

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE CIÊNCIAS RURAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**

**NÃO-NORMALIDADE MULTIVARIADA E  
MULTICOLINEARIDADE EM ANÁLISE DE TRILHA  
NA CULTURA DE MILHO**

**DISSERTAÇÃO DE MESTRADO**

**Marcos Toebe**

**Santa Maria, RS, Brasil  
2012**

**NÃO-NORMALIDADE MULTIVARIADA E  
MULTICOLINEARIDADE EM ANÁLISE DE TRILHA NA  
CULTURA DE MILHO**

**por**

**Marcos Toebe**

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Agronomia, Área de concentração em Produção Vegetal, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Agronomia**

**Orientador: Prof. Alberto Cargnelutti Filho**

**Santa Maria, RS, Brasil  
2012**

T641n Toebe, Marcos

Não-normalidade multivariada e multicolinearidade em análise de trilha na cultura de milho / por Marcos Toebe. – 2012.

107 p. ; il. ; 30 cm

Orientador: Alberto Cargnelutti Filho

Dissertação (mestrado) – Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós-Graduação em Agronomia, RS, 2012

1. *Zea mays* L. 2. Pressupostos 3. Transformações Box-Cox 4. Análise de trilha sob multicolinearidade 5. Eliminação de variáveis I. Cargnelutti Filho, Alberto II. Título

CDU 633.15

Ficha catalográfica elaborada por Cláudia Terezinha Branco Gallotti – CRB 10/1109  
Biblioteca Central UFSM

---

© 2012

Todos os direitos autorais reservados a Marcos Toebe. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Rua Elpídio Menezes, n. 70, Bairro: Camobi, CEP: 97105-110

Endereço eletrônico: m.toebe@gmail.com

---

**Universidade Federal de Santa Maria  
Centro de Ciências Rurais  
Programa de Pós-Graduação em Agronomia**

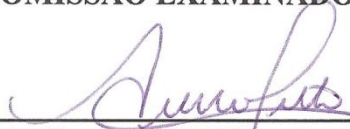
A Comissão Examinadora, abaixo assinada,  
Aprova a Dissertação de Mestrado

**NÃO-NORMALIDADE MULTIVARIADA E MULTICOLINEARIDADE  
EM ANÁLISE DE TRILHA NA CULTURA DE MILHO**

elaborada por  
**Marcos Toebe**

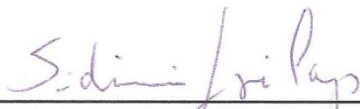
como requisito parcial para obtenção do grau de  
**Mestre em Agronomia**

**COMISSÃO EXAMINADORA:**



---

**Alberto Cargnelutti Filho, Dr. (UFSM)  
(Presidente/Orientador)**



---

**Sidinei José Lopes, Dr. (UFSM)**



---

**Alfredo do Nascimento Junior, Dr. (EMBRAPA)**

**Santa Maria, 16 de fevereiro de 2012.**

## *Dedicatória*

*A Deus*

*Ao meu orientador, **Alberto***

*Aos meus pais, **Hildo e Ilvani***

*A minha esposa, **Íris***

***Dedico este trabalho!***

## **AGRADECIMENTOS**

A Deus, pelo maravilhoso dom da vida e pela infindável provisão em minha vida.

A Universidade Federal de Santa Maria e ao Programa de Pós-Graduação em Agronomia, pela oportunidade de realização do curso de mestrado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de mestrado.

A Fundação Estadual de Pesquisa Agropecuária (FEPAGRO) e demais pesquisadores que contribuíram para a obtenção do banco de dados utilizado nessa dissertação.

Aos meus pais Hildo e Ilvani Toebe, que sempre enfrentaram todas as dificuldades e deram o seu melhor para a concretização desse sonho, aos quais não tenho palavras para agradecer por tudo o que fizeram em minha vida.

A minha esposa Íris, pelo apoio, amor, carinho, compreensão e constante incentivo.

Ao meu orientador Alberto Cargnelutti Filho, pelo companheirismo e dedicação na orientação desse trabalho e pela preocupação no meu crescimento pessoal e profissional.

Aos meus irmãos Ademir e Simone, pelos exemplos de vida.

Ao professor Lindolfo Storck, pela amizade e pelos seus sábios conselhos.

Aos professores Sidinei José Lopes e Alessandro Dal'Col Lúcio, do setor de experimentação vegetal, pelos ensinamentos e pela amizade.

Aos demais professores do curso de Agronomia e de Pós-Graduação em Agronomia, pelos preciosos ensinamentos.

Aos meus amigos e colegas do setor de experimentação vegetal: Cláudia Burin, André Luis Fick, Bruna Mendonça Alves, Ismael Márcio Neu, Régés Bellé Stefanello, Gabriele Casarotto, Giovani Facco, Glauber Stürmer, Betânia Brum, Fernando Machado Haesbaert, Diogo Vanderlei Schwertner, Daniel Santos, Rélia Brunet e Tatiani Reis da Silveira.

Aos funcionários do departamento de fitotecnia, em especial ao João Colpo e ao secretário de pós-graduação Régis Fabiano Santos.

Aos meus colegas de graduação e de pós-graduação que me apoiaram nessa caminhada e a todos os amigos que torceram pelo meu sucesso.

A todos os demais familiares, professores, pesquisadores, colegas e amigos que me apoiaram incondicionalmente e que contribuíram, direta ou indiretamente, na realização desse trabalho.

Muito Obrigado!

## RESUMO

Dissertação de Mestrado  
Programa de Pós-Graduação em Agronomia  
Universidade Federal de Santa Maria

### NÃO-NORMALIDADE MULTIVARIADA E MULTICOLINEARIDADE EM ANÁLISE DE TRILHA NA CULTURA DE MILHO

AUTOR: MARCOS TOEBE

ORIENTADOR: ALBERTO CARGNELUTTI FILHO

Local e Data da Defesa: Santa Maria, 16 de fevereiro de 2012.

A análise de trilha permite avaliar os efeitos diretos e indiretos de variáveis explicativas sobre a variável de interesse, por meio do desdobramento dos coeficientes de correlação. Para que os resultados gerados pela análise de trilha apresentem confiabilidade adequada, alguns pressupostos devem ser atendidos. Assim, os objetivos deste trabalho foram: verificar a interferência da não-normalidade multivariada e da multicolinearidade em análise de trilha na cultura de milho e, comparar métodos alternativos de estimação dos coeficientes de trilha. Foram utilizados dados de 44 ensaios de competição de cultivares de milho, conduzidos no estado do Rio Grande do Sul, entre os anos agrícolas de 2002/03 e 2004/05. Em cada cultivar, de cada ensaio, foram mensuradas sete variáveis explicativas (número de dias até o florescimento masculino, estatura de plantas, altura de inserção da espiga, posição relativa da espiga, número de plantas, número de espigas e prolificidade) e a variável principal (produtividade de grãos). Para cada ensaio, foram calculadas estatísticas descritivas e realizado o diagnóstico de normalidade uni e multivariada, por meio dos testes de Shapiro-Wilk e de Shapiro-Wilk multivariado generalizado por Royston, respectivamente. A seguir, nos dados dos ensaios que não apresentaram distribuição normal, foi realizada a transformação dos dados com a utilização da família de transformações Box-Cox. Para os dados originais e os dados transformados, foram calculados os coeficientes de correlação entre as sete variáveis explicativas (matriz de correlação  $X'X$ ) e os coeficientes de correlação de cada variável explicativa com a produtividade de grãos (matriz de correlação  $X'Y$ ). A seguir, foi realizado o diagnóstico de multicolinearidade na matriz de correlação  $X'X$ , por meio de quatro métodos: fator de inflação de variância, tolerância, número de condição e determinante da matriz. Por fim, foi realizada a análise de trilha, com a utilização do sistema de equações normais  $X'X \hat{\beta} = X'Y$ , por três formas: análise de trilha tradicional, análise de trilha sob multicolinearidade e análise de trilha tradicional, com eliminação de variáveis. A transformação de dados, a fim de obter a normalidade multivariada, contribui para a redução do grau de multicolinearidade e na estabilização das estimativas dos efeitos diretos em análise de trilha com alto grau de multicolinearidade. Os efeitos adversos do alto grau de multicolinearidade na estimativa dos efeitos diretos de análises de trilha são maiores que a não-normalidade multivariada. A análise de trilha tradicional, com eliminação de variáveis, é mais adequada que a análise de trilha sob multicolinearidade.

**Palavras-chave:** *Zea mays* L. Pressupostos. Transformações Box-Cox. Análise de trilha sob multicolinearidade. Eliminação de variáveis.

## ABSTRACT

Master Dissertation  
Programa de Pós-Graduação em Agronomia  
Universidade Federal de Santa Maria

### NON-NORMALITY MULTIVARIATE AND MULTICOLLINEARITY IN PATH ANALYSIS IN CORN

AUTHOR: MARCOS TOEBE

ADVISER: ALBERTO CARGNELUTTI FILHO

Place and Date of the Defense: Santa Maria, 16<sup>th</sup> February, 2012.

The path analysis allows evaluation of the direct and indirect effects of the explicative variables on variable of interest, through the breakdown of the correlation coefficients. In order to make the results obtained through the path analysis reliable, some assumptions must be met. Thus, the objectives of this study were to verify the normality and the multicollinearity interference in the corn path analysis and compare alternative methods for estimating the path coefficients. Data from 44 trials of corn cultivars was used, carried out in the state of Rio Grande do Sul, between the crop years 2002/03 and 2004/05. In each cultivar, of each trial, were measured (number of days until the male flowering, plant height, ear insertion height, relative position of the ear, number of plants, number of ears and prolificacy) and the main variable (grain yield). For each trial, descriptive statistics were calculated and univariate and multivariate normality diagnoses were conducted using the Shapiro-Wilk test and the Shapiro-Wilk multivariate generalized by Royston test, respectively. Thereupon, in the trials data that did not present a normal distribution, a transformation of the data by the Box-Cox family of transformations was carried out. The correlation coefficients between the seven explicative variables (correlation matrix  $X'X$ ) and the correlation coefficients of each explicative variable with the grain yield (correlation matrix  $X'Y$ ) were calculated for the original and transformed data. Then, the multicollinearity was diagnosed in the correlation matrix  $X'X$ , using four methods: variance inflation factor, tolerance, the condition number and the matrix determinant. Finally, the path analysis was performed, using the normal equations system  $X'X \hat{\beta} = X'Y$ , in three forms: traditional path analysis, path analysis under multicollinearity and traditional path analysis, with elimination of variables. The data transformation, to obtain multivariate normality, contributes to the degree of multicollinearity decrease and in the stabilization of the direct effects in path analysis with high degree of multicollinearity. The high degrees of multicollinearity adverse effects in the estimation of the direct effects in path analysis are larger than the multivariate non-normality. The traditional path analysis, with elimination of variables, is more appropriate than the path analysis under multicollinearity.

**Key words:** *Zea mays* L. Assumptions. Box-Cox transformations. Path analysis under multicollinearity. Elimination of variables.



## LISTA DE TABELAS

- Tabela 1 - Relação dos 44 ensaios de competição de cultivares de milho, conduzidos em locais do estado do Rio Grande do Sul, com a respectiva classificação da categoria e do ciclo, ano agrícola, data de semeadura, data de colheita e número de cultivares avaliadas. ....32
- Tabela 2 - Equações para a estimação dos efeitos diretos das sete variáveis explicativas (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 e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional. ....46
- Tabela 3 - Equações para a estimação dos efeitos indiretos das sete variáveis explicativas (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 e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional. ....46
- Tabela 4 - Equações para a estimação dos efeitos diretos das cinco variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional, com eliminação de variáveis. ....51
- Tabela 5 - Equações para a estimação dos efeitos indiretos das cinco variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional, com eliminação de variáveis. ....51
- Tabela 6 - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: florescimento masculino (em dias) e estatura de plantas (em cm) mensuradas em 44 ensaios de competição de cultivares de milho. ....54
- Tabela 7 - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: altura de inserção da espiga (em cm) e posição relativa da espiga (índice), mensuradas em 44 ensaios de competição de cultivares de milho. ....55
- Tabela 8 - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: número de plantas ( $ha^{-1}$ ) e número de espigas ( $ha^{-1}$ ) mensuradas em 44 ensaios de competição de cultivares de milho. ....58
- Tabela 9 - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: prolificidade (número de espigas dividido por número de plantas) e produtividade de grãos (em  $kg\ ha^{-1}$ ), mensuradas em 44 ensaios de competição de cultivares de milho. ....59

Tabela 10 - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM) em 44 ensaios de competição de cultivares de milho. ....	62
Tabela 11 - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM), em 13 ensaios de competição de cultivares de milho. Sem (Sem T.) e com (Com T.) transformação de dados, de variáveis que não se ajustaram a distribuição normal, por meio da metodologia Box-Cox e o respectivo valor de $\lambda$ utilizado. ....	63
Tabela 12 - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM), em 17 ensaios de competição de cultivares de milho. Sem (Sem T.) e com (Com T.) transformação de dados, de variáveis que não se ajustaram a distribuição normal, por meio da metodologia Box-Cox e o respectivo valor de $\lambda$ utilizado. ....	64
Tabela 13 - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 44 ensaios de competição de cultivares de milho. ....	67
Tabela 14 - Fator de inflação da variância (FIV) para as variáveis explicativas em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada, sem a necessidade de transformação dos dados. ....	69
Tabela 15 - Tolerância para as variáveis explicativas, número de condição (NC) e determinante da matriz de correlação $X'X$ (DM), em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada sem a necessidade de transformação dos dados. ....	71
Tabela 16 - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos, coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual (RES) em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada, sem a necessidade de transformação dos dados. ....	75
Tabela 17 - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 13 ensaios de competição de cultivares de milho, que apresentaram distribuição normal multivariada após a transformação de variáveis, e valores de coeficiente de correlação linear de Pearson ajustado entre os valores antes e após a transformação de variáveis (r entre grupos). ....	77

Tabela 18 - Fator de inflação da variância (FIV) para variáveis explicativas em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados. ....	78
Tabela 19 - Tolerância para as variáveis explicativas em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados. ....	80
Tabela 20 - Número de condição e determinante da matriz de correlação $X'X$ , em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados. ....	81
Tabela 21 - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados. ....	83
Tabela 22 - Coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados. ....	85
Tabela 23 - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados, e valores de coeficiente de correlação linear de Pearson ajustado entre os valores antes e após a transformação de variáveis ( $r$ entre grupos). ....	87
Tabela 24 - Fator de inflação da variância (FIV) para variáveis explicativas em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados. ....	88
Tabela 25 - Tolerância para as variáveis explicativas em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados. ....	90
Tabela 26 - Número de condição e determinante da matriz de correlação $X'X$ , em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados. ....	91
Tabela 27 - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos em 17 ensaios de competição de cultivares de milho, sem distribuição normal multivariada antes e após a transformação de dados. ....	93
Tabela 28 - Coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de variáveis. ....	94

## LISTA DE FIGURAS

- Figura 1 - Ranking dos dez países com maior produção mundial de milho no ano de 2009. Adaptado de FAO, 2009. .... 15
- Figura 2 - Séries históricas da área semeada (em milhões de hectares), da produção (em milhões de toneladas) e da produtividade média (em toneladas por hectare) de milho no Brasil. Adaptado de CONAB, 2011. .... 16
- Figura 3 - Mapa do estado do Rio Grande do Sul, com os nove municípios onde foram realizados os 44 ensaios de competição de cultivares de milho utilizados nesse trabalho. .... 33
- Figura 4 - Diagrama causal com uma única cadeia, com os efeitos diretos das variáveis explicativas (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 e PLI: prolificidade) e da variável residual (Resíduo) sobre a produtividade de grãos em milho (PROD) e as inter-relações entre as variáveis explicativas na análise de trilha tradicional..... 43
- Figura 5 - Diagrama causal com uma única cadeia, com os efeitos diretos das variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) e da variável residual (Resíduo) sobre a produtividade de grãos em milho (PROD) e as inter-relações entre as variáveis explicativas na análise de trilha tradicional com eliminação de variáveis..... 49

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	<b>15</b>
<b>2.1 A cultura do milho e os ensaios de competição de cultivares</b> .....	<b>15</b>
<b>2.2 Normalidade dos dados</b> .....	<b>18</b>
2.2.1 Normalidade univariada .....	18
2.2.2 Normalidade multivariada .....	19
2.2.3 Transformação de variáveis.....	20
<b>2.3 Multicolinearidade</b> .....	<b>20</b>
2.3.1 Diagnóstico de multicolinearidade .....	21
2.3.2 Impacto da multicolinearidade em análises multivariadas .....	23
2.3.2 Ações corretivas para redução da multicolinearidade .....	24
<b>2.4 Análise de trilha</b> .....	<b>25</b>
2.4.1 Análise de trilha tradicional.....	26
2.4.2 Análise de trilha sob multicolinearidade .....	27
2.4.3 Análise de trilha tradicional com eliminação de variáveis .....	28
2.4.4 Utilização da matriz de correlação fenotípica em análise de trilha .....	29
<b>3 METODOLOGIA</b> .....	<b>31</b>
<b>3.1 Os ensaios</b> .....	<b>31</b>
<b>3.2 Estatísticas descritivas</b> .....	<b>33</b>
<b>3.3 Diagnóstico da normalidade</b> .....	<b>34</b>
3.3.1 Normalidade univariada .....	34
3.3.2 Normalidade multivariada .....	35
3.3.3 Transformação de dados.....	35
3.3.3.1 Critérios de transformação de dados .....	36
<b>3.4 Obtenção das matrizes de correlações fenotípicas</b> .....	<b>37</b>
<b>3.5 Diagnóstico de multicolinearidade</b> .....	<b>39</b>
3.5.1 Fator de inflação da variância.....	39
3.5.2 Tolerância .....	40
3.5.3 Número de condição.....	40
3.5.4 Determinante da matriz de correlação .....	41
<b>3.6 Análise de trilha</b> .....	<b>42</b>
3.6.1 Análise de trilha tradicional.....	43
3.6.2 Análise de trilha sob multicolinearidade .....	47
3.6.3 Análise de trilha tradicional com eliminação de variáveis .....	48
<b>4 RESULTADOS E DISCUSSÃO</b> .....	<b>53</b>
<b>4.1 Considerações finais</b> .....	<b>95</b>
<b>5 CONCLUSÕES</b> .....	<b>99</b>
<b>6 REFERÊNCIAS</b> .....	<b>101</b>

# 1 INTRODUÇÃO

O milho é um dos cereais mais cultivados no mundo, devido ao seu potencial produtivo e valor nutritivo, sendo utilizado na alimentação humana e animal, em complexos industriais e como fonte de energia, entre outras áreas. O conhecimento sobre a cultura avançou, significativamente, nas últimas décadas em relação às demais culturas (FANCELLI; DOURADO NETO, 2004), contribuindo para a elevação dos patamares de produtividade e do valor nutritivo de grãos. Nesse sentido, Duvic (2005) considera que, aproximadamente, 50% do aumento da produtividade de grãos estão vinculados ao progresso do melhoramento genético, através do desenvolvimento de cultivares de alto potencial produtivo. Assim, o melhoramento de plantas é a principal estratégia para o aumento da produtividade de grãos de maneira sustentável e equilibrada (BORÉM; MIRANDA, 2009).

O melhoramento de plantas pode ser realizado de modo direto, através da seleção de plantas com maiores potenciais produtivos ou nutritivos, ou de modo indireto, via estudo da relação existente entre caracteres. A seleção indireta é aconselhada, quando a variável principal é de difícil mensuração, apresenta baixa herdabilidade ou quando se deseja a seleção precoce de plantas. Para que a seleção indireta possa ser realizada, é necessária a existência de associação entre uma ou mais variáveis com a variável de interesse. Assim, estudos de relações de caracteres são imprescindíveis no melhoramento de plantas via seleção indireta.

Os coeficientes de correlação fenotípicos e genotípicos são importantes na mensuração da associação existente entre duas variáveis. No entanto, quando um grupo de variáveis atua sobre a variável principal, os valores de coeficientes de correlação obtidos podem não ser medidas reais da associação existente entre duas variáveis, devido à interferência de outras variáveis sobre a associação. Por isso, a utilização de técnicas complementares é aconselhada na verificação da associação real existente entre duas variáveis. Nesse sentido, a análise de trilha tem sido amplamente utilizada para essa finalidade, por permitir o desdobramento dos coeficientes de correlação em efeitos diretos e indiretos sobre a variável principal, em relações de causa e efeito (CRUZ; REGAZZI, 1997; CRUZ; CARNEIRO, 2006).

De maneira semelhante a outras técnicas multivariadas de análise de dados, a análise de trilha apresenta pressupostos que devem ser atendidos para que os resultados gerados apresentem confiabilidade e, expressem de modo real os efeitos diretos e indiretos de cada variável sobre a variável principal. Entre os pressupostos mais importantes da análise de

trilha, estão a suposição de normalidade multivariada e da ausência ou baixo grau de multicolinearidade entre as variáveis explicativas.

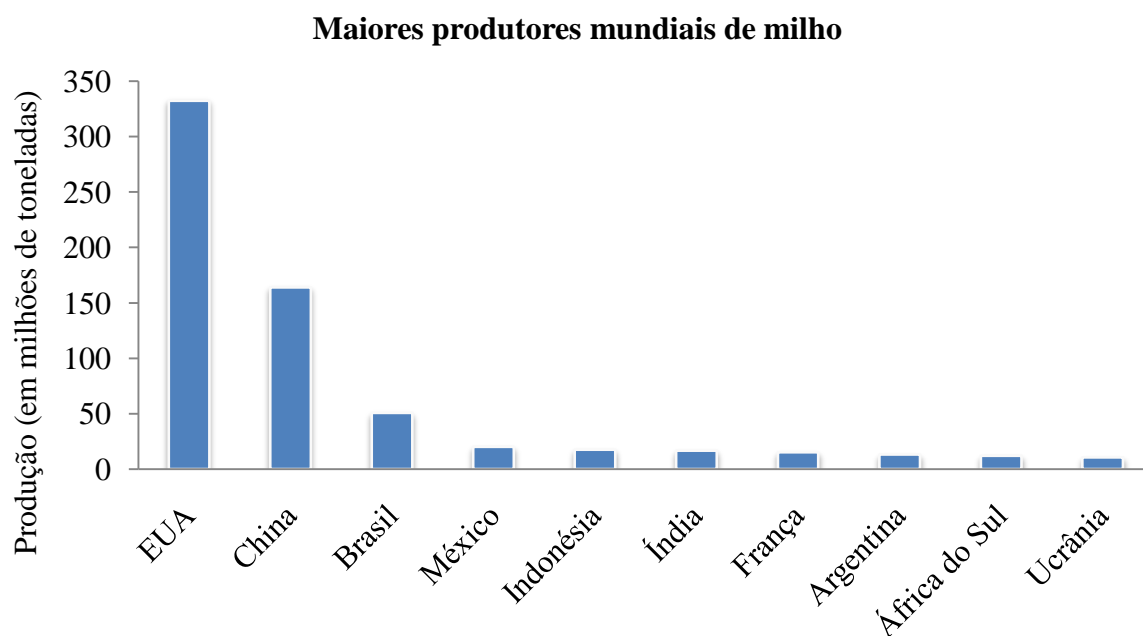
Os efeitos adversos da violação da pressuposição de normalidade multivariada em análise de trilha, ainda não foram reportados na literatura. Em relação aos efeitos nocivos da presença de elevado grau de multicolinearidade em análise de trilha, alguns estudos foram conduzidos em culturas agrícolas tais como: arroz, pimentão, canola, soja e milho. Nesse sentido, medidas corretivas, como a exclusão de variáveis explicativas altamente correlacionadas, ou alternativas, como a análise de trilha sob multicolinearidade, têm sido sugeridas. No entanto, a interferência da não-normalidade multivariada e da multicolinearidade na análise de trilha em ensaios de competição de cultivares de milho ainda não é conhecida.

Os objetivos deste trabalho foram verificar a interferência da não-normalidade multivariada e da multicolinearidade em análise de trilha na cultura de milho e, comparar métodos alternativos de estimação dos coeficientes de trilha.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A cultura do milho e os ensaios de competição de cultivares

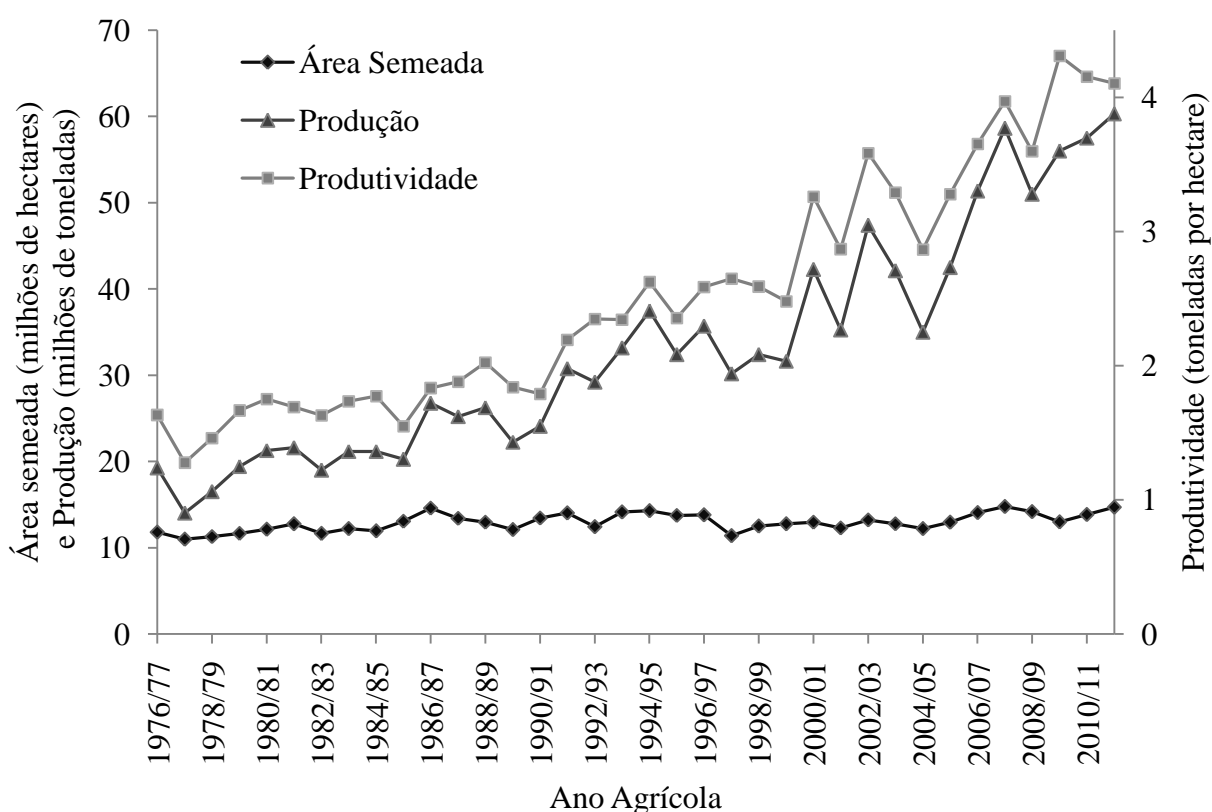
O milho, juntamente com o arroz e o trigo, são os cereais mais cultivados e consumidos em todo o mundo. De acordo com estimativas para a safra agrícola de 2011/12, a produção mundial de milho deve ser de, aproximadamente, 867,5 milhões de toneladas. Dessa forma, a produção mundial de milho supera a produção mundial de trigo (689,0 milhões de toneladas) e de arroz (460,8 milhões de toneladas) (USDA, 2011b). Os Estados Unidos da América (EUA) respondem por quase 50% da produção de milho, seguidos da China e do Brasil (Figura 1 - FAO, 2009). Em relação a produtividade de grãos de milho, em nível mundial, foi verificado um aporte de 61 kg por hectare ano<sup>-1</sup> entre 1961 e 2002 (DUVIC, 2005).



**Figura 1** - Ranking dos dez países com maior produção mundial de milho no ano de 2009. Adaptado de FAO, 2009.



No Brasil, a estimativa da produção de milho para o ano agrícola de 2011/12 é de 60,3 milhões de toneladas, com produtividade média de 4,1 toneladas por hectare, em 14,7 milhões de hectares (Figura 2 - CONAB, 2011). Considerando a série histórica da produção nacional, verifica-se que a área semeada aumentou, aproximadamente, 25% nos últimos 36 anos (de 11,8 milhões de hectares na safra 1976/77 para 14,7 milhões de hectares na safra 2011/12). Nesse mesmo período, a produção de milho aumentou 313% (de 19,3 milhões de toneladas na safra 1976/77 para 60,3 milhões de toneladas na safra 2011/12). Dessa forma, é evidenciado o aumento da produtividade por área, que foi de, aproximadamente, 256% (de 1,6 toneladas por hectares na safra 1976/77 para 4,1 toneladas por hectares na safra 2011/12), com um incremento médio de 70 kg por hectare ano<sup>-1</sup>.



**Figura 2** - Séries históricas da área semeada (em milhões de hectares), da produção (em milhões de toneladas) e da produtividade média (em toneladas por hectare) de milho no Brasil. Adaptado de CONAB, 2011.

A tendência mundial e brasileira de aumento da produtividade de grãos nas lavouras comerciais nos últimos anos está associada às melhorias no manejo da cultura e ao

melhoramento genético. Nesse sentido, Duvic (2005) e Borém e Miranda (2009) consideram que cada um destes contribui com cerca de 50% do aumento produtivo. Em relação às práticas de manejo, Duvic (2005) aponta: o controle das plantas daninhas e de insetos, a redução do espaçamento de plantas, a melhoria da mecanização agrícola e, principalmente, o aporte de fertilizantes nitrogenados sintéticos, como os principais fatores responsáveis pelo aumento da produtividade de grãos.

Considerando a contribuição do melhoramento genético, a introdução de híbridos em substituição as variedades de polinização aberta foi um dos grandes responsáveis pelo aumento produtivo. Além disso, Duvic (2005) descreve alterações realizadas em características das plantas de milho que contribuíram no aumento da produtividade de grãos: alterações que aumentaram a eficiência produtiva, como redução do ângulo de inclinação das folhas em relação ao colmo, resultando em aumento da taxa fotossintética, redução do tamanho de pendão e, maior período de enchimento de grãos; alterações que aumentaram a tolerância a fatores bióticos e abióticos, como tolerância ao aumento de densidade, redução de transpiração em períodos de estresse hídrico, redução do período entre a antese e o espigamento, redução da esterilidade e, tolerância a solos secos ou excessivamente úmidos, a baixa ou excessiva fertilidade, a baixa ou elevada temperatura, a insetos, a plantas daninhas, a doenças e a herbicidas. Todas essas alterações objetivaram a construção do ideótipo de uma planta de milho, conforme descrito por Borém e Miranda (2009). Além disso, a exploração da heterose e a utilização de técnicas modernas, como de marcadores moleculares, contribuíram no desenvolvimento de híbridos com maior potencial produtivo.

Embora as lavouras de milho do Brasil apresentem aumento constante da produtividade de grãos, os patamares ainda são consideravelmente inferiores aos de outros países. Por exemplo, a produtividade média estimada para 2011/12, de 4,1 toneladas por hectare representa apenas 44,6% da produtividade média dos EUA, de 9,2 toneladas por hectare em 2011 (USDA, 2011a). Nesse sentido, Fancelli e Dourado Neto (2004) consideram os baixos níveis tecnológicos adotados por grande parte dos produtores nacionais, como responsáveis pela baixa expressão do potencial genético das atuais cultivares. Ainda, considerando os ganhos genéticos decorrentes da substituição anual de cultivares, Storck et al. (2005) concluíram que as substituições são dinâmicas, variando entre 53 e 65% ao ano, e em alguns casos, não resultam em ganho genético para produtividade de grãos. Assim, Duvic (2005) salienta que o aumento da produtividade de grãos somente ocorre quando as práticas de manejo e os avanços do melhoramento genético se complementam.

Por isso, as empresas de melhoramento genético de milho têm buscado o desenvolvimento de cultivares que apresentem maior potencial produtivo em relação às disponibilizadas no mercado e utilizadas pelos produtores rurais. Após o desenvolvimento de novas cultivares, as empresas precisam submetê-las a diversas avaliações para identificar sua adaptação, estabilidade e contribuição no progresso genético. Nesse sentido, as redes de ensaios de competição de cultivares de milho são utilizadas para a avaliação das novas cultivares, gerando informações que podem ser utilizadas por técnicos e produtores para a escolha das cultivares a serem adotadas na produção de milho em cada região. Já as redes de ensaios de cultivares indicadas, fornecem informações importantes aos produtores sobre cultivares que já estão no mercado, em relação as suas principais características fenológicas, morfológicas e produtivas.

## **2.2 Normalidade dos dados**

A maioria das variáveis mensuradas na natureza, como as características fenológicas, morfológicas e produtivas de milho, apresenta comportamento aproximadamente simétrico, com distribuição gaussiana em forma de sino (SANTOS; FERREIRA, 2003). Assim, grande parte dos testes e procedimentos de estimação utilizados, adotam a distribuição normal dos dados como referência (PEREIRA FILHO; FERREIRA, 2011). Portanto, a normalidade dos dados é um dos pressupostos básicos em análises estatísticas inferenciais.

A interferência da não-normalidade sobre a análise estatística depende basicamente da forma de distribuição dos dados e do tamanho de amostra, sendo que o efeito nocivo sobre amostras grandes é menor (HAIR et al., 2009). Ainda, de acordo com esses autores, em análises estatísticas multivariadas, a normalidade é a suposição fundamental para que os resultados obtidos sejam válidos e confiáveis.

### **2.2.1 Normalidade univariada**

É importante assegurar que os dados utilizados apresentem distribuição normal. Inúmeros procedimentos têm sido utilizados a fim de verificar se os dados se ajustam a

distribuição normal univariada. A análise gráfica, por meio de histograma de frequência ou através de gráfico de probabilidade normal, embora permita uma visualização do comportamento dos dados de modo prático, é subjetiva e dificulta a detecção de normalidade (FERREIRA, 2008; HAIR et al., 2009). De modo indireto, pode-se verificar a normalidade dos dados, através da comparação dos valores de assimetria e curtose esperados na distribuição normal com os valores estimados nos dados amostrais, sendo a estatística de assimetria considerada mais poderosa em relação à curtose (SANTOS; FERREIRA, 2003).

Além desses modos indiretos de verificação de normalidade univariada, vários testes estatísticos foram desenvolvidos, tais como: Kolmogorov-Smirnov (KOLMOGOROV, 1933), Shapiro-Wilk (SHAPIRO; WILK, 1965), Lilliefors (LILLIEFORS, 1967) e qui-quadrado (CAMPOS, 1983). Dentre esses, o teste de Shapiro-Wilk tem sido considerado um dos mais poderosos testes de normalidade univariada (FERREIRA, 2008). Em amostras pequenas ( $n < 30$ ), esse teste é mais apropriado em relação ao teste de Kolmogorov-Smirnov (MAROCO, 2007), sendo o teste de normalidade univariada mais indicado.

### 2.2.2 Normalidade multivariada

Da mesma forma que para a detecção da normalidade univariada, diversos procedimentos são descritos na literatura para a verificação da normalidade multivariada, como: análises gráficas, testes baseados em desvios de assimetria e curtose (FERREIRA, 2008; MARDIA, 1970) e outros testes estatísticos. Entre eles, a utilização do teste de Shapiro-Wilk multivariado generalizado por Royston (ROYSTON, 1983) conforme descrito em Ferreira (2008) é recomendado como teste padrão de normalidade multivariada devido ao excelente controle do erro tipo I e do seu poder (CANTELMO; FERREIRA, 2007b).

Por outro lado, a utilização da função “*mshapiro.test*” do pacote “*mvnrmtest*”, do software R (R DEVELOPMENT CORE TEAM, 2012) na verificação da normalidade multivariada é desaconselhada em situações reais, devido ao baixo desempenho em relação ao controle do erro tipo I (CANTELMO; FERREIRA, 2007a; CANTELMO; FERREIRA, 2007b).

Em relação aos possíveis impactos da não-normalidade multivariada em análise de trilha, nenhuma aplicação em culturas agrícolas foi encontrada na literatura.

### 2.2.3 Transformação de variáveis

A transformação de dados têm sido sugerida para corrigir violações das suposições estatísticas, como a normalidade, podendo também contribuir para a melhoria das relações entre variáveis (HAIR et al., 2009). Para a transformação de dados, a família de transformações Box-Cox (BOX; COX, 1964) é amplamente utilizada, pois permite identificar a melhor transformação dos dados, com base na adição de um valor (*lambda*) que maximize o estimador de máxima verossimilhança e minimize o resíduo (KUTNER et al., 2004).

Ainda, a possibilidade de adição de uma constante, no momento da transformação dos dados, contribui para aumentar a aderência dos dados à distribuição normal quando existem muitos valores zeros, conforme aplicação prática em experimentos de abobrinha italiana (COUTO et al., 2009) e de brócolis (LÚCIO et al., 2010).

## 2.3 Multicolinearidade

De acordo com Gujarati (2006), o termo multicolinearidade foi empregado, inicialmente, por Ragnar Frisch, em 1934, e se refere a existência de relação linear perfeita ou menos que perfeita entre algumas ou todas as variáveis explicativas de um modelo de regressão. A multicolinearidade também pode ser entendida, como a extensão em que uma variável pode ser explicada por outras variáveis explicativas (HAIR et al., 2009). Os termos colinearidade e multicolinearidade se referem, respectivamente, à relação linear entre duas e mais de duas variáveis explicativas, embora, ambos são usados como sinônimos (GUJARATI, 2006; HAIR et al., 2009).

Quando a multicolinearidade é perfeita, os coeficientes de regressão são indeterminados e seus erros-padrão são infinitos (CRUZ; CARNEIRO, 2006; GUJARATI, 2006). Ainda, com alto grau de multicolinearidade, as variâncias e covariâncias são inflacionadas, e, por consequência, os intervalos de confiança dos coeficientes de regressão ou de trilha estimados são maiores, dificultando a rejeição de uma hipótese falsa, aumentando o erro tipo II (GUJARATI, 2006).

Em relação à multicolinearidade, Hair et al. (2009) atribuem três principais tarefas ao pesquisador: avaliar o grau de multicolinearidade, determinar seu impacto sobre os resultados e aplicar ações corretivas, quando necessário.

### 2.3.1 Diagnóstico de multicolinearidade

De acordo com Carvalho (1995), é importante realizar o diagnóstico de multicolinearidade, sempre que posteriormente forem realizadas análises de regressão, de trilha, de índice de seleção, de correlação parcial e de correlação canônica. Nesse sentido, vários métodos de diagnóstico do grau de multicolinearidade foram propostos na literatura.

Um primeiro indicativo de multicolinearidade é quando o coeficiente de determinação de um modelo de regressão ou de trilha é elevado e poucos coeficientes individuais são significativos pelo teste t, indicando alta variância compartilhada e pequena variância individual (GUJARATI, 2006). Variância individual ou específica é aquela que está unicamente relacionada à determinada variável explicativa, ou seja, não está associada a outras variáveis explicativas; e, a variância compartilhada ou comum é aquela que é comum a duas ou mais variáveis (HAIR et al., 2009). Outra forma de diagnóstico, é a análise dos coeficientes de correlação simples de cada par de variáveis, sendo que a alta correlação entre duas variáveis explicativas pode indicar alta colinearidade (CRUZ; CARNEIRO, 2006; VASCONCELLOS; ALVES, 2000). No entanto, a análise dos valores de coeficiente de correlação é desaconselhada para casos com mais de duas variáveis, visto que existem inter-relações entre o conjunto de variáveis explicativas (GUJARATI, 2006; HAIR et al., 2009). Assim, coeficientes de correlação elevados não, necessariamente, indicam alto grau de multicolinearidade e vice-versa.

Por isolar o efeito de outras variáveis, Farrar e Glauber (1967) propuseram, a análise de coeficientes de correlações parciais entre variáveis explicativas. No entanto, esse método sofreu críticas e mostrou-se ineficaz, pois uma determinada correlação parcial pode estar associada a diferentes padrões de multicolinearidade (WICHERS, 1975). Outra proposição sugere o uso de regressões auxiliares entre variáveis explicativas. Nesse caso, quando o coeficiente de determinação destas regressões for maior que o obtido para o modelo principal, têm-se problemas de multicolinearidade (GUJARATI, 2006).

O fator de inflação da variância (FIV) é outro indicador do grau de multicolinearidade, sendo obtido pelo inverso da tolerância. O FIV também pode ser obtido, para cada variável, na diagonal da inversa da matriz de correlação  $X'X$  (CRUZ; CARNEIRO, 2006; HAIR et al., 2009). Assim, obtem-se o FIV para a variável  $X_1$ , na linha 1 e coluna 1 da inversa da matriz de correlação  $X'X$ , o FIV da variável  $X_2$ , na linha 2 e coluna 2 e assim, sucessivamente, até  $X_n$  na linha  $n$  e coluna  $n$ . O FIV indica a quantidade de aumento da variância do coeficiente de regressão estimado, em função da presença da multicolinearidade e, a raiz quadrada de FIV expressa o erro-padrão associado ao coeficiente de regressão (FÁVERO et al., 2009). Considera-se multicolinearidade severa, quando  $FIV > 10$  (FÁVERO et al., 2009; GUJARATI, 2006; HAIR et al., 2009). Quanto maior for o erro-padrão, maior será o intervalo de confiança e maior será a dificuldade em detectar a significância estatística dos coeficientes de regressão estimados (HAIR et al., 2009).

Outro método de diagnóstico de multicolinearidade é denominado “tolerância” e indica a proporção da variação de uma variável explicativa que independe das demais variáveis explicativas (FÁVERO et al., 2009; GUJARATI, 2006; HAIR et al., 2009). A tolerância é obtida pela subtração de  $1 - R_X^2$ , em que  $R_X^2$  é o coeficiente de ajuste da regressão (coeficiente de determinação do modelo) entre a variável explicativa  $X$  e as outras variáveis explicativas presentes (FÁVERO et al., 2009). Assim, quanto menor for o valor de tolerância, maior será a quantidade de variância compartilhada entre as variáveis explicativas e maior será o grau de multicolinearidade. Como critério de decisão, adota-se tolerância menor que 0,10, como indicativo de multicolinearidade severa, embora esse valor seja considerado pequeno, pois com tolerância  $< 0,10$ , o coeficiente de determinação obtido entre as variáveis explicativas ultrapassa 0,90 (GUJARATI, 2006; HAIR et al., 2009).

O número de condição, que é a divisão do maior pelo menor autovalor da matriz de correlação  $X'X$ , tem sido largamente utilizado no diagnóstico do grau de multicolinearidade (CRUZ; CARNEIRO, 2006; GUJARATI, 2006). Número de condição menor que 100 indica multicolinearidade fraca, entre 100 e 1.000, moderada a forte e, maior que 1.000, multicolinearidade severa (MONTGOMERY; PECK, 1982).

Por fim, também tem sido considerado no diagnóstico de multicolinearidade, o determinante da matriz de correlação  $X'X$ , sendo que quanto menor for o valor do determinante, maior será o grau de multicolinearidade e maior será a probabilidade de singularidade da matriz, impedindo o cálculo da inversa da matriz de correlação  $X'X$ . Quando a multicolinearidade for perfeita, ao menos um dos autovalores será nulo e por consequência o

determinante também será nulo. No entanto, não é possível identificar nesse método, quais variáveis estão causando a multicolinearidade (CRUZ; CARNEIRO, 2006).

Os métodos mais indicados e utilizados são: fator de inflação da variância, tolerância, número de condição e determinante da matriz. A utilização de mais de um desses métodos é recomendada para a confirmação do grau de multicolinearidade, pois de acordo com Gujarati (2006), um valor pequeno de tolerância e alto de FIV, por exemplo, não necessariamente confirma um alto grau de multicolinearidade. Da mesma forma, um determinante da matriz próximo a zero indica problemas de multicolinearidade, mas não permite a verificação das variáveis causadoras (CRUZ; CARNEIRO, 2006).

### 2.3.2 Impacto da multicolinearidade em análises multivariadas

Conforme destacado anteriormente, o grau de multicolinearidade está diretamente associado ao erro-padrão nas estimativas de coeficiente de regressão e de trilha e pode resultar em estimativas de coeficientes errados ou com sinais trocados (CRUZ; CARNEIRO, 2006; GUJARATI, 2006). Além disso, a multicolinearidade dificulta a verificação da contribuição relativa de cada variável explicativa sobre a variável dependente em análises de regressão múltipla (CRUZ; CARNEIRO, 2006; HAIR et al., 2009). No método de componentes principais, Srivastava (2002) relata que, quando o número de variáveis explicativas é elevado, o grau de multicolinearidade tende a ser maior e, a matriz de correlação  $X'X$  singular.

O uso de variáveis altamente correlacionadas na análise de agrupamento, aumenta a contribuição dessas variáveis no processo de formação de grupos, mesmo que o uso das mesmas expresse informações redundantes. Assim, deve-se examinar as variáveis utilizadas no agrupamento quanto à existência de multicolinearidade (CORRAR et al., 2007; CRUZ; CARNEIRO, 2006; HAIR et al., 2009). Em estudo com cultivares de feijão, diferenças de agrupamentos foram observadas quando se possuía baixo e elevado grau de multicolinearidade (CARGNELUTTI FILHO et al., 2009) e, para a obtenção de adequados padrões de agrupamento, os autores sugerem a redução do grau de multicolinearidade.

Especificamente em análise de trilha, em que são estimados coeficientes de trilha obtidos por equações de regressão, esses são afetados pelo grau de multicolinearidade existente entre os caracteres estudados. Quando o grau de multicolinearidade é elevado, os coeficientes de trilha podem atingir valores elevados, e, por consequência, serem pouco



confiáveis e sem sentido prático em relação ao fenômeno estudado (CARVALHO, 1995). Em canola, ocorreu uma super-estimativa dos valores dos coeficientes de correlações simples e dos efeitos diretos sobre o caráter produtividade de grãos, estimados através da análise de trilha com multicolinearidade severa (COIMBRA et al., 2005). Segundo os autores, a realização da análise de trilha com grau de multicolinearidade severa produz resultados sem nenhuma importância biológica para o melhorista de plantas. Em arroz, componentes primários de rendimento apresentaram efeito direto negativo sobre a produtividade de grãos, na presença de multicolinearidade, impossibilitando a interpretação dos coeficientes de trilha estimados (SHRIVASTAVA; SHARMA, 1976). Gondim et al. (2008) salientam que muitos trabalhos desenvolvidos por pesquisadores com a aplicação de análise de trilha, não levaram em consideração o diagnóstico da multicolinearidade, desprezando os possíveis efeitos adversos dessa sobre os coeficientes estimados.

Por outro lado, Goldberger (1968 apud GUJARATI, 2006) critica a importância dada a multicolinearidade e segundo esse, deve-se observar também a micronumerosidade, referindo-se a amostras pequenas. Dessa forma, o autor salienta que, amostras pequenas e falta de variabilidade entre as variáveis explicativas, pode resultar em efeitos nocivos maiores na estimação de coeficientes de regressão, em relação a condições de alto grau de multicolinearidade.

### 2.3.2 Ações corretivas para redução da multicolinearidade

Quando em determinado conjunto de variáveis existe alto grau de multicolinearidade, algumas medidas podem ser adotadas a fim de reduzir os efeitos nocivos da mesma. Um primeiro conselho, de acordo com Gujarati (2006), é buscar identificar os padrões de associação entre as variáveis, com base em trabalhos desenvolvidos *a priori*. Ainda de acordo com o autor, a transformação de variáveis e, por consequência, a alteração nos padrões de associação pode reduzir o grau de multicolinearidade. Outra forma de reduzir o efeito da multicolinearidade sobre a estimação de coeficientes de regressão ou de trilha é através da adição de dados, visto que a interferência da multicolinearidade em amostras grandes é menor (GUJARATI, 2006). Também, a utilização de técnicas multivariadas, como a análise fatorial, pode ser uma alternativa, pois permite o agrupamento de variáveis semelhantes em um único

fator. Dessa forma, a técnica não sofre impacto do grau de multicolinearidade e contribui para sua redução (FÁVERO, 2009; GUJARATI, 2006).

Todas as medidas descritas anteriormente contribuem para a redução do grau de multicolinearidade. No entanto, a eliminação de variáveis altamente correlacionadas tem sido descrita como uma das mais efetivas formas de redução da multicolinearidade (FÁVERO et al., 2009; GUJARATI, 2006; HAIR et al., 2009). Embora essa técnica seja relativamente simples e alcance o objetivo, deve-se ter o cuidado de não eliminar variáveis importantes na predição da variável principal, resultando em erro de especificação (GUJARATI, 2006; HAIR et al., 2009). O erro de especificação se refere ao erro de previsão da variável dependente devido à omissão de uma variável-chave na análise, com distorção dos coeficientes estimados e redução do poder preditivo (HAIR et al., 2009). No entanto, Gurajati (2006) ressalta que o erro de especificação pode ocorrer tanto pela omissão de uma variável-chave, como pela permanência de uma variável redundante na análise.

Ainda, pode-se reduzir o impacto de multicolinearidade por técnicas específicas de análise, como a regressão em crista e, a análise de trilha sob multicolinearidade, cujo procedimento será descrito detalhadamente nos itens 2.4.2 e 3.6.2.

## **2.4 Análise de trilha**

O estudo das relações existentes entre caracteres é importante no processo de melhoramento de plantas via seleção indireta, quando a variável de interesse apresenta baixa herdabilidade, dificuldade de mensuração, se deseja a seleção precoce de plantas ou linhagens ou ainda, quando se deseja a seleção simultânea para mais de um caractere. Nesse sentido, o primeiro passo é a determinação da associação entre variáveis mediante a estimativa dos coeficientes de correlação simples (fenotípicos ou genotípicos). No entanto, quando se possui mais de duas variáveis, a associação medida pelo coeficiente de correlação pode não ser real. Assim, valores próximos a unidade não necessariamente indicam elevada associação linear e valores próximos a zero não necessariamente indicam falta de associação. Isso se deve ao efeito que outras variáveis exercem sobre uma determinada associação (CRUZ; REGAZZI, 1997).

Assim, embora os coeficientes de correlação sejam importantes indicativos da associação entre duas variáveis, mostrando a magnitude e o sentido da associação, técnicas

complementares devem ser utilizadas a fim de comprovar as relações de causa e efeito entre variáveis. Dessa forma, a análise de trilha vem sendo amplamente utilizada em estudos de melhoramento genético de inúmeras culturas agrícolas e forrageiras e no melhoramento genético animal. Essa técnica foi descrita inicialmente por Wright (1921, 1923) e consiste no desdobramento dos coeficientes de correlação simples em efeitos diretos e indiretos sobre a variável principal. Assim, mediante a utilização dessa técnica, é possível diagnosticar as reais relações de causa e efeito das variáveis explicativas sobre a variável principal. Variáveis que apresentam alta associação medida através do coeficiente de correlação e que possuem efeitos diretos de mesma intensidade e direção, são consideradas variáveis de causa e efeito (CRUZ; CARNEIRO, 2006; CRUZ et al., 2004; CRUZ; REGAZZI, 1997).

#### 2.4.1 Análise de trilha tradicional

A análise de trilha tradicional, com um único modelo causal é uma regressão parcial padronizada. A partir de um conjunto de variáveis explicativas, determina-se a contribuição de cada uma sobre a variável principal, através dos seus efeitos diretos e indiretos (CRUZ; CARNEIRO, 2006).

Em análise de trilha, correlacionando caracteres de tubérculos de batata, não foi verificado elevado grau de multicolinearidade, permitindo a realização da análise de trilha tradicional (SILVA et al., 2007). Em caracteres de capim-elefante também foi constatada multicolinearidade fraca, podendo estar associado ao pequeno número de variáveis explicativas (três) (DAHER et al., 2004). Em arroz, foi realizada a análise de trilha tradicional da produtividade de grãos em função dos componentes primários de rendimento com fraca multicolinearidade (MARCHEZAN et al., 2005). Para componentes de rendimento de cana-planta e cana-soca, também foi verificada presença de multicolinearidade fraca entre os caracteres (ESPÓSITO et al., 2012). Em outras culturas, como algodão (HOOPERHEIDE et al., 2007), pereira (GALARÇA et al., 2010) e espécies florestais exóticas (LORENTZ et al., 2006) também foi observada multicolinearidade fraca entre os caracteres estudados. Nesses casos, a adoção da análise de trilha tradicional é adequada (CRUZ; CARNEIRO, 2006).

No entanto, quando a matriz de correlação apresenta problemas de multicolinearidade moderada ou severa, os coeficientes de trilha estimados têm apresentado pouco sentido biológico, conforme já relatado para as culturas de pimentão (CARVALHO et al., 1999), de

soja (BIZETI et al., 2004) e de canola (COIMBRA et al., 2005). Nessas condições, a adoção de técnicas alternativas, como a análise de trilha sob multicolinearidade ou a análise de trilha tradicional, com a eliminação de variáveis altamente correlacionadas, devem ser consideradas.

#### 2.4.2 Análise de trilha sob multicolinearidade

Em função dos problemas ocasionados pela multicolinearidade sobre análises estatísticas, como na regressão múltipla e na análise de trilha, Carvalho (1995) propôs a utilização da regressão em crista ou cumeeira e, num processo similar, análise de trilha sob multicolinearidade. Conforme descrito por Cruz e Carneiro (2006), o método é baseado na estimação de coeficientes de trilha, a partir de equações parcialmente modificadas, através da adição de uma constante  $k$  nos elementos da diagonal da matriz de correlação  $X'X$ . A partir da adição desta constante positiva e de pequena magnitude, a matriz de correlação  $X'X$  passa a ser não-singular, permitindo o cálculo da matriz inversa comum (CRUZ; CARNEIRO, 2006).

Conforme Cruz e Carneiro (2006), como consequência da adição de uma constante, a soma de quadrados dos desvios é maior em relação à regressão ou análise de trilha tradicional e o coeficiente de determinação da equação é menor. Também, o estimador do modelo passa a ser tendencioso e a variância associada ao estimador decresce a medida em que se aumenta o valor de  $k$  adicionado. Pela interferência do valor de  $k$  adicionado na análise de regressão ou de trilha, alguns métodos de estimativa de  $k$  foram desenvolvidos, sendo descritos por Cruz e Carneiro (2006). Entre eles, o método visual com base no traço de crista, que permite a identificação do menor valor  $k$ , a partir do qual, ocorre uma estabilização da estimativa dos coeficientes de regressão ou de trilha, tem sido utilizado frequentemente (CRUZ; CARNEIRO, 2006). Assim, deve-se escolher um valor  $k$  que permita a estabilização dos coeficientes estimados, mas que: não reduza demasiadamente a acurácia, aumente o resíduo ou que resulte em estimadores viesados (CRUZ; CARNEIRO, 2006).

Em estudo de dezesseis caracteres da cultura de mamoeiro, Oliveira et al. (2010) constataram elevado grau de multicolinearidade e optaram pela adição da constante  $k$  na diagonal da matriz de correlação  $X'X$ , antes da análise de trilha, porém não informaram o valor da constante. Em girassol, Amorim et al. (2008) utilizaram um valor de  $k = 0,05$ . Segundo os autores, esse valor foi suficiente para a correção das distorções nas estimativas dos coeficientes de trilha, contribuindo para que o fator de inflação de variância das 13

variáveis reduzisse para valores menores do que 10. O mesmo valor de  $k$  utilizado em caracteres de girassol, foi suficiente para a estabilização dos coeficientes de trilha estimados para caracteres de soja (BIZETI et al., 2004) e de pimentão (CARVALHO et al., 1999).

Na cultura de canola, Coimbra et al. (2005) verificaram a existência de multicolinearidade severa em caracteres de uma época de semeadura. Na diagonal da matriz de correlação  $X'X$  dessa época, foi adicionado um valor de  $k = 0,278$ . Na seleção de caracteres, visando a produção de vagens maduras de amendoim, Luz et al. (2011) também adicionaram uma constante  $k$  na diagonal da matriz de correlação  $X'X$ , após detecção de problemas devido a elevada multicolinearidade. No entanto, Carvalho et al. (1999) observaram que em um dos casos de análise de trilha em pimentão, não se encontrou um valor adequado de  $k$  que reduzisse a variância dos coeficientes de trilha, sem que houvesse expressiva redução do coeficiente de determinação dos modelos de regressão. Em estudo de relações entre 20 caracteres de tomate, a utilização da constante  $k$  no intervalo de zero a um, também não foi efetiva para a estabilização da variância associada aos coeficientes de trilha estimados, sendo utilizada a análise de trilha tradicional, após a eliminação de oito caracteres altamente correlacionados (RODRIGUES et al., 2010).

#### 2.4.3 Análise de trilha tradicional com eliminação de variáveis

Além da análise de trilha sob multicolinearidade, alternativamente, pode-se realizar a análise de trilha tradicional, após a eliminação de variáveis altamente correlacionadas. Assim, realiza-se o diagnóstico de multicolinearidade e procede-se a eliminação de uma ou mais variáveis que estejam altamente associadas às demais variáveis explicativas e que apresentem elevada variância compartilhada (CRUZ; CARNEIRO, 2006).

Para realizar a eliminação de variáveis, é necessário que o pesquisador conheça previamente os padrões de associação das variáveis explicativas com a variável principal. Caso contrário, o pesquisador poderá vir a cometer erros de especificação, através da eliminação de variáveis que apresentam importância na predição da variável principal ou devido à manutenção de variáveis com informações redundantes (CRUZ; CARNEIRO, 2006; GUJARATI, 2006). Assim, para que a análise de trilha apresente funcionalidade prática é necessária a formulação prévia das relações de causa e efeito entre as variáveis buscando manter as variáveis de maior interesse (CRUZ et al., 2004). Carvalho et al. (1999) salientam

que no descarte de variáveis, não necessariamente a variável que tem maior importância econômica é a que mais explica a variável básica ou principal.

Além disso, a decomposição dos coeficientes de correlação é variável em função do conjunto de variáveis presentes no estudo (CRUZ; REGAZZI, 1997). Assim, após a exclusão de variáveis, os efeitos diretos e indiretos das variáveis explicativas que permanecem sofrem alterações quanto à magnitude e inclusive em relação à direção da associação (positiva ou negativa), cabendo ao pesquisador estar atento para realizar a correta interpretação dos coeficientes estimados.

A utilização da análise de trilha tradicional, após a eliminação de variáveis altamente correlacionadas, tem se mostrado eficiente na redução dos efeitos adversos da multicolinearidade em diversas culturas. Entre elas, citam-se os estudos realizados em arroz (SHRIVASTAVA; SHARMA, 1976), em pimentão (CARVALHO et al., 1999), em soja (CARVALHO et al., 2002; BIZETI et al., 2004), em milho (LOPES et al., 2007), em trigo (GONDIM et al., 2008) e em tomateiro (RODRIGUES et al., 2010). Ainda, Carvalho et al. (1999) e Bizeti et al. (2004) concluíram que tanto a eliminação de variáveis altamente correlacionadas, como a utilização da constante  $k$  na diagonal da matriz de correlação  $X'X$ , foram eficientes na redução dos efeitos adversos da multicolinearidade em análise de trilha de pimentão e soja, respectivamente.

#### 2.4.4 Utilização da matriz de correlação fenotípica em análise de trilha

Conforme descrito anteriormente, em estudos de relações entre caracteres, o conhecimento do grau de associação dos mesmos, por meio de correlações, possibilita identificar caracteres que possam ser usados na seleção indireta, medindo a magnitude e a direção das associações. Em análise de trilha, o uso de matrizes de correlações fenotípicas e genotípicas tem sido relatado na literatura. As correlações fenotípicas expressam as associações entre caracteres com influências de causas genéticas e ambientais, sendo diretamente mensuradas na natureza. As correlações genotípicas expressam apenas a associação de natureza herdável (CRUZ; REGAZZI, 1997) e são ocasionadas principalmente pelo pleiotropismo, definido como a propriedade pela qual um gene condiciona mais de um caráter simultaneamente e, pelo desequilíbrio de ligação, que é a associação não-aleatória entre alelos de diferentes locos (FALCONER, 1987).

Segundo Kang et al. (1983), quando os valores de coeficientes de correlações fenotípicas e genotípicas são diferentes em magnitude ou sentido, as correlações genotípicas são intrinsecamente mais úteis que as fenotípicas para indicar estratégias de seleção. No entanto, Shukla et al. (1998) e Resende (2002) afirmam que caracteres genotipicamente correlacionados, mas não fenotipicamente correlacionados, podem não ser de valor prático na seleção, pois geralmente esta é baseada no fenótipo.

Ainda, a proporção genética da variabilidade total é denominada herdabilidade no sentido amplo (BORÉM; MIRANDA, 2009). Em 16 ensaios de competição de cultivares de milho, foi constatado que para 81,25% dos casos de variáveis mensuradas, a acurácia seletiva foi alta ou muito alta (CARGNELUTTI FILHO; GUADAGNIN, 2011) e dessa forma, a herdabilidade também foi alta ou muito alta (RESENDE; DUARTE, 2007; CARGNELUTTI FILHO; STORCK, 2009). Esses resultados indicam que para caracteres mensurados em ensaios de competição de cultivares de milho, a utilização da matriz de coeficientes de correlação fenotípica ou genotípica, resulta em interpretações semelhantes dos efeitos diretos e indiretos da análise de trilha. Nesse sentido, Espósito et al. (2012) verificaram que praticamente não houve diferenças de interpretações e conclusões entre as análises de trilha baseadas nas matrizes de correlações fenotípicas ou genotípicas em cana-planta e cana-soca. Esses autores comentam ainda, que não há consenso quanto à utilização preferencial de uma delas e salientam a importância de estudos futuros nesse sentido.

Considerando a utilização paralela de coeficientes de correlação fenotípicas e genotípicas, Churata e Ayala-Osuna (1996), Silva et al. (2008), Oliveira et al. (2010) e Espósito et al. (2012), verificaram que as estimativas das correlações genotípicas foram maiores do que as fenotípicas, para todas as características de milho, capim-elefante, mamoeiro, cana-planta e cana-soca, respectivamente, evidenciando maiores contribuições dos fatores genéticos. Segundo Churata e Ayala-Osuna (1996), pode-se verificar a semelhança entre os coeficientes de correlação fenotípica e genotípica para a maioria dos caracteres em relação à magnitude e direção das associações. Já em soja, Carvalho et al. (2002) constataram maiores graus de multicolinearidade associados as matrizes de correlações genotípicas em relação às fenotípicas.

Assim, as correlações fenotípicas são em geral, mais utilizadas pelos melhoristas em estudos de análise de trilha (CABRAL et al., 2011). As correlações fenotípicas foram usadas em estudos de análise de trilha em culturas agrícolas como: canola (COIMBRA et al., 2004; COIMBRA et al., 2005), soja (BIZETI et al., 2004), espécies florestais exóticas (LORENTZ et al., 2006), trigo (GONDIM et al., 2008) e feijão (CABRAL et al., 2011).

## 3 METODOLOGIA

### 3.1 Os ensaios

Os dados utilizados nesse trabalho foram provenientes de ensaios (experimentos) de competição de cultivares de milho do estado do Rio Grande do Sul (RS), conduzidos nos anos agrícolas 2002/03, 2003/04 e 2004/05. Foram usados os dados de 44 ensaios da rede estadual, sob coordenação da Fundação Estadual de Pesquisa Agropecuária (FEPAGRO). Esses ensaios são classificados quanto à categoria (estadual e indicado) e quanto ao ciclo (superprecoce e precoce) (Tabela 1).

Os ensaios da categoria estadual contiveram cultivares ainda não indicadas aos produtores, juntamente com duas ou mais cultivares testemunhas, amplamente comercializadas no RS. Os ensaios da categoria indicado, contiveram cultivares já indicadas aos produtores. Em relação ao ciclo, são consideradas cultivares de ciclo superprecoce, aquelas que necessitam entre 780 e 830 unidades calóricas (UC) para completar o período de emergência das plantas até o pendoamento-espigamento. Já as cultivares de ciclo precoce, necessitam de 831 a 890 UC para completar o mesmo período fenológico (FANCELLI; DOURADO NETO, 2004).

Os ensaios foram conduzidos nos municípios de Capão do Leão, Carazinho, Cruz Alta, Passo Fundo, Pelotas, Rio Pardo, Roca Sales, Santa Maria e São Borja (Figura 3). Nos 44 ensaios, foi usado o delineamento de blocos ao acaso com três repetições. Cada parcela foi composta por uma cultivar semeada em duas linhas de 5m de comprimento. A densidade de semeadura e o espaçamento entre linhas variaram em função da categoria, do ciclo e do ano agrícola do ensaio. O número de cultivares semeadas em cada ensaio variou entre nove e 40.

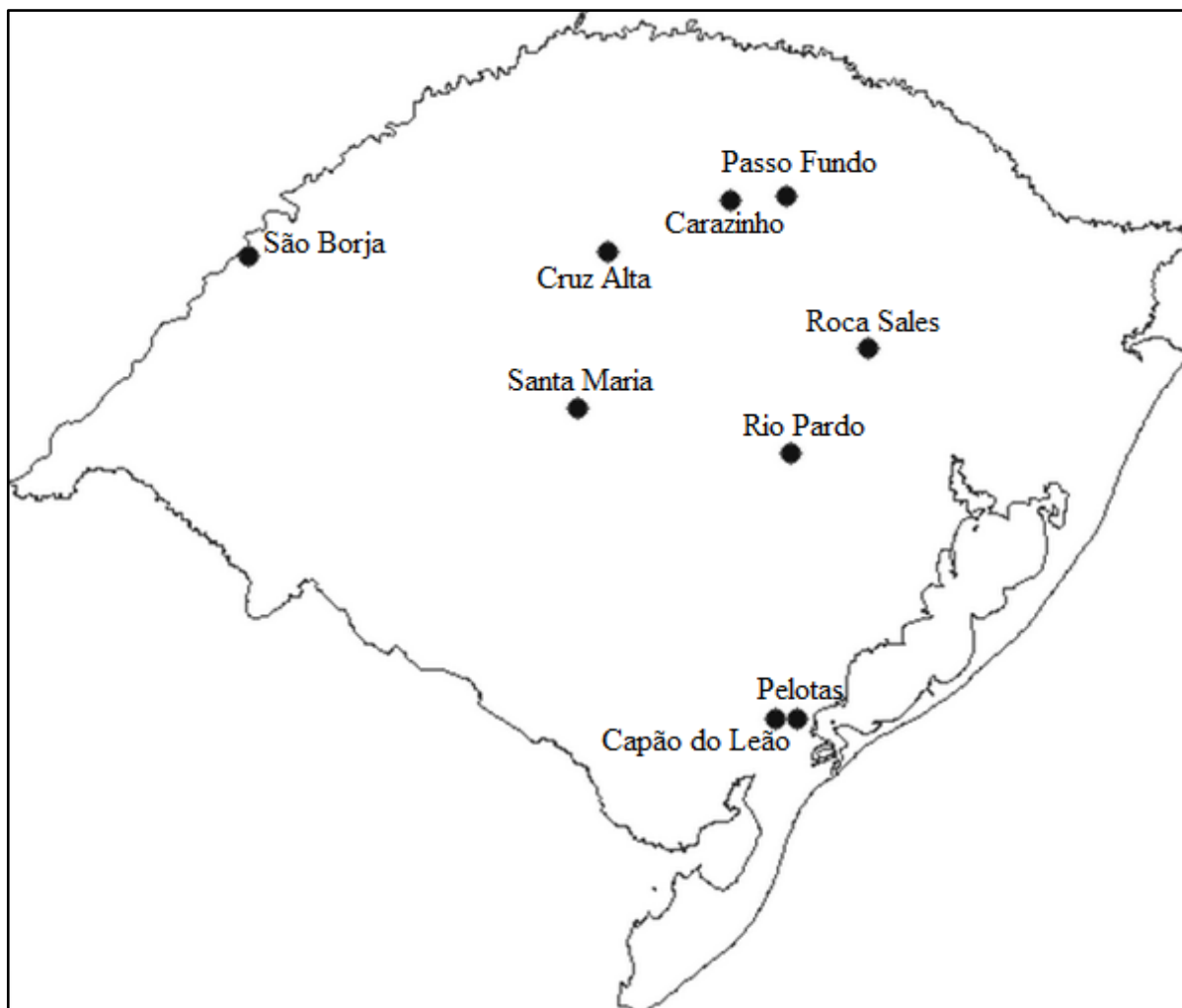
Para cada cultivar, em cada ensaio, foram mensuradas oito variáveis: número de dias até o florescimento masculino (FM, em dias); estatura de plantas na colheita (EP, em cm); altura da inserção da espiga na colheita (AIE, em cm); posição relativa da espiga (PRE, obtida pela divisão de AIE por EP); número de plantas na colheita (NP, unidades ha<sup>-1</sup>); número de espigas na colheita (NE, unidades ha<sup>-1</sup>); prolificidade (PLI, obtida pela divisão de NE por NP); e, produtividade de grãos ajustada a 13% de umidade (PROD, em kg ha<sup>-1</sup>). Foram utilizados os valores médios de cada cultivar, obtidos a partir da média das três repetições.



**Tabela 1** - Relação dos 44 ensaios de competição de cultivares de milho, conduzidos em locais do estado do Rio Grande do Sul, com a respectiva classificação da categoria e do ciclo, ano agrícola, data de semeadura, data de colheita e número de cultivares avaliadas.

Ensaio	Local	Categoria <sup>(1)</sup>	Ciclo <sup>(2)</sup>	Ano Agrícola	Data de Semeadura	Data de Colheita	Número de Cultivares
1	Passo Fundo	Estadual	Superprecoce	2002/03	03/11/2002	17/04/2003	11
2	Pelotas	Estadual	Superprecoce	2002/03	09/01/2003	16/06/2003	11
3	Rio Pardo	Estadual	Superprecoce	2002/03	26/12/2002	12/05/2003	11
4	Roca Sales	Estadual	Superprecoce	2002/03	27/09/2002	27/02/2003	11
5	São Borja	Estadual	Superprecoce	2002/03	22/11/2002	27/03/2003	11
6	Passo Fundo	Estadual	Precoce	2002/03	03/11/2002	21/04/2003	36
7	Pelotas	Estadual	Precoce	2002/03	08/01/2003	24/06/2003	36
8	Rio Pardo	Estadual	Precoce	2002/03	26/12/2002	12/05/2003	36
9	Roca Sales	Estadual	Precoce	2002/03	27/09/2002	27/02/2003	36
10	São Borja	Estadual	Precoce	2002/03	22/11/2002	28/03/2003	36
11	Passo Fundo	Indicado	Superprecoce	2002/03	03/11/2002	17/04/2003	18
12	Rio Pardo	Indicado	Superprecoce	2002/03	26/12/2002	12/05/2003	18
13	Santa Maria	Indicado	Superprecoce	2002/03	06/11/2002	17/03/2003	18
14	Passo Fundo	Indicado	Precoce	2002/03	03/11/2002	17/04/2003	27
15	Rio Pardo	Indicado	Precoce	2002/03	26/12/2002	12/05/2003	27
16	Santa Maria	Indicado	Precoce	2002/03	06/11/2002	17/03/2003	27
17	Passo Fundo	Estadual	Superprecoce	2003/04	17/10/2003	Não Informado	9
18	Pelotas	Estadual	Superprecoce	2003/04	20/12/2003	Não Informado	9
19	Rio Pardo	Estadual	Superprecoce	2003/04	28/10/2003	17/03/2004	9
20	Passo Fundo	Estadual	Precoce	2003/04	10/10/2003	Não Informado	40
21	Pelotas	Estadual	Precoce	2003/04	20/12/2003	Não Informado	40
22	Rio Pardo	Estadual	Precoce	2003/04	28/10/2003	06/04/2004	40
23	Passo Fundo	Indicado	Superprecoce	2003/04	10/10/2003	Não Informado	16
24	Pelotas	Indicado	Superprecoce	2003/04	20/12/2003	Não Informado	16
25	Rio Pardo	Indicado	Superprecoce	2003/04	29/10/2003	19/03/2004	16
26	Santa Maria	Indicado	Superprecoce	2003/04	29/10/2003	22/03/2004	16
27	Passo Fundo	Indicado	Precoce	2003/04	13/10/2003	Não Informado	27
28	Pelotas	Indicado	Precoce	2003/04	20/12/2003	Não Informado	27
29	Rio Pardo	Indicado	Precoce	2003/04	29/10/2003	13/04/2004	27
30	Santa Maria	Indicado	Precoce	2003/04	30/10/2003	22/03/2004	27
31	Capão do Leão	Estadual	Superprecoce	2004/05	26/11/2004	19/05/2005	17
32	Carazinho	Estadual	Superprecoce	2004/05	05/10/2004	22/03/2005	17
33	Cruz Alta	Estadual	Superprecoce	2004/05	30/09/2004	08/03/2005	17
34	Rio Pardo	Estadual	Superprecoce	2004/05	01/12/2004	27/04/2005	17
35	Capão do Leão	Estadual	Precoce	2004/05	26/11/2004	23/05/2005	32
36	Carazinho	Estadual	Precoce	2004/05	05/10/2004	22/03/2005	32
37	Cruz Alta	Estadual	Precoce	2004/05	30/09/2004	08/03/2005	32
38	Rio Pardo	Estadual	Precoce	2004/05	01/12/2004	27/04/2005	32
39	Cruz Alta	Indicado	Superprecoce	2004/05	30/09/2004	08/03/2005	12
40	Rio Pardo	Indicado	Superprecoce	2004/05	01/12/2004	27/04/2005	12
41	Santa Maria	Indicado	Superprecoce	2004/05	21/10/2004	18/03/2005	12
42	Cruz Alta	Indicado	Precoce	2004/05	30/09/2004	08/03/2005	30
43	Rio Pardo	Indicado	Precoce	2004/05	01/12/2004	27/04/2005	30
44	Santa Maria	Indicado	Precoce	2004/05	21/10/2004	18/03/2005	30

<sup>(1)</sup> Os ensaios da categoria estadual contêm cultivares ainda não indicadas aos produtores, juntamente com duas ou mais cultivares testemunhas, amplamente comercializadas no estado do RS. Os ensaios da categoria indicado, contêm cultivares já indicadas aos produtores. <sup>(2)</sup> Cultivares de ciclo superprecoce necessitam entre 780 e 830 unidades calóricas (UC) para completar o período de emergência das plantas até o pendramento-espigamento e cultivares de ciclo precoce, necessitam de 831 e 890 UC para completar o mesmo período fenológico (FANCELLI; DOURADO NETO, 2004).



**Figura 3** - Mapa do estado do Rio Grande do Sul, com os nove municípios onde foram realizados os 44 ensaios de competição de cultivares de milho utilizados nesse trabalho.

### 3.2 Estatísticas descritivas

Para cada variável, em cada ensaio, calcularam-se as estatísticas: mínimo, média, máximo, desvio-padrão e coeficiente de variação, com base nos valores médios de cada cultivar. Como foram usados apenas os valores médios de cada cultivar, o coeficiente de variação (em %) expressa apenas a variabilidade entre as cultivares de um mesmo ensaio para cada variável. Dessa forma, não se pode utilizar as informações de coeficiente de variação desse trabalho, para a interpretação da precisão experimental dos ensaios.

### 3.3 Diagnóstico da normalidade

#### 3.3.1 Normalidade univariada

Para cada variável, em cada ensaio, foi realizado o diagnóstico da distribuição normal univariada dos dados por meio do teste de Shapiro-Wilk (SHAPIRO; WILK, 1965). Assim, inicialmente foram estabelecidas as hipóteses do teste, sendo a hipótese nula  $H_0$ : a amostra provém de uma população normal e a hipótese alternativa  $H_1$ : a amostra não provém de uma população normal. Foi estabelecido um nível de significância  $\alpha$  do teste de 0,10. Posteriormente os valores da variável  $X$  foram dispostos em ordem crescente ( $X_1 \leq X_2 \leq \dots \leq X_n$ ). Em seguida, calculou-se a estatística  $W_{cal}$  conforme descrito por Fávero et al. (2009) por meio das equações:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} = \sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2 \quad (1)$$

$$b = \sum_{i=1}^{n/2} a_{n-i+1} \times (X_{n-i+1} - X_i) \quad (2)$$

$$W_{cal} = \frac{b^2}{\sum_{i=1}^n (X_i - \bar{X})^2}, \text{ para } i = 1, 2, \dots, n \quad (3)$$

em que:

$X_i$  são os valores da variável  $X$  dispostos em ordem crescente;

$\bar{X}$  é a média da variável  $X$ ;

$n$  é o número de observações (número de cultivares) da variável  $X$ . Caso  $n$  seja ímpar, desconsidera-se o valor da mediana; e,

$a_i$  são coeficientes tabelados para o teste de normalidade  $W$  de Shapiro-Wilk em função dos valores de  $i$  e de  $n$ .

Após o cálculo da estatística  $W_{cal}$ , utilizou-se para a interpretação, a estatística valor- $p$ , que corresponde ao nível mínimo de significância. Assim, quando o valor- $p$  foi  $\leq 0,10$  para a estatística  $W_{cal}$ , rejeitou-se  $H_0$  e concluiu-se que os dados da variável  $X$  não provém de uma população normal. Quando o valor- $p$  foi  $> 0,10$ , não rejeitou-se  $H_0$  e concluiu-se que os dados

da variável  $X$  provêm de uma população com distribuição normal. Para a aplicação do teste de normalidade univariada de Shapiro-Wilk, utilizou-se a função “*Shapiro.test()*” do software R (R DEVELOPMENT CORE TEAM, 2012).

### 3.3.2 Normalidade multivariada

Em cada ensaio, após o diagnóstico da normalidade univariada, foi realizado o diagnóstico da distribuição normal multivariada dos dados. Foi estabelecida a hipótese nula  $H_0$ : os dados amostrais seguem distribuição normal multivariada e a hipótese alternativa  $H_1$ : os dados amostrais não seguem distribuição normal multivariada. Estabeleceu-se um nível de significância  $\alpha$  de 0,10. Em seguida foi calculada a estatística do teste multivariado de normalidade de Shapiro-Wilk generalizado por Royston (ROYSTON, 1983), conforme descrito por Ferreira (2008).

Da mesma forma que para a distribuição de normalidade univariada, utilizou-se como critério de decisão a estatística valor-p. Assim, quando o valor-p calculado foi  $\leq 0,10$  para a estatística  $W$  do teste multivariado de normalidade de Shapiro-Wilk, rejeitou-se  $H_0$  e concluiu-se que os dados amostrais não seguem distribuição normal multivariada. Quando o valor-p foi  $> 0,10$ , não rejeitou-se  $H_0$  e concluiu-se que os dados amostrais seguem distribuição normal multivariada. As análises de normalidade multivariada foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2012) com programação específica desenvolvida para esse estudo<sup>1</sup>.

### 3.3.3 Transformação de dados

Após o diagnóstico de normalidade uni e multivariada dos dados amostrais, procedeu-se a transformação dos dados de variáveis que não se ajustaram a distribuição normal. Para isso, utilizou-se a família de transformações Box-Cox (BOX; COX, 1964), conforme descrito

---

<sup>1</sup> A programação utilizada no software R para o cálculo do teste multivariado de normalidade de Shapiro-Wilk, generalizado por Royston (ROYSTON, 1983), foi desenvolvida com auxílio do Professor Dr. Daniel Furtado Ferreira, do Departamento de Ciências Exatas da Universidade Federal de Lavras.

por: Draper e Smith (1981), Srivastava (2002), Kutner et al. (2004) e Fávero et al. (2009). A família de transformações é expressa pela seguinte equação:

$$X^{(\lambda)} = \begin{cases} \frac{X^{\lambda-1}}{\lambda}, & \text{para } \lambda \neq 0 \\ \ln X, & \text{para } \lambda = 0 \end{cases} \quad (4)$$

em que  $X$  é a variável que esta sendo transformada e  $\lambda$  é um parâmetro adicional e desconhecido do modelo. O objetivo é transformar a variável  $X$  em uma nova escala, mais próxima à distribuição normal, ou seja,  $X \sim N(0,1)$ .

Para a escolha da potência  $\lambda$ , foi estabelecida uma amplitude entre  $-5 \leq \lambda \leq 5$ . Os valores de  $\lambda$  desse intervalo foram inseridos na equação:

$$Z^{(\lambda)} = \begin{cases} \frac{X^{\lambda-1}}{\lambda X^{\lambda-1}}, & \text{para } \lambda \neq 0 \\ X \ln X, & \text{para } \lambda = 0 \end{cases} \quad (5)$$

em que se tem o jacobiano  $J(\lambda, z) = 1$ , com valores transformados ( $Z^{(\lambda)}$ ). Para encontrar o melhor  $\lambda$ , utilizou-se o estimador de máxima verossimilhança ( $L_{\max}$ ) que foi determinado através da equação:

$$L_{\max(\lambda)} = -\frac{n}{2} \ln \left( \frac{\hat{\sigma}_\lambda^2(z)}{n} \right) = -\frac{n}{2} \ln \left( \frac{SSE(\lambda)}{n} \right) \quad (6)$$

em que  $SSE(\lambda)$  é a soma de quadrados,  $n$  é o número de observações e  $\hat{\sigma}_\lambda^2(z)$  é a variância da variável ( $Z$ ) transformada pela potência  $\lambda$ . Assim, foi possível obter o melhor valor de  $\lambda$  que maximizasse o estimador de máxima verossimilhança ( $L_{\max}$ ) e minimizasse o resíduo. O processo descrito acima foi realizado com até 100 iterações para encontrar o melhor  $\lambda$  no intervalo  $[-5, 5]$ . As transformações dos dados foram realizadas com auxílio do software Statistica 7.0 (STATSOFT, 2005).

### 3.3.3.1 Critérios de transformação de dados

Nos ensaios que apresentaram distribuição normal multivariada ( $p > 0,10$ ), foi desconsiderado o valor-p da normalidade univariada. Ou seja, mesmo que alguma variável isolada não se ajustou a distribuição normal univariada ( $p \leq 0,10$ ), não se realizou transformação de dados, visto que, já se possuía a normalidade multivariada.

Para os demais ensaios, adotaram-se dois critérios de transformação dos dados de variáveis. Primeiro, procedeu-se a transformação dos dados das variáveis que não se ajustaram a distribuição normal univariada pelo teste de Shapiro-Wilk ( $p \leq 0,10$ ). Após a transformação dessas variáveis, testou-se a normalidade multivariada pelo teste de Shapiro-Wilk generalizado por Royston. Nos casos em que os ensaios passaram a apresentar a normalidade multivariada, não se realizou mais nenhum tipo de transformação dos dados.

Como muitos ensaios ainda não apresentavam distribuição normal multivariada, testou-se um segundo critério de transformação dos dados. Nesse critério, procedeu-se a transformação dos dados de todas as variáveis, independentemente do valor-p do teste univariado de Shapiro-Wilk. Após a transformação das variáveis desses ensaios, foi realizado o teste de normalidade univariada, com a função “*Shapiro.test()*” do software R. A seguir os valores-p da normalidade univariada antes e após a transformação dos dados foram comparados. Nos casos em que houve acréscimo do valor-p, ou seja, os dados apresentaram melhoria do ajuste em relação à distribuição normal univariada, prosseguiram-se as análises subsequentes com os dados transformados. Quando não se observou acréscimo do valor-p, mantiveram-se os dados originais nas análises subsequentes. Ao final, nesses ensaios, testou-se a normalidade multivariada pelo teste de Shapiro-Wilk generalizado por Royston.

### 3.4 Obtenção das matrizes de correlações fenotípicas

Após o diagnóstico de normalidade, em cada um dos 44 ensaios, foram calculados os coeficientes de correlação linear de Pearson para cada par das oito variáveis. Os valores de coeficientes de correlação de Pearson são idênticos aos coeficientes de correlação fenotípicos, pois foram obtidos a partir das médias de cada cultivar. O cálculo dos coeficientes de correlação foi realizado conforme descrito em Cruz e Regazzi (1997) por meio da equação:

$$r = \frac{C\hat{o}v(X,Y)}{\sqrt{\hat{V}(X)\hat{V}(Y)}} \quad (7)$$

em que:

$$C\hat{o}v(X,Y) = \frac{\sum_{i=1}^n x_i y_i}{n-1}; \quad \hat{V}(X) = \frac{\sum_{i=1}^n x_i^2}{n-1} \quad e \quad \hat{V}(Y) = \frac{\sum_{i=1}^n y_i^2}{n-1} \quad (8)$$

Sendo  $x_i = X_i - \bar{X}$  e  $y_i = Y_i - \bar{Y}$ .

Foram testadas as hipóteses  $H_0: r = 0$  (coeficiente de correlação entre as variáveis X e Y não difere de zero) *versus*  $H_1: r \neq 0$  (coeficiente de correlação entre as variáveis X e Y difere de zero). A significância das correlações foi testada por meio do teste t através da equação:

$$t = r \times \sqrt{\frac{n-2}{1-r^2}} \quad (9)$$

em que  $t$  está associado a  $n-2$  graus de liberdade, sendo  $n$  = número de cultivares em cada ensaio e o nível de significância  $\alpha$  de 0,05. Assim, quando  $|r| \geq r(\alpha = 0,05; n-2)$ , rejeitou-se  $H_0$  e concluiu-se que a correlação entre as variáveis X e Y diferiu de zero. Caso contrário, não rejeitou-se  $H_0$ .

Nos ensaios em que foi realizada a transformação de variáveis, foram calculados coeficientes de correlação antes e após a transformação dos dados. Após realizar-se o cálculo do coeficiente de correlação para cada par de variáveis, obtiveram-se as matrizes de coeficientes de correlação linear de Pearson (matrizes de correlações fenotípicas) simétricas com dimensão  $8 \times 8$  para cada ensaio. Cada matriz foi desmembrada em duas matrizes menores, sendo a matriz de correlação simétrica  $X'X$  referente às correlações entre as sete variáveis explicativas com dimensão  $7 \times 7$ , conforme detalhamento abaixo e no item 3.6. As correlações de cada variável explicativa com a produtividade de grãos foram alocadas na matriz de correlação  $X'Y$ , com dimensão  $7 \times 1$ . Assim, o diagnóstico de multicolinearidade foi realizado somente entre as variáveis explicativas (matriz de correlação  $X'X$ ) e posteriormente, foram estimados os coeficientes de trilha pelo sistema de equações normais  $X'X \hat{\beta} = X'Y$  conforme descrito no item 3.6.

Matriz de correlação X'X							Matriz de correlação X'Y
1	$r_{fm;ep}$	$r_{fm;aie}$	$r_{fm;pre}$	$r_{fm;np}$	$r_{fm;ne}$	$r_{fm;pli}$	$r_{fm;prod}$
$r_{ep;fm}$	1	$r_{ep;aie}$	$r_{ep;pre}$	$r_{ep;np}$	$r_{ep;ne}$	$r_{ep;pli}$	$r_{ep;prod}$
$r_{aie;fm}$	$r_{aie;ep}$	1	$r_{aie;pre}$	$r_{aie;np}$	$r_{aie;ne}$	$r_{aie;pli}$	$r_{aie;prod}$
$r_{pre;fm}$	$r_{pre;ep}$	$r_{pre;aie}$	1	$r_{pre;np}$	$r_{pre;ne}$	$r_{pre;pli}$	$r_{pre;prod}$
$r_{np;fm}$	$r_{np;ep}$	$r_{np;aie}$	$r_{np;pre}$	1	$r_{np;ne}$	$r_{np;pli}$	$r_{np;prod}$
$r_{ne;fm}$	$r_{ne;ep}$	$r_{ne;aie}$	$r_{ne;pre}$	$r_{ne;np}$	1	$r_{ne;pli}$	$r_{ne;prod}$
$r_{pli;fm}$	$r_{pli;ep}$	$r_{pli;aie}$	$r_{pli;pre}$	$r_{pli;np}$	$r_{pli;ne}$	1	$r_{pli;prod}$

### 3.5 Diagnóstico de multicolinearidade

O diagnóstico de multicolinearidade foi realizado em cada ensaio, com base nas variáveis explicativas (independentes). A produtividade de grãos (PROD) não foi considerada nessa etapa, pelo fato de constituir a variável principal (dependente). O diagnóstico de multicolinearidade foi realizado para três diferentes situações: 1) para as sete variáveis explicativas; 2) para as sete variáveis explicativas, após a adição de uma constante  $k$  na diagonal da matriz de correlação  $X'X$  e; 3) após a eliminação de variáveis explicativas altamente correlacionadas. O diagnóstico de multicolinearidade foi realizado mediante a utilização de quatro diferentes métodos:

#### 3.5.1 Fator de inflação da variância

O primeiro método de diagnóstico de multicolinearidade entre as variáveis explicativas, foi o fator de inflação de variância (FIV), obtido pelo inverso do índice de tolerância, conforme equação descrita por Fávero et al. (2009):

$$FIV = 1/\text{Tolerância} \quad (10)$$



Os valores de FIV para cada variável também podem ser obtidos 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). Assim, pode-se obter alternativamente o FIV para a variável FM, na linha 1 e coluna 1, o FIV da variável EP na linha 2 e coluna 2 e assim sucessivamente, até PLI na linha 7 e coluna 7 da inversa da matriz de correlação  $X'X$ . Considerou-se multicolinearidade severa quando os valores de fator de inflação de variância foram acima de 10 (GUJARATI, 2006).

### 3.5.2 Tolerância

A tolerância indica a proporção da variação de uma variável explicativa que independe das demais variáveis explicativas e foi calculada através da equação:

$$\text{Tolerância} = 1 - R_X^2 \quad (11)$$

em que:  $R_X^2$  é o coeficiente de ajuste da regressão (coeficiente de determinação do modelo) entre a variável explicativa X e as outras variáveis explicativas presentes (FÁVERO et al., 2009). Menores valores de tolerância indicaram maiores problemas de multicolinearidade entre as variáveis explicativas. Considerou-se multicolinearidade severa quando os valores de tolerância foram menores que 0,10 (FÁVERO et al., 2009; GUJARATI, 2006).

### 3.5.3 Número de condição

Foi verificado o grau de multicolinearidade na matriz de correlação  $X'X$  por meio do seu número de condição, conforme proposto por Montgomery e Peck (1982). As variáveis que apresentaram os maiores elementos nos autovetores associados aos menores autovalores foram as que mais contribuíram para a multicolinearidade (CRUZ; CARNEIRO, 2006). Assim, calculou-se a razão entre o maior e o menor autovalor da matriz de correlação  $X'X$ . Quando o número de condição (NC) resultante dessa divisão foi  $NC \leq 100$ , considerou-se

multicolinearidade fraca entre as variáveis explicativas,  $100 < NC < 1.000$ , multicolinearidade moderada a severa e  $NC \geq 1.000$ , multicolinearidade severa.

#### 3.5.4 Determinante da matriz de correlação

Examinou-se o determinante da matriz de correlação  $X'X$ , para verificar a intensidade de multicolinearidade. Quanto mais próximo de zero foi o determinante, concluiu-se que existia elevada associação linear entre as variáveis explicativas (multicolinearidade severa) e quanto mais próximo de um, concluiu-se que a multicolinearidade apresentava pouca intensidade (CRUZ; CARNEIRO, 2006).

Para o diagnóstico de multicolinearidade, foram utilizados os softwares Microsoft Office Excel e GENES (CRUZ, 2006).

Foi constatado que nos 44 ensaios houve problemas sérios de multicolinearidade entre as variáveis explicativas da matriz de correlação  $X'X$ . A partir dos diagnósticos de normalidade multivariada e de multicolinearidade, separaram-se os ensaios em três grupos (casos) e as análises posteriores foram realizadas dentro de cada caso.

#### **Caso 1 - Ensaios com distribuição normal multivariada e com multicolinearidade**

Para os ensaios do caso 1, não foi realizada transformação de variáveis, pois os mesmos apresentaram distribuição normal multivariada ( $p > 0,10$ ). Foram calculadas as matrizes de correlações fenotípicas  $X'X$  e  $X'Y$ , realizado o diagnóstico de multicolinearidade na matriz de correlação  $X'X$  e a análise de trilha mediante três formas, conforme descrição no item 3.6.

#### **Caso 2 - Ensaios com distribuição normal multivariada após a transformação de dados e com multicolinearidade**

Nesses ensaios que apresentaram distribuição normal após a transformação de dados, foram calculadas as matrizes de correlações fenotípicas  $X'X$  e  $X'Y$ , realizado o diagnóstico de multicolinearidade na matriz de correlação  $X'X$  e a análise de trilha sob três formas, conforme descrito no item 3.6. Essas análises estatísticas foram realizadas tanto antes quanto após a transformação dos dados, a fim de verificar a interferência da não-normalidade multivariada e da multicolinearidade na análise de trilha.

### **Caso 3 - Ensaios sem distribuição normal multivariada e com multicolinearidade**

Para os ensaios que não apresentaram distribuição normal multivariada, mesmo após a transformação de dados, foram calculadas as matrizes de correlações fenotípicas  $X'X$  e  $X'Y$ , realizado o diagnóstico de multicolinearidade na matriz de correlação  $X'X$  e a análise de trilha sob três formas, conforme descrito no item 3.6, tanto antes quanto após a transformação de dados de variáveis. Embora a pressuposição de normalidade multivariada não tenha sido atendida nesses ensaios, prosseguiu-se a realização das análises estatísticas a fim de verificar a interferência da transformação de dados e da multicolinearidade sobre as análises e sobre os resultados.

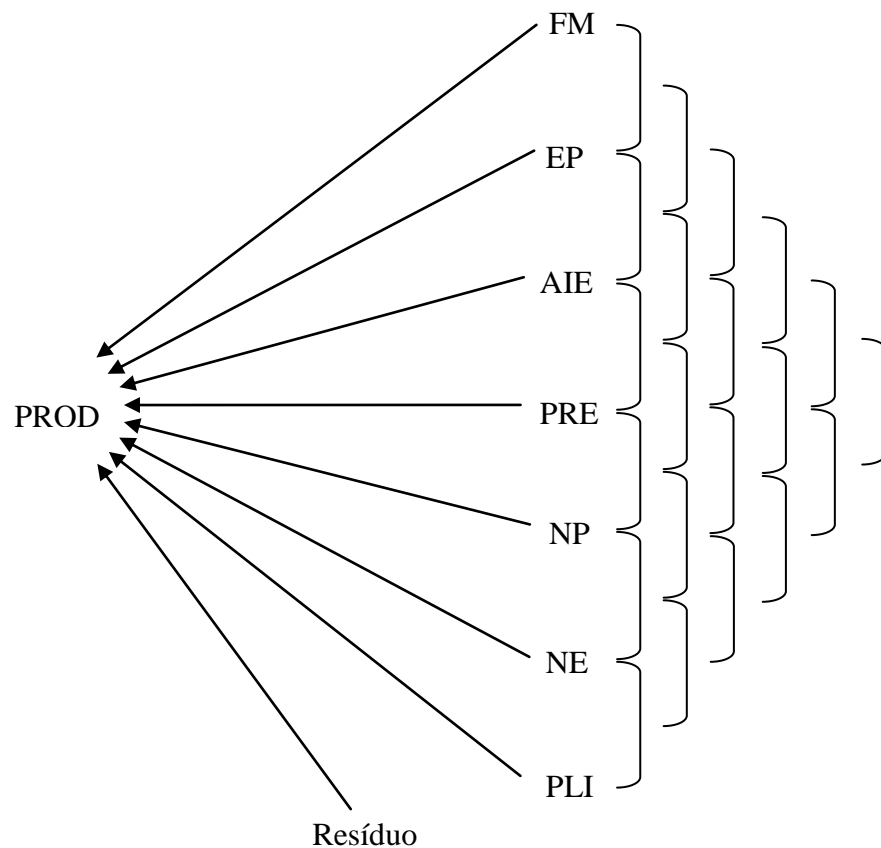
### **3.6 Análise de trilha**

Dentro de cada caso, em cada ensaio, foram realizadas análises de trilha (*path analysis*) conforme metodologia descrita por Cruz e Regazzi (1997) e Cruz e Carneiro (2006), utilizando os softwares Microsoft Office Excel e GENES (CRUZ, 2006). As variáveis explicativas foram: FM, EP, AIE, PRE, NP, NE e PLI. A variável básica ou principal foi produtividade de grãos (PROD).

Procedeu-se a estimação dos efeitos diretos e indiretos das variáveis explicativas sobre a produtividade de grãos, mediante três formas de análise de trilha. Os efeitos indiretos foram calculados em todos os métodos, mas não estão apresentados nesse trabalho.

### 3.6.1 Análise de trilha tradicional

A primeira forma de estimação dos efeitos diretos e indiretos foi realizada com base na análise de trilha tradicional, ignorando os efeitos adversos da multicolinearidade (CRUZ; CARNEIRO, 2006). Assim, calcularam-se os efeitos diretos e indiretos das sete variáveis explicativas (FM, EP, AIE, PRE, NP, NE e PLI) sobre a produtividade de grãos, o coeficiente de determinação da análise de trilha e o efeito da variável residual. Para o cálculo dos efeitos diretos e indiretos utilizou-se o diagrama causal com uma única cadeia (Figura 4), evidenciando os efeitos das variáveis explicativas (FM, EP, AIE, PRE, NP, NE e PLI) e da variável residual (Resíduo) sobre a produtividade de grãos (PROD) e as inter-relações entre as variáveis explicativas.



**Figura 4** - Diagrama causal com uma única cadeia, com os efeitos diretos das variáveis explicativas (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 e PLI: prolificidade) e da variável residual (Resíduo) sobre a produtividade de grãos em milho (PROD) e as inter-relações entre as variáveis explicativas na análise de trilha tradicional.

Inicialmente foi realizada a padronização das variáveis e estabelecido o modelo:

$$\text{PROD} = \hat{\beta}_1\text{FM} + \hat{\beta}_2\text{EP} + \hat{\beta}_3\text{AIE} + \hat{\beta}_4\text{PRE} + \hat{\beta}_5\text{NP} + \hat{\beta}_6\text{NE} + \hat{\beta}_7\text{PLI} + \text{Resíduo} \quad (12)$$

Sendo  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$  e  $\hat{\beta}_7$ , os estimadores dos efeitos diretos das variáveis FM, EP, AIE, PRE, NP, NE e PLI, respectivamente, sobre a produtividade de grãos. Assim, foi utilizado o sistema de equações normais  $X'X \hat{\beta} = X'Y$  para determinar os efeitos diretos e indiretos de cada variável sobre a produtividade, pelo método dos mínimos quadrados, sendo a matriz de correlação  $X'X$  obtida com base nos coeficientes de correlação de Pearson (fenotípicos) entre as sete variáveis explicativas:

$$X'X = \begin{bmatrix} 1 & r_{fm;ep} & r_{fm;aie} & r_{fm;pre} & r_{fm;np} & r_{fm;ne} & r_{fm;pli} \\ r_{ep;fm} & 1 & r_{ep;aie} & r_{ep;pre} & r_{ep;np} & r_{ep;ne} & r_{ep;pli} \\ r_{aie;fm} & r_{aie;ep} & 1 & r_{aie;pre} & r_{aie;np} & r_{aie;ne} & r_{aie;pli} \\ r_{pre;fm} & r_{pre;ep} & r_{pre;aie} & 1 & r_{pre;np} & r_{pre;ne} & r_{pre;pli} \\ r_{np;fm} & r_{np;ep} & r_{np;aie} & r_{np;pre} & 1 & r_{np;ne} & r_{np;pli} \\ r_{ne;fm} & r_{ne;ep} & r_{ne;aie} & r_{ne;pre} & r_{ne;np} & 1 & r_{ne;pli} \\ r_{pli;fm} & r_{pli;ep} & r_{pli;aie} & r_{pli;pre} & r_{pli;np} & r_{pli;ne} & 1 \end{bmatrix}$$

A matriz  $X'Y$  foi obtida com base nas correlações das variáveis explicativas com a produtividade de grãos, conforme já descrito no item 3.4:

$$X'Y = \begin{bmatrix} r_{fm;prod} \\ r_{ep;prod} \\ r_{aie;prod} \\ r_{pre;prod} \\ r_{np;prod} \\ r_{ne;prod} \\ r_{pli;prod} \end{bmatrix}$$

Por fim, os estimadores de efeitos diretos  $\hat{\beta}$ , foram dispostos em um vetor:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \\ \hat{\beta}_7 \end{bmatrix}$$

Com isso, o sistema de equações normais ficou estabelecido por  $X'X \hat{\beta} = X'Y$ :

$$\begin{bmatrix} 1 & \Gamma_{fm;ep} & \Gamma_{fm;aie} & \Gamma_{fm;pre} & \Gamma_{fm;np} & \Gamma_{fm;ne} & \Gamma_{fm;pli} \\ \Gamma_{ep;fm} & 1 & \Gamma_{ep;aie} & \Gamma_{ep;pre} & \Gamma_{ep;np} & \Gamma_{ep;ne} & \Gamma_{ep;pli} \\ \Gamma_{aie;fm} & \Gamma_{aie;ep} & 1 & \Gamma_{aie;pre} & \Gamma_{aie;np} & \Gamma_{aie;ne} & \Gamma_{aie;pli} \\ \Gamma_{pre;fm} & \Gamma_{pre;ep} & \Gamma_{pre;aie} & 1 & \Gamma_{pre;np} & \Gamma_{pre;ne} & \Gamma_{pre;pli} \\ \Gamma_{np;fm} & \Gamma_{np;ep} & \Gamma_{np;aie} & \Gamma_{np;pre} & 1 & \Gamma_{np;ne} & \Gamma_{np;pli} \\ \Gamma_{ne;fm} & \Gamma_{ne;ep} & \Gamma_{ne;aie} & \Gamma_{ne;pre} & \Gamma_{ne;np} & 1 & \Gamma_{ne;pli} \\ \Gamma_{pli;fm} & \Gamma_{pli;ep} & \Gamma_{pli;aie} & \Gamma_{pli;pre} & \Gamma_{pli;np} & \Gamma_{pli;ne} & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \\ \hat{\beta}_7 \end{bmatrix} = \begin{bmatrix} \Gamma_{fm;prod} \\ \Gamma_{ep;prod} \\ \Gamma_{aie;prod} \\ \Gamma_{pre;prod} \\ \Gamma_{np;prod} \\ \Gamma_{ne;prod} \\ \Gamma_{pli;prod} \end{bmatrix}$$

Posteriormente, foi calculada a inversa da matriz de correlação  $X'X$ . Assim, o modelo passou a ser:  $\hat{\beta} = X'X^{-1} \times X'Y$  e, dessa forma, foi possível calcular os efeitos diretos e indiretos de cada variável sobre a produtividade de grãos, o coeficiente de determinação da análise de trilha e o efeito da variável residual. Para calcular os efeitos diretos de cada variável explicativa sobre a produtividade de grãos foram utilizadas as equações descritas na Tabela 2. Os efeitos indiretos de cada variável sobre a produtividade de grãos, via outras variáveis explicativas, foram calculados com base nas equações apresentadas na Tabela 3.

**Tabela 2** - Equações para a estimação dos efeitos diretos das sete variáveis explicativas (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 e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional.

Efeito direto	Estimador dos efeitos diretos
FM	$\hat{\beta}_1 = (r_{fm;fm}^{(1)} \times r_{fm;prod}^{(2)}) + (r_{fm;ep} \times r_{ep;prod}) + (r_{fm;aie} \times r_{aie;prod}) + (r_{fm;pre} \times r_{pre;prod}) + (r_{fm;np} \times r_{np;prod}) + (r_{fm;ne} \times r_{ne;prod}) + (r_{fm;pli} \times r_{pli;prod})$
EP	$\hat{\beta}_2 = (r_{ep;fm} \times r_{fm;prod}) + (r_{ep;ep} \times r_{ep;prod}) + (r_{ep;aie} \times r_{aie;prod}) + (r_{ep;pre} \times r_{pre;prod}) + (r_{ep;np} \times r_{np;prod}) + (r_{ep;ne} \times r_{ne;prod}) + (r_{ep;pli} \times r_{pli;prod})$
AIE	$\hat{\beta}_3 = (r_{aie;fm} \times r_{fm;prod}) + (r_{aie;ep} \times r_{ep;prod}) + (r_{aie;aie} \times r_{aie;prod}) + (r_{aie;pre} \times r_{pre;prod}) + (r_{aie;np} \times r_{np;prod}) + (r_{aie;ne} \times r_{ne;prod}) + (r_{aie;pli} \times r_{pli;prod})$
PRE	$\hat{\beta}_4 = (r_{pre;fm} \times r_{fm;prod}) + (r_{pre;ep} \times r_{ep;prod}) + (r_{pre;aie} \times r_{aie;prod}) + (r_{pre;pre} \times r_{pre;prod}) + (r_{pre;np} \times r_{np;prod}) + (r_{pre;ne} \times r_{ne;prod}) + (r_{pre;pli} \times r_{pli;prod})$
NP	$\hat{\beta}_5 = (r_{np;fm} \times r_{fm;prod}) + (r_{np;ep} \times r_{ep;prod}) + (r_{np;aie} \times r_{aie;prod}) + (r_{np;pre} \times r_{pre;prod}) + (r_{np;np} \times r_{np;prod}) + (r_{np;ne} \times r_{ne;prod}) + (r_{np;pli} \times r_{pli;prod})$
NE	$\hat{\beta}_6 = (r_{ne;fm} \times r_{fm;prod}) + (r_{ne;ep} \times r_{ep;prod}) + (r_{ne;aie} \times r_{aie;prod}) + (r_{ne;pre} \times r_{pre;prod}) + (r_{ne;np} \times r_{np;prod}) + (r_{ne;ne} \times r_{ne;prod}) + (r_{ne;pli} \times r_{pli;prod})$
PLI	$\hat{\beta}_7 = (r_{pli;fm} \times r_{fm;prod}) + (r_{pli;ep} \times r_{ep;prod}) + (r_{pli;aie} \times r_{aie;prod}) + (r_{pli;pre} \times r_{pre;prod}) + (r_{pli;np} \times r_{np;prod}) + (r_{pli;ne} \times r_{ne;prod}) + (r_{pli;pli} \times r_{pli;prod})$

<sup>(1)</sup> Em cada multiplicação, entre parênteses, o primeiro coeficiente é obtido da inversa da matriz de correlação  $X'X^{-1}$ . <sup>(2)</sup> O segundo coeficiente de cada multiplicação, entre parênteses, é obtido da matriz de correlação  $X'Y$ .

**Tabela 3** - Equações para a estimação dos efeitos indiretos das sete variáveis explicativas (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 e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional.

Efeito Indireto	Variável com efeito direto sobre a produtividade de grãos						
	FM	EP	AIE	PRE	NP	NE	PLI
Via FM	-	$\hat{\beta}_1^{(1)} \times r_{fm;ep}^{(2)}$	$\hat{\beta}_1 \times r_{fm;aie}$	$\hat{\beta}_1 \times r_{fm;pre}$	$\hat{\beta}_1 \times r_{fm;np}$	$\hat{\beta}_1 \times r_{fm;ne}$	$\hat{\beta}_1 \times r_{fm;pli}$
Via EP	$\hat{\beta}_2 \times r_{ep;fm}$	-	$\hat{\beta}_2 \times r_{ep;aie}$	$\hat{\beta}_2 \times r_{ep;pre}$	$\hat{\beta}_2 \times r_{ep;np}$	$\hat{\beta}_2 \times r_{ep;ne}$	$\hat{\beta}_2 \times r_{ep;pli}$
Via AIE	$\hat{\beta}_3 \times r_{aie;fm}$	$\hat{\beta}_3 \times r_{aie;ep}$	-	$\hat{\beta}_3 \times r_{aie;pre}$	$\hat{\beta}_3 \times r_{aie;np}$	$\hat{\beta}_3 \times r_{aie;ne}$	$\hat{\beta}_3 \times r_{aie;pli}$
Via PRE	$\hat{\beta}_4 \times r_{pre;fm}$	$\hat{\beta}_4 \times r_{pre;ep}$	$\hat{\beta}_4 \times r_{pre;aie}$	-	$\hat{\beta}_4 \times r_{pre;np}$	$\hat{\beta}_4 \times r_{pre;ne}$	$\hat{\beta}_4 \times r_{pre;pli}$
Via NP	$\hat{\beta}_5 \times r_{np;fm}$	$\hat{\beta}_5 \times r_{np;ep}$	$\hat{\beta}_5 \times r_{np;aie}$	$\hat{\beta}_5 \times r_{np;pre}$	-	$\hat{\beta}_5 \times r_{np;ne}$	$\hat{\beta}_5 \times r_{np;pli}$
Via NE	$\hat{\beta}_6 \times r_{ne;fm}$	$\hat{\beta}_6 \times r_{ne;ep}$	$\hat{\beta}_6 \times r_{ne;aie}$	$\hat{\beta}_6 \times r_{ne;pre}$	$\hat{\beta}_6 \times r_{ne;np}$	-	$\hat{\beta}_6 \times r_{ne;pli}$
Via PLI	$\hat{\beta}_7 \times r_{pli;fm}$	$\hat{\beta}_7 \times r_{pli;ep}$	$\hat{\beta}_7 \times r_{pli;aie}$	$\hat{\beta}_7 \times r_{pli;pre}$	$\hat{\beta}_7 \times r_{pli;np}$	$\hat{\beta}_7 \times r_{pli;ne}$	-

<sup>(1)</sup> Os coeficiente  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$  e  $\hat{\beta}_7$ , são os estimadores dos efeitos diretos das variáveis FM, EP, AIE, PRE, NP, NE e PLI sobre a produtividade de grãos, respectivamente. <sup>(2)</sup> Os coeficientes de correlação entre as variáveis explicativas são obtidos na matriz de correlação  $X'X$ .

O coeficiente de determinação da análise de trilha foi obtido através do somatório dos estimadores  $\hat{\beta}$  multiplicados pelos respectivos coeficientes de correlação da matriz  $X'Y$ , conforme a equação:

$$R^2 = \hat{\beta}_1 \times r_{fm;prod} + \hat{\beta}_2 \times r_{ep;prod} + \hat{\beta}_3 \times r_{aie;prod} + \hat{\beta}_4 \times r_{pre;prod} + \hat{\beta}_5 \times r_{np;prod} + \hat{\beta}_6 \times r_{ne;prod} + \hat{\beta}_7 \times r_{pli;prod} \quad (13)$$

O efeito da variável residual da análise de trilha foi obtido por meio da equação:

$$\hat{\beta}_{RES} = \sqrt{1-R^2} \quad (14)$$

### 3.6.2 Análise de trilha sob multicolinearidade

O segundo modo de estimação dos efeitos diretos e indiretos foi realizado com base na análise de trilha sob multicolinearidade, com procedimento semelhante à análise de regressão em crista (CARVALHO, 1995; CARVALHO; CRUZ, 1996; CRUZ; CARNEIRO, 2006). Assim, mantiveram-se as sete variáveis explicativas (FM, EP, AIE, PRE, NP, NE e PLI) na estimativa dos efeitos diretos e indiretos sobre a produtividade de grãos, com o mesmo diagrama causal da análise de trilha tradicional (Figura 4). A única alteração nessa forma de análise foi a introdução de uma constante  $k$  na diagonal da matriz de correlação  $X'X$ , a fim de reduzir a elevada variância associada ao estimador de mínimos quadrados da análise de trilha (CARVALHO, 1995; CARVALHO; CRUZ, 1996).

Dessa forma, o sistema de equações normais  $X'X \hat{\beta} = X'Y$  passou a ser  $(X'X + k) \hat{\beta} = X'Y$ . Foram testados 21 valores da constante  $k$  ( $k = 0,00; 0,05; 0,10; \dots; 1,00$ ). Conforme preconizado por Carvalho e Cruz (1996) foi atribuído o menor valor de  $k$ , a partir do qual, os coeficientes de trilha estavam estabilizados. Nos 44 ensaios, após a atribuição de uma constante  $k = 0,10$ , observou-se a estabilização dos efeitos diretos e indiretos, sendo esse valor da constante atribuído a todas as análises de trilha sob multicolinearidade. Assim, o sistema de equações normais  $(X'X + k) \hat{\beta} = X'Y$  nos 44 ensaios foi:



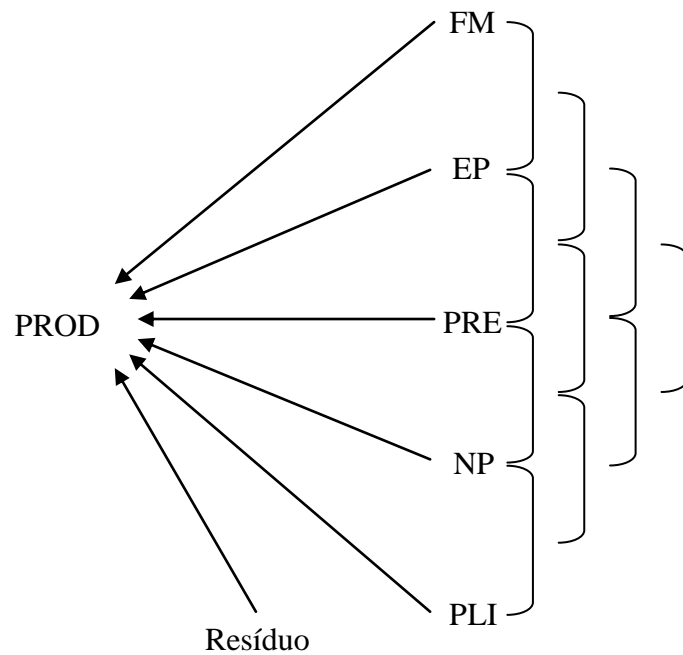
$$\begin{bmatrix}
 1,10 & r_{fm;ep} & r_{fm;aie} & r_{fm;pre} & r_{fm;np} & r_{fm;ne} & r_{fm;pli} \\
 r_{ep;fm} & 1,10 & r_{ep;aie} & r_{ep;pre} & r_{ep;np} & r_{ep;ne} & r_{ep;pli} \\
 r_{aie;fm} & r_{aie;ep} & 1,10 & r_{aie;pre} & r_{aie;np} & r_{aie;ne} & r_{aie;pli} \\
 r_{pre;fm} & r_{pre;ep} & r_{pre;aie} & 1,10 & r_{pre;np} & r_{pre;ne} & r_{pre;pli} \\
 r_{np;fm} & r_{np;ep} & r_{np;aie} & r_{np;pre} & 1,10 & r_{np;ne} & r_{np;pli} \\
 r_{ne;fm} & r_{ne;ep} & r_{ne;aie} & r_{ne;pre} & r_{ne;np} & 1,10 & r_{ne;pli} \\
 r_{pli;fm} & r_{pli;ep} & r_{pli;aie} & r_{pli;pre} & r_{pli;np} & r_{pli;ne} & 1,10
 \end{bmatrix}
 \times
 \begin{bmatrix}
 \hat{\beta}_1 \\
 \hat{\beta}_2 \\
 \hat{\beta}_3 \\
 \hat{\beta}_4 \\
 \hat{\beta}_5 \\
 \hat{\beta}_6 \\
 \hat{\beta}_7
 \end{bmatrix}
 =
 \begin{bmatrix}
 r_{fm;prod} \\
 r_{ep;prod} \\
 r_{aie;prod} \\
 r_{pre;prod} \\
 r_{np;prod} \\
 r_{ne;prod} \\
 r_{pli;prod}
 \end{bmatrix}$$

Os efeitos diretos e indiretos, o coeficiente de determinação da análise de trilha e o efeito da variável residual, em cada ensaio, foram calculados da mesma forma que para a análise de trilha tradicional (Tabelas 2 e 3 e equações 13 e 14, respectivamente). Ainda, procedeu-se o diagnóstico da multicolinearidade por meio dos métodos descritos no item 3.5, a fim de verificar o efeito da adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , sobre a multicolinearidade.

### 3.6.3 Análise de trilha tradicional com eliminação de variáveis

A terceira forma de estimação dos efeitos diretos e indiretos foi realizada com base na análise de trilha tradicional, após a eliminação de variáveis (CRUZ; CARNEIRO, 2006). Constatou-se que existia alta associação linear entre as variáveis EP, AIE e PRE e entre as variáveis NP, NE e PLI. Foram eliminadas variáveis do grupo de sete variáveis explicativas e realizado o diagnóstico de multicolinearidade por meio dos métodos descritos no item 3.5.

Devido à existência de dois subgrupos com variáveis altamente correlacionadas dentro do grupo de sete variáveis explicativas, foi necessária a eliminação de duas variáveis para a realização da análise de trilha. Dessa forma, para os 44 ensaios, a única situação em que a multicolinearidade foi fraca, foi no caso da eliminação das variáveis AIE (altamente associada à EP e PRE) e NE (altamente associada à NP e PLI). Assim, calcularam-se os efeitos diretos e indiretos de cinco variáveis explicativas (FM, EP, PRE, NP e PLI) sobre a produtividade de grãos, o coeficiente de determinação da análise de trilha e o efeito da variável residual. O diagrama causal também foi em única cadeia (Figura 5).



**Figura 5** - Diagrama causal com uma única cadeia, com os efeitos diretos das variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) e da variável residual (Resíduo) sobre a produtividade de grãos em milho (PROD) e as inter-relações entre as variáveis explicativas na análise de trilha tradicional com eliminação de variáveis.

Foi estabelecido o modelo:

$$\text{PROD} = \hat{\beta}_1 \text{FM} + \hat{\beta}_2 \text{EP} + \hat{\beta}_3 \text{PRE} + \hat{\beta}_4 \text{NP} + \hat{\beta}_5 \text{PLI} + \text{Resíduo} \quad (15)$$

Sendo  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ ,  $\hat{\beta}_3$ ,  $\hat{\beta}_4$  e  $\hat{\beta}_5$ , 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  $X'X \hat{\beta} = X'Y$ , sendo a matriz de correlação  $X'X$  com base nos coeficientes de correlação de Pearson (fenotípicos) entre as cinco variáveis explicativas:

$$X'X = \begin{bmatrix} 1 & r_{fm;ep} & r_{fm;pre} & r_{fm;np} & r_{fm;pli} \\ r_{ep;fm} & 1 & r_{ep;pre} & r_{ep;np} & r_{ep;pli} \\ r_{pre;fm} & r_{pre;ep} & 1 & r_{pre;np} & r_{pre;pli} \\ r_{np;fm} & r_{np;ep} & r_{np;pre} & 1 & r_{np;pli} \\ r_{pli;fm} & r_{pli;ep} & r_{pli;pre} & r_{pli;np} & 1 \end{bmatrix}$$

A matriz de correlação  $X'Y$  foi obtida a partir das correlações das variáveis explicativas com a produtividade de grãos:

$$X'Y = \begin{bmatrix} \Gamma_{fm;prod} \\ \Gamma_{ep;prod} \\ \Gamma_{pre;prod} \\ \Gamma_{np;prod} \\ \Gamma_{pli;prod} \end{bmatrix}$$

Os estimadores  $\hat{\beta}$  dos efeitos diretos foram dispostos em um vetor:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix}$$

Com isso, o sistema de equações normais ficou estabelecido por  $X'X \hat{\beta} = X'Y$ :

$$\begin{bmatrix} 1 & \Gamma_{fm;ep} & \Gamma_{fm;pre} & \Gamma_{fm;np} & \Gamma_{fm;pli} \\ \Gamma_{ep;fm} & 1 & \Gamma_{ep;pre} & \Gamma_{ep;np} & \Gamma_{ep;pli} \\ \Gamma_{pre;fm} & \Gamma_{pre;ep} & 1 & \Gamma_{pre;np} & \Gamma_{pre;pli} \\ \Gamma_{np;fm} & \Gamma_{np;ep} & \Gamma_{np;pre} & 1 & \Gamma_{np;pli} \\ \Gamma_{pli;fm} & \Gamma_{pli;ep} & \Gamma_{pli;pre} & \Gamma_{pli;np} & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix} = \begin{bmatrix} \Gamma_{fm;prod} \\ \Gamma_{ep;prod} \\ \Gamma_{pre;prod} \\ \Gamma_{np;prod} \\ \Gamma_{pli;prod} \end{bmatrix}$$

Posteriormente, foi calculada a inversa da matriz de correlação  $X'X$  e o modelo passou a ser:  $\hat{\beta} = X'X^{-1} \times X'Y$ . Para calcular os efeitos diretos, de cada variável sobre a

produtividade de grãos, foram utilizadas as equações descritas na Tabela 4. Os efeitos indiretos de cada variável sobre a produtividade de grãos, via outras variáveis explicativas, foram calculados com base nas equações apresentadas na Tabela 5.

**Tabela 4** - Equações para a estimação dos efeitos diretos das cinco variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional, com eliminação de variáveis.

Efeito direto	Estimador dos efeitos diretos
FM	$\hat{\beta}_1 = (r_{fm;fm}^{(1)} \times r_{fm;prod}^{(2)}) + (r_{fm;ep} \times r_{ep;prod}) + (r_{fm;pre} \times r_{pre;prod}) + (r_{fm;np} \times r_{np;prod}) + (r_{fm;pli} \times r_{pli;prod})$
EP	$\hat{\beta}_2 = (r_{ep;fm} \times r_{fm;prod}) + (r_{ep;ep} \times r_{ep;prod}) + (r_{ep;pre} \times r_{pre;prod}) + (r_{ep;np} \times r_{np;prod}) + (r_{ep;pli} \times r_{pli;prod})$
PRE	$\hat{\beta}_3 = (r_{pre;fm} \times r_{fm;prod}) + (r_{pre;ep} \times r_{ep;prod}) + (r_{pre;pre} \times r_{pre;prod}) + (r_{pre;np} \times r_{np;prod}) + (r_{pre;pli} \times r_{pli;prod})$
NP	$\hat{\beta}_4 = (r_{np;fm} \times r_{fm;prod}) + (r_{np;ep} \times r_{ep;prod}) + (r_{np;pre} \times r_{pre;prod}) + (r_{np;np} \times r_{np;prod}) + (r_{np;pli} \times r_{pli;prod})$
PLI	$\hat{\beta}_5 = (r_{pli;fm} \times r_{fm;prod}) + (r_{pli;ep} \times r_{ep;prod}) + (r_{pli;pre} \times r_{pre;prod}) + (r_{pli;np} \times r_{np;prod}) + (r_{pli;pli} \times r_{pli;prod})$

<sup>(1)</sup> Em cada multiplicação, entre parênteses, o primeiro coeficiente é obtido da inversa da matriz de correlação  $X'X^{-1}$ . <sup>(2)</sup> O segundo coeficiente de cada multiplicação, entre parênteses, é obtido da matriz de correlação  $X'Y$ .

**Tabela 5** - Equações para a estimação dos efeitos indiretos das cinco variáveis explicativas (FM: florescimento masculino, EP: estatura de plantas, PRE: posição relativa da espiga, NP: número de plantas e PLI: prolificidade) sobre a produtividade de grãos (PROD), na análise de trilha tradicional, com eliminação de variáveis.

Efeito	Variável com efeito direto sobre a produtividade de grãos				
	FM	EP	PRE	NP	PLI
Via FM	-	$\hat{\beta}_1^{(1)} \times r_{fm;ep}^{(2)}$	$\hat{\beta}_1 \times r_{fm;pre}$	$\hat{\beta}_1 \times r_{fm;np}$	$\hat{\beta}_1 \times r_{fm;pli}$
Via EP	$\hat{\beta}_2 \times r_{ep;fm}$	-	$\hat{\beta}_2 \times r_{ep;pre}$	$\hat{\beta}_2 \times r_{ep;np}$	$\hat{\beta}_2 \times r_{ep;pli}$
Via PRE	$\hat{\beta}_3 \times r_{pre;fm}$	$\hat{\beta}_3 \times r_{pre;ep}$	-	$\hat{\beta}_3 \times r_{pre;np}$	$\hat{\beta}_3 \times r_{pre;pli}$
Via NP	$\hat{\beta}_4 \times r_{np;fm}$	$\hat{\beta}_4 \times r_{np;ep}$	$\hat{\beta}_4 \times r_{np;pre}$	-	$\hat{\beta}_4 \times r_{np;pli}$
Via PLI	$\hat{\beta}_5 \times r_{pli;fm}$	$\hat{\beta}_5 \times r_{pli;ep}$	$\hat{\beta}_5 \times r_{pli;pre}$	$\hat{\beta}_5 \times r_{pli;np}$	-

<sup>(1)</sup> Os coeficiente  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$ , são os estimadores dos efeitos diretos das variáveis FM, EP, PRE, NP e PLI sobre a produtividade de grãos, respectivamente. <sup>(2)</sup> Os coeficientes de correlação entre as variáveis explicativas são obtidos na matriz de correlação  $X'X$ .

O coeficiente de determinação da análise de trilha foi obtido através do somatório dos estimadores  $\hat{\beta}$  multiplicados pelos respectivos coeficientes de correlação da matriz  $X'Y$ , conforme a equação:

$$R^2 = \hat{\beta}_1 \times r_{fm;prod} + \hat{\beta}_2 \times r_{ep;prod} + \hat{\beta}_3 \times r_{pre;prod} + \hat{\beta}_4 \times r_{np;prod} + \hat{\beta}_5 \times r_{pli;prod} \quad (16)$$

O efeito da variável residual da análise de trilha foi obtido por meio da equação:

$$\hat{\beta}_{RES} = \sqrt{1-R^2} \quad (17)$$

## 4 RESULTADOS E DISCUSSÃO

Entre os ensaios, em média, foram necessários 71,7 dias desde a emergência das plantas até o florescimento masculino (Tabela 6). Os valores mínimos observados em cada ensaio variaram de 55 a 83 dias e os valores máximos entre 59 e 94,3 dias. O ensaio 12, classificado como superprecoce, apresentou a menor média para o florescimento masculino, com 57,1 dias. Já o ensaio 20, de ciclo precoce, apresentou a maior média, necessitando em média, 86,8 dias desde a emergência até o florescimento masculino. Os valores de desvio-padrão oscilaram entre 0,77 e 5,16 dias e os valores de coeficiente de variação oscilaram entre 1,07 e 7,66%. Como as cultivares são agrupadas em ensaios de ciclo precoce e superprecoce é esperado que ocorra pequena variabilidade entre cultivares, dentro de um ensaio, em relação ao período da emergência das plantas até o florescimento masculino.

A média geral dos ensaios para a estatura de plantas na colheita foi de 207,3 cm (Tabela 6). As médias de cada ensaio variaram entre 164,5 e 272,4 cm, para os ensaios 31 e 8, respectivamente. Considerando os extremos, a menor estatura de plantas foi mensurada em uma cultivar do ensaio 10, com 137 cm e, a maior estatura de plantas, foi mensurada em uma cultivar do ensaio 8, com 294,7 cm. Assim, a estatura de plantas da cultivar mais alta foi 2,15 vezes maior em relação a cultivar mais baixa. O desvio-padrão médio dos 44 ensaios foi de 12,47 cm e, o coeficiente de variação médio, foi de 6,06%. Apenas o ensaio 2 apresentou CV superior a 10% (10,67%).

A média de altura de inserção da espiga dos 44 ensaios foi de 113,9 cm e as médias das cultivares, em cada ensaio, oscilaram entre 73,4 e 143,9 cm, respectivamente, para os ensaios 31 e 8 (Tabela 7). Esses ensaios também apresentaram as menores e maiores médias de estatura de plantas, respectivamente (Tabela 6). Em relação aos valores extremos de altura de inserção da espigas, uma cultivar do ensaio 10 apresentou 47 cm e uma cultivar do ensaio 8 teve 170 cm. Dessa forma, a amplitude máxima entre cultivares de diferentes ensaios foi de 123 cm, para a altura de inserção da espigas.

A média da posição relativa da espiga foi de 0,55, ou seja, a altura de inserção da espiga corresponde a 55% da estatura de plantas, em média (Tabela 7). Os valores mínimos e máximos foram de 0,34 e 0,72, respectivamente, para cultivares dos ensaios 10 e 30, e a média dos ensaios oscilou entre 0,40 e 0,63. Assim, pode-se considerar que em média, a altura de inserção da espigas varia de 40 a 63% da estatura de plantas.

**Tabela 6** - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: florescimento masculino (em dias) e estatura de plantas (em cm) mensuradas em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	----- Florescimento Masculino -----					----- Estatura de Plantas -----				
	Mínimo	Média	Máximo	DP	CV	Mínimo	Média	Máximo	DP	CV
1	72,0	77,3	83,0	3,20	4,14	183,3	201,1	228,3	13,02	6,47
2	60,0	62,5	67,0	1,92	3,07	150,0	181,8	215,0	19,40	10,67
3	55,7	58,2	60,3	1,62	2,78	247,7	261,6	290,3	12,65	4,84
4	78,0	82,5	84,0	1,92	2,33	180,0	194,5	210,0	10,36	5,33
5	61,0	67,4	75,0	5,16	7,66	170,0	182,7	200,0	9,86	5,40
6	64,0	75,9	82,0	3,86	5,09	171,7	217,4	246,7	15,90	7,31
7	57,0	62,6	68,0	3,78	6,04	145,0	183,3	215,0	16,43	8,96
8	57,0	60,3	63,3	1,47	2,44	239,0	272,4	294,7	13,39	4,92
9	75,0	82,4	86,0	2,75	3,34	170,0	199,6	230,0	16,23	8,13
10	62,0	68,6	75,0	4,16	6,06	137,0	181,0	205,0	14,20	7,85
11	67,0	74,0	77,0	3,07	4,15	181,7	204,1	216,7	9,15	4,48
12	55,0	57,1	59,0	1,36	2,38	231,0	248,5	263,3	10,38	4,18
13	70,7	71,8	73,7	0,77	1,07	213,3	225,4	236,3	8,09	3,59
14	67,0	76,3	81,0	2,72	3,56	186,7	214,1	240,0	13,15	6,14
15	56,3	59,5	62,3	1,60	2,69	238,7	260,6	294,3	15,88	6,09
16	70,7	75,1	83,0	2,90	3,86	197,7	224,1	262,3	15,42	6,88
17	77,0	81,2	85,0	2,43	2,99	213,3	232,2	270,0	17,52	7,55
18	61,3	64,2	69,7	2,84	4,42	196,7	214,6	240,0	12,52	5,83
19	66,0	70,2	72,0	2,05	2,92	169,7	187,5	217,7	14,28	7,62
20	82,0	86,8	93,0	2,58	2,97	215,0	234,6	256,7	11,71	4,99
21	59,0	64,4	70,0	2,48	3,85	196,7	229,5	260,0	12,84	5,59
22	65,0	70,1	75,7	3,15	4,49	185,7	206,4	228,7	10,28	4,98
23	77,7	82,0	87,7	2,75	3,35	190,0	220,7	241,7	12,02	5,45
24	58,3	62,4	69,3	2,66	4,26	206,7	225,3	245,0	11,87	5,27
25	65,0	68,5	74,0	2,68	3,91	156,3	200,2	217,7	14,15	7,07
26	68,3	69,7	71,0	0,93	1,33	185,3	200,6	208,7	6,55	3,27
27	78,7	84,2	88,0	2,54	3,02	201,7	240,0	275,0	14,63	6,10
28	60,7	64,8	70,0	2,71	4,18	180,0	221,7	251,7	17,23	7,77
29	65,0	71,0	75,0	2,50	3,52	176,3	202,3	231,3	13,95	6,90
30	70,0	71,7	73,0	0,82	1,14	176,7	201,4	233,7	12,64	6,28
31	55,0	62,5	69,0	4,24	6,78	144,0	164,5	183,0	10,78	6,55
32	73,3	76,2	81,0	2,42	3,18	146,7	166,9	186,7	9,54	5,72
33	83,0	84,2	87,3	1,12	1,33	172,2	185,3	192,2	6,16	3,32
34	61,0	64,2	67,0	2,83	4,41	149,0	168,7	190,7	10,93	6,48
35	57,0	65,4	73,0	3,77	5,76	150,7	175,3	228,7	15,66	8,93
36	74,0	80,5	89,0	3,78	4,70	153,3	172,0	190,0	10,50	6,10
37	82,0	86,1	94,3	3,09	3,59	168,9	197,0	214,4	9,27	4,71
38	65,0	66,8	69,0	1,67	2,50	170,0	205,4	241,0	14,31	6,97
39	82,0	84,0	85,7	0,98	1,17	175,6	187,5	202,2	7,35	3,92
40	61,0	63,3	65,0	2,06	3,25	185,0	198,2	215,3	8,43	4,25
41	68,0	71,7	73,7	1,94	2,71	177,7	205,8	224,3	12,55	6,10
42	83,0	84,8	90,0	1,51	1,78	172,2	194,5	210,6	10,23	5,26
43	65,0	67,1	69,0	1,74	2,59	178,3	206,3	241,7	12,79	6,20
44	70,3	73,5	76,0	1,32	1,80	199,3	223,8	252,0	14,40	6,43
Média	67,3	71,7	75,9	2,45	3,47	182,6	207,3	231,8	12,47	6,06

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

**Tabela 7** - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: altura de inserção da espiga (em cm) e posição relativa da espiga (índice), mensuradas em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	----- Altura de Inserção da Espiga -----					----- Posição Relativa da Espiga -----				
	Mínimo	Média	Máximo	DP	CV	Mínimo	Média	Máximo	DP	CV
1	96,7	106,8	123,3	7,58	7,10	0,50	0,53	0,55	0,01	1,89
2	75,0	90,5	110,0	12,54	13,86	0,44	0,50	0,54	0,03	6,00
3	123,3	136,3	151,3	8,22	6,03	0,50	0,52	0,57	0,02	3,85
4	110,0	111,8	120,0	4,05	3,62	0,55	0,58	0,61	0,02	3,45
5	67,0	73,7	83,0	4,94	6,70	0,38	0,40	0,43	0,01	2,50
6	78,3	121,4	136,7	13,36	11,00	0,46	0,56	0,60	0,03	5,36
7	65,0	93,9	120,0	13,42	14,29	0,40	0,51	0,63	0,05	9,80
8	101,3	143,9	170,0	13,33	9,26	0,42	0,53	0,58	0,03	5,66
9	90,0	113,9	140,0	13,58	11,92	0,47	0,57	0,65	0,04	7,02
10	47,0	78,1	92,0	9,90	12,68	0,34	0,43	0,49	0,03	6,98
11	90,0	107,6	116,7	6,87	6,38	0,42	0,53	0,60	0,04	7,55
12	112,7	128,6	143,7	9,05	7,04	0,45	0,52	0,58	0,03	5,77
13	115,0	126,9	140,0	6,72	5,30	0,52	0,56	0,61	0,02	3,57
14	90,0	119,4	140,0	11,64	9,75	0,48	0,56	0,60	0,03	5,36
15	115,0	139,3	167,0	12,60	9,05	0,43	0,53	0,60	0,04	7,55
16	97,3	129,8	161,3	15,28	11,77	0,49	0,58	0,65	0,04	6,90
17	101,7	128,5	170,0	18,99	14,78	0,47	0,55	0,63	0,05	9,09
18	96,7	118,7	133,3	11,05	9,31	0,49	0,55	0,59	0,03	5,45
19	84,0	105,0	133,3	14,30	13,62	0,48	0,56	0,61	0,04	7,14
20	110,0	127,8	145,0	9,61	7,52	0,50	0,54	0,58	0,02	3,70
21	106,7	127,7	146,7	10,17	7,96	0,50	0,56	0,62	0,03	5,36
22	99,3	117,4	143,0	9,74	8,30	0,52	0,57	0,65	0,03	5,26
23	103,3	119,1	133,3	8,25	6,93	0,49	0,54	0,58	0,02	3,70
24	113,3	129,4	148,3	11,72	9,06	0,53	0,57	0,63	0,03	5,26
25	90,7	114,4	133,3	11,46	10,02	0,47	0,57	0,65	0,04	7,02
26	92,7	109,6	129,3	8,23	7,51	0,50	0,55	0,67	0,04	7,27
27	113,3	134,9	161,7	12,92	9,58	0,51	0,56	0,63	0,03	5,36
28	85,0	127,2	158,3	17,28	13,58	0,47	0,57	0,64	0,04	7,02
29	94,7	118,1	140,7	12,55	10,63	0,51	0,58	0,66	0,04	6,90
30	94,0	115,3	148,3	12,47	10,82	0,48	0,57	0,72	0,04	7,02
31	57,3	73,4	93,0	9,34	12,72	0,40	0,45	0,51	0,03	6,67
32	66,7	79,8	96,7	6,40	8,02	0,40	0,48	0,52	0,03	6,25
33	95,0	112,1	131,7	8,67	7,73	0,52	0,60	0,68	0,05	8,33
34	86,3	98,4	112,7	9,36	9,51	0,52	0,58	0,64	0,04	6,90
35	68,7	81,0	104,3	9,18	11,33	0,40	0,46	0,51	0,03	6,52
36	76,7	87,1	100,0	4,92	5,65	0,44	0,51	0,58	0,03	5,88
37	106,7	122,0	142,2	8,24	6,75	0,53	0,62	0,68	0,04	6,45
38	80,0	115,8	165,0	14,63	12,63	0,47	0,56	0,68	0,04	7,14
39	102,2	114,4	127,8	7,43	6,49	0,51	0,61	0,67	0,04	6,56
40	101,3	111,4	121,7	8,02	7,20	0,52	0,56	0,60	0,03	5,36
41	103,0	118,4	129,7	7,86	6,64	0,49	0,58	0,69	0,05	8,62
42	100,0	123,5	145,0	11,29	9,14	0,57	0,63	0,69	0,04	6,35
43	99,3	121,6	150,0	10,96	9,01	0,54	0,59	0,64	0,03	5,08
44	118,3	135,9	153,0	9,96	7,33	0,55	0,61	0,66	0,03	4,92
Média	93,7	113,9	134,4	10,41	9,22	0,48	0,55	0,61	0,03	6,04

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



Em relação ao número de plantas, a média geral dos ensaios foi de 55.023 plantas ha<sup>-1</sup> e as médias dos ensaios oscilaram entre 39.248 e 63.646 plantas ha<sup>-1</sup>, para os ensaios 7 e 25, respectivamente (Tabela 8). Conforme descrito na metodologia, a densidade de semeadura variou em função da categoria, do ciclo e do ano agrícola do ensaio. Assim, a comparação do número de plantas entre ensaios de diferentes categorias, ciclos ou anos agrícolas não é adequada. No entanto, como a semeadura é realizada em excesso, com posterior desbaste de plantas até atingir a população desejada, espera-se uma pequena variação da densidade entre cultivares de um mesmo ensaio, ou de ensaios pertencentes a uma mesma categoria, ciclo e ano agrícola.

Porém, verificou-se uma elevada variabilidade do número de plantas entre ensaios, como no ensaio 7, conduzido em Pelotas, que apresentou um número de plantas por cultivar, consideravelmente menor em relação aos ensaios do mesmo grupo (ensaios 6, 8, 9 e 10) conduzidos em Passo Fundo, Rio Pardo, Roca Sales e São Borja (Tabela 8). Também dentro de um mesmo ensaio, foram verificadas amplas variações do número de plantas, como no ensaio 28, em que a amplitude entre cultivares foi de 32.083 plantas ha<sup>-1</sup>. Essa ampla variabilidade pode estar associada a problemas de ajuste da população final e a perdas de plantas ocorridas durante a condução experimental, por atuação de fatores bióticos e abióticos. O desvio-padrão médio para o número de plantas foi de 3.823 plantas ha<sup>-1</sup> e o coeficiente de variação médio foi de 7,25%.

A média geral do número de espigas foi de 51.247 espigas ha<sup>-1</sup> (Tabela 8). Assim como para o número de plantas, também foi verificada elevada variabilidade do número de espigas, que oscilou entre 23.750 e 73.750 espigas ha<sup>-1</sup>, para uma cultivar do ensaio 12 e outra do ensaio 3, respectivamente. Conforme salientado anteriormente, deve-se comparar apenas cultivares dentro de um mesmo ensaio, ou entre ensaios de um mesmo grupo, em função das diferentes densidades de semeadura utilizadas. O ensaio 19 apresentou a menor variabilidade entre cultivares para o número de espigas, com amplitude de 6.667 espigas ha<sup>-1</sup>. Já no ensaio 7, o número de espigas oscilou entre 26.667 e 71.250 espigas ha<sup>-1</sup> e assim, a amplitude entre cultivares foi de 44.583 espigas ha<sup>-1</sup>.

A variação das médias dos ensaios de um mesmo grupo de cultivares também foi elevada, como no caso dos ensaios de ciclo superprecoce do ano 2004/05, que apresentaram médias de 39.514 a 57.292 espigas ha<sup>-1</sup>, nos locais de Cruz Alta e Santa Maria, respectivamente. Além da interferência de fatores já mencionados para o número de plantas, a variabilidade do número de espigas está diretamente associada à prolificidade das cultivares. O desvio-padrão oscilou entre 2.195 e 10.854 espigas ha<sup>-1</sup> e, o coeficiente de variação médio,

foi de 10,87% (Tabela 8). A ampla variabilidade verificada em cultivares de mesmo ensaio é importante na discriminação estatística de cultivares quanto à produção de espigas por área.

Como pode ser verificado nas médias gerais, o número de espigas foi menor em relação ao número de plantas (Tabela 8). Com isso, a prolificidade média dos ensaios foi de 0,93 e, dessa forma, algumas cultivares não produziram espigas em todas as plantas (Tabela 9). A menor prolificidade média foi observada no ensaio 34 e a maior no ensaio 7, com valores de prolificidade de 0,73 e 1,09, respectivamente. Em relação aos extremos, uma cultivar do ensaio 12 apresentou uma prolificidade de apenas 0,54 e uma cultivar do ensaio 7 apresentou prolificidade de 1,53. O desvio-padrão médio foi de 0,08 do índice de prolificidade e o coeficiente de variação médio foi de 8,32% (Tabela 9).

A produtividade de grãos foi a variável que apresentou a maior variabilidade entre cultivares e ensaios. A média da produtividade de grãos foi de 7.048 kg ha<sup>-1</sup> (Tabela 9). Nas safras agrícolas 2002/03, 2003/04 e 2004/05, a produtividade de grãos de milho no estado do Rio Grande do Sul foi de 3.750, 2.700 e 1.269 kg ha<sup>-1</sup>, respectivamente (CONAB, 2011). A média da produtividade desses três anos agrícolas é de, aproximadamente, 2.573 kg ha<sup>-1</sup>. Com isso, constata-se que a média da produtividade de grãos dos ensaios foi 2,74 vezes maior que a produtividade das lavouras comerciais. Esses resultados indicam o elevado potencial de produtividade das novas cultivares que são lançadas no mercado, associado às boas práticas de manejo adotadas nas áreas experimentais da rede de ensaios de competição de cultivares de milho.

A menor produtividade de grãos foi mensurada em uma cultivar do ensaio 10, com 1.772 kg ha<sup>-1</sup> e a maior produtividade foi mensurada em uma cultivar do ensaio 23, com 13.089 kg ha<sup>-1</sup> (Tabela 9). Assim, a cultivar mais produtiva apresentou produtividade 7,39 vezes maior em relação a cultivar menos produtiva. Em relação à média dos ensaios, a amplitude máxima foi de 6.210 kg ha<sup>-1</sup>, pois o ensaio 31 apresentou produtividade média de 4.016 kg ha<sup>-1</sup> e o ensaio 25 apresentou produtividade média de 10.226 kg ha<sup>-1</sup>. A ampla variabilidade é confirmada com os elevados valores de desvio-padrão (488 kg ≤ DP ≤ 1.757 kg) e coeficiente de variação dos ensaios (7,54% ≤ CV ≤ 25,68%) (Tabela 9). Essa ampla variabilidade verificada para a produtividade de grãos é importante, pois indica a existência de cultivares com alto e baixo potencial produtivo nos ensaios, permitindo uma discriminação precisa das cultivares para identificar aquelas com maior potencial em relação a produtividade de grãos.

**Tabela 8** - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: número de plantas ( $\text{ha}^{-1}$ ) e número de espigas ( $\text{ha}^{-1}$ ) mensuradas em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	Número de Plantas					Número de Espigas				
	Mínimo	Média	Máximo	DP	CV	Mínimo	Média	Máximo	DP	CV
1	59.167	62.462	64.167	1.702	2,72	46.250	58.258	65.000	5.658	9,71
2	31.250	50.417	62.917	8.207	16,28	31.250	52.159	69.167	10.854	20,81
3	45.833	61.061	65.000	5.784	9,47	51.667	58.598	73.750	6.076	10,37
4	58.750	62.462	64.583	1.858	2,97	59.167	63.485	66.667	2.410	3,80
5	56.667	63.636	69.583	3.864	6,07	42.917	53.371	68.333	8.121	15,22
6	49.583	53.866	55.417	1.432	2,66	38.333	48.993	54.583	3.557	7,26
7	24.167	39.248	51.250	6.041	15,39	26.667	42.928	71.250	8.931	20,80
8	49.583	54.271	55.000	1.469	2,71	33.333	52.998	64.167	6.226	11,75
9	46.667	53.507	55.000	1.945	3,64	46.250	54.306	57.917	2.340	4,31
10	44.167	55.451	65.000	4.825	8,70	29.167	47.488	62.083	6.848	14,42
11	52.083	61.991	65.000	3.107	5,01	50.417	58.981	63.333	3.247	5,51
12	42.917	58.542	65.000	6.727	11,49	23.750	51.481	63.750	8.987	17,46
13	51.250	58.657	62.500	3.289	5,61	46.250	53.634	62.083	4.052	7,55
14	47.917	53.441	55.833	2.179	4,08	43.750	49.985	54.583	3.214	6,43
15	40.833	52.083	55.000	3.603	6,92	30.000	50.586	60.833	6.380	12,61
16	40.417	51.235	56.667	3.474	6,78	32.500	45.602	52.500	5.253	11,52
17	53.750	60.463	64.583	3.401	5,62	53.750	56.991	62.083	3.273	5,74
18	44.167	55.602	66.250	6.092	10,96	40.000	49.954	56.667	4.923	9,86
19	56.250	61.435	64.583	3.098	5,04	60.000	63.889	66.667	2.195	3,44
20	34.167	49.167	55.000	4.045	8,23	32.917	49.771	59.167	5.219	10,49
21	38.333	53.750	67.083	6.957	12,94	34.583	47.938	57.500	5.406	11,28
22	50.833	54.021	55.000	1.005	1,86	47.500	55.771	65.417	3.405	6,11
23	42.083	52.474	62.500	5.291	10,08	39.167	52.786	62.083	6.827	12,93
24	43.750	55.339	66.667	6.225	11,25	38.750	49.323	58.333	5.469	11,09
25	60.833	63.646	65.000	1.123	1,76	60.417	64.531	70.833	2.504	3,88
26	54.583	63.047	68.333	3.812	6,05	44.583	59.063	67.917	7.203	12,20
27	46.667	52.068	56.667	2.614	5,02	46.667	51.559	57.083	2.662	5,16
28	33.750	51.204	65.833	6.993	13,66	31.250	46.883	61.667	7.223	15,41
29	48.750	54.120	55.000	1.236	2,28	37.917	53.364	61.667	5.209	9,76
30	51.250	55.448	58.333	1.457	2,63	37.917	52.407	57.917	4.547	8,68
31	39.048	52.661	60.952	6.391	12,14	35.238	46.359	55.238	6.613	14,26
32	58.750	60.270	62.083	931	1,54	46.250	54.583	60.833	4.077	7,47
33	27.500	41.520	57.500	6.198	14,93	24.167	37.304	46.667	5.581	14,96
34	52.083	62.451	65.000	4.375	7,01	33.750	45.711	57.500	6.612	14,46
35	25.238	46.429	57.143	6.901	14,86	25.714	46.027	60.000	6.814	14,80
36	54.583	59.258	62.500	1.696	2,86	35.417	51.445	60.000	6.270	12,19
37	26.250	43.151	55.417	6.262	14,51	24.167	35.273	47.917	5.533	15,69
38	34.583	53.867	55.000	3.861	7,17	33.333	47.520	59.167	5.862	12,34
39	34.583	46.319	52.917	5.152	11,12	30.000	39.514	44.167	4.405	11,15
40	56.250	62.951	65.000	2.933	4,66	41.667	54.670	63.333	6.956	12,72
41	58.333	62.535	66.250	2.187	3,50	48.333	57.292	72.917	6.762	11,80
42	35.833	46.042	52.083	3.418	7,42	30.833	39.389	47.500	4.092	10,39
43	43.333	53.847	55.000	2.867	5,32	40.833	49.042	60.833	5.632	11,48
44	51.667	55.611	61.250	2.174	3,91	47.083	53.653	71.667	4.748	8,85
Média	45.419	55.023	60.610	3.823	7,25	39.406	51.247	60.971	5.413	10,87

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

**Tabela 9** - Mínimo, média, máximo, desvio-padrão (DP) e coeficiente de variação (CV, em %) para as variáveis: prolificidade (número de espigas dividido por número de plantas) e produtividade de grãos (em kg ha<sup>-1</sup>), mensuradas em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	----- Prolificidade -----					----- Produtividade de grãos -----				
	Mínimo	Média	Máximo	DP	CV	Mínimo	Média	Máximo	DP	CV
1	0,75	0,93	1,02	0,08	8,60	4.869	7.670	9.974	1.443	18,81
2	0,89	1,03	1,27	0,11	10,68	4.332	6.993	9.852	1.580	22,59
3	0,83	0,96	1,13	0,10	10,42	5.591	7.067	9.880	1.259	17,82
4	0,96	1,02	1,05	0,02	1,96	8.025	9.263	11.000	957	10,33
5	0,72	0,84	1,01	0,10	11,90	3.155	4.343	6.154	934	21,51
6	0,73	0,91	1,01	0,07	7,69	5.878	8.874	12.473	1.548	17,44
7	0,86	1,09	1,53	0,13	11,93	3.030	5.766	7.840	1.273	22,08
8	0,67	0,98	1,17	0,10	10,20	3.472	7.034	10.083	1.572	22,35
9	0,96	1,02	1,10	0,03	2,94	5.284	8.910	11.072	1.221	13,70
10	0,62	0,85	0,98	0,08	9,41	1.772	4.284	6.527	1.077	25,14
11	0,86	0,95	1,01	0,03	3,16	6.723	8.857	10.268	1.119	12,63
12	0,54	0,88	0,98	0,10	11,36	2.065	6.129	8.516	1.574	25,68
13	0,81	0,91	0,99	0,05	5,49	5.928	7.808	9.053	864	11,07
14	0,81	0,94	1,01	0,05	5,32	6.206	8.707	10.675	1.170	13,44
15	0,66	0,97	1,32	0,13	13,40	3.828	6.683	8.781	1.244	18,61
16	0,65	0,89	1,19	0,11	12,36	3.086	6.020	8.497	1.322	21,96
17	0,87	0,94	1,02	0,05	5,32	8.373	9.633	12.330	1.196	12,42
18	0,86	0,90	1,01	0,04	4,44	5.590	6.211	7.153	488	7,86
19	0,98	1,04	1,13	0,05	4,81	8.335	9.819	10.688	740	7,54
20	0,87	1,01	1,13	0,06	5,94	6.193	9.492	12.237	1.260	13,27
21	0,78	0,90	1,04	0,07	7,78	4.543	6.223	8.077	954	15,33
22	0,88	1,03	1,24	0,06	5,83	7.749	9.417	11.144	853	9,06
23	0,90	1,01	1,19	0,08	7,92	7.105	10.149	13.089	1.757	17,31
24	0,76	0,89	0,94	0,05	5,62	4.628	6.135	7.837	863	14,07
25	0,95	1,01	1,13	0,04	3,96	6.725	10.226	12.442	1.465	14,33
26	0,76	0,93	1,03	0,07	7,53	3.541	8.053	10.133	1.569	19,48
27	0,90	0,99	1,12	0,05	5,05	7.917	9.986	11.511	849	8,50
28	0,80	0,92	1,07	0,07	7,61	3.962	6.054	8.368	956	15,79
29	0,72	0,99	1,26	0,10	10,10	6.208	9.015	11.079	1.375	15,25
30	0,66	0,95	1,05	0,09	9,47	4.049	7.051	9.173	1.404	19,91
31	0,69	0,88	0,96	0,07	7,95	2.860	4.016	5.057	620	15,44
32	0,78	0,90	0,98	0,06	6,67	2.488	4.639	6.829	915	19,72
33	0,72	0,90	1,02	0,09	10,00	2.918	5.589	7.522	1.013	18,12
34	0,62	0,73	0,88	0,08	10,96	2.745	4.463	6.373	935	20,95
35	0,88	0,99	1,13	0,06	6,06	2.764	4.536	6.768	956	21,08
36	0,62	0,87	0,99	0,09	10,34	2.245	4.359	6.395	981	22,51
37	0,68	0,82	0,99	0,08	9,76	2.786	4.923	7.112	1.151	23,38
38	0,61	0,88	1,10	0,10	11,36	3.946	5.644	8.021	1.007	17,84
39	0,69	0,86	0,95	0,09	10,47	3.826	5.626	6.816	871	15,48
40	0,64	0,87	1,01	0,11	12,64	4.058	5.864	7.320	1.089	18,57
41	0,78	0,92	1,14	0,10	10,87	6.630	8.511	10.256	1.224	14,38
42	0,60	0,86	0,98	0,09	10,47	3.701	5.612	7.593	1.001	17,84
43	0,75	0,91	1,11	0,09	9,89	3.780	5.804	7.322	914	15,75
44	0,84	0,97	1,28	0,10	10,31	6.217	8.644	10.992	1.017	11,77
Média	0,77	0,93	1,08	0,08	8,32	4.753	7.048	9.097	1.127	16,78

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

Em relação ao diagnóstico de normalidade univariada, em 59,09% dos ensaios (26 ensaios), a hipótese nula  $H_0$ : a amostra provém de uma população normal foi rejeitada ( $p \leq 0,10$ ) para o florescimento masculino (Tabela 10). Considerando as variáveis EP, AIE e PRE, a hipótese nula foi rejeitada em 9, 11 e 9 ensaios, o que representa, respectivamente, 20,45, 25,00 e 20,45% dos ensaios. Nas variáveis NP, NE e PLI, 22, 16 e 15 ensaios, não apresentaram distribuição normal, respectivamente, ou seja, em 50,00, 36,36 e 34,09% dos ensaios, a hipótese nula foi rejeitada ( $p \leq 0,10$ ). Já a variável principal produtividade de grãos apresentou comportamento normal na maioria dos ensaios, sendo que em apenas cinco ensaios (11,36% dos ensaios), a hipótese nula foi rejeitada ( $p \leq 0,10$ ).

Com base no diagnóstico de normalidade univariada é possível concluir que existem dois grupos de variáveis quanto à distribuição dos dados. As variáveis que foram mensuradas (EP, AIE, PRE e PROD) apresentam predominância de comportamento normal. Já as variáveis provenientes de contagens (FM, NP, NE e PLI) apresentam maior ocorrência de não-normalidade, embora apenas a variável FM apresente maior quantidade de ensaios com distribuição não-normal dos dados (59,09% dos ensaios). Dessa forma, dos 352 diagnósticos de normalidade univariada realizados (44 ensaios  $\times$  8 variáveis), em 113 casos a hipótese nula foi rejeitada ( $p \leq 0,10$ ), ou seja, em 32,10% dos casos (Tabela 10). Assim, verifica-se que em variáveis fenológicas, morfológicas e produtivas de milho, há uma predominância de distribuições dos dados próximas a distribuição normal. No entanto, em aproximadamente 1/3 das variáveis, a distribuição dos dados não segue a distribuição normal, sendo necessária a utilização de famílias de transformação dos dados.

Em relação ao diagnóstico de normalidade multivariada, houve tendência a rejeição da hipótese nula ( $p \leq 0,10$ ), visto que em 30 ensaios (68,18% dos ensaios) a distribuição multivariada não foi semelhante à distribuição normal multivariada (Tabela 10). Em apenas 14 ensaios (ensaios 1, 2, 5, 13,17, 18, 19, 21, 23, 24, 27, 28, 31 e 41), foi verificado que os dados amostrais seguem distribuição normal multivariada ( $p > 0,10$ ). Conforme descrito na metodologia, nos critérios de transformação dos dados, nesses 14 ensaios não foram realizadas transformações de dados, mesmo que em alguns casos não houve ajuste normal univariado. Esses 14 ensaios compõem o caso 1, que será estudado detalhadamente a seguir.

Nos demais 30 ensaios, foram realizadas transformações dos dados, mediante a utilização da família de transformações Box-Cox. Nos ensaios 10, 12, 16, 20, 25, 26, 35, 36, 37, 39 e 44, após a transformação dos dados das variáveis que não apresentavam ajuste normal univariado ( $p \leq 0,10$ ), foi verificado que os dados amostrais passaram a se ajustar a distribuição normal multivariada ( $p > 0,10$ ) (Tabela 11).

Nos demais 19 ensaios, a hipótese nula para a normalidade multivariada foi rejeitada ( $p \leq 0,10$ ), ou seja, a transformação das variáveis que não se ajustaram a distribuição normal univariada não foi suficiente para garantir a distribuição normal multivariada. Dessa forma, na busca de mais ensaios com normalidade multivariada, utilizou-se o segundo critério de transformação. Assim, todas as variáveis desses 19 ensaios foram transformadas e mantiveram-se no estudo, variáveis transformadas, quando estas apresentavam melhoria no ajuste à distribuição normal (valor-p maior que a variável original). Quando as transformações testadas não melhoraram os ajustes, mantiveram-se os dados originais. Assim, após a utilização desse segundo critério, mais dois ensaios (7 e 32) passaram a apresentar distribuição normal multivariada (Tabela 11).

Assim, os 11 ensaios que apresentaram distribuição normal multivariada após a transformação das variáveis que não se ajustaram a distribuição normal univariada e os ensaios 7 e 32, constituíram o caso 2, que também será estudado separadamente. Por fim, os demais 17 ensaios que não se ajustaram a distribuição normal multivariada, mesmo utilizando os dois critérios de transformações descritos na metodologia, constituíram o caso 3 (Tabela 12).

Após o diagnóstico de normalidade univariada e multivariada, foram calculados os coeficientes de correlação de Pearson (fenotípicos) entre as variáveis explicativas para a obtenção da matriz de correlação  $X'X$  e das variáveis explicativas com a produtividade de grãos, para a obtenção da matriz de correlação  $X'Y$ . Com exceção dos 14 ensaios que apresentaram distribuição normal multivariada (Tabela 10), nos demais 30 ensaios, essas matrizes foram calculadas antes e após a transformação dos dados e serão novamente apresentadas nos casos 2 e 3. Os valores dos coeficientes de correlação entre as variáveis explicativas não serão apresentados. Apenas os coeficientes de correlação entre as variáveis explicativas e a produtividade de grãos (matriz de correlação  $X'Y$ ) são apresentados para cada ensaio (Tabela 13).

**Tabela 10** - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM) em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	Variáveis mensuradas <sup>(2)</sup>								NM
	FM	EP	AIE	PRE	NP	NE	PLI	PROD	
1	0,985	0,742	0,408	0,436	0,092	0,430	0,145	0,827	0,506
2	0,104	0,431	0,076	0,495	0,298	0,954	0,450	0,959	0,381
3	0,073	0,156	0,831	0,005	0,001	0,073	0,213	0,107	0,000
4	0,010	0,205	0,000	0,055	0,503	0,236	0,241	0,090	0,000
5	0,193	0,653	0,748	0,904	0,812	0,634	0,172	0,646	0,777
6	0,000	0,041	0,000	0,000	0,000	0,006	0,013	0,709	0,000
7	0,002	0,440	0,517	0,176	0,921	0,210	0,017	0,138	0,008
8	0,449	0,690	0,183	0,168	0,000	0,109	0,329	0,279	0,000
9	0,004	0,032	0,011	0,239	0,000	0,003	0,209	0,028	0,000
10	0,005	0,023	0,019	0,213	0,555	0,632	0,097	0,795	0,007
11	0,004	0,254	0,046	0,019	0,000	0,014	0,522	0,170	0,000
12	0,105	0,337	0,494	0,999	0,009	0,013	0,001	0,177	0,000
13	0,205	0,063	0,557	0,674	0,007	0,651	0,564	0,580	0,105
14	0,004	0,068	0,650	0,484	0,000	0,046	0,238	0,026	0,000
15	0,234	0,064	0,510	0,207	0,000	0,014	0,023	0,207	0,000
16	0,021	0,398	0,973	0,839	0,041	0,016	0,095	0,763	0,024
17	0,922	0,199	0,277	0,791	0,521	0,066	0,431	0,074	0,251
18	0,129	0,471	0,638	0,405	0,907	0,554	0,018	0,577	0,273
19	0,049	0,419	0,750	0,666	0,095	0,634	0,230	0,507	0,293
20	0,335	0,198	0,424	0,087	0,003	0,116	0,577	0,885	0,031
21	0,100	0,948	0,394	0,897	0,928	0,247	0,540	0,279	0,595
22	0,005	0,672	0,177	0,249	0,000	0,919	0,065	0,637	0,000
23	0,821	0,493	0,435	0,816	0,464	0,503	0,304	0,700	0,892
24	0,268	0,336	0,262	0,122	0,993	0,923	0,025	0,602	0,209
25	0,158	0,004	0,975	0,246	0,049	0,267	0,012	0,610	0,005
26	0,098	0,160	0,620	0,010	0,074	0,051	0,103	0,015	0,005
27	0,331	0,779	0,279	0,324	0,626	0,891	0,448	0,707	0,817
28	0,032	0,600	0,701	0,540	0,972	0,892	0,137	0,602	0,434
29	0,094	0,477	0,540	0,503	0,000	0,000	0,001	0,133	0,000
30	0,049	0,606	0,499	0,007	0,269	0,000	0,000	0,260	0,000
31	0,501	0,935	0,526	0,395	0,157	0,215	0,032	0,919	0,293
32	0,064	0,678	0,023	0,118	0,362	0,432	0,175	0,314	0,076
33	0,001	0,049	0,623	0,765	0,162	0,266	0,227	0,304	0,007
34	0,000	0,936	0,044	0,324	0,000	0,623	0,518	0,268	0,000
35	0,911	0,005	0,043	0,904	0,002	0,070	0,524	0,332	0,002
36	0,284	0,438	0,123	0,481	0,324	0,054	0,013	0,968	0,074
37	0,002	0,166	0,781	0,544	0,151	0,533	0,479	0,394	0,061
38	0,000	0,257	0,000	0,083	0,000	0,971	0,698	0,746	0,000
39	0,008	0,819	0,841	0,179	0,361	0,018	0,118	0,772	0,028
40	0,000	0,372	0,037	0,173	0,003	0,518	0,490	0,479	0,001
41	0,070	0,516	0,904	0,435	0,990	0,377	0,502	0,399	0,594
42	0,000	0,185	0,915	0,059	0,228	0,544	0,128	0,531	0,000
43	0,000	0,782	0,719	0,872	0,000	0,129	0,247	0,490	0,000
44	0,450	0,439	0,384	1,000	0,230	0,001	0,011	0,764	0,008

<sup>(1)</sup> Ensaios definidos 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, PROD: produtividade de grãos.

**Tabela 11** - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM), em 13 ensaios de competição de cultivares de milho. Sem (Sem T.) e com (Com T.) transformação de dados, de variáveis que não se ajustaram a distribuição normal, por meio da metodologia Box-Cox e o respectivo valor de  $\lambda$  utilizado.

Ensaio <sup>(1)</sup>		Variáveis mensuradas <sup>(2)</sup>								NM
		FM	EP	AIE	PRE	NP	NE	PLI	PROD	
7 <sup>(3)</sup>	Sem T.	0,002	0,440	0,517	0,176	0,921	0,210	0,017	0,138	0,008
	Com T.	0,002	0,442	0,527	0,176	0,986	0,874	0,889	0,547	0,138
	Valor de $\lambda$	STA <sup>(4)</sup>	1,241	1,199	-	1,604	-0,025	-1,699	1,896	
10	Sem T.	0,005	0,023	0,019	0,213	0,555	0,632	0,097	0,795	0,007
	Com T.	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 <sup>(5)</sup>	3,733	-	-	-	4,065	-	
12	Sem T.	0,105	0,337	0,494	0,999	0,009	0,013	0,001	0,177	0,000
	Com T.	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	-	
16	Sem T.	0,021	0,398	0,973	0,839	0,041	0,016	0,095	0,763	0,024
	Com T.	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	-	
20	Sem T.	0,335	0,198	0,424	0,087	0,003	0,116	0,577	0,885	0,031
	Com T.	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	-	-	-	
25	Sem T.	0,158	0,004	0,975	0,246	0,049	0,267	0,012	0,610	0,005
	Com T.	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	-	
26	Sem T.	0,098	0,160	0,620	0,010	0,074	0,051	0,103	0,015	0,005
	Com T.	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	
32 <sup>(3)</sup>	Sem T.	0,064	0,678	0,023	0,118	0,362	0,432	0,175	0,314	0,076
	Com T.	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	-	
35	Sem T.	0,911	0,005	0,043	0,904	0,002	0,070	0,524	0,332	0,002
	Com T.	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	-	-	
36	Sem T.	0,284	0,438	0,123	0,481	0,324	0,054	0,013	0,968	0,074
	Com T.	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	-	
37	Sem T.	0,002	0,166	0,781	0,544	0,151	0,533	0,479	0,394	0,061
	Com T.	0,008	0,166	0,781	0,544	0,151	0,533	0,479	0,394	0,115
	Valor de $\lambda$	-1,899	-	-	-	-	-	-	-	
39	Sem T.	0,008	0,819	0,841	0,179	0,361	0,018	0,118	0,772	0,028
	Com T.	0,008	0,819	0,841	0,179	0,361	0,380	0,118	0,772	0,121
	Valor de $\lambda$	STA	-	-	-	-	5,000	-	-	
44	Sem T.	0,450	0,439	0,384	1,000	0,230	0,001	0,011	0,764	0,008
	Com T.	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> Ensaios definidos 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, PROD: produtividade de grãos. <sup>(3)</sup> Nesses ensaios não se obteve normalidade multivariada após a transformação das variáveis que não seguiram distribuição normal univariada. Assim, todas as variáveis desses ensaios, independentemente da distribuição normal univariada, foram transformadas, e mantiveram-se os dados transformados para os casos em que a normalidade univariada apresentava melhoria, ou seja, acréscimo no valor-p. <sup>(4)</sup> Sem transformação adequada, ou seja, nenhum valor de  $\lambda$  dentro do intervalo de -5 a 5, melhorou a distribuição dos dados. <sup>(5)</sup> Valores de  $\lambda = |5|$  indicam que o melhor valor de  $\lambda$  para a transformação dos dados não foi encontrado entre  $-5 \leq \lambda \leq 5$ .



**Tabela 12** - Valor-p do teste de normalidade univariado de Shapiro-Wilk para cada variável e valor-p do teste de normalidade multivariado de Shapiro-Wilk generalizado por Royston entre as oito variáveis (NM), em 17 ensaios de competição de cultivares de milho. Sem (Sem T.) e com (Com T.) transformação de dados, de variáveis que não se ajustaram a distribuição normal, por meio da metodologia Box-Cox e o respectivo valor de  $\lambda$  utilizado.

Ensaio <sup>(1)(2)</sup>		Variáveis mensuradas <sup>(3)</sup>								NM
		FM	EP	AIE	PRE	NP	NE	PLI	PROD	
3	Sem T.	0,073	0,156	0,831	0,005	0,001	0,073	0,213	0,107	0,000
	Com T.	0,125	0,156	0,831	0,085	0,022	0,073	0,551	0,107	0,021
	Valor de $\lambda$	5,000 <sup>(4)</sup>	-	-	-5,000	5,000	STA <sup>(5)</sup>	-2,583	-	-
4	Sem T.	0,010	0,205	0,000	0,055	0,503	0,236	0,241	0,090	0,000
	Com T.	0,018	0,205	0,000	0,073	0,624	0,370	0,451	0,090	0,000
	Valor de $\lambda$	5,000	-	STA	-5,000	5,000	5,000	5,000	5,000	STA
6	Sem T.	0,000	0,041	0,000	0,000	0,000	0,006	0,013	0,709	0,000
	Com T.	0,000	0,556	0,127	0,033	0,001	0,551	0,631	0,709	0,000
	Valor de $\lambda$	STA	5,000	5,000	5,000	5,000	5,000	5,000	-	-
8	Sem T.	0,449	0,690	0,183	0,168	0,000	0,109	0,329	0,279	0,000
	Com T.	0,468	0,754	0,659	0,780	0,000	0,785	0,922	0,340	0,000
	Valor de $\lambda$	-0,094	3,109	2,695	4,129	STA	2,780	2,727	1,318	-
9	Sem T.	0,004	0,032	0,011	0,239	0,000	0,003	0,209	0,028	0,000
	Com T.	0,039	0,075	0,015	0,243	0,000	0,133	0,615	0,518	0,000
	Valor de $\lambda$	5,000	-1,374	0,178	1,427	STA	5,000	-5,000	2,879	-
11	Sem T.	0,004	0,254	0,046	0,019	0,000	0,014	0,522	0,170	0,000
	Com T.	0,017	0,772	0,377	0,091	0,004	0,201	0,954	0,259	0,010
	Valor de $\lambda$	5,000	5,000	5,000	3,665	5,000	5,000	5,000	2,898	-
14	Sem T.	0,004	0,068	0,650	0,484	0,000	0,046	0,238	0,026	0,000
	Com T.	0,031	0,312	0,729	0,484	0,003	0,182	0,490	0,180	0,021
	Valor de $\lambda$	5,000	5,000	5,000	-	5,000	5,000	3,888	5,000	-
15	Sem T.	0,234	0,064	0,510	0,207	0,000	0,014	0,023	0,207	0,000
	Com T.	0,234	0,064	0,510	0,207	0,002	0,928	0,025	0,207	0,004
	Valor de $\lambda$	-	STA	-	-	5,000	5,000	1,332	-	-
22	Sem T.	0,005	0,672	0,177	0,249	0,000	0,919	0,065	0,637	0,000
	Com T.	0,008	0,857	0,738	0,644	0,000	0,944	0,085	0,688	0,001
	Valor de $\lambda$	3,897	-1,076	-1,655	-2,921	STA	0,368	0,195	0,222	-
29	Sem T.	0,094	0,477	0,540	0,503	0,000	0,000	0,001	0,133	0,000
	Com T.	0,094	0,477	0,540	0,503	0,000	0,050	0,001	0,175	0,000
	Valor de $\lambda$	STA	-	-	-	STA	5,000	STA	5,000	-
30	Sem T.	0,049	0,606	0,499	0,007	0,269	0,000	0,000	0,260	0,000
	Com T.	0,077	0,606	0,499	0,042	0,508	0,075	0,069	0,260	0,055
	Valor de $\lambda$	5,000	-	-	-3,967	5,000	5,000	5,000	-	-
33	Sem T.	0,001	0,049	0,623	0,765	0,162	0,266	0,227	0,304	0,007
	Com T.	0,001	0,079	0,623	0,765	0,162	0,266	0,331	0,304	0,011
	Valor de $\lambda$	STA	5,000	-	-	-	-	2,763	-	-
34	Sem T.	0,000	0,936	0,044	0,324	0,000	0,623	0,518	0,268	0,000
	Com T.	0,000	0,936	0,063	0,324	0,000	0,623	0,518	0,268	0,000
	Valor de $\lambda$	STA	-	5,000	-	STA	-	-	-	-
38	Sem T.	0,000	0,257	0,000	0,083	0,000	0,971	0,698	0,746	0,000
	Com T.	0,000	0,261	0,002	0,579	0,000	0,984	0,747	0,888	0,000
	Valor de $\lambda$	STA	1,196	-0,372	-1,786	STA	1,478	1,414	0,168	-
40	Sem T.	0,000	0,372	0,037	0,173	0,003	0,518	0,490	0,479	0,001
	Com T.	0,000	0,576	0,045	0,198	0,004	0,518	0,716	0,482	0,001
	Valor de $\lambda$	STA	-1,620	2,546	3,404	5,000	-	3,149	1,420	-
42	Sem T.	0,000	0,185	0,915	0,059	0,228	0,544	0,128	0,531	0,000
	Com T.	0,000	0,537	0,949	0,066	0,536	0,670	0,728	0,729	0,001
	Valor de $\lambda$	STA	5,000	1,717	1,960	2,965	1,742	3,780	1,536	-
43	Sem T.	0,000	0,782	0,719	0,872	0,000	0,129	0,247	0,490	0,000
	Com T.	0,000	0,970	0,786	0,886	0,000	0,415	0,857	0,740	0,000
	Valor de $\lambda$	STA	-0,915	0,366	-0,153	STA	-0,856	-1,614	1,902	-

<sup>(1)</sup> Ensaios definidos na Tabela 1. <sup>(2)</sup> Nesses ensaios todas as variáveis, independentemente da distribuição normal univariada, foram transformadas, e mantiveram-se os dados transformados para os casos em que a normalidade univariada apresentava melhoria, ou seja, acréscimo no valor-p. <sup>(3)</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, PROD: produtividade de grãos. <sup>(4)</sup> Valores de  $\lambda = |5|$  indicam que o melhor valor de  $\lambda$  para a transformação dos dados não foi encontrado entre  $-5 \leq \lambda \leq 5$ . <sup>(5)</sup> Sem transformação adequada, ou seja, nenhum valor de  $\lambda$  dentro do intervalo de  $-5$  a  $5$ , melhorou a distribuição dos dados.

A variável florescimento masculino apresentou coeficiente de correlação médio de baixa magnitude e sinal inverso com a produtividade de grãos ( $r = -0,11$ ) (Tabela 13). Em progênies de meios-irmãos de seis populações de milho crioulo, Barros et al. (2010) verificaram correlação negativa entre FM e PROD, com média de  $-0,22$ . Em milho pipoca, Rangel et al. (2011) observaram correlação média de  $-0,25$ . Assim, poderia se inferir que cultivares mais produtivas são aquelas que apresentam menor ciclo da emergência ao florescimento masculino. No entanto, em apenas oito dos 44 ensaios, os coeficientes de correlação foram significativos, ou seja, diferiram de zero. Além disso, os valores de coeficiente de correlação de FM com PROD oscilaram em intensidade e direção ( $-0,69 \leq r \leq 0,55$ ). Assim, provavelmente essa variável não apresenta relação de causa e efeito com a produtividade de grãos (CRUZ; REGAZZI, 1997).

A variável estatura de plantas apresentou correlação média positiva com a produtividade de grãos ( $r = 0,20$ ) e ampla variabilidade entre os 44 ensaios ( $-0,43 \leq r \leq 0,91$ ) (Tabela 13). Em milho crioulo e em milho pipoca, Barros et al. (2010) e Rangel et al. (2011) também verificaram correlação fenotípica média positiva e semelhante à encontrada nesse estudo ( $r = 0,21$  e  $r = 0,26$ , respectivamente). Em seleção recorrente de milho, Lima Neto e Souza Júnior (2009) obtiveram correlação fenotípica positiva ( $0,28 \leq r \leq 0,36$ ). Já Churata e Ayala-Osuna (1996), Saidaiah et al. (2008), Rafiq et al. (2010) e Alvi et al. (2003), verificaram maior intensidade da associação entre a estatura de plantas e a produtividade de grãos, com coeficientes de correlação fenotípicos iguais a  $0,46$ ,  $0,50$ ,  $0,53$  e  $0,75$ , respectivamente.

Foi verificada associação média positiva de baixa magnitude ( $r = 0,13$ ) entre a altura de inserção da espiga e a produtividade de grãos e assim como para a associação entre EP e PROD, foi constatada ampla variação dos coeficientes de correlação estimados ( $-0,41 \leq r \leq 0,84$ ) (Tabela 13). Barros et al. (2010) obtiveram uma correlação média de baixa magnitude, entre a altura de inserção da espiga e a produtividade de grãos, em milho crioulo ( $r = 0,19$ ) e Rafiq et al. (2010) obtiveram um coeficiente médio de  $0,24$ . Coeficientes de correlação de maior magnitude entre AIE e PROD foram obtidos em estudos realizados por Lima Neto e Souza Júnior (2009) com  $0,27 \leq r \leq 0,31$ , por Saidaiah et al. (2008) com  $r = 0,30$ , por Churata e Ayala-Osuna (1996) com  $r = 0,50$  e por Alvi et al. (2003) com correlação de  $0,57$ . As diferenças genéticas e ambientais podem explicar as diferentes magnitudes de correlação encontradas nesses trabalhos.

A associação média entre PRE e PROD foi de baixa intensidade ( $r = 0,03$ ) e também oscilou em relação à magnitude e direção da associação ( $-0,59 \leq r \leq 0,60$ ) (Tabela 13). Em

seleção recorrente de milho, Lima Neto e Souza Júnior (2009) também obtiveram correlação fenotípica de baixa magnitude ( $0,13 \leq r \leq 0,19$ ). Com base na reduzida associação das variáveis FM, EP, AIE e PRE com PROD e na ampla oscilação dos coeficientes estimados entre os 44 ensaios, espera-se que essas variáveis não apresentam efeitos diretos consideráveis sobre a produtividade de grãos, em relações de causa e efeito (CRUZ; REGAZZI, 1997).

A associação entre o número de plantas e a produtividade de grãos foi um pouco superior às associações destacadas anteriormente, com coeficiente de correlação médio de 0,39 e em 22 ensaios, a associação foi estatisticamente significativa (Tabela 13). Embora os coeficientes tenham oscilado em magnitude e direção da associação ( $-0,73 \leq r \leq 0,87$ ), apenas três dos 44 ensaios apresentaram associações negativas.

O número de espigas foi a variável que apresentou a maior correlação com a produtividade de grãos (Tabela 13). A correlação média foi de 0,71 e em todos os ensaios, a direção da associação foi positiva indicando que, provavelmente, cultivares com maior número de espigas sejam mais produtivas, numa relação de causa e efeito (CRUZ; REGAZZI, 1997). Ainda, com exceção dos ensaios 17 e 18, todos os demais ensaios apresentaram coeficientes de correlação estatisticamente significativos entre NE e PROD ( $0,48 \leq r \leq 0,92$ ) reforçando a possibilidade de relação de causa e efeito. Churata e Ayala-Osuna (1996) e Rangel et al. (2011) também obtiveram coeficientes de correlação fenotípicos de alta magnitude entre NE e PROD, com  $r = 0,63$  e  $r = 0,50$ , respectivamente.

A variável prolificidade apresentou associação positiva com a produtividade de grãos ( $r = 0,52$ ), com intensidade um pouco inferior a observada entre NE e PROD (Tabela 13). Em 26 ensaios, essa associação foi estatisticamente significativa, e apenas no ensaio 21, a associação entre PLI e PROD foi negativa ( $r = -0,01$ ), porém não significativa. Nos demais 43 ensaios, os valores de correlação foram positivos ( $0,11 \leq r \leq 0,94$ ). A média de correlação entre PLI e PROD foi próxima à estimada em cultivares crioulas de milho por Barros et al. (2010), com  $r = 0,53$  e da estimada em populações de milho sob seleção recorrente, por Lima Neto e Souza Júnior (2009) com  $0,45 \leq r \leq 0,53$ .

Com base nos coeficientes de correlação fenotípicos estimados nos 44 ensaios, das sete variáveis explicativas com a produtividade de grãos, pode-se verificar uma maior intensidade das associações de NE, PLI e NP sobre PROD, nessa ordem de importância (Tabela 13). A análise de trilha permitirá avaliar a proporção dos coeficientes de correlação de cada variável, que de fato, atua diretamente sobre a produtividade de grãos.

**Tabela 13** - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 44 ensaios de competição de cultivares de milho.

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>						
	FM	EP	AIE	PRE	NP	NE	PLI
1	-0,18 <sup>ns(3)</sup>	0,23 <sup>ns</sup>	0,26 <sup>ns</sup>	0,12 <sup>ns</sup>	0,31 <sup>ns</sup>	0,83 *	0,81 *
2	0,55 <sup>ns</sup>	0,40 <sup>ns</sup>	0,37 <sup>ns</sup>	0,12 <sup>ns</sup>	0,87 *	0,73 *	0,11 <sup>ns</sup>
3	0,29 <sup>ns</sup>	0,86 *	0,48 <sup>ns</sup>	-0,33 <sup>ns</sup>	0,57 <sup>ns</sup>	0,73 *	0,14 <sup>ns</sup>
4	-0,16 <sup>ns</sup>	0,50 <sup>ns</sup>	0,34 <sup>ns</sup>	-0,39 <sup>ns</sup>	0,43 <sup>ns</sup>	0,62 *	0,44 <sup>ns</sup>
5	0,03 <sup>ns</sup>	0,41 <sup>ns</sup>	0,58 <sup>ns</sup>	0,51 <sup>ns</sup>	0,19 <sup>ns</sup>	0,83 *	0,89 *
6	0,28 <sup>ns</sup>	0,19 <sup>ns</sup>	0,37 *	0,49 *	-0,02 <sup>ns</sup>	0,66 *	0,67 *
7	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 *
8	-0,12 <sup>ns</sup>	0,34 *	0,40 *	0,37 *	0,38 *	0,67 *	0,66 *
9	0,44 *	0,26 <sup>ns</sup>	0,10 <sup>ns</sup>	-0,09 <sup>ns</sup>	0,61 *	0,62 *	0,14 <sup>ns</sup>
10	-0,15 <sup>ns</sup>	0,56 *	0,54 *	0,38 *	0,44 *	0,80 *	0,79 *
11	-0,30 <sup>ns</sup>	0,01 <sup>ns</sup>	-0,14 <sup>ns</sup>	-0,13 <sup>ns</sup>	0,68 *	0,78 *	0,22 <sup>ns</sup>
12	-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 *
13	-0,08 <sup>ns</sup>	0,19 <sup>ns</sup>	-0,22 <sup>ns</sup>	-0,46 <sup>ns</sup>	0,80 *	0,89 *	0,44 <sup>ns</sup>
14	0,05 <sup>ns</sup>	-0,18 <sup>ns</sup>	-0,05 <sup>ns</sup>	0,14 <sup>ns</sup>	0,17 <sup>ns</sup>	0,70 *	0,75 *
15	0,30 <sup>ns</sup>	0,49 *	0,41 *	0,10 <sup>ns</sup>	0,03 <sup>ns</sup>	0,69 *	0,65 *
16	-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 *
17	0,29 <sup>ns</sup>	0,91 *	0,84 *	0,60 <sup>ns</sup>	-0,73 *	0,10 <sup>ns</sup>	0,85 *
18	0,41 <sup>ns</sup>	0,63 <sup>ns</sup>	0,69 *	0,54 <sup>ns</sup>	0,21 <sup>ns</sup>	0,36 <sup>ns</sup>	0,26 <sup>ns</sup>
19	0,08 <sup>ns</sup>	0,44 <sup>ns</sup>	0,28 <sup>ns</sup>	0,02 <sup>ns</sup>	0,38 <sup>ns</sup>	0,71 *	0,12 <sup>ns</sup>
20	-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>
21	0,19 <sup>ns</sup>	0,10 <sup>ns</sup>	0,14 <sup>ns</sup>	0,11 <sup>ns</sup>	0,58 *	0,67 *	-0,01 <sup>ns</sup>
22	0,09 <sup>ns</sup>	0,04 <sup>ns</sup>	0,07 <sup>ns</sup>	0,06 <sup>ns</sup>	0,18 <sup>ns</sup>	0,51 *	0,47 *
23	-0,46 <sup>ns</sup>	0,24 <sup>ns</sup>	0,11 <sup>ns</sup>	-0,10 <sup>ns</sup>	0,65 *	0,84 *	0,57 *
24	-0,39 <sup>ns</sup>	0,06 <sup>ns</sup>	0,07 <sup>ns</sup>	0,07 <sup>ns</sup>	0,72 *	0,85 *	0,22 <sup>ns</sup>
25	-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>
26	-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 *
27	-0,37 <sup>ns</sup>	-0,43 *	-0,31 <sup>ns</sup>	-0,07 <sup>ns</sup>	0,26 <sup>ns</sup>	0,48 *	0,24 <sup>ns</sup>
28	0,13 <sup>ns</sup>	0,52 *	0,60 *	0,54 *	0,63 *	0,78 *	0,46 *
29	-0,20 <sup>ns</sup>	-0,15 <sup>ns</sup>	-0,15 <sup>ns</sup>	-0,09 <sup>ns</sup>	0,12 <sup>ns</sup>	0,64 *	0,58 *
30	-0,18 <sup>ns</sup>	-0,38 <sup>ns</sup>	-0,35 <sup>ns</sup>	-0,20 <sup>ns</sup>	-0,38 <sup>ns</sup>	0,83 *	0,80 *
31	-0,15 <sup>ns</sup>	0,28 <sup>ns</sup>	0,40 <sup>ns</sup>	0,42 <sup>ns</sup>	0,69 *	0,77 *	0,31 <sup>ns</sup>
32	-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 *
33	-0,52 *	-0,01 <sup>ns</sup>	-0,36 <sup>ns</sup>	-0,37 <sup>ns</sup>	0,49 *	0,75 *	0,43 <sup>ns</sup>
34	0,38 <sup>ns</sup>	0,82 *	0,61 *	0,08 <sup>ns</sup>	0,71 *	0,92 *	0,75 *
35	-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>
36	-0,56 *	-0,11 <sup>ns</sup>	0,07 <sup>ns</sup>	0,21 <sup>ns</sup>	0,62 *	0,86 *	0,85 *
37	-0,48 *	-0,03 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,32 <sup>ns</sup>	0,56 *	0,84 *	0,46 *
38	-0,09 <sup>ns</sup>	0,19 <sup>ns</sup>	0,08 <sup>ns</sup>	-0,03 <sup>ns</sup>	0,15 <sup>ns</sup>	0,79 *	0,71 *
39	-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>
40	-0,69 *	-0,27 <sup>ns</sup>	-0,37 <sup>ns</sup>	-0,29 <sup>ns</sup>	0,30 <sup>ns</sup>	0,89 *	0,79 *
41	-0,64 *	0,51 <sup>ns</sup>	-0,21 <sup>ns</sup>	-0,53 <sup>ns</sup>	0,28 <sup>ns</sup>	0,68 *	0,65 *
42	-0,10 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,20 <sup>ns</sup>	-0,03 <sup>ns</sup>	0,12 <sup>ns</sup>	0,66 *	0,58 *
43	-0,14 <sup>ns</sup>	0,02 <sup>ns</sup>	0,03 <sup>ns</sup>	0,03 <sup>ns</sup>	0,35 <sup>ns</sup>	0,66 *	0,56 *
44	-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	-0,11	0,20	0,13	0,03	0,39	0,71	0,52

<sup>(1)</sup> Ensaios definidos 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> \*Coeficiente de correlação linear de Pearson significativo a 5% de probabilidade de erro, pelo teste t, com n-2 graus de liberdade, sendo n = número de cultivares do ensaio. <sup>ns</sup> = Não significativo.

### **Caso 1 - Ensaio com distribuição normal multivariada e com multicolinearidade**

Conforme já salientado, nos 14 ensaios do caso 1, não foi realizada transformação de dados de variáveis, pois esses ensaios apresentaram distribuição normal multivariada ( $p > 0,10$ ). Após a verificação da normalidade e a obtenção de matrizes de correlação  $X'X$  e  $X'Y$ , foi realizado o diagnóstico de multicolinearidade entre as variáveis explicativas, com base na matriz de correlação  $X'X$  de cada ensaio.

Em relação ao fator de inflação de variância (FIV), nos 14 ensaios foram verificadas médias de FIV acima de 10, ou seja, foi diagnosticado elevado grau de multicolinearidade entre as sete variáveis explicativas (GUJARATI, 2006) (Tabela 14). O menor valor de FIV foi em relação a variável florescimento masculino, que em 11 ensaios apresentou  $FIV < 10$ . Todas as demais variáveis, em todos os ensaios, superaram os valores de FIV aceitáveis, indicando problemas de alto grau de multicolinearidade na matriz de correlação  $X'X$  ( $71,85 \leq FIV \leq 71.994,38$ ).

Entre as variáveis explicativas, a altura de inserção da espiga apresentou a maior média de  $FIV = 7.963,02$  (Tabela 14). Esse elevado valor de FIV, provavelmente, está associado à elevada correlação de AIE com EP e PRE. Considerando que o FIV indica o aumento da variância associada a cada coeficiente de regressão estimado em função da presença da multicolinearidade e, sua raiz expressa o erro-padrão associado ao coeficiente de regressão ou de trilha estimado, pode-se dizer que em média, para a variável AIE, o erro-padrão associado aos coeficientes estimados aumentou 89 vezes em função do alto grau de multicolinearidade. Assim, é possível concluir que a matriz de correlação  $X'X$  esta mal condicionada e, provavelmente, seja singular. Além disso, devido ao elevado erro-padrão, maior será o intervalo de confiança e maior será a dificuldade em detectar a significância estatística dos coeficientes estimados (HAIR et al., 2009).

A modificação parcial do sistema de equações normais, pela adição de uma constante  $k$  nos elementos da diagonal da matriz de correlação  $X'X$ , pode reduzir os efeitos nocivos da multicolinearidade (CRUZ; CARNEIRO, 2006). Nesse sentido, foi verificado que após a adição de uma constante  $k = 0,10$ , todas as variáveis, em todos os 14 ensaios do caso 1, passaram a apresentar FIV dentro dos limites aceitáveis ( $1,19 \leq FIV \leq 8,06$ ) (Tabela 14). A média de FIV foi mais elevada em AIE e NE, com valores médios de 7,01 e 5,87, respectivamente.

**Tabela 14** - Fator de inflação da variância (FIV) para as variáveis explicativas em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada, sem a necessidade de transformação dos dados.

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>						
	FM	EP	AIE	PRE	NP	NE	PLI
----- Para sete variáveis explicativas -----							
1	3,81	7.806,60	9.382,54	858,42	551,06	6.481,47	5.284,62
2	26,63	1.361,19	2.110,85	592,39	5.631,10	10.062,53	1.918,07
5	9,74	4.479,39	7.038,06	1.716,00	476,41	2.844,48	2.007,98
13	1,42	1.039,76	2.253,77	1.330,63	677,38	1.204,22	543,06
17	4,03	843,78	2.547,99	729,45	5.759,02	5.650,27	5.529,57
18	84,87	31.359,62	71.994,38	18.721,15	3.926,30	4.099,91	678,00
19	4,62	1.417,85	3.987,46	1.070,53	2.131,75	1.012,37	1.949,45
21	1,44	851,85	1.744,83	656,83	234,54	179,91	75,93
23	2,81	1.206,99	1.978,55	893,37	765,52	1.371,86	499,21
24	1,64	571,39	1.671,13	687,78	992,85	979,19	273,30
27	3,17	2.143,12	5.179,34	1.547,03	510,45	541,53	439,34
28	1,50	129,75	385,89	117,26	260,53	324,39	71,85
31	2,31	253,91	921,21	326,58	554,76	735,63	202,17
41	10,07	205,62	286,33	366,02	380,60	4.626,55	3.924,73
Média	11,29	3.833,63	7.963,02	2.115,25	1.632,30	2.865,31	1.671,23
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----							
1	2,57	6,01	6,83	2,88	2,68	6,68	5,96
2	1,88	4,89	6,97	3,95	4,71	7,53	3,27
5	2,40	4,74	6,90	3,52	2,25	6,82	5,27
13	1,22	3,85	6,23	4,32	4,40	6,34	3,62
17	2,14	5,62	8,06	4,49	5,42	4,65	5,70
18	4,32	5,71	7,62	4,55	6,52	5,81	2,93
19	3,07	4,51	7,79	4,58	5,69	3,21	5,75
21	1,19	3,89	6,74	3,32	6,01	4,85	2,67
23	2,19	4,22	6,40	3,57	4,27	6,62	3,14
24	1,49	3,67	7,17	3,83	5,79	5,75	2,75
27	2,47	4,30	7,49	3,36	4,68	4,86	4,07
28	1,51	3,86	7,57	3,71	5,24	6,27	2,35
31	1,62	3,59	7,57	4,37	5,08	6,20	2,55
41	2,59	3,72	4,84	5,96	1,75	6,57	5,87
Média	2,19	4,47	7,01	4,03	4,61	5,87	3,99
----- Para cinco variáveis explicativas, excluindo AIE e NE -----							
1	3,08	2,33	-	3,23	2,52	-	2,06
2	2,16	1,31	-	4,37	2,34	-	3,53
5	3,01	1,09	-	3,35	1,44	-	1,79
13	1,13	1,65	-	1,26	1,73	-	1,16
17	2,45	7,34	-	3,31	2,81	-	4,39
18	9,91	8,67	-	4,39	4,32	-	2,81
19	4,03	2,83	-	4,17	3,04	-	5,08
21	1,10	1,04	-	1,12	1,37	-	1,35
23	2,33	1,43	-	1,71	1,38	-	1,12
24	1,46	1,70	-	1,30	1,49	-	1,86
27	2,77	2,35	-	1,59	1,27	-	1,41
28	1,47	1,85	-	1,86	1,23	-	1,22
31	1,70	1,78	-	2,53	1,17	-	1,09
41	2,57	2,33	-	4,15	1,21	-	1,37
Média	2,80	2,69	-	2,74	1,95	-	2,16

<sup>(1)</sup> Ensaios definidos 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.

Conforme discutido na metodologia, foi constatada uma alta correlação de AIE com EP e PRE e de NE com NP e PLI. Com base nessas informações obtidas no diagnóstico de multicolinearidade, as variáveis AIE e NE foram eliminadas do estudo, a fim de reduzir o grau de multicolinearidade. Constatou-se que a eliminação dessas variáveis foi eficiente na redução do grau de multicolinearidade de todos os ensaios, nas cinco variáveis que permaneceram no estudo ( $1,04 \leq FIV \leq 9,91$ ) (Tabela 14 - terço inferior da tabela).

Com exceção da variável florescimento masculino, todas as outras quatro variáveis apresentaram FIV inferiores aos observados na utilização da constante  $k = 0,10$ . Pelo fato de FM não estar altamente correlacionada às demais variáveis explicativas, a eliminação de variáveis reduziu em menores escalas seus valores de FIV. As médias gerais de FIV, com base em todos os ensaios e variáveis, foram de 2.870,29, 4,60 e 2,47, respectivamente, para as sete variáveis explicativas, para as sete variáveis explicativas com adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  e para cinco variáveis explicativas, após a eliminação de AIE e NE.

Verificou-se num primeiro momento, que tanto a adição de uma constante  $k = 0,10$  nos elementos da diagonal da matriz de correlação  $X'X$ , quanto a eliminação de AIE e NE foram eficientes na redução do grau de multicolinearidade. As principais desvantagens da utilização da constante  $k$  são a redução do coeficiente de determinação da análise de trilha e a utilização de estimadores tendenciosos (CRUZ; CARNEIRO, 2006). Em relação à eliminação de variáveis, o maior problema pode estar relacionado ao erro de especificação (GUJARATI, 2006; HAIR et al., 2009). Como a variável NE apresentou a maior magnitude dos valores de coeficientes de correlação com a produtividade de grãos (Tabela 13), sua eliminação pode reduzir a capacidade preditiva das análises de trilha dos ensaios.

Em relação ao diagnóstico do grau de multicolinearidade com base na tolerância, foram confirmadas as afirmações realizadas com base no FIV. Considerando as matrizes de correlação  $X'X$  dos 14 ensaios com sete variáveis explicativas, com exceção da variável FM, todas as demais apresentaram tolerância  $\leq 0,01$  (Tabela 15). Esses valores indicam que menos de 1% da variação de uma variável independe das demais variáveis explicativas, ou seja, os coeficientes de ajuste da regressão entre uma variável explicativa com as demais superam 0,99. Dessa forma, ficou evidenciada a elevada variância compartilhada e a presença de variáveis com informações redundantes. Considerando-se o critério de decisão de que tolerância menor do que 0,10 é um indicativo de multicolinearidade severa (GUJARATI, 2006; HAIR et al., 2009), apenas a variável FM não teve alto grau de multicolinearidade (Tolerância = 0,34).

**Tabela 15** - Tolerância para as variáveis explicativas, número de condição (NC) e determinante da matriz de correlação  $X'X$  (DM), em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada sem a necessidade de transformação dos dados.

Ensaio <sup>(1)</sup>	Tolerância para as variáveis explicativas <sup>(2)</sup>							NC	DM
	FM	EP	AIE	PRE	NP	NE	PLI		
----- Para sete variáveis explicativas -----									
1	0,26	0,00	0,00	0,00	0,00	0,00	0,00	65.869,32	0,00
2	0,04	0,00	0,00	0,00	0,00	0,00	0,00	73.498,21	0,00
5	0,10	0,00	0,00	0,00	0,00	0,00	0,00	50.447,51	0,00
13	0,70	0,00	0,00	0,00	0,00	0,00	0,00	11.662,77	0,00
17	0,25	0,00	0,00	0,00	0,00	0,00	0,00	79.405,90	0,00
18	0,01	0,00	0,00	0,00	0,00	0,00	0,00	438.090,69	0,00
19	0,22	0,00	0,00	0,00	0,00	0,00	0,00	32.711,33	0,00
21	0,69	0,00	0,00	0,00	0,00	0,01	0,01	8.068,85	0,00
23	0,36	0,00	0,00	0,00	0,00	0,00	0,00	12.167,40	0,00
24	0,61	0,00	0,00	0,00	0,00	0,00	0,00	9.566,72	0,00
27	0,32	0,00	0,00	0,00	0,00	0,00	0,00	30.117,18	0,00
28	0,67	0,01	0,00	0,01	0,00	0,00	0,01	2.552,34	0,00
31	0,43	0,00	0,00	0,00	0,00	0,00	0,00	6.153,51	0,00
41	0,10	0,00	0,00	0,00	0,00	0,00	0,00	27.981,54	0,00
Média	0,34	0,00	0,00	0,00	0,00	0,00	0,00	60.592,38	0,00
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----									
1	0,39	0,17	0,15	0,35	0,37	0,15	0,17	28,65	0,01
2	0,53	0,20	0,14	0,25	0,21	0,13	0,31	40,10	0,01
5	0,42	0,21	0,14	0,28	0,44	0,15	0,19	30,65	0,02
13	0,82	0,26	0,16	0,23	0,23	0,16	0,28	26,10	0,03
17	0,47	0,18	0,12	0,22	0,18	0,22	0,18	41,30	0,01
18	0,23	0,18	0,13	0,22	0,15	0,17	0,34	35,10	0,01
19	0,33	0,22	0,13	0,22	0,18	0,31	0,17	33,18	0,01
21	0,84	0,26	0,15	0,30	0,17	0,21	0,37	25,71	0,05
23	0,46	0,24	0,16	0,28	0,23	0,15	0,32	27,17	0,03
24	0,67	0,27	0,14	0,26	0,17	0,17	0,36	29,19	0,02
27	0,40	0,23	0,13	0,30	0,21	0,21	0,25	32,98	0,02
28	0,66	0,26	0,13	0,27	0,19	0,16	0,42	35,50	0,02
31	0,62	0,28	0,13	0,23	0,20	0,16	0,39	28,78	0,02
41	0,39	0,27	0,21	0,17	0,57	0,15	0,17	31,74	0,02
Média	0,52	0,23	0,14	0,26	0,25	0,18	0,28	31,87	0,02
----- Para cinco variáveis explicativas, excluindo AIE e NE -----									
1	0,32	0,43	-	0,31	0,40	-	0,48	13,57	0,12
2	0,46	0,76	-	0,23	0,43	-	0,28	22,00	0,14
5	0,33	0,92	-	0,30	0,69	-	0,56	12,48	0,24
13	0,89	0,61	-	0,79	0,58	-	0,86	5,17	0,47
17	0,41	0,14	-	0,30	0,36	-	0,23	33,81	0,03
18	0,10	0,12	-	0,23	0,23	-	0,36	51,29	0,03
19	0,25	0,35	-	0,24	0,33	-	0,20	25,41	0,03
21	0,91	0,96	-	0,89	0,73	-	0,74	3,25	0,65
23	0,43	0,70	-	0,58	0,73	-	0,90	7,69	0,40
24	0,69	0,59	-	0,77	0,67	-	0,54	6,99	0,35
27	0,36	0,43	-	0,63	0,79	-	0,71	9,54	0,21
28	0,68	0,54	-	0,54	0,81	-	0,82	6,31	0,33
31	0,59	0,56	-	0,40	0,85	-	0,92	8,25	0,37
41	0,39	0,43	-	0,24	0,83	-	0,73	17,19	0,17
Média	0,49	0,54	-	0,46	0,60	-	0,59	15,92	0,25

<sup>(1)</sup> Ensaios definidos 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.



Após a adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , todas as variáveis, em todos os ensaios, apresentaram valores de tolerância aceitáveis, ou seja, superiores a 0,10 ( $0,12 \leq \text{tolerância} \leq 0,84$ ) (Tabela 15). A eliminação de variáveis também foi eficiente na redução da variância compartilhada e no grau de multicolinearidade ( $0,10 \leq \text{tolerância} \leq 0,96$ ). As médias gerais de tolerância, com base em todos os ensaios e variáveis, foram de 0,05, 0,27 e 0,54, respectivamente, para as sete variáveis explicativas, para as sete variáveis explicativas, com adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , e para cinco variáveis explicativas, após a eliminação de AIE e NE.

Com base na classificação do número de condição, estabelecida por Montgomery e Peck (1982), constatou-se que houve multicolinearidade severa em todos os 14 ensaios, quando se utilizou a matriz de correlação  $X'X$  com sete variáveis explicativas ( $2.552,34 \leq \text{NC} \leq 438.090,69$ ) (Tabela 15). Tanto a adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  ( $25,71 \leq \text{NC} \leq 41,30$ ), quanto a eliminação de variáveis ( $3,25 \leq \text{NC} \leq 51,29$ ) foram eficientes para a redução do número de condição. Em média, o número de condição foi de 60.592,38, 31,87 e 15,92, respectivamente, para as sete variáveis explicativas, para as sete variáveis explicativas, com adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , e para cinco variáveis explicativas, após a eliminação de AIE e NE.

Por fim, foi verificado que em todos os ensaios o determinante da matriz de correlação  $X'X$  para as sete variáveis explicativas foi igual a zero (Tabela 15), indicando elevado grau de multicolinearidade, problema de condicionamento da matriz e sua possível singularidade. Dessa forma, o cálculo da inversa da matriz de correlação  $X'X$ , necessária para a estimação dos efeitos diretos e indiretos da análise de trilha pelo sistema de equações normais  $\hat{\beta} = X'X^{-1} \times X'Y$  é dificultado (CRUZ; CARNEIRO, 2006). A adição da constante  $k = 0,10$  elevou os valores de determinante, que foram em média de 0,02, e a eliminação das variáveis AIE e NE elevou o determinante da matriz, em todos os ensaios, em relação à adição de  $k = 0,10$ , com média 0,25.

Assim, com base na análise do fator de inflação de variância, da tolerância, do número de condição e do determinante da matriz, verificou-se elevado grau de multicolinearidade entre as sete variáveis explicativas (Tabelas 14 e 15). Esses resultados indicam possíveis efeitos adversos da multicolinearidade sobre os efeitos diretos e indiretos da análise de trilha, nos 14 ensaios. Também, verificou-se que a adição da constante  $k = 0,10$  e a eliminação de variáveis, foram eficientes na redução do grau de multicolinearidade. No entanto, o menor valor de FIV associado ao maior valor de tolerância, menor valor de número de condição e maior determinante da matriz, no caso de eliminação de variáveis em relação à adição de  $k =$

0,10, sugere que a eliminação de variáveis deva ser, preferencialmente, utilizada para a redução do grau de multicolinearidade entre variáveis explicativas.

Como já salientado, o florescimento masculino não estava diretamente relacionado às demais variáveis explicativas e não apresentava indicadores de alto grau de multicolinearidade. Por isso, mesmo na análise de trilha tradicional, com alto grau de multicolinearidade entre as sete variáveis explicativas, seus efeitos diretos apresentaram coerência com os coeficientes de correlação de FM com PROD, com exceção do ensaio 41 (Tabelas 13 e 16). No entanto, a oscilação entre efeitos diretos positivos e negativos de baixa magnitude, sugere que essa variável não é uma variável de causa e efeito.

As demais seis variáveis explicativas apresentaram elevada multicolinearidade (Tabelas 14 e 15). Essas variáveis também apresentaram ampla variação dos efeitos diretos estimados. A variável estatura de plantas apresentou correlação média de 0,32 com a produtividade de grãos nos 14 ensaios do caso 1 e oscilou entre  $-0,43 \leq r \leq 0,91$  (Tabela 13). Já o efeito direto médio dessa variável sobre a produtividade de grãos foi de - 0,33. No entanto, a interpretação desse valor médio de efeito direto não apresenta sentido biológico devido a elevada amplitude dos coeficientes estimados ( $-16,68 \leq EP \leq 17,52$ ) (Tabela 16).

No caso de AIE, a média dos efeitos diretos foi de 0,19, próximo a média de coeficiente de correlação entre AIE e PROD ( $r = 0,26$ ). Mas, assim como para EP, foi verificada uma elevada oscilação dos efeitos diretos ( $-25,90 \leq AIE \leq 18,73$ ), impedindo a interpretação dos mesmos. A posição relativa de espiga apresentou comportamento similar, com efeito direto médio de 0,12, próximo a média de correlação com PROD ( $r = 0,13$ ) e oscilando entre  $-5,72 \leq PRE \leq 13,23$  (Tabela 16). A mesma situação foi observada para a variável NP com efeito direto médio de 0,37, similar a média de coeficiente de correlação com PROD ( $r = 0,42$ ) e efeitos diretos variando de  $-16,73 \leq NP \leq 10,06$ .

Em relação ao número de espigas, também se constatou elevada variabilidade entre os efeitos diretos estimados nos 14 ensaios ( $-21,94 \leq NE \leq 16,29$ ) (Tabela 16). A média de correlação entre NE e PROD, nos 14 ensaios, foi de 0,68, sendo que todos os coeficientes foram positivos e de alta magnitude, sugerindo uma relação de causa e efeito (Tabela 13). No entanto, a média dos efeitos diretos foi de - 1,42, demonstrando os efeitos adversos do alto grau de multicolinearidade sobre as estimativas dos coeficientes de trilha. Por fim, a variável PLI repetiu o comportamento das demais variáveis explicativas, com ampla variação dos efeitos diretos estimados, com média de 2,74.

As maiores oscilações foram observadas nas variáveis AIE, NE e EP (Tabela 16), que foram as variáveis que apresentaram os maiores problemas de multicolinearidade (Tabela 14).

Entre os ensaios, o ensaio 18 apresentou os maiores escores positivos e negativos de efeitos diretos. Esse ensaio teve os maiores valores de FIV e o maior número de condição (Tabelas 14 e 15), evidenciando a relação do grau de multicolinearidade com a amplitude de variação dos efeitos diretos estimados. Com base na análise de trilha tradicional, com alto grau de multicolinearidade, é possível concluir que os efeitos diretos e indiretos (não apresentados), estimados a partir de matrizes de correlação mal condicionadas, não apresentam confiabilidade e não devem ser utilizados.

A realização da análise de trilha sob multicolinearidade, com adição de um valor  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , permitiu a estabilização dos efeitos diretos e indiretos (não apresentados). Os maiores coeficientes de correlação médios obtidos nos 14 ensaios entre NE e PROD ( $r = 0,68$ ), entre PLI e PROD ( $r = 0,42$ ) e entre NP e PROD ( $r = 0,42$ ) (Tabela 13) resultaram nos maiores efeitos diretos médios de NE, PLI e NP sobre a produtividade de grãos, com coeficientes de 0,33, 0,21 e 0,16, respectivamente (Tabela 16). Além disso, nos 14 ensaios, os efeitos diretos de NE sobre PROD foram positivos ( $0,01 \leq NE \leq 0,51$ ) e em apenas um e três ensaios, os efeitos diretos de PLI e NP sobre PROD foram negativos, respectivamente. Esses resultados confirmam que NE, PLI e NP, são as variáveis que apresentam maiores efeitos diretos sobre a produtividade de grãos, nessa ordem de importância, com relação de causa e efeito. As demais variáveis apresentaram menores valores médios de efeitos diretos, oscilando entre efeitos positivos e negativos.

A análise de trilha tradicional, com eliminação de variáveis, também foi eficiente na estabilização dos efeitos diretos e indiretos (não apresentados). Em apenas dois casos, os efeitos diretos estimados superaram a unidade (Tabela 16). Como ocorreu a eliminação de variáveis, o efeito direto da variável NE foi realocado em PLI e NP. Assim, maiores escores de efeito direto foram de PLI sobre PROD (0,53) e de NP sobre PROD (0,48). O efeito direto de AIE sobre PROD passou a ser predito pela variável EP (0,23). As demais variáveis apresentaram efeitos diretos similares entre as estimativas realizadas com base na análise de trilha sob multicolinearidade e na análise de trilha tradicional, com eliminação de variáveis.

Com base nesses 14 ensaios, nota-se que tanto a eliminação de variáveis, como a análise de trilha sob multicolinearidade, com a adição de um valor  $k = 0,10$ , são eficientes para reduzir os efeitos adversos da multicolinearidade nas estimativas dos efeitos diretos das variáveis explicativas sobre a produtividade de grãos. O maior coeficiente de determinação e o menor efeito da variável residual na análise de trilha tradicional, com a eliminação de variáveis, em relação à análise de trilha sob multicolinearidade, associado à maior redução do grau de multicolinearidade (Tabelas 14 e 15), sugerem a utilização preferencial dessa técnica.

**Tabela 16** - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos, coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual (RES) em 14 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada, sem a necessidade de transformação dos dados.

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>							$R^2$	RES
	FM	EP	AIE	PRE	NP	NE	PLI		
---- Análise de trilha tradicional com sete variáveis explicativas e alto grau de multicolinearidade ----									
1	-0,38	-16,68	18,73	-5,72	1,96	-6,74	6,80	0,80	0,45
2	0,30	0,99	-1,15	0,20	5,24	-5,56	2,68	0,95	0,22
5	0,17	-9,02	11,64	-5,53	1,25	-3,14	3,35	0,92	0,29
13	0,15	-0,34	0,29	-0,38	3,00	-2,88	2,34	0,90	0,32
17	-0,02	-1,86	4,20	-2,55	10,06	-9,99	10,57	0,99	0,10
18	-0,65	17,52	-25,90	13,23	-16,73	16,29	-6,89	0,64	0,60
19	0,32	6,00	-9,14	4,29	7,59	-4,33	8,08	0,99	0,12
21	0,30	-0,80	1,23	-0,58	-3,63	3,91	-2,11	0,62	0,62
23	-0,24	5,29	-6,47	4,47	-1,03	2,01	-0,74	0,80	0,44
24	0,04	0,38	-0,13	-0,11	0,49	0,41	0,43	0,79	0,46
27	-0,08	-3,24	4,62	-2,35	-3,82	4,34	-3,58	0,39	0,78
28	-0,04	-1,13	2,41	-1,15	-2,51	3,43	-1,24	0,75	0,50
31	-0,33	0,29	-0,73	0,99	-3,17	4,25	-1,78	0,80	0,45
41	-1,18	-2,02	3,13	-3,09	6,45	-21,94	20,51	0,86	0,37
Média	-0,12	-0,33	0,19	0,12	0,37	-1,42	2,74	0,80	0,41
-Análise de trilha sob multicolinearidade, com acréscimo de valor $k = 0,10$ na diagonal da matriz $X'X$ -									
1	-0,26	0,19	0,18	-0,06	0,00	0,35	0,38	0,72	0,53
2	0,12	0,15	-0,06	-0,36	0,60	0,40	-0,07	0,86	0,37
5	-0,02	0,13	0,17	0,07	-0,13	0,35	0,45	0,85	0,39
13	0,15	-0,02	-0,09	-0,11	0,46	0,41	0,14	0,84	0,40
17	-0,03	0,25	0,23	-0,01	-0,24	0,01	0,32	0,85	0,39
18	0,08	0,30	0,23	0,02	-0,02	0,17	0,18	0,50	0,71
19	0,15	0,49	-0,05	-0,47	0,13	0,51	0,38	0,66	0,58
21	0,18	0,05	0,09	0,10	0,23	0,49	0,07	0,52	0,69
23	-0,29	0,16	0,11	0,04	0,21	0,34	0,24	0,74	0,51
24	0,00	0,22	0,02	-0,16	0,34	0,49	0,28	0,73	0,52
27	-0,13	-0,24	-0,07	0,16	0,12	0,26	0,06	0,34	0,81
28	-0,05	0,11	0,17	0,08	0,19	0,36	0,18	0,68	0,56
31	-0,30	-0,09	0,16	0,34	0,23	0,37	0,19	0,72	0,53
41	-0,50	0,23	0,13	-0,03	0,15	0,17	0,18	0,70	0,54
Média	-0,07	0,14	0,09	-0,03	0,16	0,33	0,21	0,69	0,54
--- Análise de trilha tradicional com eliminação de variáveis e reduzido grau de multicolinearidade ---									
1	-0,31	0,42	-	-0,03	0,06	-	0,74	0,76	0,49
2	0,02	0,09	-	-0,61	1,08	-	0,26	0,95	0,23
5	0,02	0,27	-	0,11	0,00	-	0,81	0,89	0,33
13	0,18	-0,13	-	-0,17	0,84	-	0,41	0,89	0,33
17	-0,02	0,37	-	0,08	-0,24	-	0,37	0,88	0,35
18	-0,08	0,64	-	0,02	0,28	-	0,31	0,52	0,69
19	0,38	0,61	-	-0,86	0,95	-	1,50	0,81	0,43
21	0,18	0,12	-	0,17	0,82	-	0,40	0,53	0,68
23	-0,36	0,30	-	0,16	0,45	-	0,46	0,77	0,48
24	0,05	0,30	-	-0,19	0,91	-	0,64	0,79	0,46
27	-0,14	-0,32	-	0,16	0,40	-	0,30	0,35	0,80
28	-0,07	0,22	-	0,16	0,52	-	0,37	0,70	0,55
31	-0,40	-0,09	-	0,56	0,54	-	0,43	0,77	0,48
41	-0,61	0,44	-	0,26	0,18	-	0,35	0,75	0,50
Média	-0,08	0,23	-	-0,01	0,48	-	0,53	0,74	0,49

<sup>(1)</sup> Ensaios definidos 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.

## **Caso 2 - Ensaios com distribuição normal multivariada após a transformação de variáveis e com multicolinearidade**

Nesses 13 ensaios que apresentaram distribuição normal multivariada após a transformação de variáveis, os maiores coeficientes de correlação foram entre NE e PROD ( $r = 0,75$ ), PLI e PROD ( $r = 0,58$ ) e entre NP e PROD ( $r = 0,49$ ) (Tabela 17). Da mesma forma que verificado para os 14 ensaios do caso 1, sugere-se que essas variáveis sejam de causa e efeito, nessa ordem de importância. A variável FM apresentou correlação negativa com PROD em 12 ensaios, mas seu coeficiente de correlação médio foi de baixa magnitude ( $r = -0,28$ ). As demais variáveis também apresentaram coeficientes de correlação com a produtividade de grãos de baixa magnitude ( $r \leq 0,16$ ).

Como algumas variáveis, em alguns ensaios, não foram transformadas, seus coeficientes de correlação com a produtividade de grãos são idênticos antes e após a transformação de dados. Os altos coeficientes de correlação ( $r \geq 0,96$ ) ajustados entre os valores de correlação estimados antes e após a transformação de variáveis e as médias semelhantes entre os dois grupos de coeficientes indicam que não houveram alterações significativas da transformação de dados sobre as matrizes de correlação (Tabela 17).

Nos 13 ensaios, o fator de inflação de variância (FIV) da variável FM foi inferior a 10, tanto antes como após a transformação de dados, ou seja, sem e com distribuição normal multivariada (Tabela 18). Assim como no caso 1, as demais variáveis explicativas apresentaram elevado valor de FIV em todos os ensaios ( $37,1 \leq FIV \leq 16.385,7$ ), sendo que as variáveis NE, PLI e AIE apresentaram as maiores médias de FIV, nessa ordem. Na média geral das seis variáveis com FIV maior que 10, foram verificadas reduções expressivas de FIV, indicando que a transformação dos dados a fim de obter a distribuição normal multivariada contribuiu na melhoria das associações entre as variáveis (HAIR et al., 2009) e na redução do grau de multicolinearidade na matriz de correlação  $X'X$ .

Assim como nos 14 ensaios do caso 1, a utilização da constante  $k = 0,10$  foi eficiente na redução de FIV em todas as variáveis dos 13 ensaios do caso 2 (Tabela 18). De modo geral, as variáveis transformadas apresentaram FIV um pouco inferiores em relação às variáveis originais nessa condição. A eliminação das variáveis AIE e NE também foi eficiente na redução do fator de inflação de variância, sendo os valores de FIV antes e após a transformação de dados similares entre si. Com exceção da variável FM, todas as demais variáveis apresentaram FIV menores quando se realizou a eliminação de variáveis em relação

à adição da constante  $k = 0,10$  na diagonal da matriz  $X'X$ , sugerindo a utilização preferencial da eliminação de variáveis para a redução do grau de multicolinearidade.

**Tabela 17** - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 13 ensaios de competição de cultivares de milho, que apresentaram distribuição normal multivariada após a transformação de variáveis, e valores de coeficiente de correlação linear de Pearson ajustado entre os valores antes e após a transformação de variáveis (r entre grupos).

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>						
	FM	EP	AIE	PRE	NP	NE	PLI
----- Coeficientes de correlação de Pearson antes da transformação de variáveis -----							
7	0,27 <sup>ns(3)</sup>	0,28 <sup>ns</sup>	0,29 <sup>ns</sup>	0,17 <sup>ns</sup>	0,61 *	0,75 *	0,53 *
10	-0,15 <sup>ns</sup>	0,56 *	0,54 *	0,38 *	0,44 *	0,80 *	0,79 *
12	-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 *
16	-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 *
20	-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>
25	-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>
26	-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 *
32	-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 *
35	-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>
36	-0,56 *	-0,11 <sup>ns</sup>	0,07 <sup>ns</sup>	0,21 <sup>ns</sup>	0,62 *	0,86 *	0,85 *
37	-0,48 *	-0,03 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,32 <sup>ns</sup>	0,56 *	0,84 *	0,46 *
39	-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>
44	-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
----- Coeficientes de correlação de Pearson após a transformação de variáveis -----							
7	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 *
10	-0,16 <sup>ns</sup>	0,52 *	0,48 *	0,38 *	0,44 *	0,80 *	0,78 *
12	-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 *
16	-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 *
20	-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>
25	-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 *
26	-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 *
32	-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 *
35	-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>
36	-0,56 *	-0,11 <sup>ns</sup>	0,07 <sup>ns</sup>	0,21 <sup>ns</sup>	0,62 *	0,87 *	0,86 *
37	-0,48 *	-0,03 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,32 <sup>ns</sup>	0,56 *	0,84 *	0,46 *
39	-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>
44	-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>(4)</sup>	0,96 *	0,99 *	1,00 *	1,00 *	0,98 *	0,98 *	0,98 *

<sup>(1)</sup> Ensaios definidos 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> \*Coeficiente de correlação linear de Pearson significativo a 5% de probabilidade de erro, pelo teste t, com n-2 graus de liberdade, sendo n = número de cultivares do Ensaio. <sup>ns</sup> = Não significativo. <sup>(4)</sup> \*Coeficiente de correlação linear de Pearson significativo a 5% de probabilidade de erro, pelo teste t, com 11 graus de liberdade.

**Tabela 18** - Fator de inflação da variância (FIV) para variáveis explicativas em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados.

Ensaio (1)	Variáveis explicativas <sup>(2)</sup>													
	FM		EP		AIE		PRE		NP		NE		PLI	
	NN <sup>(3)</sup>	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO
	----- Para sete variáveis explicativas -----													
7	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
10	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
12	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
16	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
20	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
25	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
26	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
32	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
35	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
36	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
37	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
39	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
44	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
	----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----													
7	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
10	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
12	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
16	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
20	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
25	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
26	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
32	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
35	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
36	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
37	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
39	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
44	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
	----- Para cinco variáveis explicativas, excluindo AIE e NE -----													
7	1,3	1,3	1,1	1,1	-	-	1,2	1,2	1,0	1,0	-	-	1,1	1,2
10	1,5	1,4	1,4	1,3	-	-	2,1	2,1	1,6	1,6	-	-	1,3	1,2
12	1,5	1,5	1,5	1,3	-	-	1,1	1,1	1,6	1,5	-	-	1,5	1,2
16	1,8	1,8	1,6	1,6	-	-	2,3	2,3	1,4	1,3	-	-	1,5	1,4
20	2,1	2,0	1,3	1,3	-	-	2,3	2,1	1,4	1,3	-	-	1,1	1,1
25	1,6	1,8	1,1	1,3	-	-	1,1	1,1	1,6	1,6	-	-	1,1	1,2
26	1,3	1,3	1,3	1,3	-	-	1,0	1,0	3,0	2,9	-	-	2,8	2,6
32	2,4	2,3	1,4	1,5	-	-	1,5	1,4	4,8	5,2	-	-	5,4	5,7
35	1,1	1,1	1,1	1,1	-	-	1,1	1,1	1,2	1,1	-	-	1,2	1,2
36	1,9	1,8	1,7	1,7	-	-	1,7	1,6	2,1	2,1	-	-	2,1	2,0
37	1,5	1,6	1,1	1,1	-	-	1,2	1,2	1,6	1,6	-	-	1,2	1,2
39	3,5	3,5	2,5	2,5	-	-	3,2	3,2	1,4	1,4	-	-	2,7	2,7
44	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

(1) Ensaios definidos na Tabela 1. (2) 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. (3) NN: distribuição multivariada não-normal, NO: distribuição multivariada normal após a transformação de variáveis.

Com exceção de FM, todas as demais variáveis apresentaram tolerância  $\leq 0,03$  (Tabela 19). Esses resultados estão de acordo com os altos valores de FIV verificados nessas variáveis (Tabela 18) e com os baixos valores de tolerância observados no caso 1 (Tabela 15). Assim, confirma-se a existência de multicolinearidade severa entre as variáveis EP, AIE, PRE, NP, NE e PLI e, nesse sentido, a transformação dos dados contribuiu para o aumento da tolerância média dessas variáveis (Tabela 19). No entanto, em apenas seis das 78 condições (13 ensaios  $\times$  6 variáveis com elevado grau de multicolinearidade), a tolerância após a transformação de variáveis foi superior a 0,10, ou seja, em poucos casos, a transformação de variáveis foi suficiente para a redução expressiva do grau de multicolinearidade.

Após a adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , todas as variáveis em todos os ensaios, tanto antes como após a transformação de variáveis, apresentaram valores de tolerância superiores a 0,10 ( $0,13 \leq \text{tolerância} \leq 0,83$ ) (Tabela 19). A eliminação de variáveis também foi eficiente na redução da variância compartilhada e no grau de multicolinearidade ( $0,18 \leq \text{tolerância} \leq 0,97$ ). Com exceção da variável NP no ensaio 33, em todas as variáveis com alto grau de multicolinearidade (EP, AIE, PRE, NP, NE e PLI), nos 13 ensaios, a eliminação de variáveis aumentou a tolerância em proporções maiores ao aumento verificado na adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ . Como maiores valores de tolerância indicam menores graus de multicolinearidade, verificou-se que a eliminação de variáveis tende a ser mais eficiente na redução da multicolinearidade.

Em relação ao número de condição, todos os 13 ensaios apresentaram multicolinearidade severa antes da transformação de dados ( $2.123,13 \leq NC \leq 86.389,83$ ) (Tabela 20). Após a transformação de dados, de modo geral, ocorreu uma redução dos valores de número de condição, sendo que quatro ensaios apresentaram multicolinearidade moderada a severa ( $100 \leq NC \leq 1.000$ ) e os demais nove ensaios apresentaram multicolinearidade severa ( $NC > 1.000$ ). Assim, a média de NC nos ensaios com distribuição multivariada não-normal foi de 19.008,58 e com distribuição normal multivariada de 4.362,77. Esses resultados reforçam as inferências supracitadas de que a transformação de dados pode contribuir para a redução do grau de multicolinearidade, nos casos em que a multicolinearidade é severa.

O número de condição também confirmou as inferências de que tanto a adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  ( $22,89 \leq NC \leq 38,75$ ), quanto a eliminação de variáveis ( $2,62 \leq NC \leq 25,45$ ) são eficientes para a redução do grau de multicolinearidade (Tabela 20). A eliminação das variáveis AIE e NE foi mais eficiente em reduzir o grau de multicolinearidade em relação à adição de  $k = 0,10$ , sendo que em ambas, os valores de número de condição antes e após a transformação de dados foram semelhantes.



**Tabela 19** - Tolerância para as variáveis explicativas em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados.

Ensaio ( <sup>1</sup> )	Variáveis explicativas( <sup>2</sup> )													
	FM		EP		AIE		PRE		NP		NE		PLI	
	NN( <sup>3</sup> )	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO
----- Para sete variáveis explicativas -----														
7	0,67	0,75	0,01	0,01	0,00	0,00	0,00	0,01	0,01	0,04	0,00	0,02	0,01	0,08
10	0,64	0,51	0,01	0,11	0,00	0,04	0,01	0,07	0,01	0,03	0,00	0,01	0,01	0,02
12	0,52	0,58	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,07	0,01	0,04	0,02	0,11
16	0,53	0,53	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,19	0,01	0,10	0,01	0,11
20	0,46	0,46	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,06	0,00	0,03	0,00	0,07
25	0,49	0,28	0,00	0,02	0,00	0,01	0,00	0,02	0,00	0,10	0,00	0,02	0,00	0,02
26	0,44	0,38	0,00	0,23	0,00	0,05	0,00	0,05	0,00	0,05	0,00	0,02	0,00	0,06
32	0,40	0,39	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00
35	0,80	0,86	0,00	0,02	0,00	0,01	0,00	0,03	0,00	0,04	0,00	0,04	0,01	0,14
36	0,42	0,43	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,03	0,00	0,00	0,00	0,00
37	0,51	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,03	0,03
39	0,24	0,23	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,05	0,00	0,06
44	0,74	0,68	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,02	0,00	0,01
Média	0,53	0,51	0,00	0,03	0,00	0,01	0,00	0,02	0,01	0,06	0,00	0,03	0,01	0,05
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----														
7	0,75	0,75	0,29	0,29	0,15	0,15	0,26	0,26	0,24	0,24	0,15	0,17	0,32	0,40
10	0,65	0,67	0,28	0,36	0,14	0,18	0,25	0,25	0,28	0,30	0,15	0,15	0,26	0,27
12	0,64	0,62	0,34	0,35	0,17	0,17	0,19	0,19	0,26	0,27	0,14	0,17	0,29	0,37
16	0,56	0,56	0,27	0,27	0,13	0,13	0,25	0,25	0,33	0,40	0,19	0,26	0,19	0,27
20	0,51	0,51	0,24	0,24	0,14	0,14	0,28	0,28	0,23	0,28	0,15	0,18	0,32	0,33
25	0,61	0,55	0,27	0,25	0,16	0,17	0,25	0,27	0,39	0,42	0,17	0,18	0,18	0,19
26	0,72	0,63	0,46	0,54	0,16	0,20	0,18	0,22	0,24	0,25	0,13	0,14	0,22	0,26
32	0,46	0,48	0,27	0,27	0,17	0,17	0,23	0,23	0,25	0,24	0,13	0,13	0,15	0,16
35	0,82	0,83	0,21	0,24	0,15	0,15	0,38	0,35	0,17	0,20	0,18	0,21	0,48	0,47
36	0,54	0,56	0,20	0,20	0,23	0,23	0,23	0,23	0,42	0,41	0,14	0,14	0,17	0,18
37	0,63	0,62	0,28	0,28	0,17	0,17	0,22	0,22	0,20	0,20	0,19	0,18	0,34	0,34
39	0,34	0,35	0,37	0,37	0,18	0,18	0,15	0,15	0,21	0,25	0,21	0,26	0,22	0,24
44	0,82	0,82	0,20	0,20	0,16	0,16	0,35	0,35	0,47	0,40	0,18	0,21	0,15	0,16
Média	0,62	0,61	0,28	0,30	0,16	0,17	0,25	0,25	0,28	0,30	0,16	0,18	0,25	0,28
----- Para cinco variáveis explicativas, excluindo AIE e NE -----														
7	0,79	0,80	0,88	0,88	-	-	0,82	0,82	0,96	0,97	-	-	0,87	0,86
10	0,66	0,69	0,69	0,78	-	-	0,47	0,49	0,61	0,62	-	-	0,80	0,85
12	0,65	0,65	0,69	0,76	-	-	0,94	0,94	0,61	0,69	-	-	0,69	0,84
16	0,55	0,55	0,64	0,64	-	-	0,44	0,44	0,74	0,77	-	-	0,68	0,71
20	0,47	0,50	0,77	0,78	-	-	0,44	0,47	0,73	0,78	-	-	0,87	0,87
25	0,62	0,56	0,87	0,76	-	-	0,87	0,87	0,61	0,62	-	-	0,89	0,86
26	0,75	0,74	0,77	0,79	-	-	0,96	0,95	0,33	0,35	-	-	0,36	0,38
32	0,42	0,44	0,69	0,69	-	-	0,68	0,72	0,21	0,19	-	-	0,18	0,18
35	0,88	0,90	0,88	0,91	-	-	0,88	0,89	0,85	0,87	-	-	0,81	0,83
36	0,53	0,55	0,57	0,58	-	-	0,58	0,61	0,48	0,47	-	-	0,47	0,51
37	0,65	0,64	0,93	0,92	-	-	0,81	0,80	0,61	0,61	-	-	0,81	0,81
39	0,28	0,28	0,39	0,39	-	-	0,31	0,31	0,72	0,72	-	-	0,37	0,37
44	0,88	0,89	0,89	0,87	-	-	0,76	0,76	0,65	0,58	-	-	0,74	0,65
Média	0,63	0,63	0,74	0,75	-	-	0,69	0,70	0,63	0,63	-	-	0,66	0,67

(<sup>1</sup>) Ensaios definidos 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>) NN: distribuição multivariada não-normal, NO: distribuição multivariada normal após a transformação de variáveis.

**Tabela 20** - Número de condição e determinante da matriz de correlação  $X'X$ , em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados.

Ensaio <sup>(1)</sup>	Número de Condição		Determinante da Matriz	
	NN <sup>(2)</sup>	NO	NN	NO
----- Para sete variáveis explicativas -----				
7	2.123,13	1.646,83	0,00	0,00
10	4.344,47	614,49	0,00	0,00
12	5.737,23	5.202,80	0,00	0,00
16	3.756,93	4.014,12	0,00	0,00
20	8.794,57	2.720,41	0,00	0,00
25	86.389,83	438,44	0,00	0,00
26	13.146,70	308,41	0,00	0,00
32	65.311,14	4.938,54	0,00	0,00
35	6.054,96	465,64	0,00	0,00
36	15.927,61	2.960,88	0,00	0,00
37	7.277,73	7.425,90	0,00	0,00
39	14.171,85	15.257,99	0,00	0,00
44	14.075,45	10.721,50	0,00	0,00
Média	19.008,58	4.362,77	0,00	0,00
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----				
7	26,35	26,74	0,04	0,05
10	38,75	34,98	0,02	0,03
12	28,61	28,39	0,03	0,04
16	31,74	31,62	0,02	0,03
20	28,94	28,45	0,02	0,03
25	23,26	22,89	0,03	0,04
26	31,18	28,71	0,02	0,02
32	35,59	35,38	0,01	0,01
35	24,39	22,88	0,05	0,06
36	32,40	31,61	0,02	0,02
37	28,79	28,87	0,04	0,04
39	34,03	33,72	0,02	0,02
44	27,05	27,00	0,04	0,04
Média	30,08	29,33	0,03	0,03
----- Para cinco variáveis explicativas, excluindo AIE e NE -----				
7	3,03	3,04	0,70	0,69
10	7,52	6,88	0,29	0,34
12	6,05	4,62	0,45	0,52
16	7,83	7,64	0,26	0,27
20	7,55	6,96	0,34	0,36
25	5,30	5,69	0,48	0,41
26	11,20	10,45	0,26	0,27
32	24,04	25,45	0,09	0,08
35	2,83	2,62	0,70	0,73
36	9,06	8,57	0,19	0,21
37	4,81	4,85	0,50	0,49
39	17,59	17,59	0,10	0,10
44	4,26	5,00	0,54	0,47
Média	8,54	8,41	0,38	0,38

<sup>(1)</sup> Ensaios definidos na Tabela 1. <sup>(2)</sup> NN: distribuição multivariada não-normal, NO: distribuição multivariada normal após a transformação de variáveis.

Nos 13 ensaios, o determinante da matriz de correlação  $X'X$  entre as sete variáveis explicativas foi igual a zero antes e após a transformação dos dados, indicando elevado grau de multicolinearidade (Tabela 20). A adição de  $k = 0,10$  elevou os valores de determinante, mas em pequena escala. Já a eliminação das variáveis altamente correlacionadas, permitiu o aumento considerável do determinante da matriz em todos os ensaios.

De um modo geral, o estudo do caso 2 comprovou através da análise do fator de inflação de variância, da tolerância, do número de condição e do determinante da matriz, a existência de multicolinearidade severa e, que a eliminação de variáveis é mais eficiente que a adição da constante  $k = 0,10$ , na redução do grau de multicolinearidade (Tabelas 18, 19 e 20), conforme relatado para o caso 1. A transformação de variáveis contribuiu para a redução do grau de multicolinearidade, quando a mesma foi de elevado grau. Quando o grau de multicolinearidade é reduzido, através da eliminação de variáveis ou adição da constante  $k = 0,10$ , a transformação de dados não altera expressivamente o grau de multicolinearidade.

Em relação à análise de trilha, verifica-se que o florescimento masculino apresentou efeito direto médio negativo e de baixa magnitude nas análises de trilha tradicional, sob multicolinearidade e tradicional com eliminação de variáveis (Tabela 21). Ainda, em cada ensaio, os valores de efeitos diretos estimados pelas três formas de análise de trilha foram concordantes em relação à magnitude e direção dos efeitos de FM e não sofreram alterações significativas com a transformação dos dados. Essa variável apresentou baixo grau de multicolinearidade (Tabelas 18 e 19) e, dessa forma, os coeficientes estimados, tanto pela análise de trilha com alto grau de multicolinearidade, como pelas análises de trilha alternativas, resultaram em valores estáveis, que também não foram alterados com a transformação dos dados, visto que a correlação entre os efeitos diretos, estimados antes e após a transformação de dados, foi elevada ( $r \geq 0,94$ ) (Tabela 21).

Os efeitos diretos calculados para as seis outras variáveis explicativas, em cada um dos 13 ensaios, apresentaram ampla variação nos seus coeficientes estimados ( $-23,20 \leq$  efeito direto  $\leq 29,21$ ) (Tabela 21). Esses resultados indicam que a análise de trilha realizada com alto grau de multicolinearidade não proporciona a estimação de coeficientes confiáveis. No caso 1, os efeitos diretos médios estimados com alto grau de multicolinearidade de NE e PLI sobre PROD foram, respectivamente, -1,42 e 2,74 (Tabela 16). No caso 2, para as mesmas variáveis, os valores de efeito direto estimados foram de 3,64 e -2,47 (Tabela 21). Com base nessas informações, comprovou-se que os coeficientes estimados não representam associações reais, visto que tanto NE, como PLI apresentam alta correlação com a produtividade de grãos e apresentam relação de causa e efeito.

**Tabela 21** - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados.

Ensaio ( <sup>1</sup> )	Variáveis explicativas( <sup>2</sup> )													
	FM		EP		AIE		PRE		NP		NE		PLI	
	NN( <sup>3</sup> )	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO	NN	NO
----- Análise de trilha tradicional com sete variáveis explicativas e alto grau de multicolinearidade -----														
7	0,06	0,06	-1,23	-1,85	2,13	3,06	-1,43	-2,13	2,21	-0,07	-2,18	0,91	1,70	-0,02
10	-0,27	-0,32	1,91	0,65	-2,67	-0,70	1,89	0,77	-2,03	-0,02	3,47	0,11	-1,80	0,50
12	0,28	0,32	-3,74	-5,06	6,71	8,94	-6,13	-8,16	-0,59	0,46	1,96	0,41	-0,64	0,27
16	-0,51	-0,52	2,43	2,41	-4,18	-4,14	2,54	2,52	0,55	0,31	-0,31	0,04	0,91	0,54
20	-0,41	-0,33	1,82	-1,00	-2,69	1,46	1,81	-0,57	5,41	-0,29	-6,39	1,06	4,07	-0,41
25	-0,33	-0,17	-12,85	0,28	17,92	-0,48	-13,13	0,51	-8,20	0,66	18,35	-0,92	-17,38	1,39
26	0,11	-0,04	-0,38	0,06	0,92	0,01	-0,76	0,11	-0,01	0,38	0,41	0,14	0,60	0,48
32	-0,11	-0,16	2,65	2,76	-3,28	-3,44	3,01	3,15	-6,06	-1,02	29,21	4,60	-23,20	-2,99
35	-0,25	-0,31	-5,56	2,73	7,12	-3,55	-3,41	2,06	-2,44	-0,98	2,96	1,56	-0,77	-0,23
36	-0,04	-0,02	2,22	2,24	-2,24	-2,25	2,10	2,15	-1,11	0,55	5,20	-1,92	-3,61	2,43
37	0,03	0,04	-1,63	-1,72	2,16	2,29	-1,90	-2,01	0,47	0,47	0,29	0,30	0,47	0,47
39	0,04	0,04	1,63	2,11	-1,91	-2,69	2,17	3,13	1,82	0,88	-0,97	-0,03	1,75	0,92
44	-0,29	-0,18	-0,36	-2,79	0,41	3,08	-0,47	-2,01	2,39	0,44	-4,68	-0,01	5,83	0,67
Média	-0,13	-0,12	-1,01	0,06	1,57	0,12	-1,05	-0,04	-0,58	0,14	3,64	0,48	-2,47	0,31
Correlação( <sup>4</sup> )	0,94*		0,22 <sup>ns</sup>		0,27 <sup>ns</sup>		0,42 <sup>ns</sup>		0,18 <sup>ns</sup>		0,46 <sup>ns</sup>		0,41 <sup>ns</sup>	
--- Análise de trilha sob multicolinearidade, com acréscimo de valor $k = 0,10$ na diagonal da matriz $X'X$ ---														
7	0,11	0,06	0,10	0,08	0,11	0,08	-0,01	-0,07	0,32	0,26	0,27	0,35	0,22	0,24
10	-0,23	-0,23	0,19	0,24	0,13	-0,01	0,09	0,21	-0,04	-0,03	0,29	0,27	0,37	0,40
12	0,15	0,23	0,16	0,09	0,15	0,18	-0,01	0,03	0,32	0,40	0,41	0,37	0,27	0,27
16	-0,41	-0,43	0,01	0,02	0,00	0,02	0,10	0,10	0,15	0,17	0,33	0,28	0,30	0,34
20	-0,29	-0,31	-0,04	-0,04	0,05	0,07	0,16	0,16	0,34	0,20	0,23	0,38	0,08	0,01
25	-0,27	-0,25	-0,10	-0,06	0,11	0,04	0,18	0,21	0,12	0,16	0,23	0,12	0,19	0,31
26	0,10	-0,03	0,00	0,04	0,05	0,06	0,05	0,05	0,11	0,29	0,35	0,30	0,51	0,36
32	-0,03	-0,05	0,23	0,20	0,15	0,16	0,12	0,14	-0,05	-0,10	0,36	0,42	0,42	0,41
35	-0,23	-0,23	0,03	0,16	0,09	-0,09	0,03	0,15	0,19	0,07	0,38	0,48	0,27	0,20
36	-0,09	-0,13	-0,06	-0,07	-0,11	-0,10	0,02	0,05	0,03	0,01	0,37	0,33	0,41	0,44
37	-0,04	-0,03	-0,02	-0,02	-0,04	-0,04	-0,04	-0,04	0,25	0,25	0,43	0,44	0,31	0,31
39	-0,02	-0,04	0,42	0,41	0,08	0,05	-0,13	-0,14	0,29	0,39	0,48	0,37	0,30	0,37
44	-0,23	-0,20	0,02	0,01	-0,08	-0,08	-0,20	-0,20	0,21	0,23	0,29	0,34	0,23	0,22
Média	-0,11	-0,13	0,07	0,08	0,05	0,03	0,03	0,05	0,17	0,18	0,34	0,34	0,30	0,30
Correlação	0,96*		0,94*		0,74*		0,93*		0,82*		0,55 <sup>ns</sup>		0,82*	
----- Análise de trilha tradicional com eliminação de variáveis e reduzido grau de multicolinearidade -----														
7	0,11	0,08	0,17	0,14	-	-	0,06	-0,02	0,55	0,56	-	-	0,39	0,44
10	-0,26	-0,29	0,27	0,25	-	-	0,20	0,27	0,11	0,08	-	-	0,60	0,61
12	0,17	0,26	0,28	0,23	-	-	0,12	0,20	0,62	0,73	-	-	0,54	0,50
16	-0,47	-0,48	0,00	0,00	-	-	0,15	0,16	0,37	0,35	-	-	0,65	0,61
20	-0,35	-0,41	-0,02	0,00	-	-	0,25	0,30	0,53	0,46	-	-	0,23	0,25
25	-0,32	-0,29	-0,04	-0,05	-	-	0,28	0,24	0,26	0,25	-	-	0,42	0,46
26	0,10	-0,06	0,00	0,07	-	-	0,11	0,12	0,19	0,45	-	-	0,83	0,55
32	0,00	-0,01	0,38	0,37	-	-	0,21	0,24	-0,16	-0,22	-	-	0,93	0,99
35	-0,24	-0,24	0,10	0,11	-	-	0,08	0,12	0,62	0,56	-	-	0,48	0,46
36	-0,08	-0,13	-0,22	-0,22	-	-	-0,12	-0,08	0,08	0,04	-	-	0,80	0,80
37	-0,02	-0,01	-0,06	-0,06	-	-	-0,06	-0,06	0,72	0,72	-	-	0,65	0,65
39	-0,02	-0,02	0,54	0,54	-	-	0,01	0,01	0,84	0,84	-	-	0,83	0,83
44	-0,26	-0,20	-0,05	-0,08	-	-	-0,25	-0,25	0,38	0,47	-	-	0,60	0,67
Média	-0,13	-0,14	0,11	0,10	-	-	0,08	0,10	0,39	0,41	-	-	0,61	0,60
Correlação	0,96*		0,99*				0,96*		0,96*				0,90*	

(<sup>1</sup>) Ensaios definidos 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>) NN: distribuição multivariada não-normal, NO: distribuição multivariada normal após a transformação de variáveis. (<sup>4</sup>) \*Coeficiente de correlação linear de Pearson significativo entre os efeitos diretos estimados antes e após a transformação de variáveis, a 5% de probabilidade de erro, pelo teste t, com 11 graus de liberdade. <sup>ns</sup> = Não significativo.

Como já verificado no diagnóstico de multicolinearidade, a transformação de dados contribuiu para a redução do grau de multicolinearidade. Com isso, houve uma redução da variabilidade dos efeitos diretos ( $-8,16 \leq \text{efeito direto} \leq 8,94$ ) (Tabela 21). No entanto, essa redução não foi suficiente para garantir a estabilização dos coeficientes estimados e a utilização de técnicas alternativas de análise de trilha foi necessária. Além disso, não houve correlação estatisticamente significativa entre os efeitos diretos estimados antes e após a transformação de dados, nas seis variáveis explicativas com alto grau de multicolinearidade.

Assim como no caso 1, a realização da análise de trilha sob multicolinearidade, com adição de um valor  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , permitiu a estabilização dos efeitos diretos ( $-0,41 \leq \text{efeito direto} \leq 0,51$ ) (Tabela 21). Os maiores coeficientes de correlação médios obtidos nos 13 ensaios entre NE e PROD ( $r = 0,75$ ), entre PLI e PROD ( $r = 0,58$ ) e entre NP e PROD ( $r = 0,49$ ) (Tabela 17) resultaram nos maiores efeitos diretos médios de NE, PLI e NP sobre a produtividade de grãos, com coeficientes de 0,34, 0,30 e 0,17, respectivamente (Tabela 21). Os efeitos diretos das variáveis NE e PLI sobre PROD foram positivos nos 13 ensaios e em apenas dois ensaios, os efeitos diretos de NP sobre PROD não foram positivos, indicando que essas três variáveis atuam positivamente sobre a produtividade de grãos, nessa ordem de importância. Os altos coeficientes de correlação entre os efeitos diretos estimados nos ensaios sem e com distribuição normal multivariada, ou seja, antes e após a transformação dos dados ( $0,55 \leq r \leq 0,96$ ), revelam que quando se reduz o impacto da multicolinearidade, a interferência da não-normalidade multivariada é pequena (Tabela 21).

Na análise de trilha tradicional, com eliminação das variáveis AIE e NE, também ocorreu a estabilização dos efeitos diretos ( $-0,48 \leq \text{efeito direto} \leq 0,99$ ) (Tabela 21). As variáveis que apresentaram os maiores efeitos diretos sobre a produtividade de grãos foram PLI e NP, com coeficientes médios de 0,61 e 0,29, confirmando a atuação positiva dessas variáveis sobre a produtividade de grãos. Assim, como na análise de trilha sob multicolinearidade, foi verificada alta correlação entre os efeitos diretos estimados antes e após a transformação dos dados ( $0,90 \leq r \leq 0,99$ ), sugerindo pequena interferência da não-normalidade multivariada, quando os efeitos adversos de multicolinearidade são corrigidos.

Em relação às médias dos coeficientes de determinação das análises de trilha e dos efeitos da variável residual, constatou-se que não houve diferenças entre os valores destes sem e com normalidade multivariada, nas três formas de análise de trilha (Tabela 22). A análise de trilha tradicional, com eliminação das variáveis, apresentou maiores coeficientes de determinação e menores efeitos da variável residual, em relação à análise de trilha sob multicolinearidade, com a adição de um valor  $k = 0,10$ .

**Tabela 22** - Coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual em 13 ensaios de competição de cultivares de milho que seguiram a distribuição normal multivariada após a transformação de dados.

Ensaio <sup>(1)</sup>	----- Coeficiente de determinação -----		----- Efeito da variável residual -----	
	NN <sup>(2)</sup>	NO	NN	NO
---- Análise de trilha tradicional com sete variáveis explicativas e alto grau de multicolinearidade ----				
7	0,66	0,66	0,58	0,58
10	0,82	0,82	0,42	0,43
12	0,86	0,84	0,38	0,40
16	0,80	0,79	0,45	0,46
20	0,62	0,56	0,62	0,66
25	0,47	0,44	0,73	0,75
26	0,91	0,90	0,29	0,31
32	0,85	0,84	0,39	0,40
35	0,63	0,71	0,60	0,54
36	0,80	0,82	0,45	0,42
37	0,73	0,74	0,52	0,51
39	0,98	0,98	0,14	0,16
44	0,48	0,48	0,72	0,72
Média	0,74	0,74	0,48	0,49
Análise de trilha sob multicolinearidade, com acréscimo de valor $k = 0,10$ na diagonal da matriz $X'X$				
7	0,60	0,61	0,63	0,63
10	0,76	0,75	0,49	0,50
12	0,76	0,70	0,49	0,55
16	0,72	0,72	0,53	0,53
20	0,53	0,51	0,69	0,70
25	0,37	0,38	0,79	0,78
26	0,87	0,87	0,36	0,36
32	0,68	0,71	0,56	0,54
35	0,55	0,53	0,67	0,68
36	0,74	0,76	0,51	0,49
37	0,69	0,69	0,55	0,56
39	0,89	0,88	0,32	0,35
44	0,42	0,45	0,76	0,74
Média	0,66	0,66	0,57	0,57
--- Análise de trilha tradicional com eliminação de variáveis e reduzido grau de multicolinearidade ---				
7	0,63	0,63	0,61	0,61
10	0,79	0,79	0,46	0,45
12	0,80	0,75	0,45	0,50
16	0,77	0,77	0,48	0,48
20	0,57	0,52	0,66	0,69
25	0,40	0,42	0,78	0,76
26	0,91	0,90	0,29	0,32
32	0,73	0,76	0,52	0,49
35	0,57	0,53	0,65	0,69
36	0,77	0,80	0,48	0,45
37	0,73	0,73	0,52	0,52
39	0,97	0,97	0,17	0,17
44	0,46	0,48	0,74	0,72
Média	0,70	0,70	0,52	0,53

<sup>(1)</sup> Ensaios definidos na Tabela 1. <sup>(2)</sup> NN: distribuição multivariada não-normal, NO: distribuição multivariada normal após a transformação de variáveis.

Por fim, com base nos 13 ensaios do caso 2, concluiu-se que a transformação de dados para a obtenção de distribuição normal multivariada atenua os efeitos adversos do elevado grau de multicolinearidade na análise de trilha. Quando o grau de multicolinearidade é reduzido mediante a adição de um valor  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  ou através da eliminação de variáveis, os efeitos diretos estimados antes e após a transformação de dados são similares, ou seja, o impacto da não-normalidade multivariada nessa condição não é elevado. Por ser mais eficiente na redução da multicolinearidade e apresentar maior coeficiente de determinação, a técnica de análise de trilha tradicional, com eliminação de variáveis, deve ser preferencialmente utilizada.

### **Caso 3 - Ensaios sem distribuição normal multivariada e com multicolinearidade**

Dos 44 ensaios, 14 apresentaram distribuição normal multivariada, 13 apresentaram distribuição normal multivariada após a transformação de dados e 17 não apresentaram distribuição normal multivariada, mesmo após a transformação de dados (Tabelas 10, 11 e 12). Esses 17 ensaios foram mantidos, a fim de verificar a interferência da transformação de dados e da multicolinearidade em análises de trilha, na ausência de normalidade multivariada, antes e após a transformação de dados.

Conforme já verificado nos 14 ensaios do caso 1 e nos 13 ensaios do caso 2, os maiores coeficientes de correlação no caso 3, também foram entre NE e PROD ( $r = 0,71$ ), PLI e PROD ( $r = 0,55$ ) e entre NP e PROD ( $r = 0,29$ ) (Tabela 23). As demais variáveis apresentaram coeficientes de correlação com a produtividade de grãos de baixa magnitude ( $-0,04 \leq r \leq 0,14$ ). Os coeficientes de correlação das variáveis explicativas com a produtividade de grãos, estimados sem e com transformação de dados, foram similares ( $0,95 \leq r \leq 1,00$ ), indicando que a utilização dos dados originais ou transformados resultam em matrizes de correlação semelhantes.

O fator de inflação de variância médio, dos 17 ensaios, foi de 2,3 para a variável FM sem e com transformação de dados. Para as demais variáveis explicativas, houve alto grau de multicolinearidade, com média geral de  $FIV = 1.597,8$  e amplitude de  $30,1 \leq FIV \leq 17.085,5$  (Tabela 24). O valor médio de FIV com dados transformados foi de 520,8, com amplitude de  $2,8 \leq FIV \leq 16.016,3$ . Portanto, constatou-se que, assim como verificado no estudo do caso 2, a transformação dos dados contribuiu na redução do grau de multicolinearidade, embora a pressuposição de normalidade multivariada não tenha sido atendida no caso 3.

**Tabela 23** - Valores do coeficiente de correlação linear de Pearson das variáveis explicativas com a produtividade de grãos em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados, e valores de coeficiente de correlação linear de Pearson ajustado entre os valores antes e após a transformação de variáveis (r entre grupos).

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>						
	FM	EP	AIE	PRE	NP	NE	PLI
----- Coeficientes de correlação de Pearson antes da transformação de variáveis -----							
3	0,29 <sup>ns</sup>	0,86 *	0,48 <sup>ns</sup>	-0,33 <sup>ns</sup>	0,57 <sup>ns</sup>	0,73 *	0,14 <sup>ns</sup>
4	-0,16 <sup>ns</sup>	0,50 <sup>ns</sup>	0,34 <sup>ns</sup>	-0,39 <sup>ns</sup>	0,43 <sup>ns</sup>	0,62 *	0,44 <sup>ns</sup>
6	0,28 <sup>ns</sup>	0,19 <sup>ns</sup>	0,37 *	0,49 *	-0,02 <sup>ns</sup>	0,66 *	0,67 *
8	-0,12 <sup>ns</sup>	0,34 *	0,40 *	0,37 *	0,38 *	0,67 *	0,66 *
9	0,44 *	0,26 <sup>ns</sup>	0,10 <sup>ns</sup>	-0,09 <sup>ns</sup>	0,61 *	0,62 *	0,14 <sup>ns</sup>
11	-0,30 <sup>ns</sup>	0,01 <sup>ns</sup>	-0,14 <sup>ns</sup>	-0,13 <sup>ns</sup>	0,68 *	0,78 *	0,22 <sup>ns</sup>
14	0,05 <sup>ns</sup>	-0,18 <sup>ns</sup>	-0,05 <sup>ns</sup>	0,14 <sup>ns</sup>	0,17 <sup>ns</sup>	0,70 *	0,75 *
15	0,30 <sup>ns</sup>	0,49 *	0,41 *	0,10 <sup>ns</sup>	0,03 <sup>ns</sup>	0,69 *	0,65 *
22	0,09 <sup>ns</sup>	0,04 <sup>ns</sup>	0,07 <sup>ns</sup>	0,06 <sup>ns</sup>	0,18 <sup>ns</sup>	0,51 *	0,47 *
29	-0,20 <sup>ns</sup>	-0,15 <sup>ns</sup>	-0,15 <sup>ns</sup>	-0,09 <sup>ns</sup>	0,12 <sup>ns</sup>	0,64 *	0,58 *
30	-0,18 <sup>ns</sup>	-0,38 <sup>ns</sup>	-0,35 <sup>ns</sup>	-0,20 <sup>ns</sup>	-0,38 <sup>ns</sup>	0,83 *	0,80 *
33	-0,52 *	-0,01 <sup>ns</sup>	-0,36 <sup>ns</sup>	-0,37 <sup>ns</sup>	0,49 *	0,75 *	0,43 <sup>ns</sup>
34	0,38 <sup>ns</sup>	0,82 *	0,61 *	0,08 <sup>ns</sup>	0,71 *	0,92 *	0,75 *
38	-0,09 <sup>ns</sup>	0,19 <sup>ns</sup>	0,08 <sup>ns</sup>	-0,03 <sup>ns</sup>	0,15 <sup>ns</sup>	0,79 *	0,71 *
40	-0,69 *	-0,27 <sup>ns</sup>	-0,37 <sup>ns</sup>	-0,29 <sup>ns</sup>	0,30 <sup>ns</sup>	0,89 *	0,79 *
42	-0,10 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,20 <sup>ns</sup>	-0,03 <sup>ns</sup>	0,12 <sup>ns</sup>	0,66 *	0,58 *
43	-0,14 <sup>ns</sup>	0,02 <sup>ns</sup>	0,03 <sup>ns</sup>	0,03 <sup>ns</sup>	0,35 <sup>ns</sup>	0,66 *	0,56 *
Média	-0,04	0,14	0,07	-0,04	0,29	0,71	0,55
----- Coeficientes de correlação de Pearson após a transformação de variáveis -----							
3	0,32 <sup>ns</sup>	0,86 *	0,48 <sup>ns</sup>	-0,33 <sup>ns</sup>	0,62 *	0,73 *	0,11 <sup>ns</sup>
4	-0,18 <sup>ns</sup>	0,50 <sup>ns</sup>	0,34 <sup>ns</sup>	-0,39 <sup>ns</sup>	0,44 <sup>ns</sup>	0,64 *	0,43 <sup>ns</sup>
6	0,22 <sup>ns</sup>	0,14 <sup>ns</sup>	0,28 <sup>ns</sup>	0,48 *	-0,01 <sup>ns</sup>	0,67 *	0,68 *
8	-0,13 <sup>ns</sup>	0,32 <sup>ns</sup>	0,37 *	0,32 <sup>ns</sup>	0,38 *	0,71 *	0,69 *
9	0,33 *	0,22 <sup>ns</sup>	0,09 <sup>ns</sup>	-0,11 <sup>ns</sup>	0,56 *	0,58 *	0,18 <sup>ns</sup>
11	-0,26 <sup>ns</sup>	0,05 <sup>ns</sup>	-0,07 <sup>ns</sup>	-0,11 <sup>ns</sup>	0,66 *	0,74 *	0,18 <sup>ns</sup>
14	-0,05 <sup>ns</sup>	-0,26 <sup>ns</sup>	-0,14 <sup>ns</sup>	0,08 <sup>ns</sup>	0,14 <sup>ns</sup>	0,67 *	0,72 *
15	0,30 <sup>ns</sup>	0,49 *	0,41 *	0,10 <sup>ns</sup>	0,02 <sup>ns</sup>	0,64 *	0,64 *
22	0,10 <sup>ns</sup>	0,01 <sup>ns</sup>	0,04 <sup>ns</sup>	0,05 <sup>ns</sup>	0,18 <sup>ns</sup>	0,51 *	0,47 *
29	-0,09 <sup>ns</sup>	-0,15 <sup>ns</sup>	-0,13 <sup>ns</sup>	-0,06 <sup>ns</sup>	0,14 <sup>ns</sup>	0,53 *	0,47 *
30	-0,17 <sup>ns</sup>	-0,38 <sup>ns</sup>	-0,35 <sup>ns</sup>	-0,25 <sup>ns</sup>	-0,42 *	0,85 *	0,79 *
33	-0,52 *	-0,02 <sup>ns</sup>	-0,36 <sup>ns</sup>	-0,37 <sup>ns</sup>	0,49 *	0,75 *	0,42 <sup>ns</sup>
34	0,38 <sup>ns</sup>	0,82 *	0,58 *	0,08 <sup>ns</sup>	0,71 *	0,92 *	0,75 *
38	-0,11 <sup>ns</sup>	0,20 <sup>ns</sup>	0,12 <sup>ns</sup>	-0,02 <sup>ns</sup>	0,13 <sup>ns</sup>	0,79 *	0,72 *
40	-0,70 *	-0,28 <sup>ns</sup>	-0,37 <sup>ns</sup>	-0,30 <sup>ns</sup>	0,32 <sup>ns</sup>	0,89 *	0,76 *
42	-0,10 <sup>ns</sup>	-0,30 <sup>ns</sup>	-0,19 <sup>ns</sup>	-0,03 <sup>ns</sup>	0,05 <sup>ns</sup>	0,62 *	0,54 *
43	-0,15 <sup>ns</sup>	0,04 <sup>ns</sup>	0,04 <sup>ns</sup>	0,04 <sup>ns</sup>	0,35 <sup>ns</sup>	0,70 *	0,59 *
Média	-0,05	0,13	0,07	-0,05	0,28	0,70	0,54
r entre grupos <sup>(4)</sup>	0,99 *	1,00 *	0,99 *	0,99 *	0,99 *	0,95 *	0,99 *

<sup>(1)</sup> Ensaios definidos 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> \*Coeficiente de correlação linear de Pearson significativo a 5% de probabilidade de erro, pelo teste t, com n-2 graus de liberdade, sendo n = número de cultivares do ensaio. <sup>ns</sup> = Não significativo. <sup>(4)</sup> \*Coeficiente de correlação linear de Pearson significativo a 5% de probabilidade de erro, pelo teste t, com 15 graus de liberdade. <sup>ns</sup> = Não significativo.



**Tabela 24** - Fator de inflação da variância (FIV) para variáveis explicativas em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados.

Ensaio <sup>(1)</sup>	Variáveis explicativas <sup>(2)</sup>													
	FM		EP		AIE		PRE		NP		NE		PLI	
	ST <sup>(3)</sup>	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT
	Para sete variáveis explicativas													
3	2,6	3,3	5366,0	147,0	8517,1	208,1	3190,7	78,0	617,7	603	523,3	47,2	562,3	50,0
4	1,3	1,3	2316,5	599,9	1056,2	297,5	1073,2	271,1	10625,4	412,7	17085,5	622,8	6660,8	223,5
6	5,0	4,2	327,6	31,5	744,4	48,7	206,5	15,7	307,5	189	2125,9	80,5	2112,4	84,1
8	1,1	1,1	153,9	53,4	532,7	153,3	212,8	56,9	44,3	6,6	991,1	166,2	776,9	137,7
9	2,3	2,2	216,8	186,3	452,3	398,5	184,8	168,4	1605,9	27,9	2242,4	39,4	1029,6	21,6
11	3,3	2,9	272,0	22,5	555,0	41,1	649,8	44,5	789,9	31,1	903,0	32,2	365,5	14,2
14	2,5	2,2	392,7	14,4	981,1	27,9	273,9	8,9	1811,1	63,6	4592,7	180,7	2866,1	123,8
15	1,6	1,5	433,6	481,8	923,6	1030,1	529,0	589,5	30,1	3,2	9,28	6,6	87,3	6,3
22	1,3	1,3	284,4	79,6	759,7	217,3	384,2	110,1	485,4	1519,8	5148,1	16016,3	5007,8	15556,2
29	1,2	1,3	388,7	375,7	902,7	888,3	308,4	305,0	152,9	2,8	1999,4	28,7	2191,0	32,4
30	1,1	1,1	628,7	6,6	1871,6	17,2	914,3	8,6	167,9	9,8	1559,3	32,3	2214,6	54,1
33	1,9	2,3	181,2	125,3	931,7	651,5	853,1	581,0	536,6	298,1	611,9	332,8	249,5	139,2
34	1,7	1,5	1770,2	46,9	3585,6	72,0	1824,7	37,7	78,9	80,0	349,4	347,1	206,1	209,2
38	1,5	1,4	70,6	209,8	229,0	603,9	97,3	234,8	134,0	98,6	276,6	199,7	253,9	186,1
40	7,6	8,4	2903,8	413,5	8740,0	1309,8	4528,6	625,0	504,7	36,2	3351,4	155,8	3250,3	194,8
42	1,5	1,9	366,1	119,5	1156,6	362,0	538,4	177,4	127,9	9,4	256,2	20,0	261,8	20,7
43	1,5	1,7	656,2	196,1	1388,9	420,8	337,1	101,8	935,5	274,9	4278,9	880,0	3368,9	648,4
Média	2,3	2,3	984,1	182,9	1960,5	396,9	947,5	200,9	1115,0	173,8	2728,7	1128,7	1850,9	1041,3
	Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$													
3	1,5	1,7	5,0	5,1	6,6	6,4	3,2	3,3	5,0	4,9	4,7	4,5	4,6	4,0
4	1,2	1,2	6,7	6,6	3,7	3,8	3,9	3,7	4,4	4,6	6,4	6,4	3,4	3,2
6	3,0	2,4	4,2	4,4	7,6	6,1	3,6	2,9	1,9	2,1	6,0	5,5	6,0	5,5
8	1,2	1,2	3,3	3,4	7,4	7,1	3,9	3,6	1,7	1,6	6,9	6,7	5,8	5,8
9	1,9	1,9	3,8	3,8	6,7	6,7	3,5	3,6	4,7	4,1	5,9	5,3	3,4	3,5
11	2,3	2,0	3,2	3,2	5,4	4,9	5,8	5,1	5,1	4,8	5,7	5,0	3,1	2,8
14	2,1	2,0	4,2	3,9	7,5	6,1	3,4	2,9	3,2	2,9	6,3	6,1	4,3	4,5
15	1,5	1,5	3,9	3,9	6,3	6,4	4,1	4,2	2,7	2,2	5,6	3,4	5,2	3,2
22	1,3	1,3	3,3	3,2	6,7	6,6	4,0	4,0	1,7	1,7	6,2	6,2	6,1	6,0
29	1,2	1,2	3,8	3,8	7,0	7,0	3,3	3,3	1,5	1,5	5,9	5,1	6,4	5,6
30	1,1	1,1	3,2	2,6	6,9	5,3	4,0	3,1	1,9	2,3	5,4	4,4	7,0	6,5
33	1,6	1,7	2,1	2,2	6,1	6,1	5,7	5,7	5,1	5,1	5,9	5,8	3,2	3,2
34	1,4	1,4	4,4	4,6	6,6	5,9	3,9	3,6	2,6	2,6	7,1	7,1	4,8	4,8
38	1,4	1,4	3,2	3,4	7,1	7,2	4,0	3,8	3,1	3,1	5,4	5,3	5,0	5,0
40	3,4	3,3	3,6	3,6	7,6	7,7	4,6	4,4	3,0	2,6	6,3	5,7	6,6	7,1
42	1,5	1,5	3,5	3,6	7,0	6,9	4,0	4,1	3,2	2,6	5,2	4,5	5,3	4,6
43	1,5	1,5	4,2	4,2	7,3	7,3	3,0	3,0	2,2	2,7	6,3	6,2	5,2	4,8
Média	1,7	1,7	3,9	3,8	6,7	6,3	4,0	3,8	3,1	3,0	6,0	5,5	5,0	4,7
	Para cinco variáveis explicativas, excluindo AIE e NE													
3	1,5	1,8	2,9	3,4	-	-	1,1	1,2	3,2	3,5	-	-	2,5	2,2
4	1,2	1,2	2,7	2,4	-	-	2,7	2,4	1,2	1,2	-	-	1,5	1,5
6	3,7	2,8	2,0	1,8	-	-	2,7	1,9	1,1	1,2	-	-	1,1	1,1
8	1,1	1,1	1,4	1,3	-	-	1,5	1,4	1,3	1,3	-	-	1,6	1,5
9	2,0	2,0	1,2	1,3	-	-	1,3	1,3	1,5	1,5	-	-	1,2	1,3
11	2,4	1,9	2,2	1,9	-	-	2,7	2,1	1,3	1,4	-	-	1,4	1,4
14	2,2	2,1	2,0	2,0	-	-	1,7	1,6	1,1	1,1	-	-	1,0	1,0
15	1,5	1,5	1,9	1,8	-	-	1,2	1,3	1,7	1,7	-	-	1,3	1,3
22	1,3	1,3	1,1	1,1	-	-	1,2	1,2	1,2	1,2	-	-	1,2	1,2
29	1,2	1,2	1,2	1,2	-	-	1,1	1,1	1,2	1,3	-	-	1,1	1,2
30	1,0	1,0	1,1	1,1	-	-	1,1	1,1	1,9	2,1	-	-	1,9	2,1
33	1,6	1,7	1,2	1,2	-	-	1,3	1,3	1,5	1,5	-	-	1,8	1,9
34	1,4	1,4	2,7	2,7	-	-	1,3	1,3	1,5	1,5	-	-	2,0	2,0
38	1,4	1,4	1,3	1,2	-	-	1,6	1,5	1,1	1,1	-	-	1,2	1,2
40	4,9	4,6	3,6	3,3	-	-	4,2	3,8	3,3	2,6	-	-	3,1	4,1
42	1,4	1,6	1,8	1,9	-	-	1,1	1,1	1,4	1,4	-	-	1,3	1,3
43	1,4	1,4	1,3	1,3	-	-	1,5	1,6	1,0	1,0	-	-	1,0	1,0
Média	1,8	1,7	1,8	1,8	-	-	1,7	1,6	1,6	1,6	-	-	1,5	1,6

<sup>(1)</sup> Ensaios definidos 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> ST: distribuição multivariada não-normal sem transformação de variáveis, CT: distribuição multivariada não-normal com transformação de variáveis.

A utilização da constante  $k = 0,10$  foi eficiente na redução de FIV em todas as variáveis dos 17 ensaios (Tabela 24) e assim como no caso 2, as variáveis transformadas apresentaram FIV um pouco inferiores em relação às variáveis sem transformação. A eliminação das variáveis AIE e NE também foi eficiente na redução do fator de inflação de variância, sendo os valores de FIV antes e após a transformação de variáveis similares entre si e menores em relação aos valores de FIV obtidos pela adição da constante  $k = 0,10$ .

Como já verificado nos casos 1 e 2, no caso 3, com exceção de FM, todas as demais variáveis apresentaram tolerância  $\leq 0,03$  (Tabela 25), indicando existência de multicolinearidade severa entre as variáveis EP, AIE, PRE, NP, NE e PLI. Na média geral, a transformação de dados elevou os valores de tolerância, mas em apenas algumas variáveis e ensaios, esse aumento foi satisfatório (tolerância  $> 0,10$ ). Tanto no caso de adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , como no caso de eliminação de variáveis, os valores de tolerância sem e com a transformação de dados foram similares. Além disso, a tolerância esteve acima dos limites críticos nessas duas situações ( $0,13 \leq \text{tolerância} \leq 0,99$ ). A eliminação de variáveis foi mais eficiente que a adição de  $k = 0,10$  no aumento da tolerância, ou seja, na redução do grau de multicolinearidade.

Todos os 17 ensaios apresentaram multicolinearidade severa com base no número de condição, sem a transformação de dados ( $1.763,46 \leq NC \leq 100.562,94$ ), com média de  $NC = 25.020,58$  (Tabela 26). O diagnóstico com base nos dados transformados indicou a redução média de  $NC$  para  $7.593,11$ . A adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  ( $24,59 \leq NC \leq 43,13$ ) e a eliminação de variáveis ( $3,02 \leq NC \leq 32,89$ ) foram eficientes na redução do grau de multicolinearidade. Nessas duas técnicas, os resultados de  $NC$  obtidos sem e com a transformação de dados foram similares e a eliminação de variáveis promoveu maior redução do grau de multicolinearidade.

Os valores de determinante da matriz de correlação  $X'X$ , para as sete variáveis explicativas sem e com transformação de dados, iguais a zero, comprovam o elevado grau de multicolinearidade dos 17 ensaios (Tabela 26). Os determinantes das matrizes obtidos após a eliminação de variáveis superaram os obtidos com base na adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ . Assim, com base nos quatro métodos de diagnóstico de multicolinearidade, verificou-se a existência de multicolinearidade severa nos 17 ensaios. A eliminação de variáveis foi mais eficiente que a adição de  $k = 0,10$ , na redução do grau de multicolinearidade, conforme relatado para os casos 1 e 2 e, a utilização de dados transformados contribuiu para a redução parcial do grau de multicolinearidade, quando a mesma foi severa.

**Tabela 25** - Tolerância para as variáveis explicativas em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados.

Ensaio (1)	Variáveis explicativas <sup>(2)</sup>													
	FM		EP		AIE		PRE		NP		NE		PLI	
	ST <sup>(3)</sup>	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT
----- Para sete variáveis explicativas -----														
3	0,38	0,31	0,00	0,01	0,00	0,00	0,00	0,01	0,00	0,02	0,00	0,02	0,00	0,02
4	0,75	0,77	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
6	0,20	0,24	0,00	0,03	0,00	0,02	0,00	0,06	0,00	0,05	0,00	0,01	0,00	0,01
8	0,90	0,90	0,01	0,02	0,00	0,01	0,00	0,02	0,02	0,15	0,00	0,01	0,00	0,01
9	0,44	0,46	0,00	0,01	0,00	0,00	0,01	0,01	0,00	0,04	0,00	0,03	0,00	0,05
11	0,30	0,34	0,00	0,04	0,00	0,02	0,00	0,02	0,00	0,03	0,00	0,03	0,00	0,07
14	0,40	0,45	0,00	0,07	0,00	0,04	0,00	0,11	0,00	0,02	0,00	0,01	0,00	0,01
15	0,62	0,65	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,32	0,01	0,15	0,01	0,16
22	0,76	0,78	0,00	0,01	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00
29	0,83	0,77	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,36	0,00	0,03	0,00	0,03
30	0,95	0,95	0,00	0,15	0,00	0,06	0,00	0,12	0,01	0,10	0,00	0,03	0,00	0,02
33	0,52	0,43	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
34	0,59	0,66	0,00	0,02	0,00	0,01	0,00	0,03	0,01	0,01	0,00	0,00	0,00	0,00
38	0,69	0,71	0,01	0,00	0,00	0,00	0,01	0,00	0,01	0,01	0,00	0,01	0,00	0,01
40	0,13	0,12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,01	0,00	0,01
42	0,65	0,54	0,00	0,01	0,00	0,00	0,00	0,01	0,01	0,11	0,00	0,05	0,00	0,05
43	0,69	0,59	0,00	0,01	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00
Média	0,58	0,57	0,00	0,02	0,00	0,01	0,00	0,02	0,01	0,07	0,00	0,02	0,00	0,03
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----														
3	0,66	0,60	0,20	0,20	0,15	0,16	0,31	0,31	0,20	0,20	0,21	0,22	0,22	0,25
4	0,80	0,80	0,15	0,15	0,27	0,26	0,26	0,27	0,23	0,22	0,16	0,16	0,29	0,32
6	0,33	0,42	0,24	0,23	0,13	0,16	0,28	0,34	0,54	0,47	0,17	0,18	0,17	0,18
8	0,84	0,84	0,31	0,29	0,14	0,14	0,25	0,28	0,59	0,64	0,14	0,15	0,17	0,17
9	0,53	0,52	0,26	0,26	0,15	0,15	0,28	0,28	0,21	0,25	0,17	0,19	0,29	0,29
11	0,44	0,49	0,32	0,31	0,19	0,21	0,17	0,20	0,19	0,21	0,18	0,20	0,32	0,36
14	0,49	0,50	0,24	0,26	0,13	0,16	0,30	0,34	0,31	0,35	0,16	0,16	0,23	0,22
15	0,65	0,67	0,25	0,26	0,16	0,16	0,24	0,24	0,36	0,46	0,18	0,29	0,19	0,31
22	0,75	0,75	0,31	0,32	0,15	0,15	0,25	0,25	0,60	0,60	0,16	0,16	0,17	0,17
29	0,80	0,80	0,26	0,26	0,14	0,14	0,30	0,30	0,65	0,69	0,17	0,19	0,16	0,18
30	0,88	0,88	0,32	0,38	0,15	0,19	0,25	0,32	0,52	0,44	0,18	0,23	0,14	0,15
33	0,61	0,60	0,47	0,46	0,16	0,16	0,17	0,18	0,20	0,20	0,17	0,17	0,32	0,31
34	0,70	0,70	0,23	0,22	0,15	0,17	0,26	0,28	0,38	0,38	0,14	0,14	0,21	0,21
38	0,69	0,70	0,31	0,29	0,14	0,14	0,25	0,26	0,32	0,32	0,19	0,19	0,20	0,20
40	0,30	0,30	0,28	0,28	0,13	0,13	0,22	0,23	0,33	0,38	0,16	0,18	0,15	0,14
42	0,69	0,65	0,29	0,28	0,14	0,15	0,25	0,25	0,31	0,38	0,19	0,22	0,19	0,22
43	0,67	0,67	0,24	0,24	0,14	0,14	0,33	0,33	0,45	0,37	0,16	0,16	0,19	0,21
Média	0,64	0,64	0,27	0,28	0,15	0,16	0,26	0,27	0,38	0,39	0,17	0,19	0,21	0,23
----- Para cinco variáveis explicativas, excluindo AIE e NE -----														
3	0,65	0,56	0,35	0,29	-	-	0,91	0,83	0,31	0,29	-	-	0,40	0,45
4	0,86	0,87	0,38	0,41	-	-	0,37	0,42	0,85	0,82	-	-	0,65	0,67
6	0,27	0,36	0,50	0,55	-	-	0,37	0,53	0,88	0,86	-	-	0,93	0,89
8	0,92	0,91	0,72	0,74	-	-	0,66	0,71	0,77	0,79	-	-	0,63	0,68
9	0,51	0,50	0,80	0,79	-	-	0,74	0,75	0,68	0,68	-	-	0,82	0,79
11	0,41	0,53	0,46	0,51	-	-	0,37	0,49	0,79	0,74	-	-	0,74	0,74
14	0,46	0,48	0,50	0,50	-	-	0,60	0,63	0,94	0,95	-	-	0,99	0,98
15	0,66	0,69	0,53	0,57	-	-	0,81	0,77	0,60	0,60	-	-	0,79	0,79
22	0,80	0,79	0,90	0,91	-	-	0,86	0,86	0,84	0,85	-	-	0,82	0,85
29	0,86	0,86	0,85	0,85	-	-	0,89	0,89	0,83	0,78	-	-	0,89	0,83
30	0,96	0,97	0,89	0,90	-	-	0,93	0,92	0,53	0,48	-	-	0,52	0,47
33	0,61	0,59	0,86	0,84	-	-	0,75	0,76	0,66	0,66	-	-	0,55	0,53
34	0,73	0,73	0,37	0,37	-	-	0,80	0,80	0,67	0,67	-	-	0,50	0,50
38	0,72	0,73	0,79	0,83	-	-	0,63	0,66	0,89	0,88	-	-	0,84	0,84
40	0,20	0,22	0,28	0,31	-	-	0,24	0,26	0,31	0,38	-	-	0,32	0,24
42	0,70	0,64	0,57	0,53	-	-	0,90	0,90	0,71	0,70	-	-	0,77	0,76
43	0,70	0,70	0,79	0,78	-	-	0,65	0,64	0,99	0,99	-	-	0,97	0,97
Média	0,65	0,65	0,62	0,63	-	-	0,68	0,69	0,72	0,71	-	-	0,71	0,71

(1) Ensaios definidos na Tabela 1. (2) 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. (3) ST: distribuição multivariada não-normal sem transformação de variáveis, CT: distribuição multivariada não-normal com transformação de variáveis.

**Tabela 26** - Número de condição e determinante da matriz de correlação  $X'X$ , em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de dados.

Ensaio <sup>(1)</sup>	----- Número de condição -----		----- Determinante da matriz -----	
	ST <sup>(2)</sup>	CT	ST	CT
----- Para sete variáveis explicativas -----				
3	52.304,09	1.557,25	0,00	0,00
4	100.562,94	3.807,78	0,00	0,00
6	15.978,42	632,58	0,00	0,00
8	6.677,53	1.185,46	0,00	0,00
9	12.466,96	1.937,16	0,00	0,00
11	6.524,39	282,34	0,00	0,00
14	30.110,94	1.154,61	0,00	0,00
15	5.237,57	5.777,62	0,00	0,00
22	27.313,20	81.493,86	0,00	0,00
29	10.895,77	4.049,23	0,00	0,00
30	11.606,92	249,23	0,00	0,00
33	5.886,57	4.372,12	0,00	0,00
34	27.035,32	2.370,26	0,00	0,00
38	1.763,46	2.730,12	0,00	0,00
40	80.732,39	10.606,66	0,00	0,00
42	5.993,49	1.772,71	0,00	0,00
43	24.259,98	5.103,88	0,00	0,00
Média	25.020,58	7.593,11	0,00	0,00
----- Para sete variáveis explicativas, adicionando valor $k = 0,10$ na diagonal da matriz $X'X$ -----				
3	31,45	31,47	0,02	0,02
4	30,24	29,77	0,03	0,03
6	34,47	30,21	0,01	0,03
8	37,29	35,43	0,02	0,03
9	26,40	26,33	0,03	0,03
11	28,17	24,93	0,03	0,04
14	33,43	31,62	0,02	0,03
15	28,61	28,37	0,03	0,05
22	26,64	25,62	0,04	0,04
29	25,25	26,62	0,04	0,05
30	27,38	24,59	0,04	0,05
33	30,79	30,79	0,03	0,03
34	38,58	37,68	0,02	0,02
38	27,17	26,83	0,04	0,04
40	42,75	43,13	0,01	0,01
42	28,08	27,57	0,04	0,04
43	29,21	29,06	0,03	0,03
Média	30,94	30,00	0,03	0,03
----- Para cinco variáveis explicativas, excluindo AIE e NE -----				
3	13,50	15,53	0,20	0,17
4	10,29	8,88	0,23	0,26
6	14,19	9,57	0,17	0,27
8	5,17	4,42	0,42	0,47
9	5,87	6,08	0,43	0,42
11	11,35	8,44	0,23	0,28
14	7,52	7,19	0,31	0,33
15	5,64	5,24	0,40	0,41
22	3,14	3,04	0,68	0,69
29	3,02	3,33	0,70	0,66
30	5,69	6,56	0,47	0,43
33	6,04	6,38	0,41	0,39
34	9,64	9,64	0,26	0,26
38	4,29	3,96	0,51	0,54
40	32,89	27,30	0,05	0,05
42	5,71	6,41	0,46	0,43
43	4,20	4,24	0,56	0,55
Média	8,71	8,01	0,38	0,39

<sup>(1)</sup> Ensaios definidos na Tabela 1. <sup>(2)</sup> ST: distribuição multivariada não-normal sem transformação de variáveis, CT: distribuição multivariada não-normal com transformação de variáveis.

Os altos coeficientes de correlação entre os efeitos diretos estimados sem e com transformação de dados, nas análises de trilha tradicional ( $0,56 \leq r \leq 0,92$ ), sob multicolinearidade ( $0,82 \leq r \leq 0,98$ ) e tradicional, com eliminação de variáveis ( $0,93 \leq r \leq 0,99$ ), indica pouca interferência da transformação de dados na estimativa dos efeitos diretos (Tabela 27).

Assim como nos 13 ensaios do caso 2, o florescimento masculino apresentou efeito direto médio negativo e de baixa magnitude nas análises de trilha tradicional, sob multicolinearidade e tradicional com eliminação de variáveis (Tabela 27). Os valores de efeitos diretos estimados foram concordantes entre as três formas de análise de trilha devido ao baixo grau de multicolinearidade nessa variável (Tabelas 24 e 25). Os efeitos diretos estimados para as demais seis variáveis explicativas por meio da análise de trilha tradicional apresentaram ampla variação entre os 17 ensaios ( $-51,98 \leq \text{efeito direto} \leq 66,42$ ) (Tabela 27). Esses resultados são concordantes com os obtidos nos 13 ensaios do caso 2 e a interpretação dos efeitos diretos estimados em condição de elevado grau de multicolinearidade não deve ser realizada.

De modo semelhante aos casos 1 e 2, a realização da análise de trilha sob multicolinearidade, com adição de um valor  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , permitiu a estabilização dos efeitos diretos ( $-0,42 \leq \text{efeito direto} \leq 0,56$ ) (Tabela 27). Os maiores coeficientes de correlação médios obtidos nos 17 ensaios entre NE e PROD, PLI e PROD e entre NP e PROD, resultaram nos maiores efeitos diretos médios de NE, PLI e NP sobre a produtividade de grãos, nessa ordem de importância. A eliminação das variáveis AIE e NE também permitiu a estabilização dos efeitos diretos ( $-0,39 \leq \text{efeito direto} \leq 0,96$ ) e as variáveis que apresentaram os maiores efeitos diretos sobre a produtividade de grãos foram PLI e NP, com coeficientes médios de 0,59 e 0,35, respectivamente.

A análise de trilha tradicional, com eliminação de variáveis, apresentou maior coeficiente de determinação médio em relação à análise de trilha sob multicolinearidade e menor efeito da variável residual (Tabela 28). Como já relatado nos casos 1 e 2, por reduzir de forma mais expressiva o grau de multicolinearidade, apresentar maior coeficiente de determinação e menor efeito da variável residual, a análise de trilha tradicional com eliminação de variáveis deve ser preferencialmente utilizada.

**Tabela 27** - Efeitos diretos das variáveis explicativas sobre a produtividade de grãos em 17 ensaios de competição de cultivares de milho, sem distribuição normal multivariada antes e após a transformação de dados.

Ensaio (1)	Variáveis explicativas <sup>(2)</sup>													
	FM		EP		AIE		PRE		NP		NE		PLI	
	ST <sup>(3)</sup>	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT	ST	CT
----- Análise de trilha tradicional com sete variáveis explicativas e alto grau de multicolinearidade -----														
3	-0,09	0,17	-0,66	-2,03	1,80	3,27	-1,47	-2,34	-2,52	0,72	2,63	-0,29	-2,52	0,65
4	-0,29	-0,28	5,80	-2,33	-3,64	1,87	3,90	-1,63	-51,98	-9,90	66,42	12,67	-41,14	-7,24
6	-0,12	0,01	1,24	0,31	-1,79	-0,43	1,46	0,67	0,26	0,43	-0,04	-0,29	0,70	0,96
8	-0,16	-0,16	1,46	0,40	-2,61	-0,57	1,72	0,39	-1,03	-0,19	5,40	1,52	-4,24	-0,77
9	0,25	0,17	3,15	1,16	-4,40	-1,51	2,66	0,91	-5,46	0,17	7,05	0,49	-4,57	-0,05
11	0,08	0,03	1,82	0,05	-2,45	0,13	2,64	-0,12	-0,14	0,55	1,00	0,24	-0,15	0,29
14	0,06	0,06	4,11	0,09	-7,11	-0,72	3,95	0,57	1,58	0,34	-2,13	-0,25	2,40	0,91
15	0,22	0,17	3,60	4,29	-5,00	-5,79	3,90	4,47	-1,44	-0,17	2,73	0,43	-2,05	0,18
22	-0,12	-0,13	-1,98	1,13	2,94	-2,04	-2,08	1,42	5,83	2,68	-17,98	-7,68	18,30	8,12
29	-0,01	0,02	-2,75	-1,83	3,91	2,49	-2,31	-1,49	1,82	0,24	-5,44	0,75	6,37	-0,15
30	-0,10	-0,07	-1,69	-0,28	2,63	0,16	-1,97	-0,25	-0,17	-0,32	1,40	1,30	-0,73	-0,69
33	-0,15	-0,35	-1,71	-2,14	3,71	4,91	-3,50	-4,57	-5,43	-5,23	6,39	6,10	-3,52	-3,37
34	0,04	0,09	5,26	0,64	-7,17	-0,52	5,04	0,30	-0,51	-0,65	1,92	2,28	-1,16	-1,34
38	0,12	0,10	0,06	-0,86	0,16	1,74	-0,15	-1,12	-0,26	-0,38	0,99	1,12	-0,11	-0,26
40	-0,84	-0,77	-9,11	-3,28	16,93	6,17	-11,84	-4,50	-7,66	-1,55	20,33	4,55	-19,11	-4,14
42	-0,18	-0,20	4,01	2,01	-7,28	-3,72	4,97	2,63	2,38	-0,07	-2,97	0,52	3,68	0,15
43	-0,12	-0,02	3,41	3,05	-4,82	-4,37	2,28	2,09	4,20	3,76	-8,19	-5,99	7,84	5,70
Média	-0,08	-0,07	0,94	0,02	-0,83	0,06	0,54	-0,15	-3,56	-0,56	4,68	1,03	-2,35	-0,06
Correlação <sup>(4)</sup>	0,92 *		0,56 *		0,73 *		0,71 *		0,88 *		0,88 *		0,86 *	
----- Análise de trilha sob multicolinearidade, com acréscimo de valor $k = 0,10$ na diagonal da matriz $X'X$ -----														
3	0,05	0,07	0,45	0,39	0,14	0,21	-0,36	-0,42	0,16	0,15	0,22	0,23	0,02	0,03
4	-0,11	-0,12	0,18	0,13	0,23	0,24	0,01	-0,02	0,21	0,14	0,27	0,36	0,10	0,06
6	0,00	0,05	-0,04	-0,03	0,13	0,01	0,37	0,39	0,09	0,13	0,34	0,29	0,31	0,36
8	-0,16	-0,15	0,05	0,05	0,04	0,03	0,08	0,04	0,03	0,04	0,29	0,33	0,26	0,28
9	0,25	0,18	0,17	0,13	-0,03	0,01	-0,04	-0,06	0,26	0,25	0,30	0,31	0,06	0,06
11	-0,02	0,00	0,10	0,09	0,01	0,07	-0,02	-0,04	0,34	0,36	0,41	0,39	0,15	0,16
14	0,13	0,06	-0,24	-0,21	-0,11	-0,19	0,23	0,26	-0,01	-0,02	0,32	0,30	0,43	0,41
15	0,14	0,13	0,20	0,23	0,09	0,11	0,00	0,01	-0,14	-0,11	0,38	0,26	0,22	0,31
22	-0,10	-0,09	-0,08	-0,04	-0,02	-0,09	-0,02	0,03	0,17	0,17	0,27	0,28	0,25	0,23
29	-0,05	0,02	-0,09	-0,14	-0,05	-0,06	-0,03	0,00	0,18	0,25	0,33	0,40	0,29	0,18
30	-0,10	-0,07	-0,11	-0,18	-0,09	-0,03	-0,06	-0,13	0,08	0,00	0,45	0,54	0,35	0,22
33	-0,13	-0,13	-0,03	-0,04	-0,01	-0,01	-0,03	-0,04	0,17	0,14	0,40	0,42	0,23	0,21
34	0,11	0,12	0,20	0,21	0,05	0,02	-0,08	-0,06	0,28	0,28	0,33	0,33	0,19	0,19
38	0,08	0,06	0,12	0,12	0,03	0,06	-0,05	-0,06	0,09	0,08	0,44	0,45	0,37	0,37
40	-0,21	-0,19	0,12	0,09	0,11	0,07	0,07	0,05	0,10	0,13	0,46	0,56	0,37	0,22
42	-0,09	-0,09	-0,08	-0,08	-0,05	-0,09	0,07	0,08	0,09	-0,01	0,32	0,35	0,34	0,26
43	-0,13	-0,13	0,07	0,10	0,00	-0,03	-0,03	0,00	0,19	0,19	0,31	0,30	0,27	0,31
Média	-0,02	-0,02	0,06	0,05	0,03	0,02	0,01	0,00	0,14	0,13	0,34	0,36	0,25	0,23
Correlação <sup>(4)</sup>	0,96 *		0,98 *		0,86 *		0,98 *		0,94 *		0,82 *		0,84 *	
----- Análise de trilha tradicional com eliminação de variáveis e reduzido grau de multicolinearidade -----														
3	0,04	0,08	0,63	0,61	-	-	-0,30	-0,33	0,35	0,34	-	-	0,22	0,22
4	-0,13	-0,13	0,58	0,51	-	-	0,30	0,24	0,46	0,45	-	-	0,29	0,27
6	-0,06	0,05	0,05	-0,03	-	-	0,53	0,45	0,25	0,30	-	-	0,67	0,68
8	-0,17	-0,16	0,07	0,07	-	-	0,11	0,06	0,08	0,08	-	-	0,55	0,61
9	0,27	0,18	0,16	0,16	-	-	-0,06	-0,04	0,53	0,52	-	-	0,28	0,30
11	-0,01	0,06	0,11	0,12	-	-	0,00	0,00	0,77	0,79	-	-	0,44	0,45
14	0,18	0,10	-0,37	-0,39	-	-	0,20	0,20	0,19	0,16	-	-	0,72	0,70
15	0,13	0,11	0,32	0,36	-	-	0,07	0,08	0,05	0,00	-	-	0,59	0,56
22	-0,13	-0,11	-0,11	-0,11	-	-	-0,05	-0,04	0,29	0,28	-	-	0,56	0,54
29	-0,04	0,07	-0,14	-0,18	-	-	-0,06	-0,01	0,30	0,38	-	-	0,67	0,63
30	-0,11	-0,08	-0,17	-0,20	-	-	-0,14	-0,17	0,28	0,30	-	-	0,95	0,96
33	-0,12	-0,11	-0,04	-0,05	-	-	-0,03	-0,04	0,57	0,56	-	-	0,52	0,52
34	0,13	0,13	0,22	0,22	-	-	-0,06	-0,06	0,46	0,46	-	-	0,45	0,45
38	0,10	0,08	0,16	0,17	-	-	-0,04	-0,04	0,42	0,42	-	-	0,84	0,85
40	-0,29	-0,20	0,31	0,22	-	-	0,29	0,27	0,19	0,26	-	-	0,94	0,93
42	-0,12	-0,11	-0,09	-0,14	-	-	0,04	0,00	0,34	0,22	-	-	0,71	0,64
43	-0,14	-0,14	0,08	0,09	-	-	-0,04	-0,03	0,35	0,38	-	-	0,57	0,60
Média	-0,03	-0,01	0,10	0,08	-	-	0,04	0,03	0,35	0,35	-	-	0,59	0,58
Correlação <sup>(4)</sup>	0,93 *		0,99 *		-	-	0,99 *		0,97 *		-	-	0,99 *	

<sup>(1)</sup> Ensaios definidos 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> ST: distribuição multivariada não-normal sem transformação de variáveis, CT: distribuição multivariada não-normal com transformação de variáveis. <sup>(4)</sup> \*Coeficiente de correlação linear de Pearson significativo entre os efeitos diretos estimados antes e após a transformação de variáveis, a 5% de probabilidade de erro, pelo teste t, com 15 graus de liberdade.

**Tabela 28** - Coeficiente de determinação da análise de trilha ( $R^2$ ) e efeito da variável residual em 17 ensaios de competição de cultivares de milho, que não apresentaram distribuição normal multivariada antes e após a transformação de variáveis.

Ensaio <sup>(1)</sup>	Coeficiente de determinação		Efeito da variável residual	
	ST <sup>(2)</sup>	CT	ST	CT
----- Análise de trilha tradicional com sete variáveis explicativas e alto grau de multicolinearidade -----				
3	0,90	0,96	0,32	0,21
4	0,78	0,78	0,46	0,46
6	0,70	0,69	0,55	0,55
8	0,52	0,53	0,69	0,68
9	0,60	0,45	0,63	0,74
11	0,64	0,60	0,60	0,64
14	0,73	0,65	0,52	0,59
15	0,69	0,62	0,56	0,61
22	0,37	0,31	0,79	0,83
29	0,49	0,40	0,71	0,77
30	0,77	0,82	0,48	0,43
33	0,66	0,69	0,58	0,55
34	0,92	0,91	0,28	0,29
38	0,69	0,70	0,56	0,55
40	0,96	0,92	0,21	0,28
42	0,54	0,47	0,68	0,73
43	0,50	0,60	0,71	0,63
Média	0,67	0,65	0,55	0,56
----- Análise de trilha sob multicolinearidade, com acréscimo de valor $k = 0,10$ na diagonal da matriz $X'X$ -----				
3	0,84	0,86	0,40	0,37
4	0,48	0,50	0,72	0,71
6	0,65	0,64	0,59	0,60
8	0,46	0,50	0,73	0,71
9	0,51	0,42	0,70	0,76
11	0,59	0,56	0,64	0,66
14	0,63	0,60	0,61	0,63
15	0,58	0,56	0,65	0,66
22	0,27	0,27	0,85	0,85
29	0,44	0,36	0,75	0,80
30	0,73	0,75	0,52	0,50
33	0,56	0,56	0,66	0,67
34	0,87	0,87	0,36	0,36
38	0,65	0,65	0,59	0,59
40	0,79	0,77	0,46	0,48
42	0,46	0,40	0,74	0,77
43	0,44	0,49	0,75	0,72
Média	0,58	0,57	0,63	0,64
----- Análise de trilha tradicional com eliminação de variáveis e reduzido grau de multicolinearidade -----				
3	0,88	0,89	0,35	0,32
4	0,51	0,50	0,70	0,70
6	0,69	0,69	0,55	0,56
8	0,48	0,52	0,72	0,70
9	0,53	0,44	0,69	0,75
11	0,63	0,59	0,61	0,64
14	0,67	0,64	0,57	0,60
15	0,59	0,58	0,64	0,65
22	0,29	0,29	0,84	0,84
29	0,46	0,37	0,73	0,80
30	0,77	0,76	0,48	0,49
33	0,58	0,57	0,65	0,66
34	0,90	0,90	0,32	0,32
38	0,68	0,69	0,56	0,56
40	0,83	0,78	0,41	0,46
42	0,49	0,41	0,72	0,77
43	0,46	0,52	0,73	0,69
Média	0,61	0,60	0,60	0,62

(<sup>1</sup>) Ensaios definidos na Tabela 1. (<sup>2</sup>) ST: distribuição multivariada não-normal sem transformação de variáveis, CT: distribuição multivariada não-normal com transformação de variáveis.

#### 4.1 Considerações finais

Com base nos 44 ensaios utilizados nesse trabalho, constatou-se que, na maioria das variáveis mensuradas, os dados se ajustaram à distribuição normal univariada (67,9% dos casos). No entanto, apenas 14 ensaios (31,8% dos ensaios) apresentaram distribuição normal multivariada (Tabela 10). Assim, além do diagnóstico da normalidade univariada, existe a necessidade da verificação da distribuição normal multivariada. Nos 30 ensaios que não se ajustaram a normal multivariada, a transformação de dados através da família de transformações Box-Cox foi eficiente em 13 ensaios (43,3% dos ensaios sem distribuição normal multivariada). Não foram encontrados dados na literatura, indicando as possíveis implicações da não-normalidade multivariada sobre os efeitos diretos estimados pela análise de trilha.

Em relação aos coeficientes de correlação das variáveis explicativas com a produtividade de grãos (PROD), verificou-se maior intensidade de associação do número de espigas (NE), da prolificidade (PLI) e do número de plantas (NP) sobre a produtividade de grãos, nessa ordem de importância. Os coeficientes de correlação entre as variáveis explicativas e a produtividade de grãos, estimados antes e após a transformação de dados de ensaios que não apresentaram distribuição normal multivariada foram semelhantes entre si ( $r \geq 0,95$ ) (Tabelas 17 e 23).

A utilização dos quatro métodos de diagnóstico de multicolinearidade foi eficiente na mensuração do grau de multicolinearidade entre as variáveis explicativas, na matriz de correlação  $X'X$ . Houve multicolinearidade severa nos 44 ensaios e o fator de inflação de variância juntamente com a tolerância foram importantes na mensuração da multicolinearidade de cada variável explicativa com as demais variáveis explicativas. Já o número de condição e o determinante da matriz foram importantes na mensuração do grau de multicolinearidade na matriz de correlação  $X'X$  como um todo, permitindo identificar problemas de condicionamento e de singularidade da matriz.

Verificou-se que de modo geral, a transformação de dados reduziu o grau de multicolinearidade e a variabilidade dos efeitos diretos estimados na análise de trilha tradicional, com alto grau de multicolinearidade. No entanto, quando o grau de multicolinearidade foi reduzido pela adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  ou pela eliminação das variáveis AIE e NE, a transformação de dados não alterou de modo expressivo o grau de multicolinearidade e os efeitos diretos. Nesses casos, o



grau de multicolinearidade foi pequeno antes e após a transformação de dados (Tabelas 18, 19, 20, 24, 25 e 26) e os efeitos diretos, estimados pela análise de trilha sob multicolinearidade e pela análise de trilha tradicional com eliminação de variáveis, também foram similares antes e após a transformação de dados ( $r \geq 0,55$ ) (Tabelas 21 e 27).

Entre as variáveis estudadas, apenas o florescimento masculino não apresentou multicolinearidade severa com as demais variáveis explicativas da matriz de correlação  $X'X$ . Assim, seus efeitos diretos estimados através das análises de trilha tradicional, sob multicolinearidade e tradicional com eliminação de variáveis foram similares (Tabelas 16, 21 e 27).

Todas as demais variáveis explicativas tiveram elevado grau de multicolinearidade nos 44 ensaios. Assim, os efeitos diretos estimados pela análise de trilha tradicional, com alto grau de multicolinearidade, não apresentaram sentido biológico, em ensaios com distribuição normal multivariada ( $-21,94 \leq \text{efeito direto} \leq 20,51$ ), com distribuição normal multivariada após a transformação de dados ( $-23,20 \leq \text{efeito direto} \leq 29,21$ ) e sem distribuição normal multivariada ( $-51,98 \leq \text{efeito direto} \leq 66,42$ ). Nesse sentido, Carvalho (1995) afirma que quando o grau de multicolinearidade é elevado, os coeficientes de trilha podem atingir valores elevados e pouco confiáveis, assumindo valores sem nenhum sentido prático em relação ao fenômeno estudado.

Assim como nos resultados obtidos nesse trabalho, em canola, Coimbra et al. (2005) verificaram uma super-estimativa dos efeitos diretos sobre a produtividade de grãos ( $-10,63 \leq \text{efeito direto} \leq 6,24$ ). Em arroz, Shrivastava e Sharma (1976) verificaram a ocorrência de efeitos diretos negativos de caracteres primários sobre a produtividade de grãos, devido ao elevado grau de multicolinearidade. Em feijão, Carvalho e Cruz (1996) observaram que as consequências adversas da multicolinearidade sobre a estimativa dos efeitos diretos tornam se evidentes na medida em que o grau da multicolinearidade aumenta. Na cultura da soja, Bizeti et al. (2004) verificaram estimativas de efeitos diretos entre  $-0,86$  e  $3,91$ , na presença de alto grau de multicolinearidade. Com base nos resultados desse trabalho e nos trabalhos realizados por outros pesquisadores, fica evidenciada a importância da realização do diagnóstico de multicolinearidade e a adoção de medidas que proporcionem a redução dos efeitos adversos da mesma, sobre a estimativa dos efeitos diretos por meio da análise de trilha.

A adição da constante  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , foi eficiente na redução do grau de multicolinearidade entre as variáveis explicativas em todos os ensaios. Assim, os valores de fator de inflação de variância foram menores que 10 (Tabelas 14, 18 e 24), a tolerância foi maior que 0,10 (Tabelas 15, 19 e 25), o número de condição foi menor

que 100 e o determinante da matriz de correlação foi maior que zero (Tabelas 15, 20 e 26). Assim, com base nesses quatro métodos de diagnóstico, pode-se concluir que a adição de  $k = 0,10$  é satisfatória para reduzir a multicolinearidade.

Os efeitos diretos estimados a partir da análise de trilha sob multicolinearidade, com adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$ , foram estáveis nos ensaios com distribuição normal multivariada ( $-0,50 \leq$  efeito direto  $\leq 0,60$ ), com distribuição normal multivariada após a transformação de dados ( $-0,41 \leq$  efeito direto  $\leq 0,51$ ) e sem distribuição normal multivariada ( $-0,42 \leq$  efeito direto  $\leq 0,56$ ). De modo geral, as variáveis NE, PLI e NP apresentaram os maiores efeitos diretos sobre a produtividade de grãos, nessa ordem de importância. Esses valores estão de acordo com os maiores coeficientes de correlação dessas variáveis explicativas com a produtividade de grãos. As demais variáveis explicativas apresentaram efeitos diretos de baixa magnitude.

Como ocorreu a redução da multicolinearidade e estabilização das estimativas dos efeitos diretos, concluiu-se que a análise de trilha sob multicolinearidade é adequada para a estimativa dos efeitos diretos na presença de alto grau de multicolinearidade. A utilização da análise de trilha sob multicolinearidade, com adição de uma constante  $k$  na diagonal da matriz  $X'X$ , também foi eficiente para a redução dos efeitos adversos do elevado grau de multicolinearidade nas culturas de pimentão (CARVALHO et al., 1999), de soja (BIZETI et al., 2004), de canola (COIMBRA et al., 2005), de girassol (AMORIM et al., 2008), de mamoeiro (OLIVEIRA et al., 2010) e de amendoim (LUZ et al., 2011). Mas não foi eficiente na estabilização dos efeitos diretos na cultura do tomate (RODRIGUES et al., 2010).

A eliminação das variáveis AIE e NE também contribuiu para que houvesse uma redução significativa do grau de multicolinearidade, sendo que em todos os casos, os valores de fator de inflação de variância foram menores que 10 (Tabelas 14, 18 e 24), a tolerância foi maior que 0,10 (Tabelas 15, 19 e 25), o número de condição foi menor que 100 e o determinante da matriz de correlação  $X'X$  foi maior que zero (Tabelas 15, 20 e 26). Assim como na análise de trilha sob multicolinearidade, houve uma estabilização dos efeitos diretos estimados (Tabelas 16, 21 e 27). De modo geral, os maiores efeitos diretos foram de PLI e NP sobre a produtividade de grãos, nessa ordem de importância e os efeitos diretos da variável NE estão preditos nessas duas variáveis.

A eliminação de variáveis também apresentou bons resultados na redução do grau de multicolinearidade e na estabilização dos efeitos diretos em arroz (SHRIVASTAVA; SHARMA, 1976), em pimentão (CARVALHO et al., 1999), em soja (CARVALHO et al.,

2002; BIZETI et al., 2004), em milho (LOPES et al., 2007), em trigo (GONDIM et al., 2008) e em tomateiro (RODRIGUES et al., 2010).

Por fim, a análise de trilha sob multicolinearidade é indicada quando o pesquisador não deseja excluir variáveis explicativas importantes na predição de uma variável básica. Com base nos 44 ensaios, ficou evidenciado que essa técnica pode ser utilizada com alto grau de confiabilidade nas estimativas dos efeitos diretos. A análise de trilha tradicional, com eliminação de variáveis, também é eficiente na redução da multicolinearidade e proporciona confiabilidade das estimativas de efeitos diretos. O maior problema, associado à eliminação de variáveis, é a possibilidade de se cometer erros de especificação pela exclusão de variáveis importantes na predição de uma variável básica. No entanto, a manutenção de variáveis com informações redundantes também não é recomendada. Assim, é importante o conhecimento prévio das possíveis associações entre variáveis por parte do pesquisador.

No caso particular dos 44 ensaios de competição de cultivares de milho, constatou-se que, quando eliminou-se variáveis, os efeitos diretos de NE sobre a produtividade de grãos passaram a ser explicados por PLI e NP. Assim, não houve perdas de predição nas análises de trilha. Além disso, os quatro métodos de diagnóstico de multicolinearidade evidenciaram que a eliminação de variáveis foi mais eficiente na redução do grau de multicolinearidade em relação à adição de  $k = 0,10$  na diagonal da matriz de correlação  $X'X$  e tem como vantagem o fato dos estimadores não serem viesados. Os coeficientes de determinação das análises de trilha tradicional, com eliminação de variáveis, foram superiores aos obtidos pela análise de trilha sob multicolinearidade e assim, verificou-se que a produtividade de grãos é melhor explicada mediante a utilização da análise de trilha tradicional, com eliminação de variáveis.

## **5 CONCLUSÕES**

1. A transformação de dados a fim de obter a normalidade multivariada contribui para a redução do grau de multicolinearidade e na estabilização das estimativas dos efeitos diretos em análise de trilha com alto grau de multicolinearidade.
2. Os efeitos adversos do alto grau de multicolinearidade na estimativa dos efeitos diretos de análises de trilha são maiores que a não-normalidade multivariada.
3. A análise de trilha tradicional com eliminação de variáveis é mais adequada que a análise de trilha sob multicolinearidade.



## 6 REFERÊNCIAS

ALVI, M.B. et al. 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.

AMORIM, E.P. et al. Correlações e análise de trilha em girassol. **Bragantia**, v.67, p.307-316, 2008.

BARROS, L.B.; MOREIRA, R.M.P.; FERREIRA, J.M. Phenotypic, additive genetic and environment correlations of maize landraces populations in family farm systems. **Scientia Agricola**, v.67, p.685-691, 2010.

BIZETI, H.S. et al. Path analysis under multicollinearity in soybean. **Brazilian Archives of Biology and Technology**, v.47, p.669-676, 2004.

BORÉM, A.; MIRANDA, G.V. **Melhoramento de plantas**. 5 ed. Viçosa: UFV, 2009. 529 p.

BOX, G.E.P.; COX, D.R. An analysis of transformations. **Journal of the Royal Society**, v.26, p.211-252, 1964.

CABRAL, P.D.S. et al. Análise de trilha do rendimento de grãos de feijoeiro (*Phaseolus vulgaris* L.) e seus componentes. **Revista Ciência Agronômica**, v.42, p.132-138, 2011.

CAMPOS, H. **Estatística experimental não paramétrica**. Piracicaba: ESALQ, 1983. 349 p.

CANTELMO, N.F.; FERREIRA, D.F. Desempenho de testes de normalidade multivariados avaliado por simulação Monte Carlo. **Ciência e Agrotecnologia**, v.31, p.1630-1636, 2007a.

CANTELMO, N.F.; FERREIRA, D.F. Desempenhos dos testes de normalidade multivariada do programa R e de Shapiro-Wilk de Royston avaliados por simulação Monte Carlo. **Revista de Matemática e Estatística**, v.25, p.07-17, 2007b.

CARGNELUTTI FILHO, A.; GUADAGNIN, J.P. Planejamento experimental em milho. **Revista Ciência Agronômica**, v.42, p.1009-1016, 2011.

CARGNELUTTI FILHO, A.; STORCK, L. Medidas do grau de precisão experimental em ensaios de competição de cultivares de milho. **Pesquisa Agropecuária Brasileira**, v. 44, p. 111-117, 2009.

CARGNELUTTI FILHO, A.; STORCK, L.; RIBEIRO, N.D. Agrupamento de cultivares de feijão em presença e em ausência de multicolinearidade. **Ciência Rural**, v.39, p.2409-2418, 2009.

CARVALHO, C.G.P. et al. Análise de trilha sob multicolinearidade em pimentão. **Pesquisa Agropecuária Brasileira**, v.34, p.603-613, 1999.

CARVALHO, C.G.P. et al. Correlações e análise de trilha em linhagens de soja semeadas em diferentes épocas. **Pesquisa Agropecuária Brasileira**, v.37, p.311-320, 2002.

CARVALHO, S.P. **Métodos alternativos de estimação de coeficientes de trilha e índices de seleção, sob multicolinearidade**. Viçosa: UFV, 1995. 163 p.

CARVALHO, S.P.; 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.

CHURATA, B.G.M.; AYALA-OSUNA, J.T. Correlações genotípicas, fenotípicas e de ambiente e análise de trilha em caracteres avaliados no composto de milho (*Zea mays*) arquitetura. **Ciência e Prática**, v.43, p.628-636, 1996.

COIMBRA, J.L.M. et al. Análise de trilha dos componentes do rendimento de grãos em genótipos de canola. **Ciência Rural**, v.34, p.1421-1428, 2004.

COIMBRA, J.L.M. et al. Consequências da multicolinearidade sobre a análise de trilha em canola. **Ciência Rural**, v.35, p.347-352, 2005.

CONAB - Companhia Nacional de Abastecimento. **Séries históricas: milho 1<sup>a</sup> e 2<sup>a</sup> safras**. CONAB, 2011. Disponível em: <[http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina\\_objcmsconteudos=2#A\\_objcmsconteudos](http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina_objcmsconteudos=2#A_objcmsconteudos)>. Acesso em: 27 dez. 2011.

CORRAR, L.J.; PAULO, E.; DIAS FILHO, J.M. **Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2007. 541 p.

COUTO, M.R.M. et al. Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, v.39, p.1701-1707, 2009.

CRUZ, C.D. **Programa Genes: Estatística Experimental e Matrizes**. Viçosa: UFV, 2006. 285 p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa: UFV, 2006. v.2, 585 p.

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

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: UFV, 2004. 480 p.

DAHER, R.F. et al. Análise de trilha de caracteres forrageiros do capim-elefante (*Pennisetum purpureum* Schum.). **Ciência Rural**, v.34, p.1531-1535, 2004.

DRAPER, N.R.; SMITH, H. **Applied Regression Analysis** (Wiley series in probability and mathematical statistics). 2. ed. New York: John Wiley e Sons, 1981. 709 p.

DUVIC, D.N. The contribution of breeding to yield advances in maize (*Zea mays* L.). **Advances in Agronomy**, v.86, p.83-145, 2005.

ESPÓSITO, D.P. et al. Análise de trilha usando valores fenotípicos e genotípicos para componentes do rendimento na seleção de famílias de cana-de-açúcar. **Ciência Rural**, v.42, p.38-44, 2012.

FALCONER, D.S. **Introdução à genética quantitativa**. Viçosa: UFV, 1987. 279 p.

FANCELLI, A.L.; DOURADO NETO, D. **Produção de milho**. Guaíba: Agropecuária, 2004. 360 p.

FAO - Food and Agriculture Organization of the United Nations. **FAOSTAT - Countries by commodity: Top production - Maize - 2009**. FAO, 2009. Disponível em: <<http://faostat.fao.org/site/339/default.aspx>>. Acesso em: 27 dez. 2011.



FARRAR, D.E.; GLAUBER, R.R. Multicollinearity in regression analysis: the problem revisited. **Review of Economics and Statistics**, v.49, p.92-107, 1967.

FÁVERO, L.P. et al. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009. 646 p.

FERREIRA, D.F. **Estatística Multivariada**. 1. ed. Lavras: UFLA, 2008. 662 p.

GALARÇA, S.P. et al. Correlação de Pearson e análise de trilha identificando variáveis para caracterizar porta-enxerto de *Pyrus communis* L. **Ciência e Agrotecnologia**, v.34, p.860-869, 2010.

GONDIM, T.C.O. et al. Análise de trilha para componentes do rendimento e caracteres agronômicos de trigo sob desfolha. **Pesquisa Agropecuária Brasileira**, v.43, p.487-493, 2008.

GUJARATI, D.N. **Econometria Básica**. 4. Ed. Rio de Janeiro: Campus Elsevier, 2006. 812 p.

HAIR, J.F. et al. **Análise multivariada de dados**. 6 ed. Porto Alegre: Bookman, 2009. 688 p.

HOOGERHEIDE, E.S.S. et al. Correlações e análise de trilha de caracteres tecnológicos e a produtividade de fibra de algodão. **Pesquisa Agropecuária Brasileira**, v.42, p.1401-1405, 2007.

KANG, M.S. et al. Genetic and phenotypic path analysis and heritability in sugarcane. **Crop Science**, v. 23, p.643-647, 1983.

KOLMOGOROV, A.N. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell Istituto degli Attuari**, v.4, p.83-91, 1933.

KUTNER, M.H.; NACHTSHEIM C.J.; NETER, J. **Applied linear regression models**. 4. ed. Boston: McGraw-Hill, 2004. 701 p.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality. **Biometrika**, v.62, p.399-402, 1967.

LIMA NETO, F.P.; SOUZA JÚNIOR, C.L. Number of recombinations and genetic properties of a maize population undergoing recurrent selection. **Scientia Agricola**, v.66, p.52-58, 2009.

LOPES, S.J. et al. Relações de causa e efeito em espigas de milho relacionadas aos tipos de híbridos. **Ciência Rural**, v.37, p.1536-1542, 2007.

LORENTZ, L.H.; FORTES, F.O.; LÚCIO, A.D. Análise de trilha entre as variáveis das análises de sementes de espécies florestais exóticas do Rio Grande do Sul. **Revista Árvore**, v.30, p.567-574, 2006.

LÚCIO, A.D. et al. Excesso de zeros nas variáveis observadas: estudo de caso em experimentos com brócolis. **Bragantia**, v.69, p.1036-1046, 2010.

LUZ, L.N.; SANTOS, R.C.; MELO FILHO, P.A. Correlations and path analysis of peanut traits associated with the peg. **Crop Breeding and Applied Biotechnology**, v.11, p.88-93, 2011.

MARCHEZAN, E. et al. Análise de coeficiente de trilha para os componentes de produção em arroz. **Ciência Rural**, v.35, p.1027-1033, 2005.

MARDIA, K.V. Measures of multivariate skewness and kurtosis with applications. **Biometrika**, v.57, p.519-530, 1970.

MAROCO, J. **Análise estatística com utilização do SPSS**. 3. ed. Lisboa: Edições Sílabo, 2007. 824 p.

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

OLIVEIRA, E.J. et al. Correlações genéticas e análise de trilha para número de frutos comerciais por planta em mamoeiro. **Pesquisa Agropecuária Brasileira**, v.45, p.855-862, 2010.

PEREIRA FILHO, N.A.; FERREIRA, D.F. Teste Monte Carlo de normalidade univariada. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 56 E SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 14., 2011, Maringá - PR. **Anais eletrônicos...** Maringá: UEM, 2011. Disponível em: <<http://www.rbras.org.br/rbras56/doku.php?id=anais>>. Acesso em: 15 jan. 2012.

RAFIQ, C.M. et al. Studies on heritability, correlation and path analysis in maize (*Zea mays* L.). **Journal of Agricultural Research**, v.48, p.35-38, 2010.

RANGEL, R.M.; AMARAL JÚNIOR, A.T.; FREITAS JÚNIOR, S.P. Associação entre características agronômicas e capacidade de expansão em população de milho pipoca sob seleção recorrente. **Ciência e Agrotecnologia**, v.35, p.225-233, 2011.

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.R-project.org>>.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v.37, p.182-194, 2007.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

RODRIGUES, G.B. et al. Análise de trilha de componentes de produção primários e secundários em tomateiro do grupo Salada. **Pesquisa Agropecuária Brasileira**, v.45, p.155-162, 2010.

ROYSTON, J.B. Some techniques for assessing multivariate based on the Shapiro -Wilk W. **Applied Statistics**, v.32, p.121-133, 1983.

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.

SANTOS, A.C.; FERREIRA, D. F. Definição do tamanho amostral usando simulação Monte Carlo para o teste de normalidade baseado em assimetria e curtose: I. Abordagem univariada. **Ciência e Agrotecnologia**, v.27, p.432-437, 2003.

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete sample). **Biometrika**, v.52, p.591-611, 1965.

SHRIVASTAVA, M.N.; SHARMA, K.K. Analysis of path coefficients in rice. **Zeitschrift fuer Pflanzenzuechtung**, v.77, p.174-177, 1976.

SHUKLA, S.; SINGH, K.; PUSHPENDRA. Correlation and path coefficient analysis of yield and its components in soybean (*Glycine max* (L.) Merrill). **Soybean Genetics Newsletter**, v.25, p.67-70, 1998.

SILVA, G.O. et al. Correlações entre caracteres de aparência e rendimento e análise de trilha para aparência de batata. **Bragantia**, v.66, p.381-388, 2007.

SILVA, M.A. et al. Análise de trilha em caracteres produtivos de *Pennisetum* sob corte em Itaimbé, Pernambuco. **Revista Brasileira de Zootecnia**, v.37, p.1185-1191, 2008.

SRIVASTAVA, M.S. **Methods of multivariate statistics** (Wiley series in probability and statistics). New York: John Wiley e Sons, 2002. 697 p.

STATSOFT. **Statistica 7.0 Software**. Tucksa, USA, 2005.

STORCK, L.; BISOGNIN, D.A.; CARGNELUTTI FILHO, A. Ganho genético decorrente da substituição anual de cultivares de milho. **Pesquisa Agropecuária Brasileira**, v.40, p.881-886, 2005.

USDA - United States Department of Agriculture. **Statistics by Subject: National Statistics for Corn**, 2011a. Disponível em: <[http://www.nass.usda.gov/Statistics\\_by\\_Subject/result.php?AFBDFE1E-1AFC-35DE-8A93-7FB72F0DA089&sector=CROPS&group=FIELD%20CROPS&comm=CORN](http://www.nass.usda.gov/Statistics_by_Subject/result.php?AFBDFE1E-1AFC-35DE-8A93-7FB72F0DA089&sector=CROPS&group=FIELD%20CROPS&comm=CORN)>. Acesso em: 27 dez. 2011.

USDA - United States Department of Agriculture. **World Agricultural Supply and Demand Estimates**. World Agricultural Outlook Board - WASDE - 501, 2011b. 40 p. Disponível em: <<http://www.usda.gov/oce/commodity/wasde/latest.pdf>>. Acesso em: 27 dez. 2011.

VASCONCELLOS, M.A.S.; ALVES, D. **Manual de econometria**. São Paulo: Atlas, 2000. 308 p.

WICHERS, C.R. The detection of multicollinearity: a comment. **Review of Economics and Statistics**, v.57 p.365-366, 1975.

WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**, v.20, p.557-585, 1921.

WRIGHT, S. Theory of path coefficients. **Genetics**, v.8, p.239-285, 1923.