...    | ...    | ...    | ...    | ...    | ...  | ...    |
| C15      | 0.484671 | 0.488563 | 1.96725 | 1.479058 | 0.61105 | ...  | -0.02274 |

3.2. Significance Test of Category Contribution on Qualitative Variables
Approximate p-values reflect the significance of categories on the dependency structure. Approximate p-values are presented in Table 3. Based on Table 3, the 5% significance level ($\alpha$) shows that all research categories make a significant contribution to the association between qualitative variables, so that all categories examined are worthy of attention as a basis for policy-making.

**Table 3. Approximate P-Values on Dependency Structure**

| Category | $\chi^2_{(\text{count})}$ df P-value | Category | $\chi^2_{(\text{count})}$ df P-value | Category | $\chi^2_{(\text{count})}$ df P-value |
|----------|---------------------------------|----------|---------------------------------|----------|---------------------------------|
| D1       | 6864   44  0                     | D17      | 7128   44  0                     | C1       | 7436   44  0                     |
| D2       | 7348   44  0                     | D18      | 7304   44  0                     | C2       | 5720   44  0                     |
| D3       | 7480   44  0                     | D19      | 6952   44  0                     | C3       | 7128   44  0                     |
| D4       | 7392   44  0                     | D20      | 7436   44  0                     | C4       | 5984   44  0                     |
| D5       | 7260   44  0                     | D21      | 7040   44  0                     | C5       | 7392   44  0                     |
| D6       | 7348   44  0                     | D22      | 7304   44  0                     | C6       | 7436   44  0                     |
| D7       | 7216   44  0                     | D23      | 7392   44  0                     | C7       | 7436   44  0                     |
| D8       | 7348   44  0                     | D24      | 7084   44  0                     | C8       | 7128   44  0                     |
| D9       | 7172   44  0                     | D25      | 7436   44  0                     | C9       | 7216   44  0                     |
| D10      | 7348   44  0                     | D26      | 7040   44  0                     | C10      | 7436   44  0                     |
| D11      | 6688   44  0                     | D27      | 7436   44  0                     | C11      | 6424   44  0                     |
| D12      | 7436   44  0                     | D28      | 7392   44  0                     | C12      | 7040   44  0                     |
| D13      | 7392   44  0                     | D29      | 7436   44  0                     | C13      | 7172   44  0                     |
| D14      | 7304   44  0                     | A1       | 3300   44  0                     | C14      | 7216   44  0                     |
| D15      | 7480   44  0                     | A2       | 6424   44  0                     | C15      | 7172   44  0                     |
| D16      | 7216   44  0                     | A3       | 5324   44  0                     |

Interpreting characteristics between categories of the three variables in this study can be done by finding the distance between categories based on the main coordinates of all dimensions.

The close distance between categories points out that there are high positive dependencies between these categories. While long distances have a high negative dependency, the probability of infant mortality in the sub-district area is very low. Meanwhile a high positive dependency shows that the likelihood of infant death in the sub-district area is very high.

**Table 4. Euclidean Distance Between Categories**
Based on the Euclidean distance above, we can see the smallest Euclidean distance and below the average of each variable, and the conclusions are that:
1. C2 and the A1 age group provide significant positive dependence on Sub-districts of D10, D11, D16, D17, and D22.

2. C2 and A3 age groups provide significant positive dependence on Sub-districts of D14, D18, and D21.

3. C4 and the A1 age group provide significant positive dependence on Sub-districts of D1 and D24.

4. C11 and the A1 age group provide significant positive dependence on Sub-districts of D5, D9, and D26.

5. C11 and the A2 age group provide significant positive dependence on Sub-districts of D26.

6. C11 and the A3 age group provide significant positive dependence on Sub-districts of D19.

7. C12 and group A1 provide significant positive dependence on Sub-districts of D7.

8. C14 Sepsis and A3 age groups provide significant positive dependence on Sub-districts of D8.

4. Conclusion

Based on the results and discussion, it concludes that:

1. Asphyxia and early neonatal age groups provide significant positive dependence on the Sub-districts of Bojongloa Kaler, Bojongloa Kidul, Cicendo, Cidadap, and Lengkong.

2. Asphyxia and the infant age group provide significant positive dependence on Sub-districts of Cibeunying Kidul, Cinambo, and Kiara Condong.

3. LBW and early neonatal age groups provide significant positive dependence on Sub-district of Andir and Rancasari.

4. Congenital abnormalities and early neonatal age groups provide significant positive dependence on Sub-district of Babakan Ciparay, Batununggal, and Sukajadi.

5. Congenital abnormalities and advanced neonatal age groups provide a significant positive dependence on Sub-district of Bandung Kidul.

6. Congenital abnormalities and infant age groups provide significant positive dependence on Sub-district of Coblong.

7. Pneumonia and the early neonatal age group provide a significant positive dependence on the Sub-district of Bandung Kulon.

8. Sepsis and the infant age group provide a significant positive dependence on the Sub-district of Bandung Wetan.

5. Suggestion

Elaborating the results of the analysis in this study will require more data with a broader scope, for example on a provincial or national scale.

6. Acknowledgements

The Authors acknowledge the financial support provided by Universitas Padjadjaran Fundamental Research (RFU) 2019, Number 3370/UN6.D/LT/2019.

References

[1] Rencher, C.A. (2002). Methods of Multivariate Analysis. Second Edition. A John Wiley and Sons, Inc. Publication. Canada.

[2] Pangastuti, A.2013. Mapping Laptop Brand Perceptions among Students using Multiple Correspondence Analysis, Universitas Diponegoro. Semarang.

[3] Nurfitasari, A.2017. Mapping of Criminal Acts in Madiun Area with Multiple Correspondence Analysis. Institut Pertanian Bogor.

[4] Ariyanti, D. 2017. Trend Analysis of Land Value Zones in the Surabaya Region Using Multiple Correspondence Analysis. Institut Teknologi Sepuluh Nopember: Surabaya.
[5] Beh, E.J. (2010). *Elliptical Confidence Regions for Simple Correspondence Analysis*. Journal of Statistical Planning and Inference. 140 (2010) 2582-2588.

[6] Oktari, W., Yozza, H., Yanuar, F. (2017). *Modeling the Number of Infant Mortality in Padang City in 2013 and 2014 with the Negative Binomial Regression Approach*. Jurnal Matematika Unand. Vol. VI No. 1 Hal. 74-82. Jurusan Matematika FMIPA Unand.

[7] Hajarisman, N., Mutaqin, A.K., A., A.I. (2013). *Estimation of Infant Mortality through the two-level hierarchical Poisson Bayes Regression Model*. Statistics, Vol. 13 No. 2, 81-92.

[8] Elyana, M.A.K., Srinandi, I.G.A.M. (2012). *Modeling Infant Mortality Rates using the Geographically Weighted Poisson Regression Approach in Bali Province*. Mathematical E-Journal, Vol. 1, No. 1, 94-98.

[9] Greenacre, M.J. 1984. *Theory and Application of Correspondence Analysis*. Academic Press Inc. New York.

[10] Ginanjar, I 2017. *Simplification of Correspondence Analysis to Improve Accuracy of Main Coordinates*. ITB, Bandung.

[11] Santosa, Budi (2007). *Data Mining Data Utilization Techniques for Business Purposes*. Yogyakarta: Graha Ilmu.