

**AUTORES**  
**AUTHORS**

**Marcelo Antônio MORGANO**

Mestre em química, Centro de Química de Alimentos e Nutrição Aplicada, ITAL - CP 139 CEP 13073-001 – Campinas – SP  
Fone: 2415222 - Fax: 2424585  
e-mail: morgano@ital.org.br

**Sônia Cláudia do N. de QUEIROZ**

Mestre em química, Instituto de Química, UNICAMP CP 6154 – CEP 13081-970 – Campinas – SP  
Fone: 7883044 - Fax: 7883023  
e-mail: soniacnq@iqm.unicamp.br

**Márcia Miguel Castro FERREIRA**

Profa. Dra., Instituto de Química, UNICAMP CP 6154 – CEP 13081-970 – Campinas – SP  
Fone: 7883102 - Fax: 7883023  
e-mail: marcia@iqm.unicamp.br

**RESUMO**

Neste artigo é apresentada uma aplicação de métodos estatísticos e matemáticos em dados provenientes de análises de alimentos. Um conjunto de 17 amostras de vegetais e 15 variáveis foi estudado pela aplicação dos métodos de Análise por Agrupamento Hierárquico (AAH) e Análise por Componentes Principais (ACP). Aspectos teóricos de ambos os métodos foram abordados. A análise exploratória de dados multivariados tem se mostrado de grande importância para a ciência de alimentos que necessita cada vez mais de informações rápidas sobre a qualidade de seus produtos. Dos vegetais estudados constam grãos (diferentes feijões, lentilhas e grão-de-bico), tubérculos (batata, batata doce, cenoura, beterraba, inhame) e verduras (alface, chicória, repolho, aipo, couve-flor e brócolis). Entre as variáveis, estão os teores de cinco metais (Na, K, Ca, Mg, Fe), os teores de fósforo, nitrogênio total, umidade, gordura, fibra, carboidratos, proteína, caloria, tiamina e niacina. O método de AAH agrupou as amostras em quatro grupos distintos, um contendo os grãos, dois com tubérculos e um com as verduras. A análise por componentes principais, além de formar grupos semelhantes, identificou por meio dos *loadings*, quais variáveis foram mais importantes para a separação das amostras nestes grupos. Foram analisados separadamente os grãos, as verduras e os legumes para melhor enfatizar as características de cada grupo.

**SUMMARY**

This article reports the application of statistical and mathematical methods to food analysis. Seventeen samples of vegetables and fifteen variables were analyzed by Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) and the theoretical basis of the methods was described. Exploratory data analysis has become very important in food science as a method to quickly obtain and use information on food quality. The types of vegetables studied included grains (black bean, chickpea, red lentil, brown-green lentil, azuki bean, soybean), tubers (potato, sweet potato, carrot, beet, yam) and greens (lettuce, chicory, cabbage, celery, cauliflower, broccoli). Among the variables are the content of five metals (potassium, sodium, calcium, magnesium, and iron), total nitrogen and phosphorus, moisture, carbohydrates, protein, fat, calories, dietary fiber, thiamin and niacin. HCA showed four different clusters, one with the grains, two containing the tubers and one with the greens. PCA showed similar results and the loadings indicate the main variables responsible for the sample clustering. The grains, tubers and greens were analyzed separately in order to focus the characteristics of each group.

**PALAVRAS-CHAVE**  
**KEY WORDS**

Análise por Agrupamento Hierárquico; Análise por Componentes Principais; Análise multivariada; Vegetais / Hierarchical cluster analysis; Principal component analysis; Multivariate analysis; Vegetables.

## 1. INTRODUÇÃO

A indústria de alimentos necessita cada vez mais de informações rápidas sobre a qualidade de seus produtos. A identificação dos atributos que influenciam no sabor e aroma dos alimentos é cada vez mais estudado, a fim de se obter um produto com a qualidade desejada pelos consumidores. Entretanto, em algumas situações, a quantificação da qualidade dos alimentos é difícil de se conseguir pelo uso de instrumentos analíticos. Desta forma, os cientistas têm recorrido a uma série de medidas indiretas, onde muitas vezes são necessários diversos testes para se chegar a alguma conclusão a respeito da qualidade do alimento. Isto requer um alto custo e mão-de-obra especializada. Desta forma, a quimiometria pode ser aplicada com a finalidade de reduzir o número de determinações e, conseqüentemente, o custo final do produto.

A quimiometria é uma ferramenta matemática e estatística freqüentemente utilizada para maximizar as informações que podem ser extraídas de um conjunto de dados. Usando seus recursos, os métodos de desenvolvimento de produtos alimentícios e análises de alimentos em geral são simplificados. Como exemplo de aplicações, pode-se citar: controle e monitoramento de processos (CATET *et al.*, 1990, ROBERT *et al.*, 1987); classificação de alimentos e determinação da origem geográfica (HEADLEY, HARDY, 1989, HERRANZ *et al.*, 1990, ALVAREZ, 1991, STENROOS, SIEBERT, 1984, SCARMINO *et al.*, 1982); avaliação sensorial (HOUGH *et al.*, 1996, KJOLSTA *et al.*, 1990, GERBI *et al.*, 1997, DEVER, CLIFF, 1995); determinação de fraude de alimentos (BRIANDET *et al.*, 1996) e outros. Nestes casos, o uso da quimiometria ajuda a decidir quais determinações são importantes e, assim, algumas delas podem ser suprimidas. Isto pode ser facilmente visualizado em gráficos bidimensionais, contendo grande parte das informações estatísticas pelo uso da técnica de Análise por Componentes Principais. Pode-se, também, agrupar as amostras por similaridade e representar esta informação na forma de um diagrama bidimensional denominado de dendograma. Neste caso, a técnica empregada é a análise por agrupamento hierárquico. Estes métodos têm recebido recentemente grande destaque na área de ciência de alimentos e no presente trabalho, pretende-se introduzir os seus conceitos básicos utilizando como aplicação um conjunto de dados, descrevendo a composição química de diferentes tipos de vegetais.

### 1.1 Análise por Componentes Principais – ACP

A base fundamental da maioria dos métodos modernos para tratamento de dados multivariados é a Análise por Componentes Principais - ACP (SHARAF *et al.*, 1986, BEEBE *et al.*, 1998), que consiste numa transformação da matriz de dados com o objetivo de representar as variações presentes em muitas variáveis, através de um número menor de "fatores". Constrói-se um novo sistema de eixos (denominados rotineiramente de fatores, componentes principais, variáveis latentes ou ainda autovetores) para representar as amostras, no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões. Estas novas variáveis são obtidas em ordem decrescente de quantidade de informação estatística que

descrevem, ou seja, a primeira componente principal aponta a direção de maior variação dos dados, a segunda, que é ortogonal à primeira, aponta outra direção que descreve a maior variação restante dos dados e assim por diante.

Usando a notação matricial, as componentes principais são obtidas por meio de transformações lineares conforme a equação:

$$\mathbf{X P} = \mathbf{T} \quad (1)$$

em que  $\mathbf{X}$  é a matriz original dos dados,  $\mathbf{T}$  é a matriz de *scores* que contém as coordenadas das amostras no novo sistema de eixos e  $\mathbf{P}$  é a matriz dos *loadings*, onde os elementos de cada coluna correspondem aos coeficientes das combinações lineares das variáveis originais.

O método PCA está baseado na correlação entre variáveis e na realidade agrupa aquelas que estão altamente correlacionadas. As colunas da matriz de *loadings*, correspondem aos autovetores da matriz de correlação (ou variância e covariância dependendo do pré-processamento utilizado), enquanto os autovalores representam uma fração da variância total dos dados. Assim, cada componente principal descreve uma porcentagem da variância total dos dados.

Uma das vantagens destas transformações é que ruídos experimentais podem ser eliminados, pois estes não estão correlacionados com as informações contidas na matriz de dados original. Outra vantagem é que podem ser escolhidas as variáveis originais mais importantes do ponto de vista estatístico.

Os dados experimentais originais podem não ter uma distribuição adequada para a análise, dificultando a extração de informações úteis e interpretação dos mesmos. Nestes casos, um pré-processamento nos dados originais pode ser de grande valia. Medidas em diferentes unidades e variáveis com diferentes variâncias são algumas das razões que levam a estes problemas. Os métodos de pré-processamento mais utilizados consistem basicamente em centrar na média ou auto-escalar os dados. No primeiro caso, calcula-se a média<sup>1</sup> dos valores experimentais para cada variável e subtrai-se cada valor experimental do respectivo valor médio. Auto-escalar significa centrar os dados na média e dividi-los pelo respectivo desvio-padrão<sup>1</sup>, sendo um para cada variável.

### 1.2 Análise por Agrupamento Hierárquico – AAH

Esta é uma ferramenta excelente para análise preliminar dos dados (SHARAF *et al.* 1986, BEEBE *et al.*, 1998) sendo útil para determinar a semelhança entre objetos e identificar amostras anômalas. O método relaciona as amostras de forma que as mais semelhantes são agrupadas entre si com relação às variáveis usadas no processo de agrupamento. Cada amostra é

<sup>1</sup>Para definição de média e desvio-padrão, o leitor deve consultar qualquer livro básico de estatística, por exemplo J. E. Freund & G. A. Simon, "Modern Elementary Statistics", 9ª Ed. Prentice Hall 1997.

representada como um ponto no espaço de  $m$  variáveis e, portanto, a questão fundamental reside em transformar a informação de um espaço  $m$ -dimensional para um espaço bidimensional. O parâmetro usado neste processo é a medida de distância entre as amostras. Estas distâncias, quando visualizadas através de diagramas conhecidos por dendogramas, indicam as amostras mais semelhantes entre si, ou seja, quanto menor a distância entre os pontos, maior será a semelhança entre as mesmas. Para proceder à análise hierárquica por agrupamento, é necessário estabelecer o conceito de distância entre duas amostras  $a$  e  $b$  ( $d_{ab}$ ) em um espaço de  $m$  variáveis. Há várias medidas usuais e neste trabalho será utilizada a distância euclideana, que pode ser determinada pela expressão matemática:

$$d_{ab} = \left( \sum_{j=1}^m (x_{aj} - x_{bj})^2 \right)^{1/2} \quad (2)$$

em que  $x_{aj}$  é a  $j$ -ésima coordenada da amostra  $a$ .

No método de análise por agrupamento hierárquico as distâncias entre todas as amostras são calculadas e comparadas. Como as distâncias reais entre as amostras mudam com o tipo e o tamanho do conjunto de dados, é necessário transformá-las numa escala de similaridade que é calculada segundo a expressão:

$$s_{ab} = 1 - \{ d_{ab} / d_{max} \} \quad (3)$$

onde:  $s_{ab}$  é a similaridade entre as amostras  $a$  e  $b$ ,  $d_{ab}$  é a distância entre as mesmas e  $d_{max}$  é a maior distância entre todas as amostras. Nestes diagramas, a escala varia de um (amostras idênticas) a zero (amostras sem similaridade).

Após o cálculo das similaridades as duas amostras mais próximas são conectadas formando um agrupamento. Este processo é repetido até que todas as amostras sejam conectadas

formando um único grupo. Uma vez que as amostras são conectadas pela proximidade entre elas, é necessário definir a distância entre uma amostra e um grupo ou entre grupos de amostras e há várias técnicas para tal. As mais usuais são: conexão pelo vizinho mais próximo (*single linkage* ou *nearest neighbour*); conexão pelo vizinho mais distante (*complete link* ou *farthest neighbour*); conexão pela distância média (*average link*). Quando a distância entre dois agrupamentos for obtida pela distância média, pode-se utilizar a média aritmética ou ponderada. No primeiro caso, cada agrupamento tem o mesmo peso e no último, o coeficiente de ponderação é correspondente ao número de elementos nos respectivos *clusters* (agrupamentos) e, neste caso, o método é denominado *incremental*.

## 2. METODOLOGIA

### 2.1 Amostras

Os dados utilizados neste estudo foram extraídos de uma tabela de composição de alimentos (HOLLAND *et al.*, 1994). Foram selecionadas 17 amostras de vegetais e 15 variáveis, sendo apresentadas nas Tabelas 1 e 2. Os vegetais "in natura" escolhidos foram inhame, batata, batata-doce (batatadoc), repolho, couve-flor (couveF), brócolis, alface, chicória, beterraba (beter), aipo, cenoura, grão-de-bico (GrãoBico), feijão preto (feijãoP), lentilha vermelha (lentv), lentilha verde e marrom (lentvm), feijão azuki (feijãoA) e soja. As variáveis medidas para estas amostras foram: teores de umidade (Umid), proteína (Prot), gordura (Gord), carboidrato (Carb), calorias (Kcal), nitrogênio total (N), fibra (Fibra), sódio (Na), potássio (K), cálcio (Ca), magnésio (Mg), fósforo (P), ferro (Fe), tiamina (Tiam) e niacina (Niac).

TABELA 1. Composição centesimal dos vegetais (HOLLAND *et al.*, 1994).

Vegetais	Umidade (g/100g)	Proteína (g/100g)	Gordura (g/100g)	Carboidrato (g/100g)	Calorias (Kcal/100g)	N total (g/100g)	Fibra (g/100g)
Feijão azuki	12,7	19,9	0,5	50,1	272	3,18	11,1
Feijão preto	12,5	24,9	1,4	40,8	275	3,98	17,9
Lentilha verde marrom	10,8	24,3	1,9	48,8	297	3,90	8,9
Lentilha vermelha	11,1	23,8	1,3	56,3	318	3,80	10,5
Soja	8,5	35,9	18,6	15,8	370	5,74	15,7
Batata	81,7	1,7	0,3	16,1	70	0,28	1,3
Cenoura	88,8	0,7	0,5	6,0	30	0,11	2,6
Aipo	95,1	0,5	0,2	0,9	7	0,08	1,6
Brócolis	88,2	4,4	0,9	1,8	33	0,71	2,6
Couve-flor	88,4	3,6	0,9	3,0	34	0,58	1,9
Chicória	94,3	0,5	0,6	2,8	11	0,09	0,9
Alface	95,1	0,8	0,5	1,7	14	0,13	1,3
Grão-de-bico	10,0	21,3	5,4	49,6	320	3,42	13,5
Repolho	90,1	1,7	0,4	4,1	26	0,28	2,9
Beterraba	87,1	1,7	0,1	7,6	36	0,27	2,8
Batata-doce	73,7	1,2	0,3	21,3	87	0,19	2,3
Inhame	67,2	1,5	0,3	28,2	114	0,25	3,7

**TABELA 2.** Composição de micronutrientes presentes nos vegetais (HOLLAND *et al.*, 1994).

Vegetais	Minerais (mg/100g)						Vitaminas (mg/100g)	
	Na	K	Ca	Mg	P	Fe	Tiamina	Niacina
Feijão azuki	5	1220	84	130	380	4,2	0,45	2,6
Feijão preto	40	800	150	160	370	6,3	0,42	2,0
Lentilha verde marrom	12	940	71	110	350	11,1	0,41	2,2
Lentilha vermelha	36	710	51	83	320	7,6	0,50	2,0
Soja	5	1730	240	250	660	9,7	0,61	2,2
Batata	11	320	6	14	34	0,3	0,15	0,4
Cenoura	40	240	34	9	25	0,4	0,04	0,2
Aipo	60	320	41	5	21	0,4	0,06	0,3
Brócolis	8	370	56	22	87	1,7	0,10	0,9
Couve-flor	9	380	21	17	64	0,7	0,17	0,6
Chicória	1	170	21	6	27	0,4	0,14	0,1
Alface	3	220	28	6	28	0,7	0,12	0,4
Grão-de-bico	39	1000	160	130	310	5,5	0,39	0,4
Repolho	5	270	52	8	41	0,7	0,15	1,9
Beterraba	66	380	20	11	51	1,0	0,01	0,1
Batata-doce	40	370	24	18	50	0,7	0,17	0,5
Inhame	2	380	15	15	27	0,7	0,16	0,2

## 2.2 Análise estatística

O tratamento estatístico dos dados (ACP e AAH) foi feito utilizando o programa computacional PIROUETTE 1997.

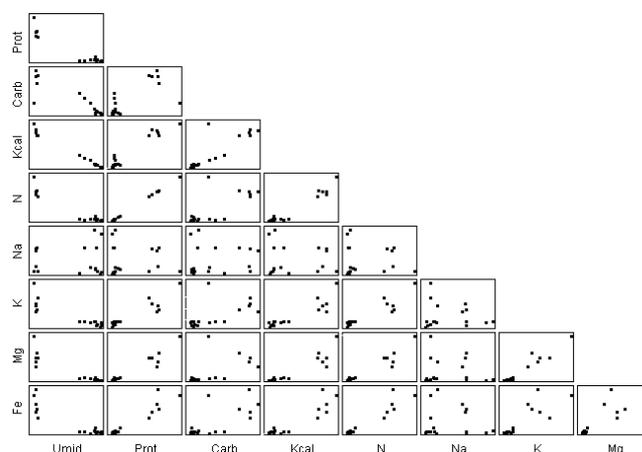
Os dados foram previamente auto-escalados, uma vez que há uma grande variação das respostas das diversas variáveis, ou seja, diferem em ordem de grandeza. Para a obtenção do dendograma foram utilizados a distância euclidiana e o método de conexão incremental, introduzidos na seção anterior.

## 3. RESULTADOS E DISCUSSÃO

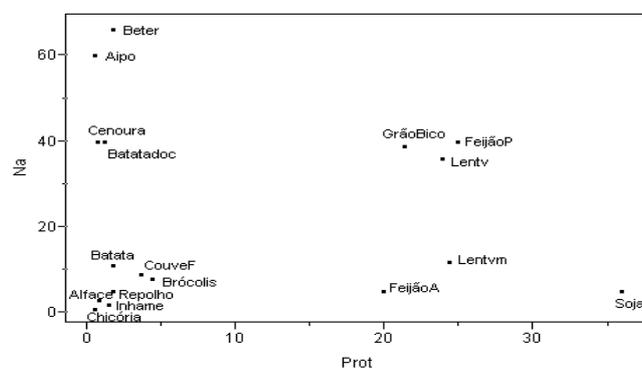
A análise exploratória foi feita utilizando as 17 amostras e 15 variáveis mencionadas anteriormente,  $X = (n=17 \times m=15)$ .

Visualizando os dados originais antes de qualquer tratamento, já é possível obter informações úteis em diagramas de variáveis *versus* variáveis. A Figura 1 mostra que há correlação entre algumas variáveis como, por exemplo, os teores de nitrogênio total e proteínas, as quantidades de calorías e umidade e as quantidades de calorías e carboidratos, indicando que os teores de nitrogênio total e proteínas terão *loadings* semelhantes na análise de componentes principais e as quantidades de calorías e umidade apresentarão *loadings* com sinais contrários, uma vez que são inversamente proporcionais entre si. Pode-se notar que nem sempre há correlações, por exemplo entre os teores de sódio e de carboidratos, teores de magnésio e ferro, teores de ferro e carboidratos. Observa-se, também, que os grãos estão separados dos outros vegetais para algumas variáveis. Um exemplo é ilustrado na Figura 2 onde os grãos apresentam um teor muito maior de proteínas, enquanto as verduras tendem a ter um menor teor de sódio. Entretanto, esta análise de variável *versus* variável nem sempre

é capaz de captar detalhes que poderão ser visualizados após as análises por ACP e AAH, nas quais as variáveis são consideradas simultaneamente.



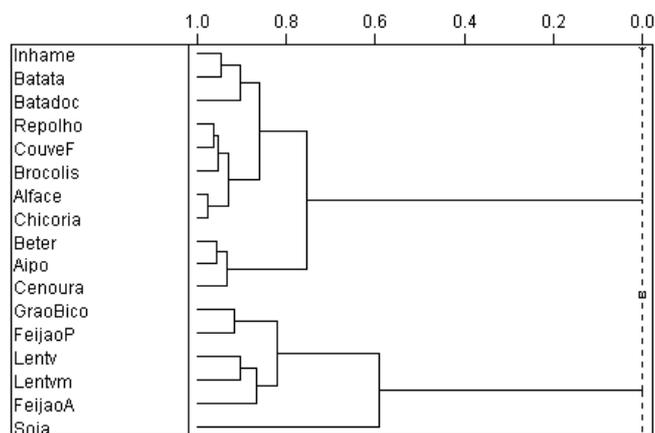
**FIGURA 1.** Gráfico de variáveis *versus* variáveis.



**FIGURA 2.** Gráfico do teor de proteínas *versus* teor de sódio.

Como pode ser observado na Figura 3, o dendograma mostra a similaridade entre os vegetais segundo as variáveis escolhidas, sendo que a escala (0,0 – 1,0) acima do dendograma representa os índices de similaridade. A maior similaridade obtida foi entre a alface e a chicória, com índice de similaridade de 0,98, o que é justificado por serem verduras do tipo folha verde com composições similares.

Observou-se, também, que existe uma separação entre dois grupos principais: grão-de-bico, feijões, lentilhas e soja formando um agrupamento e os outros vegetais formando um outro, o que é esperado, uma vez que alguns destes vegetais são grãos e outros são legumes ou verduras. No grupo dos legumes e verduras, a alface e a chicória se agrupam com a couve-flor, brócolis e repolho, com grau de similaridade de 0,93, enquanto os tubérculos, para este mesmo índice de similaridade, formam dois grupos: batata, inhame e batata-doce; beterraba, aipo e cenoura.



**FIGURA 3.** Dendograma obtido para as 17 amostras de vegetais. A escala acima do dendograma indica os índices de similaridade.

Pela Análise por Componentes Principais, Tabela 3, observou-se que 95,8% da variância dos dados foram explicados nas três primeiras componentes principais.

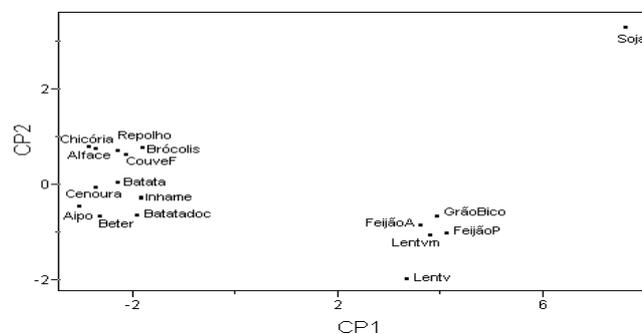
**TABELA 3.** Porcentagem de variância dos dados explicada pela análise por componentes principais.

Componente Principal	% de Variância	% de Variância Acumulada
CP1	80,1	80,1
CP2	9,1	89,2
CP3	6,6	95,8
CP4	1,6	97,3
CP5	1,1	98,9
CP6	0,8	99,3

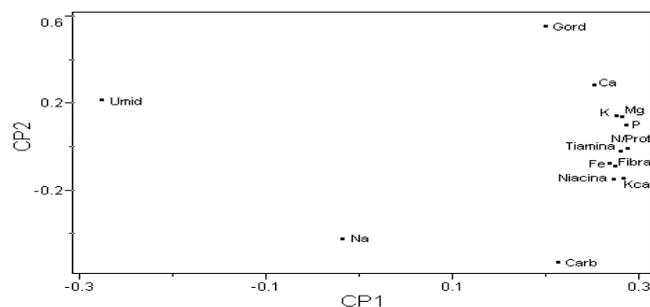
A separação entre grãos e outros vegetais foi descrita na primeira componente principal, como pode ser visualizado na Figura 4 (CP1 versus CP2), a qual descreveu cerca de 80% da variância total dos dados. Analisando conjuntamente os *scores* e os *loadings* para as duas primeiras CP nas Figuras 4 e 5 respectivamente, notou-se que os grãos apresentam *scores* positivos na CP1, enquanto os outros vegetais têm *scores* negativos. Os tubérculos incluindo o aipo têm valores negativos próximos de zero na CP2, enquanto para as verduras os *scores* são ligeiramente positivos. Quando comparados entre si, a separação entre os grãos e os outros vegetais é, em grande parte, devido aos altos teores de umidade, proteínas, carboidratos e calorias. Os grãos têm *scores* positivos, por terem um alto teor de proteínas, carboidratos, calorias e gordura, e um baixo teor de umidade, enquanto os outros vegetais têm alto teor de umidade e baixos teores de proteínas, calorias, etc.

Observou-se no dendograma e nos gráficos de *scores* que a soja está deslocada em relação aos demais grãos, com um índice de similaridade de 0,59, que é o mais baixo de todos, embora esteja incluída no agrupamento dos grãos. Isto pode ser explicado por ter altos teores de gordura e cálcio e baixos teores de carboidratos, quando comparados aos outros grãos (Figura 5).

A segunda componente principal descreveu 9% dos dados e foi responsável pela separação entre verduras e tubérculos. Isto se deve principalmente à concentração de sódio e carboidratos, que estão em maior quantidade nos tubérculos, como pode ser visualizado na Figura 5. O aipo apresentou um *score* não esperado, por ter uma concentração alta de sódio quando comparado às outras verduras.



**FIGURA 4.** Gráfico de *scores* (CP1 versus CP2) para os vegetais estudados.



**FIGURA 5.** Gráfico de *loadings* (CP1 versus CP2) das variáveis utilizadas.

Para se ter uma idéia melhor dos fatores que poderiam influenciar a separação entre tubérculos e verduras fez-se uma nova análise por componentes principais excluindo-se os grãos. Os resultados estão nas Figuras 6 e 7.

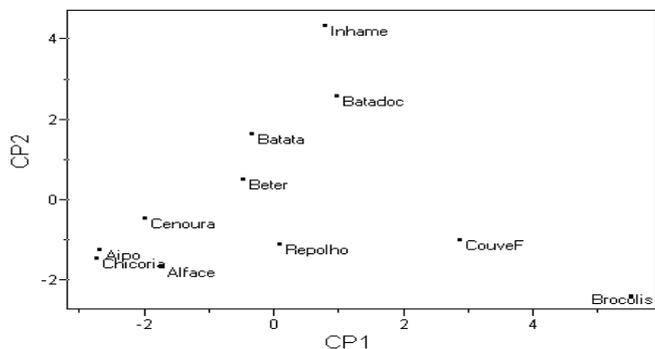


FIGURA 6. Gráfico de *scores* (CP1 versus CP2) para os legumes e verduras.

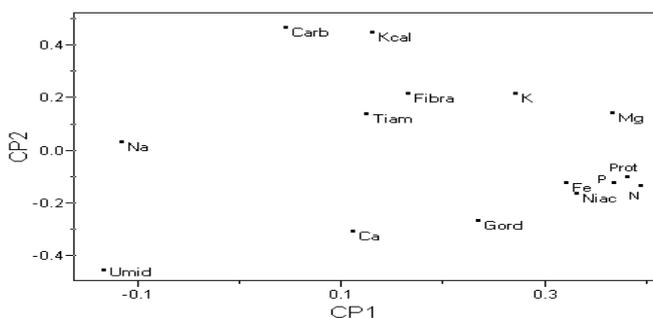


FIGURA 7. Gráfico de *loadings* (CP1 versus CP2) para os legumes e verduras.

Novamente a CP2 (contendo 28,70% da variância dos dados) separou os tubérculos das verduras e o aipo agora está entre as verduras como esperado (Figura 6). O brócolis e a couve-flor apresentaram *scores* altamente positivos na CP1 em virtude dos teores mais elevados de gordura, nitrogênio total, proteínas e fósforo, quando comparados aos outros vegetais. Os tubérculos têm *scores* positivos na CP2, principalmente por serem mais ricos em carboidratos, conseqüentemente, mais calorias e menor teor de água que as verduras. Note que no CP2, o teor de sódio tem uma contribuição quase que desprezível fazendo com que o aipo se apresente junto com as verduras.

Uma última análise foi feita, considerando apenas os grãos e os resultados estão nas Figuras 8 e 9. A primeira componente principal separou a soja dos outros grãos (Figura 8) em conseqüência dos menores teores de carboidratos e umidade e maiores teores de proteínas, gordura, fósforo, nitrogênio total, magnésio e cálcio quando comparados aos outros grãos (Figura 9). A segunda componente principal foi composta principalmente pelos teores de sódio e niacina. O grão-de-bico e feijão preto apresentaram *scores* positivos (Figura 8) por terem maior teor de sódio, enquanto o feijão azuki tem um teor maior de niacina, quando comparado aos outros grãos.

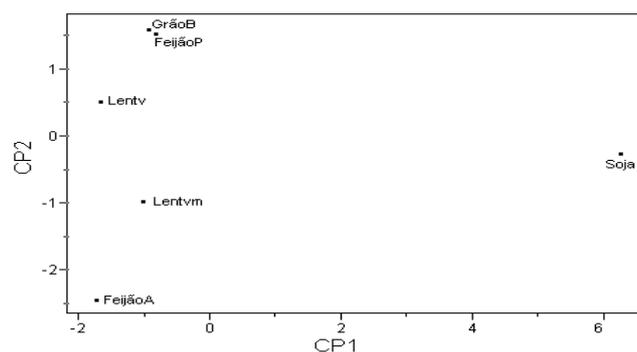


FIGURA 8. Gráfico de *scores* (CP1 versus CP2) para os grãos.

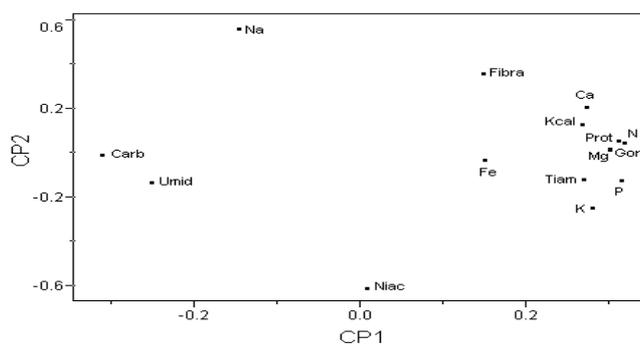


FIGURA 9. Gráfico de *loadings* (CP1 versus CP2) para os grãos.

#### 4. CONCLUSÃO

Os vegetais estudados neste trabalho, incluindo grãos, tubérculos e verduras, foram explorados com respeito à composição dos micronutrientes e teores de umidade, proteína, gordura, caloria, carboidratos, nitrogênio total e fibra, utilizando os métodos estatísticos ACP e AAH. As amostras foram agrupadas de acordo com o esperado, mostrando que a análise exploratória de dados permite que se obtenham informações rápidas e eficientes sobre a similaridade entre amostras através de visualização gráfica. Este tipo de metodologia de análise de dados é particularmente útil quando a quantidade de informações é muito grande, o que é comum ocorrer com dados reais. Com o avanço tecnológico programas computacionais nesta área tornaram-se disponíveis e a utilização deles é relativamente simples. Entretanto, é necessário que os analistas de alimentos tenham domínio sobre os seus princípios para aplicarem os métodos multivariados corretamente, de forma que equívocos sejam evitados.

#### AGRADECIMENTOS

Os autores agradecem o apoio financeiro da FAPESP.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALVAREZ, P.J.M. Application of multivariate statistical methods to the differentiation of gin brands. **J. Sci. Food Agric.**, **57**(2):263-272, 1991.
- BEEBE, K.R., PELL, R.J., SEASHOLTZ, M.B. **Chemometrics: A Practical Guide**, Wiley, New York, 1998.
- BRIANDET, R., KEMSLEY, E.K., WILSON, R.H. Approches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. **J. Sci. Food Agric.**, **71**:359-366, 1996.
- CATET, F., BERTRAND, D., ROBERT, P., MAILLOT, J., DIEUDONNE, J., ROUCH, C. Quantitative determination of sugar cane sucrose by multidimensional statistical analysis of their mid-infrared attenuated total reflectance spectra. **Appl. Spectroscopy**, **45**(2):166-170, 1990.
- DEVER, M.C., CLIFF, M.A. Analysis of variation and multivariate relationships among analytical and sensory characteristics in whole apple evaluation. **J. Sci. Food Agric.**, **69**:329-338, 1995.
- GERBI, V., ZEPPA, G., ANTONELLI, A., CARNACINI, A. Sensory characterization of wine vinegars. **Food Quality and Preference**, **8**(1):27-34, 1997.
- HEADLEY, L.M., HARDY, J.K. Classification of whiskies by principal component analysis. **J. Food Sci.**, **54**(5):1351-1354 e 1358, 1989.
- HERRANZ, A., de la SERNA, P., BARRO, C. Multivariate statistical methods applied to the differentiation of rum brands. **J. Sci. Food Agric.**, **51**:555-560, 1990.
- HOLLAND, B., WELCH, A.A., UNWIN, I.D., BUSS, D.H., PAUL, A.A., SOUTHGATE, D.A.T. **The composition of foods**, 5<sup>th</sup> ed. Royal Society of Chemistry, 1994.
- HOUGH, G., CALIFANO, A.N., BERTOLA, N.C., BEVILACQUA, A.E., MARTINEZ, E., VEJA, M.J.M., ZARITZKY, N.E. Partial Least Squares correlations between sensory and instrumental measurements of flavor and texture for reggianito grating cheese. **Food Quality and Preference**, **7**(1):47-53, 1996.
- KJOLSTAD, L., ISAKSSON, T., ROSENFELD, H.J. Prediction of sensory quality by near infrared reflectance analysis of frozen and freeze dried green peas (*Pisum sativum*). **J. Sci. Food Agric.**, **51**(2):247-260, 1990.
- PIROUETT, Multivariate Data Analysis for IBM PS Systems, version 2.1, 1997. Infometrix, Seattle, WA, USA.
- ROBERT, P., BERTRAND, D., DEVAUX, M.F., GRAPPIN, R. Multivariate analysis applied to near-infrared spectra of milk. **Anal. Chem.**, **59**:2187-2191, 1987.
- SCARMINO, I.S., BRUNS, R.E., ZAGATTO, E.A.G. Pattern recognition classification of mineral waters based on spectrochemical analysis. **Energ. Nucl. Agric.**, **4**(2):99-111, 1982.
- SHARAF, M.A., ILLMAN, D.L., KOWALSKI, B.R. **Chemometrics**, 1<sup>a</sup> ed., John Wiley & Sons, New York, 1986, 332p.
- STENROOS, L.E., SIEBERT, K.J. Application of pattern-recognition techniques to the essential oil of hops. **J. Am. Soc. Brew. Chem.**, **42**(2):54-61, 1984.