

# Dimensionamento amostral para estimação de coeficientes de correlação em híbridos de milho, safras e níveis de precisão

Marcos Toebe (1); Alberto Cargnelutti Filho (1\*); Sidinei José Lopes (1); Cláudia Burin (1); Tatiani Reis da Silveira (1); Gabriele Casarotto (2)

(1) Universidade Federal de Santa Maria (UFSM), Departamento de Fitotecnia, Avenida Roraima, 1.000, 97105-900 Santa Maria (RS), Brasil.

(2) Universidade Federal do Rio Grande do Sul (UFRGS), Departamento de Agronomia, Avenida Paulo Gama, 110, 90040-060 Porto Alegre (RS), Brasil.

(\* Autor correspondente: alberto.cargnelutti.filho@gmail.com)

Recebido: 19/set./2014; Aceito: 8/out./2014

## Resumo

O objetivo deste trabalho foi determinar o tamanho de amostra necessário para a estimação de coeficientes de correlação linear de Pearson para híbridos simples, triplo e duplo de milho em safras e níveis de precisão. Em 361, 373 e 416 plantas, respectivamente, dos híbridos simples, triplo e duplo da safra 2008/2009 e em 1.777, 1.693 e 1.720 plantas, respectivamente, dos híbridos simples, triplo e duplo da safra 2009/2010 foram mensurados 12 caracteres: altura de planta; altura de inserção e peso de espiga; número de fileiras de grãos por espiga; comprimento e diâmetro de espiga; peso e diâmetro de sabugo; massa de 100 grãos; número de grãos por espiga; comprimento e diâmetro de grãos; e produtividade de grãos. Em cada híbrido e safra foram estimados os coeficientes de correlação para os 66 pares de caracteres e determinou-se o tamanho de amostra para estimação de coeficientes de correlação em quatro níveis de precisão [amplitudes do intervalo de confiança de 95% ( $AI_{95\%}$ ) de 0,15, 0,25, 0,35 e 0,45] por meio de reamostragem com reposição. O tamanho de amostra varia entre híbridos, safras e pares de caracteres. Maior tamanho de amostra é necessário para a estimação do coeficiente de correlação entre caracteres fracamente correlacionados e menor tamanho de amostra é necessário para a estimação do coeficiente de correlação entre caracteres altamente correlacionados. Independentemente do híbrido, da safra e do par de caracteres, 375, 195 e 120 plantas são suficientes, respectivamente, para a estimação de coeficientes de correlação com  $AI_{95\%}$  máximas de 0,25, 0,35 e 0,45.

Palavras-chave: *Zea mays* L, reamostragem, planejamento experimental, relações lineares.

## Sample size in the estimation of correlation coefficients for corn hybrids in crops and accuracy levels

### Abstract

This study determined the sample size necessary for the estimation of the Pearson linear correlation coefficients for single, triple and double corn hybrids in crops and accuracy levels. In 361, 373 and 416 plants, respectively, of the single, triple and double hybrids of the 2008/2009 crop and, in 1,777, 1,693 and 1,720 plants, respectively, of the single, triple and double hybrids of the 2009/2010 crop, twelve traits were measured: plant height, ear insertion height, ear weight, number of grain rows per ear, ear length and diameter, cob weight and diameter, weight of hundred grains, number of grains per ear, grain length and grain yield. Then, in each hybrid and crop, were estimated the correlation coefficients for the 66 pairs of traits and determined the sample size necessary to estimate the correlation coefficients in four accuracy levels [amplitudes of the confidence interval of 95% ( $AI_{95\%}$ ) of 0.15, 0.25, 0.35 and 0.45], by resampling with replacement. The sample size varies among hybrids, crops and pairs of traits. Larger sample size is required to estimate the correlation coefficient between weakly correlated traits and smaller sample size is needed to estimate the correlation coefficient between highly correlated traits. Independently of hybrid, crop and pairs of traits, 375, 195 and 120 plants are sufficient, respectively, to estimate the correlation coefficients with maximum  $AI_{95\%}$  of 0.25, 0.35 and 0.45.

Key words: *Zea mays* L, resampling, experimental design, linear relationships.

## 1. INTRODUÇÃO

O milho é o cereal com maior volume de produção mundial, com produção estimada em 906,82 milhões de toneladas para a safra de 2014/2015, em área de 160,2 milhões de hectares, sendo o Brasil o terceiro maior produtor mundial (FAO, 2014). Dentre as finalidades de uso do milho, destaca-se sua utilização na alimentação humana e animal e como matéria-prima industrial, principalmente devido à quantidade de reservas acumuladas nos grãos (Fancelli & Dourado Neto, 2004). Nesse sentido, o aumento na produtividade de grãos de milho verificado nas últimas décadas é atribuído igualmente às melhorias de manejo e ao melhoramento genético (Duvick, 2005).

No melhoramento genético, a seleção de plantas pode ser realizada de modo direto ou de modo indireto, via estudo de relações lineares entre caracteres. Para o estudo de relações lineares entre caracteres, pode-se utilizar o coeficiente de correlação linear de Pearson ( $r$ ), que mede o sentido e a intensidade da relação linear entre duas variáveis aleatórias (Ferreira, 2009). O sentido da correlação pode ser positivo ou negativo, no intervalo de  $-1 \leq r \leq 1$ , sendo que a intensidade da correlação linear será maior quanto mais próximo a  $|1|$  for o  $r$ . Estudos complementares aos coeficientes de correlação, como análises de trilha e correlações canônicas, também têm sido recomendados para a seleção indireta de plantas (Cruz & Regazzi, 1997).

Para que os resultados gerados em estudos de relações lineares apresentem confiabilidade, é necessário o correto dimensionamento do tamanho de amostra (número de plantas) a ser usado para a estimação dos coeficientes de correlação. Esses coeficientes poderão ser interpretados isoladamente ou utilizados em análises complementares como, por exemplo, nas análises de trilha e correlações canônicas e, por isso, devem ser estimados com precisão. Nesse sentido, Cargnelutti Filho et al. (2010) realizaram o dimensionamento amostral para a estimação de coeficientes de correlação linear de Pearson entre caracteres de híbridos simples, triplo e duplo de milho, com base em dados de uma safra e para um nível de precisão. Os autores verificaram que para a estimação de 91 pares de caracteres, com amplitude do intervalo de confiança de 95% ( $AIC_{95\%}$ ) máxima de 0,30 foram necessárias até 300 plantas, dependendo do híbrido e do par de caracteres. Estudos de dimensionamento amostral para a estimação de coeficientes de correlação de Pearson também foram realizados em crambe (Cargnelutti Filho et al., 2011) e em mamoneira (Cargnelutti Filho et al., 2012). Ainda, Shieh (2010) avaliou as propriedades e o efeito do dimensionamento amostral sobre o coeficiente de correlação linear de Pearson, e Bonett & Wright (2000) realizaram estudos de dimensionamento amostral para a estimação de coeficientes de correlação de Pearson, de Kendall e de Spearman.

Conforme destacado acima, estudos de dimensionamento amostral para a estimação de coeficientes de correlação já foram realizados, inclusive na cultura do milho (Cargnelutti Filho et al., 2010). No entanto, destaca-se que o dimensionamento amostral, considerando híbridos, safras e níveis de precisão, é importante para possibilitar a estimação precisa dos coeficientes de correlação, amplamente utilizados em estudos de relações lineares. Assim, o objetivo deste trabalho foi determinar o tamanho de amostra necessário para a estimação de coeficientes de correlação linear de Pearson para híbridos simples, triplo e duplo de milho, em safras e níveis de precisão.

## 2. MATERIAL E MÉTODO

Foram conduzidos dois experimentos com a cultura de milho (*Zea mays* L.), nas safras agrícolas de 2008/2009 (primeiro experimento) e de 2009/2010 (segundo experimento), em área experimental localizada no município de Santa Maria, estado do Rio Grande do Sul (29°42'S, 53°49'W, a 95 m de altitude). No primeiro experimento, semeado em 26/12/2008, foram conduzidas quatro parcelas com o híbrido simples P32R21, quatro com o híbrido triplo DKB566 e quatro com o híbrido duplo DKB747. No segundo experimento, semeado em 26/10/2009, foram conduzidas 16 parcelas com o híbrido simples 30F53, 16 com o híbrido triplo DKB566 e 16 com o híbrido duplo DKB747.

Cada parcela foi composta de quatro fileiras de 6 m de comprimento, espaçadas em 0,80 m, sendo a densidade ajustada para cinco plantas por metro linear, representando uma densidade de semeadura de 62.500 plantas ha<sup>-1</sup>. Dessa forma, cada parcela foi composta por 120 plantas, totalizando 1.440 plantas no primeiro experimento (3 híbridos × 4 parcelas/híbrido × 120 plantas/parcela) e 5.760 plantas no segundo experimento (3 híbridos × 16 parcelas/híbrido × 120 plantas/parcela). Em cada safra, as parcelas dos híbridos simples, triplo e duplo foram casualizadas na área experimental. Nos dois experimentos, a adubação de base foi de 750 kg ha<sup>-1</sup> da fórmula 3-24-18 (NPK) e a adubação de cobertura foi de 300 kg ha<sup>-1</sup> de ureia com 45% de N. Os demais tratamentos culturais foram realizados de acordo com as recomendações para a cultura do milho (Fancelli & Dourado Neto, 2004).

No primeiro experimento, foram avaliadas 361, 373 e 416 plantas, respectivamente, dos híbridos simples, triplo e duplo e, no segundo experimento, 1.777, 1.693 e 1.720 plantas, respectivamente. Foram avaliadas apenas plantas que continham os 12 caracteres descritos a seguir. Em função disso, o número final de plantas avaliadas, em cada safra, diferiu entre os híbridos simples, triplo e duplo. Em cada uma das 6.340 plantas foram mensurados os seguintes caracteres: altura de planta na colheita (AP); altura de inserção de espiga (AIE); peso de espiga sem palha (PE);

número de fileiras de grãos por espiga (NF); comprimento de espiga (CE); diâmetro de espiga (DE); peso de sabugo (PS); diâmetro de sabugo (DS); massa de 100 grãos (MCG); número de grãos por espiga (NGR); comprimento de grãos (CGR), obtido pela diferença entre os diâmetros de espiga e de sabugo dividido por dois; e a produtividade de grãos (PROD), em gramas por planta. A seguir, para cada híbrido em cada experimento foram calculados os coeficientes de correlação linear de Pearson ( $r$ ) para cada um dos 66 pares de caracteres, sendo a significância do  $r$  verificada por meio do teste  $t$  de *Student* em 5% de significância.

Para cada híbrido, em cada experimento, foram planejados 199 tamanhos de amostra, sendo o tamanho de amostra inicial de dez plantas e os demais obtidos com o incremento de cinco plantas. Dessa forma, os tamanhos de amostra planejados foram de  $n = 10, 15, 20, \dots, 1.000$  plantas. Para cada tamanho de amostra planejado, foram obtidas 1.000 reamostras com reposição (Ferreira, 2009), mesmo número de reamostras utilizadas em estudos anteriores de dimensionamento amostral para estimação do coeficiente de correlação linear de Pearson (Cargnelutti Filho et al., 2010, 2012), sendo que em cada reamostra foram estimados os coeficientes de correlação linear de Pearson ( $r$ ) dos 66 pares de caracteres. Assim, para cada tamanho de amostra planejado foram obtidas 1.000 estimativas de  $r$  para cada um dos 66 pares de caracteres. Com base nessas 1.000 estimativas, foram determinados: o percentil 2,5%, a média, o percentil 97,5% e foi calculada a amplitude do intervalo de confiança de 95% ( $AIC_{95\%}$ ), pela diferença entre o percentil 97,5% e o percentil 2,5%.

Para a determinação do tamanho de amostra (número de plantas) necessário para a estimação do  $r$ , de cada um dos 66 pares de caracteres em cada híbrido e experimento, foram fixados inicialmente limites máximos de  $AIC_{95\%}$  do  $r$  iguais a 0,15 (maior precisão), 0,25, 0,35 e 0,45 (menor precisão). A seguir, partiu-se do tamanho de amostra inicial ( $n = 10$  plantas) e considerou-se como tamanho de amostra adequado ( $n$ ) o número de plantas a partir do qual a  $AIC_{95\%}$  do  $r$  foi menor ou igual ao limite máximo estabelecido para cada nível de precisão (0,15, 0,25, 0,35 ou 0,45). Os coeficientes de correlação, obtidos com os dados do primeiro experimento, e o tamanho de amostra para a  $AIC_{95\%}$  de 0,30 (nível de precisão intermediário entre 0,15 e 0,45) foram apresentados por Cargnelutti Filho et al. (2010). As análises estatísticas foram realizadas com o auxílio do programa R (R Development Core Team, 2014) e do aplicativo Microsoft Office Excel®.

### 3. RESULTADOS E DISCUSSÃO

Os coeficientes de correlação linear de Pearson ( $r$ ) apresentaram elevada variabilidade entre os 66 pares de caracteres avaliados (Tabela 1). Nesse sentido, para os híbridos

conduzidos na safra 2008/2009, o  $r$  variou de 0,02 (correlação entre DS e CGR) a 1,00 (correlação entre PE e PROD) para o híbrido simples P32R21, de -0,20 (correlação entre NF e MCG) a 1,00 (correlação entre PE e PROD) para o híbrido triplo DKB566 e de -0,16 (correlação entre NF e MCG) a 0,99 (correlação entre PE e PROD) para o híbrido duplo DKB747. Na safra 2009/2010, o coeficiente de correlação variou de -0,20 (correlação entre AIE e CE) a 1,00 (correlação entre PE e PROD) para o híbrido simples P32R21, de -0,12 (correlação entre AIE e CE) a 1,00 (correlação entre PE e PROD) para o híbrido triplo DKB566 e de -0,06 (correlação entre NF e MCG) a 0,99 (correlação entre PE e PROD) para o híbrido duplo DKB747. Variabilidade de intensidade e sentido dos coeficientes de correlação também foram verificados em 91 pares de caracteres de híbridos simples, triplo e duplo de milho avaliados em uma safra (Cargnelutti Filho et al., 2010), em 210 pares de caracteres de crame (Cargnelutti Filho et al., 2011) e em 210 pares de caracteres de híbridos de mamoneira (Cargnelutti Filho et al., 2012).

De maneira geral, verificou-se que as correlações mais fracas entre pares de caracteres apresentaram maiores amplitudes entre híbridos e safras, como pode ser constatado para as correlações entre AIE e PE ( $-0,12 \leq r \leq 0,37$ ), AIE e CE ( $-0,20 \leq r \leq 0,30$ ), AIE e MCG ( $-0,14 \leq r \leq 0,29$ ), AIE e NGR ( $-0,09 \leq r \leq 0,34$ ), AIE e PROD ( $-0,13 \leq r \leq 0,38$ ), NF e MCG ( $-0,20 \leq r \leq 0,11$ ) e entre DS e CGR ( $-0,06 \leq r \leq 0,31$ ) (Tabela 1). Por outro lado, correlações de maior intensidade entre pares de caracteres apresentaram menores oscilações entre híbridos e safras, conforme pode ser constatado, por exemplo, para as correlações entre DE e NGR ( $0,75 \leq r \leq 0,85$ ), PE e DE ( $0,81 \leq r \leq 0,86$ ), DE e PROD ( $0,81 \leq r \leq 0,86$ ), PE e NGR ( $0,90 \leq r \leq 0,95$ ), NGR e PROD ( $0,91 \leq r \leq 0,95$ ) e entre PE e PROD ( $0,99 \leq r \leq 1,00$ ).

O tamanho de amostra necessário para a estimação do  $r$ , com  $AIC_{95\%}$  menor ou igual a 0,15, apresentou elevada variabilidade entre os 66 pares de caracteres mensurados nos híbridos simples (10 plantas  $\leq n \leq 890$  plantas), triplo (10 plantas  $\leq n \leq 990$  plantas) e duplo (10 plantas  $\leq n \leq 800$  plantas) da safra 2008/2009 e também nos híbridos simples (10 plantas  $\leq n > 1.000$  plantas), triplo (10 plantas  $\leq n \leq 825$  plantas) e duplo (10 plantas  $\leq n \leq 880$  plantas) da safra 2009/2010 (Tabela 2). Os maiores tamanhos de amostra foram necessários para a estimação do coeficiente de correlação entre os caracteres NF e CE (445 plantas  $\leq n \leq 935$  plantas), NF e MCG (575 plantas  $\leq n \leq 775$  plantas), AP e DS (565 plantas  $\leq n \leq 800$  plantas), MCG e NGR (625 plantas  $\leq n \leq 845$  plantas), AIE e DS (565 plantas  $\leq n \leq 990$  plantas), DS e CGR (590 plantas  $\leq n \leq 880$  plantas) e entre MCG e CGR (585 plantas  $\leq n > 1.000$  plantas). Esses pares de caracteres apresentaram coeficientes de correlação de baixa magnitude nos híbridos simples, triplo e duplo avaliados nas safras 2008/2009 e 2009/2010 ( $-0,20 \leq r \leq 0,43$ ) (Tabela 1).

**Tabela 1.** Estimativas dos coeficientes de correlação linear de Pearson<sup>(1)</sup> entre os 66 pares de caracteres de milho dos híbridos simples P32R21, triplo DKB566 e duplo DKB747 da safra 2008/2009 e dos híbridos simples 30F53, triplo DKB566 e duplo DKB747 da safra 2009/2010

Caractere <sup>(2)</sup>	AP	AIE	PE	NF	CE	DE	PS	DS	MCG	NGR	CGR	PROD
Híbrido simples P32R21 (2008/2009) acima da diagonal e híbrido triplo DKB566 (2008/2009) abaixo da diagonal <sup>(3)</sup>												
AP		0,61	0,40	0,21	0,38	0,35	0,40	0,29	0,29	0,33	0,24	0,38
AIE	0,74		0,26	0,09	0,24	0,24	0,20	0,14	0,22	0,23	0,20	0,26
PE	0,31	0,31		0,57	0,80	0,83	0,67	0,48	0,54	0,95	0,68	1,00
NF	0,09	0,17	0,42		0,38	0,70	0,23	0,41	0,11	0,65	0,57	0,59
CE	0,28	0,26	0,84	0,26		0,56	0,73	0,28	0,30	0,79	0,50	0,78
DE	0,26	0,31	0,83	0,57	0,66		0,47	0,55	0,53	0,79	0,84	0,83
PS	0,31	0,23	0,78	0,14	0,77	0,62		0,48	0,40	0,55	0,26	0,62
DS	0,19	0,16	0,54	0,43	0,44	0,68	0,61		0,40	0,40	0,02	0,46
MCG	0,27	0,18	0,48	-0,20	0,42	0,46	0,59	0,41		0,30	0,38	0,54
NGR	0,24	0,28	0,90	0,61	0,74	0,75	0,56	0,42	0,08		0,69	0,95
CGR	0,22	0,30	0,73	0,46	0,57	0,86	0,40	0,21	0,33	0,70		0,70
PROD	0,31	0,31	1,00	0,43	0,83	0,83	0,75	0,52	0,47	0,91	0,75	
Híbrido duplo DKB747 (2008/2009) acima da diagonal e híbrido simples 30F53 (2009/2010) abaixo da diagonal												
AP		0,73	0,48	0,15	0,44	0,43	0,44	0,29	0,35	0,42	0,30	0,47
AIE	0,58		0,37	0,04	0,30	0,35	0,26	0,15	0,29	0,34	0,32	0,38
PE	0,42	-0,12		0,33	0,86	0,81	0,86	0,56	0,59	0,91	0,56	0,99
NF	0,31	0,01	0,53		0,20	0,54	0,20	0,46	-0,16	0,49	0,28	0,35
CE	0,34	-0,20	0,92	0,43		0,61	0,80	0,45	0,47	0,79	0,39	0,84
DE	0,39	-0,02	0,86	0,67	0,76		0,69	0,68	0,50	0,75	0,69	0,81
PS	0,42	-0,08	0,94	0,46	0,88	0,80		0,65	0,57	0,69	0,29	0,80
DS	0,36	0,04	0,68	0,54	0,58	0,73	0,70		0,35	0,47	-0,06	0,52
MCG	0,16	-0,14	0,49	-0,06	0,41	0,32	0,52	0,30		0,23	0,34	0,58
NGR	0,41	-0,09	0,94	0,62	0,88	0,85	0,83	0,64	0,19		0,56	0,92
CGR	0,28	-0,05	0,72	0,55	0,65	0,88	0,62	0,31	0,23	0,73		0,60
PROD	0,42	-0,13	1,00	0,54	0,91	0,86	0,92	0,67	0,48	0,94	0,72	
Híbrido triplo DKB566 (2009/2010) acima da diagonal e híbrido duplo DKB747 (2009/2010) abaixo da diagonal												
AP		0,67	0,24	0,17	0,20	0,21	0,24	0,25	0,15	0,22	0,12	0,23
AIE	0,76		-0,01	0,02	-0,12	0,00	-0,02	0,03	-0,11	0,04	-0,02	-0,01
PE	0,29	0,15		0,42	0,79	0,86	0,85	0,64	0,55	0,91	0,72	1,00
NF	0,12	0,02	0,39		0,27	0,58	0,30	0,44	-0,03	0,54	0,48	0,43
CE	0,28	0,08	0,81	0,31		0,62	0,81	0,54	0,44	0,72	0,47	0,77
DE	0,26	0,15	0,82	0,59	0,62		0,69	0,67	0,42	0,83	0,89	0,86
PS	0,32	0,15	0,86	0,33	0,84	0,68		0,67	0,57	0,69	0,49	0,82
DS	0,31	0,15	0,70	0,50	0,61	0,77	0,74		0,36	0,59	0,26	0,63
MCG	0,23	0,11	0,47	-0,06	0,47	0,35	0,54	0,35		0,20	0,32	0,54
NGR	0,22	0,12	0,90	0,48	0,68	0,79	0,66	0,62	0,09		0,72	0,92
CGR	0,09	0,08	0,57	0,40	0,34	0,76	0,31	0,18	0,20	0,60		0,73
PROD	0,27	0,15	0,99	0,39	0,77	0,82	0,79	0,66	0,44	0,92	0,61	

<sup>(1)</sup> Valor superior a |0,10|, |0,10| e |0,09|, respectivamente, para os híbridos simples P32R21, triplo DKB566 e duplo DKB747 de milho da safra 2008/2009 e superior a |0,04|, |0,04| e |0,04|, respectivamente, para os híbridos simples 30F53, triplo DKB566 e duplo DKB747 de milho da safra 2009/2010, é significativo a 5% de probabilidade, pelo teste t, com, respectivamente, 359, 371, 414, 1.775, 1.691 e 1.718 graus de liberdade. <sup>(2)</sup> AP: altura de planta na colheita; AIE: altura de inserção de espiga; PE: peso de espiga; NF: número de fileiras de grãos por espiga; CE: comprimento de espiga; DE: diâmetro de espiga; PS: peso de sabugo; DS: diâmetro de sabugo; MCG: massa de 100 grãos; NGR: número de grãos por espiga; CGR: comprimento de grãos; PROD: produtividade de grãos. <sup>(3)</sup> Os coeficientes de correlação entre os 66 pares de caracteres avaliados nos híbridos simples, triplo e duplo do experimento conduzido na safra 2008/2009 já foram previamente apresentados por Cargnelutti Filho et al. (2010).

Menores tamanhos de amostra foram necessários para a estimação do r entre caracteres com maiores magnitudes de correlação como, por exemplo, entre PE e DE (40 plantas  $\leq n \leq 100$  plantas), DE e PROD (45 plantas  $\leq n \leq 100$  plantas), PE e CE (25 plantas  $\leq n \leq 120$  plantas) e entre CE e PROD (30 plantas  $\leq n \leq 145$  plantas) (Tabela 2), que apresentaram coeficientes de correlação no intervalo de  $0,77 \leq r \leq 0,92$  (Tabela 1). Tamanhos de amostra ainda

menores foram necessários para a estimação do r entre os caracteres PE e NGR (15 plantas  $\leq n \leq 45$  plantas), NGR e PROD (10 plantas  $\leq n \leq 30$  plantas) e entre PE e PROD, sendo que para esse último par de caracteres, 10 plantas foram suficientes para a estimação do r, com  $AIC_{95\%}$  menor ou igual a 0,15, independentemente do híbrido e da safra (Tabela 2). Esses pares de caracteres apresentaram os maiores escores de correlação, sendo a correlação entre PE

**Tabela 2.** Tamanho de amostra (número de plantas) para a estimação do coeficiente de correlação linear de Pearson de 66 pares de caracteres, mensurados nos híbridos simples P32R21, triplo DKB566 e duplo DKB747 na safra 2008/2009 e nos híbridos simples 30F53, triplo DKB566 e duplo DKB747 na safra 2009/2010, para amplitude do intervalo de confiança de 95% igual a 0,15

Caractere <sup>(1)</sup>	AP	AIE	PE	NF	CE	DE	PS	DS	MCG	NGR	CGR	PROD
Híbrido simples P32R21 (2008/2009) acima da diagonal e híbrido triplo DKB566 (2008/2009) abaixo da diagonal												
AP		325	590	600	705	620	535	795	850	670	520	610
AIE	125		675	615	685	745	600	890	675	675	635	675
PE	615	600		280	105	60	290	475	370	15	180	10
NF	520	575	540		445	210	575	410	775	240	435	255
CE	660	520	70	935		290	165	790	640	125	405	135
DE	715	715	70	400	300		410	420	495	100	85	50
PS	570	605	125	565	150	250		385	565	410	670	365
DS	800	990	355	405	440	255	300		610	585	815	475
MCG	620	665	575	635	805	600	325	470		675	715	400
NGR	545	595	35	345	170	200	260	410	795		195	10
CGR	675	575	135	585	410	75	355	590	750	255		180
PROD	615	580	10	540	80	75	155	380	610	30	135	
Híbrido duplo DKB747 (2008/2009) acima da diagonal e híbrido simples 30F53 (2009/2010) abaixo da diagonal												
AP		195	525	685	525	600	480	775	620	525	640	525
AIE	235		515	680	705	555	580	765	500	590	465	510
PE	510	550		520	55	100	60	390	300	40	420	10
NF	670	585	390		715	385	615	440	575	425	615	520
CE	650	510	25	610		350	70	500	395	115	635	65
DE	595	510	40	345	160		175	260	415	190	240	100
PS	515	570	30	420	40	110		240	285	200	635	105
DS	655	635	390	370	375	235	230		575	475	800	410
MCG	745	510	635	760	715	935	580	640		675	585	360
NGR	575	545	15	290	65	70	80	485	845		465	30
CGR	595	520	145	530	305	80	250	865	>1.000	185		390
PROD	505	550	10	390	30	50	40	400	675	15	145	
Híbrido triplo DKB566 (2009/2010) acima da diagonal e híbrido duplo DKB747 (2009/2010) abaixo da diagonal												
AP		200	675	770	720	805	705	630	670	685	760	685
AIE	125		740	775	705	825	705	705	650	810	825	745
PE	585	590		520	120	45	95	300	470	25	140	10
NF	690	670	535		710	430	565	475	715	435	625	520
CE	630	640	95	655		365	85	465	565	200	495	130
DE	665	540	70	410	300		180	245	585	75	60	45
PS	560	590	85	580	70	175		280	490	200	360	135
DS	565	565	205	515	300	150	150		520	410	760	310
MCG	510	580	395	645	405	480	300	470		755	660	510
NGR	695	590	25	435	255	125	305	370	625		190	25
CGR	745	560	350	675	605	160	590	880	625	305		140
PROD	650	590	10	535	145	70	175	260	470	20	275	

<sup>(1)</sup> AP: altura de planta na colheita; AIE: altura de inserção de espiga; PE: peso de espiga; NF: número de fileiras de grãos por espiga; CE: comprimento de espiga; DE: diâmetro de espiga; PS: peso de sabugo; DS: diâmetro de sabugo; MCG: massa de 100 grãos; NGR: número de grãos por espiga; CGR: comprimento de grãos; PROD: produtividade de grãos.

e NGR ( $0,90 \leq r \leq 0,95$ ), NGR e PROD ( $0,91 \leq r \leq 0,95$ ) e entre PE e PROD ( $0,99 \leq r \leq 1,00$ ) de elevada magnitude (Tabela 1). Dessa forma, pode-se destacar que quanto maior for a intensidade da associação entre dois caracteres, menor será o tamanho de amostra necessário para a estimação da correlação e vice-versa, num determinado nível de precisão, conforme já constatado em estudos anteriores (Bonett & Wright, 2000; Cargnelutti Filho et al., 2010, 2011, 2012; Shieh, 2010).

Para condições nas quais é desejado estimar o  $r$  para cada um dos 66 pares de caracteres com  $AIC_{95\%}$  máxima de 0,25, seria necessária a mensuração de 320, 365 e 295 plantas, respectivamente, nos híbridos simples P32R21, triplo DKB566 e duplo DKB747 da safra 2008/2009 e de 375, 315 e 310 plantas, respectivamente, nos híbridos simples 30F53, triplo DKB566 e duplo DKB747 da safra 2009/2010 (Tabela 3). Assim, independentemente do híbrido, da safra e do par de caracteres, seria recomendada

**Tabela 3.** Tamanho de amostra (número de plantas) para a estimação do coeficiente de correlação linear de Pearson de 66 pares de caracteres, mensurados nos híbridos simples P32R21, triplo DKB566 e duplo DKB747 na safra 2008/2009 e nos híbridos simples 30F53, triplo DKB566 e duplo DKB747 na safra 2009/2010, para amplitude do intervalo de confiança de 95% igual a 0,25

Caractere <sup>(1)</sup>	AP	AIE	PE	NF	CE	DE	PS	DS	MCG	NGR	CGR	PROD
Híbrido simples P32R21 (2008/2009) acima da diagonal e híbrido triplo DKB566 (2008/2009) abaixo da diagonal												
AP		120	225	255	240	240	200	290	315	225	190	225
AIE	55		245	230	255	290	245	320	250	245	235	245
PE	225	235		110	40	25	120	180	135	10	70	10
NF	225	225	230		175	90	215	170	290	90	165	100
CE	230	215	35	340		125	65	290	250	50	155	55
DE	265	265	35	140	115		160	155	205	40	35	25
PS	215	215	50	205	55	95		150	210	155	240	155
DS	285	365	130	165	175	100	115		220	215	305	180
MCG	225	255	210	215	275	225	115	175		255	270	135
NGR	215	215	15	130	65	75	95	180	285		75	10
CGR	225	215	55	210	150	35	125	235	285	105		70
PROD	225	225	10	230	35	40	65	130	215	15	55	
Híbrido duplo DKB747 (2008/2009) acima da diagonal e híbrido simples 30F53 (2009/2010) abaixo da diagonal												
AP		80	200	255	200	235	185	275	220	200	260	200
AIE	95		185	250	245	205	210	290	190	205	190	185
PE	195	215		210	25	40	25	140	125	20	155	10
NF	250	220	145		250	145	225	160	205	175	230	210
CE	225	195	15	225		130	35	180	135	50	245	30
DE	215	215	20	125	60		75	95	155	70	95	40
PS	195	200	10	160	25	40		90	115	70	240	35
DS	225	240	140	160	130	75	85		225	170	295	160
MCG	270	190	230	305	275	340	205	250		245	220	130
NGR	195	210	10	120	25	30	35	175	295		165	15
CGR	225	180	65	210	120	40	100	330	375	70		150
PROD	195	215	10	145	15	25	15	160	230	10	65	
Híbrido triplo DKB566 (2009/2010) acima da diagonal e híbrido duplo DKB747 (2009/2010) abaixo da diagonal												
AP		80	235	270	300	300	250	245	240	265	275	235
AIE	50		275	300	250	315	265	265	225	275	310	275
PE	230	240		190	45	25	35	120	165	15	60	10
NF	270	250	190		275	165	205	190	250	165	230	190
CE	230	235	40	255		140	35	175	200	75	190	50
DE	250	225	25	155	115		75	90	200	35	25	25
PS	210	245	35	230	30	70		110	190	80	135	40
DS	235	220	85	185	125	55	55		195	150	275	120
MCG	210	230	150	260	150	190	120	170		280	260	195
NGR	250	240	15	155	95	45	105	125	245		70	10
CGR	255	235	125	270	210	70	220	310	230	130		55
PROD	230	240	10	190	55	35	70	100	160	15	110	

<sup>(1)</sup> AP: altura de planta na colheita; AIE: altura de inserção de espiga; PE: peso de espiga; NF: número de fileiras de grãos por espiga; CE: comprimento de espiga; DE: diâmetro de espiga; PS: peso de sabugo; DS: diâmetro de sabugo; MCG: massa de 100 grãos; NGR: número de grãos por espiga; CGR: comprimento de grãos; PROD: produtividade de grãos.

a mensuração de 375 plantas, para a estimação do  $r$  com  $AIC_{95\%}$  máxima de 0,25.

Para a estimação do  $r$  de cada um dos 66 pares de caracteres com  $AIC_{95\%}$  máxima de 0,35 seria necessária a mensuração de 165, 195 e 155 plantas, respectivamente, nos híbridos simples, triplo e duplo da safra 2008/2009 e de 190, 165 e 160 plantas, respectivamente, nos híbridos simples, triplo e duplo da safra 2009/2010 (Tabela 4). Já a mensuração de 100, 110 e 100 plantas, respectivamente, nos híbridos simples, triplo e duplo da safra 2008/2009 e de 120, 100 e

100 plantas, respectivamente, nos híbridos simples, triplo e duplo da safra 2009/2010 seria suficiente para a estimação do coeficiente de correlação de cada um dos 66 pares de caracteres com  $AIC_{95\%}$  máxima de 0,45 (Tabela 5).

Para a estimação do  $r$  em 91 pares de caracteres de milho com  $AIC_{95\%}$  máxima de 0,30, Cargnelutti Filho et al. (2010) recomendaram a mensuração de até 300 plantas, dependendo do híbrido e do par de caracteres. Em crambe, Cargnelutti Filho et al. (2011) verificaram que para a estimação do  $r$  com  $AIC_{95\%}$  máxima de 0,15 o tamanho de amostra oscilou

**Tabela 4.** Tamanho de amostra (número de plantas) para a estimação do coeficiente de correlação linear de Pearson de 66 pares de caracteres, mensurados nos híbridos simples P32R21, triplo DKB566 e duplo DKB747 na safra 2008/2009 e nos híbridos simples 30F53, triplo DKB566 e duplo DKB747 na safra 2009/2010, para amplitude do intervalo de confiança de 95% igual a 0,35

Caractere <sup>(1)</sup>	AP	AIE	PE	NF	CE	DE	PS	DS	MCG	NGR	CGR	PROD
Híbrido simples P32R21 (2008/2009) acima da diagonal e híbrido triplo DKB566 (2008/2009) abaixo da diagonal												
AP		70	120	135	130	110	105	150	160	125	105	120
AIE	30		125	125	140	140	125	165	125	140	120	125
PE	115	120		60	25	15	60	100	75	10	40	10
NF	120	125	125		85	50	120	95	135	55	90	60
CE	125	120	15	175		60	35	155	130	30	70	30
DE	140	135	25	80	60		85	90	100	25	25	15
PS	115	120	30	125	25	50		80	120	80	120	75
DS	155	195	70	85	95	50	60		125	120	160	100
MCG	125	130	110	130	155	115	65	90		120	135	80
NGR	115	115	10	75	35	45	55	90	160		40	10
CGR	120	130	35	125	75	20	70	125	165	65		35
PROD	115	120	10	125	20	25	40	70	130	10	35	
Híbrido duplo DKB747 (2008/2009) acima da diagonal e híbrido simples 30F53 (2009/2010) abaixo da diagonal												
AP		45	95	130	105	125	95	155	125	125	125	95
AIE	50		100	130	125	100	125	145	100	120	100	100
PE	105	110		105	15	25	15	70	65	10	90	10
NF	140	110	80		155	75	120	80	110	100	125	105
CE	120	100	10	120		70	20	95	85	25	125	20
DE	110	110	15	70	40		40	50	90	40	55	25
PS	100	100	10	90	10	20		55	60	40	130	25
DS	130	130	70	75	65	35	40		125	95	155	80
MCG	140	110	130	160	130	180	110	120		125	115	65
NGR	105	110	10	60	15	20	20	95	170		90	10
CGR	125	100	35	115	60	20	50	175	190	40		75
PROD	105	110	10	80	10	15	10	90	130	10	35	
Híbrido triplo DKB566 (2009/2010) acima da diagonal e híbrido duplo DKB747 (2009/2010) abaixo da diagonal												
AP		40	135	150	140	160	135	130	135	140	160	135
AIE	30		135	165	145	165	135	140	140	165	165	135
PE	110	115		95	30	15	20	65	90	10	35	10
NF	135	130	115		145	85	100	100	140	90	125	95
CE	115	115	25	125		75	20	95	110	45	100	30
DE	120	120	15	90	65		40	45	105	25	15	15
PS	110	110	20	115	20	40		55	100	40	80	25
DS	110	105	40	95	65	30	30		95	80	145	70
MCG	115	120	85	135	85	95	65	95		135	130	90
NGR	135	115	10	90	45	30	60	65	135		40	10
CGR	130	130	70	135	120	40	120	160	125	70		35
PROD	110	110	10	115	35	20	35	50	90	10	60	

<sup>(1)</sup> AP: altura de planta na colheita; AIE: altura de inserção de espiga; PE: peso de espiga; NF: número de fileiras de grãos por espiga; CE: comprimento de espiga; DE: diâmetro de espiga; PS: peso de sabugo; DS: diâmetro de sabugo; MCG: massa de 100 grãos; NGR: número de grãos por espiga; CGR: comprimento de grãos; PROD: produtividade de grãos.

entre 8 e 665 plantas, de acordo com o par de caracteres considerado. Já em 210 pares de caracteres de dois híbridos de mamoneira, Cargnelutti Filho et al. (2012) constaram que 96 plantas foram suficientes para a estimação do  $r$  com  $AIC_{95\%}$  máxima de 0,52. Os autores também verificaram que para a estimação do  $r$  com  $AIC_{95\%}$  de 0,20 o tamanho de amostra oscilou entre 10 e 661 plantas, dependendo do par de caracteres mensurados.

Segundo Bonett & Wright (2000), para a estimação do  $r$  por meio do intervalo de confiança de Fisher, com

$AIC_{95\%}$  de 0,10, foi necessário tamanho de amostra de  $n = 1.507$  e  $n = 63$  observações, respectivamente, para coeficientes de correlação de baixa ( $r = 0,10$ ) e de elevada magnitude ( $r = 0,90$ ). Já os tamanhos de amostra de  $n = 168$  e  $n = 13$  observações seriam suficientes, de acordo com os autores, para a estimação desses coeficientes com  $AIC_{95\%}$  igual a 0,30. De acordo com Shieh (2010), a utilização de tamanhos de amostra maiores reduziu o viés e a raiz do quadrado médio do erro associados às estimativas do  $r$ . O autor constatou que maiores escores da raiz do quadrado

**Tabela 5.** Tamanho de amostra (número de plantas) para a estimação do coeficiente de correlação linear de Pearson de 66 pares de caracteres, mensurados nos híbridos simples P32R21, triplo DKB566 e duplo DKB747 na safra 2008/2009 e nos híbridos simples 30F53, triplo DKB566 e duplo DKB747 na safra 2009/2010, para amplitude do intervalo de confiança de 95% igual a 0,45

Caractere <sup>(1)</sup>	AP	AIE	PE	NF	CE	DE	PS	DS	MCG	NGR	CGR	PROD
Híbrido simples P32R21 (2008/2009) acima da diagonal e híbrido triplo DKB566 (2008/2009) abaixo da diagonal												
AP		40	75	70	80	70	65	85	95	75	70	75
AIE	20		80	80	80	85	75	95	90	85	75	80
PE	80	70		40	20	10	40	60	50	10	25	10
NF	75	70	75		60	30	75	55	90	40	60	40
CE	70	70	15	100		40	25	95	80	20	50	20
DE	90	85	15	50	35		50	50	60	20	15	10
PS	70	70	20	75	15	40		50	70	50	80	45
DS	90	110	45	55	60	30	40		75	70	100	60
MCG	75	80	75	75	90	75	40	60		80	90	50
NGR	65	65	10	45	25	30	30	55	90		30	10
CGR	80	75	25	70	50	15	50	65	90	35		25
PROD	80	75	10	75	15	15	20	45	75	10	25	
Híbrido duplo DKB747 (2008/2009) acima da diagonal e híbrido simples 30F53 (2009/2010) abaixo da diagonal												
AP		30	70	85	65	70	70	95	70	70	75	70
AIE	30		70	85	80	70	80	95	65	75	65	60
PE	65	65		70	10	15	10	50	40	10	55	10
NF	85	70	50		90	50	75	55	70	60	75	65
CE	75	65	10	70		45	15	60	50	20	80	10
DE	75	70	10	45	20		25	30	55	25	35	20
PS	65	65	10	55	10	15		30	40	25	70	15
DS	70	80	35	55	45	25	25		65	60	100	55
MCG	85	65	70	95	90	105	65	80		80	80	45
NGR	65	70	10	40	10	15	10	55	100		55	10
CGR	80	70	25	65	40	15	30	95	120	30		45
PROD	65	65	10	50	10	10	10	35	70	10	25	
Híbrido triplo DKB566 (2009/2010) acima da diagonal e híbrido duplo DKB747 (2009/2010) abaixo da diagonal												
AP		30	85	85	85	85	85	75	80	85	95	85
AIE	20		85	90	80	100	80	80	80	95	100	85
PE	75	75		60	15	10	10	45	60	10	25	10
NF	85	75	65		90	50	70	65	85	50	70	70
CE	75	75	15	80		50	15	60	65	30	65	20
DE	75	75	15	55	40		20	35	70	15	15	10
PS	70	75	15	80	10	20		35	55	25	45	15
DS	75	75	30	55	45	20	20		65	50	90	45
MCG	65	70	50	85	50	60	45	60		85	75	60
NGR	80	75	10	55	30	20	40	40	85		30	10
CGR	85	75	40	85	70	25	75	100	75	40		25
PROD	75	75	10	65	20	15	25	35	55	10	40	

<sup>(1)</sup> AP: altura de planta na colheita; AIE: altura de inserção de espiga; PE: peso de espiga; NF: número de fileiras de grãos por espiga; CE: comprimento de espiga; DE: diâmetro de espiga; PS: peso de sabugo; DS: diâmetro de sabugo; MCG: massa de 100 grãos; NGR: número de grãos por espiga; CGR: comprimento de grãos; PROD: produtividade de grãos.

médio do erro estão associados a coeficientes de correlação de pequena magnitude e pequenos escores de raiz do quadrado médio do erro estão associados a coeficientes de correlação de elevada magnitude, tanto positivos como negativos.

Verificou-se no presente estudo que independentemente do híbrido, da safra e do par de caracteres, 375, 195 e 120 plantas foram suficientes, respectivamente, para a estimação do  $r$  com  $AIC_{95\%}$  máximas de 0,25, 0,35 e 0,45 (Tabela 3, Tabela 4 e Tabela 5). Assim, se for conduzido um experimento com cinco tratamentos e quatro repetições (20 parcelas no total)

e forem avaliadas dez plantas por parcela (200 plantas no total), pode-se estimar o  $r$  com  $AIC_{95\%}$  máxima de 0,35, desde que os efeitos de tratamentos e de controle local sejam suprimidos. Se, entretanto, forem avaliadas seis plantas por parcela (120 plantas no total), pode-se estimar o  $r$  de cada par de caracteres com  $AIC_{95\%}$  máxima de 0,45, desde que os efeitos de tratamentos e de controle local também sejam suprimidos. Caso o pesquisador deseje estimar o  $r$  dos 66 pares de caracteres com  $AIC_{95\%}$  máxima de 0,45, dentro de cada tratamento, utilizando quatro repetições,

deverá avaliar 30 plantas por repetição (120 plantas por tratamento), desde que o efeito de controle local seja excluído antes da estimação do  $r$ .

#### 4. CONCLUSÃO

O tamanho de amostra varia entre híbridos, safras e pares de caracteres. Maior tamanho de amostra é necessário para a estimação do coeficiente de correlação entre caracteres fracamente correlacionados e menor tamanho de amostra é necessário para a estimação do coeficiente de correlação entre caracteres altamente correlacionados.

Independentemente do híbrido, da safra e do par de caracteres, 375, 195 e 120 plantas são suficientes, respectivamente, para a estimação de coeficientes de correlação com  $AIC_{95\%}$  máximas de 0,25, 0,35 e 0,45.

#### AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelas bolsas concedidas. Aos bolsistas e voluntários, pelo auxílio na condução experimental e coleta dos dados.

#### REFERÊNCIAS

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, 65, 23-28. <http://dx.doi.org/10.1007/BF02294183>.

Cargnelutti Filho, A., Toebe, M., Burin, C., Silveira, T. R., & Casarotto, G. (2010). Tamanho de amostra para estimação do coeficiente de correlação linear de Pearson entre caracteres de milho. *Pesquisa Agropecuária Brasileira*, 45, 1363-1371. Recuperado de <http://www.scielo.br/pdf/pab/v45n12/v45n12a05.pdf>

Cargnelutti Filho, A., Lopes, S. J., Toebe, M., Silveira, T. R., & Schwantes, I. A. (2011). Tamanho de amostra para estimação do coeficiente de correlação de Pearson entre caracteres de *Crambe abyssinica*. *Revista Ciência Agronômica*, 42, 149-158. <http://dx.doi.org/10.1590/S1806-66902011000100019>.

Cargnelutti Filho, A., Lopes, S. J., Brum, B., Toebe, M., Silveira, T. R., & Casarotto, G. (2012). Tamanho de amostra para a estimação do coeficiente de correlação linear de Pearson entre caracteres de mamoneira. *Semina. Ciências Agrárias*, 33, 953-962. <http://dx.doi.org/10.5433/1679-0359.2012v33n3p953>.

Cruz, C. D., & Regazzi, A. J. (1997). Modelos biométricos aplicados ao melhoramento genético (2. ed.). Viçosa: UFV. 390 p.

Duvick, D. N. (2005). The contribution of breeding to yield advances in maize (*Zea mays* L.). *Advances in Agronomy*, 86, 83-145. [http://dx.doi.org/10.1016/S0065-2113\(05\)86002-X](http://dx.doi.org/10.1016/S0065-2113(05)86002-X).

Fancelli, A. L., & Dourado Neto, D. (2004). Produção de milho (2. ed.). Guaíba: Agropecuária. 360 p.

Food and Agriculture Organization of the United Nations – FAO. (2014). Recuperado de <http://statistics.amis-outlook.org/data/index.html#DOWNLOAD>

Ferreira, D.F. (2009). Estatística básica. 2. ed. (664p). Lavras: UFLA.

R Development Core Team. (2014). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Recuperado de <http://www.R-project.org>

Shieh, G. (2010). Estimation of the simple correlation coefficient. *Behavior Research Methods*, 42, 906-917. <http://dx.doi.org/10.3758/BRM.42.4.906>. PMID:21139158