Journal of Statistics on Health Decision

EXTENDED ABSTRACT

A20 CO-Tucker method for analysis of urine and saliva interactions during pregnancy

Francisco Silva[1], Daniela Duarte[2], Adelaide Freitas[1,3], Joana L. Pereira[4], Tânia Ascensão[5], Mariana Osório[5], Cristina Pita[5], Maria do Céu Almeida[5], Fátima Negrão[5], Ana Luísa Costa[4], Ana Maria Gil[2]

[1]Department of Mathematics, University of Aveiro, Aveiro, Portugal
[2]CICECO - Aveiro Institute of Materials, Chemistry Department, University of Aveiro, 3810-193 Aveiro, Portugal
[3]Center for Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, Aveiro, Portugal
[4]Dentistry Department, Faculty of Medicine, Institute of Paediatric and Preventive Dentistry, University of Coimbra, 3000-075 Coimbra, Portugal
[5]Maternidade Bissaya Barreto, Centro Hospitalar e Universitário de Coimbra, 3000-045 Coimbra, Portugal

Introduction

Pregnancy is a dynamic period in which women undergo metabolic adaptations indispensable to ensure adequate growth and fetal development [1]. In the present study, Nuclear Magnetic Resonance (NMR) spectroscopy was used to analyse urine and saliva from a cohort of healthy pregnancy evolution. Due to the amount of data generated by NMR, multivariate analysis is of outmost importance for data mining [2]. The resulting proton NMR spectra of urine and saliva samples were handled in order to differentiate each pregnancy trimester (from first to third trimester) through multivariate analysis, based on previous results obtained for urine [3]. Concomitantly, univariate analysis enabled the identification of four salivary and twenty-four urinary metabolites descriptive of pregnancy progression in this cohort of pregnant women with normal pregnancy evolution. The interpretation of the metabolic changes observed is one of the major challenges in metabolomics, and detecting intra- and inter- biofluid relationships between metabolites may be of valuable aid. Hence, two sets of longitudinal metabolic data matrices, measured for different variable sets (urine and saliva), were obtained. When there are several variables measured for a collection of subjects under various conditions (time or space), the structure of the data can be depicted by a "cube", and then, the so-called three-way data analysis is usually suggested. One of the most common methods for modelling three-way data sets is the three-mode model called Tucker3 [4], where the data is represented by a linear combination of orthogonal components, determined by an appropriate generalization of the principal component analysis (PCA) of multivariate data to multiway data arrays.

The main aim of this work was to study three pairs of matrices with urinary and salivary metabolites in order to probe for possible interactions between metabolites, which may confirm/aid biochemical interpretation. Each pair was associated to one of the pregnancy trimesters. Therefore, a simultaneous analysis of a sequence of paired matrices was performed, applying the recently proposed Co-Tucker method [5].

Methods

Urine and saliva samples were collected at the Maternity Bissaya Barreto, University Hospital Centre of Coimbra, for pregnant women. Sample collection procedures, preparation and NMR spectral acquisition conditions have been described elsewhere [3,6]. Unidimensional 1H NMR spectra were acquired for each sample and, for selected samples, 2D spectra were acquired to aid peak assignment. In this way, a sequence of three paired metabolic data matrices related to the seven pregnant women, measured for three timepoints, were obtained: three (7x4) data matrices of salivary metabolites for the first, second, and third pregnancy trimesters, which are paired with three (7x24) data matrices of urinary metabolites measured for the same three trimesters. Thus, a paired of three data arrays was constructed, with (7x4x3) and (7x24x3) dimensions, respectively.

All statistical analyses were performed using R software (version v.3.6.3 for Windows) with RStudio (version v.1.2.1335).

Co-inertia Analysis

Co-inertia analysis [7] is responsible for describing the co-structure/concordance between two matrices, $\boldsymbol{X}$ and $\boldsymbol{Y}$, which is provided by the cross product matrix $\boldsymbol{Y}^T\boldsymbol{D}_N\boldsymbol{X}$, where $\boldsymbol{D}_N$ denotes a diagonal matrix, $diag(1/N, 1/N, \ldots, 1/N)$ and N is the number of objects. It is worth to mentioning that if the columns of both the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are centered, then $\boldsymbol{Y}^T\boldsymbol{D}_N\boldsymbol{X}$ corresponds to the covariance matrix between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Keywords:
hree-way data analysis. Principal component analysis. Biplots. Tucker3. Co-Inertia analysis. Co-Tucker model

Corresponding author:
Francisco Silva
francisco.silva1@ua.pt

Supplementary material:
xxxx

Support statement:
xxx

Conflict of interest:
The authors declare no conflict of interest

Clinical study registration number: XXXXXXXXX

First published: 22JUN2021

In Figure 1, the construction of the cross product matrix $\boldsymbol{Y}^T\boldsymbol{D}_N\boldsymbol{X}$ for several pairs of matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, is illustrated.

Tucker3 Model

The Tucker3 model [4] is defined by a decomposition of the $(I \times J \times K)$ three-way data array $\mathbb{X} = (x_{ijk})$. In this decomposition, each element of $\mathbb{X}$ is given by the following form:

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk}, \qquad i = 1,\dots I;\ j = 1,\dots J;\ k = 1,\dots K$$

where $\boldsymbol{A} = [a_{ip}]$ is a $(I \times P)$ matrix with the coefficients for the description of the I subjects (first mode of $\mathbb{X}$) in a P-dimensional space, $\boldsymbol{B} = [b_{jq}]$ is a $(J \times Q)$ matrix with the coefficients for the description of the J variables (second mode of $\mathbb{X}$) in a Q dimensional space, and $\boldsymbol{C} = [c_{kr}]$ is a $(K \times R)$ matrix with the coefficients for the description of the K conditions (third mode of $\mathbb{X}$) in a reduced space of dimension R.

All these matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are column-wise orthonormal matrices. The element g_{prq} belongs to the $(P \times Q \times R)$ three-way core array $\mathbb{G}$ and describes the interaction among the p-th, q-th and r-th levels of the three correspondent dimensions of the array $\mathbb{X}$. The formula of computation of the array $\mathbb{G}$ can be found in [5]. The elements e_{ijk} of three-way array $\mathbb{E}$ are the approximation errors of the model of $\mathbb{X}$. In Tucker3 model, it is assumed that the first mode corresponds to the subjects, second mode to the variables, and the third mode to the conditions. The aim of this method is to summarize the principal information of the three-way data set $\mathbb{X}$ in a lower dimensional space of variables with a limited number of components in each of the three sets of entities of $\mathbb{X}$.

Co-Tucker model

The Co-Tucker method was reported before [5] and it consists of a combination of the Co-inertia analysis and the Tucker3 model. In particular, the Tucker3 model is performed on the sequences of cross product matrices which are computed using Co-inertia analysis [7]. Firstly, applying Co-inertia analysis, a sequence of K cross-covariance $(I \times J)$ matrices, given by $\boldsymbol{W_k} = \boldsymbol{Y_k^T D_N X_k}$, for $k = 1,2,..K$, are obtained (Figure 1). The new K cross-matrices form the $(I \times J \times K)$ array $\mathbb{W}$ (Figure 1). Then, the Tucker3 model is applied on $\mathbb{W}$ getting the following decomposition of each element w_{ijk} of this new array $\mathbb{W}$ (Figure 2):

The first, second and third modes are associated with the I variables of $\mathbb{X}$, the J variables of $\mathbb{Y}$ and the K

$$w_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk}, \qquad i = 1,\dots I;\ j = 1,\dots J;\ k = 1,\dots K$$

condictions simultaneous of $\mathbb{X}$ and $\mathbb{Y}$, respectively. The complexity $P \times Q \times R$ of the Co-Tucker model for the approximation $\hat{\mathbb{W}} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr}$ depends on the number of components that are chosen. The core array $\mathbb{G}$ of the correspondent Tucker3 model indicates the importance of each combination of components and allows the calculation of the proportion of variability explained $(= g_{pqr}^2 / \sum_{pqr}(g_{pqr}^2))$. After adjusting the Co-Tucker method, the main goal is the construction of the joint biplot where the projection of the components of the first and second modes in the r-th component of the third mode are represented. Thus, for each $r = 1, \dots, R$, the $(P \times Q)$ core slice G_r of $\mathbb{G}$ is decomposed via Singular Value Decomposition into $\boldsymbol{G_r} = \boldsymbol{U_r \Lambda_r V_r}$. In this manner, the coordinates of the I levels of $\mathbb{X}$, $\boldsymbol{A^*}_r = {}^{I}/_{J}^{\frac{1}{4}}\boldsymbol{A U}_r\boldsymbol{\Lambda}_r^{\frac{1}{2}}$ represented by arrows in the biplot, and the coordinates of the J levels of $\mathbb{Y}$, $\boldsymbol{B^*}_r = {}^{J}/_{I}^{\frac{1}{4}}\boldsymbol{B V}_r\boldsymbol{\Lambda}_r^{\frac{1}{2}}$ represented by points in the biplot, are projected in the same R-dimensional space. In terms of projections, directions, and proximities, the elements (points and arrows) of any joint biplots are interpreted in the similar way as the classical biplot [8]. Consequently, if the i-th level of first mode is close to the j-th level of the second mode, then there is positive (negative, resp.) interaction in the k-th positive (negative, resp.) level of the third mode (matrix $\boldsymbol{C}$). Analogously, the i-th level of first mode is far from the j-th level of the second mode, the interpretation occur in inverse way.

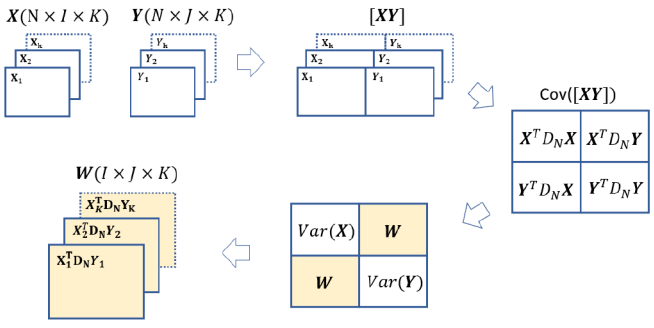

Figure 1 - Co-Tucker model - step 1: K Co-Inertia of K-cross product tables

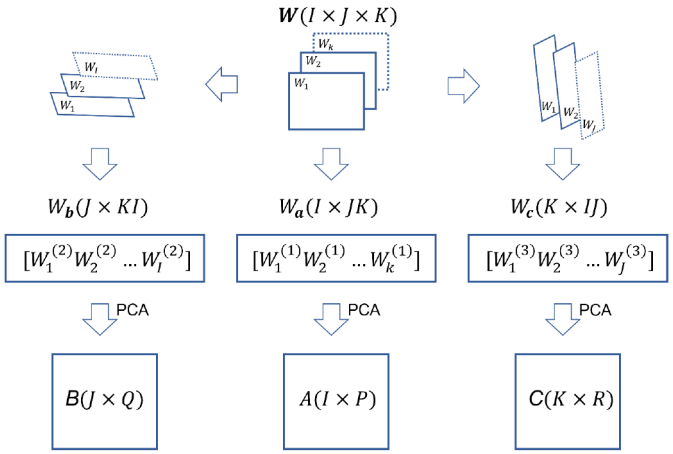

Figure 2 – Co-Tucker model – step 2: Decomposition of the new K-table (𝕎) by PCA.

Results

The decomposition model with complexity 2 × 2 × 2 (i.e., $P = Q = R = 2$) was chosen for the (4x24x3) array 𝕎 obtained from the Co-inertia analysis applied on the (7x4x3) and (7x24x3) paired data arrays. Consequently, a model with two components for the saliva mode, P_1 and P_2, two components for the urine mode, Q_1 and Q_2, and two components for the trimester mode, R_1 and R_2 was defined. This model explains 83.6% of the total variance of the data. The components of salivary and urinary modes projected onto the first component R_1, explaining 58.3% of the variance, are visually depicted into a joint biplot (Figure 3). In this biplot, the contrast between the first and third trimester with the second is highlighted (since both the first and third conditions have positive signs and the second condition has negative sign in R_1, which is represented into the matrix $\boldsymbol{C}$). Furthermore, this biplot also shows positive interactions between the N-acetyl group of glycoproteins (NAG in Figure 3) and the unassigned resonance at 2.94 ppm (U2 in Figure 3), alanine and glucose (followed by an unassigned resonance at 2.95 ppm (U3) and choline) in the first and third trimesters (positive values). As a result, when increasing the values of salivary NAG, the values of urinary U2, alanine and glucose increase in the first and third trimesters. Naturally, opposite interactions occur in the second trimester. NAG displays a negative interaction with N-acetyl neuraminic acid (N5AC in Figure 3) (followed by creatinine) in both the first and third trimesters, and a positive interaction in the second trimester.

A second interaction occurs between urea and guanidoacetate (GAA) (followed by threonine) and 2-ketoglutarate (2-KG) (represented by X2-KG in Figure 3). This interaction is positive with GAA in the first and third trimesters, and negative in the second trimester. The interaction of urea with 2-KG is positive in the second trimester, and negative in the first and third trimesters.

NAG and urea reflect saliva-urine interaction variability associated with R_1 (larger arrows in the biplot). The remainder urinary methabolites that are located near of the centre of the biplot exhibit low interactions with salivary metabolites.

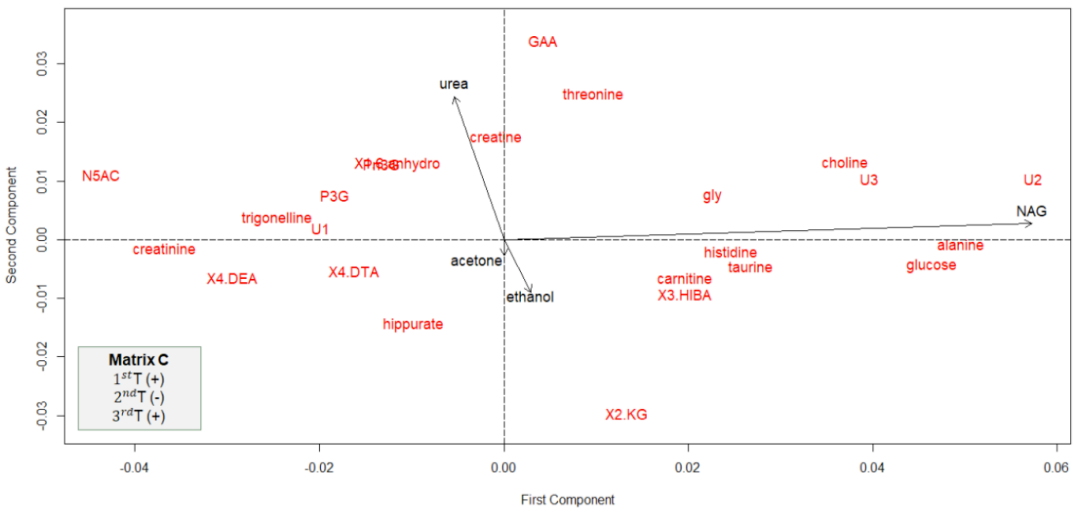

Figure 3 - Joint biplot: First vs second component of salivary mode and first vs second component of urinary mode (58.3% of the total variance).

Discussion

The positive correlation between the salivary N-acetyl group of glycoproteins (NAG) and urinary glucose in the first and third trimesters of pregnancy could be explained since glucose could be one of the constituents of glycoproteins. Moreover, alanine is a gluconeogenic amino acid (meaning that it is precursor of glucose biosynthesis), which would be consistent with the positive interaction observed here between glucose and alanine. However, interpretation of the positive choline interaction with salivary NAG could not be advanced at this stage and needs further analysis. In addition, the positive correlation between urinary N5Ac (the most common sialic acid found in urine), and salivary NAG could be explained with basis on the importance of the former as a main component of glycoproteins. Furthermore, a positive creatinine interaction with salivary NAG was also found. Creatinine is formed from creatine, which is obtained from the diet, or from the conversion of GAA, however, at this stage no explanation for the interaction with NAG could be advanced.

The positive interaction found between salivary urea and GAA could be explained by an urea cycle deregulation throughout pregnancy. In addition, the interaction of 2ketogluratate (2-KG), an intermediate of the tricarboxylic acid (TCA) cycle (an important pathway in energy metabolism), with salivary urea could not be explained at this stage.

The new correlations unveiled in this study, including those still requiring putative interpretation, will trigger further and novel biochemical assessment of pregnancy metabolism.

Ethics committee and informed consent

The current research was approved by the Ethics Committee from the University Hospital Centre of Coimbra (CHUC-091-17, dated 25 June 2018), and subjects gave their informed consent before they enrolled in the study.

Acknowledgements

This work was developed within the scope of the project CICECO-Aveiro Institute of Materials (UIDB/50011/2020 and UIDP/50011/2020) and the Center for Research and Development in Mathematics and Applications (CIDMA, University of Aveiro: reference UIDB/04106/2020), financed by national funds through the Portuguese Foundation for Science and Technology (FCT)/MEC. We also acknowledge the Portuguese National NMR Network (RNRMN), supported by FCT funds and also FCT for the SFRH/BD/119509/2016 grant (D.D.).

References

1. Hadden DR, McLaughlin C. Normal and abnormal maternal metabolism during pregnancy. Seminars in Fetal & Neonatal Medicine. 2009;14(2):66–71. https://doi.org/1 0.1016/j.siny.2008.09.004
2. Trygg J, Holmes E, Lundstedt T. Chemometrics in Metabonomics. Journal of Proteome Research. 2007;6: 469 – 479. https://doi.org/10.1021/pr060594q
3. Diaz SO, Barros AS, Goodfellow BJ, Duarte IF, Carreira IM, Galhano E, Pita C, Almeida MC, Gil AM. Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine. Journal of Proteome Research. 2013;12(2):969–979. https://doi.org/10.1021/pr301022e
4. Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966 Sep;31(3):279–311. https://doi.org/10.1007/BF02289464
5. Mendes S, Fernández-Gómez MJ, Marques SC, Pardal MÂ, Azeiteiro UM, GalindoVillardón MP. CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables. Journal of Applied Statistics. 2017;44(15):2729–2755. https://doi.org/10.1080/02664763.2016.1261815
6. Duarte D, Castro B, Pereira JL, Marques JF, Costa AL, Gil AM. Evaluation of Saliva Stability for NMR Metabolomics: Collection and Handling Protocols. Metabolites. 2020;10(12):515. https://doi.org/10.3390/metabo10120515
7. Dolédec S, Chessel D. Co-inertia analysis: an alternative method for studying species– environment relationships. Freshwater Biology, 1994;31(3):277–294. https://doi.org/10.1111/j.1365-2427.1994.tb01741.x
8. Gabriel KR. The biplot-graphic display of matrices with application to principal component analysis. Biometrika. 1971;58:453–467. https://doi.org/10.2307/2334381