Journal of Statistics on Health Decision

EXTENDED ABSTRACT

Evolution of diet patterns over time in European countries from 1963 to 2013: an exploratory analysis using PCA for compositional data vectors.

Mariana Pinto[1], Adelaide Freitas[1,2], Marco Costa[2,3]

[1]Department of Mathematics, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[2]Center for Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[3]Águeda School of Technology and Management (ESTGA), University of Aveiro, 3810-193 Aveiro, Portugal

Introduction:

The study of the evolution of the diet of populations, namely in Europe, is crucial for decision-making by public authorities to promote healthy nutrition and prevent diet related noncommunicable diseases [1]. In fact, for instance, European Commission food law impacts the transformative potential of alternative proteins [2]. Composition food consumption in terms of each macronutrient, i.e., protein (animal and vegetal), fat and carbohydrate, has been usually investigated separately by nutrient [3]. However, these macronutrients constitute the whole diet of each individual and so, add up to a constant sum (in general 100% or 1 if the constitution of macronutrients in a diet is evaluated in percentage or proportion, respectively). Hence, it seems crucial that any analysis of diet compositions must take the constrained or compositional nature of the data. A set of several procedures has been developed in Statistics and applied in many fields to handle with this type of data structure, i.e., with compositional data [4,5].

Compositional data are multivariate observations that represent quantitative descriptions of the parts of some whole, conveying only relative information between parts. In mathematical notation, a p-multivariate observation corresponds to a sample unit described by a vector with p components. When all the components of this vector are positive numbers and contain relative information of parts of a whole (e.g, proportions, percentages), the nature of multivariate observation is compositional. This means that the important information between components is given by ratio rather difference. For instance, while the difference between 0.05 and 0.10 and between 0.45 and 0.50 are the same, in a compositional perspective, the number 0.05 should be considered as half of 0.10 while 0.45 forming a fraction 0.9 of 0.50. In mathematical notation, a D-part compositional observation is expressed as $\mathbf{x} = [x_1 x_2 \cdots x_D]$ with $x_j > 0$, for $j = 1, 2, ..., D$, and subject to the so-called unit-sum constraint, $\Sigma_{j=1}^{D} x_j = 1$.

In this paper, composition food consumption by both country and decade will be analyzed. In this context, a $(p \times D)$-multivariate observation is described by p variables each one with D-part compositional components. This type of multivariate observation is referred to as a (p-dimensional) compositional data vector as, for instance, the quantity of four macronutrients (p=4: animal protein, vegetal protein, fat, carbohydrate) across six years (D=6: 1963, 1973, 1983, 1993, 2003, 2013) by country). In [6], the operations in the space of the p-dimensional compositional data vectors were deduced: $\bigoplus$ for the "addition" (called perturbation) and $\bigotimes$ for the multiplication by a scalar (called powering) (for details see Equation 14 and 15 in [7]).

Recently, principal component analysis (PCA) [8] for modeling compositional data vectors was developed in [7]. In this study, this statistical procedure is explored to reduce the dimensionality of the evolution of diet patterns of ten countries in Europe during five decades. The main focus is on a temporal composition of the consumption in the diet of the four macronutrients.

Methods:

Daily caloric supply derived from each of the four macronutrients by country has been produced by Our World in Data and are freely accessible in [9]. For this study, data related to ten European countries (see Table 2) 10-by-10 years from 1963 to 2013 were collected. Hence, the data set includes ten observations featured by four 6-dimensional compositional variables: animal protein (**U1**), vegetal protein (**U2**), fat (**U3**), carbohydrate (**U4**). Given the structure of the data and following [7], a set of instructions in R were constructed to perform the PCA method on the correlation matrix for compositional data vectors (for more details see [10]).

Keywords:
Compositional data vectors, Diet pattern, Macronutrient shares, Principal component analysis.

Corresponding author:
Adelaide Freitas
adelaide@ua.pt

Conflict of interest:
The authors declare no conflict of interest.

First published: 20JUL2022

Results:

The unit-loading and the variance of each principal component (PC) are displayed in Table 1 and the compositional scores of PC1 for the ten observations are shown in Table 2. PC1 accounts for 60.9% of the total variance and

$$PC1 = 0.475 \bigotimes \mathbf{U1} \bigoplus (-0.526) \bigotimes \mathbf{U2} \bigoplus 0.545 \bigotimes \mathbf{U3} \bigoplus (-0.448) \bigotimes \mathbf{U4}.$$

Table 1 - Coefficients of the PCs and the percentage of variance explained by each PC.

Components	PC1	PC2	PC3	PC4
Animal protein	0.475	-0.562	0.285	0.614
Vegetal protein	-0.526	-0.421	-0.661	0.330
Fat	0.545	-0.400	-0.464	-0.572
Carbohydrate	-0.448	-0.589	0.515	-0.432
Variance (%)	60.9	31.7	5.2	2.2

Table 2 - Score of the first PC.

Country	PC1 scores (%)
Finland	[24.0 24.2 21.0 13.6 8.8 8.3]
France	[9.0 18.8 16.7 19.5 22.1 13.9]
Germany	[36.7 19.8 12.4 13.6 9.6 7.9]
Greece	[11.1 12.6 11.9 14.7 18.4 31.4]
Italy	[5.9 6.3 18.5 23.9 21.2 24.2]
Norway	[28.5 36.3 13.2 8.1 7.3 6.5]
Portugal	[3.5 4.7 8.5 13.9 27.3 42.1]
Spain	[2.8 4.3 10.3 18.4 36.2 28.1]
Sweden	[40.1 19.8 14.3 8.7 9.1 8.0]
United Kingdom	[40.8 23.3 15.0 10.4 5.0 5.6]

Then, considering the coefficient signals, PC1 represents a balance of consumption of animal protein and fat comparatively with vegetal protein and carbohydrate such that higher the score of PC1, the greater the fraction consumption of animal protein and fat compared with those of vegetal protein and carbohydrate.

It is clearly observed in Table 2 that there existed a higher percentage of consumption of animal protein and fat in the first decade in Finland, Germany, Norway, Sweden and United Kingdom, while this percentage is higher, in general, for the other countries in the last two decades.

Discussion:

PCA can be applied on data sets that can be arranged into compositional vector. In terms of macronutrient diet pattern, PCA revealed different shares of animal protein and fat between countries further Mediterranean region and further north in Europe from 1963 to 2013.

Acknowledgements:

This work was developed within the scope of the Center for Research and Development in Mathematics and Applications (CIDMA, University of Aveiro: references UIDB/04106/2020 and UIDP/04106/2020), financed by national funds through the Portuguese Foundation for Science and Technology (FCT)/MEC.

References:

1. Breda J, Castro SLN, Whiting S, Williams J, Jewell J, Engesveen K, Wickramasingh K. Towards better nutrition in Europe: Evaluating progress and defining future directions, Food Policy, 2020, 96, 101887. https://doi.org/10.1016/j.foodpol.2020.101887
2. Lähteenmäki-Uutela A, Rahikainen M, Lonkila A, Yang B. Alternative proteins and EU food law. Food Control, 2021; 130, 108336. https://doi.org/10.1016/j.foodcont.2021.108336
3. Trinh HT, Morais J, Thomas-Agnan C, Simioni M. Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: New insights using compositional data analysis, Statistical Methods in Medical Research. 2019;28(8):2305–2325. https://doi.org/10.1177/0962280218770223
4. Aitchison J. The Statistical Analysis of Compositional Data, London: Chapman & Hall, 1986
5. Filzmoser T, Hron K, Matthias T. Applied compositional data analysis with worked examples in R, Springer, 2018. https://doi.org/10.1007/978-3-319-96422-5
6. Gallo M. Coda in three-way arrays and relative sample space. Electron J Appl Stat Anal, 2012;5(3):400–405.
7. Wang H, Shangguan L, Guan, R, Billard L. Principal component analysis for compositional data vectors, Comput Stat;2015;30(4):1079-1096. https://doi.org/10.1007/s00180-015-0570-1
8. Jolliffe IT. Principal component analysis, 2nd ed. Springer, New York. 2002.
9. Ritchie H, Roser M. Diet Compositions. Published online at OurWorldInData.org. 2017. Retrieved from: https://ourworldindata.org/diet-compositions
10. Gamelas C. Análise de Componentes Principais em vetores de dados composicionais. Dissertação de mestrado, Universidade de Aveiro, 2020.