DOI: 10.1590/1980-549720200063

BRIEF COMMUNICATION / *COMUNICAÇÃO BREVE*

Multivariate estimate of eating patterns: is the whole different from the parts?

Estimativa multivariada de padrões alimentares: o todo é diferente da reunião das partes?

Iolanda Karla Santana dos Santos[I,II], Wolney Lisbôa Conde[I], Alicia Matijasevich Manitto[III]

ABSTRACT: ***Objective:*** To describe the correlations between eating patterns for the years 2007 to 2012, and for each year of the period from 2007 to 2012. ***Method:*** Cross-sectional study with data from the System of Surveillance of Risk and Protection Factors to Chronic Diseases by Telephone Survey with the selection of 167,761 individuals aged 18 to 44 years old. Eating patterns were identified with a Principal Component Analysis. To compare the effects of the extraction and the estimate of eating patterns among different surveys we conducted the following analyzes: in the first, we used the total data set for the years from 2007 to 2012; in the second, the patterns were estimated in each annual set of data for the period from 2007 to 2012. Steps 1 and 2 were performed with no rotation, with Varimax rotation and with Promax rotation. After extracting the patterns, standardized scores with zero mean were generated for each pattern. The association between the patterns generated in the analyzes was estimated by the Pearson correlation coefficient (*r*). ***Results:*** In the non-rotated analyzes, the components retained in the set presented correlations that were higher than 0.90, with the retained patterns in each year. In the rotated analyzes, only the first component had correlations that were higher than 0.90. ***Conclusion:*** Estimates of eating patterns either segmented — year by year — or in general — all of the years — showed high correlation and consistency between the patterns identified when in the same data pool.

Keywords: Epidemiology. Surveillance. Eating behavior.

[I]School of Public Health, Universidade de São Paulo – São Paulo (SP), Brazil.
[II]Fundação Universidade Federal do ABC – Santo André (SP), Brazil.
[III]School of Medicine, Universidade de São Paulo – São Paulo (SP), Brazil.
Corresponding author: Iolanda Karla Santana dos Santos. Faculdade de Saúde Pública, Universidade de São Paulo. Av. Dr. Arnaldo, 715, Cerqueira César, CEP: 01246-904, São Paulo, SP, Brazil. E-mail: iolanda.santos@usp.br
Conflict of interests: nothing to declare – **Financial support:** none.

RESUMO: ***Objetivo:*** Descrever as correlações entre padrões alimentares para o conjunto de anos de 2007 a 2012 e para cada ano do mesmo período. ***Método:*** Estudo transversal com dados do Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, com seleção de 167.761 indivíduos de 18 a 44 anos. Os padrões alimentares foram identificados com Análise de Componentes Principais. Para comparar os efeitos da extração e a estimativa de padrões alimentares entre diferentes inquéritos, conduzimos as seguintes análises: na primeira usamos o conjunto total de dados para os anos de 2007 a 2012; na segunda, os padrões foram estimados em cada conjunto anual de dados para o período de 2007 a 2012. As etapas 1 e 2 foram realizadas sem rotação, com rotação *Varimax* e rotação *Promax*. Após a extração dos padrões, foram calculados escores padronizados com média zero. A associação entre os padrões gerados nas análises foi estimada pelo coeficiente de correlação de Pearson (*r*). ***Resultados:*** Nas análises sem rotação, os componentes retidos no conjunto apresentaram correlações superiores a 0,90 com os padrões retidos em cada ano. Nas análises com rotação, apenas o primeiro componente apresentou correlações superiores a 0,90. ***Conclusão:*** As estimativas de padrões alimentares de forma segmentada — ano a ano — ou de forma geral — todos os anos — apresentam altas correlação e consistência entre os padrões identificados quando no mesmo *pool* de dados.

Palavras-chave: Epidemiologia. Vigilância. Comportamento alimentar.

INTRODUCTION

An analysis of dietary patterns is preferable to describing diets by type of food or nutrient, because food consumption is determined by multiple factors and food choices and their nutrients does not occur randomly[1]. Considering that food consumption is not random and there is a correlation between food and nutrients, the study of diets using patterns has been widespread[1].

In general, analyzes that compare dietary patterns, estimated through multivariate analysis, between two or more surveys, are conducted in each period separately, which makes it difficult to compare dietary patterns. This is because the composition and the order of importance in the explanation of the variability are modified according to how the data set is treated. Alternatively, it is possible to estimate the patterns in the total set of surveys and then calculate the scores for each pattern according to the periods, or other strata in the data set.

The aim of this study was to describe the correlations between dietary patterns for the set of years from 2007 to 2012 and for each year in the same period.

METHODS

This was a cross-sectional study with data from the Surveillance System for Risk and Protection Factors for Chronic Diseases by Telephone Survey (*Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico* - Vigitel) from

2007 to 2012. Aspects related to Vigitel's research methodology are available in official publications[2].

In this study, 167,761 individuals aged 18 to 44 years old were selected. The food consumption variables selected were: weekly frequency of consumption of beans, vegetables, raw vegetables, cooked vegetables, red meat, chicken, fruits, soft drinks or artificial juice, milk, daily vegetable consumption and consumption of visible fat.

Dietary patterns were identified with the Principal Component Analysis (PCA). PCA is a factor analysis that reduces data into patterns based on the correlations between the variables[3]. The first main component corresponds to the direction of greatest variance, and the other components are orthogonal to the previous components[4]. Rotations are used in order to improve the interpretation of the extracted components. Varimax rotation of the orthogonal matrix maximizes the variation between the factorial loads, and the components remain not correlated. Promax oblique matrix rotation rotates the axes so that the vertices can have angles other than 90 degrees. In this type of rotation, the probability of some association between the components cannot be ruled out[5].

To compare the effects of extraction and the estimation of dietary patterns between different surveys, we conducted the following analyzes:

- in the first, we used Vigitel's total data set for the years 2007 to 2012;
- in the second, the patterns were estimated in each Vigitel annual data set for the period from 2007 to 2012.

Steps 1 and 2 described above were performed with no rotation, with Varimax rotation and with Promax rotation. In the analysis, the components with eigenvalues> 1.0 were retained, according to the Kaiser rule[5]. We considered the number of patterns retained in the first stage. After extracting the patterns, standardized scores were calculated with an average of zero for each one, so that each individual received a standardized value that represented their adherence to each of the patterns analyzed. The patterns were named according to their order of retention, that is, the first pattern was named CP1, the second CP2, and so on. The association between the patterns generated in the analyzes described above was estimated by Pearson's correlation coefficient (r). The analyzes were conducted using the Stata program (Stata Corporation, College Station, United States).

Vigitel was approved by the National Human Research Ethics Commission of the Ministry of Health[2]. For Vigitel, free and informed consent was obtained orally at the time of telephone contact with the interviewees. The present study was assessed and approved by the Research Ethics Committee of the School of Public Health of the Universidade de São Paulo under Report number 1,885,826 of January 5, 2017.

This article comes from the master's dissertation of the author, Iolanda Karla Santana dos Santos. It is entitled *Patterns of food consumption and physical activity based on data from VIGITEL* and was presented to the Graduate Program in Nutrition in Public Health of the School of Public Health from the Universidade de São Paulo.

RESULTS

Table 1 shows the correlations between the patterns retained in the 2007 to 2012 set and for each year of the same period, with no rotation, and with the Varimax and Promax rotations. In the analyzes with no rotations, the components retained in the 2007 to 2012 set showed correlations greater than 0.90 with the retained patterns in each year, separately. In the rotational analyzes, only the first component showed correlations greater than 0.90 in all of the years.

DISCUSSION

Our results indicate that:

- PCA analysis can be used in time series data sets with the same sample structure;
- depending on the purpose of the study, it is not advisable to use Varimax or Promax rotation after retaining the components.

Table 1. Correlations of principal components (PC) not rotated and rotated retained for the set of years 2007–2012 and retained in analysis for each year. Surveillance System for Risk and Protection Factors for Chronic Diseases by Telephone Survey, 2007 - 2012.

	Monitoring Year						
	2007	2008	2009	2010	2011	2012	Total
No rotation							
CP1	0.9996	0.9997	0.9999	0.9997	0.9997	0.9997	0.9989
CP2	0.9972	0.9976	0.9990	0.9957	0.9975	0.9976	0.9968
CP3	0.9947	0.9910	0.9751	0.9936	0.9950	0.9902	0.9888
CP4	0.9704	0.9487	0.9783	0.9849	0.9897	0.9845	0.9742
Varimax							
CP1	0.9995	0.9999	0.9991	0.9999	1.000	0.9998	0.9989
CP2	-0.0556	-0.0626	-0.0336	-0.0275	0.9992	0.0186	0.1218
CP3	0.3539	0.5064	0.1683	0.5039	0.1508	0.1186	0.3120
CP4	0.9760	0.0801	0.0632	0.9897	0.1500	0.9998	0.5249
Promax							
CP1	0.9995	0.9999	0.9994	0.9998	1.000	0.9998	0.9989
CP2	-0.6927	-0.6423	0.9874	-0.6629	0.9978	0.9953	0.0542
CP3	0.0357	0.0234	0.9963	0.0431	-0.0651	-0.0722	0.1657
CP4	0.0802	0.9869	0.9991	0.0802	0.9990	0.9994	0.6680

In this study, with six years of monitoring and pattern retention with eigenvalues > 1.0, the correlations between the retained patterns in the set and for each year with no rotation were greater than 0.90, showing high internal consistency. Regarding the patterns that did not remain in the comparative analyzes, some pairs showed correlations below 0.90.

In an expanded analysis (data not shown can be requested from the authors), in which we included all of the years of monitoring, the correlations between some of the patterns extracted from the set and equivalent patterns extracted from the databases, separated by year, were less than 0.90. This lower association is exactly the reflection of changes in the consumption of diet components, distributed among the population and relevant to the interpretation of changes in dietary patterns. In this case, without analyzing the databases together, it would be impossible to interpret the changes in dietary patterns that occurred in the period.

CONCLUSION

PCA and other multivariate techniques contribute widely to time series studies, and their interpretation becomes more effective without the use of statistical adjustments such as vector rotation, which is more useful when pursuing other objectives.

REFERENCES

1. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol [Internet] 2002; 13(1): 3-9. Available from: http://journals.lww.com/co-lipidology/Fulltext/2002/02000/Dietary_pattern_analysis__a_new_direction_in.2.aspx
2. Brasil. Ministério da Saúde. VIGITEL Brasil 2012. Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico. Estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2012. Brasília: Ministério da Saúde; 2013.
3. Olinto MTA. Padrões Alimentares. In: KAC G, SICHIERI R, GIGANTE DP, eds. Epidemiologia Nutricional. 20ª ed. Rio de Janeiro: Fiocruz/Atheneu; 2007. p. 213-25.
4. Lyra W da S, Silva EC da, Araújo MCU de, Fragoso WD, Veras G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. Quím Nova 2010; 33(7): 1594-7. https://doi.org/10.1590/S0100-40422010000700030
5. Jolliffe IT. Principal Component Analysis. 2ª ed. Nova York: Springer; 2002.

Received on: 07/29/2019
Revised on: 11/11/2019
Accepted on: 11/18/2019

Authors' contributions: Iolanda Karla Santana dos Santos: study design, literature review, data analysis and interpretation, preparation of the first version of the article, critical review of the text, approval of the final version of the article. Wolney Lisbôa Conde: study design, analysis and interpretation of the data, critical revision of the text, approval of the final version of the article. Alicia Matijasevich Manitto: study design, data analysis and interpretation, critical revision of the text, approval of the final version of the article.