

**Análise de agrupamentos em banco de dados de séries históricas de preço de venda de alguns produtos agropecuários praticados no Rio Grande do Sul em 2014**

**Application of cluster analysis technique in database of historical series of sales price of some agricultural products performed in Rio Grande do Sul in 2014**

DOI:10.34117/bjdv5n7-027

Recebimento dos originais: 14/05/2019

Aceitação para publicação: 13/06/2019

**José Cardoso Sobrinho**

Doutorado em Engenharia Agrícola pela Universidade Federal de Viçosa – UFV  
 Instituição: Universidade Federal de Santa Maria – UFSM – Departamento Didático do Colégio Politécnico  
 Endereço: Avenida Roraima, nº1000, Campus UFSM, Prédio 70- CEP – 97105-900 - Bairro Camobi – Santa Maria - RS  
 Email: jcardosos@smail.ufsm.br

**Diogo João Cardoso**

Doutorando em Ciência da Computação pela Universidade Federal de Pelotas – UFPEL.  
 Instituição: Universidade Federal de Pelotas – UFPEL  
 Centro de Desenvolvimento Tecnológico (CDTec)  
 Endereço: Campus Porto – UFPEL  
 Rua Gomes Carneiro, 1  
 CEP 96010-610 – Pelotas – RS – Brasil  
 Email: diogo.j.cardoso@gmail.com

**Henrique Fernando Lidorio**

Mestrando em Engenharia Agrícola pela Universidade Federal de Santa Maria – UFSM  
 Instituição: Universidade Federal de Santa Maria – UFSM  
 Centro de Ciências Rurais – CCR  
 Endereço: Avenida Roraima, nº1000, Campus UFSM, Prédio 70- CEP – 97105-900 - Bairro Camobi – Santa Maria - RS  
 Email: henrique.fernando@outlook.com

**RESUMO**

Objetivou-se avaliar por meio de estatística multivariada, especificamente análise de agrupamentos, o banco de dados de séries históricas dos preços em R\$.kg-1 dos seguintes produtos: arroz irrigado, cordeiro, feijão preto, leite, soja, soja (sem Royalties), boi gordo, sorgo granífero, leitão, sorgo forragem, maçã Fuji grande, maçã Fuji miúda, maçã gala (grande), maçã gala (miúda), trigo em grão ensacado e trigo a granel – nacional, produzidos e comercializados no Rio Grande do Sul de março de 2004 até dezembro de 2014. Utilizou-se o software Statistica para a realização da análise de agrupamentos (cluster). Foram utilizados preços publicados em sites especializados das variáveis no agronegócio brasileiro. Apresentou-se o dendrograma considerando o vizinho mais próximo e foi considerada a distância euclidiana como medida de dissimilaridade. Na análise do dendrograma verificou-

se a existência de dois grandes grupos homogêneos de variáveis distintas. O primeiro grupo é formado pelas variáveis: boi gordo e cordeiro vivo, isto ocorreu porque estas variáveis possuem valores superiores às demais. No segundo grupo pode-se verificar a formação de dois subgrupos, um deles pelas variáveis maçãs (todas), acrescidas das variáveis leitão e feijão preto; outro composto pelas demais variáveis que apresentavam baixa dissimilaridade.

**Palavras-Chave:** Análise multivariada, Produtos agrícolas, Conglomerados

## ABSTRACT

Aimed to evaluate by multivariate statistics, cluster analysis, the database of historical series of the prices is in R\$.kg-1 of the products: irrigated rice, lamb, black beans, milk, soy, soy (without royalties), cattle, grain sorghum, piglet, forage sorghum, large Fuji apple, small Fuji apple, gala apple (large), gala apple (small), bagged grain wheat, national bulk wheat, produced and marketed in RS-BRAZIL from March 2004 to December 2014. It was used the Statistica software to perform the cluster analysis. The prices used were published in specialized sites of the variables in the Brazilian agribusiness. It was presented the dendrogram considering the nearest neighbor and as cluster algorithm of the data and it was considered the Euclidean distance as dissimilarity measure. In the dendrogram analysis it was verified the existence of two large homogeneous groups of distinctive variables. The first group is consisted by variables of: fat lamb and live cattle, this happened because these variables have higher values than the others. In the second group it can be seen the formation of two subgroups, one is composed by variables apples (all of them), plus the variables piglet and black beans, the other is composed by the other variables that presented low dissimilarity.

**Keywords:** Multivariate analysis, Agricultural products, Cluster

## 1 INTRODUÇÃO

Nas últimas décadas, o ambiente empresarial tornou-se mais competitivo e volátil, exigindo das empresas uma necessidade eminente de adaptação às constantes alterações do mercado tendo em vista obter e manter sua vantagem competitiva (CALLADO, 2013). O agronegócio é um parâmetro que contribui para o crescimento das empresas agrícolas, rurais e não rurais, em todas as regiões do país, independentemente do estado da federação onde os dados são coletados. Entende-se por agronegócio, a totalidade das operações de produtos e distribuição de suprimentos agrícolas, das operações de produção nas unidades agrícolas, do armazenamento, do processamento e da distribuição de produtos agrícolas. As áreas contempladas pela Engenharia Agrícola, além de suas especificidades estão inseridas no contexto do agronegócio brasileiro. A área de tecnologia pós-colheita envolve toda a produção dos grãos e ou sementes desde a colheita, beneficiamento, secagem, armazenamento e até a comercialização de grãos, onde certamente há grande interação entre a área e setores e o agronegócio. A técnica denominada análise de agrupamentos, também conhecida como análise de conglomerados, classificação ou *cluster*, tem como objetivo dividir os elementos da

amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características (MINGOTI, 2013). Neste contexto o presente trabalho objetivou avaliar por meio de estatística multivariada (agrupamentos), o banco de dados de séries históricas dos preços de produtos agrícolas no RS.

## 2 MATERIAL E MÉTODOS

O banco de dados de séries históricas dos preços em R\$.kg-1 dos seguintes produtos: arroz irrigado, cordeiro, feijão preto, leite, soja, soja (sem Royalties), boi gordo, sorgo granífero, leitão, sorgo forragem, maçã Fuji grande, maçã Fuji miúda, maçã gala (grande), maçã gala (miúda), trigo em grão ensacado e trigo a granel – nacional, produzidos e comercializados no Rio Grande do Sul de março de 2004 até dezembro de 2014. Os dados foram levantados em sites especializados na área de agronegócio e lançados em planilha no aplicativo Statistica versão 9.0 licenciado para uso na UFSM. Para realizar o dentrograma (árvore) foi considerando o vizinho mais próximo e estimou-se a distância euclidiana como medida de dissimilaridade. No método do vizinho mais próximo a similaridade entre dois conglomerados é definida pelos dois elementos mais parecidos entre si (Sneath, 1957). Em cada estágio do processo de agrupamento, os dois conglomerados que são mais similares com relação à distância foram combinados em um único cluster.

A distância Euclidiana entre dois elementos foi expressa por:

$$d(X_l, X_k) = \left( \sum_{i=1}^p (X_{il} - X_{ik})^2 \right)^{0,5} \quad (1)$$

em que

$d(X_l, X_k)$  - distância entre os dois elementos  $X_l$  e  $X_k$ , com  $l \neq k$ .

$X_{il}$  - representa o valor observado da variável  $i$  medida no elemento  $l$ .

$X_{ik}$  - representa o valor observado da variável  $i$  medida no elemento  $k$ .

## 3 RESULTADOS E DISCUSSÃO

Justifica-se a inserção deste trabalho nos grupos e subgrupos, mesmo com diferenças biológicas entre si, porque a análise de variância multivariada avalia a diferença de variância

entre os grupos. Quando à região envolvida, é indiferente, uma vez que foi realizado no estado do Rio Grande do Sul, assim como poderia ser realizado em qualquer estado da federação. Em relação ao público alvo, acredita-se que a ciência não contenha grupo específico para atendimento, ou seja, deve atender a todos e quaisquer pesquisadores, estudantes de ensino fundamental, médio, graduação e pós graduação. Interpreta-se um dendrograma pela similaridade admitindo que duas amostras próximas devam também ter valores muito próximos para as variáveis medidas, isto significa que elas devem estar muito próximas matematicamente no espaço multidimensional, então, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas. No dendrograma da Figura 1, hierarquizou-se esta similaridade de modo que se pode ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto das 16 variáveis utilizadas no presente trabalho. Na Figura 1 apresenta-se o dendrograma das variáveis avaliadas. Verificou-se a existência de dois grandes grupos homogêneos de variáveis distintas. O primeiro grupo foi formado pelas variáveis: boi gordo e cordeiro vivo, isto ocorreu porque estas variáveis possuem valores em R\$.kg-1 superiores às demais variáveis. No segundo grupo, pode-se verificar a formação de dois subgrupos, um deles pelas variáveis maçãs (todas), acrescidas das variáveis leitão e feijão preto; outro composto pelas demais variáveis que apresentavam baixa dissimilaridade.

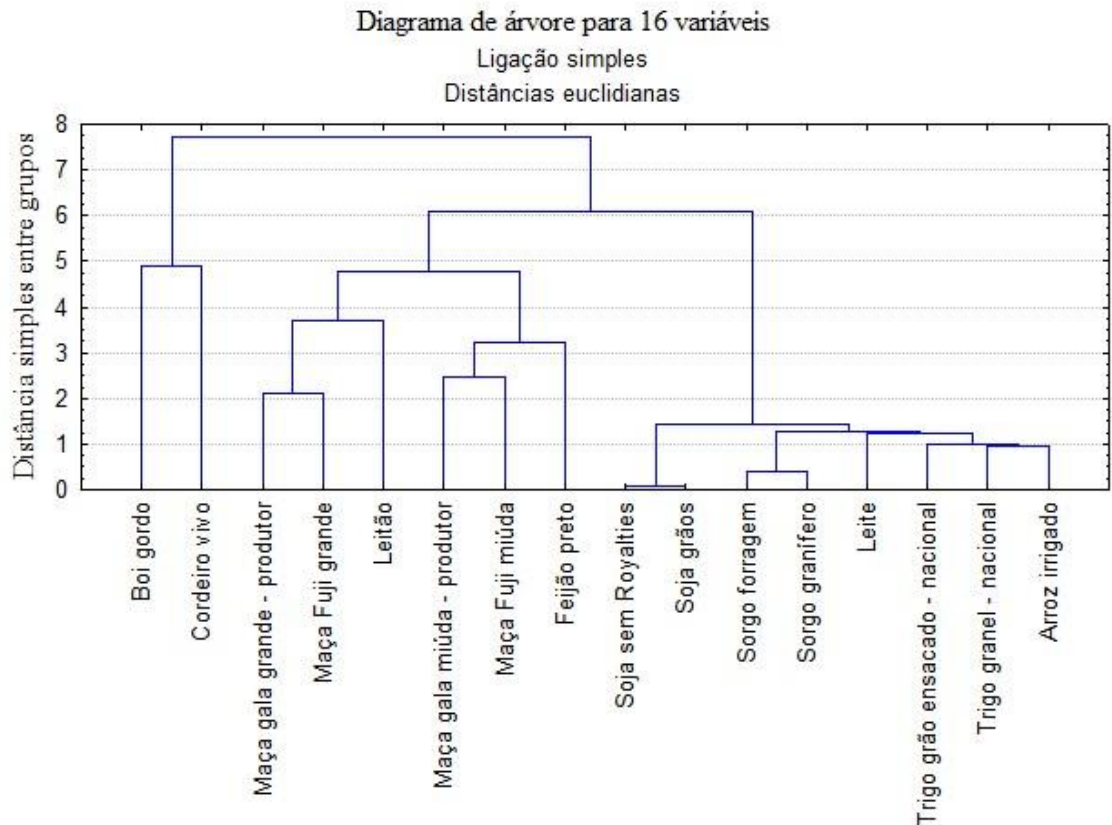


Figura 1 – Dendrograma das 16 variáveis feito por ligação simples e distâncias euclidianas.

#### 4 CONCLUSÕES

Na análise do dendrograma, é possível verificar dois grandes grupos de variáveis, um composto por boi gordo e cordeiro vivo e outro pelas demais 14 variáveis. Neste há dois sub-grupos, em um deles a distância simples entre os grupos não chegou a 2 e outro sub-grupo cuja distância simples entre grupos variou entre 2 e tendendo a cinco pela esquerda.

#### REFERÊNCIAS

CALLADO, A.L.C; SOARES, K.R., Análise da utilização de indicadores de desempenho no contexto das agroindústrias. **XX Congresso Brasileiro de Custos** – Uberlândia, MG, Brasil, 18 a 20 de novembro de 2013.

MINGOTI, S.A.; Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. 3ª Edição. Belo Horizonte, Editora UFMG. 2013, 297p..

SNEATH, P.H.; The application of computer to taxonomy. **Journal of General Microbiology**, 17. P. 201-226, 1957.