

Agrupamento de cultivares de feijão em presença e em ausência de multicolinearidade

Cluster in common bean cultivars in the presence and absence of multicollinearity

Alberto Cargnelutti Filho^{1*} Lindolfo Storck¹ Nerinéia Dalfollo Ribeiro¹

RESUMO

O objetivo deste trabalho foi estudar o efeito da multicolinearidade em análise de agrupamento de cultivares de feijão (*Phaseolus vulgaris* L.), com base em características de produção, fenologia e morfologia, comparando os padrões obtidos em presença e em ausência de multicolinearidade. Para tanto, foram planejados dois conjuntos de dados, cada um com 10 indivíduos e nove variáveis, para demonstrar o efeito das diferentes escalas de medidas e da multicolinearidade no agrupamento. Também foram usados os dados de 13 cultivares de feijão avaliadas em nove experimentos conduzidos em Santa Maria, Estado do Rio Grande do Sul (latitude 29°42'S, longitude 53°49'W e 95m de altitude), entre os anos agrícolas de 2000/2001 e 2004/2005. Nos experimentos, foram avaliados os caracteres produtividade de grãos, número de vagens por planta, número de sementes por vagem, massa de cem grãos, população final de plantas, número de dias da emergência ao florescimento, número de dias da emergência à colheita, altura de inserção da primeira vagem, altura de inserção da última vagem e grau de acamamento. Os padrões de agrupamento de cultivares de feijão, com base em características de produção, fenologia e morfologia, diferem em presença e em ausência de multicolinearidade. O efeito da multicolinearidade deve ser contornado para proporcionar padrões de agrupamento adequados.

Palavras-chave: *Phaseolus vulgaris* L., distância euclidiana, ligação completa.

ABSTRACT

The aim of this research was to study the effect of multicollinearity in cluster analysis of common bean cultivars (*Phaseolus vulgaris* L.) based on productive, phenologic and morphologic traits, comparing the standard obtained in the presence and absence of multicollinearity. Thus, it was planned two data sets, each with ten individuals and nine variables, to demonstrate the effect of different scales and measures of

multicollinearity in the cluster. Thirteen common bean cultivars were evaluated in nine experiments conducted in Santa Maria, in Rio Grande do Sul State, Brazil (latitude 29°42'S, longitude 53°49'W and altitude 95m) in the agricultural years from 2000/2001 to 2004/2005. The following characters were evaluated in the experiments: grain yield, number of pods per plant, number of seeds per pod, weight of 100 grains, plants final population, number of days of the emergency to flowering, number of days of the emergency to harvest, height of first pod insertion, height of the final pod insertion and degree of the down. Standart cluster of common bean cultivars based on productive, phenologic and morphologic traits, differ in the presence and absence of multicollinearity. The effect of multicollinearity must be controlled to provide standards to appropriate cluster.

Key words: *Phaseolus vulgaris* L., euclidian distance, complete linkage.

INTRODUÇÃO

Em programas de melhoramento de feijão, a análise de agrupamento, com base em diferenças morfológicas, fisiológicas e moleculares, tem sido utilizada para identificar os genótipos divergentes e com maior probabilidade de sucesso nos cruzamentos. A análise de agrupamento apresenta a finalidade de reunir, por algum critério de classificação, os genitores em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre os grupos. Há diversas medidas de dissimilaridade, entre os pares de genótipos, que são utilizadas nos diversos métodos de agrupamento. As diferentes combinações entre essas medidas e os métodos podem levar a distintos

¹Departamento 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.

padrões de agrupamento (CRUZ & REGAZZI, 1997; MINGOTI, 2005; MANLY, 2008).

Em estudos de divergência genética em feijão, comparações entre algumas medidas de dissimilaridade (MALUF & FERREIRA, 1983) e entre alguns métodos de agrupamento (BONETT et al., 2006; CARGNELUTTI FILHO et al., 2008) têm sido realizadas, revelando distintos níveis de concordância. Em análise de agrupamento, a padronização das variáveis tem sido recomendada para minimizar o efeito das diferentes escalas de medidas, fazendo com que todas as variáveis exerçam a mesma importância na definição dos grupos (CRUZ & REGAZZI, 1997; BARROSO & ARTES, 2003; HAIR et al., 2005; CORRAR et al., 2007).

A multicolinearidade é definida como o grau em que qualquer efeito de variável pode ser previsto ou explicado por outras variáveis e é um pressuposto básico em análise de agrupamento. A multicolinearidade entre as variáveis pode alterar os padrões de agrupamento, pelo fato dessas variáveis multicolineares serem implicitamente ponderadas com maior peso (HAIR et al., 2005; CORRAR et al., 2007). A multicolinearidade atua como um processo de ponderação não visível para o observador, mas que afeta a análise. Por essa razão, é necessário examinar as variáveis usadas em análise de agrupamento quanto à multicolinearidade (BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003; HAIR et al., 2005; CORRAR et al., 2007).

O efeito de multicolinearidade entre os caracteres em análise de trilha tem sido destacado na cultura do pimentão (CARVALHO et al., 1999) e da canola (COIMBRA et al., 2005). De maneira geral, esses estudos revelam que, em presença de multicolinearidade moderada a severa, a análise de trilha produz resultados sem nenhuma importância biológica. No entanto, estudos dessa natureza, para a cultura do feijão, não foram encontrados na literatura. Assim, o objetivo deste trabalho foi estudar o efeito da multicolinearidade em análise de agrupamento de cultivares de feijão (*Phaseolus vulgaris* L.), com base em características de produção, fenologia e morfologia, na região da Depressão Central do Estado do Rio Grande do Sul, comparando os padrões obtidos em presença e em ausência de multicolinearidade.

MATERIAL E MÉTODOS

Foram definidas as médias para dois bancos de dados (Casos 1 e 2, Tabela 1) com 10 indivíduos (1, 2, ..., 10) e nove variáveis (X1, X2, ..., X9). A definição dessas médias permite visualizar e evidenciar o efeito

das diferentes escalas de medidas e do grau de multicolinearidade entre as variáveis na análise de agrupamento. Também foram utilizadas as médias de dados reais (Caso 3, Tabela 2) de 13 cultivares de feijão avaliadas em nove experimentos conduzidos no delineamento aleatorizado em blocos, com três repetições, na Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul (latitude 29°42'S, longitude 53°49'W e 95m de altitude). Os experimentos foram conduzidos nos anos agrícolas de 2000/2001 a 2004/2005, em duas épocas de cultivo: safra (semeadura em setembro ou outubro) e safrinha (semeadura em janeiro ou fevereiro).

Em cada experimento, foram avaliados os caracteres produtividade de grãos, número de vagens por planta, número de sementes por vagem, massa de cem grãos, população final de plantas, número de dias da emergência ao florescimento, número de dias da emergência à colheita, altura de inserção da primeira vagem, altura de inserção da última vagem e grau de acamamento. Neste estudo, considerou-se a média de cada um desses caracteres.

Em cada banco de dados, os escores foram padronizados a fim de obter uma variável com média zero e desvio padrão um. Foi determinada a matriz de coeficientes de correlação linear de Pearson entre as variáveis, e foi realizado o diagnóstico de multicolinearidade (CRUZ, 2001) conforme critério de MONTGOMERY & PECK (1981).

Foram determinadas as matrizes de distância euclidiana entre os indivíduos (casos 1 e 2) e entre as cultivares (caso 3). Essas matrizes, em escala relativa, foram utilizadas como medida de dissimilaridade para a análise de agrupamento dos indivíduos e das cultivares pelo método hierárquico da ligação completa (vizinho mais distante), adotando-se o nível de 60% de pareceria para a formação dos grupos (CRUZ, 2001; CRUZ & CARNEIRO, 2003).

Nos três casos, fez-se o agrupamento em situações extremas (sem padronização e com multicolinearidade, com padronização e com multicolinearidade e com padronização e sem multicolinearidade), além de algumas situações intermediárias nos casos 1 e 2, ou seja, com padronização e com diferentes graus de multicolinearidade.

A fim de validar os agrupamentos, fez-se a análise de variância univariada e teste F para verificar se as médias dos grupos diferem. Nos casos em que se tem mais de dois grupos, a conclusão do teste F, sobre a validade dos agrupamentos, é aproximada, pois informa somente que o contraste entre a maior e a menor

Tabela 1 - Média, média padronizada e matriz de coeficientes de correlação de Pearson entre as variáveis X1, X2, ..., X9 de 10 indivíduos (caso 1 e caso 2).

Indivíduo	X1	X2	X3	X4	X5	X6	X7	X8	X9
-----Média - caso 1-----									
1	21,3	19,7	20,0	20,7	19,0	24,7	19,3	23,7	100,0
2	21,7	22,3	20,3	21,7	23,0	21,3	21,3	23,3	91,3
3	21,7	21,7	21,3	23,7	23,3	20,0	22,7	26,3	85,7
4	24,3	24,3	24,0	26,7	24,3	21,7	24,3	25,3	21,3
5	23,0	26,3	23,7	25,7	23,3	25,7	22,3	27,7	20,3
6	206,7	206,7	210,7	207,7	175,0	219,7	207,0	192,3	82,3
7	223,3	215,0	195,3	193,3	201,0	218,0	213,3	194,0	72,7
8	196,0	201,3	196,0	193,0	209,3	217,7	200,0	200,3	6,0
9	220,7	218,0	208,3	211,7	194,3	195,0	222,7	221,7	5,0
10	198,0	201,3	214,7	216,7	213,7	220,7	228,7	210,3	4,0
-----Média padronizada - caso 1-----									
Indivíduo	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	-0,956	-0,979	-0,966	-0,977	-0,981	-0,926	-0,972	-0,962	1,261
2	-0,952	-0,953	-0,963	-0,967	-0,938	-0,960	-0,953	-0,966	1,046
3	-0,952	-0,959	-0,953	-0,946	-0,935	-0,973	-0,939	-0,934	0,908
4	-0,926	-0,933	-0,925	-0,914	-0,925	-0,956	-0,923	-0,945	-0,679
5	-0,939	-0,912	-0,928	-0,925	-0,935	-0,916	-0,943	-0,919	-0,704
6	0,922	0,929	1,006	0,980	0,690	1,000	0,874	0,824	0,824
7	1,091	1,014	0,847	0,829	0,968	0,983	0,936	0,842	0,588
8	0,814	0,874	0,854	0,826	1,057	0,981	0,805	0,909	-1,057
9	1,064	1,045	0,981	1,021	0,896	0,756	1,028	1,135	-1,081
10	0,834	0,874	1,047	1,074	1,104	1,010	1,087	1,015	-1,106
-----Média - caso 2-----									
Indivíduo	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	96,7	98,7	92,3	93,7	94,3	96,0	94,3	93,0	22,0
2	87,7	96,0	94,3	92,7	95,0	101,0	98,3	95,7	24,7
3	94,7	91,7	95,3	86,7	92,0	98,0	97,0	94,3	23,3
4	20,7	20,0	20,3	22,0	20,0	20,7	22,0	21,3	25,0
5	20,3	22,0	21,3	19,3	20,3	20,0	18,0	19,7	25,7
6	85,3	84,0	74,0	81,3	78,7	78,3	76,7	75,3	204,7
7	88,0	81,3	74,7	80,3	81,3	77,0	79,7	76,7	222,0
8	6,0	5,0	5,0	6,0	6,0	4,0	4,0	4,0	225,7
9	5,0	5,0	4,0	4,0	5,0	5,0	6,0	6,0	219,7
10	4,0	5,0	6,0	6,0	4,0	4,0	6,0	5,0	199,7
-----Média padronizada - caso 2-----									
Indivíduo	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	1,084	1,130	1,076	1,102	1,079	1,063	1,049	1,073	-0,967
2	0,871	1,067	1,126	1,077	1,096	1,179	1,145	1,139	-0,940
3	1,037	0,965	1,150	0,928	1,023	1,110	1,114	1,105	-0,954
4	-0,713	-0,730	-0,702	-0,673	-0,717	-0,692	-0,671	-0,680	-0,937
5	-0,722	-0,682	-0,677	-0,740	-0,709	-0,709	-0,766	-0,719	-0,930
6	0,815	0,783	0,624	0,795	0,702	0,650	0,631	0,640	0,850
7	0,879	0,719	0,642	0,770	0,765	0,620	0,702	0,675	1,022
8	-1,060	-1,084	-1,080	-1,069	-1,055	-1,082	-1,099	-1,102	1,059
9	-1,084	-1,084	-1,104	-1,119	-1,079	-1,058	-1,052	-1,053	0,999
10	-1,107	-1,084	-1,055	-1,069	-1,103	-1,082	-1,052	-1,078	0,800
-----Correlação (caso 1 na diagonal superior, caso 2 na diagonal inferior)-----									
	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1		0,9993*	0,9933*	0,9927*	0,9872*	0,9903*	0,9949*	0,9934*	-0,3676 ^{ns}
X2	0,9958*		0,9962*	0,9955*	0,9906*	0,9932*	0,9966*	0,9960*	-0,3872 ^{ns}
X3	0,9902*	0,9959*		0,9998*	0,9907*	0,9950*	0,9982*	0,9966*	-0,4004 ^{ns}
X4	0,9960*	0,9993*	0,9943*		0,9903*	0,9929*	0,9987*	0,9973*	-0,4111 ^{ns}
X5	0,9958*	0,9990*	0,9976*	0,9987*		0,9924*	0,9935*	0,9932*	-0,4390 ^{ns}
X6	0,9891*	0,9962*	0,9995*	0,9948*	0,9978*		0,9926*	0,9887*	-0,3699 ^{ns}
X7	0,9913*	0,9961*	0,9990*	0,9956*	0,9983*	0,9992*		0,9978*	-0,4087 ^{ns}
X8	0,9915*	0,9969*	0,9995*	0,9959*	0,9986*	0,9997*	0,9997*		-0,4353 ^{ns}
X9	-0,3354 ^{ns}	-0,3773 ^{ns}	-0,4237 ^{ns}	-0,3653 ^{ns}	-0,3798 ^{ns}	-0,4188 ^{ns}	-0,4017 ^{ns}	-0,4118 ^{ns}	

* Significativo a 5% de probabilidade de erro, pelo teste t, com oito graus de liberdade, ^{ns} Não significativo.

Tabela 2 - Média, média padronizada e matriz de coeficientes de correlação de Pearson entre os caracteres produtividade de grãos, em t ha⁻¹ (PROD), número de vagens por planta (NVP), número de sementes por vagem (NSV), massa de cem grãos, em g (MCG), população final de plantas, em plantas m⁻² (POP), número de dias da emergência ao florescimento (FLOR), número de dias da emergência à colheita (CICLO), altura de inserção da primeira vagem, em cm (APV), altura de inserção da última vagem, em cm (AUV), e grau de acamamento ⁽¹⁾ (ACA) de 13 cultivares de feijão, avaliadas em nove experimentos ⁽²⁾ (caso 3).

Cultivar	PROD	NVP	NSV	MCG	POP	FLOR	CICLO	APV	AUV	ACA
1 - 'Carioca'	1,498	11,49	4,23	20,38	50,56	37,44	78,22	14,55	43,73	7,33
2 - 'Diamante Negro'	1,247	9,98	4,61	17,23	60,78	37,59	80,48	18,00	44,32	5,41
3 - 'TPS Nobre'	1,550	12,77	4,73	17,54	57,00	37,78	78,22	14,54	39,07	4,85
4 - 'Guapo Brilhante'	1,825	13,17	4,44	18,65	61,33	35,81	75,22	13,01	37,42	3,41
5 - 'Guateian 6662'	1,226	11,00	4,52	15,19	54,85	39,11	77,07	14,18	40,26	7,70
6 - 'IAPAR 44'	1,036	9,63	4,37	15,58	63,74	39,04	79,74	18,77	44,30	6,96
7 - 'Macanudo'	1,468	13,21	4,37	18,15	51,89	35,85	74,70	13,23	39,70	7,00
8 - 'Macotaço'	1,361	12,71	4,41	17,55	52,52	37,85	78,15	14,64	43,11	5,96
9 - 'Minuano'	1,682	14,80	4,25	17,37	54,30	36,26	75,70	13,69	38,88	7,11
10 - 'Pérola'	1,253	8,95	4,18	22,31	47,70	39,04	80,96	19,31	53,42	6,19
11 - 'Rio Tibagi'	1,138	10,48	4,63	14,47	57,04	37,67	78,41	15,04	40,23	6,41
12 - 'TPS Bionobre'	1,530	11,61	4,43	19,43	60,52	37,00	76,33	16,07	45,45	5,74
13 - 'TPS Bonito'	1,406	12,36	3,97	19,68	50,33	37,44	76,56	13,88	35,91	6,41
Média padronizada										
Cultivar	PROD	NVP	NSV	MCG	POP	FLOR	CICLO	APV	AUV	ACA
1 - 'Carioca'	0,433	-0,131	-0,803	1,108	-1,013	-0,077	0,278	-0,357	0,387	0,983
2 - 'Diamante Negro'	-0,694	-1,032	1,049	-0,337	1,048	0,056	1,426	1,285	0,517	-0,674
3 - 'TPS Nobre'	0,670	0,636	1,642	-0,192	0,286	0,222	0,278	-0,361	-0,645	-1,153
4 - 'Guapo Brilhante'	1,907	0,876	0,203	0,313	1,160	-1,539	-1,247	-1,090	-1,011	-2,396
5 - 'Guateian 6662'	-0,793	-0,424	0,595	-1,270	-0,147	1,419	-0,306	-0,532	-0,381	1,302
6 - 'IAPAR 44'	-1,645	-1,244	-0,118	-1,090	1,645	1,353	1,050	1,651	0,512	0,665
7 - 'Macanudo'	0,298	0,902	-0,108	0,085	-0,744	-1,506	-1,510	-0,986	-0,505	0,696
8 - 'Macotaço'	-0,183	0,605	0,071	-0,189	-0,617	0,289	0,240	-0,314	0,249	-0,196
9 - 'Minuano'	1,264	1,854	-0,714	-0,273	-0,259	-1,140	-1,002	-0,766	-0,687	0,792
10 - 'Pérola'	-0,669	-1,647	-1,043	1,992	-1,588	1,353	1,671	1,905	2,530	-0,005
11 - 'Rio Tibagi'	-1,186	-0,732	1,122	-1,603	0,294	0,123	0,372	-0,125	-0,389	0,186
12 - 'TPS Bionobre'	0,579	-0,053	0,164	0,672	0,995	-0,476	-0,682	0,365	0,767	-0,387
13 - 'TPS Bonito'	0,021	0,390	-2,061	0,784	-1,058	-0,077	-0,569	-0,675	-1,344	0,186
Correlação										
	PROD	NVP	NSV	MCG	POP	FLOR	CICLO	APV	AUV	ACA
PROD	1,00	0,80*	-0,12 ^{ns}	0,39 ^{ns}	-0,04 ^{ns}	-0,79*	-0,69*	-0,65*	-0,41 ^{ns}	-0,48 ^{ns}
NVP	0,80*	1,00	-0,10 ^{ns}	-0,01 ^{ns}	-0,11 ^{ns}	-0,76*	-0,82*	-0,84*	-0,70*	-0,14 ^{ns}
NSV	-0,12 ^{ns}	-0,10 ^{ns}	1,00	-0,61*	0,57*	0,06 ^{ns}	0,15 ^{ns}	0,02 ^{ns}	-0,08 ^{ns}	-0,34 ^{ns}
MCG	0,39 ^{ns}	-0,01 ^{ns}	-0,61*	1,00	-0,54 ^{ns}	-0,14 ^{ns}	0,05 ^{ns}	0,14 ^{ns}	0,43 ^{ns}	-0,18 ^{ns}
POP	-0,04 ^{ns}	-0,11 ^{ns}	0,57*	-0,54 ^{ns}	1,00	-0,07 ^{ns}	0,02 ^{ns}	0,19 ^{ns}	-0,15 ^{ns}	-0,42 ^{ns}
FLOR	-0,79*	-0,76*	0,06 ^{ns}	-0,14 ^{ns}	-0,07 ^{ns}	1,00	0,77*	0,66*	0,52 ^{ns}	0,35 ^{ns}
CICLO	-0,69*	-0,82*	0,15 ^{ns}	0,05 ^{ns}	0,02 ^{ns}	0,77*	1,00	0,86*	0,70*	0,05 ^{ns}
APV	-0,65*	-0,84*	0,02 ^{ns}	0,14 ^{ns}	0,19 ^{ns}	0,66*	0,86*	1,00	0,83*	0,06 ^{ns}
AUV	-0,41 ^{ns}	-0,70*	-0,08 ^{ns}	0,43 ^{ns}	-0,15 ^{ns}	0,52 ^{ns}	0,70*	0,83*	1,00	0,12 ^{ns}
ACA	-0,48 ^{ns}	-0,14 ^{ns}	-0,34 ^{ns}	-0,18 ^{ns}	-0,42 ^{ns}	0,35 ^{ns}	0,05 ^{ns}	0,06 ^{ns}	0,12 ^{ns}	1,00

⁽¹⁾ Escala de notas entre um a nove (1 = planta ereta; 9 = planta acamada). ⁽²⁾ 1) Safra-2000/2001, 2) Safrinha-2001, 3) Safra-2001/2002, 4) Safrinha-2002, 5) Safra-2002/2003, 6) Safrinha-2003, 7) Safra-2003/2004, 8) Safrinha-2004 e 9) Safra-2004/2005. * Significativo a 5% de probabilidade de erro, pelo teste t, com 11 graus de liberdade, ^{ns} Não significativo.

média difere. Foram representadas as médias dos grupos em um gráfico de perfis de médias (BARROSO & ARTES, 2003). As análises foram realizadas com o auxílio da planilha eletrônica Office Excel e do programa Genes (CRUZ, 2001).

RESULTADOS E DISCUSSÃO

Os valores das variáveis X1, X2, ..., X8 (Tabela 1 - caso 1) sugerem a formação do grupo 1 (indivíduos 1, 2, 3, 4 e 5) e do grupo 2 (indivíduos 6, 7,

8, 9 e 10), com valores baixos e altos, respectivamente. Por outro lado, os valores da variável X9 sugerem a formação de quatro grupos de indivíduos: grupo 1, com os indivíduos 1, 2 e 3 (valores altos), grupo 2, com os indivíduos 4 e 5 (valores médios baixos), grupo 3, com os indivíduos 6 e 7 (valores médios altos) e o grupo 4, com os indivíduos 8, 9 e 10 (valores baixos). De maneira geral, as variáveis X1, X2, ..., X8 apresentam maior amplitude de variação quando comparadas à variável X9, caracterizando diferentes escalas de medidas e exercendo maior importância na definição dos grupos. Assim, a padronização das variáveis é um procedimento adequado para minimizar o efeito das diferentes escalas de medidas das variáveis, 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).

Os coeficientes de correlação linear de Pearson entre as variáveis X1, X2, ..., X8 foram significativos e de alta magnitude (Tabela 1 - caso 1), o que sugere a presença de multicolinearidade e que o uso de todas as variáveis na análise de agrupamento não é um procedimento adequado, devido a essas variáveis multicolineares, em análise de agrupamento, serem implicitamente ponderadas com maior peso (BARROSO & ARTES, 2003; CRUZ & CARNEIRO, 2003; HAIR et al., 2005; CORRAR et al., 2007). Diante dessas considerações, parece coerente que, para a análise de agrupamento, seja mantida a variável X9 e mais uma seja escolhida entre as oito primeiras variáveis. Fica evidente que, com base nessas duas variáveis, quatro grupos de indivíduos representam a realidade desse banco de dados.

O diagnóstico de multicolinearidade revelou que o número de condição (NC) oscilou entre 5186162 (situação 1 - multicolinearidade severa) a 2 (situação 6 - multicolinearidade fraca) (Figura 1). Na situação 1, o agrupamento foi realizado com base em todas as variáveis não padronizadas e, nesse caso, o efeito da escala de medida e da multicolinearidade age em sentido favorável à separação dos indivíduos em dois grupos de acordo com as variáveis X1, X2, ..., X7 e X8, e as diferenças existentes em relação a X9 não são evidenciadas.

Em um dendrograma, grande mudança de nível indica a união de indivíduos ou grupos heterogêneos (BARROSO & ARTES, 2003). Assim, na situação 1, o grupo 1 (indivíduos 1, 2 e 3) e o grupo 2 (indivíduos 4 e 5) são caracterizados como menos dissimilares que a realidade. O mesmo acontece com o grupo 3 (indivíduos 6 e 7) e 4 (indivíduos 8, 9 e 10). A percentagem de dissimilaridade entre esses grupos,

destacados acima, se aproxima um pouco mais da realidade com a padronização das variáveis (situação 2). No entanto, ainda há o efeito da multicolinearidade que impede a expressão do verdadeiro grau de dissimilaridade entre os grupos. Da situação 2 até a situação 6, há redução gradativa do número de condição e aumento gradativo da percentagem de dissimilaridade entre os grupos com a exclusão de algumas variáveis, ilustrando, visualmente, a atuação da multicolinearidade na análise de agrupamento. Na situação 6, os efeitos da escala de medida e da multicolinearidade foram contornados, respectivamente, pela padronização e pela eliminação de variáveis, permitindo a separação dos indivíduos em quatro grupos, com base em 60% de parença, proporcionando o agrupamento esperado (real), conforme demonstrado anteriormente.

Em relação à variável X9, nas situações 1, 2, 3, 4 e 5, o valor do teste F da análise de variância (1,4) foi não significativo ($P > 0,05$), o que revela que a variância entre grupos não é superior à variância dentro de grupos. Isso significa que o agrupamento não está adequado, pois as médias dos grupos não diferem, diferentemente do que acontece na situação 6, em que as médias dos grupos diferem.

Em relação ao caso 2, os valores das variáveis X1, X2, ..., X8, altamente correlacionadas, sugerem a formação de quatro grupos de indivíduos: grupo 1 (indivíduos 1, 2 e 3, com valores altos), grupo 2 (indivíduos 6 e 7, com valores médios altos), grupo 3 (indivíduos 4 e 5, com valores médios baixos) e o grupo 4 (indivíduos 8, 9 e 10, com valores baixos) (Tabela 1 - caso 2). Já os valores da variável X9 sugerem a formação do grupo 1 (indivíduos 1, 2, 3, 4 e 5) e do grupo 2 (indivíduos 6, 7, 8, 9 e 10), com valores baixos e altos, respectivamente. De maneira semelhante ao caso 1, a padronização e a eliminação de variáveis são procedimentos adequados para contornar, respectivamente, os efeitos da escala de medida e da multicolinearidade entre as variáveis (CORRAR et al., 2007) e proporcionar um dendrograma que evidencie a formação de quatro grupos de indivíduos.

O que difere, em relação ao caso 1, é que, na situação 1 (agrupamento com base em todas as variáveis não padronizadas), o efeito da escala de medida age em sentido favorável à formação de dois grupos, devido à maior amplitude de variação de X9 em relação as demais, e o efeito da multicolinearidade atua favoravelmente para a separação dos indivíduos em quatro grupos de acordo com as variáveis X1, X2, ..., X7 e X8 (Tabela 1 - caso 2). Assim, com base em 60% de parença, já na situação 1 quatro grupos são evidenciados (Figura 2), provavelmente, pela maior

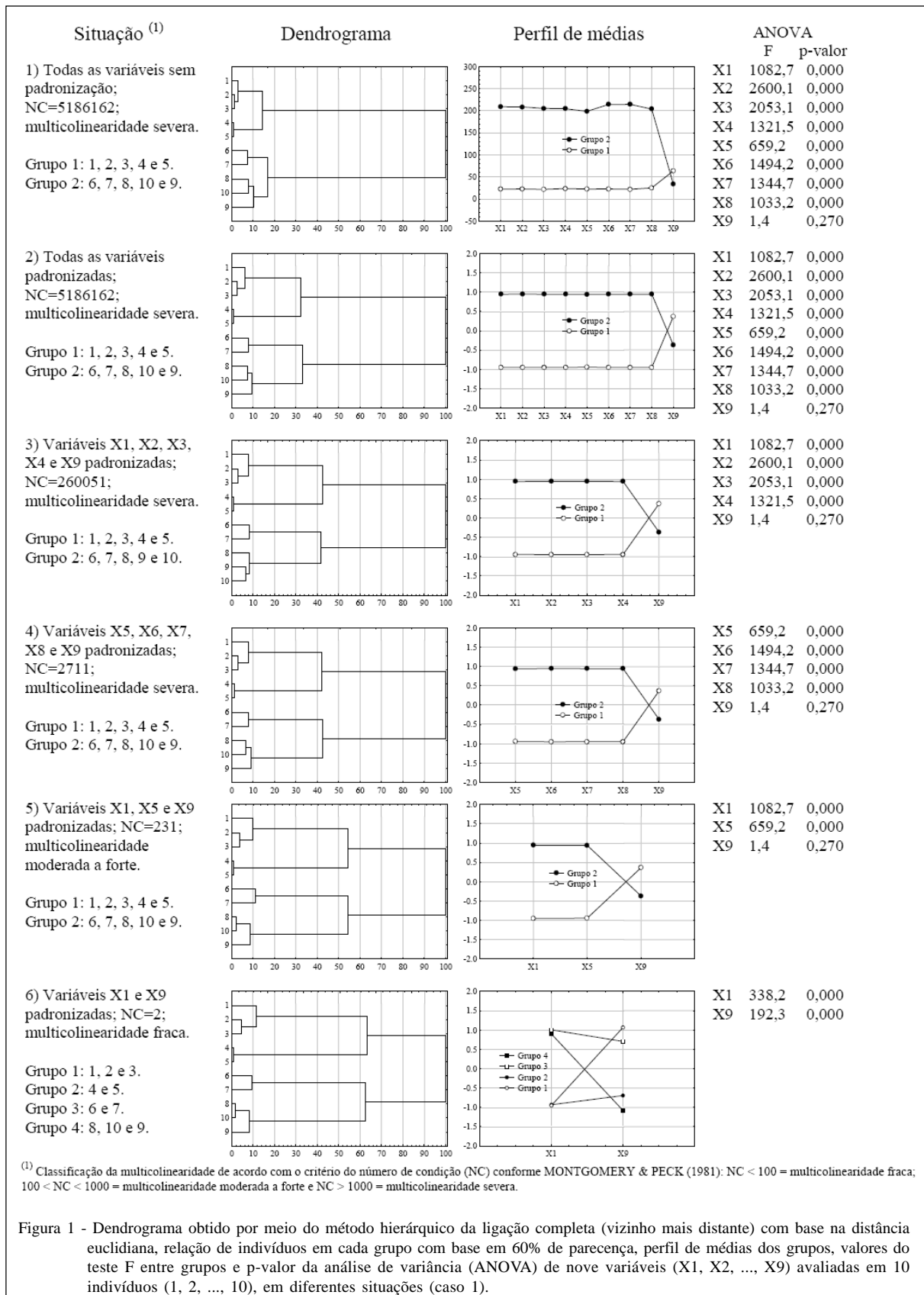
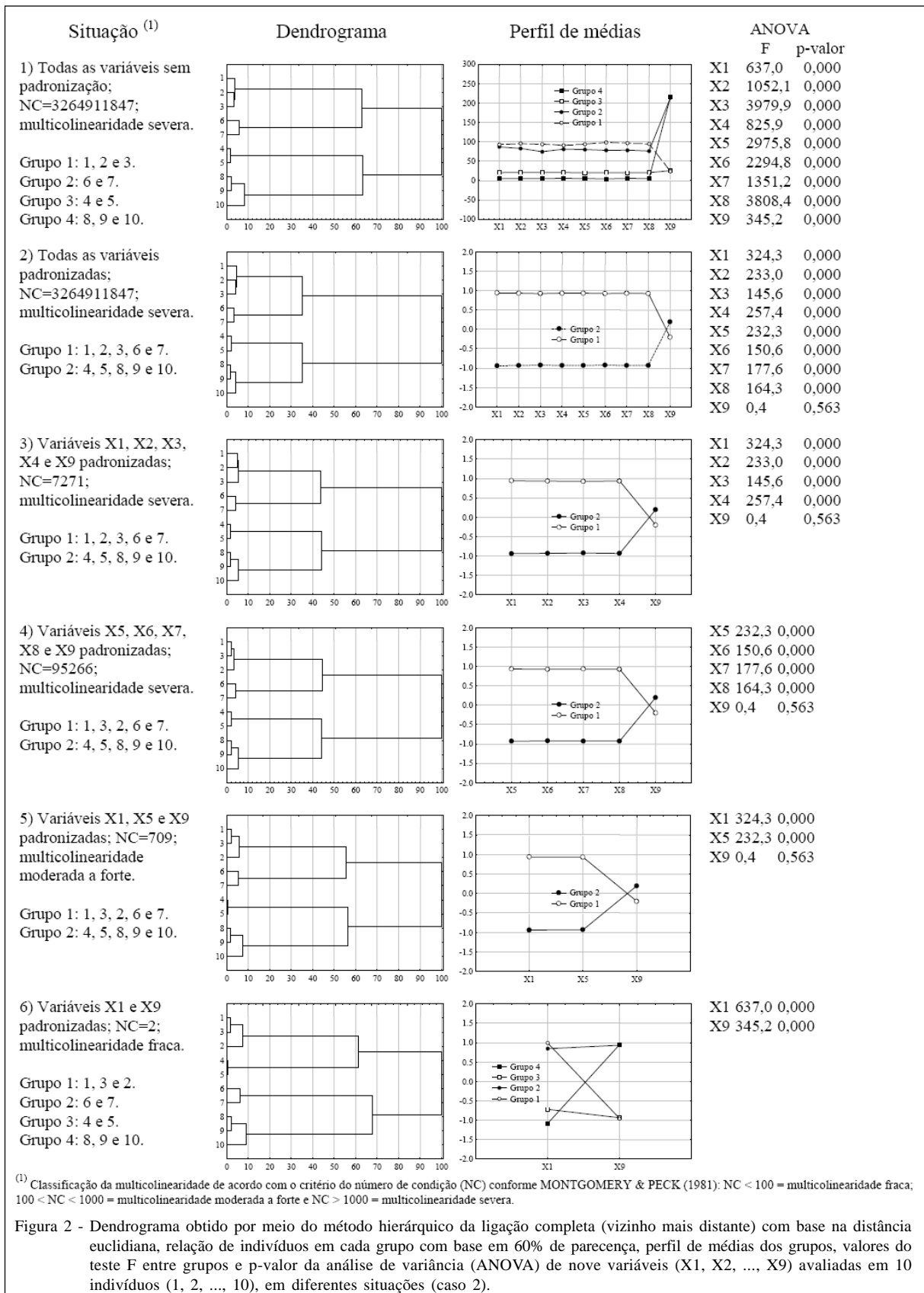


Figura 1 - Dendrograma obtido por meio do método hierárquico da ligação completa (vizinho mais distante) com base na distância euclidiana, relação de indivíduos em cada grupo com base em 60% de parença, perfil de médias dos grupos, valores do teste F entre grupos e p-valor da análise de variância (ANOVA) de nove variáveis (X1, X2, ..., X9) avaliadas em 10 indivíduos (1, 2, ..., 10), em diferentes situações (caso 1).



influência das variáveis X1, X2, ..., X7 e X8 quando comparadas a X9, que atua no sentido da formação de dois grupos. Salienta-se que há somente a variável X9 atuando na formação de dois grupos e ainda que a sua escala de medidas é cerca de duas vezes superior à escala das variáveis X1, X2, ..., X7 e X8. Com o aumento dessa relação, haveria direcionamento no sentido da formação de dois grupos.

Apesar da similaridade dos dendrogramas da situação 1 e da situação 6, o primeiro foi gerado em condições inadequadas (diferentes escalas de medidas e multicolinearidade severa entre as variáveis) e suas semelhanças são meramente casuais. Já na situação 6 os efeitos da escala de medida e da multicolinearidade foram contornados, respectivamente, pela padronização e eliminação de variáveis, permitindo a separação dos indivíduos em quatro grupos, com base em 60% de parença, proporcionando o agrupamento esperado (Figura 2). Nas situações intermediárias a essas duas, o comportamento é semelhante ao caso 1. Os P-valores do teste F da análise de variância validam o agrupamento formado na situação 6 como adequado à realidade desses dados (Figura 2).

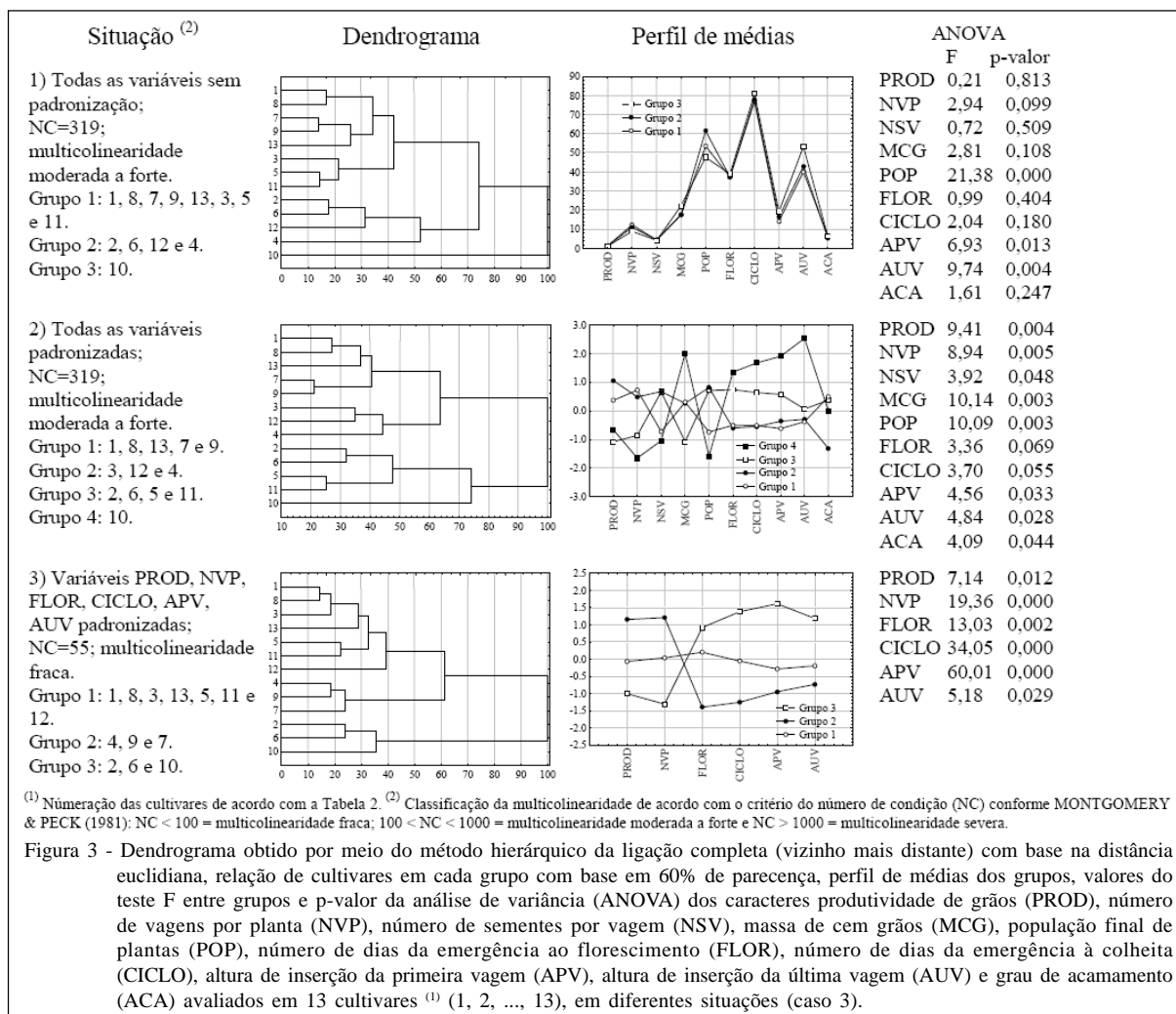
Do ponto de vista prático, destaca-se que, havendo multicolinearidade entre as variáveis, mesmo que seu impacto não seja percebido em uma análise de agrupamento, o seu efeito deve ser contornado para que o agrupamento expresse a realidade dos indivíduos que estão sendo agrupados. Diferentes cenários formados pela combinação de distintos número de indivíduos, número de variáveis, diferentes graus das escalas de medidas e graus de multicolinearidade, além do agrupamento por diversos métodos, com base em diferentes medidas de dissimilaridade poderiam ser simulados e estudados antes dessas informações serem generalizadas. Ainda, outras formas de contornar a multicolinearidade devem ser investigadas, pelo fato de o descarte de variáveis que causam multicolinearidade ser um procedimento que pode excluir variáveis importantes em um programa de melhoramento genético. Nesse sentido, a distância generalizada de Mahalanobis (D^2), entre os pares de genótipos, tem sido amplamente utilizada como medida de dissimilaridade nos métodos de agrupamento. A distância D^2 considera a existência de correlações entre os caracteres analisados por meio da matriz de variâncias e covariâncias residuais, porém, para ser calculada, necessita de experimentos com repetições e ainda que exista distribuição normal multidimensional e homogeneidade da matriz de covariâncias residuais, o que restringe o seu uso (CRUZ & REGAZZI, 1997; CRUZ & CARNEIRO, 2003).

Os dados do caso 3 revelam a dificuldade de estabelecer, *a priori*, qual a separação adequada das cultivares, diferentemente dos casos anteriores (Tabela 2). No entanto, o que deve ser considerado é que, existindo diferentes escalas de medidas e presença de multicolinearidade, estas devem ser contornadas para evitar agrupamentos inadequados. Apesar da dificuldade de visualização, caso existam esses problemas, estes atuarão gerando diferentes padrões de agrupamento não condizentes com a realidade.

Usando 60% de parença como critério para a separação das cultivares de feijão, houve a formação de três grupos em presença do efeito de diferentes escalas de medida das variáveis e da multicolinearidade (Figura 3 - situação 1). Após a padronização das variáveis, houve a formação de quatro grupos de cultivares (Figura 3 - situação 2) e, após a minimização do efeito da escala de medida e da multicolinearidade (Figura 3 - situação 3), novamente, três grupos foram formados, porém as cultivares pertencentes em cada grupo foram diferentes. Portanto, pode-se inferir que os padrões de agrupamento de cultivares de feijão diferem em presença e em ausência de multicolinearidade. Na situação 3 (Figura 3), o p-valor das variáveis utilizadas para o agrupamento revelou a consistência deste. Outro aspecto que reforça a adequabilidade desse agrupamento é que a média de cada cultivar foi estimada com base em 27 observações (três repetições x nove experimentos), o que o torna uma estimativa confiável, caracterizando que existe homogeneidade dentro do grupo e heterogeneidade entre os grupos. A identificação de cultivares divergentes, com base em apenas um experimento, poderá ser pouco eficiente devido à variabilidade de resposta das cultivares frente às condições ambientais (interação cultivares x ambientes), e o uso dos dados de sete e seis experimentos, respectivamente, para os métodos de Ward e de Tocher, foi considerado suficiente para o agrupamento de cultivares de feijão (CARGNELUTTI FILHO et al., 2009).

CONCLUSÕES

Os padrões de agrupamento de cultivares de feijão (*Phaseolus vulgaris* L.), com base em características de produção, fenologia e morfologia, na região da Depressão Central do Estado do Rio Grande do Sul, diferem em presença e em ausência de multicolinearidade. O efeito da multicolinearidade deve ser contornado para proporcionar padrões de agrupamento adequados.



AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de Produtividade em Pesquisa aos autores deste trabalho.

REFERÊNCIAS

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

BONETT, L.P. et al. Divergência genética em germoplasma de feijoeiro comum coletado no estado do Paraná, Brasil. **Semina**, v.27, p.547-560, 2006.

CARGNELUTTI FILHO, A. et al. Número necessário de experimentos para a análise de agrupamento de cultivares de feijão. **Ciência Rural**, v.39, p.371-378, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782009000200009&lng=pt&nrm=iso>. Acesso em: 26 jun. 2009. doi: 10.1590/S0103-84782008005000059.

CARGNELUTTI FILHO, A. et al. Comparação de métodos de agrupamento para o estudo da divergência genética em cultivares de feijão. **Ciência Rural**, v.38, p.2138-2145, 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782008000800008&lng=pt&nrm=iso>. Acesso em: 26 jun. 2009. doi: 10.1590/S0103-84782008000800008.

CARVALHO, C.G.P. et al. Análise de trilha sob multicolinearidade em pimentão. **Pesquisa Agropecuária Brasileira**, v.34, p.603-613, 1999. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X1999000400011&lng=pt&nrm=iso>. Acesso em: 26 jun. 2009. doi: 10.1590/S0100-204X1999000400011.

COIMBRA, J.L.M. et al. Consequências da multicolinearidade sobre a análise de trilha em canola. **Ciência Rural**, v.35, p.347-352, 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782005000200015&lng=pt&nrm=iso>. Acesso em: 26 jun. 2009. doi: 10.1590/S0103-84782005000200015.

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**: versão Windows: aplicativo computacional em genética e estatística. Viçosa: UFV, 2001. 648p.
- 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.
- HAIR, J.F. et al. **Análise multivariada de dados**. 5.ed. Porto Alegre: Bookman, 2005. 593p.
- MALUF, W.R.; FERREIRA, P.E. Análise multivariada da divergência genética em feijão-vagem (*Phaseolus vulgaris* L.). **Horticultura Brasileira**, v.1, p.31-34, 1983.
- 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, 1981. 504p.