

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

Ismael Mario Márcio Neu

**NÚMEROS DE ENSAIOS, OBSERVAÇÕES E VARIÁVEIS NA
CORRELAÇÃO CANÔNICA EM CENTEIO**

Santa Maria, RS, Brasil
2022

Ismael Mario Márcio Neu

**NÚMEROS DE ENSAIOS, OBSERVAÇÕES E VARIÁVEIS NA
CORRELAÇÃO CANÔNICA EM CENTEIO**

Tese apresentada ao Curso de Pós-Graduação em
Agronomia, da Universidade Federal de Santa Maria
(UFSM, RS), como requisito parcial para a obtenção
do título de **Doutor em Agronomia**.

Orientador: Prof. Dr. Alberto Cargnelutti Filho

Santa Maria, RS, Brasil
2022

O presente estudo foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Neu, Ismael Mario Márcio

Números de ensaios, observações e variáveis na correlação canônica em centeio / Ismael Mario Márcio Neu.- 2022.

136 p.; 30 cm

Orientador: Alberto Cargnelutti Filho

Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Agronomia, RS, 2022

1. Análise multivariada 2. Dimensionamento experimental 3. Multicolinearidade 4. Análise de correlação canônica 5. Cereal de inverno I. Cargnelutti Filho, Alberto II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

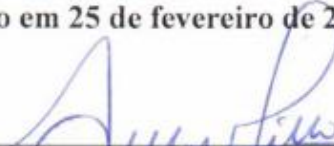
Declaro, ISMAEL MARIO MÁRCIO NEU, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Tese) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Ismael Mario Márcio Neu

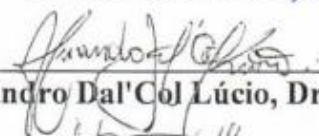
**NÚMEROS DE ENSAIOS, OBSERVAÇÕES E VARIÁVEIS NA
CORRELAÇÃO CANÔNICA EM CENTEIO**

Tese apresentada ao Curso de Pós-Graduação em
Agronomia, da Universidade Federal de Santa Maria
(UFSM, RS), como requisito parcial para a obtenção
do título de **Doutor em Agronomia**.

Aprovado em 25 de fevereiro de 2022:



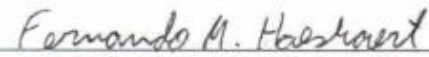
Alberto Cargnelatti Filho, Dr. (UFSM)
(Presidente/Orientador)



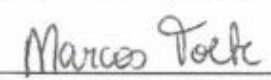
Alessandro Dal'Col Lúcio, Dr. (UFSM)



Betânia Brum de Bortolli, Dra. (UTFPR)



Fernando Machado Haesbaert, Dr. (UFT)



Marcos Toebe, Dr. (UFSM)

Santa Maria, RS, Brasil
2022

AGRADECIMENTOS

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

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela concessão da bolsa de doutorado.

A minha família, em especial aos meus pais, Ildemar e Claudete, irmãs Adriana e Júlia, pelo constante incentivo, apoio, amor, carinho e por tudo que fizeram e que são na minha vida.

Ao meu orientador Alberto Cargnelutti Filho, pela orientação deste trabalho e pela preocupação com o meu crescimento pessoal.

Aos professores do Programa de Pós-Graduação em Agronomia, pelos ensinamentos para o meu aperfeiçoamento e aos professores membros da banca, pelas sugestões de melhoria ao trabalho.

Aos colegas do Setor de Experimentação Agrícola pela parceria e amizade, que além de colegas e amigos, assumiram o papel de segunda família.

Ao Departamento de Fitotecnia e aos funcionários, pela estrutura para condução dos experimentos e apoio nas atividades de campo e administrativas.

Aos demais familiares, professores, pesquisadores e amigos que me apoiaram e contribuíram, direta ou indiretamente, na realização desse trabalho.

A todos, meu sincero,

Muito obrigado!

RESUMO

NÚMEROS DE ENSAIOS, OBSERVAÇÕES E VARIÁVEIS NA CORRELAÇÃO CANÔNICA EM CENTEIO

AUTOR: Ismael Mario Márcio Neu

ORIENTADOR: Prof. Dr. Alberto Cargnelutti Filho

As técnicas multivariadas são ferramentas importantes de análise na compreensão do objeto de estudo. Para maior confiabilidade nas interpretações de modelos é necessário o correto dimensionamento amostral. Pouco tem sido explorado sobre o número de medições ou ensaios necessários no diagnóstico de multicolinearidade e na correlação do primeiro par canônico e a relação entre o número de observações e variáveis para a estimativa da correlação no primeiro par canônico. Assim, os objetivos deste estudo foram determinar: o número de ensaios necessário para estimar o grau de multicolinearidade pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*) e a estimativa da correlação do primeiro par canônico; e a relação entre o número de observações e de variáveis para a estimativa da correlação canônica. Foram conduzidos oito ensaios de uniformidade com duas cultivares de centeio, com a avaliação de caracteres morfológicos e produtivos. Para o estudo do número de ensaios para o diagnóstico do grau de multicolinearidade em cada grupo de variáveis, foram planejados casos – diferentes combinações de caracteres – e realizado o diagnóstico de multicolinearidade pelos indicadores *NC* e *FIV*. A análise de repetibilidade e o número de ensaios para estimar a multicolinearidade foi realizada por cinco métodos em cada caso, cultivar e grupo de variáveis. Na determinação do número de ensaio para a estimação da correlação do primeiro par canônico, foram considerados cenários, constituídos por casos com o mesmo número de caracteres combinados em cada grupo de variáveis. Para cada cenário e cultivar, foi realizada a análise de repetibilidade por cinco métodos: análise de variância, componentes principais e análise estrutural, com base nas matrizes de correlação e de variâncias e covariâncias. Em seguida, foi determinado o número de ensaios para diferentes níveis de precisão. Para estudar a relação entre o número de observações e variáveis ($n:p$) para estimar a correlação do primeiro par canônico, 1.000 amostras com distribuição normal multivariada foram simuladas em 100 tamanhos de amostra ($n:p = 1, 2, 3, \dots, 100$), três épocas de semeadura e duas cultivares. Em seguida, com as estimativas de médias em cada tamanho de amostra planejado, foi determinada a relação $n:p$ por meio de dois modelos de regressão segmentados com platô. Há a necessidade de três e 16 ensaios para o diagnóstico de multicolinearidade e quatro e oito ensaios para estimar a correlação do primeiro par canônico nas cultivares BRS Progresso e Temprano, respectivamente, com precisão mínima de 95%. Enquanto que, em análise de correlação canônica, é recomendado, a utilização de no mínimo de 27 observações para cada variável.

Palavras-chave: *Secale cereale* L. Análise multivariada. Multicolinearidade. Dimensionamento amostral.

ABSTRACT

NUMBERS OF TRIALS, OBSERVATIONS AND VARIABLES IN THE CANONICAL CORRELATION IN RYE

AUTHOR: Ismael Mario Márcio Neu
ADVISOR: Prof. Dr. Alberto Cargnelutti Filho

Multivariate techniques are important analysis tools for understanding the object of study. For greater reliability in the interpretations of models, the correct sample size is necessary. Little has been explored about the number of measurements or tests needed in the diagnosis of multicollinearity and the correlation of the first canonical pair and the relationship between the number of observations and variables for estimating the correlation in the first canonical pair. Thus, the objectives of this study were to determine: the number of trials necessary to estimate the degree of multicollinearity by the condition number (*NC*) and variance inflation factor (*FIV*) indicators and the estimate of the correlation of the first canonical pair; and the relationship between the number of observations and variables for estimating the canonical correlation. Eight uniformity tests were conducted with two rye cultivars, with the evaluation of morphological and productive characters. To study the number of tests for the diagnosis of the degree of multicollinearity in each group of variables, cases were planned – different combinations of characters – and the diagnosis of multicollinearity was performed using the *NC* and *FIV* indicators. The repeatability analysis and the number of trials to estimate multicollinearity were performed by five methods in each case, cultivar and group of variables. In determining the number of trials for estimating the correlation of the first canonical pair, scenarios were considered, consisting of cases with the same number of characters combined in each group of variables. For each scenario and cultivar, repeatability analysis was performed by five methods: analysis of variance, principal components and structural analysis, based on correlation and variance and covariance matrices. Then, the number of trials for different levels of accuracy was determined. To study the relationship between the number of observations and variables (*n:p*) to estimate the correlation of the first canonical pair, 1,000 samples with multivariate normal distribution were simulated in 100 sample sizes (*n:p* = 1, 2, 3, . . ., 100), three sowing dates and two cultivars. Then, with the averages' estimates in each planned sample size, was determined the *n:p* relationship through two segmented regression models with plateau. There is need of three and 16 trials for the diagnosis of multicollinearity and of four and eight trials for estimating the first pair correlation canonical in the cultivars BRS Progresso and Temprano, respectively, with a minimum accuracy of 95%. While, in canonical correlation analysis, it is recommended the use at least 27 observations for each variable.

Keywords: *Secale cereale* L. Multivariate analysis. Multicollinearity. Sample sizing.

LISTA DE FIGURAS

- Figura 1 – Croqui de localização e instalação de oito ensaios de uniformidade com duas cultivares de centeio (*Secale cereale* L.) em cinco épocas de semeadura, conduzidos em Santa Maria – RS, Safra 2016..... 50
- Figura 2 – Relação entre as estimativas de multicolinearidade obtidas pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*) em combinações de caracteres morfológicos e produtivos em cultivares de centeio (*Secale cereale* L.) conduzidos em diferentes épocas de semeadura em Santa Maria – RS, Safra 2016..... 81
- Figura 3 – Médias e diagrama ‘boxplot’ de correlação do primeiro par canônico obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em três épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016..... 105
- Figura 4 – Histogramas de densidades da correlação do primeiro par canônico obtidas em simulações em diferentes relações entre o número de observações (tamanho de amostra) e de variáveis em ensaio conduzido na primeira época com a cultivar BRS Progresso [A] e em três épocas e duas cultivares de centeio (*Secale cereale* L.) [B], ensaios conduzidos em Santa Maria – RS, Safra 2016..... 107
- Figura 5 – Amplitude do intervalo de confiança de 95% (AIC 95%) de estimativas de correlação do primeiro par canônico obtidas em simulações em diferentes relações entre o número de observações (tamanho de amostra) e de variáveis em ensaio conduzido em três épocas de semeadura e duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016..... 108
- Figura 6 – Médias e diagrama ‘boxplot’ do índice de redundância ($X | Y =$ fração total da variância do primeiro grupo contabilizada pelas variáveis canônicas do segundo grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016..... 109
- Figura 7 – Médias e diagrama ‘boxplot’ do índice de redundância ($Y | X =$ fração total da variância do segundo grupo contabilizada pelas variáveis canônicas do

primeiro grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016. 110

Figura 8 – Amplitude do intervalo de confiança de 95% (AIC_{95%}) de estimativas de índice de redundância ($X | Y =$ fração total da variância do primeiro grupo contabilizada pelas variáveis canônicas do segundo grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016. 111

Figura 9 – Médias e amplitude do intervalo de confiança de 95% (AIC 95%) das estimativas de correlação do primeiro par canônico, relação entre o número de observações e de variáveis (n:p) determinado pelo método de modelo quadrático com resposta em platô (MQRP), estimativa da correlação canônica estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura com a cultivar BRS Progresso (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016. 112

Figura 10 – Médias e amplitude do intervalo de confiança de 95% (AIC 95%) das estimativas de correlação do primeiro par canônico, relação entre o número de observações e de variáveis (n:p) determinado pelo método de modelo quadrático com resposta em platô (MQRP), estimativa da correlação canônica estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura com a cultivar Temprano (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016. 113

LISTA DE TABELAS

Tabela 1 – Métodos utilizados para determinação do coeficiente de repetibilidade em diferentes culturas	37
Tabela 2 – Datas de semeadura, de emergência, de florescimento e de colheita (avaliação), estande de plantas (EP = número de plantas m ⁻²) e número de dias entre a semeadura e a colheita (Ciclo) em duas cultivares de centeio (<i>Secale cereale</i> L.), em oito ensaios de uniformidade (E1, E2, E3, E4, E5, E6, E7 e E8), conduzidos em Santa Maria – RS, Safra 2016.....	50
Tabela 3 – Identificação (ID), sigla e descrição de caracteres definidos em cada grupo de variáveis (morfológicas e produtivas) em duas cultivares de centeio (<i>Secale cereale</i> L.), em oito ensaios de uniformidade (E1, E2, E3, E4, E5, E6, E7 e E8), conduzidos em Santa Maria – RS, Safra 2016.....	52
Tabela 4 – Demonstração de esquema de combinações (Comb.) de oito de caracteres morfológicos (X ¹) arranjados em combinações de 2, 3, 4, 5, 6 e 7 caracteres em diferentes casos de estudo com cultivares de centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.....	54
Tabela 5 – Demonstração de esquema de combinações (Comb.) de oito de caracteres produtivos (Y ¹) arranjados em combinações de 2, 3, 4, 5, 6 e 7 caracteres em diferentes casos de estudo com cultivares de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.....	55
Tabela 6 – Número de combinações (Comb.) de caracteres em cada caso, grupo de variáveis, ensaio, cultivar e total em cenários com os casos formados pelas combinações de caracteres do grupo de variáveis morfológicas (MORF) e produtivas (PROD) de centeio (<i>Secale cereale</i> L.), em ensaios conduzidos em Santa Maria – RS, Safra 2016.....	59
Tabela 7 – Médias e coeficiente de variação de estimativas em caracteres morfológicos avaliados em oito ensaios de uniformidade (E1, E2, ..., E8) na cultura do centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.....	66
Tabela 8 – Médias e coeficiente de variação de estimativas em caracteres produtivos avaliados em oito ensaios de uniformidade (E1, E2, ..., E8) na cultura do centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.....	68

Tabela 9 – Correlações entre pares de caracteres morfológicos ¹ de centeio conduzidos diferentes épocas de semeadura com cultivares de centeio (<i>Secale cereale</i> L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.....	70
Tabela 10 – Correlações entre pares de caracteres produtivos ¹ de centeio em ensaios conduzidos diferentes épocas de semeadura com cultivares centeio (<i>Secale cereale</i> L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.	71
Tabela 11 – Grau de multicolinearidade estimado pelos indicadores número de condição (<i>NC</i>) e fator de inflação da variância (<i>FIV</i>) em caracteres morfológicos e produtivos de duas cultivares de centeio (<i>Secale cereale</i> L.), considerando todos os oito e seis caracteres (eliminação de caracteres com $FIV_j > 10$), ensaios conduzidos em Santa Maria – RS, Safra 2016.....	74
Tabela 12 – Estimativa do número de condição (<i>NC</i>) e do maior valor do fator de inflação da variância (<i>FIV</i>) em 28 combinações (Comb.) no caso 2, com oito caracteres morfológicos combinados dois a dois ($C_{(8,2)}$) em ensaios de uniformidade ¹ (E1, E2, ..., E8) com cultivares de centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.	77
Tabela 13 – Estimativa do número de condição (<i>NC</i>) e do maior valor do fator de inflação da variância (<i>FIV</i>) em 28 combinações (Comb.) no caso 2, de oito caracteres produtivos combinados dois a dois ($C_{(8,2)}$) em ensaios de uniformidade ¹ (E1, E2, ..., E8) com cultivares de centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.	78
Tabela 14 – Estimativa do número de condição (<i>NC</i>) e do maior valor do fator de inflação da variância (<i>FIV</i>) em oito combinações (Comb.) no caso 7, de oito caracteres morfológicos e oito caracteres produtivos combinados sete a sete ($C_{(8,7)}$) em ensaios de uniformidade ¹ (E1, E2, ..., E8) com cultivares de centeio centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.	79
Tabela 15 – Estatísticas descritivas de mínimos (Mín), média (m), mediana (Med), máximo (Máx) e amplitude (Amp = Máx - Mín) de número de condição (<i>NC</i>) e fator de inflação da variância (<i>FIV</i>), e percentual de combinações com problemas de multicolinearidade (PCMD) em casos (C) de combinações, grupos de variáveis e cultivares de centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.	80

Tabela 16 – Coeficientes de repetibilidade (<i>rptb</i>) e de determinação (R^2) do grau de multicolinearidade obtido pelos indicadores número de condição (<i>NC</i>) e fator de inflação da variância (<i>FIV</i>), em diferentes métodos para diferentes grupos de caracteres ¹ combinados em dois grupos de variáveis e em duas cultivares de centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.	85
Tabela 17 – Número de medições ou de ensaios associado a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) para estimar a multicolinearidade pelo indicador número de condição (<i>NC</i>) para diferentes caracteres combinados em dois grupos de variáveis e em duas cultivares de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.....	88
Tabela 18 – Número de medições ou de ensaios associado a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) para estimar a multicolinearidade pelo indicador fator de inflação da variância (<i>FIV</i>) para diferentes caracteres combinados em dois grupos de variáveis e em duas cultivares de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.	89
Tabela 19 – Estimativa de correlações canônicas para o primeiro par canônico obtidas para o caso com a combinação de dois caracteres morfológicos (X) e dois caracteres produtivos (Y) – total de 784 combinações, em oito ensaios de uniformidade ¹ (E1, E2, ..., E8) em centeio (<i>Secale cereale</i> L.), conduzidos em Santa Maria – RS, Safra 2016.....	95
Tabela 20 – Mínimo, média, máximo, percentis ($P_{2,5\%}$ e $P_{97,5\%}$), amplitude do intervalo de confiança ($AIC_{95\%} = P_{97,5\%} - P_{2,5\%}$) e percentual de combinações com correlações canônicas significativas pelo teste Qui-Quadrado em nível de 5% de significância (PCCS), de estimativas obtidas em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicos e produtivos em cultivares de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.	96
Tabela 21 – Coeficientes de repetibilidade (<i>rptb</i>) e de determinação (R^2) para a estimativa da correlação canônica para o primeiro par canônico por diferentes métodos em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicos (MORF) e produtivos (PROD) em duas cultivares	

de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.	98
Tabela 22 – Número de ensaios associados a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) de estimativas da correlação canônica para o primeiro par por diferentes métodos em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicos (MORF) e produtivos (PROD) em duas cultivares de centeio (<i>Secale cereale</i> L.), Santa Maria – RS, Safra 2016.	100
Tabela 23 – Estimativas da relação entre o número de observações e de variáveis (n:p) determinados por dois métodos de regressão com resposta em platô, estimativa da correlação do primeiro par canônico estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura e duas cultivares de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.	114
Tabela 24 – Estimativas da relação entre o número de observações e de variáveis (n:p) determinados por dois métodos de regressão com resposta em platô, para a estimativa do índices de redundâncias em análise de correlação canônica, estimativa do índice para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura e duas cultivar de centeio (<i>Secale cereale</i> L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.	115

SUMÁRIO

1	INTRODUÇÃO.....	21
1.1	Contexto.....	21
1.2	Justificativa.....	24
1.3	Hipóteses.....	25
1.4	Objetivo geral e objetivos específicos.....	25
1.4.1	Objetivo geral:.....	25
1.4.2	Objetivos específicos:.....	26
2	REVISÃO BIBLIOGRÁFICA.....	27
2.1	Cultura do canteiro.....	27
2.2	Multicolinearidade.....	29
2.3	Indicadores de multicolinearidade.....	30
2.4	Análise de repetibilidade.....	32
2.4.1	Contextualização.....	32
2.4.2	Métodos de estimativa do coeficiente de repetibilidade.....	33
2.4.3	Estudos de repetibilidade.....	36
2.4.4	Precisão da estimativa do coeficiente de repetibilidade.....	38
2.5	Número de medições ou ensaios.....	38
2.5.1	Contextualização.....	38
2.5.2	Método de estimação do número de medições ou ensaios.....	39
2.5.3	Estudos de número de medições ou ensaios.....	39
2.6	Análise de correlação canônica.....	40
2.7	Relação entre tamanho de amostra e número de variáveis.....	44
2.8	Modelos de regressão.....	46
2.8.1	Modelo linear segmentado com resposta platô.....	47
2.8.2	Modelo quadrático segmentado com resposta platô.....	47
3	MATERIAL E MÉTODOS.....	49
3.1	Instalação do experimento e coleta de dados.....	49
3.2	Análise de dados.....	51
3.2.1	Caracteres morfológicos e produtivos.....	51
3.2.2	Correlações entre caracteres.....	52
3.2.3	Combinações de caracteres.....	53
3.2.4	Diagnóstico de multicolinearidade.....	55

3.3	Estudo 1 – Número de medições para o diagnóstico de multicolinearidade.....	56
3.3.1	Casos com combinações de caracteres morfológicos e produtivos	56
3.3.2	Estimativa do coeficiente de repetibilidade	57
3.3.3	Número de medições ou ensaios.....	58
3.4	Estudo 2 – Número de ensaios para a estimação da correlação canônica.....	58
3.4.1	Cenários com caracteres combinados	58
3.4.2	Estimativa da correlação canônica	60
3.4.3	Estimativa do coeficiente de repetibilidade	60
3.4.4	Determinação do número de medições ou de ensaios	61
3.5	Estudo 3 – Relação entre o número de observações e de variáveis em correlação canônica.....	61
3.5.1	Análise de correlação canônica e reamostragem.....	61
3.5.2	Determinação da relação entre o número de observações e de variáveis	62
3.6	Aplicativos e softwares estatísticos.....	63
4	RESULTADOS E DISCUSSÃO	65
4.1	Estatística descritiva de caracteres morfológicos e produtivos.....	65
4.1.1	Caracteres morfológicos	65
4.1.2	Caracteres produtivos.....	67
4.1.3	Considerações gerais sobre os caracteres em centeio	68
4.2	Correlação linear entre caracteres	70
4.3	Multicolinearidade em caracteres de centeio	73
4.3.1	Grau de multicolinearidade em grupos de variáveis	73
4.3.2	Grau de multicolinearidade estimado em combinações de caracteres.....	75
4.3.3	Considerações sobre a multicolinearidade em caracteres de centeio	83
4.4	Repetibilidade e número de ensaios para diagnóstico de multicolinearidade.....	84
4.4.1	Coeficiente de repetibilidade para o diagnóstico de multicolinearidade	84
4.4.2	Número de ensaios para o diagnóstico de multicolinearidade	87
4.4.3	Considerações sobre a repetibilidade e número de ensaios para o diagnóstico da multicolinearidade	91
4.5	Repetibilidade e número de ensaios para estimação da correlação canônica	94
4.5.1	Estimativa da correlação do primeiro par canônico.....	94
4.5.2	Coeficiente de repetibilidade para estimação da correlação canônica.....	97
4.5.3	Número de ensaios para estimação da correlação canônica	99

4.5.4	Considerações sobre a repetibilidade e número de ensaio para a correlação canônica	102
4.6	Relação entre tamanho de amostra e número de variáveis em correlação canônica	104
4.6.1	Análise de correlação canônica e reamostragem	104
4.6.2	Determinação da relação entre o número de observações e de variáveis	111
4.6.3	Considerações parciais sobre a relação entre o tamanho de amostra e o número de variáveis.....	115
5	CONSIDERAÇÕES FINAIS	119
6	CONCLUSÕES.....	121
	REFERÊNCIAS.....	123
	APÊNDICE A – ROTINA EM R PARA DETERMINAÇÃO DA RELAÇÃO NÚMERO DE OBSERVAÇÕES E DE VARIÁVEIS EM CORRELAÇÃO CANÔNICA	133

1 INTRODUÇÃO

1.1 CONTEXTO

As técnicas estatísticas multivariadas são ferramentas importantes que auxiliam na compreensão de determinado fenômeno, por considerarem todos os efeitos das características e não cada uma de maneira isolada, como ocorre em técnicas univariadas. São técnicas que se popularizaram com a popularização dos microcomputadores.

As análises de regressão múltipla, de variância multivariada, de componentes principais, de trilha e de correlações canônicas podem ser citadas como sendo as técnicas mais utilizadas dentre as multivariadas. Quando empregadas corretamente, as técnicas multivariadas se tornam ferramentas importantíssimas na compreensão do fenômeno em estudo e, conseqüentemente, na elaboração de inferências, utilização na interpretação dos resultados e na tomada de decisão. Para o seu correto emprego, devem ser respeitados os pressupostos de cada modelo, bem como, a utilização de banco de dados representativos da população, o uso de experimento planejado com o uso de dimensionamento amostral adequado e controle de todas as etapas de condução e avaliação.

Independentemente da técnica utilizada, as estimativas dos parâmetros do modelo podem ser obtidas com menor ou maior grau de confiabilidade. Neste sentido, o dimensionamento amostral tem a sua importância fundamental na obtenção de estimativas sem vieses dos parâmetros. Para estimativas de maior confiabilidade, é necessário a utilização de um número mínimo de observações, o suficiente para a redução do intervalo de confiança para a estimativa do parâmetro e que seja capaz de capturar toda a variabilidade da característica analisada.

Mais de uma variável é considerada simultaneamente na análise multivariada e a relação entre os números de observações (tamanho de amostra) e de variáveis utilizados no modelo também deve ser observada para que estimativas de maior confiabilidade sejam obtidas. Em técnicas multivariadas, têm-se generalizações e recomendações práticas da necessidade mínima de dez observações para cada variável utilizada na análise, não havendo consenso para a indicação de um número adequado de observações para cada variável. Estudos realizados em diversas técnicas multivariadas, não apontam para uma única indicação para o relação entre o número de observações e de variáveis, sendo também constatado que a relação de dez observações para cada variável pode ser insuficiente para a obtenção de estimativas confiáveis dos parâmetros.

Na Agronomia e ciências correlatas, os estudos são conduzidos em multi-ambientes de produção, com o objetivo de verificar a existência ou não de variabilidade entre os ambientes, identificar genótipo que possua superioridade, adaptabilidade e estabilidade entre os ambientes. Esse conhecimento é importante para a indicação de algum material genético para cultivo com ou sem restrição de área, época, tecnologia ou qualquer outro fator que possa estar ligado com o perfeito desenvolvimento e expressão genética da cultura. O ambiente pode ser compreendido como o local geográfico, a época de semeadura, o ano ou qualquer outro evento que permita a distinção de uma condição para a outra. Estimativas de maior confiabilidade e representatividade da população são obtidas quando da utilização de um maior número de ensaios (ambientes), no entanto, quanto maior for o número, maior será o custo de implantação, condução e avaliação, sem que haja um ganho na precisão na mesma proporção.

Assim como em análises univariadas, nos modelos multivariados, em sua grande maioria, a validade dos resultados depende do atendimento dos pressupostos de normalidade, linearidade, homogeneidade e heterocedasticidade. Nos modelos multivariados, além desses pressupostos, o diagnóstico do grau de multicolinearidade entre as variáveis preditoras e, conseqüentemente, o uso de técnicas que reduzam o seu grau também são fundamentais para obtenção de menor imprecisão e vieses nas estimativas.

Na ocorrência de colinearidade, duas variáveis estão altamente correlacionadas entre si e a incorporação de ambas no modelo, pode não trazer um aumento significativo no grau de explicação do modelo, e sim, resultar em estimativas sem sentido biológico. De modo semelhante, quando mais de duas variáveis estão fortemente relacionadas entre si, tem-se a multicolinearidade.

Assim, também o diagnóstico do grau de multicolinearidade deve ser realizado para que estimativas não viesadas sejam obtidas. Quando verificada a presença de multicolinearidade em grau elevado, há a necessidade de contorná-la a níveis que, conforme cada indicador, não seja de magnitude o suficiente para causar sérios problemas, como o viés das estimativas dos parâmetros e a interpretação equivocada ou sentido dos resultados. A eliminação de variáveis que estão causando a multicolinearidade ou a adição de uma constante à diagonal da matriz de correlação estão entre as alternativas mais usuais e empregadas pelos pesquisadores para a redução do grau de multicolinearidade. Apesar de ser uma prática de fácil execução para a redução do grau de multicolinearidade, houve o custo para a avaliação da característica excluída. Mais uma vez, o correto planejamento experimental e uso de correto dimensionamento amostral são fundamentais para otimização de recurso e obter estimativas fidedignas da população e com alta precisão.

Vários são os indicadores disponíveis na literatura que podem ser utilizados para determinação do grau de multicolinearidade. Dentre eles, o número de condição e fator de inflação da variância são os frequentemente utilizados. Estudos com estes indicadores têm apontado para a necessidade de tamanho mínimo de amostra para a estimativa do grau de multicolinearidade e ocorrência de baixa precisão na estimativa do grau de multicolinearidade quando determinado a partir de tamanho amostral insuficiente. Também a utilização de tamanhos de amostra diferentes no diagnóstico da multicolinearidade foi verificada quando a estimação foi realizada por distintos indicadores e em diferentes grupos de variáveis e cultivares.

Os indicadores número de condição e fator de inflação da variância utilizam a matriz de correlação linear de Pearson, podendo esta ser obtida de dados provenientes de um único ou de múltiplos ambientes. Pode ser que há pouca variabilidade entre ambientes e poucas medições ou ensaios podem ser necessárias para a estimação do grau de multicolinearidade com maior precisão. Caso contrário, maior número de medições deverá ser utilizado para a obtenção do mesmo grau de confiabilidade.

O estudo de repetibilidade vem sendo utilizado com frequência para determinar o número de ensaios em médias e/ou correlações em caracteres de várias culturas. Esses estudos têm sido utilizados para a determinação do número de ensaios em características agronômicas e a análise de repetibilidade pode também ser útil na determinação do número de medições ou ensaios necessários no diagnóstico do grau de multicolinearidade e na estimativa de parâmetros em métodos multivariados, como na análise de correlação canônica.

Elevada repetibilidade, ou seja, coeficiente de repetibilidade próximo a unidade, significa que poucas medições ou número de ensaios são suficientes para a estimação da característica com alto grau de confiabilidade. Por outro lado, estimativas de coeficiente próximo de zero indicam que um maior número de medições ou ensaios é necessário na estimação.

Realizar o estudo para determinar o número mínimo de medições ou de ensaios para a estimação da multicolinearidade e para a correlação do primeiro par canônico pode permitir que estes possam ser estimados com maior grau de confiança em caracteres de centeio. Assim como verificar a relação ideal entre o número de observações e de variáveis na obtenção de parâmetros sem vieses em análise de correlação canônica com caracteres de centeio. Bem como, verificar a relação entre o número de observações e variáveis para servirem de referência a outras culturas. Determinações podem ser úteis no planejamento experimental à cultura do centeio, bem como material de referência a outras culturas. Apesar de o presente estudo utilizar

somente duas cultivares de centeio, elas são de distintos propósitos. Sendo assim, os resultados podem ser utilizados tanto em genótipos com o interesse em produção de grãos quanto para plantas forrageiras e de cobertura de solo.

Os resultados encontrados no presente estudo quanto às estimativas do grau de multicolinearidade por meio de dois indicadores e estimativas de parâmetros obtidos na análise de correlação canônica serão de enorme valia ao planejamento experimental. Com eles, o pesquisador poderá realizar o planejamento com maior segurança em relação ao número mínimo de ensaios necessário e a utilização do tamanho de amostra em função do número de variáveis no diagnóstico do grau de multicolinearidade e na obtenção de estimativas de parâmetros em análise de correlação canônica. Diretamente, esses estudos fornecerão para a cultura do centeio, mas que, indiretamente, poderão ser utilizadas a outras culturas. Assim, o estudo foi realizado com intuito de contribuir com o planejamento e análise experimental, na determinação do número de medições ou de ensaios necessários no diagnóstico do grau de multicolinearidade e estimação da correlação do primeiro par canônico e na indicação da relação entre o número de observações e de variáveis em análise de correlação canônica.

1.2 JUSTIFICATIVA

Estimativas do grau de multicolinearidade e de parâmetros da correlação canônica de baixa confiabilidade podem estar sendo geradas não tão somente devido à insuficiência amostral, mas também, pela inadequada utilização de número de medições ou ensaios e da relação do número de observações e de variáveis. Existem poucos de estudos relacionados à recomendação do número mínimo de medições e/ou de ensaios necessários para a obtenção de estimativas precisas dos indicadores de multicolinearidade e da correlação do primeiro par canônico, bem como, a identificação da relação adequada entre o número de observações (tamanho de amostra) e de variáveis em análise de correlação canônica.

Poucos são os estudos envolvendo a multicolinearidade e o dimensionamento amostral na análise de correlação canônica na cultura do centeio. Para as demais culturas, a maioria das inferências foram realizadas relacionando a interpretação dos parâmetros das análises com o grau de multicolinearidade e o número de observações utilizado nos estudos. Grande parte do conhecimento envolvendo multicolinearidade, tamanho de amostra e número de variáveis é oriunda de inferências realizadas a partir de resultados em estudos com objetivo diferente ao de verificar o tamanho de amostra e o número de variáveis, sendo assim, generalizações ou recomendações práticas. Não há consenso na literatura da relação ideal entre o número de

observações e de variáveis, nem tampouco, pesquisas nessa abordagem com a cultura do centeio.

1.3 HIPÓTESES

Há necessidade de mais de um ensaio para estimar o grau de multicolinearidade pelos indicadores número de condição e fator de inflação da variância em caracteres morfológicos e produtivos de centeio.

Há variabilidade no número de ensaios necessários para o diagnóstico de multicolinearidade entre cultivares de centeio, épocas de semeadura, grupos de variáveis e indicador.

Mais de um ensaio é necessário na estimação da correlação do primeiro par canônico na cultura do centeio.

Há variabilidade na necessidade do número de ensaios entre as cultivares e épocas de semeadura para estimar a correlação do primeiro par canônico.

As estimativas da correlação entre as variáveis canônicas no primeiro par são influenciadas pelo tamanho de amostra e número de variáveis.

Há variabilidade na relação entre o número de observações (tamanho de amostra) e de variáveis para estimação da correlação entre as variáveis canônicas no primeiro par para as variáveis canônicas obtidas pelos modelos linear e quadrático com resposta em platô.

Há variabilidade na relação entre o número de observações e de variáveis para a estimação da correlação entre as variáveis canônicas no primeiro par entre cultivares e épocas de semeadura.

1.4 OBJETIVO GERAL E OBJETIVOS ESPECÍFICOS

1.4.1 Objetivo geral:

Determinar o número de ensaios necessário para a estimação do grau de multicolinearidade - número de condição (*NC*) e fator de inflação de variância (*FIV*) - em matriz de correlação linear, e para a estimativa da correlação do primeiro par canônico, bem como, determinar a relação entre o número de observações (tamanho de amostra) e de variáveis necessária para a condução de análise de correlação canônica e obtenção de estimativas confiáveis de parâmetros, em ensaios com duas cultivares de centeio.

1.4.2 Objetivos específicos:

Determinar o número mínimo de medições ou de ensaios necessário para estimar o grau de multicolinearidade pelos indicadores número de condição e fator de inflação da variância em matrizes de correlação linear de caracteres morfológicos e produtivos de centeio.

Verificar a variabilidade no número de ensaios necessário para o diagnóstico de multicolinearidade entre cultivares de centeio, épocas de semeadura, grupos de variáveis e indicador.

Estimar o número de ensaios necessário para estimar a correlação do primeiro par canônico na cultura do centeio.

Verificar a variabilidade no número de ensaios para estimar a correlação do primeiro par canônico entre cultivares e épocas de semeadura de centeio.

Verificar se há variabilidade nas estimativas da correlação entre as variáveis canônicas no primeiro par em função do tamanho de amostra e número de variáveis.

Comparar os valores da relação entre o número de observações e de variáveis necessário para estimar a correlação entre as variáveis canônicas no primeiro par.

Verificar se há variabilidade nas estimativas de relação entre o tamanho de amostra e o número de variáveis para estimação da correlação entre as variáveis canônicas no primeiro par entre cultivares de centeio e épocas de semeadura.

2 REVISÃO BIBLIOGRÁFICA

2.1 CULTURA DO CENTEIO

O centeio (*Secale cereale* L.), pertencente à família Poaceae (BUSHUK, 2001), é uma cultura de ciclo anual com porte variando entre 1,20 e 1,80 m de altura e com espigas longas, duas flores férteis e dois grãos por espigueta (BAIER, 1994). O cereal possui grande importância na cadeia produtiva de alimentos em países da Europa. A Alemanha lidera o *ranking* mundial de produção de centeio, seguidos por Polônia e Rússia, com produções de 3,9; 2,9; e 2,1 milhões de toneladas, respectivamente (PARIONA, 2017). No Brasil, a produção é de aproximadamente 9,2 mil toneladas e produtividade média de 1.731 kg ha⁻¹ (IBGE, 2021), com o cultivo predominante nas regiões dos Estados do Paraná e Rio Grande do Sul.

A cultura foi trazida ao Brasil na época colonial, com o aumento de área cultivada até meados da década de 80, com declínio de área cultivada em favor a outras culturas, como por exemplo, o trigo. O material genético disponível no mercado possui histórico de lançamento e registro recente no MAPA – Ministério da Agricultura, Pecuária e Abastecimento. Cinco cultivares de centeio encontram-se com registro no MAPA (MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO, 2021). As cultivares BR 1, BRS Serrano e BRS Progresso foram desenvolvidas e registradas nos anos de 1999, 2005 e 2013, respectivamente (EMBRAPA, 2013; EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 1999, 2005), a cultivar IPR 89 em 2000 (INSTITUTO AGRONÔMICO DO PARANÁ, 2000) e a cultivar Temprano no ano 2010 (ATLÂNTICA SEMENTES S.A., 2010). Todas elas, com exceção da Temprano, foram selecionadas e desenvolvidas tendo como propósito a produção de grãos.

Apesar de pouca explorada, a cultura tem grande potencial agrônomo de ser utilizada devido às suas características morfológicas, nutricionais e propósitos de uso. Ela possui capacidade de adaptar-se às condições de frio intenso e solos com problemas de fertilidade (MORRISON, 2016; SCHLEGEL, 2013) e pode ser cultivada tanto para a produção de grãos, quanto como planta forragem e/ou de cobertura de solo. Em estudo comparando o centeio com outras culturas, ela foi a que teve o melhor desempenho de produção de grãos em condições de menor temperatura e clima árido em relação a outros cereais de inverno: cevada (*Hordeum vulgare* L.), triticale (*Triticosecale Wittmack*) e trigo (*Triticum aestivum* L.) (SCHITTENHELM; KRAFT; WITTICH, 2014).

Os grãos são ótimas fontes proteica e energética, com teores de 13% de proteína e 80% de carboidratos em matéria seca (MORI; NASCIMENTO JUNIOR; MIRANDA, 2013). A característica de oferta precoce de forragem permite o seu uso ainda no final do outono (PAULINO; CARVALHO, 2004), época em que outros cereais forrageiros ainda não estão no ponto ideal de pastejo.

A cultura possui características que a tornam interessante como integrante de sistemas de rotações de culturas. Além dos benefícios proporcionados às características físico-químicas do solo, favorece no manejo de pragas, doenças e plantas invasoras durante o inverno. O centeio possui recomendação de uso como planta de cobertura de solo (ABOU CHEHADE *et al.*, 2021; CRAWFORD; WILLIAMS; WORTMAN, 2018; ZIECH *et al.*, 2015), auxiliando na manutenção do conteúdo de água no solo (BASCHÉ *et al.*, 2016) e exerce efeito alelopático ou de retardamento na germinação de plantas espontâneas (ABOU CHEHADE *et al.*, 2021; CRAWFORD; WILLIAMS; WORTMAN, 2018).

O volume de matéria fresca produzido varia de acordo com o ambiente, principalmente devido às características climáticas e de solo. Em experimento conduzido em Passo Fundo/RS, com semeadura em 18 de maio de 2011, foi observada a necessidade de 170 dias para completar o ciclo, com produção de 13.448 kg ha⁻¹ de massa seca de parte aérea (LEHMEN *et al.*, 2014). Em ensaios conduzidos em Santa Maria, as cultivares BRS Progresso e Temprano, produziram em média, 28.870 e 24.056 kg ha⁻¹ de massa fresca no estágio de florescimento pleno da cultura (CHAVES *et al.*, 2018). Em Pato Branco/PR, as semeaduras realizadas até maio apresentaram elevada capacidade de produção e manutenção da produção (51,4 kg ha⁻¹ dia⁻¹), com maiores rendimentos de biomassa e com maior período de utilização da pastagem, quando da realização de semeadura precoce (semeaduras realizadas em: 11 de março, 08 de abril, 06 de maio e 03 de junho de 2009) (FERRAZZA *et al.*, 2013). Em Canoinhas/SC, em ensaios com consórcios de gramíneas anuais de inverno manejadas em sistema de cortes, com produção total de 3.794 kg ha⁻¹ em solteiro e 5.218 kg ha⁻¹ em consórcio (centeio + aveia + azevém) (HANISCH *et al.*, 2012).

As diferenças entre as estimativas de produção de massa podem ser devidas ao local de condução dos ensaios e ao material genético utilizado nos experimentos, tendo especialmente a época de semeadura uma forte influência sobre a produtividade. A época preferencial de semeadura da cultura do centeio está de acordo com aptidão da cultivar. Semeaduras de cultivares para a produção de grãos são recomendadas nos meses de junho e julho e para formação de pastagens entre os meses de março e maio (BAIER, 1994).

2.2 MULTICOLINEARIDADE

Nas Ciências Agrárias, assim como nas demais áreas do conhecimento, as técnicas de análises multivariadas são ferramentas poderosas nas mãos dos pesquisadores no auxílio da compreensão do fenômeno de estudo, no planejamento e na tomada de decisões. Diferentemente das técnicas univariadas, as multivariadas utilizam múltiplas variáveis e por este motivo, possuem superioridade de capacidade em capturar a variabilidade do objeto de estudo, fornecendo uma compreensão do todo e não de maneira particionada ou delimitada a condicionantes impostas na análise individual de cada característica.

Para obtenção de estimativas de maior confiabilidade, assim como em técnicas de análise univariada, as pressuposições do modelo matemático necessitam ser atendidas. Ainda no caso multivariado, também a multicolinearidade necessita ser investigada e se constatada em nível elevado, há a necessidade de realizar a execução de algum procedimento que proporcione a redução do grau (HAIR *et al.*, 2009). O termo multicolinearidade foi usado inicialmente como referência a casos em que as variáveis independentes estão relacionadas entre si de forma perfeita e menos que perfeita (GUJARATI; PORTER, 2011) e pode ser compreendida como sendo a relação linear de uma variável com outras variáveis (ALIN, 2010). Ou seja, corresponde ao grau em que uma variável pode ser explicada pelas outras variáveis do modelo (HAIR *et al.*, 2009).

A multicolinearidade reduz o desempenho e a predição do modelo na maioria dos métodos lineares. Alto grau de multicolinearidade em conjunto de variáveis independentes significa que as mesmas compartilham informações entre si (DORMANN *et al.*, 2013) e como consequência, há aumento da variância dos parâmetros estimados causado pelo alto grau de multicolinearidade entre as variáveis independentes (FIGUEIREDO FILHO *et al.*, 2011).

O aumento da variância dos parâmetros impossibilita a interpretação do efeito real das variáveis envolvidas no modelo analisado. Isso se deve especialmente a diminuição da confiabilidade das estimativas dos parâmetros dos modelos (ALIN, 2010). Estatísticas superestimadas e presença excessiva de falsos positivos podem verificadas em presença de multicolinearidade, bem como, podem ser resultantes dos vieses da magnitude das estimativas dos coeficientes e dos fatores de inflação de variância (GOODHUE; LEWIS; THOMPSON, 2017). Quando duas ou mais variáveis preditoras estão fortemente correlacionadas, a inversão matricial no ajuste de regressão múltipla é impossibilitada e, conseqüentemente, podem ocorrer distorções nas estimativas dos coeficientes (FARRAR; GLAUBER, 1967).

Estimativas viesadas dos parâmetros de modelos multivariados são obtidos na situação de elevado grau de multicolinearidade. Em análise de correlação canônica, assim como em outras técnicas multivariadas, estimativas dos parâmetros se tornam de baixa confiabilidade em condição de multicolinearidade (HAIR *et al.*, 2009) ou de inadequada interpretação dos resultados (ALVES; CARGNELUTTI FILHO; BURIN, 2017; DORMANN *et al.*, 2013; TOEBE; CARGNELUTTI FILHO, 2013). Em regressões múltiplas sob várias condições de multicolinearidade com dois, quatro e seis preditores, os estudos indicaram que a multicolinearidade não está relacionada ao erro Tipo I, mas sim com o aumento do erro Tipo II, estando a colinearidade e o número de preditores do modelo relacionado com a obtenção de parâmetros com vieses (LAVERY *et al.*, 2019).

Em estudo de análise de trilha sob multicolinearidade em caracteres de milho, foram verificados erros de estimação dos coeficientes de efeitos diretos e resultados sem interpretação biológica (TOEBE; CARGNELUTTI FILHO, 2013). Em análise de correlação canônica com caracteres de milho, em condição de elevada multicolinearidade, houve a superestimação da variabilidade dos coeficientes canônicos, quando comparado com as análises realizadas logo após a redução do grau de multicolinearidade com a eliminação de variáveis (ALVES; CARGNELUTTI FILHO; BURIN, 2017). Em trigo também foi observada imprecisão nas estimativas dos coeficientes do índice de seleção clássico para ganhos simultâneos de produção de grãos e componentes primários, sob condição de multicolinearidade (CARVALHO; CRUZ; CARVALHO, 1999).

Estudos envolvendo o dimensionamento amostral para o diagnóstico do grau de multicolinearidade têm sido realizados em centeio e outras culturas. Em centeio, há a recomendação de utilizar 116 e 99 plantas em caracteres morfológicos e produtivos de centeio, respectivamente, para a realização do diagnóstico pelos indicadores número de condição e fator de inflação da variância (NEU, 2019; NEU *et al.*, 2021).

2.3 INDICADORES DE MULTICOLINEARIDADE

Há uma diversidade de métodos ou testes que permitem a detecção do grau de multicolinearidade. Dentre eles, a magnitude de correlação entre variáveis predictoras, testes de hipóteses e indicadores de multicolinearidade. Neste último grupo, o número de condição (*NC*) e o fator de inflação de variância (*FIV*) podem ser usados como ferramentas de identificação ou mensuração do grau de multicolinearidade. De maneira mais empírica, o coeficiente de correlação linear de Pearson pode identificar as variáveis fortemente relacionadas entre si.

Níveis de correlações entre variáveis superiores a $|r| \geq 0,7$ devem ser analisados com maior atenção, pois nestes casos, as estimativas dos parâmetros da análise multivariada podem ser comprometidas devido à presença de elevado grau de multicolinearidade entre as variáveis (DORMANN *et al.*, 2013).

A estimativa do grau de multicolinearidade pelo indicador número de condição (NC) é obtida a partir da relação entre o maior autovalor ($\lambda_{máx}$) e menor autovalor (λ_{min}) da matriz de correlação (CRUZ; CARNEIRO, 2006; GUJARATI; PORTER, 2011). Sendo assim, tem-se:

$$NC = \lambda_{máx}/\lambda_{min} \quad (1)$$

Autovalor(es) próximo(s) de zero tem por significado a forte associação entre ao menos um par de variáveis. Como regra prática, o indicador divide o nível de multicolinearidade em classes: fraca ($NC \leq 100$); moderada à forte ($100 < NC \leq 1.000$); e, severa ($NC > 1.000$) multicolinearidade (MONTGOMERY *et al.*, 2012). A multicolinearidade é tolerada quando diagnosticada sob fraco grau, sendo necessário a adoção de alguma medida de redução do grau quando $NC > 100$.

O diagnóstico do grau de multicolinearidade pelo indicador fator de inflação de variância (FIV) é estimado por meio de múltiplas regressões, de cada variável explicativa com as demais, sendo o R_j^2 o coeficiente de determinação múltiplo de x_j sobre as demais variáveis explicativas, conforme equação a seguir:

$$FIV_j = 1/(1 - R_j^2) \quad (2)$$

Em condição de ausência de multicolinearidade, o FIV possui valor igual a um ($FIV = 1$). Por outro lado, o FIV assumirá valores muito altos quanto maior for o grau de multicolinearidade. Como regra prática, tem-se problema de multicolinearidade quando o valor de FIV_j for maior a dez ($FIV_j > 10$) (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009), o que acontece quando R_j^2 for maior que 0,90 (GUJARATI; PORTER, 2011). Este indicador possui a vantagem em relação ao NC , de ser possível identificar a variável causadora de multicolinearidade, pois fornece a estimativa de inflação da variância para cada variável.

O FIV é a proporção de aumento da inflação da variância da variável quando comparada em relação à variância se não houvesse multicolinearidade. O valor deste indicador pode ser

interpretado como a velocidade com a qual as variâncias e covariâncias aumentam, devido a presença da multicolinearidade (GUJARATI; PORTER, 2011).

O diagnóstico do grau de multicolinearidade em características de centeio foi encontrado na literatura. Grau de multicolinearidade severo em características relacionadas à fatores climáticos e de solo que afetam a germinação e o estabelecimento da cultura foi observado e a eliminação de variáveis foi eficiente na redução do grau (WILSON; BAKER; ALLAN, 2013). A estimação da quantidade de foragem de centeio em função de variáveis climáticas foi realizada por meio de análise de regressão linear e a eliminação de duas variáveis foi necessária para diminuir a multicolinearidade (PENG *et al.*, 2016). Em estudo de determinação do tamanho de amostra para o diagnóstico do grau de multicolinearidade por diferentes indicadores utilizando diferentes combinações de caracteres produtivos, o grau foi classificado como fraco ou moderado à forte, dependendo de quais caracteres estavam presentes nas combinações (NEU, 2019; NEU *et al.*, 2021).

2.4 ANÁLISE DE REPETIBILIDADE

2.4.1 Contextualização

O coeficiente de repetibilidade (*rptb*) tem sido utilizado em estudos de melhoramento genético como indicador de relação entre as medidas em um mesmo indivíduo, podendo as avaliações serem repetidas no espaço ou no tempo (CRUZ; REGAZZI, 1997). Estatisticamente, o coeficiente *rptb* de uma dada característica pode ser conceituado como sendo a correlação entre as medidas em um mesmo indivíduo (ABEYWARDENA, 1972; CRUZ; REGAZZI, 1997). Fornece a previsão com que determinado valor se repete em sucessivas medições. Na literatura, o mesmo termo pode estar referenciado por coeficiente intra-classe e mede a relação das variâncias entre grupos e dentro de grupos.

O *rptb* possui valor variando entre zero e um ($0 \leq rptb \leq 1$), com repetibilidade máxima para valores próximos a unidade, verificada quando a característica se manifesta com muita constância entre as sucessivas medições em um mesmo indivíduo (CRUZ; REGAZZI, 1997). Quando $rptb \geq 0,50$ é presumido que a maior proporção da variância total é de origem genética, indicando também uma menor influência dos fatores ambientais sobre a característica (LIRA JÚNIOR; FERNANDES BEZERRA; LEDERMAN, 2010). Assim, poucas mensurações são necessárias para obtenção da estimativa do caractere analisado quando o *rptb* for de alta magnitude (CRUZ; REGAZZI, 1997). Os mesmos autores afirmam que a estimativa

desse coeficiente expressa a proporção da variância total que é explicada pelas variações proporcionadas pelo genótipo e pelas alterações permanentes atribuídas ao ambiente comum. Ou seja, expressa a variação de uma dada característica que é devida ao indivíduo, com supressão da variação causada pelo erro experimental.

O estudo da repetibilidade pode também ser ferramenta útil para verificar consistências nas avaliações de um observador e auxiliar na detecção de mudanças nos efeitos da característica avaliada (HARPER, 1994). Uma das principais aplicações da análise é determinar o número de medições necessárias para estimar o valor real da característica analisada, levando em conta a precisão desejada, o custo mínimo e a mão-de-obra (CRUZ; REGAZZI, 1997). Isso porque o uso da análise de repetibilidade permite a determinação do número de medições ou de ensaios necessários para obtenção de estimativas de uma dada característica com certo nível de precisão. Uma vez conhecido o número mínimo de medições, recursos na condução de ensaios podem ser otimizados com a redução de tempo, mão de obra e custo (AZEVEDO *et al.*, 2016; CRUZ; REGAZZI, 1997), sem que a precisão seja afetada de maneira significativa.

Não foram encontrados na literatura estudos de repetibilidade com características agronômicas na cultura do centeio, nem tampouco, para o diagnóstico de multicolinearidade e estimativa da correlação do primeiro par canônico. Em outras culturas, estudos de repetibilidade em diversas características têm sido realizados com diferentes metodologias, com o *rptb* estimado pela análise de variância (AOV) e pelos métodos de componentes principais (PC) e análise estrutural (SA), ambas com base em matrizes de correlação (R) e variância e covariância (S) (CRUZ; REGAZZI, 1997).

2.4.2 Métodos de estimativa do coeficiente de repetibilidade

A estimativa do coeficiente de repetibilidade (*rptb*) pode ser realizada por distintos métodos, conforme descritos em Cruz e Regazzi (1997):

- AOV: análise de variância;
- PCR: componentes principais com base na matriz de correlação;
- PCS: componentes principais com base na matriz de covariância;
- SAR: análise estrutural determinada com base no autovalor teórico da matriz de correlação;
- SAS: análise estrutural determinada com base no autovalor teórico da matriz de variância e covariância.

2.4.2.1 Método de análise de variância

Para o método de análise de variância (AOV) com dois fatores de variação, o modelo matemático pode ser definido por:

$$Y_{ij} = m + F_i + A_j + \varepsilon_{ij} \quad (3)$$

Sendo:

- Y_{ij} a estimativa da variável resposta associada a observação da i-ésima fonte de variação e j-ésimo ambiente;
- m a média geral;
- F_i o efeito da i-ésima fonte de variação, sob influência do ambiente;
- A_j o efeito do ambiente na j-ésima medição; e,
- ε_{ij} o erro experimental estabelecido pelos efeitos do ambiente na j-ésima medição da i-ésima fonte de variação.

Em seguida, após obtidos os quadrados médios das fontes de variação, o coeficiente de repetibilidade (*rptb*) pode ser estimado em função da variância de cada fonte de variação, de acordo com a seguinte expressão (CRUZ; REGAZZI, 1997):

$$r = \frac{\widehat{\sigma}_G}{\widehat{\sigma} + \widehat{\sigma}_G} \quad (4)$$

Onde:

- $\widehat{\sigma}_G$ é a variância atribuída aos efeitos do genótipo sob influência do ambiente;
- $\widehat{\sigma}$ é a variância residual.

2.4.2.2 Método de componentes principais

Duas são as possibilidades de se obter as estimativas do coeficiente de repetibilidade (*rptb*) pelo método de componentes principais: a matriz de correlação (R) e a matriz de covariâncias (S). No método de componentes principais com base na matriz de correlação (PCR), a estimativa de *rptb* consiste na obtenção das estimativas dos autovalores e os

autovetores normalizados da matriz de correlação (ABEYWARDENA, 1972; CRUZ; REGAZZI, 1997):

$$rptb = \frac{\lambda_k}{\sum_j \lambda_j} \quad (j = 1, 2, \dots, \eta) \quad (5)$$

Onde:

- λ_k é o autovalor associado ao autovetor que apresenta elementos de mesmo sinal e semelhante magnitude;
- λ_j é o j-ésimo autovalor.

O primeiro autovalor associado ao primeiro autovetor é utilizado na estimação de *rptb*, mas está sob influência do número de ensaios utilizados na análise, e por esse motivo, a estimativa de *rptb* pode ser obtido por:

$$rptb = \frac{\widehat{\lambda}_1 - 1}{\eta - 1} \quad (6)$$

Onde:

- η é o número de ensaios;
- $\widehat{\lambda}_1$ é o autovalor de matriz de correlação (**R**) associado ao autovetor que possui os elementos de mesmo sinal e semelhante magnitude.

A estimativa pelo método de componente principal com base na matriz de covariância (PCS) pode ser realizada utilizando a matriz de covariâncias (**S**), conforme equação:

$$rptb = \frac{\widehat{\lambda}_1 - \widehat{\sigma}_Y^2}{\widehat{\sigma}_Y^2 (\eta - 1)} \quad (7)$$

Onde: $\widehat{\lambda}_1$ é o autovalor de matriz de covariância (**S**) associado ao autovetor que possui os elementos de mesmo sinal e semelhante magnitude.

2.4.2.3 Método de análise estrutural

Pelo método da análise estrutural, também existem duas maneiras possibilidades para a obtenção da estimativa de coeficiente de repetibilidade (*rptb*), ou seja, pode-se estimar o *rptb* utilizando a matriz de correlação (SAR) e a matriz de covariâncias (SAS). A estimativa de *rptb* por meio do método de análise estrutural com base na matriz de correlação (SAR) pode ser obtida pela expressão:

$$rptb = \frac{\alpha' \hat{S}\alpha - 1}{\eta - 1} \quad (8)$$

Sendo $\alpha' = [1/\sqrt{\eta}, \dots, 1/\sqrt{\eta}]$ o autovetor com elementos paramétricos associados ao maior autovalor de R.

Pelo método da análise estrutural com base na matriz de covariância (SAS), a estimativa de *rptb* pode ser obtida pela expressão:

$$rptb = \frac{\alpha' \hat{R}\alpha - \widehat{\sigma}_Y^2}{\widehat{\sigma}_Y^2 (\eta - 1)} \quad (9)$$

2.4.3 Estudos de repetibilidade

O coeficiente de repetibilidade (*rptb*) tem sido utilizado na estimativa de médias de características nas mais diversas culturas e por diferentes metodologias (Tabela 1). Conforme apontado pela literatura, há variabilidade na estimativa de *rptb* entre os métodos, sendo recomendado a utilização de mais de um método para maior confiabilidade. Isso porque maiores valores de estimativas de *rptb* para determinada característica é obtida ora por um, ora por outro método.

Em estudo com capim-tanzânia (*Panicum maximum* Jacq.), maiores estimativas de *rptb* foram obtidas ao ser estimado pelo método da análise de variância (AOV) (FERNANDES *et al.*, 2017). Enquanto que, em outro estudo com a mesma cultura, menores estimativas de *rptb* para algumas características foram obtidas pelos métodos da análise de variância (AOV) e análise estrutural com base na matriz de correlações (SAR) e matriz de variâncias e covariâncias (SAS) (CARGNELUTTI FILHO *et al.*, 2004). Os métodos PCS e AOV, respectivamente, estimaram os maiores e os menores valores de *rptb* em características de soja (*Glycine max*)

(DUARTE; FERREIRA; SILVA, 2022). Em capim-elefante (*Pennisetum spp.*) (CAVALCANTE *et al.*, 2012; SOUZA *et al.*, 2017), couve (*Brassica oleracea* var. *acephala*) (AZEVEDO *et al.*, 2016), braquiária (*Urochloa brizantha*) (TORRES *et al.*, 2015) e em morangueiro (*Fragaria x ananassa*) (DIEL *et al.*, 2020), as estimativas de *rptb* em todas as características pelo método AOV foram sempre inferiores às estimativas obtidas pelos métodos análise estrutural e componentes principais (AZEVEDO *et al.*, 2016).

Tabela 1 – Métodos utilizados para determinação do coeficiente de repetibilidade em diferentes culturas

Cultura	Autores	Métodos ¹				
		AOV	PCR	PCS	SAR	SAS
Braquiária: (<i>Urochloa brizantha</i>)	Torres <i>et al.</i> (2015)	X	X	X	X	-
(<i>Urochloa ruziziensis</i>)	Souza Sobrinho <i>et al.</i> (2010)	X	X	X	X	-
Capim-elefante: (<i>Pennisetum spp.</i>)	Cavalcante <i>et al.</i> (2012)	X	X	X	X	-
	Souza <i>et al.</i> (2017)	X	X	X	X	-
Capim-tanzânia: (<i>Panicum maximum</i> Jacq.)	Cargnelutti Filho <i>et al.</i> (2004)	X	X	X	X	X
	Fernandes <i>et al.</i> (2017)	X	X	X	X	-
Ciriguela: (<i>Spondias purpurea</i> L.)	Lira Júnior <i>et al.</i> (2014)	X	X	X	X	X
Couve: (<i>Brassica oleracea</i> var. <i>acephala</i>)	Azevedo <i>et al.</i> (2016)	X	X	X	X	-
	Brito <i>et al.</i> (2019)	X	-	-	-	-
Morangueiro: (<i>Fragaria x ananassa</i>)	Andrade Júnior <i>et al.</i> (2020)	X	-	-	-	-
	Diel <i>et al.</i> (2020)	X	X	X	X	-
Pitangueira: (<i>Eugenia uniflora</i> L.)	Lira Júnior; Fernandes Bezerra e Lederman (2010)	X	X	X	X	X
Soja: (<i>Glycine max</i>)	Matsuo <i>et al.</i> (2012)	X	X	X	X	X
	Duarte <i>et al.</i> (2022)	X	X	X	X	X
Trigo: (<i>Triticum aestivum</i>)	Pagliosa <i>et al.</i> (2014)	X	X	X	X	

¹AOV: análise de variância; PCR: componentes principais com base na matriz de correlação; PCS: componentes principais com base na matriz de covariância; SAR: análise estrutural determinada com base no autovalor teórico da matriz de correlação; SAS: análise estrutural determinada com base no autovalor teórico da matriz de covariância.

Em estudo em características forrageiras de braquiária (*Urochloa brizantha*), o método PCR forneceu estimativas mais precisas de *rptb* (TORRES *et al.*, 2015). Isso se deve pelo fato do método por componentes principais considerar o comportamento cíclico da característica, com os elementos do autovetor de mesmo sinal e magnitudes, expressando a tendência dos

genótipos em manter suas posições nas sucessivas medições (ABEYWARDENA, 1972). Em capim-tanzânia (*Panicum maximum* Jacq.), maiores estimativas de *rptb* para algumas características foram obtidas pelo método PCR e para outras características, pelo método CPS (CARGNELUTTI FILHO *et al.*, 2004). As maiores estimativas de *rptb* foram obtidas pelo método PCS em estudos com o capim-elefante (*Pennisetum* spp.) (CAVALCANTE *et al.*, 2012), em trigo (*Triticum aestivum*) (PAGLIOSA *et al.*, 2014) e no morangueiro (*Fragaria x ananassa*) (DIEL *et al.*, 2020).

2.4.4 Precisão da estimativa do coeficiente de repetibilidade

A precisão da estimativa do coeficiente de repetibilidade (*rptb*) pode ser obtida por meio da determinação do coeficiente de determinação (R^2). Assim, a precisão da estimativa do coeficiente de repetibilidade pode ser obtida pela expressão (CRUZ; REGAZZI, 1997):

$$R^2 = \frac{\eta \times rptb}{1 + rptb \times (\eta - 1)} \quad (10)$$

Sendo:

- R^2 o coeficiente de determinação.
- η o número de ensaios utilizados na análise de repetibilidade;
- *rptb* o coeficiente de repetibilidade.

2.5 NÚMERO DE MEDIÇÕES OU ENSAIOS

2.5.1 Contextualização

A determinação do número de medições ou de ensaios pode ser realizada considerando o coeficiente de repetibilidade (*rptb*) de uma determinada característica em função da precisão desejada. A estimativa do coeficiente de repetibilidade pode variar de 0 a 1, e altos coeficientes permitem que o valor real para uma dada característica seja predito com apenas algumas medidas (CRUZ; REGAZZI, 1997).

Para valores do *rptb* de alta magnitude, poucas medições são necessárias para estimar uma determinada característica. Entretanto, quando o *rptb* for de baixa magnitude, maior número de medições é requerido para obter o mesmo nível de precisão. Isso ocorre por que o

aumento no número de medições, minimiza a influência do ambiente ou acrescenta maior variabilidade genética e assim, resulta em maior precisão (ANDRADE JÚNIOR *et al.*, 2020).

2.5.2 Método de estimação do número de medições ou ensaios

O número de medições ou de ensaios (η_m) para a característica em estudo pode ser obtido pela expressão (CRUZ; REGAZZI, 1997):

$$\eta_m = \frac{R^2 (1 - rptb)}{(1 - R^2) rptb} \quad (11)$$

Sendo:

- η_m o número mínimo de medições ou de ensaios;
- $rptb$ o coeficiente de repetibilidade;
- R^2 o coeficiente de determinação.

2.5.3 Estudos de número de medições ou ensaios

Estudos com o objetivo de determinar o número de ensaios necessário para o diagnóstico de multicolinearidade e a correlação do primeiro par canônico em centeio não foram encontrados. No entanto, a determinação do número de medições ou de ensaios tem sido realizada para características qualitativas e morfoagronômicas em outras culturas como braquiária, capim-elefante, capim-tanzânia, ciriguela, couve, morangueiro e pitangueira, por exemplo.

Nestes estudos, foi observado a necessidade de diferentes números de medições entre os caracteres analisados. Em trigo (*Triticum aestivum*) foi observado a necessidade de no mínimo de 11 espigas para a estimativa com alta confiabilidade da porcentagem de grãos germinados na espiga (PAGLIOSA *et al.*, 2014). Em braquiária (*Urochloa brizantha*) foi verificado a necessidade de duas medições para que seja quantificada as características de fibra em detergente neutro, celulose e sílica, enquanto que para teor de proteína bruta, fibra em detergente ácido e digestibilidade in vitro da matéria orgânica requerem quatro medições com precisão mínima de 80% (TORRES *et al.*, 2015). Em genótipos de capim-elefante (*Pennisetum* spp.) foi constatado que na previsão do valor real das variáveis massa de forragem, altura da

planta, comprimento e largura da folha, diâmetro do colmo, clorose e índice de área foliar, três ciclos de avaliação são suficientes ($R^2 = 90\%$) e sete avaliações para o caractere comprimento do entrenó (CAVALCANTE *et al.*, 2012). Em outro estudo com a mesma cultura, foi verificada a necessidade de no mínimo três cortes para os caracteres número de afilhos e largura de lâmina foliar, seis cortes para os caracteres altura de plantas e diâmetro de colmo, e nove cortes para o caractere produtividade de matéria seca ($R^2 = 80\%$) (SOUZA *et al.*, 2017).

2.6 ANÁLISE DE CORRELAÇÃO CANÔNICA

As relações entre variáveis dependentes e independentes são verificadas constantemente. No caso univariado, são utilizadas as análises de regressão linear e análise de variância (ANOVA) quando se dispõem de variáveis quantitativas e qualitativas, respectivamente. Quando se dispõem de duas ou mais variáveis independentes, é necessária uma abordagem multivariada e podem ser utilizadas a análise de regressão múltipla e análise multivariada de variâncias (MANOVA). No entanto, a utilização destas análises está limitada ao tipo e a possibilidade de haver somente uma variável dependente.

A análise de correlação canônica é uma técnica que apresenta o menor número de restrições e ela possibilita o estudo de correlações entre os grupos de variáveis dependentes e independentes (HAIR *et al.*, 2009). A ideia inicial da análise foi de verificar a relação linear entre dois grupos de variáveis e foi idealizada por Harold Hotelling (1935; 1936).

O principal objetivo da análise de correlação canônica é encontrar a combinação linear no primeiro (\mathbf{X}) e no segundo (\mathbf{Y}) grupo, contendo as p e q variáveis em cada grupo, respectivamente, de tal maneira que a correlação entre as duas combinações seja máxima (DUNHAM; KRAVETZ, 1975; HAIR *et al.*, 2009; HOTELLING, 1935; 1936).

Cada grupo de variáveis pode ser expresso por:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \dots \\ \mathbf{X}_p \end{bmatrix} \text{ e } \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \dots \\ \mathbf{Y}_q \end{bmatrix} \quad (12)$$

sendo p e q os números de variáveis presentes no primeiro e no segundo grupo, respectivamente, e o máximo de combinações possíveis é igual ao número de variáveis do menor conjunto de

dados contendo p e q variáveis, ou seja, $s = \min\{p, q\}$ (DUNHAM; KRAVETZ, 1975). As variáveis canônicas U e V , definidas por:

$$\begin{cases} U = a_1 X_1 + a_2 X_2 + \dots + a_p X_p \\ V = b_1 Y_1 + b_2 Y_2 + \dots + b_q Y_q \end{cases} \quad (13)$$

com as combinações lineares de \mathbf{X} e \mathbf{Y} encontradas matricialmente por $U = \mathbf{a}^t \mathbf{X}$ e $V = \mathbf{b}^t \mathbf{Y}$, onde $\mathbf{a} = (a_1, a_2, \dots, a_p)$ e $\mathbf{b} = (b_1, b_2, \dots, b_q)$ são vetores não nulos das combinações lineares.

Por definição:

$$\begin{cases} \text{Var}(U) = \text{Cov}(\mathbf{a}^t \mathbf{X}) = \mathbf{a}^t \mathbf{S}_{11} \mathbf{a} \\ \text{Var}(V) = \text{Cov}(\mathbf{b}^t \mathbf{Y}) = \mathbf{b}^t \mathbf{S}_{22} \mathbf{b} \\ \text{Cov}(U, V) = \mathbf{a}^t \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b} = \mathbf{a}^t \mathbf{S}_{12} \mathbf{b} \end{cases} \quad (14)$$

sendo \mathbf{S}_{11} , \mathbf{S}_{22} e \mathbf{S}_{12} as matrizes de covariâncias amostrais entre os caracteres morfológicos, produtivos e de morfológicos com produtivos, respectivamente.

As correlações entre as variáveis canônicas U e V , em número máximo de s , são estimadas por:

$$\text{Corr}(U, V) = \rho_{U,V} = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}} = \frac{\mathbf{a}^t \mathbf{S}_{12} \mathbf{b}}{\sqrt{\mathbf{a}^t \mathbf{S}_{11} \mathbf{a}} \sqrt{\mathbf{b}^t \mathbf{S}_{22} \mathbf{b}}} \quad (15)$$

A primeira correlação canônica ($\rho_1 = \rho_{U_1, V_1}$) será a máxima relação entre as variáveis U e V . A correlação canônica também pode ser expressa em função das raízes dos autovalores ($\rho_j = \sqrt{\lambda_j}$) (DUNHAM; KRAVETZ, 1975). A segunda correlação canônica também será máxima, mas independente da primeira correlação canônica. O mesmo raciocínio é válido para as demais correlações, até a última correlação canônica (ρ_{U_s, V_s}). Assim, existirão no máximo $s - 1$ pares de variáveis canônicas independentes do par de máxima correlação (ρ_1) que expressam a variância total dos dois grupos de variáveis.

A máxima correlação (ρ_1), obtida entre as combinações de U_1 e V_1 , corresponde ao primeiro autovalor da matriz \mathbf{M}_1 .

$$\begin{aligned}\mathbf{M}_1 &= \mathbf{S}_{11}^{-1/2} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{11}^{-1/2} \\ \mathbf{M}_2 &= \mathbf{S}_{22}^{-1/2} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1/2}\end{aligned}\quad (16)$$

O k-ésimo par de variáveis canônicas pode ser descrito por:

$$\begin{aligned}U_k &= \mathbf{e}_k^t \mathbf{S}_{11}^{-1/2} \mathbf{X} = \mathbf{a}_k^t \\ V_k &= \mathbf{f}_k^t \mathbf{S}_{22}^{-1/2} \mathbf{Y} = \mathbf{b}_k^t\end{aligned}\quad (17)$$

e tem-se que ρ_k é o k-ésimo autovalor da matriz \mathbf{M}_1 e \mathbf{e}_k é o respectivo autovetor e \mathbf{f}_k é o k-ésimo autovetor da matriz \mathbf{M}_2 .

Para que a análise de correlação canônica tenha sentido, identificando as combinações entre \mathbf{X} e \mathbf{Y} , é necessário que $\mathbf{S}_{12} \neq \mathbf{0}$, pois se $\mathbf{S}_{12} = \mathbf{0}$, implica que $\text{Cov}(U, V) = \mathbf{a}^t \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b} = \mathbf{a}^t \mathbf{S}_{12} \mathbf{b} = \mathbf{0}$, para todos \mathbf{a} e \mathbf{b} e conseqüentemente, todas as correlações canônicas serão também nulas. Assim, o teste de significância para todas as correlações canônicas é dado por:

$$\begin{aligned}H_0: \rho_1 = \rho_2 = \dots = \rho_k = \dots = \rho_s = 0 \\ H_1: \rho_i \neq 0, \text{ para algum } i (1 \leq i \leq s)\end{aligned}$$

A estatística do teste segue a distribuição Qui-Quadrado ($\chi^2_{[\alpha;v]}$) com $v = pq$ graus de liberdade, com correção de Bartlett (BARTLETT, 1947; 1941):

$$\chi_c^2 = \left[n - 1 - \frac{1}{2} (p + q + 1) \right] \ln \left[\prod_{i=1}^s (1 - \rho_i^2) \right] \quad (18)$$

Se a hipótese é rejeitada, o teste é realizado para as demais k-ésimas correlações, com as hipóteses dados por:

$$\begin{aligned}H_0: \rho_k > 0; \rho_{k+1} = \dots = \rho_k = \dots = \rho_s = 0 \\ H_1: \rho_i \neq 0, \text{ para algum } i (k + 1 \leq i \leq s)\end{aligned}$$

e $v = (p - k)(q - k)$ graus de liberdade e com estatística do teste Qui-Quadrado:

$$\chi_c^2 = \left[n - 1 - \frac{1}{2} (p + q + 1) \right] \ln \left[\prod_{i=k+1}^s (1 - \rho_i^2) \right] \quad (19)$$

O teste de hipótese considera o tamanho de amostra (n) e o número de variáveis em cada grupo (p e q) e problemas na confiabilidade dos resultados podem ocorrer com a adoção de dimensionamento amostral insuficiente.

Relações podem não ser detectadas em condição de tamanhos de amostra insignificantes e amostras muito grandes indicarão que há significância estatística na maioria dos casos, sem necessariamente de fato, possuírem significância prática (HAIR *et al.*, 2009). O uso da correlação canônica somente faz sentido se os vetores de medida das variáveis de dois grupos, X e Y, são independentes ou não correlacionados (CRUZ; REGAZZI, 1997). Em teste para verificar a independência entre os dois grupos de variáveis, foi observado que desde que o número de caracteres não exceda a 12, há o controle da taxa de erro tipo I e elevado poder de teste para tamanhos de amostra pequenos (25, 30 e 50) (BRUM *et al.*, 2016). Em regressão múltipla, foi verificado que a multicolinearidade causou o aumento do erro Tipo II (LAVERY *et al.*, 2019).

Alguns estudos realizaram inferências sobre tamanho de amostra em análises multivariadas e na detecção de multicolinearidade. Para tanto, há a necessidade de utilização de maiores tamanhos de amostra para diagnóstico de multicolinearidade (OLIVOTO *et al.*, 2017; SARI *et al.*, 2018), com ocorrência de erros de estimação dos coeficientes de efeitos diretos e resultados sem interpretação biológica em análise de trilha (TOEBE; CARGNELUTTI FILHO, 2013) e superestimação da variabilidade dos coeficientes canônicos, quando comparado com as análises realizadas logo após a redução do grau de multicolinearidade com a eliminação de variáveis (ALVES; CARGNELUTTI FILHO; BURIN, 2017). Em regressão múltipla, a multicolinearidade está relacionada com o número de preditores do modelo e consequentemente, na obtenção de parâmetros com vieses (LAVERY *et al.*, 2019). As variáveis colineares entre si podem ser irrelevantes para determinar uma relação quando a interpretação da análise for realizada por pesos canônicos (HAIR *et al.*, 2009).

Estudos utilizando a técnica de análise de correlação canônica em situações de presença e/ou ausência de multicolinearidade não foram encontradas na cultura do centeio. Mas essa abordagem foi realizada em culturas como de soja e milho. Em populações segregantes de soja, não foram observados problemas de multicolinearidade quando utilizada a análise de correlação canônica (PEREIRA *et al.*, 2017). Entretanto, em caracteres morfológicos e produtivos de milho, foi necessária a eliminação de variáveis (ALVES *et al.*, 2016) para reduzir o grau de multicolinearidade e em estudo com presença e ausência de multicolinearidade, foi observado que há superestimação dos coeficientes canônicos (ALVES; CARGNELUTTI FILHO; BURIN, 2017).

A interpretação de resultados na análise de correlação canônica é bastante complexa, com verificação dos pares de variáveis canônicos significativos e então, baseada nas estimativas de pesos canônicos, cargas canônicas ou cargas canônicas cruzadas de variáveis canônicas

significativas. Por ser análise dependente de computação devido aos vários cálculos matriciais, o usuário deve fazer uso do método disponível no aplicativo utilizado para a análise e dar preferência pelas cargas canônicas cruzadas, seguido pela carga canônica e como último recurso, os pesos canônicos (HAIR *et al.*, 2009). Os pesos canônicos podem ser mais adequados quando a previsão for o objetivo de estudo, enquanto que as correlações podem explicar as relações entre variáveis canônicas e grupos de variáveis (ALPERT; PETERSON, 1972).

Entretanto, não somente a significância das variáveis canônicas deve ser observada no momento da interpretação, mas também, o total de variância captada pelas variáveis canônicas. Durante a interpretação, a utilização de correlação canônica ao quadrado (R^2_{ci}) tem sido muito utilizada. Apesar de ser de fácil acesso, deve ser atentado para o fato de que R^2_{ci} representa a variância compartilhada pelas composições lineares dos conjuntos de variáveis canônicas de X e Y, e não a variância extraída dos conjuntos de variáveis originais (ALPERT; PETERSON, 1972).

Com a popularização da análise de correlação canônica, propostas de interpretações foram sendo discutidas na comunidade científica. O índice de redundância foi proposto para superar o viés e a incerteza inerentes ao uso de R^2_{ci} como uma medida de variância compartilhada (LAMBERT; DURAND, 1975). O índice fornece uma medida resumo da habilidade de um conjunto de variáveis de um grupo (X, por exemplo) em explicar a variação nas variáveis do outro grupo (Y, por exemplo), sendo a medida de redundância análoga à estatística R^2 da regressão múltipla (ALPERT; PETERSON, 1972; HAIR *et al.*, 2009). Pode ser compreendido como sendo o critério para avaliar a eficácia na previsão das relações entre grupos, informando a quantidade total de variância extraída pelo conjunto de variáveis preditoras (STEWART; LOVE, 1968), pois o índice não é sensível a eventuais excessos computados nos pesos nas variáveis por possíveis vieses (ALPERT; PETERSON, 1972). E devido a essa característica, os mesmos autores recomendam validação cruzada (repetidas divisões de amostras e replicação) como um teste empírico de inferência estatística. O uso da redundância se torna interessante, pois as estimativas do índice não aumentam necessariamente com a adição ou remoção de variável no conjunto de dados analisado (DUNHAM; KRAVETZ, 1975).

2.7 RELAÇÃO ENTRE TAMANHO DE AMOSTRA E NÚMERO DE VARIÁVEIS

A determinação da relação entre os números de observações e de variáveis na cultura do centeio não tem sido explorada. Em alguns estudos, foi explorada, principalmente, a

determinação do tamanho de amostra para a determinação da média de caracteres (BANDEIRA *et al.*, 2018a; BANDEIRA *et al.*, 2018b) e a estimação do grau de multicolinearidade em grupos de variáveis (NEU, 2019; NEU *et al.*, 2021).

Em outras culturas, a determinação do tamanho de amostra tem sido realizada para a estimação de coeficientes de correlação linear de Pearson em milho (TOEBE *et al.*, 2015), análise de trilha em híbridos de milho (TOEBE *et al.*, 2017b) e determinação do número de observações para a estimativa de parâmetros em modelo em regressão (CARGNELUTTI FILHO; TOEBE, 2020). Nesses estudos, a utilização de um maior número de plantas na análise proporcionou melhoria na precisão e, conseqüentemente, maior confiabilidade das estimativas. Porém, os ganhos de precisão não são significativos a partir de determinado tamanho de amostra.

Quando realizada estudo envolvendo dimensionamento amostral na determinação do grau de multicolinearidade, foi constatado haver a necessidade de diferentes tamanhos de amostras para estimar o grau de multicolinearidade por diferentes indicadores do grau de multicolinearidade em caracteres morfológicos e produtivos de centeio. Há a necessidade, no mínimo, de 102 observações para diagnóstico realizado pelo número de condição ou fator de inflação da variância em caracteres produtivos (NEU, 2019; NEU *et al.*, 2021) e pelo menos, 176 e 85 plantas em caracteres morfológicos quando diagnosticado pelo número de condição ou fator de inflação da variância, respectivamente (NEU, 2019).

Em outras culturas, foram realizadas inferências relacionando o tamanho de amostra suficiente para a detecção do grau de multicolinearidade em genótipos de milho (OLIVOTO *et al.*, 2017) e cultivar de tomate cereja (SARI *et al.*, 2018). Nestes estudos, os autores apontaram que tamanhos menores não estimam o grau de forma eficiente, havendo a necessidade de utilização de tamanhos de amostra maiores para a detecção do grau de multicolinearidade. Ainda no estudo com caracteres em tomate cereja, foi verificado há uma tendência de redução na estimativa do menor autovalor e incremento na estimativa do maior autovalor da matriz de correlação quando da utilização de pequenas amostras (SARI *et al.*, 2018), afetando a estimativa do número de condição (relação entre o maior e o menor autovalor).

O tamanho de amostra tem efeito no poder da análise multivariada e na estimativa de seus coeficientes. Para tanto, em condições de tamanhos de amostra elevados, ocorre redução da probabilidade de erro do intervalo de confiança em regressão múltipla (BONETT; WRIGHT, 2011) e quando se aumenta o número de covariáveis em análise de covariância, maiores tamanhos são necessários (SHIEH, 2017).

Indicações do número mínimo de observações (tamanho de amostra) e número de variáveis necessário na análise de correlação canônica não são consensuais entre os trabalhos científicos. A recomendação prática indica para a necessidade de relação mínima de dez observações para cada variável em análises multivariadas (HAIR *et al.*, 2009). Estudos abordando a indicação da relação entre o tamanho de amostra e o número de variáveis necessário para alcançar resultados confiáveis e estáveis dos coeficientes e correlações canônicas, foi recomendado como a relação devendo ser 20:1 (BARCIKOWSKI; STEVENS, 1975) e 42:1 a 68:1, considerando as duas maiores correlações canônicas (ALEXANDER, 2000; BARCIKOWSKI; STEVENS, 1975).

Em análise de regressão logística, estudos com simulações apontaram que não ocorreram problemas na precisão e significância dos coeficientes de regressão para relações maiores ou iguais a 10:1 (PEDUZZI *et al.*, 1996; WYNANTS *et al.*, 2015). Já em outro estudo, foi observado ser de fraca a estimação dos parâmetros em relação 10:1, sendo necessário mais estudos sobre o tema (VAN SMEDEN *et al.*, 2016).

A ideia de que não existe uma regra única para a relação ideal entre o tamanho de amostra e o número de variáveis também é levantada pelos pesquisadores. Ao realizar a análise, deve ser considerado o número de preditores, a magnitude dos prováveis coeficientes de regressão com base na literatura e as correlações entre os preditores para determinar o tamanho de amostra (COURVOISIER *et al.*, 2011).

Em análise de regressão linear utilizando dados de ensaio clínico, 36,7 indivíduos por preditor são necessários para prever a função pulmonar em mulheres afro-americanas (RILEY *et al.*, 2019). Em simulações, são adequados de 15 a 25 observações por parâmetro (GREEN, 1991) e mínimo de duas observações para estimar os coeficientes de regressão com acurácia e respectivos intervalos de confiança (AUSTIN; STEYERBERG, 2015). Indicação de tamanho de amostra em função do número de variáveis também foi realizada para pequenas ($n = 50 + 8 \times p$, sendo o número de variáveis) e grandes amostras ($n = 104 + p$) (GREEN, 1991), mas que em condições simuladas, as recomendações não apresentam desempenho suficiente para avaliar corretamente a significância do preditor (LAVERY *et al.*, 2019).

2.8 MODELOS DE REGRESSÃO

Os modelos de regressão são comumente utilizados, nas Ciências Agrárias ou em outras áreas, como ferramenta de descrição e/ou previsão no comportamento da característica em estudo. Dentre eles, o modelo linear segmentado com resposta platô (MLRP) e modelo

quadrático segmentado com resposta platô (MQRP) tem sido utilizados na literatura na determinação de tamanho ótimo de parcela, como por exemplo, nas culturas de maracujazeiro (PEIXOTO; FARIA; DE MORAIS, 2011), rabanete (SILVA *et al.*, 2012), mamoeiro (SCHMILDT *et al.*, 2016) e batata doce (RODRÍGUEZ *et al.*, 2018). Na grande maioria dos estudos, os modelos têm sido utilizados em estudos para a determinação do tamanho ótimo de parcela, mas, eles foram também empregados na determinação do tamanho de amostra do grau de multicolinearidade pelos indicadores número de condição e fator de inflação da variância em caracteres morfológicos e produtivos de centeio (NEU, 2019; NEU *et al.*, 2021).

2.8.1 Modelo linear segmentado com resposta platô

Este modelo pertence aos modelos segmentados e possui duas retas conectadas por um ponto em comum, sendo a segunda reta de coeficiente angular nulo (SCHABENBERGER; PIERCE, 2001). Este modelo descreve uma reta crescente ou decrescente de acordo com o coeficiente angular β_1 e outra reta horizontal ao eixo das abcissas, com altura P (platô). O método possui a função:

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i & \text{se } X_i \leq X_c \\ P + \varepsilon_i & \text{se } X_i > X_c \end{cases} \quad (20)$$

Sendo: Y_i é a variável dependente; X_i é a variável independente; β_0 é o intercepto do modelo linear; β_1 é coeficiente angular; ε_i é o erro associado à i -ésima observação; P é o platô; e X_c é o ponto crítico correspondente a união das duas retas. O ponto crítico, que satisfaz a condição, é estimado por:

$$X_c = (\hat{P} - \hat{\beta}_0) / \hat{\beta}_1 \quad (21)$$

Onde \hat{P} , $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimativas dos parâmetros P, β_0 e β_1 , respectivamente.

2.8.2 Modelo quadrático segmentado com resposta platô

De maneira similar ao anterior, o modelo quadrático segmentado com resposta platô possui duas partes, sendo uma delas quadrática e a outra, linear e com coeficiente angular nulo. O modelo possui função:

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i & \text{se } X_i \leq X_c \\ P + \varepsilon_i & \text{se } X_i > X_c \end{cases} \quad (22)$$

em que: Y_i é a variável dependente; X_i é a variável independente; β_0 , β_1 e β_2 são parâmetros do modelo; ε_i é o erro associado à i -ésima observação; P é o platô; e X_c é o ponto crítico e correspondente a união entre a equação quadrática e o platô. A união entre as duas equações é dada por:

$$X_c = -\widehat{\beta}_1/2 \times \widehat{\beta}_2 \quad (23)$$

e o valor correspondente da variável Y_i ao ponto X_i , é:

$$Y_n = \widehat{\beta}_0 - \widehat{\beta}_1^2/4 \times \widehat{\beta}_2 \quad (24)$$

onde: $\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$ são estimativas dos parâmetros do modelo.

3 MATERIAL E MÉTODOS

3.1 INSTALAÇÃO DO EXPERIMENTO E COLETA DE DADOS

Foram conduzidos oito ensaios de uniformidade sem aplicação de tratamentos (experimentos em branco), em diferentes épocas de semeadura para duas cultivares de centeio (*Secale cereale* L.), durante a safra de 2016. A área experimental pertence ao Departamento de Fitotecnia da Universidade Federal de Santa Maria, localizada a 29°42'S, 53°49'O 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 Bruno-Acinzentado distrófico típico (SANTOS *et al.*, 2018).

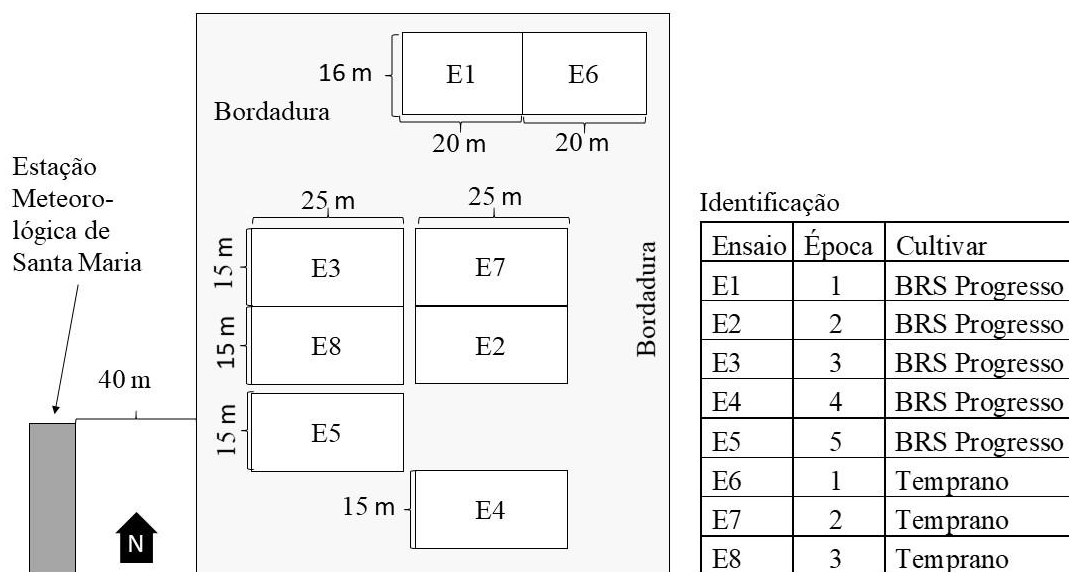
O preparo convencional do solo com gradagens foi realizado em toda a área experimental, de maneira homogênea. A fertilidade do solo foi corrigida com a aplicação de 500 kg ha⁻¹ de adubo com a formulação 5-20-20 (NPK) – 25 kg ha⁻¹ de N, 100 kg ha⁻¹ de P e 100 kg ha⁻¹ de K.

Foram semeadas duas cultivares: a cultivar BRS Progresso, que tem como propósito a produção de grãos (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013); e, a cultivar Temprano, que tem o propósito para a cobertura de solo e planta forrageira (ATLÂNTICA SEMENTES S.A., 2010). As épocas de semeaduras foram inicialmente planejadas para atenderem a recomendação de implantação de ambas as cultivares, ou seja, de março a julho.

As sementes de cada cultivar foram semeadas, à lanço, em área de 320 m² (20 m × 16 m) na primeira época de semeadura (Figura 1), enquanto que nas demais épocas, cada cultivar foi instalada em área de 375 m² (25 m × 15 m). Em ambas cultivares e nas épocas de semeadura foi usada a densidade de 455 sementes m⁻².

O estande de plantas foi determinado com a contagem do número de plantas em cinco pontos, em área formada por um quadrado de 1 m de aresta, amostrados aleatoriamente dentro de cada ensaio aos quinze dias após a semeadura (Tabela 2). A adubação de cobertura foi realizada quando as plantas estavam entre os estádios com três (V3) e quatro (V4) folhas desenvolvidas, na quantidade de 25 kg ha⁻¹ de N. Os demais tratamentos culturais foram realizados conforme necessidade e recomendações de manejo para a cultura de centeio (BAIER, 1994).

Figura 1 – Croqui de localização e instalação de oito ensaios de uniformidade com duas cultivares de centeio (*Secale cereale* L.) em cinco épocas de semeadura, conduzidos em Santa Maria – RS, Safra 2016.



Fonte: Adaptado de NEU (2019).

Tabela 2 – Datas de semeadura, de emergência, de florescimento e de colheita (avaliação), estande de plantas (EP = número de plantas m⁻²) e número de dias entre a semeadura e a colheita (Ciclo) em duas cultivares de centeio (*Secale cereale* L.), em oito ensaios de uniformidade (E1, E2, E3, E4, E5, E6, E7 e E8), conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Época	Ensaio	Datas				EP	Ciclo
			Semeadura	Emergência	Florescimento	Avaliação		
BRS Progresso	1	E1	03/05	09/05	08/08	24/10	365	174
BRS Progresso	2	E2	25/05	01/06	26/08	10/11	263	169
BRS Progresso	3	E3	07/06	17/06	15/09	19/11	262	165
BRS Progresso	4	E4	22/06	29/06	22/09	27/11	204	158
BRS Progresso	5	E5	04/07	10/07	29/09	02/12	344	151
Temprano	1	E6	03/05	10/05	22/09	07/11	284	188
Temprano	2	E7	25/05	02/06	03/10	18/11	204	177
Temprano	3	E8	07/06	20/06	10/10	24/11	369	170

Fonte: NEU (2019).

As avaliações foram realizadas para cada cultivar e época de semeadura. Ao total, oito ensaios de uniformidade foram avaliados, constituídos por cinco épocas de semeadura da cultivar BRS Progresso (E1, E2, E3, E4 e E5) e três épocas de semeadura da cultivar Temprano

(E6, E7 e E8). Foram coletadas de forma aleatória e avaliadas 100 plantas em cada ensaio de uniformidade (cultivar × época de semeadura), com exceção dos ensaios E4 e E8 (90 plantas), em total 780 plantas. Em cada colmo de plantas, foram avaliadas características morfológicas e produtivas, totalizando 1.960 colmos (193; 370; 242; 169; 162; 185; 392; e, 247 colmos nos ensaios E1, E2, ..., E7, e E8, respectivamente).

Em cada colmo, foram realizadas:

- contagens dos números de nós, espiguetas e grãos espiga⁻¹;
- mensurações dos comprimentos de colmo, pedúnculo e de espiga (em centímetros); e,
- pesagens das massas (frescas e secas) de colmo, pedúnculo, espiga e de grãos (em gramas).

As pesagens das massas foram realizadas em balança analítica. Em seguida, baseados nos dados coletados em cada um dos 1960 colmos, foram calculados os valores médios ou total planta⁻¹, caracterizando as estimativas de caracteres morfológicos e produtivos utilizadas nas análises para cada planta em cada ensaio. O caractere número de colmos planta⁻¹ foi definido pela contagem do número de colmos utilizado para a obtenção da média ou total planta⁻¹ e classificado como sendo variável morfológica.

3.2 ANÁLISE DE DADOS

3.2.1 Caracteres morfológicos e produtivos

As estimativas de caracteres obtidas pela média ou somatório de características de dados coletados a campo foram alocados em dois grupos de variáveis – caracteres morfológicos e produtivos (Tabela 3). No total, foram obtidos 16 caracteres e utilizado no presente estudo, sendo destes, oito caracteres pertencentes do grupo das variáveis morfológicas e outros oito caracteres ao grupo das variáveis produtivas.

Em cada ensaio, foram obtidas as estimativas de médias de cada caractere do grupo de variáveis morfológicas e produtivas. As médias de caracteres em cada cultivar foram estimadas (cinco e três ensaios na cultivar BRS Progresso e Temprano, respectivamente) e comparadas pelo teste z a 5% de probabilidade de erro.

Tabela 3 – Identificação (ID), sigla e descrição de caracteres definidos em cada grupo de variáveis (morfológicas e produtivas) em duas cultivares de centeio (*Secale cereale* L.), em oito ensaios de uniformidade (E1, E2, E3, E4, E5, E6, E7 e E8), conduzidos em Santa Maria – RS, Safra 2016.

ID	Grupo	Sigla	Descrição
X1	Morfológico	APPM	altura de planta, em cm (média de colmos planta ⁻¹), definido como sendo a porção entre a base da planta até a última espiguetas.
X2	Morfológico	CCPM	comprimento de colmo, em cm (média de colmos planta ⁻¹), definido como sendo a porção entre a base da plantas até a inserção da folha bandeira no colmo.
X3	Morfológico	CHPM	comprimento de pedúnculo, em cm (média de colmos planta ⁻¹), definido como sendo a porção entre a inserção da bandeira e espiga no pedúnculo.
X4	Morfológico	MFPM	massa fresca de parte aérea, em gramas (média de massa de parte aérea).
X5	Morfológico	MFPT	massa fresca total de parte aérea, em gramas (soma das massas de colmos, folhas, pedúnculo e espigas).
X6	Morfológico	NCPT	número de colmos planta ⁻¹ , obtido por contagem (soma do número de colmo planta ⁻¹ = principal + número de afilhos).
X7	Morfológico	NNPM	número de nós por colmo, obtido por contagem (número de nós da planta dividido pelo número de colmos da planta).
X8	Morfológico	RFMM	relação entre a massa fresca de colmos + folhas + pedúnculos e a massa fresca total de parte aérea, em gramas.
Y1	Produtivo	CEPM	comprimento de espiga, em cm (médias de comprimento de espigas planta ⁻¹).
Y2	Produtivo	M100	massa média de 100 grãos, em gramas.
Y3	Produtivo	MGPT	massa de grãos planta ⁻¹ , em gramas (soma de massa de grãos planta ⁻¹).
Y4	Produtivo	NGPM	número de grãos espiga ⁻¹ , obtido por contagem (divisão do número de grãos da planta pelo número de espigas da planta).
Y5	Produtivo	NGPT	número de grãos planta ⁻¹ , obtido por contagem (soma de grãos em todas espigas planta ⁻¹).
Y6	Produtivo	NSPM	número de espiguetas espiga ⁻¹ , obtido por contagem (divisão do número de espiguetas da planta pelo número de espigas da planta).
Y7	Produtivo	NSPT	número de espiguetas planta ⁻¹ , obtido por contagem (soma de todas as espiguetas da planta).
Y8	Produtivo	RGPM	relação entre a massa de grãos e a massa fresca total de parte aérea.

Fonte: Autor.

3.2.2 Correlações entre caracteres

As correlações lineares de Pearson entre os caracteres foram estimadas em cada grupo de variáveis, ensaio e cultivar. A estimativa da correlação entre os pares de caracteres X_i e $X_{i'}$, foi realizada conforme a expressão (BUSSAB; MORETIN, 2013):

$$r_{X_i X_{i'}} = \frac{Cov(X_i X_{i'})}{\sqrt{var X_i} \sqrt{var X_{i'}}} \quad (25)$$

Sendo:

- $Cov(X_i X_{i'})$ a covariância entre os caracteres X_i e $X_{i'}$;
- $var X_i$ a variância do caractere X_i ;
- $var X_{i'}$ a variância do caractere $X_{i'}$.

Em seguida, para cada par de caracteres X_i e $X_{i'}$, a significância das correlações foi avaliada pelo teste t de *Student* ao nível de 5% de significância ($\alpha = 0,05$), considerando:

$$t_c = \frac{r_{X_i X_{i'}} \sqrt{n-2}}{\sqrt{1 - r_{X_i X_{i'}}^2}} \quad (26)$$

Sendo:

- $r_{X_i X_{i'}}$ a correlação estimada entre o par de caracteres X_i e $X_{i'}$; e,
- n o tamanho de amostra ou observações utilizadas na estimação de $r_{X_i X_{i'}}$.

A hipótese nula ($H_0: r_{X_i X_{i'}} = 0$) foi rejeitada ao nível α de probabilidade de erro se $|t_c| \geq t_{(\alpha; n-2)}$.

Foram obtidas oito matrizes de correlações em cada grupo de variáveis. Foram apresentadas as oito matrizes de correlação de cada par de variáveis entre as épocas de semeadura para cada cultivar (oito ensaios).

3.2.3 Combinações de caracteres

Após definidos os caracteres dentro de cada grupo de variáveis, foram planejados casos de estudos, obtidos pelas combinações de p_0 caracteres combinados p_i a $p_i - C_{(p_0, p_i)}$. Definiu-se como p_0 sendo o número máximo de caracteres avaliados ($p_0 = 8$ caracteres) dentro do grupo de variáveis, p_i como o número de caracteres combinados na i -ésima combinação de caracteres dentro de um mesmo grupo de variáveis, com i variando de 2 à $(p_0 - 1)$. Ou seja, no primeiro caso ($C_{(p_0, p_i)} = C_{(8, 2)} = 28$ combinações), identificado por caso 2, foram consideradas todas as

combinações de caracteres dois a dois. Nas combinações seguintes, com acréscimo de um caractere, foram consideradas as combinações com três ($C_{(8,3)}$), quatro ($C_{(8,4)}$) até a última combinação com ($p_0 - 1 = 7$) caracteres ($C_{(8,7)}$).

No total, foram obtidas 596 combinações, sendo 246 combinações pertencentes ao grupo de variáveis morfológicas e outras 246 combinações pertencentes ao grupo de variáveis produtivas. Dentro de cada grupo de variáveis, os casos contendo 2, 3, 4, 5, 6 e 7 combinações obtidas, respectivamente, pelas combinações de $p_i = 2, 3, 4, \dots, 7$ caracteres, foram compostos de 28, 56, 70, 56, 28 e 8 combinações.

Tabela 4 – Demonstração de esquema de combinações (Comb.) de oito de caracteres morfológicos (X^1) arranjados em combinações de 2, 3, 4, 5, 6 e 7 caracteres em diferentes casos de estudo com cultivares de centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Comb.	Caso						
	2	3	4	5	6	7	
1	X1;X2	X1;X2;X3	X1;X2;X3;X4	X1;X2;X3;X4;X5	X1;X2;X3;X4;X5;X6	X1;X2;X3;X4;X5;X6;X7	
2	X1;X3	X1;X2;X4	X1;X2;X3;X5	X1;X2;X3;X4;X6	X1;X2;X3;X4;X5;X7	X1;X2;X3;X4;X5;X6;X8	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
7	X1;X8	X1;X3;X4	X1;X2;X4;X6	X1;X2;X3;X5;X8	X1;X2;X3;X5;X6;X7	X1;X3;X4;X5;X6;X7;X8	
8	X2;X3	X1;X3;X5	X1;X2;X4;X7	X1;X2;X3;X6;X7	X1;X2;X3;X5;X6;X8	X2;X3;X4;X5;X6;X7;X8	
9	X2;X4	X1;X3;X6	X1;X2;X4;X8	X1;X2;X3;X6;X8	X1;X2;X3;X5;X7;X8		
10	X2;X5	X1;X3;X7	X1;X2;X5;X6	X1;X2;X3;X7;X8	X1;X2;X3;X6;X7;X8		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
27	X6;X8	X2;X4;X5	X1;X4;X5;X7	X1;X3;X5;X6;X7	X2;X4;X5;X6;X7;X8		
28	X7;X8	X2;X4;X6	X1;X4;X5;X8	X1;X3;X5;X6;X8	X3;X4;X5;X6;X7;X8		
29	-	X2;X4;X7	X1;X4;X6;X7	X1;X3;X5;X7;X8			-
30	-	X2;X4;X8	X1;X4;X6;X8	X1;X3;X6;X7;X8			-
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
55	-	X5;X7;X8	X2;X6;X7;X8	X3;X5;X6;X7;X8			-
56	-	X6;X7;X8	X3;X4;X5;X6	X4;X5;X6;X7;X8			-
57	-	-	X3;X4;X5;X7	-	-	-	
58	-	-	X3;X4;X5;X8	-	-	-	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
68	-	-	X4;X5;X7;X8	-	-	-	
69	-	-	X4;X6;X7;X8	-	-	-	
70	-	-	X5;X6;X7;X8	-	-	-	
Total	28	56	70	56	28	8	

¹Caracteres definidos na Tabela 3.

Fonte: Autor.

Tabela 5 – Demonstração de esquema de combinações (Comb.) de oito de caracteres produtivos (Y¹) arranjados em combinações de 2, 3, 4, 5, 6 e 7 caracteres em diferentes casos de estudo com cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Comb.	Caso						
	2	3	4	5	6	7	
1	Y1;Y2	Y1;Y2;Y3	Y1;Y2;Y3;Y4	Y1;Y2;Y3;Y4;Y5	Y1;Y2;Y3;Y4;Y5;Y6	Y1;Y2;Y3;Y4;Y5;Y6;Y7	
2	Y1;Y3	Y1;Y2;Y4	Y1;Y2;Y3;Y5	Y1;Y2;Y3;Y4;Y6	Y1;Y2;Y3;Y4;Y5;Y7	Y1;Y2;Y3;Y4;Y5;Y6;Y8	
:	:	:	:	:	:	:	:
7	Y1;Y8	Y1;Y3;Y4	Y1;Y2;Y4;Y6	Y1;Y2;Y3;Y5;Y8	Y1;Y2;Y3;Y5;Y6;Y7	Y1;Y3;Y4;Y5;Y6;Y7;Y8	
8	Y2;Y3	Y1;Y3;Y5	Y1;Y2;Y4;Y7	Y1;Y2;Y3;Y6;Y7	Y1;Y2;Y3;Y5;Y6;Y8	Y2;Y3;Y4;Y5;Y6;Y7;Y8	
9	Y2;Y4	Y1;Y3;Y6	Y1;Y2;Y4;Y8	Y1;Y2;Y3;Y6;Y8	Y1;Y2;Y3;Y5;Y7;Y8		-
10	Y2;Y5	Y1;Y3;Y7	Y1;Y2;Y5;Y6	Y1;Y2;Y3;Y7;Y8	Y1;Y2;Y3;Y6;Y7;Y8		-
:	:	:	:	:	:	:	:
27	Y6;Y8	Y2;Y4;Y5	Y1;Y4;Y5;Y7	Y1;Y3;Y5;Y6;Y7	Y2;Y4;Y5;Y6;Y7;Y8		-
28	Y7;Y8	Y2;Y4;Y6	Y1;Y4;Y5;Y8	Y1;Y3;Y5;Y6;Y8	Y3;Y4;Y5;Y6;Y7;Y8		-
29	-	Y2;Y4;Y7	Y1;Y4;Y6;Y7	Y1;Y3;Y5;Y7;Y8			-
30	-	Y2;Y4;Y8	Y1;Y4;Y6;Y8	Y1;Y3;Y6;Y7;Y8			-
:	:	:	:	:	:	:	:
55	-	Y5;Y7;Y8	Y2;Y6;Y7;Y8	Y3;Y5;Y6;Y7;Y8			-
56	-	Y6;Y7;Y8	Y3;Y4;Y5;Y6	Y4;Y5;Y6;Y7;Y8			-
57	-	-	Y3;Y4;Y5;Y7				-
58	-	-	Y3;Y4;Y5;Y8				-
:	:	:	:	:	:	:	:
68	-	-	Y4;Y5;Y7;Y8				-
69	-	-	Y4;Y6;Y7;Y8				-
70	-	-	Y5;Y6;Y7;Y8				-
Total	28	56	70	56	28		8

¹Caracteres definidos na Tabela 3.

Fonte: Autor.

3.2.4 Diagnóstico de multicolinearidade

3.2.4.1 Multicolinearidade sem e com eliminação de caracteres em grupos de variáveis

Foi realizado o diagnóstico da multicolinearidade pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*), em dois grupos de variáveis e em cada ensaio (época de semeadura × cultivar), com todos os oito caracteres morfológicos e oito caracteres produtivos, conforme definido na Tabela 3. O diagnóstico foi realizado utilizando os caracteres dentro de cada grupo de variáveis. Em seguida, em cada grupo de variáveis e ensaio, foi eliminado o caractere que apresentou a estimativa de *FIV* superior a dez ($FIV_j > 10$). Repetiu-se o diagnóstico de multicolinearidade com a exclusão do caractere com a maior estimativa do

grau em cada grupo de variáveis obtida na análise anterior, até atendido o critério de $FIV_j \leq 10$ (GUJARATI; PORTER, 2011).

3.2.4.2 Multicolinearidade em combinações de caracteres

O grau de multicolinearidade foi estimado pelos indicadores número de condição (NC) e fator de inflação de variância (FIV), utilizando a matriz de correlação linear de Pearson entre caracteres em cada combinação dentro de cada caso, grupo de variáveis (Tabela 3) e ensaio (épocas de semeadura \times cultivares) (Tabela 1). As combinações dentro de cada caso, contendo 2, 3, 4, 5, 6 e 7 caracteres combinados foram definidas no Capítulo 3.2.3.

O NC foi estimado utilizando a matriz de correlação linear de Pearson em cada grupo de variáveis (morfológicos e produtivos). Ao realizar o diagnóstico, obteve-se um único valor para indicador NC , sendo este, utilizado para interpretar o grau de multicolinearidade.

Também foi estimado o grau de multicolinearidade pelo indicador FIV nas mesmas matrizes de correlação em cada combinação em que foi realizado o diagnóstico pelo NC . O indicador FIV fornece a estimativa de inflação para cada caractere, ou seja, obtêm-se uma estimativa para cada variável. Devido a essa característica, este indicador possibilita identificar o caractere que está causando a multicolinearidade.

Como o indicador FIV apresenta estimativa de inflação da variância para cada caractere, foi considerado a maior estimativa entre as j -ésimas estimativas de FIV , ou seja, $FIV_{máx} = \text{máximo}(FIV_j)$ para cada combinação.

3.3 ESTUDO 1 – NÚMERO DE MEDIÇÕES PARA O DIAGNÓSTICO DE MULTICOLINEARIDADE

3.3.1 Casos com combinações de caracteres morfológicos e produtivos

Em cada grupo de variáveis morfológicas e produtivas (Tabela 3), foram planejados seis casos formados pelas combinações de oito caracteres ($p_0 = 8$ caracteres) combinados p_i a p_i ($C_{(p_0, p_i)}$, com $p_i = 2, 3, 4, 5, 6$ e 7 caracteres). Ou seja, os seis casos identificados de 2, 3, 4, 5, 6 a 7 contêm as combinações de 2, 3, 4, 5, 6 e 7 caracteres, respectivamente, conforme descrito no Capítulo 3.2.3.

Em seguida, foram obtidas as estimativas do grau de multicolinearidade pelos indicadores número de condição (NC) e fator de inflação da variância (FIV) utilizando os

caracteres combinados conforme definido em cada combinação e caso. Portanto, as estimativas foram realizadas em cada combinação de caracteres dentro de cada caso, grupos de variáveis, época de semeadura e cultivar.

3.3.2 Estimativa do coeficiente de repetibilidade

O estudo de repetibilidade (*rptb*) para o diagnóstico de multicolinearidade pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*) foi realizado em cada caso (caracteres combinados), dentro de cada grupo de variáveis (caracteres morfológicos e produtivos) e em cada cultivar. As diferentes combinações de caracteres, dentro de cada caso, foram consideradas como sendo os ‘sujeitos’ observados e as épocas de semeadura consideradas como as medidas repetidas. Assim, as estimativas de *rptb* foram realizadas com os dados provenientes de cinco e três épocas de semeadura (ensaios) nas cultivares BRS Progresso e Temprano, respectivamente.

Como exemplo, para a estimativa de *rptb* da multicolinearidade pelo indicador *NC* com as combinações de oito caracteres dois a dois – caso 2 com total de 28 combinações, foram consideradas as 140 (28 combinações × cinco ensaios) e 84 estimativas de *NC* (28 combinações × três ensaios) em caracteres morfológicos na cultivar BRS Progresso e Temprano, respectivamente. O mesmo número de estimativas, para cada cultivar e caso foram obtidas para o grupo de variáveis produtivas, pois este grupo de variáveis também foi composto por total de oito caracteres. Portanto, para cada grupo de variáveis, foram obtidas 140; 280; 350; 280; 140; e 40 estimativas de multicolinearidade pelo indicador *NC* na cultivar BRS Progresso e 84; 168; 210; 168; 84; e 24 estimativas de multicolinearidade pelo indicador *NC* na cultivar Temprano, para os casos com 2, 3, 4, 5, 6 e 7 caracteres combinados, respectivamente.

A estimativa do coeficiente de repetibilidade (*rptb*) para a estimativa do grau de multicolinearidade foi realizada pelos métodos (CRUZ; REGAZZI, 1997): análise de variância (AOV); componentes principais com base na matriz de correlação (PCR); componentes principais com base na matriz de variância e covariância (PCS); análise estrutural determinada com base no autovalor teórico da matriz de correlação (SAR); e, análise estrutural determinada com base no autovalor teórico da matriz de variância e covariância (SAS), conforme equações descritas no Capítulo 2.4.2. Para estimativa de *rptb* pelo método AOV, foi considerado o modelo da Equação 3, considerando a combinação de caracteres a fonte de variação e o ensaio como sendo o ambiente.

O coeficiente de repetibilidade (*rptb*) pelo método AOV foi estimado pela expressão Equação 4 (CRUZ; REGAZZI, 1997). As demais estimativas de *rptb* pelos métodos PCR, PCS, SAR e SAS, foram obtidas conforme as Equações 6 a 9. Também foi determinada a precisão na estimativa de *rptb* do grau de multicolinearidade com a determinação do coeficiente de determinação (R^2) em distintos métodos, de acordo com a Equação 10, sendo *rptb* o coeficiente de repetibilidade e η o número de ensaios utilizados na estimação da repetibilidade, com $\eta = 5$ e $\eta = 3$ quando a estimativa foi realizada com dados da cultivar BRS Progresso e Temprano, respectivamente.

3.3.3 Número de medições ou ensaios

A determinação do número mínimo de medições ou de ensaios (η_m) para a estimação do grau de multicolinearidade foi realizada utilizando as estimativas do coeficiente de repetibilidade em cada cultivar, grupo de variáveis, caso, indicador (*NC* e *FIV*), e em cinco níveis de precisão – coeficientes de determinações ($R^2 = 0,80; 0,85; 0,90; 0,95; 0,99$). A equação para a determinação de η_m foi definida conforme Equação 11, mostrada no Capítulo 2.5.2.

3.4 ESTUDO 2 – NÚMERO DE ENSAIOS PARA A ESTIMAÇÃO DA CORRELAÇÃO CANÔNICA

Para o estudo do número de ensaios para a estimação da correlação canônica do primeiro par canônico (ρ_{U_1, V_1}), foram obtidas as estimativas do parâmetro em cenários obtidos pelas combinações de caracteres em cada grupo de variáveis. Ou seja, as análises de correlação canônica foram realizadas em todos os oito ensaios – cinco épocas de semeadura com a cultivar BRS Progresso (ensaios E1, E2, E3, E4 e E5) e outras três épocas de semeadura com a cultivar Temprano (ensaios E6, E7 e E8) (Tabela 2), contendo p e q caracteres em cada grupo de variáveis, sendo p e q o número de caracteres combinados em cada cenário.

3.4.1 Cenários com caracteres combinados

Foram formados cenários contendo os casos dos dois grupos de variáveis (caracteres morfológicos e produtivos), conforme apresentado na Tabela 3. As formas de obtenção de cada

caso e as respectivas combinações entre os caracteres em cada grupo de variáveis foram realizadas conforme descritas no Capítulo 3.2.3.

Em cada grupo de variáveis, foram obtidos seis casos, obtidos pelas combinações de p_0 caracteres combinados p_i a $p_i - C_{(p_0, p_i)}$, sendo p_0 o número máximo de caracteres avaliados ($p_0 =$ oito caracteres) e p_i o número de caracteres combinados ($i = 2, 3, \dots, [p_0 - 1]$ caracteres combinados). Portanto, para os casos com 2, 3, 4, 5, 6 e 7 caracteres combinados, tem-se o total de 28, 56, 70, 56, 28 e 8 combinações de caracteres, respectivamente.

Em seguida, a análise de correlação canônica foi realizada com os dois grupos de variáveis, tendo os caracteres morfológicos no primeiro grupo (X) e no segundo grupo (Y) os caracteres produtivos. Assim, foram planejados cenários compostos pelos casos contendo as combinações de caracteres morfológicos no primeiro e pelos casos contendo as combinações de caracteres produtivos no segundo grupo de variáveis (caso no grupo de variáveis morfológicas \times caso no grupo de variáveis produtivas). Assim, foram formados seis cenários: 2×2 ; 3×3 ; 4×4 ; 5×5 ; 6×6 ; e 7×7 (Tabela 6), com 784; 3.136; 4.900; 3.136; 784; e, 64 combinações em cada cenário, respectivamente, totalizando 12.804 combinações em cada ensaio.

Tabela 6 – Número de combinações (Comb.) de caracteres em cada caso, grupo de variáveis, ensaio, cultivar e total em cenários com os casos formados pelas combinações de caracteres do grupo de variáveis morfológicas (MORF) e produtivas (PROD) de centeio (*Secale cereale* L.), em ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cenário	Caso		Nº Comb. Caracteres Caso ⁻¹		Nº de Comb. Ensaio ⁻¹	Cultivar ²		Total de Comb.
	MORF	PROD	MORF	PROD		BRSP	TEMP	
2×2	2	2	28	28	784	3.920	2.352	6.272
3×3	3	3	56	56	3.136	15.680	9.408	25.088
4×4	4	4	70	70	4.900	24.500	14.700	38.950
5×5	5	5	56	56	3.136	15.680	9.408	25.088
6×6	6	6	28	28	784	3.920	2.352	6.272
7×7	7	7	8	8	64	320	192	512
Total	-	-	-	-	12.804	64.020	38.412	102.432

¹Casos descritos no Capítulo 3.2.3. ²Cultivar: BRS Progresso (BRSP) e Temprano (TEMP).

Fonte: Autor.

3.4.2 Estimativa da correlação canônica

A análise de correlação canônica tem por objetivo encontrar a combinação linear no primeiro grupo de variáveis (X – caracteres morfológicos) e no segundo grupo de variáveis (Y – caracteres produtivos), que maximize a correlação entre as duas variáveis canônicas (HAIR *et al.*, 2009), contendo p e q variáveis, respectivamente. O máximo de combinações lineares corresponde ao mínimo entre p e q , ou seja, $s = \text{mín}\{p, q\}$ e a primeira correlação canônica (ρ_{U_1, V_1}) entre as primeiras variáveis canônicas (U_1 e V_1) expressa a máxima combinação linear entre o grupo de variáveis morfológicas e produtivas.

Em cada cenário foram considerados os casos com igualdade de número de caracteres combinados em cada grupo de variáveis (Tabela 4). Para cada combinação, foi realizada a análise de correlação canônica e anotado o valor da estimativa da correlação do primeiro par canônico. Ao total, foram obtidas 102.432 estimativas de correlações canônicas para o primeiro par canônico (oito ensaios \times 12.804 combinações).

3.4.3 Estimativa do coeficiente de repetibilidade

O estudo de repetibilidade ($rptb$) para a estimativa da correlação para o primeiro par canônico (ρ_{U_1, V_1}) foi realizado em cada cenário, cultivar e pelos métodos análise de variância (AOV), componentes principais com base na matriz de correlação (PCR); componentes principais com base na matriz de covariância (PCS), análise estrutural, determinada com base no autovalor teórico da matriz de correlação (SAR) e análise estrutural determinada com base no autovalor teórico da matriz de covariância (SAS) (CRUZ; REGAZZI, 1997), com equações descritas no Capítulo 2.4.2.

Para a obtenção da estimativa de $rptb$ em cada cultivar e cenário, as combinações foram consideradas como os indivíduos observados e as épocas de semeadura foram consideradas como sendo as medidas repetidas e elas variaram de acordo com a cultivar. Os ensaios E1, E2, E3, E4 e E5 correspondem às épocas de semeadura na cultivar BRS Progresso e ensaios E6, E7 e E8 na cultivar Temprano (Tabela 2). Para estimativa de $rptb$ pelo método AOV, foi considerado o modelo da Equação 3, considerando a combinação de caracteres em cada cenário a fonte de variação e o ensaio como sendo o ambiente.

O coeficiente de repetibilidade ($rptb$) pelo método AOV foi estimado pela Equação 4 (CRUZ; REGAZZI, 1997). As demais estimativas de $rptb$ pelos métodos PCR, PCS, SAR e SAS, foram obtidos conforme as Equações 6 a 9. Também foi determinada a precisão na

estimativa de $rptb$ da correlação canônica (ρ_{U_1, V_1}) para o primeiro par canônico com a determinação do coeficiente de determinação (R^2) em distintos métodos, de acordo com a Equação 10, sendo $rptb$ o coeficiente de repetibilidade e η o número de ensaios utilizados na estimação da repetibilidade, com $\eta = 5$ e $\eta = 3$ quando estimativa realizada na cultivar BRS Progresso e Temprano, respectivamente.

3.4.4 Determinação do número de medições ou de ensaios

O número de medições ou de ensaios (η_m) para estimar a correlação do primeiro par canônico foi realizado para cada estimativa do coeficiente de repetibilidade em cada cultivar, grupo de variáveis, cenário e em cinco níveis de precisão – coeficientes de determinações ($R^2 = 0,80; 0,85; 0,90; 0,95; 0,99$), conforme Equação 11 definida no Capítulo 2.5.2.

3.5 ESTUDO 3 – RELAÇÃO ENTRE O NÚMERO DE OBSERVAÇÕES E DE VARIÁVEIS EM CORRELAÇÃO CANÔNICA

3.5.1 Análise de correlação canônica e reamostragem

A análise de correlação canônica foi realizada com dados de seis ensaios cultivados em três épocas de semeadura e com duas cultivares de centeio. Ou seja, foram considerados os ensaios avaliados em três épocas de semeadura na cultivar BRS Progresso (ensaios E1, E2 e E3) e outras três, na cultivar Temprano (E6, E7 e E8) (Tabela 2). Foram utilizados seis caracteres morfológicos – primeiro grupo (X), e seis caracteres produtivos – segundo grupo (Y), com exceção no ensaio E8 (terceira época de semeadura da cultivar Temprano) no grupo de variáveis produtivas. Neste ensaio, foram considerados seis e cinco caracteres nos grupos X e Y, respectivamente.

O número de caracteres em cada grupo de variáveis foi definido após o diagnóstico de multicolinearidade em cada grupo de variáveis e ensaio. Para a análise foram considerados somente os caracteres em que após o diagnóstico de multicolinearidade, a maior estimativa do fator de inflação da variância foi inferior ou igual a dez ($FIV_{m\acute{a}x} \leq 10$) (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009), sendo os demais caracteres eliminados, conforme descrito no Capítulo 3.2.4.1 e permanecendo os caracteres conforme apresentados no Capítulo 4.3.1. Por este motivo, foram utilizados seis caracteres em cada grupo de variáveis e época de semeadura,

com exceção no grupo de variáveis produtivas da cultivar Temprano (ensaio E8). Logo, em todos os ensaios, o primeiro grupo foi composto por $p = 6$ caracteres e o segundo grupo por $q = 6$ caracteres, mas com $q = 5$ caracteres no ensaio E8.

Em seguida, foram planejados 100 tamanhos de amostra (n) para o estudo de relações entre o número de observações (tamanho de amostra) e de variáveis ($n:p$). O primeiro tamanho de amostra planejado foi composto por $n = p + q$ observações e os tamanhos seguintes, foram obtidos com o acréscimo de $p + q$ observações, até o último tamanho, contendo $100 \times (p + q)$ observações. Para $p = q = 6$ caracteres, foram planejados os tamanhos de amostras $n = 12, 24, 36, 48, 60, \dots, 1.200$ observações, e no ensaio E8 com $p = 6$ e $q = 5$ caracteres, foram planejados os tamanhos de amostras $n = 11, 22, 33, 44, 55, \dots, 1.100$ observações, correspondendo as relações $n:p = 1, 2, 3, 4, 5, \dots, 100$.

Posteriormente, para cada ensaio (três épocas \times três cultivares), foram estimadas as médias e as matrizes de variância e covariância de cada grupo de variáveis (X e Y). Com estas estimativas, foram geradas amostras multivariadas com distribuição normal ($R = 1.000$ reamostras) em cada tamanho de amostra planejado, utilizando a função ‘mvnorm’ do pacote MASS (VENABLES; RIPLEY, 2002). Em seguida, a normalidade multivariada foi verificada ao nível de 5% de significância, pela função ‘mvShapiro.Test’ do pacote mvShapiroTest (GONZALEZ-ESTRADA; VILLASENOR-ALVA, 2013) e realizado o diagnóstico de multicolinearidade pelo fator de inflação da variância (FIV). A análise de correlação canônica foi realizada quando atendidos os critérios de normalidade (valor- $p > 0,05$) e de multicolinearidade ($FIV \leq 10$). Caso contrário, nova amostra multivariada foi simulada.

Em cada análise, foram anotadas as estimativas da correlação do primeiro par canônico e o índice de redundância. Por fim, foram obtidas as estimativas de média, desvio-padrão e percentis de 2,5% e 97,5% entre as R estimativas em cada tamanho de amostra planejado, época e cultivar.

3.5.2 Determinação da relação entre o número de observações e de variáveis

A determinação da relação entre o número de observações (tamanho de amostra) e de variáveis para estimar a correlação do primeiro par canônico e índice de redundância foi realizada por meio de dois métodos de regressão segmentado com resposta em platô. Foram utilizados o modelo linear segmentado com resposta em platô (MLRP) e modelo quadrático segmentado com resposta em platô (MQRP), conforme Equações 21 e 22, definidas no Capítulo 2.8. Em seguida, com base nas estimativas médias de correlação do primeiro par canônico, foi

realizado a análise de regressão, sendo as estimativas das médias da correlação do primeiro par canônico consideradas como sendo as variáveis dependentes (Y) e a relação n:p como sendo a variável independente (X). O mesmo procedimento foi realizado para o índice de redundância.

Para cada análise, após obtidos os parâmetros dos modelos, foram calculadas as estimativas da relação n:p, correspondendo à união das funções segmentadas (linear ou quadrática com outra de coeficiente angular nulo ou platô), conforme as Equações 21 e 23 para os modelos MLRP e MGRP, respectivamente. Para cada análise, também foi estimado o valor da variável resposta para a relação n:p estimada ($Y_{n:p}$) e o coeficiente de determinação ajustado (R^2_a).

3.6 APLICATIVOS E SOFTWARES ESTATÍSTICOS

Os dados coletados das avaliações foram tabulados em planilha eletrônica Excel[®]. As análises de determinação do número de ensaios para o diagnóstico de multicolinearidade e para estimar a correlação do primeiro par canônico e a determinação da relação entre o número de observações e de variáveis para a correlação canônica foram executadas no Software R (R TEAM CORE, 2021). A representação gráfica também foi realizada utilizando pacotes de funções executadas no Software R.

Os algoritmos dos coeficientes de repetibilidade e o número de ensaios pelos diferentes métodos foram realizados conforme descrito em Cruz e Regazzi (1997). No estudo da relação entre o número de observações e de variáveis, foram utilizadas amostras multivariadas obtidas com auxílio da função ‘mvrnorm’ do pacote MASS (VENABLES; RIPLEY, 2002) e a confirmação da normalidade multivariada foi realizada com o auxílio da função ‘mvShapiro.Test’ do pacote mvShapiroTest (GONZALEZ-ESTRADA; VILLASENOR-ALVA, 2013).

A análise de correlação canônica foi realizada utilizando a função ‘cca’, presente na biblioteca em R ‘yacca’ (BUTTS, 2018). Na apresentação dos resultados em gráfico, foram utilizadas funções presentes na biblioteca ‘ggplot2’ (WICKHAM, 2016).

4 RESULTADOS E DISCUSSÃO

Os resultados apresentados neste capítulo são oriundos de oito ensaios conduzidos com duas cultivares de centeio (BRS Progresso e Temprano). Inicialmente, foram semeados ensaios em cinco épocas com ambas as cultivares. No entanto, não foram considerados os ensaios semeados em ensaios no final de junho (ensaio E4: 22/06 – quarta época) e início de julho (ensaio E5: 04/07 – quinta época) com a cultivar Temprano, por causa do não desenvolvimento e/ou enchimento de grãos. Assim, foram considerados todos os ensaios semeados com cultivar BRS Progresso (ensaios E1, E2, E3, E4 e E5, correspondendo as semeaduras na primeira, segunda, terceira, quarta e quinta época, respectivamente) e ensaios semeados na primeira, segunda e terceira época com a cultivar Temprano (ensaios E6, E7 e E8, respectivamente).

4.1 ESTATÍSTICA DESCRITIVA DE CARACTERES MORFOLÓGICOS E PRODUTIVOS

4.1.1 Caracteres morfológicos

As médias dos caracteres entre cultivares foi testada pelo teste z em nível de 5% de significância e pôde ser verificado superioridade da cultivar BRS Progresso para alguns caracteres, enquanto que em outros caracteres, maiores médias ocorreram na cultivar Temprano (Tabela 7). De maneira geral, as médias dos caracteres morfológicos na cultivar BRS Progresso foram inferiores às médias da cultivar Temprano em todos os caracteres, com exceção de massa fresca de parte aérea (MFPM), com superioridade em 19,46%. Entre as médias dos caracteres comprimento de colmo (CCPM) e massa fresca total de parte aérea (MFPT) não foram verificadas diferenças significativas a 5% de significância pelo teste t de *Student* entre as cultivares.

A cultivar Temprano superou as médias da cultivar BRS Progresso em 2,24; 4,80; 22,41; 2,04; e, 4,17% para altura de planta (APPM), comprimento de pedúnculo (CHPM), número de colmos planta⁻¹ (NCPT), número de nós por colmo (NNPM) e relação entre a massa fresca de colmos + folhas + pedúnculos e a massa fresca total de parte aérea (RFMM), respectivamente. Não pode ser observado alta variabilidade nas estimativas dos caracteres morfológicos entre as épocas de semeadura. Maiores valores do coeficiente de variação foram verificados para os caracteres MFPM, MFPT e NCPT, mas mantendo a magnitude dentro de cada cultivar e/ou época de semeadura.

Tabela 7 – Médias e coeficiente de variação de estimativas em caracteres morfológicos avaliados em oito ensaios de uniformidade (E1, E2, ..., E8) na cultura do centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Época	Ensaio ²	Caracteres morfológicos ¹							
			APPM	CCPM	CHPM	MFPM	MFPT	NCPT	NNPM	RFMM
----- Média -----										
BRS Progresso	1	E1	147,07	88,40	48,07	3,56	6,97	1,93	4,94	0,70
BRS Progresso	2	E2	158,76	103,79	40,33	5,23	19,43	3,70	4,99	0,68
BRS Progresso	3	E3	156,38	103,05	39,97	4,63	11,25	2,42	5,15	0,64
BRS Progresso	4	E4	157,26	100,68	43,84	3,92	7,44	1,88	4,90	0,61
BRS Progresso	5	E5	152,32	95,94	45,86	3,21	5,33	1,62	4,59	0,64
Média cultivar	-	-	154,30b ³	98,32a	43,61b	4,11a	10,14a	2,32b	4,91b	0,65b
Temprano	1	E6	145,40	87,53	46,57	3,16	5,63	1,85	4,54	0,72
Temprano	2	E7	164,03	104,03	46,93	3,42	13,34	3,92	5,21	0,76
Temprano	3	E8	164,53	110,21	41,83	3,34	8,60	2,74	5,30	0,78
Média cultivar	-	-	157,76a	100,26a	45,23a	3,31b	9,21a	2,84a	5,01a	0,75a
----- Coeficiente de variação (%) -----										
BRS Progresso	1	E1	7,82	11,13	10,64	31,96	51,66	39,15	9,00	8,58
BRS Progresso	2	E2	7,00	9,04	14,59	27,02	56,88	47,75	9,46	10,01
BRS Progresso	3	E3	9,62	12,27	14,93	27,85	60,06	45,18	9,39	9,61
BRS Progresso	4	E4	9,66	9,71	17,44	30,19	60,19	47,78	9,80	8,11
BRS Progresso	5	E5	11,44	13,32	16,03	37,30	70,02	52,48	12,68	9,67
Temprano	1	E6	10,59	13,67	12,73	34,87	42,71	37,15	11,82	7,59
Temprano	2	E7	8,90	11,69	13,74	34,48	63,35	56,95	9,14	9,20
Temprano	3	E8	10,12	9,34	25,30	33,48	57,50	56,91	9,78	8,82

¹Caracteres definidos na Tabela 3. ²Ensaio detalhado na Tabela 2. ³Médias com diferentes letras na coluna (entre cultivares) diferem entre si pelo teste z em nível de 5% de significância.

Fonte: Autor.

Nos caracteres MFPM, MFPT e NCPT, as estimativas do coeficiente de variação (CV, em percentagem) são classificadas como sendo altas (CV entre 20,0% e 30,0%) e muito altas (CV superior a 30,0%), segundo a classificação de Pimentel-Gomes (1990). Entre os ensaios para esses três caracteres, alto CV foi observado em caractere MFPM avaliado na cultivar BRS Progresso em ensaio semeado em final de maior (25/05 – segunda época; CV = 27,02%) e início de junho (07/06 – terceira época; CV = 27,85%). Nos demais ensaios de MFPM e todos os oito para MFPT e NCPT, o CV foi classificado como muito alto ($30,19\% \leq CV \leq 70,02\%$). O maior valor foi observado para o caractere MFPT (CV = 70,02%) em ensaio semeado no início de julho (04/07 – quinta época) com a cultivar BRS Progresso. O caractere CHPM apresentou alta estimativa de CV = 25,30% em um único ensaio com semeadura em início de junho (07/06 –

terceira época) com a cultivar Temprano. Nas demais estimativas, os valores de CVs foram de baixos a médios ($CV \leq 17,44\%$), independentemente da época de semeadura e cultivar.

4.1.2 Caracteres produtivos

De maneira semelhante ao observado no grupo de caracteres morfológicos, diferenças nas médias de caracteres produtivos entre as duas cultivares foram verificadas pelo teste z em nível de 5% de significância (Tabela 8), com exceção para o caractere comprimento de espiga (CEPM). A cultivar BRS Progresso apresentou as maiores médias de massa média de 100 grãos (M100), massa de grãos planta⁻¹ (MGPT), número de grãos espiga⁻¹ (NGPM), número de grãos planta⁻¹ (NGPT) e relação entre a massa de grãos e a massa fresca total de parte aérea (RGPM). Para estes caracteres, a cultivar BRS Progresso superou em 69,53; 96,39; 36,23; 14,38; e, 85,56% as estimativas de médias para a cultivar Temprano. Por outro lado, as maiores médias de número de espiguetas espiga⁻¹ (NSPM) e número de espiguetas planta⁻¹ (NSPT) foram observadas na cultivar Temprano, com estimativas de 10,40 e 35,89% superiores às médias da cultivar BRS Progresso, respectivamente.

Três dos oito caracteres produtivos apresentaram estimativas muito altas de CV ($CV > 30,0\%$) em todos os ensaios, independentemente de ensaio (época de semeadura \times cultivar). Maiores estimativas do coeficiente de variação (CV, em percentagem) foram observados nos caracteres MGPT, NGPT e NSPT, com estimativas entre 55,29 e 96,48%; entre 49,51 e 79,18%; e, entre 40,32 e 63,07%, respectivamente. Os caracteres CEPM e NSPM apresentaram CV médios (CV entre 10 e 20%), com exceção na primeira e quinta época de semeadura com as cultivares Temprano e BRS Progresso, respectivamente. Alta variabilidade para M100 e RGPM foi observada na segunda e terceira época de semeadura com a cultivar Temprano ($42,59\% \leq CV \leq 56,84\%$).

Tabela 8 – Médias e coeficiente de variação de estimativas em caracteres produtivos avaliados em oito ensaios de uniformidade (E1, E2, ..., E8) na cultura do centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Época	Ensaio ²	Caracteres produtivos ¹							
			CEPM	M100	MGPT	NGPM	NGPT	NSPM	NSPT	RGPM
----- Média -----										
BRS Progresso	1	1	10,60	2,44	1,66	34,31	66,54	29,08	56,89	0,24
BRS Progresso	2	2	14,64	2,72	4,75	47,79	179,34	36,95	137,46	0,25
BRS Progresso	3	3	13,36	2,39	3,29	56,74	135,73	36,23	87,32	0,30
BRS Progresso	4	4	12,73	2,17	2,33	55,83	105,89	34,85	65,84	0,31
BRS Progresso	5	5	10,52	1,87	1,50	46,38	76,41	29,97	49,49	0,28
Média cultivar	-	-	12,36a ³	2,32a	2,71a	48,05a	112,92a	33,39b	79,68b	0,27a
Temprano	1	6	11,30	1,57	1,02	36,27	65,46	31,45	56,95	0,18
Temprano	2	7	13,07	1,27	2,01	37,29	149,16	40,74	161,45	0,14
Temprano	3	8	12,49	1,26	1,08	31,93	79,64	38,57	106,22	0,12
Média cultivar	-	-	12,28a	1,37b	1,38b	35,27b	98,72b	36,86a	108,26a	0,15b
----- Coeficiente de variação (%) -----										
BRS Progresso	1	1	18,66	16,62	55,29	27,58	49,51	19,51	46,65	22,53
BRS Progresso	2	2	13,22	27,40	64,71	32,92	64,81	16,23	51,60	28,06
BRS Progresso	3	3	14,40	19,49	64,51	25,28	54,50	13,79	49,11	21,96
BRS Progresso	4	4	16,63	21,34	62,97	24,59	54,03	17,05	52,36	16,03
BRS Progresso	5	5	21,90	30,12	68,82	27,25	61,98	22,14	61,33	24,61
Temprano	1	6	22,40	28,07	55,72	41,31	52,05	24,42	40,32	29,76
Temprano	2	7	18,52	54,85	96,48	42,40	79,18	15,11	63,07	53,87
Temprano	3	8	15,26	42,59	87,67	51,75	67,37	11,80	61,81	56,84

¹Caracteres produtivos definidos na Tabela 3. ²Ensaio detalhado na Tabela 2. ³Médias com diferentes letras na coluna (entre cultivares) diferem entre si pelo teste z em nível de 5% de significância.

Fonte: Autor.

4.1.3 Considerações gerais sobre os caracteres em centeio

A cultivar Temprano apresentou estimativas médias iguais ou superiores à cultivar BRS Progresso em caracteres morfológicos, com exceção da MFPM. Apesar da média de massa fresca de parte aérea por colmo (MFPM) na Temprano ser inferior que na BRS Progresso, ela conseguiu apresentar maior estimativa média de massa fresca de parte aérea por planta (MFPT), possivelmente por apresentar também o maior número de colmos por planta (NCPT). Já a cultivar BRS Progresso apresentou maior número de caracteres com as maiores médias em caracteres produtivos (M100, MGPT, MGPM, NGPT e RGPM). Esses foram obtidos por meio de mensuração de características ligadas aos grãos. A cultivar BRS Progresso tem recomendação de uso para a produção de grãos (EMBRAPA, 2013), enquanto que a cultivar Temprano possui o propósito de produção de massa para uso como planta forrageira e/ou como

cobertura de solo (ATLÂNTICA SEMENTES S.A., 2010) e essas características podem ter contribuído para que maior parte das médias de caracteres morfológicos fosse observada na cultivar Temprano e médias de caracteres produtivos na cultivar BRS Progresso.

No entanto, as estimativas de altura de planta nas cultivares BRS Progresso e Temprano encontradas nesse estudo foram superiores aos valores de duas populações segregantes de centeio de inverno (populações A e B) obtidos de ensaios conduzidos na Alemanha e Polônia (A = 117,9 cm; B = 115,6 cm) (MIEDANER *et al.*, 2012), mas inferiores à estimativa média em avaliações com a cultivar BRS Serrano (160 cm) (LEHMEN *et al.*, 2014). Este último trabalho foi conduzido em Passo Fundo com semeadura em 11 de maio e o ciclo da BRS Serrano (170 dias) foi muito similar ao ciclo da BRS Progresso (169 dias) semeada na segunda época de semeadura (25/05). A massa de parte aérea na cultivar BRS Progresso reduziu da segunda à quinta época de semeadura. Esse mesmo efeito foi observado com a massa de matéria seca na cultivar BRS Serrano, apresentando estimativas de 7.724 kg ha⁻¹ – semeadura em 11/03, e 4.324 kg ha⁻¹ – semeadura em 03/06 (FERRAZZA *et al.*, 2013).

Já em caracteres produtivos, a BRS Progresso foi a cultivar que apresentou as maiores estimativas de massa média de 100 grãos (M100), massa de grãos planta⁻¹ (MGPT), número de grãos espiga⁻¹ (NGPM), número de grãos planta⁻¹ (NGPT) e relação entre massa média de grãos planta⁻¹ e massa fresca total de planta (RGPM). Para as condições de condução do ensaio, a BRS Progresso apresentou as maiores estimativas de caracteres relacionados aos grãos. Esse comportamento está relacionado com a aptidão da cultivar BRS Progresso que tem como propósito a produção de grãos (BAIER, 1994).

A estimativa da média de MGPT na cultivar BRS Progresso foi semelhante às estimativas de rendimentos de grãos em estudo realizado com duas populações segregantes de centeio de inverno em ensaios conduzidos na Alemanha e Polônia (MIEDANER *et al.*, 2012). No presente estudo, a cultivar apresentou média 2,71 g planta⁻¹, correspondendo a estimativa de rendimento de 7.793,96 Mg ha⁻¹ (2,71 g planta⁻¹ × população média de 287,6 plantas m⁻²), com oscilação entre estimativas de 6.059,0 à 12.492,5 Mg ha⁻¹, nos ensaios E1 e E2, respectivamente.

4.2 CORRELAÇÃO LINEAR ENTRE CARACTERES

Diferentes valores de estimativas de correlação entre pares de caracteres foram observados entre as épocas de semeadura dentro de uma mesma cultivar e entre as cultivares dentro de uma mesma época, tanto em caracteres morfológicos (Tabela 9), como em caracteres produtivos (Tabela 10). Quando estimadas a menor e a maior correlação entre pares de caracteres dentro de uma mesma cultivar, pôde ser observado que as magnitudes das correlações foram muito semelhantes e significativas em nível de 5% de probabilidade de erro. No entanto, em outros ensaios, foi verificada significância das correlações a 5% de probabilidade de erro pelo teste *t* de *Student* em ensaios para algumas das épocas. Ou seja, em uma ou outra época de semeadura, a correlação foi significativa entre dois caracteres e em outras épocas de semeadura, foi verificada correlação nula (correlação não significativa).

Tabela 9 – Correlações entre pares de caracteres morfológicos¹ de centeio conduzidos diferentes épocas de semeadura com cultivares de centeio (*Secale cereale* L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.

	APPM	CCPM	CHPM	MFPM	MFPT	NCPT	NNPM	RFMM
(continua)								
----- Época 1 ² -----								
APPM	1,00	0,85*	0,45*	0,44*	0,12 ^{ns}	-0,17 ^{ns}	0,28*	0,15 ^{ns}
CCPM	0,88*	1,00	-0,07 ^{ns}	0,14 ^{ns}	-0,12 ^{ns}	-0,28*	0,43*	0,14 ^{ns}
CHPM	0,58*	0,14 ^{ns}	1,00	0,42*	0,28*	0,09 ^{ns}	-0,18 ^{ns}	0,06 ^{ns}
MFPM	0,60*	0,35*	0,48*	1,00	0,63*	0,13 ^{ns}	-0,05 ^{ns}	0,27*
MFPT	0,25*	0,08 ^{ns}	0,29*	0,53*	1,00	0,82*	-0,17 ^{ns}	0,15 ^{ns}
NCPT	-0,31*	-0,25*	-0,18 ^{ns}	-0,29*	0,60*	1,00	-0,26*	0,00 ^{ns}
NNPM	0,49*	0,61*	-0,04 ^{ns}	0,34*	0,15 ^{ns}	-0,09 ^{ns}	1,00	-0,06 ^{ns}
RFMM	-0,05 ^{ns}	0,00 ^{ns}	-0,01 ^{ns}	-0,21*	-0,20*	-0,01 ^{ns}	0,00 ^{ns}	1,00
----- Época 2 -----								
APPM	1,00	0,83*	0,45*	0,61*	0,32*	0,06 ^{ns}	0,26*	0,14 ^{ns}
CCPM	0,84*	1,00	-0,09 ^{ns}	0,46*	0,26*	0,07 ^{ns}	0,50*	0,15 ^{ns}
CHPM	0,54*	0,02 ^{ns}	1,00	0,24*	0,05 ^{ns}	-0,08 ^{ns}	-0,39*	0,04 ^{ns}
MFPM	0,49*	0,14 ^{ns}	0,59*	1,00	0,48*	0,03 ^{ns}	0,30*	0,02 ^{ns}
MFPT	0,15 ^{ns}	-0,05 ^{ns}	0,28*	0,37*	1,00	0,86*	0,09 ^{ns}	0,00 ^{ns}
NCPT	-0,05 ^{ns}	-0,09 ^{ns}	0,04 ^{ns}	-0,03 ^{ns}	0,89*	1,00	-0,06 ^{ns}	0,02 ^{ns}
NNPM	0,23*	0,43*	-0,27*	0,04 ^{ns}	-0,07 ^{ns}	-0,09 ^{ns}	1,00	0,18 ^{ns}
RFMM	0,17 ^{ns}	0,47*	-0,37*	-0,47*	-0,27*	-0,07 ^{ns}	0,24*	1,00
----- Época 3 -----								
APPM	1,00	0,89*	0,50*	0,62*	0,31*	0,00 ^{ns}	0,46*	0,10 ^{ns}
CCPM	0,74*	1,00	0,06 ^{ns}	0,49*	0,22*	-0,04 ^{ns}	0,57*	0,10 ^{ns}
CHPM	0,80*	0,19 ^{ns}	1,00	0,32*	0,23*	0,07 ^{ns}	-0,05 ^{ns}	0,03 ^{ns}

Tabela 9 – Correlações entre pares de caracteres morfológicos¹ de centeio conduzidos diferentes épocas de semeadura com cultivares de centeio (*Secale cereale* L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.

	APPM	CCPM	CHPM	MFPM	MFPT	NCPT	NNPM	RFMM
	(conclusão)							
MFPM	0,68*	0,27*	0,70*	1,00	0,49*	0,03 ^{ns}	0,25*	0,18 ^{ns}
MFPT	0,13 ^{ns}	0,05 ^{ns}	0,12 ^{ns}	0,14 ^{ns}	1,00	0,86*	-0,11 ^{ns}	0,17 ^{ns}
NCPT	-0,23*	-0,11 ^{ns}	-0,24*	-0,33*	0,85*	1,00	-0,28*	0,12 ^{ns}
NNPM	0,19 ^{ns}	0,39*	-0,05 ^{ns}	0,19 ^{ns}	-0,03 ^{ns}	-0,11 ^{ns}	1,00	0,02 ^{ns}
RFMM	-0,27*	0,14 ^{ns}	-0,54*	-0,52*	-0,04 ^{ns}	0,23*	-0,11 ^{ns}	1,00
	----- Época 4 -----							
APPM	1,00	0,84*	0,77*	0,56*	0,28*	-0,01 ^{ns}	0,25*	0,22*
CCPM	-	1,00	0,33*	0,38*	0,20 ^{ns}	-0,02 ^{ns}	0,43*	0,27*
CHPM	-	-	1,00	0,41*	0,17 ^{ns}	-0,03 ^{ns}	-0,06 ^{ns}	0,15 ^{ns}
MFPM	-	-	-	1,00	0,57*	0,07 ^{ns}	0,19 ^{ns}	0,04 ^{ns}
MFPT	-	-	-	-	1,00	0,82*	0,06 ^{ns}	-0,03 ^{ns}
NCPT	-	-	-	-	-	1,00	-0,03 ^{ns}	-0,08 ^{ns}
NNPM	-	-	-	-	-	-	1,00	0,34*
RFMM	-	-	-	-	-	-	-	1,00
	----- Época 5 -----							
APPM	1,00	0,87*	0,65*	0,71*	0,38*	0,08 ^{ns}	0,46*	0,07 ^{ns}
CCPM	-	1,00	0,21*	0,51*	0,24*	0,03 ^{ns}	0,62*	0,01 ^{ns}
CHPM	-	-	1,00	0,54*	0,32*	0,06 ^{ns}	-0,03 ^{ns}	0,16 ^{ns}
MFPM	-	-	-	1,00	0,53*	0,13 ^{ns}	0,29*	-0,03 ^{ns}
MFPT	-	-	-	-	1,00	0,89*	0,05 ^{ns}	-0,05 ^{ns}
NCPT	-	-	-	-	-	1,00	-0,05 ^{ns}	-0,09 ^{ns}
NNPM	-	-	-	-	-	-	1,00	-0,10 ^{ns}
RFMM	-	-	-	-	-	-	-	1,00

¹Caracteres definidos na Tabela 3. ²Épocas de semeadura detalhadas na Tabela 2. *Significativo em nível de 5% de significância pelo teste *t* de Student. ^{ns} Não significativo.

Fonte: Autor.

Tabela 10 – Correlações entre pares de caracteres produtivos¹ de centeio em ensaios conduzidos diferentes épocas de semeadura com cultivares centeio (*Secale cereale* L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.

	CEPM	M100	MGPT	NGPM	NGPT	NSPM	NSPT	RGPM
	(continua)							
	----- Época 1 ² -----							
CEPM	1,00	0,34*	0,56*	0,70*	0,53*	0,87*	0,52*	-0,03 ^{ns}
M100	0,17 ^{ns}	1,00	0,49*	0,17 ^{ns}	0,26*	0,34*	0,33*	0,18 ^{ns}
MGPT	0,50*	0,48*	1,00	0,52*	0,96*	0,53*	0,90*	0,22*
NGPM	0,79*	0,08 ^{ns}	0,59*	1,00	0,54*	0,69*	0,30*	0,41*
NGPT	0,47*	-0,03 ^{ns}	0,83*	0,66*	1,00	0,49*	0,92*	0,20*
NSPM	0,88*	0,09 ^{ns}	0,47*	0,79*	0,50*	1,00	0,54*	0,02 ^{ns}
NSPT	0,34*	-0,02 ^{ns}	0,75*	0,36*	0,89*	0,42*	1,00	0,00 ^{ns}
RGPM	0,24*	0,62*	0,61*	0,52*	0,36*	0,21*	0,13 ^{ns}	1,00

Tabela 10 – Correlações entre pares de caracteres produtivos¹ de centeio em ensaios conduzidos diferentes épocas de semeadura com cultivares centeio (*Secale cereale* L.): BRS Progresso (diagonal superior) e Temprano (diagonal inferior), em ensaios conduzidos em Santa Maria – RS, Safra 2016.

	CEPM	M100	MGPT	NGPM	NGPT	NSPM	NSPT	RGPM
(conclusão)								
----- Época 2 -----								
CEPM	1,00	-0,11 ^{ns}	0,42*	0,50*	0,41*	0,64*	0,39*	0,00 ^{ns}
M100	0,35*	1,00	0,16 ^{ns}	-0,13 ^{ns}	-0,15 ^{ns}	0,02 ^{ns}	-0,12 ^{ns}	0,40*
MGPT	0,41*	0,67*	1,00	0,55*	0,93*	0,50*	0,89*	0,36*
NGPM	0,48*	0,40*	0,62*	1,00	0,60*	0,83*	0,35*	0,51*
NGPT	0,28*	0,14 ^{ns}	0,79*	0,58*	1,00	0,47*	0,93*	0,22*
NSPM	0,74*	0,19 ^{ns}	0,32*	0,52*	0,32*	1,00	0,35*	0,38*
NSPT	0,24*	-0,01 ^{ns}	0,59*	0,18 ^{ns}	0,84*	0,32*	1,00	0,06 ^{ns}
RGPM	0,30*	0,84*	0,76*	0,70*	0,39*	0,19 ^{ns}	0,08 ^{ns}	1,00
----- Época 3 -----								
CEPM	1,00	0,34*	0,29*	0,45*	0,22*	0,65*	0,18 ^{ns}	0,03 ^{ns}
M100	0,17 ^{ns}	1,00	0,38*	0,17 ^{ns}	0,13 ^{ns}	0,26*	0,11 ^{ns}	0,35*
MGPT	0,12 ^{ns}	0,59*	1,00	0,39*	0,95*	0,37*	0,90*	0,23*
NGPM	0,25*	0,67*	0,59*	1,00	0,38*	0,82*	0,14 ^{ns}	0,55*
NGPT	0,08 ^{ns}	0,28*	0,90*	0,46*	1,00	0,33*	0,95*	0,16 ^{ns}
NSPM	0,72*	-0,08 ^{ns}	-0,02 ^{ns}	0,01 ^{ns}	-0,02 ^{ns}	1,00	0,21*	0,22*
NSPT	0,06 ^{ns}	-0,24*	0,33*	-0,30*	0,55*	0,24*	1,00	-0,04 ^{ns}
RGPM	0,03 ^{ns}	0,82*	0,66*	0,87*	0,45*	-0,24*	-0,29*	1,00
----- Época 4 -----								
CEPM	1,00	0,51*	0,49*	0,68*	0,37*	0,83*	0,35*	0,15 ^{ns}
M100	-	1,00	0,45*	0,12 ^{ns}	0,10 ^{ns}	0,29*	0,12 ^{ns}	0,38*
MGPT	-	-	1,00	0,42*	0,91*	0,41*	0,89*	0,28*
NGPM	-	-	-	1,00	0,44*	0,81*	0,31*	0,28*
NGPT	-	-	-	-	1,00	0,39*	0,97*	0,13 ^{ns}
NSPM	-	-	-	-	-	1,00	0,35*	0,05 ^{ns}
NSPT	-	-	-	-	-	-	1,00	0,06 ^{ns}
RGPM	-	-	-	-	-	-	-	1,00
----- Época 5 -----								
CEPM	1,00	0,36*	0,48*	0,77*	0,45*	0,76*	0,41*	0,05 ^{ns}
M100	-	1,00	0,51*	0,43*	0,26*	0,28*	0,20 ^{ns}	0,70*
MGPT	-	-	1,00	0,49*	0,94*	0,41*	0,89*	0,35*
NGPM	-	-	-	1,00	0,45*	0,75*	0,31*	0,34*
NGPT	-	-	-	-	1,00	0,40*	0,96*	0,19 ^{ns}
NSPM	-	-	-	-	-	1,00	0,44*	0,07 ^{ns}
NSPT	-	-	-	-	-	-	1,00	0,09 ^{ns}
RGPM	-	-	-	-	-	-	-	1,00

¹Caracteres definidos na Tabela 3. ²Épocas de semeadura detalhadas na Tabela 2. *Significativo em nível de 5% de significância pelo teste *t* de Student. ^{ns} Não significativo.

Fonte: Autor.

Altas magnitudes de correlações entre determinados pares de caracteres morfológicos e produtivos na cultivar BRS Progresso (r_{PROG}) e Temprano (r_{TEMP}) foram observadas. Entre caracteres morfológicos, o NCPT está fortemente relacionado com MFPT ($0,82 \leq r_{\text{PROG}} \leq 0,89$; $0,60 \leq r_{\text{TEMP}} \leq 0,89$); APPM com CCPM ($0,83 \leq r_{\text{PROG}} \leq 0,89$; $0,74 \leq r_{\text{TEMP}} \leq 0,88$), com CHPM ($0,45 \leq r_{\text{PROG}} \leq 0,77$; $0,54 \leq r_{\text{TEMP}} \leq 0,80$) e com MFPM ($0,44 \leq r_{\text{PROG}} \leq 0,71$; $0,49 \leq r_{\text{TEMP}} \leq 0,68$). As demais correlações em todos os ensaios de uma mesma cultivar foram não significativas em todos ou ao menos em um ensaio (época de semeadura).

Maior número de correlações significativas foram verificados entre pares de caracteres produtivos. Da maneira semelhante ao observado nas correlações entre pares de caracteres morfológicos, também para os caracteres produtivos foram observadas correlações significativas em todos os ensaios, em outros momentos, somente em um ou outro ensaio dentro da mesma cultivar. De maneira geral, pode ser observado que os caracteres do grupo de variáveis produtivas estão mais fortemente relacionados uns com os outros. A época de semeadura pode também ter efeito sobre as estimativas das correlações, principalmente na magnitude e em sua significância. Foi verificado, principalmente, nos pares de caracteres de magnitude intermediária, que dependendo do ensaio, as correlações podem ou não ser significativas a 5% de probabilidade de erro pelo teste *t* de *Student*.

4.3 MULTICOLINEARIDADE EM CARACTERES DE CENTEIO

4.3.1 Grau de multicolinearidade em grupos de variáveis

Em diagnósticos realizados com todos os caracteres morfológicos e os caracteres produtivos (Tabela 3), foram verificadas moderada à forte ($100 < NC \leq 1.000$) e severa multicolinearidade ($NC > 1.000$) pelo indicador número de condição (MONTGOMERY *et al.*, 2012) e presença em nível prejudicial pelo fator de inflação da variância ($FIV_j > 10$) (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009) em cada grupo de variáveis e em todos os ensaios (Tabela 11). O maior grau de multicolinearidade foi verificado em diagnóstico realizado com os oito caracteres morfológicos ($NC = 1.079,55$; $FIV_{\text{máx}} = 184,70$) e oito caracteres produtivos ($NC = 1.207,95$; $FIV_{\text{máx}} = 155,89$), em ensaio com a semeadura da cultivar BRS Progresso realizada no início de julho (E5: 04/07 – quinta época). Considerando somente os ensaios com a cultivar Temprano, as maiores estimativas foram observadas em ensaio na primeira época (ensaio E6) para os grupos de variáveis morfológicas ($NC = 872,73$; $FIV_{\text{máx}} = 157,76$) e produtivas ($NC = 396,27$; $FIV_{\text{máx}} = 49,07$).

Tabela 11 – Grau de multicolinearidade estimado pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*) em caracteres morfológicos e produtivos de duas cultivares de centeio (*Secale cereale* L.), considerando todos os oito e seis caracteres (eliminação de caracteres com $FIV_j > 10$), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Ensaio ¹	<i>NC</i>	<i>FIV</i> – Fator de inflação da variância							
			----- Caracteres morfológicos ² -----							
			NCPT	NNPM	CCPM	CHPM	APPM	MFPT	MFPM	RFMM
BRS Progresso	1	528,04	12,37	1,38	83,06	23,19	107,19	19,12	8,75	1,30
		3,64	1,15	1,31	1,36	1,26	-	-	1,37	1,11
	2	310,48	15,04	1,87	38,77	15,91	55,59	19,70	5,67	1,09
		5,86	1,04	1,82	1,57	1,49	-	-	1,54	1,08
	3	694,24	19,19	1,68	85,59	23,18	120,95	25,27	8,15	1,08
		5,92	1,13	1,67	1,87	1,14	-	-	1,52	1,05
	4	886,74	11,79	1,50	71,47	46,71	169,03	17,24	8,75	1,32
		5,19	1,02	1,46	1,54	1,40	-	-	1,37	1,20
	5	1.079,55	18,00	1,77	103,15	37,62	184,70	24,57	7,24	1,10
		6,94	1,04	1,74	2,05	1,56	-	-	1,95	1,07
Temprano	6	872,73	9,76	1,80	97,13	28,01	157,76	12,33	8,56	1,20
		6,98	1,15	1,79	1,74	1,43	-	-	1,77	1,08
	7	360,24	19,54	1,43	50,23	18,14	74,65	22,86	5,94	2,11
		8,30	1,02	1,42	1,88	1,80	-	-	2,12	2,08
	8	816,59	14,74	1,49	52,69	61,79	146,91	13,41	5,75	2,11
		10,14	1,13	1,40	1,54	2,49	-	-	2,41	1,86
			----- Caracteres produtivos -----							
				CEPM	NSPT	NSPM	NGPT	NGPM	MGPT	RGPM
BRS Progresso	1	623,25	5,31	59,09	14,75	73,94	20,23	43,91	1,82	4,19
		25,01	5,29	1,52	4,61	-	3,75	-	1,82	1,34
	2	425,88	2,21	37,28	8,01	61,86	19,77	36,42	2,43	4,61
		25,61	2,00	1,24	4,68	-	5,20	-	2,25	1,64
	3	708,82	1,98	66,93	10,71	113,68	21,67	42,72	2,64	4,88
		34,67	1,96	1,07	5,89	-	6,67	-	2,61	1,55
	4	913,82	4,61	90,80	10,44	122,93	16,73	24,99	1,72	5,27
		28,13	4,43	1,16	5,89	-	4,37	-	1,65	2,04
	5	1.207,95	3,81	98,13	12,84	155,89	19,90	32,53	2,73	5,43
		23,98	3,75	1,30	3,10	-	4,01	-	2,71	2,50
Temprano	6	396,27	6,26	26,56	10,89	49,07	20,80	18,52	4,53	7,57
		45,09	6,05	1,25	5,77	-	7,59	-	4,37	2,85
	7	259,72	2,66	13,74	3,71	34,41	9,90	19,26	10,00	11,91
		53,54	2,53	1,16	2,94	-	4,48	-	9,74	5,80
	8	177,16	2,60	4,80	3,24	23,69	9,48	20,74	12,43	6,45
		8,40	2,44	1,18	2,41	-	1,96	-	-	1,86

¹Ensaios detalhados na Tabela 2. ²Caracteres definidos na Tabela 3.

Fonte: Autor.

Entre os ensaios de uma mesma cultivar, as menores estimativas de multicolinearidade foram observadas em caracteres morfológicos em ensaios semeados no final de maio (segunda época) com a cultivar BRS Progresso (ensaio E2: $NC = 310,48$; $FIV_{máx} = 55,59$) e com a cultivar Temprano (ensaio E7: $NC = 360,24$; $FIV_{máx} = 74,65$). Já em diagnóstico realizado em caracteres produtivos, as menores estimativas foram observadas em ensaios de segunda época de semeadura com a cultivar BRS Progresso (ensaio E2: $NC = 425,88$; $FIV_{máx} = 61,86$) e terceira época com a cultivar Temprano (ensaio E8: $NC = 177,16$; $FIV_{máx} = 23,69$).

Independentemente da época de semeadura, cultivar ou grupo de variáveis, foi verificada a existência de multicolinearidade em elevado grau quando da utilização de todos os oito caracteres. O fator de inflação da variância (FIV) proporciona a identificação do caractere que está causando a multicolinearidade, representando a proporção da inflação em relação a situação de ausência de multicolinearidade (GUJARATI; PORTER, 2011). Depois de identificar os caracteres responsáveis por causarem a multicolinearidade em cada grupo de variáveis e proceder a eliminação de dois caracteres em cada grupo de variáveis morfológicas e produtivas nos ensaios em ambas as cultivares, com exceção do ensaio semeadado em início de junho (E8: 07/06 – terceira época) com a cultivar Temprano, foi verificada a ocorrência de multicolinearidade fraca ($NC \leq 100$) (MONTGOMERY *et al.*, 2012) e estimativas de FIV_j menores ou iguais a dez ($FIV_j \leq 10$) (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009).

No ensaio E8 somente a eliminação de dois caracteres não foi suficiente para o atendimento do critério de $FIV_j \leq 10$. Assim, houve a necessidade de eliminação também do caractere RGPM ($NC = 53,33$ e $FIV_j = 11,77$). Após a eliminação do caractere RGPM no ensaio E8, foram observadas estimativas $NC = 8,40$ e $FIV_{máx} = 2,44$. Assim, eliminações dos caracteres APPM e MFPT no grupo de variáveis morfológicas e dos caracteres NGPT e MGPT no grupo de variáveis produtivas foram suficientes para a redução no grau de multicolinearidade em todos os ensaios (época de semeadura \times cultivar), exceto na terceira época da Temprano (ensaio E8) no grupo de variáveis produtivas.

4.3.2 Grau de multicolinearidade estimado em combinações de caracteres

A determinação do grau de multicolinearidade pelos indicadores número de condição (NC) e fator de inflação da variância (FIV) foram realizadas em diferentes combinações de caracteres nos dois grupos de variáveis e em todos os ensaios (época de semeadura \times cultivar). De maneira geral, foram observadas variabilidades nas estimativas do grau de multicolinearidade nos diferentes casos obtidos pelas combinações de oito caracteres

morfológicos ou produtivos e em duas cultivares de centeio. As menores estimativas em ambos os grupos de variáveis foram observadas nas combinações do Caso 2 ($C_{(p0,pi)} = C_{(8,2)} = 28$ combinações), ou seja, combinações de oito caracteres dois a dois em grupos de variáveis morfológicas (Tabela 12) e produtivas (Tabela 13). Entre os casos, de maneira geral, as maiores estimativas puderam ser verificadas nas combinações do caso 7 ($C_{(p0,pi)} = C_{(8,7)} = 8$ combinações), ou seja, em combinações contendo o maior número de caracteres – sete caracteres (Tabela 14). Independentemente de cultivar, grupo de variável e indicador, há aumento na estimativa do grau de multicolinearidade à medida que os casos são constituídos pelas combinações de maior número de caracteres (Tabela 15). A visão geral das estimativas de multicolinearidade pelos indicadores *NC* e *FIV* em cada combinação em cada caso, grupo de variáveis, época de semeadura e cultivar podem ser visualizadas na Figura 2.

Em todas as combinações e ensaios no Caso 2, fraca multicolinearidade ($NC \leq 100$) (MONTGOMERY *et al.*, 2012) e estimativas de $FIV_j \leq 10$ (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009) foram verificadas no grupo de variáveis morfológicas. Em combinações com caracteres produtivos, também foi verificada fraca multicolinearidade pelo indicador *NC*, porém, com existência de multicolinearidade pelo indicador *FIV* ($FIV_j > 10$) em quatro das 224 combinações (E1, E5 e duas em E3).

Foi observado comportamento semelhante das estimativas de multicolinearidade em combinações de caracteres morfológicos no Caso 2, com estimativas médias de 2,61 em ambas as cultivares ($1,0 \leq NC_{BRSP} \leq 16,6$; $1,0 \leq NC_{TEMP} \leq 17,6$) pelo indicador *NC* e amplitude de 15,6 e 16,6 as cultivares BRS Progresso (BRSP) e Temprano (TEMP), respectivamente. Amplitudes de magnitudes semelhantes pelo indicador *FIV* também puderam ser verificadas em diagnósticos realizados em combinações com caracteres morfológicos na cultivar BRS Progresso (3,7) e Temprano (3,9), assim como, para as respectivas médias de 1,31 ($1,0 \leq FIV_{BRSP} \leq 4,7$) e 1,30 ($1,0 \leq FIV_{TEMP} \leq 4,9$).

Contudo, divergência de magnitude para as estimativas médias e de amplitude de multicolinearidade entre as duas cultivares puderam ser observadas em combinações de caracteres do grupo de variáveis produtivas. Maiores amplitudes de multicolinearidade estimadas por *NC* (77,3) e *FIV* (18,8) foram observadas entre as estimativas em combinações de caracteres produtivos de BRS Progresso, correspondendo a amplitudes de aproximadamente 4,11 e 4,27 vezes maiores às amplitudes obtidas entre as combinações de caracteres produtivos na cultivar Temprano, respectivamente. Para esse mesmo grupo de variáveis e caso, foram observadas médias de 5,85 ($1,0 \leq NC_{BRSP} \leq 78,3$) e 3,87 ($1,0 \leq NC_{TEMP} \leq 19,8$) por *NC* e médias de 2,08 ($1,0 \leq FIV_{BRSP} \leq 20,1$) e 1,58 ($1,0 \leq FIV_{TEMP} \leq 5,5$) por *FIV*.

Tabela 12 – Estimativa do número de condição (*NC*) e do maior valor do fator de inflação da variância (*FIV*) em 28 combinações (Comb.) no caso 2, com oito caracteres morfológicos combinados dois a dois ($C_{(8,2)}$) em ensaios de uniformidade¹ (E1, E2, ..., E8) com cultivares de centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Comb.	Caso 2 ²	Número de condição (<i>NC</i>)								Fator de Inflação da Variância (<i>FIV</i>)							
		BRS Progresso					Temprano			BRS Progresso					Temprano		
		E1	E2	E3	E4	E5	E6	E7	E8	E1	E2	E3	E4	E5	E6	E7	E8
1	X1 ³ × X2	12,0	11,1	16,6	11,8	14,8	15,1	11,4	6,6	3,5	3,3	4,7	3,5	4,2	4,3	3,4	2,2
2	X1 × X3	2,6	2,6	3,0	7,8	4,6	3,7	3,4	8,9	1,3	1,3	1,3	2,5	1,7	1,5	1,4	2,8
3	X1 × X4	2,6	4,1	4,3	3,6	6,0	3,9	2,9	5,2	1,2	1,6	1,6	1,5	2,0	1,5	1,3	1,8
4	X1 × X5	1,3	2,0	1,9	1,8	2,2	1,7	1,4	1,3	1,0	1,1	1,1	1,1	1,2	1,1	1,0	1,0
5	X1 × X6	1,4	1,1	1,0	1,0	1,2	1,9	1,1	1,6	1,0	1,0	1,0	1,0	1,0	1,1	1,0	1,1
6	X1 × X7	1,8	1,7	2,7	1,7	2,7	2,9	1,6	1,5	1,1	1,1	1,3	1,1	1,3	1,3	1,1	1,0
7	X1 × X8	1,3	1,3	1,2	1,6	1,2	1,1	1,4	1,7	1,0	1,0	1,0	1,1	1,0	1,0	1,0	1,1
8	X2 × X3	1,2	1,2	1,1	2,0	1,5	1,3	1,0	1,5	1,0	1,0	1,0	1,1	1,0	1,0	1,0	1,0
9	X2 × X4	1,3	2,7	2,9	2,2	3,1	2,1	1,3	1,7	1,0	1,3	1,3	1,2	1,4	1,1	1,0	1,1
10	X2 × X5	1,3	1,7	1,6	1,5	1,6	1,2	1,1	1,1	1,0	1,1	1,1	1,0	1,1	1,0	1,0	1,0
11	X2 × X6	1,8	1,2	1,1	1,0	1,1	1,7	1,2	1,3	1,1	1,0	1,0	1,0	1,0	1,1	1,0	1,0
12	X2 × X7	2,5	3,0	3,7	2,5	4,3	4,1	2,5	2,3	1,2	1,3	1,5	1,2	1,6	1,6	1,2	1,2
13	X2 × X8	1,3	1,4	1,2	1,7	1,0	1,0	2,8	1,3	1,0	1,0	1,0	1,1	1,0	1,0	1,3	1,0
14	X3 × X4	2,4	1,6	1,9	2,4	3,4	2,8	3,8	5,7	1,2	1,1	1,1	1,2	1,4	1,3	1,5	2,0
15	X3 × X5	1,8	1,1	1,6	1,4	1,9	1,8	1,8	1,3	1,1	1,0	1,1	1,0	1,1	1,1	1,1	1,0
16	X3 × X6	1,2	1,2	1,2	1,1	1,1	1,4	1,1	1,6	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,1
17	X3 × X7	1,4	2,3	1,1	1,1	1,1	1,1	1,8	1,1	1,0	1,2	1,0	1,0	1,0	1,0	1,1	1,0
18	X3 × X8	1,1	1,1	1,1	1,3	1,4	1,0	2,2	3,3	1,0	1,0	1,0	1,0	1,0	1,0	1,2	1,4
19	X4 × X5	4,3	2,8	2,9	3,6	3,2	3,3	2,2	1,3	1,6	1,3	1,3	1,5	1,4	1,4	1,2	1,0
20	X4 × X6	1,3	1,1	1,1	1,2	1,3	1,8	1,1	2,0	1,0	1,0	1,0	1,0	1,0	1,1	1,0	1,1
21	X4 × X7	1,1	1,9	1,7	1,5	1,8	2,0	1,1	1,5	1,0	1,1	1,1	1,0	1,1	1,1	1,0	1,0
22	X4 × X8	1,8	1,0	1,4	1,1	1,1	1,5	2,7	3,1	1,1	1,0	1,0	1,0	1,0	1,0	1,3	1,4
23	X5 × X6	10,0	13,4	13,1	10,4	16,4	4,0	17,6	12,4	3,0	3,9	3,8	3,1	4,6	1,6	4,9	3,6
24	X5 × X7	1,4	1,2	1,2	1,1	1,1	1,4	1,1	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
25	X5 × X8	1,3	1,0	1,4	1,1	1,1	1,5	1,7	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,1	1,0
26	X6 × X7	1,7	1,1	1,8	1,1	1,1	1,2	1,2	1,3	1,1	1,0	1,1	1,0	1,0	1,0	1,0	1,0
27	X6 × X8	1,0	1,0	1,3	1,2	1,2	1,0	1,2	1,6	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,1
28	X7 × X8	1,1	1,5	1,0	2,0	1,2	1,0	1,6	1,3	1,0	1,0	1,0	1,1	1,0	1,0	1,1	1,0

¹Ensaio detalhados na Tabela 2. ²Caso definido conforme descrito no Capítulo 3.2.3. ³Caracteres definidos na Tabela 3.

Fonte: Autor.

Tabela 13 – Estimativa do número de condição (NC) e do maior valor do fator de inflação da variância (FIV) em 28 combinações (Comb.) no caso 2, de oito caracteres produtivos combinados dois a dois ($C_{(8,2)}$) em ensaios de uniformidade¹ (E1, E2, ..., E8) com cultivares de centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Comb.	Caso 2 ²	Número de Condição (NC)								Fator de Inflação da Variância (FIV)							
		BRS Progresso					Temprano			BRS Progresso					Temprano		
		E1	E2	E3	E4	E5	E6	E7	E8	E1	E2	E3	E4	E5	E6	E7	E8
1	Y1 ³ ×Y2	2,0	1,2	2,0	3,1	2,1	1,4	2,1	1,4	1,1	1,0	1,1	1,3	1,1	1,0	1,1	1,0
2	Y1×Y3	3,6	2,5	1,8	3,0	2,8	3,0	2,4	1,3	1,5	1,2	1,1	1,3	1,3	1,3	1,2	1,0
3	Y1×Y4	5,7	3,0	2,6	5,2	7,5	8,5	2,8	1,7	2,0	1,3	1,2	1,9	2,4	2,6	1,3	1,1
4	Y1×Y5	3,3	2,4	1,6	2,2	2,6	2,8	1,8	1,2	1,4	1,2	1,1	1,2	1,2	1,3	1,1	1,0
5	Y1×Y6	13,9	4,5	4,7	10,7	7,2	15,9	6,8	6,2	4,0	1,7	1,7	3,2	2,3	4,5	2,2	2,1
6	Y1×Y7	3,2	2,3	1,4	2,1	2,4	2,0	1,6	1,1	1,4	1,2	1,0	1,1	1,2	1,1	1,1	1,0
7	Y1×Y8	1,1	1,0	1,1	1,3	1,1	1,6	1,9	1,1	1,0	1,0	1,0	1,0	1,0	1,1	1,1	1,0
8	Y2×Y3	2,9	1,4	2,2	2,6	3,1	2,8	5,1	3,8	1,3	1,0	1,2	1,3	1,3	1,3	1,8	1,5
9	Y2×Y4	1,4	1,3	1,4	1,3	2,5	1,2	2,4	5,0	1,0	1,0	1,0	1,0	1,2	1,0	1,2	1,8
10	Y2×Y5	1,7	1,4	1,3	1,2	1,7	1,1	1,3	1,8	1,1	1,0	1,0	1,0	1,1	1,0	1,0	1,1
11	Y2×Y6	2,0	1,0	1,7	1,8	1,8	1,2	1,5	1,2	1,1	1,0	1,1	1,1	1,1	1,0	1,0	1,0
12	Y2×Y7	2,0	1,3	1,2	1,3	1,5	1,0	1,0	1,6	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,1
13	Y2×Y8	1,5	2,3	2,1	2,2	5,7	4,3	11,4	10,0	1,0	1,2	1,1	1,2	2,0	1,6	3,4	3,0
14	Y3×Y4	3,2	3,5	2,3	2,4	2,9	3,9	4,3	3,9	1,4	1,4	1,2	1,2	1,3	1,5	1,6	1,5
15	Y3×Y5	44,6	28,8	40,3	21,1	35,1	10,8	8,4	19,8	11,6	7,7	10,6	5,8	9,3	3,2	2,6	5,5
16	Y3×Y6	3,2	3,0	2,2	2,4	2,4	2,8	1,9	1,0	1,4	1,3	1,2	1,2	1,2	1,3	1,1	1,0
17	Y3×Y7	19,9	17,8	19,9	17,7	17,0	7,0	3,9	2,0	5,5	5,0	5,5	4,9	4,8	2,3	1,5	1,1
18	Y3×Y8	1,6	2,1	1,6	1,8	2,1	4,1	7,2	4,8	1,1	1,1	1,1	1,1	1,1	1,6	2,3	1,8
19	Y4×Y5	3,4	4,0	2,2	2,6	2,6	4,8	3,8	2,7	1,4	1,6	1,2	1,2	1,2	1,8	1,5	1,3
20	Y4×Y6	5,5	11,1	10,2	9,5	7,0	8,5	3,2	1,0	1,9	3,3	3,1	2,9	2,3	2,7	1,4	1,0
21	Y4×Y7	1,9	2,1	1,3	1,9	1,9	2,1	1,4	1,9	1,1	1,1	1,0	1,1	1,1	1,1	1,0	1,1
22	Y4×Y8	2,4	3,1	3,5	1,8	2,0	3,1	5,6	14,8	1,2	1,4	1,4	1,1	1,1	1,4	2,0	4,2
23	Y5×Y6	3,0	2,8	2,0	2,3	2,3	3,0	1,9	1,0	1,3	1,3	1,1	1,2	1,2	1,3	1,1	1,0
24	Y5×Y7	23,1	28,5	40,0	78,3	51,6	17,3	11,8	3,4	6,3	7,6	10,5	20,1	13,4	4,8	3,5	1,4
25	Y5×Y8	1,5	1,6	1,4	1,3	1,5	2,1	2,3	2,6	1,0	1,1	1,0	1,0	1,0	1,1	1,2	1,3
26	Y6×Y7	3,4	2,1	1,5	2,1	2,6	2,5	1,9	1,6	1,4	1,1	1,0	1,1	1,2	1,2	1,1	1,1
27	Y6×Y8	1,0	2,2	1,6	1,1	1,2	1,5	1,5	1,6	1,0	1,2	1,1	1,0	1,0	1,0	1,0	1,1
28	Y7×Y8	1,0	1,1	1,1	1,1	1,2	1,3	1,2	1,8	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,1

¹Ensaio detalhado na Tabela 2. ²Caso definido conforme descrito no Capítulo 3.2.3. ³Caracteres definidos na Tabela 3.

Fonte: Autor.

Tabela 14 – Estimativa do número de condição (*NC*) e do maior valor do fator de inflação da variância (*FIV*) em oito combinações (Comb.) no caso 7, de oito caracteres morfológicos e oito caracteres produtivos combinados sete a sete ($C_{(8,7)}$) em ensaios de uniformidade¹ (E1, E2, ..., E8) com cultivares de centeio centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Comb.	CNP ²	Cultivar BRS Progresso					Cultivar Temprano		
		E1	E2	E3	E4	E5	E6	E7	E8
----- Número de condição (<i>NC</i>) em caracteres morfológicos ³ -----									
1	RFMM	436,0	306,3	666,7	791,8	1067,5	791,1	350,0	683,5
2	NNPM	525,6	290,5	650,9	843,4	1007,9	795,0	359,8	763,4
3	NCPT	517,7	299,5	659,0	848,5	1056,0	815,1	352,2	786,6
4	MFPT	508,4	273,8	619,7	810,2	986,4	802,1	338,2	811,2
5	MFPM	398,4	257,1	452,6	520,9	736,7	678,2	301,9	543,7
6	CHPM	85,2	107,1	134,8	91,1	140,3	79,0	105,4	76,7
7	CCPM	89,2	87,9	117,3	82,8	130,2	70,1	114,7	89,6
8	APPM	87,8	85,3	108,1	75,6	117,2	60,7	111,3	73,2
----- Número de condição (<i>NC</i>) em caracteres produtivos -----									
1	RGPM	615,2	392,3	688,1	855,4	1170,8	370,8	226,9	137,4
2	NSPT	296,5	247,3	271,1	172,4	209,1	149,3	140,4	120,4
3	NSPM	317,3	258,9	406,2	501,1	349,4	207,8	222,4	145,8
4	NGPT	281,1	189,7	186,0	147,8	144,7	123,0	80,7	61,7
5	NGPM	318,2	226,3	340,8	279,0	294,2	169,3	159,0	114,0
6	MGPT	505,2	331,6	485,7	697,4	717,9	311,4	126,3	56,8
7	M100	585,7	422,8	613,7	870,5	955,2	356,8	171,0	109,3
8	CEPM	539,5	376,4	652,9	765,6	1050,9	345,4	232,8	173,3
----- Fator de inflação da variância (<i>FIV</i>) em caracteres morfológicos -----									
1	RFMM	91,6	55,4	117,8	154,6	182,9	143,9	73,7	132,7
2	NNPM	106,9	54,5	119,8	166,8	183,0	157,8	74,6	139,8
3	NCPT	106,8	55,3	117,5	164,3	184,7	149,1	74,6	146,4
4	MFPT	107,2	54,2	116,7	165,4	184,3	148,6	74,4	145,6
5	MFPM	82,7	53,6	94,9	113,3	149,3	144,5	65,6	119,5
6	CHPM	19,1	19,4	24,2	17,0	24,6	11,9	22,8	14,7
7	CCPM	19,1	19,2	24,2	16,8	24,5	11,7	22,8	14,7
8	APPM	19,1	19,2	24,4	16,9	24,5	11,6	22,8	14,7
----- Fator de inflação da variância (<i>FIV</i>) em caracteres produtivos -----									
1	RGPM	73,9	58,9	113,0	118,2	155,8	47,8	34,2	23,5
2	NSPT	41,6	33,7	42,7	24,9	29,3	18,4	18,5	18,1
3	NSPM	45,9	45,6	78,9	84,9	56,1	33,1	30,9	19,9
4	NGPT	33,8	27,2	27,9	20,0	18,6	14,6	10,2	12,3
5	NGPM	42,0	36,2	61,8	47,9	47,2	25,1	18,9	18,1
6	MGPT	56,9	46,2	74,3	98,1	89,3	38,7	18,3	12,1
7	M100	60,0	56,0	88,3	116,0	120,6	36,5	22,5	14,7
8	CEPM	73,7	58,1	112,3	118,0	153,6	48,8	33,5	23,3

¹Ensaios detalhados na Tabela 2. ²Caractere não presente na combinação de oito caracteres sete a sete ($C_{(p\theta, p)} = C_{(8,7)} = 8$ combinações). ³Caracteres em cada grupo de variáveis estão definidos na Tabela 3.

Fonte: Autor.

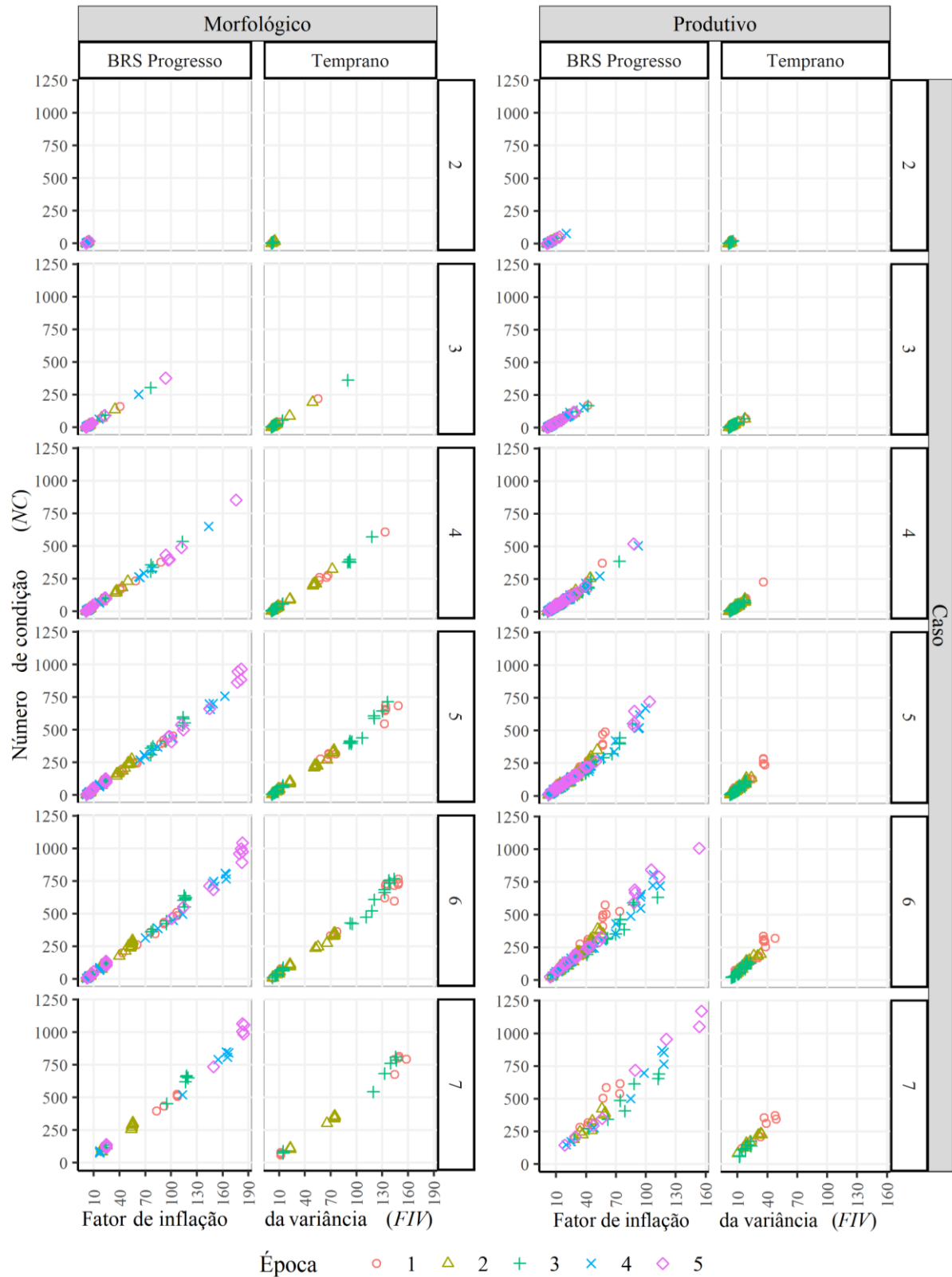
Tabela 15 – Estatísticas descritivas de mínimos (Mín), média (\bar{m}), mediana (Med), máximo (Máx) e amplitude (Amp = Máx - Mín) de número de condição (NC) e fator de inflação da variância (FIV), e percentual de combinações com problemas de multicolinearidade (PCMD) em casos (C) de combinações, grupos de variáveis e cultivares de centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

C ¹	N ^o Comb. ²	NC - Número de Condição						FIV - Fator de Inflação da Variância					
		Mín	\bar{m}	Med	Máx	Amp	PCMD ³	Min	\bar{m}	Med	Máx	Amp	PCMD ³
----- Cultivar BRS Progresso – Grupo morfológico ⁴ -----													
2	28	1,0	2,61	1,45	16,6	15,6	0,00	1,0	1,31	1,03	4,7	3,7	0,00
3	56	1,1	11,07	3,2	377,1	376,0	1,79	1,0	3,32	1,37	93,7	92,7	3,57
4	70	1,5	37,3	6,43	853,0	851,6	8,00	1,0	9,17	2,04	175,5	174,4	14,29
5	56	2,7	100,2	22,52	965,2	962,5	23,21	1,3	22,15	5,4	181,5	180,2	35,71
6	28	3,6	223,81	94,05	1043,7	1040,1	49,29	1,4	45,94	19,16	182,9	181,6	67,86
7	8	75,6	426,78	352,37	1067,5	991,9	80,00	16,8	82,39	69,05	184,7	167,9	100,00
----- Cultivar BRS Progresso – Grupo produtivo -----													
2	28	1,0	5,85	2,25	78,3	77,3	0,00	1,0	2,08	1,17	20,1	19,1	3,57
3	56	1,5	18,95	6,61	170,3	168,9	2,86	1,0	5	2,1	42,0	41,0	13,57
4	70	2,6	50,57	27,67	519,5	516,9	14,57	1,2	10,87	5,77	92,3	91,1	35,14
5	56	7,1	120	69,54	723,0	716,0	40,71	2,0	21,6	13,26	103,8	101,9	63,21
6	28	23,6	251,1	198,46	1008,5	984,9	75,71	3,9	39,53	33,03	153,5	149,6	87,86
7	8	144,7	468,25	384,32	1170,8	1026,2	100,00	18,6	67,18	57,49	155,8	137,1	100,00
----- Cultivar Temprano – Grupo morfológico -----													
2	28	1,0	2,61	1,6	17,6	16,6	0,00	1,0	1,3	1,06	4,9	3,9	0,00
3	56	1,1	11,19	3,72	360,9	359,8	1,79	1,0	3,34	1,47	90,0	89,0	3,57
4	70	1,5	36,91	7,63	608,4	606,9	7,14	1,0	9,1	2,27	133,0	131,9	14,29
5	56	4,5	97,66	27,39	713,7	709,3	19,05	1,6	21,7	7,09	148,4	146,7	36,31
6	28	7,0	216,94	74,24	767,5	760,5	41,67	1,8	44,83	14,6	148,9	147,1	69,05
7	8	60,7	414,69	351,12	815,1	754,4	75,00	11,6	80,77	74,47	157,8	146,1	100,00
----- Cultivar Temprano – Grupo produtivo -----													
2	28	1,0	3,87	2,33	19,8	18,8	0,00	1,0	1,58	1,19	5,5	4,5	0,00
3	56	1,8	11,25	7,68	70,4	68,6	0,00	1,1	3,24	2,4	18,1	17,0	2,98
4	70	3,7	27,03	19,83	228,4	224,7	0,95	1,4	6,1	4,72	35,8	34,4	14,76
5	56	8,4	57,51	43,19	287,7	279,3	17,26	2,4	10,46	8,52	37,2	34,7	39,88
6	28	18,0	107,79	104,38	336,3	318,2	51,19	3,9	16,59	17,27	47,8	43,9	75,00
7	8	56,8	179,67	154,14	370,8	314,0	87,50	10,2	24,67	21,2	48,9	38,6	100,00

¹Caso: descrito no Capítulo 3.2.3. ²Número de combinações. ³Percentual de combinações com multicolinearidade detectada (PCMD): $NC > 100$ (MONTGOMERY *et al.*, 2012) e $FIV_j > 10$ (GUJARATI; PORTER, 2011; HAIR *et al.*, 2009). ⁴Caracteres definidos na Tabela 3.

Fonte: Autor.

Figura 2 – Relação entre as estimativas de multicolinearidade obtidas pelos indicadores número de condição (NC) e fator de inflação da variância (FIV) em combinações de caracteres morfológicos e produtivos em cultivares de centeio (*Secale cereale* L.) conduzidos em diferentes épocas de semeadura em Santa Maria – RS, Safra 2016.



Estimativas de valores de mínimos, médios e máximos de multicolinearidade obtidas por NC ou FIV aumentaram gradativamente, em casos com maior número de caracteres presentes em cada combinação. De maneira geral, o incremento foi na proporção de duas à quatro vezes no valor médio da multicolinearidade, quando comparado com o valor médio estimado no caso com um caractere a menos na combinação. Ou seja, na média, para cada caractere adicionada à matriz de correlação, a multicolinearidade pode ser aumentada em duas a quatro vezes.

Em combinações no Caso 7, constituídos pelas combinações de sete caracteres, estimativas de fraca ($NC \leq 100$), moderada à forte ($100 < NC \leq 1.000$) e severa ($NC > 1.000$) multicolinearidade (MONTGOMERY *et al.*, 2012) foram verificadas. Valores de $NC \leq 100$ foram observados em oito de 40 e em seis de 24 combinações no grupo de variáveis morfológicas na BRS Progresso (ensaios E1, E2 e E4) e Temprano (ensaios E6 e E8), respectivamente. Já em caracteres produtivos com a cultivar Temprano, foi constatado valores de $NC \leq 100$ somente em três de 24 combinações. Severa multicolinearidade ($NC > 1.000$) foi diagnosticada em três e duas combinações de caracteres morfológicos e produtivos em ensaio semeador em início de julho (ensaio E5: 04/07 – quinta época de semeadura) com a cultivar BRS Progresso. No entanto, independentemente do grupo de variáveis e de cultivar, multicolinearidade foi detectada em todas as combinações no Caso 7 pelo indicador FIV ($FIV_j > 10$).

Com a ocorrência de maiores estimativas de multicolinearidade em combinações em casos de maior número de caracteres combinados, foram observadas estimativas de médias de 426,78 ($75,6 \leq NC_{BRSP} \leq 1.067,5$) e 414,69 ($60,7 \leq NC_{TEMP} \leq 815,13$) entre as combinações de caracteres morfológicos na BRS Progresso e Temprano pelo indicador NC , correspondendo a amplitudes de 991,9 e 754,4, respectivamente. Em combinações de caracteres produtivos na BRS Progresso e Temprano, as médias foram de 384,32 ($144,7 \leq NC_{BRSP} \leq 1.170,8$) e 179,67 ($56,8 \leq NC_{TEMP} \leq 370,8$) com amplitudes de 1.026,2 e 314, respectivamente.

Ainda para estimativas de multicolinearidade, porém obtidas pelo indicador FIV , foram observadas médias de 82,39 ($16,8 \leq FIV_{BRSP} \leq 184,7$) e 80,77 ($60,7 \leq FIV_{TEMP} \leq 157,8$) e amplitudes de 167,9 e 146,1 em caracteres morfológicos nas cultivares BRS Progresso e Temprano, respectivamente. Em caracteres produtivos, foram observadas estimativas médias de 67,18 ($18,6 \leq FIV_{BRSP} \leq 155,8$) e 24,67 ($10,2 \leq FIV_{TEMP} \leq 48,9$) e amplitudes de 137,1 e 38,6, respectivamente.

Ausência de problemas devidos à multicolinearidade é apontada para valores de $NC \leq 100$ – multicolinearidade fraca (MONTGOMERY *et al.*, 2012) e/ou $FIV \leq 10$ (GUJARATI;

PORTER, 2011; HAIR *et al.*, 2009). De maneira geral, maior percentual de combinações diagnosticadas com multicolinearidade foi obtido pelo indicador *FIV*, apesar de observada ausência de multicolinearidade por ambos os indicadores em todas 140 e 84 combinações de caracteres de cultivares BRS Progresso e Temprano, respectivamente do Caso 2, com exceção em 3,57% das combinações de caracteres produtivos na cultivar BRS Progresso ($1,0 \leq FIV \leq 20,1$).

Por outro lado, em 100% das combinações com sete caracteres (Caso 7) foram diagnosticadas com multicolinearidade quando realizada pelo indicador *FIV* em ambas as cultivares e grupos de variáveis. Pelo indicador *NC*, somente em combinações de caracteres produtivos na cultivar BRS Progresso foi detectada moderada à forte ou severa multicolinearidade em todas as combinações.

4.3.3 Considerações sobre a multicolinearidade em caracteres de centeio

A ocorrência de níveis elevados de multicolinearidade foram verificados em diagnósticos onde foram incluídos todos os oito caracteres morfológicos e produtivos (Tabela 3). A eliminação dos caracteres morfológicos altura de planta (APPM) e massa fresca de parte aérea planta⁻¹ (MFPT) e dos caracteres produtivos massa de grãos planta⁻¹ (MGPT) e número de grãos planta⁻¹ (NGPT) foi eficiente na redução do grau de multicolinearidade. Essa prática tem sido reportada em estudos com as culturas da aveia preta (MEIRA *et al.*, 2019), crotalária (TOEBE *et al.*, 2017a) e milho (ALVES *et al.*, 2016a; ALVES *et al.*, 2016b; OLIVOTO *et al.*, 2017; TOEBE *et al.*, 2017b; TOEBE; CARGNELUTTI FILHO, 2013), com redução do grau de multicolinearidade de níveis com $NC \geq 1.000$ ou $FIV > 10$ para valores de $NC \leq 100$ e $FIV \leq 10$.

Os estudos com aveia branca (MEIRA *et al.*, 2019) e crotalária (TOEBE *et al.*, 2017a) apontaram a ocorrência de alta correlação dos caracteres eliminados com os demais. Em centeio, também foi verificado níveis elevados de correlação linear (*r*) entre os caracteres eliminados e os demais caracteres morfológicos e/ou produtivos. No grupo de variáveis morfológicas e entre os ensaios em cada cultivar foram observadas correlações não significativas pelo teste *t* de *Student* em nível de 5% de probabilidade de erro ou de magnitude inferior a 0,38 entre os pares de caracteres APPM e MFPT na cultivar BRS Progresso ($0,12^{ns} \leq r \leq 0,38^*$)¹ e Temprano ($0,13^{ns} \leq r \leq 0,25^*$). No entanto, correlações de alta magnitudes ($r \leq$

¹ *Significativo nível de 5% de probabilidade de erro pelo teste *t* de *Student*. ^{ns}Não significativo.

0,89) foram verificadas de APPM e MFPT com os demais caracteres. Já no grupo de variáveis produtivas foi verificada alta magnitude das correlações entre os dois caracteres eliminados (MGPT e NGPT), com estimativas de $0,91 \leq r \leq 0,96$ e $0,79 \leq r \leq 0,90$ na BRS Progresso e Temprano, respectivamente.

4.4 REPETIBILIDADE E NÚMERO DE ENSAIOS PARA DIAGNÓSTICO DE MULTICOLINEARIDADE

4.4.1 Coeficiente de repetibilidade para o diagnóstico de multicolinearidade

Os valores dos coeficientes de repetibilidade (*rptb*) para a estimativa do grau de multicolinearidade obtidas pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*), por diferentes metodologias e diferentes agrupamentos de variáveis morfológicas e produtivas, em duas cultivares de centeio estão apresentados na Tabela 16. De maneira geral, independentemente do indicador, os valores de *rptb* da multicolinearidade, estimados pelo método da análise de variância (AOV) foram inferiores aos valores de *rptb* obtidos por meio dos demais métodos utilizados no estudo. No presente estudo, menores coeficientes de repetibilidade por AOV foram observados nos Casos 2 e 7. Também pode ser observado a diminuição no ajuste do modelo por AOV a medida que maior foi o número de caracteres combinados, independente do indicador, grupo de variáveis e cultivar.

Estudos de repetibilidade para estimação de multicolinearidade não foram encontrados. No entanto, menores estimativas de *rptb* pelo método AOV também foram encontrados em estudos com outras culturas. Neles, menores valores de *rptb* foram obtidos com características em soja (*Glycine max*) (DUARTE; FERREIRA; SILVA, 2022; MATSUO *et al.*, 2012), em parâmetros genéticos em capim-elefante (*Pennisetum spp.*) (CAVALCANTE *et al.*, 2012; SOUZA *et al.*, 2017), couve (*Brassica oleracea* var. *acephala*) (AZEVEDO *et al.*, 2016), braquiária (*Urochloa brizantha*) (TORRES *et al.*, 2015) e em morangueiro (*Fragaria x ananassa*) (DIEL *et al.*, 2020).

Tabela 16 – Coeficientes de repetibilidade (*rptb*) e de determinação (R^2) do grau de multicolinearidade obtido pelos indicadores número de condição (*NC*) e fator de inflação da variância (*FIV*), em diferentes métodos para diferentes grupos de caracteres¹ combinados em dois grupos de variáveis e em duas cultivares de centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Caso ²	Cultivar BRS Progresso								Cultivar Temprano							
	Morfológico				Produtivo				Morfológico				Produtivo			
	<i>NC</i>		<i>FIV</i>		<i>NC</i>		<i>FIV</i>		<i>NC</i>		<i>FIV</i>		<i>NC</i>		<i>FIV</i>	
	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2	<i>rptb</i>	R^2
----- AOV – Análise de variância -----																
2	0,923	0,984	0,922	0,983	0,817	0,957	0,813	0,956	0,650	0,848	0,630	0,836	0,498	0,748	0,478	0,733
3	0,859	0,968	0,860	0,968	0,849	0,966	0,851	0,966	0,874	0,954	0,876	0,955	0,746	0,898	0,762	0,906
4	0,838	0,963	0,852	0,966	0,885	0,975	0,869	0,971	0,892	0,961	0,894	0,962	0,622	0,832	0,683	0,866
5	0,825	0,959	0,842	0,964	0,880	0,973	0,858	0,968	0,871	0,953	0,883	0,958	0,484	0,738	0,567	0,797
6	0,812	0,956	0,828	0,960	0,831	0,961	0,806	0,954	0,832	0,937	0,855	0,947	0,409	0,675	0,524	0,768
7	0,792	0,950	0,803	0,953	0,707	0,924	0,696	0,919	0,785	0,916	0,816	0,930	0,438	0,701	0,612	0,825
----- PCR – Componentes principais com base na matriz de correlação -----																
2	0,948	0,989	0,950	0,990	0,885	0,975	0,883	0,974	0,659	0,853	0,641	0,843	0,556	0,790	0,539	0,778
3	0,983	0,997	0,984	0,997	0,879	0,973	0,880	0,973	0,965	0,988	0,965	0,988	0,753	0,901	0,767	0,908
4	0,979	0,996	0,977	0,995	0,917	0,982	0,902	0,979	0,960	0,986	0,955	0,984	0,687	0,868	0,712	0,881
5	0,980	0,996	0,978	0,996	0,933	0,986	0,915	0,982	0,969	0,989	0,966	0,988	0,612	0,826	0,642	0,844
6	0,983	0,997	0,982	0,996	0,937	0,987	0,917	0,982	0,978	0,992	0,979	0,993	0,575	0,803	0,635	0,839
7	0,988	0,998	0,991	0,998	0,961	0,992	0,953	0,990	0,992	0,997	0,996	0,999	0,621	0,831	0,757	0,903
----- PCS – Componentes principais com base na matriz de covariância -----																
2	0,954	0,991	0,957	0,991	0,872	0,972	0,871	0,971	0,702	0,876	0,691	0,870	0,512	0,759	0,494	0,746
3	0,991	0,998	0,991	0,998	0,873	0,972	0,877	0,973	0,973	0,991	0,973	0,991	0,747	0,898	0,764	0,907
4	0,988	0,998	0,986	0,997	0,914	0,981	0,903	0,979	0,967	0,989	0,961	0,987	0,724	0,887	0,730	0,890
5	0,989	0,998	0,988	0,998	0,932	0,986	0,918	0,982	0,977	0,992	0,972	0,991	0,727	0,889	0,710	0,880
6	0,992	0,998	0,992	0,998	0,942	0,988	0,927	0,985	0,987	0,995	0,986	0,995	0,722	0,886	0,768	0,908
7	0,994	0,999	0,996	0,999	0,969	0,994	0,954	0,990	0,992	0,997	0,995	0,998	0,706	0,878	0,841	0,941
----- SAR – Análise estrutural com base na matriz de correlação -----																
2	0,948	0,989	0,950	0,990	0,882	0,974	0,881	0,974	0,653	0,850	0,634	0,838	0,549	0,785	0,531	0,773
3	0,983	0,997	0,984	0,997	0,877	0,973	0,878	0,973	0,965	0,988	0,965	0,988	0,751	0,900	0,765	0,907
4	0,979	0,996	0,977	0,995	0,916	0,982	0,902	0,979	0,960	0,986	0,955	0,984	0,680	0,864	0,706	0,878
5	0,980	0,996	0,978	0,996	0,933	0,986	0,914	0,982	0,969	0,989	0,966	0,988	0,599	0,818	0,631	0,837
6	0,983	0,997	0,982	0,996	0,937	0,987	0,917	0,982	0,978	0,992	0,979	0,993	0,558	0,791	0,623	0,832
7	0,988	0,998	0,991	0,998	0,961	0,992	0,953	0,990	0,992	0,997	0,996	0,999	0,605	0,821	0,753	0,902
----- SAS – Análise estrutural com base na matriz de covariância -----																
2	0,923	0,984	0,922	0,983	0,817	0,957	0,813	0,956	0,650	0,848	0,630	0,836	0,498	0,748	0,478	0,733
3	0,859	0,968	0,860	0,968	0,849	0,966	0,851	0,966	0,874	0,954	0,876	0,955	0,746	0,898	0,762	0,906
4	0,838	0,963	0,852	0,966	0,885	0,975	0,869	0,971	0,892	0,961	0,894	0,962	0,622	0,832	0,683	0,866
5	0,825	0,959	0,842	0,964	0,880	0,973	0,858	0,968	0,871	0,953	0,883	0,958	0,484	0,738	0,567	0,797
6	0,812	0,956	0,828	0,960	0,831	0,961	0,806	0,954	0,832	0,937	0,855	0,947	0,409	0,675	0,524	0,768
7	0,792	0,950	0,803	0,953	0,707	0,924	0,696	0,919	0,785	0,916	0,816	0,930	0,438	0,701	0,612	0,825

¹Caracteres definidos na Tabela 3. ²Casos definidos conforme descrito no Capítulo 3.2.3.

Fonte: Autor.

Nos demais métodos, a repetibilidade do grau de multicolinearidade aumentou à medida que maior número de caracteres foram combinados. Ou seja, é esperado uma maior regularidade na estimativa do grau de multicolinearidade entre as sucessivas medições, quanto maior for o número de caracteres utilizados na estimativa do grau de multicolinearidade. Obter altas estimativas de repetibilidade é interessante, pois indica que é possível estimar o valor real de uma determinada característica com um número relativamente pequeno de medições (CARGNELUTTI FILHO *et al.*, 2004).

Quanto maior for a estimativa do *rptb*, maior será a previsibilidade de ocorrer nas medições subsequentes valores muito próximos às estimativas de eventos anteriores (CRUZ; REGAZZI, 1997). Ou seja, maior será a previsibilidade de ocorrer valores de estimativas de multicolinearidade muito próximos aos valores observados em diagnósticos realizados anteriormente com outro conjunto de dados.

Maiores valores de repetibilidade foram observados nos casos de combinações com sete caracteres (Caso 7). No entanto, apesar de a estimativa da multicolinearidade apresentar alta previsibilidade entre as sucessivas mensurações em condição de utilização de um maior número de caracteres, deve ser atentado para o fato de que, os casos contendo maior número de caracteres combinados apresentam também as maiores estimativas médias de multicolinearidade (Tabela 15). Ao mesmo tempo que se consegue obter estimativas de multicolinearidade mais precisas, maiores são os valores de multicolinearidade tanto pelo indicador número de condição (*NC*) e fator de inflação da variância (*FIV*), necessitando que seja tomada alguma medida para a diminuição do grau de multicolinearidade. Os coeficientes de repetibilidade para estimação do grau de multicolinearidade foram muito similares pelos dois indicadores (*NC* e *FIV*), indicando que ambos são concordantes na estimação da multicolinearidade.

Maior atenção deve ser dada na estimação do grau de multicolinearidade entre grupos de variáveis. Pois, maiores valores do *rptb* para estimação do grau de multicolinearidade foram verificados em caracteres do grupo de variáveis morfológicas, padrão que se repetiu em ambas as cultivares. Esse fato pode estar relacionado com a maior variabilidade dos dados em caracteres produtivos, estando estes, por tempo mais prolongado sob efeito do ambiente.

Para esse grupo de variáveis, as estimativas de *rptb* foram obtidas com maior acurácia, expressas pelos valores de coeficientes de determinação (R^2), com valores de R^2 superiores a 0,934 e 0,836 nas estimações nas cultivares BRS Progresso e Temprano, respectivamente. Enquanto que, em caracteres dos grupos de variáveis produtivas, os valores de R^2 foram superiores a 0,836 e 0,675. Obtenção de altos valores de *rptb* e R^2 é interessante, pois indica

que o modelo matemático utilizado foi eficiente em ajustar na determinação da repetibilidade (CAVALCANTE *et al.*, 2012). Valores de R^2 superiores 82% foram observados em estudos de comprimento do hipocótilo e comprimento do epicótilo na cultura da soja por todos os métodos avaliados (MATSUO *et al.*, 2012). Em estudo de soja (*Glycine max*) sob condições de disponibilidade hídrica, maiores valores foram observados para o caractere altura ($rptb = 0,73$; $R^2 = 0,94$) no experimento I determinados pelo método CPS e caractere comprimento radicular ($rptb = 0,83$; $R^2 = 0,96$) no segundo experimento (DUARTE; FERREIRA; SILVA, 2022).

Foi observada maior precisão nas estimativas de $rptb$ em grupos de variáveis para a cultivar BRS Progresso em relação a Temprano. A maior precisão pode estar relacionada com o número de ensaios utilizados na obtenção das estimativas de $rptb$ na cultivar BRS Progresso ($\eta = 5$ ensaios) e Temprano ($\eta = 3$ ensaios). Assim, a cultivar ou o número de ensaios utilizados na estimação de $rptb$ podem ter influência sobre as estimativas da repetibilidade do NC e FIV .

De maneira geral, maiores valores de repetibilidade para a estimação do grau de multicolinearidade foram observadas quando realizados pelos métodos de componentes principais (PCR e PCS). Similarmente, altos valores de $rptb$ foram verificados pelos métodos de componentes principais em estudos com braquiária (*Urochloa brizantha*) (TORRES *et al.*, 2015), capim-tanzânia (*Panicum maximum* Jacq.) (CARGNELUTTI FILHO *et al.*, 2004), capim-elefante (*Pennisetum* spp.) e morangueiro (*Fragaria x ananassa*) (DIEL *et al.*, 2020). Maiores estimativas de $rptb$ e acurácia foram obtidas pelo método PCS em estudo para determinar o número mínimo de espigas necessário para a avaliação adequada da porcentagem de brotação de grãos na espiga de trigo (*Triticum aestivum*) (PAGLIOSA *et al.*, 2014).

4.4.2 Número de ensaios para o diagnóstico de multicolinearidade

Menores números de medições ou de ensaios (η_m) necessários para o diagnóstico da multicolinearidade foram calculados para os casos, grupos de variáveis, cultivares e métodos em que apresentaram os maiores coeficientes de repetibilidade ($rptb$), conforme apresentados na Tabela 16. De maneira geral, maior número de medições ou de ensaios para a estimação da multicolinearidade pelo indicador número de condição (NC – Tabela 17) e fator de inflação da variância (FIV – Tabela 18), foram observados nos casos contendo menor número de caracteres combinados e para maiores precisões (R^2 – coeficiente de determinação) na estimativa de $rptb$ ($R^2 = 0,99$).

Tabela 17 – Número de medições ou de ensaios associado a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) para estimar a multicolinearidade pelo indicador número de condição (NC) para diferentes caracteres combinados em dois grupos de variáveis e em duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Caso ¹	Cultivar BRS Progresso										Cultivar Temprano									
	Variáveis morfológicas ²					Variáveis produtivas					Variáveis morfológicas					Variáveis produtivas				
	Coeficiente de determinação (R^2)										Coeficiente de determinação (R^2)									
	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99
----- AOV – Análise de variância -----																				
2	0,3	0,5	0,8	1,6	8,3	0,9	1,3	2,0	4,3	22,2	2,2	3,1	4,8	10,2	53,3	4,0	5,7	9,1	19,2	99,8
3	0,7	0,9	1,5	3,1	16,3	0,7	1,0	1,6	3,4	17,6	0,6	0,8	1,3	2,7	14,2	1,4	1,9	3,1	6,5	33,7
4	0,8	1,1	1,7	3,7	19,1	0,5	0,7	1,2	2,5	12,9	0,5	0,7	1,1	2,3	12,0	2,4	3,4	5,5	11,5	60,1
5	0,8	1,2	1,9	4,0	21,0	0,5	0,8	1,2	2,6	13,5	0,6	0,8	1,3	2,8	14,7	4,3	6,0	9,6	20,2	105,5
6	0,9	1,3	2,1	4,4	22,9	0,8	1,1	1,8	3,9	20,1	0,8	1,1	1,8	3,8	19,9	5,8	8,2	13,0	27,4	142,9
7	1,1	1,5	2,4	5,0	26,0	1,7	2,3	3,7	7,9	40,9	1,1	1,6	2,5	5,2	27,1	5,1	7,3	11,5	24,4	127,0
----- PCR – Componentes principais com base na matriz de correlação -----																				
2	0,2	0,3	0,5	1,0	5,4	0,5	0,7	1,2	2,5	12,9	2,1	2,9	4,7	9,8	51,2	3,2	4,5	7,2	15,2	79,1
3	0,1	0,1	0,2	0,3	1,7	0,6	0,8	1,2	2,6	13,7	0,1	0,2	0,3	0,7	3,6	1,3	1,9	3,0	6,2	32,5
4	0,1	0,1	0,2	0,4	2,1	0,4	0,5	0,8	1,7	9,0	0,2	0,2	0,4	0,8	4,2	1,8	2,6	4,1	8,7	45,2
5	0,1	0,1	0,2	0,4	2,0	0,3	0,4	0,6	1,4	7,1	0,1	0,2	0,3	0,6	3,2	2,5	3,6	5,7	12,0	62,7
6	0,1	0,1	0,2	0,3	1,7	0,3	0,4	0,6	1,3	6,7	0,1	0,1	0,2	0,4	2,3	3,0	4,2	6,6	14,0	73,1
7	0,0	0,1	0,1	0,2	1,2	0,2	0,2	0,4	0,8	4,0	0,0	0,0	0,1	0,2	0,8	2,4	3,5	5,5	11,6	60,4
----- PCS – Componentes principais com base na matriz de covariância -----																				
2	0,2	0,3	0,4	0,9	4,7	0,6	0,8	1,3	2,8	14,5	1,7	2,4	3,8	8,1	42,0	3,8	5,4	8,6	18,1	94,3
3	0,0	0,1	0,1	0,2	0,9	0,6	0,8	1,3	2,8	14,4	0,1	0,2	0,3	0,5	2,8	1,4	1,9	3,1	6,5	33,6
4	0,0	0,1	0,1	0,2	1,2	0,4	0,5	0,8	1,8	9,3	0,1	0,2	0,3	0,7	3,4	1,5	2,2	3,4	7,2	37,8
5	0,0	0,1	0,1	0,2	1,1	0,3	0,4	0,7	1,4	7,2	0,1	0,1	0,2	0,4	2,3	1,5	2,1	3,4	7,1	37,2
6	0,0	0,0	0,1	0,2	0,8	0,2	0,3	0,6	1,2	6,1	0,1	0,1	0,1	0,3	1,4	1,5	2,2	3,5	7,3	38,1
7	0,0	0,0	0,1	0,1	0,6	0,1	0,2	0,3	0,6	3,1	0,0	0,0	0,1	0,2	0,8	1,7	2,4	3,7	7,9	41,2
----- SAR – Análise estrutural com base na matriz de correlação -----																				
2	0,2	0,3	0,5	1,0	5,5	0,5	0,8	1,2	2,5	13,2	2,1	3,0	4,8	10,1	52,5	3,3	4,7	7,4	15,6	81,4
3	0,1	0,1	0,2	0,3	1,7	0,6	0,8	1,3	2,7	13,9	0,1	0,2	0,3	0,7	3,6	1,3	1,9	3,0	6,3	32,9
4	0,1	0,1	0,2	0,4	2,1	0,4	0,5	0,8	1,7	9,1	0,2	0,2	0,4	0,8	4,2	1,9	2,7	4,2	8,9	46,6
5	0,1	0,1	0,2	0,4	2,0	0,3	0,4	0,6	1,4	7,1	0,1	0,2	0,3	0,6	3,2	2,7	3,8	6,0	12,7	66,2
6	0,1	0,1	0,2	0,3	1,7	0,3	0,4	0,6	1,3	6,7	0,1	0,1	0,2	0,4	2,3	3,2	4,5	7,1	15,0	78,3
7	0,0	0,1	0,1	0,2	1,2	0,2	0,2	0,4	0,8	4,0	0,0	0,0	0,1	0,2	0,8	2,6	3,7	5,9	12,4	64,6
----- SAS – Análise estrutural com base na matriz de covariância -----																				
2	0,3	0,5	0,8	1,6	8,3	0,9	1,3	2,0	4,3	22,2	2,2	3,1	4,8	10,2	53,3	4,0	5,7	9,1	19,2	99,8
3	0,7	0,9	1,5	3,1	16,3	0,7	1,0	1,6	3,4	17,6	0,6	0,8	1,3	2,7	14,2	1,4	1,9	3,1	6,5	33,7
4	0,8	1,1	1,7	3,7	19,1	0,5	0,7	1,2	2,5	12,9	0,5	0,7	1,1	2,3	12,0	2,4	3,4	5,5	11,5	60,1
5	0,8	1,2	1,9	4,0	21,0	0,5	0,8	1,2	2,6	13,5	0,6	0,8	1,3	2,8	14,7	4,3	6,0	9,6	20,2	105,5
6	0,9	1,3	2,1	4,4	22,9	0,8	1,1	1,8	3,9	20,1	0,8	1,1	1,8	3,8	19,9	5,8	8,2	13,0	27,4	142,9
7	1,1	1,5	2,4	5,0	26,0	1,7	2,3	3,7	7,9	40,9	1,1	1,6	2,5	5,2	27,1	5,1	7,3	11,5	24,4	127,0

¹Casos definidos conforme descrito no Capítulo 3.2.3. ²Caracteres em cada grupo de variáveis foram definidos na Tabela 3.

Fonte: Autor.

Tabela 18 – Número de medições ou de ensaios associado a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) para estimar a multicolinearidade pelo indicador fator de inflação da variância (FIV) para diferentes caracteres combinados em dois grupos de variáveis e em duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Caso ¹	Cultivar BRS Progresso										Cultivar Temprano									
	Variáveis morfológicas ²					Variáveis produtivas					Variáveis morfológicas					Variáveis produtivas				
	Coeficiente de determinação (R^2)										Coeficiente de determinação (R^2)									
	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99
----- AOV – Análise de variância -----																				
2	0,3	0,5	0,8	1,6	8,4	0,9	1,3	2,1	4,4	22,8	2,4	3,3	5,3	11,2	58,2	4,4	6,2	9,8	20,8	108,1
3	0,7	0,9	1,5	3,1	16,2	0,7	1,0	1,6	3,3	17,3	0,6	0,8	1,3	2,7	14,0	1,2	1,8	2,8	5,9	30,9
4	0,7	1,0	1,6	3,3	17,2	0,6	0,9	1,4	2,9	15,0	0,5	0,7	1,1	2,3	11,7	1,9	2,6	4,2	8,8	45,9
5	0,7	1,1	1,7	3,6	18,5	0,7	0,9	1,5	3,1	16,4	0,5	0,7	1,2	2,5	13,1	3,1	4,3	6,9	14,5	75,7
6	0,8	1,2	1,9	3,9	20,6	1,0	1,4	2,2	4,6	23,9	0,7	1,0	1,5	3,2	16,8	3,6	5,1	8,2	17,3	89,9
7	1,0	1,4	2,2	4,7	24,3	1,8	2,5	3,9	8,3	43,3	0,9	1,3	2,0	4,3	22,3	2,5	3,6	5,7	12,1	62,9
----- PCR – Componentes principais com base na matriz de correlação -----																				
2	0,2	0,3	0,5	1,0	5,2	0,5	0,7	1,2	2,5	13,1	2,2	3,2	5,0	10,7	55,5	3,4	4,9	7,7	16,3	84,8
3	0,1	0,1	0,1	0,3	1,6	0,5	0,8	1,2	2,6	13,5	0,1	0,2	0,3	0,7	3,5	1,2	1,7	2,7	5,8	30,1
4	0,1	0,1	0,2	0,5	2,4	0,4	0,6	1,0	2,1	10,7	0,2	0,3	0,4	0,9	4,7	1,6	2,3	3,6	7,7	40,1
5	0,1	0,1	0,2	0,4	2,2	0,4	0,5	0,8	1,8	9,3	0,1	0,2	0,3	0,7	3,5	2,2	3,2	5,0	10,6	55,1
6	0,1	0,1	0,2	0,3	1,8	0,4	0,5	0,8	1,7	8,9	0,1	0,1	0,2	0,4	2,1	2,3	3,3	5,2	10,9	57,0
7	0,0	0,1	0,1	0,2	0,9	0,2	0,3	0,4	0,9	4,8	0,0	0,0	0,0	0,1	0,4	1,3	1,8	2,9	6,1	31,8
----- PCS – Componentes principais com base na matriz de covariância -----																				
2	0,2	0,3	0,4	0,9	4,5	0,6	0,8	1,3	2,8	14,7	1,8	2,5	4,0	8,5	44,2	4,1	5,8	9,2	19,4	101,2
3	0,0	0,1	0,1	0,2	0,9	0,6	0,8	1,3	2,7	13,9	0,1	0,2	0,2	0,5	2,7	1,2	1,7	2,8	5,9	30,5
4	0,1	0,1	0,1	0,3	1,4	0,4	0,6	1,0	2,0	10,6	0,2	0,2	0,4	0,8	4,0	1,5	2,1	3,3	7,0	36,6
5	0,0	0,1	0,1	0,2	1,2	0,4	0,5	0,8	1,7	8,8	0,1	0,2	0,3	0,5	2,8	1,6	2,3	3,7	7,8	40,4
6	0,0	0,0	0,1	0,2	0,8	0,3	0,4	0,7	1,5	7,8	0,1	0,1	0,1	0,3	1,4	1,2	1,7	2,7	5,7	30,0
7	0,0	0,0	0,0	0,1	0,4	0,2	0,3	0,4	0,9	4,8	0,0	0,0	0,0	0,1	0,5	0,8	1,1	1,7	3,6	18,7
----- SAR – Análise estrutural com base na matriz de correlação -----																				
2	0,2	0,3	0,5	1,0	5,2	0,5	0,8	1,2	2,6	13,4	2,3	3,3	5,2	11,0	57,2	3,5	5,0	7,9	16,8	87,3
3	0,1	0,1	0,1	0,3	1,6	0,6	0,8	1,3	2,6	13,8	0,1	0,2	0,3	0,7	3,5	1,2	1,7	2,8	5,8	30,4
4	0,1	0,1	0,2	0,5	2,4	0,4	0,6	1,0	2,1	10,8	0,2	0,3	0,4	0,9	4,7	1,7	2,4	3,7	7,9	41,2
5	0,1	0,1	0,2	0,4	2,2	0,4	0,5	0,8	1,8	9,3	0,1	0,2	0,3	0,7	3,5	2,3	3,3	5,3	11,1	57,9
6	0,1	0,1	0,2	0,4	1,8	0,4	0,5	0,8	1,7	8,9	0,1	0,1	0,2	0,4	2,1	2,4	3,4	5,4	11,5	59,8
7	0,0	0,1	0,1	0,2	0,9	0,2	0,3	0,4	0,9	4,9	0,0	0,0	0,0	0,1	0,4	1,3	1,9	2,9	6,2	32,4
----- SAS – Análise estrutural com base na matriz de covariância -----																				
2	0,3	0,5	0,8	1,6	8,4	0,9	1,3	2,1	4,4	22,8	2,4	3,3	5,3	11,2	58,2	4,4	6,2	9,8	20,8	108,1
3	0,7	0,9	1,5	3,1	16,2	0,7	1,0	1,6	3,3	17,3	0,6	0,8	1,3	2,7	14,0	1,2	1,8	2,8	5,9	30,9
4	0,7	1,0	1,6	3,3	17,2	0,6	0,9	1,4	2,9	15,0	0,5	0,7	1,1	2,3	11,7	1,9	2,6	4,2	8,8	45,9
5	0,7	1,1	1,7	3,6	18,5	0,7	0,9	1,5	3,1	16,4	0,5	0,7	1,2	2,5	13,1	3,1	4,3	6,9	14,5	75,7
6	0,8	1,2	1,9	3,9	20,6	1,0	1,4	2,2	4,6	23,9	0,7	1,0	1,5	3,2	16,8	3,6	5,1	8,2	17,3	89,9
7	1,0	1,4	2,2	4,7	24,3	1,8	2,5	3,9	8,3	43,3	0,9	1,3	2,0	4,3	22,3	2,5	3,6	5,7	12,1	62,9

¹Casos definidos conforme descrito no Capítulo 3.2.3. ²Caracteres em cada grupo de variáveis foram definidos na Tabela 3.

Fonte: Autor.

Foi observado que, independentemente da precisão, parece haver a necessidade de utilização de maior número de ensaios no diagnóstico de multicolinearidade com caracteres produtivos. Já entre as cultivares, há a necessidade de maiores números de ensaios quando multicolinearidade diagnosticada em caracteres da cultivar Temprano.

No diagnóstico de multicolinearidade realizado com um único ensaio, independentemente de método utilizado para a determinação de *rptb*, do caso e grupo de variáveis, é esperada precisão superior a 80% na estimativa da multicolinearidade pelos indicadores *NC* em caracteres morfológicos e produtivos na cultivar BRS Progresso, com exceção no Caso 7 quando o *rptb* foi estimado pelos métodos AOV ou SAS. Já em caracteres morfológicos na cultivar Temprano, um único ensaio não é suficiente para obtenção $R^2 = 0,80$ no Casos 2 quando estimados os *rptb* de multicolinearidade pelos métodos PCR, PCS e SAR e nos Casos 2 e 7 pelos métodos AOV e SAS. Já em caracteres produtivos, utilizando como referência as estimativas de *rptb* obtidas pelos métodos AOV e SAS, há a necessidade de dois (Caso 3) a nove (Caso 6) ensaios para a estimação da multicolinearidade, enquanto que pelos demais métodos, dois a cinco ensaios são suficientes para precisão mínima de 80%.

Um único ensaio também é suficiente em diagnóstico realizado pelo indicador *FIV* em ambos os grupos de variáveis na BRS Progresso com precisão superior 80%, com exceção para os Casos 6 e 7 quando diagnóstico realizado com caracteres produtivos. Com o mesmo nível de precisão, um único ensaio é somente suficiente na determinação da multicolinearidade pelo indicador *FIV* em caracteres morfológicos da cultivar Temprano, exceto para combinação do Caso 2. Desejando precisão mínima de 80% em nas estimativas de *FIV* em caracteres produtivos, são necessários de dois a cinco ensaios.

Diferentes valores de η_m são obtidos pelos diferentes métodos utilizados para a estimação de *rptb*, com maiores estimativas obtidas pelos métodos de componentes principais (PCR e PCS). Estes métodos consideram o comportamento cíclico da característica, com os elementos do autovetor de mesmo sinal e magnitudes, expressando a tendência dos genótipos em manter suas posições nas sucessivas medições (ABEYWARDENA, 1972). Método baseado em componentes principais são recomendados pela maior acurácia na estimação da repetibilidade (ABEYWARDENA, 1972) e o método baseado na matriz de correlação (PCR) forneceu estimativas mais precisas de *rptb* em braquiária (*Urochloa brizantha*) (TORRES *et al.*, 2015). Diante disso e pelos elevados valores de *rptb* e R^2 , as inferências do número de medições ou ensaios para o diagnóstico de multicolinearidade foram realizadas pelo método PCR. As estimativas de η_m obtidas pelos métodos PCR e PCS são muito semelhantes entre si,

com as mesmas recomendações quando os valores das Tabelas Tabela 17 e Tabela 18 são arredondadas para valores inteiros.

Um único ensaio é suficiente para o diagnóstico de multicolinearidade pelos indicadores *NC* e *FIV* em caracteres morfológicos com precisão mínima de 95% em ambas as cultivares e em todas as combinações de caracteres, exceto na Temprano e combinações com dois caracteres (Caso 2). Já em caracteres produtivos na BRS Progresso e com $R^2 \geq 0,95$, há a necessidade de até três ensaios para o diagnóstico de multicolinearidade pelos indicadores *NC* e *FIV*. Por outro lado, são necessários de sete a 16 e seis a 17 ensaios para o diagnóstico em caracteres produtivos na cultivar Temprano pelos indicadores *NC* (Tabela 17) e *FIV* (Tabela 18), respectivamente.

Menores números de ensaios são necessários para a estimação da multicolinearidade em grupos de variáveis de centeio quando relacionados com as recomendações do número de ensaios necessários para a obtenção de estimativas de características agronômicas em outras culturas. Foi relatado que, genótipos de capim-elefante (*Pennisetum* spp.), sete ciclos de avaliações seriam suficientes para estimar o comprimento do entrenó e três ciclos para as variáveis massa de forragem, altura da planta, comprimento e largura da folha, diâmetro do colmo, clorose e índice de área foliar com precisão de 90% ($R^2 = 0,90$) (CAVALCANTE *et al.*, 2012). Em outro estudo com a mesma espécie vegetal, foi recomendado a necessidade mínima de seis medições para a estimação dos caracteres altura de plantas e diâmetro de colmo, e nove cortes para o caractere produtividade de matéria seca (SOUZA *et al.*, 2017). Para estimar da porcentagem de grãos germinados em espiga de trigo (*Triticum aestivum*), foi verificada a necessidade de no mínimo de 11 avaliações (PAGLIOSA *et al.*, 2014). Em características de soja (*Glycine max*), maiores medições seriam necessárias, quando o objetivo for a predição do comprimento do hipocótilo, comprimento do epicótilo e do internódio ($\eta_m = 15$; $R^2 = 0,90$) e para o comprimento do pecíolo da primeira folha trifoliolada ($\eta_m = 21$; $R^2 = 0,85$) (MATSUO *et al.*, 2012).

4.4.3 Considerações sobre a repetibilidade e número de ensaios para o diagnóstico da multicolinearidade

O número de ensaios necessário para estimar a multicolinearidade obtido nesse estudo serve de referencial para planejamentos futuros de novos ensaios com a cultura do centeio e demais culturas. Não é objetivo deste trabalho a indicação de um único número de ensaio. O pesquisador deverá adotar o número de ensaios de acordo com o nível de precisão almejado,

bem como, analisar o comportamento da repetibilidade com base em ensaios anteriores ou trabalhos publicados na comunidade científica.

Há variabilidade na estimativa da repetibilidade da multicolinearidade de acordo com o método utilizado. O método por componentes principais (PC) é recomendado para determinação da repetibilidade com maior precisão (ABEYWARDENA, 1972) e no presente estudo, este método com base na matriz de correlação (PCR) e na matriz de variância e covariância (PCS) apresentou as maiores estimativas de *rptb*. O método PCR também forneceu estimativas mais precisas de *rptb* de características em braquiária (TORRES *et al.*, 2015).

Não foram encontrados trabalhos que avaliaram a repetibilidade para a estimação da multicolinearidade em caracteres, nem tampouco de características morfológicas ou produtivas na cultura do centeio. Em outras culturas e para diferentes características agrônômicas, as estimativas de repetibilidade foram de magnitudes semelhantes às encontradas no presente estudo para a determinação do grau de multicolinearidade.

Em caracteres morfofisiológicos de genótipos de capim-elefante (*Pennisetum spp.*), a partir de dados obtidos durante seis ciclos de avaliação, os caracteres massa de matéria seca de forragem ($rptb \geq 0,81$; $R^2 \geq 0,96$) e clorose foliar ($rptb = 0,84$; $R^2 = 0,97$) se destacaram com os maiores estimativas de *rptb* (CAVALCANTE *et al.*, 2012). Os mesmos autores evidenciaram a influência ambiental sobre a variável comprimento do entrenó, devido a estimativa de *rptb* de baixa magnitude ($0,57 \leq rptb \leq 0,61$; $0,84 \leq R^2 \leq 0,85$), caracterizando baixa regularidade entre medidas repetidas. Valores de baixa magnitude de *rptb* também foram verificados para o caractere comprimento do hipocótilo, variando de 0,345 a 0,793 pelos métodos AOV e CPS, respectivamente; e de 0,478 (AOV) a 0,914 (CPR, CPS e SAS) para o comprimento do epicótilo (MATSUO *et al.*, 2012).

Em cirigueleira (*Spondias purpurea* L.), a repetibilidade variou de 0,47 a 0,58 para o caractere número de frutos (LIRA JÚNIOR *et al.*, 2014). Estimativas medianas de *rptb* foram verificadas em características de capim-tanzânia (*Panicum maximum* Jacq.), com valores de 0,42 a 0,55 e coeficiente de determinação (R^2) de alta magnitude ($0,81 \leq R^2 \leq 0,95$), sugerindo que o número de colheitas foi suficiente para fornecer alta precisão na seleção de genótipos (FERNANDES *et al.*, 2017). Para estimar da porcentagem de grãos germinados em espiga de trigo (*Triticum aestivum*), foi verificado a necessidade de no mínimo de onze avaliações ($rptb = 0,90$) pelo método AOV e acurácia esperada de $R^2 = 0,99$ (PAGLIOSA *et al.*, 2014). No mesmo trabalho, quando estimado por PCR ou PCS, a necessidade diminui para sete e três medições para o mesmo nível de acurácia, respectivamente.

Houve variabilidade no número mínimo de medições ou de ensaios para o diagnóstico de multicolinearidade em caracteres morfológicos e produtivos de centeio. Variabilidade verificada também na estimação de diferentes características em outras culturas. A predição do desempenho de genótipos de Capim-tanzânia (*Panicum maximum* Jacq.) com acurácia de 80% em todas as características avaliadas (massas de matéria seca total, folhas e colmo e altura de planta) podem ser obtidas com oito cortes (CARGNELUTTI FILHO *et al.*, 2004). Em capim-elfante (*Pennisetum spp.*), três medições foram suficientes para avaliar características em caracteres massa de forragem, altura da planta, comprimento e largura da folha, diâmetro do colmo, clorose e índice de área foliar com $R^2 = 90\%$; enquanto que, para o comprimento do entrenó, são necessários no mínimo sete medições para se obter o mesmo nível de precisão (CAVALCANTE *et al.*, 2012). Em braquiária (*Urochloa brizantha*) foi verificado a necessidade de duas medições para as determinações de fibra em detergente neutro, celulose e sílica, enquanto que para proteína bruta, fibra em detergente ácido e digestibilidade in vitro da matéria orgânica requerem quatro medições com confiabilidade de 80% (TORRES *et al.*, 2015).

Diferentes números de ensaios foram estimados para diferentes níveis de precisão. Cabe ao pesquisador analisar o material e mão-de-obra disponível para a execução do experimento e auxiliado por informações coletadas em ensaios anteriores ou de estudos com a cultura, optar pelo nível de precisão desejado e então, fazer uso da informação do número de ensaios. O pesquisador também deve realizar a busca por informações de tamanho de amostra, tamanho de parcela, relações entre caracteres, diagnóstico de multicolinearidade e demais informações sobre a cultura que possam interferir na precisão experimental e nas estimativas dos parâmetros que serão utilizados na interpretação dos resultados.

Recomendação de utilização de diferentes números de ensaios foi identificado no estudo para realizar o diagnóstico de multicolinearidade em diferentes grupos de variáveis e cultivares. No entanto, adoção de diferentes números de ensaios dentro de um mesmo experimento não se torna prático e, portanto, há a necessidade de padronização para um único valor. Assim como na cultura do centeio, ou em qualquer outra cultura, a adoção de um mesmo número de ensaio se faz necessário e para o presente estudo, a recomendação foi realizada em função da recomendação da cultivar que há necessidade de maior número de ensaios. A recomendação utilizando como referência a cultivar de necessidade de maior número de ensaios se faz necessário, pois nesta condição, se terá a certeza de que a precisão mínima será atendida todas as cultivares e grupos de variáveis.

Em caracteres na cultivar BRS Progresso, três ensaios são suficientes para estimação do grau de multicolinearidade pelos indicadores *NC* e *FIV*, com precisão de 95%. Já em caracteres

na cultivar Temprano, são necessários no mínimo 16 ensaios para o mesmo nível de precisão ($R^2 = 0,95$).

Assim, conclui-se que:

1. Um ensaio não é eficiente para estimar o grau de multicolinearidade pelos indicadores número de condição e fator de inflação da variância em caracteres morfológicos e produtivos de centeio.
2. Há variabilidade nas estimativas do número de medições entre cultivares e números de caracteres combinados em cada grupo de variáveis.
3. Não há variabilidade no número de ensaios para realizar o diagnóstico de multicolinearidade pelos indicadores número de condição e fator de inflação de variância.
4. São necessários no mínimo três e 16 ensaios para o diagnóstico de multicolinearidade em caracteres na cultivar BRS Progresso e Temprano, respectivamente, com precisão mínima de 95%.

4.5 REPETIBILIDADE E NÚMERO DE ENSAIOS PARA ESTIMAÇÃO DA CORRELAÇÃO CANÔNICA

4.5.1 Estimativa da correlação do primeiro par canônico

Na Tabela 19 estão apresentadas algumas das combinações das estimativas de correlações canônicas para o primeiro par canônico (ρ_{U_1, V_1}) para cada ensaio no cenário 2×2 (28 combinações de caracteres do Caso 2 para o grupo de variáveis morfológicas \times 28 combinações de caracteres do Caso 2 para o grupo de variáveis produtivas). Não foram apresentadas todas as composições, pois em cada ensaio para este cenário foram obtidas total de 784 combinações e a apresentação parcial foi utilizada para exemplificar a organização dos dados para este cenário. Ao total, para esse cenário, foram obtidas 5.984 estimativas de ρ_{U_1, V_1} (784 combinações \times oito ensaios), sendo destas, 3.920 e 2.352 combinações referentes aos ensaios com as cultivares BRS Progresso e Temprano, respectivamente.

Tabela 19 – Estimativa de correlações canônicas para o primeiro par canônico obtidas para o caso com a combinação de dois caracteres morfológicos (X) e dois caracteres produtivos (Y) – total de 784 combinações, em oito ensaios de uniformidade¹ (E1, E2, ..., E8) em centeio (*Secale cereale* L.), conduzidos em Santa Maria – RS, Safra 2016.

Combinação ²	Grupo de Variáveis ³		Cultivar							
	Morfológico	Produtivo	BRS Progresso					Temprano		
	X	Y	E1	E2	E3	E4	E5	E6	E7	E8
1	X1 x X2	Y1 x Y2	0,648	0,473	0,642	0,663	0,749	0,762	0,727	0,695
2	X1 x X2	Y1 x Y3	0,647	0,384	0,654	0,601	0,749	0,764	0,730	0,620
3	X1 x X2	Y1 x Y4	0,662	0,364	0,649	0,670	0,769	0,762	0,778	0,772
4	X1 x X2	Y1 x Y5	0,647	0,359	0,649	0,595	0,749	0,764	0,724	0,576
5	X1 x X2	Y1 x Y6	0,659	0,361	0,654	0,690	0,787	0,762	0,734	0,580
6	X1 x X2	Y1 x Y7	0,649	0,353	0,647	0,595	0,749	0,763	0,713	0,562
7	X1 x X2	Y1 x Y8	0,649	0,360	0,642	0,634	0,758	0,771	0,750	0,734
8	X1 x X2	Y2 x Y3	0,436	0,367	0,396	0,230	0,370	0,344	0,509	0,555
9	X1 x X2	Y2 x Y4	0,578	0,396	0,492	0,634	0,698	0,588	0,639	0,695
10	X1 x X2	Y2 x Y5	0,449	0,369	0,383	0,254	0,388	0,344	0,512	0,561
11	X1 x X2	Y2 x Y6	0,625	0,392	0,579	0,707	0,742	0,669	0,694	0,616
12	X1 x X2	Y2 x Y7	0,396	0,331	0,358	0,203	0,383	0,265	0,472	0,549
13	X1 x X2	Y2 x Y8	0,295	0,335	0,357	0,291	0,451	0,128	0,543	0,561
14	X1 x X2	Y3 x Y4	0,575	0,309	0,458	0,635	0,698	0,585	0,624	0,684
15	X1 x X2	Y3 x Y5	0,445	0,334	0,416	0,253	0,349	0,346	0,498	0,446
16	X1 x X2	Y3 x Y6	0,628	0,338	0,567	0,693	0,739	0,667	0,687	0,464
17	X1 x X2	Y3 x Y7	0,436	0,370	0,438	0,223	0,348	0,340	0,505	0,509
18	X1 x X2	Y3 x Y8	0,448	0,348	0,345	0,357	0,405	0,383	0,552	0,527
19	X1 x X2	Y4 x Y5	0,568	0,263	0,429	0,634	0,698	0,597	0,618	0,686
20	X1 x X2	Y4 x Y6	0,648	0,302	0,553	0,701	0,769	0,674	0,723	0,722
21	X1 x X2	Y4 x Y7	0,588	0,267	0,438	0,633	0,704	0,585	0,622	0,685
22	X1 x X2	Y4 x Y8	0,641	0,355	0,506	0,722	0,764	0,647	0,617	0,704
23	X1 x X2	Y5 x Y6	0,628	0,311	0,554	0,690	0,738	0,668	0,658	0,361
24	X1 x X2	Y5 x Y7	0,446	0,229	0,286	0,326	0,338	0,353	0,375	0,465
25	X1 x X2	Y5 x Y8	0,450	0,252	0,253	0,348	0,367	0,345	0,552	0,525
26	X1 x X2	Y6 x Y7	0,625	0,310	0,554	0,691	0,737	0,673	0,651	0,338
27	X1 x X2	Y6 x Y8	0,629	0,379	0,565	0,717	0,744	0,671	0,713	0,641
28	X1 x X2	Y7 x Y8	0,394	0,208	0,200	0,300	0,332	0,260	0,558	0,524
29	X1 x X3	Y1 x Y2	0,484	0,440	0,511	0,561	0,680	0,635	0,529	0,625
30	X1 x X3	Y1 x Y3	0,494	0,408	0,528	0,495	0,679	0,639	0,531	0,518
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
782	X7 x X8	Y6 x Y7	0,245	0,510	0,353	0,255	0,305	0,348	0,236	0,299
783	X7 x X8	Y6 x Y8	0,980	0,967	0,944	0,964	0,955	0,934	0,970	0,960
784	X7 x X8	Y7 x Y8	0,980	0,968	0,944	0,964	0,955	0,933	0,969	0,960

¹Ensaios definidos na Tabela 2. ²Combinações obtidas em cada cenário, conforme descrito no Capítulo 3.4.1.

³Caracteres presentes em cada grupo de variáveis foram descritos na Tabela 3.

Fonte: Autor.

Variabilidade nas estimativas da correlação canônica para o primeiro par (ρ_{U_1, V_1}) podem ser observadas nos diferentes cenários formados pelas combinações de casos com caracteres morfológicos e produtivos (Tabela 20). Maiores amplitudes são observadas em cenários contendo combinações de dois caracteres morfológicos e outros dois caracteres produtivos em cada grupo. No cenário 2×2 , tem-se 0,756 e 0,798 de amplitude do intervalo de confiança de 95% ($AIC_{95\%}$) nas estimativas de ρ_{U_1, V_1} nas cultivares BRS Progresso e Temprano, respectivamente. Enquanto que em cenários formados pela combinação de maior número de caracteres em cada grupo de variáveis, o $AIC_{95\%}$ foi reduzido de magnitude. Em cenário contendo os casos com sete caracteres combinados em cada grupo de variáveis (7×7), o $AIC_{95\%}$ foi de 0,001 e 0,011 nas cultivares BRS Progresso e Temprano, respectivamente.

Tabela 20 – Mínimo, média, máximo, percentis ($P_{2,5\%}$ e $P_{97,5\%}$), amplitude do intervalo de confiança ($AIC_{95\%} = P_{97,5\%} - P_{2,5\%}$) e percentual de combinações com correlações canônicas significativas pelo teste Qui-Quadrado em nível de 5% de significância (PCCS), de estimativas obtidas em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicos e produtivos em cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cenário ¹	Nº combinações ensaio ⁻¹	Mínimo	Percentil 2,5%	Média	Percentil 97,5%	Máximo	$AIC_{95\%}$	PCCS (%)
----- Cultivar BRS Progresso -----								
2×2	784	0,112	0,227	0,676	0,983	0,994	0,756	93,14
3×3	3136	0,211	0,450	0,870	0,990	1,000	0,541	99,65
4×4	4900	0,401	0,741	0,958	1,000	1,000	0,295	100,00
5×5	3136	0,754	0,958	0,987	1,000	1,000	0,042	100,00
6×6	784	0,960	0,982	0,994	1,000	1,000	0,018	100,00
7×7	64	0,982	0,990	0,998	1,000	1,000	0,001	100,00
----- Cultivar Temprano -----								
2×2	784	0,090	0,182	0,690	0,979	0,993	0,798	92,30
3×3	3136	0,191	0,561	0,875	0,991	1,000	0,430	99,53
4×4	4900	0,532	0,817	0,952	1,000	1,000	0,183	100,00
5×5	3136	0,735	0,929	0,981	1,000	1,000	0,071	100,00
6×6	784	0,939	0,972	0,992	1,000	1,000	0,028	100,00
7×7	64	0,982	0,989	0,997	1,000	1,000	0,011	100,00

¹Cenários descritos no Capítulo 3.4.1.

Fonte: Autor.

Maiores quantidades de caracteres em cada grupo de variáveis podem também ser responsáveis por aumentar a probabilidade de ocorrência de ρ_{U_1, V_1} significativos pelo teste Qui-Quadrado em nível de 5% de significância. Em cenários contendo quatro ou mais caracteres combinados, a totalidade das estimativas de ρ_{U_1, V_1} foram significativas em nível de 5% de

significância. Entretanto, nos demais cenários, somente em 93,14% e 92,30% no cenário 2×2 e em 99,65% e 99,53% das combinações no cenário 3×3 , a $\rho_{U_1V_1}$ foi significativa em análises realizadas com os caracteres combinados para as cultivares BRS Progresso e Temprano, respectivamente.

4.5.2 Coeficiente de repetibilidade para estimação da correlação canônica

Maiores diferenças de estimativa do coeficiente de repetibilidade (*rptb*) para a estimação da correlação do primeiro par canônico ($\rho_{U_1V_1}$) foram observadas principalmente entre cultivares (Tabela 21), com estimativas levemente superiores em análises na BRS Progresso. Comparando as estimativas de *rptb* entre cultivares e cenários, não pode ser observado nenhum padrão. No cenário 2×2 , a maior estimativa de *rptb* para $\rho_{U_1V_1}$ ocorreu quando determinada pelo método AOV, com *rptb* de 0,889, e com menor estimativa no cenário 5×5 (*rptb* = 0,790) na cultivar BRS progresso. Já na cultivar Temprano, maiores e menores estimativas foram encontradas nos cenários 3×3 (0,853) e 5×5 (0,644), respectivamente. Altas estimativas de *rptb* indica que é a estimação do valor real de uma determinada característica é possível com um número relativamente pequeno medições (CARGNELUTTI FILHO *et al.*, 2004). Porém, valores de baixa magnitude de *rptb* também podem ocorrer. Em estudo realizado com características de soja, foi verificado estimativas de *rptb* variando de 0,345 a 0,793 pelos métodos AOV e CPS, respectivamente; e de 0,478 (AOV) a 0,914 (CPR, CPS e SAS) para o comprimento do epicótilo (MATSUO *et al.*, 2012).

Concordâncias nas estimativas de *rptb* para $\rho_{U_1V_1}$ puderam ser verificadas quando o coeficiente foi estimado pelos métodos de componentes principais (PC) e de análise estrutural (AS). O cenário 5×5 também foi o responsável pela obtenção de menores estimativas de *rptb* em ambas as cultivares pelo método de componentes principais, tanto em análise com base na matriz de correlação (PCR), como na matriz de covariância (PCS). O método da análise estrutural, com base na matriz de correlação (SAR) e matriz de covariâncias (SAS) resultou em menores estimativas de *rptb* somente na cultivar BRS Progresso no cenário 5×5 , enquanto que na cultivar Temprano, ocorreu no cenário 6×6 .

Maiores estimativas de *rptb* pelos métodos PCR, PCS, SAR e SAS foram observados em cenários de casos com menores e/ou maiores números de caracteres combinados. Não pode ser observado tendência de aumento ou diminuição na estimativa de *rptb*, nos cenários com o menor ou o maior número de caracteres combinados.

Tabela 21 – Coeficientes de repetibilidade ($rptb$) e de determinação (R^2) para a estimativa da correlação canônica para o primeiro par canônico por diferentes métodos em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicas (MORF) e produtivos (PROD) em duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cenário ¹	Caso		Nº de Comb.	Cultivar							
	MORF	PROD		BRS Progresso		Temprano		BRS Progresso		Temprano	
				$rptb$	R^2	$rptb$	R^2	$rptb$	R^2	$rptb$	R^2
----- AOV ² -----											
2 × 2	2	2	784	0,915	0,982	0,839	0,940	-	-	-	-
3 × 3	3	3	3136	0,906	0,980	0,858	0,948	-	-	-	-
4 × 4	4	4	4900	0,885	0,975	0,811	0,928	-	-	-	-
5 × 5	5	5	3136	0,808	0,955	0,663	0,855	-	-	-	-
6 × 6	6	6	784	0,849	0,966	0,661	0,854	-	-	-	-
7 × 7	7	7	64	0,854	0,967	0,836	0,939	-	-	-	-
Média	-	-		0,869	0,971	0,778	0,911	-	-	-	-
				----- PCR -----				----- PCS -----			
2 × 2	2	2	784	0,920	0,983	0,839	0,940	0,921	0,983	0,840	0,940
3 × 3	3	3	3136	0,931	0,985	0,862	0,949	0,931	0,985	0,861	0,949
4 × 4	4	4	4900	0,921	0,983	0,813	0,929	0,919	0,983	0,815	0,930
5 × 5	5	5	3136	0,851	0,966	0,693	0,871	0,865	0,970	0,706	0,878
6 × 6	6	6	784	0,886	0,975	0,716	0,883	0,890	0,976	0,702	0,876
7 × 7	7	7	64	0,935	0,986	0,844	0,942	0,944	0,988	0,837	0,939
Média	-	-		0,907	0,980	0,795	0,919	0,911	0,981	0,793	0,919
				----- SAR -----				----- SAS -----			
2 × 2	2	2	784	0,920	0,983	0,839	0,940	0,915	0,982	0,839	0,940
3 × 3	3	3	3136	0,930	0,985	0,861	0,949	0,906	0,980	0,858	0,948
4 × 4	4	4	4900	0,921	0,983	0,811	0,928	0,885	0,975	0,811	0,928
5 × 5	5	5	3136	0,851	0,966	0,685	0,867	0,808	0,955	0,663	0,855
6 × 6	6	6	784	0,886	0,975	0,712	0,881	0,849	0,966	0,661	0,854
7 × 7	7	7	64	0,935	0,986	0,842	0,941	0,854	0,967	0,836	0,939
Média	-	-		0,907	0,980	0,792	0,918	0,869	0,971	0,778	0,911

¹Cenários descritos no Capítulo 3.4.1. ²Método AOV: análise de variância; PCR: componentes principais usando a matriz de covariância; PCS: componentes principais usando a matriz de correlação; SAR: análise estrutural usando a matriz de correlação; SAS: análise estrutural usando a matriz de covariância.

Fonte: Autor.

Independentemente dos métodos, as estimativas de $rptb$ para obter os valores de ρ_{U_1, V_1} foram obtidas com alta precisão, com os coeficientes de determinação (R^2) superiores a 0,950 e 0,844 (método AOV) na cultivar BRS Progresso e Temprano, respectivamente. Desconsiderando o método AOV, as estimativas de $rptb$ foram obtidas com precisão superior a 95,5% e 85,4% para as mesmas cultivares, ambas obtidas pelo método de análise estrutural (SAR e SAS). Obtenção de altos valores de R^2 na determinação de $rptb$ indica eficiência no modelo matemático utilizado (CAVALCANTE *et al.*, 2012).

De maneira geral, os métodos de componentes principais foram os que melhor conseguiram estimar o *rptb*. A estimativa de *rptb* foi realizada com precisões superiores 96,6% na cultivar BRS Progresso e 87,1% na cultivar Temprano. Entre os métodos PCR e PCS, o R^2 observado foi muito semelhante. Dentre os métodos, o PCR pode ser uma ótima opção para ser utilizado na estimativa de *rptb* de ρ_{U_1, V_1} .

Estes métodos consideram o comportamento cíclico da característica, com os elementos do autovetor de mesmo sinal e magnitudes, expressando a tendência dos genótipos em manter suas posições nas sucessivas medições (ABEYWARDENA, 1972). Método baseado em componentes principais são recomendados pois estimam a reprodutibilidade com maior acurácia (ABEYWARDENA, 1972) e o método PCR forneceu estimativas mais precisas de *rptb* em braquiária (*Urochloa brizantha*) (TORRES *et al.*, 2015),

Altos valores de *rptb* ($rptb \geq 0,90$) e R^2 ($R^2 \geq 0,99$) foram observados em estudo para determinar por meio de repetibilidade, o número mínimo de espigas necessário para a avaliação adequada da porcentagem de brotação de grãos na espiga de trigo (*Triticum aestivum*) (PAGLIOSA *et al.*, 2014). Em características de soja, estimativas de R^2 superiores a 82% foram observadas em todos os métodos utilizados no estudo de repetibilidade (MATSUO *et al.*, 2012).

4.5.3 Número de ensaios para estimação da correlação canônica

No geral, é necessário maior número de medições ou ensaios (η_m) para a estimativa da correlação do primeiro par canônico (ρ_{U_1, V_1}), à medida que aumenta o número de caracteres combinados em cada grupo de variáveis, até o cenário composto por seis caracteres combinados (Tabela 22). Considerando o menor nível de precisão ($R^2 = 0,80$), o η_m variou de 0,5 a 1,1 quando utilizado o método de análise de variância (AOV) na estimativa do coeficiente de repetibilidade (*rptb*) na cultivar BRS Progresso.

Tabela 22 – Número de ensaios associados a diferentes coeficientes de determinação ($R^2 = 0,80, 0,85, 0,90, 0,95$ e $0,99$) de estimativas da correlação canônica para o primeiro par por diferentes métodos em cenários formados pelas combinações de casos com caracteres do grupo de variáveis morfológicos (MORF) e produtivos (PROD) em duas cultivares de centeio (*Secale cereale* L.), Santa Maria – RS, Safra 2016.

Cenário	Caso		Cultivar BRS Progresso					Cultivar Temprano				
	MORF	PROD	0,80	0,85	0,90	0,95	0,99	0,80	0,85	0,90	0,95	0,99
----- AOV – Análise de variância -----												
2 × 2	2	2	0,5	0,7	1,1	2,4	12,4	0,8	1,1	1,8	3,7	19,5
3 × 3	3	3	0,5	0,7	1,1	2,4	12,6	0,7	1,0	1,6	3,3	17,1
4 × 4	4	4	0,6	0,8	1,3	2,7	14,2	1,0	1,4	2,2	4,7	24,5
5 × 5	5	5	1,1	1,5	2,4	5,0	26,3	2,2	3,1	5,0	10,5	54,8
6 × 6	6	6	1,1	1,5	2,4	5,0	26,1	2,2	3,1	4,9	10,3	53,6
7 × 7	7	7	0,9	1,3	2,0	4,3	22,3	0,9	1,3	2,0	4,2	21,9
Média	-	-	0,8	1,1	1,7	3,6	19,0	1,3	1,8	2,9	6,1	31,9
----- PCR – Componente principal com base na matriz de correlação -----												
2 × 2	2	2	0,3	0,5	0,8	1,6	8,6	0,8	1,1	1,7	3,6	18,9
3 × 3	3	3	0,3	0,4	0,7	1,4	7,4	0,6	0,9	1,4	3,0	15,9
4 × 4	4	4	0,3	0,5	0,8	1,6	8,5	0,9	1,3	2,1	4,4	22,8
5 × 5	5	5	0,7	1,0	1,6	3,3	17,3	1,8	2,5	4,0	8,4	43,9
6 × 6	6	6	0,5	0,7	1,2	2,4	12,7	1,6	2,2	3,6	7,5	39,2
7 × 7	7	7	0,3	0,4	0,6	1,3	6,8	0,7	1,0	1,7	3,5	18,3
Média	-	-	0,4	0,6	0,9	2,0	10,2	1,1	1,5	2,4	5,1	26,5
----- PCS – Componente principal com base na matriz de covariância -----												
2 × 2	2	2	0,3	0,5	0,8	1,6	8,5	0,8	1,1	1,7	3,6	18,9
3 × 3	3	3	0,3	0,4	0,7	1,4	7,3	0,6	0,9	1,5	3,1	16,0
4 × 4	4	4	0,4	0,5	0,8	1,7	8,8	0,9	1,3	2,0	4,3	22,5
5 × 5	5	5	0,6	0,9	1,4	3,0	15,5	1,7	2,4	3,7	7,9	41,2
6 × 6	6	6	0,5	0,7	1,1	2,4	12,3	1,7	2,4	3,8	8,0	41,9
7 × 7	7	7	0,2	0,3	0,5	1,1	5,9	0,8	1,1	1,8	3,7	19,3
Média	-	-	0,4	0,6	0,9	1,9	9,7	1,1	1,5	2,4	5,1	26,6
----- SAR – Análise estrutural com base na matriz de correlação -----												
2 × 2	2	2	0,4	0,5	0,8	1,8	9,2	0,8	1,1	1,7	3,7	19,1
3 × 3	3	3	0,4	0,6	0,9	2,0	10,2	0,7	0,9	1,5	3,1	16,4
4 × 4	4	4	0,5	0,7	1,2	2,5	12,9	0,9	1,3	2,1	4,4	23,1
5 × 5	5	5	1,0	1,3	2,1	4,5	23,5	2,0	2,9	4,6	9,7	50,4
6 × 6	6	6	0,7	1,0	1,6	3,4	17,6	2,0	2,9	4,6	9,7	50,7
7 × 7	7	7	0,7	1,0	1,5	3,3	16,9	0,8	1,1	1,8	3,7	19,4
Média	-	-	0,6	0,9	1,4	2,9	15,1	1,2	1,7	2,7	5,7	29,8
----- SAS – Análise estrutural com base na matriz de covariância -----												
2 × 2	2	2	0,4	0,5	0,8	1,8	9,2	0,8	1,1	1,7	3,7	19,1
3 × 3	3	3	0,4	0,6	0,9	2,0	10,2	0,7	0,9	1,5	3,1	16,4
4 × 4	4	4	0,5	0,7	1,2	2,5	12,9	0,9	1,3	2,1	4,4	23,1
5 × 5	5	5	1,0	1,3	2,1	4,5	23,5	2,0	2,9	4,6	9,7	50,4
6 × 6	6	6	0,7	1,0	1,6	3,4	17,6	2,0	2,9	4,6	9,7	50,7
7 × 7	7	7	0,7	1,0	1,5	3,3	16,9	0,8	1,1	1,8	3,7	19,4
Média	-	-	0,6	0,9	1,4	2,9	15,1	1,2	1,7	2,7	5,7	29,8

Fonte: Autor.

Quando o *rptb* foi estimado pelo método de componentes principais, com base na matriz de correlação (PCR) ou na matriz variância e covariância (PCS), o η_m variou de 0,3 a 0,7 e entre 0,4 e 1,0 quando utilizado *rptb* foi obtido pelo método de análise estrutural com base na matriz de correlação (SAR) ou matriz de variância e covariância (SAS). Na cultivar Temprano, os valores de η_m foram um pouco superiores aos encontrados para a cultivar BRS Progresso. Portanto, independentemente do nível de precisão aceito na estimativa de ρ_{U_1, V_1} , maior número de medições ou de ensaios é necessário em ensaio com a cultivar Temprano. A cultivar tem como propósito de uso como planta forrageira e/ou planta de cobertura de solo e portanto, novos estudos devem ser conduzidos, com a finalidade de verificar a necessidade de maior número de ensaios está relacionado com a cultivar e/ou o seu propósito de uso.

O método da AOV apresentou maiores necessidade de η_m . No entanto, esse método apresentou os menores valores de coeficiente determinação (R^2) na estimativa do coeficiente de repetibilidade (*rptb*), quando comparado com os outros métodos (Tabela 21). O método de componentes principais (PCR ou PCS), resultou em menores valores de η_m , mas com acurácia na estimação de *rptb*. Diante disso, é preferível que seja utilizado o método que conseguiu estimar o *rptb* com maior precisão.

A superioridade nas estimativas de *rptb* pelo método de componentes principais também foi verificado em estudo com (*Urochloa brizantha*) (TORRES *et al.*, 2015) e foi recomendado para a determinação da repetibilidade com maior acurácia (ABEYWARDENA, 1972). Diante disso, as inferências do número de medições ou ensaios foram realizadas utilizando os valores de η_m calculados utilizando as estimativas de *rptb* pelo método PCR.

Pouca variabilidade nas estimativas de η_m foi percebida entre os métodos PCR e PCS, dentro de uma mesma cultivar, cenário e nível de precisão. Objetivando precisão mínima de 85% nas estimativas de ρ_{U_1, V_1} , um único ensaio é suficiente para que seja estimado a correlação do primeiro par canônico em caracteres da BRS Progresso, independentemente do número de caracteres em cada grupo de variáveis. Já na cultivar Temprano, para que se obtenha as estimativas de ρ_{U_1, V_1} com precisão mínimo de 85%, são necessários de dois a três ensaios, dependendo do número de caracteres em cada grupo de variáveis.

Diante da necessidade de aumentar a precisão nas estimativas de ρ_{U_1, V_1} , maior número de ensaios são necessários. Para precisão mínima de 90%, são necessários de dois a três ensaios na cultivar BRS Progresso, e de dois a quatro na cultivar Temprano. Em ambas as cultivares, maiores estimativas de η_m foram observadas em cenários contendo cinco e seis caracteres

(cenários 5×5 e 6×6), com necessidade de 18 e 44 medições ou ensaios para estimar ρ_{U_1, V_1} com nível de precisão de 99% em BRS Progresso e Temprano, respectivamente.

Na prática, realizar 44 medições, ou seja, conduzir 44 ensaios se torna trabalhoso e ineficaz devido ao alto investimento de recursos financeiros, de material e mão-de-obra. Assim, para otimização de recursos, menor número de ensaios podem ser utilizados, desde que o pesquisador esteja ciente que estará sacrificando a precisão ao realizar a análise de correlação canônica e obtenção das estimativas de ρ_{U_1, V_1} , em prol de utilizar menor número de ensaios e redução no custo de implantação e coleta dos dados de campo. Cabe ao pesquisador utilizar as informações deste estudo, associados a interpretações de trabalhos prévios e decidir pelo número de medições ou de ensaios, observando o número de caracteres em cada grupo de variáveis e a precisão almejada. Nesse processo, o pesquisador deverá fazer uma consulta a literatura e identificar os caracteres associados entre si para evitar que planeje avaliar um grupo de variáveis e no passo anterior a realização da análise de correlação canônica, necessite realizar o diagnóstico de multicolinearidade e necessite a eliminação de um ou mais caracteres.

Considerando hipoteticamente que se deseja obter precisão mínima de 95% nas estimativas de ρ_{U_1, V_1} , dois ensaios são suficientes na cultivar BRS Progresso, com os grupos contendo dois, três, quatro ou sete caracteres. Já, para a estimativa da correlação do primeiro par canônico com cinco e seis caracteres em cada grupo de variáveis, quatro e três ensaios são necessários, respectivamente. Na cultivar Temprano, as estimativas de ρ_{U_1, V_1} com o mesmo nível de precisão podem ser obtidas ao ser utilizado quatro ensaios com dois, três e sete caracteres em cada grupo de variáveis, ou cinco, nove e oito ensaios em cenários com quatro, cinco e seis caracteres, respectivamente.

4.5.4 Considerações sobre a repetibilidade e número de ensaio para a correlação canônica

Pode ser observado dependência na estimativa da correlação do primeiro par canônico (ρ_{U_1, V_1}) com o número de caracteres utilizados em cada grupo de variáveis. Desejando obter maior precisão nas estimativas de ρ_{U_1, V_1} , há a necessidade de utilização de maior número de medições ou ensaios (η_m) em cenários com cinco a seis caracteres, em cada grupo de variáveis. Não foram encontrados na literatura, trabalhos que estimaram a repetibilidade para a estimativa do primeiro par canônico e que relacionasse com o número de variáveis.

As inferências para determinar o η_m neste estudo foram realizadas pelo método de componentes principais utilizando a matriz de correlação (PCR). A escolha pelo método é devido a sua superioridade em estimar o coeficiente de repetibilidade ($rptb$), com precisão superior (R^2) a 96,6% e 87,1% na cultivar BRS Progresso e Temprano, respectivamente. O método de componentes principais (PC) tem sido recomendado para estimar a repetibilidade com maior precisão (ABEYWARDENA, 1972). Em estudos com outras culturas, o método PCR também forneceu estimativas mais precisas de $rptb$ de características em braquiária (TORRES *et al.*, 2015) e ($R^2 \geq 0,96$) em capim-elefante (*Pennisetum spp.*) (CAVALCANTE *et al.*, 2012).

O maior número de ensaios deve ser priorizado em ensaios conduzidos com a cultivar Temprano, com valores de η_m um pouco superiores às encontradas na cultivar BRS Progresso para estimar a correlação canônica (ρ_{U_1, V_1}). Foi observado também a necessidade de maiores η_m em níveis de precisões mais elevados. Desejando obter estimativas de ρ_{U_1, V_1} com precisão mínima de 95%, são necessárias de dois a quatro ensaios na cultivar BRS Progresso e de quatro a nove ensaios na Temprano.

Os resultados encontrados do número de ensaios para estimar a correlação do primeiro par canônico se assemelham com alguns trabalhos que determinaram o número de ensaios para estimar as estimativas de características em outras culturas. Ao determinar o número de medições necessárias para a predição do desempenho de genótipos de Capim-tanzânia (*Panicum maximum* Jacq.), oito cortes seriam suficientes com 80% de exatidão na estimação real valor da característica avaliada (CARGNELUTTI FILHO *et al.*, 2004). Três a sete medições foram recomendados para estimação de características de capim-elefante (*Pennisetum spp.*) com precisão de $R^2 = 90\%$ (CAVALCANTE *et al.*, 2012) e duas a quatro medições para características de braquiária (*Urochloa brizantha*) com confiabilidade de 80% (TORRES *et al.*, 2015). Em trigo (*Triticum aestivum*), uma única medição seria suficiente para estimar da porcentagem de grãos germinados com acurácia de 90% ($R^2 = 0,90$) e maior número de medições devem ser realizadas quando maior precisão ($R^2 = 0,99$) for almejada ($\eta_m = 7$) (PAGLIOSA *et al.*, 2014). No entanto, em características de soja (*Glycine max*), seriam necessários 15 avaliações para a predição dos comprimentos do hipocótilo, do epicótilo e do internódio ($R^2 = 0,90$) e 21 avaliações para o comprimento do pecíolo da primeira folha trifoliolada ($R^2 = 0,85$) (MATSUO *et al.*, 2012).

Para a estimativa da correlação do primeiro par canônico, o número de medições ou de ensaios varia com o número de variáveis, a cultivar e com o nível de precisão, devendo o

pesquisador adotar um valor, estando ciente do nível de precisão a ser alcançado na análise. Na hipótese de ser almejada precisão mínima de 95% nas estimativas de ρ_{U_1, V_1} , dois ensaios são suficientes na cultivar BRS Progresso, com os grupos contendo dois, três, quatro ou sete caracteres. Já, para a estimativa da correlação do primeiro par canônico com cinco e seis caracteres em cada grupo de variáveis, quatro e três ensaios são necessários. Na cultivar Temprano, as estimativas de ρ_{U_1, V_1} com o mesmo nível de precisão podem ser obtidas ao serem utilizados quatro ensaios com dois, três e sete caracteres, em cada grupo de variáveis; ou, cinco, nove e oito ensaios em cenários com quatro, cinco e seis caracteres, respectivamente.

Assim, conclui-se que:

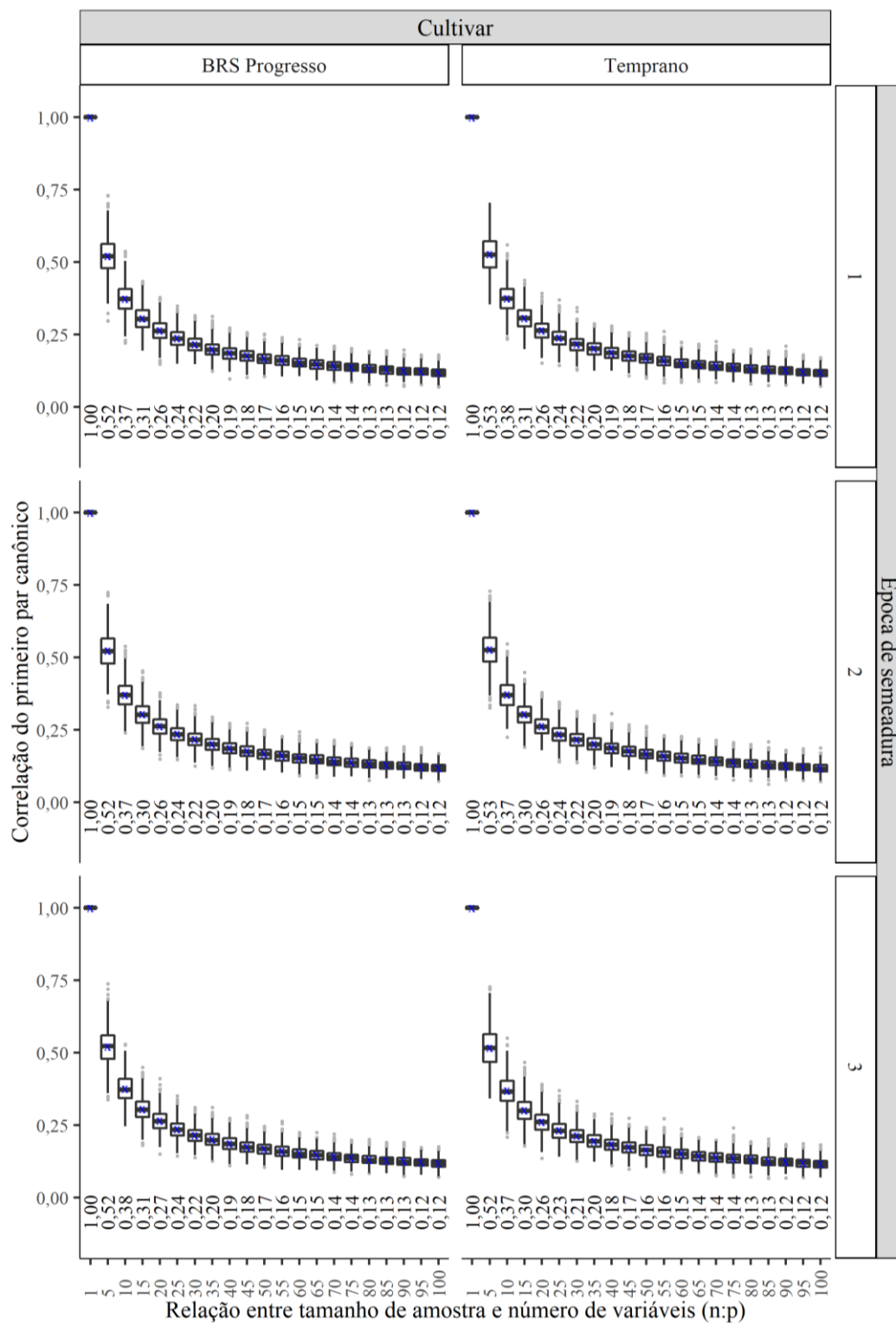
1. Há variabilidade na estimativa do número de medições da correlação do primeiro par canônico entre cultivares e cenário (números de caracteres combinados em cada grupo de variáveis).
2. Mais de um ensaio é necessário na estimação da correlação do primeiro par canônico na cultura do centeio.
3. Maior número de ensaios é necessário para estimar a da correlação do primeiro par canônico em variáveis da cultivar Temprano.
4. Com precisão mínima de 95%, até quatro e oito ensaios são necessários para estimar a correlação do primeiro par canônico nas cultivares BRS Progresso e Temprano, respectivamente.

4.6 RELAÇÃO ENTRE TAMANHO DE AMOSTRA E NÚMERO DE VARIÁVEIS EM CORRELAÇÃO CANÔNICA

4.6.1 Análise de correlação canônica e reamostragem

A magnitude das estimativas da correlação do primeiro par canônico (ρ_{U_1, V_1}) foram influenciadas pela relação entre o número de observações (tamanho de amostra) e de variáveis (n:p) (Figura 3). Padrões nas estimativas médias de ρ_{U_1, V_1} foram observadas em simulações com dados de ensaios de três épocas de semeadura e duas cultivares de centeio. Conforme os resultados simulados, não pode ser observado variação nas estimativas de (ρ_{U_1, V_1}) em função de características de cultivar e época de semeadura.

Figura 3 – Médias e diagrama ‘boxplot’ de correlação do primeiro par canônico obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em três épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

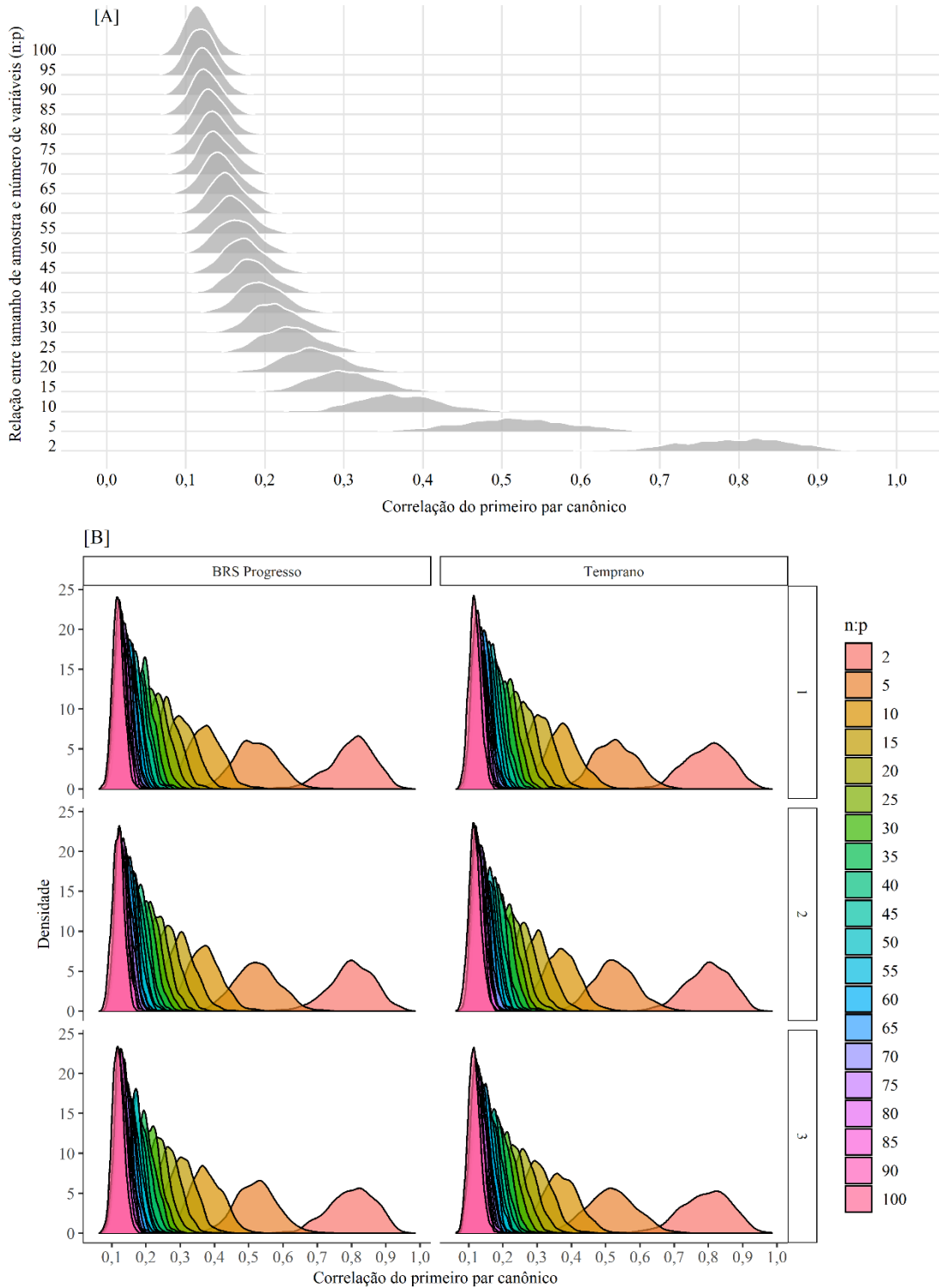


Fonte: Autor.

Maiores estimativas de ρ_{U_1, V_1} foram observadas em simulações com menores relações n:p, com estimativa $\rho_{U_1, V_1} = 1,00$ em relação n:p = 1. À medida que houve acréscimos de número de observações (maior relação n:p), menores foram as estimativas médias do parâmetro ρ_{U_1, V_1} , com tendência de estabilização em altas relações n:p. Ao mesmo tempo que menores estimativas de ρ_{U_1, V_1} foram observadas em simulações de maiores relações, menores amplitudes do intervalo de confiança ($IAC_{95\%}$) e dispersões das estimativas de ρ_{U_1, V_1} foram verificadas (Figuras Figura 4 e Figura 5). Esses resultados evidenciam que o tamanho de amostra e o número de variáveis possuem efeito sobre a estimativa de ρ_{U_1, V_1} e que as estimativas em correlação canônica podem estar sendo obtidas com menor precisão com condição de menor relação entre o número de observações e de variáveis.

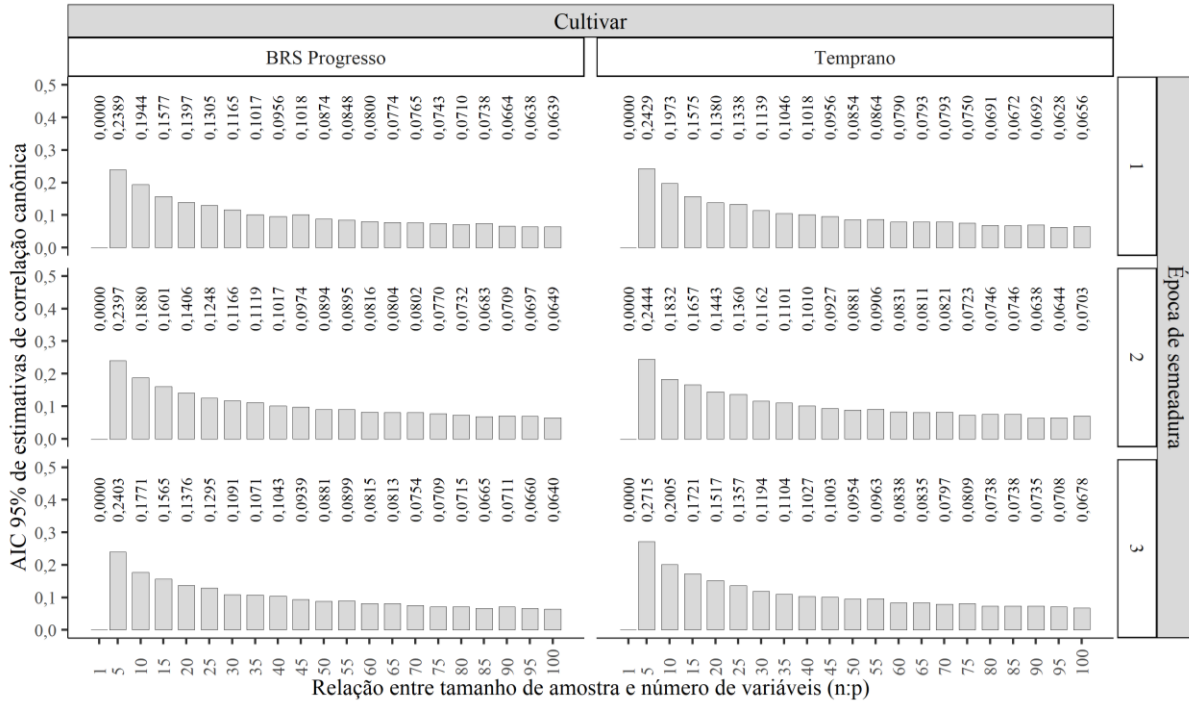
De maneira semelhante ao comportamento das estimativas do parâmetro ρ_{U_1, V_1} , o de redundância ou a fração total da variância de um grupo variáveis contabilizada pelas variáveis canônicas do outro grupo, apresentaram maiores estimativas em simulações com menores relações n:p e menores estimativas em simulações com maiores relações n:p (Figuras Figura 6 e Figura 7). Em média, aproximadamente 55% da variância nas variáveis do primeiro grupo de variáveis foram explicadas pelas variáveis canônicas do segundo grupo e vice-versa, em análise de correlação canônica realizada em todos os ensaios com a cultivar BRS Progresso e relação n:p = 1. Em dados com a terceira época da cultivar Temprano, o índice de redundância variou de 0,54 a 0,59 para a mesma relação n:p. Independentemente da cultivar e da época de semeadura, a média da amplitude do intervalo de confiança de 95% ($AIC_{95\%}$) para a quantidade de variância total nas variáveis do primeiro grupo explicada pelas variáveis canônicas do segundo grupo foi superior a 0,33. Enquanto que na relação n:p = 2, a $AIC_{95\%}$ média foi inferior a 0,12, com diminuição gradativa em simulações com relações maiores n:p.

Figura 4 – Histogramas de densidades da correlação do primeiro par canônico obtidas em simulações em diferentes relações entre o número de observações (tamanho de amostra) e de variáveis em ensaio conduzido na primeira época com a cultivar BRS Progresso [A] e em três épocas e duas cultivares de centeio (*Secale cereale* L.) [B], ensaios conduzidos em Santa Maria – RS, Safra 2016.



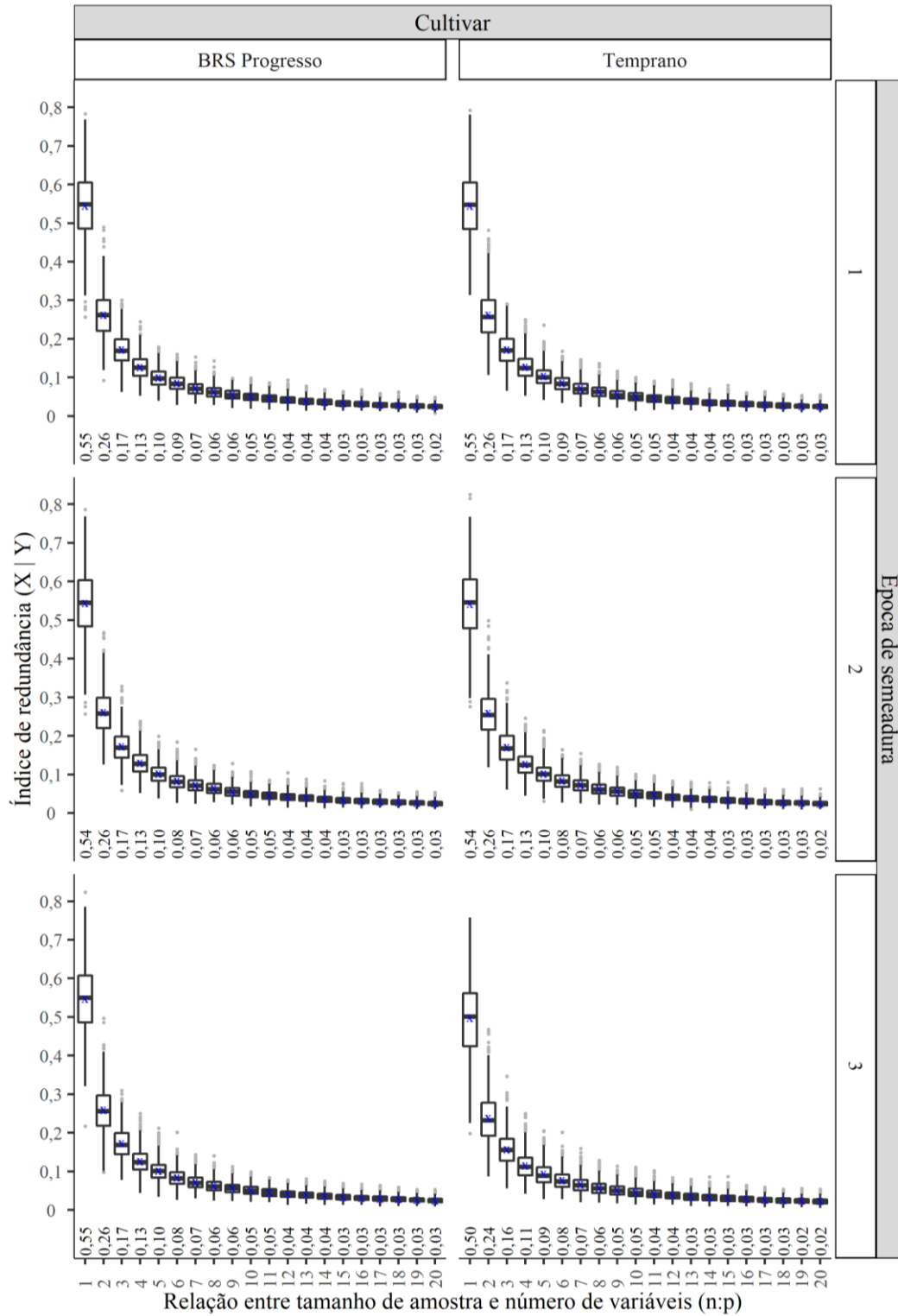
Fonte: Autor.

Figura 5 – Amplitude do intervalo de confiança de 95% (AIC 95%) de estimativas de correlação do primeiro par canônico obtidas em simulações em diferentes relações entre o número de observações (tamanho de amostra) e de variáveis em ensaio conduzido em três épocas de semeadura e duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



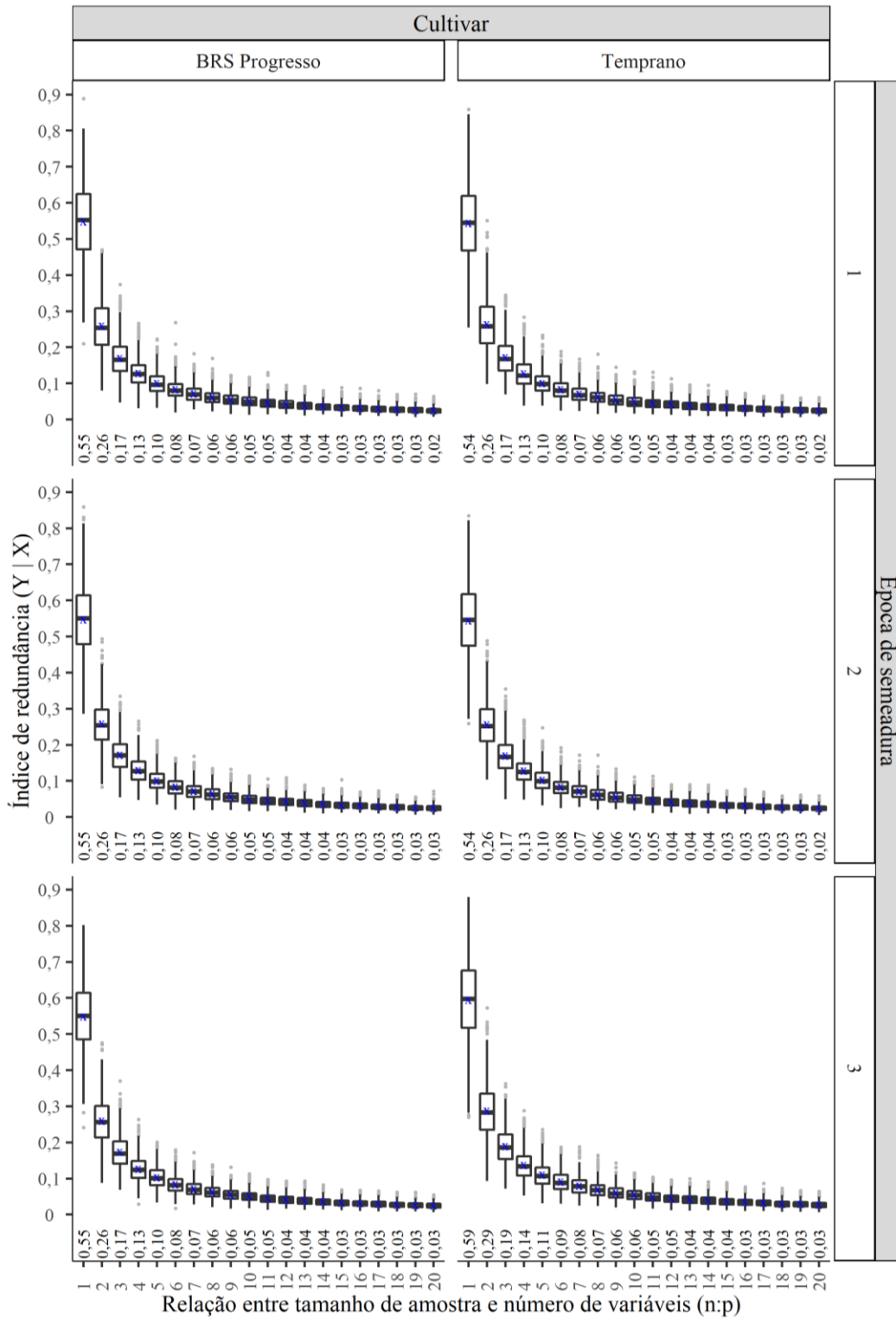
Fonte: Autor.

Figura 6 – Médias e diagrama ‘boxplot’ do índice de redundância ($X | Y =$ fração total da variância do primeiro grupo contabilizada pelas variáveis canônicas do segundo grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



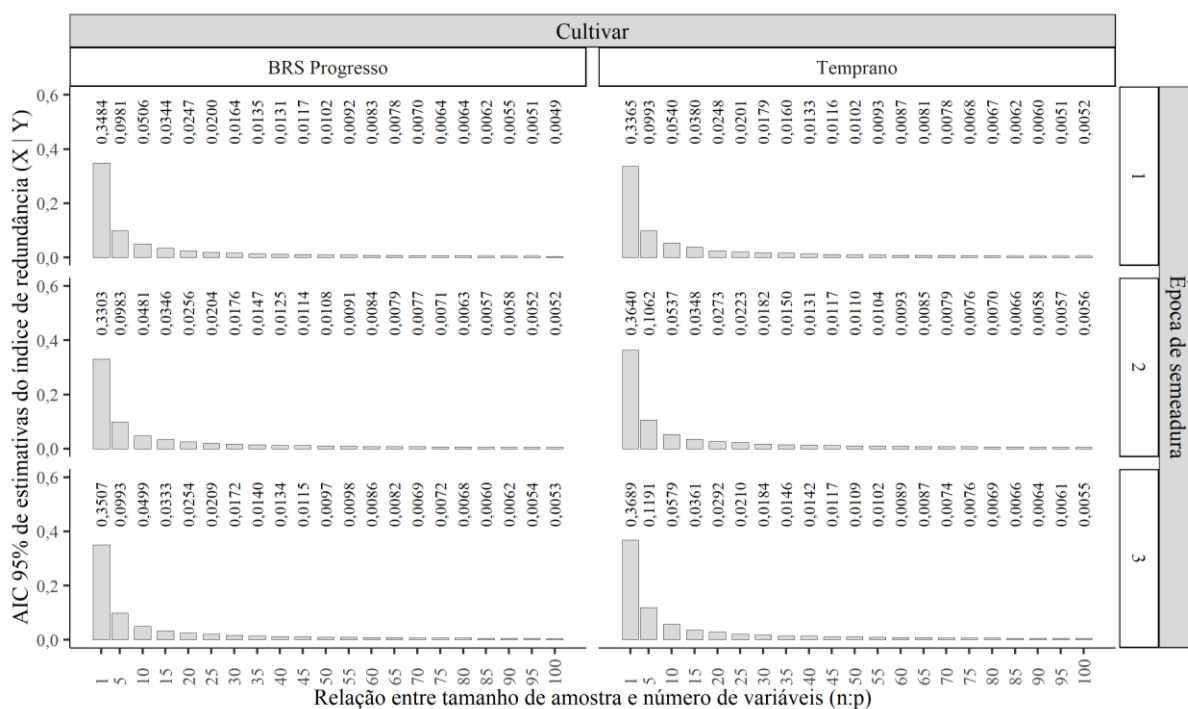
Fonte: Autor.

Figura 7 – Médias e diagrama ‘boxplot’ do índice de redundância ($Y | X =$ fração total da variância do segundo grupo contabilizada pelas variáveis canônicas do primeiro grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



Fonte: Autor.

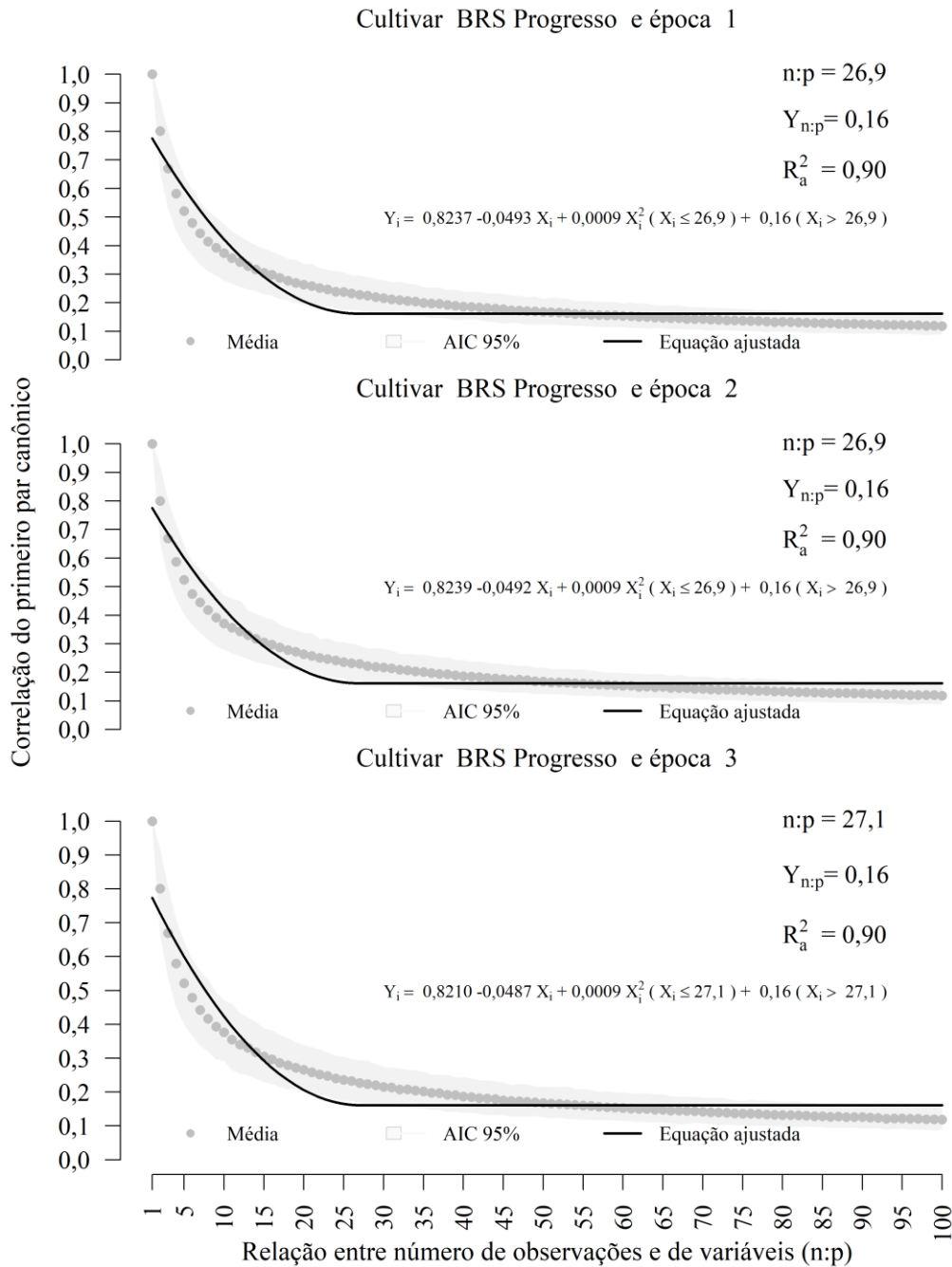
Figura 8 – Amplitude do intervalo de confiança de 95% (AIC_{95%}) de estimativas de índice de redundância (X | Y = fração total da variância do primeiro grupo contabilizada pelas variáveis canônicas do segundo grupo) obtidos em 1.000 simulações em diferentes relações entre o número de observações e de variáveis em análises realizadas em ensaios conduzidos em duas épocas de semeadura e de duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



4.6.2 Determinação da relação entre o número de observações e de variáveis

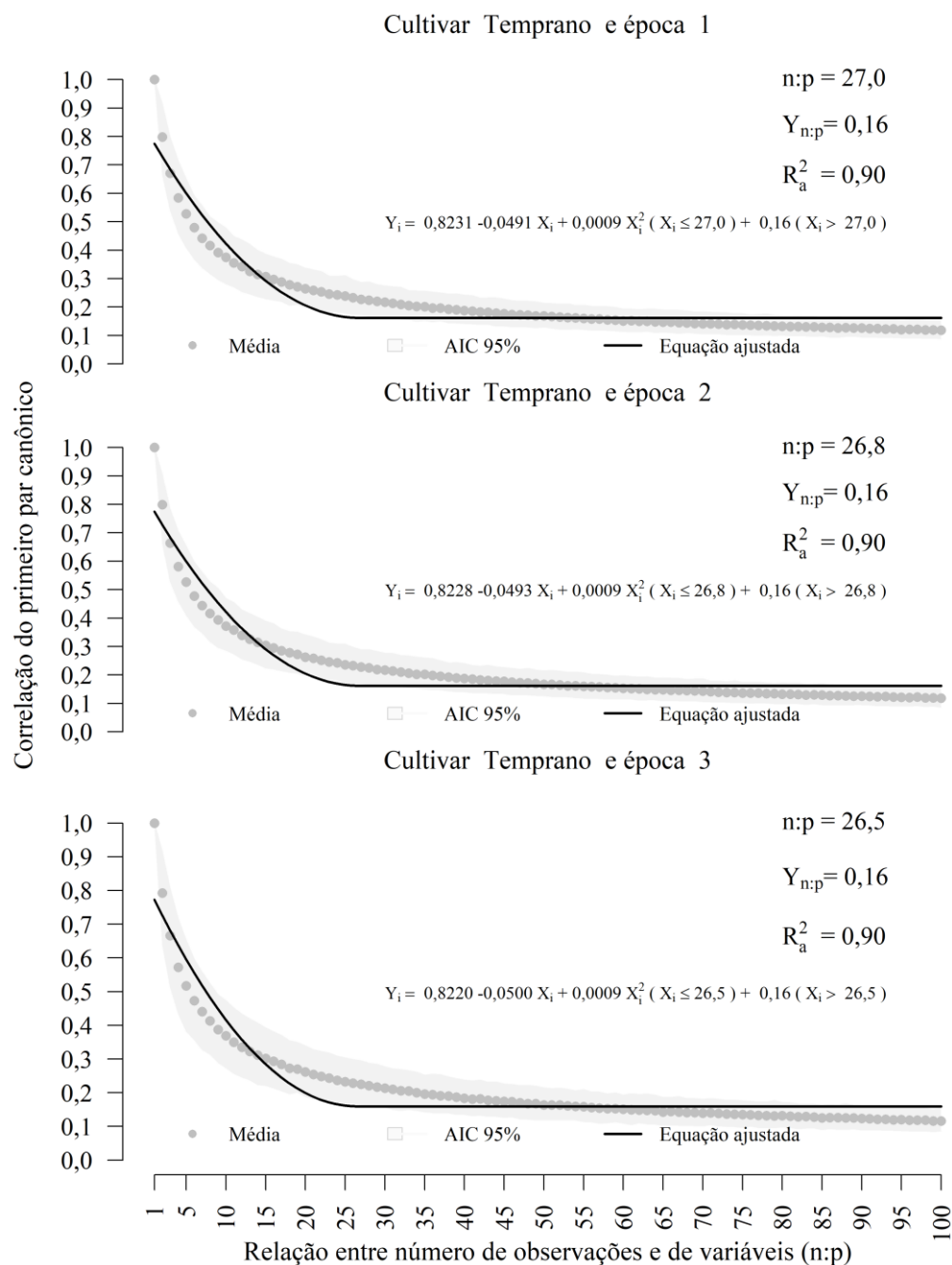
Os valores da relação entre o número de observações e de variáveis (n:p) para estimar a correlação do primeiro par canônico (ρ_{U_1, V_1}), diferiram apenas entre os métodos utilizados (Figuras Figura 9 e Figura 10, Tabela 23). De acordo com os resultados obtidos em simulações, foi constatado que relação n:p muito semelhantes é requerida na análise de correlação canônica realizadas com caracteres provenientes de três épocas de semeadura e duas cultivares de centeio.

Figura 9 – Médias e amplitude do intervalo de confiança de 95% (AIC 95%) das estimativas de correlação do primeiro par canônico, relação entre o número de observações e de variáveis (n:p) determinado pelo método de modelo quadrático com resposta em platô (MQRP), estimativa da correlação canônica estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R_a^2), estimado em três épocas de semeadura com a cultivar BRS Progresso (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



Fonte: Autor.

Figura 10 – Médias e amplitude do intervalo de confiança de 95% (AIC 95%) das estimativas de correlação do primeiro par canônico, relação entre o número de observações e de variáveis (n:p) determinado pelo método de modelo quadrático com resposta em platô (MQRP), estimativa da correlação canônica estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R_a^2), estimado em três épocas de semeadura com a cultivar Temprano (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.



Fonte: Autor.

Tabela 23 – Estimativas da relação entre o número de observações e de variáveis (n:p) determinados por dois métodos de regressão com resposta em platô, estimativa da correlação do primeiro par canônico estimada para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura e duas cultivares de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Época ¹	MLRP ²			MQRP ³		
		n:p	$Y_{n:p}$	R^2_a	n:p	$Y_{n:p}$	R^2_a
BRS Progresso	1	17,6	0,17	0,86	26,9	0,16	0,90
BRS Progresso	2	17,6	0,17	0,86	26,9	0,16	0,90
BRS Progresso	3	17,6	0,17	0,86	27,1	0,16	0,90
Temprano	1	17,6	0,17	0,86	27,0	0,16	0,90
Temprano	2	17,6	0,17	0,86	26,8	0,16	0,90
Temprano	3	17,5	0,16	0,86	26,5	0,16	0,90
Média	-	17,6	-	-	26,9	-	-

¹Detalhes de épocas de semeadura na Tabela 2. ²MLRP: Modelo linear com resposta em platô. ³MQRP: Modelo quadrático com resposta em platô.

Fonte: Autor.

Entre os métodos, o modelo linear com resposta em platô (MLRP) estimou a necessidade de menor relação n:p para estimar a ρ_{U_1, V_1} , comparado com o modelo quadrático com resposta em platô (MQRP). A estimativa média da relação n:p pelo método MLRP foi de 17,6 ($17,5 \leq n:p \leq 17,6$). Pelo método MQRP, o valor médio da relação n:p foi 52,84% ($n:p = 26,9$) superior à estimativa média obtida pelo método MLRP, com explicação de 90% da variação de ρ_{U_1, V_1} , enquanto que o coeficiente de determinação ajustado (R^2_a) do método MLRP foi de 0,86. O maior poder de explicação do método MQRP em relação ao método MLRP pode ser utilizado como modelo de previsão ρ_{U_1, V_1} em função da relação n:p e também como critério para determinar a relação n:p.

Superioridade de média da relação n:p determinado pelo método MQRP foi confirmada pelo teste *t* de *Student* para amostras independentes a 5% de probabilidade de erro ($t = 116,26$; $GL = 5,2158$; $\text{valor-p} \leq 0,0000$). Diante da maior explicação e maior média da relação n:p, o método MQRP foi utilizado na determinação da relação entre o número de observações e de variáveis para a correlação do primeiro par canônico.

A determinação da relação entre o número de observações e de variáveis para a estimativa de índice de redundância na análise de correlação canônica apresentou valores inferiores aos necessários para a estimação da correlação do primeiro par canônico (Tabela 24). Independentemente da cultivar e época de semeadura, a relação de $n:p = 5,2$ foi observada pelo método MLRP, com coeficiente de determinação próximo de 0,88. O método MQRP apresentou maior explicação da variação do índice de redundância, com $R^2_a = 0,91$ para a

quantidade de variância da variável do primeiro grupo explicada pelas variáveis canônicas do segundo grupo ($X | Y$) e $0,87 \leq R^2_a \leq 0,88$ para a quantidade de variância explicada da variável no segundo grupo pelas variáveis canônicas do primeiro grupo ($Y | X$). Para este parâmetro, 6,5 observações para cada variável seriam suficientes, segundo o método MQRP.

Tabela 24 – Estimativas da relação entre o número de observações e de variáveis (n:p) determinados por dois métodos de regressão com resposta em platô, para a estimativa do índices de redundâncias em análise de correlação canônica, estimativa do índice para a relação n:p ($Y_{n:p}$) e coeficiente de determinação ajustado (R^2_a), estimado em três épocas de semeadura e duas cultivar de centeio (*Secale cereale* L.), ensaios conduzidos em Santa Maria – RS, Safra 2016.

Cultivar	Época ¹	MLRP ²			MQRP ³		
		n:p	$Y_{n:p}$	R^2_a	n:p	$Y_{n:p}$	R^2_a
----- Índice de redundância – X Y -----							
BRS Progresso	1	5,2	0,02	0,88	6,5	0,02	0,91
BRS Progresso	2	5,2	0,02	0,88	6,6	0,02	0,91
BRS Progresso	3	5,2	0,02	0,88	6,4	0,02	0,91
Temprano	1	5,2	0,02	0,88	6,6	0,02	0,91
Temprano	2	5,2	0,02	0,88	6,5	0,02	0,91
Temprano	3	5,2	0,01	0,88	6,5	0,01	0,91
----- Índice de redundância – Y X -----							
BRS Progresso	1	5,2	0,02	0,88	5,2	0,02	0,88
BRS Progresso	2	5,2	0,02	0,88	5,2	0,02	0,88
BRS Progresso	3	5,2	0,02	0,88	5,2	0,02	0,88
Temprano	1	5,2	0,02	0,88	5,2	0,02	0,88
Temprano	2	5,2	0,02	0,87	5,2	0,02	0,87
Temprano	3	5,2	0,02	0,88	5,2	0,02	0,88

¹Detalhes de épocas de semeadura na Tabela 2. ²MLRP: Modelo linear com resposta em platô. ³MQRP: Modelo quadrático com resposta em platô.

Fonte: Autor.

4.6.3 Considerações parciais sobre a relação entre o tamanho de amostra e o número de variáveis

O tamanho de amostra e o número de variáveis utilizadas em análise de correlação canônica possuem efeito na estimativa da correlação do primeiro par canônico (ρ_{U_1, V_1}). Maiores estimativas de ρ_{U_1, V_1} foram verificadas em simulações de menores relações entre o número de observações e de variáveis (n:p), com médias de ρ_{U_1, V_1} oscilando de 1,00 à 0,12, em simulações com a relação entre o número de observações e de variáveis (n:p) de 1 à 100 (Figura 3).

A maior amplitude do intervalo de confiança de 95% ($AIC_{95\%}$) foi verificado nas simulações com menor relação n:p, gerando imprecisão na obtenção da estimativa de correlação do primeiro par canônico. Maiores e menores amplitudes do intervalo de confiança em menores e maiores relações n:p já eram esperadas, pois o tamanho de amostra impacta na precisão da estimativa da característica avaliada. O mesmo foi verificado em estudo de determinação do tamanho de amostra para a estimação de coeficientes de correlação linear de Pearson (TOEBE *et al.*, 2015) e análise de trilha em híbridos de milho (TOEBE *et al.*, 2017b) e determinação do número de observações para a estimativa de parâmetros em modelo em regressão (CARGNELUTTI FILHO; TOEBE, 2020). O número de observações ou tamanho de amostra possui relevância na obtenção de estimativas das características em estudo e na análise de correlação canônica, pode ser verificado o mesmo, relacionando o tamanho de amostra e o número de variáveis na estimativa dos parâmetros.

Pelas simulações, em relação n:p = 1, o primeiro par de variáveis canônicas revelou estar fortemente correlacionado ($\rho_{U_1, V_1} = 1$). Nessa relação, para nem todas as 1.000 simulações foi possível estimar a estatística Qui-Quadrado (χ^2). Somente em 48,4% das amostras, o valor da estatística pôde ser estimado e com estes, o valor médio da estatística foi de 132,57. Enquanto nas demais relações, a estatística apresentou valores médios entre 29,25 e 36,99.

Não foram encontrados trabalhos com a cultura do centeio ou trabalhos utilizando a mesma metodologia para que os resultados pudessem ser comparados. No entanto, em análise de regressão múltipla, foi apontado que houve redução da probabilidade de erro na estimação do intervalo de confiança dos parâmetros em condições de maiores tamanhos de amostra (BONETT; WRIGHT, 2011). Os resultados deste estudo foram divergentes da recomendação generalizada para a análise multivariada, onde há a recomendação prática da necessidade de relação mínima de dez observações para cada variável (HAIR *et al.*, 2009).

Apesar de ser determinado haver a necessidade de relações de 5,2 a 6,5 para estimar o índice de redundância em análise de correlação canônica, essa relação pode ser insuficiente para estimar a correlação do primeiro par canônico. O uso de menores relações deve ser evitadas, embora haja recomendação para determinado parâmetro, como por exemplo, o índice de redundância. Dentro da possibilidade de material, espaço e mão-de-obra, uso de recomendações com base em parâmetros que necessitem a maior relação entre o número de observações e de variáveis deve ser preferido no planejamento experimental.

Indicações de diferentes relações foram realizadas em simulações em análise de regressão logística, sendo apontado que em relações maiores ou iguais a 10:1, não ocorreram

problemas na precisão e significância dos coeficientes de regressão (PEDUZZI *et al.*, 1996; WYNANTS *et al.*, 2015). No entanto, em outro estudo envolvendo a mesma abordagem, os autores alertam que em relação 10:1, os parâmetros podem estar sendo estimados com baixa confiabilidade (VAN SMEDEN *et al.*, 2016).

A relação entre o número de observações e de variáveis para a estimação da correlação do primeiro par canônico pelo método MQRP quando utilizada como referência na cultura do centeio apresenta valores semelhantes aos resultados de trabalhos de abordagens semelhantes. Foi apontado haver a necessidade de 20 observações para cada variável para a obtenção de estimativas confiáveis e estáveis dos coeficientes e correlações canônicos (BARCIKOWSKI; STEVENS, 1975), enquanto que, considerando as duas maiores correlações canônicas, seriam necessário de 42 a 68 observações para cada variável (ALEXANDER, 2000; BARCIKOWSKI; STEVENS, 1975).

De acordo com o presente estudo, são necessárias aproximadamente 27 observações ($n:p = 26,9 \cong 27$ observações) para cada variável a ser utilizada na análise de correlação canônica. Ou seja, supondo que o pesquisador deseja realizar a análise com cinco variáveis no primeiro grupo ($p = 5$) e seis variáveis no segundo grupo ($q = 6$), seriam necessárias de $n = (p + q) \times n:p = (5 + 6) \times 27 = 297$ observações.

Assim, conclui-se que:

1. Há variabilidade nas estimativas da correlação do primeiro par canônico em função do tamanho de amostra e número de variáveis e essa variabilidade diminui à medida que maior número de observações em relação ao número de variáveis são utilizados.
2. Diferentes estimativas de relação entre o número de observações e de variáveis são obtidas pelos métodos de modelo linear com regressão em platô e modelo quadrático com regressão em platô, com maior explicação e estimativas obtidas pelo modelo quadrático com resposta em platô.
3. A mesma relação entre o número de observações e de variáveis pode ser utilizada em análise de correlação canônica, independentemente de cultivar e época de cultivo.
4. É recomendado utilizar 27 observações para cada variável presente na análise de correlação canônica em ensaios com a cultura do centeio.

5 CONSIDERAÇÕES FINAIS

Multicolinearidade em níveis elevados foram verificadas em diagnósticos com todos os oito caracteres morfológicos e produtivos e a eliminação dos caracteres morfológicos altura de planta e massa fresca de parte aérea planta⁻¹ e dos caracteres produtivos massa de grãos planta⁻¹ e número de grãos planta⁻¹ foram eficientes na redução do grau de multicolinearidade. Estes caracteres, altamente correlacionados com os demais caracteres dentro do mesmo grupo de variáveis, foram responsáveis pelo alto grau de multicolinearidade. A eliminação dos caracteres foi eficiente e a magnitude da correlação pode ser utilizado como o primeiro indicador a ser utilizado para identificar os caracteres causadores da multicolinearidade.

Estimativas confiáveis dos parâmetros é desejada e maior precisão pode ser conseguida com o adequado dimensionamento amostral. Além do tamanho de amostra, deve ser utilizado também adequado o número de ensaios no diagnóstico do grau de multicolinearidade. Alta concordância no grau de multicolinearidade entre ensaios foi observada no estudo, apresentando elevada repetibilidade entre as estimativas de multicolinearidade entre os ensaios. Entretanto, quantidades diferentes de ensaios entre as cultivares são necessários.

A multicolinearidade foi estimada por meio dos indicadores número de condição e fator de inflação da variância, estimado a partir da matriz de correlação linear. Diferenças nas estimativas de repetibilidade pode estar associada às correlações divergentes e as diferentes respostas de cada cultivar entre as épocas. Assim, há a necessidade de no mínimo três e 16 ensaios para o diagnóstico de multicolinearidade em caracteres na cultivar BRS Progresso e Temprano, respectivamente, com precisão mínima de 95%.

Na análise de correlação canônica, mais de um ensaio é necessário para estimar a correlação do primeiro par canônico. Para este parâmetro, elevada concordância para as estimativas da correlação do primeiro par canônico foi observada entre os ensaios. Desejando precisão mínima de 95%, são necessários de dois a quatro ensaios na cultivar BRS Progresso e de quatro a nove ensaios na Temprano para estimação da correlação do primeiro par canônico. De maneira semelhante à necessidade para estimar o grau de multicolinearidade, foi verificado a necessidade de maior número de ensaios em análises com a cultivar Temprano. Outros estudos devem ser realizados para verificar se esse padrão pode ser atribuído a característica da cultivar. A cultivar BRS Progresso possui aptidão para a produção de grãos e cultivar Temprano, uso como planta forrageira e/ou planta de cobertura de solo.

As simulações em análise de correlação canônica com diferentes relações entre o tamanho de amostra e o número de variáveis, apontaram que estimativas de menor precisão

podem ser obtidas quando se utiliza baixa relação entre o número de observações e de variáveis. Em menores relações $n:p$, as estimativas foram obtidas em uma amplitude do intervalo maior, gerando maior incerteza no verdadeiro valor para a característica.

Diferenças na estimativa dos parâmetros não foram verificadas entre cultivares e épocas de semeadura. As similaridades nas estimativas entre os ensaios podem ser interpretadas como positivas, tanto pela não necessidade de diferentes relações em diferentes cultivares, bem como, credibilidade ao presente estudo.

As correlações, em média, possuem suas estimativas reduzidas à medida que se aumenta a relação $n:p$, mas com menor amplitude do intervalo de confiança de 95%. O modelo quadrático com resposta em platô foi o modelo que melhor ajustou a estimativa da correlação canônica em função da relação entre o número de observações e de variáveis e é recomendado o emprego de aproximadamente 27 observações para cada variável utilizada na análise de correlação canônica em ensaios com a cultura do centeio.

6 CONCLUSÕES

Conclui-se que:

1. Um ensaio não é eficiente para estimar o grau de multicolinearidade pelos indicadores número de condição e fator de inflação da variância em caracteres morfológicos e produtivos e a correlação do primeiro par canônico em centeio.
2. Há variabilidade nas estimativas do número de medições entre cultivares e números de caracteres combinados em cada grupo de variáveis.
3. Não há variabilidade no número de ensaios para realizar o diagnóstico de multicolinearidade pelos indicadores número de condição e fator de inflação de variância.
4. São necessários no mínimo três e 16 ensaios para o diagnóstico de multicolinearidade em caracteres na cultivar BRS Progresso e Temprano, respectivamente, com precisão mínima de 95%.
5. Há variabilidade na estimativa do número de medições da correlação do primeiro par canônico entre cultivares e cenário (números de caracteres combinados em cada grupo de variáveis).
6. Maior número de ensaios é necessário para estimar a da correlação do primeiro par canônico em variáveis da cultivar Temprano.
7. Com precisão mínima de 95%, até quatro e oito ensaios são necessários para estimar a correlação do primeiro par canônico nas cultivares BRS Progresso e Temprano, respectivamente.
8. Há variabilidade nas estimativas da correlação do primeiro par canônico em função do tamanho de amostra e número de variáveis e essa variabilidade diminui à medida que maior número de observações em relação ao número de variáveis são utilizados.
9. A mesma relação entre o número de observações e de variáveis pode ser utilizado em análise de correlação canônica, independentemente de cultivar e época de cultivo.
10. É recomendado utilizar 27 observações para cada variável presente na análise de correlação canônica em ensaios com a cultura do centeio.

REFERÊNCIAS

ABEYWARDENA, V. An application of principal component analysis in genetics. **Journal of Genetics**, [s. l.], v. 61, n. 1, p. 27–51, 1972. Disponível em: <http://link.springer.com/10.1007/BF02984099>. Acesso em: 29 nov. 2021.

ABOU CHEHADE, L. *et al.* Rye (*Secale cereale* L.) and squarrose clover (*Trifolium squarrosum* L.) cover crops can increase their allelopathic potential for weed control when used mixed as dead mulch. **Italian Journal of Agronomy**, [s. l.], v. 16, n. 4, p. 1–11, 2021. Disponível em: <https://agronomy.it/index.php/agro/article/view/1869>. Acesso em: 15 nov. 2021.

ALEXANDER, E. D. **Using canonical correlation to explore relationships between sets of variables: An applied example with interpretive suggestions**. Paper presented at the Annual Meeting of the Southwest Educational Research Association. Dallas, TX: [s. n.], 2000. Disponível em: <https://eric.ed.gov/?id=ED445082>. Acesso em: 2 fev. 2022.

ALIN, A. Multicollinearity. **Wiley Interdisciplinary Reviews: Computational Statistics**, [s. l.], v. 2, n. 3, p. 370–374, 2010. Disponível em: <http://doi.wiley.com/10.1002/wics.84>. Acesso em: 1 maio 2017.

ALPERT, M. I.; PETERSON, R. A. On the interpretation of canonical analysis. **Journal of Marketing Research**, [s. l.], v. 9, n. 2, p. 187, 1972. Disponível em: <https://www.jstor.org/stable/3149953?origin=crossref>. Acesso em: 4 fev. 2022.

ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, [s. l.], v. 22, n. 6, p. 711–728, 2013. Disponível em: http://www.schweizerbart.de/papers/metz/detail/22/82078/Koppen_s_climate_classification_map_for_Brazil?af=crossref. Acesso em: 20 set. 2019.

ALVES, B. M. *et al.* Correlações canônicas entre caracteres agronômicos e nutricionais proteicos e energéticos em genótipos de milho. **Revista Brasileira de Milho e Sorgo**, [s. l.], v. 15, n. 2, p. 171–185, 2016. Disponível em: <http://www.bibliotekevirtual.org/index.php/2013-02-07-03-02-35/2013-02-07-03-03-11/1982-rbms/v15n02/20296-correlacoes-canonicas-entre-caracteres-agronicos-e-nutricionais-proteicos-e-energeticos-em-genotipos-de-milho.html>. Acesso em: 4 maio 2017.

ALVES, B. M. *et al.* Linear relations among phenological, morphological, productive and protein-nutritional traits in early maturing and super-early maturing maize genotypes. **Journal of Cereal Science**, [s. l.], v. 70, p. 229–239, 2016. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0733521016301199>. Acesso em: 10 jun. 2017.

ALVES, B. M.; CARGNELUTTI FILHO, A.; BURIN, C. Multicollinearity in canonical correlation analysis in maize. **Genetics and Molecular Research**, [s. l.], v. 16, n. 1, p. 1–14, 2017. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/28363005>. Acesso em: 11 maio 2017.

ANDRADE JÚNIOR, V. C. *et al.* Repeatability and heritability of production characters in strawberry fruits. **Horticultura Brasileira**, [s. l.], v. 38, n. 1, p. 89–93, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-05362020000100089&tlng=en. Acesso em: 26 nov. 2021.

ATLÂNTICA SEMENTES S.A. **Centeio: Temprano**. Concessão: 2010.

AUSTIN, P. C.; STEYERBERG, E. W. The number of subjects per variable required in linear regression analyses. **Journal of Clinical Epidemiology**, [s. l.], v. 68, n. 6, p. 627–636, 2015. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2014.12.014>. Acesso em: 23 jan. 2022.

AZEVEDO, A. M. *et al.* Estudo da repetibilidade genética em clones de couve. **Horticultura Brasileira**, [s. l.], v. 34, n. 1, p. 54–58, 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-05362016000100054&lng=pt&tlng=pt. Acesso em: 28 dez. 2021.

BAIER, A. C. **Centeio**. Passo Fundo - RS: EMBRAPA Trigo, 1994. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/164511/1/FL-06193.pdf>. Acesso em: 20 set. 2019.

BANDEIRA, C. T. *et al.* Sample size to estimate the mean of morphological traits of rye cultivars in sowing dates and evaluation times. **Semina: Ciências Agrárias**, [s. l.], v. 39, n. 2, p. 521–532, 2018. Disponível em: <http://www.uel.br/revistas/uel/index.php/semagrarias/article/view/29999>. Acesso em: 17 dez. 2018.

BANDEIRA, C. T. *et al.* Sample sufficiency for estimation of the mean of rye traits at flowering stage. **Journal of Agricultural Science**, [s. l.], v. 10, n. 3, p. 178–186, 2018. Disponível em: <http://www.ccsenet.org/journal/index.php/jas/article/view/72446>. Acesso em: 17 dez. 2018.

BARCIKOWSKI, R. S.; STEVENS, J. P. A Monte Carlo study of the stability of canonical correlations, canonical weights and canonical variate-variable correlations. **Multivariate Behavioral Research**, [s. l.], v. 10, n. 3, p. 353–364, 1975. Disponível em: http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr1003_8. Acesso em: 2 fev. 2022.

BARTLETT, M. S. Multivariate analysis. **Supplement to the Journal of the Royal Statistical Society**, [s. l.], v. 9, n. 2, p. 176, 1947. Disponível em: <http://www.jstor.org/stable/2984113>. Acesso em: 8 maio 2017.

BARTLETT, M. S. The statistical significance of canonical correlations. **Biometrika**, [s. l.], v. 32, n. 1, p. 29, 1941. Disponível em: <https://www.jstor.org/stable/2332247?origin=crossref>. Acesso em: 4 fev. 2022.

BASCHE, A. D. *et al.* Soil water improvements with the long-term use of a winter rye cover crop. **Agricultural Water Management**, [s. l.], v. 172, p. 40–50, 2016. Disponível em: <http://dx.doi.org/10.1016/j.agwat.2016.04.006>. Acesso em: 17 dez. 2018.

BONETT, D. G.; WRIGHT, T. A. Sample size requirements for multiple regression interval estimation. **Journal of Organizational Behavior**, [s. l.], v. 32, n. 6, p. 822–830, 2011. Disponível em: <http://doi.wiley.com/10.1002/job.717>. Acesso em: 16 maio 2019.

BRUM, B. *et al.* Likelihood ratio test between two groups of castor oil plant traits. **Ciência Rural**, [s. l.], v. 46, n. 7, p. 1158–1164, 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782016000701158&lng=en&tlng=en. Acesso em: 8 maio 2017.

BUSHUK, W. Rye production and uses worldwide. **Cereal Foods World**, [s. l.], v. 46, n. 2, p. 70–73, 2001. Disponível em: http://www.agmrc.org/media/cms/bushuk_C8B79BAB55BB0.pdf. Acesso em: 13 nov. 2016.

BUSSAB, W. de O.; MORETIN, P. A. **Estatística Básica**. 8. ed. São Paulo: Saraiva, 2013.

BUTTS, C. T. **Yacca: Yet another canonical correlation analysis package**. Versão 1.1.1. [S. l.: s. n.], 2018. Disponível em: <https://cran.r-project.org/package=yacca>. Acesso em: 17 jan. 2021.

CARGNELUTTI FILHO, A. *et al.* Análise de repetibilidade de caracteres forrageiros de genótipos de *Panicum maximum*, avaliados com e sem restrição solar. **Ciência Rural**, [s. l.], v. 34, n. 3, p. 723–729, 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782004000300011&lng=pt&tlng=pt. Acesso em: 25 nov. 2021.

CARGNELUTTI FILHO, A.; TOEBE, M. Reference sample size for multiple regression in corn. **Pesquisa Agropecuária Brasileira**, [s. l.], v. 55, n. e01400, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2020000102100&tlng=en. Acesso em: 25 nov. 2021.

CARVALHO, S. P. de; CRUZ, C. D.; CARVALHO, C. G. P. de. Estimating gain by use of a classic selection index under multicollinearity in wheat (*Triticum aestivum*). **Genetics and Molecular Biology**, [s. l.], v. 22, n. 1, p. 109–113, 1999. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47571999000100021&lng=en&tlng=en. Acesso em: 20 dez. 2021.

CAVALCANTE, M. *et al.* Coeficiente de repetibilidade e parâmetros genéticos em capim-elefante. **Pesquisa Agropecuária Brasileira**, [s. l.], v. 47, n. 4, p. 569–575, 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2012000400013&lng=pt&tlng=pt. Acesso em: 18 set. 2019.

CHAVES, G. G. *et al.* Tamanho de parcela e número de repetições para avaliação de caracteres vegetativos em centeio. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, [s. l.], v. 13, n. 3, p. 1–9, 2018. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782016000100044&lng=pt&nrm=iso&tlng=en. Acesso em: 17 dez. 2018.

COURVOISIER, D. S. *et al.* Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. **Journal of Clinical Epidemiology**, [s. l.], v. 64, n. 9, p. 993–1000, 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0895435610004245>. Acesso em: 2 fev. 2022.

CRAWFORD, L. E.; WILLIAMS, M. M.; WORTMAN, S. E. An early-killed rye (*Secale cereale*) cover crop has potential for weed management in edamame (*Glycine max*). **Weed Science**, [s. l.], v. 66, n. 4, p. 502–507, 2018. Disponível em: https://www.cambridge.org/core/product/identifider/S004317451800005X/type/journal_article. Acesso em: 25 nov. 2021.

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

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

DIEL, M. I. *et al.* Repeatability coefficients and number of measurements for evaluating traits in strawberry. **Acta Scientiarum. Agronomy**, [s. l.], v. 42, n. e43357, p. 1–9, 2020. Disponível em: <http://www.periodicos.uem.br/ojs/index.php/ActaSciAgron/article/view/43357>. Acesso em: 16 jan. 2021.

DORMANN, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, [s. l.], v. 36, n. 1, p. 27–46, 2013. Disponível em: <http://doi.wiley.com/10.1111/j.1600-0587.2012.07348.x>. Acesso em: 1 maio 2017.

DUARTE, A. B.; FERREIRA, D. de O.; SILVA, F. L. da. Repeatability and the optimal number of measurements for screening of soybean cultivars under water deficit. **REVISTA CIÊNCIA AGRONÔMICA**, [s. l.], v. 53, n. 2, p. 1–13, 2022. Disponível em: <http://ccarevista.ufc.br/seer/index.php/ccarevista/issue/view/81>. Acesso em: 18 maio 2022.

DUNHAM, R. B.; KRAVETZ, D. J. Canonical correlation analysis in a predictive system. **The Journal of Experimental Education**, [s. l.], v. 43, n. 4, p. 35–42, 1975. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/00220973.1975.10806349>. Acesso em: 2 fev. 2022.

EMBRAPA. **Centeio: BRS Progresso**. Concessão: 2013.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Centeio: BR 1**. Concessão: 1999.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Centeio: BRS Serrano**. Concessão: 2005.

FARRAR, D. E.; GLAUBER, R. R. Multicollinearity in regression analysis: The problem revisited. **The Review of Economics and Statistics**, [s. l.], v. 49, n. 1, p. 92–107, 1967. Disponível em: <http://www.jstor.org/stable/1937887>. Acesso em: 30 abr. 2017.

FERNANDES, F. D. *et al.* Repeatability, number of harvests, and phenotypic stability of dry matter yield and quality traits of Panicum maximum Jacq. **Acta Scientiarum. Animal Sciences**, [s. l.], v. 39, n. 2, p. 149–155, 2017. Disponível em: <http://periodicos.uem.br/ojs/index.php/ActaSciAnimSci/article/view/32915>. Acesso em: 27 nov. 2021.

FERRAZZA, J. M. *et al.* Dinâmica de produção de forragem de gramíneas anuais de inverno em diferentes épocas de semeadura. **Ciência Rural**, Fortaleza, Ceará, v. 43, n. 7, p. 1174–1181, 2013. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782013000700006&lng=pt&tlng=pt. Acesso em: 9 jul. 2017.

FIGUEIREDO FILHO, D. *et al.* O que fazer e o que não fazer com a regressão: Pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (MQO). **Revista Política Hoje**, [s. l.], v. 20, n. 1, p. 44–99, 2011. Disponível em: <https://periodicos.ufpe.br/revistas/politica hoje/article/view/3808>. Acesso em: 12 jan. 2019.

GONZALEZ-ESTRADA, E.; VILLASENOR-ALVA, J. A. **MvShapiroTest: Generalized Shapiro-Wilk test for multivariate normality**. Versão 1.0. [S. l.: s. n.], 2013. Disponível em: <https://cran.r-project.org/package=mvShapiroTest>. Acesso em: 12 fev. 2022.

GOODHUE, D. L.; LEWIS, W.; THOMPSON, R. Multicollinearity and measurement error statistical blind spot: correcting for excessive false positives in regression and PLS. **MIS Quarterly**, [s. l.], v. 41, n. 3, p. 667–684, 2017. Disponível em: <https://misq.org/a-multicollinearity-and-measurement-error-statistical-blind-spot-correcting-for-excessive-false-positives-in-regression-and-pls.html>. Acesso em: 6 jan. 2019.

GREEN, S. B. How many subjects does it take to do a regression analysis?. **Multivariate Behavioral Research**, [s. l.], v. 26, n. 3, p. 499–510, 1991. Disponível em: http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr2603_7. Acesso em: 28 jan. 2022.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: AMGH Editora Ltda, 2011.

HAIR, J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre, Brasil: Bookman, 2009-. ISSN 18733476.

HANISCH, A. L. *et al.* Consórcios de gramíneas anuais de inverno com e sem fertilização. **Revista Agropecuária Catarinense**, [s. l.], v. 25, n. 3, p. 51–53, 2012. Disponível em: <https://publicacoes.epagri.sc.gov.br/RAC/article/view/663>. Acesso em: 25 nov. 2021.

HARPER, D. G. C. Some comments on the repeatability of measurements. **Ringling & Migration**, [s. l.], v. 15, n. 2, p. 84–90, 1994. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/03078698.1994.9674078>. Acesso em: 26 dez. 2021.

HOTELLING, H. Relations between two sets of variates. **Biometrika**, [s. l.], v. 28, n. 3/4, p. 321–337, 1936. Disponível em: <https://www.jstor.org/stable/2333955?origin=crossref>. Acesso em: 9 maio 2017.

HOTELLING, H. The most predictable criterion. **Journal of Educational Psychology**, [s. l.], v. 26, n. 2, p. 139–142, 1935. Disponível em: <https://psycnet.apa.org/record/1935-03006-001>. Acesso em: 14 set. 2021.

IBGE. **Sistema de Recuperação Automática - SIDRA**. [S. l.], 2021. Disponível em: <https://sidra.ibge.gov.br/Tabela/1612>. Acesso em: 14 nov. 2021.

INSTITUTO AGRONÔMICO DO PARANÁ. **Centeio: IP 89**. Concessão: 2000.

LAMBERT, Z. V.; DURAND, R. M. Some precautions in using canonical analysis. **Journal of Marketing Research**, [s. l.], v. 12, n. 4, p. 468–475, 1975. Disponível em: <http://journals.sagepub.com/doi/10.1177/002224377501200411>. Acesso em: 7 jul. 2020.

LAVERY, M. R. *et al.* Number of predictors and multicollinearity: What are their effects on error and bias in regression?. **Communications in Statistics - Simulation and Computation**, [s. l.], v. 48, n. 1, p. 27–38, 2019. Disponível em: <https://doi.org/10.1080/03610918.2017.1371750>. Acesso em: 27 jan. 2022.

LEHMEN, R. I. *et al.* Rendimento, valor nutritivo e características fermentativas de

silagens de cereais de inverno. **Ciência Rural**, [s. l.], v. 44, n. 7, p. 1180–1185, 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782014000701180&lng=pt&tlng=pt. Acesso em: 3 fev. 2019.

LIRA JÚNIOR, J. S. *et al.* Repetibilidade da produção, número e peso de fruto em cirigueleira (*Spondias purpurea* L.). **Revista Brasileira de Fruticultura**, [s. l.], v. 36, n. 1, p. 214–220, 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-29452014000100025&lng=pt&tlng=pt. Acesso em: 28 dez. 2021.

LIRA JÚNIOR, J. S.; FERNANDES BEZERRA, J. E.; LEDERMAN, I. E. Repetibilidade da produção, número e peso de frutos de seleções de pitanga roxa. **Acta Agronômica**, [s. l.], v. 59, n. 1, p. 103–110, 2010. Disponível em: <http://www.redalyc.org/articulo.oa?id=169916223011>. Acesso em: 26 nov. 2021.

MATSUO, É. *et al.* Análise da repetibilidade em alguns descritores morfológicos para soja. **Ciência Rural**, [s. l.], v. 42, n. 2, p. 189–196, 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782012000200001&lng=pt&tlng=pt. Acesso em: 18 maio 2022.

MEIRA, D. *et al.* Multivariate analysis revealed genetic divergence and promising traits for indirect selection in black oat. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, [s. l.], v. 14, n. 4, p. 1–7, 2019. Disponível em: <http://www.agraria.pro.br/ojs32/index.php/RBCA/article/view/v14i4a6514>. Acesso em: 8 maio 2021.

MIEDANER, T. *et al.* Genetic architecture of complex agronomic traits examined in two testcross populations of rye (*Secale cereale* L.). **BMC Genomics**, [s. l.], v. 13, n. 1, p. 1–13, 2012. Disponível em: <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-706>. Acesso em: 12 dez. 2021.

MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Registro Nacional de Cultivares**. [s. l.], 2021. Disponível em: <https://sistemas.agricultura.gov.br/snpc/cultivarweb/>. Acesso em: 14 dez. 2021.

MONTGOMERY, D. C. *et al.* **Introduction to linear regression analysis**. 5. ed. Hoboken, NJ, USA, New Jersey: John Wiley & Sons, Inc., 2012. (821).

MORI, C.; DO NASCIMENTO JUNIOR, A.; DE MIRANDA, M. Z. **Aspectos econômicos e conjunturais da cultura do centeio**: Documentos 142. Passo Fundo, Brasil: Embrapa Trigo, 2013. Disponível em: http://www.cnpt.embrapa.br/biblio/do/p_do142.htm. Acesso em: 14 dez. 2018.

MORRISON, L. A. Cereals: Domestication of the cereal grains. *In*: **ENCYCLOPEDIA OF FOOD GRAINS**. 2. ed. [S. l.]: Elsevier Academic Press, 2016. p. 86–98. *E-book*. Disponível em: <http://dx.doi.org/10.1016/B978-0-12-394437-5.00007-3>. Acesso em: 28 abr. 2021.

NEU, I. M. M. *et al.* Tamanho de amostra para avaliação do grau de multicolinearidade em caracteres produtivos de centeio. **Ciência e Natura**, [s. l.], v. 43, n. e38, p. 1–22, 2021. Disponível em: <https://periodicos.ufsm.br/cienciaenatura/article/view/41667>. Acesso em: 25 nov. 2021.

NEU, I. M. M. **Tamanho de amostra para avaliar a multicolinearidade em caracteres de cultivares de centeio**. 2019. 77 f. - Dissertação (Mestrado) - Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria, Santa Maria, RS, 2019.

OLIVOTO, T. *et al.* Multicollinearity in path analysis: A simple method to reduce its effects. **Agronomy Journal**, [s. l.], v. 109, n. 1, p. 131–142, 2017. Disponível em: <https://dl.sciencesocieties.org/publications/aj/abstracts/109/1/131>. Acesso em: 6 dez. 2018.

OLIVOTO, T. *et al.* Optimal sample size and data arrangement method in estimating correlation matrices with lesser collinearity: A statistical focus in maize breeding. **African Journal of Agricultural Research**, [s. l.], v. 12, n. 2, p. 93–103, 2017. Disponível em: <http://academicjournals.org/journal/AJAR/article-abstract/5DB554B62359>. Acesso em: 9 maio 2017.

PAGLIOSA, E. S. *et al.* Repeatability of pre-harvest sprouting in wheat. **American Journal of Plant Sciences**, [s. l.], v. 05, n. 11, p. 1607–1613, 2014. Disponível em: <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ajps.2014.511174>. Acesso em: 10 mar. 2018.

PARIONA, A. **The top 10 rye producing countries of the world**. [S. l.], 2017. Disponível em: <https://www.worldatlas.com/articles/the-top-10-rye-producing-countries-of-the-world.html>. Acesso em: 3 jan. 2019.

PAULINO, V. T.; DE CARVALHO, D. D. Pastagens de inverno. **Revista Científica Eletrônica De Agronomia**, [s. l.], v. 3, n. 5, p. 1–6, 2004. Disponível em: http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/aJw7z9ZzHjfTqFf_2013-4-26-12-15-21.pdf. Acesso em: 10 mar. 2018.

PEDUZZI, P. *et al.* A simulation study of the number of events per variable in logistic regression analysis. **Journal of Clinical Epidemiology**, [s. l.], v. 49, n. 12, p. 1373–1379, 1996. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0895435696002363>. Acesso em: 2 fev. 2022.

PEIXOTO, A. P. B.; FARIA, G. A.; DE MORAIS, A. R. Modelos de regressão com platô na estimativa do tamanho de parcelas em experimento de conservação in vitro de maracujazeiro. **Ciência Rural**, [s. l.], v. 41, n. 11, p. 1907–1913, 2011. Disponível em: <http://www.scielo.br/pdf/cr/v41n11/a16711cr4630.pdf>. Acesso em: 30 abr. 2019.

PENG, J. L. *et al.* A yield estimation model of forage rye based on climate data by locations in South Korea using general linear model. **Journal of The Korean Society of Grassland and Forage Science**, [s. l.], v. 36, n. 3, p. 205–214, 2016. Disponível em: <http://jksgfs.or.kr/journal/article.php?code=44770>. Acesso em: 3 out. 2019.

PEREIRA, E. M. *et al.* Canonical correlations between agronomic traits and seed physiological quality in segregating soybean populations. **Genetics and Molecular Research**, [s. l.], v. 16, n. 2, p. 1–11, 2017. Disponível em: <http://www.funpecrp.com.br/gmr/year2017/vol16-2/pdf/gmr-16-02-gmr.16029547.pdf>. Acesso em: 13 jun. 2019.

R TEAM CORE. **R: A language and environment for statistical computing**. Versão 4.1.1. Vienna, Áustria: R Foundation for Statistical Computing, 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 10 ago. 2021.

RILEY, R. D. *et al.* Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. **Statistics in Medicine**, [s. l.], v. 38, n. 7, p. 1262–1275, 2019. Disponível em: <http://doi.wiley.com/10.1002/sim.7993>. Acesso em: 16 jan. 2021.

RODRÍGUEZ, R. A. *et al.* Tamanho ótimo de parcela y número de repeticiones para evaluar el rendimiento de boniato con mulch y suelo descubierto. **Agrociencia Uruguay**, [s. l.], v. 22, n. 1, p. 90–97, 2018. Disponível em: <http://www.fagro.edu.uy/agrociencia/index.php/directorio/article/view/1342>. Acesso em: 30 abr. 2019.

SANTOS, H. G. dos *et al.* **Sistema brasileiro de classificação de solos**. 5. ed. Brasília, DF: Embrapa: Embrapa Solos, 2018.

SARI, B. G. *et al.* Interference of sample size on multicollinearity diagnosis in path analysis. **Pesquisa Agropecuária Brasileira**, [s. l.], v. 53, n. 6, p. 769–773, 2018. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2018000600769&lng=en&tlng=en. Acesso em: 6 dez. 2018.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil sciences**. [S. l.]: CRC Press, 2001.

SCHITTENHELM, S.; KRAFT, M.; WITTICH, K.-P. Performance of winter cereals grown on field-stored soil moisture only. **European Journal of Agronomy**, [s. l.], v. 52, p. 247–258, 2014. Disponível em: <http://dx.doi.org/10.1016/j.eja.2013.08.010>. Acesso em: 7 set. 2019.

SCHLEGEL, R. H. J. Chapter 3: Physiology. *In*: RYE: GENETICS, BREEDING AND CULTIVATION. Boca Raton, USA: CRC Press, 2013. p. 51–84.

SCHMILDT, E. R. *et al.* Optimum plot size and number of replications in papaya field experiment. **Revista Brasileira de Fruticultura**, [s. l.], v. 38, n. 2, 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-29452016000200603&lng=en&tlng=en. Acesso em: 30 abr. 2019.

SHIEH, G. Power and sample size calculations for contrast analysis in ANCOVA. **Multivariate Behavioral Research**, [s. l.], v. 52, n. 1, p. 1–11, 2017. Disponível em: <http://dx.doi.org/10.1080/00273171.2016.1219841>. Acesso em: 15 maio 2019.

SILVA, L. F. de O. *et al.* Tamanho ótimo de parcela para experimentos com rabanetes. **Revista Ceres**, [s. l.], v. 59, n. 5, p. 624–629, 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-737X2012000500007&lng=pt&nrm=iso&tlng=en. Acesso em: 5 abr. 2014.

SOUZA, Y. P. de *et al.* Repeatability and minimum number of evaluations for morpho-agronomic characters of elephant-grass for energy purposes. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, [s. l.], v. 12, n. 3, p. 391–397, 2017. Disponível em: http://www.agraria.pro.br/ojs-2.4.6/index.php?journal=agraria&page=article&op=view&path%5B%5D=agraria_v12i3a5456. Acesso em: 18 set. 2019.

STEWART, D.; LOVE, W. A general canonical correlation index. **Psychological Bulletin**, [s. l.], v. 70, n. 3, Pt.1, p. 160–163, 1968. Disponível em:

<http://doi.apa.org/getdoi.cfm?doi=10.1037/h0026143>. Acesso em: 4 fev. 2022.

TOEBE, M. *et al.* Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*. **Bragantia**, [s. l.], v. 76, n. 1, p. 45–53, 2017. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0006-87052017000100045&lng=pt&tlng=pt. Acesso em: 6 dez. 2018.

TOEBE, M. *et al.* Sample size for estimation of direct effects in path analysis of corn. **Genetics and Molecular Research**, [s. l.], v. 16, n. 2, p. 1–23, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28407183>. Acesso em: 13 jan. 2019.

TOEBE, M. *et al.* Sample size in the estimation of correlation coefficients for corn hybrids in crops and accuracy levels. **Bragantia**, [s. l.], v. 74, n. 1, p. 16–24, 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0006-87052015000100016&lng=pt&tlng=pt. Acesso em: 28 jan. 2019.

TOEBE, M.; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, [s. l.], v. 48, n. 5, p. 466–477, 2013. Disponível em: <http://www.scielo.br/pdf/pab/v48n5/02.pdf>. Acesso em: 25 mar. 2017.

TORRES, F. E. *et al.* Minimum number of measurements for accurate evaluation of qualitative traits in *Urochloa brizantha*. **Journal of Agronomy**, [s. l.], v. 14, n. 3, p. 180–184, 2015. Disponível em: <http://dx.doi.org/10.3923/ja.2015.180.184>. Acesso em: 18 set. 2019.

VAN SMEDEN, M. *et al.* No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. **BMC Medical Research Methodology**, [s. l.], v. 16, n. 1, p. 1–13, 2016. Disponível em: <http://dx.doi.org/10.1186/s12874-016-0267-3>. Acesso em: 2 fev. 2022.

VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. New York: Springer, 2002. Disponível em: <https://www.stats.ox.ac.uk/pub/MASS4/>.

WICKHAM, H. **Ggplot2: Elegant graphics for data analysis**. New York: Springer-Verlag, 2016. Disponível em: <https://ggplot2.tidyverse.org>. Acesso em: 10 mar. 2014.

WILSON, M. L.; BAKER, J. M.; ALLAN, D. L. Factors affecting successful establishment of aerially seeded winter rye. **Agronomy Journal**, [s. l.], v. 105, n. 6, p. 1868–1877, 2013. Disponível em: <https://www.agronomy.org/publications/aj/abstracts/105/6/1868>. Acesso em: 11 set. 2019.

WYNANTS, L. *et al.* A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. **Journal of Clinical Epidemiology**, [s. l.], v. 68, n. 12, p. 1406–1414, 2015. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2015.02.002>. Acesso em: 2 fev. 2022.

ZIECH, A. R. D. *et al.* Proteção do solo por plantas de cobertura de ciclo hibernar na região Sul do Brasil. **Pesquisa Agropecuária Brasileira**, [s. l.], v. 50, n. 5, p. 374–382, 2015. Disponível em: <http://www.scielo.br/pdf/pab/v50n5/0100-204X-pab-50-05-00374.pdf>. Acesso em: 10 mar. 2018.

APÊNDICE A – ROTINA EM R PARA DETERMINAÇÃO DA RELAÇÃO NÚMERO DE OBSERVAÇÕES E DE VARIÁVEIS EM CORRELAÇÃO CANÔNICA

```
#####  
### Análise: Determinação de relação n:p                                     ###  
### - Análise de correlação canônica                                       ###  
### - Reamostras normal multivariadas                                       ###  
#####  
library(dplyr)  
library(MASS)           # amostras normais multivariadas  
library(mvShapiroTest) # teste normalidade multivariada  
library(yacca)         # análise de correlação canônica  
#####  
# Função auxiliar: detecção multicolinearidade – Fator de Inflação da Variância  
#-----  
marquardt_vif <- function(correl){  
  # Marquardt (1970). Generalized inverses, ridge regression, biased linear estimation  
  # and nonlinear estimation. Technometrics, v. 12, p. 591-256.  
  xname <- colnames(correl)  
  A <- diag(eigen(correl)$values)  
  A1 <- solve(A)  
  V <- eigen(correl)$vectors  
  Vt <- t(V)  
  R1 <- V%*%A1%*%Vt  
  VIF <- as.vector(diag(R1))  
  names(VIF) <- xname  
  return(VIF)  
}  
#####  
# Load dados e seleção caracteres  
#-----  
load(file = './Dados/001_BD_Centeio2016_DrINeu.RData')  
load(file = './Dados/CCA/B105_caract_ausencia_colin_cultivares.RData')  
RyeDr2019 <- RyeDr2019 %>%  
  select(c('Ensaio', 'Epoca', 'Cultivar',  
    all_of(CMorf), all_of(CProd))) %>%  
  filter(Epoca %in% 1:3)  
caracteres <- c('Morf', 'Prod')  
#####  
# Definição de parâmetros de reamostragem  
#-----  
# número de reamostragens  
R <- 1000  
# relações entre tamanho de amostra e número de variáveis em cada grupo  
np <- seq(from = 1, to = 100, by = 1)  
# tamanhos de amostra planejados  
n_plan <- np*(length(CMorf) + length(CProd))
```

```
#####
# Objetos para saída de resultados das análises
#-----
# nomes de colunas das tabela
id_coef <- c('correlacao', "Chi2aprox", "valorp", 'sharedVarCV')
# Coef. correlação canônica
cancor <-
  array(data = NA,
        dim = c(R, length(n_plan), length(id_coef)),
        dimnames = list(1:R, paste0('n_', n_plan), id_coef))
# Índice de redundância
redundancia <-
  array(data = NA,
        dim = c(R, length(n_plan), length(caracteres)),
        dimnames = list(1:R, paste0('n_', n_plan), caracteres))
#####
# Reamostragem e simulação
#-----
# Cultivar: BRS Progresso (Prog) e Ensaio E1
dados <- RyeDr2019 %>% filter(Cultivar == 'Prog', Epoca == 1)

for (m in length(n_plan):1){
  i <- 1
  #####
  # Seleção dados para cada grupo de variáveis
  #-----
  X <- dados %>% select(all_of(CMorf))
  Y <- dados %>% select(all_of(CProd))

  #####
  # R reamostragens para cada np
  #-----
  while (i <= R){
    # data frame Grupo 1
    df1 <-
      as.data.frame(
        mvrnorm(
          n = n_plan[m],
          mu = apply(X, 2, mean),
          Sigma = cov(X),
          empirical = FALSE))
    # data frame Grupo 2
    df2 <-
      as.data.frame(
        mvrnorm(
          n = n_plan[m],
          mu = apply(Y, 2, mean),
          Sigma = cov(Y),
          empirical = FALSE))

```



```

#####
# Teste de normalidade multivariada
#-----
norm_df1 <- tryCatch(mvShapiro.Test(as.matrix(df1))[[2]], error=function(e)
return(NULL))
norm_df2 <- tryCatch(mvShapiro.Test(as.matrix(df2))[[2]], error=function(e)
return(NULL))

#####
# Verifica se normalidade multivariada foi atendida
#-----
if (!is.null(norm_df1) & !is.null(norm_df2) & is.numeric(norm_df1) &
is.numeric(norm_df2)){
  if(as.numeric(norm_df1) >= 0.05 & as.numeric(norm_df2) >= 0.05){
    critério_norm <- TRUE
  }else{
    critério_norm <- FALSE
  }
}

#####
# Continuação da análise se normalidade multivariada atendida
#-----
if(critério_norm & !any(is.na(cor(df1))) & !any(is.na(cor(df2)))){

#####
# Diagnóstico de multicolinearidade - VIF
#-----
vif_df1 <- tryCatch(marquardt_vif(cor(df1)), error=function(e) return(NULL))
vif_df2 <- tryCatch(marquardt_vif(cor(df2)), error=function(e) return(NULL))

if(!is.null(vif_df1) & !is.null(vif_df2)){

  if(max(vif_df1) <= 10){
    critério_colin_df1 <- TRUE
  }else{
    critério_colin_df1 <- FALSE
  }

  if(max(vif_df2) <= 10){
    critério_colin_df2 <- TRUE
  }else{
    critério_colin_df2 <- FALSE
  }
}
}

```

```

#####
# Análise de CANCOR se critério de multicolinearidade atendida
#-----
if(critério_colin_df1 & critério_colin_df2){

#####
# Análise de correlação canônica
#-----
analise <- tryCatch(
  cca(
    x = df1,
    y = df2,
    test = "Bartlett",
    sign = 0.05
  ),
  error=function(e) return(NULL)
)

#####
# Verifica se analise não nula e preenche objetos de resultados se analise
#-----
if(!is.null(analise)){

#####
# CANCOR e significância
#-----
cancor[i, m, 1] <- analise[['corr']][[1]]
cancor[i, m, 2] <- analise[['sigtest']][1, 3]
cancor[i, m, 3] <- analise[['sigtest']][1, 6]

#####
# Índice de redundância
#-----
redundancia[i, m, 1] <- analise[["xrd"]]
redundancia[i, m, 2] <- analise[["yrd"]]

#####
# Incrementa i se critérios atendidos
#-----
i <- i + 1

}
}
}
}

```