Received: 7 September 2020 | Revised: 14 January 2021 | Accepted: 27 January 2021
DOI: 10.1002/cmm4.1152

SPECIAL ISSUE PAPER

WILEY

Isolated and structured families of models for stochastic symmetric matrices

Cristina Dias[1,2] | **Carla Santos**[2,3] | **João Tiago Mexia**[2,4]

[1]Polytechnic Institute of Portalegre, Portalegre, Portugal
[2]Center of Mathematics and Applications (CMA), Caparica, Portugal
[3]Polytechnic Institute of Beja, Beja, Portugal
[4]Faculty of Sciences and Technology, Nova University of Lisbon, Lisbon, Portugal

Correspondence
Cristina Dias, Polytechnic Institute of Portalegre, Campus Politécnico,10, 7300-555 Portalegre, Portugal.
Email: cpsd@ipportalegre.pt

Funding information
Fundação para a Ciência e a Tecnologia, I.P., Grant/Award Number: UIDB/00297/2020

Abstract

Stochastic symmetric matrices with a dominant eigenvalue, $\boldsymbol{\alpha}$,can be written as the sum of $\lambda\boldsymbol{\alpha}\boldsymbol{\alpha}^t$ (where λ is the first eigenvalue), with a symmetric error matrix $\boldsymbol{E}$. The information in the stochastic matrix will be condensed in its structured vectors, $\lambda\boldsymbol{\alpha}$, and the sum of square of residues, V. When the matrices of a family correspond to the treatments of a base design, we say the family is structured. The action of the factors, which are considered in the base design, on the structure vectors of the family matrices will be analyzed. We use ANOVA (Analysis of Variance) and related techniques, to study the action under linear combinations of the components of structure vectors of the m matrices of the model. Orthogonal models with m treatments are associated to orthogonal partitions. The hypothesis to be tested, on the action of the factors in the base design, will be associated to the spaces in the orthogonal partitions. We will show how to carry out transversal and longitudinal analysis for families of stochastic symmetric matrices with dominant eigenvalue associated to orthogonal models.

KEYWORDS
base design, models, structured families, symmetric stochastic matrix

1 | INTRODUCTION

The use of the pair given by a dominant eigenvalue and the corresponding eigenvector, when they exist, to condense the information in a symmetric matrix showed to be useful when considering series of studies, see.[1-7]

These studies are matrix triplets constituted by a data matrix and two weight matrices. The columns of the data matrix correspond to variables measured in objects, which are associated to the rows. The matrix triplets $(\boldsymbol{X}_j, \boldsymbol{D}_j, \dot{\boldsymbol{D}}_j)$, $i, j = 1, \dots, k$, are condensed in the $\boldsymbol{A}_j = \boldsymbol{X}_j\dot{\boldsymbol{D}}_j\boldsymbol{D}_j\boldsymbol{X}^t_{j,}$, $i, j = 1, \dots, k$, if the studies are on the same objects or $\boldsymbol{B}_j = \boldsymbol{X}^t_j\boldsymbol{D}_j\boldsymbol{X}_j\dot{\boldsymbol{D}}_j$, $i, j = 1, \dots, k$, of the studies use on the same variables. We next obtain the inner product $\boldsymbol{A}_j \mid \boldsymbol{A}_{j\prime} = tr(\boldsymbol{A}_j, \boldsymbol{A}^t_{j\prime})$, $i, j = 1, \dots, k$, or $\boldsymbol{B}_j \mid \boldsymbol{B}_{j\prime} = \mathrm{tr}(\boldsymbol{B}_j, \boldsymbol{B}^t_{j\prime})$, $i, j = 1, \dots, k$, where tr indicates trace. These are the matrices whose information is condensed whenever they have a leading eigenvalue.

Besides singular sets of studies, we can consider structured families whose series correspond to the treatments of a base design, see References 8,9.

We now present a general formulation of the condensation of information in a symmetric matrix with a dominant eigenvalue, both for singular matrices and structured families whose matrices correspond to the treatments of a base design. This opens a wide range of possible applications.The models will be derived from the spectral analysis of their mean matrices $\boldsymbol{\mu}$, and are given by

$$\boldsymbol{M} = \boldsymbol{\mu} + \overline{\boldsymbol{E}} = \sum_{j=1}^{k} \lambda_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^t + \overline{\boldsymbol{E}}, \tag{1}$$

with $(\lambda_j, \boldsymbol{\alpha}_j)$ are the pairs (eigenvalues, eigenvectors) of $\boldsymbol{\mu}$ and $\overline{\boldsymbol{E}}$ is a symmetric stochastic matrix.[10,11] To avoid considering a possible large number of small eigenvalues we assume that $\lambda_1 \geq \dots \geq \lambda_k$.

With $\boldsymbol{\beta}_j = \lambda_j \boldsymbol{\alpha}_j, j = 1, \dots, k$, the information contained in $\boldsymbol{M}$ can be condensed in a structure vector

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1^t \dots \boldsymbol{\beta}_k^t]^t,$$

and, a sum of square of residues

$$V = \|\boldsymbol{M}\|^2 - \|\boldsymbol{\beta}\|^2, \tag{2}$$

where $\| \|$ indicates the Euclidean norm both for matrices and vectors.

In the next section we consider these models. Next, in Section 3, we study structured families of symmetric stochastic matrices. The matrices in these families correspond to the treatments of a base design with fixed effects. We then study the action of the factors in the base design on the structure vectors of the matrices. When we may assume that the first eigenvalue of the matrices is dominant, we may lighten our treatment restricting it to the first structured vector of the matrices.[9,12]

2 | MODELS

For $\boldsymbol{M}$ a symmetrical $n \times n$ stochastic matrix, with mean matrix $\boldsymbol{\mu}$, we assume the model

$$\boldsymbol{M} = \boldsymbol{\mu} + \overline{\boldsymbol{E}}, \tag{3}$$

where $\overline{\boldsymbol{E}} = \frac{1}{2}(\boldsymbol{E} + \boldsymbol{E}^t)$ and $\boldsymbol{E}$ is a symmetric error matrix, with vec $(\boldsymbol{E})$ normal with null mean vector and covariance matrix $\sigma^2 \boldsymbol{I}_{n^2}$. Now, $\boldsymbol{\mu}$ will be symmetrical, so, if it has rank k it will have the pairs $(\lambda_i, \boldsymbol{\alpha}_i)$ of eigenvalues and eigenvectors, $i = 1, \dots, k$.

Let $(\theta_i, \boldsymbol{\gamma}_i)$, $i = 1, \dots, n$ be the pairs of eigenvalues and eigenvectors of matrix $\boldsymbol{M}$ ordered according to $\theta_1 \geq \dots \geq \theta_n$. Following Reference 6, we estimate $\boldsymbol{\beta}_i = \lambda_i^{1/2} \boldsymbol{\alpha}_i$ by $\widetilde{\boldsymbol{\beta}}_i = \theta_i^{1/2} \boldsymbol{\gamma}_i$, $i = 1, \dots, k$. With $\boldsymbol{m}_1 \dots \boldsymbol{m}_n$ the column vectors of matrix $\boldsymbol{M}$, we have

$$\widetilde{\boldsymbol{\beta}}_i = \boldsymbol{M}\boldsymbol{\gamma}_i = \boldsymbol{M}^t \boldsymbol{\gamma}_i = \begin{bmatrix} \boldsymbol{m}_1^t \boldsymbol{\gamma}_i \\ \vdots \\ \boldsymbol{m}_n^t \boldsymbol{\gamma}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1^t \boldsymbol{m}_1 \\ \vdots \\ \boldsymbol{\gamma}_i^t \boldsymbol{m}_n \end{bmatrix} \tag{4}$$

$$= (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t) \begin{bmatrix} \boldsymbol{m}_1 \\ \vdots \\ \boldsymbol{m}_n \end{bmatrix},$$

$$= (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\, \boldsymbol{Z}, i = 1, \dots, k, \tag{5}$$

with $\otimes$, indicating Kronecker matrix product, and $\boldsymbol{Z} = \text{vec}(\boldsymbol{M})$.

Then, with $\boldsymbol{\eta}$ and $\boldsymbol{\beta}_i$ the mean vectors of $\boldsymbol{Z}$ and $\widetilde{\boldsymbol{\beta}}_i$, $i = 1, \dots, k$ we have

$$\boldsymbol{\beta}_i = (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\, \boldsymbol{\eta}, \quad i = 1, \dots, k, \tag{6}$$

while the covariance matrix of $\widetilde{\boldsymbol{\beta}}_i$, $i = 1, \dots, k$, will be

$$\sum(\widetilde{\boldsymbol{\beta}}_i) = (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\, \boldsymbol{L}(\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i), i = 1, \dots, k, \tag{7}$$

with $\boldsymbol{L} = \boldsymbol{\Sigma}(\boldsymbol{Z})$.

Following Reference 11. the $\widetilde{\boldsymbol{\beta}}_1 \ldots \widetilde{\boldsymbol{\beta}}_k \left[\boldsymbol{\gamma}_1 \ldots \boldsymbol{\gamma}_k\right]$ if they are good estimators (for small k) of the $\boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_k[\boldsymbol{\alpha}_1 \ldots \boldsymbol{\alpha}_k]$, we will have

$$\overline{\boldsymbol{E}} = \boldsymbol{M} - \sum_{i=1}^{k} \boldsymbol{\beta}_i \boldsymbol{\alpha}_i^t \approx \boldsymbol{M} - \sum_{i=1}^{k} \widetilde{\boldsymbol{\beta}}_i \boldsymbol{\gamma}_i^t = \boldsymbol{M} - \sum_{i=1}^{k} (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\boldsymbol{Z}\boldsymbol{\gamma}_i^t, \tag{8}$$

as well as

$$\boldsymbol{R} = \text{vec}(\overline{\boldsymbol{E}}) \approx \boldsymbol{Z} - \text{vec}\left(\sum_{i=1}^{k} (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\boldsymbol{Z}\boldsymbol{\gamma}_i^t\right). \tag{9}$$

Since

$$vec\left(\sum_{i=1}^{k} (\boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t)\boldsymbol{Z}\boldsymbol{\gamma}_i^t\right) = \boldsymbol{W}\boldsymbol{Z}, \tag{10}$$

with

$$\boldsymbol{W} = \sum_{i=1}^{k} (\boldsymbol{\gamma}_i \otimes \boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t), \tag{11}$$

we get

$$\boldsymbol{R} = (\boldsymbol{I}_{n^2} - \boldsymbol{W})\boldsymbol{Z}, \tag{12}$$

so

$$\Sigma(\boldsymbol{R}) = \sigma^2(\boldsymbol{I}_{n^2} - \boldsymbol{W})\boldsymbol{L}(\boldsymbol{I}_{n^2} - \boldsymbol{W}^t), \tag{13}$$

while $\boldsymbol{R}$ can be considered as a residue vector, and

$$\widetilde{\boldsymbol{\beta}} = [\widetilde{\boldsymbol{\beta}}_1^t \ldots \widetilde{\boldsymbol{\beta}}_k^t]^t, \tag{14}$$

will be the adjusted global structure vector.[13,14]

Vectors $\boldsymbol{R}$ and $\widetilde{\boldsymbol{\beta}}$ play a key role in inference. Since these vectors are not independent, we have to obtain a homoscedastic residue vector $\boldsymbol{AR}$ independent from the vector $\widetilde{\boldsymbol{\beta}} = \boldsymbol{BZ}$, with

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_n \otimes \boldsymbol{\gamma}_i^t \\ \vdots \\ \boldsymbol{I}_n \otimes \boldsymbol{\gamma}_k^t \end{bmatrix}.$$

Applying the Gram–Schmidt orthogonalization to the column vectors of

$$\Sigma(\boldsymbol{R}; \widetilde{\boldsymbol{\beta}}) = (\boldsymbol{I}_{n^2} - \boldsymbol{W})\boldsymbol{L}\boldsymbol{B}^t. \tag{15}$$

After applying the Gram–Schmidt method to the column vectors of matrix $(\boldsymbol{I}_{n^2} - \boldsymbol{W})\boldsymbol{L}\boldsymbol{B}^t$, which has a characteristic r, one obtains r orthonormalized vectors $\boldsymbol{z}_1, \ldots \boldsymbol{z}_r$. Let $\boldsymbol{\delta}_i$, $i = 1, \ldots, n^2$, be the vectors with n^2 components of which $n^2 - 1$ are null and the ith component is one. Applying the Gram–Schmidt method to the vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_r, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_{n^2-a}$, one obtains not only $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_r$ but also the line vectors of matrix $\boldsymbol{A}: a_1, \ldots, a_{n^2-r}$.

Being the covariance matrix of $\boldsymbol{AR}$ given by

$$\Sigma(\boldsymbol{AR}) = \sigma^2\boldsymbol{A}(\boldsymbol{I}_{n^2} - W)\boldsymbol{L}(\boldsymbol{I}_{n^2} - \boldsymbol{W}^t)\boldsymbol{A}^t,$$

if $\Sigma(\boldsymbol{AR})$ has rank g, it will have positive eigenvalues $v_1, \ldots, v_g$, (since covariance matrices do not have negative eigenvalues), associated to eigenvectors $\xi_1, \ldots, \xi_g$. Thus, with

$$\boldsymbol{G} = \boldsymbol{D}\left(v_1^{-1/2}, \ldots, v_g^{-1/2}\right)[\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_g]^t,$$

where

$$\boldsymbol{D}\left(v_1^{-1/2}, \ldots, v_g^{-1/2}\right),$$

is the diagonal matrix with principal elements $v_1^{-1/2}, \ldots, v_g^{-1/2}$, we have $\Sigma(\boldsymbol{GAR}) = \sigma^2 \boldsymbol{I}_g$. So,

$$\dot{\boldsymbol{R}} = \boldsymbol{GAR} = (\dot{\boldsymbol{R}}_1, \ldots, \dot{\boldsymbol{R}}_g), \tag{16}$$

is an homoscedastic residues vector. So, when normality is assumed, $\dot{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{\beta}}$ will be independent and normal. We are led to test the hypothesis, see Reference 14

$$H_{ok} : \dot{\boldsymbol{R}}_1, \ldots, \dot{\boldsymbol{R}}_g \ i.i.d. \sim N(0, \sigma^2).$$

Now, when H_o holds, the statistic,

$$\widetilde{\mathcal{F}} = \frac{g\dot{\boldsymbol{R}}_0^2}{\sum_{j=1}^{g} \dot{\boldsymbol{R}}_j^2 - g\dot{\boldsymbol{R}}_0^2}, \tag{17}$$

where $\dot{\boldsymbol{R}}_0 = \frac{1}{g}\sum_{j=1}^{g} \dot{\boldsymbol{R}}_j$, will be the quotient of two independent central chi-squares with, 1 and $g-1$ degrees of freedom, and we will use this statistic to test H_o. If $\bar{f}_{q/2}$ and $\bar{f}_{1-q/2}$ are quantiles for probabilities $q/2$ and $1 - q/2$, for that quotient, we may use $\left[\bar{f}_{q/2};\bar{f}_{1-q/2}\right]$ as the q level acceptance region for the model. The correspondent rejection region will be $\left[0;\bar{f}_{q/2}\right] \cup [\bar{f}_{1-q/2};+\infty[$. When H_o does not hold, the numerator and denominator of $\mathcal{F}$ have non-centrality parameters δ_1 and δ_2 There will be alternatives in which δ_1 predominates over δ_2 (δ_2 predominates over δ_1) and in which $\mathcal{F}$ tends to take larger (smaller) values than when H_o holds. When the hypothesis is not rejected, we may estimate σ^2 by $\widetilde{\sigma}^2 = \frac{\|\dot{\boldsymbol{R}}\|^2}{g}$.

In practice we can adjust the model for increasing values of k till it is not rejected. We point out that, $\bar{f}_p = f_{1,\ g-1,\ p}$ with $f_{1,g-1,p}$ is the pth quantile for the central $\mathcal{F}$ distribution with 1 and $g-1$ degrees of freedom.

3 | STRUCTURED FAMILIES

In this work after considering isolate models we study structured families. A first example of such families is that of multiregression designs, see References 15,16. So, for each treatment of a base design we have a linear regression on the same variables.

The matrices of values of controlled variables and the variance of the error, are assumed to be the same for the different regressions, see Reference 15. The inference for this family of regressions is centered on the vectors of coefficients or, more generally, on estimable vectors, leading to interesting results, see.[15-21] These models, in a structured family, correspond to the treatments of a base design with fixed effects. The most interesting case is when the absence of effects and interactions for the factors in the base design are associated to the spaces of an orthogonal partition

$$R^d = \boxplus_{j=1}^{m} \varpi_j.$$

Let the g_j row vectors of matrices $\boldsymbol{A}_j$ constitute an orthonormal basis for $\varpi_j, j = 1, \ldots, m$, then, we have the sum of squares

$$S_j = \|\boldsymbol{A}_j\boldsymbol{Y}\|^2, j = 1, \ldots, m, \tag{18}$$

where $\boldsymbol{Y}$ is a vector whose components correspond to the treatments of the base design. For instance, if the models in structured family are for symmetrical stochastic matrices with dominant first eigenvalues, we may be interested in the action of the factors on the base design, of their first structure vectors for which we have the estimators

$$\widetilde{\boldsymbol{\beta}}_1(h) = (\widetilde{\boldsymbol{\beta}}_{1,1}(h) \ \dots \ \widetilde{\boldsymbol{\beta}}_{1,n}(h)), h = 1, \dots, d. \tag{19}$$

Then we carry out Anova-like analysis for the

$$\boldsymbol{Z}(l) = (\widetilde{\boldsymbol{\beta}}_{1,l}(1) \ \dots \ \widetilde{\boldsymbol{\beta}}_{1,l}(d)), l = 1, \dots, n, \tag{20}$$

this is for the vectors of homologue components of the estimated first structure vectors.[22,23] Then, we have a transversal analysis. Another possibility is to carry out a longitudinal analysis on the vectors of linear combinations

$$\mathbf{Z}(\boldsymbol{c}) = (\boldsymbol{c}^t\widetilde{\boldsymbol{\beta}}_1(1), \dots, \boldsymbol{c}^t\widetilde{\boldsymbol{\beta}}_1(d)), \tag{21}$$

thus, the components of $\mathbf{Z}(\boldsymbol{c})$ will be contrast on the components of the $\widetilde{\boldsymbol{\beta}}_1(1), \dots, \boldsymbol{c}^t\widetilde{\boldsymbol{\beta}}_1(d)$.

To avoid repetitions, we represent by $\dot{\boldsymbol{z}}$ the vector which we carry the analysis. We then have the sums of squares

$$S_j = \|\boldsymbol{A}_j\dot{\boldsymbol{z}}\|^2, j = 1, \dots, m, \tag{22}$$

we now point out that, see Reference 11, the hypothesis $H_{0,j}$ associated to $\varpi_j, j = 1, \dots, m$ may be written as

$$H_{0,j} : \boldsymbol{A_j}\boldsymbol{\eta} = 0_{g_j}, j = 1, \dots, m, \tag{23}$$

with $\boldsymbol{\eta}$ the mean vector of $\boldsymbol{Z}$. This hypothesis holds if and only if $\boldsymbol{\eta} \in \omega_j, j = 1, \dots, m$ with $\boldsymbol{\omega}_j$ the orthogonal complement of $\varpi_j \, j = 1, \dots, m$.

In general, we use the sum of the sum of squares of higher-order interactions to estimate the error. Let $\mathcal{D}$ be the set of indexes of these interactions then with

$$\begin{cases} S = \sum_{j \in \mathcal{D}} S_j \\ g = \sum_{j \in \mathcal{D}} g_j \end{cases} \tag{24}$$

we have the test statistics

$$\mathcal{F}_j = \frac{g}{g_j} \frac{S_j}{S}, j \notin \mathcal{D} \tag{25}$$

with g_j and g degrees of freedom see References 13,14.

When there are u factors, with $J_1 \dots J_u$ levels in the base design, there are $2^u \subset \bar{\bar{u}} = \{1, \dots, u\}$ of sets of factor indexes in a fixed effects model corresponds to the general mean value, and if $\#(\varphi) = 1 \ [>1]$, φ will correspond to the effects of levels, (interactions between levels) for the factor (factors), with index (indexes) in φ.

Then, to obtain the sum of squares for the effects and interactions we have, see Reference 11, the matrices

$$\boldsymbol{A}(\varphi) = \otimes_{l=1}^{u} \boldsymbol{A}_l(\varphi); \varphi \subseteq \bar{\bar{u}}, \tag{26}$$

where $\otimes$ indicates the Kronecker matrix product and

$$\boldsymbol{A}_l(\varphi) = \begin{cases} \frac{1}{\sqrt{J_l}} \mathbf{1}_{J_l}^t, l \notin \varphi \\ \boldsymbol{T}_{J_l} \ , l \in \varphi \end{cases}, l = 1, \dots, u; \varphi \subseteq \bar{\bar{u}}, \tag{27}$$

with $\boldsymbol{T}_{J_l}$ obtained deleting the first row equal to $\frac{1}{\sqrt{J_l}}\mathbf{1}_{J_l}^t$ from a orthogonal matrix when $l \in \varphi$. Then $\boldsymbol{A}(\varphi)$ has rank $J_l \times J_l$, with

$$g(\varphi) = \prod_{l\in\varphi}(J_l - 1); \varphi \subseteq \bar{\bar{u}}, \tag{28}$$

the degrees of freedom for the hypothesis associated to φ.

The order of a factor interaction is the number of factors taken in it minus one, so now we may take

$$D_h = \{\varphi, \#(\varphi) \geq h\}, \tag{29}$$

to obtain

$$\begin{cases} S = \sum\limits_{\#(\varphi)\geq h} S(\varphi) \\ g = \sum\limits_{\#(\varphi)\geq h} g(h) \end{cases} \tag{30}$$

with

$$S(\varphi) = \|\boldsymbol{A}(\varphi)\boldsymbol{Z}\|^2; \varphi \leq \bar{\bar{u}}, \tag{31}$$

we then have the statistics

$$\mathcal{F} = \frac{g}{g(\varphi)}\frac{S(\varphi)}{S}, \#(\varphi) < h, . \tag{32}$$

with $g(\varphi)$ and g degrees of freedom.

As an alternative we may consider the case in which we have estimators $\widetilde{\sigma}_h^2$ for $h = 1, \ldots, d$. This is, we did not reject the homoscedasticity of vector $\boldsymbol{R}_h$, $h = 1, \ldots, d$ for any matrix in the family. Then, we can carry the Bartlett Chi-square test for

$$H_{0,j} : \sigma_1^2 = \ldots = \sigma_d^2 = \sigma^2, j = 1, \ldots, m. \tag{33}$$

If this hypothesis is not rejected, we can use the estimator

$$\widetilde{\sigma}^2 = \frac{1}{d}\sum_{h=1}^{d}\widetilde{\sigma}_h^2, \tag{34}$$

with g_h the number of components of $\dot{\boldsymbol{R}}_h$, $h = 1, \ldots, d$ and

$$\bar{g} = \sum_{h=1}^{d} g_h. \tag{35}$$

We use the statistic.

$$\mathcal{F}_j = \frac{1}{g_j}\frac{S_j}{\widetilde{\sigma}^2}, j = 1, \ldots, m, . \tag{36}$$

for testing all the $H_{0,j}$, $j = 1, \ldots, m$. We now have g_j, $j = 1, \ldots, m$ and $\bar{g}$ degrees of freedom for $\mathcal{F}_j$, $j = 1, \ldots, m$.

4 | CONCLUSIONS

In this paper we present models for symmetric stochastic matrices, showing how to adjust and validate them in the general case, using the estimator

$$\widetilde{\boldsymbol{\beta}}_i = \theta_i^{1/2}\boldsymbol{\gamma}_i, i = 1, \ldots, k.$$

These models provided the basis for making inference for isolated matrices and structured families of matrices. We study the action of the factors in the base design on the structured vectors of the matrices, and we show that when the first eigenvalue of the matrices is dominant, we may lighten our treatment restricting it to the first structured vector of the matrices. Finally, we show how to carry out transversal and longitudinal analysis for families of stochastic symmetric matrices with dominant eigenvalue associated to orthogonal models.

The information condensation based on the pair of a dominant eigenvalue and the corresponding eigenvector, is now given without restricting it to series of studies.

ACKNOWLEDGEMENTS

This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

ORCID

Cristina Dias https://orcid.org/0000-0001-6350-5610
Carla Santos https://orcid.org/0000-0002-0077-1249

REFERENCES

1. Escoufier Y. Le traitement des variables vectorielles. *Biometrics*. 1973;29(5):751-760.
2. Escoufier YL, Hermier H. A propos de la comparaison graphique des matrices de variance. *Biom J*. 1978;20(5):477-483.
3. Lavit C. *Analyse Conjointe de Tableaux Quantitatifs*. Masson, Paris: Collection Méthods+ Programmes; 1988.
4. Lavit C, Escoufier Y, Sabatier R, Traissac P. The ACT (STATIS method). *Comput Stat Data Anal*. 1994;18:97-119.
5. Oliveira MM, Mexia JT. Tests for the rank of Hilbert-Schmidt product matrices. *Adv Data Sci Classif*. 1998;2:619-625.
6. Oliveira MM, Mexia JT. F tests for hypothesis on the structurevectors of series. *Discuss Math Biom Lett*. 1999;19(2):345-353.
7. Oliveira MM, Mexia JT. Multiple comparations for rank one common structures. *Biom Lett*. 1999;36(2):159-167.
8. Oliveira MM, Mexia JT. ANOVA like analysis of matched series of studies with a common structure. *J Stat Plan Infer*. 2007;137:1862-1870.
9. Oliveira MM, Mexia JT. Modelling series of studies with a common structure. *Comput Stat Data Anal*. 2007;51:5876-5885.
10. Areia A. *Séries Emparelhadas de Estudos* [Ph.D. thesis]. Évora: University of Évora; 2009.
11. Dias C. *Models and Families Of Models for Symmetric Stochastic Matrices* [Ph.D. thesis]. Évora: University of Évora;.2013.
12. Oliveira MM, Mexia J. F tests for hypothesis on the structure vectors of series. *Discuss Math*. 1999;19(2):345-353.
13. Mexia JT. Best linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in presence of controlled Heterocedasticity. *Comp Stat Data Anal*. 1990;10(3):271-281.
14. Mexia JT. *Introdução à Inferência Estatística Linear. Centro de Estudos de Matemática Aplicada*. Lisboa: Edições Universitárias Lusófonas; 1995.
15. Carvalho F, Mexia JT, Santos C, Nunes C. Inference for types and structured families of commutative orthogonal block structures. *Metrika*. 2015;78:337-372.
16. Moreira EE, Ribeiro AB, Mateus E, Mexia JT, Ottosen LM. Regressional modelling of electrodialytic removal of Cu, Cr and as from CCA timber waste: application to sawdust. *Wood Sci Technol*. 2005a;39(4):291-309.
17. Mexia, J.T. 1987. *Multi-treatment regression designs*. Trabalhos de Investigação, No 1. Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
18. Moreira EE, Ribeiro AB, Mateus E, Mexia JT, Ottosen LM. Regressional modelling of electrodia- lytic removal of Cu, Cr and as from CCA timber waste: application to wood chips. *Listy Biometryczne*. 2005b;42(1):11-23.
19. Moreira E, Mexia JT. Multiple regression models with cross nested orthogonal base model. Paper presented at: Proceedings of the 56th Session of the ISI 2007; 2007. International Statistical Institute, Lisboa.
20. Moreira E. *Família estruturada de modelos com base ortogonal: teoria e aplicações* [Ph.D. thesis], Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa (in Portuguese); 2008.
21. Cantarinha A. *Resultados Assintóticos para Famílias Estruturadas de Modelos Colectivos. Aplicação aos Fogos Florestais em Portugal Continental* [Ph.D. thesis]. Universidade de Évora; 2012.
22. Ito PK. *Robustness of anova and macanova test procedures*. In: Krishnaiah PR, ed. *Handbook of Statistics1*. North Holland: Amsterdam, Netherlands; 1980:199-236.
23. Scheffé H. *The Analysis of Variance*. New York, NY: John Wiley & Sons; 1959.

AUTHOR BIOGRAPHIES

Cristina Dias completed a degree in Applied Mathematics in University of Évora, Portugal, in 2000 concluded the Master of Applied Mathematics in University of Évora, Portugal, and in 2013 PhD in Applied Mathematics in University of Évora, Portugal. She attended Graduate Diploma in Data Analysis in Social Sciences in University Institute of Lisbon, Portugal in 2015. She began her professional career in 1996 as a teacher of secondary public education and entered university education in the year 2000, where she teaches linear algebra and analytic geometry; Statistic; Probability and statistics; Quantitative methods; Mathematics; Mathematics Applied to Social Sciences; Operational Research, to Engineering and Business Sciences courses. Her current research work is in the areas of Multivariate Data Analysis, Generalized Linear Models, Mixed Linear Models, Computational Statistics, Life Sciences and Environment Statistics, Statistical Modeling in Educational Sciences, Linear Analysis and Linear Inference with five main lines as Member of the Portuguese Society of Statistics, Portuguese Society of Mathematics, Member of the Portuguese Operational Research Association.

Carla Santos is an Adjunct Professor at the Department of Mathematics and Physical Sciences of the Polytechnic Institute of Beja and member of Center of Mathematics and Applications, FCT- New University of Lisbon. She is PhD in Mathematics from UBI, MSc. in Applied Mathematics from the University of Évora, Portugal (UE), and BSc. in Mathematics from UE. Her main research interests are in Applied Mathematics and Probability and Statistics, more specifically Statistical Inference in Linear Models. She has published research papers in these fields and also in Mathematical Education.

João Tiago Praça Nunes Mexia was born in Lisbon in June of 1939 from a Catholic family including his parents and his two sisters. His first degree was in forestry, from Universidade Técnica de Lisboa (Lisbon University of Technology). He concentrated a great part of his activities in 19 PhD supervisions and 12 PhD co-supervisions. His research was centered on Linear Statistical Inference, with 63 papers published in international journals. As pertaining remarks, we point out the support he received in the first years at FCT/UNL, from the then Head of the Mathematics Department, Professor António César de Freitas. Moreover, his colleagues and students have been a constant encouragement in his work. He has more than 10 co-authors. In the last 10 years he has actively cooperated with Polish statisticians, namely Professors Tadeusz Calinski, Stanislaw Mejza, and Roman Zmyślony. As a high point in his academic career he was the "promoter" of the Honoris Causa Degree awarded by Universidade Nova de Lisboa to Professor C. R. Rao in 2006. *Biography of Professor João Tiago Praça Nunes Mexia*. https://www.researchgate.net/publication/267177055_Biography_of_Professor_Joao_Tiago_Praca_Nunes_Mexia (accessed June 13, 2020).

How to cite this article: Dias C, Santos C, Mexia JT. Isolated and structured families of models for stochastic symmetric matrices. *Comp and Math Methods*. 2021;e1152. https://doi.org/10.1002/cmm4.1152