

Parte I - Métodos em epidemiologia nutricional

12 - Padrões alimentares: análise de componentes principais

Maria Teresa Anselmo Olinto

SciELO Books / SciELO Livros / SciELO Libros

OLINTO, MTA. Padrões alimentares: análise de componentes principais. In: KAC, G., SICHIERI, R., and GIGANTE, DP., orgs. *Epidemiologia nutricional* [online]. Rio de Janeiro: Editora FIOCRUZ/Atheneu, 2007, pp. 213-225. ISBN 978-85-7541-320-3. Available from SciELO Books <<http://books.scielo.org>>.



All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

Padrões Alimentares: análise de componentes principais

Maria Teresa Anselmo Olinto

O perfil de morbidade e mortalidade da população brasileira tem passado por transformações que podem ser atribuídas às transições demográfica, nutricional e epidemiológica. A transição demográfica tem se caracterizado pelo aumento da proporção de pessoas com mais de 60 anos na estrutura populacional. A transição nutricional pode ser detectada pela elevação da prevalência de indivíduos apresentando sobrepeso e obesidade, com todas as suas conseqüências sobre o perfil de morbi-mortalidade. Já a transição epidemiológica destaca-se nesse quadro pela redução relativa de óbitos por doenças infecciosas e pelo aumento das mortes por Doenças Crônicas Não Transmissíveis (DCNT) – doenças cardiovasculares, neoplasias, diabetes *mellitus* e doenças respiratórias crônicas. Assim, estudos epidemiológicos têm sido realizados para identificar fatores etiológicos que possam reduzir ou atenuar a carga produzida pelo envelhecimento da população, pelas alterações nutricionais e pela presença de DCNT. Entre os principais fatores de risco modificáveis, estão os padrões alimentares.

Padrão alimentar pode ser definido como “o conjunto ou grupos de alimentos consumidos por uma dada população” (Garcia, 1999: 17). Na epidemiologia nutricional, esse conjunto ou grupos de alimentos são detectados por meio de métodos estatísticos de redução e/ou agregação de componentes.

Tradicionalmente, a epidemiologia nutricional dedicou-se ao estudo do efeito de nutrientes ou de um alimento específico sobre desfechos de saúde. Há evidências de que alimentos e nutrientes agem sinergicamente no risco de várias doenças crônicas (Jacobs & Steffen, 2003). A proposta de utilizar o padrão alimentar como exposição em epidemiologia nutricional supera limitações desde a incapacidade na detecção de pequenos efeitos de simples nutrientes até a dificuldade na avaliação das intercorrelações e das interações entre os nutrientes – por exemplo, o efeito da interação entre sódio e cálcio sobre a densidade mineral óssea (Mizushima et al., 1999).

Além disso, padrões alimentares expressam melhor a complexidade envolvida no ato de se alimentar, uma vez que pessoas não consomem de forma isolada alimentos ou nutrientes. Para a epidemiologia, que tem a população como foco, o estudo dos padrões alimentares pode melhor subsidiar a proposição de medidas efetivas de promoção da saúde por meio da alimentação (Newby et al., 2004).

A identificação de padrões alimentares pode ser feita *a priori* ou *a posteriori* (Newby et al., 2004). Na definição de padrões alimentares *a priori*, são propostos índices que permitem avaliar a qualidade da dieta com base em critérios conceituais de nutrição saudável e de diretrizes e recomendações nutricionais. Há índices resultantes de escores da totalização de nutrientes, da totalização de alimentos ou ainda índices que resumem a adesão

dos indivíduos a uma determinada diretriz dietética (Kant et al., 2000; Fitzgerald, Dewer & Veugelers, 2002; Kennedy, Olls & Callsson, 1995; Kant, 2004).

Na definição de padrões alimentares *a posteriori*, parte-se de dados empíricos de alimentos que são agregados com base em análise estatística, com posterior avaliação, ou seja, identificação de padrão alimentar. O presente capítulo aborda a identificação de padrões dietéticos por meio de métodos *a posteriori*. Esses métodos incluem: a escolha do instrumento para avaliar o consumo alimentar; a definição do tamanho de amostra; a coleta das informações; a análise estatística dos dados e a interpretação dos resultados com a definição de nomes para os padrões alimentares.

Na maioria dos estudos de padrão alimentar, os instrumentos utilizados para avaliar o consumo alimentar são o Questionário de Frequência Alimentar (QFA) e o Registro Dietético (RD) (Kant, 2004). Ambos os instrumentos apresentam vantagens e desvantagens (Thompson & Byers, 1994). O instrumento utilizado pode interferir na detecção da relação entre dieta e doença. Em algumas situações, há evidências de que o RD pode ser preferível ao QFA (Freedman et al., 2006). Além disso, no estudo dos padrões alimentares há a alternativa de utilizar dados existentes sobre disponibilidade de alimentos para determinadas populações (Sichieri, Castro & Moura, 2003).

Os métodos de avaliação de consumo estão apresentados de forma detalhada no capítulo 10, “Métodos de avaliação do consumo de alimentos”. Na utilização de QFA para a determinação dos padrões alimentares, deve-se ter atenção especial à definição do número e à forma de apresentação dos alimentos no QFA, assim como ao período de referência da informação. A eficiência é a melhor regra para definir o número de alimentos a serem incluídos em um QFA. Deve-se contrabalançar o tempo gasto para coletar os dados sobre os alimentos consumidos com a qualidade da informação obtida. Também se ressalta que o tempo de aplicação de um QFA terá repercussões no cronograma de execução da investigação e no tamanho da equipe necessária para aplicar o questionário na população. Na bibliografia, identificam-se estudos com QFA compostos de lista com menos de vinte alimentos e até de lista com 277 alimentos (Kumagai et al., 1999; Sevak et al., 2004).

Para a análise, interpretação e identificação dos padrões alimentares, considera-se ideal que cada alimento seja apresentado de forma isolada no QFA, ou seja, que não se incluam dois alimentos no mesmo item – por exemplo: ‘pão/bolo’. Entretanto, considerando-se a possibilidade de menor eficiência na aplicação de um questionário muito longo, recorre-se, na maioria das vezes, ao uso de QFA que contemplam no mesmo item dois ou três alimentos. Recomenda-se que o agrupamento desses alimentos seja realizado de acordo com o objetivo do estudo e suas hipóteses e, também, considerando as características nutricionais ou funcionais dos alimentos. O período de coleta de consumo alimentar pode ser referente ao último mês, ano etc. (Willett, 1998). No entanto, padrões alimentares provenientes de QFA de período curto de observação podem não contemplar as variações sazonais da produção de alimentos e da própria alimentação.

Procedimentos Estatísticos para a Redução dos Dados

Em epidemiologia nutricional, os dois métodos estatísticos mais utilizados para derivar padrões são a análise de agrupamento (*cluster*) e a análise fatorial. Em recente revisão (Newby et al., 2004) foram identificados 58 artigos com o emprego de análise fatorial e 35 com análise de agrupamento na identificação de padrões alimentares – artigos publicados desde 1980. A análise de agrupamento assemelha-se à análise fatorial em seu objetivo de avaliar a estrutura e de reduzir ou agrupar dados. No entanto, diferem no sentido de que a primeira agrega indivíduos (objetos) e a segunda, prioritariamente, agrega variáveis (Hair et al., 2005).

A análise fatorial reduz os dados em padrões baseados nas inter-relações (correlações) entre as variáveis, no caso, os itens alimentares. A análise de agrupamento reduz os dados e forma grupos com base em uma medida de

similaridade em termos de distância euclidiana através de sucessivos agrupamentos de pares próximos. Assim, ao mesmo tempo que os objetos ou os indivíduos são agregados em subgrupos relativamente homogêneos, ela maximiza a heterogeneidade entre os grupos, afastando os elementos mais distantes. Esta técnica é adequada para três situações: explorar padrões alimentares quando se suspeita que a amostra não é homogênea; quando inexistem as propriedades psicométricas requeridas para a análise fatorial ou, ainda, quando a intenção do investigador é manter todos os itens alimentares propostos no instrumento, ou seja, não excluir itens. No procedimento de análise fatorial, excluem-se itens-fator por apresentarem saturação insuficiente.

Conforme o objetivo da investigação científica, a análise fatorial pode ser exploratória ou confirmatória. Na análise exploratória, procura-se descrever e resumir dados, agrupando-se as variáveis que são correlacionadas. Este tipo de análise é indicado na geração de hipóteses e é utilizado para o estágio investigatório de processos subjacentes para identificar qual o aspecto que os itens agrupados manifestam. Não há necessidade de se conhecer, *a priori*, o número de fatores. A análise confirmatória, a qual envolve técnicas estatísticas mais sofisticadas, é utilizada em estágios avançados de uma investigação científica, ou seja, no teste de uma teoria ou hipótese, partindo-se de um número conhecido de fatores. No caso dos padrões alimentares, seria utilizada em uma etapa posterior à identificação dos padrões alimentares – por exemplo, na elaboração de um instrumento de avaliação dietética.

Na análise fatorial, os métodos utilizados para a redução de um grande número de variáveis a um número menor são a Análise de Componentes Principais (ACP) e a Análise de Fator Comum (AFC). Tais procedimentos baseiam-se em modelos matemáticos diferentes. A ACP analisa toda a variância, tanto a compartilhada como a exclusiva, e pressupõe que não existe erro. A AFC analisa unicamente a variância compartilhada; a variância exclusiva (única) não é incluída, e alguma variância do erro é admitida (Dancey & Reidy, 2006). A análise de componentes principais é uma técnica exploratória e a análise de fator comum tem sido utilizada para testar hipóteses, ou seja, no sentido confirmatório. Na prática, esses procedimentos podem ser utilizados nos mesmos tipos de dados oferecendo resultados similares, principalmente para grande conjunto de participantes. Para a identificação dos padrões alimentares, a análise de componentes principais tem sido utilizada com mais frequência.

Análise de Componentes Principais (ACP)

A análise de componentes principais foi desenvolvida por Hotelling em 1933, mas, posteriormente, seu uso foi muito ampliado com o desenvolvimento da informática e dos pacotes estatísticos (López-Valcárcel, 1991). Os objetivos da Análise de Componentes Principais (ACP) são: 1) descrever de forma sintética grande número de variáveis e 2) obter índices sintéticos, os componentes, que são as dimensões subjacentes que se identificam e podem ser nomeados.

As variáveis agrupadas em cada fator são mais fortemente correlacionadas entre si do que com as variáveis pertencentes aos outros fatores. Portanto, este procedimento possibilita que os itens de alimentos contidos no instrumento de avaliação de consumo alimentar sejam agrupados com base no grau de correlação entre eles. As etapas de análise incluem a preparação da matriz de correlação, a extração de um conjunto de fatores da matriz de correlação, a determinação do número de fatores e a rotação dos fatores para aumentar a sua interpretabilidade. A interpretação e a denominação dos fatores, no caso dos padrões alimentares, depende do significado de cada combinação das variáveis (itens de alimentos) observadas no fator e, principalmente, daqueles itens com maior carga fatorial.

Há diversas regras para estimar o tamanho da amostra (n) necessário para garantir a realização das etapas da análise fatorial. Em uma das regras, selecionam-se no mínimo dez indivíduos para cada item alimentar contido no instrumento. Em outra, o número de indivíduos deve ser no mínimo cinco vezes maior do que o número de itens presentes no instrumento (Pestana & Gageiro, 2005; Hair et al., 2005). Portanto, sendo K = número de itens de alimentos no instrumento, n é obtido da seguinte forma:

$$\begin{array}{l} \text{Se } 5 < K \leq 15 \Rightarrow n = 10 \times K \\ \text{Se } K > 15 \Rightarrow n = 5 \times K \end{array}$$

Por exemplo, para um QFA que contenha vinte itens alimentares (QFA-20), o tamanho da amostra deve ser de no mínimo cem pessoas com informações válidas em todos os itens alimentares. Como os estudos epidemiológicos incluem amostras com grande número de pessoas, eles possibilitam a aplicação de QFA com listas de alimentos bem completas.

Os dados do QFA coletados na amostra, após a codificação ter sido conferida e a dupla digitação realizada com a análise de consistência, devem ser transferidos para um pacote estatístico que permita a realização da ACP. Salienta-se que, quando se aplica um QFA quantitativo, será necessário converter as porções dos alimentos para gramas ou mililitros. Para o procedimento correto da análise de componentes principais, todos os itens de alimentos devem estar com a mesma unidade de medida. A utilização de QFA qualitativo, composto apenas da lista de alimentos e uma escala de frequência de consumo de cada item alimentar, é o método mais simplificado para a coleta, digitação e análise dos dados. Nesse caso, por exemplo, o QFA pode conter uma escala *likert* com as seguintes categorias: não-consumo = 0; consumo uma vez na semana = 1; consumo duas vezes na semana = 2; consumo três vezes na semana = 3; consumo de quatro a cinco vezes na semana = 4; e consumo de seis a sete vezes na semana = 5. Com esse tipo de escala qualitativa, não há necessidade de conversão de porções para uma unidade de medida (gramas), e os valores podem ser digitados diretamente no banco de dados e utilizados para a ACP.

Especificamente, nos procedimentos de análise dos dados deve-se iniciar com a avaliação da aplicabilidade do método de ACP. Para isso, utilizam-se os testes estatísticos de Kaiser-Meyer-Olkin (KMO) e o teste de Esfericidade de Bartlett. O KMO verifica a existência e o peso das correlações parciais. Para que o modelo fatorial seja adequado, o quociente entre o coeficiente de correlação simples e o coeficiente de correlação parcial deve se aproximar de 1, indicando um denominador de correlação parcial pequeno. Não são aceitos valores abaixo de 0,6; neste caso, deve-se utilizar outro tipo de análise. O teste de Esfericidade de Bartlett testa a hipótese nula de que a matriz de correlações é igual à matriz de identidade. Um valor de p igual ou menor a 0,05 indica que os dados produzem uma matriz de identidade, concluindo-se que o modelo fatorial é adequado para a análise dos dados. Salienta-se que o teste de esfericidade de Bartlett é muito sensível ao tamanho da amostra; portanto, em amostras grandes tende-se a rejeitar a hipótese nula. Assim, na análise de banco de dados provenientes de estudos epidemiológicos é preferível utilizar o KMO.

A seguir serão apresentados resultados da análise de componentes principais com base em um QFA com vinte itens alimentares (QFA-20) aplicados em uma amostra de 1.026 mulheres adultas. Esse QFA é resultante de uma versão resumida daquele utilizado no estudo de Alves e colaboradores (2006), o qual continha setenta itens alimentares. A Tabela 1 apresenta as estatísticas de KMO e de Bartlett para a aceitação da análise fatorial nos dados provenientes do QFA-20. Observa-se que os requisitos exigidos para a análise fatorial se cumprem com KMO de 0,730 (existem pequenas correlações parciais) e com o teste de esfericidade de Bartlett significativo ($p < 0,001$).

Tabela 1 – Parâmetros estatísticos para aceitação de análise fatorial (KMO e teste de esfericidade de Bartlett) para o QFA-20 (n = 1.023)

Kaiser - Meyer - Olkin	0,730
Teste de esfericidade de Bartlett Aprox. qui-quadrado	2291,322
Graus de liberdade	190
Significância estatística	< 0,001

Uma vez avaliada a aplicabilidade do método fatorial no conjunto de dados, deve-se observar, também, o percentual da variância de cada item explicada por todos os fatores juntos. Esse percentual, denominado comunalidade, pode ser interpretado como um indicador de confiabilidade do item. As comunalidades são calculadas antes e depois da extração dos fatores. Na ACP, as comunalidades iniciais são sempre iguais a 1, porque o método utiliza a variância total. Após a extração dos fatores, os resultados das comunalidades dos itens devem variar de 0 a 1, sendo 0 quando não explicam nenhuma variância e 1 quando explicam toda a sua variância.

A Tabela 2 apresenta as comunalidades para cada item de alimento presente no QFA-20. Os resultados mostram que as proporções das variâncias explicadas pelos fatores (comunalidades) são todas adequadas (próximas e superiores a 0,4). Por exemplo, a comunalidade do item 2 igual a 0,64 indica que ele explica 64% da variância total deste com relação aos fatores extraídos. No entanto, perde 36% do poder explicativo com a inter-relação com os demais itens.

Tabela 2 – Comunalidades iniciais e após a extração dos fatores por análise de componentes principais (QFA-20)

Item alimentar	Inicial	Extração
13	1,000	0,512
11	1,000	0,407
12	1,000	0,485
1	1,000	0,590
10	1,000	0,495
2	1,000	0,640
19	1,000	0,371
15	1,000	0,292
16	1,000	0,508
18	1,000	0,488
3	1,000	0,407
9	1,000	0,360
6	1,000	0,495
4	1,000	0,417
5	1,000	0,467
17	1,000	0,343

Tabela 2 – Comunalidades iniciais e após a extração dos fatores por análise de componentes principais (QFA-20 - continuação)

Item alimentar	Inicial	Extração
8	1,000	0,324
14	1,000	0,383
20	1,000	0,378

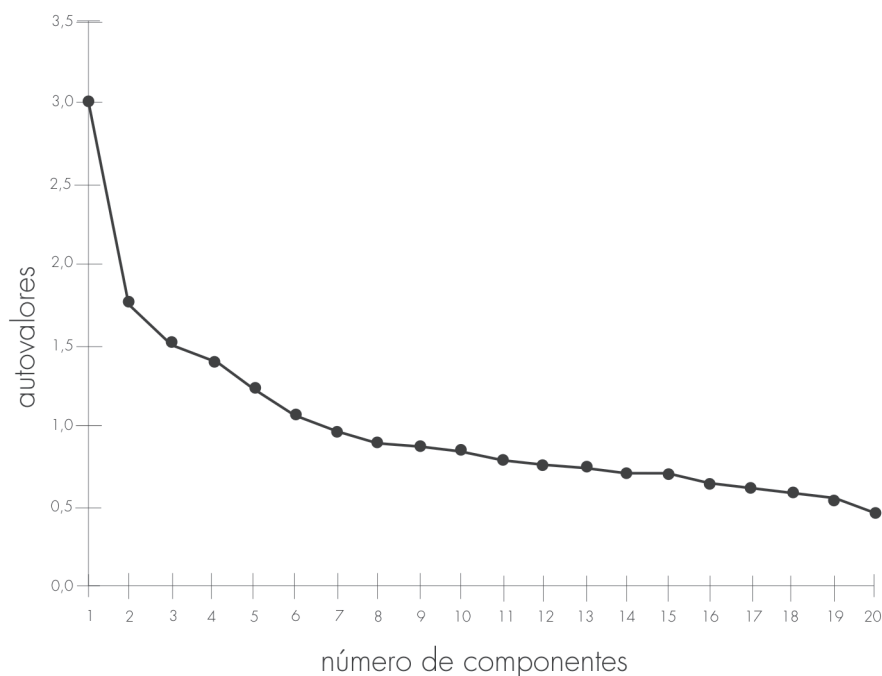
No próximo passo da análise, observam-se os resultados referentes à variância total explicada (Tabela 3). Cada autovalor representa o total da variância explicada pelo fator (componente). Os autovalores estão ordenados por tamanho, e a soma dos autovalores é igual ao número de variáveis na análise. A porcentagem da variância total atribuída ao fator 1 é 15,05% e ao fator 2 é 8,79%, e o acumulado para esses dois fatores é de 23,84%. Os autovalores representam a proporção da variância que cada fator é capaz de reter. Cada fator (componente) que apresenta autovalor maior do que 1 é retido. Na Tabela 3, os cinco fatores que apresentam claramente autovalor > 1 encontram-se grifados. No fator 6, o autovalor foi igual a 1,070, não sendo retido na extração. Os cinco fatores explicam 44,43% da variabilidade dos vinte itens originais.

Tabela 3 – Autovalores e porcentagem da variância total explicada antes da rotação dos fatores

Componente	Autovalores iniciais		
	Total	% da variância	% Acumulado
1	3,010	15,050	15,050
2	1,759	8,793	23,843
3	1,501	7,506	31,350
4	1,388	6,938	38,287
5	1,230	6,148	44,435
6	1,070	5,352	49,788
7	0,956	4,778	54,566
8	0,893	4,463	59,029
9	0,870	4,350	63,379
10	0,835	4,175	67,554
11	0,787	3,934	71,488
12	0,749	3,745	75,233
13	0,744	3,720	78,953
14	0,710	3,548	82,501
15	0,695	3,474	85,975
16	0,623	3,113	89,088
17	0,611	3,057	92,144
18	0,590	2,948	95,092
19	0,527	2,635	97,727
20	0,455	2,273	100,000

Além da utilização dos autovalores e do percentual da variância acumulada para definir o número de fatores a serem retidos, utiliza-se o teste gráfico de Cattell (*scree plot*). Ele é determinado fazendo-se o gráfico dos autovalores em relação ao número de fatores em sua ordem de extração. Na Figura 1, observa-se que a explicação fatorial com cinco fatores é adequada. Salienta-se que os pontos no maior declive indicam o número apropriado de fatores a serem retidos.

Figura 1 – Gráfico de sedimentação de Cattell da análise de componentes principais (QFA-20)



A Tabela 4 mostra o total da variância explicada após a rotação dos fatores. Embora o percentual de cada componente apresente alteração, salienta-se que o percentual acumulado dos fatores extraídos permanece o mesmo. Conforme referido anteriormente, o percentual acumulado pode contribuir para a definição do número de fatores a serem retidos. A Tabela 4 apresenta apenas os valores dos cinco fatores retidos. Entretanto, há divergências quanto ao valor satisfatório.

Tabela 4 – Autovalores e porcentagem da variância total explicada depois da rotação dos fatores

Componente	Autovalores iniciais		
	Total	% da variância	% acumulado
1	2,114	10,568	10,568
2	1,851	9,257	19,825
3	1,761	8,807	28,632
4	1,719	8,597	37,229
5	1,441	7,207	44,435

A rotação de fatores é importante para gerar uma estrutura de fatores facilmente interpretável; no entanto, não melhora o grau de ajuste em relação aos dados. Há três procedimentos de rotação ortogonal (Varimax, Quartimax e Equamax) e dois procedimentos de rotação oblíqua (Oblimin direto e Promax). A rotação oblíqua pressupõe a existência de correlação entre os fatores resultantes. O caso mais simples de rotação é a ortogonal. A maioria dos estudos exploratórios de padrão alimentar utiliza a rotação ortogonal, ou seja, busca modelos alimentares protótipos. A vantagem da rotação ortogonal é que os fatores resultantes não são relacionados e, portanto, podem ser incluídos juntos em uma posterior análise de regressão multivariada, sem afetar o resultado de qualquer dos fatores.

Entre os vários estudos realizados recentemente para a determinação de padrões alimentares, a análise de componentes com rotação ortogonal Varimax é a mais utilizada. O objetivo da rotação é maximizar as cargas fatoriais maiores e minimizar as cargas menores. Permite, também, a visualização dos resultados da matriz, tornando-os de mais fácil interpretação. A rotação ortogonal Varimax faz com que cada fator somente apareça com as saturações altas e com o menor número de variáveis (itens) possíveis.

Tabela 5 – Matriz da estrutura fatorial rotada

Alimento	Item	Fator				
		1	2	3	4	5
Couve	13	0,708				
Brócolis	12	0,662				
Abóbora e cenoura	10	0,644				
Agrião e alface	11	0,631				
Balas, sobremesas e doces	16		0,711			
Chocolate	18		0,692			
Presunto, mortadela, salame e copa	9		0,551			
Queijo	3	0,341	0,526			
Leite integral	1			0,766		
Leite desnatado	2			-0,740		
Açúcar	19			0,576		
Nata	15			0,471		
Melão, melancia	6				0,701	
Manga, pêra e pêssego	5				0,668	
Caqui, uva	4				0,627	
Sorvete	17				0,481	
Massa integral	7					0,721
Soja	14					0,613
Açúcar mascavo	20					0,570
Peixes, sardinhas, atum enlatado	8					0,379

A Tabela 5 apresenta a matriz de correlação rotada com os itens alimentares e os respectivos alimentos que eles representam. Cada item aparece com as respectivas saturações (cargas) nos fatores extraídos. No exemplo do

QFA-20, foram extraídos cinco fatores (padrões). Utilizando-se a rotação ortogonal Varimax, tem-se uma boa visualização das cargas de saturação de cada item alimentar nos fatores extraídos. Foram mantidos na matriz os alimentos cuja carga do fator apresenta-se superior a 0,30. Com essa carga já seria possível obter um poder estatístico de 80% e um nível de significância de 0,05 em uma amostra de 350 pessoas (Hair et al., 2005).

O exemplo mostrado na Tabela 5 apresenta fatores claros com índices limpos. Mas, às vezes, os resultados podem aparecer com: 1) alguns itens que não saturam em nenhum fator por não atingir o valor mínimo estabelecido para carga (usualmente, 0,30); 2) itens que saturam em mais de um fator e 3) fatores com menos de três itens. Para cada uma dessas ocorrências, o investigador deve tomar decisões durante o planejamento da análise, ou seja, no próprio programa estatístico definem-se os parâmetros a serem seguidos.

Cabe destacar que manter e apresentar no padrão alimentar itens com cargas negativas é uma decisão a ser tomada tendo em vista o objetivo de investigação. Observa-se, por exemplo, que no fator 3 aparece o item 2 (leite desnatado) com carga negativa (Tabela 6). Ao excluir-se o item alimentar com carga negativa daquele padrão, tem-se um modelo alimentar protótipo, ou seja, apenas com os alimentos que são consumidos. Em contrapartida, a opção de manter no padrão alimentar um item com associação inversa com os outros alimentos do fator tem a vantagem de abarcar a complexidade envolvida nos hábitos alimentares. Em termos de saúde pública, essa informação pode subsidiar estratégias para recomendações nutricionais na população em estudo. Há, também, a possibilidade de um item negativo ser testado em outros fatores (ver alpha de Cronbach a seguir).

Ao final, deve-se proceder à análise de consistência interna dos fatores. O alpha de Cronbach é uma das medidas mais utilizadas para avaliar a consistência interna de um grupo de variáveis. Os resultados do alpha de Cronbach variam de 0 a 1. Valores menores do que 0,6 são inaceitáveis quando se busca um construto homogêneo em termos de validação de escalas. No entanto, na busca de um padrão ou conjunto de elementos que manifestem um perfil de agregação, sem os requisitos de validar um determinado construto, os índices podem ser mais flexíveis. Os itens do mesmo fator que apresentarem correlação negativa no fator devem ser excluídos da análise, ou então verifica-se a possibilidade de um melhor agrupamento com outro fator, por meio do valor de alpha.

Nesse exemplo, em que o item 2 apresenta uma correlação negativa alta no fator 3 e diminui o alpha do fator, verificamos que, se incluído no fator 5, ele torna-se positivo e acrescenta consistência estatística ao fator, além de congruência explicativa à sua interpretação. Além disso, deve-se testar a presença de cada item no fator. Por exemplo, se na exclusão de um item o alpha de Cronbach passa de 0,75 (razoável) para 0,8 (bom), seria melhor excluir este item do fator. Esse procedimento deve ser realizado com cada fator separadamente, e quando for necessário excluir mais de um item, é conveniente eliminar um de cada vez. Salienta-se que o investigador deve considerar que, embora alguns itens não contribuam para aumentar significativamente o alpha, eles apresentam validade de construto na melhor explicação do fator e, portanto, sua exclusão não será a decisão mais acertada.

A Tabela 6 (página seguinte) apresenta um resumo dos procedimentos realizados para testar a consistência dos padrões. O alpha de Cronbach para o fator 1 é considerado razoável, e a retirada de qualquer um dos quatro itens alimentares reduz a consistência do fator. O mesmo ocorreu com o fator 2, embora o alpha de 0,552 seja mais fraco. Para o fator 3, observa-se que com a retirada do item 2 houve um aumento do alpha (de - 0,316 aumenta para 0,373), ou seja, aumenta a consistência do fator. Nesse caso, embora não seja possível observar na Tabela 6, há uma inversão de sinal tornando positivos os outros três itens alimentares remanescentes do fator. O fator 4 apresenta um alpha de 0,554, e a retirada de algum fator, ou mesmo o acréscimo do item 2, não melhora sua consistência. Ao final, observa-se que a inclusão do item 2 no fator 5, além de ter aumentado o alpha de Cronbach (de 0,338 aumenta para 0,386), tornou a carga desse item positiva.

Tabela 6 – Valores alpha de Cronbach: teste dos itens alimentares para consistência interna dos fatores

Fator	Alpha de Cronbach	Número de itens*	Item que, se deletado, aumenta o alpha	Mudança de item 2 do fator 3 para o fator 5
1	0,650	4	Nenhum	
2	0,552	4	Nenhum	
3	-0,316	4	Item 2	0,373
4	0,506	4	Nenhum	
5	0,338	4	Nenhum	0,385

* Todos os itens de cada fator foram testados.

Após a verificação da consistência interna, analisa-se cada fator e tenta-se explicar, teoricamente, porque esses itens se agruparam e quais os elementos comuns entre eles. Na nova configuração do fator 5, por exemplo, observa-se que a contribuição do item 2 em termos explicativos do fator é coerente com o conteúdo dos itens. O padrão alimentar (fator 5) era composto de massa integral, soja, açúcar mascavo e peixes, portanto, há coerência na inclusão do item 2, ou seja, leite desnatado. Entretanto, a presença desse item não estava adequada ao fator 3, o qual incluía leite integral, açúcar e nata.

A busca da explicação teórica para os grupamentos formados permite que se atribuam nomes aos fatores. Na maioria dos estudos, os padrões alimentares são rotulados de acordo com o item alimentar de maior carga de saturação ou com a composição nutricional dos alimentos do fator. Há, também, denominação de fatores de acordo com a cultura (tradicional) ou área geográfica (Mediterrâneo). No estudo realizado no Rio de Janeiro, as cargas para o feijão e o arroz justificaram a denominação do padrão alimentar tradicional (Sichieri, 2002). Já no estudo de padrões alimentares das mulheres do sul do Brasil, os fatores foram denominados considerando-se dois critérios: a relação com as DCNT (padrão alimentar de risco ou padrão alimentar protetor) e o custo da porção dos alimentos em cada fator (Alves et al., 2006).

O Quadro 1 apresenta um exemplo para denominação dos padrões alimentares identificados com base no QFA-20 aplicando-se análise de componentes principais com rotação ortogonal Varimax. Neste exemplo, os nomes foram atribuídos de acordo com grupo de alimentos e características nutricionais. Portanto, os nomes dos padrões alimentares identificados foram: ‘vegetais’, ‘doces e embutidos’, ‘calóricos’, ‘frutas’ e ‘saúáveis’.

Quadro 1 – Exemplo de denominação para os cinco padrões alimentares identificados

Alimento	Padrão alimentar (Fatores)				
	1	2	3	4	5
Couve Brócolis Abóbora e cenoura Agrião e alface	Vegetais				
Balas, sobremesas e doces Chocolate Presunto, mortadela, salame e copa Queijo		Doces e embutidos			
Leite integral Açúcar Nata			Calóricos		
Melão, melancia Manga, pêra e pêssego Caqui, uva Sorvete				Frutas	
Massa integral Soja Açúcar mascavo Peixes, sardinhas, atum enlatado Leite desnatado					Saudáveis

Considerações Finais

A análise fatorial permite reduzir um grande número de variáveis alimentares em um número menor de fatores, identificando estruturas subjacentes, ou seja, comportamentos não observáveis diretamente, que sintetizam a complexa interação de fatores genéticos, ambientais e individuais.

Neste tipo de análise, os padrões são derivados sem uma hipótese *a priori*. Assim, eles podem ser adequados para uma amostra particular, mas não necessariamente representam padrões de dietas ideais. Além disso, por serem específicos para uma amostra, os padrões derivados empiricamente podem não se repetir em outras populações, o que dificulta as comparações entre os estudos. As especificidades, entretanto, refletem o comportamento alimentar real de uma população e podem fornecer informações úteis para a elaboração de diretrizes nutricionais (Jacques & Tucker, 2001; Hu, 2002; Newby et al., 2004).

A identificação de padrões alimentares é útil na investigação da relação entre dieta e doença, principalmente quando a causa está vinculada a várias características da dieta. Porém, não seria a opção mais adequada para doenças causadas por um nutriente específico, como, por exemplo, ácido fólico e defeito do tubo neural (Hu, 2002).

Uma crítica freqüente na área da epidemiologia nutricional ao uso desta abordagem refere-se à subjetividade envolvida nas diversas etapas da análise para a redução dos fatores que vão desde a elaboração do QFA até a denominação dos fatores, a saber: quais e quantos alimentos serão incluídos no QFA; como os alimentos serão listados (individualmente ou em grupos); se em grupos, qual o critério para agrupar os alimentos; como serão inseridas as variáveis no banco de dados; qual o número de fatores a ser retido e, finalmente, que critérios serão utilizados para atribuir nomes aos padrões alimentares (Martinez et al., 1998; Newby et al., 2004). Nenhuma dessas decisões é inócua, e qualquer mudança será de difícil controle e detecção, levando a diferentes resultados.

Por exemplo, a escolha dos alimentos na lista do QFA é crucial e primordial na identificação de padrões alimentares, mas não menos importante é a escolha do número de fatores a serem extraídos.

Enfim, embora haja incertezas e subjetividade nas decisões durante os processos estatísticos de redução de dados, eles têm sido fundamentais para que se possa trabalhar com a enorme quantidade de informação proveniente dos instrumentos de coleta de consumo alimentar. Além disso, a subjetividade na condução deste tipo de análise não deve ser motivo para abandoná-la, sendo recomendados maior detalhamento na descrição de todas as decisões tomadas e mais estudos de validação.

Referências

- ALVES, A. L. S. et al. Padrões alimentares de mulheres adultas residentes em área urbana no sul do Brasil. *Revista de Saúde Pública*, 40(5): 865-873, 2006.
- DANCEY, C. P. & REIDY, J. *Estatística sem Matemática para Psicologia: usando SPSS para Windows*. Porto Alegre: Artmed, 2006.
- FITZGERALD, A. L.; DEWER, R. A. & VEUGELERS, P. J. Diet quality and cancer incidence in Nova Scotia, Canada. *Nutrition and Cancer*, 43: 127-132, 2002.
- FREEDMAN, L. S. et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *International Journal of Epidemiology*, 35(4): 1.011-1.021, 2006.
- GARCIA, R. W. D. *A Comida, a Dieta, o Gosto: mudanças na cultura alimentar urbana*, 1999. Tese de Doutorado, São Paulo: Universidade de São Paulo.
- HAIR, J. F. et al. *Análise Multivariada de Dados*. Porto Alegre: Artmed, 2005.
- HU, F. B. Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology*, 13(1): 3-9, 2002.
- JACOBS, D. R. J. & STEFFEN, L. M. Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *American Journal of Clinical Nutrition*, 78, suppl. 3: 508S-513S, 2003.
- JACQUES, P. F. & TUCKER, K. L. Are dietary patterns useful for understanding the role of diet in chronic disease? *American Journal of Clinical Nutrition*, 73(1): 1-2, 2001.
- KANT, A. K. Dietary patterns and health outcomes. *Journal of the American Dietetic Association*, 104(4): 615-635, 2004.
- KANT, A. K. et al. A prospective study of diet quality and mortality in women. *Jama*, 283: 2.109-2.115, 2000.
- KENNEDY, E. T.; OHLS, J. & CALLSON, S. The Healthy Eating Index: design and applications. *Journal of the American Dietetic Association*, 95: 1.103-1.108, 1995.
- KUMAGAI, S. et al. Effect of food intake on all-cause mortality in the community elderly: a 7-year longitudinal study. *Journal of Nutrition, Health and Aging*, 3: 29-33, 1999.
- LÓPEZ-VALCÁRCEL, B. G. *Análisis Multivariante: aplicación al ámbito sanitario*. Barcelona: SG Editores, 1991.
- MARTINEZ, M. E. et al. Invited commentary. Factor analysis and the search for objectivity. *American Journal of Epidemiology*, 148(1): 17-19, 1998.

- MIZUSHIMA, S. et al. Preventive nutritional factors in epidemiology: interaction between sodium and calcium. *Clinical and Experimental Pharmacology and Physiology*, 26(7): 573-575, 1999.
- NEWBY, P. K. et al. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutrition Reviews*, 62(5): 177-203, 2004.
- PESTANA, M. H. & J. N. GAGEIRO. *Análise de Dados para Ciências Sociais: a complementaridade do SPSS*. Lisboa: Edições Sílabo, 2005.
- SEVAK, L. et al. Validation of a food frequency questionnaire to assess macro and micro nutrients intake among South Asians in the United Kingdom. *European Journal of Nutrition*, 43(3): 160-168, 2004.
- SICHERI, R. Dietary patterns and their associations with obesity in the Brazilian city of Rio de Janeiro. *Obesity Research*, 10(1): 42-48, 2002.
- SICHERI, R.; CASTRO, J. F. & MOURA, A. S. Fatores associados ao padrão de consumo alimentar da população brasileira urbana. *Cadernos de Saúde Pública*, 19, supl. 1: S47-S53, 2003.
- THOMPSON, F. & BYERS, T. Dietary Assessment Resource Manual: official publications of the American Institute of Nutrition. *Journal of Nutrition*, 124(11s): 2.245S-2.317S, 1994.
- WILLETT, W. C. *Nutritional Epidemiology*. Oxford: Oxford University Press, 1998.