SIMPLE ALGORITHM TO CONSTRUCT CIRCULAR CONFIDENCE REGIONS IN CORRESPONDENCE ANALYSIS USING R

Karunia Eka Lestari [1*]**, Marsah Rahmawati Utami**[2]**, Mokhammad Ridwan Yudhanegara**[3]

[1,3]*Departement of Mathematics Education, Faculty of Teacher Training and Education,*
Universitas Singaperbangsa Karawang
[2]*Departement of Pharmacy, Faculty of Health Science, Universitas Singaperbangsa Karawang*
JL. H. S Ronggowaluyo Teluk Jambe Timur, Karawang, 41361, Indonesia

Corresponding author e-mail: [1]* *karunia@fkip.unsika.ac.id*

Abstract. Correspondence analysis has been widely applied in various fields as a graphical method to depict the association structure between two categorical random variables on a low-dimensional plot. This study built a simple algorithm to determine the principal coordinates and construct the circular confidence regions on the correspondence plot. In this algorithm, the determination of the standard residual matrix and the principal coordinates is built directly from the contingency table (without calculating a correspondence matrix). The algorithm was developed using R and applied to data on Covid-19 cases in West Java.

Keywords: correspondence analysis, categorical data, circular confidence region, contingency table.

https://ojs3.unpatti.ac.id/index.php/barekeng/   *barekeng.math@yahoo.com*

1. INTRODUCTION

Data visualization is an essential and critical aspect in the big data era. It helps effectively communicate data to readers, so it should be clear and simple. Various developments are performed to get a graphical method that is simpler and more informative. One of the exciting methods to be developed and researched is correspondence analysis (CA). CA is a graphical method that displays the association structure between two categorical random variables on a low-dimensional plot [1]. Therefore, CA is known as a method for reducing dimensions.

The reduction process involves the decomposition of singular values from the standard residual matrix that represents the values of the dependencies between variables. The decomposition of the standard residual matrix by singular-value decomposition (SVD) produces two orthogonal matrices and a diagonal matrix of singular values as a scale factor. These three matrices play a role in determining the vectors that represent each row and column category, called the principal coordinates of the rows and columns. These coordinates are depicted on a low-dimensional space and known as a correspondence plot [2]. The exciting aspect about CA is how this method can transform data in a table to produce a data visualization. The transformation process involves matrix operations, the role of eigenvalues, and matrix decomposition.

In the last five years, the investigation of principal coordinates has attracted the attention of many researchers [1], [3]–[6]. In contrast, examination of the relative position of a coordinate point to the origin tends to be excluded [7], [8]. Therefore, this study aims to identify the relative position of a principal coordinate to the origin by constructing statistical inference of circular confidence regions.

2. RESEARCH METHODS

The research method is divided into three stages: initiation, development, and case studies. A literature study related to the circular confidence regions was carried out at the initiation stage. In the development stage, we generated an algorithm to construct the circular confidence regions. This algorithm was built under R code and applied to real data as a case study. The data used is data on Covid-19 cases in West Java obtained through the website www.pikobar.jabarprov.go.id. Data analysis was performed using correspondence analysis to investigate the dependence structure between involved variables. The following notations are summarized from several literatures for determining principal coordinates in correspondence analysis [1]–[3], [9]–[11].

2.1 Correspondence Analysis

Consider two categorical random variables X and Y. Suppose X consists of I category, denoted $X_1, X_2, \cdots, X_I$, and Y consists of $\mathcal{J}$ category, denoted $Y_1, Y_2, \cdots, Y_\mathcal{J}$. If a random sample of n individuals or objects is selected, which is classified based on the categories of variables X and Y. The results of the cross-classification can be presented in the $I \times \mathcal{J}$ two-way contingency table. In matrix notation, such a contingency table is expressed as a cross-tabulation matrix $\mathbf{N} = (n_{ij})$. The (i,j)th element of $\mathbf{N}$ is denoted by n_{ij}, so that $n = \sum_{i=1}^{I} \sum_{j=1}^{\mathcal{J}} n_{ij}$. Let the ith row marginal frequency and the jth column marginal frequency respectively as $n_{i\cdot} = \sum_{j=1}^{\mathcal{J}} n_{ij}$ and $n_{\cdot j} = \sum_{i=1}^{I} n_{ij}$. The vector of row (and column) marginal frequencies are denoted by

$$\mathbf{b} = (n_{1\cdot}, \ n_{2\cdot}, \cdots, n_{I\cdot})^{\mathrm{T}} \quad \text{and} \quad \mathbf{k} = (n_{\cdot 1}, \ n_{\cdot 2}, \cdots, n_{\cdot \mathcal{J}})^{\mathrm{T}}. \tag{1}$$

Here, the non-singular matrices $\mathbf{D}_b$ and $\mathbf{D}_k$, respectively, represent the diagonal matrix of the marginal frequencies of rows and columns,

$$\mathbf{D}_b = \text{diag}(\mathbf{b}) \quad \text{and} \quad \mathbf{D}_k = \text{diag}(\mathbf{k}). \tag{2}$$

The association between rows and columns in a contingency table are represented by a standard residual matrix $\mathbf{S} = (s_{ij})$, where $s_{ij} = \frac{n_{ij} - \frac{n_{i\cdot} n_{\cdot j}}{n}}{\sqrt{n_{i\cdot} n_{\cdot j}}}$. The element s_{ij} represents the dependency of $p_{ij} \neq p_{i\cdot} p_{\cdot j}$. In the matrix operation expressed as

$$\mathbf{S} = \mathbf{D}_b^{-1/2} (\mathbf{N} - n^{-1} \mathbf{b} \mathbf{k}^{\mathrm{T}}) \mathbf{D}_k^{-1/2}. \tag{3}$$

The principal coordinates of row $\mathbf{F}$ and column $\mathbf{G}$ are determined by

$$\mathbf{F} = \sqrt{n}\, \mathbf{D}_b^{-1/2} \mathbf{U} \mathbf{D}_\sigma \quad \text{and} \quad \mathbf{G} = \sqrt{n}\, \mathbf{D}_k^{-1/2} \mathbf{V} \mathbf{D}_\sigma, \tag{4}$$

where $\mathbf{D}_\sigma$, $\mathbf{U}$, and $\mathbf{V}$ are obtained from the SVD of $\mathbf{S}$ and respectively referred to as the diagonal matrix of the singular values of $\mathbf{S}$, the matrix in which each column is the left and right singular vectors of $\mathbf{S}$.

Each column of matrices $\mathbf{F}$ and $\mathbf{G}$ represent the principal axes or dimensions. The first column of X and Y are covered the largest variance, the second column covers the second largest variance and so on, such that the low-dimensional plot absorbs the larger variance. Mapping the column vectors of $\mathbf{F}$ and $\mathbf{G}$ produces a correspondence plot [12].

2.2 Circular Confidence Regions

Some researchers have used circular confidence regions to identify whether a category is statistically consistent with what would be expected based on the independence hypothesis [8], [13]–[17]. These regions are derived algebraically based on singular values and vectors from the transformed contingency table. The confidence regions can also be considered parametric since they are based on Pearson's chi-square statistic assumptions.

Based on Equation (3), the difference between the observed cell frequency and the estimated cell frequency (under the assumption that H_0 is independent) is zero. It shows the configuration position of the coordinate points representing each category on the contingency table centred at the origin. The coordinate points close to the origin indicate that the category represented by the coordinate point does not explain the association between variables. Therefore, a measure is needed to find out how close (or far) a coordinate is from the origin so that it is said to contribute significantly in describing the association between categorical variables. It can be known through the circular confidence regions construction [7]. This construction is important to determine which categories contribute significantly to the structure of dependence between variables. Lebart [9] defined circular confidence regions for the principal coordinates of row ($r_{i,\alpha}$) and column ($r_{j,\alpha}$) as follows

$$r_i = \sqrt{\frac{n \cdot \chi_\alpha^2}{n_{i.}}} \ \text{dan} \ r_j = \sqrt{\frac{n \cdot \chi_\alpha^2}{n_{.j}}}. \tag{5}$$

Here χ_α^2 is the $(1-\alpha)$th percentile of the chi-square distribution with degrees of freedom 2. For the two-dimensional plot on $\mathbb{R}^2$ with a significance level $\alpha = 5\%$, the value of $\chi_\alpha^2 = 5.99$ is obtained. Additionally, $r_{i,\alpha}$ and $r_{j,\alpha}$ are the radius length of the $100\,(1-\alpha)\%$ circular confidence region for the ith row and jth column categories in the contingency table, respectively.

Taking into the theory underlying the construction of the $100\,(1-\alpha)\%$ confidence region of a coordinate point in a correspondence plot, one can estimate the p-value of this point in relation to its proximity to the origin. The p-value can be used to assess the statistical significance of the row (or column) categories on the association between variables. The approximation is determined and derived algebraically based on the circular region [14].

The p-value approximation is determined by firstly formulating the null hypothesis and its alternatives. Considering the relative distance of a row (or column) principal coordinate from the origin on a correspondence plot reflects the variation of the category corresponding to that coordinate under the assumption that H_0 is independent. Thus, a contribution to the chi-square statistic of the ith row category or jth column category can be determined by considering proximity from the origin. The following is the formulation of the null and alternative hypotheses

$$H_0: f_{im} = 0, \ vs. \ H_1: f_{im} \neq 0, \tag{6}$$
$$H_0: g_{im} = 0, \ vs. \ H_1: g_{im} \neq 0. \tag{7}$$

where $m = 1, 2, \cdots, \mathcal{M}^*$.

Based on the hypothesis above, consider the following statistics

$$X_{i\mathcal{M}^*}^2 = np_{i.} \sum_{m=1}^{\mathcal{M}^*} f_{im}^2, \tag{8}$$
$$X_{j\mathcal{M}^*}^2 = np_{.j} \sum_{m=1}^{\mathcal{M}^*} g_{jm}^2 \tag{9}$$

Generally, the plot dimensions used to represent the association between the row and column categorical variables are less than $\mathcal{M}^* = \min\{I, J\} - 1$. The plot dimensions used are commonly $D = 2$ on

$\mathbb{R}^2$ or $D = 3$ on $\mathbb{R}^3$. Therefore, the p-value approximation of the principal coordinates of the ith row and jth column in the m-dimensional correspondence plot to be calculated by

$$p - value_{iD} = P\{\chi_\alpha^2 > \chi_{iD}^2\} \approx P\{\chi^2 > np_{i\cdot} \textstyle\sum_{m=1}^D f_{im}^2\} \tag{10}$$
$$p - value_{jD} = P\{\chi_\alpha^2 > \chi_{jD}^2\} \approx P\{\chi^2 > np_{\cdot j} \textstyle\sum_{m=1}^D g_{jm}^2\} \tag{11}$$

Equations (10) and (11) consider the proximity of a coordinate point from the origin and ignore the principal inertia for each axis in the correspondence map. Based on the sample in the contingency table, this p-value represents the opportunity to get the coordinates of the ith line as the observed extreme point from the contingency table. Thus, a p-value less than a certain significance level provides evidence that the category has a statistically significant contribution in describing the association structure since a certain point in the configuration is assumed inconsistent.

3. RESULTS AND DISCUSSION

This section presents the results obtained at the development stage and case studies. The results at the development stage are in the form of algorithms and R code to determine the circular confidence regions in the correspondence analysis. The algorithm is applied to the real data. Data on Covid-19 cases in West Java Province was used as a case study.

3.1. Simple Algorithm for Circular Confidence Regions

From equations (3), (4), and (5), the following algorithm is obtained

Algorithm 1. *Circular confidence regions construction*

Step 1: Calculate the standard residual matrix, **S**.
Step 2: Determine the SVD of the standard residual matrix, $\mathbf{S} = \mathbf{U}\mathbf{D}_\sigma\mathbf{V}^\mathrm{T}$.
Step 3: Determine the principal coordinates of row **F** and column **G**.
Step 4: Construct circular confidence regions for the principal coordinates of row $r_{i,\alpha}$ and column $r_{j,\alpha}$.
Step 5: Plotting correspondence plot and circular confidence regions.

Algorithm 1 was obtained by modifying Beh and Lombardo's algorithm [14]. The main difference between this algorithm and previous research lies in determining the standard residual matrix and the principal coordinates built directly from the contingency table (without involving the correspondence matrix). The advantages of working from this contingency table include (1) reducing long computational processes; (2) minimizing the rounding process such that computational errors can be minimized and the principal coordinates produced are more precise than the classical method; and (3) the algorithm is simpler. The algorithm is presented in R code with input parameters as follows

(1) Contingency table or cross-tabulation matrix **N.**
(2) `M`, the number of dimensions considered to construct circular confidence regions or calculate p-values. From an algebraic perspective, `M` is the rank of the cross-tabulation matrix **N**. On the other side, from the correspondence analysis point of view, `M` referred to the minimum dimensions required to obtain an optimal correspondence plot. It is usually set as $M = \min(I, J) - 1$, where I and J are the number of rows and columns of **N**.
(3) `alpha`, significance level considered for constructing circular confidence regions. By default, set `alpha=0.05`.
(4) `scaleplot`, a function that scales the principal axis to provide the best visualization of principal coordinates or circular confidence regions. By default, set `scaleplot=1.2`.
(5) `a1` and `a2`, dimensions used to construct a two-dimensional correspondence plot. By default, set `a1=1` and `a2=2` to construct a plot on $\mathbb{R}^2$.
(6) `cols`, the colors used to draw principal coordinates, labels, and circular confidence regions. By default, the red is choose to the row features, and blue to the column features.

Here is an R script to construct circular confidence regions by applying Algorithm 1.

```r
#Input parameters of contingency table
N<-matrix(c(10554,2349,1747,4903,4812,1012,1779,477,5653,1570,608,2178,2171,638,2937,
            6330,2281,602,4213,920,14052,2467,20293,28089,3914,1805,572,27511,5677,639
            1,24123,19430,4296,8819,11570,15545,7497,7275,10651,6101,10548,34682,37987
            ,12983,3438,26162,6452,23307,8530,58950,52717,6420,10096,3350,138,66,14,50
            1,987,189,237,90,274,184,124,449,120,388,1578,83,205,88,400,248,213,85,383
            ,1085,125,434,95),ncol = 3)
dimnames(N)<-list(paste(c("Bogor","Sukabumi","Cianjur","Bandung","Garut",
            "Tasikmalaya","Ciamis","Kuningan","Cirebon","Majalengka","Sumedang",
            "Indramayu","Subang","Purwakarta","Karawang","Bekasi","West        Bandung",
            "Pangandaran","Bogor City","Sukabumi City", "Bandung City","Cirebon City ",
            "Bekasi   City","Depok   City","Cimahi   City","Tasikmalaya   City","Banjar
            City")),paste(c("Isolated","Healed", "Died")))
city<- dimnames(N)[[1]]                #row category label
case<- dimnames(N)[[2]]                #columncategory label

##Contingency table
N

#1# CALCULATE THE STANDARD RESIDUAL MATRIX
n <- sum(N)                            #number of observation
I <- nrow(N)                           #number of row
J <- ncol(N)                           #number of column
b <- as.matrix(apply(N, 1, sum))       #row marginal frequencies vector
k <- as.matrix(apply(N, 2, sum))       #column marginal frequencies vector
Db<-diag(b[1:I])                       #diagonal matrix of row marginal frequencies
Dk<-diag(k[1:J])                       #diagonal matrix of column marginal frequencies

##The standard residual matrix
S<-(diag((b^(-0.5))[1:I]))%*%(N-((1/n)*(b %*% t(k))))%*%(diag((k^(-0.5))[1:J]))
S

#2# DETERMINE THE SVD OF THE STANDARD RESIDUAL MATRIX
dns <- svd(S)
dns

#3# DETERMINE THE PRINCIPAL COORDINATES OF ROW AND COLUMN
f<-((n^(0.5))*(diag((b^(-0.5))[1:I])))%*%(dns$u)%*%(diag(dns$d))
f
g<-((n^(0.5))*(diag((k^(-0.5))[1:J])))%*%(dns$v)%*%(diag(dns$d))
g

#4# CONSTRUCT CIRCULAR CONFIDENCE REGIONS
#Input parameter M
M =min(nrow(N), ncol(N))-1
#Input parameter alpha
alpha<-0.05

##Inertia
Inersia.Principal<- diag(t(f[, 1:min(I-1, J-1)])%*% ((1/n)*Db)%*% f[, 1:min(I-1,J-1)])
Inertia.Total<- sum(Inersia.Principal)
Percentage.Inertia <- (Inersia.Principal/Inertia.Total) * 100
Total.Perc.Inertia.M <- sum(Inersia.Principal [1:M])

##Circular convidence regions
chi.squared.stat<- qchisq(1-alpha, df = (I - 1) * (J - 1))
radii <- sqrt(qchisq(1 - alpha, 2)/(b))
radij <- sqrt(qchisq(1 - alpha, 2)/(k))

#Approximate p-values
pvalrowcircle <- vector(mode = "numeric", length = I)
pvalcolcircle <- vector(mode = "numeric", length = J)

for (i in 1:I){
 #P-value approximation for the circular confidence regions of the row category
  pvalrowcircle[i] <- 1 - pchisq(b[i] * (f[i, 1]^2 + f[i, 2]^2),df = (I-1)*(J-1))
}
```

```r
for (j in 1:J){
  # P-value approximation for the circular confidence regions of the column category
  pvalcolcircle[j] <- 1 - pchisq(k[i] * (g[j, 1]^2 + g[j, 2]^2),df = (I - 1) * (J - 1))
}

#5# PLOTTING CORRESPONDENCE PLOT AND CIRCULAR CONFIDENCE REGIONS

par(pty = "s")

#Input parameter scaleplot,a1,a2,cols
scaleplot<-1.2
a1<-1
a2<-2
cols=c(2,4)

##Correspondence plot
plot(0, 0, pch = " ", xlim = scaleplot * range(f[, 1:M], g[, 1:M]),
     ylim = scaleplot * range(f[, 1:M], g[, 1:M]),
     xlab = paste("Principal Axis", a1, "(",round(Percentage.Inertia[a1],digits = 2),
"%)"),
     ylab = paste("Principal Axis", a2, "(",round(Percentage.Inertia[a2], digits = 2),
"%)"))
text(f[,1], f[,2], labels = city, adj = 0, col = cols[1])
points(f[, a1], f[, a2], pch = "*", col = cols[1])
text(g[,1], g[,2], labels = case, adj = 1, col = cols[2])
points(g[, a1], g[, a2], pch = "#", col = cols[2])
abline(h = 0, v = 0)

##Circular confidence regions
title(main = paste(100 * (1 - alpha), "% Confidence Regions"))
symbols(f[,a1], f[,a2], circles = radii, add = T, fg = cols[1])
symbols(g[,a1], g[,a2], circles = radij, add = T, fg = cols[2])

#SUMMARRY#
summ.name <- c("Radius","P-value-circle")
row.summ <- cbind(radii,pvalrowcircle)
col.summ <- cbind(radij,pvalcolcircle)
dimnames(row.summ) <- list(paste(city), paste(summ.name))
dimnames(col.summ) <- list(paste(case), paste(summ.name))

list(Row.Summary = round(row.summ, digits = 4), Column.Summary =
       round(col.summ, digits = 4))
```

3.2.   Case Study

This case study presented the analysis of Covid-19 data in West Java Province. The data records the total confirmed cases of Covid-19 in 27 Cities/Regencies of West Java on July 28, 2021. The analysis involves two categorical variables, covering Cities/Regencies in West Java, with a nominal scale consisting of 27 categories, and the other variable is confirmed Covid-19 cases, which has a nominal scale and consists of three categories. The attributes of each variable are presented in Table 1

Table 1. Categorical variables attributes

Variable	Category			
City/Regency	1= Bogor 2= Sukabumi 3= Cianjur 4= Bandung 5= Garut 6= Tasikmalaya 7= Ciamis	8= Kuningan 9= Cirebon 10= Majalengka 11= Sumedang 12= Indramayu 13= Subang 14= Purwakarta	15= Karawang 16= Bekasi 17= West Bandung 18= Pangandaran 19= Bogor 20= Sukabumi 21= Bandung City	22= Cirebon City 23= Bekasi City 24= Depok City 25= Cimahi City 26= Tasikmalaya City 27= Banjar City
Confirmed Case of Covid-19	1= Isolated 2= Healed 3= Died			

By running the previous R code on the Covid-19 data obtained the correspondence plot and circular confidence regions as follows
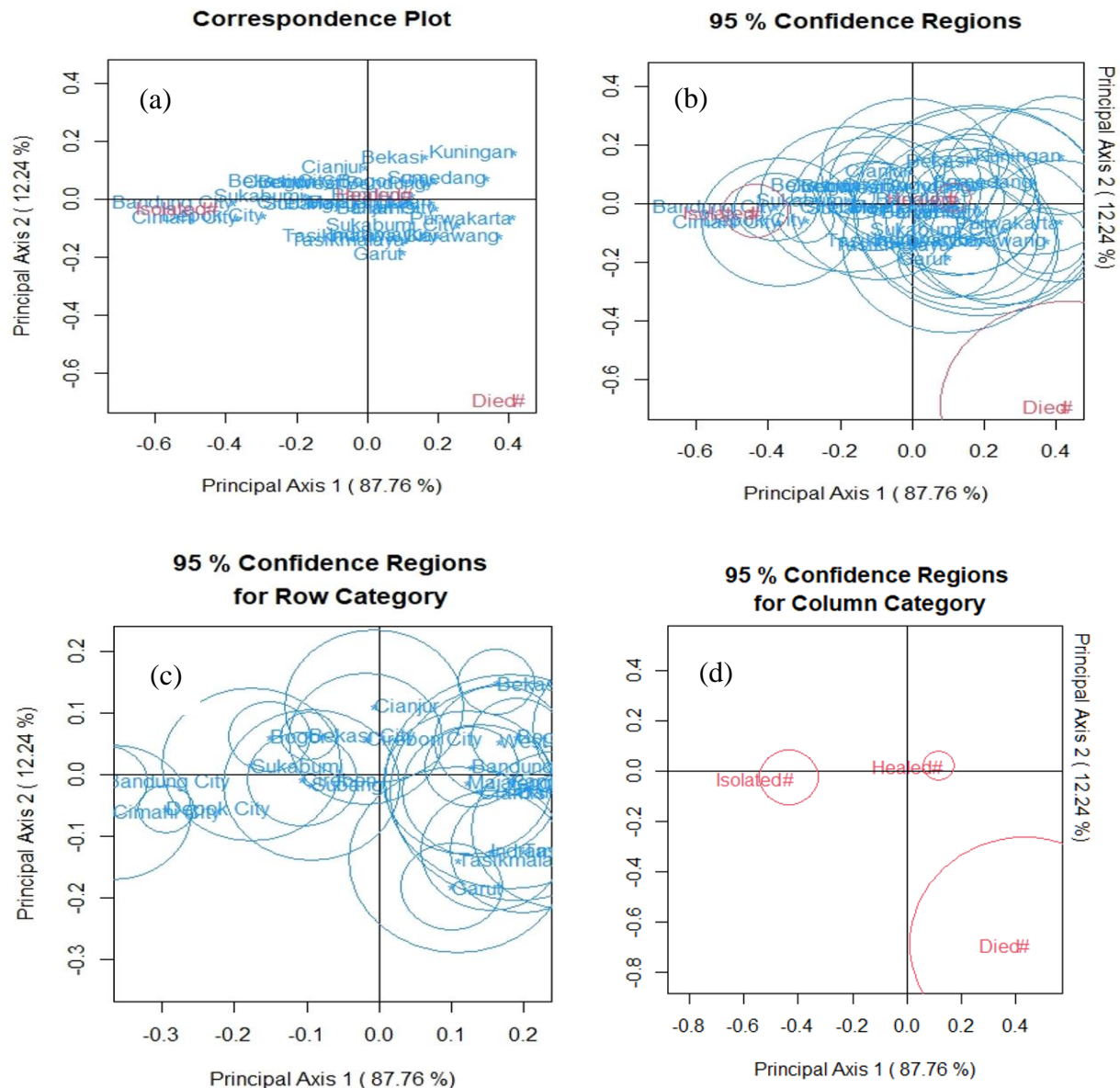

Figure 1. Correspondence Plot and Circular Confidence Regions for the Covid-19 Data in West Java

The association structure between Cities/Regencies in West Java based on the type of confirmed cases of Covid-19 is visualized through Figure 1(a). The figure suggests that the coordinates which represent the "died" category are located farthest from the origin $(0,0)$. It shows that the "died" column category has the largest contribution to the association structure among variables. If the `scaleplot=0.5` is set up, we can observe the "Bandung" coordinate is closest to the "healed" coordinates than other coordinates, and the coordinates with relatively close to "isolated" are the "Bandung City", "Depok City" and "Cimahi City" coordinates". It indicates that there is a strong association between the number of recovered cases and Bandung Regency and the number of isolated cases in Bandung City, Depok City and Cimahi City.

How close (or far) a coordinate from the origin such that it is said to contribute significantly in describing the association between categorical variables can be known by its circular confidence region as shown in Figure 1(b). If its circular confidence region contains the origin $(0,0)$, then the category represented by this coordinate does not significantly explain the association between variables.

The circular confidence regions for each row and column principal coordinate are presented separately in Figures 1(c) and 1(d). It addressed to get a better interpretation about which categories contribute to the

association structure between variables. Figure 1(c) shows that the confidence region of the "Cirebon City" coordinate contains the point $(0,0)$. It indicates that "Cirebon City" does not contribute significantly to the association structure such that if the Covid-19 cases in Cirebon City are excluded, it will not affect the analysis. Furthermore, Figure 1(d) shows that the confidence region of each column category does not contain a point $(0,0)$. It indicates that each of these categories contributes significantly to the association structure among variables. In addition, this figure exhibit that the coordinates of the "isolated", "healed", and "died" categories lie on different quadrants. Therefore, the column categories can be said to be mutually independent.

Generally, the observed categories significantly contribute to the association between variables, except for the "Cirebon City" category, which is not significant. For more details, the p-value for each category is presented in Table 2.

Table 2. Radius and p-value of circular confidence regions for the Covid-19 data in West Java

Category	Radius	*p-value*	Category	Radius	*p-value*	Category	Radius	*p-value*
Row Category								
Bogor	0,0125	0,0000	Majalengka	0,0254	0,0000	Bogor City	0,0140	0,0000
Sukabumi	0,0272	0,0000	Sumedang	0,0274	0,0000	Sukabumi City	0,0280	0,0000
Cianjur	0,0271	0,0001	Indramayu	0,0212	0,0000	Bandung City	0,0126	0,0000
Bandung	0,0142	0,0000	Subang	0,0267	0,0149	Cirebon City	0,0233	**0,8891**
Garut	0,0154	0,0000	Purwakarta	0,0228	0,0000	Bekasi City	0,0087	0,0000
Tasikmalaya	0,0330	0,0000	Karawang	0,0124	0,0000	Depok City	0,0086	0,0000
Ciamis	0,235	0,0000	Bekasi	0,0116	0,0000	Cimahi City	0,0239	0,0000
Kuningan	0,0222	0,0000	West Bandung	0,0197	0,0000	Tasikmalaya City	0,0220	0,0000
Cirebon	0,0167	0,0000	Pangandaran	0,0381	0,0000	Banjar City	0,0386	0,0000
Column Category								
Isolated	0,0068	0,0000	Healed	0,0036	0,0000	Died	0,0261	0,0000

4. CONCLUSIONS

The circular confidence regions algorithm proposed is built directly from the contingency table (without involving the correspondence matrix). Some of the advantages of working from a contingency table include:

(1) Reduce long computational processes.
(2) Minimize the rounding process to reduce computational errors and obtain the principal coordinates that are more precise than the classical method.
(3) The algorithm is simpler than in previous literature.

AKNOWLEDGEMENT

REFERENCES

[1]   M. Greenacre, *Correspondence Analysis in Practice*. Boca Raton: CRC Press Taylor & Francis Group, 2017.
[2]   K. E. Lestari, "Structure of Contingency Table Using Cardano and Cardano-Ferrari Formulas on Correspondence Analysis," Institut Teknologi Bandung, 2021.
[3]   I. Ginanjar, U. S. Pasaribu, and A. Barra, "Simplification of correspondence analysis for more precise calculation which one

qualitative variables is two categorical data," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 3, pp. 1983–1991, 2016.

[4] Y. Shi, L. Zhang, K. A. Do, C. B. Peterson, and R. R. Jenq, "APCoA: Covariate adjusted principal coordinates analysis," *Bioinformatics*, vol. 36, no. 13, pp. 4099–4101, 2020, doi: 10.1093/bioinformatics/btaa276.

[5] K. E. Lestari, U. S. Pasaribu, S. W. Indratno, and H. Garminia, "Generating roots of cubic polynomials by Cardano's approach on correspondence analysis," *Heliyon*, vol. 6, no. 6, p. e03998, 2020, doi: 10.1016/j.heliyon.2020.e03998.

[6] K. E. Lestari, U. S. Pasaribu, S. W. Indratno, and H. Garminia, "The comparative analysis of dependence for three-way contingency table using Burt matrix and Tucker3 in correspondence analysis," *J. Phys. Conf. Ser.*, vol. 1245, no. 1, 2019, doi: 10.1088/1742-6596/1245/1/012056.

[7] K. E. Lestari, U. S. Pasaribu, S. W. Indratno, and H. Garminia, "The Reliability of Crash Car Protection Level Based on the Circle Confidence Region on the Correspondence Plot," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 598, no. 1, 2019, doi: 10.1088/1757-899X/598/1/012061.

[8] A. D'Ambra, P. Amenta, and E. J. Beh, "Confidence regions and other tools for an extension of correspondence analysis based on cumulative frequencies," *AStA Adv. Stat. Anal.*, no. 0123456789, 2020, doi: 10.1007/s10182-020-00382-5.

[9] E. J. Beh and R. Lombardo, *Correspondence analysis theory, practice, and new strategies*. Wess Sussex: John Wiley & Sons, Ltd, 2014.

[10] K. E. Lestari, U. S. Pasaribu, and S. W. Indratno, "Graphical depiction of three-way association in contingency table using higher-order singular value decomposition Tucker3," *J. Phys. Conf. Ser.*, vol. 1280, no. 2, 2019, doi: 10.1088/1742-6596/1280/2/022035.

[11] U. S. Pasaribu, K. E. Lestari, S. W. Indratno, H. Garminia, R. K. N. Sari, and J.- Paul, "Some Properties of the Scaled Burt Matrix on Multiple Correspondence Analysis 2 . A Preliminary Theory," 2021.

[12] L. Doey and J. Kurta, "Correspondence Analysis applied to psychological research," *Tutor. Quant. Methods Psychol.*, vol. 7, no. 1, pp. 5–14, 2011, doi: 10.20982/tqmp.07.1.p005.

[13] T. J. Ringrose, "Bootstrap confidence regions for correspondence analysis," *J. Stat. Comput. Simul.*, vol. 82, no. 10, pp. 1397–1413, 2012, doi: 10.1080/00949655.2011.579968.

[14] E. J. Beh and R. Lombardo, "Confidence regions and approximate p-values for classical and non symmetric correspondence analysis," *Commun. Stat. - Theory Methods*, vol. 44, no. 1, pp. 95–114, 2015, doi: 10.1080/03610926.2013.768665.

[15] L. Palla, A. Chapman, E. Beh, G. Pot, and E. Almiron-Roig, "Where do adolescents eat less-healthy foods? Correspondence analysis and logistic regression results from the UK national diet and nutrition survey," *Nutrients*, vol. 12, no. 8, pp. 1–19, 2020, doi: 10.3390/nu12082235.

[16] A. N. Chapman, E. J. Beh, and L. Palla, "Application of Correspondence Analysis to Graphically Investigate Associations between Foods and Eating Locations," *Stud. Health Technol. Inform.*, vol. 235, no. 22, pp. 166–170, 2017, doi: 10.3233/978-1-61499-753-5-166.

[17] R. Lombardo and E. J. Beh, "Variants of simple correspondence analysis," *R J.*, vol. 8, no. 2, pp. 167–184, 2016, doi: 10.32614/rj-2016-039.