

A comparison of three statistical methods applied in the identification of eating patterns

Comparação de três métodos estatísticos aplicados na identificação de padrões alimentares

Diana Barbosa Cunha ¹
Renan Moritz Varnier Rodrigues de Almeida ²
Rosângela Alves Pereira ¹

Abstract

This work aimed to compare the results of three statistical methods applied in the identification of dietary patterns. Data from 1,009 adults between the ages of 20 and 65 (339 males and 670 females) were collected in a population-based cross-sectional survey in the Metropolitan Region of Rio de Janeiro, Brazil. Information on food consumption was obtained using a semi-quantitative food frequency questionnaire. A factor analysis, cluster analysis, and reduced rank regression (RRR) analysis were applied to identify dietary patterns. The patterns identified by the three methods were similar. The factor analysis identified "mixed", "Western", and "traditional" eating patterns and explained 35% of the data variance. The cluster analysis identified "mixed" and "traditional" patterns. In the RRR, the consumption of carbohydrates and lipids were included as response variables and again "mixed" and "traditional" patterns were identified. Studies comparing these methods can help to inform decisions as to which procedures best suit a specific research scenario.

Food Consumption; Food Habits; Statistical Factor Analysis

Introduction

The human diet is extremely complex. When energy and nutrient intakes are used to study the effects of diet on the development of disease, it is often difficult to observe associations between specific outcomes and diet. The analysis of food consumption through an examination of dietary patterns is an alternative strategy to understand the relationships among dietary elements. The identification of eating patterns yields a more coherent model in the study of eating habits that makes it possible to identify population subgroups at risk of disease and to propose well-grounded dietary guidelines ^{1,2,3}.

Dietary patterns can be identified by *a priori* hypothesis-oriented methods (such as indexes and scores based on food guides and nutritional recommendations) or by exploratory statistical procedures that analyze the food intake data co-variation structure and reveal dietary patterns, which are interpreted *a posteriori* ^{3,4}. Two approaches have been frequently used: the factor analysis ^{5,6,7,8,9,10} and the cluster analysis ^{11,12,13,14}. Recently, the method known as reduced rank regression analysis (RRR) or maximum redundancy analysis has been applied for the same purpose ^{15,16,17,18,19}. This approach is both *a priori* hypothesis-oriented and an exploratory statistical method, since its application requires the definition of response variables, based on the scientific knowledge of the disease

¹ Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

² Programa de Engenharia Biomédica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

Correspondence

R. A. Pereira
Departamento de Nutrição Social Aplicada, Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro.
Av. Carlos Chagas Filho 373, Rio de Janeiro, RJ 21941-590, Brasil.
roapereira@ufrj.br

physiology, and the subsequent analysis aims to explore the data²⁰.

In the present study, three methods (factor analysis, cluster analysis and RRR) are compared and applied to the identification of dietary patterns in a sample of adults living in a low socioeconomic neighborhood in the greater metropolitan region of Rio de Janeiro, Brazil.

Material and methods

The analysis included male and female subjects between the ages of 19 and 65; data were obtained in a population-based cross-sectional study carried out in the municipality of Duque de Caxias in the state of Rio de Janeiro, Brazil. Subject selection was based on a three-stage random sampling process (census tract, residences and individuals). The calculated sample size was 1,125 households and was based on a 14.5% prevalence rate of “extreme poverty” (defined as a monthly per capita income equal to one quarter of the monthly minimum wage, or approximately US\$ 30 in 2005) and an acceptable maximum error of 5%.

Data were collected in household interviews conducted between May and December 2005. All subjects signed a free consent form, and the research was approved by the Institutional Research Board of the Institute of Social Medicine at the State University of Rio de Janeiro. Detailed information about sampling and data collection methods can be found in Salles-Costa et al.²¹.

Food consumption was estimated using a semi-quantitative food frequency questionnaire (FFQ) that had been validated for the Rio de Janeiro area adult population²² which included 82 foods, three servings sizes and eight options for reporting the frequency of food intake (ranging from “never or almost never” to “more than three times per day”).

In the first step of the analysis, the reported intake frequencies of the 82 food items in the FFQ were converted to daily consumption frequencies; these food items were grouped into 21 categories according to their nutritional characteristics and usual frequency of consumption in the study population (Table 1). Some foods were not included in any group and maintained isolated for different reasons; for example eggs, soft drinks and sweetened fruit juices were kept separate because of their nutritional composition; rice, beans and bread constituted three different categories because they are major staple foods in the regional diet and consumed in large quantities by the studied population.

Factor analysis

Factor analysis is “a set of statistical techniques that are applied with the aim of representing or describing a number of initial variables by a smaller number of hypothetical variables”²³ (p. 223). Thus, a smaller set of variables (called “factors” or “components”) can be identified from the correlation structure of a given set of variables; these factors represent the variance of the original data^{24,25}.

The Bartlett Test of Sphericity (BTS) and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) were used to evaluate whether the data were suitable for a factor analysis. The principal component analysis (PCA) method was used for factor extraction, and the factors were orthogonally rotated using the varimax procedure in order to improve the interpretation of the results^{25,26}.

The procedure generated factor loadings for each variable (in this case the food groups) related to each factor (or dietary pattern) identified. These loadings measure the correlation between the identified dietary pattern and the original variables (or food groups)²⁶. Thus, factor loadings with positive values indicate that the variable is associated with the eating pattern, and negative values show that the food group is inversely related to the pattern. Larger factor loading values indicate a greater contribution of that food group to the dietary pattern. Food items were retained in the pattern if the factor loading value was equal to or above 0.30. The communalities represent the variance of each item explained by all factors combined. In this study, a minimum communality value equal to or greater than 0.25 was considered to be acceptable²³.

The analysis also estimates *eigenvalues*, or the proportion of the total data variance that can be explained by each factor. The decision to consider a dietary pattern was based on the *scree test*²⁷, a chart that depicts the eigenvalues and their corresponding extracted factors (Figure 1). The points were plotted on a chart, and factors with eigenvalues located before the inflection point of the plotted line were retained.

Factor scores are standardized variables that represent the adherence of the subject to each factor^{24,25}. Those values were saved for subsequent analysis of the association between dietary patterns and a given outcome. Most statistical software programs are able to perform a factor analysis. In this study, the analysis was performed with the *Data Reduction – Factor Analysis* procedure in the SPSS version 16.0 package, selecting the following options: in ex-

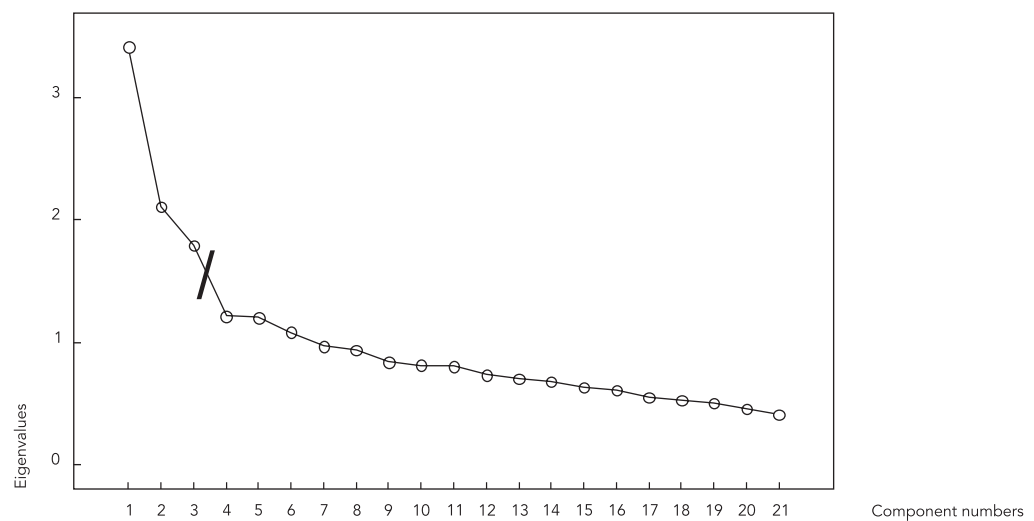
Table 1

Food groups considered for the identification of dietary patterns of adults (N = 1,009) living in a low socioeconomic neighborhood in the Rio de Janeiro Metropolitan Area, Brazil, 2005.

Groups	Foods
Cereals	Polenta, corn and spaghetti
Bread	French bread or loaf bread
Rice	Rice
Beans	Black beans
Fish and shrimps	Fresh fish, shrimps, sardines and tuna fish
Vegetables	Chayote, okra, cucumber, beets, carrots, onion, garlic, peppers, tomatoes, zucchini, pumpkin and green beans
Leafy vegetables	Cauliflower, lettuce, cabbage and chicory
Roots	Manioc, cassava flour, yams and baked potatoes
Fruits	Orange, tangerine, banana, pears, pineapple, apples, guava, melon or watermelon, avocado, mango, passion fruit, grapes
Juices	Juices
Cakes, biscuits	Cakes, biscuits (salted or sweet)
Soft drinks	Soft drinks
Milk and dairy	Milk, yogurt, cheese, cream cheese, other dairy
Meat	Chicken, beef, ground beef, pork, hamburger, viscera
Eggs	Eggs
Sausages	Sausages
Caffeinated beverages	Yerba mate and coffee
Sweets	Caramels, ice cream, pudding, flan, other sweets, chocolate in powder form or in bars
Sugar	Added sugar
Snacks and fast food	Fried chips, other salty snacks, pizza and popcorn
Sauces and fats	Mayonnaise, bacon, butter or margarine

Figure 1

Scree plot test to define the factors to be retained in the model (factor analysis).



traction, principal components; in rotation, *varimax* (SPSS Inc., Chicago, USA).

Cluster analysis

Cluster analysis is another method used to identify dietary patterns. Instead of grouping factors, this method is based on the clustering of individuals according to regularities in their food intake. Thus, the aim of the method is to assign individuals to subgroups (or clusters) in which food consumption is relatively homogeneous; the intra-individual variability inside the cluster is supposedly small, while the between-groups variability is important because of the differences in food intake among subgroups^{1,28,29}. In this procedure, individuals were classified in a pre-defined number of clusters by means of the Euclidean distance metric. The initial cluster seeds were followed by repeated comparisons between the means of the initial clusters and subsequent updates of the cluster groupings and means. Subjects were moved between mutually exclusive clusters and new means were computed until the distances between the observations within a cluster were minimized relative to the distances between the clusters^{30,31,32}.

Most statistical software packages are able to perform a cluster analysis. In this study, the *Classify-K-Means Cluster* procedure in the SPSS program (version 16.0) was used with 20 maximum iterations. This procedure is particularly useful for large samples. The analysis was performed for two and three clusters, and the most interpretable solution was compared with the other methods of dietary pattern identification.

Reduced rank regression analysis (RRR)

The RRR (also called the maximum redundancy analysis method), introduced in nutritional epidemiology by Hoffmann et al.¹⁹, is similar to the factor analysis. However, the RRR includes two different kinds of variables: predictors and responses. The factor analysis cannot always satisfactorily identify dietary patterns that are predictors of disease, and the inclusion of response variables in the RRR approach represents an advance in the study of the effects of diet on the development of chronic disease. The RRR simultaneously encompasses both hypothesis-oriented and explanatory approaches². Usually, the predictor variables are the intakes of specific foods or food groups and the response variables are defined by biomarkers or other mediating variables (e.g., nutrients or ratios of nutrients) that are presumed to be important for the development of the disease under study¹⁹. Unlike

the factor analysis (which yields factors that explain the maximum variation in the consumption of food groups), the RRR analysis yields a linear combination of food groups that explain the maximum variation in the response variables^{2,33,34}.

In an RRR, the number of response variables limits the number of dietary patterns that can be identified^{19,35}. In this study, the 21 food groups mentioned above were the predictor variables in the RRR, while the RRR response variables were the consumption of carbohydrates and fat (both in grams per day). These response variables were chosen because of their known importance to weight variation^{36,37,38}. The RRR procedure calculates X and Y-scores¹⁵; in the present analysis, the X-scores were derived from the intakes of each food group (the predictor variables) and the Y-scores were based on the intakes of carbohydrates and fat (the response variables). The RRR was completed with the PLS procedure in SAS for Windows, version 9.1 (SAS Inst., Cary, USA). The SAS code for the RRR procedure is shown in the Table 2. Cronbach's alpha value was estimated to ascertain the extracted patterns' internal consistency, and, finally, the patterns were named based on the interpretation of the data.

To compare the methods, the means of the X-scores (RRR) and factor scores (factor analysis) were calculated and analyzed according to the identified cluster solutions.

Results

A total of 1,253 individuals in the examined age range were interviewed; 222 individuals (17% of the total data set) were excluded because they reported implausible energy consumption [less than 500kcal/day (n = 5) or more than 6,000kcal/day (n = 217)]. Another 22 subjects were excluded because of incomplete information. The final sample consisted of 1,009 people; 34% (n = 339) were male and 66% (n = 670) were female. The average age of the participants was 39 (standard deviation – SD = 12 years), and there was no statistically significant difference in mean age by sex (Student's t test; p-value = 0.73).

Table 3 describes the dietary patterns identified through the factor, cluster and RRR analysis. The factor analysis indicated that three factors should be retained. These factors were: "mixed" (characterized by the consumption of cereals, fish and shrimp, vegetables, fruits, eggs, meats and caffeinated beverages), "traditional" (characterized by the consumption of rice and beans, breads, sauces, sugar and fats) and "Western" (characterized by a higher intake of juices, cakes

Table 2

SAS code for the reduced rank regression analysis (RRR) applied in the identification of dietary patterns.

SAS code	Meaning
proc pls method=RRR	Invokes the <i>pls</i> procedure and indicates that the data will be analyzed by RRR
nfac=2 varss details;	Specifies the number of factors to be extracted
model GLICI LIPID	Defines the response variables
=grupo1-grupo21;	Indicates the predictors variables
output out=pattern xscore=scorex yscore=scorey; run;	Two factors and two response scores will be produced

Table 3

Dietary patterns identified by factor, cluster, and reduced rank regression (RRR) analysis. Adults (N = 1,009) living in a low socioeconomic neighborhood in Rio de Janeiro Metropolitan Area, Brazil, 2005.

	Mixed pattern	Traditional pattern	Western pattern
Factor analysis	Cereals, fish and shrimps, leafy greens, vegetables, roots, meat, eggs, sausage, caffeinated beverages	Rice, beans, bread, sugar, sauces, fats	Fast-foods, soft drinks, juice, milk and dairy, sweets, cakes and cookies
Cluster analysis	Cereals, fish and shrimps, leafy greens, vegetables, roots, meat, eggs, sausage, caffeinated beverages, soft drinks, juice, milk and dairy, sweets, cakes and cookies, sauces and fats, fast-foods	Rice, beans, bread, sugar	
RRR	Cereals, leafy greens, vegetables, roots, meat, eggs, sausage, caffeinated beverages, soft drinks, juice, milk and dairy, sweets, cakes and cookies, sauces and fats, fast-foods, sugar	Rice, beans, bread, caffeinated beverages	

and cookies, soft drinks, dairy products, pastries, snacks and fast food).

The solution of the cluster analysis identified two patterns, as follows: “mixed” (characterized by the consumption of fruit, juices, cakes and cookies, soft drinks, dairy products, pastries, snacks, fast food, sauces and fats) and “traditional” (characterized by the consumption of rice, beans, bread and sugar). In this analysis, 202 individuals were associated with the “mixed” cluster and 807 subjects were included in the “traditional” cluster.

The two patterns defined by the RRR were the “mixed” (including cereals, leafy greens, vegetables, roots, meat, eggs, sausage, caffeinated beverages, soft drinks, juice, milk and dairy products, sweets, cakes and cookies, sauces and fats, fast foods and sugar) and the “traditional” (which included caffeinated beverages, bread, rice and beans, and which was inversely associated with

the consumption of cereals, fish and shrimps, meat and eggs).

Table 4 shows the factor loadings of patterns obtained from the factor analysis and RRR, the variation in food groups and nutrient sources explained by each of the dietary patterns and the Cronbach’s alpha values. The “mixed” dietary pattern identified by the factor analysis explained 16% of the intake variance; the “Western” pattern explained 10% of the intake variance and the “traditional” pattern explained 9% of the intake variance. These patterns together explained 35% of the total variance in food consumption.

The patterns identified by the RRR explained, together, 17% of the data variance. The “mixed” pattern explained 54% of the carbohydrate intake variance and 67% of the fat intake variance. The “traditional” pattern explained 60% of the carbohydrate intake variance and 72% of the fat intake variance. Together, the response vari-

Table 4

Dietary patterns, factor loadings, proportion of variance explained and Cronbach's alpha for factor and reduced rank regression (RRR) analysis. Adults (N = 1,009) living in a low socioeconomic neighborhood in the Rio de Janeiro Metropolitan Area, Brazil, 2005.

Food groups	Factor analysis			RRR	
	Factor 1 (Mixed)	Factor 2 (Western)	Factor 3 (Traditional)	Factor 1 (Mixed)	Factor 2 (Traditional)
Cereals	0.38	-0.15	0.29	0.20	-0.20
Bread	-0.03	0.34	0.50	0.18	0.23
Rice	-0.07	-0.15	0.67	0.01	0.22
Beans	0.01	-0.08	0.66	0.08	0.19
Fish and shrimps	0.59	0.15	-0.15	0.25	-0.39
Vegetables	0.65	0.13	-0.20	0.16	-0.07
Leafy vegetables	0.65	0.10	-0.34	0.19	-0.15
Roots	0.54	-0.03	0.17	0.17	-0.07
Fruits	0.49	0.45	-0.18	0.27	0.07
Juices	0.17	0.45	-0.08	0.17	0.11
Cakes and biscuits	-0.09	0.48	-0.08	0.48	0.35
Soft drinks	0.02	0.51	0.12	0.19	0.10
Milk and dairy	0.11	0.62	-0.10	0.27	-0.13
Meat	0.52	0.39	0.05	0.32	-0.44
Eggs	0.40	-0.02	0.21	0.09	-0.30
Sausage	0.48	0.09	0.19	0.18	-0.34
Caffeinated beverages	0.36	-0.11	0.29	0.04	0.04
Sweets	0.01	0.70	0.19	0.25	-0.09
Sugar	0.04	0.07	0.33	-0.02	-0.00
Snacks and fast food	0.11	0.53	0.28	0.29	-0.21
Sauces and fats	0.10	0.23	0.46	0.15	-0.12
	Proportion of variance explained (%)				
Factors	16	10	9	11	5
Total	16	26	35	11	17
Fat	-	-	-	67	72
Carbohydrates	-	-	-	54	60
Fat + carbohydrates	-	-	-	60	66
Cronbach's alpha	0.64	0.55	0.48	0.56	0.47

ables explained 60% of the variability of the first pattern and 66% of the variability of the second pattern.

The Cronbach's alpha values identified for the three methods were acceptable: the values in the factor analysis were 0.64, 0.55 and 0.48, for "mixed", "Western" and "traditional" patterns, respectively. In the cluster analysis, the Cronbach's alpha values observed for "mixed" and "traditional" patterns were 0.67 and 0.44, respectively. In the RRR, the Cronbach's alpha values were 0.56 for the "mixed" pattern and 0.47 for the "traditional" pattern (Table 5). For the three analyses, there was no indication that

the removal of any food group would improve the Cronbach's alpha.

The means and 95% confidence intervals of the factor scores obtained from the factor analysis and of the X-scores from the RRR were calculated for the two cluster solutions (Figure 2). The "mixed" cluster presented the highest mean factor scores for the "mixed" pattern and the lowest mean scores for the "traditional" pattern. On the other hand, in the "traditional" cluster, the factor scores were less impressive, but the highest scores were observed for the "traditional" pattern identified in the factor analysis. In the RRR patterns, in the "mixed" cluster were observed

Table 5

Median intake frequency of food groups according to the clusters of dietary patterns (and respective Cronbach's alpha). Adults (N = 1,009) living in a low socioeconomic neighborhood in the Rio de Janeiro Metropolitan Area, Brazil, 2005.

Food groups	Cluster 1 (Mixed) [n = 202]	Cluster 2 (Traditional) [n = 807]
Cereals	1.27	0.91
Bread	1.72	1.81
Rice	5.21	5.95
Beans	1.99	2.46
Fish and shrimps	0.56	0.23
Vegetables	12.28	5.06
Leafy vegetables	5.04	1.74
Roots	2.22	1.25
Fruits	5.22	1.81
Juices	1.18	0.70
Cakes and biscuits	2.43	1.75
Soft drinks	0.98	0.75
Milk and dairy	2.06	1.04
Meat	2.01	1.26
Eggs	0.78	0.48
Sausage	0.54	0.29
Caffeinated beverages	2.93	2.05
Sweets	1.10	0.86
Sugar	2.80	2.96
Snacks and fast food	0.64	0.54
Sauces and fats	1.51	1.45
Cronbach's alpha	0.67	0.44

for the “mixed” RRR pattern. The highest scores in the “traditional” cluster were observed for the “traditional” pattern, although the mean was insignificant and nearly zero.

Discussion

This study found that factor, cluster and RRR analyses identified two comparable patterns: the “mixed” and the “traditional”. The “mixed” patterns yielded by the three methods presented common foods groups such as cereals, leafy greens, vegetables, roots, meat, eggs, sausage and caffeinated beverages; similarly, the “traditional” patterns described by the three methods shared foods, like rice, beans and bread. The factor analysis also allowed for the identification of a third pattern, denominated as “Western”, that included some of the items assigned to the “mixed” patterns derived by the cluster and the RRR analysis, for example, fast-foods, soft drinks, juice, milk and dairy, sweets, cakes and cookies.

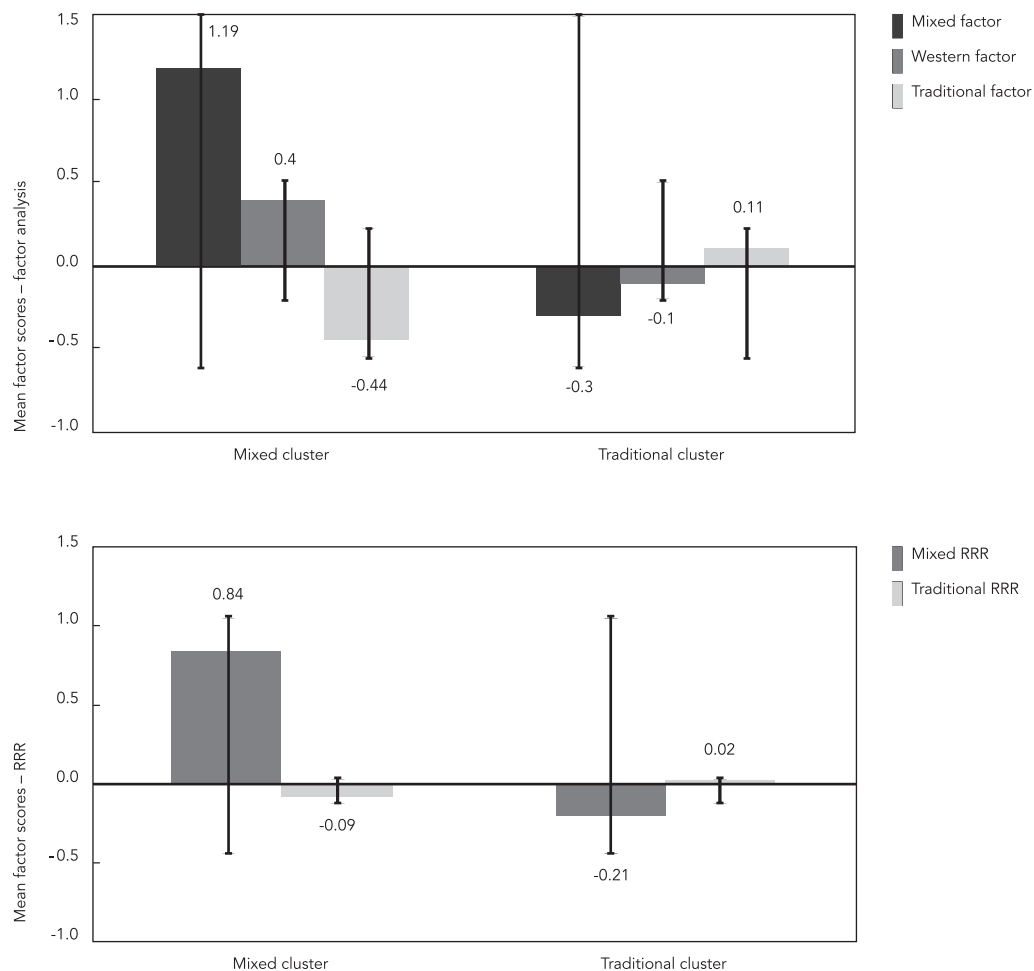
Although there are many published scientific articles discussing the identification of dietary

patterns, few studies directly compare these three methods. Some studies compared the results of factor and cluster analysis^{24,39,40,41,42}. Additionally, factor analysis was compared with RRR by Di Bello et al.⁴³ and Hoffman et al.³⁴. However, like in many other fields of knowledge, comparative studies are very useful as researchers choose the most appropriate method for a specific scenario.

Other studies have also observed similar dietary patterns when comparing factor and cluster analysis^{24,29,31,39,42}. Lozada et al.³⁹ carried out a cross-sectional study involving 477 women aged 12-19 years; like the present study, they compared the means of extracted factor scores in each identified cluster in order to characterize cluster versus factor results. The authors also concluded that the dietary patterns obtained from the two methods were highly comparable. Newby et al.²⁹ analyzed the dietary patterns of 459 healthy subjects enrolled in the Baltimore Longitudinal Study of Aging. The authors compared cluster and factor analysis in relation to plasma lipid biomarkers and observed a high degree of compatibility between the patterns.

Figure 2

Mean (and 95% confidence interval – 95%CI) of factor scores for patterns derived by factor and reduced rank regression (RRR) analysis according to the clusters identified in cluster analysis. Adults (N = 1,009) living in a low socioeconomic neighborhood in the Rio de Janeiro Metropolitan Area, Brazil, 2005.



Cluster analysis has the advantage of categorizing each subject exclusively into one pattern. Factor analysis, however, attributes factor score values based on the extracted dietary patterns of all subjects and can make it difficult to translate results to the individual level. The concept of assigning a *factor score* is less intuitive than assigning individuals to a subgroup; on the other hand, factors are traditionally estimated by procedures that guarantee their statistical independence (such as orthogonal rotation), which allows them to be used in linear regression or other types of modeling³¹. Also, factor analysis appears to be reproducible over time and across different di-

etary assessment methods⁴⁴. In this study, factor analysis was able to correctly reveal the diversity and specificity of the dietary patterns; it identified the “Western” pattern in addition to the “mixed” and “traditional” patterns.

In the present study, RRR explained less of the consumption variability than factor analysis, since RRR focuses on the variation of the response variables that can be explained by the dietary patterns. The RRR allows the inclusion of an *a priori* hypothesis through the response variables; therefore, this approach allows researchers to examine biologically important intermediate variables in order to identify and interpret

the associations between diet and disease. However, it is necessary to include response variables that are presumably predictive of the health outcomes under study, and this may not always be feasible³⁰. Additionally, Schulze & Hoffman²⁰ pointed out that it is unlikely that the RRR would be able to identify dietary patterns that are related to all of the potential pathways involved in the relationship between diet and disease.

The use of one of the three compared procedures involves some degree of subjectivity, since some decisions are made arbitrarily, e.g., the definition of the variables to be included in a model, the number of factors that should be retained and the name associated with each pattern. This freedom of action and the fact that the identified dietary patterns are specific for a particular data set make it harder to compare between different studies²⁸.

In summary, the factor analysis is more formalized and more commonly used in studies that are trying to identify dietary patterns. Cluster analysis has the advantage of generating food patterns that are mutually exclusive. RRR represents an interesting alternative approach for the study of dietary patterns, since, by including pre-defined variables, prior information is taken into account in the study. This makes RRR a powerful tool for the confirmation or rejection of a hypothesis⁴⁵.

The assumptions and objectives of a particular study must affect the selection of the statistical technique that will be used to identify food patterns. Studies of the identification of dietary patterns are ongoing, and analyses that compare and improve these methodologies are essential to provide technical and scientific bases for deciding which method is suitable in each specific research scenario.

Resumo

O objetivo deste estudo foi comparar três métodos estatísticos empregados na identificação de padrões alimentares. Foram utilizados dados de 1.009 adultos de 20 a 65 anos de idade (339 homens e 670 mulheres) coletados em estudo transversal de base populacional desenvolvido na Região Metropolitana do Rio de Janeiro, Brasil. Para a identificação de padrões alimentares foram aplicadas a análise fatorial, análise de cluster e a reduced rank regression analysis (RRR). Os três métodos identificaram padrões alimentares similares. A análise fatorial identificou os padrões "misto", "ocidental" e "tradicional", que explicaram 35% da variância total. Pela análise de cluster, foram identificados dois padrões "misto" e "tradicional". Para a RRR, os consumos de carboidratos e de lipídeos foram incluídos como variáveis respostas e também foram identificados os padrões "misto" e "tradicional". A comparação desses métodos auxilia na definição do procedimento mais adequado para um cenário de pesquisa específico.

Consumo de Alimentos; Hábitos Alimentares; Análise Fatorial

Contributors

D. B. Cunha participated in the statistical analysis, results interpretation, and manuscript writing and R. M. V. R. Almeida in the data analysis and interpretation, and manuscript writing. R. A. Pereira was responsible for the study conception, data analysis and interpretation, manuscript writing and revising.

Acknowledgments

This work was partly supported by the National Research Council (CNPq, grant n°. 503139; 2003-3) and by the National Cancer Institute of Brazil. D. B. Cunha received a fellowship from the Graduate Studies Coordinating Board (CAPES) during her Masters degree work.

References

1. Jinlin F, Binyou W, Terry C. A new approach to the study of diet and risk of type 2 diabetes. *J Postgrad Med* 2007; 53:139-43.
2. van Dam RM. New approaches to the study of dietary patterns. *Br J Nutr* 2005; 93:573-4.
3. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002; 13:3-9.
4. Sieri S, Krogh V, Pala V, Muti P, Micheli A, Evangelista A, et al. Dietary patterns and risk of breast cancer in the ORDET cohort. *Cancer Epidemiol Biomarkers Prev* 2004; 13:567-72.
5. Jackson M, Walker S, Simpson C, McFarlane-Anderson N, Bennett F. Are food patterns associated with prostate cancer in Jamaican men: a preliminary report. *Infect Agent Cancer* 2009; 4 Suppl 1:S5.
6. Gustaw-Rothenberg K. Dietary patterns associated with Alzheimer's disease: population based study. *Int J Environ Res Public Health* 2009; 6:1335-40.
7. Neumann AICP, Martins IS, Marcopito LF, Araújo EAC. Padrões alimentares associados a fatores de risco para doenças cardiovasculares entre residentes de um município brasileiro. *Rev Panam Salud Pública* 2007; 22:329-39.
8. Schulze MB, Fung TT, Manson JE, Willett WC, Hu FB. Dietary patterns and changes in body weight in women. *Obesity (Silver Spring)* 2006; 14:1444-53.
9. Sichieri R, Castro JE, Moura AS. Fatores associados ao padrão de consumo alimentar da população brasileira urbana. *Cad Saúde Pública* 2003; 19 Suppl 1:S47-53.
10. Marchioni DML, Latorre MRDO, Eluf-Neto J, Wünsch-Filho V, Fisberg RM. Identification of dietary patterns using factor analysis in an epidemiological study in Sao Paulo. *São Paulo Med J* 2005; 123:124-7.
11. Samieri C, Jutand MA, Feart C, Capuron L, Letenneur L, Barberger-Gateau P. Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health. *J Am Diet Assoc* 2008; 108:1461-71.
12. Carrera PM, Gao X, Tucker KL. A study of dietary patterns in the Mexican-American population and their association with obesity. *J Am Diet Assoc* 2007; 107:1735-42.
13. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres R, Tucker KL. Dietary patterns and changes in body mass index and waist circumference in adults. *Am J Clin Nutr* 2003; 77:1417-25.
14. Quatromoni PA, Copenhafer DL, D'Agostino RB, Millen BE. Dietary patterns predict the development of overweight in women: The Framingham Nutrition Studies. *J Am Diet Assoc* 2002; 102:1239-46.
15. Liese AD, Weis KE, Schulz M, Toozee JA. Food intake patterns associated with incident type 2 diabetes: the Insulin Resistance Atherosclerosis Study. *Diabetes Care* 2009; 32:263-8.
16. McNaughton SA, Mishra GD, Brunner EJ. Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. *Diabetes Care* 2008; 31:1343-8.
17. Nöthlings U, Murphy SP, Wilkens LR, Boeing H, Schulze MB, Bueno-de-Mesquita HB, et al. A food pattern that is predictive of flavonol intake and risk of pancreatic cancer. *Am J Clin Nutr* 2008; 88:1653-62.
18. McCann SE, McCann WE, Hong CC, Marshall JR, Edge SB, Trevisan M, et al. Dietary patterns related to glycemic index and load and risk of premenopausal and postmenopausal breast cancer in the Western New York Exposure and Breast Cancer Study. *Am J Clin Nutr* 2007; 86:465-71.
19. Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 2004; 159:935-44.
20. Schulze MB, Hoffmann K. Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. *Br J Nutr* 2006; 95:860-9.
21. Salles-Costa R, Pereira RA, Vasconcellos MTL, Veiga GV, Marins VMR, Jardim BC, et al. Associação entre fatores socioeconômicos e insegurança alimentar: estudo de base populacional na Região Metropolitana do Rio de Janeiro, Brasil. *Rev Nutr* 2008; 21 Suppl:99s-109s.
22. Sichieri R, Everhart J. Validity of a Brazilian food frequency questionnaire against dietary recalls and estimated energy intake. *Nutr Res* 1998; 18:1649-59.
23. Olinto MTA. Padrões alimentares: análise de componentes principais. In: Kac G, Sichieri R, Gigante DP, organizadores. *Epidemiologia nutricional*. Rio de Janeiro: Editora Fiocruz/Editora Atheneu; 2007. p. 213-25.
24. Hearty AP, Gibney MJ. Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults. *Br J Nutr* 2009; 101:598-608.
25. Norusis MJ. *The SPSS guide to data analysis*. Chicago: SPSS Inc.; 1990.
26. Basilevsky AT. *Statistical factor analysis and related methods: theory and application*. New York: Wiley-Interscience; 1994.
27. Cattell R. The screen test for the number of factors. *Multivariate Behav Res* 1966; 1:245-76.
28. Schulze MB, Hu FB. Dietary patterns and risk of hypertension, type 2 diabetes mellitus, and coronary heart disease. *Curr Atheroscler Rep* 2002; 4:462-7.
29. Newby PK, Muller D, Tucker KL. Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *Am J Clin Nutr* 2004; 80:759-67.
30. Michels KB, Schulze MB. Can dietary patterns help us detect diet-disease associations? *Nutr Res Rev* 2005; 18:241-8.
31. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* 2004; 62:177-203.
32. Greenwood DC, Cade JE, Draper A, Barrett JH, Calvert C, Greenhalgh A. Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *Eur J Clin Nutr* 2000; 54:314-20.

33. Nettleton JA, Steffen LM, Schulze MB, Jenny NS, Barr RG, Bertoni AG, et al. Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr* 2007; 85:1615-25.
34. Hoffmann K, Boeing H, Boffetta P, Nagel G, Orfanos P, Ferrari P, et al. Comparison of two statistical approaches to predict all-cause mortality by dietary patterns in German elderly subjects. *Br J Nutr* 2005; 93:709-16.
35. Hoffmann K, Zyriax BC, Boeing H, Windler E. A dietary pattern derived to explain biomarker variation is strongly associated with the risk of coronary artery disease. *Am J Clin Nutr* 2004; 80:633-40.
36. Mueller-Cunningham WM, Quintana R, Kasim-Karakas SE. An ad libitum, very low-fat diet results in weight loss and changes in nutrient intakes in postmenopausal women. *J Am Diet Assoc* 2003; 103:1600-6.
37. Sherwood NE, Jeffery RW, French SA, Hannan PJ, Murray DM. Predictors of weight gain in the Pound of Prevention Study. *Int J Obes Relat Metab Disord* 2000; 24:395-403.
38. Lissner L, Heitmann BL, Bengtsson C. Low-fat diets may prevent weight gain in sedentary women: prospective observations from the population study of women in Gothenburg, Sweden. *Obes Res* 1997; 5:43-8.
39. Lozada AL, Flores M, Rodríguez S, Barquera S. Patrones dietarios en adolescentes mexicanas. Una comparación de dos métodos. *Encuesta Nacional de Nutrición, 1999. Salud Pública Méx* 2007; 49:263-73.
40. Bamia C, Orfanos P, Ferrari P, Overvad K, Hundborg HH, Tjønneland A, et al. Dietary patterns among older Europeans: the EPIC-Elderly study. *Br J Nutr* 2005; 94:100-13.
41. Kant AK, Graubard BI, Schatzkin A. Dietary patterns predict mortality in a national cohort: the National Health Interview Surveys, 1987 and 1992. *J Nutr* 2004; 134:1793-9.
42. Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D, Trichopoulou A. Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. *Eur J Clin Nutr* 2003; 57:1378-85.
43. DiBello JR, Kraft P, McGarvey ST, Goldberg R, Campos H, Baylin A. Comparison of 3 methods for identifying dietary patterns associated with risk of disease. *Am J Epidemiol* 2008; 168:1433-43.
44. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* 1999; 69:243-9.
45. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. *Am J Clin Nutr* 2009; 90:1075-83.

Submitted on 12/Sep/2009

Final version resubmitted on 12/Feb/2010

Approved on 01/Apr/2010