

ANÁLISE MULTIVARIADA DE DADOS QUALITATIVOS DE INFORMAÇÃO NUTRICIONAL DA CONCENTRAÇÃO DE FRUTAS EM PRODUTOS ALIMENTÍCIOS: VALORES FORMAIS E VALORES REAIS

Giovani Glaucio de Oliveira Costa¹

¹Departamento de Tecnologias e Linguagens-Universidade Federal Rural do Rio de Janeiro-Nova Iguaçu.
Rio de Janeiro giovaniglaucio@ufrj.br

Resumo

Atualmente existe certo desacordo entre o que deve ser formalmente exposto pelas indústrias com relação a composição de frutas de seus produtos e o que realmente é oferecido ao consumidor final. Os produtos têm foto de fruta na embalagem, cor de fruta, sabor de fruta, etc., mas não contém fruta. Essa foi a constatação do Instituto de Defesa do Consumidor em relação a investigação de 18 produtos alimentícios industrializados, avaliados em sua última pesquisa. O presente artigo objetiva prosseguir no estudo do Idec para investigar se o perfil de “propaganda enganosa” atingi como um todo a amostra em estudo ou se este perfil se segmenta com variações diferenciadas entre os produtos, revelando uma estrutura com configurações distintas, coexistentes num mesmo espaço de análise. Para atingir o objetivo, optou-se por reduzir a dimensionalidade do espaço de análise, para que se possa perceber com mais clareza e rapidez as configurações coexistentes, e assim realizar um estudo também estrutural. Como as variáveis da pesquisa são qualitativas, a técnica multivariada adequada ao intento é a Análise de Homogeneidade. A conclusão a que se chegou com a estruturação do problema é que realmente a grande maioria dos produtos investigados (83%) está no perfil geral do Idec: “uso de imagens de frutas nas embalagens desproporcional ao conteúdo” e “desinformação do percentual de frutas contidas nas embalagens”. Mas existe, sim, latente na base de dados, um subperfil que apresenta variação do perfil geral: o Perfil 2. Estas configurações de produtos alimentícios, basicamente os refrescos em pó, usam “imagens de frutas nas embalagens proporcionais ao conteúdo” e informam o “percentual de frutas contidas nas embalagens”. Esta revelação da investigação é importante às tomadas de decisão com respeito às políticas adotadas no sentido da defesa dos direitos do consumidor e valorização da qualidade de vida.

Palavras-chave: *produtos industrializados, informação nutricional da concentração e percentual de frutas, análise estrutural, análise de homogeneidade.*

1 INTRODUÇÃO

Atualmente existe certo desacordo entre o que deve ser formalmente exposto pelas indústrias com relação a composição de frutas de seus produtos e o que realmente é oferecido ao consumidor final. Os produtos têm foto de fruta na embalagem, cor de fruta, sabor de fruta, etc., mas não contém fruta. Essa foi a constatação do Instituto de Defesa do Consumidor (IDEC, 2011)

em relação a investigação de 18 produtos alimentícios industrializados, avaliados em sua última pesquisa.

A análise levou em consideração se os apelos relacionados a frutas nos rótulos estão condizentes com o conteúdo. E revelou que as situações *“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”* e *“desinformação do percentual de frutas contidos nas embalagens”* são tendências gerais no setor, isto é, constituem *“regras de comportamento”* no setor no setor.

Segundo o Idec (2011), é difícil para o leigo entender que o produto tem sabor de morango, mas não tem morango, por exemplo. E, ao usar com destaque imagens de frutas nas embalagens, as empresas estão induzindo o consumidor a achar que está comprando um produto mais natural e saudável, o que pode ser entendido como *“propaganda enganosa”*.

O Idec (2011) analisou embalagens de três fabricantes em cada uma das seis categorias verificadas. No grupo de iogurtes, foram avaliados *Batavo Kissy Morango*, *Leco Light Lactive* e *Danoninho Iogurte*. No de refrescos em pó, *Tang*, *La Fruta* e *Camp*. Entre os isotônicos, *Gatorade*, *Taeq* e *Marathon*. Na classificação néctar de fruta, *Top Fruit*, *Maguary* e *Del Valle-Kapo*. Na categoria gelatina, *Frutop*, *Dr. Oetker* e *Royal*. E na de sorvetes, o sabor napolitano da *Sorveteria Kidelícia de Sabor*, *Nestlé* e *Kibon Balance*.

Segundo o Conselho Federal de Nutricionistas (CFN, 2011), o que a indústria faz ao usar e abusar da embalagem das frutas é no mínimo violação dos direitos do consumidor, enganoso e antiético. Para o CFN (2011), quem consome muitos produtos industrializados pode acabar ultrapassando o percentual máximo de aditivos químicos recomendados pela FAO e o uso efetivo de frutas no conteúdo destes produtos pode minimizar os efeitos da industrialização na saúde do consumidor.

Contudo, a Associação Brasileira de embalagem contemporiza alegando que o uso de frutas na embalagem confere uma conotação mais saudável ao produto, mas que o recurso também é usado para facilitar a identificação do sabor. O órgão destaca que a indústria deve seguir normas legais e que eventuais mudanças têm que ser discutidas com o poder público e o consumidor, para o que é formal na norma, não estejam divergentes com o que realmente é oferecido.

O IDEC (2011) e o CFN (2011) defendem, no entanto mudanças imediatas: divulgação do percentual de frutas dos produtos, a proibição de imagens de frutas nas embalagens desproporcionais ao conteúdo e atitude proativa com relação ao atendimento total às necessidades do consumidor.

Segundo pesquisa do Idec (2011), 62% dos consumidores consideram difícil saber se há frutas no alimento e que gostariam de regulamentação obrigando essa informação. Um grupo de 25% dos entrevistados defende a mudança para evitar propaganda enganosa.

O levantamento do Idec (2011), como se relatou em parágrafo anterior, levou em consideração se os apelos relacionados a frutas nos rótulos estão condizentes com o conteúdo. E revelou que as situações *“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”* e *“desinformação do percentual de frutas contidos nas embalagens”* são tendências gerais no setor, isto é, constituem *“regras de comportamento”* no setor no setor. Contudo, como podem existir exceções à regra, é oportuno se investigar se existe variação do comportamento geral constatado e em que grau e que grupos de produtos apresentam dispersões semelhantes.

O que se constata com a análise das tabelas de frequências triviais já realizadas pelo Idec (2011), que à nível geral, existe uma diferença entre o que as indústrias alimentícias informam no conteúdo das embalagens sobre o sabor da fruta, cor da fruta e o conteúdo propriamente dito. Mas num possível prosseguindo na análise, poder-se-ia questionar se os produtos investigados possuem o perfil de *“propaganda enganosa”* (uso de imagens de frutas nas embalagens desproporcionais ao conteúdo) como um todo homogêneo, ou se este perfil se segmenta com variações diferenciadas entre os produtos, se revelando uma estrutura com configurações distintas, coexistentes num mesmo espaço de análise.

Este artigo se propõe oportunamente a prosseguir na análise dos dados da pesquisa do Idec (2011), com a incumbência de reduzir a dimensionalidade do espaço de análise, para que se possam perceber com mais clareza e rapidez se existem grupos de produtos alimentícios com graus diferenciados de *“propaganda enganosa”*, se distanciamento das normas de regulamentação, e assim realizar um estudo também estrutural.

O presente artigo levanta o *“problema de pesquisa”* de que o perfil geral dos produtos alimentícios investigados aponta para *“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”* e *“desinformação do percentual de frutas contidas nas embalagens”*, mas questiona que pode haver certo nível de dispersão quanto este perfil geral evidenciado. Um olhar mais depurado, mas detalhado da estrutura do problema, pode revelar que nem todas as indústrias têm exatamente o mesmo grau de desproporcionalidade quanto à estrutura geral e, é muito importante que se detecte esse fato e que se dissemine esta informação ou constatação no setor, na sociedade e no meio científico.

Na elaboração de um plano de pesquisa que planeja uma coleta extensa de informações, a seleção dos métodos de análise de dados é uma tarefa (entre outras) de importância decisiva, quaisquer que seja o campo em estudo. É essencial garantir que os instrumentos que venham a ser

selecionados permitam operacionalizar as orientações sistematizadas no modelo de análise, e sejam consistentes com a natureza da informação escolhida.

Todavia, muitas das vezes quando o pesquisador elabora esse plano de pesquisa quantitativa, utilizando o método estatístico, ele se depara com questionários onde se não toda a maioria das variáveis é qualitativa, isto é, as questões que envolvem o questionário têm como respostas, muitas das vezes estimuladas, a forma categorial. As técnicas tradicionais de Análise Multivariada pressupõem que as variáveis sejam quantitativas e que sigam à Curva Normal. O analista, então, tem utilizado como alternativa o recurso analítico trivial do exame de tabelas de frequência, tabelas de contingência com testes de independência do qui-quadrado.

No estudo do Idec (2011), constata-se que todas as variáveis da pesquisa são variáveis qualitativas nominais, categoriais e o instituto, então, se orientou como de usual, para a análise de dados, na aplicação da descrição trivial dos dados.

A Análise de Homogeneidade ou Análise de Correspondência Múltipla, técnica recente e inovadora, pouco difundida ainda principalmente no Brasil, permite que os pesquisadores ao utilizar questionários que envolvam perguntas com respostas nominais, saiam do campo descritivo trivial e desenvolvam técnicas multivariadas para o tratamento avançado dos múltiplos indicadores relacionais envolvidos na pesquisa. A Análise de Homogeneidade trata múltiplas variáveis categoriais e realiza análises estruturais, a partir da redução de dimensionalidade do espaço de partida (base de dados original).

A Análise de Homogeneidade é o instrumento analítico que privilegia as situações nas quais é necessário lidar conceitualmente e metodologicamente com objetos de estudo de configuração complexa, com muitos indicadores inter-relacionados e que por simplificação, maior nitidez e rapidez, é feita uma redução de dimensionalidade.

Portanto, a Análise de Homogeneidade é um recurso de análise apropriado para se atingir o objetivo deste artigo e que será explorado e prestigiado no desenvolvimento das seções a seguir.

2 MATERIAL E MÉTODOS

A HOMALS foi formalizada por GIFÍ (1996) que também foi responsável pela operacionalização dos seus procedimentos no *software* SPSS. O método de Análise de Homogeneidade (HOMALS-Homogeneity Analysis by Means of Least Square), também conhecida como Análise de Correspondência Múltipla (ACM), permite resumir um grande número de variáveis qualitativas (ou tratadas como tal) de uma base de dados de estudo em um pequeno número de variáveis quantitativas, facilitando o estudo das relações entre as diversas características existentes num determinado espaço de análise (CARVALHO, 2004).

Em sentido restrito, a Análise de Homogeneidade é um método de análise quantitativa para variáveis nominais e tem por objetivo atingir a solução ótima na quantificação de dados qualitativos (GREENACRE, 1984). Em sentido amplo, a Análise de Homogeneidade refere-se a um conjunto de critérios para analisar dados multivariados, com o objetivo de aperfeiçoar a homogeneidade das variáveis (GREENACRE, 1984).

A Análise de Homogeneidade integra os métodos de “*Optimal Scaling*” (GIFI, 1996). Estes métodos consistem, na prática, numa intervenção sobre as categorias com o objetivo de proceder a sua quantificação (GIFI, 1996).

A HOMALS viabiliza uma Análise de Homogeneidade por via de um método algébrico do tipo “*Alternating Least Squares*” (GIFI, 1996). O algoritmo da HOMALS é, então, do tipo ALS, o que significa que em cada uma das iterações vão sendo determinadas, alternativamente, estimativas quantitativas para os objetos e categorias da base de dados (GIFI, 1996). Este processo é um procedimento de transformação ótima porque vai sendo minimizada uma função perda, cuja convergência corresponderá às quantificações das categorias e dos objetos. Essa função perda é a média do quadrado dos comprimentos das “linhas” (distância *euclidiana*) que ligam as observações às categorias das variáveis, podendo ser interpretada como o nível de ajuste do modelo: quanto menor, melhor o ajuste do modelo e vice-versa (GIFI, 1996). Estas quantificações têm a propriedade de guardarem e resumirem as características essenciais das categorias e objetos de análise da matriz de dados de partida, mantendo a variação de informação do espaço original intacta no espaço gerado de menores dimensões (GIFI, 1996).

A Análise de Homogeneidade consiste em analisar simultaneamente uma população de n indivíduos a partir de j variáveis nominais ou das categorias, ou tratadas como tal, que compõe estas variáveis (CARVALHO, 2004). A determinação das relações entre as categorias é dada pela forma como estas incidem sobre os casos, se junto ou separadamente (CARVALHO, 2004).

Para duas categorias j e k quaisquer, o critério de comparação entre elas é dado pela sua distância, calculada segundo a métrica qui-quadrado através da seguinte equação (GIFI, 1996):

$$d^2_{j,k} = n[(n_j - n_{j,k})/n_{j,k} + (n_k - n_{j,k})/n_{j,k}] \quad \text{Equação 1}$$

$n_{j,k}$ é a frequência dos casos que apresentaram as categorias j e k simultaneamente, n_j é a frequência dos casos que apresentaram somente a categoria j e n_k , somente a categoria k .

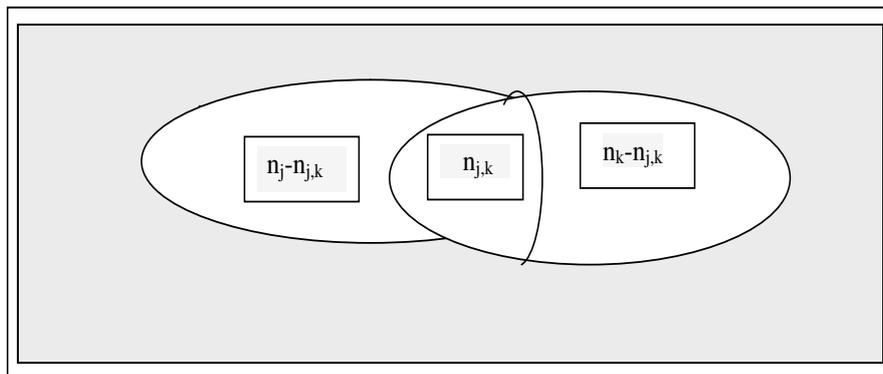
A equação 1 pode ser interpretada da seguinte forma: o quadrado da distância entre as categorias j e k é dado pela proporção de casos que pertencem somente à categoria j , mais a

proporção daquelas que pertencem somente à categoria k . A distância entre as categorias cresce quanto menor for à prevalência de casos que compartilhem ambas as categorias, ou melhor, quanto maior for à exclusividade entre elas. Se as categorias j e k compartilham muitos objetos, elas estão mais próximas, pois $n_{j,k} \rightarrow \infty$ e as proporções contidas na expressão 1 tendem a zero, conseqüentemente $d_{n,k}^2$ tende à zero. Entretanto, se as categorias j e k compartilham poucos objetos, elas estão mais distantes, pois $n_j - n_{j,k} \rightarrow \infty$ e as proporções contidas na fórmula 1 tendem ao infinito, conseqüentemente $d_{n,k}^2$ tende ao infinito.

A análise das relações é feita em termos de localização desses valores no espaço, mais especificamente, no plano cartesiano, que possibilita uma comparação simplificada das múltiplas correlações existentes (GIFI, 1996).

A Figura 1 pode facilitar o entendimento da interpretação da equação 1.

Figura 1- Relação entre as Categorias j e k



Assim, é possível dizer, de acordo com a equação 1, que (GIFI,1996):

Categorias pertencentes a uma mesma variável estarão obrigatoriamente separadas no espaço;

Categorias que são comuns a maiorias dos indivíduos se situarão próximas;

Categorias pouco freqüentes se localizarão bem separadas das demais.

A HOMALS constitui, portanto, num conjunto de critérios para analisar dados multivariados, com o objetivo de se atingir a homogeneidade das variáveis (GREENACRE, 1984). Fala-se em homogeneidade no sentido em que a proximidade de certo número de categorias de diferentes variáveis induz a presença de indivíduos na base de dados que compartilham tendencialmente as mesmas características, isto é, tem o mesmo perfil. Todos os diferentes núcleos de homogeneidade correspondem grupos ou configurações de indivíduos com perfis

distintos, mas que coexistem, com maior ou menor proximidade, no mesmo espaço (CARVALHO, 2004).

A descrição das configurações ou núcleos pode contemplar dois vetores analíticos (CARVALHO, 2004):

Identificação do perfil de cada grupo definido;

Observação do posicionamento relativo dos vários grupos que permita detectar a existência de relação de associação ou de oposição.

Os resultados ou *outputs* da HOMALS incluem (CARVALHO,2004):

Estatísticas- valores próprios (variâncias explicadas de eixos estruturantes), medidas de discriminação ou importância das variáveis, quantificações das categorias e escores de objetos;

Imagens gráficas(planos bidimensionais) que permitem a visualização da informação contida na matriz de entrada num espaço de menores dimensões.

Segundo Carvalho (2004), que apresenta o método, como prosseguimento analítico de validação dos resultados da Análise de Homogeneidade, Pode-se fazer a aproximação da configuração topológica delineada pela esta com a construção de tipologias classificatórias da Análise de *Clusters*, articulando as potencialidades das duas técnicas numa única análise de dados.

A classificação dos objetos aos *clusters* é determinada em função das distâncias que existem entre os objetos, relativamente a um conjunto de variáveis definidas. Na articulação HOMALS/Análise de *Clusters*, a Análise de *Clusters* toma como variáveis de *input*, para a classificação, as dimensões que sustentam o plano em estudo na Análise de Homogeneidade¹.

Quando se faz referência às dimensões enquanto novas variáveis, isso já pressupõe que depois de realizada a HOMALS e tomada à decisão sobre quantas dimensões interpretar, foram salvos na base de dados os escores de objetos por dimensão. Com este procedimento, serão acrescentadas à base de dados novas variáveis, em número igual ao indicado para as dimensões.

As dimensões são transformadas em novas variáveis com características particularmente interessantes: variáveis compostas e contínuas, que têm enorme potencial analítico. As dimensões são constituídas pelos escores de objetos e estes, algebricamente são determinadas a partir das múltiplas variáveis qualitativas que figuram na matriz de *input* e que definem o perfil de cada indivíduo. Daí afirmar-se que as dimensões são variáveis compostas, garantindo que a multidimensionalidade foi preservada. Quanto à sua natureza, e apesar do *input* ser qualitativas, as dimensões são variáveis contínuas.

Para realizar a Análise de *Clusters* começa-se por recorrer-se ao Método *K-Means Clusters*. Assim, será utilizada como variáveis de *input* as dimensões consideradas para a análise e solicitar o

número de grupos que se quer formar, correspondente ao número de configurações reveladas no gráfico das categorias da Análise de Homogeneidade.

Para garantir que a partição dos indivíduos nos *clusters* gerados pela análise fique definida como nova variável é necessário salvar a referida variável de classificação (de *clusters*) na base de dados. Em seguida, vem à caracterização dos grupos obtidos, a fim de efetuar a correspondência destes (os *clusters*) com os perfis previamente definidos via HOMALS (as configurações).

Na prática, a variável de classificação (os *clusters*), salva na base de dados, classifica cada objeto em um dos clusters gerados (é tipológica). Contudo, o método de *clusters* só diz quem é cada indivíduo, a que grupo pertence, mas não informa **como** eles são. A associação com a informação da HOMALS (com as configurações) vai justamente traçar o perfil de todos os clusters gerado pela Análise de *Clusters*, revelando **quem** ou **como** são cada grupo ou perfil ou *clusters*.

Usa-se o Método *K-Means*, um **método não hierárquico**, pois geralmente já se sabe, através da análise do plano da análise da HOMALS, o número de grupos que a Análise de *Clusters* deve segmentar o espaço de análise original.

Caso não se tenha a informação prévia de quantos grupos deve a Análise de *Clusters* segmentar o espaço de análise ou para validar se aquela que se considera a melhor partição é consistente com as conclusões retiradas do plano em estudo da HOMALS, sugere-se, antes, usar um **método hierárquico** de aglomeração, que indique o número de *clusters* ideal para a classificação do espaço de análise.

Para se concluir quanto ao número de *clusters*, pode-se usar a leitura do dendograma, estatística gerada pelo método hierárquico. Em casos em que o número de objetos da base de dados é alto, sugere-se usar um gráfico com as distâncias dos coeficientes de fusão (*Screen Plots*), estatística fornecida pelo *output* da análise. Este gráfico relaciona números de *clusters* em função das distâncias ou coeficientes de fusão. Observam-se neste gráfico a partir de quantos *clusters* as distâncias deixam de ser tão significativas. Os resultados da realização de um método de classificação hierárquico permite, assim, validar a solução sugerida pela HOMALS.

Classificando os casos, é, então, feita a correspondência entre a solução dos *clusters* determinados e os perfis configurados pelo plano da HOMALS. Este procedimento pode ser conseguido pelo cruzamento de cada uma das variáveis usadas na definição dos perfis, via HOMALS, com a variável na base de dados que define os clusters.

Para tanto, são retomadas as variáveis de *input* da HOMALS que mais discriminam nas dimensões retidas e efetuar-se o cruzamento com a variável de *clusters* salva na base de dados. A partir das percentagens mais elevadas para cada um dos *clusters* pode-se fazer uma síntese interpretativa de suas principais características.

Para validar graficamente a consistência desta classificação, pode ainda proceder-se à projeção dos *clusters* no plano das categorias. Do ponto de vista técnico, consiste numa representação gráfica na qual estão simultaneamente representadas as variáveis que configuram os perfis, via HOMALS, como variáveis ativas, e a variável que identifica a tipologia de segmentação, tendo esta o status de variável passiva. O objetivo é observar a disposição dos *clusters* formados pela Análise de *Clusters* com a realização da Análise de Homogeneidade.

A representação gráfica deve evidenciar a tipologia de segmentação coerente com os resultados desde logo obtidos com a realização da Análise de Homogeneidade.

A exemplificação sistematizada evidencia como a gestão complementar entre a HOMALS e a Análise de Clusters pode de fato, desempenhar um papel importante na operacionalização de grupos, na perspectiva de definir tipologias.

Assim, num primeiro momento, afere-se sobre a configuração topológica do espaço de análise, o que implica a operar com múltiplas variáveis qualitativas que interagem de forma sistêmica. Este processo analítico resulta na identificação de diferentes perfis. Num segundo momento, parte-se para o ensaio de tipologias dos objetos do espaço de análise. De cada tipo gerado pelo método não hierárquico de aglomeração é traçado um perfil coerente com o evidenciado pela Análise de Homogeneidade. É este encadeamento operatório que se considera ter uma concretização possível da articulação da Análise de Homogeneidade com a Análise de Clusters.

A par do suporte quantitativo, parte-se para a atribuição de sentido a esses elementos numéricos. Contemplando todos estes elementos, isto é, com a atribuição de sentido as estatísticas geradas, a interpretação incidirá sobre as dimensões e os planos, realizando um processo de redução e estruturação do espaço de análise.

O método da HOMALS, basicamente, então, tem as seguintes fases:

1ª)Preparação da Matriz de Input;

2ª)Identificação das Dimensões mais Importantes;

3ª)Identificação das Variáveis mais Importantes;

4ª)Identificação das Categorias mais Importantes;

5ª)Estruturação do Espaço de Análise em Planos.

6ª)Articulação da Análise de Homogeneidade com a Análise de Clusters.

3 REDUÇÃO DE DIMENSIONALIDADE E ANÁLISE ESTRUTURAL DO PROBLEMA

O objetivo deste estudo é realizar uma análise estrutural quanto à homogeneidade de categorias das variáveis e dos produtos alimentícios envolvidos no levantamento do IDEC (2011).

Antes deve ser feita uma redução da dimensão do problema, para facilitar a estruturação. A Análise de Homogeneidade é a técnica que se propõe a esse intento.

Depois de codificadas as categorias das variáveis, é obtida a “matriz de input” de dados para o SPSS e realização da HOMALS. Os resultados desta operação se encontram na Tabela 3.

Tabela 1- Amostra de Produtos Alimentícios com o Levantamento “uso de imagens de frutas nas embalagens desproporcional ao conteúdo” e “desinformação do percentual de frutas contidos nas embalagens”

Código	Produto	Tipo	Embalagem	Ingredientes	Info%
1	Batavo Kissy Morango	Iogurte(IO)	EMB-S	ING-S	INFO-N
2	Leco Light Lective	Iogurte(IO)	EMB-S	ING-S	INFO-N
3	Danoninho Iogurte	Iogurte(IO)	EMB-S	ING-S	INFO-N
4	Tang	Refresco em Pó(RP)	EMB-S	ING-S	INFO-S
5	La Frutta	Refresco em Pó(RP)	EMB-S	ING-S	INFO-S
6	Camp	Refresco em Pó(RP)	EMB-S	ING-S	INFO-S
7	Gatorate	Isotônico(ISO)	EMB-N	ING-N	INFO-N
8	Taeq	Isotônico(ISO)	EMB-S	ING-N	INFO-N
9	Marathon	Isotônico(ISO)	EMB-N	ING-N	INFO-N
10	Top Fruit	Néctar(NÉC)	EMB-S	ING-S	INFO-N
11	Maguary	Néctar(NÉC)	EMB-S	ING-S	INFO-N
12	Del Valle-Kapo	Néctar(NÉC)	EMB-S	ING-S	INFO-N
13	Frutop	Gelatina(GEL)	EMB-S	ING-N	INFO-N
14	Dr.Oetker	Gelatina(GEL)	EMB-N	ING-N	INFO-N
15	Royal	Gelatina(GEL)	EMB-S	ING-N	INFO-N
16	Napolitano	Sorvete(SOR)	EMB-S	ING-S	INFO-N
17	Napolitano Nestlé	Sorvete(SOR)	EMB-S	ING-N	INFO-N
18	Kibon Balance Napolitano	Sorvete(SOR)	EMB-S	ING-N	INFO-N

Fonte: Instituto Brasileiro de Defesa do Consumidor(Idec).

Tabela 2- Definição das Variáveis e Categorias

Variáveis	Descrição	Categorias
Tipo	Tipo do produto	Iogurte (IO)-1 Refresco em Pó (RP)-2 Isotônico (ISO)-3 Néctar (NÉC)-4 Gelatina (GEL)-5 Sorvete (SOR)-6
Embalagem	Usa imagem da fruta na embalagem?	EMB-S -1 e EMB-N -2
Ingredientes	Tem fruta nos ingredientes?	ING-S -1 e ING-N -2
Info%	Informa o percentual da fruta?	INFO-S -1 e INFO-N -2

Fonte: Instituto Brasileiro de Defesa do Consumidor(Idec)

Tabela 3- Matriz de Input para a Análise de Homogeneidade

Produto	Tipo	Embalagem	Ingrediente	Info%
1	1	1	1	2
2	1	1	1	2
3	1	1	1	2
4	2	1	1	1
5	2	1	1	1
6	2	1	1	1
7	3	2	2	2
8	3	1	2	2
9	3	2	2	2
10	4	1	1	2
11	4	1	1	2
12	4	1	1	2
13	5	1	2	2
14	5	2	2	2
15	5	1	2	2
16	6	1	1	2
17	6	1	2	2
18	6	1	2	2

O desenvolvimento e os resultados da Análise de Homogeneidade dos dezoito produtos da amostra quanto ao *“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”* e *“desinformação do percentual de frutas contidos nas embalagens”* segue adiante. Todos os cálculos e procedimentos da HOMALS realizados foram feitos junto ao pacote estatístico SPSS.

Identificação Das Dimensões Mais Importantes

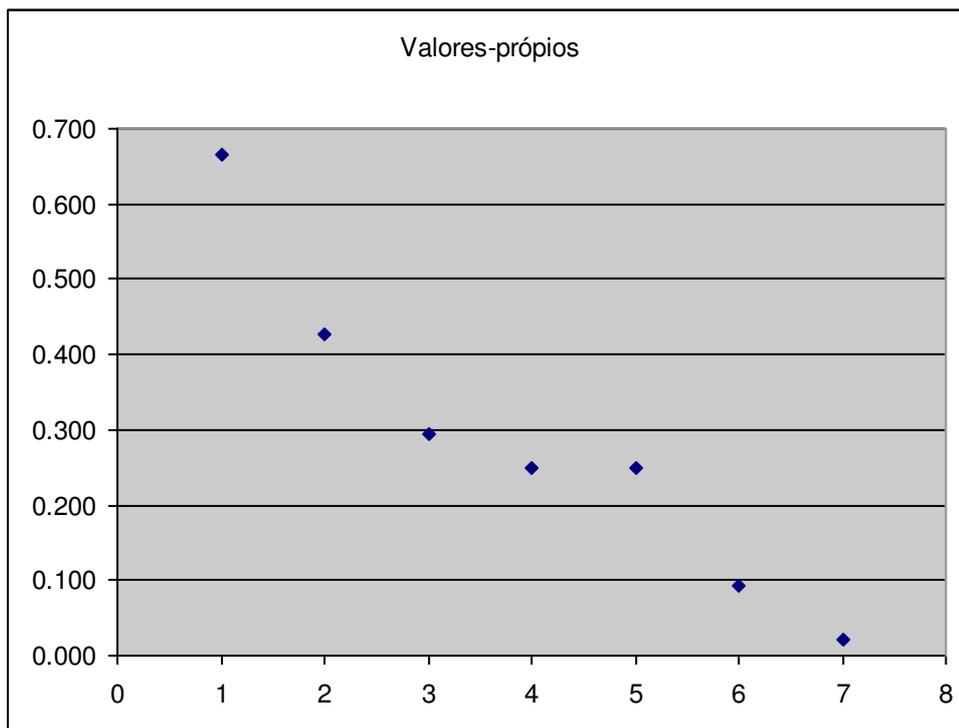
Nesta seção, serão selecionadas as dimensões mais importantes para a estruturação do espaço de análise.

Apresentam-se os valores próprios para o espaço *“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”* e *“desinformação do percentual de frutas contido nas embalagens”*. Foi usado um número elevado de dimensões, a fim de ser possível analisar o comportamento dos valores próprios. Solicitou-se uma solução com 7 dimensões, número de categorias (12) menos o número de variáveis(5), que corresponde ao número máximo que é possível solicitar no pacote estatístico utilizado neste estudo de caso, o SPSS. O SPSS forneceu os valores próprios das 3 primeiras dimensões somente por razões computacionais, mas esse resultado não afeta a análise ora proposta.

Tabela 4- Distribuição dos Valores Próprios

Dimensões	Valores-próprios
1	0.666
2	0.426
3	0.294
4	0.250
5	0.250
6	0.094
7	0.020

Figura 2 - Representação dos Valores Próprios



Pelo que se pode observar da Tabela 4 e Figura 2, é evidente a superioridade das duas primeiras dimensões na explicação do espaço de partida. As duas dimensões juntas representam 1,092 da informação da base de dados original do problema, isto é, 54,6% da variação total do espaço de análise. Portanto, com somente os primeiros eixos estruturantes já se têm uma boa representação das informações da base de partida.

IDENTIFICAÇÃO DAS VARIÁVEIS MAIS IMPORTANTES

Neste momento, vão-se analisar as variáveis mais importantes em cada uma das dimensões mais importantes. Este procedimento permitirá nomear as dimensões e interpretá-las.

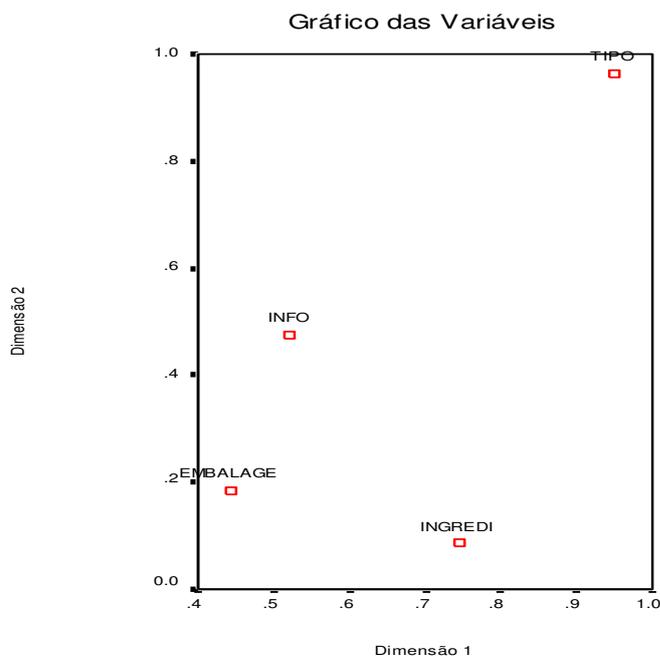
Tabela 5-Medidas de Discriminação

Variáveis	Dimensão 1	Dimensão 2
Tipo	0.951	0.962
Embalagem	0.445	0.184
Ingredientes	0.747	0.086
Info%	0.521	0.474

A Tabela 5 apresenta as medidas de discriminação das dimensões retidas para o problema na fase anterior. As variáveis mais importantes em cada uma das dimensões retidas foram sombreadas em cada uma das dimensões da Tabela 5. Estas são as variáveis mais importantes em cada solução considerada.

Todas as variáveis foram consideradas importantes para discriminação da dimensão 1. Na dimensão 2, somente a variável Ingredientes (Tem frutas nos ingredientes?) não foi contemplada como importante para a discriminação.

Figura 3- Gráfico das Variáveis



As variáveis Ingredientes ((Tem fruta nos ingredientes?)), Embalagem (Usa imagem de fruta na embalagem?), Info%(Informa o percentual de fruta?) estão próximas da dimensão 1. As variáveis Embalagem (Usa imagem de fruta na embalagem?) e Info%(Informa o percentual de fruta?) também estão próximas da dimensão 2. A variável Ingredientes (Tem fruta nos ingredientes?) está distante da dimensão 2. A variável Tipo (Tipo do produto) está na diagonal do plano, evidenciado que discrimina as duas variáveis.

Por existirem muitas variáveis que simultaneamente discriminam as duas dimensões retidas fica difícil usá-las para dá uma nomeação às variáveis.

Identificação Das Categorias Mais Importantes

A consulta das medidas de discriminação é essencial mas não é suficiente, pois não permite, por si só, esclarecer sobre o que se passa com cada uma das categorias. Não será pelo fato de certa variável ter associada uma medida de discriminação elevada que todas as suas categorias serão necessárias, importantes e diferenciadoras dos “cases”. É necessário avaliar as quantificações das categorias, tarefa que pode ser realizada com a observação de sua disposição relativa.

Assim, conjugando os resultados das medidas de discriminação com a análise das quantificações das categorias das variáveis já selecionadas, identificam-se as categorias a privilegiar na leitura das configurações do espaço de produtos alimentícios quanto ao “*uso de imagens de frutas nas embalagens desproporcional ao conteúdo*” e “*desinformação do percentual de frutas contidos nas embalagens*”.

Nesta etapa, é oportuno também avaliar como internamente cada dimensão atua sobre as suas categorias importantes: associando-as ou opondo-as.

Para descrever a dimensão 1, pode-se proceder à análise da Figura 4 .

Figura 4 -Quantificação das Categorias

Categorias	Frequências	Tipo	
		1	2
Iogurte(IO)	3	-0,450	-1,078
Refresco em pó(RP)	3	-1,614	1,540
Isotônico(ISO)	3	1,313	0,839
Néctar(NÉC)	3	-0,450	-1,085
Gelatina(GEL)	3	0,954	0,300
Sorvete(SOR)	3	0,247	-0,515

Categorias	Frequências	Embalagem	
		1	2
Sim, usa imagem da fruta na embalagem(BEM-S)	15	-0,298	-0,192
Não, não usa imagem da fruta na embalagem(BEM-N)	3	1,492	0,959

Categorias	Frequências	Ingredientes	
		1	2
Sim, tem fruta nos ingredientes(ING-S)	10	-0,773	-0,262
Não, não tem fruta nos ingredientes(ING-N)	8	0,966	0,327

Categorias	Frequências	Info%	
		1	2
Sim, informa o percentual de frutas(INFO-S)	3	-1,614	1,540
Não, não informa o percentual de frutas(INFO-S)	15	0,323	-0,308

Primeiro, considera-se todas as variáveis importantes para esta dimensão. Somente vão-se considerar as quantificações das categorias destes indicadores. Segundo, consideram-se na coluna da dimensão 1 em cada variável assinalada anteriormente as categorias afastadas de zero, separando-as em dois blocos: as de **sinal negativo**, em negrito, vão para o da dimensão $1 < 0$, e as de **sinal positivo**, para a dimensão $1 > 0$. Os resultados desta operação então registrados no Quadro 1. Portanto, só vão ser consideradas para a análise as categorias com quantificação altas, isto é, as mais importantes, as que promovem a discriminação dos grupos. Para a formação do Quadro 2, segue-se raciocínio análogo

Quadro 1- Dimensão 1-Tipos de Discriminação

Variáveis	Quantificação das Categorias (QC)	
	QC>0	QC<0
Tipo Embalage Ingredi Info%	Isotônico (ISO) Gelatina (GEL) Sorvete (SOR) Não, não usa imagem da fruta na embalagem (BEM-N) Não, não tem fruta nos ingredientes(ING-N) Não, não informa o percentual de frutas(INFO-S)	Iogurte (IO) Refresco em pó (RP) Néctar (NÉC) Sim, usa imagem da fruta na embalagem (BEM-S) Sim, tem fruta nos ingredientes(ING-S) Sim, informa o percentual de frutas(INFO-S)

A dimensão associa os isotônicos, gelatinas, sorvetes, que não usam imagem da fruta na embalagem, não têm fruta nos ingredientes e não informam o percentual de frutas. Em oposição a este grupo, coloca o iogurte, refresco em pó, néctar, que usam imagem da fruta na embalagem, têm frutas nos ingredientes e informam o percentual de frutas.

Quadro 2- Dimensão 2- Tipos de Discriminação

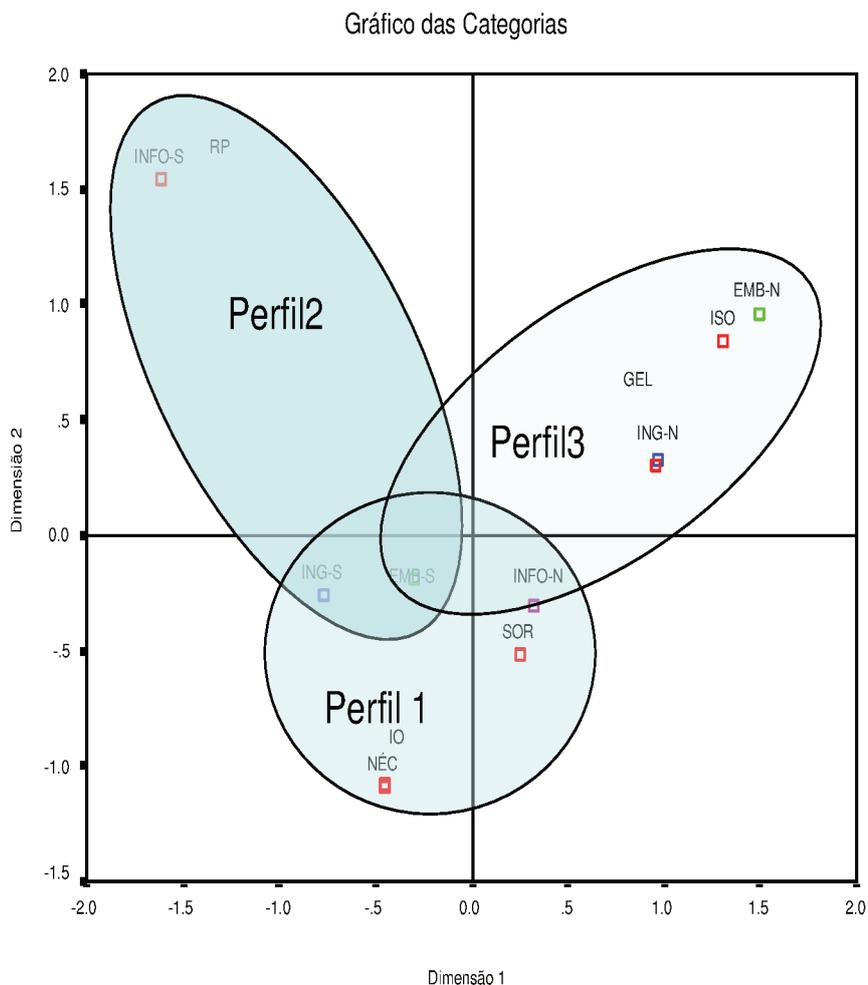
Variáveis	Quantificação das Categorias (QC)	
	QC>0	QC<0
Tipo Embalage Info%	Refresco em pó (RP) Isotônico (ISO) Gelatina (GEL) Não, não usa imagem da fruta na embalagem (BEM-N) Sim, informa o percentual de frutas (INFO-S)	Iogurte (IO) Néctar (NÉC) Sorvete (SOR) Sim, usa imagem da fruta na embalagem (BEM-S) Não, não informa o percentual de frutas (INFO-S)

A dimensão 2 associa o refresco em pó, isotônico, gelatina, que não usam imagem da fruta na embalagem e que informam o percentual de frutas. Em oposição a este grupo, coloca o iogurte, néctar, sorvete, que usam imagem de fruta na embalagem e não informam o percentual de frutas.

Estruturação Do Espaço De Supermercados

Até agora se investigaram as dimensões mais importantes, as variáveis mais importantes e as categorias mais importantes para a topologia do espaço de análise. Os próximos passos da análise consistirão da estruturação do espaço de categorias das variáveis e dos “cases” no que tange ao “uso de imagens de frutas nas embalagens desproporcional ao conteúdo” e “desinformação do percentual de frutas contidos nas embalagens”, no sentido que se procurarão evidências de configurações latentes, não tão evidentes no espaço de partida.

Figura 5 – Gráfico das Categorias



A Figura 5 deixa nítido três perfis distintos coexistem na base de dados analisada.

O *Perfil 1* é formado pelos produtos nas categorias iogurte, néctar, sorvete, que usam fruta na embalagem, usam fruta nos ingredientes e não informam o percentual de fruta. Está presente nos 3º e 4º quadrantes.

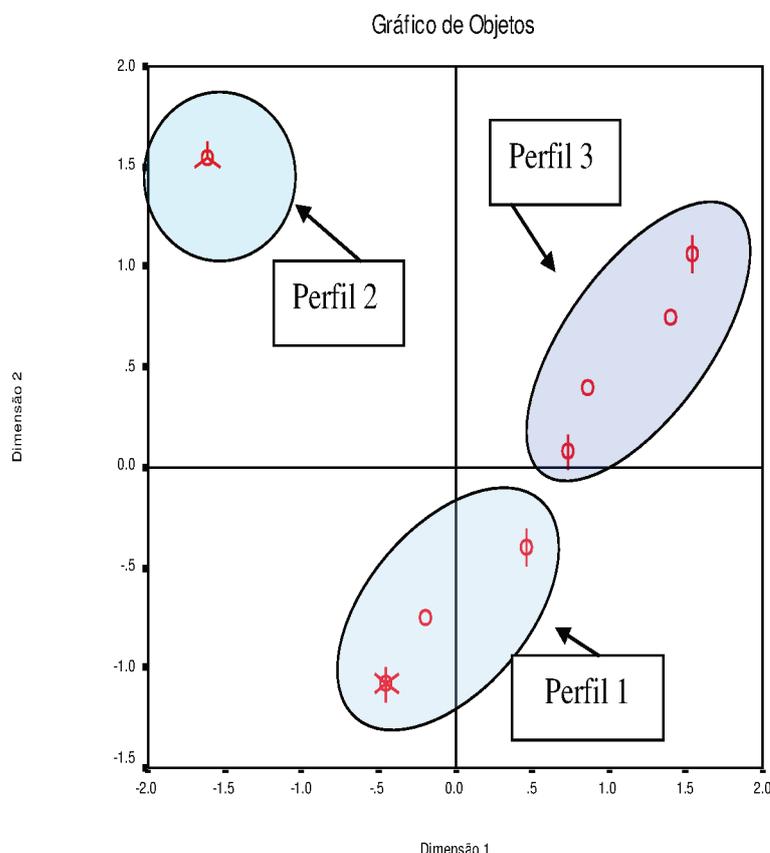
O *Perfil 2* é constituído pelo refresco em pó, que usam fruta na embalagem, usam fruta no ingrediente e informam o percentual de fruta. Está presente nos 2º e 3º quadrantes.

O *Perfil 3* reúne os isotônicos e gelatinas, que usam e não usam frutas na embalagem, que não usam frutas nos ingredientes e não informam o percentual de fruta. Está presente nos 1º, 3º e 4º quadrantes.

Enfim, chegou-se à estruturação em perfis da base de dados. Análise mostrou, como inicialmente se questionou neste artigo, que existe variação em torno do comportamento geral constatado e que três grupos de produtos alimentícios possuem na base de dados perfis distintos, mas que coexistem. As suspeitas levantadas parecem ter tido fundamento.

O interesse agora é saber quem faz parte de cada grupo evidenciado na Figura 5. Esta etapa analítica pode ser alcançada pela evidência do gráfico de objetos.

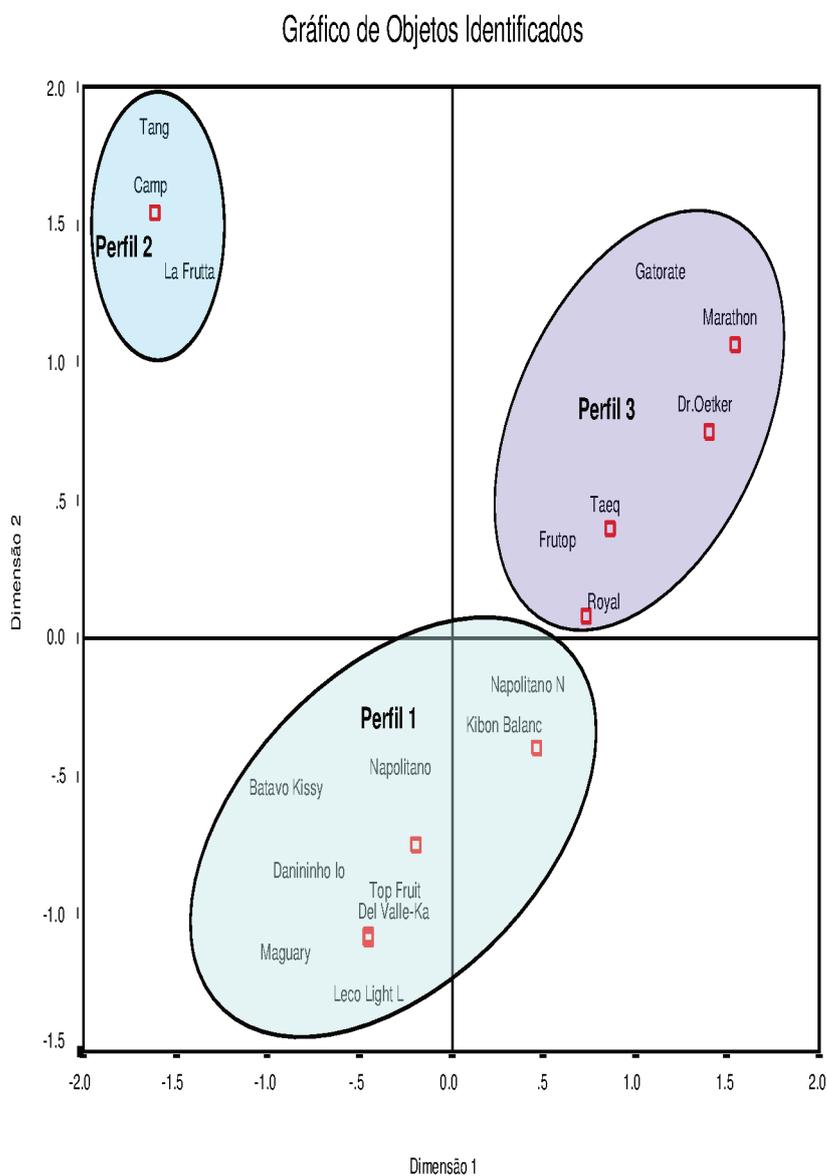
Figura 6 – Gráfico de Objetos



A Figura 6 evidencia três grupos de produtos alimentício. O *Perfil 1* com 9 produtos (3º e 4º quadrantes). O *Perfil 2* com 3 produtos alimentícios (2º quadrante) e finalmente o *Perfil 3*, com 6 produtos alimentícios (1º quadrante). Verifica-se que a maior parte dos “cases” da base de dados estão nos *Perfis 1 e 3*.

É importante neste estágio da análise obter-se a informação de quem são os perfis da Figura 6, principalmente os *Perfis 1 e 3*. Esta informação pode ser obtida através do gráfico de objetos identificados, a Figura 7.

Figura 7 – Gráfico de Objetos Identificados



Então, pela Figura 7, constata-se que existe o *Perfil 1* que é formado pelos produtos alimentícios *Batavo Kissy Morango*, *Leco Light Lective*, *Danoninho Iogurte*, *Top Fruit*, *Maguary*, *Del Valle-Kapo*, *Sorvete Napolitano da Kidelícia de Sabor*, *Sorvete Napolitano da Nestlé* e *Sorvete Kibon Balance Napolitano*. O *Perfil 2* é constituído pelos produtos alimentícios *Tang*, *La Fruta* e *Camp*. Finalmente, o *Perfil 3* é composto pelos produtos alimentícios *Gatorate*, *Taeq*, *Marathon*, *Frutop*, *Dr. Oetker* e *Royal*.

Associando as informações dos gráficos de categorias e objetos têm-se a estruturação completa do espaço de partida.

O *Perfil 1* é formado pelos produtos alimentícios *Batavo Kissy Morango*, *Leco Light Lective*, *Danoninho Iogurte*, *Top Fruit*, *Maguary*, *Del Valle-Kapo*, *Sorvete Napolitano da Kidelícia de Sabor*, *Sorvete Napolitano da Nestlé* e *Sorvete Kibon Balance Napolitano* e tem a característica de ser formado pelos produtos nas categorias iogurte, néctar, sorvete, que usam fruta na embalagem, usam fruta nos ingredientes e não informam o percentual de fruta. Está presente nos 3º e 4º quadrantes. Constitui **50%** dos “cases” da base de dados.

O *Perfil 2* é constituído pelos produtos alimentícios *Tang*, *La Fruta* e *Camp* nas categorias refresco em pó, que usam fruta na embalagem, usam fruta no ingrediente e informam o percentual de fruta. Está presente nos 2º e 3º quadrantes. Constitui **17%** dos “cases” da base de dados.

Finalmente, o *Perfil 3* é composto pelos produtos alimentícios *Gatorate*, *Taeq*, *Marathon*, *Frutop*, *Dr. Oetker* e *Royal* nas categorias de isotônicos e gelatinas, que divergem quanto ao uso de frutas na embalagem, que não usam frutas nos ingredientes e não informam o percentual de fruta. Está presente nos 1º, 3º e 4º quadrantes. Constitui **33%** dos “cases” da base de dados.

Portanto, como já havia sido revelado pelo levantamento do IDEC (2011), a grande maioria dos produtos investigados (83%) está no perfil geral de **“uso de imagens de frutas nas embalagens desproporcional ao conteúdo”** e **“desinformação do percentual de frutas contidos nas embalagens”**.

Contudo, como este estudo problematizou, existe exceção à regra, isto é, um *subperfil* que apresenta variação do perfil geral: o *Perfil 2*. Esta configuração de produtos alimentícios, basicamente os refrescos em pó, usa **“imagens de frutas nas embalagens proporcional ao conteúdo”** e informa o **“percentual de frutas contidos nas embalagens”**. Esta informação é importante às tomadas de decisão com respeito às políticas adotadas no sentido da defesa dos direitos do consumidor.

O gráfico das categorias informou as características marcantes de todos os perfis ou configuração coexistentes no mesmo espaço. O gráfico de objetos revelou os produtos base

de dados com perfis homogêneos, inclusive identificando-os. Associou-se os dois gráficos e se obteve cada grupo de produtos da amostra investigada segundo o seu perfil.

Para validar os resultados encontrados na Análise de Homogeneidade que foi finalizada nesta etapa, pode-se recorrer a Articulação da Análise de Homogeneidade com a Análise de Clusters e se potencializar a confiabilidade dos resultados encontrados, permitindo se tecer teorias sobre o problema para fins de tomadas de decisão mais seguras. È que se propõe a seção seguinte.

Articulação Da Homals Com Análise De Clusters

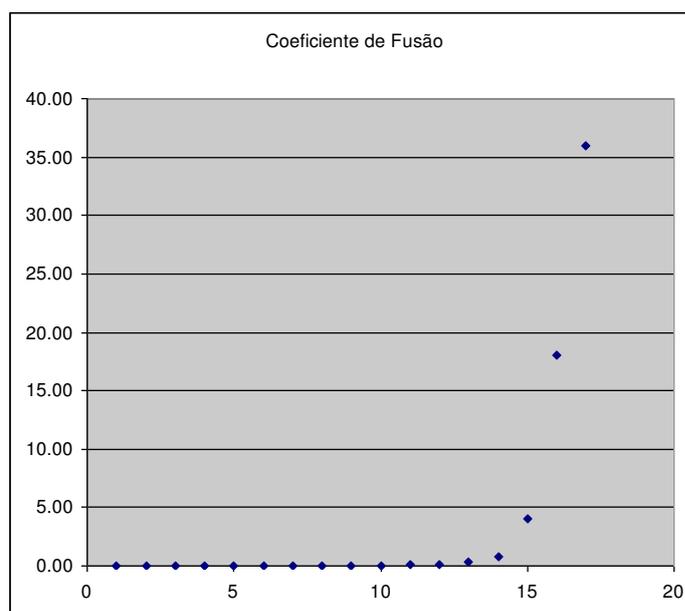
Nos parágrafos posteriores, vai-se associar as configurações identificadas na HOMALS com os clusters gerados na Análise de Clusters, para validar os resultados da Análise de Homogeneidade. Para tanto foi realizada uma análise classificatória dos produtos da amostra aplicando a Análise de Clusters, em dois estágios:

Método Hierárquico, para validar o número de clusters a ser segmentada a base de dados;

Método não hierárquico K-Means, para a classificação propriamente dita.

Aplicando o método hierárquico, foram gerados os coeficientes de fusão, para qual foi construído o Gráfico da Figura 8. Verifica-se, neste gráfico, que para mais de três grupos, as distâncias deixam de ser significativas. Pode observar-se que o declive é bastante menos acentuado a partir do quarto agrupamento. Portanto, a análise de clusters hierárquico sugere para o agrupamento dos produtos da amostra **três** clusters.

Figura 8 – Gráfico dos Coeficientes de Fusão



Concluída esta demonstração, retoma-se a solução otimizada com a realização do método de *K-Means não hierárquico*. Classificados os “cases”, é então necessário fazer a correspondência entre a solução de três clusters e os três perfis configurados pelo plano da HOMALS. Esta operação pode ser realizada através do cruzamento de cada uma das variáveis usadas na definição dos perfis, com a aplicação da HOMALS, com os clusters. Esta operação encontra-se na Tabela 6.

Tabela 6- Caracterização dos *Clusters*

Categorias das Variáveis	Clusters 1	Clusters 2	Clusters 3
Iogurte-IO	33	0	0
Refresco em Pó-RP	0	100	0
Isotônico-ISSO	0	0	50
Néctar-NÉC	33	0	0
Gelatina-GEL	0	0	50
Sorvete-SOR	33	0	0
Usa a Imagem da Fruta na Embalagem-EMB-S	100	100	50
Não Usa a Imagem da Fruta na Embalagem-EMB-N	0	0	50
Tem Frutas nos Ingredientes-ING-S	78	100	0
Não Tem Frutas nos Ingredientes-ING-N	22	0	100
Informa o Percentual da Fruta-INFO-S	0	100	0
Não Informa o Percentual da Fruta-INFO-N	100	0	100

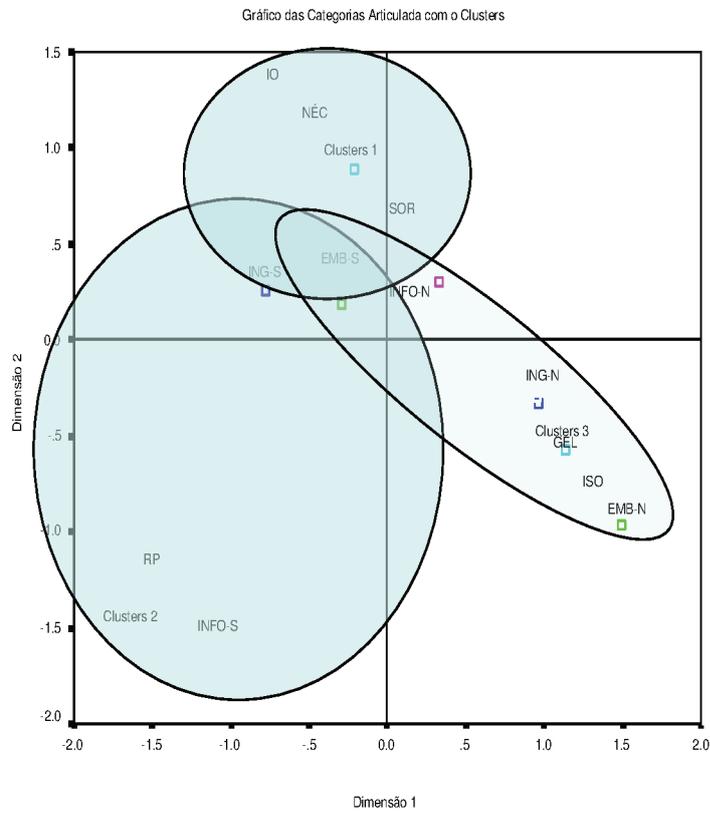
O *Clusters 1* é formado pelos produtos nas categorias iogurte, néctar, sorvete, que usam fruta na embalagem, usam fruta nos ingredientes e não informam o percentual de fruta. O *Clusters 2* é constituído pelo refresco em pó, que usam fruta na embalagem, usam fruta no ingrediente e informam o percentual de fruta. O *Clusters 3* reúne os isotônicos e gelatinas, que usam e não usam frutas na embalagem, que não usam frutas nos ingredientes e não informam o percentual de fruta. Pode-se observar que estas características de clusters gerados pela Análise de Clusters é exatamente igual ao perfil gerado pelo gráfico das categorias da Análise de Homogeneidade.

Para validar graficamente a consistência desta classificação, pode-se ainda proceder-se à projeção simultânea dos três clusters da Análise de Clusters com os três perfis da Análise de Homogeneidade num mesmo plano bidimensional.

Do ponto vista técnico, consiste numa representação gráfica na qual estão ao mesmo tempo representadas as variáveis que configuram os perfis, como variáveis **ativas**, e a variável que identifica os clusters, como variável **passiva**. A representação permitirá verificar que a posição ocupada por cada um dos três clusters está muito próxima dos três perfis da HOMALS.

Conforme evidencia a representação gráfica (Figura 9) os clusters da Análise de Clusters estão bem próximos do perfis da Análise de Homogeneidade.

Figura 9- Gráfico das Categorias Articulas com o Clusters



Portanto, a tipologia de segmentação afigura-se coerente com os resultados topológicos obtidos com a realização da Análise de Homogeneidade, imprimindo confiabilidade aos resultados.

4 CONCLUSÃO

A pesquisa do Idec deu um passo inicial no entendimento de como os produtos alimentícios de uma amostra considerada se comprometem com informação dada ao consumidor no que concerne ao uso de imagens de frutas nas embalagens e o seu efetivo conteúdo e à informação do percentual de frutas presentes nas embalagens comercializadas.

O Idec mostra que os dezoito produtos da amostra apresentam foto de fruta na embalagem, cor de fruta, sabor de fruta, mas não tem fruta. Além disso, não informam o percentual de frutas contidas no conteúdo da mercadoria. É patente o baixo grau de comprometimento do setor com o tema.

O presente artigo levantou a questão adicional dos produtos investigados possuírem o perfil de “propaganda enganosa” como um todo homogêneo, ou se este perfil se segmenta

com variações diferenciadas entre os produtos, revelando uma estrutura com configurações distintas, coexistentes num mesmo espaço de análise.

O que se sinalizou é que podem existir exceções à regra e que seria oportuno se investigar se há variação do comportamento geral constatado, em que grau e que grupos de produtos apresentam dispersões semelhantes.

A conclusão a que se chegou com o estudo é que realmente a grande maioria dos produtos investigados (83%) está no perfil geral do Idec: “*uso de imagens de frutas nas embalagens desproporcional ao conteúdo*” e “*desinformação do percentual de frutas contidos nas embalagens*”. Mas existe, sim, latente na base de dados, um *subperfil* que apresenta variação do perfil geral: o *Perfil 2*. Esta configuração de produtos alimentícios, basicamente os refrescos em pó, usam “*imagens de frutas nas embalagens proporcional ao conteúdo*” e informam o “percentual de frutas contidos nas embalagens”.

Esta revelação da investigação é importante às tomadas de decisão com respeito às políticas adotadas no sentido da defesa dos direitos do consumidor.

ABSTRACT: *Currently there is disagreement between what should be formally exposed by the industries regarding the composition of their products and what is actually offered to the final consumer. The products have photo of fruit on the packaging, fruit color, fruit flavour, etc., but contains no fruit. This was the finding of the Office of consumer protection in relation to the investigation of 18 industrialized food products valued at their latest survey. This article aims to further the study of Idec to investigate whether the profile of "false advertising" reach as a whole the sample under study or if this profile segments with different variations between products, revealing a structure with different configurations, coexisting in the same space. To achieve the goal, it was decided to reduce the dimensionality of the space of analysis in order to understand more clearly and quickly coexisting, and thus settings should undertake a study as well as structural variables are qualitative research, multivariate techniques appropriate to the intent is the analysis of homogeneity. The conclusion that has been reached with the structuring of the problem is that actually the vast majority of the products investigated (83%) are in the General profile of Idec: "use of images of fruit on the packaging the content" and disproportionate "misinformation of the percentage of fruit contained in packaging". But there is, Yes, latent in the database, a sub-profile that presents the variation of General profile: the Profile 2. This setting of food products, basically the powdered juices, use "pictures of fruit on the packaging is in proportion to the content" and tell the "percentage of fruits contained in packaging". This revelation of the investigation is important to decision making with respect to the policies adopted to protect consumer rights and enhancement of quality of life.*

Key-words: *industrial products, nutritional information of concentration and percentage of fruits, structural analysis, analysis of homogeneity.*

REFERÊNCIAS

CARVALHO, HELENA. Análise Multivariada de Dados Qualitativos- **Utilização da HOMALS com SPSS**. Lisboa: Edições Silabo, 2004.

- GEER, JOHN. **Multivariate Analysis of Categorical Data: Theory**. USA: Sage Publications, 1993.
- GIFI, ALBERT. **Nonlinear Multivariate Analysis**. England: John Wiley & Sons, 1996.
- GREENACRE, MICHEL. **Theory and Applications of Correspondence Analysis**. London: Academic Press, 1984.
- INSTITUTO BRASILEIRO DE DEFESA DO CONSUMIDOR. **Pesquisa de Sustentabilidade das Dez Maiores Redes de Supermercados do Brasil**. 2011.

Submetido em 11 dez. 2012; Aceito para publicação em 14 abr. 2014.