



Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

APLICAÇÃO DE ANÁLISE DE CLUSTER EM VALORES NUTRICIONAIS REFERENTES À TIPOS DE CARNES

Samara Rilda de S.BEZERRA¹, Edwirde Luiz SILVA¹, Aldreany P.Araújo SILVA¹

¹ Departamento de Estatística, Universidade Estadual da Paraíba-UEPB, Campus I, Campina Grande-PB. E-mail: samaujp@yahoo.com.br. Telefone: (83)8161-4261.

¹ Departamento de Estatística, Universidade Estadual da Paraíba-UEPB, Campus I, Campina Grande-PB. E-mail: edwirde@uepb.edu.br. Telefone: (83)8756-7007.

¹ Departamento de Química, Universidade Estadual da Paraíba-UEPB, Campus I, Campina Grande-PB. E-mail: Telefone: (83)8161-4261.

RESUMO

Aos dados foi aplicada a Análise de Agrupamento, utilizando o coeficiente de semelhança entre pares de tipos de carnes, com suas respectivas características, considerando a correlação de Pearson distância. Esta análise em sua maior parte é realizada com objetivo de se tratar da heterogeneidade dos dados. Em vez de lidar com um grupo de observações amplamente divergentes, se divide o grupo em subconjuntos mais homogêneos não é a mesma coisa de encontrar agrupamento que ocorram naturalmente. Ao agrupamento, utiliza-se o método das k-médias nos cinco tipos de carnes (Bife, coração, Marisco Enlatado, Siri Enlatada; o camarão enlatado). O agrupamento aglomerativo consistindo de cinco etapas do processo Iterativo. Nos agrupamentos hierárquicos utilizar-se uma medida bastante comum, que a correlação cofenética, que mede o grau de preservação das distâncias emparelhadas pelo dendrograma resultante do agrupamento em relação às distâncias originais. A matriz de distâncias recuperadas é denominada de matriz cofenética. A correlação cofenética foi relativamente alta, logo o agrupamento teve boa qualidade. À medida que se afasta da origem aglomeram-se o primeiro do grupo ao grupo II, o Bife. Os resultados foram avaliados por estatística descritiva utilizando-se medida de tendência central e de variabilidade de dados (desvio-padrão e coeficiente de variação) entre os nutrientes. Foi aplicada a técnica de multivariada: Análise de agrupamentos. Tais análises mostram 2 grupos de carnes.

PALAVRAS CHAVE: carnes, correlação, dendrograma.

1 INTRODUÇÃO

A análise de agrupamento foi processada segundo a metodologia proposta por Sneath & Sokal (1973), e foi aplicada aos dados, utilizando-se como coeficiente de semelhança entre pares de tipos de carnes consideradas a correlação de Pearson distância, que é um tipo de proximidade, mas não necessariamente de similaridade. Basicamente, quanto menor as correlações mais similares eles são segundo as características consideradas. A estratégia de agrupamento adotada foi a do vizinho mais próximo.



Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

Etapa 0. Inicia-se com todos os objetos em agrupamento separados (isto é, 5 agrupamentos com cada tipo de carne). Represente esses agrupamentos por C_1, \dots, C_5 . Nesse passo inicial, a distância entre dois agrupamento é definida como a distância entre dois objetos neles contidos; isto é,

$$d_{C_i C_j} = d_{ij}$$

sendo $t = 1$ um índice do processo iterativo.

Etapa 1. Encontre a menor distância entre dois agrupamentos quaisquer. Represente esses dois agrupamentos mais próximos por C_i e C_j .

Etapa 2. Combina-se a distância C_i e C_j para formar um novo agrupamento denominado C_{n+1} .

Etapa 3. Defina-se a distância entre o novo agrupamento C_{n+t} e todos os agrupamentos C_k da seguinte maneira:

$$d_{C_{(n+t)} C_{(k)}} = \min\{d_{C_{(i)} C_{(k)}}, d_{C_{(j)} C_{(k)}}\}$$

Etapa 4. Adiciona-se o agrupamento C_{n+t} como um novo agrupamento e remova os agrupamentos C_i e C_j . Considere $t = t + 1$.

Etapa 5. Volte à etapa e continue até que reste um agrupamento.

Para ajudar a escolha do melhor método, se define o coeficiente de correlação “cofenética” (r_c) que ainda hoje é a medida de validação mais utilizada pelos taxonomistas numéricos. Esta medida dá-nos a relação entre cada valor da matriz de semelhanças e um valor obtido a partir do dendrograma, significando, em última instância, na medida em que o dendrograma resultante da aplicação de um método hierárquico representa os valores da matriz de semelhanças/distâncias. Mais precisamente, a correlação entre os elementos da matriz de distância (ou semelhanças) e os correspondentes coeficientes de função, ou seja, as distâncias (ou semelhanças) a que indivíduos se juntam pela primeira vez para formar grupos. Embora este método de validação seja apropriado, sobretudo quando se utiliza um método hierárquico, aglomerativo, que referiu a sua sensibilidade ao tamanho dos



Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

grupos como razão suficiente para não ser aceite como justificação direta e final da técnica utilizada.

Outro procedimento muito vulgarizado é o da utilização de vários critérios de agrupamentos e a comparação posterior dos resultados obtidos. Se os resultados forem semelhantes, é possível concluir que, em princípios, qualquer método será de confiança e existem, de fato, grupos entre indivíduos observados.

O grau de convergência dos vários critérios de agregação pode ser aferido através de uma tabela de contingência indicado o número de indivíduos que se agrupam no mesmo cluster, nas soluções resultantes da utilização de diferentes critérios de agregação, para o mesmo número de clusters. Poder-se-á ainda apresentar a no mesmo quadrado o nível de significância de um teste para a independência dos resultados de cada par de critérios de agregação.

Tabela 3– Comparação dos resultados de aplicação de diferentes métodos de agregação para uma participação em 2 grupos de 5 tipos diferentes tipos de carnes.

	Ward	Ligação Simples	Ligação Completa
Distância Euclidiana	0,9356	0,9343	0,9376
Raiz Quadrada da Distância Euclidiana	0,7091	0,8355	0,7067



Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

3 RESULTADOS E DISCUSSÃO

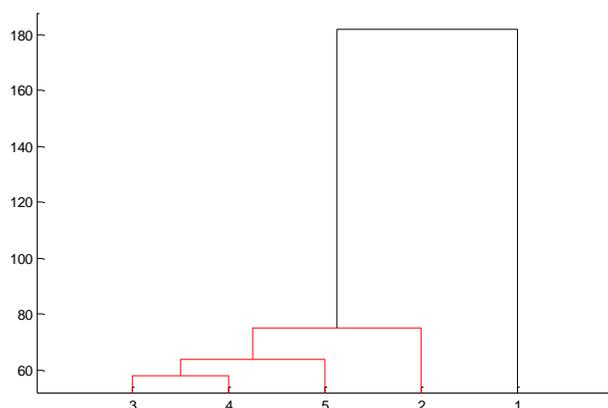
Agrupamento, utilizando o método das k-médias, os n=5 tipos de carnes, mensurados para energia (cal), proteínas (g), gordura (g), cálcio (mg) e ferro (mg) (Bussab et al., 1990). Analisa-se os resultados alcançados, considerando k=2 grupos. Os dados experimentos são mostrados na Tabela 1.

Tabela 1– Valores nutricionais referentes á diversos tipos de carnes

Alimentos	Energia	Proteínas	Gordura	Cálcio	Ferro
Bife	340	20	28	9	2,6
Coração	160	26	5	14	5,9
Marisco	45	7	1	74	5,4
Enlatado					
Siri Enlatado	90	14	2	38	0,8
Camarão	110	23	1	98	2,6

Fonte: Bussab (1990).

Figura 1. Dendrograma dos tipos de carnes segundo as características





Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

A primeira formação de grupo por semelhança ocorre à distância inferior a 60 no gráfico, entre as carnes: Bife, coração, Marisco Enlatado, Siri Enlatada; o segundo, apenas o camarão enlatado.

Para os agrupamentos hierárquicos pode-se utilizar uma medida bastante comum, que é a correlação cofenética. Segundo FERREIDA (2008) o coeficiente de correlação cofenético mede o grau de preservação das distâncias emparelhadas pelo dendrograma resultante do agrupamento em relação às distâncias originais (Sneath & Sokal, 1973). A ideia é obter a correlação entre as distâncias observadas entre os objetos e as distâncias recuperadas da análise de agrupamento. A recuperação das distâncias entre os pares de objetos é feita a partir do dendrograma. Os $n(n-1)/2$ pares de distâncias, originais e recuperadas, são correlacionadas para a obtenção da correlação cofenética. A matriz de distâncias recuperadas é denominada de matriz cofenética. Como a correlação cofenética foi relativamente alta (0.9343), pode-se afirmar que o agrupamento teve boa qualidade.

À medida que se afasta da origem aglomeram-se o primeiro do grupo ao grupo II, o Bife.

4 CONCLUSÃO

O trabalho identificou similaridade dos tipos de carnes, revelando as que apresentam as mesmas características. O resultado obtido revelou que determinados tipos de carnes apresentam características nutricionais semelhantes, tais como proteínas, cálcio e ferro. As formações dos dois grupos de carnes foram: Bife, coração, Marisco Enlatado, Siri Enlatada; e a segunda formação apresentou apenas um elemento no segundo grupo, o camarão enlatado.



Encontro Nacional de Educação, Ciência e Tecnologia/UEPB

REFERÊNCIAS

LATTIN, J. M., CARROLL, J. D; GREEN, P.E. Análise de dados multivariado. São Paulo: Cengage Learning, 2011

FERREIRA, D.F. Estatística multivariada. 1.ed. Lavras: Editora UFLA, 2008. p.389-391.

REIS, Elizabeth. Estatística multivariada Aplicada. Edição Silabo. 2 ed. revista e corrigida. Lisboa-Portugal, 2001

MINGOTI, Suelli. Análise de dados através de métodos estatísticos multivariada: uma abordagem aplicada. Belo Horizonte: Editora UFMG,2005.

HAIR, Joseph F., ANDERSON, Rolp E.,BLACK,Willian. Análise multivariada de dados. 5 ed. Tradução. Adonei Schup Sant'anna e Ancelmo coves neto.Porto Alegre. Brookman, 2005.