Test treatment effect differences in repeatedly measured symptoms with binary values: The matched correspondence analysis approach

Se-Kang Kim¹

Published online: 19 February 2020
© The Psychonomic Society, Inc. 2020

Abstract

When a continuous variable is measured twice, paired t test can be used to examine the statistical difference between two time points. However, when several related but dichotomously scored (0, 1) variables are measured twice, it would not be reasonable to use paired t test (or chi-squared test) to examine the related binary variable differences. Therefore, the present study introduces a novel statistical approach, called *matched correspondence analysis* (matched CA), which tests the related binary value differences between two time points. Matched CA was originally designed to study between-group comparisons (e.g., gender) in two contingency tables of the same size, with the same row and column quantities. However, unlike the original matched CA, the present study applies matched CA to the analysis of within-group matched matrices (e.g., at admission and at discharge) and examines the related binary value differences between two time points. To test the stability of parameter estimates, permutation and bootstrapping methods are used, and the pros and cons of within-group matched CA are discussed.

Keywords Matched correspondence analysis · Related psychiatric symptoms with binary values · BMI · EDI-2

When patients are repeatedly measured on their symptoms before and after treatment, if the measures are continuous, paired t test can be used to study the difference between the before and after treatment effects; in this example, two continuous variables (e.g., body mass index scores and Eating Disorder Inventory 2 scores in the present study) can easily be measured for differences with paired t tests. However, when the data consist of interrelated variables scored with binary indicators, such as 0 = *symptom absence* and 1 = *symptom presence*, conducting a separate paired t test for each repeatedly measured binary variable is not recommended, due to violation of the Gaussian assumption and inflated Type I errors. Hence, the present study introduces a novel approach, called *matched correspondence analysis* (CA), to test the differences in the repeatedly measured binary indicators. The application of matched CA to repeated binary measures is a novel approach introduced in the present study, which will describe binary clinical measures as an empirical example to illustrate matched CA.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-019-01328-9>) contains supplementary material, which is available to authorized users.

✉ Se-Kang Kim
sekim@fordham.edu

¹ Department of Psychology, Fordham University, The Bronx, NY, USA

Objectives

The example data set used in the present study includes two kinds of repeatedly measured data; one is continuous and the other is binary. To test treatment efficacy, we can easily conduct paired t tests for continuous measures; however, we cannot conduct paired t tests with (0, 1) binary data. McNemar's chi-squared test may be used for each dichotomous variable to test the difference between two time points, but if the dichotomous variables are interrelated, the McNemar's test results would be biased. Moreover, multiple comparisons can inflate Type I errors. Although Bonferroni correction could be used to control the Type I error rate, power would substantially decrease. Therefore, the present study introduces a new variant of correspondence analysis techniques, called the correspondence analysis (CA) of matched matrices (hereafter called *matched CA*). This technique analyzes related-multiple binary indicators, without the statistical complications just described. The present study is the first one to test the statistical differences in the related binary variables between two time points using matched CA, and the first that allows comparisons among variables (in terms of directions and magnitudes in their scale values) without inflated Type I errors.

This study was designed to (1) include a brief introduction of matched CA with a Greenacre's classical example

(Greenacre, 2017), (2) compare novel aspects of the present study with Greenacre's original matched CA, (3) show how to prepare an appropriate matrix for matched CA, (4) extract two types of dimensions (e.g., sum and difference) by matched CA, (5) determine the statistical significance of the dimensions with a permutation test, and (6) determine, on the basis of significant dimensions, whether the differences (being examined) are statistically significant with a bootstrap method. Finally, an empirical example illustrates these steps and interpret the results.

A brief introduction of Greenacre's matched CA

Matched CA was originally designed for the analysis of comparisons between two groups of the same size (or between subjects), with the same row and column entities, so as to optimally scale the similarities and the differences between the groups. Its classical example is in gender comparison (Greenacre, 2017). Greenacre illustrated the utility of matched CA with separate male and female matrices that each consisted of 24 countries as rows and four work-related categories (*full-time*, *part-time*, *stay at home*, and *don't know/unsure/missing*) as columns. The female and the male matrices were denoted by $\mathbf{A}_{24 \times 4}$ and $\mathbf{B}_{24 \times 4}$, respectively. To apply matched CA, $\mathbf{A}$ is stacked on top of $\mathbf{B}$ and the two matrices are concatenated in the following way (Greenacre, 2017):

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix}_{48 \times 8} \quad (1)$$

This concatenated matrix is called a *block-circulant* or "ABBA" matrix. The ABBA matrix is 48×8 , repeating the rows and columns twice, and the maximum dimensionality of this concatenated matrix is 7 because its dimensionality is determined by $\min(48 - 1, 8 - 1)$. Unlike ordinary CA, matched CA decomposes the total inertia into between- and within-group inertias, and the between-group inertias are given by the eigenvalues of the sum (e.g., female + male) dimensions and correspond to the between-group (e.g., country) effects. Whereas, the within-group inertias are given by the eigenvalues of the difference (e.g., female – male) dimensions and correspond to the within-group female–male differences. The sum dimensions characterize the similarities between females and males, whereas the difference dimensions describe the differences between females and males.

To identify either sum or difference dimensions, it is necessary to examine the patterns of coordinate signs in dimensions. Since the concatenated matrix, ABBA, is being analyzed here, there are two sets of identical coordinates up to possible sign changes. If there is no sign change in the dimensions, these dimensions are defined as the sum dimensions.

But, if the sign changes appear in the dimensions, they are defined as difference dimensions. Especially in the present study, the interest is in the difference dimensions, because the coordinates of the difference dimensions, in fact, represent the differences in the binary measurements.

Novel aspects in the study as compared to Greenacre's matched CA

In the present study, unlike Greenacre's between-group matched CA, matched CA is applied within groups that are repeatedly measured at two different time points. In the previous example, A and B were male and female matrices, but in the present study A is replaced with a Time 1 measurement matrix, and B with a Time 2 measurement matrix. A and B consist of the same group of people but are measured at both Time 1 and Time 2. As was explained previously, these measurements consist of (0, 1) binary values. As is shown in Eq. 1, the same concatenated format ABBA can be used to conduct matched CA of the repeatedly measured binary data.

When the Greenacre's between-matched CA paradigm (Greenacre, 2003, 2017) is applied to repeated measures, specifically matching individuals into their age groups, a researcher may lose between-individual matching (within the age groups). However, with the age-group matching, a researcher can gain the important information about differences in the measuring entities across different age groups.

Multiple CA of related binary indicators When interrelated binary indicator variables are analyzed, one may think of using multiple correspondence analysis (MCA) to analyze them. However, MCA is analogous to the principal component analysis (PCA) of the indicator variables and analyzes a rectangular dataset in which the rows represent individuals ($i = 1, \dots, I$) and columns the variables ($j = 1, \dots, J$), and $I \gg J$ (Beh & Lombardo, 2014; Le Roux & Rouanet, 2010; Lebart, Morineau, & Warwick, 1984). For example, a part of archival data being used in the present study includes 5,000 eating disorder patients who were repeatedly measured on eight psychiatric symptoms that are scored dichotomously (0 = *symptom absence* or 1 = *symptom presence*) before and after treatment. If this data set is converted into a contingency table, it constructs a $5,000 \times 16$ (= 8 symptom indicators before treatment + 8 symptom indicators after treatment) matrix for MCA. However, such an analysis may optimally scale 5,000 patient points and symptom indicators before and after treatment but does not optimally scale the differences in the symptom indicators between before- and after-treatment.

Utility of matched CA over MCA Matched CA is needed to optimally scale the binary score differences between two time points. Specifically, the coordinates of the *difference*

dimensions represent the differences between before and after treatment. In the example of the eating disorder patients, if the coordinates were negative, they would represent an improvement of the symptoms. If the coordinates were positive, they would represent a deterioration. Note that the positive values of the coordinates represent worsened symptoms and the mean of coordinates is set to be zero. If the coordinates were zero, the symptoms would remain the same. The next few sections introduce matched matrices with example data and singular value decomposition (SVD) of the matched matrices that is fundamental to optimal scaling of the differences in the symptom indicators measured at two time points.

Prepare an appropriate matrix for matched CA

Discretizing ages In this study, statistical differences between binary symptom indicators are tested before and after treatment. Specifically, using matched CA, we can examine how age groups differ in the psychiatric symptom indicators between two time points. To do so, ages are discretized into ten groups. Since individuals are matched into their age groups, we may lose between-individual matching within the age groups. To minimize such loss, the number of (age) groups should be maximized. Depending on the character of any particular study, a researcher will try to maximize the number of groups, while maintaining a reasonable number of frequencies in each age group, which will minimize the loss of the individual matching in newly generated (age) groups. Considering both the number of groups (which minimizes loss of between-individual matching) and reasonable frequency numbers for each cell, a total of ten groups was determined to be optimal in the present study. Our contingency table generated from the original data set is a 10 (age groups) $\times$ 8 (psychiatric symptom-presence indicators) matrix. Note that we intentionally do not include “0” (= *symptom absence*) indicators that are redundant score categories for examining the treatment effects.

Generating a concatenated matrix Let $\mathbf{B}_{10 \times 8}$ be a symptom indicator matrix before treatment (simply Before) and $\mathbf{A}_{10 \times 8}$ be a symptom indicator matrix after treatment (simply After). For matched CA, we need to stack $\mathbf{A}$ on top of $\mathbf{B}$ and then concatenate them, creating the following arrangement:

$$\mathbf{C}_{20 \times 16} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \quad (2)$$

This new concatenated matrix $\mathbf{C}$ is of dimension 20×16 , repeating the rows and columns twice. Having 20 rows and 16 columns, the maximum dimensionality of $\mathbf{C}$ is 15, since dimensionality = $\min(20 - 1, 16 - 1)$. The 15 dimensions are split into two sets; one for the (After + Before) sum

dimensional spaces and the other for the (After – Before) difference dimensional spaces.

Extract sum and difference dimensions from the concatenated matrix $\mathbf{C}$

Recall that $\mathbf{A}$ and $\mathbf{B}$ are both of dimension 10×8 and two separate *singular value decompositions* (SVDs) can be performed, one for the sum ($\mathbf{A} + \mathbf{B}$) and another for the difference ($\mathbf{A} - \mathbf{B}$) (Greenacre, 2003). Before we apply the SVDs to analysis of the concatenated matrix, the utility of the SVD algorithm to extract the best (in terms of least squares) dimensions will briefly be introduced.

SVDs of $(\mathbf{A} + \mathbf{B})$ and of $(\mathbf{A} - \mathbf{B})$. We conduct an SVD of ($\mathbf{A} + \mathbf{B}$) and of ($\mathbf{A} - \mathbf{B}$), so that

$$[\mathbf{A} + \mathbf{B}]_{10 \times 8} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad [\mathbf{A} - \mathbf{B}]_{10 \times 8} = \mathbf{X} \mathbf{L} \mathbf{Y}^T \quad (3)$$

where $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$, $\mathbf{X}^T \mathbf{X} = \mathbf{Y}^T \mathbf{Y} = \mathbf{I}$ (to constrain these sums of products to 1), and $\mathbf{\Sigma}$ and $\mathbf{L}$ are diagonal matrices consisting of the singular values of the sum matrix and difference matrix, respectively. Then, the concatenated matrix of Eq. 2 can be decomposed as follows:

$$\begin{aligned} \mathbf{C}_{20 \times 16} &= \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{U} & \mathbf{X} \\ \mathbf{U} & -\mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{L} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{V} & \mathbf{Y} \\ \mathbf{V} & -\mathbf{Y} \end{bmatrix}^T \end{aligned} \quad (4)$$

In Eq. 4, the left and right singular vectors are orthonormal. It is important to note that the same sign in the singular vectors in $[\mathbf{V}, \mathbf{V}]$ correspond to the sum ($\mathbf{A} + \mathbf{B}$) dimensions and the change in sign of the singular vectors in $[\mathbf{Y}, -\mathbf{Y}]$ correspond to the ($\mathbf{A} - \mathbf{B}$) difference dimensions. The constant $1/\sqrt{2}$ is multiplied to the left and the right singular vectors to ensure the correct normalization, such as and $\mathbf{X}^T \mathbf{X} = \mathbf{Y}^T \mathbf{Y} = \mathbf{I}$. For example, focusing on the normalization of $\mathbf{X}$,

$$\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{X} \\ -\mathbf{X} \end{bmatrix}^T \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{X} \\ -\mathbf{X} \end{bmatrix} = \frac{1}{2} \mathbf{X}^T \mathbf{X} + \frac{1}{2} \mathbf{X}^T \mathbf{X} = \mathbf{I},$$

whereas, for the normalization of $\mathbf{Y}$,

$$\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Y} \\ -\mathbf{Y} \end{bmatrix}^T \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Y} \\ -\mathbf{Y} \end{bmatrix} = \frac{1}{2} \mathbf{Y}^T \mathbf{Y} + \frac{1}{2} \mathbf{Y}^T \mathbf{Y} = \mathbf{I} \quad (5)$$

This demonstration can be easily applied to $\mathbf{U}$ and $\mathbf{V}$ to show $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$.

Sum and difference dimensions embedded in $\mathbf{C}$ Although the SVD results for the sum ($\mathbf{A} + \mathbf{B}$) and the difference ($\mathbf{A} - \mathbf{B}$) matrices do not appear specifically in Eq. 4, they are embedded, and sum and difference dimensions are arranged in

descending order according to the magnitude of the corresponding singular values. In the SVD of the concatenated matrix $\mathbf{C}$ (see Eqs. 3 and 4), the sum and the difference singular vectors, which correspond to sum or difference dimensions, can be easily identified. The left/right singular vectors corresponding to the sum dimensions have two identical vectors stacked on top of each other, whereas the left/right singular vectors corresponding to the *difference* dimensions have positive vectors stacked on top of the negative of the same vectors. Accordingly, the sum-of-squares of the elements of the concatenated matrix $\mathbf{C}$ can be decomposed into the sum and difference components;

$$\sum_i \sum_j (a_{ij} + b_{ij})^2 + \sum_i \sum_j (a_{ij} - b_{ij})^2$$

where a and b are elements in the matrix $\mathbf{A}$ and $\mathbf{B}$, respectively. Similarly, the total eigenvalue can be decomposed into the eigenvalues (inertia) sum and into the eigenvalues (inertia) difference.

Normalizing $\mathbf{C}$ for SVD The separate SVD of the sum ($\mathbf{A} + \mathbf{B}$) and the difference ($\mathbf{A} - \mathbf{B}$) is demonstrated in the concatenated matrix $\mathbf{C}$. However, before applying an SVD, a standardization of $\mathbf{C}$ is required for simple CA because the rows and columns are differently weighted, the difference being the relative proportion of their respective margins. Several components are involved in the standardization and their notation is now introduced. The concatenated matrix $\mathbf{C}$ is first converted to the correspondence matrix, $\mathbf{P} = (1/n)\mathbf{C}$, where $n = \mathbf{1}^T \mathbf{C} \mathbf{1}$ is a grand total of $\mathbf{C}$ and $\mathbf{1}$ = a vector of 1s with a length of 20. The i th row and j th column proportions is defined by $r_i = \sum_{j=1}^J p_{ij}$ and $c_j = \sum_{i=1}^I p_{ij}$, respectively ($i = 1, \dots, I = 20$; $j = 1, \dots, J = 16$), so that $\mathbf{r} = \mathbf{P} \mathbf{1}$ and $\mathbf{c} = \mathbf{P}^T \mathbf{1}$. We also define the diagonal matrices of the row and column proportions by $\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$, respectively. The subsequent definitions and results are given in terms of these relative quantities $\mathbf{P} = \{p_{ij}\}$, $\mathbf{r} = \{r_i\}$ and $\mathbf{c} = \{c_j\}$, whose elements add up to 1 in each case. The standardized matrix $\mathbf{C}$ is as follows:

$$\mathbf{C}_{20 \times 16} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{r} \mathbf{c}^T) \mathbf{D}_c^{-1/2} \quad (6)$$

Interpretation of the differences The advantage of matched CA is to separate the sum and difference dimensions from the total dimensional space. The difference dimensional coordinates are of particular interest. These coordinates are assumed to characterize the “true” differences between the binary symptom indicators (column categories) after the initial symptom influences on eating disorder patients are controlled; the difference dimensional coordinates satisfy this requirement. If the coordinates are statistically significant and positive,

symptoms become deteriorated even after the treatment. If the coordinates are not significant, there is no treatment effect, and if they are statistically significant and negative, there is a significant improvement on the symptoms after treatment.

Determine the statistical significance of dimensions and coordinates

A permutation technique is used to test the statistical significance of the eigenvalues (often referred to as principal inertias in the CA literature) for the extracted dimensions. The permutation-based analyses resemble bootstrap analyses since they rely on randomizations of the observed data. The primary difference is that, whereas bootstrap analyses typically seek to quantify the sampling distribution of some statistic computed from the data, permutation analyses typically seek to quantify the null distribution that one expects to see “purely by chance.” We will consider the permutation test in the context of CA, such that the data come in a pair of categorical variables $\{(R_i, C_j)\}$, for $i = 1, \dots, I$ and $j = 1, \dots, J$, where R_i = the i th row category and C_j = the j th column category. This permutation is used to evaluate statistical significance of dimensions in this study and the bootstrapping is used to test statistical stability of the coordinates in statistically significant dimensions (tested by a permutation test) because the matched CA and the classical approach to CA do not provide any test statistics (e.g., standard errors of coordinates) to assess the statistical significance of the coordinates.

Empirical data and measurements

Sample

The sample analyzed here consists of 5,193 female patients for which archival data was made available. The patients were consecutively enrolled in treatment and met *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV TR*; American Psychiatric Association, 2000) criteria for a principal eating disorder (ED) diagnosis at the Remuda Ranch Programs for Eating Disorders in Wickenburg, Arizona. The Institutional Review Board approval was granted for analyzing this archival data set. The age range of the patients was from 12 to 68 years ($M = 22$, $SD = 8.1$), and the ages were discretized into ten age groups, labeled as can be seen in Table 1.

The sample consisted of 93.6% Caucasians, 2.7% mixed/unknowns, 2.1% Hispanics, 0.9% Asians, 0.7% African Americans, and 0.2% Native Americans. A team of clinicians interviewed all patients within two days of admission and gathered detailed information about their

Table 1 Patient age groups: Descriptive statistics

Indicator	Age Group	Sample Size	%
1	14–15	639	12.3
2	16–17	973	18.7
3	18–19	829	16
4	20–21	691	13.3
5	22–23	446	8.6
6	24–25	295	5.7
7	26–27	257	4.9
8	28–29	206	4
9	30–39	544	10.5
10	40–68	313	6

background and symptoms using proprietary structured formats. The interview included an assessment of various clinical characteristics including age of disorder onset and illness duration. To facilitate the objective diagnoses, all patients received admission drug screens and completed extensive and comprehensive psychological testing. With input from multiple data sources, a team psychiatrist and psychologist reached consensus and assigned admission eating disorder and comorbid psychiatric diagnoses.

Measures: Eating Disorder Inventory 2, BMI, and psychiatric comorbidity

Patients were measured on the Eating Disorder Inventory 2 (EDI-2), which consists of 11 subscales, BMI, and psychiatric comorbidity during admission (Before) to inpatient care and again approximately 50 days later at discharge (After) from the program. The EDI-2 (Garner, 1991) is a self-report measure of symptoms related to anorexia nervosa or bulimia nervosa that is frequently used to assess eating disorder psychopathology. The EDI-2 consists of 91 items (rated on a five-point Likert Scale from 0 to 4) organized into 11 subscales. Research has shown that all 11 EDI-2 subscales show significant test–retest reliability coefficients, ranging from .81 to .89 in an eating disorder group (Thiel & Paul, 2006). The mean scores and standard deviations of the 11 EDI-2 subscale scores were: for Before ($M = 100.62$ and $SD = 42.78$) and for After ($M = 60.40$ and $SD = 39.74$). Note that the lower EDI-2 scores the Before mean score represent an improvement of eating disorder symptoms. The mean scores and standard deviations of BMI were: for Before ($M = 18.45$ and $SD = 3.86$) and for After ($M = 20.10$ and $SD = 2.90$). The higher After BMI scores (than the Pre mean score) represent an improvement of body weights after treatment.

The following eight psychiatric diagnoses were also assessed at Before and at After and originally coded dichotomously in the archival data set in which “1” was for the symptom presence and “0” for the symptom absence: (1) *major*

depressive disorder (MD), (2) *dysthymia* (DY), (3) *depression not otherwise specified* (DP), (4) *post traumatic stress disorder* (PT), (5) *obsessive compulsive disorder* (OC), (6) *generalized anxiety disorder* (GA), (7) *anxiety disorder not otherwise specified* (AD), and (8) *social phobia* (SP). In total there were 16 categories, but for the present study, we will analyze only the symptom presence categories.

Evaluate statistical significance of BMI and EDI scores between before and after

BMI and EDI scores are continuous, so to examine whether the BMI and EDI scores increased after treatment, paired t test were conducted to test the statistical differences between Before and After. Then, the paired t test results were compared with those from the binary psychiatric symptom indicators, to study whether the paired t test results and the results from psychiatric symptom score comparisons (between Before and After) are consistent, assuming that treatment is effective for both BMI/EDI and psychiatric comorbidity. Notably, increased BMI scores after treatment have been used as a criterion to release patients with eating disorders.

Evaluate statistical significance of dimensions

We conduct a permutation test to evaluate the statistical significance of the principal inertias (or eigenvalues) for the extracted dimensions. The permutation procedures are described as follows to generate random contingency tables and to calculate the empirical p value:

- Randomly shuffle each cell count in the rows or columns to create a randomly generated table from the original (observed) contingency table;
- Cross-tabulate the permuted data to obtain a random contingency table that has the same dimension as the original concatenated table (20×16), $\tilde{C}$;
- Repeat steps (a) and (b) 10,000 times to create 10,000 randomly generated contingency tables, $\{\tilde{C}^{(1)}, \tilde{C}^{(2)}, \dots, \tilde{C}^{(10,000)}\}$;
- Perform a CA on each of the randomly generated contingency tables $\tilde{C}^{(\#)}$ to create a sampling distribution of random inertias in dimension k , $\{\tilde{\lambda}_k^{(\#)}\}$, where $\# = 1, 2, \dots, 10,000$ (permutation number) and $k = 1, \dots, 15$ (dimension number);
- From the sampling distribution consisting of 10,000 random principal inertias, $\{\tilde{\lambda}_k^{(1)}, \tilde{\lambda}_k^{(2)}, \dots, \tilde{\lambda}_k^{(10,000)}\}$, compare them with an observed principal inertia λ_k ;

- (f) Count any $\{\tilde{\lambda}_k^{(\#)} > \lambda_k\}$ and divide it by 10,000 to calculate an empirical P -value for λ_k ;
- (g) Finally, plot the simulated mean inertias, the 95th percentile inertias, and observed inertias over 15 dimensions.

Evaluating the statistical significance of dimensional coordinates

Since the data consist of repeated measures before and after treatment, the aim was to study psychiatric comorbidity score changes after treatment. The comorbid measures were dichotomously (0, 1) scored, and we rely on the matched CA approach to optimally scale the (After – Before) differences in the symptom indicators. Matched CA estimates two types of dimensions: one type for the After + Before sum effects (sum dimensions), and the other for the After – Before difference effects (difference dimensions).

Examining the coordinates of a difference dimension

The coordinates of the difference dimensions were assumed to represent real symptom difference scores after controlling the initial symptom influences on the patients. Usually, several difference dimensions appear, so the first set of difference coordinates is interpreted, because the first difference dimension accounts for the largest amount of total inertia caused by After – Before differences. In the present study, a test of the statistical significance of the first difference-dimensional inertia and its coordinates was undertaken. Among the dimensional coordinates, only the statistically significant ones were included for further investigation, and their significance was determined from the 95% bootstrap empirical confidence intervals.

Bootstrap empirical confidence intervals (BECI)

Bootstrapping is useful particularly in cases in which one would like to extract a statistic and the sampling distribution of that statistic is not available in closed form (e.g., for frequentist error bars). Bootstrapping provides us with a range of values we would expect for this statistic, given the degree of variation in the dataset, assuming that the sample (the observed contingency table, in the present case) contains variability commensurate with the variability one would get by sampling new data sets from the same (finite) population. According to the recommendation from Efron and Tibshirani (1993), 2,000 bootstrap-concatenated (20×16) tables were generated with replacement, each bootstrap table is analyzed by (simple) CA. Then, a coordinate of indicator t on difference dimension k (denoted by $\varnothing_{t(k)}^{(b)}$) had 2,000 replicates ($b = 1, \dots, 2,000$), $\varnothing_{t(k)}^{(\cdot)} = (\varnothing_{t(k)}^{(1)}, \varnothing_{t(k)}^{(2)}, \dots, \varnothing_{t(k)}^{(2,000)})$, which constituted a sampling distribution of $\varnothing_{t(k)}$. The mean,

$\bar{\varnothing}_{t(k)} = (1/2,000) \sum_{b=1}^{2,000} \varnothing_{t(k)}^{(b)}$, and standard deviation of the sampling distribution, $\sqrt{\sum_{b=1}^B (\varnothing_{t(k)}^{(b)} - \bar{\varnothing}_{t(k)})^2 / (B-1)}$ and $B = 2,000$ were computed, and the standard deviation was in fact a bootstrap standard error of $\varnothing_{t(k)}$. Because the $\varnothing_{t(k)}$ sampling distribution was not necessarily normal, the Gaussian approach could not be automatically applied for constructing confidence intervals. Therefore, the BECI approach is proposed here. To construct the 95% empirical confidence interval for $\varnothing_{t(k)}$, the 2.5th percentile value was chosen as the upper tail of the $\varnothing_{t(k)}$ sampling distribution, $\varnothing_{t(k)}^{(\text{UP})}$, and the 97.5th percentile as the lower tail of the distribution, $\varnothing_{t(k)}^{(\text{LO})}$, which is the 95% BECI for $\varnothing_{t(k)}$: $(\varnothing_{t(k)}^{(\text{LO})}, \varnothing_{t(k)}^{(\text{UP})})$.

No need for the confidence ellipses in the study Several researchers (e.g., Beh & Lombardo, 2014; Greenacre, 2010) have described the visualization of the bootstrap confidence ellipses for category coordinates to determine their stability in a two-dimensional space. However, the present study was not intended to test the stability of category coordinates in a two-dimensional map; rather, the aim was to evaluate the significance of the cardinal values of coordinates in a given dimension. This is because the main interest was to evaluate the statistical significance of difference-dimensional coordinates. Specifically, the significant coordinates (of symptom presence indicators in the first difference dimension) represent, depending on their directions, an improvement or deterioration of the symptoms after treatment. For example, positive coordinates of the difference-dimension would represent deteriorated symptoms. A zero (or statistically nonsignificant) coordinate would reflect no change in symptoms, and a negative coordinate would represent an improvement of the symptoms.

Results

Paired t tests for BMI and EDI-2 scores

BMI

Since ED patients were repeatedly measured on their BMI both before and after treatment, and because the BMI scores were continuous, paired t tests were conducted to examine whether the After BMI scores increased significantly as compared to the Before BMI scores. Q–Q plots for the Before and After difference scores were examined for the BMI and EDI-2 subscale scores, which indicated that the distributions of the difference scores were approximately normal. A summary of these results can be found in Table 2.

Table 2 Age group changes in body mass index (BMI Δ)

Age Groups	BMI Δ	SD	df	p Value
14–15	2.29	1.63	636	< .001
16–17	1.91	1.86	970	< .001
18–19	1.33	1.45	821	< .001
20–21	1.24	1.69	681	< .001
22–23	1.42	1.49	439	< .001
24–25	1.42	1.44	290	< .001
26–27	1.56	1.54	250	< .001
28–29	1.57	1.86	202	< .001
30–39	1.63	1.42	531	< .001
40–68	1.98	1.56	310	< .001

These results indicate that there was a statistically significant increase in BMI after treatment.

EDI-2

ED patients were also repeatedly measured on EDI-2 both before and after treatment, and because the EDI scores were continuous, paired *t* tests were performed to examine whether there was a statistically significant decrease (Δ) (which would represent an improvement in comorbidity) in After as compared to the Before EDI-2 scores. This test was performed for all ten age groups (see Table 3).

Each subscale difference (11 subscales) was also examined between Before and After across the ten age groups, and the results were all statistically significant with all *p* values < .01, indicating that there were significant improvements in individual EDI-2 subscales, as well.

Since BMI and EDI-2 scores were continuous measures from Before and After, score differences between Before and After could easily be tested with paired *t* tests. However, the psychiatric symptoms were repeatedly measured with binary (0, 1) responses, and it was not possible to conduct

Table 3 Age group changes in Eating Disorder Inventory 2 (EDI Δ) scores

Age Groups	EDI Δ	SD	df	p Value
14–15	– 30.77	42.33	636	< .001
16–17	– 39.77	41.88	972	< .001
18–19	– 41.18	41.12	828	< .001
20–21	– 44.12	40.06	690	< .001
22–23	– 45.7	39.99	445	< .001
24–25	– 48.45	40.73	294	< .001
26–27	– 42.74	40.02	256	< .001
28–29	– 44.94	45.38	205	< .001
30–39	– 38.66	40.98	543	< .001
40–68	– 27.97	36.12	312	< .001

ordinary paired *t* tests to study the differences. Hence, the following section will show how testing the binary response differences in the symptom measures using the matched CA approach described in the previous sections. Because the binary indicator differences were examined in the context of the dimensional scale values (or coordinates) estimated by matched CA, first we needed to test the statistical significance of the CA dimensions.

Empirical *p* values for principal inertias

Utilizing the permutation simulation, the empirical *p* values were calculated for all 15 dimensions from the concatenated frequency table; to four decimal places, they were .0000, .0000, .0000, .7485, .9997, .9894, .9701, .8852, .9931, .9976, .9857, .9995, .9997, and .9610, respectively. This shows that the first four dimensions were statistically significant at $\alpha = .05$. The simulation results included the mean inertias calculated from 10,000 simulated inertias and the 95th percentile simulated inertias along with the observed inertias. These results were summarized in Figure 1.

Examine associations among binary variables

To examine the relationships among binary symptom indicators, Cramer's *V*s was estimated at Before and After. The range of Cramer's *V* was between .01 and .30. Most Cramer's *V* values were significant at $\alpha = .05$. However, the associations between Major Depression and Anxiety Disorder NOS (*p* = .27 at Before

Fig. 1 Results from 10,000 simulated principal inertia mean values and the observed principal inertias. Observed = principal inertias from the original data; Sim. Mean = average of 10,000 simulated principal inertias; Sim. 95% = 95th percentile of the 10,000 simulated principal inertias. This simulated inertia plot indicates that the first four dimensions were statistically significant.

and $p = .34$ at After), between PTSD and Social Phobia ($p = .14$ at Before only), and between Anxiety Disorder NOS and Social Phobia ($p = .96$ at Before and $p = .70$ at After) were not statistically significant at $\alpha = .05$. In sum, except these three pairs of the symptoms, all other symptoms were significantly interrelated both Before and After. Hence, pairwise comparisons between significantly related binary variables using a conventional statistical method, using chi-squared test or McNemar's test, would not be recommended.

Test differences in binary responses at before and after

Matched CA of the concatenated (20×16) contingency table was performed and fifteen dimensions for rows and columns were identified. As the first section of these Results shows, only the first four dimensions were found to be statistically significant. An examination of the coordinate signs revealed that the first dimension was classified as the first sum dimension, the second dimension as the first difference dimension, the third dimension as the second sum dimension, and the fourth dimension as the third sum dimension. The first two dimensions were retained for further investigation because they accounted for the largest amount of total inertia. The first sum dimension (Dimension 1) accounted 59.6% of the total inertia, whereas the first difference dimension (Dimension 2) accounted for 16.6% of total inertia. To test the statistical significance of the coordinates, we can calculate the bootstrap standard errors and the 95% BECI's for the coordinates along the first two dimensions.

Age-level coordinates

Table 4 summarizes the original coordinates of the age groups (rows), the mean coordinates estimated from 2,000 bootstrapped

coordinates, the standard errors (SEs), and the 95% BECI's. As is shown in Table 4, most coordinates of Sum Dimension 1 (except age groups 18–19 and 20–21) are statistically significant, whereas for Difference Dimension 2, the coordinates of age groups 12–15, 16–17, 20–21, and 30–39 were statistically significant, on the basis of the 95% BECI results. Note that the age group coordinates of the difference dimension were supposed to represent treatment efficacy after treatment for each age group when all (eight) psychiatric symptoms were aggregated. However, none of the coordinates of the difference dimension were negative, implying no treatment efficacy for psychiatric comorbidity in any age group. Furthermore, comorbidity worsened in age groups 12–15, 16–17, 20–21, and 30–39; this is represented by their positive and significant coordinates.

Misleading interpretation of the sum-dimensional coordinates of age levels In the sum dimension, the coordinate value, 0.46, of the 22–23 age group (which was an aggregated scale value between Before and After treatment, estimated by the CA) indicates a deterioration in their comorbidity. Similarly, the coordinate of 0.5 for ages 24–25, the coordinate of 0.85 for ages 26–27, the coordinate of 0.63 for ages 28–29, the coordinate of 1.32 for ages 30–39, and the coordinate of 1.19 for ages 40–68 all represent a deterioration of the comorbidity, because these coordinates were positive and significant. However, ages 12–15 and 16–17 indicate an increase in their comorbidity, because the coordinates were significant and negative (-1.78 and -1.25). Coordinates from the 10 (age groups) $\times$ 8 (After symptom-presence indicators) contingency table were also calculated, to identify similarities or differences between the sum dimension and the After symptom coordinates. The After symptom coordinates are in parentheses next to the sum dimension coordinates in Table 4. As is shown in the table, the sum dimensional coordinates and the coordinates estimated from the After

Table 4 95% bootstrap empirical confidence intervals (BECIs) for the patient age groups

Age	Sum Dimension 1					Difference Dimension 2				
	Cord	SE	Mean	Lower	Upper	Cord	SE	Mean	Lower	Upper
12–15	$-1.78(-1.77)$	0.13	-1.76	-2	-1.5	1.09	0.32	0.96	0.29	1.54
16–17	$-1.25(-1.21)$	0.1	-1.23	-1.43	-1.04	1.03	0.35	0.91	0.1	1.5
18–19	0.13 (0.09)	0.14	0.13	-0.15	0.4	1	0.55	0.92	-0.36	1.81
20–21	0.2 (0.18)	0.14	0.19	-0.07	0.47	0.9	0.32	0.82	0.14	1.41
22–23	0.46 (0.39)	0.17	0.45	0.11	0.78	1	0.48	0.9	-0.15	1.78
24–25	0.5 (0.36)	0.23	0.49	0.07	0.93	0.93	0.51	0.83	-0.24	1.76
26–27	0.85 (0.93)	0.23	0.84	0.39	1.27	1.06	0.54	0.93	-0.19	1.9
28–29	0.63 (0.46)	0.28	0.62	0.07	1.15	1.07	0.78	0.96	-0.69	2.36
30–39	1.32 (1.42)	0.16	1.3	0.98	1.62	0.9	0.35	0.8	0.05	1.43
40–68	1.19 (1.25)	0.2	1.17	0.77	1.58	1.08	0.48	0.96	-0.03	1.85
% Var	59.6					16.6				

Cord = coordinate; SE = bootstrap standard error estimates; The values in parentheses = coordinates estimated from only After 10×8 cross-tabulation; Mean = the mean coordinates computed from 2,000 bootstrap coordinates; Lower = the 2.5th percentile in the 95% bootstrap empirical confidence interval; Upper = the 97.5th percentile. The statistically meaningful coordinates are bolded; especially in Difference Dimension 2, the red-highlighted positive coordinates (12–15, 16–17, 20–21, and 30–39) represent deteriorated psychiatric comorbidity in these age groups.

symptoms are very similar. This implies that the proportions in the distribution of the symptom-presence indicators for the sum dimension were very similar to those for the After symptom dimension across age groups. Neither the sum coordinates nor the coordinates of the After symptoms represent differences in treatment efficacy after controlling the initial psychiatric symptom effects on the ED patients. Therefore, if a researcher uses the (sum or After symptom) dimensional coordinates to assess treatment efficacy for psychiatric comorbidity after treatment, their assessment would be misleading.

Interpreting the difference-dimensional coordinates of the age groups The difference-dimensional coordinates adequately represent treatment efficacy changes after controlling for the initial psychiatric symptom effects on the patients. For Table 4, the summary of Difference Dimension 2 shows that the coordinates for ages 12–15, 16–17, 20–21, and 30–39 were positive and statistically significant, implying a deterioration of comorbidity. On the other hand, for the remaining age groups, the coordinates are positive but not statistically significant, indicating that there is no improvement; that is, psychiatric comorbidity still remains.

Psychiatric symptom coordinates

Table 5 summarizes the bootstrap results for the column (eight symptom indicators) dimensional coordinates. These are assumed to represent the treatment effect for each symptom indicator when the age groups are aggregated. Likewise, Dimension 1 corresponds to the first sum dimension of the symptom indicators and Dimension 2 reflects the first difference dimension of the symptom indicators.

Misleading interpretation of the sum-dimensional coordinates of psychiatric symptoms When examining the treatment efficacy changes after treatment, one may interpret the Dimension 1 coordinates as reflecting that the psychiatric symptoms, DP (depression not otherwise specified), OC (obsessive compulsive disorder), and AD (anxiety disorder not otherwise specified) improved after treatment because their scale values were negative and statistically significant. Also, the coordinates from the 10 (age groups) $\times$ 8 (After symptom indicators) contingency table were estimated in order to show (dis) similarities in the sum dimension coordinates. As is shown in Table 5, the sum dimensional coordinates and the coordinates estimated from the After symptoms were very similar, implying that the proportions in the distribution of symptom presence indicators for the sum dimension were very similar to those of the After symptom dimension across the age groups. Neither the sum coordinates nor the coordinates of the After symptoms represented true treatment efficacy changes in comorbidity after controlling for the initial psychiatric symptom effects on the ED patients. Therefore, the sum or After symptom dimensional coordinates would give a

misleading interpretation of the treatment efficacy for psychiatric comorbidity after treatment.

Interpreting the difference-dimensional coordinates of psychiatric symptoms The coordinates of the difference dimension (Dimension 2) represent the true treatment efficacy changes after ED treatment. According to the difference dimensional coordinates, OC (obsessive compulsive disorder), GA (generalized anxiety disorder), and SP (social phobia) worsened, since their coordinate values were positive and statistically significant. MD (major depressive disorder), DY (dysthymia), PT (post traumatic stress disorder), and AD (anxiety disorder not otherwise specified) showed no improvement since their coordinates were not statistically significant and assumed to be zero (sustained the same as before treatment). Only DP (depression not otherwise specified) showed improvement, because its coordinate was negative and statistically significant. In summary, the results based on the difference-dimensional coordinates generally indicate no improvement or deterioration in psychiatric symptoms after treatment, except for DP.

Discussion

In the study, matched CA was introduced in order to test the statistical differences in the related binary psychiatric measures between two time points. Testing differences in the binary values (which were interrelated) at two time points across ten different age groups would not be possible with conventional analytical methods (e.g., paired *t* test, McNemar test, or logistic regression paradigms) or simple/multiple correspondence analysis. Thus, the use of matched CA was introduced in order to test for such differences. By implementing this approach, the coordinate values of the difference dimension were considered as true differences between the before- and after-treatment effects in the psychiatric symptom indicators. Interestingly, the paired *t* test results from BMI and EDI-II were different from the matched CA results from the binary psychiatric symptom indicators. The paired *t* test results imply that the eating disorder patients were improved in their body weights (measured by BMI) and severity of eating disorders (measured by EDI-2), whereas psychiatric comorbidity was hardly improved. These contrasting results indicate different treatments may be required to treat psychiatric comorbidity for the eating disorder patients.

Pros and cons of matched CA

Pro 1 Researchers could mistakenly use the traditional approach to CA of the 10 (age groups) $\times$ 16 (psychiatric symptom-presence indicators before and after treatment) data but doing so does not represent the real differences in the symptom indicators after treatment. Hence, clinical evaluations based on these results would be misleading. Only matched CA of the data

Table 5 95% bootstrap empirical confidence interval (BECIs) results for psychiatric symptom-presence indicators

Symptom	Sum Dimension 1					Difference Dimension 2				
	Cord	SE	Mean	Lower	Upper	Cord	SE	Mean	Lower	Upper
MD	0.24 (0.27)	0.09	0.24	0.06	0.41	0.12	0.19	0.09	− 0.29	0.46
DY	2.42 (2.41)	0.19	2.4	2.03	2.78	0.62	1.02	0.61	− 1.67	2.45
DP	− 1.04 (− 1.02)	0.12	− 1.03	− 1.26	− 0.79	− 0.86	0.28	− 0.77	− 1.28	− 0.18
PT	1.81 (1.66)	0.17	1.79	1.44	2.13	1.24	0.78	1.11	− 0.62	2.46
OC	− 0.35 (− 0.49)	0.18	− 0.34	− 0.7	− 0.01	1.42	0.38	1.29	0.49	1.99
GA	0.42 (0.48)	0.15	0.42	0.11	0.71	1.43	0.36	1.3	0.51	1.88
AD	− 1.08 (− 1.16)	0.11	− 1.07	− 1.28	− 0.84	0.11	0.23	0.1	− 0.36	0.54
SP	− 0.04 (− 0.23)	0.31	− 0.05	− 0.64	0.54	2.93	0.85	2.63	0.82	4.02
% Var	59.6					16.6				

MD = major depressive disorder, DY = dysthymia, DP = depression not otherwise specified, PT = post traumatic stress disorder, OC = obsessive compulsive disorder, GA = generalized anxiety disorder, AD = anxiety disorder not otherwise specified, and SP = social phobia. Cord = coordinates; SE = bootstrap standard error estimates; The values in parentheses = coordinates estimated from only After 10 × 8 cross-tabulation; Mean = the mean coordinates computed from 2,000 bootstrap coordinates; Lower = the 2.5th percentile in the 95% bootstrap empirical confidence interval; Upper = the 97.5th percentile. The statistically meaningful coordinates were bolded; especially in Difference Dimension 2, the negative coordinate represented the improved symptom (DP) and the positive ones deteriorated psychiatric symptoms (OC, GA, and SP), respectively.

provides the appropriate measures—the difference-dimensional coordinates—that reflect the true differences in the symptoms after treatment since these coordinates are calculated after controlling the initial symptom effects (before treatment) on the patients. Unlike the traditional parametric multivariate methods (e.g., factor analysis, structural equation modeling, or hierarchical linear modeling), there is no restriction of sample sizes considering the number of input variables, but it is important to keep reasonable counts in each cell in a contingency table in order to avoid biased results because of unbalanced cell frequencies.

Pro 2 In the study, matched CA was applied in order to analyze the within-group data, but the matched CA paradigm can be extended to analysis of within- and between-group data. Let's assume: A = females at Time 1; B = males at Time 1; C = females at Time 2; and D = males at Time 2. This two block circulant matrices ABBA and CDDC for the gender differences are nested within another “ABBA” (or “CDDC”) style block circulant matrix for the two time points in order to examine the gender effects nested within time. Regarding the matrix format, CDDC is stacked on top of ABBA and then they are concatenated in the following pattern (e.g., Greenacre, 2003, 2017; Greenacre & Korneliusson, 2015):

$$\begin{bmatrix} \text{C D} & \text{A B} \\ \text{D C} & \text{B A} \\ \text{A B} & \text{C D} \\ \text{B A} & \text{D C} \end{bmatrix}$$

With this matrix, a researcher can examine both gender and time differences.

Cons However, for using matched CA, one has to construct a two-way concatenated table from the original data set. If the data contain natural categories (e.g., gender, race/ethnicity, region or country), it is simple to generate a two-way table. However, as is

shown in the example, when a continuous variable is involved, a researcher needs to discretize it. One way is to use *z* scores; a researcher first converts scores of the continuous variable into *z* scores and make them ordered categories. In the previous study by Kim and Frisby (2019), they used the *z* scores less than − 1 to create the low level in intelligence; the *z* score within − 1 and 1 to create the middle level; and the *z* score larger than 1 to make the high level. Or, as shown in the present study, ages can be discretized into several groups, considering reasonably balanced distributions (e.g., approximately equal) of counts in each age group. As was explained previously, age group matching may cause loss of between-individual matching (within the age groups), and a research may want to generate as many as categories to minimize the loss (M. J. Greenacre, personal communication, June 2, 2018). However, for the age grouping, a researcher has to consider the number of frequencies included in each category; if too many categories are generated, there would be many empty cells in a table. Thus, a caution is required to balance the number of generated categories and reasonable cell counts in a category (e.g., minimum of five observations in a category).

Conclusion

The hope here is that researchers can use the version of matched CA introduced here for their research when they utilize between-matching (e.g., gender), within-matching (e.g., age), or both between- and within-matching (e.g., gender and age) for their analysis. All necessary R codes and data used in the study are included in [Appendix](#), so that researchers may use the R codes for their research or may replicate the present results.

Acknowledgements I thank W. Wu and two anonymous reviewers for their helpful comments to enhance the quality of this study. In addition, I sincerely appreciate my colleague, L. Feurstahler, for her willingness to proofread this article.

References

- American Psychiatric Association. (2000). DSM-IV: Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.
- Beh, E. J., & Lombardo, R. (2014). Correspondence analysis: Theory, practice and new strategies. West Sussex, UK: Wiley. doi:<https://doi.org/10.1002/9781118762875>
- Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. New York, NY: Chapman & Hall.
- Garner, D. M. (1991). Eating Disorder Inventory-2 Manual. Odessa, FL: Psychological Assessment Resources.
- Greenacre, M. J. (2003). Singular value decomposition of matched matrices. *Journal of Applied Statistics*, 30, 1101–1113. doi:<https://doi.org/10.1080/026647.6032000107132>
- Greenacre, M. J. (2010). Biplots in practice. Madrid, Spain: Fundación BBVA.
- Greenacre, M. J. (2017). Correspondence analysis in practice (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Greenacre, M. J., & Korneliussen, T. (2015). *Total inertia decomposition in tables with a factorial structure*. Paper presented at the 2015 CARME (Correspondence Analysis and Related Methods) conference, Naples, Italy.
- Le Roux, B., & Rouanet, H. (2010). Multiple correspondence analysis. Thousand Oaks, CA: Sage.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). Multivariate descriptive statistical analysis. New York, NY: Wiley.
- Kim, S.-K., & Frisby, C. L. (2019). Gaining from discretization of continuous data: The correspondence analysis biplot approach. *Behavior Research Methods*, 51, 589–601. doi:<https://doi.org/10.3758/s13428-018-1161-1>
- Thiel, A., & Paul, T. (2006). Test–retest reliability of the Eating Disorder Inventory 2. *Journal of Psychosomatic Research*, 61, 567–569. doi:[https://doi.org/10.1016/S0022-3999\(06\)00505-8](https://doi.org/10.1016/S0022-3999(06)00505-8)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.