

# NIASRA

NATIONAL INSTITUTE FOR APPLIED  
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research  
Australia***

**University of Wollongong, Australia**

**Working Paper**

21-23

**Profile Transformations for Reciprocal  
Averaging and Singular Value Decomposition**

Ting-Wu Wang, Eric J. Beh, Rosaria Lombardo, and Ian  
W. Renner

*Copyright © 2023 by the National Institute for Applied Statistics Research Australia, UOW.  
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,  
Wollongong NSW 2522, Australia T: +61 2 42215076. E: [karink@uow.edu.au](mailto:karink@uow.edu.au)

# Profile Transformations for Reciprocal Averaging and Singular Value Decomposition

Ting-Wu Wang<sup>1\*</sup>, Eric J. Beh<sup>2,3</sup>, Rosaria Lombardo<sup>4</sup> and Ian  
W. Renner<sup>1</sup>

<sup>1\*</sup>School of Information & Physical Sciences, University of  
Newcastle, University Dr, Callaghan, 2308, NSW, Australia.

<sup>2</sup>National Institute for Applied Statistics Research Australia  
(NIASRA), University of Wollongong, Northfields Ave,  
Wollongong, 2522, NSW, Australia.

<sup>3</sup>Centre for Multi-Dimensional Data Visualisation (MuViSU),  
Stellenbosch University, Matieland, 7602, South Africa.

<sup>4</sup>Department of Economics, University of Campania, via Gran  
Priorato di Malta, Capua, 81043, Campania, Italy.

\*Corresponding author(s). E-mail(s): [ting-wu.wang@uon.edu.au](mailto:ting-wu.wang@uon.edu.au);

## Abstract

Power transformations of count data, including for cell frequencies of a contingency table, have been well understood for nearly 100 years with much of the attention focused on the square root transformation. Over the past 15 years or so this topic has been the focus of some new insights into areas of correspondence analysis where two forms of power transformation have been discussed. One type considers the impact of raising the joint proportions of the cell frequencies of a table to a known power while the second type examines the power transformation of the relative distribution of the cell frequencies. While the foundations of the graphical features of correspondence analysis rest with the numerical algorithms of reciprocal averaging, and other analogous techniques, discussions of the role of power transformations in reciprocal averaging have not been described. Therefore, this paper examines this link where a power transformation is applied to the cell frequencies of a two-way contingency table. In doing so, we show that reciprocal averaging can be performed under such a transformation to obtain row and

column scores that provide the maximum association between the variables and the greatest discrimination between the categories. Finally, we discuss the connection between performing reciprocal averaging and singular value decomposition under this type of power transformation. The R function, `powerRA.exe()` is included in the Appendix to this paper and calculates the row and column scores, and their maximum correlation, under a power transformation of the cell frequencies.

**Keywords:** Reciprocal Averaging, Canonical Correlation Analysis, Power Transformations

## 1 Introduction

Transforming a random variable is commonly adopted to overcome various problems with the data, including to stabilise the variance when it is linked to the expectation. This is particularly the case for a Poisson random variable,  $Y$ , with expectation and variance  $\lambda$ , and dates back to [Bartlett \(1936\)](#) who considered the square root transformation,  $\sqrt{Y}$ . [Bartlett \(1947\)](#) showed improvements could be made to stabilise the variance by using the transformation  $\sqrt{Y + 1/2}$  while [Anscombe \(1948\)](#) showed further improvements could be gained using  $\sqrt{Y + 3/8}$ . [Anscombe \(1953\)](#) then showed that  $Y^{2/3}$  yielded an asymptotically zero skewness coefficient and was “highly successful in normalizing the distribution even if  $[\lambda]$  is as low as 4 . . . but at the cost of a non-constant variance” (p. 229); [Anscombe \(1953\)](#) stated that the variance of  $Y^{2/3}$  is  $(4/9)\lambda^{1/3}$ . A decade later, the Box-Cox power transformation was developed as a strategy to improve the normality of random variables ([Box and Cox, 1964](#)). Another transformation that is frequently used to accommodate for the non-constant variance of a Poisson variable was proposed by [Freeman and Tukey \(1950\)](#) and is of the form  $\sqrt{Y} + \sqrt{Y + 1}$ . More recently, [Yu \(2009\)](#) showed that  $(\sqrt{Y} + \sqrt{Y + 1})/2$  has improved variance stabilisation properties for small  $\lambda$ .

While the square root transformation may be the most common power transformation considered throughout the statistical and allied literature, it is

not the only transformation that one may consider. A more general transformation, and one that is the focus of this paper, involves  $Y^\delta$  for some given and real value of  $\delta$ . [Bishop et al. \(2007, Example 14.6-3\)](#) examined the inferential aspects of this transformation and showed that  $Y^\delta$  has an expectation of  $\lambda^\delta$  and a variance of  $\delta^2 \lambda^{2\delta-1}$  so that, for all  $\delta$ , the residual  $Z_\delta = (Y^\delta - \lambda^\delta) / (\delta \lambda^{\delta-1/2})$  is an asymptotically standard normal random variable. One may note, therefore, that using this result the variance of  $Y^{2/3}$  is exactly what [Anscombe \(1953\)](#) stated. Note also that, for a sample of size  $n$ , the sum-of-squares of the  $Z_\delta$  values gives the Cressie-Read family of divergence statistics ([Cressie and Read, 1984](#); [Read and Cressie, 1988](#)) which is a chi-squared random variable with  $n - 1$  degrees of freedom, for  $\delta \in (-\infty, \infty)$ .

Our attention in this paper will focus on a two-way contingency table where its cell frequencies (like  $Y$  above) are raised to the power of  $\delta$ , although we shall be relaxing any need to impose a Poisson (or any other) distribution assumption on the frequencies. In doing so, our core focus is to develop an objective strategy for determining scores for each row and column category that best discriminates the categories while also maximising the correlation between the variables. To do this we shall be using reciprocal averaging (RA), a technique that is akin to the scoring procedures described by, for example, [Hirschfeld \(1935\)](#) and [Hill \(1974\)](#) and is related to the dual scaling method of [Nishisato \(1980, 2007\)](#) and to the singular value and eigen-decompositions that are commonly used in the context of correspondence analysis (CA); see, for example, [Greenacre \(1984, 2017\)](#), [Lebart et al. \(1984\)](#), [Beh \(2004\)](#) and [Beh and Lombardo \(2014, 2021\)](#). By adopting an RA strategy to determine the scores that are calculated for each of the categories of the contingency table, a clear interpretation of their meaning can be made in terms of the power transformation of the profiles.

To discuss the impact of applying a power transformation to the cell frequencies on the row and column scores obtained using RA, this paper is divided into seven further sections. Section 2 gives an overview of the classical approach

## 4 Profile Transformations for RA and SVD

to RA that makes use of the *profile* of each row and column category of the contingency table. Three types of transformation for a contingency table are discussed in Section 3. The first two we briefly discuss are the “power family 1” and “power family 2” transformations described Greenacre (2009). The third transformation involves raising the cell frequencies of the contingency table to a power  $\delta$ ; it is this transformation that we shall give the remainder of our attention to. This is certainly the case for Section 4 which outlines the RA algorithm for this third transformation that produces the one-dimensional set of row and column scores for a given value of  $\delta$  and is fitted with the `powerRA.exe()` function described in the Appendix. Section 4 also uses canonical correlation analysis to show that the correlation between these scores is maximised for a given power transformation. Section 5 expands upon the one-dimensional RA algorithm described in Section 4 by showing that a multi-dimensional solution can be obtained through matrix decomposition. We demonstrate the applicability of this method of RA in Section’s 6 and 7. Section 6 examines the changes in the row and column scores under the transformation of the cell counts by studying the ground-breaking data set of Selikoff (1981) that established the dangers of occupational exposure of asbestos fibre’s. Section 7 studies the monthly distribution of tropical depressions, storms and hurricanes appearing in the Hurricane Databases (HURDAT) of the US National Hurricane Center between 1851 – 2021. Some final remarks are left for Section 8.

## 2 An Overview of Hill’s Reciprocal Averaging

### 2.1 On Profiles

The process of RA has been described in various ways since Horst (1935, p. 370) first coined the phrase *method of reciprocal averaging* to describe the method outlined in Richardson and Kuder (1933). While these contributions provide a written description of the technique, Hirschfeld (1935) presented what is essentially the same idea but as a detailed mathematical description of the method for a two-way contingency table. Such a description would form

the basis of the method outlined by Hill (1974) and has been discussed in the context of CA in the texts of Greenacre (1984, Section 4.2), Lebart et al. (1984, Section 5.2.1), Gifi (1990, pp. 106 – 118), Nishisato (2007, Section 2.2), Beh and Lombardo (2014, Section 3.5), Greenacre (2017, pp. 111 – 112) and Nishisato et al. (2021, Chapters 7 & 8). We now provide an overview of RA which is based on the centred profile of a category as a measure of its departure from the hypothesis of independence between the two categorical variables. This will help to give some context to the developments outlined in the following sections which focus on the RA of the power transformation of these profiles.

Suppose we have an  $I \times J$  contingency table  $\mathbf{N}$  with sample size  $n$ . Denote the  $(i, j)$ th cell frequency of  $\mathbf{N}$  by  $n_{ij}$  so that the  $(i, j)$ th element of the corresponding matrix of cell proportions,  $\mathbf{P}$ , is  $p_{ij} = n_{ij}/n$ , for  $i = 1, 2, \dots, I$  and  $j = 1, 2, \dots, J$ . Denote  $p_{i\bullet} = \sum_{j=1}^J p_{ij}$  and  $p_{\bullet j} = \sum_{i=1}^I p_{ij}$  to be the contingency table's  $i$ th row and  $j$ th column marginal proportion, respectively.

When examining the association between the variables of an  $I \times J$  contingency table, RA calculates the row scores,  $\mathbf{a} = (a_1, a_2, \dots, a_I)^T$ , and the column scores,  $\mathbf{b} = (b_1, b_2, \dots, b_J)^T$ , that maximise the association between their variables and ensures the greatest discrimination between the categories. Here we shall focus on a one-dimensional solution to the RA problem and then establish how eigen-decomposition (ED) and singular value decomposition (SVD) can be used to determine a multi-dimensional set of orthogonal solutions.

The foundations of RA rest with (using the parlance of CA) the profile of each category. The  $i$ th centred row profile is defined as

$$\left( \frac{p_{i1}}{p_{i\bullet}} - p_{\bullet 1}, \frac{p_{i2}}{p_{i\bullet}} - p_{\bullet 2}, \dots, \frac{p_{iJ}}{p_{i\bullet}} - p_{\bullet J} \right) \quad (1)$$

while the  $j$ th centred column profile is

$$\left( \frac{p_{1j}}{p_{\bullet j}} - p_{1\bullet}, \frac{p_{2j}}{p_{\bullet j}} - p_{2\bullet}, \dots, \frac{p_{Ij}}{p_{\bullet j}} - p_{I\bullet} \right). \quad (2)$$

6 *Profile Transformations for RA and SVD*

When there is complete independence between the row and column variables of  $\mathbf{N}$ , the elements of all profiles will be zero.

The row and column scores can be determined by solving the following two equations

$$\lambda \mathbf{a} = (\mathbf{R}^{-1} \mathbf{P} - \mathbf{1}_I \mathbf{c}^T) \mathbf{b} \quad (3)$$

$$\lambda \mathbf{b} = (\mathbf{C}^{-1} \mathbf{P}^T - \mathbf{1}_J \mathbf{r}^T) \mathbf{a}. \quad (4)$$

Here,  $\mathbf{1}$  denotes a vector of 1's of length specified by its subscript, and  $\mathbf{r} = \text{vec}(p_{i\bullet})^T$ ,  $\mathbf{R} = \text{diag}(p_{i\bullet})$  while  $\mathbf{c} = \text{vec}(p_{\bullet j})^T$  and  $\mathbf{C} = \text{diag}(p_{\bullet j})$ . After a number of iterations cycling through (3) and (4), the scores will converge to a desired level to give an optimal solution which maximises the correlation  $\lambda = \mathbf{a}^T \mathbf{P} \mathbf{b}$  between  $\mathbf{a}$  and  $\mathbf{b}$ . The solutions to  $\mathbf{a}$  and  $\mathbf{b}$  are centred at zero with a variance of one so that they have the property  $\mathbf{a}^T \mathbf{R} \mathbf{a} = 1$  and  $\mathbf{b}^T \mathbf{C} \mathbf{b} = 1$ , respectively.

Beh and Lombardo (2014, Section 3.5.2), for example, showed that when pre-multiplying both sides of (3) by  $\mathbf{R}^{1/2}$  and (4) by  $\mathbf{C}^{1/2}$ , the solution to the triplet  $(\mathbf{a}, \mathbf{b}, \lambda)$  found via RA can also be determined from the one-dimensional solution to the SVD of the matrix of Pearson's residuals

$$\mathbf{Z} = \mathbf{R}^{-1/2} (\mathbf{P} - \mathbf{r} \mathbf{c}^T) \mathbf{C}^{-1/2}.$$

While RA provides a one-dimensional solution of  $\mathbf{a}$  and  $\mathbf{b}$ , the SVD of  $\mathbf{Z}$  provides the analyst with a multi-dimensional solution. Such a solution is obtained from

$$\mathbf{Z} = \mathbf{A} \mathbf{\Lambda} \mathbf{B}^T$$

where  $\mathbf{A}$  is the  $I \times M$  column matrix of left singular vectors while  $\mathbf{B}$  is the  $J \times M$  column matrix of right singular vectors; the first column of  $\mathbf{A}$  and  $\mathbf{B}$  are weighted versions of  $\mathbf{a}$  and  $\mathbf{b}$ , respectively, which are obtained from the RA of the contingency table. These matrices are constrained such that  $\mathbf{A}^T \mathbf{A} = \mathbf{I}_M$

and  $\mathbf{B}^T \mathbf{B} = \mathbf{I}_M$ . The singular values of  $\mathbf{Z}$  are the elements of the  $M \times M$  diagonal matrix,  $\mathbf{\Lambda}$ , where the (1, 1)th element is equivalent to  $\lambda$ .

A feature obtained from the SVD of  $\mathbf{Z}$  is that Pearson's chi-squared statistic can be expressed in terms of its elements, and the elements of  $\mathbf{\Lambda}$ , such that

$$X^2 = n \cdot \text{trace}(\mathbf{Z}\mathbf{Z}^T) = n \cdot \text{trace}(\mathbf{Z}^T \mathbf{Z}) = n \cdot \text{trace}(\mathbf{\Lambda}^2) .$$

### 3 Three Power Transformations

#### 3.1 Two Known Power Transformations

There are times when one may wish to apply a power transformation to the elements of a contingency table, or to some alternative form of the data. [Greenacre \(2009\)](#), [Cuadras and Cuadras \(2006\)](#) and [Beh and Lombardo \(2023\)](#) describe when it may be appropriate to do so from a CA perspective and [Greenacre \(2009\)](#) considered two types of power transformation. The first type we describe is his “power family 2” transformation and it also lies at the heart of the methods discussed by [Cuadras and Cuadras \(2006\)](#) and [Beh and Lombardo \(2023\)](#). Therefore, the profiles defined by (1) and (2) can be amended to incorporate the power transformation of its elements so that, for the  $i$ th centred row profile,

$$\left( \left( \frac{p_{i1}}{p_{i\bullet}} \right)^\delta - p_{\bullet 1}^\delta, \left( \frac{p_{i2}}{p_{i\bullet}} \right)^\delta - p_{\bullet 2}^\delta, \dots, \left( \frac{p_{iJ}}{p_{i\bullet}} \right)^\delta - p_{\bullet J}^\delta \right) \quad (5)$$

and, for the  $j$ th centred column profile,

$$\left( \left( \frac{p_{1j}}{p_{\bullet j}} \right)^\delta - p_{1\bullet}^\delta, \left( \frac{p_{2j}}{p_{\bullet j}} \right)^\delta - p_{2\bullet}^\delta, \dots, \left( \frac{p_{Ij}}{p_{\bullet j}} \right)^\delta - p_{I\bullet}^\delta \right) \quad (6)$$

for some given value of  $\delta$ . [Greenacre \(2009\)](#) showed that  $\delta \rightarrow 0$  leads to his log-ratio analysis (LRA). Comparisons could then be made by observing differences in the results between classical CA (when  $\delta = 1$ ) and LRA. At its



core, the discussion of Cuadras and Cuadras (2006) also involves (5) and (6) which they referred to as their parametric version of CA. Part of their discussion was to show what happens when  $\delta = 1/2$  leading to their Hellinger Distance Decomposition (HDD) method and comparisons could be then made of this method with CA. Therefore, such an approach to CA involves taking the square-root of the profile elements and the interested reader is invited to peruse the pages of Escofier (1978), Domenges and Volle (1979), Rao (1995a,b, 1997), Nakayama et al. (1998) and Beh et al. (2018) for discussions involving the square-root of the profile elements and the Hellinger distance in CA.

There are many benefits of the LRA and HDD methods of Greenacre (2009) and Cuadras and Cuadras (2006), and further comparisons were undertaken by Cuadras et al. (2006). Beh and Lombardo (2023) showed that both are special cases of a more general framework where the Cressie-Read family of divergence statistics (Cressie and Read, 1984; Read and Cressie, 1988) is used as the measure of association for the variables. They showed that LRA and HDD have, at their numerical heart, the modified log-likelihood ratio statistic and the Freeman-Tukey statistic, respectively. Beh et al. (2023) show the link between RA and the power transformed profiles given by (5) and (6).

### 3.2 A New Power Transformation

While (5) and (6) show that each profile element is raised to a power  $\delta$ , at its core, they both involve  $p_{ij}^\delta$  not  $n_{ij}^\delta$ ; this is the second type of transformation (briefly) considered by Greenacre (2009) and is his “power family 1” transformation. One may also note that (5) is expressed in terms of  $p_{i\bullet}^\delta$ . However,  $p_{i\bullet}^\delta$  is not the summation of  $p_{ij}^\delta$  across the all of the columns. Nor does  $p_{\bullet j}^\delta$ , a term present in (6), involve the summation (across the rows) of the  $p_{ij}^\delta$  values. Rather,  $p_{i\bullet}^\delta$  is the power transformation of the  $i$ th marginal proportion of the contingency table while  $p_{\bullet j}^\delta$  is the power transformation of the  $j$ th marginal proportion. In both cases, a unitary transformation is applied to the cell proportions ( $p_{ij}^1$ , for  $i = 1, 2, \dots, I$  and  $j = 1, 2, \dots, J$ ). This suggest

a third, new, transformation may be considered. While the null hypothesis of independent is preserved when  $\delta = 1$  so that

$$H_0 : p_{ij} = p_{i\bullet} p_{\bullet j}$$

it is not true that

$$H_{0|\delta} : p_{ij}^\delta = (p_{i\bullet} p_{\bullet j})^\delta$$

for any given value of  $\delta \neq 1$ . Therefore, the third transformation we now discuss may be viewed as being more closely related to the “power family 1” than to “power family 2” but is nonetheless different to both. To accommodate for the discrepancy in the null hypothesis,  $H_{0|\delta}$ , we define an alternative null hypothesis involving the power transformation of the cell frequencies so that

$$H_{0|\delta} : p_{ij}(\delta) = p_{i\bullet}(\delta) p_{\bullet j}(\delta) \quad (7)$$

where  $p_{ij}(\delta)$  is the joint proportion after each element of  $n_{ij}$  is raised to the power of  $\delta$ . That is,  $p_{ij}(\delta) = n_{ij}^\delta / n_\delta$  where  $n_\delta = \sum_{i=1}^I \sum_{j=1}^J n_{ij}^\delta$ . In (7),  $p_{i\bullet}(\delta)$  and  $p_{\bullet j}(\delta)$  are defined as

$$p_{i\bullet}(\delta) = \sum_{j=1}^J p_{ij}(\delta) = \frac{1}{n_\delta} \sum_{j=1}^J n_{ij}^\delta$$

and

$$p_{\bullet j}(\delta) = \sum_{i=1}^I p_{ij}(\delta) = \frac{1}{n_\delta} \sum_{i=1}^I n_{ij}^\delta,$$

respectively.

With this alternative transformation defined, we now turn our attention to deriving the RA algorithm.

## 4 Power Transformation and Reciprocal Averaging

### 4.1 The Algorithm

Consider the power transformation of the  $(i, j)$ th cell frequency,  $n_{ij}^\delta$ , for all  $i = 1, 2, \dots, I$  and  $j = 1, 2, \dots, J$ . Then, under this particular transformation and its associated null hypothesis of (7), the  $i$ th centred row profile is

$$\left( \frac{p_{i1}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet 1}(\delta), \frac{p_{i2}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet 2}(\delta), \dots, \frac{p_{iJ}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet J}(\delta) \right) \quad (8)$$

while

$$\left( \frac{p_{1j}(\delta)}{p_{\bullet j}(\delta)} - p_{1\bullet}(\delta), \frac{p_{2j}(\delta)}{p_{\bullet j}(\delta)} - p_{2\bullet}(\delta), \dots, \frac{p_{Ij}(\delta)}{p_{\bullet j}(\delta)} - p_{I\bullet}(\delta) \right) \quad (9)$$

is the  $j$ th centred column profile.

Denote  $a_i(\delta)$  to be the  $i$ th row score for some value of  $\delta \in (-\infty, \infty)$  and  $b_j(\delta)$  to be the  $j$ th column score for this value of  $\delta$ . The RA of the power transformed centred profiles involves solving  $a_i(\delta)$  and  $b_j(\delta)$  such that

$$\begin{aligned} \lambda(\delta) a_i(\delta) &= \left( \frac{p_{i1}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet 1}(\delta) \right) b_1(\delta) + \dots + \left( \frac{p_{iJ}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet J}(\delta) \right) b_J(\delta) \\ &= \sum_{j=1}^J \left( \frac{p_{ij}(\delta)}{p_{i\bullet}(\delta)} - p_{\bullet j}(\delta) \right) b_j(\delta) \end{aligned} \quad (10)$$

and

$$\begin{aligned} \lambda(\delta) b_j(\delta) &= \left( \frac{p_{1j}(\delta)}{p_{\bullet j}(\delta)} - p_{1\bullet}(\delta) \right) a_1(\delta) + \dots + \left( \frac{p_{Ij}(\delta)}{p_{\bullet j}(\delta)} - p_{I\bullet}(\delta) \right) a_I(\delta) \\ &= \sum_{i=1}^I \left( \frac{p_{ij}(\delta)}{p_{\bullet j}(\delta)} - p_{i\bullet}(\delta) \right) a_i(\delta). \end{aligned} \quad (11)$$

Here

$$\sum_{i=1}^I p_{i\bullet}(\delta) a_i(\delta) = 0, \quad \sum_{i=1}^I p_{i\bullet}(\delta) a_i^2(\delta) = 1 \quad (12)$$

and

$$\sum_{j=1}^J p_{\bullet j}(\delta) b_j(\delta) = 0 \quad \sum_{j=1}^J p_{\bullet j}(\delta) b_j^2(\delta) = 1 \quad (13)$$

while

$$\lambda(\delta) = \sum_{j=1}^J \sum_{i=1}^I p_{ij}(\delta) a_i(\delta) b_j(\delta) \quad (14)$$

is the correlation between the set of row scores  $\mathbf{a}_\delta = (a_1(\delta), \dots, a_I(\delta))^T$  and column scores,  $\mathbf{b}_\delta = (b_1(\delta), \dots, b_J(\delta))^T$ .

The RA algorithm for solving the triplet  $(\mathbf{a}_\delta, \mathbf{b}_\delta, \lambda(\delta))$  can be expressed in matrix notation. To do so, we define  $\mathbf{P}_\delta$  to be the  $I \times J$  matrix of  $p_{ij}(\delta) = n_{ij}^\delta/n_\delta$  elements. We also define  $\mathbf{r}_\delta = \text{vec}(p_{i\bullet}(\delta))^T$  and  $\mathbf{R}_\delta = \text{diag}(p_{i\bullet}(\delta))$  while  $\mathbf{c}_\delta = \text{vec}(p_{\bullet j}(\delta))^T$  and  $\mathbf{C}_\delta = \text{diag}(p_{\bullet j}(\delta))$ . Then (10) and (11) are elements of

$$\lambda(\delta) \mathbf{a}_\delta = (\mathbf{R}_\delta^{-1} \mathbf{P}_\delta - \mathbf{1}_I \mathbf{c}_\delta^T) \mathbf{b}_\delta \quad (15)$$

and

$$\lambda(\delta) \mathbf{b}_\delta = (\mathbf{C}_\delta^{-1} \mathbf{P}_\delta^T - \mathbf{1}_J \mathbf{r}_\delta^T) \mathbf{a}_\delta, \quad (16)$$

respectively, and the matrix form of the right-hand side of (12) and (13) is

$$\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta = 1 \quad \text{and} \quad \mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta = 1, \quad (17)$$

respectively.

The iterative steps involved in solving for  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  here are akin to the steps used in solving the classical RA problem described by, for example, Hill (1974) and Beh and Lombardo (2014, Section 3.5). Also, the above method and classical RA are equivalent when  $\delta = 1$ . This can be verified since  $n_1 = \sum_{i=1}^I \sum_{j=1}^J n_{ij}^1 = n$ , so that  $p_{ij}(1) = n_{ij}/n = p_{ij}$  and the  $i$ th row and  $j$ th

column marginal proportions are  $p_{i\bullet}(1) = \sum_{j=1}^J n_{ij}/n = p_{i\bullet}$  and  $p_{\bullet j}(1) = \sum_{i=1}^I n_{ij}/n = p_{\bullet j}$ , respectively. Therefore, the matrix  $\mathbf{P}_1$  contains the  $p_{ij}$  elements, for  $i = 1, 2, \dots, I$  and  $j = 1, 2, \dots, J$  and so is equivalent to  $\mathbf{P}$  while  $\mathbf{R}_1 = \mathbf{R}$  and  $\mathbf{C}_1 = \mathbf{C}$ . The R function `powerRA.exe()` is included in the Appendix to this paper and calculates the row and column scores, and their maximum correlation, using this method of RA.

We do note that other emendations have been proposed to the traditional approach to RA discussed in Section 2. For example, Nishisato (1984) proposed the *method of reciprocal medians* to help “mitigate the problem of extreme weights” (p. 143) while Nishisato et al. (2021, Chapter 8) discussed the *reciprocal geometric averaging* and *reciprocal harmonic averaging* algorithms for  $n_{ij}$  and so there is scope for generalising these RA methods for  $n_{ij}^\delta$ .

## 4.2 Maximising the Correlation of the Scores

We can show using canonical correlation analysis that  $\lambda^2(\delta)$ , defined by squaring (14), yields the maximum (positive) correlation between  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$ . To do so we note that this correlation can be expressed as

$$\lambda(\delta) = \frac{\text{Cov}(\mathbf{a}_\delta, \mathbf{b}_\delta)}{\sqrt{\text{Var}(\mathbf{a}_\delta)}\sqrt{\text{Var}(\mathbf{b}_\delta)}} = \frac{\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta}{\sqrt{\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta} \sqrt{\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta}}.$$

Note that at this stage we are not yet requiring that  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  are subject to any particular property. Therefore, the squared correlation can be expressed as

$$\lambda^2(\delta) = (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-1} (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta)^2 (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1}. \quad (18)$$

We can now maximise  $\lambda^2(\delta)$  by differentiating (18) with respect to the two sets of scores. When differentiated with respect to  $\mathbf{a}_\delta$ , we get

$$\frac{\partial}{\partial \mathbf{a}_\delta} \lambda^2(\delta) = \left[ \frac{\partial}{\partial \mathbf{a}_\delta} (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-1} (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta)^2 \right] (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1}$$

$$\begin{aligned}
&= \left\{ - (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-2} (2\mathbf{R}_\delta \mathbf{a}_\delta) (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta)^2 + \right. \\
&\quad \left. 2 (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-1} (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) (\mathbf{P}_\delta \mathbf{b}_\delta) \right\} (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1} \\
&= 2 (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-1} (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) (\mathbf{P}_\delta \mathbf{b}_\delta) (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1} - \\
&\quad 2 (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-2} (\mathbf{R}_\delta \mathbf{a}_\delta) (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta)^2 (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1}
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{a}_\delta} \lambda^2(\delta) &= 2 (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta)^{-2} \left( (\mathbf{a}_\delta^T \mathbf{R}_\delta \mathbf{a}_\delta) (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) \mathbf{P}_\delta \mathbf{b}_\delta - \right. \\
&\quad \left. (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta)^2 \mathbf{R}_\delta \mathbf{a}_\delta \right) (\mathbf{b}_\delta^T \mathbf{C}_\delta \mathbf{b}_\delta)^{-1}.
\end{aligned}$$

It is at this point that we substitute into this derivative the property of  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  given by (17) yielding

$$\frac{\partial}{\partial \mathbf{a}_\delta} \lambda^2(\delta) = 2 (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) (\mathbf{P}_\delta \mathbf{b}_\delta - (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) \mathbf{R}_\delta \mathbf{a}_\delta).$$

Setting this derivative to zero gives

$$\mathbf{P}_\delta \mathbf{b}_\delta = (\mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta) \mathbf{R}_\delta \mathbf{a}_\delta \tag{19}$$

while substituting (17) into (18) gives

$$\lambda(\delta) = \mathbf{a}_\delta^T \mathbf{P}_\delta \mathbf{b}_\delta \tag{20}$$

which is the matrix form of (14). We can confirm that this is the maximum correlation between the row and column scores for the given value of  $\delta$  since

$$\frac{\partial^2}{\partial \mathbf{a}_\delta^2} \lambda^2(\delta) = -2\lambda(\delta) \mathbf{R}_\delta < 0$$

for all  $\delta \in (-\infty, \infty)$ . To show that the row scores are those obtained via RA, (19) reduces to

$$\mathbf{P}_\delta \mathbf{b}_\delta = \lambda(\delta) \mathbf{R}_\delta \mathbf{a}_\delta \quad (21)$$

so that

$$\begin{aligned} \lambda(\delta) \mathbf{a}_\delta &= \mathbf{R}_\delta^{-1} \mathbf{P}_\delta \mathbf{b}_\delta \\ &= (\mathbf{R}_\delta^{-1} \mathbf{P}_\delta - \mathbf{1}_I \mathbf{c}_\delta^T) \mathbf{b}_\delta \end{aligned}$$

which is just (15). Following a similar derivation, but one that maximises the squared correlation (18) with respect to  $\mathbf{b}_\delta$ , leads to (16). Therefore, obtaining the row and column scores,  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  using the RA algorithm described above produces the maximum possible correlation for a given  $\delta$  – defined by (14) – of these scores.

## 5 Solving the $M$ -Dimensional Problem

### 5.1 One-dimensional Solution via Matrix Decomposition

Rather than solving for  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  iteratively we can instead use matrix decomposition. Here we shall describe the link between the RA procedure described in Section 4 and the role of ED and SVD for determining the solution to the triplet  $(\mathbf{a}_\delta, \mathbf{b}_\delta, \lambda(\delta))$ .

The matrix decomposition approach to obtaining an RA solution to the triplet is to solve for  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  by first pre-multiplying both sides of (15) by  $\lambda(\delta) \mathbf{R}_\delta^{1/2}$  such that

$$\begin{aligned} \lambda^2(\delta) \left( \mathbf{R}_\delta^{1/2} \mathbf{a}_\delta \right) &= \lambda(\delta) \mathbf{R}_\delta^{1/2} (\mathbf{R}_\delta^{-1} \mathbf{P}_\delta - \mathbf{1}_I \mathbf{c}_\delta^T) \mathbf{b}_\delta \\ &= \left[ \mathbf{R}_\delta^{-1/2} (\mathbf{P}_\delta - \mathbf{r}_\delta \mathbf{c}_\delta^T) \mathbf{C}_\delta^{-1/2} \right] \left( \lambda(\delta) \mathbf{C}_\delta^{1/2} \mathbf{b}_\delta \right) \quad (22) \end{aligned}$$

while pre-multiplying (16) by  $\mathbf{C}_\delta^{1/2}$  gives

$$\lambda(\delta) \mathbf{C}_\delta^{1/2} \mathbf{b}_\delta = \left[ \mathbf{C}_\delta^{-1/2} (\mathbf{P}_\delta - \mathbf{r}_\delta \mathbf{c}_\delta^T)^T \mathbf{R}_\delta^{-1/2} \right] \left( \mathbf{R}_\delta^{1/2} \mathbf{a}_\delta \right). \quad (23)$$

Substituting (23) into (22) and simplifying gives

$$\left( \mathbf{Z}_\delta \mathbf{Z}_\delta^T - \lambda^2(\delta) \mathbf{I}_I \right) \left( \mathbf{R}_\delta^{1/2} \mathbf{a}_\delta \right) = \mathbf{0}_I \quad (24)$$

where

$$\mathbf{Z}_\delta = \mathbf{R}_\delta^{-1/2} (\mathbf{P}_\delta - \mathbf{r}_\delta \mathbf{c}_\delta^T) \mathbf{C}_\delta^{-1/2} \quad (25)$$

which is the matrix of standardised residuals for the elements of  $\mathbf{P}_\delta$ . Equation (24) shows that the solution to  $\mathbf{a}_\delta$  can be obtained from pre-multiplying the first eigen-vector of  $\mathbf{Z}_\delta \mathbf{Z}_\delta^T$  by  $\mathbf{R}_\delta^{-1/2}$ , while the correlation  $\lambda(\delta)$  is equivalent to the square root of its largest eigen-value. Therefore,  $\lambda(\delta)$  is the absolute value of this correlation. The solution to  $\mathbf{b}_\delta$  can be found by solving

$$\left( \mathbf{Z}_\delta^T \mathbf{Z}_\delta - \lambda^2(\delta) \mathbf{I}_J \right) \left( \mathbf{C}_\delta^{1/2} \mathbf{b}_\delta \right) = \mathbf{0}_J$$

so that it can be obtained by pre-multiplying the largest eigen-vector of  $\mathbf{Z}_\delta^T \mathbf{Z}_\delta$  by  $\mathbf{C}_\delta^{-1/2}$ .

Rather than obtaining the solution to  $\mathbf{a}_\delta$  and  $\mathbf{b}_\delta$  through two ED's, they can be determined by applying a SVD to (25). By using such an approach multiple orthogonal solutions to these scores can also be obtained and we shall discuss this further in Section 5.2.

## 5.2 $M$ -dimensional Reciprocal Averaging via Eigen-Decomposition

By considering the power transformation of the cell frequencies that is outlined in Section 3, solving the two RA formulae of (15) and (16) produces an  $M$ -dimensional solution that can be obtained by applying a SVD



to  $\mathbf{Z}_\delta$ , where  $M = \min(I, J) - 1$ . To show this, we start by examining how they can be obtained from an ED involving  $\mathbf{Z}_\delta$ . To do so, we define  $\mathbf{A}_\delta = (\mathbf{a}_{\delta(1)}, \dots, \mathbf{a}_{\delta(M)})$  to be the  $I \times M$  column matrix of row scores where  $\mathbf{a}_{\delta(m)} = (a_{1(m)}(\delta), a_{2(m)}(\delta), \dots, a_{I(m)}(\delta))^T$  and  $\mathbf{A}_\delta^T \mathbf{R}_\delta \mathbf{A}_\delta = \mathbf{I}_I$ . We also define  $\mathbf{B}_\delta = (\mathbf{b}_{\delta(1)}, \dots, \mathbf{b}_{\delta(M)})$  to be the  $J \times M$  column matrix of column scores where  $\mathbf{b}_{\delta(m)} = (b_{1(m)}(\delta), b_{2(m)}(\delta), \dots, b_{J(m)}(\delta))^T$  and  $\mathbf{B}_\delta^T \mathbf{C}_\delta \mathbf{B}_\delta = \mathbf{I}_J$ . Denote the  $M \times M$  diagonal correlation matrix between  $\mathbf{A}_\delta$  and  $\mathbf{B}_\delta$  by  $\mathbf{\Lambda}_\delta = \mathbf{A}_\delta^T \mathbf{P}_\delta \mathbf{B}_\delta$  where  $\lambda_m(\delta)$  is the  $(m, m)$ th element. Then the RA formulae for obtaining a solution to each of the  $M$  row and column scores is to iteratively solve

$$\lambda_m(\delta) \mathbf{a}_{\delta(m)} = (\mathbf{R}_\delta^{-1} \mathbf{P}_\delta - \mathbf{1}_I \mathbf{c}_\delta^T) \mathbf{b}_{\delta(m)} \quad (26)$$

and

$$\lambda_m(\delta) \mathbf{b}_{\delta(m)} = (\mathbf{C}_\delta^{-1} \mathbf{P}_\delta^T - \mathbf{1}_J \mathbf{r}_\delta^T) \mathbf{a}_{\delta(m)}. \quad (27)$$

By following a derivation that is similar to what is described in Section 5.1 we obtain the following two ED equations

$$(\mathbf{Z}_\delta \mathbf{Z}_\delta^T - \lambda_m^2(\delta) \mathbf{I}_I) (\mathbf{R}_\delta^{1/2} \mathbf{a}_{\delta(m)}) = \mathbf{0}_I \quad (28)$$

$$(\mathbf{Z}_\delta^T \mathbf{Z}_\delta - \lambda_m^2(\delta) \mathbf{I}_J) (\mathbf{C}_\delta^{1/2} \mathbf{b}_{\delta(m)}) = \mathbf{0}_J \quad (29)$$

where  $\mathbf{Z}_\delta$  is defined by (25). Therefore,  $\lambda_m^2(\delta)$  is the  $m$ th largest eigen-value of  $\mathbf{Z}_\delta \mathbf{Z}_\delta^T$  and  $\mathbf{Z}_\delta^T \mathbf{Z}_\delta$  so that  $\lambda_m(\delta) = \mathbf{a}_{\delta(m)}^T \mathbf{P}_\delta \mathbf{b}_{\delta(m)}$ . Note that this expression of  $\lambda_m(\delta)$  is similar in form to Beh et al. (2023, eq. (15)) however when  $\delta \neq 1$  the two produce different correlation values due to the different ways the power transformation is applied. When  $\delta = 1$ ,  $\lambda_1(1)$  and the correlation of Beh et al. (2023, eq. (15)) are both equivalent to (20). We can also see that the  $m$ th eigen-vector of  $\mathbf{Z}_\delta \mathbf{Z}_\delta^T$  is  $\tilde{\mathbf{a}}_{\delta(m)} = \mathbf{R}_\delta^{1/2} \mathbf{a}_{\delta(m)}$  while  $\tilde{\mathbf{b}}_{\delta(m)} = \mathbf{C}_\delta^{1/2} \mathbf{b}_{\delta(m)}$  is the

$m$ th eigen-vector of  $\mathbf{Z}_\delta^T \mathbf{Z}_\delta$ . These eigen-vectors have the property

$$\tilde{\mathbf{a}}_{\delta(m)}^T \tilde{\mathbf{a}}_{\delta(m')} = \begin{cases} 1, & m = m' \\ 0, & m \neq m' \end{cases}$$

and

$$\tilde{\mathbf{b}}_{\delta(m)}^T \tilde{\mathbf{b}}_{\delta(m')} = \begin{cases} 1, & m = m' \\ 0, & m \neq m' \end{cases}.$$

Thus the vector of row and column scores,  $\mathbf{a}_{\delta(m)}$  and  $\mathbf{b}_{\delta(m)}$ , can be obtained from the components of the SVD of  $\mathbf{Z}_\delta$  such that

$$\mathbf{a}_{\delta(m)} = \mathbf{R}^{-1/2} \tilde{\mathbf{a}}_{\delta(m)} \quad (30)$$

$$\mathbf{b}_{\delta(m)} = \mathbf{C}^{-1/2} \tilde{\mathbf{b}}_{\delta(m)}. \quad (31)$$

### 5.3 $M$ -dimensional Reciprocal Averaging via SVD

Since the solution to  $\mathbf{a}_{\delta(m)}$  and  $\mathbf{b}_{\delta(m)}$  can be obtained from the ED of  $\mathbf{Z}_\delta \mathbf{Z}_\delta^T$  and  $\mathbf{Z}_\delta^T \mathbf{Z}_\delta$ , respectively, they can also be determined by applying a SVD to  $\mathbf{Z}_\delta$ . That is, one may consider

$$\mathbf{Z}_\delta = \tilde{\mathbf{A}}_\delta \mathbf{\Lambda}_\delta \tilde{\mathbf{B}}_\delta^T$$

where the  $I \times M$  matrix  $\tilde{\mathbf{A}}_\delta$  contains  $\tilde{\mathbf{a}}_{\delta(m)}$  as its  $m$ th column and is the matrix of left singular vectors of  $\mathbf{Z}_\delta$ . Similarly, the  $J \times M$  matrix  $\tilde{\mathbf{B}}_\delta$  contains  $\tilde{\mathbf{b}}_{\delta(m)}$  as its  $m$ th column and is the matrix of right singular vectors of  $\mathbf{Z}_\delta$ . The  $M \times M$  matrix  $\mathbf{\Lambda}_\delta$  is diagonal and contains the singular values,  $\lambda_m(\delta)$  for  $m = 1, 2, \dots, M$ , of  $\mathbf{Z}_\delta$ . These matrices have the property that  $\tilde{\mathbf{A}}_\delta^T \tilde{\mathbf{A}}_\delta = \mathbf{I}_I$  and  $\tilde{\mathbf{B}}_\delta^T \tilde{\mathbf{B}}_\delta = \mathbf{I}_J$ . Applying a SVD in this manner is the typical way of determining the multi-dimensional set of orthogonal row scores and column scores for the classical approach to RA (when  $\delta = 1$ ) and is a standard tool used in the CA

literature; see, for example, Nishisato (1980), Lebart et al. (1984), Greenacre (1984, 2017), Beh (2004) and Beh and Lombardo (2012, 2014, 2019, 2021) for a variety of comprehensive discussions of the role of SVD in CA. It is also easy to perform the necessary calculations in the R programming environment because of the versatility of `svd()` that is included as a base function.

Assessing the overall departure from independence between the row and column categories when a power transformation is applied to the cell frequencies can be made by denoting the measure of association such that

$$\begin{aligned}\phi^2(\delta) &= \text{trace}(\mathbf{Z}_\delta^T \mathbf{Z}_\delta) \\ &= \text{trace} \left( \left( \tilde{\mathbf{A}}_\delta \boldsymbol{\Lambda}_\delta \tilde{\mathbf{B}}_\delta^T \right)^T \tilde{\mathbf{A}}_\delta \boldsymbol{\Lambda}_\delta \tilde{\mathbf{B}}_\delta^T \right) \\ &= \text{trace} \left( \tilde{\mathbf{B}}_\delta \boldsymbol{\Lambda}_\delta \tilde{\mathbf{A}}_\delta^T \tilde{\mathbf{A}}_\delta \boldsymbol{\Lambda}_\delta \tilde{\mathbf{B}}_\delta^T \right) \\ &= \text{trace} \left( \tilde{\mathbf{B}}_\delta \boldsymbol{\Lambda}_\delta^2 \tilde{\mathbf{B}}_\delta^T \right).\end{aligned}$$

When  $\tilde{\mathbf{B}}_\delta$  is full rank  $\tilde{\mathbf{B}}_\delta \tilde{\mathbf{B}}_\delta^T = \mathbf{I}_J$  so that

$$\phi^2(\delta) = \text{trace}(\boldsymbol{\Lambda}_\delta^2) = \sum_{m=1}^M \lambda_m^2(\delta).$$

Therefore, the measure of association can be quantified from the sum-of squares of the singular values of  $\mathbf{Z}_\delta$ . For example,  $\phi^2(1) = X^2/n$  is Pearson's phi-squared statistic of a two-way contingency table and  $X^2$  is Pearson's statistic.

## 6 Application: Selikoff's Asbestos Data

### 6.1 Preliminary Analysis

Consider the data in Table 1 that originates from the studies by the American chest physician Irving Selikoff. After noticing unusual cases amongst the workers from a local asbestos factory in the 1950's, he and his team undertook a comprehensive examination of the potential causes of the illness amongst New

**Table 1:** *Cross tabulation of Selikoff's studies on the link between occupational exposure to asbestos and the diagnosed grade of asbestosis.*

Occupational exposure (years)	Asbestosis grade diagnosed				Total
	None	Grade 1	Grade 2	Grade 3	
0 - 9	310	36	0	0	346
10 -19	212	158	9	0	379
20-29	21	35	17	4	77
30-39	25	102	49	18	194
40+	7	35	51	28	121
Total	575	366	126	50	1117

York construction workers. From his study of 1117 workers, he established that there was a link between the number of years of occupational exposure to asbestos fibres and the severity of asbestosis that the worker was diagnosed with. Due to the nature of his findings, Selikoff concluded that workers who had been exposed to asbestos fibres for 20 years were highly likely to be diagnosed with severe cases of asbestosis and referred to this as his “20-year rule”. However, the results were not published until 1981 (Selikoff, 1981). The impact of this finding had been felt internationally and was described in detail, in the context of categorical data analysis, by Beh and Smith (2011), Tran et al. (2012) and Beh and Lombardo (2014, Section 1.4).

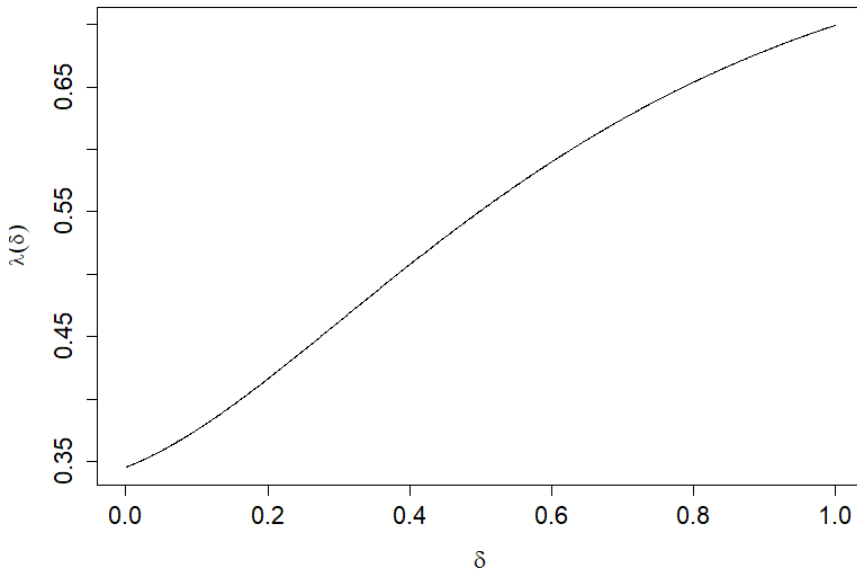
We shall therefore apply the RA and SVD scoring techniques described in Sections 4 and 5 to Table 1 where the sample size is  $n = 1117$ . It cross-classifies 5 different lengths of time that a worker was (occupationally) exposed to asbestos (in decadal intervals) and four grades (in severity) of asbestosis that the workers were diagnosed with. A chi-squared test of independence of Table 1 gives a Pearson statistic of 648.81 with  $(5 - 1)(4 - 1) = 12$  degrees of freedom. Therefore, with a p-value that is less than 0.001, there is clear evidence of the existence of a statistically significant association between the years of exposure to asbestos and the diagnosed level of asbestosis.

## 6.2 The One-Dimensional Solution

Since there exists a statistically significant association between the variables of Table 1 we can examine the nature of this association by determining the scores for the row and column categories using the RA approach described in Section 4. In doing so, the one-dimensional solution to the set of row scores,  $\mathbf{a}_\delta = (a_1(\delta), \dots, a_5(\delta))^T$ , the column scores,  $\mathbf{b}_\delta = (b_1(\delta), \dots, b_4(\delta))^T$  and their correlation,  $\lambda(\delta)$ , are summarised in Table 2 for three values of  $\delta$ ;  $\delta = 0.001$ ,  $\delta = 0.5$  and  $\delta = 1$ . The choice of  $\delta \approx 0$  is made to reflect a transformation close to, but not equal, to zero. Note that a zero power transformation will result in all transformed cell frequencies being 1 so that departures of these cells from  $1/(IJ)$  are assessed. We also consider  $\delta = 1$  to demonstrate that a unitary transformation yields the same solution as the classical method of RA and  $\delta = 0.5$  is chosen to compare the results with those obtained when a near zero and a unitary power transformation are considered. These values have been calculated using the RA algorithm described in Section 4 and by applying a SVD to the  $\mathbf{Z}_\delta$  matrix of Table 1 and rescaling the matrices of left and right singular vectors using (30) and (31). For the scores obtained using

**Table 2:** *Comparison of the row and column scores obtained from the RA and the SVD of  $\mathbf{Z}_\delta$  for Table 1; for  $\delta = 0.001, 0.5$  and  $1$ .*

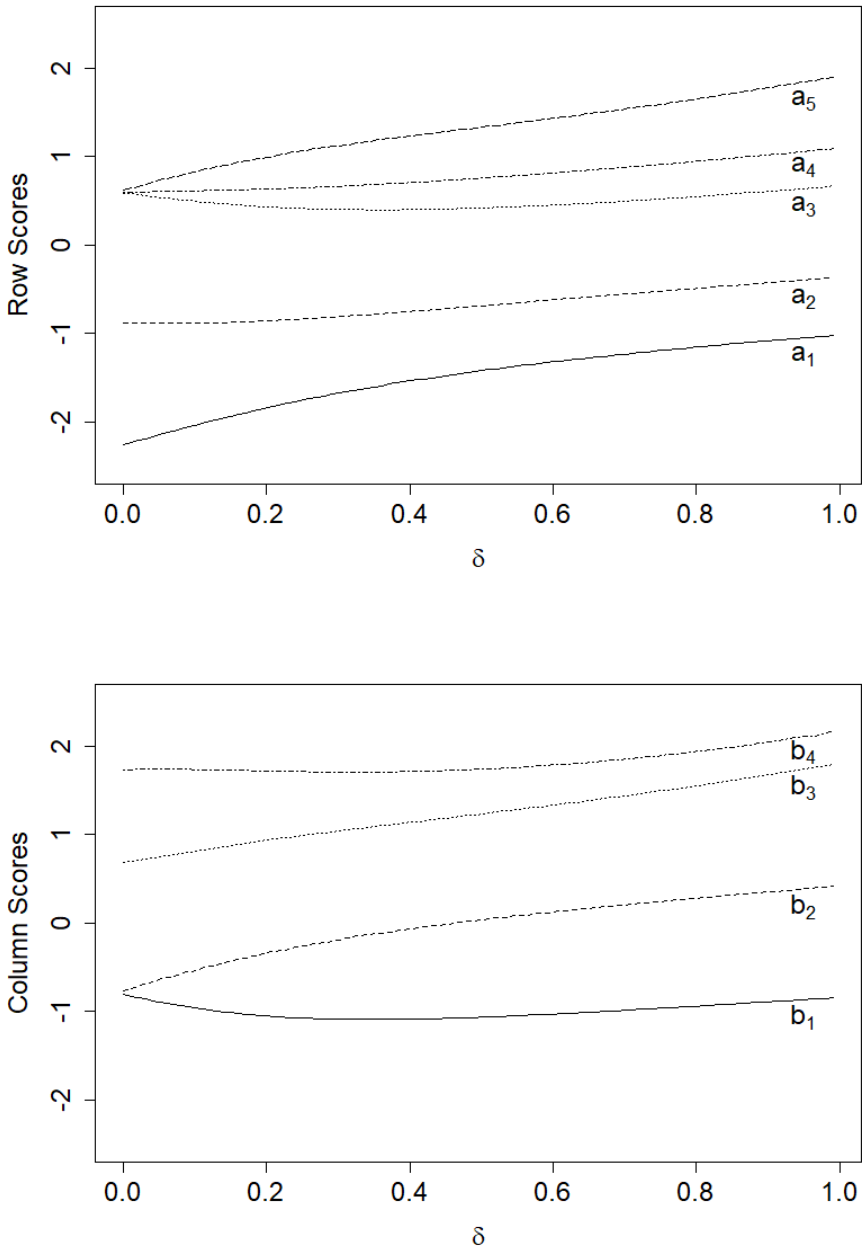
	$\delta = 1$		$\delta = 0.5$		$\delta \approx 0$	
	RA	SVD	RA	SVD	RA	SVD
Row scores						
$a_1(\delta)$	-1.0229006	-1.0228997	-1.4321469	-1.4321471	-2.2773040	-2.2773154
$a_2(\delta)$	-0.3684099	-0.3684108	-0.6924097	-0.6924096	-0.8763991	-0.8763842
$a_3(\delta)$	0.6684296	0.6684290	0.4183102	0.4183102	0.5979584	0.5979566
$a_4(\delta)$	1.0929230	1.0929224	0.7543871	0.7543872	0.5990797	0.5990779
$a_5(\delta)$	1.9012795	1.9012811	1.3263083	1.3263081	0.6012666	0.6012647
Column scores						
$b_1(\delta)$	-0.8467648	-0.8467646	-1.06740098	-1.06740100	-0.7890030	-0.7890042
$b_2(\delta)$	0.4155945	0.4155940	0.02886595	0.02886599	-0.7846205	-0.7846216
$b_3(\delta)$	1.7993315	1.7993319	1.22954011	1.22954011	0.6681646	0.6681715
$b_4(\delta)$	2.1613281	2.1613289	1.74030643	1.74030637	1.7349748	1.7349696
$\lambda(\delta)$	0.6994048	0.6994048	0.5507601	0.5507601	0.3455012	0.3455012



**Fig. 1:**  $\lambda(\delta)$  versus  $\delta \in (0, 1]$

RA, they were calculated using the R function `powerRA.exe()`. The scores from the SVD of  $\mathbf{Z}_\delta$  were obtained using the `svd()` function.

We can visually assess the impact of the transformation described in Section 4 by illustrating the change in the row scores, the column scores and their correlation for changes in  $\delta$ . Fig. 1 shows the impact on the correlation  $\lambda(\delta)$  for  $\delta \in (0, 1]$ . It shows that, for Table 1,  $\lambda(\delta)$  linearly increases from its minimum of 0.346 at  $\delta \approx 0$  to its maximum of 0.699 at  $\delta = 1$ . One can also observe changes in the row and column scores for changes in  $\delta$ . The first plot of Fig. 2 shows such changes and highlights that, for the row categories, there is virtually no difference in the profiles of the third, fourth and fifth row categories of Table 1 when  $\delta = 0.001$ . However, for this value of  $\delta$ , there is quite a big difference between the transformed profiles of these three categories and the first and second categories which are themselves different. As  $\delta$  approaches 1, so that one is getting closer to performing the classical approach to RA,



**Fig. 2:** The impact on the configuration of similar or different row (top) and column (bottom) profile distributions as  $\delta$  changes; for  $\delta \in (0, 1]$

the difference in the transformed profiles for the first two categories remains fairly constant, although the profiles of the third, fourth and fifth row profiles become more different under the transformation. A similar shift in the four column scores can also be observed from the second plot of Fig. 2. It shows that there is very little difference in the transformed profile of the first and second column categories for  $\delta = 0.001$ . However, as  $\delta \rightarrow 1$ , the scores for these column categories become increasingly different while the profile of the third and fourth column categories become increasingly similar. While not shown in the top figure of Fig. 2, as  $\delta$  approaches 2 and beyond, the fifth row category diverges away from zero. The same effect can be observed for the third and fourth column categories. These results highlight that applying a power transformation to the cells of Table 1 yields a very different set of scores as  $\delta$  changes.

While this section has highlighted the equivalency of the one-dimensional solution to  $\mathbf{a}_\delta$ ,  $\mathbf{b}_\delta$  and  $\lambda(\delta)$  when applying the method of RA described above and a SVD to  $\mathbf{Z}_\delta$ , we can also obtain a multi-dimensional solution to these quantities. We now turn our attention to examine this issue.

### 6.3 The Multi-Dimensional Solution

The solution to  $\mathbf{a}_\delta$ ,  $\mathbf{b}_\delta$  and their correlation  $\lambda(\delta)$ , using RA under any power transformation of the cells of a contingency table can be extended to the  $M$ -dimensional case by applying a full rank SVD to  $\mathbf{Z}_\delta$ . We now do so for Table 1 with  $\delta = 0.001, 0.5$  and  $1$  and obtain  $M = \min(5, 4) - 1 = 3$  orthogonal solutions. Therefore, if we were to visualise the association we would need at most three-dimensions to capture all of the association structure that exists between the variables of the contingency table.

Table 3 gives the solution to

$$\begin{aligned}\mathbf{a}_{\delta(m)} &= (a_{1(m)}(\delta), a_{2(m)}(\delta), \dots, a_{5(m)}(\delta))^T \\ \mathbf{b}_{\delta(m)} &= (b_{1(m)}(\delta), b_{2(m)}(\delta), \dots, b_{4(m)}(\delta))^T\end{aligned}$$



and  $\lambda_{(m)}(\delta)$  for  $m = 1, 2, 3$  and our three values of  $\delta$ . This table shows that applying the power transformation to the profile elements, as shown by (8) and (9), has an impact on the value of the row and column scores, and their correlation, as  $\delta$  changes.

**Table 3:** *The  $M$ -dimensional solution determined by SVD under three power transformations.*

	$m = 1$			$m = 2$			$m = 3$		
	$\delta = 1$	$\delta = 0.5$	$\delta \approx 0$	$\delta = 1$	$\delta = 0.5$	$\delta \approx 0$	$\delta = 1$	$\delta = 0.5$	$\delta \approx 0$
Row scores									
$a_{1(m)}(\delta)$	-1.0229	-1.4321	-2.2773	0.9485	1.3158	1.5180	-0.3899	0.1906	0.0020
$a_{2(m)}(\delta)$	-0.3684	-0.6924	-0.8764	-0.9168	-1.2145	-1.9735	0.8663	0.0257	0.0012
$a_{3(m)}(\delta)$	0.6684	0.4183	0.5980	-0.5817	-0.3287	0.2402	-2.7298	-2.1516	-1.6601
$a_{4(m)}(\delta)$	1.0929	0.7544	0.5991	-0.7385	-0.4548	0.2407	-0.5844	1.2277	0.5798
$a_{5(m)}(\delta)$	1.9013	1.3263	0.6013	1.7136	1.0943	0.2406	1.0754	-0.0882	1.0773
Column scores									
$b_{1(m)}(\delta)$	-0.8468	-1.0674	-0.789	0.4717	0.6792	0.2842	-0.0556	-0.2244	-1.3022
$b_{2(m)}(\delta)$	0.4156	0.0289	-0.7846	-1.3397	-1.2037	0.2822	0.2906	0.6754	1.305
$b_{3(m)}(\delta)$	1.7993	1.2295	0.6682	0.8788	-0.1133	-1.6748	-1.9635	-1.735	-0.0024
$b_{4(m)}(\delta)$	2.1613	1.7403	1.735	2.1673	1.8212	1.2893	3.46	1.6907	-0.0019
$\lambda_{(m)}(\delta)$	0.6994	0.5508	0.3455	0.2986	0.1525	0.1866	0.0501	0.0225	0.0001

## 7 Application: HURDAT

### 7.1 Preliminary Analysis

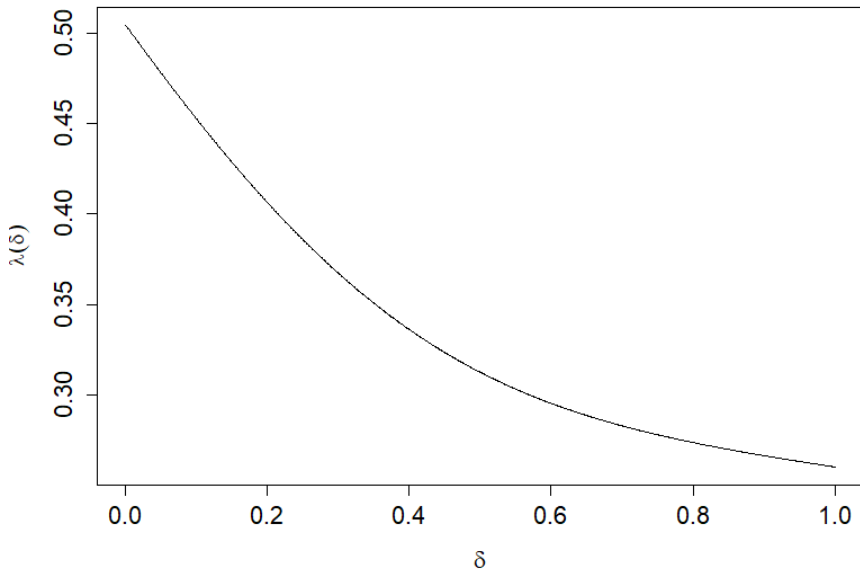
Table 4 shows the distribution of tropical depressions, tropical storms and all five categories of hurricanes occurring in the North Atlantic basin between 1851 and 2021. This data comes from the Hurricane Databases (HURDAT) of the National Hurricane Center (USA) and is available online at <https://www.nhc.noaa.gov/data/>. The hurricane categories are determined using the Saffir-Simpson Hurricane Wind Scale and are classified as Category 1 through to Category 5 where each category has a maximum (one minute) sustained wind speed ten metres above the surface of between 119–153km/h (Category 1), 154–177km/h (Category 2), 178–208km/h (Category 3), 209–251km/h (Category 4) and 252km/h or higher (Category 5), respectively. Tropical depressions (TD)

**Table 4:** *Cross-tabulation of the hurricane strength based on the Saffir-Simpson Hurricane Wind Scale and the month they occurred, for 1851 – 2021.*

Month	Saffir-Simpson Hurricane Wind Scale							Total
	TD	TS	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	
January	0	2	2	0	0	0	0	4
February	0	1	0	0	0	0	0	1
March	0	0	0	1	0	0	0	1
April	3	3	0	0	0	0	0	6
May	5	30	5	0	0	0	0	40
June	21	72	24	12	3	0	0	132
July	44	65	39	16	7	3	2	176
August	41	148	80	68	60	46	13	456
September	70	220	129	92	63	57	17	648
October	30	147	81	56	25	18	6	363
November	10	48	23	8	4	3	0	96
December	1	9	3	0	0	0	0	13
Total	225	745	386	253	162	127	38	1936

are defined as cyclones with a maximum sustained wind speed of 61km/h or less near the center of the storm while the speed of Tropical Storms (TS) falls between 62km/h and 88km/h; for a more complete description of how the data is collected see <https://www.nhc.noaa.gov/aboutsshws.php>. Data that is similar in type given by Table 4 were also reported in, for example, Cry (1965) for hurricanes recorded between 1871 and 1963.

A chi-squared test of independence of Table 4 gives a Pearson chi-squared statistic of 192.88 and with  $(12 - 1)(7 - 1) = 66$  degrees of freedom results in a p-value that is less than 0.001. This indicates that a statistically significant association exists between the months of the year and the category of storms that were formed. This should be of no surprise since the Atlantic hurricane season is between June 1 and November 30. However, while such a period was initiated in 1965 to capture the most likely period in which hurricanes would hit landfall, more recent research suggests that, with the impact of a changing global climate, questions are being asked of its reliability; see, for example, Kossin (2008) and Truchelut et al. (2022). With the existence of such an association in Table 4, we shall investigate the impact of applying a power

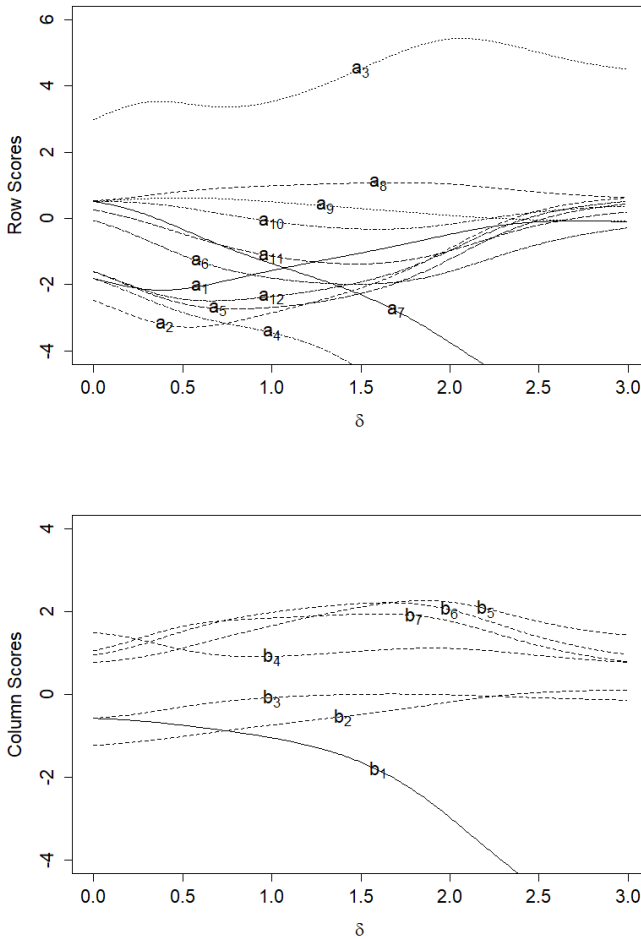


**Fig. 3:**  $\lambda(\delta)$  versus  $\delta \in (0, 1]$

transformation to its cell counts on the row and column scores calculated using RA.

## 7.2 Power Transformation of RA for HURDAT

The RA algorithm has been applied to Table 4 over a range of delta values to observe the change and the impact of power transformation on the association. The resulting effect is visualised in Fig. 3 showing the change in the correlation  $\lambda(\delta)$  for  $\delta \in (0, 1]$ . A decrease in correlation from  $\lambda(\delta) = 0.5043953$  to 0.2599614 is observed as  $\delta$  increases from approximately 0 to 1. Rather than observing an increased association as  $\delta$  increases like in Fig. 1, the association between the month and the category of hurricanes that occurs decreases as  $\delta$  approaches 1. This indicates that more association can be captured as a benefit of decreasing the value of power transformation compared to the traditional method of RA when  $\delta = 1$ .



**Fig. 4:** The impact of power transformation on the configuration of similar or different row (top) and column (bottom) profile distributions as  $\delta$  changes.

The row scores  $a_1, \dots, a_{12}$  representing each of the 12 months of Table 4 are calculated using RA and appear in the upper plot of Fig. 4 for  $\delta \in (0, 3]$ . It shows that, for  $\delta$  values in this interval, a power transformation applied to the cell counts has a fairly similar impact on the scores for the months August ( $a_8$ ), September ( $a_9$ ) and October ( $a_{10}$ ). We can also see a similar feature is observed for the row categories January ( $a_1$ ), February ( $a_2$ ), May ( $a_5$ ) and December ( $a_{12}$ ) indicating that the power transformation applied to

the cell counts of these months leads to fairly similar scores at each value of  $\delta$ ; note that it is these months that have relatively few hurricane strength storms compared to the northern hemisphere summer months. For both sets of row categories, as  $\delta$  increases, their scores approach zero. This suggests that there is effectively very little difference in the relative impact on the scores by applying a power transformation to the cell counts of June ( $a_6$ ) and November ( $a_{11}$ ). However, as  $\delta$  increases, the RA score for April ( $a_4$ ) and July ( $a_7$ ) diverge away from zero. Interestingly, both months have very different profiles with May observing only six tropical level wind speeds while July observes 176 recorded storms with 62% of them (or 109 of the 176 storms) being only a tropical depression or tropical storm. This highlights the increasing importance of April and July on the association structure between the variables. This may be because both months have very different profiles when compared with the other row categories and this is accentuated as a power transformation (of  $\delta > 1$ ) is applied to the cells counts of Table 4. Interestingly, the score for March ( $a_3$ ) is consistently different to all other scores calculated and the top plot of Fig. 4 shows it to have a consistently greater score than any other row score for  $\delta \in (0, 1]$ . This feature arises because of the single Category 2 storm recorded which lies outside of the regular hurricane season for the Atlantic basin (June 1 – November 30).

The column scores,  $b_1, \dots, b_7$  are depicted in the lower plot of Fig. 4 and represent the categories ordered by the strength of the hurricanes from low (Tropical Depression) to high (Category 5), respectively. We find a similar general pattern in the effect of power transformation on the scores calculated for Category 3 ( $b_5$ ), Category 4 ( $b_6$ ) and Category 5 ( $b_7$ ) wind speeds. At the lower power transformations (for  $\delta$  less than about 1), the pattern observed in the scores of Category 2 ( $b_4$ ) experiences a very different effect compared to the general pattern for  $b_5, b_6$  and  $b_7$ . All column scores tend to approximately zero as  $\delta$  increases except for the score associated with a TD ( $b_1$ ). The column score

diverges negatively instead of positively as observed in the scores obtained from the analysis of Selikoff's asbestos data; see Section 6.

## 8 Discussion

Strategies that numerically and visually describe the association between categorical variables often involve performing analytical techniques such as RA and CA (Nishisato, 2007; Lebart et al., 1984; Greenacre, 1984, 2017; Beh and Lombardo, 2014; Nishisato et al., 2021). Investigating objective scoring methods for the categories of a contingency table while adopting a power transformation provides a deeper understanding of how the profiles of the categories compare and, more broadly, of the association that exists between the variables. The method of RA described in this paper considers a third type of power transformation that involves raising the cell frequencies to a power  $\delta$  and complements the two discussed by Greenacre (2009). By expanding the solutions to the multi-dimensional case, such scores can be obtained by considering a matrix of modified Pearson's residuals – see (25) – and applying SVD to it. The equivalence of the scores between RA and the SVD of  $\mathbf{Z}_\delta$  can be obtained by reweighting the left and right singular vectors by diagonal matrices involving  $\mathbf{r}_\delta$  and  $\mathbf{c}_\delta$  and the application to Selikoff's classic asbestos data verifies the equivalence of the two approaches.

There are various avenues of further research that can be undertaken that use and extend the method of RA described in this paper. One such avenue involves determining the most suitable value of  $\delta$ . The choice of  $\delta$  may be made subject to a wide range of criteria such as

- ensuring the correlation is maximised across a range of  $\delta$  values. Like Greenacre (2009), the first example considered values of  $\delta$  ranging between 0 and 1 (although, strictly speaking, we consider  $\delta = 0.001$  as the value of the lower limit). However, any value of  $\delta > 0$  may be considered and we extended this interval to  $(0, 3]$  for the analysis of the HURDAT data. Both analyses show that the correlation between the row and column scores is

maximised for different values of  $\delta$ . For example, the maximum correlation for Table 1 was achieved at  $\delta = 1$ , while the maximum correlation of Table 4 approached 1 as  $\delta \rightarrow 0$ . Therefore, there is no clear general choice that may be recommended since the value of  $\delta$  depends on the structure of the data, on the association between the two variables and other factors.

- visually maximising the association between the variables in a low-dimensional plot if the scoring method outline in this paper is used as the basis for the visualisation of association (such as when performing CA, dual scaling and other analogous techniques). Such a choice was a point of discussion by Cuadras and Cuadras (2006, p. 72). It was also raised by Beh et al. (2023) and Beh and Lombardo (2023) although, for their scaling approach – which is equivalent to the “power family 2” of Greenacre (2009) – they also discussed other criteria on which  $\delta$  may be chosen,
- ensuring that an association does exist between the variables. Preliminary investigations reveal that there are values of  $\delta$  where there is no statistically significant association between the variables (when Pearson’s chi-squared statistic, say, is used as a goodness-of-fit measure). In the context of better understanding the structure of the association between categorical variables, and visually exploring the nature of the association (using, for example, CA) such a choice of  $\delta$  must be avoided. This property is directly related to the family of Cressie-Read divergence statistics and the “power family 2” transformation (Beh and Lombardo, 2023) but also has potential implications for the power transformation considered in this paper. Such an issue was discussed by Wang and Beh (2022).

Another possible way in which the RA method outlined above can be extended concerns the presence of over-dispersion that exists in the data; a property that has been discussed in the context of categorical data analysis by Haberman (1973) and Agresti (2013, p. 80), while Beh and Lombardo (2020, 2023) examined the issue from a CA perspective. Determining the appropriate

power transformation that minimises the presence of over-dispersion when performing RA is another possible avenue for future consideration that will help to gain a deeper understanding, and resolve potential issues, when analysing the association between categorical variables.

## 9 Statements and Declarations

### 9.1 Compliance with Ethical Standards

Ethical approval is not required for this research. Information and data involved in this study are available in the public domain.

### 9.2 Conflict of Interests

The authors declare that they have no conflict interests.

### 9.3 Funding

No funding was used in the preparation of this manuscript.

### 9.4 Data Availability Statements

The authors confirm that all data generated or analysed during this study are included in this article.

## References

- Agresti, A. (2013). *Categorical Data Analysis (3rd ed)*. Wiley.
- Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254.
- Anscombe, F. J. (1953). Reply to Hotelling’s “New light on the correlation coefficient and its transforms”. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15(2):229 – 230.
- Bartlett, M. S. (1936). The square root transformation in analysis of variance. *Supplement to the Journal of the Royal Statistical Society*, 3(1):68–78.



- Bartlett, M. S. (1947). The use of transformations. *Biometrics*, 3(1):39–52.
- Beh, E. J. (2004). Simple correspondence analysis: a bibliographic review. *International Statistical Review*, 72(2):257–284.
- Beh, E. J. and Lombardo, R. (2012). A genealogy of correspondence analysis. *Australian & New Zealand Journal of Statistics*, 54(2):137–168.
- Beh, E. J. and Lombardo, R. (2014). *Correspondence Analysis: Theory, Practice and New Strategies*. Wiley.
- Beh, E. J. and Lombardo, R. (2019). A genealogy of correspondence analysis: Part 2 – The variants. *Electronic Journal of Applied Statistical Analysis*, 12(2):552–603.
- Beh, E. J. and Lombardo, R. (2020). Five strategies for accommodating overdispersion in correspondence analysis. In Imaizumi, T., Okada, A., Miyamoto, S., Sakaori, F., Yamamoto, Y., and Vichi, M., editors, *Advanced Studies in Classification and Data Science*, pages 117 – 129. Springer, Singapore.
- Beh, E. J. and Lombardo, R. (2021). *An Introduction to Correspondence Analysis*. Wiley.
- Beh, E. J. and Lombardo, R. (2023). Correspondence analysis using the cressie-read family of divergence statistics. *International Statistical Review*, page (in press).
- Beh, E. J., Lombardo, R., and Alberti, G. (2018). Correspondence analysis and the Freeman–Tukey statistic: A study of archaeological data. *Computational Statistics & Data Analysis*, 128:73–86.
- Beh, E. J., Lombardo, R., and Wang, T.-W. (2023). Power transformations and reciprocal averaging. In Beh, E. J., Lombardo, R., and Clavel, J. G., editors, *Analysis of Categorical Data from Historical Perspectives: Essays in Honor of Shizuhiko Nishisato*, page (to appear). Springer, Singapore.
- Beh, E. J. and Smith, D. R. (2011). Real world occupational epidemiology, Part 1: Odds ratios, relative risk, and asbestosis. *Archives of Environmental & Occupational Health*, 66:119 – 123.

- Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer Science & Business Media.
- Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.
- Cressie, N. A. C. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B (Methodology)*, 46:440 – 464.
- Cry, G. (1965). Tropical cyclones of the north atlantic ocean: Tracks and frequencies of hurricanes and tropical storms, 1871 – 1963. *US Department of Commerce, Technical Paper No. 55:148* pages.
- Cuadras, C., Cuadras, D., and Greenacre, M. (2006). A comparison of methods for analyzing contingency tables. *Communications in Statistics–Simulation and Computation*, 35:447–459.
- Cuadras, C. M. and Cuadras, D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, 417:64–74.
- Domenges, D. and Volle, M. (1979). Analyse factorielle sphérique: Une exploration. *Annales de l'Inséé*, 35:3 – 84.
- Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. *Revue de Statistique Appliquée*, 26(4):29 – 37.
- Freeman, M. F. and Tukey, J. W. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, 21(4):607–611.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley.
- Greenacre, M. (2009). Power transformations in correspondence analysis. *Computational Statistics & Data Analysis*, 53:3107–3116.
- Greenacre, M. (2017). *Correspondence Analysis in Practice (3rd ed)*. Chapman & Hall/CRC.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press.

- Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics*, 29:205–220.
- Hill, M. O. (1974). Correspondence analysis: a neglected multivariate method. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 23:340–354.
- Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31:520–524.
- Horst, P. (1935). Measuring complex attitudes. *The Journal of Social Psychology*, 6:369–374.
- Kossin, J. P. (2008). Is the North Atlantic hurricane season getting longer? *Geophysical Research Letters*, 35, L23705:3 pages.
- Lebart, L., Morineau, A., and Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley.
- Nakayama, T., Naito, K., and Fujikoshi, Y. (1998). Stability of correspondence analysis and its alternative using Hellinger distance for contingency table. *International Journal of Mathematical and Statistical Science*, 7(1):97 – 119.
- Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press.
- Nishisato, S. (1984). Dual scaling by reciprocal medians. In *Atti della XXXII Riunione Scientifica della Società Italiana di Statistica*, pages 141 – 147. Società Italiana di Statistica, Roma.
- Nishisato, S. (2007). *Multidimensional Nonlinear Descriptive Analysis*. Chapman & Hall/CRC.
- Nishisato, S., Beh, E. J., Lombardo, R., and Clavel, J. G. (2021). *Modern Quantification Theory: Joint Graphical Display, Biplots, and Alternatives*. Springer Nature.
- Rao, C. (1995a). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Questiúó*, 19:23 – 63.

- Rao, C. (1995b). The use of Hellinger distance in graphical displays of contingency table data. In Tiit, E.-M., Kollo, T., and Niemi, H., editors, *New Trends in Probability and Statistics, Volume 3: Multivariate Statistics and Matrices in Statistics*, pages 143 – 161. VSP BV, The Netherlands.
- Rao, C. (1997). An alternative to correspondence analysis using Hellinger distance. In Fang, K. and Hickernell, F., editors, *Contemporary Multivariate Analysis and its Applications*, pages 1 – 19. Hong Kong.
- Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag.
- Richardson, M. W. and Kuder, G. F. (1933). Making a rating scale that measures. *Personnel Journal*, 12:36 – 40.
- Selikoff, I. J. (1981). Household risks with inorganic fibers. *Bulletin of the New York Academy of Medicine*, 57:947 – 961.
- Tran, D., Beh, E. J., and Smith, D. R. (2012). Real world occupational epidemiology, Part 3: An aggregate data analysis of Selikoff’s “20-year rule”. *Archives of Environmental & Occupational Health*, 67:243 – 248.
- Truchelut, R. E., Klotzbach, P. J., Staehling, E. M., Wood, K. M., Halperin, D. J., Schreck III, C., and Blake, E. S. (2022). Earlier onset of North Atlantic hurricane season with warming oceans. *Nature Communications*, 13, 4646:8 pages.
- Wang, T.-W. and Beh, E. J. (2022). Comparison of power transformation approaches to reciprocal averaging. In *24th International Conference on Computational Statistics (COMPSTAT2022)*. 23–26 August, 2022. Bologna, Italy.
- Yu, G. (2009). Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. *Statistics & Probability Letters*, 79(14):1621–1629.

## Appendix: The `powerRA.exe` function

This appendix contains the R function `powerRA.exe` which performs the RA algorithm described by (15) and (16) so that the resulting vector of row scores,  $\mathbf{a}_\delta$ , and column scores,  $\mathbf{b}_\delta$ , are subject to (17).

The function `powerRA.exe` performs these calculations using the following arguments:

- `data` - the two-way contingency table of size  $I \times J$ , where  $I > 2$  and  $J > 2$ ,
- `delta` - the power of the transformation that is applied to the cell frequencies of the contingency table. By default, `delta = 1` leading to the same solution as the traditional approach to RA,
- `rho.ini` - the initial value of the correlation between the row and column scores. By default this argument is set to `rho.ini = 1`,
- `acc` - the number of decimal places required for defining the convergence of the algorithm. By default the function performs the RA algorithm so that convergence is achieved to 6 decimal places.

```
powerRA.exe <- function(data, delta = 1, rho.ini = 1, acc = 6) {
#####
# Some basics                                     #
#####

I <- nrow(data)
J <- ncol(data)
n <- sum(data)
P <- data^delta/sum(data^delta)
pidot <- apply(P, 1, sum)
pdotj <- apply(P, 2, sum)
R <- diag(pidot, I, I)
C <- diag(pdotj, J, J)

#####
# The algorithm                                   #
#####

a.ini <- c(1:nrow(data)) # Initial value of the vector of
                        # row scores
b.ini <- c(1:ncol(data)) # Initial value of the vector of
                        # column scores
```

```

# The first iteration of the row scores, columns scores, and
# lambda value

a.old <- a.ini/sqrt(t(a.ini)**R**a.ini)[1,1]
b.old <- (1/rho.ini)*(solve(C)**t(P) - (rep(1, times = J)**%
  t(pidot)))**a.old
b.old <- b.old/sqrt(t(b.old)**C**b.old)[1,1]
lamb.old <- (t(a.old)**P**b.old)[1,1]

# The iterative step of the algorithm - it converges when the
# value of \delta differs from its updated value by less than
# acc decimal places

counter <- 0

repeat{
  a.new <- (1/lamb.old)*(solve(R)**P - (rep(1, times = I)
    **t(pidotj)))**b.old
  a.new <- a.new/sqrt(t(a.new)**R**a.new)[1,1]

  b.new <- (1/lamb.old)*(solve(C)**t(P) -
    (rep(1, times = J)**t(pidot)))**a.new
  b.new <- b.new/sqrt(t(b.new)**C**b.new)[1,1]

  lamb.new <- (t(a.new)**P**b.new)[1,1]

  counter <- counter + 1

  lamb.comp <- abs(lamb.old - lamb.new)
  if (lamb.comp < 10^(-1*acc)) break

  a.old <- a.new
  b.old <- b.new
  lamb.old <- lamb.new
}

#####
# The numerical output . . . #
#####

list(iterations = round(counter, acc),
      a = round(a.new, acc),
      b = round(b.new, acc),
      lamb = round(lamb.new, acc))
}

```

To demonstrate the applicability of the function `powerRA.exe`, define Table 1 as the R object `asbestos.dat` such that

38 *Profile Transformations for RA and SVD*

```

> asbestos.dat <- matrix(c(310, 212, 21, 25, 7, 36, 158, 35,
+   102, 35, 0, 9, 17, 49, 51, 0, 0, 4, 18, 28), nrow = 5)
> dimnames(asbestos.dat) <- list(paste(c("0-9", "10-19", "20-29",
+   "30-39", "40+")), paste(c("None", "Grade 1", "Grade 2",
+   "Grade 3")))
> asbestos.dat
      None Grade 1 Grade 2 Grade 3
0-9   310     36     0     0
10-19 212    158     9     0
20-29  21     35    17     4
30-39  25    102    49    18
40+    7     35    51    28
>

```

We can show that the function performs the classical RA on `asbestos.dat` by using the default values for its arguments as follows; however, we define `acc = 10` so that convergence is achieved after 10 decimal places:

```

> powerRA.exe(asbestos.dat, acc = 10)
$iterations
[1] 6

$a
[,1]
[1,] -1.0229006
[2,] -0.3684099
[3,]  0.6684296
[4,]  1.0929230
[5,]  1.9012795

$b
[,1]
[1,] -0.8467648
[2,]  0.4155945
[3,]  1.7993315
[4,]  2.1613281

$lamb
[1] 0.6994048

>

```

and are just those values in the second column Table 2. We can also replicate the values given in Table 2 for  $\delta = 0.5$  and  $\delta = 0.001$  using `powerRA.exe` such that

```

> powerRA.exe(asbestos.dat, delta = 0.5, acc = 10)
$iterations

```

```
[1] 5
```

```
$a
```

```
[,1]
```

```
[1,] -1.4321469
```

```
[2,] -0.6924097
```

```
[3,] 0.4183102
```

```
[4,] 0.7543871
```

```
[5,] 1.3263083
```

```
$b
```

```
[,1]
```

```
[1,] -1.06740098
```

```
[2,] 0.02886595
```

```
[3,] 1.22954011
```

```
[4,] 1.74030643
```

```
$lamb
```

```
[1] 0.5507601
```

```
>
```

```
> powerRA.exe(asbestos.dat, delta = 0.001, acc = 10)
```

```
$iterations
```

```
[1] 8
```

```
$a
```

```
[,1]
```

```
[1,] -2.2773040
```

```
[2,] -0.8763991
```

```
[3,] 0.5979584
```

```
[4,] 0.5990797
```

```
[5,] 0.6012666
```

```
$b
```

```
[,1]
```

```
[1,] -0.7890030
```

```
[2,] -0.7846205
```

```
[3,] 0.6681646
```

```
[4,] 1.7349748
```

```
$lamb
```

```
[1] 0.3455012
```

```
>
```