Ciência Rural, Santa Maria, v.41, n.9, p.1503-1508, set, 2011
ISSN 0103-8478

Consistência do padrão de agrupamento de cultivares de milho

Clustering pattern consistency of corn cultivars

Alberto Cargnelutti Filho[I] José Paulo Guadagnin[II]

RESUMO

*O objetivo deste trabalho foi avaliar a consistência do padrão de agrupamento obtido a partir da combinação de duas medidas de dissimilaridade e quatro métodos de agrupamento, em cenários formados por combinações de número de cultivares e número de variáveis, com dados reais de cultivares de milho (**Zea mays** L.) e com dados simulados. Foram usados os dados reais de cinco variáveis mensuradas em 69 experimentos de competição de cultivares de milho, cujo número de cultivares avaliadas oscilou entre 9 e 40. A fim de investigar os resultados com maior número de cultivares e de variáveis, foram simulados, sob distribuição normal padrão, 1.000 experimentos para cada um dos 54 cenários formados pela combinação entre o número de cultivares (20, 30, 40, 50, 60, 70, 80, 90 e 100) e o número de variáveis (5, 6, 7, 8, 9 e 10). Foram realizadas análises de correlação, de diagnóstico de multicolinearidade e de agrupamento. A consistência do padrão de agrupamento foi avaliada por meio do coeficiente de correlação cofenética. Há decréscimo da consistência do padrão de agrupamento com o acréscimo do número de cultivares e de variáveis. A distância euclidiana proporciona maior consistência no padrão de agrupamento em relação à distância de Manhattan. A consistência do padrão de agrupamento entre os métodos aumenta na seguinte ordem: Ward, ligação completa, ligação simples e ligação média entre grupo.*

Palavras-chave: ***Zea mays** L., medidas de dissimilaridade, métodos de agrupamento, coeficiente de correlação cofenética.*

ABSTRACT

*The objective of this research was to evaluate the clustering pattern consistency obtained from the combination of the two dissimilarity measures and four clustering methods, in scenarios consist of combinations number of cultivars and number of variables, with real data in corn cultivars (**Zea mays** L.) and simulated data. We used real data from five variables measured in 69 trials involving corn cultivars, the number of cultivars ranged between 9 and 40. In order to investigate the results with more cultivars and variables, were simulated under the standard normal distribution, 1,000 experiments for each of the 54 scenarios formed by the combination among the number of cultivars (20, 30, 40, 50, 60, 70, 80, 90 and 100) and the number of variables (5, 6, 7, 8, 9 and 10). Analyses of correlation, diagnoses of multicollinearity ans cluster were carried out. Clustering pattern consistency was evaluated by the cophenetic correlation coefficient. There is a decrease of clustering pattern consistency with the increase in the number of cultivars and variable. The euclidean distance provides greater clustering pattern consistency in relation to Manhattan distance. The clustering pattern consistency among the methods increases as follows: Ward's, complete linkage, single linkage and average linkage between groups.*

Key words: ***Zea mays** L., dissimilarity measures, clustering methods, cophenetic correlation coefficient.*

INTRODUÇÃO

Em programas de melhoramento de plantas, a identificação de indivíduos (cultivares, linhagens, clones, variedades e híbridos) divergentes, por meio de análise de agrupamento, tem sido utilizada. A análise de agrupamento apresenta a finalidade de reunir, por algum critério de classificação, os indivíduos em

[I]Departamento de Fitotecnia, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), 97105-900, Santa Maria, RS, Brasil. E-mail: cargnelutti@pq.cnpq.br. Autor para correspondência.
[II]Fundação Estadual de Pesquisa Agropecuária (FEPAGRO), Porto Alegre, RS, Brasil.

Recebido para publicação 01.08.10 Aprovado em 29.06.11 Devolvido pelo autor 02.08.11
CR-3908

grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre os grupos. Padrões de agrupamento distintos são obtidos, a partir da combinação entre as diversas medidas de dissimilaridade entre os pares de indivíduos e os diversos métodos de agrupamento (CRUZ & REGAZZI, 1997; CRUZ & CARNEIRO, 2003; MINGOTI, 2005; MANLY, 2008).

O coeficiente de correlação linear de Pearson entre os elementos da matriz de dissimilaridade (matriz de distâncias entre os indivíduos, obtida a partir dos dados originais) e os elementos da matriz cofenética (matriz de distâncias entre os indivíduos, obtida a partir do dendrograma) é denominado coeficiente de correlação cofenética. Esse coeficiente pode ser utilizado para avaliar a consistência do padrão de agrupamento de métodos de agrupamentos hierárquicos, sendo que valores próximos à unidade indicam melhor representação (BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003).

A consistência do padrão de agrupamento de 13 cultivares de feijão (***Phaseolus vulgaris*** L.), obtida a partir da combinação de oito medidas de dissimilaridade (euclidiana, euclidiana padronizada, euclidiana média, euclidiana média padronizada, quadrado da distância euclidiana, quadrado da distância euclidiana padronizada, Mahalanobis e Mahalanobis padronizada) e oito métodos de agrupamento (ligação simples, ligação completa, Ward, mediana, ligação média dentro de grupo, ligação média entre grupo, Gower e Centróide), foi avaliada por meio do coeficiente de correlação cofenética (CARGNELUTTI FILHO et al., 2010). Nesse estudo, as cultivares foram agrupadas de acordo com seis variáveis e os autores concluíram que maior consistência nos padrões de agrupamento é verificada com o método da ligação média entre grupo, obtido a partir da matriz de distância euclidiana.

Não foram encontradas na literatura comparações entre os coeficientes de correlação cofenética obtidos a partir das distâncias euclidiana e Manhattan, combinadas com os métodos hierárquicos da ligação simples (vizinho mais próximo), da ligação completa (vizinho mais distante), da ligação média entre grupo e de Ward, muito utilizados em publicações e disponíveis em diversos softwares estatísticos, com variação do número de cultivares de milho (dados reais) e ainda com variação do número de cultivares e do número de variáveis (dados simulados). Assim, é importante fazer essas comparações com base em dados reais (cenários restritos) e ainda ampliar as inferências por meio de simulação de cenários extremos.

O objetivo deste trabalho foi avaliar a consistência do padrão de agrupamento obtido a partir da combinação de duas medidas de dissimilaridade e quatro métodos de agrupamento, em cenários formados por combinações de número de cultivares e número de variáveis, com dados reais de cultivares de milho (***Zea mays*** L.) e com dados simulados.

MATERIAL E MÉTODOS

Foram usados os dados do número de dias da semeadura até 50% do florescimento masculino (DF), das estaturas de plantas (EP) e de espigas (EE) na colheita, em cm, da população final (POP), em plantas ha^{-1}, e da produtividade de grãos (PROD), em kg ha^{-1}, de 69 experimentos de competição de cultivares de milho. Os experimentos foram realizados no Estado do Rio Grande do Sul e classificados em 12 grupos de experimentos conforme a categoria (estadual e indicado), o ciclo (precoce e superprecoce) e o ano agrícola (2002/2003, 2003/2004 e 2004/2005) (Tabela 1). Em todos os experimentos, as unidades experimentais que continham as cultivares foram casualizadas conforme o delineamento em blocos ao acaso com três repetições, sendo as unidades experimentais constituídas de duas fileiras com 5m de comprimento e espaçamento entre 0,7m e 0,9m entre fileiras.

Inicialmente, em cada experimento, foi calculada a média das variáveis DF, EP, EE, POP e PROD das três repetições de cada cultivar. A normalidade dessas 345 séries de dados médios (69 experimentos x 5 variáveis) foi verificada por meio do teste de Shapiro-Wilk e foi realizada a padronização, a fim de obter uma nova variável com média zero e desvio padrão um. A seguir, em cada experimento, foi determinada a matriz de coeficientes de correlação linear de Pearson entre as variáveis padronizadas (matriz fenotípica) e realizado o diagnóstico de multicolinearidade (CRUZ, 2006), conforme critério de MONTGOMERY & PECK (1982). De acordo com MONTGOMERY & PECK (1982), a matriz pode apresentar multicolinearidade fraca (NC<100), moderada a forte (100<NC<1.000) ou severa (NC>1.000).

Em seguida, em cada experimento, determinaram-se as matrizes de distância euclidiana (E) e de Manhattan ou quarteirão (*city block*) (M) entre as cultivares (BARROSO & ARTES, 2003; CRUZ, 2006; FERREIRA, 2008). Essas matrizes de distância foram utilizadas como medida de dissimilaridade para a análise de agrupamento das cultivares por meio dos seguintes métodos hierárquicos: ligação simples (vizinho mais próximo), ligação completa (vizinho mais distante), ligação média entre grupo e Ward (CRUZ & REGAZZI, 1997; BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003; CRUZ, 2006; FERREIRA, 2008). Ao final, foram

Tabela 1 - Número de experimentos e número de cultivares de milho, avaliadas em três anos agrícolas no Estado do Rio Grande do Sul, em diferentes grupos.

Grupo de experimento				Experimentos (locais)	Cultivares (em cada local)
Número	Categoria[1]	Ciclo	Ano agrícola		
1	Estadual	Precoce	2002/2003	7	36
2	Estadual	Precoce	2003/2004	8	40
3	Estadual	Precoce	2004/2005	5	32
4	Estadual	Superprecoce	2002/2003	5	11
5	Estadual	Superprecoce	2003/2004	6	9
6	Estadual	Superprecoce	2004/2005	6	17
7	Indicado	Precoce	2002/2003	5	27
8	Indicado	Precoce	2003/2004	6	27
9	Indicado	Precoce	2004/2005	5	30
10	Indicado	Superprecoce	2002/2003	5	18
11	Indicado	Superprecoce	2003/2004	6	16
12	Indicado	Superprecoce	2004/2005	5	12
Total				69	

[1] Na categoria dos experimentos estaduais, as cultivares avaliadas foram aquelas ainda não indicadas aos produtores e as indicadas foram estudadas na categoria dos experimentos indicados.

obtidos 552 dendrogramas resultantes da combinação de 69 experimentos, duas distâncias e quatro métodos de agrupamento. Para avaliar a consistência dos 552 agrupamentos, ou seja, verificar a capacidade do dendrograma em reproduzir as matrizes de dissimilaridade (E e M), calculou-se o coeficiente de correlação cofenética - CCC (BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003). Foi considerado mais consistente o agrupamento que apresentou maior escore de CCC.

Após, foram simulados 54.000 experimentos (54 cenários x 1.000 experimentos por cenário). Os 54 cenários foram formados pela combinação entre o número de cultivares (20, 30, 40, 50, 60, 70, 80, 90 e 100) e o número de variáveis (5, 6, 7, 8, 9 e 10). Em cada experimento de cada cenário, os dados de cada variável foram simulados com distribuição normal com média zero e desvio padrão um. A seguir, em cada um dos 54.000 experimentos, foram realizados os mesmos procedimentos de análise descritos anteriormente. As análises estatísticas foram realizadas no programa R (R DEVELOPMENT CORE TEAM, 2011) e na planilha eletrônica Office Excel.

RESULTADOS E DISCUSSÃO

A média do valor P do teste de Shapiro-Wilk das 345 séries de dados reais (69 experimentos x 5 variáveis) analisadas foi de 0,34. Em 254 casos (73,62%), os dados se ajustaram à distribuição normal (P>0,05). Assim, as simulações de dados realizadas neste estudo, sob distribuição normal, devem refletir bem o comportamento dessas variáveis.

As médias e desvios padrões das variáveis DF, EP, EE, POP e PROD, das cultivares avaliadas nos 69 experimentos de competição de cultivares de milho foram, respectivamente, 74±8 dias, 204±31cm, 113±24cm, 53.717±7.619 plantas ha^{-1}, 6.785±2.712kg ha^{-1}. Essas variáveis apresentam diferentes escalas de medidas e, como consequência, têm importância diferenciada na definição dos grupos. Assim, a padronização das variáveis é um procedimento adequado para minimizar o efeito das diferentes escalas de medidas, fazendo com que todas as variáveis exerçam importância equivalente na definição dos grupos (CRUZ & REGAZZI, 1997; BARROSO & ARTES, 2003; HAIR et al., 2005; CORRAR et al., 2007).

De acordo com o critério apresentado por MONTGOMERY & PECK (1982), a matriz de coeficientes de correlação linear de Pearson apresentou uma multicolinearidade fraca, pois o número de condição (NC) oscilou entre 4 e 78 (Tabela 2) e a média foi igual a 19. Em presença de multicolinearidade, o uso de todas as variáveis na análise de agrupamento não é um procedimento adequado, pois os caracteres multicolineares são implicitamente ponderados com maior peso (BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003; HAIR et al., 2005; CORRAR et al., 2007). Dessa forma, não sendo detectada a multicolinearidade, foram utilizadas as cinco variáveis na análise de agrupamento, sendo, nessas condições, considerada uma análise adequada (CARGNELUTTI FILHO et al., 2009).

Tabela 2 - Número de cultivares (C), número de experimentos (NE), valor mínimo (mn), valor máximo (mx) e média do número de condição (NC) e dos coeficientes de correlação cofenética entre as matrizes de distância euclidiana e Manhattan e a matriz cofenética dos métodos de agrupamento hierárquicos[1], calculados com base em cinco variáveis avaliadas em 69 experimentos de competição de cultivares de milho.

C	NE	----------NC----------			----------LS----------			----------LC----------			----------LM----------			----------WA----------		
		mn	mx	m	mn	mx	m	mn	mx	m	mn	mx	m	mn	mx	m
					---------- Distância euclidiana ----------											
9	6	13	78	37	0,546	0,897	0,708	0,637	0,881	0,768	0,669	0,913	0,800	0,617	0,899	0,749
11	5	9	63	34	0,496	0,882	0,746	0,588	0,903	0,738	0,602	0,912	0,799	0,545	0,755	0,660
12	5	9	77	31	0,500	0,807	0,684	0,626	0,765	0,698	0,654	0,830	0,756	0,604	0,741	0,691
16	6	8	14	11	0,680	0,803	0,741	0,510	0,747	0,644	0,734	0,828	0,786	0,507	0,685	0,599
17	6	6	32	17	0,587	0,848	0,756	0,592	0,843	0,746	0,676	0,867	0,792	0,546	0,782	0,649
18	5	8	18	11	0,635	0,804	0,729	0,522	0,746	0,619	0,687	0,838	0,783	0,500	0,757	0,636
27	11	4	31	18	0,487	0,815	0,669	0,433	0,767	0,622	0,672	0,848	0,753	0,422	0,611	0,511
30	5	6	25	13	0,497	0,741	0,654	0,477	0,682	0,564	0,639	0,816	0,740	0,471	0,547	0,522
32	5	6	17	10	0,520	0,732	0,667	0,558	0,774	0,671	0,690	0,801	0,757	0,440	0,735	0,547
36	7	9	52	21	0,512	0,857	0,741	0,488	0,804	0,666	0,640	0,867	0,780	0,500	0,790	0,612
40	8	6	30	12	0,584	0,714	0,663	0,430	0,683	0,574	0,650	0,784	0,722	0,439	0,567	0,495
					---------- Distância de Manhattan ----------											
9	6	13	78	37	0,464	0,921	0,702	0,648	0,911	0,748	0,668	0,923	0,785	0,558	0,879	0,704
11	5	9	63	34	0,376	0,786	0,693	0,559	0,886	0,686	0,569	0,893	0,768	0,541	0,871	0,685
12	5	9	77	31	0,575	0,805	0,701	0,621	0,757	0,693	0,690	0,813	0,752	0,641	0,775	0,700
16	6	8	14	11	0,663	0,795	0,712	0,503	0,691	0,616	0,733	0,798	0,766	0,544	0,688	0,624
17	6	6	32	17	0,572	0,873	0,736	0,541	0,813	0,687	0,633	0,855	0,774	0,541	0,802	0,623
18	5	8	18	11	0,597	0,780	0,714	0,482	0,799	0,676	0,649	0,817	0,764	0,522	0,783	0,649
27	11	4	31	18	0,514	0,811	0,639	0,511	0,717	0,611	0,658	0,828	0,724	0,437	0,626	0,531
30	5	6	25	13	0,477	0,760	0,635	0,507	0,734	0,569	0,663	0,786	0,718	0,464	0,557	0,505
32	5	6	17	10	0,537	0,759	0,657	0,495	0,755	0,572	0,667	0,776	0,729	0,466	0,727	0,547
36	7	9	52	21	0,517	0,854	0,750	0,603	0,810	0,696	0,653	0,865	0,799	0,436	0,777	0,547
40	8	6	30	12	0,498	0,706	0,629	0,487	0,666	0,589	0,603	0,750	0,699	0,408	0,578	0,504

[1] LS: ligação simples (vizinho mais próximo), LC: ligação completa (vizinho mais distante), LM: ligação média entre grupo, WA: Ward.

O coeficiente de correlação cofenética (CCC) oscilou entre 0,376 e 0,923 (Tabela 2). A média dos 552 valores de CCC formados pela combinação de 69 experimentos, duas distâncias e quatro métodos de agrupamento foi de 0,675, o que revela variabilidade na consistência do padrão de agrupamento entre os experimentos, matrizes de distância e métodos de agrupamento. Avaliando a consistência do padrão de agrupamento em cultivares de feijão, CARGNELUTTI FILHO et al. (2010) encontraram resultados semelhantes, ou seja, valores de CCC entre 0,2437 e 0,9221 e a média dos 576 valores de CCC formados pela combinação de nove experimentos, oito distâncias e oito métodos de agrupamento foi de 0,6733.

Entre os 54.000 experimentos simulados, os valores mínimo, máximo e médio do número de condição (NC) foram, respectivamente, 1, 87 e 4, e do coeficiente de correlação cofenética (CCC), 0,216, 0,903 e 0,510. Com exceção da média do CCC, esses resultados revelam semelhança no comportamento entre os dados reais e os simulados, o que sugere que as inferências com base nesses dois conjuntos de dados (reais e simulados) sejam similares. O menor escore médio de CCC nos dados simulados (0,510) em relação aos dados reais (0,675) pode estar associado aos diferentes cenários (combinação de número de cultivares e de variáveis) dos dados reais e simulados.

De maneira geral, os coeficientes de correlação cofenética (CCC) obtidos a partir da combinação das medidas de dissimilaridade (euclidiana e Manhattan) e dos métodos de agrupamento (ligação simples, ligação completa, ligação média entre grupo e Ward) diminuem com o acréscimo do número de cultivares (Tabela 2 e Figura 1). Esse comportamento é mais evidente com os dados simulados (Figura 1) em relação aos dados reais (Tabela 2). Ainda, em todas as combinações entre as distâncias, os métodos de agrupamento e o número de cultivares, a consistência do agrupamento diminuiu com o acréscimo do número de variáveis (Figura 1). Portanto, esses resultados

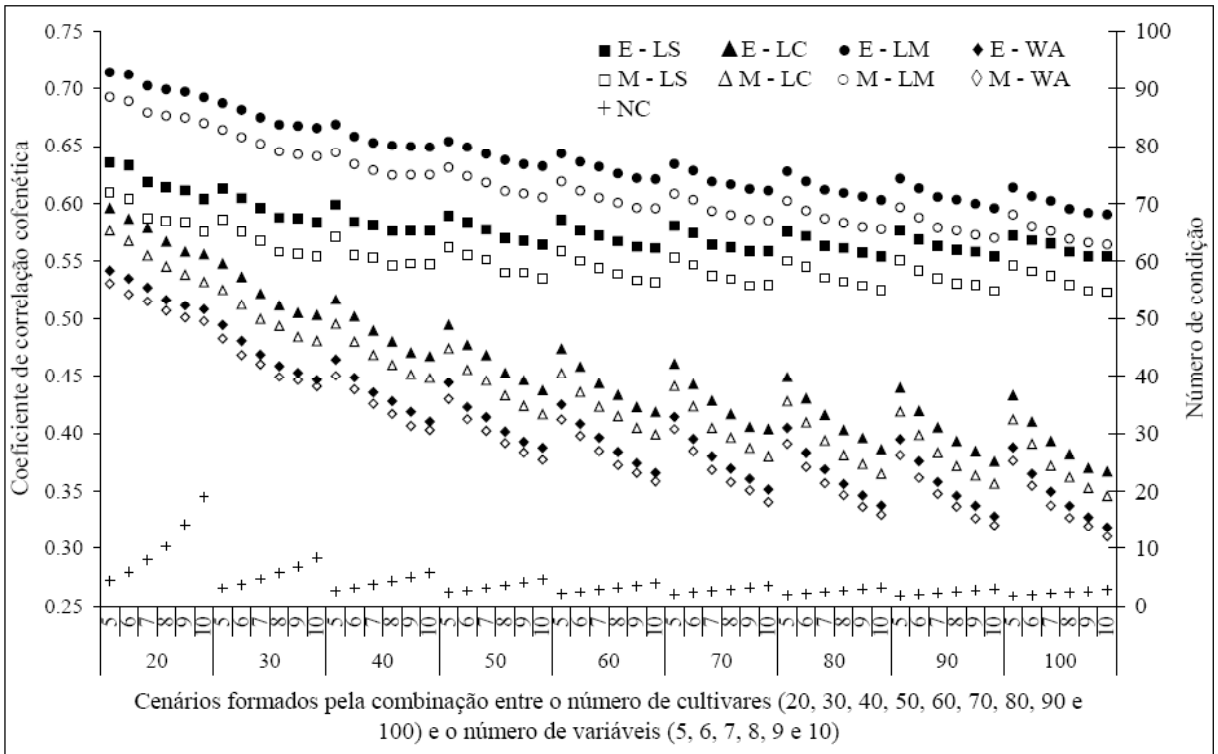

Figura 1 - Média do número de condição e dos coeficientes de correlação cofenética entre as matrizes de distância euclidiana (E) e Manhattan (M) e a matriz cofenética dos métodos de agrupamento hierárquicos da ligação simples (vizinho mais próximo, LS), da ligação completa (vizinho mais distante, LC), da ligação média entre grupo (LM) e de Ward (WA) de 1.000 experimentos simulados em cada um dos 54 cenários.

evidenciam que agrupamentos menos consistentes são obtidos nos cenários formados por maior número de cultivares e de variáveis, justificando o menor valor médio do CCC nos dados simulados.

Nas combinações formadas entre o número de cultivares e os métodos de agrupamento, de modo geral, a distância euclidiana proporcionou agrupamentos mais consistentes em relação à distância de Manhattan (Tabela 2). Esse comportamento manteve-se com a variação do número de variáveis, sendo que a relação entre as duas distâncias ficou mais evidenciada nos resultados dos experimentos simulados (Figura 1). Portanto, para este tipo de dado e para a obtenção de agrupamentos mais consistentes, a distância euclidiana deve ser a preferida. Embora a distância de Manhattan não tenha sido investigada por CARGNELUTTI FILHO et al. (2010), os autores também encontraram superioridade na consistência do padrão de agrupamento calculada a partir da distância euclidiana.

Maior consistência no padrão de agrupamento foi obtida com o método da ligação média entre grupo, com diminuição gradativa de consistência na seguinte ordem: ligação simples, ligação completa e Ward, nos cenários formados pela combinação de distintos números de cultivares, números de variáveis e medidas de dissimilaridade (euclidiana e Manhattan) (Tabela 2 e Figura 1), o que revela que o método da ligação média entre grupo deve ser o preferido. Esses resultados concordam com CARGNELUTTI FILHO et al. (2010) e SOKAL & ROHLF (1962).

Do ponto de vista prático, os resultados deste trabalho e os de CARGNELUTTI FILHO et al. (2010) revelaram agrupamentos mais consistentes formados a partir da matriz de distância euclidiana, utilizando o método da ligação média entre grupo, também definido como *unweighted pair-group average* (UPGMA).

CONCLUSÃO

Independentemente das medidas de dissimilaridade (euclidiana e Manhattan) e dos métodos de agrupamento (ligação simples, ligação completa, ligação média entre grupo e Ward), há decréscimo da consistência do padrão de agrupamento de cultivares de milho com o acréscimo do número de cultivares e de variáveis.

Independentemente do número de cultivares, do número de variáveis e dos métodos de

agrupamento, a distância euclidiana proporciona maior consistência no padrão de agrupamento de cultivares de milho em relação à distância de Manhattan.

Independentemente do número de cultivares, do número de variáveis e das medidas de dissimilaridade, a consistência do padrão de agrupamento de cultivares de milho dos métodos aumenta na seguinte ordem: Ward, ligação completa, ligação simples e ligação média entre grupo.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de Produtividade em Pesquisa para Alberto Cargnelutti Filho; à Fundação Estadual de Pesquisa Agropecuária; e aos pesquisadores, pela realização dos ensaios de competição de cultivares de milho no Estado do Rio Grande do Sul.

REFERÊNCIAS

BARROSO, L.P.; ARTES, R. **Análise multivariada**. Lavras: UFLA, 2003. 151p.

CARGNELUTTI FILHO, A. et al. Agrupamento de cultivares de feijão em presença e em ausência de multicolinearidade. **Ciência Rural**, v.39, p.2409-2418, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782009000900005&lng=pt&nrm=iso>. Acesso em: 31 jul. 2010. doi: 10.1590/S0103-84782009000900005.

CARGNELUTTI FILHO, A. et al. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de dissimilaridade e métodos de agrupamento. **Pesquisa Agropecuária Brasileira**, v.45, p.236-243, 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2010000300002&lng=pt&nrm=iso>. Acesso em: 31 jul. 2010. doi: 10.1590/S0100-204X2010000300002.

CORRAR, L.J. et al. **Análise multivariada para os cursos de administração, ciências contábeis e economia.** São Paulo: Atlas, 2007. 542p.

CRUZ, C.D. **Programa genes**: análise multivariada e simulação. Viçosa: UFV, 2006. 175p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2003. 585p.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: UFV, 1997. 390p.

FERREIRA, D.F. **Estatística multivariada**. Lavras: UFLA, 2008. 662p.

HAIR, J.F. et al. **Análise multivariada de dados**. 5.ed. Porto Alegre: Bookman, 2005. 593p.

MANLY, B.J.F. **Métodos estatísticos multivariados: uma introdução**. 3.ed. Porto Alegre: Bookman, 2008. 229p.

MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: UFMG, 2005. 297p.

MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York: John Wiley & Sons, 1982. 504p.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna, 2011. Disponível em: <http://www.R-project.org>. Acesso em: 31 jul. 2010.

SOKAL, R.R.; ROHLF, F.J. The comparison of dendrograms by objective methods. **Taxon**, v.11, p.33-40, 1962.