

Estatística

Tamanho de amostra para avaliação do grau de multicolinearidade em caracteres produtivos de centeio

Sample size for evaluation the of multicollinearity degree in productive traits of rye

Ismael Mario Márcio Neu¹, Alberto Cargnelutti Filho¹, Cláudia Marques de Bem¹, Jéssica Andiará Kleinpaul¹, Cirineu Tolfo Bandeira¹

¹ Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós-Graduação em Agronomia. Santa Maria, RS, Brasil

RESUMO

Os objetivos deste trabalho foram determinar o tamanho de amostra (número de plantas) necessário para a estimação dos indicadores do grau de multicolinearidade - número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV) - em caracteres produtivos de centeio e verificar a variabilidade do tamanho de amostra entre os indicadores. Foram conduzidos cinco e três ensaios de uniformidade (cultivares BRS Progresso e Temprano, respectivamente) e avaliados sete caracteres produtivos em 780 plantas. Vinte e um casos foram obtidos a partir de sete caracteres, combinados cinco a cinco. Para cada caso, foram planejados 197 tamanhos de amostra (20, 25, 30, ..., 1.000 plantas) e para cada tamanho foram realizadas 2.000 reamostragens, com reposição. Para cada reamostra foram determinados o NC, DET e FIV e calculada a média das 2.000 estimativas de cada indicador do grau de multicolinearidade. Após, foi determinado o tamanho de amostra, por meio de: máxima curvatura modificado e modelos linear e quadrático segmentado com resposta platô. Houve superioridade do modelo quadrático segmentado com platô no ajuste do grau de multicolinearidade em função do tamanho de amostra para todos os indicadores. Há a necessidade de maior tamanho de amostra para a detecção da multicolinearidade quando o diagnóstico for realizado pelo DET e de tamanhos de amostra superiores a 101, 258 e 102 plantas quando o diagnóstico for realizado pelo número de condição, determinante e fator de inflação da variância, respectivamente.

Palavras-chave: Secale cereale L; Dimensionamento amostral; Correlação; Análise multivariada

ABSTRACT

The objectives of this work were to determine the sample size (number of plants) necessary to estimate the indicators of the of multicollinearity degree - condition number (CN), determinant of the correlation matrix (DET), and variance inflation factor (VIF) - in productive traits of rye and to verify the variability of the sample

size between the indicators. Five and three uniformity trials were conducted with the cultivars BRS Progresso and Temprano, respectively, and seven productive traits were evaluated in 780 plants. Twenty-one cases were obtained from seven traits, combined five to five. In each case, 197 sample sizes were planned (20, 25, 30, ..., 1,000 plants) and in each size 2,000 resampling were performed, with replacement. For each resample the CN, DET and FIV were determined and the average among 2,000 estimates of each indicator of the multicollinearity degree was calculated. Then, for each case and indicator, the sample size was determined through three models: models of maximum modified curvature, segmented linear with plateau response, and segmented quadratic with plateau response. There was superiority the quadratic model segmented with plateau in adjusting the degree of multicollinearity according to the sample size for all indicators. There is a need greater sample size to detect multicollinearity when diagnosed by DET and for sizes larger than 101, 258 and 102 plants when diagnosing for the number of conditions, determinant and inflation factor performed, respectively.

Keywords: *Secale cereale* L; Sampling design; Correlation; Multivariate analysis

1 INTRODUÇÃO

Em programas de melhoramento de plantas são almejados genótipos com desempenhos superiores aos existentes. Nesse processo de seleção, as técnicas estatísticas uni e multivariadas podem ser utilizadas como ferramentas auxiliares. Essas técnicas estatísticas, além de necessitarem atender as suposições básicas de normalidade, homocedasticidade e linearidade, necessitam que a multicolinearidade seja investigada (HAIR et al., 2009). Estes autores mencionam que a presença de multicolinearidade em grau elevado superestima os parâmetros e aumenta o erro padrão das mesmas, gerando estimativas não confiáveis. Em estudo de Toebe e Cargnelutti Filho (2013), foram relatados erros de estimação dos coeficientes de efeitos diretos e, conseqüentemente, resultados sem sentido biológico e sem interpretação em análise de trilha, em presença de alto grau de multicolinearidade.

Para determinar o grau de multicolinearidade, com confiabilidade, o tamanho de amostra (número de plantas) deve ser suficiente. A reamostragem tem sido utilizada na determinação do tamanho de amostra para a estimação da média em caracteres morfológicos e produtivos de feijão de porco (CARGNELUTTI FILHO et al., 2018b); estimação da média e mediana de caracteres morfológicos e produtivos de linho (CARGNELUTTI FILHO et al., 2018a) e caracteres de três híbridos de canola (CARGNELUTTI FILHO et al., 2018c); estimação da média e coeficiente de variação em quatro espécies de crotalária (TOEBE et al., 2018); e estimação de efeitos diretos sobre a produtividade de híbridos de milho (TOEBE et al., 2017b).

Inferências na área agrônômica quanto ao tamanho de amostra necessário para o diagnóstico do grau de multicolinearidade foram realizadas em dados de genótipos de milho (OLIVOTO et al., 2017). A interferência da multicolinearidade em análise de trilha, em tomate cereja, foi estudada por Sari et al. (2018). Esses autores constataram haver interferência de amostras pequenas no diagnóstico do grau de multicolinearidade.

Supõe-se que é possível determinar o tamanho de amostra (número de plantas) suficiente para o diagnóstico do grau de multicolinearidade e que esse tamanho difere entre os seguintes indicadores: número de condição, determinante e fator de inflação da variância. Assim, os objetivos deste trabalho foram determinar o tamanho de amostra (número de plantas) necessário para a estimação dos indicadores do grau de multicolinearidade - número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV) - em caracteres produtivos de centeio e verificar a variabilidade do tamanho de amostra entre os indicadores.

2 MATERIAL E MÉTODOS

Oito ensaios de uniformidade foram conduzidos com a cultura de centeio (*Secale cereale* L.), constituídos por cinco épocas de semeadura para a cultivar BRS Progresso (ensaios E1, E2, E3, E4 e E5) e três épocas de semeadura para a cultivar Temprano (ensaios E6, E7 e E8) (Tabela 1). Esses ensaios foram conduzidos em área experimental localizada em Santa Maria - RS (29°42'S, 53°49'W e 95 m de altitude). O clima da região é classificado por Köppen como Cfa - subtropical úmido, com verões quentes e sem estação seca definida (ALVARES et al., 2013). O solo da região é classificado como Argissolo Vermelho distrófico arênico (SANTOS et al., 2018).

A área experimental foi preparada de forma homogênea e a fertilidade do solo corrigida com a aplicação de 500 kg ha⁻¹ de fertilizante (formulação 5-20-20 de NPK). Foram semeadas duas cultivares de centeio, sendo a cultivar BRS Progresso de propósito para a produção de grãos e a cultivar Temprano para a cobertura de solo e planta forrageira. As sementes de cada cultivar foram semeadas, a lanço, em área de 320 m² (20

m × 16 m) na primeira época, e nas demais épocas, cada cultivar foi instalada em área de 375 m² (25 m × 15 m).

Tabela 1 - Datas de semeadura, de emergência, de florescimento e de avaliação, número de plantas avaliadas por época de semeadura e número de dias entre a semeadura e a avaliação (Ciclo), em duas cultivares de centeio (*Secale cereale* L.), avaliadas em oito ensaios de uniformidade, Safra 2016, Santa Maria, RS.

| Cultivar | Época | Ensaio | Semeadura | Emergência | Florescimento | Avaliação | Nº de plantas | Ciclo |
|---------------|-------|--------|------------|------------|---------------|------------|---------------|-------|
| BRS Progresso | 1 | E1 | 03/05/2016 | 09/05/2016 | 08/08/2016 | 24/10/2016 | 100 | 174 |
| BRS Progresso | 2 | E2 | 25/05/2016 | 01/06/2016 | 26/08/2016 | 10/11/2016 | 100 | 169 |
| BRS Progresso | 3 | E3 | 07/06/2016 | 17/06/2016 | 15/09/2016 | 19/11/2016 | 100 | 165 |
| BRS Progresso | 4 | E4 | 22/06/2016 | 29/06/2016 | 22/09/2016 | 27/11/2016 | 90 | 158 |
| BRS Progresso | 5 | E5 | 04/07/2016 | 10/07/2016 | 29/09/2016 | 02/12/2016 | 100 | 151 |
| Temprano | 1 | E6 | 03/05/2016 | 10/05/2016 | 22/09/2016 | 07/11/2016 | 100 | 188 |
| Temprano | 2 | E7 | 25/05/2016 | 02/06/2016 | 03/10/2016 | 18/11/2016 | 100 | 177 |
| Temprano | 3 | E8 | 07/06/2016 | 20/06/2016 | 10/10/2016 | 24/11/2016 | 90 | 170 |

As épocas de semeadura foram de março a julho afim de atender a recomendação de implantação para as cultivares (BAIER, 1994). Em todos os ensaios de uniformidade foi usada a densidade de 455 sementes m⁻². A adubação de cobertura foi realizada quando as plantas estavam entre os estádios de três e quatro folhas desenvolvidas, na quantidade de 25 kg ha⁻¹ de nitrogênio. Os demais tratos culturais foram realizados conforme as recomendações de manejo para a cultura de centeio (BAIER, 1994).

Em cada ensaio de uniformidade foram coletadas plantas de forma aleatória, quando estas estavam em estágio de maturação fisiológica dos grãos. As plantas foram armazenadas em local com ventilação natural e foram avaliados os caracteres produtivos das mesmas. Nos ensaios E4 e E8 foram avaliadas 90 plantas e 100 plantas nos demais ensaios. Assim, os dados das 780 plantas avaliadas em oito ensaios foram considerados como sendo a amostra mestre.

Em cada planta foram avaliados os seguintes caracteres produtivos: massa fresca de parte aérea da planta, em gramas (MFP - soma de massas de colmos e folhas planta⁻¹), sendo desconsiderada a massa de espiga; massa seca de parte aérea da planta, em

gramas (MSP - soma de massas secas de colmos e folhas planta⁻¹), obtido após a secagem em estufa de circulação forçada de MFP, em temperatura de 65°±5°C, até peso constante; número de espiguetas espiga⁻¹ (NSE - média do número de espiguetas espiga⁻¹); número de grãos espiga⁻¹ (NGE - número médio de grãos espiga⁻¹); número de grãos espiguetas⁻¹ (NGS - número médio de grãos espiguetas⁻¹); massa de grãos planta⁻¹, em gramas (MGP - total de massa de grãos planta⁻¹); e, massa média de 100 grãos, em gramas (M100).

O diagnóstico de multicolinearidade foi realizado na amostra mestre, com os sete caracteres, bem como, para as combinações de sete caracteres seis a seis e cinco a cinco, com base nos indicadores: número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV) (Tabelas 2 e 3). O NC foi obtido pela relação entre o maior autovalor ($\lambda_{\text{máx}}$) e o menor autovalor ($\lambda_{\text{mín}}$) da matriz de correlação ($\text{NC} = \lambda_{\text{máx}}/\lambda_{\text{mín}}$) (GUJARATI; PORTER, 2011) e classificado em fraca ($\text{NC} \leq 100$), moderada a forte ($100 < \text{NC} \leq 1.000$) e severa multicolinearidade ($\text{NC} > 1.000$) (MONTGOMERY; PECK, 1982). O FIV corresponde à proporção da contribuição da variância do *j*-ésimo caractere, quando comparada a situação de ausência de multicolinearidade ($\text{FIV} = 1$). O FIV foi estimado, conforme a equação (1) (HAIR et al., 2009; GUJARATI; PORTER, 2011, FARAWAY, 2016):

$$\text{FIV}_j = 1/(1 - R_j^2) \quad (1)$$

Sendo: R_j^2 o coeficiente de determinação múltiplo de determinada variável com as demais variáveis explicativas. Problema devido à multicolinearidade pode existir quando o valor de DET for inferior a 0,00001 (FIELD, 2009) e quando FIV_j for maior ou igual a dez (HAIR et al., 2009; GUJARATI; PORTER, 2011). O NC e o DET são indicadores com interpretação para todas as variáveis, enquanto que o FIV tem a vantagem de informar a inflação de variância para cada variável e neste estudo, para cada conjunto de caracteres, foi considerado o maior valor de FIV ($\text{FIV} = \text{máximo FIV}_j$).

Tabela 2 - Grau de multicolinearidade da amostra mestre (n = 780 plantas) obtido pelos indicadores número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV) em sete caracteres produtivos e em sete caracteres combinados seis a seis, avaliados em oito ensaios de uniformidade de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS

| ----- Caracteres ----- | | | | | | | NC | Classe ⁽¹⁾ | DET | VIF ⁽²⁾ |
|--|------|------|-----|-----|-----|------|--------|-----------------------|--------|--------------------|
| ----- Todos os sete caracteres ----- | | | | | | | | | | |
| MFP | MSP | NGE | NGS | NSE | MGP | M100 | 378,72 | Moderada à forte | 0,0001 | 59,13 |
| ----- Sete caracteres combinados seis a seis ----- | | | | | | | | | | |
| MFP | MSP | M100 | MGP | NGE | NGS | - | 279,25 | Moderada à forte | 0,0007 | 46,76 |
| MFP | MSP | M100 | MGP | NGE | NSE | - | 376,77 | Moderada à forte | 0,0004 | 59,02 |
| MFP | MSP | M100 | MGP | NGS | NSE | - | 361,99 | Moderada à forte | 0,0004 | 58,34 |
| MFP | MSP | M100 | NGE | NGS | NSE | - | 300,46 | Moderada à forte | 0,0005 | 59,12 |
| MFP | MSP | MGP | NGE | NGS | NSE | - | 368,63 | Moderada à forte | 0,0002 | 58,63 |
| MSP | M100 | MGP | NGE | NGS | NSE | - | 45,55 | Fraca | 0,0041 | 9,77 |
| MFP | M100 | MGP | NGE | NGS | NSE | - | 36,63 | Fraca | 0,0051 | 8,00 |

⁽¹⁾Multicolinearidade de moderada à forte ($100 < NC \leq 1.000$); Multicolinearidade fraca ($NC \leq 100$) (MONTGOMERY; PECK, 1982).

⁽²⁾Estimativa do maior valor do fator de inflação da variância.

Na determinação do tamanho de amostra para a estimação dos indicadores do grau de multicolinearidade (NC, DET e FIV), foram considerados os 21 casos obtidos pela combinação de sete caracteres cinco a cinco ($C_{(7,5)}$). Foi optado por essa combinação, devido à proporcionalidade entre os casos de fraca e de moderada à forte multicolinearidade, visto que na combinação de sete caracteres seis a seis, somente dois casos apresentaram multicolinearidade fraca e cinco casos de moderada à forte (Tabela 2).

Em cada um dos 21 casos, foram planejados 197 tamanhos de amostra. O menor tamanho de amostra planejado foi de 20 plantas e nos tamanhos seguintes, foram acrescentadas cinco plantas, até obter o maior tamanho que foi de 1.000 plantas. Assim, para cada caso, foram planejados os tamanhos de amostra de 20, 25, 30, ..., 1.000 plantas. Para cada tamanho de amostra foram realizadas 2.000 reamostragens, com reposição. Em cada reamostragem foram estimados o NC, DET e FIV (maior valor de FIV). Após, para cada tamanho de amostra planejado, calculou-se a média das 2.000 estimativas de cada indicador de multicolinearidade (NC, DT e FIV). A opção pela reamostragem com reposição possibilita planejar tamanhos superiores ao tamanho da amostra mestre, que nesse estudo foi de n = 780 plantas.

Tabela 3 - Grau de multicolinearidade para a amostra mestre (n = 780 plantas) obtida pelos indicadores número de condição (NC), determinante (DET) e fator de inflação da variância (FIV) em 21 casos obtidos pela combinação de sete caracteres produtivos, combinados cinco a cinco, avaliadas em oito ensaios de uniformidade de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS

| Caso | ----- Caracteres ----- | | | | | NC | Classe ⁽¹⁾ | DET | VIF ⁽²⁾ |
|------------------------------------|------------------------|-----|-----|------|------|--------|-----------------------|--------|--------------------|
| 1 | MFP | MSP | NGE | NGS | NSE | 291,78 | Moderada à forte | 0,0008 | 58,45 |
| 2 | MFP | MSP | NGE | NGS | M100 | 201,51 | Moderada à forte | 0,0043 | 45,76 |
| 3 | MFP | MSP | NGE | NGS | MGP | 247,16 | Moderada à forte | 0,0014 | 45,33 |
| 4 | MFP | MSP | NGE | NSE | M100 | 299,75 | Moderada à forte | 0,0024 | 59,02 |
| 5 | MFP | MSP | NGE | NSE | MGP | 367,97 | Moderada à forte | 0,0009 | 58,52 |
| 6 | MFP | MSP | NGE | M100 | MGP | 266,41 | Moderada à forte | 0,0032 | 46,75 |
| 7 | MFP | MSP | NGS | NSE | M100 | 295,40 | Moderada à forte | 0,0023 | 58,30 |
| 8 | MFP | MSP | NGS | NSE | MGP | 354,71 | Moderada à forte | 0,0008 | 57,75 |
| 9 | MFP | MSP | NGS | M100 | MGP | 249,89 | Moderada à forte | 0,0031 | 46,12 |
| 10 | MFP | MSP | NSE | M100 | MGP | 356,07 | Moderada à forte | 0,0007 | 57,97 |
| Mínimo | | | | | | 201,51 | - | 0,0007 | 45,33 |
| Média | | | | | | 239,07 | - | 0,0020 | 53,40 |
| Máximo | | | | | | 367,97 | - | 0,0043 | 59,02 |
| Desvio-padrão | | | | | | 54,26 | - | 0,0013 | 6,40 |
| Coeficiente de variação (CV), em % | | | | | | 18,51 | - | 64,46 | 11,98 |
| 11 | MFP | NGE | NGS | NSE | M100 | 25,41 | Fraca | 0,0315 | 6,23 |
| 12 | MFP | NGE | NGS | NSE | MGP | 34,23 | Fraca | 0,0107 | 7,99 |
| 13 | MFP | NGE | NGS | M100 | MGP | 28,19 | Fraca | 0,0327 | 6,00 |
| 14 | MFP | NGE | NSE | M100 | MGP | 35,07 | Fraca | 0,0237 | 7,85 |
| 15 | MFP | NGS | NSE | M100 | MGP | 33,11 | Fraca | 0,0216 | 8,00 |
| 16 | MSP | NGE | NGS | NSE | M100 | 31,98 | Fraca | 0,0249 | 7,84 |
| 17 | MSP | NGE | NGS | NSE | MGP | 42,74 | Fraca | 0,0088 | 9,77 |
| 18 | MSP | NGE | NGS | M100 | MGP | 28,37 | Fraca | 0,0331 | 5,89 |
| 19 | MSP | NGE | NSE | M100 | MGP | 43,67 | Fraca | 0,0194 | 9,57 |
| 20 | MSP | NGS | NSE | M100 | MGP | 41,44 | Fraca | 0,0177 | 9,75 |
| 21 | NGE | NGS | NSE | M100 | MGP | 23,84 | Fraca | 0,0405 | 4,79 |
| Mínimo | | | | | | 23,84 | - | 0,0088 | 5,89 |
| Média | | | | | | 33,46 | - | 0,0241 | 7,61 |
| Máximo | | | | | | 43,67 | - | 0,0405 | 9,75 |
| Desvio-padrão | | | | | | 6,86 | - | 0,0098 | 1,70 |
| Coeficiente de variação (CV), em % | | | | | | 20,50 | - | 40,75 | 22,36 |
| Mínimo geral | | | | | | 23,84 | - | 0,0007 | 5,89 |
| Média geral | | | | | | 157,08 | - | 0,0135 | 29,41 |
| Máximo geral | | | | | | 367,97 | - | 0,0405 | 59,02 |
| Desvio-padrão geral | | | | | | 137,84 | - | 0,0133 | 23,85 |
| Coeficiente de variação (CV), em % | | | | | | 87,75 | - | 98,18 | 81,11 |

⁽¹⁾Multicolinearidade de moderada à forte ($100 < NC \leq 1.000$); Multicolinearidade fraca ($NC \leq 100$) (MONTGOMERY; PECK, 1982).

⁽²⁾Maior estimativa do fator de inflação da variância.

Foram ajustados três modelos de regressão: método da máxima curvatura modificado (MMCM), modelo linear segmentado com resposta platô (MLRP) e modelo quadrático segmentado com resposta platô (MQRP). Nesses três modelos, foi ajustada a média do indicador (NC, DET, ou FIV) (variável dependente, Y_i) em função dos tamanhos de amostra planejados (variável independente, X_i). Para cada caso, indicador e modelo ($21 \times 3 \times 3 = 189$ situações) foram determinados o tamanho de amostra (n), o grau de multicolinearidade correspondente ao n obtido pelo ajuste de regressão ($Y_{(n)}$) e o coeficiente de determinação ajustado (R^2_a).

Foram determinados os coeficientes a e b para MMCM, pela expressão em (2):

$$Y_i = a/X_i^b + \varepsilon_i \quad (2)$$

Sendo: X_i a variável independente, ou seja, os tamanhos de amostra planejados (20, 25, 30, ..., 1.000 plantas) e Y_i a variável dependente referente a média das 2.000 estimativas de cada indicador do grau de multicolinearidade.

O tamanho de amostra (n) foi determinado conforme equação (3) (MEIER; LESSMAN, 1971) e a estimativa de multicolinearidade correspondente ao n ($Y_{(n)}$) conforme equação (4):

$$n = [a^2 b^2 (2b + 1)/(b + 2)]^{1/(2b+2)} \quad (3)$$

$$Y_{(n)} = a/n^b \quad (4)$$

Sendo: a e b os parâmetros do modelo.

Para MLRP e MQRP foram consideradas as funções (5) e (6), respectivamente:

$$Y_i = \begin{cases} a + b X_i + \varepsilon_i & \text{se } X_i \leq n \\ P + \varepsilon_i & \text{se } X_i > n \end{cases} \quad (5)$$

$$Y_i = \begin{cases} a + b X_i + c X_i^2 + \varepsilon_i & \text{se } X_i \leq n \\ P + \varepsilon_i & \text{se } X_i > n \end{cases} \quad (6)$$

Sendo: X_i a variável independente, ou seja, os tamanhos de amostra planejados (20, 25, 30, ..., 1.000 plantas); Y_i a variável dependente referente a média das 2.000 estimativas de cada indicador do grau de multicolinearidade; a , b e c os parâmetros dos modelos; ε_i é o erro associado a i -ésima observação; P é o platô; e , n é a estimativa do tamanho de amostra e o ponto de união entre as duas funções.

Para o MLRP, o n foi determinado considerando a junção entre as duas retas e para MQRP foi calculado conforme equação (7):

$$n = -\hat{b}/(2\hat{c}) \quad (7)$$

Sendo: \hat{a} , \hat{b} e \hat{c} as estimativas dos parâmetros do modelo.

Para a estimativa do grau de multicolinearidade correspondente ao n obtido pelo ajuste de regressão ($Y_{(n)}$), foi considerado a estimativa do platô (\hat{P}) para MLRP, enquanto que para MQRP, a estimativa foi obtida conforme (8):

$$Y_i = \hat{a} - \hat{b}^2/(4\hat{c}) \quad (8)$$

Sendo: \hat{a} , \hat{b} e \hat{c} as estimativas dos parâmetros do modelo MQRP.

As matrizes de correlação de Pearson, com as menores, maiores e as correlações médias para cada par de caracteres, obtidas com as observações da amostra mestre foram apresentadas graficamente (WEI; SIMKO, 2017). As estimativas do tamanho de amostra (n), do grau de multicolinearidade obtido no ajuste do modelo para n ($Y_{(n)}$) e do coeficiente de determinação ajustado (R^2_a) para cada indicador e modelo foram apresentadas em diagrama de caixa. Devido à inviabilidade de representação do ajuste do grau de multicolinearidade de cada indicador em função do tamanho de amostra para todos os casos e modelos (21 casos \times 3 indicadores \times 3 modelos = 189), foram apresentados graficamente somente os ajustes dos casos com menor e maior grau para cada indicador para

o MQRP. As análises estatísticas foram realizadas no software R (R Team Core, 2018).

3 RESULTADOS E DISCUSSÃO

Foi observada multicolinearidade moderada à forte ($100 < \text{número de condição (NC)} \leq 1.000$) (MONTGOMERY; PECK, 1982), quando estimada a partir de sete caracteres produtivos de centeio (Tabela 1) e amostra mestre ($n = 780$ plantas, provenientes de oito ensaios), com estimativas de 378,72, 0,0001 e 59,13 pelos indicadores NC, determinante da matriz de correlação (DET) e fator de inflação da variância (FIV), respectivamente (Tabela 2). Elevada multicolinearidade também foi observada em cinco das sete combinações, quando sete caracteres foram combinados seis a seis. Enquanto que nas outras duas combinações, foi observado fraca multicolinearidade ($\text{NC} \leq 100$) (MONTGOMERY; PECK, 1982).

A variação da magnitude do grau de multicolinearidade também foi observada entre os 21 casos, quando da combinação de sete caracteres cinco a cinco, mas com certa proporcionalidade entre o número de casos de fraca e moderado à forte multicolinearidade (Tabela 3). Foi constatada variabilidade das estimativas de multicolinearidade pelos três indicadores, com coeficientes de variação (CV) de 87,75; 98,18 e 81,11%, respectivamente, para os indicadores NC, DET e FIV (Tabela 3). Essa alta variabilidade entre os 21 casos de multicolinearidade, obtida pelos três indicadores, está associada com a ocorrência de duas classes de multicolinearidade. Dentre os 21 casos, 52,38% apresentaram fraco grau de multicolinearidade ($\text{NC} \leq 100$) e 47,62% de moderado à forte ($100 < \text{NC} \leq 1.000$) (MONTGOMERY; PECK, 1982).

Menores variabilidades de multicolinearidade foram observadas entre os casos de mesma classe. Considerando os onze casos de multicolinearidade fraca ($\text{NC} \leq 100$), o CV foi de 20,50, 40,75 e 22,36% para os indicadores NC, DET e FIV,

respectivamente. Enquanto que ao considerar os dez casos de multicolinearidade moderada à forte ($100 < NC \leq 1.000$), o CV foi de 18,51, 64,46 e 11,98% para NC, DET e FIV, respectivamente. Pelo fato da combinação de sete caracteres cinco a cinco, com base na amostra mestre, terem apresentado uma proporção entre os casos de fraca e de moderada à forte multicolinearidade, essa combinação entre os caracteres produtivos em centeio foi utilizada como referência para a determinação do tamanho de amostra.

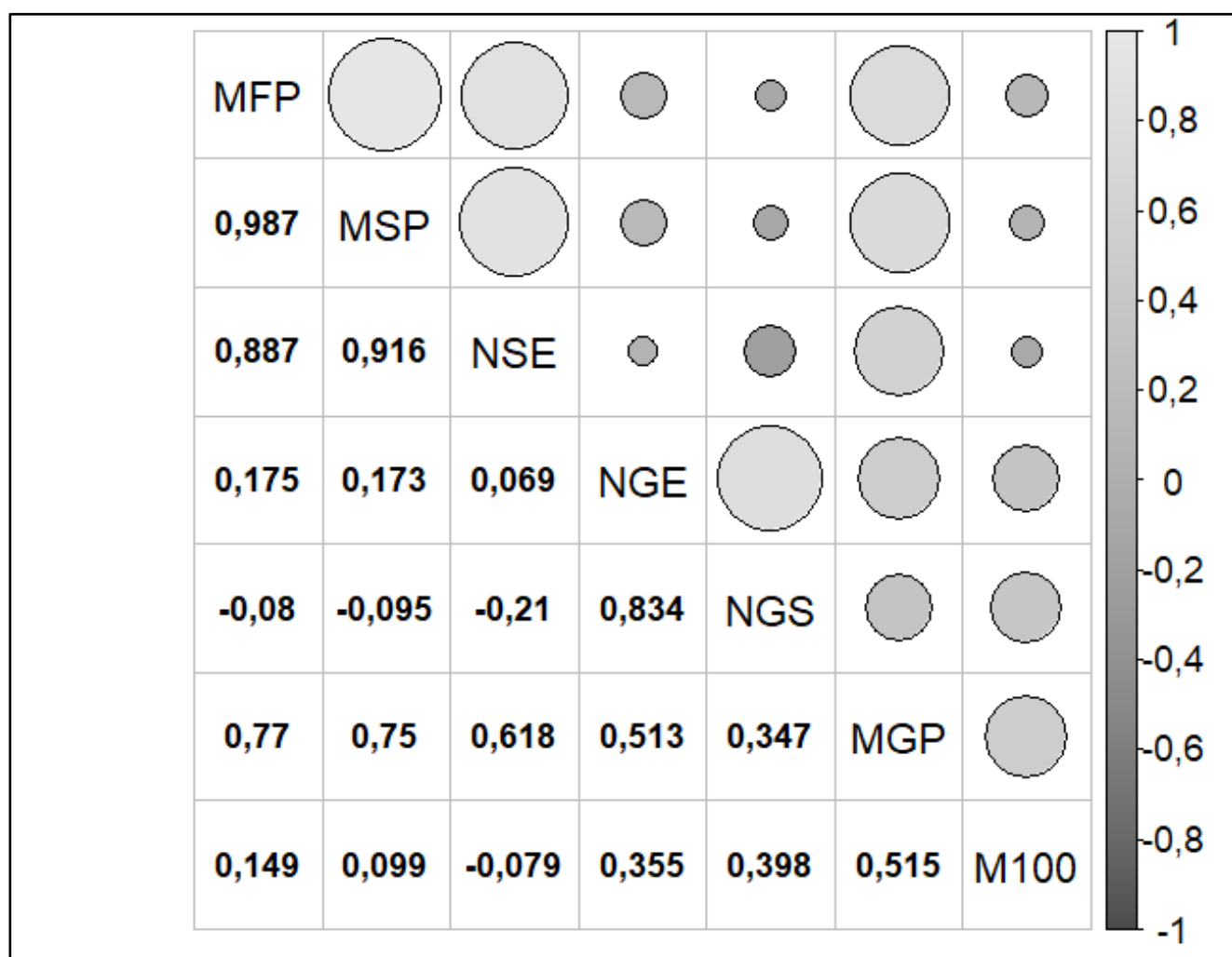
A magnitude da correlação entre variáveis pode ser um indicativo de multicolinearidade (HAIR et al., 2009). As correlações entre os caracteres massa fresca (MFP) e seca (MSP) de parte aérea da planta, número de grãos espiga⁻¹ (NGE), número de grãos espiguteta⁻¹ (NGS) e massa de grãos planta⁻¹ (MGP) podem ter contribuído para o aumento no grau de multicolinearidade. Esses caracteres estiveram presentes nos 10 casos classificados com multicolinearidade moderada à forte. Nestes 10 casos, o diagnóstico apresentou estimativas de $201,51 \leq NC \leq 356,07$; $0,0007 \leq DET \leq 0,0032$ e $45,33 \leq FIV \leq 59,02$. Já para os casos de fraca multicolinearidade, foram observadas estimativas de $23,84 \leq NC \leq 43,67$; $0,0088 \leq DET \leq 0,0405$ e $4,79 \leq FIV \leq 9,77$.

As correlações entre os caracteres produtivos obtidos a partir da amostra mestre estão apresentadas na Figura 1. Altas correlações podem ser verificadas entre os grupos de caracteres MFP, MSP e número de espiguetas espiga⁻¹ (NSE) ($r \geq 0,887$), entre NGE e NGS ($r = 0,834$) e entre MGP e todos os caracteres ($0,513 \leq r \leq 0,770$), exceto NGS ($r = 0,347$).

Os caracteres MFP, MSP e NSE apresentam baixa correlação com os caracteres NGE, NGS e massa média de 100 grãos (M100) ($|0,069| \leq r \leq |0,175|$). Sabendo que a correlação pode ser um indicativo de multicolinearidade (HAIR et al., 2009) e ao observar a Tabela 3 (caracteres combinados que compuseram cada caso) e as correlações entre os caracteres de centeio (Figura 1), fica evidenciado

que a variabilidade nas estimativas do grau de multicolinearidade foi em decorrência da magnitude da relação entre os caracteres.

Figura 1 - Matriz de correlação linear de Pearson em caracteres produtivos(1) de centeio, calculados com base na amostra mestre (n = 780 plantas), avaliados em oito ensaio de uniformidade de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS



(1)MFP: massa fresca de parte aérea; MSP: massa seca de parte aérea da planta; NSE: número de espiguetas espiga⁻¹; NGE: número de grãos espiga⁻¹; NGS: número de grãos espiguetas⁻¹; MGP: massa de grãos planta⁻¹; M100: massa média de 100 grãos, em gramas.

Os dez casos classificados como de moderada à forte multicolinearidade, estimada a partir da amostra mestre, foram constituídos pelos caracteres MFP e MSP ($r = 0,987$). Também estão presentes os caracteres NGE e NGS, que são fortemente correlacionados entre si ($r = 0,834$) e o caractere MGP com MFP ($r =$

0,770) e MSP ($r = 0750$). Sendo assim, visando reduzir custos, não é aconselhado conduzir ensaios com caracteres produtivos de centeio nas mesmas combinações dos casos 1 ao 10 (Tabela 3). Conduzir experimentos contendo caracteres que possam causar problema devido à multicolinearidade, proporcionará custos desnecessários pela demanda de tempo, mão de obra e recurso financeiro, e a não utilização de todos os caracteres em análises multivariadas. Pois, como constatado por Toebe e Cargnelutti Filho (2013), o alto grau de multicolinearidade causa erros de estimação dos coeficientes de efeitos diretos e, conseqüentemente, resultados sem sentido biológico e sem interpretação em análise de trilha.

Foi verificada variabilidade nas estimativas de multicolinearidade em caracteres de centeio por meio três indicadores em 21 casos. Em estudo que verificou o relacionamento entre produtividades de grãos e componentes de produtividade e caracteres morfológicos, foram observadas estimativas de FIV entre 1,37 e 452 (NOURAEIN, 2019). Dentre os caracteres avaliados pelos autores e que apresentaram problemas devido à alta multicolinearidade, o número de espiguetas espiga⁻¹ apresentou FIV = 13. Vale ressaltar que o indicador FIV informa sobre a inflação da variância para cada caractere, e aqui nesse estudo, são apresentados somente os valores de maior valor em cada diagnóstico. Estudos de multicolinearidade em caracteres de centeio diagnosticados por NC e DET não foram encontrados. Em outras culturas, foram verificadas estimativas de distintas magnitudes, a depender dos caracteres avaliados. Estimativas de multicolinearidade obtidas por DET, com valores entre 0,00002 e 0,02500 foram encontradas em diferentes tamanhos de amostra em tomate cereja (SARI et al., 2018) e estimativas de NC > 1.000 em caracteres de crotalária (TOEBE et al., 2017a) e de híbridos de milho (TOEBE et al., 2017c).

A estimativa do grau de multicolinearidade ($NC_{(n)}$, $DET_{(n)}$ e $FIV_{(n)}$) foi superior ao verificado em cada caso para a amostra mestre, independentemente do modelo

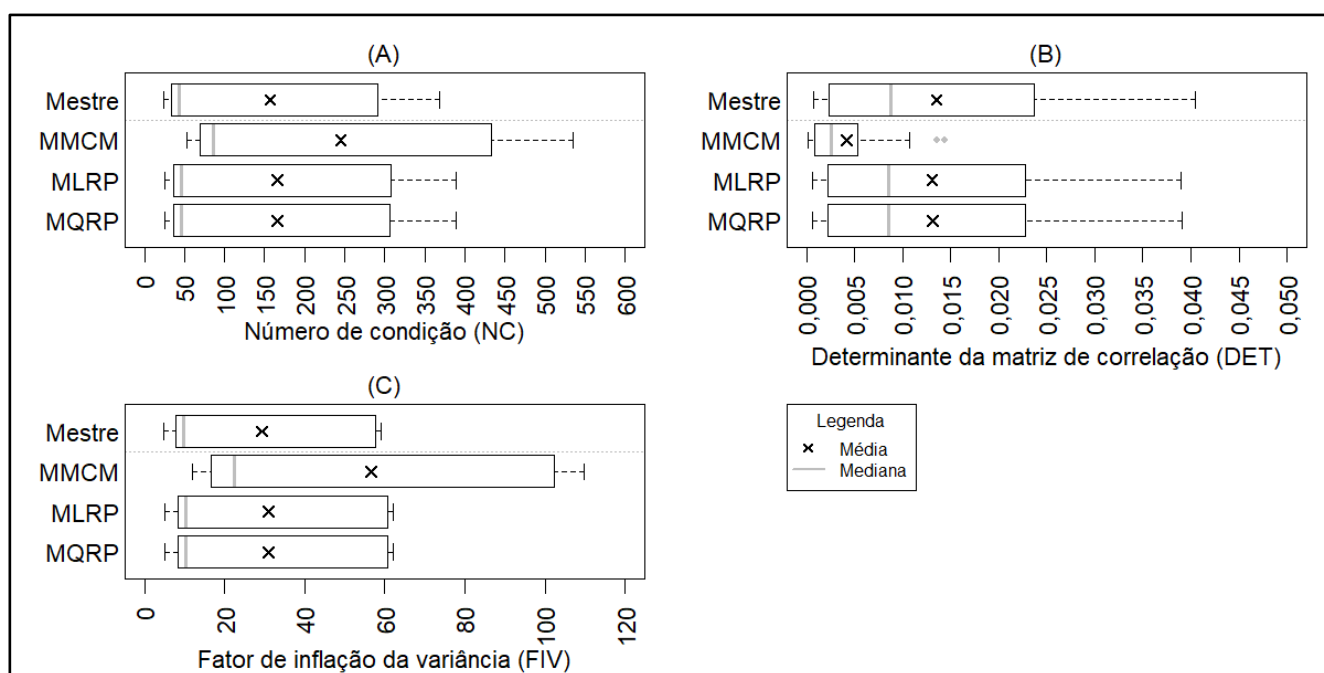
(Tabela 4 e Figura 2). No entanto, superestimativas do grau de multicolinearidade para os indicadores NC e FIV e subestimativas para o DET em relação à amostra mestre foram de maiores magnitudes quando o ajuste realizado pelo MMCM. Este comportamento foi verificado em todos os casos. Por exemplo, os valores médios de NC, DET e FIV na amostra mestre foram de 157,08, 0,0135 e 29,41, respectivamente.

Tabela 4 – Mínimo (Mín), Média (M), máximo (Máx), desvio-padrão (DP) e coeficiente de variação (CV) do grau de multicolinearidade da amostra mestre (n = 780 plantas) obtido por três indicadores e de coeficiente de determinação ajustado (R^2_a), tamanho de amostra (n) e estimativas de multicolinearidade ajustadas por três modelos, em caracteres produtivos de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS

| | Coeficiente de determinação ajustado (R^2_a) | | | | | Tamanho de amostra (n) | | | | | Estimativa do grau de multicolinearidade | | | | |
|---------------------|--|------|------|------|-------|------------------------|-----|-----|-------|--------|--|--------|--------|--------|--------|
| | Mín | M | Máx | DP | CV(%) | Mín | M | Máx | DP | CV(%) | Mín | M | Máx | DP | CV(%) |
| | Número de condição (NC) | | | | | | | | | | | | | | |
| Mestre | - | - | - | - | - | - | - | - | - | - | 23,84 | 157,08 | 367,97 | 137,84 | 87,75 |
| MMCM ⁽¹⁾ | 0,60 | 0,66 | 0,75 | 0,04 | 6,79 | 5 | 34 | 75 | 27,13 | 80,89 | 53,06 | 245,14 | 535,17 | 195,59 | 79,79 |
| MLRP | 0,85 | 0,86 | 0,88 | 0,01 | 0,99 | 66 | 79 | 101 | 9,78 | 12,43 | 25,41 | 166,07 | 388,85 | 145,43 | 87,57 |
| MQRP | 0,89 | 0,90 | 0,91 | 0,01 | 0,87 | 82 | 101 | 131 | 13,78 | 13,71 | 25,39 | 165,94 | 388,60 | 145,32 | 87,57 |
| | Determinante da matriz de correlação (DET) | | | | | | | | | | | | | | |
| Mestre | - | - | - | - | - | - | - | - | - | - | 0,0007 | 0,0135 | 0,0405 | 0,0328 | 98,02 |
| MMCM | 0,64 | 0,79 | 0,89 | 0,06 | 8,16 | < 1 | < 1 | < 1 | 0,00 | 103,99 | 0,0002 | 0,0042 | 0,0144 | 0,0045 | 106,99 |
| MLRP | 0,83 | 0,87 | 0,90 | 0,02 | 1,90 | 111 | 185 | 282 | 48,82 | 26,39 | 0,0007 | 0,0131 | 0,0390 | 0,0129 | 98,23 |
| MQRP | 0,86 | 0,91 | 0,94 | 0,02 | 1,96 | 153 | 258 | 387 | 70,11 | 27,16 | 0,0007 | 0,0131 | 0,0391 | 0,0129 | 98,23 |
| | Fator de inflação da variância (FIV) | | | | | | | | | | | | | | |
| Mestre | - | - | - | - | - | - | - | - | - | - | 4,79 | 29,41 | 59,02 | 23,85 | 81,11 |
| MMCM | 0,60 | 0,65 | 0,74 | 0,04 | 6,05 | 1 | 7 | 14 | 5,63 | 81,18 | 11,98 | 56,53 | 109,72 | 43,08 | 76,22 |
| MLRP | 0,85 | 0,86 | 0,88 | 0,01 | 0,89 | 66 | 80 | 101 | 9,99 | 12,50 | 5,15 | 30,95 | 62,07 | 25,11 | 81,12 |
| MQRP | 0,88 | 0,90 | 0,91 | 0,01 | 0,78 | 82 | 102 | 133 | 13,82 | 13,53 | 5,14 | 30,93 | 62,02 | 25,09 | 81,12 |

⁽¹⁾MMCM: Método da máxima curvatura modificado; MLRP: Modelo linear segmentado com resposta platô; MQRP: Modelo quadrático segmentado com resposta platô

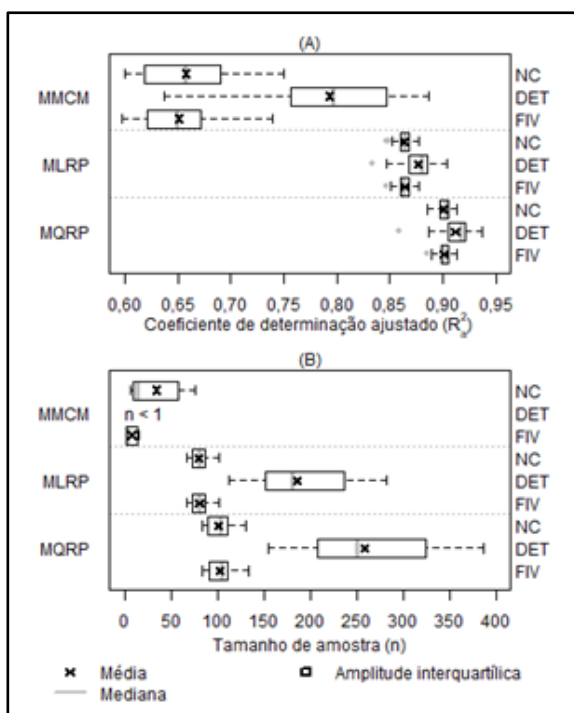
Figura 2 - Multicolinearidade em caracteres produtivos de centeio estimada pelo número de condição (A), determinante da matriz de correlação (B) e fator de inflação da variância (C) para a amostra mestre (n = 780 plantas) e estimativas obtidas com ajuste de três modelos não-lineares - método da máxima curvatura modificado (MMCM), modelo linear segmentado com resposta platô (MLRP) e modelo quadrático com resposta platô (MQRP), em caracteres produtivos de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS



Os valores médios estimados pelos modelos para o indicador NC foram superestimados em 56,06, 5,72 e 5,64% e em 92,19, 5,23 e 5,15% para FIV por meio do ajuste por MMCM, MLRP e MQRP, respectivamente. Enquanto que para o indicador DET, as estimativas para esses mesmos modelos foram subestimadas em 68,89, 3,38 e 3,24%, respectivamente.

Os três modelos apresentaram variabilidade entre os indicadores quanto à explicação do grau de multicolinearidade em função do tamanho de amostra, com coeficientes de determinação ajustado (R^2_a) superiores a 0,60 (Figura 3). Dentre os três modelos, o método da máxima curvatura modificado apresentou os menores valores de R^2_a e a maior variabilidade com $0,60 \leq R^2_a \leq 0,89$.

Figura 3 - Coeficiente de determinação ajustado (A) e tamanho de amostra (B) para os indicadores número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV), estimados pelo método da máxima curvatura modificado (MMCM), modelo linear segmentado com resposta platô (MLRP), modelo quadrático com resposta platô (MQRP) e em caracteres produtivos avaliados em oito ensaios de uniformidade de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS.



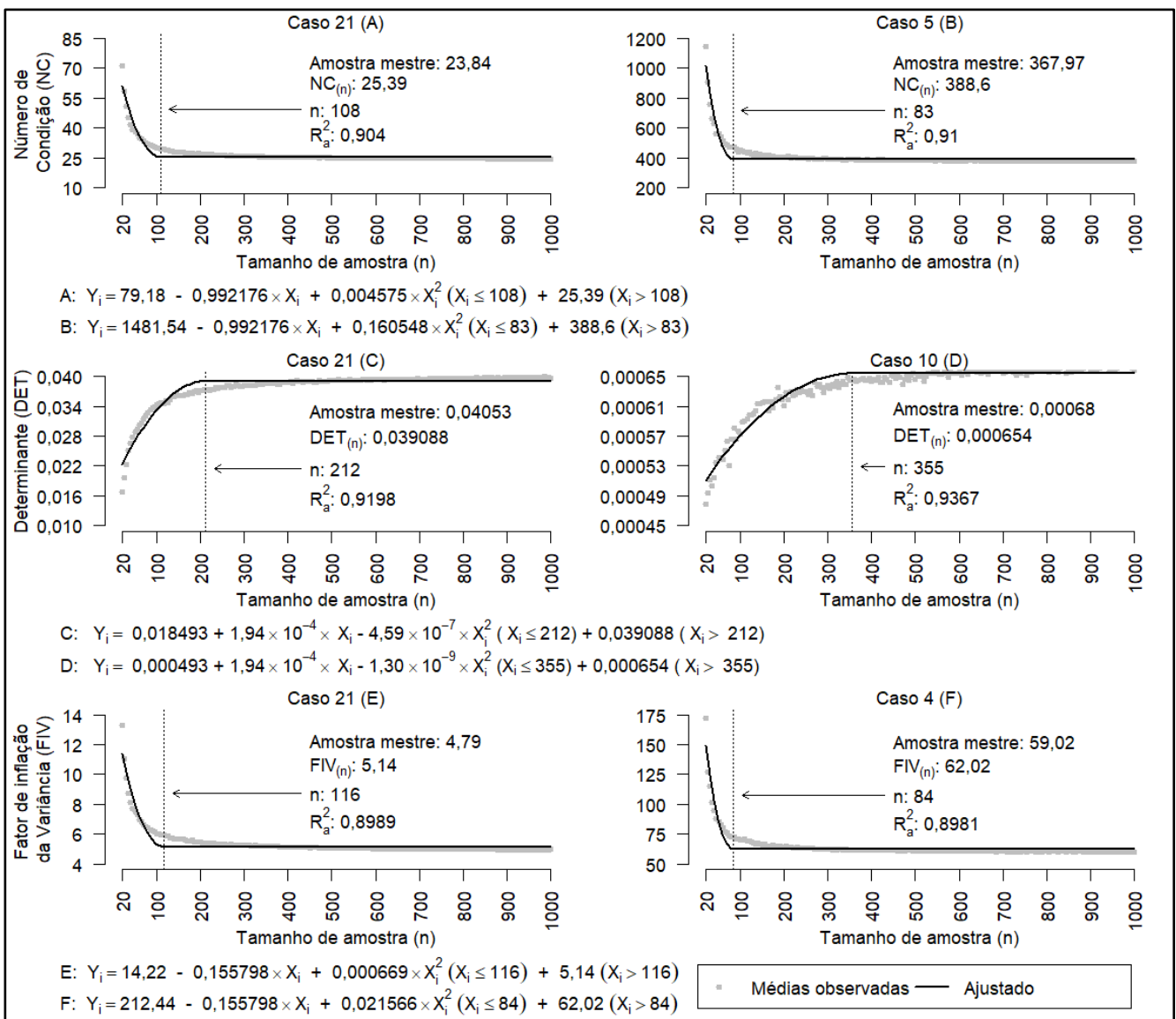
Havendo interesse em realizar a análise de regressão do grau de multicolinearidade em função do tamanho de amostra, recomenda-se utilizar tanto o modelo linear com resposta platô (MLRP), quanto o modelo quadrático com resposta platô (MQRP). Estes dois modelos apresentaram $0,83 \leq R^2_a \leq 0,94$, com R^2_a médio maior ao ajustar o grau de multicolinearidade em função do tamanho de amostra por meio de MQRP.

Considerando os valores de R^2_a superiores a 0,86 e com estimativas do grau de multicolinearidade de 3,24 a 5,64% superiores às encontradas na amostra mestre, optou-se em utilizar o MQRP para a determinação do tamanho de amostra (n). O MLRP também poderia ser utilizado, pois no ajuste, apresentou R^2_a próximos aos valores para o MQRP. Para o DET, o MMCM apresentou tamanhos de amostra (n) não práticos ($n \leq 1$).

O n necessário para o diagnóstico do grau de multicolinearidade entre caracteres produtivos de centeio, obtido por meio do MQRP, apresentou variabilidade entre os 21 casos

(Tabela 4 e Figuras 3 e 4). Considerando os valores médios dos 21 casos, o tamanho de amostra estimado foi de 101, 258 e 102 para os indicadores NC, DET e FIV, respectivamente.

Figura 4 - Tamanho de amostra estimado pelo modelo quadrático segmentado com resposta platô (MQRP) pelos indicadores número de condição (NC), determinante da matriz de correlação (DET) e fator de inflação da variância (FIV), estimativa do grau de multicolinearidade da amostra mestre (n = 780 plantas), estimativa do grau obtida com o ajuste (NC_(n), DET_(n) e FIV_(n)) e coeficiente de determinação ajustado (R²_a), em casos de menor (A, C e E) e maior (B, D e F) grau de multicolinearidade, respectivamente, provenientes de caracteres produtivos avaliados em oito ensaios de uniformidade de centeio (*Secale cereale* L.), Safra 2016, Santa Maria, RS



Considerando o caso 21 (caso de menor grau de multicolinearidade), foi estimado o n de 108, 212 e 116 plantas para NC, DET e FIV, respectivamente (Figura 4). Para os indicadores NC e FIV, o n foi superior e para DET o n foi inferior ao tamanho de amostra médio. Já o caso de maior grau não foi o mesmo entre os indicadores e o n apresentou resposta inversa ao verificado no caso menor grau de multicolinearidade. Os maiores níveis de multicolinearidade diagnosticados por NC foram observados no caso 5 (NC = 367,97) com $n = 83$, DET no caso 10 (DET = 0,000654) com $n = 355$ e por FIV no caso 4 (FIV = 59,02) com estimativa de n de 84. Segundo estes resultados, seriam necessários menores n em situações de grau de moderado a forte e maiores n em situações de fraca multicolinearidade.

Estes resultados comprovam que a detecção do grau de multicolinearidade é dependente do tamanho de amostra. A divergência no tamanho de amostra entre os casos envolvendo caracteres que resultaram em multicolinearidade de grau fraco e moderado a forte está relacionada com o grau de correlação entre os caracteres. Para Hair et al. (2009), a magnitude da correlação linear pode servir como um indicativo de presença de multicolinearidade entre os caracteres independentes.

Não foram encontrados trabalhos de tamanho de amostra para o diagnóstico do grau de multicolinearidade com a cultura do centeio para comparar com o n determinado por MQRP neste estudo. Já em estudos a partir de banco de dados de milho e tomate cereja, foram realizadas observações sobre o tamanho de amostra. Nestes há indicações de que tamanhos de amostra insuficientes podem estimar incorretamente o grau de multicolinearidade (OLIVOTO et al., 2017; SARI et al., 2018). Em estudo realizado com a cultura de tomate cereja, seriam necessários tamanhos de amostra superiores a 45 plantas, com 5% de probabilidade de erro, para estimar o indicador DET (SARI et al., 2018). Os autores Sari et al. (2018) destacam que em situações de tamanho de amostra superiores a 135 plantas, não haveria interferência da amostra no diagnóstico do grau de multicolinearidade.

O grau de multicolinearidade foi determinado utilizando tamanhos de amostra distintos. Em centeio, foram utilizados 64 acessos de todo o mundo em duas repetições (NOURAEIN, 2019). Em outras culturas, foram utilizadas 54 parcelas em híbridos transgênicos (ALVES et al., 2016) e 108 parcelas de híbridos de ciclo precoce (ALVES et al., 2017), 96 parcelas de meio-irmãos de couve (AZEVEDO et al., 2016), 66 parcelas de milho de ciclo superprecoce (ALVES et al., 2017) e 32 parcelas em girassol (FOLLMANN et al., 2019). Em crotalaria foram

utilizadas 100 plantas (TOEBE et al., 2017a) e em estudo do impacto do tamanho da amostra sobre o diagnóstico da multicolinearidade em análise de trilha em tomate cereja (347 plantas) (SARI et al., 2018).

Os resultados obtidos ao estimar o tamanho de amostra em caracteres produtivos de centeio e a variabilidade entre os cenários demonstram que não somente o tamanho de amostra possui influência na determinação do grau de multicolinearidade. O pesquisador também deverá planejar o experimento para a necessidade de realizar o diagnóstico de multicolinearidade, de acordo com a escolha de indicador e a magnitude das relações entre os caracteres. Podendo assim, utilizar o tamanho de amostra adequado em função do grau de multicolinearidade esperado para o conjunto de caracteres.

Independentemente do grau de multicolinearidade, o tamanho de amostra deve ser maior quando for diagnosticado pelo indicador DET ($n = 258$). No diagnóstico utilizando os indicadores NC e FIV, tamanhos de amostra compostas por 101 e 102 plantas podem ser utilizados, respectivamente. Havendo estudos em que foi realizado o diagnóstico do grau de multicolinearidade, deve-se utilizar tamanhos de amostra ou número de plantas maiores caso seja verificado que o conjunto de caracteres apresentou fraco grau de multicolinearidade ao utilizar o NC e FIV e moderado a forte grau quando da utilização de DET.

Pouco conhecimento científico de relacionamento entre caracteres e de dimensionamento amostral se dispõem para a cultura do centeio. Sugere-se que outros trabalhos sejam realizados para o diagnóstico do grau de multicolinearidade para outros caracteres, bem como, a determinação do tamanho de amostra e a utilização de outros modelos.

4 CONCLUSÕES

O tamanho de amostra não varia entre os indicadores número de condição (NC) e fator de inflação de variância (FIV), sendo necessário maior número de plantas quando o diagnóstico for realizado pelo determinante da matriz de correlação (DET). São necessários tamanhos de amostra distintos, para um mesmo indicador, quando o grau de multicolinearidade for de fraco ou de moderado a forte grau. Independentemente do grau de multicolinearidade entre os caracteres produtivos de centeio, há necessidade de tamanhos

superiores a 101, 258 e 102 plantas quando a detecção for realizada pelo número de condição, determinante da matriz de correlação ou fator de inflação da variância, respectivamente.

REFERÊNCIAS

ALVARES CA, STAPE JL, SENTELHAS PC, de MORAES GONÇALVES JL, SPAROVEK G. Köppen's climate classification map for Brazil. **Meteorol. Zeitschrift**. 2013;22(6):711-28.

ALVES BM, CARGNELUTTI FILHO A, BURIN C. Multicollinearity in canonical correlation analysis in maize. **Genet. Mol. Res**. 2017;16(1):1-14. <https://doi.org/10.4238/gmr16019546>.

ALVES BM, CARGNELUTTI FILHO A, BURIN C, TOEBE M. Correlações canônicas entre caracteres agronômicos e nutricionais proteicos e energéticos em genótipos de milho. **Rev. Bras. Milho e Sorgo**. 2016;15(2):171-85. <https://doi.org/10.18512/1980-6477/rbms.v15n2p171-185>.

AZEVEDO AM, SEUS R, GOMES CL, FREITAS EM de, CANDIDO DM, SILVA DJH da, et al. Correlações genotípicas e análise de trilha em famílias de meios-irmãos de couve de folhas. **Pesqui. Agropecuária Bras**. 2016;51(1):35-44. <https://doi.org/10.1590/S0100-204X2016000100005>.

BAIER AC. Centeio. Passo Fundo - RS: EMBRAPA Trigo; 1994.
<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/164511/1/FL-06193.pdf>.

CARGNELUTTI FILHO A, ALVES BM, SANTOS GO, WARTHA CA, KLEINPAUL JA, SILVEIRA DL. Sample size to estimate the mean and median of traits in flax. **Brazilian J. Agric. Sci**. 2018a;13(1):1-7. <https://doi.org/10.5039/agraria.v13i1a5492>.

CARGNELUTTI FILHO A, ALVES BM, TOEBE M, BURIN C, WARTHA CA. Sample size to estimate the mean of traits in jack bean. **Brazilian J. Agric. Sci**. 2018b;13(1):1-7. <https://doi.org/10.5039/agraria.v13i1a5505>.

CARGNELUTTI FILHO A, WARTHA CA, KLEINPAUL JA, NEU IMM, SILVEIRA DL. Sample size to estimate the mean and median of traits in canola. **J. Agric. Sci**. 2018c;10(11):123-37. <https://doi.org/10.5539/jas.v10n11p123>.

FIELD A. **Descobrimos a estatística utilizando o SPSS**. 2 ed. Porto Alegre: Artmed; 2009.

FARAWAY J. **faraway: Functions and Datasets for Books by Julian Faraway**. R package version 1.0.7, 2016. <https://CRAN.R-project.org/package=faraway>.

FOLLMANN DN, CARGNELUTTI FILHO A, SANTOS MS dos, COSTA VO, PLAUTZ ÉN, SCOPEL JVF, et al. Correlations and path analysis in sunflower grown at lower elevations. **J. Agric. Sci**. 2019;11(2):445-53. <https://doi.org/10.5539/jas.v11n2p445>.

GUJARATI DN, PORTER DC. **Econometria básica**. 5 ed. Porto Alegre: AMGH Editora Ltda; 2011.

HAIR JF, BLANCK WC, BABIN BJ, ANDERSON RE, TATHAN RL. **Análise multivariada de dados**. 6 ed. Porto Alegre: Bookman; 2009.

MEIER VD, LESSMAN KJ. Estimation of optimum field plot shape and size for testing yield in *Crambe abyssinica* Hochst. **Crop Science**. 1971;11(5):648-650.

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

NOURAEIN M. Elucidating seed yield and components in rye (*Secale cereale* L.) using path and correlation analyses. **Genet. Resour. Crop Evol.** 2019;66(7), 1533–1542. <https://doi.org/10.1007/s10722-019-00813-6>.

OLIVOTO T, NARDINO M, CARVALHO IR, FOLLMANN DN, FERRARI M, PELEGRIN AJ de, et al. Optimal sample size and data arrangement method in estimating correlation matrices with lesser collinearity: A statistical focus in maize breeding. **African J. Agric. Res.** 2017;12(2):93-103. <https://doi.org/10.5897/AJAR2016.11799>.

R TEAM CORE. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing; 2018. <https://www.r-project.org>.

SANTOS HG dos, JACOMINE PKT, ANJOS LHC dos, OLIVEIRA VÁ de, LUMBRERAS JF, COELHO MR, et al. **Sistema brasileiro de classificação de solos**. 5 ed. Brasília: Embrapa Solos; 2018.

SARI BG, LÚCIO AD, OLIVOTO T, KRYSCZUN DK, TISCHLER AL, DREBES L. Interference of sample size on multicollinearity diagnosis in path analysis. **Pesqui. Agropecuária Bras.** 2018;53(6):769-73. <https://doi.org/10.1590/s0100-204x2018000600014>.

TOEBE M, CARGNELUTTI FILHO A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesqui. Agropecuária Bras.** 2013;466-77. <https://doi.org/10.1590/S0100-204X2013000500002>.

TOEBE M, BANDEIRA CT, FORTES SKG, CARVALHO JO de, TARTAGLIA F de L, TAMBARA AL, et al. Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*. **Bragantia**. 2017a;76(1):45-53. <https://doi.org/10.1590/1678-4499.653>.

TOEBE M, CARGNELUTTI FILHO A, STORCK L, LÚCIO AD. Sample size for estimation of direct effects in path analysis of corn. **Genet. Mol. Res.** 2017b;16(2):1-23. <https://doi.org/10.4238/gmr16029523>.

TOEBE M, CARGNELUTTI FILHO A, STORCK L, LÚCIO, AD. Direct effects on scenarios and types of path analyses in corn hybrids. **Genet. Mol. Res.** 2017c;16(1);1-15. <https://doi.org/10.4238/gmr16019529>.

TOEBE M, MACHADO LN, TARTAGLIA FL, de CARVALHO JO, BANDEIRA CT, CARGNELUTTI FILHO A. Sample size for estimating mean and coefficient of variation in species of crotalarias. **An. Acad. Bras. Cienc.** 2018;90(2):1705-15. <https://doi.org/10.1590/0001-3765201820170813>.

WEI T, SIMKO, V. R **package "corrplot": Visualization of a Correlation Matrix**. Version 0.84, 2017. <https://github.com/taiyun/corrplot>.

CONTRIBUIÇÕES DE AUTORIA

1 – Ismael Mario Márcio Neu

Discente de doutorado em Agronomia, Universidade Federal de Santa Maria.

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

Contribuição: concepção, investigação, metodologia e análises estatísticas, escrita e edição.

2 – Alberto Cargnelutti Filho

Doutor em Agronomia.

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

Contribuição: concepção, investigação, metodologia, revisão e edição, supervisão.

3 – Cláudia Marques de Bem

Doutora em Agronomia.

<https://orcid.org/0000-0002-6326-8720> – claudia_debem@hotmail.com

Contribuição: investigação, revisão e edição.

4 – Jéssica Andiana 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.

5 – Cirineu Tolfo Bandeira

Mestre em Agronomia.

<https://orcid.org/0000-0001-5019-5959> – cirineutolfobandeira@gmail.com

Contribuição: investigação, revisão e edição.

COMO CITAR ESTE ARTIGO

NEU, I. M. M.; CARGNELUTTI FILHO, A; BEM, C. M.; KLEINPAUL, J. A.; BANDEIRA, C. T. Tamanho de amostra para avaliação do grau de multicolinearidade em caracteres produtivos de centeio. *Ciência e Natura*, Santa Maria, v. 43, e38, p. 1 – 22, 2021. Disponível em: <https://doi.org/10.5902/2179460X41667>. Acesso em: dia, mês (abreviado), ano.