

# Não normalidade multivariada e multicolinearidade na análise de trilha em milho

Marcos Toebe<sup>(1)</sup> e Alberto Cargnelutti Filho<sup>(1)</sup>

<sup>(1)</sup>Universidade Federal de Santa Maria, Departamento de Fitotecnia, Avenida Roraima, nº 1.000, Camobi, CEP 97105-900 Santa Maria, RS. E-mail: m.toebe@gmail.com, alberto.cargnelutti.filho@gmail.com

Resumo – O objetivo deste trabalho foi avaliar a interferência da não normalidade multivariada e da multicolinearidade na análise de trilha, em milho. Foram utilizados os dados de 13 ensaios de competição de cultivares de milho. Foram mensuradas a variável principal (produtividade de grãos) e sete variáveis explicativas (número de dias até o florescimento, estatura de plantas, altura de inserção da espiga, posição relativa da espiga, número de plantas, número de espigas e prolificidade), em cada cultivar. Procedeu-se, então, à transformação dos dados e ao diagnóstico de normalidade univariada e multivariada. Antes e após a transformação de dados, foram calculados os coeficientes de correlação e realizado o diagnóstico de multicolinearidade. A análise de trilha foi realizada por três métodos: tradicional; sob condições de multicolinearidade (análise de trilha em crista); e tradicional com eliminação de variáveis. A transformação de dados reduz o grau de multicolinearidade e a variabilidade das estimativas dos efeitos diretos, na análise de trilha tradicional com alto grau de multicolinearidade. A multicolinearidade exerce maior impacto sobre a estimativa dos efeitos diretos nas análises de trilha do que a não normalidade multivariada. A análise de trilha tradicional com eliminação de variáveis é mais adequada do que a análise de trilha em crista.

Termos para indexação: *Zea mays*, análise em crista, eliminação de variáveis, transformações Box-Cox.

## Multivariate nonnormality and multicollinearity in path analysis in corn

Abstract – The objective of this work was to evaluate the effect of multivariate nonnormality and multicollinearity in the path analysis of corn. We used data from 13 corn cultivar competition trials. The response variable (grain yield) and seven explanatory variables (number of days to tasseling, plant height, ear height, relative ear position, number of plants, number of ears and prolificity) were measured in each cultivar. Then, data transformation and the univariate and multivariate normality diagnosis were proceeded. The correlation coefficients were calculated and the diagnosis of multicollinearity was performed, before and after data transformation. The path analysis was done according to three methods: traditional; under multicollinearity (ridge path analysis); and traditional with variable elimination. Data transformation reduces the degree of multicollinearity and the variability of the direct effects, in the traditional path analysis with high multicollinearity. Multicollinearity exerts more impact on the estimation of the direct effects in path analysis than multivariate nonnormality. The traditional path analysis with elimination of variables is more appropriate than the ridge path analysis.

Index terms: *Zea mays*, ridge analysis, elimination of variables, Box-Cox transformations.

## Introdução

Pesquisas têm sido realizadas na cultura de milho, a fim de identificar a direção e a intensidade das relações lineares entre variáveis (caracteres). Além do estudo das correlações fenotípicas, genotípicas e ambientais, trabalhos complementares com o uso de análise de trilha vêm sendo realizados (Carvalho et al., 2001; Ahmad & Saleem, 2003; Alvi et al., 2003; Mohammadi et al., 2003; Saidaiah et al., 2008; Bello et al., 2010; Kumar et al., 2011; Pavan et al., 2011; Selvaraj &

Nagarajan, 2011; Toebe & Cargnelutti Filho, 2013), para identificar as verdadeiras relações de causa e efeito entre os caracteres (Cruz & Regazzi, 1997).

Para que os resultados da análise de trilha sejam fidedignos, é necessário que as pressuposições do modelo sejam atendidas, como por exemplo, a normalidade dos dados (Hair et al., 2009). Em caso de não atendimento da normalidade, pode-se realizar a transformação de dados por meio da família de transformações Box-Cox (Box & Cox, 1964), utilizada de forma eficiente com a cultura de abobrinha

italiana (Couto et al., 2009). Além do diagnóstico de normalidade, é importante verificar o grau de multicolinearidade entre as variáveis explicativas (Cruz & Carneiro, 2006). Caso a análise de trilha seja realizada com alto grau de multicolinearidade, as estimativas dos efeitos diretos e indiretos podem ser viesadas, com valores em módulo superiores a 1 e, portanto, sem sentido biológico, conforme já verificado nas culturas de pimentão (Carvalho et al., 1999), milho (Carvalho et al., 2001; Toebe & Cargnelutti Filho, 2013), soja (Bizeti et al., 2004) e canola (Coimbra et al., 2005). Na presença de elevado grau de multicolinearidade, tem sido recomendado fazer a análise de trilha sob multicolinearidade (análise de trilha em crista), com acréscimo de um valor  $k$  aos elementos da diagonal da matriz de correlação, ou realizar a análise de trilha tradicional, com a eliminação de variáveis altamente correlacionadas (Cruz & Carneiro, 2006).

Em 14 ensaios de milho, com distribuição normal multivariada e elevado grau de multicolinearidade entre as variáveis explicativas, Toebe & Cargnelutti Filho (2013) constataram que a realização da análise de trilha tradicional, com eliminação de variáveis altamente correlacionadas, foi mais adequada do que a realização da análise de trilha em crista, para a estimação precisa dos efeitos diretos das variáveis explicativas sobre a produtividade de grãos. No entanto, não foram encontrados estudos que mostrassem o efeito da não normalidade multivariada sobre a estimação dos coeficientes de trilha, ou que avaliassem os possíveis efeitos da não normalidade multivariada de acordo com os níveis de multicolinearidade e as técnicas de análise de trilha utilizadas.

O objetivo deste trabalho foi avaliar a interferência da não normalidade multivariada e da multicolinearidade, na análise de trilha em milho.

## Material e Métodos

Os dados utilizados neste estudo foram provenientes de 13 ensaios (experimentos) de competição de cultivares de milho, realizados no Estado do Rio Grande do Sul, nos anos agrícolas 2002/2003, 2003/2004 e 2004/2005 (Tabela 1). Esses dados foram obtidos de Relatórios, Atas e Resumos das Reuniões Técnicas Anuais de Milho e Sorgo, de 2003, 2004, 2005, coordenados pela Fundação Estadual de Pesquisa Agropecuária (Fepagro), do Estado do Rio Grande do Sul (Tabela 1). O delineamento de blocos ao acaso foi

utilizado nos 13 ensaios, com uso de três repetições. Cada parcela foi composta por uma cultivar, semeada em duas linhas de 5 m de comprimento, com 0,7 ou 0,8 m entre linhas. O número de cultivares nos ensaios variou entre 12 e 40.

Em cada cultivar, de cada ensaio, foram mensuradas sete variáveis explicativas – número de dias até o florescimento masculino (FM), estatura de plantas à colheita (EP), altura da inserção da espiga à colheita (AIE), posição relativa da espiga (PRE, obtida pela divisão de AIE por EP), número de plantas à colheita (NP), número de espigas à colheita (NE), e, prolificidade (PLI, obtida pela divisão de NE por NP) –, além da variável principal produtividade de grãos, que foi ajustada a 13% de umidade. Para cada variável mensurada, foram utilizadas as médias obtidas a partir das três repetições utilizadas nos ensaios.

Para cada variável, em cada ensaio, foi realizado o diagnóstico de normalidade univariada dos resíduos, por meio do teste de Shapiro-Wilk (Shapiro & Wilk, 1965). Em seguida, em cada ensaio, foi realizado o diagnóstico de normalidade multivariada (NM) dos resíduos, por meio do teste de normalidade de Shapiro-Wilk generalizado por Royston (1983), conforme descrito por Ferreira (2008). Em seguida, procedeu-se à transformação dos dados das variáveis que não se ajustaram à distribuição normal ( $p \leq 0,10$ ). Para isso, utilizou-se a família de transformações Box-Cox (Box & Cox, 1964), com valores de potência  $\lambda$  no intervalo de  $-5 \leq \lambda \leq 5$ . Após a transformação dos dados, procedeu-se novamente ao diagnóstico da normalidade univariada e multivariada, a fim de se verificar a eficiência da transformação dos dados. Nos ensaios 1 e 8, não se obteve normalidade multivariada após a transformação das variáveis. Assim, todas as variáveis desses dois ensaios foram transformadas, independentemente da distribuição normal univariada, e mantiveram-se os dados transformados nos casos em que representavam acréscimo no valor-p da normalidade univariada.

Em seguida, foram calculados os coeficientes de correlação linear de Pearson entre as oito variáveis mensuradas, antes (distribuição multivariada não normal) e depois da transformação de dados (distribuição multivariada normal), nos 13 ensaios. Esses coeficientes foram alocados em uma matriz de correlação simétrica de dimensão  $8 \times 8$ , que foi posteriormente dividida em duas matrizes menores. Uma foi denominada de matriz de correlação simétrica

$X'X$ , de dimensão  $7 \times 7$ , na qual foram alocados os coeficientes de correlação obtidos entre as sete variáveis explicativas apenas; a outra foi denominada matriz de correlação  $X'Y$ , de dimensão  $7 \times 1$ , na qual foram alocados os coeficientes de correlação de cada uma das sete variáveis explicativas com a variável dependente (produtividade de grãos).

O diagnóstico de multicolinearidade foi realizado em cada ensaio (antes e após a transformação de dados), com base na matriz de correlação  $X'X$ , tendo-se considerado as seguintes situações: as sete variáveis explicativas; as sete variáveis explicativas após a adição de uma constante  $k$  à diagonal da matriz de correlação  $X'X$ ; e após a eliminação de variáveis explicativas altamente correlacionadas.

O diagnóstico de multicolinearidade foi realizado por meio de dois métodos: fator de inflação de variância (FIV) e número de condição (NC). No primeiro método, o FIV de cada variável explicativa foi obtido na diagonal da inversa da matriz de correlação  $X'X$ , ou seja, na diagonal da matriz  $X'X^{-1}$  (Cruz & Carneiro, 2006; Hair et al., 2009). De forma alternativa, o FIV pode ser obtido pelo inverso da tolerância ( $FIV = 1/\text{tolerância}$ ), em que a tolerância indica a proporção da variação de uma variável explicativa que independe das demais variáveis explicativas. Assim,  $\text{Tolerância} = 1 - R_x^2$ , em que,  $R_x^2$  é o coeficiente de

ajuste da regressão (coeficiente de determinação do modelo) da variável explicativa  $X$ , em função das demais variáveis explicativas presentes (Fávero et al., 2009). Considerou-se haver multicolinearidade severa quando os valores de FIV foram maiores que 10 (Hair et al., 2009). No método NC, verificou-se o grau de multicolinearidade na matriz de correlação  $X'X$ , por meio do NC que representa a razão entre o maior e o menor autovalor da matriz de correlação  $X'X$  (Montgomery & Peck, 1982). Quando o NC resultante dessa divisão foi menor ou igual a 100, considerou-se haver multicolinearidade fraca entre as variáveis explicativas; para  $100 < NC < 1.000$ , considerou-se haver multicolinearidade moderada a severa; e para  $NC \geq 1.000$ , considerou-se multicolinearidade severa.

Em cada ensaio (antes e após a transformação de dados), realizaram-se análises de trilha quanto à variável principal produtividade de grãos, em função das variáveis explicativas (FM, EP, AIE, PRE, NP, NE e PLI), conforme metodologia descrita por Cruz & Regazzi (1997) e Cruz & Carneiro (2006). Estimaram-se os efeitos diretos e indiretos das variáveis explicativas sobre a produtividade de grãos, mediante três métodos de análise de trilha: tradicional; em condições de multicolinearidade (análise de trilha em crista); e tradicional com eliminação de variáveis. Nos três métodos, considerou-se que cada variável

**Tabela 1.** Resultados dos ensaios de competição de cultivares de milho, no Estado do Rio Grande do Sul, utilizados no presente estudo<sup>(1)</sup>.

Ensaio	Ano de publicação	Autor	Número do trabalho <sup>(2)</sup>	Páginas	Local	Categoria <sup>(3)</sup>	Ciclo	Ano	Cultivares
1	2003	Guadagnin, J.P. et al. (2003a)	1	1-6	Pelotas	Estadual	Precoce	2002/03	36
2	2003	Guadagnin, J.P. et al. (2003a)	1	1-6	São Borja	Estadual	Precoce	2002/03	36
3	2003	Guadagnin, J.P. et al. (2003b)	8	1-5	Rio Pardo	Indicado	Superprecoce	2002/03	18
4	2003	Guadagnin, J.P. et al. (2003c)	7	1-5	Santa Maria	Indicado	Precoce	2002/03	27
5	2004	Guadagnin, J.P. et al. (2004a)	3	1-6	Passo Fundo	Estadual	Precoce	2003/04	40
6	2004	Guadagnin, J.P. et al. (2004b)	8	1-5	Rio Pardo	Indicado	Superprecoce	2003/04	16
7	2004	Guadagnin, J.P. et al. (2004b)	8	1-5	Santa Maria	Indicado	Superprecoce	2003/04	16
8	2005	Guadagnin, J.P. et al. (2005a)	13	62	Carazinho	Estadual	Superprecoce	2004/05	17
9	2005	Guadagnin, J.P. et al. (2005b)	13	63	Capão do Leão	Estadual	Precoce	2004/05	32
10	2005	Guadagnin, J.P. et al. (2005b)	13	63	Carazinho	Estadual	Precoce	2004/05	32
11	2005	Guadagnin, J.P. et al. (2005b)	13	63	Cruz Alta	Estadual	Precoce	2004/05	32
12	2005	Guadagnin, J.P. et al. (2005c)	14	64	Cruz Alta	Indicado	Superprecoce	2004/05	12
13	2005	Guadagnin, J.P. et al. (2005d)	15	65	Santa Maria	Indicado	Precoce	2004/05	30

<sup>(1)</sup>Os dados foram provenientes dos relatórios, atas e resumos das Reuniões Técnicas Anuais de Milho e Sorgo (RTAMS), realizadas nos anos de 2003, 2004 e 2005, pela Fundação Estadual de Pesquisa Agropecuária (Fepagro) do Estado do Rio Grande do Sul. Os trabalhos publicados nas RTAMS de 2003 e 2004 estão disponibilizados em formato digital (Cd-Rom), e os trabalhos publicados nas RTAMS de 2005 estão em formato impresso. <sup>(2)</sup>Os trabalhos foram numerados de acordo com sua disposição nas publicações. <sup>(3)</sup>Os ensaios da categoria estadual contêm cultivares ainda não indicadas aos produtores. Os ensaios da categoria "indicado" contêm cultivares já indicadas aos produtores.

explicativa exerce efeito direto sobre a produtividade de grãos, e atua indiretamente pelos seus efeitos nas demais variáveis explicativas.

Na análise de trilha tradicional, realizou-se a estimação dos efeitos diretos e indiretos, tendo-se desconsiderado os possíveis efeitos adversos da multicolinearidade. Para isso, inicialmente, realizou-se a padronização das variáveis e estabeleceu-se o modelo da análise de trilha: produtividade de grãos =  $\hat{\beta}_1 FM + \hat{\beta}_2 EP + \hat{\beta}_3 AIE + \hat{\beta}_4 PRE + \hat{\beta}_5 NP + \hat{\beta}_6 NE + \hat{\beta}_7 PLI + \text{resíduo}$ , em que  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ ,  $\hat{\beta}_3$ ,  $\hat{\beta}_4$ ,  $\hat{\beta}_5$ ,  $\hat{\beta}_6$  e  $\hat{\beta}_7$  são os estimadores dos efeitos diretos das variáveis FM, EP, AIE, PRE, NP, NE e PLI, respectivamente. A seguir, utilizou-se o sistema de equações normais  $X'X\hat{\beta} = X'Y$ , para a obtenção dos efeitos diretos e indiretos de cada variável explicativa sobre a produtividade de grãos, conforme Cruz & Regazzi (1997) e Cruz & Carneiro (2006).

Na análise de trilha em condições de multicolinearidade ou análise de trilha em crista (Carvalho & Cruz, 1996; Cruz & Carneiro, 2006), mantiveram-se as sete variáveis explicativas (FM, EP, AIE, PRE, NP, NE e PLI), para a estimação dos efeitos diretos e indiretos sobre a produtividade de grãos. No entanto, adicionou-se uma constante  $k$  à diagonal da matriz de correlação  $X'X$ , a fim de reduzir a variância associada ao estimador de mínimos quadrados da análise de trilha (Carvalho & Cruz, 1996). Dessa forma, o sistema de equações normais  $X'X\hat{\beta} = X'Y$  passou a ser  $(X'X + k)\hat{\beta} = X'Y$ . Testou-se a adição de 21 valores da constante  $k$  ( $k = 0,00, 0,05, 0,10, \dots, 1,00$ ) e escolheu-se o menor valor de  $k$  a partir do qual os coeficientes de trilha se estabilizaram, conforme preconizado por Carvalho & Cruz (1996).

Na análise de trilha tradicional com eliminação de variáveis, constatou-se elevada associação linear entre as variáveis EP, AIE e PRE e entre as variáveis NP, NE e PLI. Nos 13 ensaios, verificou-se que a redução a níveis satisfatórios de multicolinearidade somente foi obtida após a exclusão das variáveis AIE e NE. Assim, estas duas variáveis explicativas foram excluídas da análise de trilha de cada ensaio e, posteriormente, calcularam-se os efeitos diretos e indiretos de cinco variáveis explicativas (FM, EP, PRE, NP e PLI) sobre a produtividade de grãos. O modelo da análise de trilha foi o seguinte: produtividade de grãos =  $\hat{\beta}_1 FM + \hat{\beta}_2 EP + \hat{\beta}_3 PRE + \hat{\beta}_4 NP + \hat{\beta}_5 PLI + \text{resíduo}$ , em que  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ ,  $\hat{\beta}_3$ ,  $\hat{\beta}_4$  e  $\hat{\beta}_5$  são os

estimadores dos efeitos diretos das variáveis FM, EP, PRE, NP e PLI, respectivamente, sobre a produtividade de grãos. Da mesma forma que na análise de trilha tradicional, foi utilizado o sistema de equações normais  $X'X\hat{\beta} = X'Y$ , para a obtenção dos efeitos diretos e indiretos de cada variável sobre a produtividade de grãos.

Para cada ensaio e método de análise de trilha, foram calculados o coeficiente de determinação e o efeito da variável residual da análise de trilha, antes e após a transformação de dados. Maiores detalhes em relação à obtenção dos efeitos diretos, coeficientes de determinação, efeito da variável residual e diagramas causais das análises de trilha são fornecidos em Toebe & Cargnelutti Filho (2013). As análises estatísticas foram realizadas com auxílio do aplicativo Microsoft Office Excel e dos programas Genes (Cruz, 2006), Statistica 7.0 (StatSoft, Tulsa, OK, EUA) e R (R Development Core Team, Viena, Áustria).

## Resultados e Discussão

O número de dias até o florescimento masculino (FM), o número de plantas (NP), o número de espigas (NE) e a prolificidade (PLI) foram as variáveis que violaram a pressuposição de normalidade ( $p \leq 0,10$ ) com maior frequência (Tabela 2). As demais variáveis apresentaram distribuição normal, na maioria dos ensaios. A estatura de plantas (EP), a altura da inserção da espiga (AIE), a posição relativa da espiga (PRE) e a produtividade de grãos não se ajustaram à distribuição normal apenas em 3, 3, 2 e 1 dos 13 ensaios avaliados, respectivamente. Os 13 ensaios apresentaram distribuição multivariada não normal, antes da transformação de dados, e distribuição multivariada normal ( $p > 0,10$ ) após a transformação. Hair et al. (2009) destacam que o atendimento da pressuposição de normalidade univariada contribui para a obtenção da distribuição normal multivariada. Assim, o diagnóstico de normalidade e a transformação de dados univariados podem representar uma alternativa mais simples, em comparação ao diagnóstico de normalidade e a transformação multivariada de dados.

Nos 13 ensaios, a média do coeficiente de correlação entre FM e produtividade de grãos foi negativa e de baixa magnitude, com valor próximo aos obtidos por Kumar et al. (2011) e por Pavan et al. (2011). A média do coeficiente de correlação entre EP e produtividade de grãos foi positiva, porém de magnitude menor que

**Tabela 2.** Significância do teste de normalidade univariado de Shapiro-Wilk para cada variável e do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston (NM) entre as oito variáveis, em 13 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	Box-Cox <sup>(2)</sup>	Variável								NM
		FM	EP	AIE	PRE	NP	NE	PLI	PROD	
1	Sem transformação	0,002	0,440	0,517	0,176	0,921	0,210	0,017	0,138	0,008
	Com transformação	0,002	0,442	0,527	0,176	0,986	0,874	0,889	0,547	0,138
	Valor de $\lambda$ <sup>(3)</sup>	STA <sup>(4)</sup>	1,241	1,199	-	1,604	-0,025	-1,699	1,896	
2	Sem transformação	0,005	0,023	0,019	0,213	0,555	0,632	0,097	0,795	0,007
	Com transformação	0,006	0,932	0,787	0,213	0,555	0,632	0,908	0,795	0,283
	Valor de $\lambda$	2,764	5,000	3,733	-	-	-	4,065	-	
3	Sem transformação	0,105	0,337	0,494	0,999	0,009	0,013	0,001	0,177	0,000
	Com transformação	0,105	0,337	0,494	0,999	0,136	0,933	0,594	0,177	0,397
	Valor de $\lambda$	-	-	-	-	5,000	3,171	5,000	-	
4	Sem transformação	0,021	0,398	0,973	0,839	0,041	0,016	0,095	0,763	0,024
	Com transformação	0,021	0,398	0,973	0,839	0,983	0,847	0,095	0,763	0,345
	Valor de $\lambda$	STA	-	-	-	5,000	5,000	STA	-	
5	Sem transformação	0,335	0,198	0,424	0,087	0,003	0,116	0,577	0,885	0,031
	Com transformação	0,335	0,198	0,424	0,199	0,795	0,116	0,577	0,885	0,451
	Valor de $\lambda$	-	-	-	5,000	5,000	-	-	-	
6	Sem transformação	0,158	0,004	0,975	0,246	0,049	0,267	0,012	0,610	0,005
	Com transformação	0,158	0,307	0,975	0,246	0,104	0,267	0,168	0,610	0,242
	Valor de $\lambda$	-	5,000	-	-	5,000	-	-5,000	-	
7	Sem transformação	0,098	0,160	0,620	0,010	0,074	0,051	0,103	0,015	0,005
	Com transformação	0,106	0,160	0,620	0,528	0,277	0,557	0,103	0,971	0,318
	Valor de $\lambda$	-1,761	-	-	-5,000	5,000	4,554	-	3,315	
8	Sem transformação	0,064	0,678	0,023	0,118	0,362	0,432	0,175	0,314	0,076
	Com transformação	0,064	0,678	0,023	0,118	0,402	0,914	0,578	0,314	0,135
	Valor de $\lambda$	STA	-	STA	-	5,000	5,000	5,000	-	
9	Sem transformação	0,911	0,005	0,043	0,904	0,002	0,070	0,524	0,332	0,002
	Com transformação	0,911	0,569	0,656	0,904	0,431	0,615	0,524	0,332	0,933
	Valor de $\lambda$	-	-1,799	-2,200	-	3,342	2,382	-	-	
10	Sem transformação	0,284	0,438	0,123	0,481	0,324	0,054	0,013	0,968	0,074
	Com transformação	0,284	0,438	0,123	0,481	0,324	0,540	0,412	0,968	0,542
	Valor de $\lambda$	-	-	-	-	-	3,511	4,285	-	
11	Sem transformação	0,002	0,166	0,781	0,544	0,151	0,533	0,479	0,394	0,061
	Com transformação	0,008	0,166	0,781	0,544	0,151	0,533	0,479	0,394	0,115
	Valor de $\lambda$	-1,899	-	-	-	-	-	-	-	
12	Sem transformação	0,008	0,819	0,841	0,179	0,361	0,018	0,118	0,772	0,028
	Com transformação	0,008	0,819	0,841	0,179	0,361	0,380	0,118	0,772	0,121
	Valor de $\lambda$	STA	-	-	-	-	5,000	-	-	
13	Sem transformação	0,450	0,439	0,384	1,000	0,230	0,001	0,011	0,764	0,008
	Com transformação	0,450	0,439	0,384	1,000	0,230	0,028	0,394	0,764	0,343
	Valor de $\lambda$	-	-	-	-	-	-0,816	-2,822	-	

<sup>(1)</sup>Ensaio definido na Tabela 1. <sup>(2)</sup>Transformação ou não de dados, por meio da família de transformações Box-Cox, das variáveis que não se ajustaram à distribuição normal, com o respectivo valor de potência  $\lambda$  utilizado. <sup>(3)</sup>Valores de  $\lambda = [5]$ , melhor valor de  $\lambda$  para a transformação dos dados não foi encontrado entre  $-5 \leq \lambda \leq 5$ . <sup>(4)</sup>Sem transformação adequada, nenhum valor de  $\lambda$  dentro do intervalo de -5 a 5 melhorou a distribuição dos dados. FM, florescimento masculino; EP, estatura de plantas; AIE, altura de inserção da espiga; PRE, posição relativa da espiga; NP, número de plantas; NE, número de espigas; PLI, prolificidade; PROD, produtividade de grãos.

a verificada por Alvi et al. (2003), Mohammadi et al. (2003), Saidaiah et al. (2008), Bello et al. (2010) e Kumar et al. (2011). As variáveis AIE e PRE também apresentaram reduzida associação com a produtividade de grãos, com valores de correlação semelhantes aos descritos por Alvi et al. (2003), Saidaiah et al. (2008) e Bello et al. (2010). Em razão da ampla oscilação verificada nos 13 ensaios quanto ao sentido e à intensidade das associações das variáveis FM, EP, AIE e PRE com a produtividade de grãos, é provável que essas variáveis não apresentem relação de causa e efeito (Cruz & Regazzi, 1997). No entanto, as variáveis NE, PLI e NP apresentaram os maiores coeficientes de correlação com a produtividade de grãos, o que sugere possível relação de causa e efeito (Cruz & Regazzi, 1997).

A fim de verificar a interferência da transformação de dados na magnitude e direção dos coeficientes de correlação estimados entre as variáveis explicativas e a produtividade de grãos, estimou-se a correlação entre os 13 coeficientes de correlação (13 ensaios) obtidos antes e os 13 coeficientes obtidos após a transformação de dados (Tabela 3). Constatou-se que os 13 coeficientes de correlação de cada variável explicativa com a produtividade de grãos, obtidos antes e após a transformação de dados, foram similares ( $0,96 \leq r \leq 1,00$ ). Isso é indicativo de que a transformação de dados não alterou a intensidade e a direção das associações lineares das variáveis explicativas com a produtividade de grãos.

Nos 13 ensaios, o diagnóstico de multicolinearidade com base no fator de inflação de variância (FIV), realizado na matriz de correlação  $X'X$  com as sete variáveis explicativas, indicou que a variável FM não apresentou elevada correlação com as demais variáveis explicativas ( $FIV < 10$ ) (Tabela 4). As demais variáveis explicativas (EP, AIE, PRE, NP, NE e PLI) apresentaram elevado grau de multicolinearidade, em todos os ensaios ( $37,1 \leq FIV \leq 16.385,7$ ). Nessas seis variáveis explicativas, altamente correlacionadas, verificaram-se reduções expressivas do FIV após a transformação dos dados. Assim, a transformação contribuiu para a redução do grau de multicolinearidade na matriz de correlação  $X'X$ . Segundo Hair et al. (2009), a transformação de dados, além de contribuir para corrigir violações da normalidade, pode contribuir para a melhoria das relações entre variáveis. Isso pode favorecer a redução do grau de multicolinearidade

entre as variáveis explicativas que compõem a matriz  $X'X$ . Após a adição da constante  $k = 0,10$ , na diagonal da matriz de correlação  $X'X$ , todas as variáveis, nos 13 ensaios, apresentaram reduzido grau de multicolinearidade ( $FIV < 10$ ). A eliminação das variáveis AIE e NE também foi eficiente na redução do FIV, que apresentou valores similares, antes e após a transformação dos dados.

Com base no número de condição (NC), houve multicolinearidade severa, nos 13 ensaios, quando foi utilizada a matriz de correlação  $X'X$  com as sete variáveis explicativas ( $2.123 \leq NC \leq 86.390$ ) (Tabela 5). Após a transformação dos dados, o número de condição diminuiu na maioria dos ensaios ( $308 \leq NC \leq 15.258$ ). Esses resultados sugerem que a transformação de dados melhora o condicionamento da matriz de correlação  $X'X$  e pode reduzir os efeitos adversos do alto grau de multicolinearidade. No entanto, essa redução não foi suficiente para atingir níveis satisfatórios de multicolinearidade ( $NC \leq 100$ ). Tanto a adição da constante  $k = 0,10$  à diagonal da matriz de correlação  $X'X$  ( $23 \leq NC \leq 39$ ), quanto à eliminação de variáveis ( $3 \leq NC \leq 25$ ) foram eficientes para a redução do grau de multicolinearidade na matriz de correlação  $X'X$ , conforme também verificado por Toebe & Cargnelutti Filho (2013).

Em geral, a utilização da análise de trilha tradicional, após a eliminação das variáveis AIE e NE, resultou na redução do grau de multicolinearidade (menores valores de FIV e NC) e em maior poder preditivo da análise de trilha (maior coeficiente de determinação e menor efeito da variável residual), em comparação à análise de trilha sob multicolinearidade, adicionada de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  (Tabelas 4 e 5). Assim, sugere-se, o uso preferencial da análise de trilha tradicional, após a eliminação de variáveis correlacionadas, o que pode em alguns casos representar economia de trabalho, em razão da necessidade de mensuração de um conjunto menor de variáveis explicativas. No presente estudo, isso não ocorreu, uma vez que as variáveis AIE e NE são requeridas para a obtenção das variáveis PRE e PLI.

A variável FM apresentou efeito direto médio negativo e de baixa magnitude sobre a produtividade de grãos, nas análises de trilha tradicional, na análise em crista e na análise tradicional com eliminação de variáveis (Tabela 6), em conformidade com os valores de correlação entre essas variáveis (Tabela 3). Os 13 valores de efeitos diretos de FM sobre a produtividade

de grãos estimados antes e após a transformação de dados foram similares entre si ( $r = 0,94$ ), o que indica reduzido efeito da transformação de dados na estimação dos efeitos diretos de FM sobre a variável principal (Tabela 6). As demais seis variáveis explicativas apresentaram elevado grau de multicolinearidade (Tabelas 4 e 5); os efeitos diretos dessas variáveis sobre a produtividade de grãos, na análise de trilha tradicional

com alto grau de multicolinearidade, apresentaram ampla variação ( $-23,20 \leq \text{efeitos diretos} \leq 29,21$ ) (Tabela 6). As variáveis NE, PLI e AIE foram as variáveis que apresentaram maiores amplitudes de efeitos diretos sobre a produtividade de grãos. Essas variáveis também apresentaram maiores escores de FIV (Tabela 4), que são indicativos de que a maior amplitude dos efeitos diretos estimados está relacionada

**Tabela 3.** Coeficientes de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos, em 13 ensaios de competição de cultivares de milho, e coeficientes de correlação linear de Pearson entre os 13 coeficientes obtidos antes e os 13 coeficientes obtidos após a transformação de variáveis ( $r$  entre grupos).

Ensaio <sup>(1)</sup>	Variável <sup>(2)</sup>						
	FM	EP	AIE	PRE	NP	NE	PLI
Antes da transformação de variáveis							
1	0,27 <sup>ns</sup>	0,28 <sup>ns</sup>	0,29 <sup>ns</sup>	0,17 <sup>ns</sup>	0,61*	0,75*	0,53*
2	-0,15 <sup>ns</sup>	0,56*	0,54*	0,38*	0,44*	0,80*	0,79*
3	-0,12 <sup>ns</sup>	0,08 <sup>ns</sup>	0,15 <sup>ns</sup>	0,10 <sup>ns</sup>	0,72*	0,82*	0,63*
4	-0,58*	0,07 <sup>ns</sup>	-0,03 <sup>ns</sup>	-0,10 <sup>ns</sup>	0,16 <sup>ns</sup>	0,78*	0,70*
5	-0,37*	0,14 <sup>ns</sup>	0,14 <sup>ns</sup>	0,08 <sup>ns</sup>	0,67*	0,67*	0,26 <sup>ns</sup>
6	-0,27 <sup>ns</sup>	0,04 <sup>ns</sup>	0,26 <sup>ns</sup>	0,30 <sup>ns</sup>	0,13 <sup>ns</sup>	0,51*	0,46 <sup>ns</sup>
7	-0,17 <sup>ns</sup>	0,03 <sup>ns</sup>	-0,04 <sup>ns</sup>	-0,06 <sup>ns</sup>	0,80*	0,92*	0,94*
8	-0,29 <sup>ns</sup>	0,28 <sup>ns</sup>	0,29 <sup>ns</sup>	0,13 <sup>ns</sup>	0,60*	0,75*	0,75*
9	-0,33 <sup>ns</sup>	0,16 <sup>ns</sup>	0,16 <sup>ns</sup>	0,07 <sup>ns</sup>	0,51*	0,68*	0,32 <sup>ns</sup>
10	-0,56*	-0,11 <sup>ns</sup>	0,07 <sup>ns</sup>	0,21 <sup>ns</sup>	0,62*	0,86*	0,85*
11	-0,48*	-0,03 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,32 <sup>ns</sup>	0,56*	0,84*	0,46*
12	-0,32 <sup>ns</sup>	0,47 <sup>ns</sup>	-0,41 <sup>ns</sup>	-0,59*	0,35 <sup>ns</sup>	0,82*	0,50 <sup>ns</sup>
13	-0,27 <sup>ns</sup>	0,05 <sup>ns</sup>	-0,16 <sup>ns</sup>	-0,38*	0,21 <sup>ns</sup>	0,49*	0,36 <sup>ns</sup>
Média	-0,28	0,16	0,07	0,00	0,49	0,75	0,58
Após a transformação das variáveis							
1	0,22 <sup>ns</sup>	0,25 <sup>ns</sup>	0,22 <sup>ns</sup>	0,10 <sup>ns</sup>	0,61*	0,78*	0,54*
2	-0,16 <sup>ns</sup>	0,52*	0,48*	0,38*	0,44*	0,80*	0,78*
3	-0,12 <sup>ns</sup>	0,08 <sup>ns</sup>	0,15 <sup>ns</sup>	0,10 <sup>ns</sup>	0,66*	0,76*	0,52*
4	-0,58*	0,07 <sup>ns</sup>	-0,03 <sup>ns</sup>	-0,10 <sup>ns</sup>	0,22 <sup>ns</sup>	0,74*	0,70*
5	-0,37*	0,14 <sup>ns</sup>	0,14 <sup>ns</sup>	0,08 <sup>ns</sup>	0,61*	0,67*	0,26 <sup>ns</sup>
6	-0,27 <sup>ns</sup>	0,05 <sup>ns</sup>	0,26 <sup>ns</sup>	0,30 <sup>ns</sup>	0,13 <sup>ns</sup>	0,51*	0,53*
7	-0,38 <sup>ns</sup>	0,12 <sup>ns</sup>	0,00 <sup>ns</sup>	-0,07 <sup>ns</sup>	0,87*	0,93*	0,89*
8	-0,29 <sup>ns</sup>	0,28 <sup>ns</sup>	0,29 <sup>ns</sup>	0,13 <sup>ns</sup>	0,62*	0,77*	0,77*
9	-0,33 <sup>ns</sup>	0,18 <sup>ns</sup>	0,16 <sup>ns</sup>	0,07 <sup>ns</sup>	0,48*	0,68*	0,32 <sup>ns</sup>
10	-0,56*	-0,11 <sup>ns</sup>	0,07 <sup>ns</sup>	0,21 <sup>ns</sup>	0,62*	0,87*	0,86*
11	-0,48*	-0,03 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,32 <sup>ns</sup>	0,56*	0,84*	0,46*
12	-0,32 <sup>ns</sup>	0,47 <sup>ns</sup>	-0,41 <sup>ns</sup>	-0,59*	0,35 <sup>ns</sup>	0,81*	0,50 <sup>ns</sup>
13	-0,27 <sup>ns</sup>	0,05 <sup>ns</sup>	-0,16 <sup>ns</sup>	-0,38*	0,21 <sup>ns</sup>	0,52*	0,35 <sup>ns</sup>
Média	-0,30	0,16	0,07	-0,01	0,49	0,74	0,58
r entre grupos <sup>(3)</sup>	0,96*	0,99*	1,00*	1,00*	0,98*	0,98*	0,98*

<sup>(1)</sup>Ensaio definido na Tabela 1. <sup>(2)</sup>FM, florescimento masculino; EP, estatura de plantas; AIE, altura de inserção da espiga; PRE, posição relativa da espiga; NP, número de plantas; NE, número de espigas; PLI, prolificidade. <sup>(3)</sup>Representa o coeficiente de correlação entre os 13 valores estimados antes e os 13 valores estimados após a transformação de dados, para cada variável, com 11 graus de liberdade. <sup>ns</sup>Não significativo. \*Significativo a 5% de probabilidade, pelo teste t, com n-2 graus de liberdade, em que n = número de cultivares do ensaio.

à presença de alto grau de multicolinearidade e que os efeitos diretos, em módulo, tendem a ser elevados na presença desta. Portanto, é provável que os elevados efeitos diretos e indiretos da análise de trilha em milho,

**Tabela 4.** Diagnóstico de multicolinearidade para as variáveis explicativas, com base no fator de inflação da variância, em 13 ensaios de competição de cultivares de milho com distribuição multivariada não normal (NN) e distribuição multivariada normal (NO), após a transformação de variáveis.

Ensaio <sup>(1)</sup>	FM		EP		AIE		PRE		NP		NE		PLI	
	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO
Sete variáveis explicativas														
1	1,5	1,3	173,4	136,4	423,9	325,2	202,3	154,6	154,4	27,7	284,1	46,5	102,2	12,7
2	1,6	2,0	159,3	8,8	442,2	24,3	170,6	15,0	144,0	32,2	380,1	95,9	181,9	42,3
3	1,9	1,7	326,3	297,2	919,8	849,0	803,5	744,0	67,4	15,1	171,3	28,5	59,2	9,5
4	1,9	1,9	244,1	262,8	717,4	782,3	236,9	257,1	67,0	5,2	155,0	10,5	149,9	9,2
5	2,2	2,2	790,7	256,1	1.762,4	561,4	504,9	161,6	453,2	17,7	774,3	32,3	282,1	13,9
6	2,0	3,6	2.203,7	52,1	4.313,5	89,4	2.413,2	47,9	3.434,5	10,1	16.385,7	46,1	15.145,6	45,2
7	2,3	2,7	263,2	4,4	1.452,6	21,7	1.285,9	19,1	613,4	19,3	2.650,4	60,6	873,5	17,4
8	2,5	2,6	323,7	337,5	620,5	648,4	414,9	433,6	497,6	29,0	11.006,8	395,4	7.338,4	263,9
9	1,2	1,2	866,7	60,6	1.374,3	108,1	337,9	33,1	427,3	26,3	392,3	25,1	82,2	7,3
10	2,4	2,3	359,6	359,6	300,6	300,5	293,7	293,9	160,1	30,2	2.847,7	461,9	2.058,2	332,9
11	2,0	2,0	614,1	625,1	1.158,4	1.178,4	834,3	849,3	72,6	72,4	79,3	79,5	37,1	37,1
12	4,2	4,3	531,7	586,1	1.535,2	1.730,3	2.098,6	2.329,0	255,4	19,2	255,1	19,1	203,6	16,7
13	1,4	1,5	1.740,5	1.518,4	2.228,8	1.955,8	725,7	636,2	211,4	22,2	1.097,3	65,6	1.370,5	93,5
Média	2,1	2,2	661,3	346,5	1.326,9	659,6	794,0	459,6	504,5	25,1	2.806,1	105,2	2.145,0	69,4
Sete variáveis explicativas, adicionadas de valor k = 0,10 à diagonal da matriz X'X														
1	1,3	1,3	3,5	3,5	6,6	6,6	3,9	3,8	4,1	4,2	6,5	5,9	3,1	2,5
2	1,5	1,5	3,5	2,8	7,3	5,7	4,0	4,0	3,5	3,3	6,8	6,7	3,9	3,7
3	1,6	1,6	2,9	2,8	5,8	5,8	5,2	5,2	3,9	3,8	7,0	5,9	3,5	2,7
4	1,8	1,8	3,7	3,7	7,5	7,5	4,0	4,0	3,0	2,5	5,4	3,9	5,3	3,7
5	2,0	1,9	4,1	4,1	7,3	7,2	3,6	3,6	4,4	3,6	6,5	5,6	3,1	3,0
6	1,6	1,8	3,8	3,9	6,2	5,8	4,0	3,7	2,6	2,4	6,0	5,4	5,7	5,4
7	1,4	1,6	2,2	1,9	6,2	5,0	5,6	4,5	4,2	4,0	7,9	7,2	4,5	3,9
8	2,2	2,1	3,7	3,7	5,9	5,9	4,3	4,3	3,9	4,2	7,7	7,6	6,5	6,4
9	1,2	1,2	4,7	4,2	6,7	6,5	2,6	2,8	6,0	5,0	5,6	4,8	2,1	2,1
10	1,9	1,8	5,0	5,0	4,4	4,4	4,3	4,3	2,4	2,5	7,2	7,1	5,8	5,7
11	1,6	1,6	3,5	3,5	5,8	5,8	4,5	4,5	5,0	5,0	5,4	5,4	3,0	3,0
12	2,9	2,9	2,7	2,7	5,5	5,5	6,5	6,5	4,8	3,9	4,9	3,8	4,5	4,2
13	1,2	1,2	5,0	5,0	6,1	6,1	2,9	2,9	2,1	2,5	5,4	4,7	6,5	6,2
Média	1,7	1,7	3,7	3,6	6,2	6,0	4,3	4,2	3,8	3,6	6,3	5,7	4,4	4,0
Cinco variáveis explicativas, excluídas AIE e NE														
1	1,3	1,3	1,1	1,1	-	-	1,2	1,2	1,0	1,0	-	-	1,1	1,2
2	1,5	1,4	1,4	1,3	-	-	2,1	2,1	1,6	1,6	-	-	1,3	1,2
3	1,5	1,5	1,5	1,3	-	-	1,1	1,1	1,6	1,5	-	-	1,5	1,2
4	1,8	1,8	1,6	1,6	-	-	2,3	2,3	1,4	1,3	-	-	1,5	1,4
5	2,1	2,0	1,3	1,3	-	-	2,3	2,1	1,4	1,3	-	-	1,1	1,1
6	1,6	1,8	1,1	1,3	-	-	1,1	1,1	1,6	1,6	-	-	1,1	1,2
7	1,3	1,3	1,3	1,3	-	-	1,0	1,0	3,0	2,9	-	-	2,8	2,6
8	2,4	2,3	1,4	1,5	-	-	1,5	1,4	4,8	5,2	-	-	5,4	5,7
9	1,1	1,1	1,1	1,1	-	-	1,1	1,1	1,2	1,1	-	-	1,2	1,2
10	1,9	1,8	1,7	1,7	-	-	1,7	1,6	2,1	2,1	-	-	2,1	2,0
11	1,5	1,6	1,1	1,1	-	-	1,2	1,2	1,6	1,6	-	-	1,2	1,2
12	3,5	3,5	2,5	2,5	-	-	3,2	3,2	1,4	1,4	-	-	2,7	2,7
13	1,1	1,1	1,1	1,1	-	-	1,3	1,3	1,5	1,7	-	-	1,3	1,5
Média	1,8	1,7	1,4	1,4	-	-	1,6	1,6	1,9	1,9	-	-	1,9	1,9

<sup>(1)</sup>Ensaio definidos na Tabela 1. FM, florescimento masculino; EP, estatura de plantas; AIE, altura de inserção da espiga; PRE, posição relativa da espiga; NP, número de plantas; NE, número de espigas; PLI, prolificidade.



obtidos por Ahmad & Saleem (2003) e Selvaraj & Nagarajan (2011), estejam associados ao elevado grau de multicolinearidade entre as variáveis explicativas.

Elevados efeitos diretos e indiretos, em análises de trilha realizada com alto grau de multicolinearidade, também foram observados nas culturas de pimentão

**Tabela 5.** Diagnóstico de multicolinearidade, com base no número de condição, e coeficiente de determinação e efeito da variável residual da análise de trilha, em 13 ensaios de competição de cultivares de milho, com distribuição multivariada não normal (NN) e distribuição multivariada normal (NO), após a transformação de variáveis.

Ensaio <sup>(1)</sup>	Número de condição		Coeficiente de determinação		Efeito da variável residual	
	NN	NO	NN	NO	NN	NO
Para as sete variáveis explicativas						
1	2.123	1.647	0,66	0,66	0,58	0,58
2	4.344	614	0,82	0,82	0,42	0,43
3	5.737	5.203	0,86	0,84	0,38	0,40
4	3.757	4.014	0,80	0,79	0,45	0,46
5	8.795	2.720	0,62	0,56	0,62	0,66
6	86.390	438	0,47	0,44	0,73	0,75
7	13.147	308	0,91	0,90	0,29	0,31
8	65.311	4.939	0,85	0,84	0,39	0,40
9	6.055	466	0,63	0,71	0,60	0,54
10	15.928	2.961	0,80	0,82	0,45	0,42
11	7.278	7.426	0,73	0,74	0,52	0,51
12	14.172	15.258	0,98	0,98	0,14	0,16
13	14.075	10.722	0,48	0,48	0,72	0,72
Média	19.009	4.363	0,74	0,74	0,48	0,49
Sete variáveis explicativas, adicionadas de valor k = 0,10 à diagonal da matriz X'X						
1	26	27	0,60	0,61	0,63	0,63
2	39	35	0,76	0,75	0,49	0,50
3	29	28	0,76	0,70	0,49	0,55
4	32	32	0,72	0,72	0,53	0,53
5	29	28	0,53	0,51	0,69	0,70
6	23	23	0,37	0,38	0,79	0,78
7	31	29	0,87	0,87	0,36	0,36
8	36	35	0,68	0,71	0,56	0,54
9	24	23	0,55	0,53	0,67	0,68
10	32	32	0,74	0,76	0,51	0,49
11	29	29	0,69	0,69	0,55	0,56
12	34	34	0,89	0,88	0,32	0,35
13	27	27	0,42	0,45	0,76	0,74
Média	30	29	0,66	0,66	0,57	0,57
Cinco variáveis explicativas, excluídas AIE e NE						
1	3	3	0,63	0,63	0,61	0,61
2	8	7	0,79	0,79	0,46	0,45
3	6	5	0,80	0,75	0,45	0,50
4	8	8	0,77	0,77	0,48	0,48
5	8	7	0,57	0,52	0,66	0,69
6	5	6	0,40	0,42	0,78	0,76
7	11	10	0,91	0,90	0,29	0,32
8	24	25	0,73	0,76	0,52	0,49
9	3	3	0,57	0,53	0,65	0,69
10	9	9	0,77	0,80	0,48	0,45
11	5	5	0,73	0,73	0,52	0,52
12	18	18	0,97	0,97	0,17	0,17
13	4	5	0,46	0,48	0,74	0,72
Média	9	8	0,70	0,70	0,52	0,53

<sup>(1)</sup>Ensaio definidos na Tabela 1.



(Carvalho et al., 1999), milho (Carvalho et al., 2001; Toebe & Cargnelutti Filho, 2013), soja (Bizeti et al., 2004) e canola (Coimbra et al., 2005). Portanto, deve ser evitada a interpretação dos efeitos diretos e indiretos estimados em análises de trilha com alto grau de multicolinearidade, uma vez que eles não expressam valores reais de relações de causa e efeito entre variáveis.

Quando os dados foram transformados, verificou-se que o FIV diminuiu nas seis variáveis explicativas altamente correlacionadas (EP, AIE, PRE, NP, NE, PLI) (Tabela 4). Nessas mesmas variáveis, verificou-se que houve redução da amplitude dos efeitos diretos obtidos na análise de trilha tradicional com alto grau de multicolinearidade ( $-8,16 \leq \text{efeitos diretos} \leq 8,94$ ) (Tabela 6). No entanto, essa redução não foi suficiente para estabilizar os coeficientes estimados, e a utilização de outros métodos de análise de trilha foi necessária. Além disso, nas seis variáveis explicativas com alto grau de multicolinearidade, não houve correlação significativa entre os efeitos diretos estimados antes e após a transformação de dados, o que indica que, nessa condição, a transformação de dados altera a estimação dos efeitos diretos. A amplitude dos efeitos diretos, estimados antes e após a transformação de dados, indica que os mesmos efeitos não apresentam sentido biológico e não devem ser considerados na identificação das relações de causa e efeito.

Nos 13 ensaios (antes e após a transformação de dados), observou-se a estabilização dos efeitos diretos e indiretos após a adição de uma constante  $k = 0,10$  (Tabela 6), cujo valor foi atribuído a todas as análises de trilha em condições de multicolinearidade (análise de trilha em crista). Essa técnica de análise de trilha também foi eficiente na estabilização dos efeitos diretos estimados nas culturas de pimentão (Carvalho et al., 1999), milho (Carvalho et al., 2001; Toebe & Cargnelutti Filho, 2013), soja (Bizeti et al., 2004) e canola (Coimbra et al., 2005). Os maiores coeficientes de correlação de NE, PLI e NP com a produtividade de grãos (Tabela 3) resultaram nos maiores efeitos diretos sobre essa variável (Tabela 6), o que confirma relações de causa e efeito (Cruz & Regazzi, 1997). Verificou-se também que os efeitos diretos, estimados nos 13 ensaios antes e após a transformação de dados, foram positivamente correlacionados ( $r \geq 0,55$ ), o que indica que quando a multicolinearidade é contornada, a transformação de dados não altera de modo expressivo os escores de efeitos diretos.

Na análise de trilha tradicional, com eliminação das variáveis AIE e NE, também ocorreu a estabilização dos efeitos diretos (Tabela 6). Essa técnica de análise de trilha resultou em menores escores de FIV e NC, maiores coeficientes de determinação e menores efeitos da variável residual da análise de trilha, em comparação à análise de trilha em crista. Em 14 ensaios de milho, nos quais se verificou o atendimento da pressuposição de normalidade multivariada, Toebe & Cargnelutti Filho (2013) também concluíram que esse método de análise de trilha foi o mais adequado para a estimação dos efeitos diretos. Esse método de análise também foi eficiente na redução da multicolinearidade e na estabilização dos efeitos diretos e indiretos em pimentão (Carvalho et al., 1999) e em soja (Bizeti et al., 2004). Nos 13 ensaios, as variáveis PLI e NP apresentaram os maiores efeitos diretos sobre a produtividade de grãos, com provável relação de causa e efeito. Ainda, verificou-se alta correlação entre os 13 efeitos diretos estimados antes e os 13 efeitos diretos estimados após a transformação dos dados, para cada variável ( $r \geq 0,90$ ), o que sugere pequena interferência da não normalidade multivariada, quando os efeitos adversos de multicolinearidade são contornados. Os efeitos indiretos foram calculados para os três métodos de análise de trilha, mas não foram apresentados.

## Conclusões

1. A transformação de dados reduz o grau de multicolinearidade e a variabilidade das estimativas dos efeitos diretos, na análise de trilha tradicional com alto grau de multicolinearidade.
2. A multicolinearidade exerce maior impacto sobre a estimativa dos efeitos diretos nas análises de trilha do que a não normalidade multivariada.
3. A análise de trilha tradicional com eliminação de variáveis é mais adequada que a análise de trilha em crista.

## Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), por bolsas concedidas; ao Engenheiro Agrônomo José Paulo Guadagnin, da Fundação Estadual de Pesquisa Agropecuária, pela coordenação dos ensaios e cessão dos dados de avaliação de cultivares de milho, no Estado do Rio Grande do Sul.

## Referências

- AHMAD, A.; SALEEM, M. Path coefficient analysis in *Zea mays* L. **International Journal of Agriculture and Biology**, v.5, p.245-248, 2003.
- ALVI, M.B.; RAFIQUE, M.; TARIQ, M.S.; HUSSAIN, A.; MAHMOOD, T.; SARWAR, M. Character association and path coefficient analysis of grain yield and yield components maize (*Zea mays* L.). **Pakistan Journal of Biological Sciences**, v.6, p.136-138, 2003. DOI: 10.3923/pjbs.2003.136.138.
- BELLO, O.B.; ABDULMALIQ, S.Y.; AFOLABI, M.S.; IGE, S.A. Correlation and path coefficient analysis of yield and agronomic characters among open pollinated maize varieties and their  $F_1$  hybrids in a diallel cross. **African Journal of Biotechnology**, v.9, p.2633-2639, 2010.
- BIZETI, H.S.; CARVALHO, C.G.P. de; SOUZA, J.R.P. de; DESTRO, D. Path analysis under multicollinearity in soybean. **Brazilian Archives of Biology and Technology**, v.47, p.669-676, 2004. DOI: 10.1590/S1516-89132004000500001.
- BOX, G.E.P.; COX, D.R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B - Statistical Methodological**, v.26, p.211-252, 1964.
- CARVALHO, C.G.P. de; BORSATO, R.; CRUZ, C.D.; VIANA, J.M.S. Path analysis under multicollinearity in  $S_0 \times S_0$  maize hybrids. **Crop Breeding and Applied Biotechnology**, v.1, p.263-270, 2001.
- CARVALHO, C.G.P. de; OLIVEIRA, V.R.; CRUZ, C.D.; CASALI, V.W.D. Análise de trilha sob multicolinearidade em pimentão. **Pesquisa Agropecuária Brasileira**, v.34, p.603-613, 1999. DOI: 10.1590/S0100-204X1999000400011.
- CARVALHO, S.P. de; CRUZ, C.D. Diagnosis of multicollinearity: assessment of the condition of correlation matrices used in genetic studies. **Brazilian Journal of Genetics**, v.19, p.479-484, 1996.
- COIMBRA, J.L.M.; BENIN, G.; VIEIRA, E.A.; OLIVEIRA, A.C. de; CARVALHO, F.I.F.; GUIDOLIN, A.F.; SOARES, A.P. Consequências da multicolinearidade sobre a análise de trilha em canola. **Ciência Rural**, v.35, p.347-352, 2005. DOI: 10.1590/S0103-84782005000200015.
- COUTO, M.R.M.; LÚCIO, A.D.; LOPES, S.J.; CARPES, R.H. Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, v.39, p.1701-1707, 2009. DOI: 10.1590/S0103-84782009005000110.
- CRUZ, C.D. **Programa Genes: estatística experimental e matrizes**. Viçosa: UFV, 2006. 285p.
- CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: Ed. UFV, 2006. v.2, 585p.
- CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: Ed. UFV, 1997. 390p.
- FÁVERO, L.P.; BELFIORE, P.; SILVA, F.L. da; CHAN, B.L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009. 646p.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: UFLA, 2008. 662p.
- HAIR, J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Análise multivariada de dados**. 6.ed. Porto Alegre: Bookman, 2009. 688p.
- KUMAR, T.S.; REDDY, D.M.; REDDY, K.H.; SUDHAKAR, P. Targeting of traits through assessment of interrelationship and path analysis between yield and yield components for grain yield improvement in single cross hybrids of maize (*Zea mays* L.). **International Journal of Applied Biology and Pharmaceutical Technology**, v.2, p.123-129, 2011.
- MOHAMMADI, S.A.; PRASANNA, B.M.; SINGH, N.N. Sequential path model for determining interrelationships among grain yield and related characters in maize. **Crop Science**, v.43, p.1690-1697, 2003. DOI: 10.2135/cropsci2003.1690.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York: J. Wiley, 1982. 504p.
- PAVAN, R.; LOHITHASWA, H.C.; WALL, M.C.; GANGASHETTY, P.; SHEKARA, B.G. Correlation and path coefficient analysis of grain yield and yield contributing traits in single cross hybrids of maize (*Zea mays* L.). **Electronic Journal of Plant Breeding**, v.2, p.253-257, 2011.
- ROYSTON, J.B. Some techniques for assessing multivariate normality based on the Shapiro-Wilk  $W$ . **Applied Statistics**, v.32, p.121-133, 1983. DOI: 10.2307/2347291.
- SAIDAIAH, P.; SATYANARAYANA, E.; KUMAR, S.S. Association and path coefficient analysis in maize (*Zea mays* L.). **Agricultural Science Digest**, v.28, p.79-83, 2008.
- SELVARAJ, C.I.; NAGARAJAN, P. Interrelationship and path-coefficient studies for qualitative traits, grain yield and other yield attributes among maize (*Zea mays* L.). **International Journal of Plant Breeding and Genetics**, v.5, p.209-223, 2011. DOI: 10.3923/ijpb.2011.209.223.
- SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, v.52, p.591-611, 1965.
- STATSOFT. **Statistica**. Version 7.0. Tucksas: StatSoft, 2005.
- TOEBE, M.; CARGNELUTTI FILHO, A. Multicollinearity in path analysis of maize (*Zea mays* L.). **Journal of Cereal Science**, v.57, p.453-462, 2013. DOI: 10.1016/j.jcs.2013.01.014.

Recebido em 27 de novembro de 2012 e aprovado em 17 de abril de 2013