Bayesian Analysis (2010) **5**, Number 3, pp. 519–532

Correspondence Analysis with Incomplete Paired Data using Bayesian Imputation

Jules J. S. de Tibeiro* and Duncan J. Murdoch†

Abstract. In this paper we consider the analysis of incomplete tables using correspondence analysis. We focus on a dataset concerning congenital heart disease (Fraser and Hunter 1975), in which the data forms a square table, but only a symmetrized version of the off- diagonal entries was reported. We use Markov chain Monte Carlo (MCMC) on a hierarchical Bayes model to estimate the underlying rates, and use correspondence analysis to study the relationships in the completed table.

Keywords: correspondence analysis, missing data, Markov chain Monte Carlo

1 Introduction

Correspondence analysis (CA) is a statistical technique to display the structural relationships in a two-way table of counts (Benzécri 1973, 1992; Lébart et al. 1984; Greenacre 1984).

The main tool is the biplot, a plot of points corresponding to each row (or each column, or both). As in other biplot techniques, CA offers the remarkable feature of jointly representing individuals and variables. A table with independent rows and columns would have all points plotted near the origin; departures from the origin measure departures from independence in terms of their contributions to a χ^2 statistic testing independence. Not only does one gain insight in the relationship amongst individuals and amongst variables, but one can also find an indication of which variables are important in the description of which individuals (Gordon 1999). (We will give more details about CA in section 2 below.)

In this paper we consider the CA of part of a dataset collected by Fraser and Hunter (1975) on congenital cardiovascular defects. A congenital cardiovascular defect occurs when the heart or blood vessels near the heart don't develop normally before birth. Congenital cardiovascular defects are present in about 1 percent of live births. The goal of Fraser and Hunter (1975) was to reveal etiologic relations among cardiac lesions. Sibships in which two or more children had dissimilar cardiac lesions were culled from reported studies and other sources. Fraser and Hunter (1975) collected information on 13 lesions; for simplicity we consider the subset of the 7 most frequent ones.

The Fraser and Hunter (1975) data (Table 1) is incomplete. There are no records of

*Sect. Sciences, Université de Moncton, Shippagan, Nouveau-Brunswick, Canada, mailto:jules.de.tibeiro@umcs.ca

†Dept. of Stat. and Act. Sci., University of Western Ontario, London, Ontario, Canada, mailto:murdoch@stats.uwo.ca

 DOI:10.1214/10-BA520

Table 1: Observed data on pairs of cardiac malformations. Actual counts X_{ij} were not observed; the table shows values $k_{ij} = X_{ij} + X_{ji}$. Values on the diagonal were unobserved and counted as zero in the column totals.

	ToF	*VSD*	*PS*	*TGV*	*PDA*	*AS*	*ASD*
ToF	.	13	19	10	4	1	1
VSD	.	.	3	5	3	3	6
PS	.	.	.	2	0	1	1
TGV	.	.	.	.	4	1	2
PDA	.	.	.	.	.	2	0
AS	.	.	.	.	.	.	2
ASD	.	.	.	.	.	.	.

cases where two siblings had the *same* defect. Additionally, the pairings are unordered, so only the top off- diagonal entries in the table are recorded, representing the total number of pairs occurring in either order.

CA is a natural method for investigating dependencies in contingency tables, but it requires complete tables. Our goal in this paper is to describe how to deal with the missing data and the lack of ordering in datasets like the Fraser and Hunter (1975) data.

de Leeuw and van der Heijden (1988) handled the missing data problem iteratively: missing values were initialized, then recalculated assuming the independence model until convergence was achieved. In this way a complete table was imputed in which the missing entries had minimal contributions to the chi-square calculations. van der Heijden et al. (1989) showed how a similar approach could be used to deal with departures from other models, such as the symmetry model and the quasi-symmetry model. de Tibeiro (1996) used a minimal trace criterion to impute the missing entries in the Fraser and Hunter (1975) dataset. Ben Salem (1992) proposed a sufficient condition for the existence of finite values of missing data satisfying the minimum trace criterion.

Several other papers have also considered the Fraser and Hunter (1975) dataset. MacGibbon (1983) used log-linear modelling and proposed a new method of outlier detection. Dinwoodie and MacGibbon (2004) used Markov chain Monte Carlo (MCMC) methods to approximate exact p-values for tests of quasi-independence. Dinwoodie et al. (2004) computed the p-values exactly using generating functions.

In contrast to these earlier papers, in this paper we will describe a Bayesian approach for handling the missing data. The missing values will be imputed as medians of their posterior distributions. Posterior simulations will be used to illustrate the uncertainty in the imputation. Like the other imputation methods, the Bayesian method will allow standard CA in the presence of missing data, rather than providing a formal test of independence or quasi-independence. Unlike them, it will incorporate and display the uncertainty due to the missing data.

The remainder of the paper is organized as follows. In section 2, we give an overview

of CA. Section 3 describes a Bayesian approach to reconstruct the diagonal for the purpose of CA. In section 4, we apply CA on the cardiology data using this approach.

2 Overview of Two-Way Correspondence Analysis

We provide here a short overview of CA. For more details, see Benzécri (1992), Greenacre and Hastie (1987), Lébart et al. (1984, 1997), and Greenacre (1984, 2007).

As mentioned above, CA attempts to display the structural relationships in a two-way table of counts in coordinates representing the rows and columns. Because our approach is chiefly graphical, following the French tradition of Benzécri (1973), we consider CA as an adaptation to categorical data of Principal Component Analysis (PCA). CA displays a low-dimensional projection of the data (e.g., into a plane). It does this for two variables or categories simultaneously, thus revealing associations (and oppositions) between them.

Similar to PCA, CA provides eigenvalues that are squared singular values (called principal inertias in CA), percentages of explained variance (percentages of inertia), factor loadings (correlations with principal axes), and communalities (percentages of explained inertia for individual rows or columns).

Contrary to the loglinear analysis approach (Bishop et al. 1975), no model is assumed in CA. However, the metric is based on the assumption that the variance of an entry is proportional to its mean, as it would be in a Poisson model. CA is essentially a method for a weighted least- squares approximation of a frequency matrix. The projections are the central output of CA.

Let I rows and J columns be collected into the $I \times J$ matrix $\mathbf{N}$ with elements k_{ij}. (For convenience of notation we will assume $I \geq J$ in this general discussion; in our cardiac data, $I = J$). Let k_{i+} and k_{+j} denote the sum of the ith row and jth column, respectively, and $k_{++} = \sum_i \sum_j k_{ij}$ denote the grand total of $\mathbf{N}$. The mass of the jth column is defined as $c_j = k_{+j}/k_{++}$, and likewise the mass of the ith row is $r_i = k_{i+}/k_{++}$. All of CA is based on the so-called *correspondence matrix* of relative frequencies $\mathbf{P}$ with entries $p_{ij} = k_{ij}/k_{++}$ and the matrix $\mathbf{S}$ with elements $s_{ij} = (p_{ij} - r_i c_j)/\sqrt{r_i c_j}$. We note here that this makes CA invariant to rescaling of the original matrix $\mathbf{N}$.

The *total inertia* $\Phi^2 = \chi^2/k_{++}$ is known as *Pearson's mean-square contingency coefficient*, where χ^2 is Pearson's chi-squared statistic of the data matrix, i.e. the sum of squares of the matrix $\mathbf{S}$.

$$\Phi^2 = \mathrm{tr}(\mathbf{S}\mathbf{S}^t) \tag{1}$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(p_{ij} - r_i c_j\right)^2 / (r_i c_j) \tag{2}$$

The data in a contingency table can be used to check for association of two categorical

variables corresponding to the rows and columns of $\mathbf{P}$. The assumption of independence is $E(p_{ij}) = r_i c_j$, $i = 1, \cdots, I$; $j = 1, \cdots, J$. The usual chi-square statistic for testing independence may be written in several forms:

$$\begin{aligned} \chi^2 &= k_{++}\text{tr}(\mathbf{S}\mathbf{S}^t) = k_{++}\text{tr}(\mathbf{S}^t\mathbf{S}) && (3) \\ &= k_{++}\sum_{i=1}^{I}\sum_{j=1}^{J}(p_{ij} - r_i c_j)^2/(r_i c_j) && (4) \\ &= k_{++}\sum_{i=1}^{k}\lambda_i^2 && (5) \end{aligned}$$

where $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_k^2 > 0$ are the nonzero eigenvalues of $\mathbf{S}\mathbf{S}^t$ and k is its rank. The biplot is a projection of the rows or columns of $\mathbf{S}\mathbf{S}^t$ onto the span of the eigenvectors corresponding to λ_1^2 and λ_2^2. When $\lambda_1^2 + \lambda_2^2$ constitutes a substantial proportion of $\sum_{i=1}^{k}\lambda_i^2$, the biplot will display most of the variability contributing to χ^2.

See the references listed above (especially Benzécri (1992) and Greenacre (2007)) for extensive discussions of the interpretation of biplots. Briefly, we can say the following. The points in the biplot correspond to coordinates of the two eigenvectors associated with λ_1 and λ_2, so the orientation of axes is arbitrary. Points appearing near the origin correspond to rows or columns in which the assumption of independence is approximately valid. If the point representing one row or column falls far from the origin, it will generally be opposed by one or more points in the opposite direction; that opposition indicates the way in which independence is violated: the table deviates from independence largely in the entries corresponding to these points. The actual linear combinations of s_{ij} values contributing to each point are not important, but the relative positions and the shape of the overall plot may be.

In the heart defect data shown in Table 1, k_{ij} is the number of cases concerning two consanguineous subjects, in which one is diagnosed with i, the other one with j. The observed table is symmetrical. The diagonal is missing because Fraser and Hunter (1975) excluded the cases where the two subjects presented the same malformation, only counting the cases where they have distinct malformations. Due to this fact, the table does not yield very well to CA. If the missing values are replaced with a *zero diagonal*, the non-responses of Table 1 result in a strong lack of independence between the factors. Figure 1 shows a biplot of this table. The lack of independence is evident in the fact that no columns are plotted near the $(0, 0)$ location at the centre of the display. The first axis displays most of the variation (59% versus 19% for the second axis); this simply reflects the fact that *ToF* and *PS* have such high counts.

There is no usable information in this plot. We might hope to see better information if we had plotted the original X_{ij} counts; the remainder of this paper presents a Bayesian approach to impute them.

3 A Bayesian Approach

Due in part to recent significant computational advances such as Markov chain Monte Carlo (MCMC), Bayesian analysis is becoming a practical choice for dealing with complicated statistical scenarios, such as random effect or mixed models, classification, missing values, image processing, signal detection, etc.

The aim of this section is to describe a Bayesian approach to imputing the missing data in our table of counts, and an MCMC method to approximate the posterior distribution. The correspondence analysis of the results is described in the next section.

3.1 The Statistical Model

Let X_{ij} be the (unobserved) count of pair (i, j), with order taken to be significant. In our actual dataset, the full table is not observed. Instead, we observe totals $k_{ij} = X_{ij} + X_{ji}$ for $i \neq j$, and have no information on X_{ii}. However, we begin by modelling a complete version of the observed table. At the end the missing values will be imputed by looking at their posterior distribution under a hierarchical Bayes model conditional on the observations.

We assume the over-parametrized log-linear model

$$X_{ij} \sim \text{Poisson}(\mu_{ij})$$

where

$$\log \mu_{ij} = \alpha_i + \alpha_j + \delta_{ij} \tag{6}$$

This model differs from the standard loglinear symmetry model (Agresti 1990, p. 353) in two ways. First, there are no identifiability constraints on the parameters. Omitting the identifiability constraints simplifies the model considerably: in the prior model used in the absence of data, the α_i and δ_{ij} terms are independent. Second, our model allows $\delta_{ij} \neq \delta_{ji}$. Allowing for a lack of symmetry in the δ_{ij} values also simplifies the model: the X_{ij} terms are independent in the prior. These two simplifications make our model easier to interpret.

The parameters α_i and δ_{ij} are modelled differently from each other. We believe that there will be large differences in α_i, as these control differences in the marginal responses in the rows or columns of our observation matrix: in our example, differences in the rates of cardiac malformations. However, we don't assume prior knowledge of the relative frequency. Thus our prior distribution on each is

$$\alpha_i \sim N(0, 10^6)$$

independently across i. This represents a diffuse but proper prior distribution. (An improper prior for the α_i terms would behave similarly in the presence of sufficient data; an advantage of using a proper prior is that it guarantees a proper posterior regardless of the amount of data present.) The α_j term in (6) is not modelled separately, as α_j is equal to α_i with $i = j$.

The parameters δ_{ij} measure departures from an independence model. We believe that the different pairings don't follow a pure independence model, but we don't expect the departures to be large. Thus we assume

$$\delta_{ij} \sim N(0, \sigma^2)$$

where the standard deviation σ is given the prior distribution

$$\sigma \sim \text{Unif}(0, 10)$$

The upper limit of 10 on the distribution of σ serves two purposes. First, it provides an upper bound on the variability of δ_{ij}, matching our belief that this variability should not be large. A value of δ_{ij} larger than 10 would make a huge change to the model, contrary to our belief. The bound on δ_{ij} guarantees that the posterior distribution will be proper. The use of a uniform prior on the standard deviation is a "default" choice recommended by Spiegelhalter et al. (2004, p. 173), who discuss a number of alternatives which may be more appropriate in specific situations. A welcome consequence of the upper limit on σ is that it allows us to "borrow strength" from the observed k_{ij} values to estimate the unobserved X_{ij} values.

We used WinBUGS (Spiegelhalter et al. 2003) to carry out the MCMC fit to this model. Once translated into the WinBUGS language, the specification above is sufficient for WinBUGS to automatically construct a Markov chain using Gibbs sampling, whose steady-state distribution is the posterior distribution of the parameters and the unobserved data. Unfortunately, this chain mixes very slowly. Since $k_{ij} = X_{ij} + X_{ji}$ is observed but X_{ij} is not, the posterior distribution of X_{ij} is highly negatively correlated with that of X_{ji}. Gibbs samplers work best when the unknown components are relatively uncorrelated, so we reformulated the model to reduce this correlation as follows:

$$\begin{aligned} \overline{\delta_{ij}} &= (\delta_{ij} + \delta_{ji})/2 \\ \epsilon_{ij} &= (\delta_{ij} - \delta_{ji})/2 \\ \nu_{ij} &= \ln 2 + \alpha_i + \alpha_j + \overline{\delta_{ij}} \end{aligned}$$

for $i < j$. In this parametrization ν_{ij} is used as the log of the mean for off-diagonal k_{ij} values, i.e. for $i < j$

$$k_{ij} \sim \text{Poisson}(\exp(\nu_{ij}))$$

We note that this model is not quite identical to the model presented earlier in this section, where the mean for k_{ij} would be $\exp(\alpha_i + \alpha_j)[\exp(\delta_{ij}) + \exp(\delta_{ji})]$ instead of $2\exp[\alpha_i + \alpha_j + (\delta_{ij} + \delta_{ji})/2]$ as here. For the small values of δ_{ij} we expect, these should be very close. The diagonal terms δ_{ii} were left unchanged in the model.

In this new formulation, the prior for δ_{ij} becomes

$$\begin{aligned} \overline{\delta_{ij}} &\sim N(0, \sigma^2/2) \\ \epsilon_{ij} &\sim N(0, \sigma^2/2) \\ \delta_{ij} &= \overline{\delta_{ij}} + \epsilon_{ij} \\ \delta_{ii} &\sim N(0, \sigma^2) \end{aligned}$$

Posterior draws of X_{ij} are based on the relation

$$X_{ij} \,|\, k_{ij}, \mu_{ij}, \mu_{ji} \sim \text{Binomial}[k_{ij}, \mu_{ij}/(\mu_{ij} + \mu_{ji})]$$

3.2 MCMC and Correspondence Analysis

Once samples are drawn from the posterior distribution, further decisions need to be made before correspondence analysis can be carried out. There are a number of possible matrices which could be used in CA.

One choice would be the imputed X_{ij} matrix; another would be the full k_{ij} matrix including the imputed diagonal. Alternatively, the μ_{ij} matrix or the symmetrized $(\mu_{ij} + \mu_{ji})/2$ matrix could be used. The posterior distribution for each of these is approximated by MCMC, giving a distribution of possible correspondence analyses of the problem.

The most natural choice for a single estimated matrix to use in CA would be the mean or median of the posterior distribution. Because the model is loglinear, the posterior distributions for the matrix entries are highly skewed to the right, and we chose to use the posterior median. It is clear that the posterior medians for k_{ij}, μ_{ij} and X_{ij} are all symmetric matrices, since our prior is symmetric and there is no information in the data to break the symmetry. The matrix entries are not all equally uncertain: non-diagonal entries of k_{ij} are known by observation, with uncertainty only in the missing diagonal entries. The μ_{ij} matrix also includes uncertainty about the actual mean, and X_{ij} includes uncertainty in how the observations are distributed between symmetric pairs.

In actual practice, we would normally do our analysis based on the k_{ij} values. CA is mainly a descriptive technique, and this comes closest to describing the observed data. In cases where CA is used for inference about the underlying model (e.g. to declare that certain characteristics are related in some way in the general population), the plots based on μ_{ij} values would be helpful to show the uncertainty in those inferences.

3.3 Uncertainty in Imputation

The Bayesian analysis produces posterior distributions for the unknown quantities, not just estimates. How should these be incorporated into the analysis?

Greenacre (1984, sec. 8.1) addresses uncertainty in CA using jackknife and bootstrap replications. This is roughly equivalent (albeit using a frequentist approach) to using

Table 2: Posterior medians of $(X_{ij} + X_{ji})/2$ and $(\mu_{ij} + \mu_{ji})/2$.

	k_{ij}						
	ToF	*VSD*	*PS*	*TGV*	*PDA*	*AS*	*ASD*
ToF	9.0	6.5	9.5	5.0	2.0	0.5	0.5
VSD	.	5.0	1.5	2.5	1.5	1.5	3.0
PS	.	.	1.0	1.0	0.0	0.5	0.5
TGV	.	.	.	2.0	2.0	0.5	1.0
PDA	.	.	.	.	0.0	1.0	0.0
AS	.	.	.	.	.	0.0	1.0
ASD	.	.	.	.	.	.	0.0

	$(\mu_{ij} + \mu_{ji})/2$						
	ToF	*VSD*	*PS*	*TGV*	*PDA*	*AS*	*ASD*
ToF	9.3	6.4	7.8	4.7	1.9	1.0	1.0
VSD	.	5.1	1.7	2.5	1.4	1.3	2.1
PS	.	.	0.9	1.1	0.4	0.5	0.5
TGV	.	.	.	1.9	1.3	0.6	0.8
PDA	.	.	.	.	0.4	0.5	0.3
AS	.	.	.	.	.	0.3	0.5
ASD	.	.	.	.	.	.	0.3

the posterior distribution of the symmetrized mean matrix. The idea is to use the estimated matrix to determine the projection, and then to display samples from the posterior distribution of $(\mu_{ij} + \mu_{ji})/2$ by projecting their biplots into the same plane.

4 Application to Cardiology Data

In this section we report the results of the Bayesian analysis described in section 3. We used WinBUGS to run a Markov chain on the reformulated model for 1000 steps of burn-in and a further 50,000 steps to obtain properties of the posterior distribution. We found that WinBUGS performed very well on this model, with fast mixing.

We used the posterior median of $(X_{ij} + X_{ji})/2$ (which is equal to $k_{ij}/2$ with no uncertainty when $i \neq j$) as our imputed data, and the posterior median of $(\mu_{ij} + \mu_{ji})/2$ as our estimate of the posterior mean matrix. Table 2 shows these results. Only the entries on and above the diagonal are shown; the entries below the diagonal mirror those above.

The eigenvalues and the percentages of inertia (in parentheses) explained by the

Table 3: Diagonal estimates using de Tibeiro (1996) method compared with posterior medians.

	ToF	*VSD*	*PS*	*TGV*	*PDA*	*AS*	*ASD*
Minimum trace	29.1	9.5	8.0	4.0	2.8	4.5	0.0
Median of k_{ii}	9	5	1	2	0	0	0

non-trivial eigenvalues are:

$$\lambda_1 = 0.0633 \ (47\%); \quad \lambda_2 = 0.0531 \ (40\%);$$
$$\lambda_3 = 0.0085 \ (6.4\%); \quad \lambda_4 = 0.0071 \ (5.5\%);$$
$$\lambda_5 = 0.0017 \ (1.2\%); \quad \lambda_6 = 0.00004 \ (0.03\%).$$

The sum of the eigenvalues is equal to the χ^2-coefficient of the table of posterior medians divided by the grand total. Here the sum of the first two most significant dimensions accounts for about 87% of χ^2/k_{++}. In other words, two axes out of a possible five will summarize 87% of the departure from independence. This suggests that a planar representation will provide a good visual summary of the data. See Figure 2. While the $(\mu_{ij} + \mu_{ji})/2$ entries are not integer-valued, we may still formally carry out the CA calculations, and have done so in this plot.

The axes of a CA are defined by *contrasts*. Figure 2 reveals that axis 1 is mainly a contrast between *PS* and *ToF*, because they have the largest contributions in that dimension. This suggests an unusual relation between those two malformations, and examination of the original table shows that the *ToF*-*PS* entry of 19 is unusually large, given that the other *PS* entries are all 3 or less, and the other *ToF* entries are all 13 or less.

The second axis consists primarily of the contrast between *PS* and *ToF* versus the other malformations. Thus most of the χ^2 statistic in (3) may be attributed to the differences between the three groups consisting of the singletons *ToF* and *PS*, and the remaining malformations.

We also applied the minimum trace method of de Tibeiro (1996) to the full dataset. The first seven diagonal entries are shown in Table 3. As can be seen, the minimum trace criterion generally gives larger estimates of the missing diagonal entries. Partly this reflects the skewness of the distribution: the posterior medians are smaller than the posterior means. It may also be explained by the fact that the minimum trace is trying to minimize the χ^2 statistic, rather than estimating values that are consistent with the observed data.

To understand the uncertainty in this correspondence analysis, we produced Figure 3. Here 1000 simulations from the posterior are shown by displaying the locations of the projections of each of the first four malformations onto copies of Figure 2. We have

also included contours of the posterior density for each malformation.

The original aim of Fraser and Hunter (1975) was to describe the relations among the cardiac malformations. From Figure 2 we would conclude that the first axis is determined almost exclusively by the opposition between *ToF* and *PS*, while the second axis opposes this pair to all the other malformations. However, Figure 3 shows that the observation about the first axis may not hold for the imputed data, in that there is considerable uncertainty about the location of *ToF* along the first axis, while the second axis is more stable. Indeed, a closer examination of the row and column contributions revealed that the first axis is "inverse" in the sense of Benzécri (1992, p. 412) which would make it difficult to interpret: it would suggest that possession of *ToF* or *PS* in one sibling would convey a protection against the same malformation in the other sibling.

In conclusion, we can say that our Bayesian approach to filling in the missing entries gave insight into the interpretation of the biplot that would not normally be available. In addition, it allows straightforward display of the uncertainty in the imputed results.

Acknowledgements

Part of this work was undertaken while the first author was visiting the Department of Statistical and Actuarial Sciences of The University of Western Ontario (Canada). The first author wishes to thank Professor Ian MacLeod of the Department of Statistical and Actuarial Sciences for his support in the computation and the analysis of the results. Both authors wish to thank Prof. J. P. Benzécri for his helpful comments on symmetrical tables. The second author was supported by an NSERC Discovery Grant.

References

Agresti, A. (1990). *Categorical Data Analysis*. Wiley. 523

Ben Salem, K. (1992). "Le critère de la trace minima pour estimer les données manquantes dans un tableau de correspondance." *Les Cahiers de l'Analyse des Données*, 17: 97–112. 520

Benzécri, J. P. (1973). *L'Analyse des Données, Tome 2: L'Analyse des Correspondences*. Dunod. 519, 521

— (1992). *Correspondence Analysis Handbook*. Marcel Dekker. 519, 521, 522, 528

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis*. MIT Press. 521

de Leeuw, J. and van der Heijden, P. G. M. (1988). "Correspondence Analysis of Incomplete Contingency Tables." *Psychometrika*, 53: 223–233. 520

de Tibeiro, J. (1996). "Sur les traits associés par paires : malformations cardiaques congénitales chez des enfants ayant mêmes parents." *Les Cahiers de l'Analyse des Données*, 21: 45–52. 520, 527

Dinwoodie, I. and MacGibbon, B. (2004). "Exact Analysis of a Paired Sibling Study." *Computational Statistics*, 19: 525–534. 520

Dinwoodie, I. H., Matusevich, L. F., and Mosteig, E. (2004). "Transform Methods for the Hypergeometric Distribution." *Statistics and Computing*, 14: 287–297. 520

Fraser, F. C. and Hunter, A. D. W. (1975). "Etiologic Relations Among Categories of Congenital Heart Malformations." *The American Journal of Cardiology*, 36: 793–796. 519, 520, 522, 528

Gordon, A. D. (1999). *Classification*. Chapman and Hall, 2nd edition. 519

Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press. 519, 521, 525

Greenacre, M. and Hastie, T. (1987). "The Geometric Interpretation of Correspondence Analysis." *Journal of the American Statistical Association*, 82: 437–447. 521

Greenacre, M. J. (2007). *Correspondence Analysis in Practice*. Chapman and Hall/CRC, second edition. 521, 522

Lébart, L., Morineau, A., and Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. John Wiley & Sons. 519, 521

Lébart, L., Salem, A., and Berry, L. (1997). *Exploring Textual Data*. Kluwer. 521

MacGibbon, B. (1983). "A Log-linear Model of a Paired Sibling Study." In Chaubey, Y. and Dwivedi, T. D. (eds.), *Proceedings of Statistics '81 Canada Conference*, 193–197. 520

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley. 524

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2003). *WinBUGS Version 1.4 Users Manual*. MRC Biostatistics Unit, Cambridge. 524

van der Heijden, P. G. M., de Falguerolles, A., and de Leeuw, J. (1989). "A Combined Approach to Contingency Table Analysis Correspondence Analysis and Log-Linear Analysis." *Applied Statistics*, 38: 249–292. 520

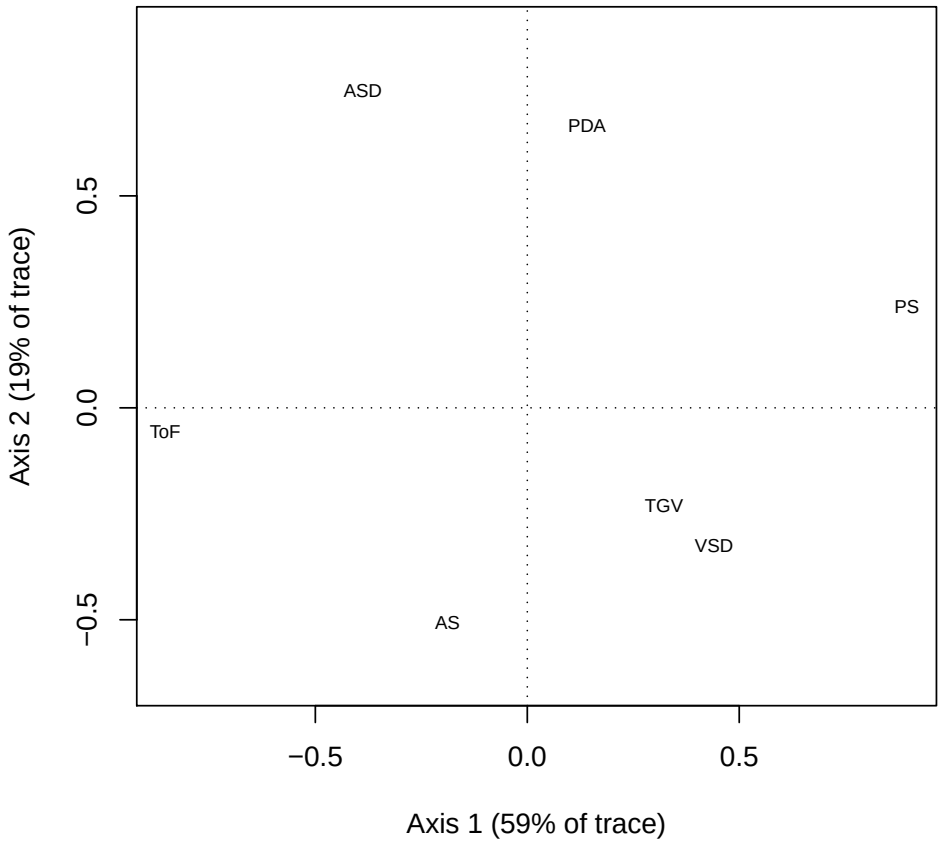

Figure 1: Biplot of the columns of N based on replacing the missing diagonal with structural zeros.

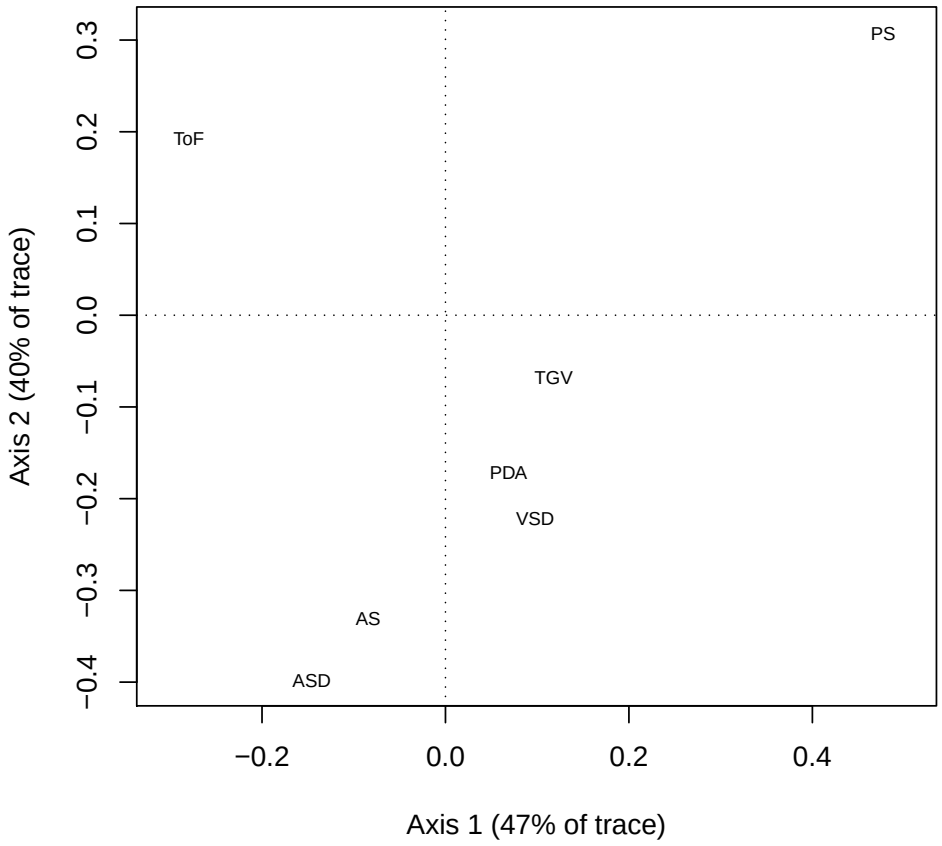

Figure 2: Biplot based on symmetrized posterior median of the μ_{ij} matrix.

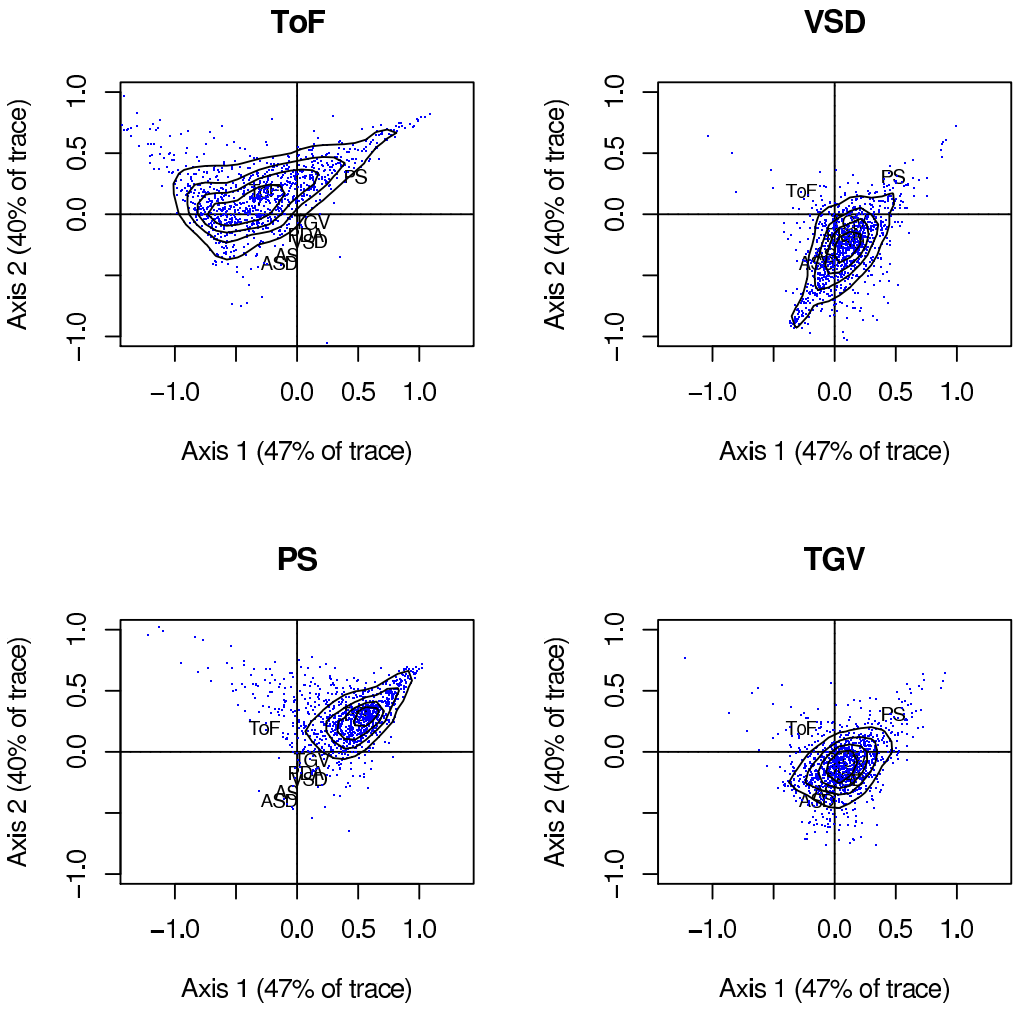

Figure 3: Uncertainty in the biplot of the symmetrized median.