Graphical depiction of three-way association in contingency table using higher-order singular value decomposition Tucker3

K E Lestari*, U S Pasaribu and S W Indratno
Institut Teknologi Bandung, Bandung, Indonesia

*Corresponding author’s email: karunia.eka.lestari@s.itb.ac.id

Abstract. The analysis of the association between the variables of a three-way table requires a very different approach to the conventional correspondence analysis techniques. The conventional technique involves by collapsing three-way table into two-way table form. By doing so, no information about three-way interaction among variables. To included interaction among three variables at once, we consider higher-order singular value decomposition (HOSVD) Tucker3. Through the HOSVD Tucker3 decomposition, we can derive marginal of row, column, and tube categories. Using those marginals, we can easily visualize dependence of the three categories. As a case study, Tucker3 is applied to household data which containing information on race, educational attainment and employment status.

1. Introduction

The graphical display of data has been a fundamentally important aspect of mathematics, especially statistics. It is reflected in many university and training courses throughout the world. In particular, an introduction to the fundamentals of statistics often stars with a first view at data visualisation. Many introductory statistics textbooks include a variety of different graphical techniques for numerical and categorical data [1-3]. Also, many of the graphical methods that have been developed are fairly simple to construct and provide a quick and intuitive summary of data.

One of the interesting methods to be studied and developed is correspondence analysis. Correspondence analysis (CA) is a graphical method for representing information of categorical variable in contingency table [4]. CA provides an intuitive visual observation of the association between variables at the category level. At present, various data can be accessed easily by anyone and anytime. The problem more to be complex when the data consisting of three categorical variables. The data collected from a numbers of observations based on three categorical variabel are organised into a three-way contingency table. The characteristic of this data is a data cube, also called a box by [5]. Algebraically, the data in the three-way contingency table is a representation of the cube matrix.

The analysis of the association between the variables of a three-way table requires a very different approach to the conventional correspondence analysis techniques. The traditional correspondence analysis involves by reducing three-way contingency table into a two-way form. By doing so, no information about three-way interaction among variables. In the last few decades, Tucker3 considered a very important method to solve the decomposition problems for three-way contingency table. Tucker3 originally proposed by [6]. Various analyses have been performed in different fields, see [7-10, 2].
respect to three-way contingency table, in this paper, we used the Tucker3 for the exploration of data and graphical depiction of association of three categorical variables.

This paper is organized as follows. Section 2 describes the graphical depiction method for three-way contingency table using Tucker3. In section 3, we used the household data for a case study. The data consists three categorical variables, i.e. race, educational attainment and employment status. The summary and future works are presented in section 4. The novelty is we derive marginal of row, column, and tube categories by matricization from HOSVD Tucker3.

2. Methods

In this section, we present the CA3 as a graphical depiction method for three-way contingency table. This table contains the counts (frequencies) of items for a cross-classification of three categorical variables. Let \(X\), \(Y\), and \(Z\) be three categorical variables observed on \(n\) individuals/objects, with \(I\), \(J\), and \(K\) categories, respectively. This observations are organised into a three-way contingency table that consists of \(I\) row, \(J\) column, and \(K\) tube categories, respectively. The row categories are associated with variable \(X\), the column categories are associated with variable \(Y\), and the tube categories are associated with variable \(Z\).

**Table 1. Three-way contingency table form.**

| Category | Tube 1 | | | Tube K |
|----------|--------| | | Column 1 | Column 2 | \(\cdots\) | Column 1 | Column 2 | \(\cdots\) | Column 1 | Column 2 | \(\cdots\) | Column 1 | Column 2 | \(\cdots\) | Column 1 |
| Row 1    | \(n_{111}\) | \(n_{121}\) | \(\cdots\) | \(n_{1J1}\) | \(n_{112}\) | \(\cdots\) | \(n_{1J2}\) | \(\cdots\) | \(n_{11K}\) | \(n_{12K}\) | \(\cdots\) | \(n_{1JK}\) |
| Row 2    | \(n_{211}\) | \(n_{221}\) | \(\cdots\) | \(n_{2J1}\) | \(n_{212}\) | \(\cdots\) | \(n_{2J2}\) | \(\cdots\) | \(n_{21K}\) | \(n_{22K}\) | \(\cdots\) | \(n_{2JK}\) |
| Row 3    | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) |

Denote \(\mathcal{R}^{I \times J \times K}\) to be the space of real functions depending on \(i, j, \) and \(k\) for \(i = 1, \cdots, I, j = 1, \cdots, J\) dan \(k = 1, \cdots, K\). Let \(\mathcal{N}\) be the three-way contingency table of size \(I \times J \times K\), where the \((i, j, k)\)th cell frequency is \(n_{ijk}\). Data visualization of the three-way contingency table is shown in Figure 1.

**Figure 1.** Visualization of data cube [2].

Denote \(\mathcal{P} = \left(\frac{N}{n}\right)\) to be the cube matrix of joint relative frequencies with \(p_{ijk}\) its \((i, j, k)\)th element so that \(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} i p_{ijk} = 1\). Let \(p_{i.} = \sum_{j=1}^{J} \sum_{k=1}^{K} i p_{ijk}\), \(p_{.i.} = \sum_{j=1}^{J} \sum_{k=1}^{K} i p_{ijk}\), and \(p_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} i p_{ijk}\). Let \(p_{ij.} = \sum_{k=1}^{K} i p_{ijk}\), \(p_{.ij} = \sum_{k=1}^{K} i p_{ijk}\), and \(p_{i.j} = \sum_{k=1}^{K} i p_{ijk}\). The association structure (dependency) between the three categorical variables, \(X, Y\), and \(Z\) in a three-way contingency table is represented by the cube matrix \(\mathcal{S} = (s_{ijk})\). The element of \(s_{ijk}\) can be expressed in term of deviations from the three-way independence model such that

\[
\mathcal{S} = (s_{ijk}), \text{ where } s_{ijk} = \frac{p_{ijk} - p_{i..}p_{.j.}p_{..}}{p_{i..}p_{.j.}p_{..}}
\]
The vectors obtained from columns of $\mathbf{S}$ are not always orthogonal. As a consequence, the dependency values of the row, column, and tube categories cannot always be depicted to Cartesian coordinates which mutually orthogonal [11]. Therefore, new bases are required which are linear combinations of row, column or tube vectors of $\mathbf{S}$, called as principal coordinates. In order to obtain these new bases, $\mathbf{S}$ is decomposed using Tucker3. This decomposition begins with reordering the elements of $\mathbf{S}$ based on column, row, or tube into a matrix, say $\mathbf{S}(1)$, $\mathbf{S}(2)$, or $\mathbf{S}(3)$. This process is called matricization, also known as unfolding or flattening [12, 13]. Figure 2 illustrates the matricization of $\mathbf{S}$.

![Figure 2](image.png)

**Figure 2.** The process of reordering the elements of $\mathbf{S}$ into a matrix. $\mathbf{S}(1)$, $\mathbf{S}(2)$, and $\mathbf{S}(3)$ respectively called as column, row, and tube fibers.

Before further discussion of Tucker3, we defined operations that will be used on Tucker3 as follows.

**Definition 1.** Let $\mathbf{S} \in \mathbb{R}^{I \times J \times K}$, $\mathbf{U} \in \mathbb{R}^{P \times I}$, $\mathbf{V} \in \mathbb{R}^{Q \times J}$ and $\mathbf{W} \in \mathbb{R}^{R \times K}$. The $n$-mode product of a cube matrix $\mathbf{S}$ with a matrix $\mathbf{U}$ or $(\mathbf{V}$ or $\mathbf{W})$, for $n = 1, 2, 3$, defined as

$$\mathbf{S} \times_1 \mathbf{U} = \sum_{i=1}^{I} s_{ijk} u_{pi}, \quad \mathbf{S} \times_2 \mathbf{V} = \sum_{j=1}^{J} s_{ijk} v_{qj}, \quad \text{and} \quad \mathbf{S} \times_3 \mathbf{W} = \sum_{k=1}^{K} s_{ijk} w_{rk}$$

By the definition, each mode-$n$ fiber is multiplied by the matrix $\mathbf{U}$ ($\mathbf{V}$, or $\mathbf{W}$). The idea can also be expressed in term of $\mathbf{S}(n)$:

$$\mathbf{S} = \mathbf{G} \times_n \mathbf{U} \iff \mathbf{S}(n) = \mathbf{U} \mathbf{G}(n), \quad \text{dengan} \quad n = 1, 2, 3$$

(2)

For distinct modes in a series of multiplications, the order of the multiplication is irrelevant [13], i.e.,

$$\mathbf{S} \times_m \mathbf{U} \times_n \mathbf{V} = \mathbf{S} \times_n \mathbf{V} \times_m \mathbf{U} \quad (m \neq n)$$

(3)

If the modes are the same, then

$$\mathbf{S} \times_n \mathbf{U} \times_n \mathbf{V} = \mathbf{S} \times_n (\mathbf{VU})$$

(4)

Meanwhile, the tensor multiplication of the matrices $\mathbf{U}$ and $\mathbf{V}$ is defined as follows

**Definition 2.** Let $\mathbf{U} = (u_{pi}) \in \mathbb{R}^{P \times I}$ and $\mathbf{V} = (v_{qj}) \in \mathbb{R}^{Q \times J}$. The tensor product or Kronecker product of matrices $\mathbf{U}$ and $\mathbf{V}$, denoted by $\mathbf{U} \otimes \mathbf{V}$, and defined by
\[ U \otimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1i}V \\ u_{21}V & u_{22}V & \cdots & u_{2i}V \\ \vdots & \vdots & \ddots & \vdots \\ u_{pi}V & u_{p2}V & \cdots & u_{pi}V \end{pmatrix} \]

where \( U \otimes V \in \mathbb{R}^{PQ \times J} \).

After obtaining the matrix \( S_1, S_2, \) and \( S_3 \), the matrices are decomposed using SVD to obtain orthogonal matrices \( U, V \) and \( W \) which represented the rows, columns, and tubes category. Furthermore, we determined the core \( \mathcal{G} \) which represented the association or interaction between each category, such that \( \mathcal{S} \) can be decomposed as a multiplication of \( \mathcal{G}, U, V \) and \( W \). This decomposition is known as Tucker3 decomposition. Mathematically, \cite{14} formulated Tucker3 as a representation of higher-order SVD (HOSVD) as given in Theorem 1 and Theorem 2 below.

**Theorem 1.** (Van Loan, 2015) Let \( \mathcal{S} \in \mathbb{R}^{I \times J \times K} \). If \( S_1 = U D_1 A^T, S_2 = V D_2 B^T, \) and \( S_3 = W D_3 C^T, \) respectively, are SVDs of \( S_1, S_2, S_3 \), where \( U \in \mathbb{R}^{I \times P}, V \in \mathbb{R}^{J \times Q}, \) \( W \in \mathbb{R}^{K \times R}, \) such that \( \mathcal{G} = S_1 U^T \times_2 V^T \times_3 W^T \in \mathbb{R}^{P \times Q \times R} \), then

\[ \mathcal{S} = \mathcal{G} \times_1 U \times_2 V \times_3 W \]

is the higher-order SVD of \( \mathcal{S} \).

**Theorem 2.** (Van Loan, 2015) Let \( \mathcal{S} \in \mathbb{R}^{I \times J \times K} \). If \( S_1 = U D_1 A^T, S_2 = V D_2 B^T, \) and \( S_3 = W D_3 C^T, \) respectively, are SVDs of \( S_1, S_2, S_3 \), where \( U \in \mathbb{R}^{I \times P}, V \in \mathbb{R}^{J \times Q}, \) \( W \in \mathbb{R}^{K \times R}, \) and \( \mathcal{S} = \mathcal{G} \times_1 U \times_2 V \times_3 W \), then

\[
\begin{align*}
S_1 &= U G_1 (W \otimes V)^T \\
S_2 &= V G_2 (W \otimes U)^T \\
S_3 &= W G_3 (V \otimes U)^T
\end{align*}
\]

Based on Tucker3, the row coordinates (\( F \)) are defined as

\[ F = U G_{(P \times Q)} (\approx f_{iqr} = \sum_{p=1}^{P} u_{ip} g_{pqr}) \quad (5) \]

Equation (5) are referred to as principal coordinates since the principal axes are expressed as a linear combination of the row components and the core elements [2]. Furthermore, the \((j, k)\)th pair of column-tube categories will be represented by a single point. Hence, the column-tube coordinates are defined as follows

\[ H = (V \otimes W) G_{(Q \times R)} (\approx h_{ijk} = \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} v_{jq} w_{kr}) \quad (6) \]

Equations (5) and (6), can be used to calculate the total inertia of the three-way contingency table:

\[ \text{Total inersia} = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{p=1}^{P} p \cdot j \cdot p \cdot k \cdot h_{ijk}^2 \]

\[ = \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{r=1}^{R} h_{ijk}^2 f_{iqr} \quad (7) \]

The percentage contribution of the \( jk \)th column-tube category on the \( p \)th axis can be quantified by

\[ \text{Cont}_{jk,p} = 100 \times \frac{h_{ijk}^2}{\text{Total inersia}} \quad (8) \]

**3. Results and Discussion**

Table 2 presents the household data consists three categorical variables, i.e. race, educational attainment and employment status that observed in 134,859 individuals. There are three categories for the race: white (R1), black or African American (R2), and Asian (R3). The educational attainment included four categories: less than a high school diploma (C1), high school graduate and no college (C2), some college
or associate degree (C3) and bachelor’s degree and higher (C4). While the employment status consists of two categories, i.e. employment (T1) and unemployment (T2). Data obtained from Bureau of Labor Statistics website, www.bls.gov.

Table 2. Employment status of the civilian noninstitutional population 25 years and over by educational attainment and race.

| Category                  | Employment     | Unemployment |
|---------------------------|----------------|--------------|
|                           | less than a high school diploma | high school graduate and no college | some college or associate degree | bachelor’s degree and higher |
|                           | less than a high school diploma | high school graduate and no college | some college or associate degree | bachelor’s degree and higher |
| White                     | 7.690          | 26.710       | 28.388       | 42.662         | 461 | 1.127 | 987 | 902 |
| Black or African American | 1.038          | 4.889        | 5.230        | 4.874         | 148 | 412   | 321 | 182 |
| Asian                     | 481            | 1.474        | 1.375        | 5.261         | 22  | 41    | 45  | 148 |

By using Equation (1), the cube matrix of $\mathcal{S}$ is obtained in Figure 3. The columns on $\mathcal{S}$ indicate dependency values between race, educational attainment, and employment status. In order to be depicted on the Cartesian coordinates, $\mathcal{S}$ be decomposed using Tucker3, as in Figure 3.

Figure 3. Tucker3 decomposition of $\mathcal{S}$ for contingency table in Table 2.

Next, using Equation (5) and (6), we have the row ($\mathcal{F}$) and column-tube ($\mathcal{H}$) coordinates as follows
Figure 4. The row (F) and column-tube (H) coordinates on the pth axis, where p = 1, 2, ..., 8 for row category and p = 1, 2, 3 for column-tube category.

These principal coordinates are depicted on the pth axis, are referred to as correspondence plot. The correspondence plot on the first axis (R) and a second axis (R^2) illustrated in Figure 5 and 6. The figures are the graphical depiction of the association between race, educational attainment, and employment status. Figure 5 shows that the white race (R1) has strong associations with a bachelor’s degree or higher and unemployment (C4T2). It is reflected by the position of R1 and C4T2 coordinate points are close to each other (see the blue circle marked). On the other hand, the position of R2 is far away from CIT2, it is indicated that the Black or African American race (R2) has a lower association with less than a high school diploma and unemployment (C1T2). Furthermore, the red circle marked shown that the Asian race (R3) has a strong association with employment status (T1) based on all educational attainment categories (C1-C4). This indicates that the majority of the Asian race population has an employment, regardless of their educational attainment.

![Correspondence plot on R](image1.png)

![Correspondence plot on R^2](image2.png)

Figure 5. (a) Correspondence plot on R and (b) Correspondence plot on R^2.

The quality of the plot is identified based on the percentage of variance which is contained in each dimension of the plot. The percentage of variance on each principal coordinate is determined by the inertia, as summarized in Table 3.
Table 3. Inertia on each principal coordinate for household data in Table 2

| Label | Category | Axis 1 | Axis 2 |
|-------|----------|--------|--------|
| R1    | White    | 3.16   | 0.10   |
| R2    | Black or African American | 82.86 | 0.16 |
| R3    | Asian    | 2.40   | 2.29   |
| C1T1  | Less than a high school diploma and employment | 0.04 | 2.20 |
| C1T2  | Less than a high school diploma and unemployment | 48.40 | 0.07 |
| C2T1  | High school graduate, no college and employment | 0.95 | 0.28 |
| C2T2  | High school graduate, no college and unemployment | 28.45 | 0.01 |
| C3T1  | Some college or associate degree and employment | 0.46 | 0.01 |
| C3T2  | Some college or associate degree and unemployment | 9.04 | 0.35 |
| C4T1  | Bachelor’s degree or higher and employment | 0.95 | 1.19 |
| C4T2  | Bachelor’s degree or higher and unemployment | 2.03 | 3.71 |

Based on Table 3, the quality of plots in Figures 5 and 6, respectively, are 88.41% and 90.95%. Thus, Figure 6 provided more information about the association between of three categorical variables in the contingency table. The contribution of the row and column-tube categories to the association depicted along the first and second axes is reported in Table 4.

Table 4. Contributions of row and column-tube categories to the total inertia in first and second axis.

| Label | Category | Axis 1 | Axis 2 |
|-------|----------|--------|--------|
| R1    | White    | 3.16   | 0.10   |
| R2    | Black or African American | 82.86 | 0.16 |
| R3    | Asian    | 2.40   | 2.29   |
| C1T1  | Less than a high school diploma and employment | 0.04 | 2.20 |
| C1T2  | Less than a high school diploma and unemployment | 48.40 | 0.07 |
| C2T1  | High school graduate, no college and employment | 0.95 | 0.28 |
| C2T2  | High school graduate, no college and unemployment | 28.45 | 0.01 |
| C3T1  | Some college or associate degree and employment | 0.46 | 0.01 |
| C3T2  | Some college or associate degree and unemployment | 9.04 | 0.35 |
| C4T1  | Bachelor’s degree or higher and employment | 0.95 | 1.19 |
| C4T2  | Bachelor’s degree or higher and unemployment | 2.03 | 3.71 |

Table 4 shows us that the most important contributors to the first axis are black or African American (R2) and less than a high school diploma and unemployment (C1T2). As in Figure 5 that R2 and C1T2 are depicted farthest from the origin. The categories that give the greatest contribution to total inertia in the second axis are Asian (R3) and a bachelor's degree or higher and unemployment (C4T2).

4. Conclusion

The role of HOSVD in Tucker3 decomposition is to derive marginal of row, column, and tube categories. Using those marginals, we can easily visualize dependence of the three categories. The association (dependency) between each category of variables is represented by the core $G$. Characteristics of core $G$ obtained from Tucker3 are not diagonal (superdiagonal). Superdiagonalization of the core $G$ is impossible, but it is possible to make $G$ that consist as many of zero elements. However, many of zero elements in the $G$ core will make calculations easier. In the future work, we will simplify the core structure in some way so that most of the elements of $G$ are zeros.

5. References

[1] Greenacre M J 1984 Theory and application of correspondence analysis (London: Academic Press)
[2] Beh E J and Lombardo R 2014 Correspondence analysis theory, practice, and new strategies (West Sussex: John Wiley & Sons, Ltd)
[3] Watkins J C 2015 An introduction to the science of statistics (Newyork: Wiley)
[4] Rencher A C and Christensen W F 2012 Methods of multivariate analysis third edition (New Jersey: John Wiley & Sons Inc)
[5] Kroonenberg P M 2008 Applied Multiway Data Analysis (Newyork: John Wiley & Sons Inc)
[6] Tucker L R 1966 Some mathematical notes on three-mode factor analysis *Psychometrika*, 31 3 279-311

[7] Kroonenberg P M 1985 *J. Applied Psychological Measurement* 9 83

[8] Rocci R and Ten Berge J M F 2002 *J. Psychometrika* 67 351

[9] Smilde A, Bro R, and Geladi P 2004 *Multi-way analysis applications in the chemical sciences* (West Sussex: John Wiley & Sons Ltd)

[10] Gallo M and Buccianti A 2013 Weighted principal component analysis for compositional data: application example for the water chemistry of the Arno river (Tuscany, central Italy) *Environmetrics* 24 4 269-277

[11] Ginanjar I, Pasaribu U S, and Indratno S W 2014 *Proc. Int. Conf. on Technology, Informatics, Management, Engineering & Environment* vol 2 (Bandung: Indonesia/IEEE)

[12] Martin C D M 2004 *Tensor decompositions* (California: American Institute of Mathematics)

[13] Kolda, T G and Bader B W 2009 Tensor decompositions and applications *SIAM review* 51 3 455-500

[14] Van Loan C F 2015 *Structured matrix computations from structured tensors* (Italia: Cornell University)

**Acknowledgments**
The authors thanks to Dr. Hanni Garminia Y who have provided direction about Tucker3, we also thank the editor. Work supported by BOPTN FMIPA ITB.