第 48 卷 第 5 期 2021 年 5 月
湖 南 大 学 学 报（自 然 科学版）
Journal of Hunan University（Natural Sciences）
Vol. 48. No. 5. May 2021

Open Access Article

Some Properties of the Scaled Burt Matrix on Multiple Correspondence Analysis

Udjianna Sekteria Pasaribu[1], Karunia Eka Lestari[2,*], Sapto Wahyu Indratno[1], Hanni Garminia[3], R. R. Kurnia Novita Sari[1]

[1] Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung, Indonesia

[2] Department of Mathematics Education, Universitas Singaperbangsa Karawang, Karawang, Indonesia

[3] Algebra Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung, Indonesia

Abstract: Multiple correspondence analysis (MCA) is well-known in statistics as a data analysis technique for multiple categorical variables. This method detects and represents underlying structures in a data set by representing data as points in a low-dimensional space. MCA is performed by applying the simple correspondence analysis (CA) algorithm to either an indicator matrix or a Burt matrix formed from these variables. Furthermore, the Burt matrix is scaled and undertaken eigendecomposition to get coordinates, which depicts the association's nature among variables. This study re-proposed the scale matrix of the Burt matrix, whose elements are the scale values of the categories of a variable, then so-called the scaled Burt matrix. While some researchers are interested in many MCA applications, we convenient our attention to exploring the properties of the scaled Burt matrix from a matrix algebraic perspective. These properties are derived mathematically to investigate the link between the Burt matrix and its scale matrix in representing the variables' associations.

Keywords: Burt matrix, categorical data analysis, indicator matrix, multiple correspondence analysis, scale matrix.

多重對應分析的尺度伯特矩陣的一些性質

摘要：多重對應分析（馬華）在統計領域是眾所周知的，是一種用於多個類別變量的數據分析技術。該方法通過將數據表示為低維空間中的點來檢測和表示數據集中的底層結構。通過將簡單對應分析（認證機構）算法應用於由這些變量形成的指標矩陣或伯特矩陣，可以執行馬華。此外，對伯特矩陣進行縮放並進行特徵分解以獲得坐標，該坐標描述了變量之間的關聯性質。這項研究重新提出了伯特矩陣的比例矩陣，其元素是變量類別的比例值，然後稱為縮放的伯特矩陣。雖然一些研究人員對許多馬華應用感興趣，但我們將注意力集中在從矩陣代數的角度探索縮放伯特 矩陣的屬性。這些屬性是通過數學推導得出的，以研究伯特矩陣與其標度矩陣之間的聯繫，以表示變量的關聯。

关键词：伯特矩陣，分類數據分析，指標矩陣，多重對應分析，比例尺矩陣。

Received: February 20, 2021 / Revised: March 19, 2021 / Accepted: April 18, 2021 / Published: May 28, 2021

Fund Project: The Indonesian Ministry of Research, Technology, and Higher Education, the Doctoral Dissertation Research Grant (PDD 2020-2021)

About the authors: Udjianna Sekteria Pasaribu, Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia; Karunia Eka Lestari, Department of Mathematics Education, Universitas Singaperbangsa Karawang, Karawang, Indonesia; Sapto Wahyu Indratno, Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia; Hanni Garminia, Algebra Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia; R.R. Kurnia Novita Sari, Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia

Corresponding authors Udjianna Sekteria Pasaribu, udjianna@math.itb.ac.id; Karunia Eka Lestari, karunia@staff.unsika.ac.id

1. Introduction

Jean-Paul Benzécri proposed multiple correspondence analysis (MCA) in the early 1960s as graphical data analysis for categorical variables. MCA is a statistical technique for uncovering latent structures in large or more complex datasets, including multidimensional categorical data [1]. This technique is practically applied to the simple correspondence analysis (CA) algorithm to multivariate categorical data that involves transforming such table into a two-way from through coded in an indicator matrix or a Burt matrix form [2]. In brief, MCA extends the CA by providing the ability to analyze a table containing some measure of correspondence among the rows and columns for more than two variables [3].

MCA is widely used in social sciences, behavioral science, material science, engineering, and biomedical research for graphically depicting the association between more than two categorical variables. Goodwill and Meloy [4] used MCA as a multidimensional scaling method to visualize the association among indicators for lone-actor terrorist attacks. Lestari et al. [5] utilized MCA to establish the reliability of crash car protection by describing circle confidence regions for each coordinate in the MCA plot. Greenacre. [6], Yudhanegara and Lestari [7] examine the utility of MCA in clustering a mixed-scale data set. Royan and Royan [8] applied MCA to explore the diabetic foot screening procedures data by investigating the relationships among the risk status classification of the post-screening decisions. Fred et al. [9] used MCA to derive the different impact dimensions of projects on biodiversity among Uganda communities. Brunette et al. [10] realize MCA to identify economic perspectives of forest adaptation to climate change. Beh and Lombardo [11] briefly explore the development, literature, and possible MCA research opportunities.

The analysis of the association between the variables of a two-way contingency table may be considered a particular case of MCA. In practice, any two-way contingency table can be obtained from a multi-way table by considering the product of the indicator matrix of one variable with the indicator matrix of another variable [11]. If there are n individuals observed based on m categories, the mathematical indicator will be $n \times m$in size. In case the number of respondents or categories is large, the indicator matrix requires large memories. It is one practical reason that the MCA is rarely performed using this matrix. Hence, many applications of this method commonly use the Burt matrix.

The Burt matrix is scaled and decomposed to get coordinates, which depicts the association among variables in the low-dimensional space. The scaling on the Burt matrix yields a matrix whose elements are the scale values of the categories of a variable, then so-called the scaled Burt matrix. While some researchers are interested in many MCA applications, we convenient our attention to exploring the properties of the scaled Burt matrix from a matrix algebraic perspective. This study examines the scaled matrix elements' characteristics through an algebraic approach to find the link between the Burt matrix and its scale matrix in representing the variables associations. It contributes to the development of scientific theories and practices relating to MCAs. Some findings which the novelty of this study are written in theorem form.

This paper is organized as follows. The preliminary theory of CA is briefly described in Section 2. In Section 3, we elaborate on expanding CA into MCA and investigating the scaled Burt matrix element. Some properties of this matrix are also presented. A case study is put forward in Section 4. A Summary and future works are presented as a conclusion in the last section.

2. A Preliminary Theory

Consider two categorical variables X_1 and X_2 , where X_1 consists of I categories, and X_2 consists of J categories. According to the I row and J columns, let N be an $I \times J$ two-way contingency table that cross-classifies n individuals. CA's main idea is to reduce the matrix's dimensionality and visualize the association between variables in a low-dimensional subspace, usually a two- or three-dimensional plot [12]. This plot represented the data as a set of points on the perpendicular coordinate axes.

For a simple example, suppose the 3×2 contingency table of Labor data reflecting a cross-classification of the race (X_1) and employment status (X_2) from a survey of 15 individuals. The contingency table (Table 1) has three categories of race as rows (e.g., white, black, and Asian) and two categories of employment status as columns (e.g., employment and unemployment). The data obtained are in Table 2.

Table 1 Labor data for 15 American civilians 25 years and over by race and employment status, along with the percentages of employment status in each race (in parentheses)

Status / Race	**Employment**	**Unemployment**	**Sum**
White	4 (66.7%)	2 (33.3%)	6
Black	1 (20%)	4 (80%)	5
Asian	3 (75%)	1 (25%)	4
Sum	8	7	15

This table can be considered in two different views: a set of rows or columns. To illustrate this point, each row in Table 1 is a set of frequencies reflecting the respective race, while each column reflects the two levels of employment status. If we want to compare the race, we should consider the different numbers of individuals in a total was have in each race. Otherwise, if we want to compare the two employment status levels visually, we should consider the number of individuals in each status.

When analyzing frequency data, it is sometimes better to reexpress the data as a set of percentages. For example, each race involves a different number of civilians and corresponds to a different base as far as the frequencies of the types of employment status are concerned. The four civilians in the white race, compare to the three in the Asian race, can be judged only concerning the number of civilians in these respective races. As percentages, they turn out to be quite different: 4 out of 6 is 66.7%, while 3 out of 4 is 75%. The visualization of the relative frequencies in Table 1 gives a more accurate comparison of the employment of people of different races.

The description above shows that the concept of a set of relative frequencies or a *profile* is fundamental to CA. Such sets or vectors of relative frequencies have special geometric features because each set's elements add up to 1 (or 100%). In analyzing a frequency table, relative frequencies can be computed for rows (row profiles) or columns (columns profiles) by dividing their frequencies by their total. Consider Table 1, the row profiles for these data: the profile of white is $[4/6 \quad 2/6]$. It is referred to as the profile of the white race across the type of employment status. Similarly, the profile of the Asian race across the type of employment status is $[3/4 \quad 1/4]$, concentrated mostly in the employment, as is the white race. In contrast, the black race has a profile of $[1/5 \quad 4/5]$, concentrated mostly in unemployment. These profiles can be depicted as points in a profile space (Fig. 1a). Similarly, the column profiles for these data: the profile of employment across the race is $[4/8 \quad 1/8 \quad 3/8]$, concentrated mostly in the Asian race, while the profile of unemployment is $[2/7 \quad 4/7 \quad 1/7]$, concentrated mostly in the Black race, as shown in Fig. 1c.

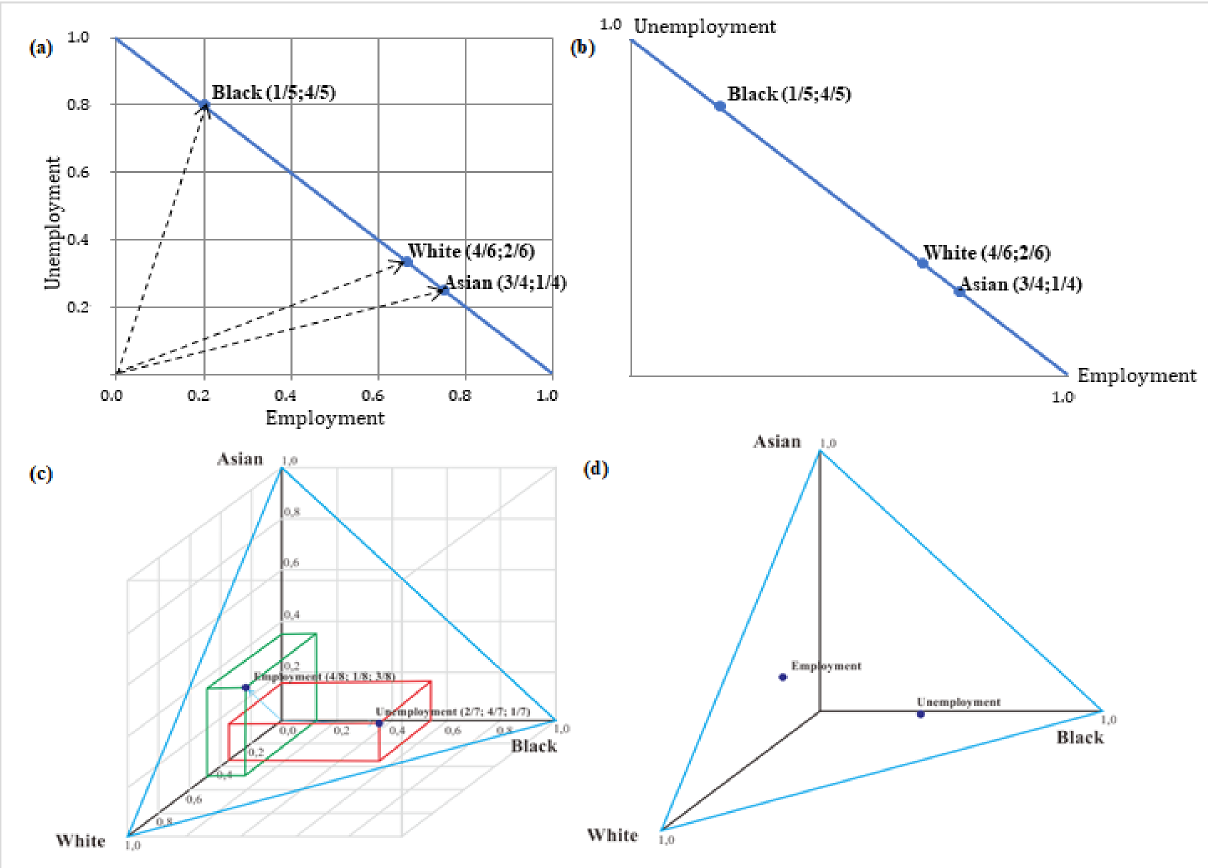

Fig. 1 The plot of the row and column profiles: (a) The row profiles in two-dimensional space; (b) The position of the three rows profile lies on a line; (c) Column profiles in three-dimensional space; (d) The two-column profile points lie precisely on an equilateral triangle

The profile points in two-dimensional space lie on a line (one-dimension) that joints the unit points $[1 \quad 0]$ and $[0 \quad 1]$ on the two axes, as shown in Fig. 1b. While the points in three-dimensional space lie precisely on a flat triangle (two-dimension) that joints the unit points $[1 \quad 0 \quad 0]$, $[0 \quad 1 \quad 0]$, and $[0 \quad 0 \quad 1]$ on the three respective axes, as in Fig. 1d. Each side is rescaled to be of length 1 and can be calibrated accordingly on a linear scale from 0 to 1.

3. Expansion into Multiple Correspondence Analysis

Consider two categorical variables X_1 and X_2, where X_1 consists of I categories, and X_2 consists of J categories. According to the I row and J columns, let N is an $I \times J$ two-way contingency table that cross-classifies n individuals. CA's

As data tables increase in size (e.g., more than two variables), it becomes more difficult to make simple graphical displays as in Fig. 1. One approach is to rearrange the multi-way frequency table as a two-way table to apply CA later. This approach well-known as multiple correspondence analysis (MCA). The expansion of CA into MCA is explained in the next section.

MCA's fundamental idea is that two or more categorical variables can be recoded as dummy variables in an indicator matrix or as a concatenation of categories-by-categories in a Burt matrix. The interpretation of the data from these both alternative codings of MCA is similar. The expansion of a simple contingency table into an indicator and Burt matrix is visualized in Fig. 2.

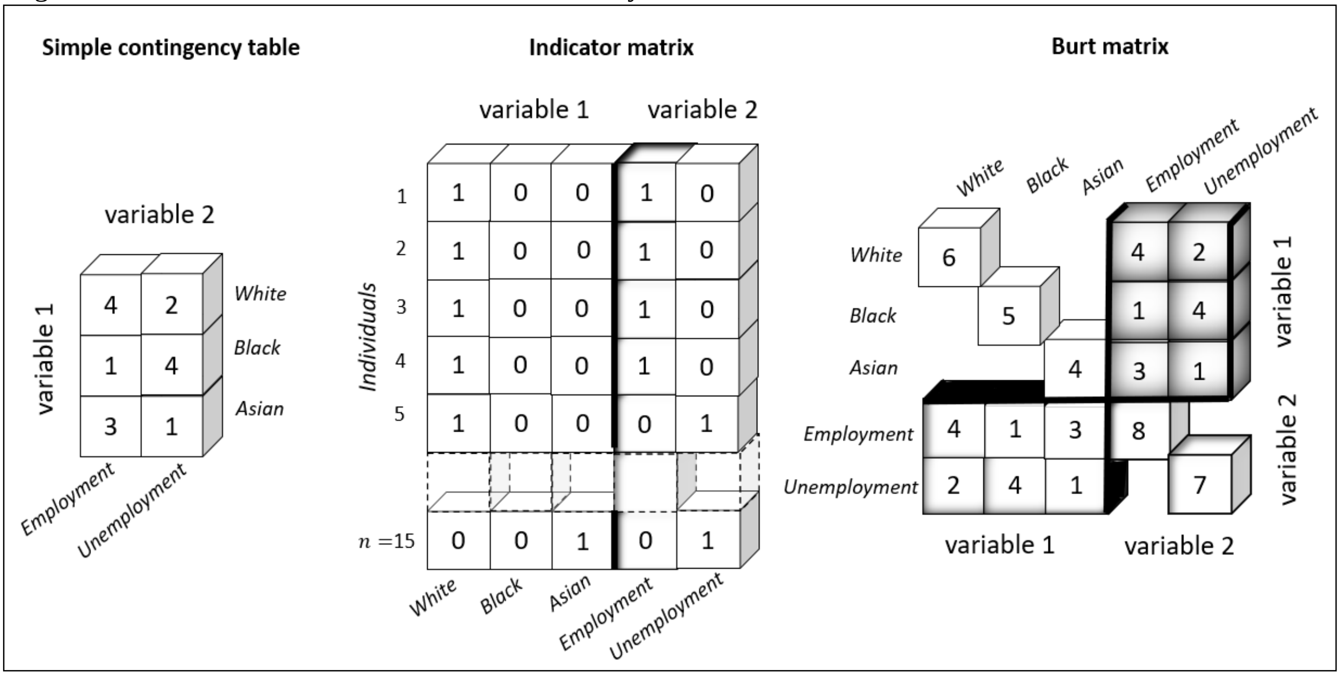

Fig. 2 Rearranging data from a simple contingency table into an indicator and Burt matrix

Considering the data in Table 1, it can recode into an indicator matrix form, which has as many rows as several individuals and as many columns as several categories. An alternative coding of the data is the Burt matrix, a square symmetric matrix of categories-by-categories. This matrix consists of all two-way contingency tables of pairs of variables, including the block diagonal of each variable's marginal frequencies. A brief description of the indicator and the Burt matrix is explained below to understand CA's expansion into MCA.

3.1. The Burt Matrix Construction

Suppose $X_1,\ X_2, \cdots,\ X_q$ are q categorical variables for n individuals, where variable k has j_k categories for $k = 1, 2, \cdots, q$. Note that for the $I \times J$ contingency table, I is referred to as j_1 that is the number of categories of X_1, and J is referred to as j_2 (the number of categories of X_2). The total number of categories under consideration is $m = \sum_{k=1}^{q} j_k$. Let X_k be the indicator matrix for the k-th variable, where X_k is a binary $n \times j_k$ matrix with precisely one nonzero element in each row i indicating in which category of variable k observation i falls for $i = 1, 2, \cdots, n$. Thus, $X = \left(x_{ij_k}\right)$, where $x_{ij_k} = 1$ if the subject i selects category k of variable j, and $x_{ij_k} = 0$ otherwise. In previous literature, X_k was called a block matrix [1] or submatrix. The concatenating these block or sub-matrices leads to the $n \times m$ super-indicator matrix, which is

$$\underset{(n\times m)}{X} = \left(\underset{(n\times j_1)}{X_1}\middle|\underset{(n\times j_2)}{X_2}\middle|\cdots\middle|\underset{(n\times j_k)}{X_k}\middle|\cdots\middle|\underset{(n\times j_q)}{X_q}\right). \qquad (1)$$

However, if the sample size is considerable, the indicator matrix can consist of thousands or even many more rows [11], [13]. It is the reason why the MCA involves summarising the data in the Burt matrix form. The Burt matrix B derived by considering its indicator matrix form and has the following block structure [5]:

$$\underset{(n\times m)}{B} = X^T X = \begin{pmatrix} \underset{(j_1\times j_1)}{D_1} & \underset{(j_1\times j_2)}{N_{12}} & \cdots & \underset{(j_1\times j_q)}{N_{1q}} \\ \underset{(j_2\times j_1)}{N_{12}^{\mathrm{T}}} & \underset{(j_2\times j_2)}{D_2} & \cdots & \underset{(j_2\times j_q)}{N_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(j_q\times j_1)}{N_{1q}^{\mathrm{T}}} & \underset{(j_q\times j_2)}{N_{2q}^{\mathrm{T}}} & \cdots & \underset{(j_q\times j_q)}{D_q} \end{pmatrix} \qquad (2)$$

Here, $\underset{(j_k\times j_{k'})}{N_{kk'}}$ is the two-way contingency table formed from the k-th and k'-th variables $\left(k \neq k'\right)$, and $\underset{(j_{k'}\times j_k)}{N_{kk'}^{\mathrm{T}}}$ is the transpose of $\underset{(j_k\times j_{k'})}{N_{kk'}}$. Denote n_{ik} as the element of $\underset{(j_k\times j_{k'})}{N_{kk'}}$, and $D_k = diag(n_{\cdot j_k})$ to be a diagonal matrix of column marginal frequencies of the

k-th variables.

The association between the q variables visualizes in a reduced dimension space by correspondence plot. It is constructed by first performing an eigendecomposition (ED) on the scaled Burt matrix. Define B^* as the scaled Burt matrix, whose elements are the scale value of the categories of a variable [11]. The scaled Burt matrix can be computed by

$$B^* = \frac{1}{q^2 n} D^{-1} B, \quad (3)$$

where $D = \text{diag}\left(\frac{n_{\cdot j_k}}{n}\right)$ is the diagonal matrix of the column proportions of the k-th variable, such that

$$\text{ED}(B^*) = U \Lambda_B U^T. (4)$$

Here, Λ_B is a diagonal matrix of the eigenvalues λ_ℓ^B for $\ell = 1, 2, \cdots, L$, that is $\Lambda_B = \text{diag}\left(\lambda_\ell^B\right)$. The matrix U contains the eigenvector of B^*. The left-hand side is the eigendecomposition of the scaled Burt matrix, and the right-hand side is the matrices resulting from the decomposition. For example, consider the 3×2 contingency table in Table 1, then $n = 15$, and $q = 2$ variables with $j_1 = 3 = I$ and $j_2 = 2 = J$ categories. Then we obtain:

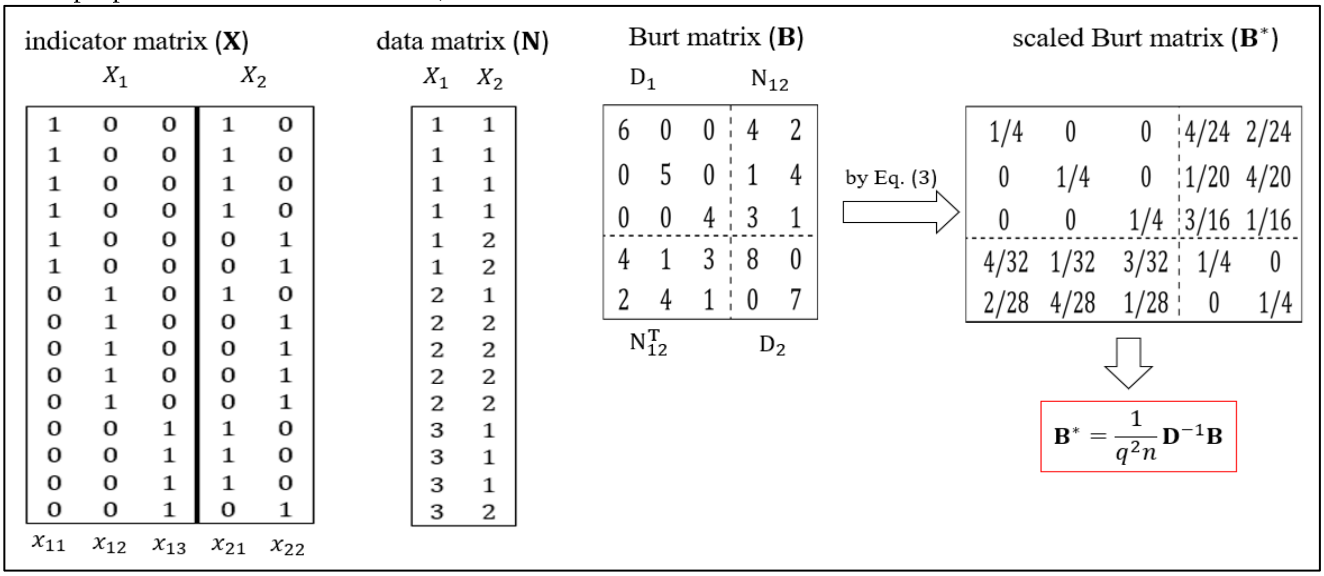

Fig. 3 Scale Burt matrix process

The matrices N and X are indeed related codings of data. In the indicator matrix approach, associations between variables are exposed by calculating the chi-square distance between different categories of the variables and between the individuals (or respondents). These associations are depicted graphically as "maps", which simplify the interpretation of the structures in the data. The number of individuals with the same characteristics for each category can be known directly through each column's marginal frequency and denoted by $n_{\cdot j_k} = \sum_{i=1}^{n} x_{i j_k}$. Oppositions between rows and columns are then maximized to uncover the underlying dimensions best able to describe the central oppositions in the data. The first axis is the most important dimension, the second axis the second most important, and so on. The number of axes to be retained for analysis is determined by calculating modified eigenvalues.

Analyzing the indicator matrix allows the direct representation of individuals as points in geometric space. On the other hand, analyzing the Burt matrix is a more natural generalization of simple correspondence analysis. Individuals or groups of individuals can be added as additional points to the graphical display. The Burt matrix is a real symmetric matrix of all two-way cross-classification between the categorical variables, a square matrix of size $m \times m$ and an analogy to the covariance matrix of continuous variables. Fig. 4 presents an MCA plot using three matrix approaches comprehensively.

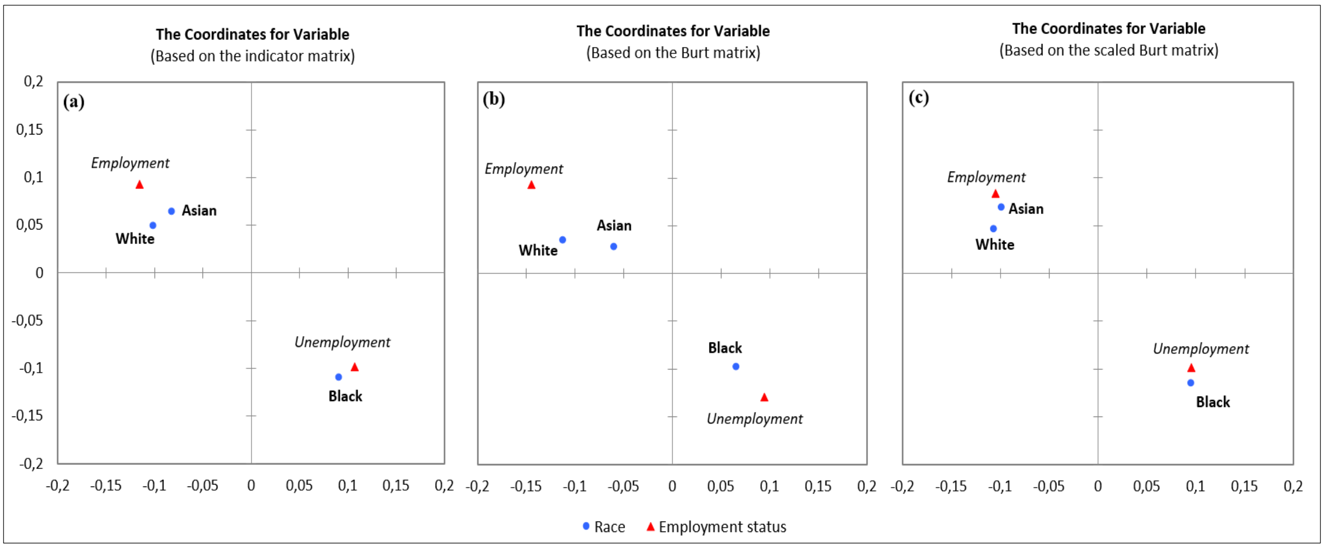

Fig. 4 MCA's plots for the variable by the indicator matrix (a), the Burt matrix (b), and the scaled Burt matrix approach (c). These plots display the association between race and employment status with a similar interpretation. The scaled Burt matrix is akin to a profiled version of the Burt matrix

Generally, MCA is characterized by the optimal scaling of categorical variables. This scaling can be undertaken using generalized singular value decomposition or eigendecomposition [11]. The link between B and B^*, including some of the properties of B^*, will be investigated in the following subsection. The research hypothesis related to this link is that the scaled Burt matrix elements represent the association of the variables and are associated with the proportion of the number of categories with the number of categorical variables.

3.2. The Scaled Burt Matrix Properties

In this section, the relationship between B and B^* by observing the elements of B^* is presented. Theorem 1a shows that the elements of B^* are the conditional probability for each pair of a categorical variable.

Theorem 1: Suppose $\underset{(\mathrm{n}\times\mathrm{m})}{X}$ is a super-indicator matrix of q categorical variables. The Burt matrix and its scale are defined by $B = X^TX$, and $B^* = \frac{1}{\mathrm{q^2 n}}D^{-1}B$, then:

the element of B^* is $\mathrm{b}^*_{\mathrm{kk'}} = \frac{\mathrm{n_{kk'}}}{\mathrm{q^2\, n_{\cdot j_k}}}$, for $\mathrm{k, k'} = 1, 2, \cdots, \mathrm{q}$.

the diagonal element of B^* is equal to $\frac{1}{\mathrm{q^2}}$.

Proof (1a): Since $B^* = \frac{1}{\mathrm{q^2 n}}D^{-1}B$, we have

$$B^* = \frac{1}{q^2 n}\begin{pmatrix} \underset{(j_1\times j_1)}{D_1^{-1}D_1} & \underset{(j_1\times j_2)}{D_1^{-1}N_{12}} & \cdots & \underset{(j_1\times j_q)}{D_1^{-1}N_{1q}} \\ \underset{(j_2\times j_1)}{D_2^{-1}N_{12}^{\mathrm{T}}} & \underset{(j_2\times j_2)}{D_2^{-1}D_2} & \cdots & \underset{(j_2\times j_q)}{D_2^{-1}N_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(j_q\times j_1)}{D_q^{-1}N_{1q}^{\mathrm{T}}} & \underset{(j_q\times j_2)}{D_q^{-1}N_{2q}^{\mathrm{T}}} & \cdots & \underset{(j_q\times j_q)}{D_q^{-1}D_q} \end{pmatrix},$$

where $\underset{(j_k\times j_k)}{D_k^{-1}} = \mathrm{diag}\left(\frac{n}{n_{\cdot j_k}}\right)$ with $n_{\cdot j_k} = \sum_{i=1}^{n} x_{ij_k}$ is the number of observations on the j-th category of k-th variable and $\sum_{j=1}^{q} j_k = m$. Suppose that $N^*_{kk'} = \left(\frac{n_{kk'}}{n_{\cdot j_k}}\right)$, since $D_k = \mathrm{diag}(n_{\cdot j_k})$ and $N_{kk'} = (n_{kk'})$, then

$$B^* = \frac{1}{q^2 n}\left[n\begin{pmatrix} \underset{(j_1\times j_1)}{N^*_{11}} & \underset{(j_1\times j_2)}{N^*_{12}} & \cdots & \underset{(j_1\times j_q)}{N^*_{1q}} \\ \underset{(j_2\times j_1)}{N^*_{21}} & \underset{(j_2\times j_2)}{N^*_{22}} & \cdots & \underset{(j_2\times j_q)}{N^*_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(j_q\times j_1)}{N^*_{q1}} & \underset{(j_q\times j_2)}{N^*_{q2}} & \underset{(j_1\times j_1)}{N^*_{11}} & \underset{(j_q\times j_q)}{N^*_{qq}} \end{pmatrix}\right]$$

or $B^* = (b^*_{kk'})$, with $b^*_{kk'} = \frac{n_{kk'}}{q^2 n_{\cdot j_k}}$.

QED.

Since $N^*_{kk'} = \frac{n_{kk'}}{n_{\cdot j_k}}$, the theorem above shows that the scaled Burt matrix element is $\frac{1}{q^2}$ times the bivariate conditional probability values for each submatrix element. The quantity of $\frac{1}{q^2}$ expresses the number of submatrices formed from q variables.

Proof (1b): By previous definition $N^*_{kk} = \frac{n_{kk}}{n_{\cdot j_k}} = 1$, then $\underset{(j_k\times j_k)}{N^*_{kk}} = \underset{(j_k\times j_k)}{I_k}$ where I_k is an identity matrix of size $j_k \times j_k$. According to Proof (1a), we obtained:

$$B^* = \frac{1}{q^2}\begin{pmatrix} \underset{(j_1\times j_1)}{I_1} & \underset{(j_1\times j_2)}{N^*_{12}} & \cdots & \underset{(j_1\times j_q)}{N^*_{1q}} \\ \underset{(j_2\times j_1)}{N^*_{21}} & \underset{(j_2\times j_2)}{I_2} & \cdots & \underset{(j_2\times j_q)}{N^*_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(j_q\times j_1)}{N^*_{q1}} & \underset{(j_q\times j_2)}{N^*_{q2}} & \cdots & \underset{(j_q\times j_q)}{I_q} \end{pmatrix}.$$

It is clear that $\frac{1}{q^2}$ has been absorbed into the diagonal submatrices, then the diagonal element of $B^*$is$\frac{1}{q^2}$.

QED.

The diagonal elements of B^*, which have the same value, indicate that each variable's category has the same proportion to be chosen. The next study will investigate the sum of elements on each submatrix of B^*, as follows.

Theorem 2: Suppose $\underset{(n\times m)}{X}$ is an indicator matrix of q categorical variables. The Burt matrix and its scale are defined by $B =$ $X^T XB$ — = XTX and B∗ = 1q2nD − 1 B, then:

the sum of the $j_k \times j_{k'}$ submatrix elements of B^* is equal to $\frac{j_k}{q^2}$.

the sum of elements of the $j_k \times j_{k'}$ submatrix on B^* is equal to $\frac{j_k}{q}$ for any k.

the sum of elements of the $j_k \times j_{k'}$ submatrix of B^* is equal to the sum of elements of the $j_h \times j_{h'}$ submatrix of B^*, for $j_k = j_h$.

The total elements of B^* are equal to $\frac{m}{q}$.

Proof (2a): Suppose that $\underset{(j_k\times j_{k'})}{N^*_{kk'}} = \left(\frac{n_{kk'}}{n_{\cdot j_k}}\right)$, is a submatrix of B^*. From Theorem 1a,

$$\underset{(j_k\times j_{k'})}{N^*_{kk}} = \frac{1}{q^2}\begin{pmatrix} \frac{n_{11}}{n_{\cdot j_1}} & \frac{n_{12}}{n_{\cdot j_1}} & \cdots & \frac{n_{1k'}}{n_{\cdot j_1}} \\ \frac{n_{21}}{n_{\cdot j_2}} & \frac{n_{22}}{n_{\cdot j_2}} & \cdots & \frac{n_{2k'}}{n_{\cdot j_2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{k1}}{n_{\cdot j_k}} & \frac{n_{k2}}{n_{\cdot j_k}} & \cdots & \frac{n_{kk'}}{n_{\cdot j_k}} \end{pmatrix}.$$

Then, the sum of elements of $\underset{(j_k\times j_{k'})}{N^*_{kk'}}$ is

$$\sum_{i=1}^{k}\sum_{i'=1}^{k'}\frac{n_{ii'}}{q^2 n_{\cdot j_i}} = \frac{1}{q^2}\left[\left(\sum_{i'=1}^{k'}\frac{n_{1k'}}{n_{\cdot j_1}}\right) + \cdots + \left(\sum_{i'=1}^{k'}\frac{n_{kk'}}{n_{\cdot j_k}}\right)\right]$$
$$= \frac{1}{q^2}[(1) + (1) + \cdots + (1)] = \frac{j_k}{q^2}.$$

QED.

Theorem 2a shows that the sum of elements of each submatrix of size $j_k \times j_{k'}$is the ratio of the number categories of k-th variable and the square of the number of variables. This ratio manifests the proportion of the number of categories and the number of submatrices on B^*. The difference between the Theorems 2a and 2b is in terms of the number of sub-matrices of B^* that are considered. Theorem 2a calculates the sum of elements from one submatrix of size $j_k \times j_{k'}$, while Theorem 2b calculates the sum of the elements of some sub-matrices with row size is j_k.

Proof (2b): If there are q categorical variable, then according to Theorem 2a, the sum of elements of the $j_k \times j_{k'}$ submatrix on B^* for any k is

$$\sum_{h=1}^{q}\left(\sum_{i=1}^{k}\sum_{i'=1}^{k'}\frac{n_{ii'}}{q^2 n_{\cdot j_i}}\right) = q\left(\frac{j_k}{q^2}\right) = \frac{j_k}{q}.$$

QED.

Theorem 2b provides a simple formulation to calculate the sum of the $j_k \times j_{k'}$ submatrix elements on B^* for any k. In this formula, the sum of submatrix elements is expressed as a proportion of the number of categories from the k-th variable and the number of variables.

Proof (2c): Suppose that $\underset{(j_k\times j_{k'})}{N^*_{kk'}}$ and $\underset{(j_h\times j_{h'})}{N^*_{hh'}}$ be the submatrix of B^*. According to Theorem 2a, the sum of elements of $\underset{(j_k\times j_{k'})}{N^*_{kk'}}$ and $\underset{(j_h\times j_{h'})}{N^*_{hh'}}$ is $\frac{j_k}{q^2}$ and $\frac{j_h}{q^2}$, respectively. Since $j_k = j_h$, then the sum of elements of $\underset{(j_k\times j_{k'})}{N^*_{kk'}}$ and $\underset{(j_h\times j_{h'})}{N^*_{hh'}}$ are equal.

QED.

Theorem 2c implies that the sum of elements for any submatrix on B depends on the number of categories of variables. Thus, two or more variables with the same number of categories will have the sum of submatrices elements with the same quantity. The last evaluation was undertaken on the total number of the scaled Burt matrix elements, as below.

Proof (2d): Since $B^* = (b^*_{kk'})$, where $b^*_{kk'} = \frac{n_{kk'}}{q^2\, n_{\cdot j_k}}$, for $k, k' = 1, 2, \cdots, q$ (Theorem 1a). By applying Theorem 2a and 2b, the total elements of $B^*$are

$$\sum_{k=1}^{q}\left(\frac{j_k}{q}\right) = \frac{\sum_{k=1}^{q} j_k}{q} = \frac{m}{q}.$$

where m is the total numbers of categories of q variables.

QED.

The last statement implies that the scaled Burt matrix's total elements depend on the number of variables and the total number of categories. Furthermore, the sum of these elements is expressed as a proportion of the total number of categories and the number of variables. Thus, the total elements of this matrix will increase as the number of categories enhance.

4. Case Study

Consider the contingency table given in Table 2, originally obtained from the United States Bureau of Labor Statistics website [15] and analyzed using three-way correspondence analysis Tucker3 by Lestari et al. [14]. The data consists of 134877 American civilians and three categorical variables; *educational attainment* (X_1), *race* (X_2), and *employment status* (X_3). The *educational attainment* reflects the educational background of the civilians. It consists of four categories ($j_1 = 4$): less than a high school diploma, high school graduate and no college, some college or associate degree, and bachelor's degree and higher. The *race* was specified into three categories ($j_2 = 3$); white, black, and Asia. The *employment status* is divided into two categories ($j_3 = 2$); employment and unemployment. The Burt matrix B and its scaled B^* for the table above, derived by its indicator matrix form, are displayed in Fig. 5.

Table 2 Labor data for 134877 American civilians 25 years and over by educational attainment, race, and employment status

Categories	**Employment**			**Unemployment**			**Marginal**
	White	**Black**	**Asian**	**White**	**Black**	**Asian**	**(X_1)**
Less than a high school diploma	7690	1038	481	461	148	22	9840
High school graduate and no college	26710	4889	1474	1,127	421	41	34662
Some college or associate degree	28388	5230	1375	987	321	45	36346
Bachelor's degree and higher	42662	4874	5261	902	182	148	54029
Marginal (X_2)	108927	17103	8847				134877
Marginal (X_3)		130072			4805		

(a)

$$\mathbf{B} = \begin{bmatrix} 9840 & 0 & 0 & 0 & 8151 & 1186 & 503 & 9209 & 631 \\ 0 & 34662 & 0 & 0 & 27837 & 5310 & 1515 & 33073 & 1589 \\ 0 & 0 & 36346 & 0 & 29375 & 5515 & 1420 & 34993 & 1353 \\ 0 & 0 & 0 & 54029 & 43564 & 5056 & 5409 & 52797 & 1232 \\ 8151 & 27837 & 29375 & 43564 & 108927 & 0 & 0 & 105450 & 3477 \\ 1186 & 5310 & 5515 & 5056 & 0 & 17103 & 0 & 16031 & 1072 \\ 503 & 1515 & 1420 & 5409 & 0 & 0 & 8847 & 8591 & 256 \\ 9209 & 33073 & 34993 & 52797 & 105450 & 16031 & 8591 & 130072 & 0 \\ 631 & 1589 & 1353 & 1232 & 3477 & 1072 & 256 & 0 & 4805 \end{bmatrix}$$

(b)

$$\mathbf{B}^* = \frac{1}{9} \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{8151}{9840} & \frac{1186}{9840} & \frac{503}{9840} & \frac{9209}{9840} & \frac{631}{9840} \\ 0 & 1 & 0 & 0 & \frac{27837}{34662} & \frac{5310}{34662} & \frac{1515}{34662} & \frac{33073}{34662} & \frac{1589}{34662} \\ 0 & 0 & 1 & 0 & \frac{29375}{36346} & \frac{5515}{36346} & \frac{1420}{36346} & \frac{34993}{36346} & \frac{1353}{36346} \\ 0 & 0 & 0 & 1 & \frac{43564}{54029} & \frac{5056}{54029} & \frac{5409}{54029} & \frac{52797}{54029} & \frac{1232}{54029} \\ \frac{8151}{108927} & \frac{27837}{108927} & \frac{29375}{108927} & \frac{43564}{108927} & 1 & 0 & 0 & \frac{105450}{108927} & \frac{3477}{108927} \\ \frac{1186}{17103} & \frac{5310}{17103} & \frac{5551}{17103} & \frac{5056}{17103} & 0 & 1 & 0 & \frac{16031}{17103} & \frac{1072}{17103} \\ \frac{503}{8847} & \frac{1515}{8847} & \frac{1420}{8847} & \frac{5409}{8847} & 0 & 0 & 1 & \frac{8591}{8847} & \frac{256}{8847} \\ \frac{9209}{130072} & \frac{33073}{130072} & \frac{34993}{130072} & \frac{52797}{130072} & \frac{105450}{130072} & \frac{16031}{130072} & \frac{8591}{130072} & 1 & 0 \\ \frac{631}{4805} & \frac{1589}{4805} & \frac{1353}{4805} & \frac{1232}{4805} & \frac{3477}{4805} & \frac{1072}{4805} & \frac{256}{4805} & 0 & 1 \end{bmatrix}$$

Fig. 5 Burt matrix construction: (a) The Burt matrix corresponding to the data in Table 2; (b) The scaled Burt matrix elements by applying Theorem 1a

Easily to verify the elements of this matrix by comparing the calculations of $B^* = \frac{1}{q^2 n} D^{-\mathbf{1}} B$. The results show that Theorem 1 accomplished. Furthermore, Theorem 2a ensure that the sum of element of the top-red marked submatrix is $\frac{4}{9}$, since

$$\sum_{k=1}^{4} \sum_{k'=1}^{3} \frac{n_{kk'}}{q^2 n_{\cdot j_k}} = \frac{1}{3^2} \left[\left(\sum_{\ell=1}^{3} \frac{n_{1k'}}{n_{\cdot j_1}} \right) + \left(\sum_{\ell=1}^{3} \frac{n_{2k'}}{n_{\cdot j_2}} \right) + \left(\sum_{\ell=1}^{3} \frac{n_{3k'}}{n_{\cdot j_3}} \right) + \left(\sum_{\ell=1}^{3} \frac{n_{4k'}}{n_{\cdot j_4}} \right) \right]$$

$$= \frac{1}{3^2} \left[\left(\frac{8{,}151 + 1{,}186 + 503}{9840} \right) + \left(\frac{27{,}837 + 5{,}310 + 1{,}515}{34{,}662} \right) \right.$$

$$+\left(\frac{29,375+5,515+1,420}{36,346}\right) + \left(\frac{43,564+5,056+5,409}{54,029}\right)\Bigg]$$

$$=\frac{4}{9}$$

where 4 refers to the number of categories for X_1, and 9 - to the squared of the number of variables. The position of the four-row profiles of this submatrix can be plotted in three-dimensional space as given in Fig. 6a.

The profile points lie precisely in the plane defined by the triangle that joins the coordinates $[1/9 \quad 0 \quad 0]$, $[0 \quad 1/9 \quad 0]$, and $[0 \quad 0 \quad 1/9]$. Each side is considered to have a length of $1/9$ (Fig. 6b). This quantity obtained is obtained from q^{-2}, where q is the number of variables. Additionally, the two red marked submatrices imply that the scaled Burt matrix is not symmetric as the original Burt matrix.

Now, let pay attention to the top three submatrices in Fig. 4b. Let these matrices be denoted as $\underset{(4\times4)}{D_1}$, $\underset{(4\times3)}{N_{12}}$, and $\underset{(4\times2)}{N_{13}}$, respectively. Based on Theorem 2b, the sum of these submatrix elements is $\frac{4}{3}$, since

$$\sum_{h=1}^{3}\left(\sum_{k=1}^{4}\sum_{k'=1}^{3}\frac{n_{kk'}}{q^2 n_{\cdot j_k}}\right) = \sum_{k=1}^{4}\sum_{k'=1}^{4}\frac{n_{kk'}}{3^2 n_{\cdot j_k}} + \sum_{k=1}^{4}\sum_{k'=1}^{3}\frac{n_{kk'}}{3^2 n_{\cdot j_k}} + \sum_{k=1}^{4}\sum_{k'=1}^{2}\frac{n_{kk'}}{3^2 n_{\cdot j_k}}$$

$$=\frac{4}{9}+\frac{4}{9}+\frac{4}{9}$$

$$=\frac{4}{3}$$

This result leads us to get the conclusion that the total elements of B^*in Fig. 4b is equal to 3, since

$$\sum_{k=1}^{3}\left(\frac{j_k}{3}\right) = \frac{4}{3}+\frac{3}{3}+\frac{2}{3} = \frac{9}{3} = 3.$$

It shows that the total elements in the scaled Burt matrix do not depend on the number of individuals n, but only depend on the number of variables q and categories of variables m.

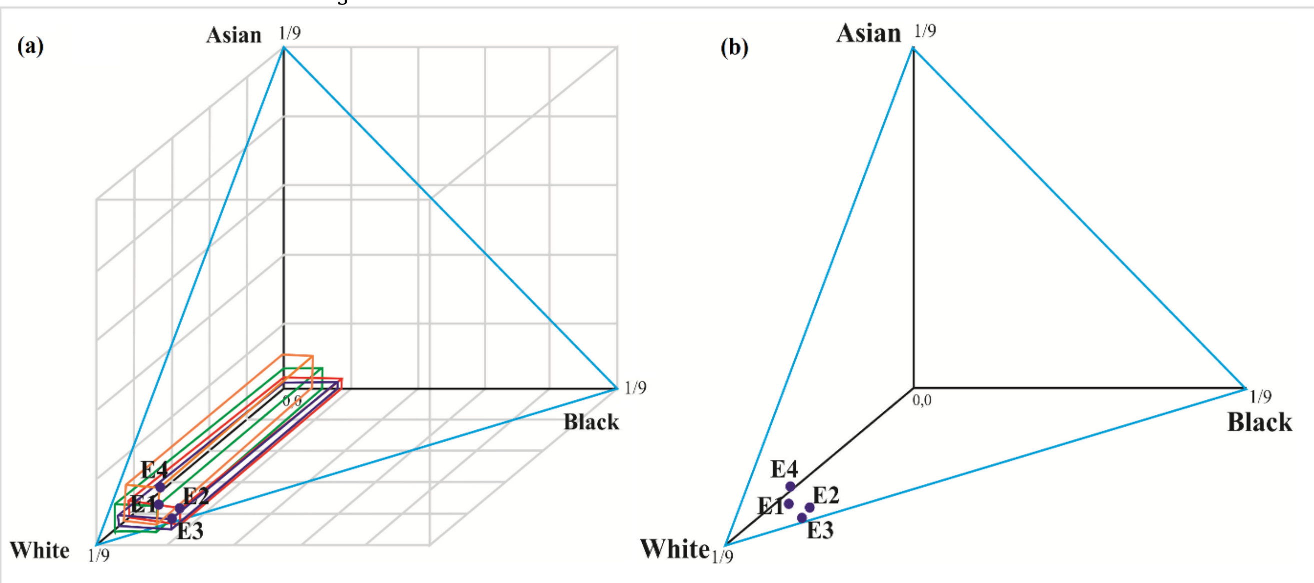

Fig. 6 Row profiles of the top-red marked submatrices on the scaled Burt matrix for labor data in Table 2

5. Conclusion

This study re-proposed the scale matrix of the Burt matrix, whose elements are the scale values of the categories of a variable, then so-called the scaled Burt matrix. This scaled matrix is then analyzed by eigendecomposition to get coordinates, which depicts the association's nature among variables. This study aims to identify the characteristics of the scaled matrix elements through an algebraic approach to find the link between the Burt matrix and its scale matrix in representing the variables associations. The results show that the elements of this matrix represent the association of the variables. The investigation of the sum of the matrix elements, both for each submatrix or overall, yields fascinating values. It is still related to the proportion of the number of categories with the number of categorical variables (as hypothesized). For example, the sum of elements for any submatrix on the scaled Burt matrix depends on the number of categories of variables. The results lead to the conclusion that the scaled Burt matrix is akin to a profiled version of the Burt matrix. The advantage of deals with a scale matrix is to standardize the independent features present in the data in a fixed range. It is performed to handle highly varying magnitudes or values, or units. This study's results are an early stage that still provides some open problems to be explored in future work.

Acknowledgment

This research was supported by the Indonesian Ministry of Research, Technology, and Higher Education, through the Doctoral Dissertation Research Grant (PDD 2020-2021). The authors thank the anonymous reviewers for providing constructive comments to improve the earlier version of the manuscript.

References

[1] HJELLBREKKE J. *Multiple correspondence analysis for the social sciences.* Taylor & Francis Group, London, 2019.

[2] BEH E. J., & LOMBARDO R. Multiple and multi-way correspondence analysis. *Advanced Review*, 2019, 11(5): 1-11. https://doi.org/10.1002/wics.1464

[3] YANG Y., POUYANFAR S., and TIAN H. IF-MCA: importance factor-based multiple correspondence analysis formultimedia data analytics. *IEEE Transactions on Multimedia*, 2018, 20: 1024-1032. https://doi.org/10.1109/TMM.2017.2760623

[4] GOODWILL A., & MELOY J. R. Visualizing the relationship among indicators for lone actor terrorist attacks: multidimensional scaling ang the TRAP-18. *Behavioral Sciences & the Law*, 2019, 37(5): 522-539. https://doi.org/10.1002/bsl.2434

[5] LESTARI K. E., PASARIBU U. S., INDRATNO S. W., and GARMINIA H. The reliability of crash car protection level based on the circle confidence region on the correspondence plot. *IOP Conference Series: Materials Science and Engineering*, 2019, 598: 012061. https://doi.org/10.1088/1757-899X/598/1/012061

[6] GREENACRE M. J. *Use of correspondence analysis in clustering a mixed-scale data set with missing data.* Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, 2019. https://doi.org/10.13140/RG.2.2.23439.43684

[7] YUDHANEGARA M. R., & LESTARI K. E. Clustering for multi-dimensional data set: a case study on educational data. *Journal of Physics: Conference Series*, 2019, 1280: 042025. https://doi.org/10.1088/1742-6596/1280/4/042025

[8] ROVAN V. U., & ROVAN J. An exploration of diabetic foot screening procedures data by a multiple correspondence analysis. *Slovenian Journal of Public Health*, 2017, 56: 65-73. https://doi.org/10.1515/sjph-2017-0009

[9] FRED R. M., MWAURA F., OGWAL F., MASIGA M., AKULLO M., and OKURUT T. O. Mitigating impacts of projects on biodiversity conservation in Uganda. *Journal of Ecosystem and Ecography*, 2017, 7: 232-235. https://doi.org/10.4172/2157-7625.1000232

[10] BRUNETTE M., BOURKE R., HANEWINKEL M., and YOUSEFPOUR R. Adaptation to climate change in forestry: a multiple correspondence analysis. *Forests*, 2018, 9(1): 20. https://doi.org/10.3390/f9010020

[11] BEH E. J., & LOMBARDO R. Multiple and multi-way correspondence analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2019, 11(5): e1464. https://doi.org/10.1002/wics.1464

[12] LESTARI K. E., PASARIBU U. S., INDRATNO S. W., and GARMINIA H. Generating roots of cubic polynomials by Cardano's approach on correspondence analysis. *Heliyon*, 2020, 6(6): e03998. https://doi.org/10.1016/j.heliyon.2020.e03998

[13] LESTARI K. E., PASARIBU U. S., INDRATNO S. W., and GARMINIA H. The comparative analysis of dependence for three-way contingency table using Burt matrix and Tucker3 in correspondence analysis. *Journal of Physics: Conference Series*, 2019, 1245: 012056. https://doi.org/10.1088/1742-6596/1245/1/012056

[14] LESTARI K. E., PASARIBU U. S., and INDRATNO S. W. Graphical depiction of three-way association in contingency table using higher-order singular value decomposition Tucker3. *Journal of Physics: Conference Series*, 2019, 1280: 022035. https://doi.org/10.1088/1742-6596/1280/2/022035

[15] UNITED STATES BUREAU OF LABOR STATISTICS. http://www.bls.gov/

参考文:

[1] HJELLBREKKE J. 社會科學的多重對應分析。泰勒和弗朗西斯集團，倫敦，2019。

[2] BEH E. J., 和 LOMBARDO R. 多路和多路對應分析。高級評論，2019，11(5)：1-11。https://doi.org/10.1002/wics.1464

[3] YANG Y.，POUYANFAR S. 和 TIAN H. 中頻-馬華：基於重要性因子的多媒體數據分析多重對應分析。電氣工程師學會多媒體彙刊，2018，20：1024-1032。https://doi.org/10.1109/TMM.2017.2760623

[4] GOODWILL A. 和 MELOY J. R. 形象化顯示了獨立演員恐怖襲擊指標之間的關係：多維縮放和陷阱-18。行為科學與法律，2019，37(5)：522-539。https://doi.org/10.1002/bsl.2434

[5] LESTARI K. E.，PASARIBU U. S.，INDRATNO S. W. 和 GARMINIA H. 基於對應圖上的圓置信區域的防撞車保護等級的可靠性。眼壓會議系列：材料科學與工程，2019，598：012061. https://doi.org/10.1088/1757-899X/598/1/012061

[6] GREENACRE M. J. 在將缺失數據的混合規模數據集聚類中使用對應分析。巴塞羅那蓬蓬法布拉大學經濟與商業系，2019. https://doi.org/10.13140/RG.2.2.23439.43684

[7] YUDHANEGARA M. R. 和 LESTARI K. E. 多維數據集的聚類：教育數據的案例研究。物理學雜誌：會議系列，2019，1280: 042025. https://doi.org/10.1088/1742-6596/1280/4/042025

[8] ROVAN V. U. 和 ROVAN J. 通過多重對應分析探索糖尿病足篩查程序數據。斯洛文尼亞公共衛生雜誌，2017，56：65-73。https://doi.org/10.1515/sjph-2017-0009

[9] FRED R. M.，MWAURA F.，OGWAL F.，MASIGA M.，AKULLO M. 和 OKURUT T. O。減輕項目對烏干達生物多樣性保護的影響。生態系統與生態學報，2017，7：232-235。https://doi.org/10.4172/2157-7625.1000232

[10] BRUNETTE M.，BOURKE R.，HANEWINKEL M. 和 YOUSEFPOUR R. 適應林業中的氣候變化：多重對應分析。森林，2018，9(1): 20. https://doi.org/10.3390/f9010020

[11] BEH E. J. 和 LOMBARDO R. 多向和多向對應分析。威利跨學科評論：計算統計，2019，11(5)：e1464。https://doi.org/10.1002/wics.1464
[12] LESTARI K. E.，PASARIBU U. S.，INDRATNO S. W. 和 GARMINIA H. 通過卡爾達諾的對應分析方法生成三次多項式的根。赫利永，2020，6（6）：e03998。https://doi.org/10.1016/j.heliyon.2020.e03998
[13] LESTARI K. E.，PASARIBU U. S.，INDRATNO S. W. 和 GARMINIA H. 在對應分析中使用伯特矩陣和塔克 3 對三向列聯表的依賴性進行比較分析。物理學雜誌：會議系列，2019, 1245: 012056. https://doi.org/10.1088/1742-6596/1245/1/012056
[14] LESTARI K. E.，PASARIBU U. S. 和 INDRATNO S. W. 使用高階奇異值分解塔克 3 在列聯表中進行三向關聯的圖形描述。物理學雜誌：會議系列，2019，1280：022035. https://doi.org/10.1088/1742-6596/1280/2/022035
[15] 美國勞工統計局。http://www.bls.gov/