

Larissa de Eleterio Lima

A comparação de técnicas multivariadas na
identificação do padrão de consumo alimentar

Niterói - RJ, Brasil

6 de Julho de 2015

Larissa de Eleterio Lima

**A comparação de técnicas
multivariadas na identificação do
padrão de consumo alimentar**

Trabalho de Conclusão de Curso

Monografia para obtenção do grau de Bacharel em Estatística
pela Universidade Federal Fluminense.

Orientador: Jony Arrais Pinto Junior

Niterói - RJ, Brasil

6 de Julho de 2015

Larissa de Eleterio Lima

**A comparação de técnicas multivariadas na
identificação do padrão de consumo alimentar**

Monografia de Projeto Final II de Graduação sob o título “*A comparação de técnicas multivariadas na identificação do padrão de consumo alimentar*”, defendida por Larissa de Eleterio Lima e aprovada em 6 de Julho de 2015, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Jony Arrais Pinto Junior

Orientador

Departamento de Estatística – UFF

Profa. Dra. Ludmilla Viana Jacobson

Departamento de Estatística – UFF

Profa. Dra. Jessica Quintanilha Kubrusly

Departamento de Estatística – UFF

Niterói, 6 de Julho de 2015

Lima, Larissa de Eleterio

A comparação de técnicas multivariadas na identificação do padrão de consumo alimentar / Larissa de Eleterio Lima; Jony Arrais Pinto Junior, orientador. Niterói, 2015. 59 f. : il.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2015.

1. Análise de padrões alimentares. 2. Análise fatorial. 3. Análise de componentes principais. 4. Análise de cluster. I. Pinto Junior, Jony Arrais, orientador. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD -

Resumo

A alimentação tem um papel significativo na identificação de doenças e tem sido largamente estudado, no entanto, a maior parte dos estudos desenvolvidos tem abordado mais frequentemente a relação de alimentos ou nutrientes isoladamente do que o efeito de padrões alimentares. De fato, os indivíduos não comem nutrientes isoladamente e existe uma relação entre os alimentos, por isso é importante avaliar os possíveis efeitos de interação entre eles.

Desse modo, esse trabalho objetiva identificar padrões alimentares utilizando três diferentes técnicas multivariadas. Esse trabalho de identificação tem com propósito servir de base para o estudo entre o perfil alimentar e sua associação com características socioeconômicas, estilo de vida e doenças crônicas.

No entanto, nem sempre a análise de todos os alimentos conjuntamente é fácil, portanto neste estudo os alimentos foram divididos em 11 grupos alimentares e foram aplicadas as técnicas de análise fatorial e análise de componentes principais em cada um dos grupos. Após isso, elegeu-se as variáveis auxiliares formadas a partir da análise de componentes principais uma vez que os perfis formados pelas duas técnicas foram semelhantes. Finalmente, através das variáveis auxiliares obtidas com a análise dos grupos alimentares, aplicou-se novamente a análise fatorial e a análise de componentes principais, além da análise de cluster, e foram obtidos os padrões alimentares, que distinguiram-se em saudável e não saudável.

Verificou-se que as três análises produziram padrões alimentares semelhantes entre si, mostrando então serem as três técnicas adequadas. No entanto, a análise de componentes principais mostrou algumas relações que as outras duas técnicas não revelaram. Relações essas que se aproximam mais dos padrões encontrados na literatura.

Palavras-chaves: **Análise de padrões alimentares, análise fatorial, análise de componentes principais, análise de cluster.**

Dedicatória

Dedico este trabalho à minha querida avó, que me ensinou a viver corajosamente a vida.

Agradecimentos

Quero agradecer à Deus que com graça irresistível me levou por um caminho incrível.

Agradecer também aos meus pais, Marcio e Eliane, que me educaram e ensinaram o valor da vida, do estudo e do trabalho honesto. À minha tia Izabel que foi essencial nas minhas escolhas pessoais.

Ao professor Jony que durante três anos me orientou com paciência e dedicação. À professora Jéssica Kubrusly que ministrou o melhor curso de inferência, e, principalmente ao professor Adrian Pizzinga que mudou completamente minha forma de enxergar o mundo e a mim mesma.

Não me esquecendo dos meus amigos que tornaram leve o que antes era insuportável. À Camila, Daiane, Júlio, Andressa e Sérgio, sem vocês esse curso não teria a mínima razão! Vocês trouxeram toda motivação e felicidade que precisava todas as manhãs. Vocês foram durante esses quatro anos meu combustível! De coração muito obrigada!

Ao Calvin, grande amigo e incentivador das grandes discussões filosóficas. À Camila, Evellyn e Rebecca que me trazem sempre de volta à simplicidade da vida. À Laninha, Thyago, Thamirys e Fabrício que com paciência compartilharam comigo do melhor que tinham em si, a essência de Cristo. À Bruna e Yuri que me ensinaram o valor do perdão e da mudança.

Aos meus líderes de ministério Ricardo e Elisa que me ensinaram a não ter medo do amanhã porque Deus já está lá. E Rodrigo e Alixe que demonstraram através de seu testemunho a plenitude de Cristo. Ao pastor Marco Antônio, cujos sermões foram minhas fiéis companhias em momentos de estudo e de reflexão.

Quero agradecer também à uma família muito especial que demonstrou muito amor e cuidado para com a minha família e me inspiraram profundamente. À Eliana, Alzeir, Airan, Luciano, Lene e o mais novo integrante: Lucas. Meu sincero agradecimento à todos vocês.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 9
2	Objetivos	p. 12
3	Materiais e Métodos	p. 13
3.1	Banco de Dados	p. 13
3.2	Análise de componentes principais	p. 14
3.3	Análise Fatorial	p. 17
3.4	Análise de cluster	p. 21
4	Análise de Resultados	p. 26
4.1	Perfis intermediários	p. 26
4.2	Padrões de consumo alimentar	p. 43
5	Conclusão	p. 50
	Referências	p. 52
	Anexo A – Lista de Siglas	p. 53

Lista de Figuras

1	Scree plot	p. 16
2	Silhouette plot	p. 25
3	Gráfico geral dos alimentos	p. 26
4	Gráficos de perfis intermediários	p. 41
5	Gráficos de outros perfis intermediários	p. 42
6	Resultados gráficos finais	p. 49

Lista de Tabelas

1	Grupo de Cereais	p. 29
2	Grupo de Leguminosas	p. 31
3	Grupo de Leite e Derivados	p. 32
4	Grupo da Carne vermelha e ovos	p. 33
5	Grupo da Carne Branca e peixe	p. 34
6	Grupo de Doces	p. 35
7	Grupo das Hortaliças	p. 36
8	Grupo das Frutas	p. 37
9	Grupo de Bebidas não alcoólicas	p. 38
10	Grupo de Bebidas alcoólicas	p. 39
11	Grupo de Fast-Food	p. 39
12	Resultados Finais	p. 43
13	Resultados análise de cluster	p. 44

1 Introdução

A análise tradicional da ingestão de alimentos geralmente procura por uma associação entre a ingestão de nutrientes individuais e doenças crônicas não transmissíveis, como as doenças cardiovasculares, diabetes e obesidade medida por meio de indicadores antropométricos. A análise de padrões alimentares pode contribuir para o entendimento desta associação, uma vez que o padrão global de alimentos que uma pessoa come é mais importante para uma dieta saudável do que se concentrar em alimentos individuais ou nutrientes individuais.

Estudos prospectivos têm analisado as associações entre a ingestão de nutrientes ou alimentos e o risco de ocorrência de doenças cardiovasculares, obesidade e outras comorbidades e outros poucos estudos têm avaliado a relação do padrão geral de dieta com o risco de ocorrência de morbidades.

Hu et al. (2004) [2] aborda a associação entre o padrão alimentar e o risco de doença cardiovascular. Neste trabalho os autores utilizam informações dietéticas recolhidas em 1984 de 71.768 mulheres com idade entre 38 e 63 anos sem um histórico de doença cardiovascular ou diabetes. Foi realizada a análise fatorial e identificou dois principais padrões alimentares Saudável e Não saudável. Calculou-se as pontuações para cada participante para cada padrão e prospectivamente examinadas suas associações com o risco de acidente vascular cerebral (AVC).

Também Gimeno et al. (2011) [10], avaliou o padrão de consumo alimentar e sua associação a variáveis socioeconômicas e doenças crônicas não transmissíveis. Este estudo teve como objetivo identificar e analisar as associações entre nutrientes ingeridos pelos moradores de Ribeirão Preto, São Paulo. Foram entrevistadas 930 pessoas, incluindo homens e mulheres com mais de 30 anos. A análise fatorial foi aplicado para identificar os padrões de consumo de alimentos, utilizando-se um questionário. Identificaram-se quatro padrões: (a)obesogênico : associado com mais atividade física, maior escolaridade e idade acima de 40 anos; (b) saudável : associada ao sexo feminino, indivíduos sem excesso de

peso, adultos mais velhos, obesidade central, mais atividade física e nível socioeconômico mais elevado; (c) misturados : sem excesso de peso, sexo feminino, e adultos jovens; e (d) Populares: associado com ausência de hipercolesterolemia e inferior renda familiar. Os resultados destacam a necessidade de incentivar a ingestão de alimentos saudáveis, a fim de evitar doenças crônicas não transmissíveis.

Conceitualmente, o exame dos padrões gerais de dieta são os que mais se aproximam do mundo real, em que as pessoas não comem nutrientes isolados, mas sim refeições que consistem em uma variedade de alimentos com combinações complexas de nutrientes.

Na literatura existem algumas formas de identificação do padrão de consumo alimentar de uma população, as mais comuns são: Análise de componentes principais e análise fatorial. Entretanto, é difícil a observação de um padrão ao se considerar todo o conjunto de dados, é proposto então, na literatura, que os alimentos sejam divididos em grupos alimentares de modo a alocar os alimentos dentro desses grupos segundo suas composições. Esta proposta objetiva superar a dificuldade de encontrar um padrão de consumo quando observadas todas as variáveis ao mesmo tempo, e assim identificar perfis intermediários de consumo e a partir destes obter o padrão alimentar.

É comum, também, na literatura a abordagem da análise de padrões alimentares em populações brasileiras, mas essas pesquisas geralmente trabalham com dados locais, de cidades ou municípios. A proposta deste trabalho é expandir os resultados encontrados em microregiões para o território nacional ou pelo menos para regiões geográficas específicas, nesse caso sudeste, sul e nordeste.

Por isso, para o recolhimento desses dados, foi aplicado um questionário que é composto por cento e trinta e oito variáveis, oriundas do Estudo Longitudinal de Saúde do Adulto. O Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil) é um estudo de coorte multicêntrico e multidisciplinar com a participação de 15.105 funcionários efetivos de seis instituições públicas de pesquisa e ensino superior, que objetiva investigar a incidência e os fatores de risco para doenças crônicas, principalmente as cardiovasculares e o diabetes.

O objetivo principal, deste trabalho, é comparar os perfis encontrados pelas diferentes técnicas multivariadas, e compreender as relações entre os nutrientes, identificando as concordâncias e discordâncias entre as técnicas. Além de prospectivamente servir como material auxiliar na avaliação do risco para doenças crônicas não transmissíveis como as doenças cardiovasculares e obesidade.

Para a identificação dos padrões de consumo alimentar pretende-se utilizar três

técnicas de análise multivariada: a análise de componentes principais, a análise fatorial e análise de cluster. Essas três técnicas foram escolhidas em particular por serem adequadas ao tratamento de variáveis quantitativas, nesse caso a medida de energia em calorias presente em cada alimento, e por transformarem grupos originais de variáveis muito grandes em subgrupos menores e descorrelatados entre si. Que permitem a possibilidade de examinar e comparar os diferentes perfis formados pelas três diferentes técnicas.

Foi utilizado o software estatístico R 2.15.2 (R Core Team(2013)) para a aplicação destas análises.

Este trabalho de conclusão de curso está organizado de forma a apresentar um capítulo introdutório em que são expostos a motivação deste estudo e os métodos que serão aplicados. Um capítulo seguinte referente aos objetivos deste trabalho, onde são expostos quais as metas propostas para este trabalho. Um terceiro para as técnicas de análise multivariada, para formalizar quais técnicas foram empregadas na análise da base de dados. Um quarto capítulo que apresenta os resultados da análise fatorial e da análise de componentes principais aplicadas em cada um dos grupos alimentares. Também um capítulo para os resultados finais obtidos, e por fim um último capítulo contendo as conclusões.

2 Objetivos

Neste trabalho, o objetivo principal é estudar as similaridades e as diferenças entre três técnicas de análise multivariada a partir dos perfis formados pelas mesmas. Os objetivos secundários de interesse são:

- Uma vez que existe uma dificuldade de trabalhar com toda a base de dados, por isso o ideal é definir grupos alimentares e alocar os alimentos de acordo com sua adequação neste mesmo grupo.
- Transformar os perfis extraídos dos grupos alimentares em variáveis auxiliares utilizando as técnicas de análise de componentes principais e a análise fatorial.
- A partir das variáveis auxiliares utilizadas, aplicar as três técnicas multivariadas e entender a associação entre os nutrientes e assim encontrar padrões alimentares.
- Escolher, se possível, uma técnica mais adequada.

3 Materiais e Métodos

3.1 Banco de Dados

Como explicitado no capítulo introdutório, diversas técnicas de análise multivariada são utilizadas para definição do padrão de consumo alimentar, como pode ser visto, por exemplo, Sichieri et al. (2003) [7], que utiliza a análise de componentes principais para entender os fatores associados ao padrão de consumo alimentar da população brasileira urbana. Neste trabalho serão utilizadas três técnicas para montar perfis alimentares: análise de componentes principais, análise fatorial e análise de cluster.

Para aplicar as três técnicas multivariadas, foram utilizados os dados dos 15.105 participantes do Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil) dispersos nas regiões brasileiras Nordeste, Sudeste e Sul. As variáveis correspondem à energia obtida em cada um dos cento e trinta e oito alimentos originalmente presentes no questionário do estudo ELSA-BRASIL.

Essa base de dados de cento e trinta e oito variáveis foi dividida em onze grupos alimentares para facilitar a interpretação das análises. O número de dimensões foram escolhidas de maneira subjetiva de acordo com a interpretabilidade obtida por cada uma delas de acordo com o especialista na área de nutrição.

O ELSA-Brasil foi concebido por grupos de pesquisa com experiência no campo da epidemiologia das doenças crônicas no Brasil, constituído por seis instituições de pesquisa e ensino superior das regiões Nordeste (UFBA), Sudeste (Fiocruz, USP, UFMG e UFES) e Sul (UFRGS).

Para a operacionalização, foram constituídos Centros de Investigação (CIs) em cada instituição executora. Na Fiocruz-Rio de Janeiro, o CI-RJ contou com o grupo de 1784 participantes voluntários, homens e mulheres entre 35 e 74 anos de idade.

Neste capítulo serão apresentados os métodos referentes a cada uma das técnicas que serão utilizadas para a definição dos perfis alimentares de interesse neste trabalho

3.2 Análise de componentes principais

Proposta por Karl Pearson (1901) [6] a análise de componentes principais procura explicar a estrutura de variâncias-convariâncias de um grupo de p variáveis a partir de p combinações lineares das mesmas. Seus objetivos gerais são a redução de dimensão dos dados e a busca por interpretabilidade.

Apesar de serem necessárias p componentes principais para reproduzir toda a variabilidade do conjunto, quase sempre, essa variabilidade pode ser bem representada por número menor de k componentes principais. Desse modo tem-se quase tanta informação a partir das k componentes principais quanto teria-se se fossem utilizadas as p componentes principais originais, Jackson (1991) [3].

A análise de componentes principais, por vezes, revela relações a princípio insuspeitas, mas também interpretáveis que uma análise mais simplificada dos dados não mostraria.

Para realizar a análise de componentes principais tomamos um vetor aleatório $\mathbf{X} = (X_1, \dots, X_p)^t$, tal que, $\text{Var}(\mathbf{X}) = \Sigma$. Extraem-se então os autovalores $\lambda_1, \dots, \lambda_p$ e os autovetores $\mathbf{e}_1, \dots, \mathbf{e}_p$ desta mesma matriz Σ . Após isso, tem-se então a matriz ortogonal $P = (\mathbf{e}_1, \dots, \mathbf{e}_p)$ e o vetor $\mathbf{Y} = P^t \mathbf{X}$ será o vetor com as componentes principais de Σ , Jonhson (2010) [4].

Uma outra alternativa à utilização da matriz de covariâncias do vetor \mathbf{X} , isto é Σ , é a obtenção das componentes principais a partir da matriz de correlações ρ , ou equivalentemente, utilizar variáveis padronizadas. Os passos para obtenção das componentes a partir da matriz ρ são análogos. Veja,

$$\mathbf{Z} = \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}), \text{ onde } \Sigma = \text{diag}[\sigma_{11}, \dots, \sigma_{pp}],$$

em que cada

$$\mathbf{Z}_i = \frac{(X_i - \mu_i)}{\sigma_i},$$

Dessa forma,

$$\text{Var}(Z_i) = \text{Var}\left(\frac{X_i - \mu_i}{\sigma_i}\right) = \text{Var}\left(\frac{X_i}{\sigma_i}\right) = \frac{\sigma_i^2}{\sigma_i^2} = 1,$$

Logo,

$$\rho_{(p \times p)} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{p1} & \cdot & \cdot & 1 \end{bmatrix}.$$

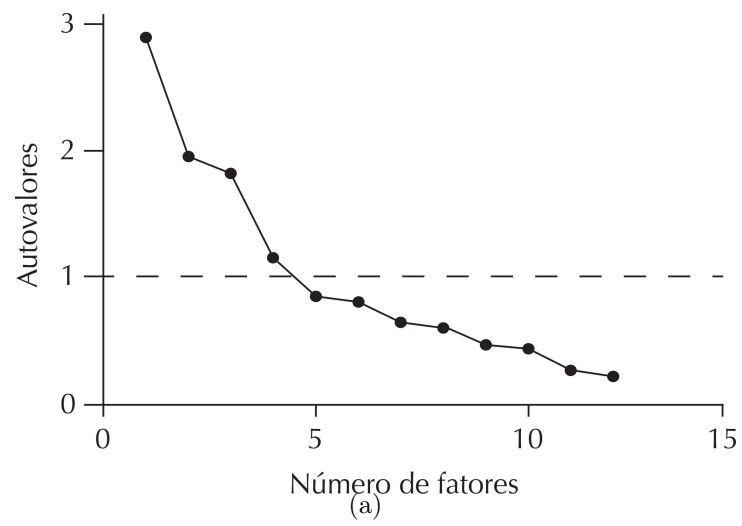
Em geral, as componentes principais obtidas a partir da matriz Σ são diferentes das obtidas a partir da matriz ρ . Quando as variáveis são medidas em escalas diferentes é mais natural estimar as componentes principais a partir da matriz de correlações ρ . Nesse estudo, as variáveis são medidas em mesma escala, isto é, em calorias.

Na prática, as componentes principais são obtidas a partir de S ou de R, isto é, a partir da matriz de covariâncias amostrais ou a partir da matriz de correlações amostrais.

Outro ponto importante a ser considerado, é a escolha do número de componentes a serem utilizados. O ideal é buscar o menor número possível de componentes. Geralmente, esse número equivale ao número de autovalores maiores que 1, isso porque o que se procura é explicar a maior parte da variabilidade dos dados com o menor conjunto de variáveis possíveis. Observe graficamente na Figura 1.

O critério de Kaiser é outra forma para determinar o número de fatores que representará o conjunto de variáveis latentes de cada grupo, Esse método determina que o número de fatores deva ser igual ao número de autovalores maiores ou iguais à média das variâncias das variáveis analisadas. Na situação em que a análise de componentes principais é feita sobre a matriz de correlação (variáveis padronizadas), esse critério corresponde à exclusão de fatores com autovalores inferiores a 1 (um). Nesses casos, o valor 1 corresponde à variância de cada variável padronizada e, conseqüentemente, esse critério descarta os fatores que tenham um grau de explicação inferior ao de uma variável isolada, ou seja, qualquer fator individual deve explicar a variância de pelo menos uma variável, se o mesmo for mantido para interpretação, e por isso os fatores com autovalores menores que 1 são desconsiderados.

Figura 1: Scree plot



3.3 Análise Fatorial

A análise fatorial é a principal e a mais antiga técnica de análise multivariada. A idéia fundamental foi proposta por Charles Spearman (1904)[9], para entender problemas relacionados à psicologia educacional, na tentativa de definir inteligência. Seu desenvolvimento e principalmente, a sua utilização, foram limitados durante muitos anos, devido à complexidade dos cálculos envolvidos. Com o advento do processamento de dados computadorizado, o uso e interesse pela análise fatorial foi renovado e retomado.

Segundo Thurstone (1947) [11], a análise fatorial é um conjunto de métodos estatísticos que, em certas situações, permite interpretar o comportamento de um número relativamente grande de variáveis observadas, em termos de um número relativamente pequeno de variáveis latentes ou fatores. As variáveis são agrupadas por meio de suas correlações, ou seja, aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo. Cada grupo de variáveis representará um fator.

A análise fatorial e a análise de componentes principais são usadas para a mesma finalidade, ou seja, expressar a informação contida numa série de observações em um menor número de dimensões, e o relacionamento entre ambas é grande. Mas existem dissimilaridades entre as técnicas, por exemplo, a análise fatorial é utilizada para variáveis que são difíceis de serem observadas diretamente enquanto a análise de componentes principais tem como foco a redução de dimensão.

Para a definição do modelo fatorial, tem-se o vetor de variáveis observáveis \mathbf{X} , com $E(\mathbf{X}) = \boldsymbol{\mu}$ e $\text{Var}(\mathbf{X}) = \Sigma$. Segundo Johnson (2001) [4], o modelo fatorial postula que \mathbf{X} é linearmente dependente de variáveis não observáveis $F_1, \dots, F_m (m < p)$, chamados fatores comuns, e p fontes adicionais de variação, chamadas erros ou, algumas vezes, fatores específicos.

Em particular, o modelo da análise de fatores pode ser escrito como:

$$X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \dots + l_{im}F_m + e_i, \quad i = 1, \dots, p \quad e \quad m < p. \quad (3.1)$$

Em Notação matricial,

$$\mathbf{X}_{p \times 1} - \boldsymbol{\mu}_{p \times 1} = L_{p \times m} F_{m \times 1} + \boldsymbol{\varepsilon}_{p \times 1},$$

em que,

$$L = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ l_{p1} & \cdot & \cdot & l_{pm} \end{bmatrix},$$

$$F = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ l_{p1} & \cdot & \cdot & l_{pm} \end{bmatrix},$$

$$E = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \cdot \\ \cdot \\ \varepsilon_{p1} \end{bmatrix}.$$

Os coeficientes l_{ij} são chamados cargas fatoriais da i -ésima variável do j -ésimo fator. Dessa forma, a matriz \mathbf{L} é a matriz de cargas fatoriais. Note que o i -ésimo fator específico ε_i , é associado à i -ésima variável resposta X_i . O modelo fatorial implica certas relações de covariância a respeito de \mathbf{F} e $\boldsymbol{\varepsilon}$ que podem ser checadas:

- $E(\mathbf{F}) = \mathbf{0}_{(m \times 1)}$, $\text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}^t) = \mathbf{I}_{(m \times m)}$,
- $E(\boldsymbol{\varepsilon}) = \mathbf{0}_{(p \times 1)}$, $\text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t) = \Psi_{(p \times p)} = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \Psi_p \end{bmatrix}$,
- $\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}, \mathbf{F}^t) = \mathbf{0}_{(p \times m)}$.

Essas suposições e as relações descritas no modelo anterior constituem o modelo de fatores ortogonais. O modelo de fatores ortogonais implica uma estrutura de covariâncias para \mathbf{X} , os passos para obter essa estrutura podem ser expressos da seguinte forma:

Primeiro manipular a seguinte expressão,

$$\begin{aligned}
(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t &= (\mathbf{LF} + \boldsymbol{\varepsilon})(\mathbf{LF} + \boldsymbol{\varepsilon})^t, \\
(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t &= (\mathbf{LF} + \boldsymbol{\varepsilon})((\mathbf{LF})^t + \boldsymbol{\varepsilon}^t), \\
(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t &= (\mathbf{LF})(\mathbf{LF})^t + \boldsymbol{\varepsilon}(\mathbf{LF})^t + (\mathbf{LF})\boldsymbol{\varepsilon}^t + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t.
\end{aligned}$$

Verificada a igualdade acima, substitui-se na seguinte expressão:

$$\begin{aligned}
\Sigma &= Cov(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t, \\
\Sigma &= \mathbf{LE}(\mathbf{FF}^t)\mathbf{L}^t + E(\boldsymbol{\varepsilon}\mathbf{F}^t)\mathbf{L}^t + \mathbf{LE}(\mathbf{F}\boldsymbol{\varepsilon}^t) + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t), \\
\Sigma &= \mathbf{LL}^t + \Psi.
\end{aligned}$$

Dessa forma a estrutura de covariâncias do modelo ortogonal podem ser descritas:

1. $Cov(\mathbf{X}) = \mathbf{LL}^t + \Psi$.

Em que,

$$\begin{aligned}
Var(X_i) &= l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \Psi_i, \\
Cov(X_i, X_k) &= l_{i1}l_{k1} + l_{i2}l_{k2} + \dots + l_{im}l_{km}.
\end{aligned}$$

2. $Cov(\mathbf{X}, \mathbf{F}) = \mathbf{L}$.

Em que,

$$Cov(X_i, F_j) = l_{ij}.$$

Estabelecidas as estruturas de covariâncias do modelo de fatores ortogonais, e verificadas as suposições do vetor \mathbf{F} , o próximo passo é estimar as cargas dos fatores, no caso $\mathbf{L} = (l_1, \dots, l_m)^t$. Existem três métodos comuns para se estimar as cargas fatoriais, são elas o método das componentes principais, o método dos fatores principais e o método da máxima verossimilhança. A seguir vai-se discorrer a respeito de cada uma delas.

- Método das componentes principais:

O método das componentes principais pode ser sistematizado da seguinte forma:

1º passo) Extrair os autovalores e autovetores normalizados da matriz de correlações ρ . A saber: $(\lambda_i, \varepsilon_i)$, $i = 1, \dots, p$

2º passo) Selecionar os m autovalores e autovetores normalizados da matriz de correlações maiores que 1.

3º passo) Estimar as matrizes L e Ψ .

A saber: $\mathbf{L} = (\lambda_1 \varepsilon_1^2, \dots, \lambda_m \varepsilon_m^2)$ e $\Psi = \text{diag}(\rho - \mathbf{L}\mathbf{L}^t)$.

- Método dos fatores principais:

Este método se comporta como uma modificação do citado anteriormente. Este método objetiva um refinamento das estimativas de L e Ψ obtidos pelo método das componentes principais.

Se a matriz de correlação for adequadamente descrita pelo modelo fatorial $\rho = \mathbf{L}\mathbf{L}^t + \Psi$, Onde: $\mathbf{L}\mathbf{L}^t = \rho - \Psi =$

$$\begin{bmatrix} h_1^2 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & h_2^2 & \dots & \rho_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{p1} & \cdot & \cdot & h_p^2 \end{bmatrix}$$

Em que,

$h_i^2 = 1 - \Psi_i$, $i = 1, \dots, p$ são as communalidades.

E os m fatores comuns podem ser usados para determinar os elementos fora da diagonal principal de ρ e as communalidades da diagonal, ou seja, $\rho_{ii} = 1 = h_i^2 + \Psi$. Se a contribuição do fator específico Ψ_i for removida da diagonal, ou seja, se o valor um for substituído por h_i^2 , a matriz resultante será $\rho - \Psi = \mathbf{L}\mathbf{L}^t$.

Do ponto de vista algébrico, o método tem por base obter um conjunto de fatores, de modo que o mais importante fator comum (fator principal) seria o fator comum, com o máximo de contribuição para a communalidade total, o segundo mais importante (segundo fator principal) seria o fator comum com a segunda maior contribuição para a communalidade total, e assim por diante, até que toda communalidade tenha sido analisada.

- Método da máxima verossimilhança:

O método da Máxima Verossimilhança para Análise Fatorial foi introduzido em 1940, e não era muito usado anteriormente, por suas dificuldades computacionais. Atualmente, no entanto, já se encontram procedimentos rápidos e eficientes para a obtenção dos estimadores por este método. Mas a principal vantagem de utilizar o Método da

Máxima Verossimilhança para Análise Fatorial é, talvez, a possibilidade de se desenvolverem testes de hipóteses, com o objetivo de testar a adequacidade do modelo, o que não é possível através dos outros dois métodos apresentados anteriormente, devido às suas características não estatísticas.

Além das suposições habituais do modelo fatorial, é necessário a suposição de que os vetores aleatórios \mathbf{F} e $\boldsymbol{\varepsilon}$ têm distribuição normal multivariada com vetores de média zero e com matrizes de covariâncias Ψ . Das suposições habituais e da suposição de normalidade, segue que \mathbf{F} e $\boldsymbol{\varepsilon}$ são mutuamente independentes. Como \mathbf{X} é expresso em termos de \mathbf{F} e $\boldsymbol{\varepsilon}$, tem-se que o vetor das variáveis observáveis é também normal, com média zero e matriz de variâncias e covariância Σ . Portanto, as condições para a aplicação do Método da Máxima Verossimilhança na Análise Fatorial são:

- a) X tem uma distribuição normal multivariada, com vetor de médias zero e matriz de covariância Σ ;
- b) $X = \Lambda F + \varepsilon$, onde $E(F) = E(\varepsilon) = 0$, $Var(F) = I$, $Var(\varepsilon) = \Psi = \text{diag}(\Psi^2)$, $Cov(F, \varepsilon) = 0$;
- c) $\Sigma = \Lambda \Lambda^2 + \Psi$;
- d) As colunas de X são amostras aleatórias de tamanho n das variáveis em questão e, portanto, a matriz de covariância amostral S tem distribuição de Wishart.

O caminho a ser tomado será o de maximizar a função de verossimilhança de X , com respeito aos m elementos de Λ e os p elementos de Ψ da seguinte forma. Admitindo que X tenha distribuição normal com média μ e variância Σ , a função de verossimilhança será da seguinte forma.

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}} \frac{1}{|\Sigma|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum (X_i - \mu)^t \Sigma^{-1} (X_i - \mu)\right\}$$

Essa maximização é feita em termos numéricos para um valor de m fixo. A vantagem do método da máxima verossimilhança é a produção de estimativas mais precisas. A dificuldade do método é a suposição de normalidade.

3.4 Análise de cluster

Análise de Cluster são agrupamentos homogêneos que basicamente seguem uma sequência de regras, ou algoritmo, para agrupar objetos sem inferência de probabilidade

a priori dos grupos. A partir de um conjunto de unidades amostrais, os quais são caracterizados segundo variáveis, obtém-se um algoritmo que possibilite reunir os indivíduos, tal que exista homogeneidade dentro do grupo e heterogeneidade entre os grupos.

Muitos algoritmos de clusterização se aplicam em dados organizados numa matriz de dissimilaridade, onde o elemento da coluna j e linha i da matriz é o número $d(i, j)$ representando a distância entre os objetos i e j . Veja,

$$\begin{bmatrix} 0 & \cdot & \cdot & \cdot \\ d(2, 1) & 0 & \cdot & \cdot \\ d(3, 1) & d(3, 2) & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ d(n, 1) & \cdot & \cdot & 0 \end{bmatrix}$$

Para que uma função d seja uma distância é necessário e suficiente que as seguintes condições sejam satisfeitas, para quaisquer objetos i, j, k :

1. $d(i, j) \geq 0$,
2. $d(i, i) = 0$,
3. $d(i, j) = d(j, i)$ (simetria),
4. $d(i, j) \leq d(i, k) + d(k, j)$ (desigualdade triangular).

A propriedade (1) implica que todos os elementos da matriz de dissimilaridade são não negativos, a propriedade (2) implica que a diagonal da matriz de dissimilaridade é formada por zeros. A propriedade (3), por sua vez, implica que a matriz de dissimilaridade é simétrica com relação a diagonal e por isso, só registramos nela os elementos abaixo da diagonal.

Assim, qualquer função que satisfaz as quatro propriedades acima é chamada de distância. As mais importantes funções nesta categoria são:

- Distância Euclidiana: $\sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$,
- Distância de Manhattan: $|x_{i1} - x_{j1}| + \dots + |x_{ip} - x_{jp}|$,
- Distância de Minkowski: $\sqrt[q]{(x_{i1} - x_{j1})^q + \dots + (x_{ip} - x_{jp})^q}$,

onde $q \geq 1$. Logo, a distância de Minkowski generaliza tanto a distância euclidiana (caso especial onde $q = 2$) quanto a distância de Manhattan (caso especial onde $q = 1$).

Para ilustrar como os agrupamentos são formados na análise de cluster, considere os objetos como pontos em um espaço p dimensional, com cada uma das p variáveis, representado por um dos eixos do espaço. Um sistema de p coordenadas dimensionais é definido agora no espaço pelos valores das variáveis por objeto. Pode-se, então, definir clusters como regiões contínuas que aparecem na massa relativamente grande de pontos no espaço, isto é, regiões com alta densidade de pontos, separadas de outras por regiões com pequena massa relativa (baixa densidade de pontos).

Para encontrar os agrupamentos naturais de indivíduos é possível utilizar técnicas hierárquicas, técnicas não hierárquicas, agrupamentos fuzzy, redes neurais, entre outras. Neste caso, serão utilizados os métodos não-hierárquicos, ou por particionamento, para avaliar a formação dos perfis.

O método por particionamento foi desenvolvido para agrupar elementos em K grupos, sendo K a quantidade de grupos definida previamente. Nem todos valores de K apresentam grupos satisfatórios, sendo assim, aplica-se o método várias vezes para diferentes valores de K , escolhendo os resultados que apresentem melhor interpretação dos grupos ou uma melhor representação gráfica. A idéia central dos métodos por particionamento é escolher uma partição inicial dos elementos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição.

Os métodos por particionamento mais conhecidos são o método k-médias e o método k-medóides.

O método k-means toma um parâmetro de entrada, K , e particiona um conjunto de N elementos em K grupos. Algumas características desse método são:

- Sensibilidade a ruídos, uma vez que um elemento com um valor extremamente alto pode distorcer a distribuição dos dados;
- Tendência a formar grupos esféricos;
- O número de grupos é o mesmo durante todo o processo;
- Inadequado para descobrir grupos com formas não convexas ou de tamanhos muito diferentes.

O método k-medoid utiliza o valor médio dos elementos em um grupo como um ponto referência, chamado de medóide. Esse é o elemento mais centralmente localizado em um grupo. A estratégia básica é encontrar K grupos em N elementos e, arbitrariamente, encontrar um elemento representativo (medóide) para cada grupo. Cada elemento

remanescente é agrupado com o medóide ao qual ele é mais similar. A estratégia, então, iterativamente, troca um dos medóides por um dos não medóides enquanto a qualidade do agrupamento resultante é melhorada.

Algumas características desse método são:

- Independente da ordem, os resultados serão os mesmos;
- Tendência a encontrar grupos esféricos;
- Processamento mais custoso que o k-means;
- Não aplicável à grandes bases de dados, pois o custo de processamento é alto;
- Mais robusto do que o k-means na presença de ruídos porque o medóide é menos influenciado pelos ruídos do que a média.

Uma forma de otimizar o método k-medoid para grandes bases de dados é considerar uma porção dos dados como uma amostra representativa, e escolher os medóides dessa amostra. Se a amostra é selecionada aleatoriamente, ela deverá representar bem o conjunto de dados originais, apresentando bons resultados

Para validar os agrupamentos produzidos pela análise de cluster, é comum a utilização de um recurso gráfico o silhouette. Por definição considera-se que os dados foram agrupados por meio de qualquer técnica, como k-médias.

E para cada i , $a(i)$ será a dissimilaridade média de i , com todos os outros dados dentro do mesmo cluster. Podemos interpretar $a(i)$ como o quão bem i é atribuído ao seu cluster.

Defina também $b(i)$ como a menor dissimilaridade média de i a qualquer outro cluster, de que não pertence. O cluster com a menor dissimilaridade média é dito ser o aglomerado vizinho do i porque é o próximo melhor cluster que se ajusta para o ponto i .

Dessa forma, pode-se definir uma silhouette:

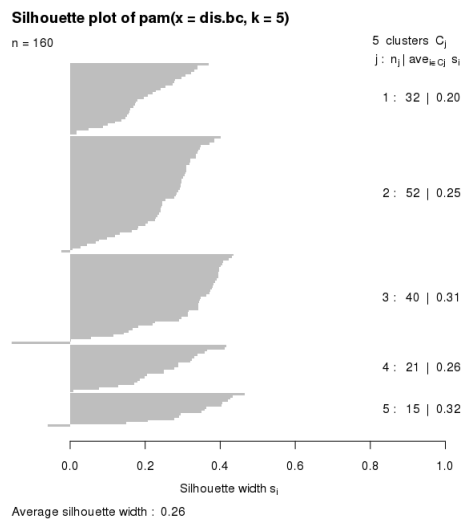
$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]}, \text{ de forma que } 1 \leq s(i) \leq 1$$

Esse método auxilia na avaliação da qualidade do ajuste. Quanto mais próximo de 1 mais bem ajustado os elementos estão dentro do cluster. Como componentes do silhouette tem-se:

- O grau de ajustamento de cada um dos elementos dentro de um cluster, destacados no gráfico como $n_j | ave_{i \in C_j}$. Que mede se os elementos dentro do clusters estão bem ajustados.
- O grau de ajustamento de todos os clusters dentro do espaço, destacados no gráfico como *Average silhouette width*. Que mede o quanto os clusters estão bem definidos no espaço. Se a análise feita está bem ajustada.

Na Figura 2 tem expressos os valores do grau de ajustamento de cada cluster, á direita do gráfico. Nessa figura, cada barra acinzentada corresponde á uma unidade amostral, seja ela variável ou indivíduo. Na parte inferior do gráfico, tem-se também o grau de ajustamento de todos os clusters no espaço.

Figura 2: Silhouette plot



No próximo capítulo serão verificados os resultados da aplicação de cada uma das técnicas de análise multivariada citadas nesse capítulo.

4 Análise de Resultados

4.1 Perfis intermediários

Teoricamente, os padrões de consumo permitem o entendimento de construções sociais e econômicas. Portanto, é da natureza dos padrões de consumo alimentar a sua maior ou menor capacidade de geração da saúde. Dessa forma, o padrão de consumo de alimentos expressa situações reais de disponibilidade de alimentos e de condições diferenciadas de inserção ou não das populações nos diferentes cenários sociais, segundo Schieri et al. (2003) [8].

Entretanto, existem diversas formas de identificar o padrão de consumo alimentar de uma população. Uma das quais consiste em avaliar a composição e a relação de todos os alimentos entre si, isto é, fora de seus grupos alimentares. Veja na Figura 4.1 a representação deste método.

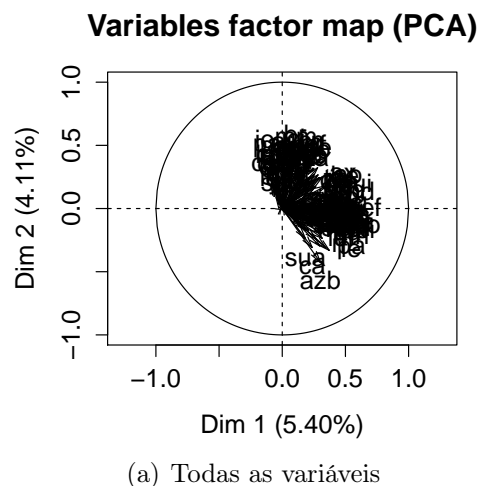


Figura 3: Gráfico geral dos alimentos

Note que, o entendimento do padrão de consumo alimentar neste trabalho tem por objetivo refletir as preferências de consumo alimentar de indivíduos para assim reconhecer

padrões de consumos que posteriormente possam ser associados a incidência de doenças crônicas não transmissíveis, como diabetes e obesidade. No entanto, na Figura 4.1 não é clara a distinção de padrão algum, impossibilitando a identificação de um padrão de dieta. Por isso, é recomendável a divisão dos alimentos, componentes da dieta individual, em grupos alimentares, Olinto (2009) [5].

Para a identificação do padrão de consumo alimentar, as cento e trinta e oito variáveis que caracterizavam o consumo dos indivíduos foram agrupadas em onze grupo alimentares para facilitar a interpretação das componentes principais (e fatores) extraídas.

Os grupos alimentares estipulados foram:

- Leguminosas: lentilha, feijão, feijoada, milho, noz.
- Cereais: arroz branco, arroz integral, aveia, batata, batata frita, farinha, farofa, macarrão instantâneo, milho verde, pão, pão light, pão integral, bolo sem recheio, biscoito doce sem recheio, biscoito salgado
- Leite e derivados: iogurte, iogurte light, leite desnatado, leite integral, leite semi-desnatado, manteiga, margarina, queijos amarelos, queijos brancos, requeijão, requeijão light.
- Doces: bala, barra de cereal, biscoito doce com recheio, bolo com recheio, chocolate, chocolate em pó, gelatina, geléia, mel, pão doce, picolé de fruta, pudim, salada de frutas com complementos, Sorvete
- Bebidas: água de coco, café com açúcar, café com adoçante, café sem açúcar, chá mate com açúcar, chá mate com adoçante, chá mate sem açúcar, extrato de soja, refrigerante, refrigerante light, suco artificial com açúcar, suco artificial com adoçante, suco artificial sem açúcar, suco industrializado com açúcar, suco industrializado com adoçante, suco industrializado sem açúcar, suco natural com açúcar, suco natural com adoçante, suco natural sem açúcar.
- Bebidas alcoólicas: bebidas destiladas, cerveja, vinho branco, vinho tinto
- Fast Food: cachorro quente, hambúrguer, maionese, maionese light, pão de queijo, pipoca, pizza, salgado assado, salgado frito, sopa instantânea
- Carve Vermelha e ovo: frios light, bacon, bucho, fígado, carne bovina com osso, carne bovina sem osso, carne de porco, estrogonofe, ovo frito e ovos cozidas.

- Carne branca e peixe: Frango, frango frito, peito frango, camarão, carangueijo, peixe, peixe frito, sardinha
- Frutas: abacate, abacaxi, banana, caqui, goiaba, laranja, maçã, mamão, manga, melancia, melão, morango, pêssego, salada de frutas sem complementos, uva
- Hortaliças: abobrinha, abóbora, alface, alho, beterraba, brócolis, cebola, cenoura, chicória, couve, couve flor, quiabo, sopa de legumes, tomate, vagem.

No apêndice encontram-se as abreviações feitas para cada alimento para que o nome da variável não interferisse na visualização gráfica.

Dessa forma, uma vez estabelecidos os grupos alimentares, é recomendável aplicar-se uma análise que reduza dimensões em cada um dos grupos alimentares formados para extrair subperfis interpretáveis como visto por Almeida (2010) [1]. Dessa forma, é possível visualizar simultaneamente a correlação que os alimentos dentro de um grupo alimentar tem entre si e sua correlação com o grupo alimentar como um todo. Além disso, obtendo o valor dos coeficientes de cada variável é possível identificar qual variável está mais associada com cada dimensão, e quanto maior for o valor do coeficiente de um alimento maior será a contribuição daquele alimento no perfil de consumo formado por aquela dimensão.

Por isso, após a divisão dos alimentos em grupos foi aplicada a análise de componentes principais e a análise fatorial em cada um dos grupos alimentares formados para além de seguir passos consoantes aos sugeridos pela literatura, comparar os perfis intermediários formados por meio das duas técnicas a fim de verificar se os resultados obtidos são semelhantes entre si. Vale ressaltar que a escolha do número de componentes principais e de fatores utilizados, para identificar os perfis formados, foi determinada pelo especialista em nutrição Dr. Paulo Cesar Castro, de acordo com a necessidade de interpretação de cada grupo alimentar. Isto é, algumas componentes principais (e fatores) acrescentavam em termos de variação explicada, mas não eram interpretáveis e por isso não foram consideradas.

Cereais

Na nutrição é conhecido que todos os tipos de cereias são boas fontes de vitaminas e minerais e têm naturalmente pouca gordura. Os cereais são os grãos comestíveis de diversas plantas como trigo, cevada, milho. Para este grupo alimentar os coeficientes obtidos por meio da análise fatorial e da análise de componentes principais estão apresentados na tabela 1:

Tabela 1: Grupo de Cereais

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
azb	0,24	-0,55	0,51	-0,44
azi	0,03	0,43	-0,20	0,54
av	0,04	0,33	-0,15	0,49
bt	0,44	-0,04	0,49	0,22
btf	0,31	-0,16	0,44	-0,02
fa	0,35	-0,07	0,43	0,14
fo	0,39	0,01	0,42	0,27
mc	0,31	-0,07	0,39	0,12
mci	0,15	-0,08	0,22	-0,002
md	0,43	0,11	0,40	0,39
pal	-0,11	0,28	-0,29	0,35
pa	0,34	-0,29	0,52	-0,16
pai	-0,01	0,39	-0,23	0,53
po	0,38	-0,09	0,46	0,12
bds	0,23	0,02	0,26	0,20
bs	0,27	-0,01	0,32	0,19
bo	0,34	-0,05	0,42	0,16

Na tabela 1 os resultados estão dispostos da seguinte forma, na primeira coluna os alimentos, nas segunda e terceira colunas os escores da análise fatorial e nas quarta e quinta colunas os coeficientes da análise de componentes principais. Isto é, negrito estão sinalizados os valores mais significativos em cada fator ou componente. Dessa forma, será destacado quais as variáveis mais expressivas em cada fator ou componente. Nas tabelas que se seguem, tanto na formação de perfis quanto na formação de padrões, a disposição será sempre semelhante a essa.

Outro detalhe importante a ser destacado é que, como citado anteriormente, a escolha do número de fatores ou componentes foi determinado de acordo com a interpretabilidade. No entanto, em cada um dos grupos alimentares, ao aplicar-se a técnica de análise de componentes principais, o gráfico produzido por essa análise mostra o valor da variação explicada pela primeira e pela segunda componente. Como mostra Figura 4.

Perceba que a primeira componente principal foi caracterizada pelo consumo de arroz branco, batata, batata frita, farinha, farofa, macarrão, mandioca, pão francês, polenta e bolo. Sendo, então, a primeira componente principal responsável pela formação de um grupo pouco restrito com seu consumo de calorias.

Já a segunda componente principal formou um grupo de indivíduos mais restritos

em seu consumo de cereais. Indivíduos com altas pontuações na segunda componente principal consomem arroz integral, aveia, pão light e pão integral. Indicando uma clara categorização entre indivíduos com uma dieta saudável e indivíduos com uma alimentação pouco criteriosa.

A análise fatorial e análise de componentes principais produziram resultados semelhantes,

Uma observação de interesse a ser feita é, a partir de agora as componentes principais e os fatores receberão uma nomenclatura que será definida de acordo com o seguinte critério, as três primeiras letras do grupo alimentar a que a componente/fator pertence e as abreviação da técnica utilizada. Por exemplo, no grupo alimentar de cereais extraímos dois fatores e duas componentes, portanto tem-se quatro variáveis *cerfa1* (primeiro fator obtido pela análise fatorial), *cerfa2* (segundo fator obtido pela análise fatorial), *cerpc1* (primeira componente obtida pela análise de componentes principais), *cerpc2* (segunda componente obtida pela análise de componentes principais).

Leguminosas

As leguminosas são um grupo de alimentos compostas pelo feijão, lentilha, pela ervilha e pelo grão-de-bico. A sua importância na alimentação se dá por conter alimentos ricos em ferro, além de serem fontes de vitaminas do complexo B, potássio e fibras. Para este grupo observamos os seguintes resultados.

Tabela 2: Grupo de Leguminosas

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
mi	0,07	0,30	0,40	0,49
fe	0,62	-0,09	-0,54	0,55
fei	0,32	0,05	-0,35	0,66
lt	0,01	0,40	0,54	0,41
no	-0,09	0,30	0,61	0,19

A primeira componente principal (legpc1) caracterizou o consumo de feijão e feijoada acrescentando escores negativos para indivíduos com esse perfil, e escores positivos para indivíduos com consumo da lentilha, milho e noz. O grupo de leguminosas, a princípio, é o menos auto-explicativo, pois consiste de em dois grupos que não distinguem nenhum hábito alimentar particular ilustrando apenas o óbvio que quem consome feijão consome também feijoada e que os consumidores de milho apresentam consumo significativo de lentilha.

Note que a análise de componentes principais obteve resultados semelhantes aos encontrados na análise fatorial. Mesmo que a primeira componente principal pontue negativamente indivíduos com o perfil encontrado no primeiro fator. Isto é, consumidores de feijão e feijoada no primeiro fator teriam altas pontuações positivas, e na primeira componente principal teriam altas pontuações negativas.

Leite e derivados

O leite e seus derivados, como queijo, iogurte, requeijão, entre outros, são importantes fontes de proteínas, vitaminas e a principal fonte de cálcio da alimentação. Para este grupo foram obtidos os resultados expressos na tabela 3:

Na Tabela 3 tem-se coeficientes mais expressivos no consumo de iogurte light, leite integral e requeijão na primeira componente principal (leipc1), sendo esse grupo caracterizado por indivíduos em dieta saudável. O segundo fator (leipc2) reflete um perfil de indivíduos preocupados com uma alimentação menos saudável porque apesar de ingerirem iogurte light, por exemplo, seguem sendo a alimentos calóricos como a

Tabela 3: Grupo de Leite e Derivados

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
io	0,44	0,02	-0,60	0,17
iol	-0,13	0,01	0,40	0,32
led	0,41	-0,16	-0,51	0,18
lei	-0,38	-0,16	0,59	-0,04
les	0,04	0,76	-0,04	0,10
mt	-0,04	-0,01	0,24	0,40
mr	-0,11	-0,03	0,17	-0,21
qa	0,03	0,01	0,24	0,55
qb	0,32	-0,01	-0,26	0,54
rq	-0,07	0,01	0,37	0,45
rql	0,35	0,01	-0,49	0,25

manteiga, queijo amarelo e o requeijão. Ou seja, um perfil de consumidores menos seletivo em sua dieta de laticíneos.

A análise fatorial formou perfis um pouco diferentes, o primeiro fator caracterizou indivíduos consumidores de iogurt, leite desnatado e requeijão light, enquanto o segundo fator caracterizou os consumidores de leite desnatado. Nesse caso, a análise fatorial não distinguiu um padrão saudável ou não saudável diferente da análise de componentes principais.

Carne vermelha e ovo

O grupo das carnes e ovos é composto por alimentos ricos em proteínas de fonte animal, além de ferro, zinco, vitaminas. Para este grupo obteve-se os seguintes resultados os resultados expressos na tabela 4:

Na Tabela 4 verifica-se que a primeira componente principal (carvpc1) reflete um grupo de indivíduos que consome bacon, bucho, fígado, carne bovina com osso, carne bovina sem osso, carne de porco, ovo frito, ovo cozido e linguiça. A segunda componente principal (carvpc2) apresenta as mesmas variáveis, no entanto com coeficientes de valor negativo, indicando serem inversamente proporcionais aos indivíduos caracterizados pela primeira componente principal. Esse fator caracterizou-se, também, pelo consumo de frios light e estrogonoff. Indivíduos com altos escores nessa componente encaixam-se no grupo de manutenção de uma dieta saudável.

Verifique que segue a concordância dos perfis formados pela análise fatorial e pela análise de componentes principais.

Tabela 4: Grupo da Carne vermelha e ovos

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
fl	-0,08	-0,10	-0,03	0,64
ba	0,15	0,47	0,49	0,02
bu	0,48	0,18	0,54	-0,28
fi	0,48	0,08	0,47	-0,34
cbo	0,36	0,21	0,49	-0,24
bos	0,14	0,24	0,39	0,25
cp	0,09	0,48	0,46	0,10
es	0,05	0,11	0,23	0,53
of	0,41	0,20	0,55	-0,08
oc	0,37	0,05	0,42	-0,12
li	0,20	0,37	0,51	0,14
pr	0,17	0,27	0,46	0,43

Carne Branca e Peixe

As carnes consideradas brancas são as provenientes de peixes e aves. As carnes brancas são fontes de ômega 3 e ômega 6, além de ácidos graxos, substâncias que melhoram o funcionamento do organismo e contribuem para a boa nutrição.

Tabela 5: Grupo da Carne Branca e peixe

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
fr	0,15	0,49	0,55	-0,41
frf	0,06	0,64	0,51	-0,48
pfr	0,14	0,31	0,44	-0,31
cm	0,16	0,03	0,45	0,62
cg	0,08	0,08	0,41	0,57
pe	0,83	0,11	0,65	0,15
pef	0,41	0,30	0,63	-0,06
sr	0,18	0,09	0,34	0,13

A análise de componentes principais distinguiu em sua primeira componente (frpepc1) o peixe cozido, peixe frito, peito de frango, frango frito, além de camarão e de carangueijo. Sua segunda componente (frpepc2) caracterizou um grupo formado por camarão e carangueijo. Interessante que, nesse subgrupo não é possível distinguir uma comparação entre um grupo mais saudável que outro. Verifica-se apenas grupos de pessoas que preferem comer peixe, outro que preferem comer frango e um outro que consome regularmente frutos do mar.

Já na análise fatorial o primeiro fator (frpefa1) discriminou indivíduos que consomem peixe cozido e peixe frito. O segundo fator (frpefa2) discriminou indivíduos que consomem frango e frango frito e peito de frango.

Doces

Os doces contêm açúcar, que é um tipo de carboidrato simples. Se esse açúcar for consumido em excesso pode provocar problemas à saúde, como o aumento de peso, aumento do colesterol sanguíneo, dores abdominais, entre outros. Como este estudo é parte de uma análise de morbidade a avaliação do padrão de consumo de doces é muito importante na determinação de fatores de risco.

Da análise de componentes principais são extraídos dois grupos. Um primeiro (docpc1) formado por bala, biscoito doce recheado, bolo recheado, chocolate, chocolate em pó, pudim, pirulito e sorvete. Sendo esse grupo caracterizado por seu consumo de

Tabela 6: Grupo de Doces

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
bl	0,30	0,15	0,43	-0,16
bc	0,04	0,08	0,18	0,44
bdr	0,37	0,04	0,38	-0,40
bor	0,40	0,08	0,52	-0,15
cho	0,52	0,05	0,55	-0,22
cep	0,26	0,05	0,36	-0,14
ge	0,05	0,12	0,25	0,43
gl	0,11	-0,01	0,35	0,45
me	0,01	0,00	0,21	0,53
pd	0,16	0,07	0,29	-0,07
pi	0,06	0,53	0,34	0,01
pu	0,36	0,11	0,57	0,09
sfc	0,13	0,12	0,29	0,05
sv	0,30	0,43	0,54	-0,07

doces industrializados, indicando ser um grupo de indivíduos pouco preocupados com dieta. Da segunda componente principal (docpc2) formado pela geléia, gelatina e mel. Como o primeiro grupo destacou-se pelo consumo de doces calóricos e industrializados expressando um grupo pouco preocupado com dieta, este segundo grupo consumidor de geléia, gelatina e mel mostra-se mais preocupado em manter hábitos alimentares saudáveis e equilibrados.

Note que houve um pequena discordância entre o perfil formado pela segunda componente principal e o perfil formado pelo segundo fator. O segundo fator remove o peso que o biscoito sem recheio, a gelatina, a geléia e o mel tem e atribui um valor expressivo ao pirulito e ao sorvete no segundo fator.

Hortaliças

Popularmente conhecidas como verduras e legumes, as hortaliças são alimentos ricos em vitaminas, minerais e fibras. O Brasil é um país que tem o privilégio de cultivar centenas de espécies de hortaliças e de oferecer boa parte delas a baixo custo. Cheirosos, coloridos, saborosos, esses vegetais ocupam lugar de destaque nas bancas das feiras e gôndolas de supermercados, o ano todo.

Tabela 7: Grupo das Hortaliças

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
ao	0,29	0,19	0,42	-0,36
ar	0,37	0,33	0,58	-0,16
al	0,07	0,60	0,48	0,64
be	0,36	0,17	0,46	-0,23
br	0,58	0,08	0,54	-0,19
cn	0,32	0,33	0,54	-0,04
ci	0,22	0,44	0,52	0,36
co	0,28	0,39	0,54	0,18
cof	0,66	0,04	0,56	-0,28
so	0,19	0,13	0,29	-0,23
to	0,08	0,46	0,42	0,52
vg	0,41	0,25	0,55	-0,21

Veja que na Tabela 7 as hortaliças distinguiram-se dois grupos. Um formado pela abóbora, abobrinha, alface, berinjela, brócolis, cenoura, chicória, couve, couve-flor e vagem, e outro formado pela alface, chicória e tomate.

No grupo das hortaliças a análise fatorial e a análise de componentes principais também formaram perfis próximos entre si, apesar de o segundo fator da análise fatorial acrescentar mais hortaliças do que a segunda componente principal descreveu.

Frutas

A palavra fruta tem muitos significados diferentes, dependendo do contexto em que se utiliza. Em botânica, um fruto é o ovário e sementes amadurecidas de uma planta em flor, e em muitas espécies, o fruto incorpora o ovário e os tecidos circundantes. Os frutos são assim, em botânica, os meios pelos quais as plantas florescem e disseminam as sementes. Na cozinha, quando os alimentos são denominados frutos, o termo é mais utilizado para designar os frutos de plantas comestíveis, doces e carnudos, e que incluem, como exemplo, ameixas, maçãs e laranjas.

Tabela 8: Grupo das Frutas

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
ab	0,21	0,07	0,32	-0,26
ax	0,16	0,36	0,50	0,18
bn	0,59	0,09	0,44	-0,55
cu	0,20	0,11	0,45	0,01
go	0,26	0,07	0,38	-0,27
la	0,46	0,15	0,48	-0,38
ma	0,31	0,19	0,44	-0,16
mm	0,18	0,30	0,43	0,12
mg	0,28	0,15	0,45	-0,23
ml	0,16	0,45	0,48	0,12
mao	0,05	0,52	0,46	0,33
mo	0,02	0,13	0,34	0,41
ps	0,11	0,15	0,44	0,26
sfs	0,06	0,17	0,24	0,25
uva	0,12	0,17	0,46	0,25

No grupo das frutas não há uma clara distinção de perfis. O primeiro formado pela banana, maçã, laranja, abacaxi, mamão, melão, melancia, caqui, morango, pêssego, uva e manga. Uma mistura de frutas de época e frutas populares e por isso caracterizando um grupo de consumidores regulares de frutas e talvez até mesmo consumidores que usem as frutas como substitutivos para doces industrializados. A segunda componente principal caracterizou apenas o consumo de morango.

Note que o primeiro fator caracterizou o consumo de laranja, maçã e banana. E o segundo fator caracterizou o consumo de abacaxi, melão e melancia.

Bebidas não alcoólicas

Tabela 9: Grupo de Bebidas não alcoólicas

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
ca	-0,48	-0,50	-0,62	0,11
cd	0,55	0,04	0,57	0,31
csa	-0,11	0,45	0,15	-0,57
cha	-0,16	-0,10	-0,31	0,05
chd	0,31	-0,04	0,36	0,29
chs	-0,05	0,29	0,15	-0,47
ex	0,02	0,06	0,09	-0,10
rf	-0,19	-0,12	-0,38	0,12
sna	-0,24	-0,22	-0,47	0,17
sns	-0,11	0,34	0,14	-0,56
snd	0,43	-0,09	0,44	0,40
sia	-0,19	-0,06	-0,39	0,13
sis	-0,07	0,14	0,07	-0,37
sid	0,31	-0,04	0,38	0,28
sua	-0,18	-0,11	-0,39	0,17
sud	0,19	-0,04	0,23	0,21
sus	-0,03	0,03	0,02	-0,15

Na Tabela 9 a primeira componente principal (bebpc1) define consumidores de café com adoçante, chá com adoçante, suco natural com adoçante, suco industrializado com adoçante. A segunda componente principal (bebpc2) formado pelos que consomem café sem açúcar, chá sem açúcar, suco natural sem açúcar. Grupo que se mostra mais próximo de uma definição de saudável. Mas note que para indivíduos com altos scores negativos na segunda componente principal indicam consumo de bebidas com açúcar. São elas café com açúcar, suco industrializado com açúcar, chá com açúcar, e suco natural com açúcar. Grupo esse também marcado pela despreocupação com o teor calórico dos alimentos.

No grupo das bebidas não alcoólicas a análise fatorial e a análise de componentes principais formaram perfis concordantes.

Bebidas alcoólicas

Na Tabela 10 verificamos que a análise fatorial produziu três grupos. Tem-se os consumidores de vinho branco (alcpc1), e em outro extremo os consumidores de vinho tinto e um terceiro grupo formado pelos consumidores de cerveja e destilados. Os dois primeiros grupos parecem caracterizar um consumo esporádico da bebida, como as duas

Tabela 10: Grupo de Bebidas alcoólicas

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
beb	0,06	0,61	0,14	0,77
cv	0,02	0,45	0,13	0,74
vb	0,81	0,00	0,83	0,05
vti	-0,09	0,19	-0,58	0,42

variáveis estão muito afastadas entre si e do terceiro grupo (alpcp2) isto indica que os consumidores de vinho branco consomem preferencialmente vinho branco e raramente ou nunca as outras bebidas alcoólicas. O mesmo raciocínio vale para o vinho tinto.

No grupo das bebidas alcoólicas a análise fatorial e a análise de componentes principais formaram perfis concordantes.

Fast-Food

Fast-Food é um termo em inglês que significa, em uma tradução livre, comida rápida. São considerados Fast-Food alimentos preparados num pequeno intervalo de tempo, que ficam pré-prontos no estoque do estabelecimento.

Na maior parte das vezes, os alimentos de Fast-Food são desprovidos de nutrientes básicos para o bom funcionamento do corpo humano, e fartos em gorduras e açúcares. A grande quantidade de gordura presente nesses alimentos pode elevar os níveis de colesterol, aumentando o risco de doenças coronárias. Além disso, o açúcar presente nesses alimentos pode ter uma ligação direta com doenças cardíacas e diabetes.

Tabela 11: Grupo de Fast-Food

Alimento	Coeficientes análise Fatorial		Coeficientes análise Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
cq	0,10	0,53	0,53	0,40
ha	0,14	0,45	0,53	0,32
mn	0,16	0,24	0,41	-0,19
mnl	0,02	0,10	0,12	0,55
paq	0,20	0,11	0,33	-0,24
pp	0,06	0,18	0,26	0,21
pz	0,17	0,38	0,52	0,27
sa	0,65	0,16	0,64	-0,38
saf	0,67	0,16	0,65	-0,41
soi	0,05	0,11	0,17	0,16

A análise de componentes principais identificou consumidores de cachorro-quente,

hamburguer, maionese, pizza, salgado assado e salgado frito. Aqueles que fazem lanches rápidos em rodoviárias ou o substituem pelo almoço. A segunda componente principal identificou um elemento isolado que seria a maionese light. Aparentemente quem pertence a este último grupo não consome outros tipos de alimentos fast-food, indicando ser esse um grupo de pessoas saudáveis e que esporadicamente consomem a maionese light.

No entanto, o primeiro fator identificou um grupo de consumidores de salgado assado e salgado frito (fasfa1). O segundo fator (fasfa2) identificou um grupo de consumidores de cachorro-quente, hambúrguer, maionese e pizza.

Conclusão sobre perfis intermediários

Dessa forma, construídos os perfis intermediários a partir da análise de componentes principais e da análise fatorial, os perfis encontrados, a partir das duas técnicas, foram muito semelhantes com exceção de alguns grupos alimentares, como o grupo de leguminosas onde os resultados dos dois primeiros fatores foram representados por apenas uma componente principal, em que esta componente atribuía escores positivos para indivíduos com perfil semelhante ao do segundo fator e escores negativos para indivíduos com perfil semelhante ao do primeiro fator. Também, no caso do grupo de leite e derivados onde os perfis formados pela análise fatorial foram diferentes dos obtidos através da análise de componentes principais, mas a análise de componentes principais distinguiu melhor os perfis do que a análise fatorial que produziu dois fatores pouco interpretáveis. Em outros grupos alimentares as diferenças foram muito pequenas, onde uma técnica incluía uma variável que a outra não tinha incluso, mas que não alterava a interpretabilidade.

Das cento e trinta e oito variáveis originais tem-se agora dezessete variáveis auxiliares, que refletem as relações dentro de cada grupo alimentar. Por exemplo, o grupo de cereais, que tinham dezessete diferentes tipos de grãos (variáveis originais) reduziu-se a dois grupos interpretáveis (variáveis auxiliares).

Figura 4: Gráficos de perfis intermediários

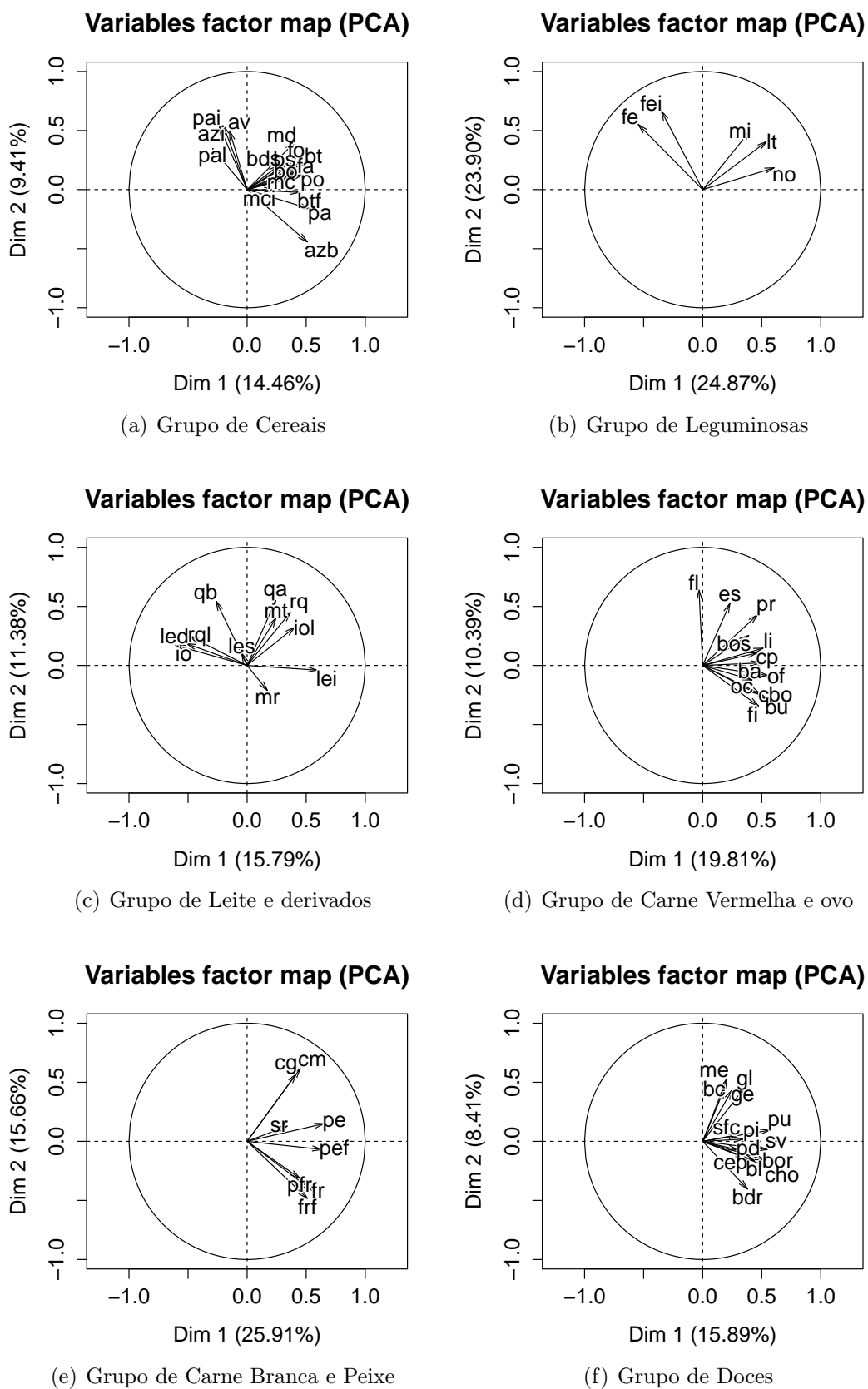
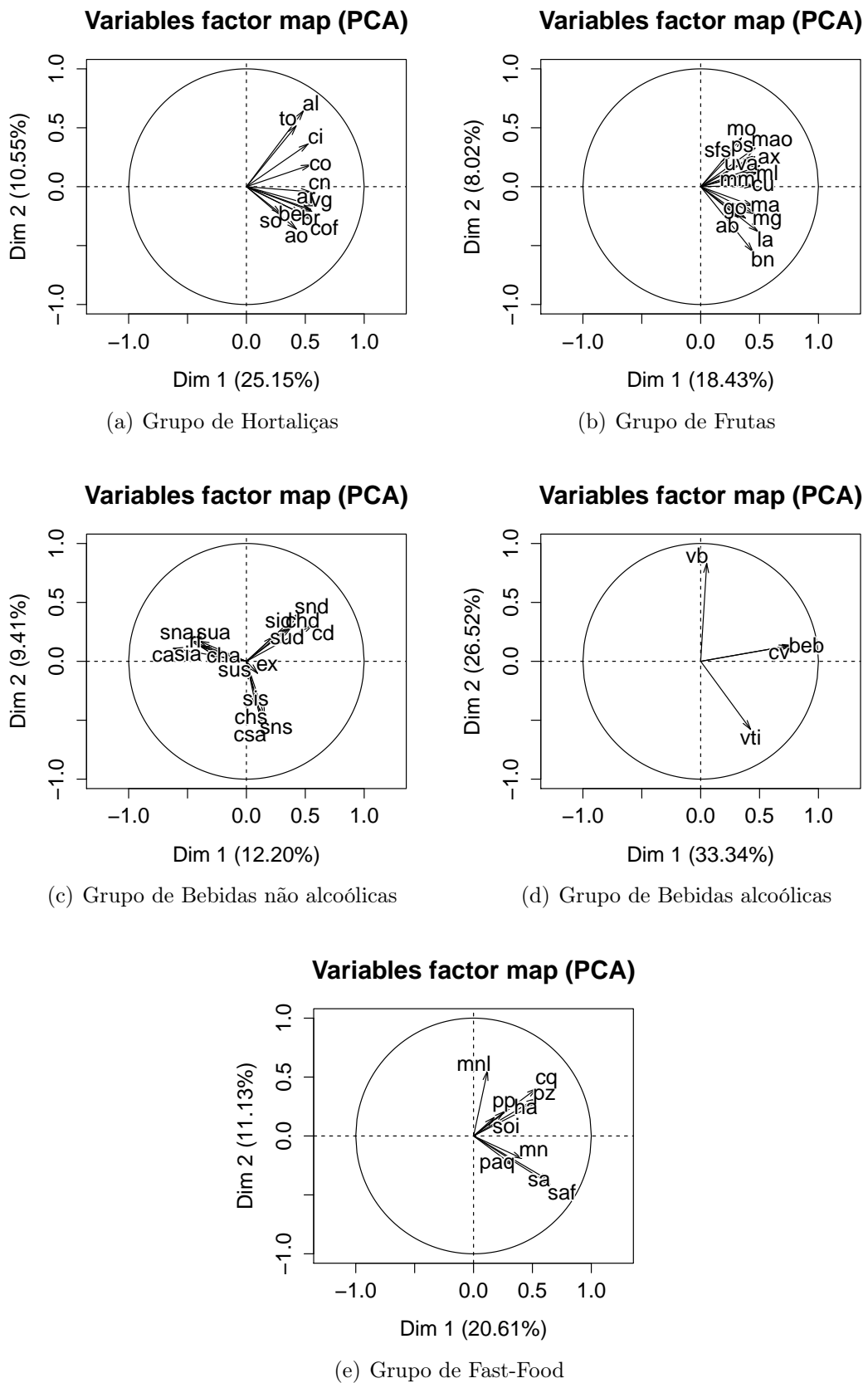


Figura 5: Gráficos de outros perfis intermediários



4.2 Padrões de consumo alimentar

Estabelecidos os perfis intermediários na seção anterior, o intuito agora é agrupar as variáveis auxiliares formadas utilizando novamente as técnicas multivariadas de análise de componentes principais, análise fatorial além da análise de cluster. Dessa forma será possível finalmente discernir padrões de consumo entre os grupos alimentares propostos, uma vez que as análises anteriores revelaram relações entre as variáveis dentro de cada um dos grupos alimentares estabelecidos.

Tabela 12: Resultados Finais

Alimento	Perfis da análise Fatorial		Perfis da análise de Componentes Principais	
	Fator 1	Fator 2	Componente 1	Componente 2
cer1	0,66	0,05	0,76	-0,17
cer2	0,04	0,50	0,08	0,59
leg1	-0,15	0,35	-0,21	0,61
leg2	0,51	0,25	0,63	0,14
doc1	0,51	0,06	0,50	0,21
doc2	-0,11	0,35	-0,10	0,46
lei1	0,38	-0,19	0,45	-0,32
lei2	0,22	0,19	0,16	0,47
carv1	0,68	0,08	0,75	-0,04
carv2	0,16	-0,11	0,00	0,25
frpe1	0,38	0,36	0,51	0,29
frpe2	-0,05	0,07	-0,11	0,24
bebe1	-0,33	0,19	-0,41	0,36
bebe2	0,12	-0,10	0,17	-0,25
alc1	0,18	0,03	0,23	0,00
alc2	0,05	-0,06	0,07	-0,12
frut1	0,21	0,58	0,33	0,51
frut2	-0,07	0,00	-0,18	0,27
hort1	0,14	0,57	0,23	0,56
hort2	-0,07	-0,07	-0,09	-0,08
fast1	0,63	-0,11	0,57	0,05
fast2	-0,02	0,03	-0,06	0,15

Tabela 13: Resultados análise de cluster

	cluster	Vizinho	Ajuste
leg2	1	3	0.59
carnevo1	1	3	0.58
fast1	1	3	0.56
cer1	1	3	0.51
doc1	1	3	0.45
frpe1	1	2	0.38
leider1	1	3	0.02
doc2	2	3	0.54
cer2	2	3	0.53
leg1	2	3	0.50
leider2	2	3	0.47
hort1	2	3	0.45
bebe1	2	3	0.29
frut2	2	3	0.28
frut1	2	1	0.25
frpe2	2	3	0.19
carv2	2	3	0.16
alc2	3	2	0.71
hort2	3	2	0.58
bebe2	3	1	0.48
alc1	3	1	0.39
fast2	3	2	0.02

Análise fatorial

Na análise fatorial, obteve-se os seguintes padrões de consumo alimentar. O primeiro fator é composto das variáveis cer1, leg2, doc1, lei1, carv1, frpe1, fast1.

Mas antes de interpretar o conjunto dessas variáveis, é importante relembrar suas interpretações individuais para então explicar qual o padrão alimentar formado por essas variáveis.

- A variável cer1 caracteriza um grupo pouco restrito com seu consumo de calorias, pois consomem arroz branco, batata, batata frita, farinha, farofa, macarrão, pão francês, polenta e bolo.
- A variável de leguminosas, leg2, caracteriza os consumidores de lentilhas, milho e feijão e feijoada.
- A variável de doces, doc1, caracteriza o consumo de doces industrializados, como bala, biscoito recheado, pudim, sorvete, pirulito, chocolate em pó.
- A variável de leite e derivados, lei1, caracteriza o consumo de iogurte light, leite integral, requeijão.
- A variável de carne vermelha, carv1, caracteriza um grupo de indivíduos que consomem carnes com alto teor de gordura, por exemplo, bacon, linguiça, carne de porco, bucho, carne bovina com osso.
- A variável de carne branca e peixe, frpe1, discriminou indivíduos que consomem peixe cozido, peixe frito, frango, frango frito e peito de frango.
- A variável de fast-food, fast1, identificou um grupo de consumidores de salgado assado, salgado frito, além de hambúrguer, cachorro-quente e pizza.

Assim, como verificado acima, as variáveis cer1, doc1, lei1, carv1 e fast1 caracterizam um grupo pouco restrito com seu consumo de calorias e pouco preocupados em manter uma dieta saudável. Pode-se concluir portanto, que esse primeiro fator caracteriza indivíduos com maus hábitos alimentares e que procuram também corrigir esses hábitos ingerindo produtos que sejam considerados mais saudáveis, como iogurt light e o frango.

Já o segundo fator foi caracterizado pelas variáveis cer2, leg1, doc2, frut1, hort1, frpe1.

Aqui novamente será lembrada as interpretações individuais das variáveis, para que auxiliem a visão geral do padrão.

- A variável de cereais cer2, caracterizou o consumo de cereais integrais como arroz integral, mandioca, aveia, pão light e pão integral.
- Já a variável de leguminosas, leg1, caracterizou o consumo de feijão e feijoada atribuindo escores negativos para os indivíduos que consomem esses alimentos e caracterizou com escores positivos os consumidores de milho, lentilha e noz.
- A variável de doces, doc2, caracterizado pelo consumo de geléia, mel, gelatina e biscoito doce sem recheio.
- A variável de frutas, fru1, caracteriza um grupo formado pela maioria das frutas desse grupo alimentar, no caso abacaxi, ameixa, banana, goiaba, laranja, maçã, mamão, manga, morango, pêssego e uva.
- A variável de carne vermelha, carv1, caracteriza um grupo de indivíduos que consomem carnes com alto teor de gordura, por exemplo, bacon, linguiça, carne de porco, bucho, carne bovina com osso.
- A variável de carne branca e peixe, frpe1, discriminou indivíduos que consomem peixe cozido, peixe frito, frango, frango frito e peito de frango.
- A variável de hortaliças, horpca1, caracterizado pelo consumo de abobrinha, brócolis, berinjela, couve-flor, cenoura e vagem.

Dessa forma, como as variáveis de cer2, doc2, fru1, horpca1, frpe1, hor1, pode-se concluir portanto, que esse segundo fator caracteriza indivíduos com bons hábitos alimentares e que procuram manter uma alimentação saudável.

Análise de componentes principais

Na análise de componentes principais, encontrou-se os seguintes padrões de consumo alimentar. A primeira componente é composta pelas variáveis cer1, leg2, doc1, lei1, carv1, frpe1, fast1, fru1. Como as variáveis citadas anteriormente caracterizam um grupo pouco restrito com seu consumo de calorias e pouco preocupados com a manutenção de um dieta saudável. Pode-se concluir portanto, que essa primeira componente principal caracteriza indivíduos com pouca ou nenhuma restrição alimentar, com maus hábitos alimentares. Perfil semelhante ao produzido pelo primeiro fator da análise fatorial.

A segunda componente principal foi caracterizada pelas variáveis cer2, leg1, doc2, fru1, hor1. Como as variáveis citadas anteriormente caracterizam um grupo preocupado em manter uma dieta saudável, pode-se concluir portanto, que essa segunda componente caracteriza indivíduos com bons hábitos alimentares e que procuram manter uma alimentação saudável. Perfil semelhante ao produzido pelo segundo fator da análise fatorial.

Análise de cluster

Na Tabela 13 tem-se as seguintes informações, na segunda coluna temos o agrupamento ao qual a variável pertence, na terceira temos qual o agrupamento mais próximo ao seu, isto é, qual o outro grupo mais próximo a que a variável perteceria, e na quarta coluna temos o grau de pertencimento da variável no seu agrupamento.

Veja que o primeiro cluster agrupou em si as mesmas variáveis que definiram o primeiro fator, são elas cer1, leg2, doc1, lei1, carv1, frpe1, fast1, fru1, como verifica-se na Tabela 12. Observe também que as variáveis auxiliares mais bem ajustadas nesse primeiro cluster são as variáveis carv1, fast1 e cer1 com um grau de ajustamento maiores que 0.50. Note que essas variáveis caracterizam bem o consumo de alimentos calóricos e uma alimentação não balanceada.

O segundo cluster agrupou as mesmas variáveis que definiram melhor o segundo fator, a saber, cer2, leg1, doc2, fru1, fru2, hor1, lei2, bebe1, frpe2, carv2. Apenas as variáveis fru2, lei2, bebe1, frpe2, carv2 foram incluídas pela análise de cluster e não estavam presentes na análise fatorial nem na análise de componentes principais, mas perceba que seus graus de ajustamento não são bons, sendo eles menores que 0.30. Mas vale explicar quais perfis alimentares cada uma dessas variáveis caracterizam. A variável fru2 representa indivíduos que consomem salada de fruta sem complemento, a variável lei2 representa indivíduos que consomem iogurte, queijo amarelo, requeijão e manteiga, a variável bebe1 representa indivíduos que tem por hábito ingerir bebidas com adoçantes, a variável carv2 os consumidores de frios light e presunto, e a variável frpe2 os consumidores de camarão e caragueijo.

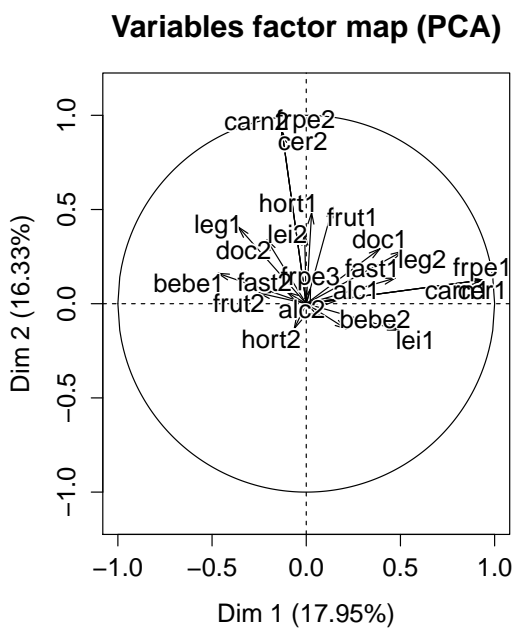
O terceiro cluster mostrou ter agrupado as outras variáveis que não se ajustaram bem nos outros grupos, note que essas variáveis também não se ajustaram bem na análise fatorial e análise de componentes principais. Agruparam-se nesse terceiro cluster as variáveis alc2, hort2, bebe2, alc1, fast2, são esses os indivíduos que tem por hábitos ingerir vinho branco ou vinho tinto, comer fast-food como hambúrguer e cachorro quente, bebem bebidas sem açúcar e comem alface, cenoura e couve-flor. Interessante notar na Figura 6, que ao observar o gráfico de silhouette, este terceiro cluster, teve um grau de ajustamento

razoável, indicando que o mesmo estaria bem definido.

Conclusão sobre os padrões de consumo alimentar

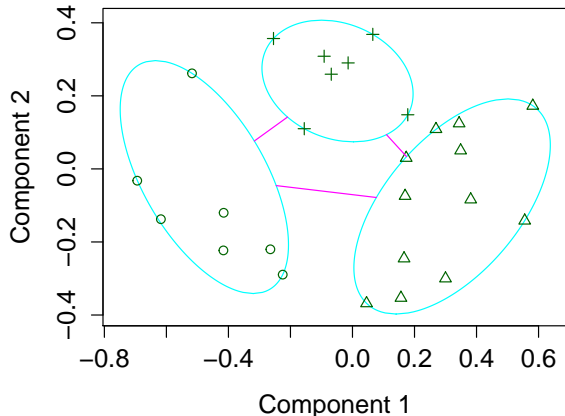
Os perfis formados pela análise fatorial, pela análise de componentes principais e pela análise de cluster foram semelhantes. Eles distinguiram dois grupos, um grupo não-saudável caracterizado pelo primeiro fator, pela primeira componente principal e pelo primeiro cluster. E outro grupo saudável caracterizado pelo segundo fator, pela segunda componente principal e pelo segundo cluster. No entanto, vale ressaltar que a análise de cluster levou em consideração algumas variáveis que atrapalhavam a visualização dos padrões em questão, como no caso do segundo cluster que incluiu variáveis a mais do que as encontradas pela análise fatorial e pela análise de componentes principais. Também o terceiro cluster que agrupou as variáveis restantes que não foram significativas em nenhum outro cluster.

Figura 6: Resultados gráficos finais



(a) Componentes principais

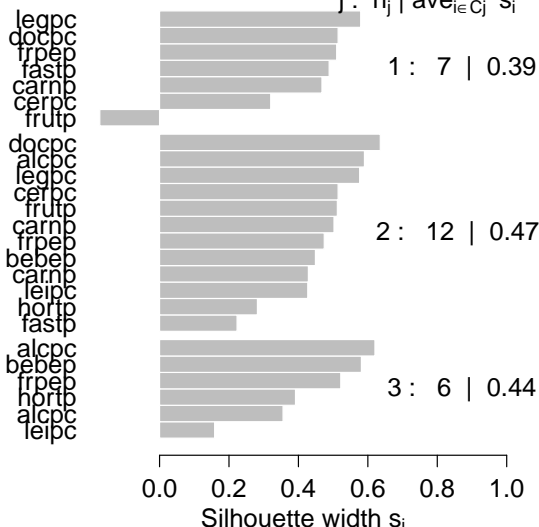
`clusplot(pam(x = cl.svarcoord[, 1:2], k =`



These two components explain 100 % of the |
(b) Cluster Partição de Medóides

Silhouette plot of pam(x = cl.s\$

n = 25 3 clusters C_j
j : n_j | $\text{ave}_{i \in C_j} s_i$



5 Conclusão

Neste trabalho, a proposta era abordar a associação entre os alimentos a fim de obter um padrão alimentar, usando informações dietéticas. Mas não somente encontrar um padrão alimentar, como discutir semelhanças e diferenças encontradas a partir da comparação dos resultados dos padrões alimentares obtidos por meio de três diferentes técnicas, no caso análise fatorial, análise de componentes principais e análise de cluster.

Para isso, das cento de trinta e oito variáveis de alimentação foram alocadas em grupos alimentares e dessa forma, foram formados os perfis intermediários a partir da análise de componentes principais e da análise fatorial. Então, como os perfis encontrados pelas duas técnicas foram semelhantes, optou-se por utilizar os dados obtidos a partir da análise de componentes principais apenas. Depois de estabelecidos os perfis intermediários, os mesmos foram reagrupados em um único grupo alimentar e utilizando novamente as técnicas multivariadas de análise de componentes principais, análise fatorial e agora a análise de cluster, finalmente chegou-se aos padrões de consumo.

A três técnicas separaram-se em dois padrões interpretáveis, mesmo que a análise de cluster tenha formado três grupos, apenas dois deles eram interpretáveis. Os dois padrões de consumidores encontrados foram saudável e não saudável. Como esperado e confirmado pela literatura.

No entanto, uma observação deve ser feita, a análise de componentes principais mostrou, durante todo o processo de análise, relações entre os alimentos que a análise fatorial não pode descrever. Isto é, algumas variáveis que contribuíam na interpretabilidade do perfil encontrado foram incluídas pela análise de componentes principais, mas a análise fatorial não conseguiu descrever essas mesmas variáveis em seus fatores. Muito disso se deve à natureza da própria técnica que, diferente da análise fatorial que foi formulada para medir variáveis não-observáveis diretamente, a análise de componentes principais extraiu relações imperceptíveis pela análise fatorial, pelo fato de a base de dados utilizada ser medida em uma escala numérica contínua e observável, nesse caso a kcal. O que é

interessante, uma vez que na literatura nutricional a grande maioria dos artigos da área se utiliza largamente da análise fatorial e não da análise de componentes principais para descrever os perfis alimentares.

Uma vez estabelecidos os padrões de consumo alimentar, o próximo passo desta discussão é a partir deles investigar como os mesmos se associam a doenças crônicas não transmissíveis, como as doenças cardiovasculares, por exemplo.

Uma outra proposta é utilizar modelos de equações estruturais para modelar variáveis não-observáveis não consideradas nessa análise mas que podem influenciar na construção dos padrões.

Referências

- [1] D. C. Almeida. A comparison of three statistical methods applied in the identification of eating patterns. *Cad. Saúde Pública, Rio de Janeiro*, 26:2138–2148, 2010.
- [2] E. R. Frank Hu. Prospective study of major dietary patterns and risk of coronary heart disease in men. *American Society for Clinical Nutrition*, 72:912–921, 2000.
- [3] J. E. Jackson. *A User's Guide To Principal Components*. Wiley-Interscience Publication, 1991.
- [4] D. W. Johnson, Richard A.; Wichern. *Applied multivariate statistical analysis*. Pearson Prentice Hall, 2007.
- [5] M. Olinto. Dietary patterns and their associations. *Cad. Saúde Pública*, 25(6):1297–1306, 2009.
- [6] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:26–31, 1901.
- [7] A. M. Rosely Sichieri, Joelma Castr. Fatores associados ao padrão de consumo alimentar da população brasileira urbana. *Cad. Saúde Pública, Rio de Janeiro*, 19:S47–S53, 2003.
- [8] R. Sichieri. Dietary patterns and their associations with obesity in the brazilian city of rio de janeiro. *Obesity research*, 1(1), 2002.
- [9] C. Spearman. "general intelligence", objectevely determined and measures. 15:201–293, 1904.
- [10] R. F. Suely Gimeno; Solange Andreoni. Assessing food dietary intakes in japanese-brazilians using factor analysis. *Cad. Saúde Pública*, 26(11):2157–2167, 2010.
- [11] L. L. Thurstone. Multiple-factor analysis; a development and expansion of the vectors of mind. *University of Chicago Press*, 19(1):535, 1947.

ANEXO A – Lista de Siglas

- 1) azb:arroz branco
- 2) azi:arroz integral
- 3) av :aveia
- 4) bt :batata
- 5) btf:batata frita
- 6) fa :farinha
- 7) fo :farofa
- 8) mc :macarrão
- 9) mci:macarrão instantâneo
- 10) md :milho verde
- 11) pal:pão light
- 12) pa :pão
- 13) pai:pão integral
- 14) po :bolo sem recheio
- 15) bds:biscoito doce sem recheio
- 16) bs :biscoito salgado
- 17) bo :bolo
- 18) mi :milho

- 19) fe :feijão
- 20) fei:feijoadada
- 21) lt :lentilha
- 22) no :noz
- 23) io :iogurte
- 24) iol:iogurte light
- 25) led:leite desnatado
- 26) lei:leite integral
- 27) les:leite semi-desnatado
- 29) mt :manteiga
- 30) mr :margarina
- 31) qa :queijos amarelos
- 32) qb :queijos brancos
- 33) rq :requeijão
- 34) rql:requeijão light
- 35) bl :bala
- 36) bc :barra de cereal
- 37) bdr:biscoito doce com recheio
- 38) bor:bolo com recheio
- 39) cho:chocolate
- 40) cep:chocolate em pó
- 41) ge :gelatina
- 42) gl :geléia
- 43) me :mel

- 44) pd :pão doce
- 45) pi :picolé de fruta
- 46) pu :pudim
- 47) sfc:salada de frutas com complementos
- 48) sv :sorvete
- 49) ca :café com açúcar
- 50) cd :café com adoçante
- 51) csd:café sem açúcar,
- 52) cha:chá mate com açúcar
- 53) chd:chá mate com adoçante
- 54) chs:chá mate sem açúcar
- 55) ex :extrato de soja
- 56) rf :refrigerante
- 57) sna:suco artificial com açúcar
- 58) snd:suco artificial com adoçante
- 59) sns:suco artificial sem açúcar
- 60) sia:suco industrializado com açúcar
- 61) sid:suco industrializado com adoçante
- 62) sis:suco industrializado sem açúcar
- 63) sua:suco natural com açúcar
- 64) sud:suco natural com adoçante
- 66) sus:suco natural sem açúcar
- 67) beb:bebidas destiladas
- 68) cv :cerveja

- 69) vb :vinho branco
- 70) vti:vinho tinto
- 71) fl :frios light
- 72) ba :bacon
- 73) bu :bucho
- 74) fi :fígado
- 75) cbo:carne bovina com osso
- 76) bos:carne bovina sem osso
- 77) cp :carne de porco
- 78) es :estrogonofe
- 79) of :ovo frito
- 80) oc :ovo cozido
- 81) li :linguiça
- 82) pr :presunto
- 83) fr :Frango
- 84) frf:frango frito
- 85) prf:peito frango
- 86) cm :camarão
- 87) cg :carangueijo
- 88) pe :peixe
- 89) pef:peixe frito
- 90) sr :sardinha
- 91) ao :abobrinha
- 92) ar :abóbora

- 93) al :alface
- 94) be :beterraba
- 95) br :brócolis
- 96) cn :cenoura
- 97) ci :chicória
- 98) co :couve
- 99) cof:couve flor
- 100) so :sopa de legumes
- 101) to :tomate
- 102) vg :vagem
- 103) ab :abacate
- 104) ax :abacaxi
- 105) bn :banana
- 106) cu :caqui
- 107) go :goiaba
- 108) la :laranja
- 109) ma :maçã,
- 110) mm :mamão
- 111) mg :manga
- 112) ml :melancia
- 113) mao:melão
- 114) mo :morango
- 115) ps :pêssego
- 116) sfs:salada de frutas sem complementos

- 117) uv :uva
- 118) cq :cachorro quente
- 119) ha :hambúrguer
- 120) mn :maionese
- 121) mnl:maionese light
- 122) paq:pão de queijo
- 123) pp :pipoca
- 124) pz :pizza
- 125) sa :salgado assado
- 126) saf :salgado frito
- 127) soi :sopa instantânea