

Estatística

Número suficiente de repetições para análise de trilha em milho

Sufficient number of replications for path analysis in maize

Alberto Cargnelutti Filho¹, Marcos Toebe¹, Bruna Mendonça Alves¹, Ismael Mario Marcio Neu¹, Jéssica Andiana Kleinpaul¹

¹Universidade Federal de Santa Maria, Santa Maria, RS, Brasil

RESUMO

Supõe-se que o número de repetições interfere nas estimativas dos coeficientes da análise de trilha. Assim, o objetivo deste trabalho foi determinar o número suficiente de repetições para a análise de trilha de caracteres de cultivares de milho. Foi conduzido um experimento com 15 cultivares de milho no delineamento de blocos ao acaso com nove repetições, sendo mensuradas sete variáveis. Foram organizados 511 arquivos de dados (matrizes) formados pelas combinações das nove repetições, em grupos de 1, 2, 3, 4, 5, 6, 7, 8 e 9 repetições. Em cada matriz, contendo as médias de 15 cultivares das sete variáveis, foram estimados os coeficientes de correlação linear de Pearson, realizado o diagnóstico de multicolinearidade, a análise de trilha e construídos diagramas de dispersão. O número suficiente de repetições para a análise de trilha, foi determinado a partir das estimativas dos parâmetros do modelo quadrático de resposta com platô. Com o aumento do número de repetições há melhoria na precisão das estimativas dos coeficientes da análise de trilha, porém os ganhos em precisão diminuem gradativamente. Seis repetições são suficientes para realizar a análise de trilha de caracteres agrônômicos de cultivares de milho e pode ser adotado como referência para o planejamento de experimentos futuros.

Palavras-chave: *Zea mays* L; Análise de correlação; Modelo quadrático de resposta com platô

ABSTRACT

The number of replications is assumed to interfere in the estimates of the path analysis coefficients. Thus, the objective of this work was to determine the sufficient number of replications for the path analysis of traits in maize cultivars. An experiment was conducted with 15 maize cultivars in a complete randomized block design with nine replications, and seven variables were measured. Then, 511 data files (matrices) formed by all combinations of the nine replications were organized, in groups of 1, 2, 3, 4, 5, 6, 7, 8 and 9 replications. In each matrix, containing the averages of 15 cultivars for the seven variables, Pearson's linear correlation coefficients were estimated, the multicollinearity diagnostics and path analysis were performed and dispersion diagrams were constructed. The sufficient number of replications for the path analysis was

determined from the parameter estimates of the quadratic response plateau model. With the replications number increases, the accuracy of the path analysis coefficient estimates improves, but the gains in accuracy gradually decrease. Six replications are sufficient to perform the path analysis of agronomic traits of maize cultivars and can be used as a reference for designing future experiments.

Key words: *Zea mays* L; Correlation analysis; Quadratic response plateau model

1 INTRODUÇÃO

O número de repetições é um aspecto importante a ser considerado em experimentos realizados a campo e tem sido um questionamento comum de pesquisadores. As repetições são necessárias para estimar o erro experimental. Quanto maior o número de repetições melhor será a estimativa da média, pois o erro padrão da média é fornecido por s/\sqrt{b} , em que s é o desvio padrão do erro experimental e b é o número de repetições (BANZATTO; KRONKA, 2013; STORCK *et al.*, 2016). O uso de número apropriado de repetições, por possibilitar uma boa estimativa do erro experimental, contribui para aumentar a precisão das estimativas das médias e, ao mesmo tempo, aumentar o poder dos testes estatísticos (CARGNELUTTI FILHO *et al.*, 2018). Estudos sobre o número de repetições têm sido realizados em milho (CARGNELUTTI FILHO *et al.*, 2010; NESI *et al.*, 2010; CARGNELUTTI FILHO; GUADAGNIN, 2011; MENDOZA; BUITRAGO, 2015; CARGNELUTTI FILHO *et al.*, 2018) e milho pipoca (CATAPATTI *et al.*, 2008), revelando ganhos promissores de precisão experimental com o aumento do número de repetições.

As relações lineares entre um conjunto de caracteres podem ser investigadas por meio do coeficiente de correlação linear de Pearson e por procedimentos estatísticos complementares, tais como, a análise de trilha. Esta análise permite desdobrar os coeficientes de correlação em efeitos diretos e indiretos de variáveis explicativas sobre uma variável principal e identificar se há associação linear de causa e efeito, possibilitando a identificação de caracteres para a seleção indireta (CRUZ *et al.*, 2012; CRUZ *et al.*, 2014). A análise de trilha tem sido utilizada para avaliar os efeitos diretos e indiretos de caracteres de cultivares de milho sobre a produtividade de grãos (TOEBE; CARGNELUTTI FILHO, 2013 a, b; MASON *et al.*, 2019) e sobre a produção de silagem (CREVELARI *et al.*, 2018).

A análise de trilha é comumente realizada a partir de dados sem repetição, ou sem delineamento experimental. Portanto, um aspecto importante a ser investigado é a interferência do número de repetições na precisão das estimativas dos coeficientes das análises de trilha. Além disso, é importante definir o número suficiente de repetições para estimação desses coeficientes. Número insuficiente de repetições pode gerar estimativas imprecisas. Por outro lado, o aumento demasiado do número de repetições pode significar desperdício de tempo, mão de obra e recursos financeiros (CARGNELUTTI FILHO *et al.*, 2018). A questão importante a ser investigada é o número suficiente ou mínimo de repetições para que se tenha melhor precisão das estimativas dos coeficientes da análise de trilha, considerando que a partir de determinado número de repetições o ganho em precisão pode ser insignificante.

Diante desse cenário de desconhecimento da interferência do número de repetições sobre os resultados da análise de trilha, o objetivo deste trabalho foi determinar o número suficiente de repetições para a execução da análise de trilha com caracteres agrônômicos associados a produtividade de grãos em cultivares de milho (*Zea mays* L.).

2 MATERIAL E MÉTODOS

2.1 Experimento de campo

Foi conduzido um experimento com 15 cultivares de milho (*Zea mays* L.), no delineamento blocos completos ao acaso com nove repetições (b), no ano agrícola 2012/2013. As cultivares foram as integrantes da rede estadual de avaliação de cultivares, coordenada pela Fundação Estadual de Pesquisa Agropecuária (FEPAGRO). O experimento foi conduzido em área do Departamento de Fitotecnia, da Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul (latitude 29°42'S, longitude 53°49'W e 95 m de altitude). As unidades experimentais (parcelas) foram constituídas de duas fileiras com 5 m de comprimento, espaçadas em 0,8 m. A densidade foi de 62.500 plantas ha⁻¹.

Em cada unidade experimental foi contado o número de dias da semeadura até 50% do florescimento masculino (FM) e o número de dias da semeadura até 50% do florescimento feminino (FF). Na colheita, com base em todas as plantas da unidade experimental, foram mensuradas as alturas de planta (AP) e de espiga (AE), contado o número de plantas (NP) e o número de espigas (NE) e determinada a produtividade de grãos (PROD), corrigida para 13% de umidade.

2.2 Análise de dados

Primeiramente, foi construído um diagrama de dispersão com as 135 observações (15 cultivares \times 9 repetições ou blocos) dos caracteres FM, FF, AP, AE, NP, NE e PROD, a fim de investigar, visualmente, a relação entre eles. A partir do experimento referência, isto é, com todas as repetições ($b = 9$), formaram-se novos arquivos de dados por combinações de 1, 2, 3, ..., $b-1$ repetições. Assim, foi obtido um arquivo formado com as 9 repetições [experimento referência; $C_{(9,9)}=1$ arquivo] e os demais pela combinação das 9 repetições em grupos de uma [$C_{(9,1)}=9$ arquivos], duas [$C_{(9,2)}=36$ arquivos], três [$C_{(9,3)}=84$ arquivos], quatro [$C_{(9,4)}=126$ arquivos], cinco [$C_{(9,5)}=126$ arquivos], seis [$C_{(9,6)}=84$ arquivos], sete [$C_{(9,7)}=36$ arquivos] e oito [$C_{(9,8)}=9$ arquivos] repetições, totalizando 511 arquivos de dados, para realizar as análises quanto ao número suficiente de repetições para a análise de trilha em milho.

Para cada arquivo, foi calculada a média das variáveis FM, FF, AP, AE, NP, NE e PROD entre as repetições de cada cultivar, obtendo-se uma matriz de médias com dimensões de 15 linhas (cultivares) e 7 colunas (variáveis). A seguir, foi determinada a matriz de coeficientes de correlação linear de Pearson entre os pares de variáveis. Na matriz com as variáveis FM, FF, AP, AE, NP e NE, foi realizado o diagnóstico de multicolinearidade (HAIR *et al.*, 2009; CRUZ *et al.*, 2014), por meio do número de condição (NC) e fator de inflação de variância (FIV). Conforme Montgomery e Peck (1982) há multicolinearidade fraca entre as variáveis quando $NC \leq 100$; moderada a forte quando $100 < NC < 1.000$; e severa quando $NC \geq 1.000$. Considerou-se haver multicolinearidade severa quando o valor de FIV foi maior que 10 (HAIR *et al.*, 2009; CRUZ *et al.*, 2014). O NC é um indicador com interpretação para todas as variáveis, enquanto

que o FIV tem a vantagem de informar a inflação de variância para cada variável e, por isso, nesse estudo, considerou-se o maior valor de FIV entre as variáveis. Para cada arquivo foi realizada a análise de trilha da variável principal PROD em função das variáveis explicativas FM, FF, AP, AE, NP e NE, totalizando 511 análises de trilha, conforme metodologia descrita em Cruz *et al.* (2012) e Cruz *et al.* (2014).

A fim de avaliar, visualmente, a variabilidade das estimativas dos coeficientes de correlação linear de Pearson e dos efeitos diretos das variáveis FM, FF, AP, AE, NP e NE sobre a PROD, bem como do coeficiente de determinação, da variável residual, do número de condição e FIV, em função do número de repetições, foram construídos diagramas de dispersão. Optou-se por apenas discutir os efeitos indiretos e não apresentá-los devido a limitação de espaço.

Para determinar o número suficiente de repetições para a análise de trilha, foram obtidas as estimativas dos parâmetros do modelo quadrático de resposta com platô das variáveis Y (coeficientes de correlação linear de Pearson entre FM *versus* PROD, FF *versus* PROD, AP *versus* PROD, AE *versus* PROD, NP *versus* PROD e NE *versus* PROD) em função da variável X (número de repetições). Nesse modelo matemático, o ponto de início do platô é denominado de X_s (número suficiente de repetições), que é localizado por meio da primeira derivada da equação quadrática, isto é, pelo ponto de máximo ou de mínimo. Quando X é maior ou igual a X_s , este ponto se torna uma constante, transformando-se numa reta com inclinação zero, denominada platô, cujo modelo é dado por $Y = p$, em que $p = \text{platô}$; e quando X é menor que X_s , este ponto se encontra na parte quadrática do modelo ($Y = a + bX + cX^2 + \text{erro}$), sendo descrito por uma parábola (GONÇALVES *et al.*, 2012). Assim, obteve-se seis valores de X_s , ou seja, um valor para cada variável. Como as demais estatísticas relacionadas à análise de trilha dependem dos coeficientes de correlação linear de Pearson, considerou-se a média dos seis valores de X_s , arredondada para o inteiro superior, como sendo o número suficiente de repetições para a análise de trilha. As análises estatísticas foram realizadas com auxílio dos programas Microsoft Office Excel 2007, GENES (CRUZ, 2016) e R (R Core Team, 2019).

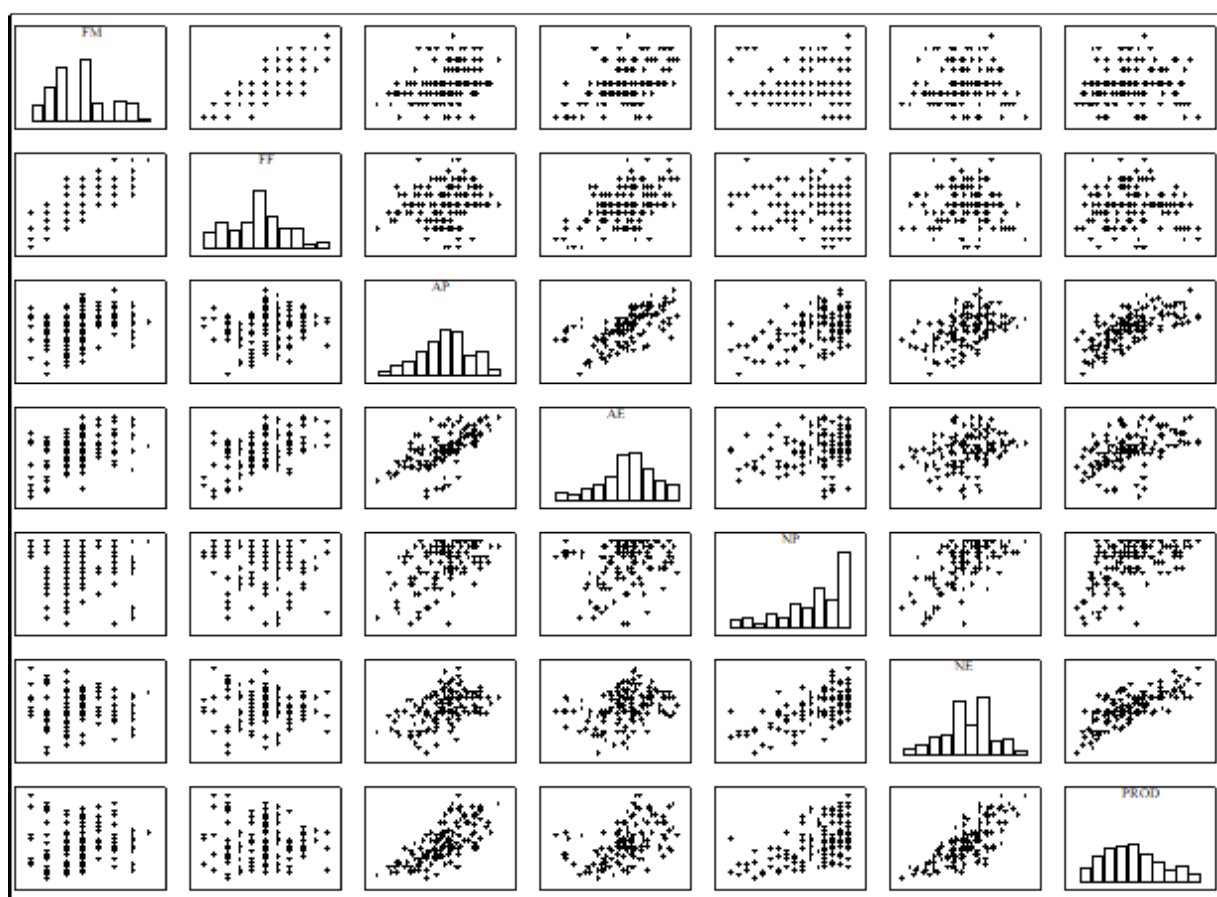
3 RESULTADOS E DISCUSSÃO

O diagrama de dispersão entre os caracteres número de dias da semeadura até 50% do florescimento masculino (FM), número de dias da semeadura até 50% do florescimento feminino (FF), alturas de planta (AP) e de espiga (AE), números de plantas (NP) e de espigas (NE) e a produtividade de grãos (PROD), sugere alguns padrões de linearidade (Figura 1). Percebe-se padrão linear entre FM *versus* FF, AP *versus* AE, NP *versus* NE, e ainda entre AP, AE, NP e NE com a PROD (Figura 1 e Tabela 1). Assim, o estudo dessas relações lineares por meio dos coeficientes de correlação linear de Pearson (r) e a análise de trilha de FM, FF, AP, AE, NP e NE sobre a PROD é adequado (CRUZ *et al.*, 2012; CRUZ *et al.*, 2014).

O diagnóstico de multicolinearidade, nas 511 matrizes de coeficientes de correlação linear de Pearson (r) entre as variáveis FM, FF, AP, AE, NP e NE, revelou que o número de condição (NC) oscilou entre 17 e 117, com média 46, e o fator de inflação de variância (FIV) entre 3 e 14, com média 7 (Figura 2). Devido a limitação de espaço para apresentação das estimativas de NC e de FIV das 511 matrizes de r , obtidas a partir dos dados de 9, 36, 84, 126, 126, 84, 36, 9 e 1 arquivos formados com 1, 2, 3, 4, 5, 6, 7, 8 e 9 repetições, respectivamente, optou-se pelo seguinte: na figura 2 foram plotadas as 511 estimativas e na Tabela 1 foram apresentadas somente as dez estimativas das situações extremas, ou seja, as estimativas obtidas com os nove arquivos de uma repetição e com um arquivo de nove repetições. Assim, considerando apenas as nove matrizes de r , obtidas a partir dos arquivos formados com apenas uma repetição ou bloco, isto é, a primeira matriz obtida somente com os dados do bloco 1 (b1), a segunda matriz obtida somente com os dados do bloco 2 (b2) e assim, sucessivamente, até a nona matriz obtida somente com os dados do bloco 9 (b9), o NC oscilou entre 20,361 e 69,609 e o FIV entre 3,264 e 13,206. Já na matriz de r obtida com os dados das nove repetições ou blocos (arquivo referência) o NC foi 53,168 e o FIV foi 7,028 (Tabela 1). Esses valores indicam multicolinearidade fraca a moderada, o que permite a realização adequada da análise de trilha (MONTGOMERY; PECK, 1982; CRUZ *et al.*, 2012; CRUZ *et al.*, 2014) e, conseqüentemente, a investigação do número suficiente de repetições. Aspectos promissores da análise de trilha de

componentes de produção de milho superdoce (ENTRINGER *et al.*, 2014) e de caracteres quantitativos sobre a capacidade de expansão do grão em milho-pipoca (CABRAL *et al.*, 2016) têm sido destacados.

Figura 1 – Matriz ilustrativa da distribuição de frequência (na diagonal) e diagramas de dispersão entre os caracteres número de dias da semeadura até 50% do florescimento masculino (FM), número de dias da semeadura até 50% do florescimento feminino (FF), altura de planta (AP), altura de espiga (AE), número de plantas (NP) e número de espigas (NE) sobre a produtividade de grãos (PROD) (fora da diagonal) de 15 cultivares de milho, avaliadas em nove repetições (n = 135)



Foram realizadas 511 análises de trilha da variável principal PROD em função das variáveis explicativas FM, FF, AP, AE, NP e NE e os resultados foram apresentados na Tabela 1 e nas Figuras 2 e 3. Devido a limitação de espaço para apresentação de todos os resultados

dessas 511 análises de trilha na forma de tabelas, optou-se pelo seguinte: na Tabela 1 foram apresentados os resultados de apenas dez análises de trilha, ou seja, as nove análises de trilha realizadas a partir dos arquivos formados com apenas uma repetição ou bloco, isto é, a primeira análise de trilha foi realizada, somente com os dados do bloco 1 (b1), a segunda análise de trilha foi realizada somente com os dados do bloco 2 (b2) e assim, sucessivamente, até a nona análise de trilha que foi realizada somente com os dados do bloco 9 (b9). Na Tabela 1 também apresentou-se o outro extremo, ou seja, a análise de trilha realizada a partir dos dados das nove repetições ou blocos (arquivo referência). Os resultados dessas mesmas dez análises de trilha foram novamente apresentados nas Figuras 2 e 3, juntamente, com os resultados das demais 501 análises de trilha, sendo 36, 84, 126, 126, 84, 36 e 9 realizadas, respectivamente, a partir dos arquivos com 2, 3, 4, 5, 6, 7 e 8 repetições ou blocos.

Tabela 1 – Coeficientes de correlação (r) e efeitos diretos e indiretos do número de dias da semeadura até 50% do florescimento masculino (FM), número de dias da semeadura até 50% do florescimento feminino (FF), altura de planta (AP), altura de espiga (AE), número de plantas (NP) e número de espigas (NE) sobre a produtividade de grãos (PROD) de 15 cultivares de milho, avaliadas no ano agrícola 2012/2013, em Santa Maria, Rio Grande do Sul. Coeficiente de determinação (R^2), efeito da variável residual, número de condição (NC) e fator de inflação de variância (FIV). Análise de trilha da PROD = função (FM, FF, AP, AE, NP, NE) realizada para cada repetição ou bloco (b) separadamente e em conjunto, ou seja, com as nove repetições (Geral)

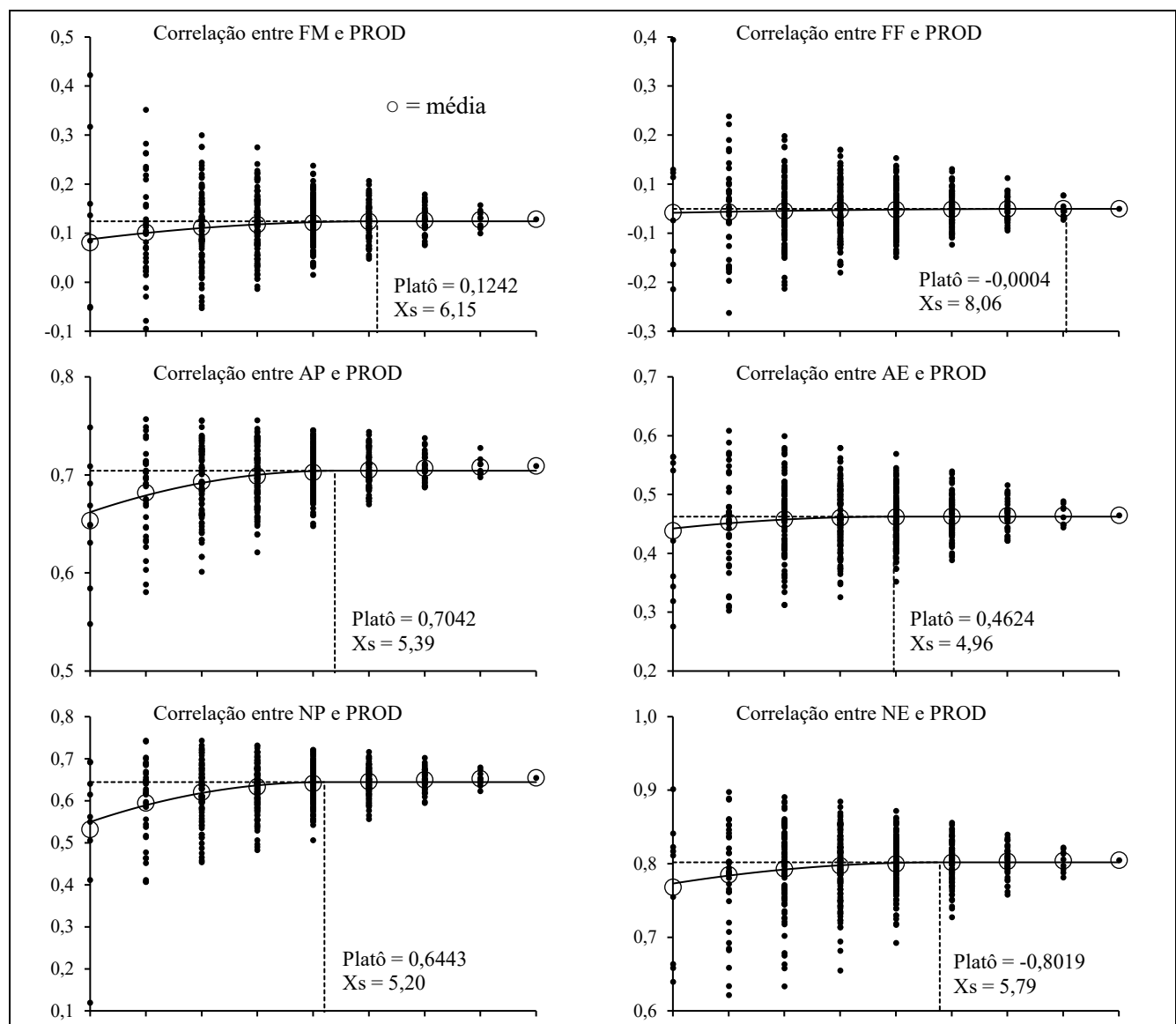
Efeitos sobre a PROD	b1	b2	b3	b4	b5	b6	b7	b8	b9	Geral
Direto de FM	0,040	0,064	0,333	0,147	0,182	0,062	-0,152	0,245	-0,268	-0,058
Indireto de FM via FF	-0,098	-0,530	-0,407	-0,107	-0,121	-0,090	-0,091	-0,278	0,099	-0,174
Indireto de FM via AP	0,147	0,071	0,139	0,083	0,091	0,029	0,069	0,082	0,017	0,204
Indireto de FM via AE	0,058	0,442	0,046	-0,072	0,014	0,026	0,252	0,229	-0,012	0,122
Indireto de FM via NP	-0,017	-0,054	-0,028	0,074	0,000	-0,003	0,017	0,016	0,079	0,032
Indireto de FM via NE	0,030	-0,043	0,002	-0,177	0,151	-0,157	0,042	0,128	-0,077	0,003
r entre FM × PROD	0,161	-0,050	0,085	-0,052	0,317	-0,131	0,137	0,423	-0,162	0,129
Direto de FF	-0,103	-0,774	-0,553	-0,137	-0,176	-0,114	-0,115	-0,308	0,133	-0,201
Indireto de FF via FM	0,038	0,044	0,245	0,114	0,125	0,049	-0,120	0,221	-0,200	-0,050

Continua...

Continuação...										
Indireto de FF via AP	0,125	0,004	0,070	0,119	-0,006	0,015	0,043	0,066	0,119	0,118
Indireto de FF via AE	0,056	0,422	0,044	-0,090	0,013	0,047	0,282	0,220	-0,029	0,130
Indireto de FF via NP	-0,018	0,113	0,027	0,022	0,003	-0,001	0,002	0,015	0,056	0,029
Indireto de FF via NE	-0,034	-0,055	0,003	-0,142	0,115	-0,084	-0,012	0,130	-0,102	-0,025
r entre FF × PROD	0,064	-0,247	-0,164	-0,114	0,074	-0,086	0,080	0,344	-0,024	0,000
Direto de AP	0,323	0,157	0,501	0,551	0,238	0,144	0,178	0,169	0,457	0,485
Indireto de AP via FM	0,018	0,029	0,093	0,022	0,070	0,013	-0,059	0,119	-0,010	-0,025
Indireto de AP via FF	-0,040	-0,019	-0,077	-0,030	0,005	-0,012	-0,028	-0,121	0,034	-0,049
Indireto de AP via AE	0,054	0,552	0,045	-0,119	0,019	0,059	0,213	0,217	-0,020	0,106
Indireto de AP via NP	0,018	-0,144	0,102	-0,079	-0,020	0,015	0,035	0,049	-0,188	-0,171
Indireto de AP via NE	0,257	0,095	0,084	0,203	0,338	0,491	0,245	0,259	0,375	0,362
r entre AP × PROD	0,631*	0,669*	0,749*	0,548*	0,649*	0,709*	0,584*	0,691*	0,649*	0,709*
Direto de AE	0,096	0,900	0,086	-0,180	0,031	0,106	0,390	0,344	-0,040	0,190
Indireto de AE via FM	0,024	0,031	0,179	0,059	0,084	0,016	-0,098	0,163	-0,078	-0,037
Indireto de AE via FF	-0,060	-0,362	-0,288	-0,069	-0,075	-0,051	-0,083	-0,197	0,095	-0,137
Indireto de AE via AP	0,180	0,096	0,264	0,365	0,143	0,080	0,098	0,107	0,228	0,271
Indireto de AE via NP	-0,004	-0,204	0,017	-0,047	-0,008	0,006	0,009	0,027	-0,038	-0,062
Indireto de AE via NE	0,083	0,079	0,018	0,233	0,379	0,408	0,107	0,120	0,177	0,240
r entre AE × PROD	0,319	0,541*	0,276	0,361	0,554*	0,564*	0,421	0,564*	0,344	0,465
Direto de NP	0,059	-0,500	0,295	-0,216	-0,044	0,021	0,062	0,061	-0,300	-0,282
Indireto de NP via FM	-0,012	0,007	-0,032	-0,050	0,001	-0,008	-0,041	0,066	0,071	0,007
Indireto de NP via FF	0,031	0,175	-0,050	0,014	0,012	0,005	-0,003	-0,077	-0,025	0,021
Indireto de NP via AP	0,101	0,045	0,173	0,202	0,107	0,102	0,101	0,135	0,287	0,294
Indireto de NP via AE	-0,007	0,366	0,005	-0,039	0,006	0,029	0,057	0,154	-0,005	0,042
Indireto de NP via NE	0,390	0,027	0,158	0,500	0,425	0,544	0,440	0,303	0,664	0,574
r entre NP × PROD	0,562*	0,120	0,550*	0,412	0,506	0,693*	0,616*	0,641*	0,692*	0,655*
Direto de NE	0,628	0,204	0,188	0,703	0,695	0,730	0,660	0,445	0,873	0,715
Indireto de NE via FM	0,002	-0,013	0,003	-0,037	0,039	-0,013	-0,010	0,070	0,024	0,000
Indireto de NE via FF	0,006	0,208	-0,010	0,028	-0,029	0,013	0,002	-0,090	-0,016	0,007
Indireto de NE via AP	0,132	0,073	0,226	0,159	0,116	0,097	0,066	0,098	0,197	0,245
Indireto de NE via AE	0,013	0,349	0,008	-0,060	0,017	0,059	0,063	0,093	-0,008	0,064
Indireto de NE via NP	0,037	-0,066	0,249	-0,154	-0,027	0,016	0,041	0,042	-0,228	-0,226
r entre NE × PROD	0,817*	0,755*	0,664*	0,640*	0,812*	0,902*	0,823*	0,658*	0,841*	0,805*
R ²	0,781	0,874	0,805	0,606	0,759	0,837	0,820	0,641	0,850	0,816
Variável residual	0,219	0,126	0,195	0,394	0,241	0,163	0,180	0,359	0,150	0,184
NC	69,609	38,920	24,297	21,424	24,108	42,758	20,361	43,972	35,170	53,168
FIV	13,206	6,352	4,416	3,665	3,264	5,039	4,019	6,416	7,579	7,028

* Significativo a 5% de probabilidade de erro pelo teste t, com 13 graus de liberdade.

Figura 2 – Estimativas do platô e do número de repetições para a análise de trilha (X_s), obtidas a partir do modelo quadrático de resposta com platô de Y [coeficientes de correlação linear de Pearson entre os caracteres número de dias da semeadura até 50% do florescimento masculino (FM), número de dias da semeadura até 50% do florescimento feminino (FF), altura de planta (AP), altura de espiga (AE), número de plantas (NP) e número de espigas (NE) versus produtividade de grãos (PROD) de 15 cultivares de milho, avaliadas em nove repetições, no ano agrícola 2012/2013, em Santa Maria, Rio Grande do Sul] em função da variável X (número de repetições). Estimativas do número de condição e fator de inflação de variância



Continua...

Continuação

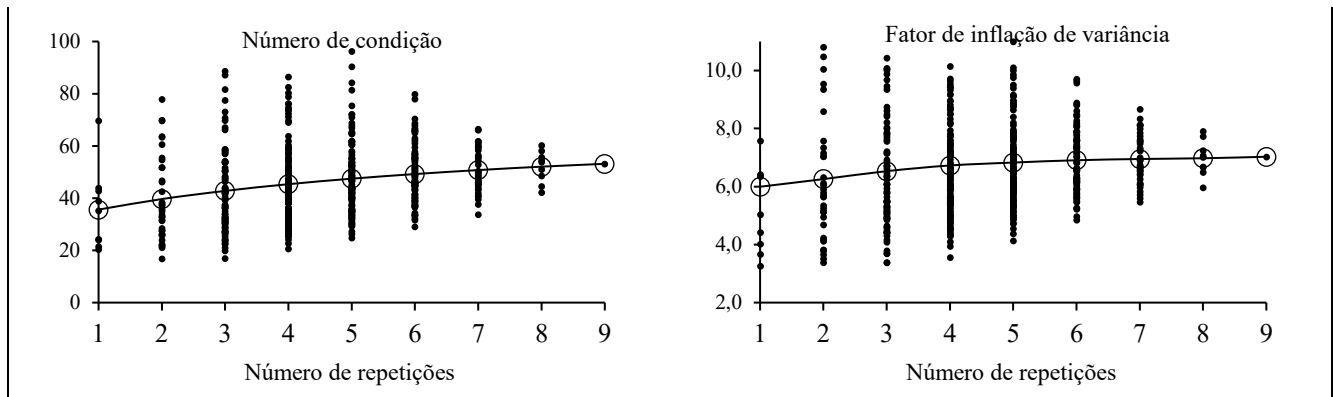
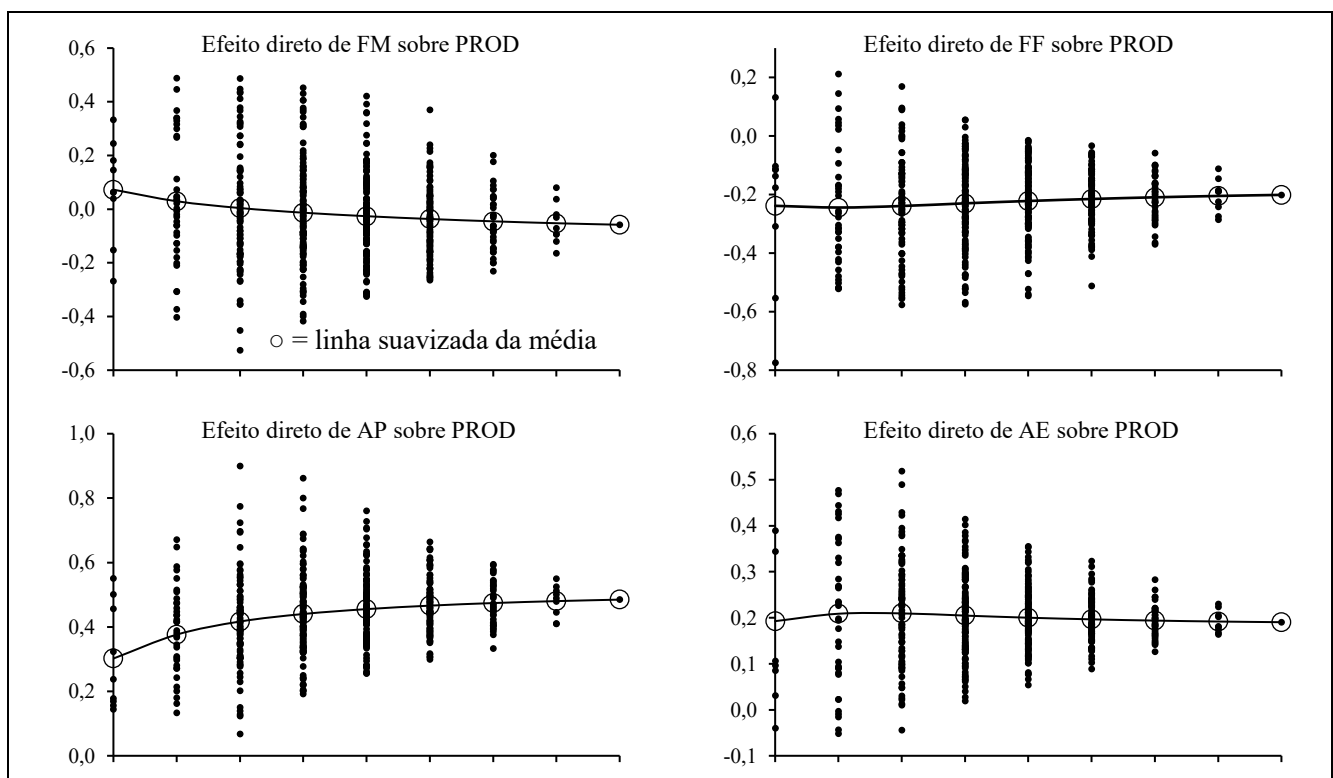
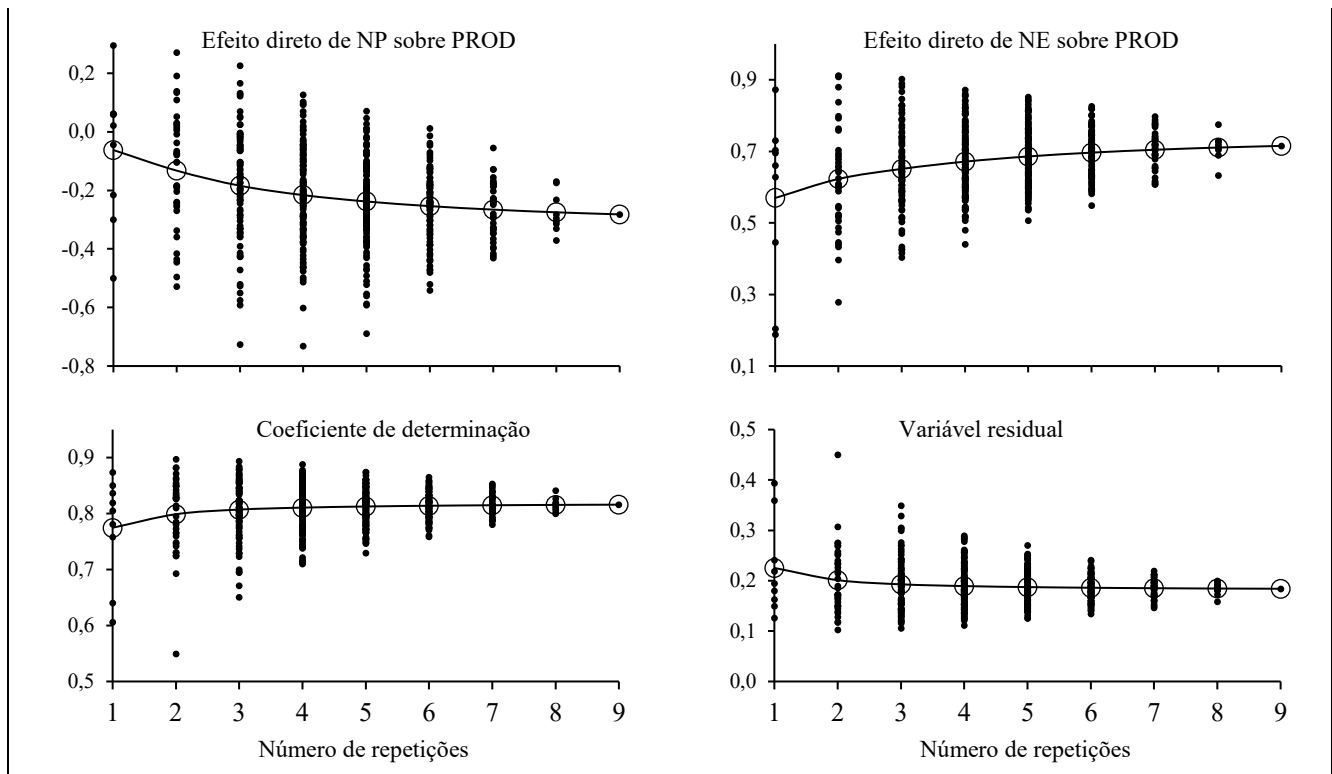


Figura 3 - Diagramas de dispersão das estimativas dos efeitos diretos dos caracteres número de dias da semeadura até 50% do florescimento masculino (FM), número de dias da semeadura até 50% do florescimento feminino (FF), altura de planta (AP), altura de espiga (AE), número de plantas (NP) e número de espigas (NE) sobre a produtividade de grãos (PROD) de 15 cultivares de milho, avaliadas no ano agrícola 2012/2013, em Santa Maria, Rio Grande do Sul, do coeficiente de determinação e da variável residual, em função do número de repetições



Continua...

Continuação...



Os coeficientes de correlação linear de Pearson (r), obtidos a partir de análise com apenas uma repetição, oscilaram de -0,162 a 0,423 entre FM e PROD; -0,247 a 0,344 entre FF e PROD; 0,548 a 0,749 entre AP e PROD; 0,276 a 0,564 entre AE e PROD; 0,120 a 0,693 entre NP e PROD; e 0,640 a 0,902 entre NE e PROD (Tabela 1). A variação entre as nove repetições (nove blocos), para cada par de caracteres, sugere diferenças entre as repetições em relação ao coeficiente de correlação, o que leva a distintos resultados da análise de trilha, quando a mesma for realizada com apenas uma repetição. Portanto, o número suficiente de repetições para a análise de trilha mereceu ser investigado, a fim de obter resultados com confiabilidade.

Entre as nove análises de trilha, realizadas com apenas uma repetição ou bloco, ou seja, a primeira análise somente com os dados do bloco 1 (b1), a segunda análise somente com os dados do bloco 2 (b2) e assim, sucessivamente, até a nona análise com somente os dados do bloco 9 (b9), em sete (b1, b4, b5, b6, b7, b8 e b9) os coeficientes de correlação linear de Pearson (r) entre a variável explicativa número de espigas (NE) e a variável principal produtividade de grãos (PROD) foram positivos ($0,640 \leq r \leq 0,902$) com efeito direto de mesmo sinal e magnitude similar sobre PROD ($0,445 \leq \text{efeito direto} \leq 0,873$), o que confirma relação de causa e efeito do NE

sobre a PROD. No entanto, em duas situações (b2 e b3) os valores altos de r (0,755 e 0,664) e baixos dos efeitos diretos (0,204 e 0,188) demonstram que a correlação entre NE e PROD é devida aos efeitos indiretos das outras variáveis (Tabela 1). Portanto, esses resultados comprovam que análises de trilha a partir de dados de apenas uma repetição (um bloco), levam a resultados distintos. Raciocínio semelhante pode ser aplicado para visualizar os distintos resultados das análises de trilha das variáveis explicativas FM, FF, AP, AE e NP sobre a PROD, a partir dos dados de apenas uma repetição.

As análises de correlação e de trilha, realizadas com os dados das nove repetições ou blocos (geral), ou seja, com os dados dos nove blocos juntos, revelaram que houve associação linear positiva entre NE e PROD ($r = 0,805$, $p \leq 0,05$) e o efeito direto foi de mesmo sinal e magnitude semelhante (0,715). Além disso, os efeitos indiretos de NE via FM (0,000), FF (0,007); AP (0,245); AE (0,064) e NP (-0,226) sobre PROD foram desprezíveis (Tabela 1). Portanto, pode-se inferir que há associação linear de causa e efeito entre NE e PROD e que as plantas com maior número de espigas apresentam maior produtividade de grãos. Em magnitude levemente inferior, mas não menos importante, constatou-se também que houve associação linear de causa e efeito entre AP e PROD ($r = 0,709$, $p \leq 0,05$; e efeito direto de AP sobre PROD de 0,485).

Observou-se associação linear positiva entre NP e PROD ($r = 0,655$, $p \leq 0,05$), mas o efeito direto de NP sobre PROD (-0,282) foi de magnitude inferior e de sinal contrário, sendo a correlação ($r = 0,655$) explicada, principalmente, pelo elevado efeito indireto de NP via NE (0,574). Constatou-se também correlação moderada entre AE e PROD ($r = 0,465$), embora não significativa ($p > 0,05$), mas com efeito direto de mesmo sinal e magnitude um pouco inferior (0,190). Assim, de maneira geral, as variáveis NE, AP, NP, AE estão associadas positivamente com PROD, sendo o NE e AP, nessa ordem, as variáveis mais fortemente correlacionadas com a PROD. Por outro lado, os coeficientes de correlação entre FM e PROD ($r = 0,129$) e FF e PROD ($r = 0,000$) foram de baixa magnitude e não significativos ($p > 0,05$) e os respectivos efeitos diretos de FM (-0,058) e FF (-0,201) sobre PROD foram desprezíveis e de sinal contrário, o que demonstra inexistência de associação linear de causa e efeito das variáveis FM e FF com a PROD (Tabela 1).

Diante desses resultados, pode-se inferir que as variáveis NP e AE, apresentaram correlações positivas moderadas com a PROD. Já o NE e AP apresentaram forte relacionamento linear, o que significa, respectivamente, que as plantas com maior número de espigas

apresentam maior produtividade de grãos e que as plantas mais altas são mais produtivas. Em híbridos de milho de ciclo precoce e tardio com genes de tolerância a seca, Mason *et al.* (2019) verificaram que o número de espigas por metro de fileira foi a única variável que apresentou correlação significativa ($0,34 \leq r \leq 0,55$) e efeito direto positivo ($0,54 \leq \text{efeito direto} \leq 0,55$) sobre a produtividade de grãos nas duas situações de interesse (híbridos de milho de ciclo precoce e tardio com genes de tolerância a seca). Na avaliação de híbridos de milho para silagem, Crevelari *et al.* (2018) verificaram correlação genotípica positiva, de fraca intensidade e não significativa, do número de espigas com a produtividade de grãos ($r = 0,18$) e com a produção de massa fresca ($r = 0,34$) na maturidade de silagem. Os autores verificaram que, nesse caso, o número total de espigas não apresentou relação de causa e efeito com a produção de massa fresca. Utilizando dados de 14 experimentos de híbridos de milho, Toebe e Cargnelutti Filho (2013 a) verificaram correlação geral de $r = 0,68$ ($0,10 \leq r \leq 0,89$) entre o número de espigas e a produtividade de grãos. Na análise de trilha em crista, os autores constataram que o número de espigas foi a única variável com efeito direto positivo sobre a produtividade de grãos em todos os experimentos ($0,01 \leq \text{efeito direto} \leq 0,51$). Quando esta variável foi excluída do estudo, verificou-se que a prolificidade passou a exercer alto efeito direto sobre a produtividade de grãos (efeito direto médio de 0,53). Em outro estudo com 13 experimentos, Toebe e Cargnelutti Filho (2013 b) verificaram alta correlação entre o número de espigas e a produtividade de grãos ($0,49 \leq r \leq 0,92$) e efeito direto positivo ($0,12 \leq \text{efeito direto} \leq 0,48$). Diferentemente do presente estudo, nos experimentos analisados por Toebe e Cargnelutti Filho (2013 a, b) não foram evidenciadas relações de causa e efeito entre a altura de plantas e a produtividade de grãos.

Os efeitos diretos e indiretos, estimados por meio da análise de trilha, dependem dos coeficientes de correlação linear de Pearson (r). Portanto, é importante investigar o número de repetições a partir do qual o valor de r estabiliza e, como consequência, a inclusão de dados de mais repetições não altera os resultados da análise de trilha. Por meio do modelo quadrático de resposta com platô das variáveis Y (r entre FM e PROD; FF e PROD; AP e PROD; AE e PROD; NP e PROD; e NE e PROD) em função da variável X (número de repetições), foi constatado que a partir de 6,15; 8,06; 5,39; 4,96; 5,20; e 5,79 repetições o valor de r estabilizou (platô) em 0,1242; -0,0004; 0,7042; 0,4624; 0,6443; e 0,8019, respectivamente (Figura 2). Assim, a média dos seis valores de

Xs (5,92), arredondada para o inteiro superior (seis repetições), foi considerada como sendo o número suficiente de repetições para a análise de trilha.

Experimentos com seis repetições tem sido incentivados por Resende e Duarte (2007) e Cargnelutti Filho *et al.* (2018) e o aumento do número de repetições tem evidenciado melhoria na precisão experimental (CARGNELUTTI FILHO *et al.*, 2010; NESI *et al.*, 2010; CARGNELUTTI FILHO; GUADAGNIN, 2011; MENDOZA; BUITRAGO, 2015). Nesse sentido, com base em 286 experimentos, Cargnelutti Filho *et al.* (2010) verificaram que a utilização de 5,35 repetições garantiria à predição da produtividade de grãos de cultivares de milho, com 85% de exatidão no prognóstico do valor real. Caso fossem utilizadas seis repetições, essa predição chegaria ao valor de 86,4%. Ao avaliar a interferência do número de repetições sobre estatísticas de precisão experimental em milho, Cargnelutti Filho *et al.* (2018) verificaram que a precisão experimental aumenta com o acréscimo do número de repetições, mas com redução gradativa do percentual de ganho sobre esta precisão. Nesse sentido, os autores destacam que o uso de seis repetições pode ser considerado como referência para experimentos com cultivares de milho. Por outro lado, Nesi *et al.* (2010) verificaram a necessidade de uso de 11 repetições para a diferenciação de cultivares de milho em relação a produtividade de grãos. Já em milho pipoca, Catapatti *et al.* (2008) verificaram que o número de diferenças significativas entre médias de genótipos aumentou quando foram utilizadas mais de quatro repetições, sendo máxima em seis repetições (maior número de repetições do estudo). Estudos indicando o número de repetições para a análise de trilha em milho não foram encontrados na literatura. Contudo, na avaliação dos coeficientes de correlação e dos efeitos diretos e indiretos da análise de trilha de componentes produtivos sobre a produção de silagem em híbridos de milho, Crevelari *et al.* (2018) conduziram experimentos em dois locais no delineamento de blocos ao acaso com seis repetições. Esse número de repetições está em concordância com os resultados do presente trabalho.

As amplitudes de variação das estimativas dos coeficientes de correlação linear de Pearson, dos efeitos diretos das variáveis FM, FF, AP, AE, NP e NE sobre a PROD, do coeficiente de determinação, da variável residual, do número de condição e do fator de inflação de variância reduziram com o acréscimo do número de repetições (Figuras 2 e 3). Padrão semelhante foi observado para os efeitos indiretos (dados não apresentados). Então, mesmo que essas medidas possam ser realizadas com apenas uma repetição, pode-se inferir que há aumento da

confiabilidade dos resultados com o acréscimo do número de repetições e que os resultados da análise de trilha a partir de seis, sete, oito ou nove repetições são similares.

Seis repetições servem como referência para o planejamento de experimentos para avaliação de cultivares de milho. Sugere-se, mais estudos com base em cenários formados pela combinação de números de indivíduos (cultivares), números de variáveis e número de repetições, na cultura do milho e em outras culturas. É importante considerar que poucas repetições podem levar a menor confiabilidade nas inferências e muitas repetições podem ser inviáveis em relação à execução do experimento.

4 CONCLUSÕES

Com o aumento do número de repetições há melhoria na precisão das estimativas dos coeficientes da análise de trilha, porém os ganhos em precisão diminuem, gradativamente.

Seis repetições são suficientes para realizar a análise de trilha de caracteres agronômicos de cultivares de milho e pode ser adotado como referência para o planejamento de experimentos futuros.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Processo 304652/2017-2) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelas bolsas concedidas; e aos alunos bolsistas e voluntários, pelo auxílio na coleta de dados.

REFERÊNCIAS

BANZATTO DA, KRONKA SN. **Experimentação agrícola**. 4 ed. Jaboticabal: FUNEP; 2013.

CABRAL PDS, AMARAL JÚNIOR AT, FREITAS ILJ, RIBEIRO RM, SILVA TRC. Relação causa e efeito de caracteres quantitativos sobre a capacidade de expansão do grão em milho-pipoca. **Rev. Ciênc. Agron.** 2016;47(1):108-117.

CARGNELUTTI FILHO A, TOEBE M, ALVES BM, KLEINPAUL JA, NEU IMM. Number of replicates and experimental precision statistics in corn. **Pesq. agropec. bras.** 2018;53(11):1213-1221.

CARGNELUTTI FILHO A, GUADAGNIN JP. Planejamento experimental em milho. **Rev. Ciênc. Agron.** 2011;42(4):1009-1016.

CARGNELUTTI FILHO A, STORCK L, GUADAGNIN JP. Número de repetições para a comparação de cultivares de milho. **Cienc. Rural.** 2010;40(5):1023-1030.

CATAPATTI TR, GONÇALVES MC, SILVA NETO MR, SOBROZA R. Tamanho de amostra e número de repetições para avaliação de caracteres agronômicos em milho-pipoca. **Ciênc. agrotec.** 2008;32(3):855-862.

CREVELARI JA, DURÃES NNL, BENDIA LCR, VETTORAZZI JCF, ENTRINGER GC, FERREIRA JÚNIOR JA, *et al.* Correlations between agronomic traits and path analysis for silage production in maize hybrids. **Bragantia.** 2018;77(2):243-252.

CRUZ CD. Genes Software - extended and integrated with the R, Matlab and Selegen. **Acta Sci. Agron.** 2016;38(4):547-552.

CRUZ CD, CARNEIRO PCS, REGAZZI AJ. **Modelos biométricos aplicados ao melhoramento genético:** volume 2. 3 ed. Viçosa: UFV; 2014.

CRUZ CD, REGAZZI AJ, CARNEIRO PCS. **Modelos biométricos aplicados ao melhoramento genético:** volume 1. 4 ed. Viçosa: UFV; 2012.

ENTRINGER GC, SANTOS PHAD, VETTORAZZI JCF, CUNHA KS, PEREIRA MG. Correlação e análise de trilha para componentes de produção de milho superdoce. **Ceres.** 2014;61(3):356-361.

GONÇALVES RP, CHAVES LM, SAVIAN TV, SILVA FF, PAIXÃO CA. Ajuste de modelos de platô de resposta via regressão isotônica. **Cienc. Rural.** 2012;42(2):354-359.

HAIR JF, BLACK WC, BABI BJ, ANDERSON RE, TATHAM RL. **Análise multivariada de dados.** Porto Alegre: Bookman; 2009.

MASON S, KMAIL Z, GALUSHA T, JUKIĆ Ž. Path analysis of drought tolerant maize hybrid yield and yield components across planting dates. **J. Cent. Eur. Agric.** 2019;20(1):194-207.

MENDOZA RG, BUITRAGO IC. Selección de estadísticos para la estimación de la precisión experimental en ensayos de maíz. **Agron. Mesoam.** 2015;26(1):55-63.

MONTGOMERY DC, PECK EA. **Introduction to linear regression analysis.** New York: John Wiley & Sons; 1982.

NESI CN, BÓ HCD, GUIDONI AL, BRINGHENTI C. Número mínimo de repetições em experimentos de competição de híbridos de milho. **Rev. Ciênc. Agrovet.** 2010;9(1):74-81.

R Core Team. **R: A language and environment for statistical computing.** Vienna, Áustria: R Foundation for Statistical Computing, 2019.

RESENDE MDV, DUARTE JB. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesq. Agropec. Trop.** 2007;37(3):182-194.

STORCK L, GARCIA DC, LOPES SJ, ESTEFANEL V. **Experimentação vegetal**. Santa Maria: UFSM; 2016.

TOEBE M, CARGNELUTTI FILHO A. Multicollinearity in path analysis of maize (*Zea mays* L.). **J Cereal Sci.** 2013a;57:453-462.

TOEBE M, CARGNELUTTI FILHO A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesq. agropec. bras.** 2013b;48(5):466-477.

CONTRIBUIÇÕES DE AUTORIA

1 – Alberto Cargnelutti Filho

Doutor em Agronomia, Universidade Federal de Santa Maria

<https://orcid.org/0000-0002-8608-9960> – alberto.cargnelutti.filho@gmail.com

Contribuição: concepção, investigação, metodologia e análises estatísticas, redação da versão inicial, revisão e edição, supervisão.

2 – Marcos Toebe

Doutor em Agronomia, Universidade Federal de Santa Maria

<https://orcid.org/0000-0003-2033-1467> – m.toebe@gmail.com

Contribuição: investigação, revisão e edição.

3 – Bruna Mendonça Alves

Doutora em Agronomia, Universidade Federal de Santa Maria

<https://orcid.org/0000-0002-9741-9021> – brunamalves11@gmail.com

Contribuição: investigação, revisão e edição.

4 – Ismael Mario Marcio Neu

Discente de doutorado em Agronomia, Universidade Federal de Santa Maria

<https://orcid.org/0000-0002-9186-2532> – ismaelmmneu@hotmail.com

Contribuição: investigação, revisão e edição.

5 – Jéssica Andiar Kleinpaul

Discente de doutorado em Agronomia, Universidade Federal de Santa Maria

<https://orcid.org/0000-0001-7550-6012> – kleinpauljessica@gmail.com

Contribuição: investigação, revisão e edição.

COMO CITAR ESTE ARTIGO

CARGNELUTTI FILHO, A; TOEBE, M; ALVES, B. M.; NEU, I. M. M.; KLEINPAUL, J. A. Número suficiente de repetições para análise de trilha em milho. **Ciência e Natura**, Santa Maria, v. 43, e32, p. 1-18, 2021. Disponível em: <https://doi.org/10.5902/2179460X40936>. Acesso em: dia, mês (abreviado), ano.