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

Jaqueline Sgarbossa

PRESSUPOSTOS MULTIVARIADOS E EFEITO DOS PARÂMETROS DO MODELO MATEMÁTICO EM ANÁLISES MULTIVARIADAS PARA ENSAIOS COM A AVEIA

Santa Maria, RS
2023

Jaqueline Sgarbossa

PRESSUPOSTOS MULTIVARIADOS E EFEITO DOS PARÂMETROS DO MODELO EM ANÁLISES MULTIVARIADAS PARA ENSAIOS COM A AVEIA

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

Orientador: Prof. Dr. Alessandro Dal'Col Lúcio

Santa Maria, RS
2023

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

Sgarbossa, Jaqueline
Pressupostos multivariados e efeito dos parâmetros do modelo matemático em análises multivariadas para ensaios com a aveia / Jaqueline Sgarbossa.- 2023.
106 p.; 30 cm

Orientador: Aelssandro Dal'Col Lúcio
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Agronomia, RS, 2023

1. Análises multivariadas 2. Pressupostos estatísticos 3. Parâmetros do modelo matemático 4. Relações lineares 5. Análise de trilha I. Dal'Col Lúcio, Aelssandro 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 BibliotecaCentral. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, JAQUELINE SGARBOSSA, 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 consequênciaslegais.

Jaqueline Sgarbossa

PRESSUPOSTOS MULTIVARIADOS E EFEITO DOS PARÂMETROS DO MODELO EM ANÁLISES MULTIVARIADAS PARA ENSAIOS COM A AVEIA

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 17 de fevereiro de 2023:

____
Alessandro Dal'Col Lúcio, Dr. (UFSM)
(Presidente/Orientador)

____
Braulio Otomar Caron, Dr. (UFSM)

____
Diego Nicolau Follmann, Dr. (UFSM)

____
José Antonio Gonzalez da Silva, Dr. (UNIJUI)

____
Fernando Machado Haesbaert, Dr. (UFT)

Santa Maria, RS
2023

A Deus, pelo dom da vida, aos meus pais Jaime (in memoriam) e Nadir, a minha irmã Grasiela e a minha sobrinha Rafaella, por serem meu espelho e inspiração diária, dedico-lhes este trabalho.

AGRADECIMENTOS

Agradeço primeiramente à Deus, pela oportunidade de existência e por sempre me proporcionar saúde, força, calma e perseverança.

Aos meus pais pelo apoio incondicional, confiança depositada e por me proporcionarem educação e humildade.

A minha mãe Nadir, pelo carinho, cuidado e amor incondicional, você é exemplo de força e superação. Obrigado por me incentivar a continuar, mesmo diante das dificuldades.

Ao meu pai Jaime (*in memoriam*), pelo amor, alegria e felicidade que proporcionou a nossa família, durante a sua passagem. Sua falta é sentida todos os dias, você permanece vivo em nossos corações.

A minha irmã Grasiela e a minha sobrinha Rafaella pelo amor, carinho, incentivo e alegria que proporcionaram e proporcionam em minha vida.

Aos demais familiares e amigos, sou extremamente grata pelo apoio, compreensão, incentivo, preocupação, pelas orações e por ter vocês em minha vida.

Ao professor Alessandro Dal'Col Lúcio, pela oportunidade e confiança em mim depositada. Obrigado pelo apoio, carinho, ensinamentos e sobretudo pela compreensão. Sou extremamente grata por conhecer um ser humano como o senhor!

Ao professor José Antonio Gonzalez da Silva por me acolher, incentivar e compreender, pelos ensinamentos e carinho. Sou muito feliz por ter lhe conhecido!

A Coordenação de Aperfeiçoamento de Pessoa de Nível Superior, pelo auxílio financeiro através da concessão de bolsa de Doutorado. A Universidade Federal de Santa Maria e ao Programa de Pós-Graduação em Agronomia, pela possibilidade de realização desse trabalho.

Aos professores, Alessandro Dal'Cól Lúcio, José Antonio Gonzalez da Silva, Braulio Otomar Caron, Fernando Machado Haesbaert e Diego Nicolau Follmann, pela disponibilidade de participar da banca de defesa.

Aos amigos: Maria, Mariane, Lana, Darlei, Jéssica, Guilherme e Claiton pela ajuda incondicional, compreensão, incentivo, por dividirem os momentos de angústia e pela eterna amizade. Aos amigos dos grupos de pesquisa em Sistemas Técnicos de Produção Agrícola (UNIJUI) e em Experimentação Agrícola (UFSM). Obrigado por tudo!

A todas as pessoas que contribuíram para a realização deste trabalho e, sobretudo, torceram por mim.

Vocês possuem minha eterna gratidão!

A esperança é um ingrediente indispensável à vida!

(Augusto Cury)

RESUMO

PRESSUPOSTOS MULTIVARIADOS E EFEITO DOS PARÂMETROS DO MODELO MATEMÁTICO EM ANÁLISES MULTIVARIADAS PARA ENSAIOS COM A AVEIA

AUTORA: Jaqueline Sgarbossa
ORIENTADOR: Alessandro Dal’Col Lúcio

A aveia é um dos principais cereais de inverno cultivados no mundo, utilizada na alimentação humana e animal, cobertura do solo, produção de palhada e rotação de culturas no sistema plantio direto. Com o intuito de potencializar os sistemas de produção de aveia, têm sido empregadas técnicas estatísticas para estudar as relações lineares entre caracteres, a fim de identificar caracteres que direta ou indiretamente favoreçam a seleção de genótipos superiores, entre estas técnicas destacam-se a correlação linear e análise de trilha. Ao proceder análises multivariadas como a análise de trilha, alguns pressupostos estatísticos devem ser atendidos, a fim de evitar a obtenção de resultados viesados. Além disso, ao trabalhar-se com esta técnica, os parâmetros do modelo matemático referentes ao delineamento experimental e tratamentos são desconsiderados, utilizando-se observações médias, sem estratificar os possíveis efeitos. Sendo assim, este estudo foi desenvolvido com o intuito de analisar as implicações da remoção dos parâmetros do modelo matemático sob os resultados de análises de correlação de Pearson e análise de trilha, em ensaios com a cultura da aveia branca, cultivada em diferentes anos e estratificando cenários agrícolas (com e sem o uso de fungicida). Os experimentos foram conduzidos no período de 2015 a 2019, no município de Augusto Pestana, Rio Grande do Sul, Brasil. O delineamento experimental utilizado foi de blocos completos ao acaso, sendo os tratamentos caracterizados por cultivares de aveia e aplicações de fungicida, com três repetições. Para cada ano, cenário e grupo de dados foi realizado diagnóstico de multicolinearidade, calculados os coeficientes de correlação de Pearson e realizada análise de trilha. A ocorrência de multicolinearidade gera a obtenção e coeficientes de trilha viesados e sem intepretação biológica, independentemente do ambiente e grupo de dados analisados. A remoção dos parâmetros do modelo matemático altera a capacidade explicativa dos caracteres em relação a variância na produtividade, para todos os ambientes, cenários e tipos de análises de trilha procedidas. Retirar os efeitos dos parâmetros do modelo, resulta em alterações na direção e magnitude (>50%) nos coeficientes de trilha independentemente do ambiente, cenário e tipo de análise de trilha procedida.

Palavras-chave: *Avena sativa*. Análise de trilha. Multicolinearidade. Relações lineares.

ABSTRACT

MULTIVARIATE ASSUMPTIONS AND EFFECT OF MATHEMATICAL MODEL PARAMETERS IN MULTIVARIATE ANALYSIS FOR OATS TESTS

AUTHOR: Jaqueline Sgarbossa
ADVISOR: Alessandro Dal'Col Lúcio

Oat is one of the main winter cereals grown in the world, used in human food and animal feed, ground cover, straw production, and crop rotation in the no-tillage system. In order to enhance the oat production systems, statistical techniques have been used to study the linear relationships between characters, in order to identify characters that directly or indirectly favor the selection of superior genotypes, among these techniques the linear correlation stands out, and path analysis. When performing multivariate analyses such as path analysis, some statistical assumptions must be met to avoid obtaining biased results. Furthermore, when working with this technique, the parameters of the mathematical model referring to the experimental design and treatments are disregarded, using only average observations, without stratifying the possible effects. Therefore, this study was developed with the aim of analyzing the implications of removing the parameters from the mathematical model on the results of Pearson correlation analysis and path analysis, in field trials with the oat crop, cultivated in different years and stratifying agricultural scenarios (with and without the use of fungicide). The experiments were conducted from 2015 to 2019, in the municipality of Augusto Pestana, Rio Grande do Sul, Brazil. The experimental design used was complete randomized blocks, with treatments characterized by oat cultivars and fungicide applications, with three replications. For each year, scenario, and data group, a multicollinearity diagnosis was performed, Pearson's correlation coefficients were calculated, and a path analysis was performed. The occurrence of multicollinearity generates biased path coefficients without biological interpretation, regardless of the environment and data group analyzed. Removing parameters from the mathematical model changes the explanatory capacity of characters in relation to yield variance, for all environments, scenarios, and types of path analysis performed. Removing the effects of model parameters results in changes in direction and magnitude (>50%) in the path coefficients regardless of the environment, scenario, and type of path analysis performed.

Keywords: *Avena sativa*. Linear relationships. Multicollinearity. Path analysis.

LISTA DE FIGURAS

LISTA DE TABELAS

ARTIGO I

ARTIGO II

SUMÁRIO

1 INTRODUÇÃO

1. 1 CONTEXTO

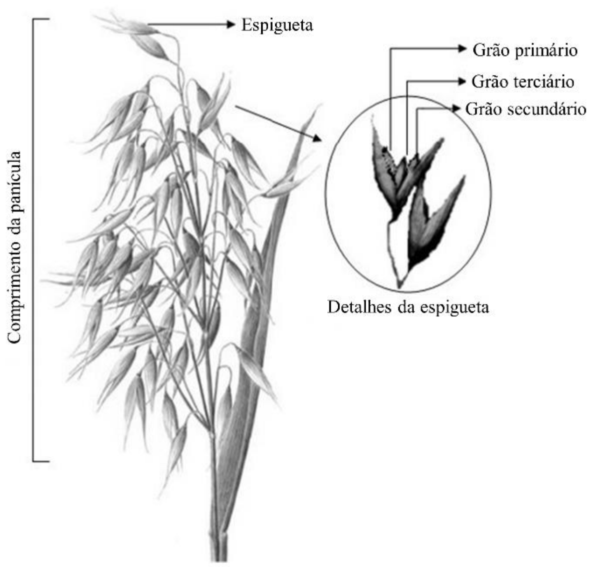
A aveia é um dos principais cerais de inverno cultivados no mundo, podendo ser cultivada com o intuito de produção de grãos para o consumo humano, matéria prima para a indústria de cosméticos e insumos da indústria química, consumo animal na forma de grãos ou forragem, seja para pastejo ou produção de feno e silagem, cobertura do solo, adubação verde, produção de palhada e ou rotação de culturas visando o sistema plantio direto (BORTOLINI; MORAES; CARVALHO, 2005; CASTRO; COSTA; NETO, 2012; FLOSS et al., 2007; FONTANELI et al., 2009; MEINERZ et al., 2011). Atualmente existe uma demanda crescente do cereal para o consumo humano, pois fornece aporte energético e nutricional equilibrado, aminoácidos, ácido graxos, vitaminas e sais minerais essenciais ao organismo humano e, sobretudo, por ser fonte de fibras alimentares (ACHLEITNER et al., 2008; GUTKOSKI et al., 2009).

No mundo são cultivados cerca de 9.772.003 hectares, com produção de 25.181.805 t de grãos e produtividade média de 2.576 kg ha^{-1}. O continente Europeu é responsável por 59,8% do total produzido, seguido pela América (30,4%), Oceania (4,6%), Asia (4,5%) e África (0,7%). O Canadá se destaca como o maior produtor mundial de aveia, sendo responsável por 18,17% do total produzido, seguindo por Rússia (16,41%), Polônia (6,46%), Espanha (5,47%), Finlândia (4,81%), Austrália (4,54%), Reino Unido (4,09%), Estados Unidos (3,77%), Brasil (3,57%) e Suécia (3,21%) (FAO, 2022). No Brasil a aveia se destaca como a segunda cultura de inverno com maior área cultivada, com aumento de 8% na área plantada entre as safras 2021 e 2022, produção de 1.262,6 t de grãos e produtividade de 2.321 kg ha^{-1}. O Rio Grande do Sul, se destaca como maior produtor nacional de aveia, sendo responsável por 71,20% da área destinada ao cultivo de aveia (387.600 ha), produção de 937.200 t de grãos e produtividade de 2.418 kg ha^{-1} (CONAB, 2022)

Pesquisas têm sido desenvolvidas a fim de aprimorar os sistemas de produção das aveias, seja por meio de programas de melhoramento genéticos direcionados ao desenvolvimento de novas cultivares ou ao aprimoramento dos manejos e tratos culturais (AZEVEDO et al., 2022; KLEIN et al., 2019; MANTAI et al., 2021a, 2021b; MEIRA et al., 2019a, 2019b, 2019c; MOURA et al., 2021; REGINATTO et al., 2021a, 2021b). Para obtenção de genótipos superiores é fundamental proceder seleção eficiente, que muitas vezes pode ser trabalhosa e demorada, quando realizada especificamente sob o caráter de interesse. Essa dificuldade pode ser superada pela seleção de materiais com base em seus componentes de rendimento e outros caracteres adaptativos, que indiretamente favoreçam o caráter de interesse principal (FALCONER; MACKAY, 1997).

Entre os componentes de rendimento que podem ser utilizados para a seleção indireta de materiais com maior potencial produtivo, podemos destacar os componentes de panícula: comprimento da panícula, massa da panícula, número de espiguetas da panícula, número de grãos da panícula, massa de grãos da panícula e índice de colheita (Figura 1). Além disso, nos programas de melhoramento genético, compreender a magnitude e a direção da associação entre caracteres é fundamental para elaborar estratégias apropriadas de seleção, que viabilizem a obtenção de genótipos superiores. Entre as técnicas mais empregadas para estimar estas associações, destaca-se a correlação linear de Pearson, que indica a força de associação linear entre um par de caracteres (PEARSON, 1920).

Figura 1 – Detalhes da panícula de aveia branca, base de análise para decomposição dos componentes da inflorescência.



Fonte: (MANTAI et al., 2016).

No entanto, quando busca-se compreender a relação de mais de dois caracteres, o coeficiente de correlação de Pearson não apresenta a verdadeira direção e magnitude das inter-relações, resultando na impossibilidade de identificar se as associações são de causa ou efeito (ALIYU; AHMED; MAGAJI, 2000). Assim, a análise de trilha (AT) é utilizada quando há uma característica dependente (de interesse) e inter-relações entre os caracteres explicativos. Esse método permite o particionamento dos coeficientes de correlação linear em efeitos diretos e indiretos de inúmeros caracteres considerados explicativos em relação a um único caractere dependente (OLIVOTO et al., 2017a). No melhoramento genético de plantas esta técnica tem se mostrado muito útil por revelar se associações são oriundas da influência direta do caráter independente sobre a variável dependente ou ocorrem em virtude de influências que o caráter independente exerce via outro (s) caráter (s), característica que fornece informações que auxiliam na seleção indireta (BELLO et al., 2010; NARDINO et al., 2016).

Embora a AT apresente magnitude e o sentido de inter-relações entre caracteres explicativos para um carácter dependente, baseia-se essencialmente nos princípios da regressão múltipla. Assim, quando dois ou mais possíveis caracteres explicativos são altamente correlacionados, é difícil estimar individualmente as relações de cada caractere explicativo, em virtude da associação entre eles e por contribuir coletivamente para explicar as relações lineares. Tal aspecto é definido como multicolinearidade. Quando esta condição está presente em níveis moderados a severos, a variância associada e estimadores dos coeficientes de trilha atingem valores elevados, tornando as estimativas pouco confiáveis e inconsistentes com a expectativa biológica (CRUZ; REGAZZI; CARNEIRO, 2012).

Adicionalmente, pesquisas tem evidenciado que o desempenho produtivo da cultura da aveia é influenciado pelas características genéticas dos materiais utilizados e do ambiente de cultivo (BENIN et al., 2003a; HOLLAND et al., 2000). Outro fator condicionante a produção da aveia é a ocorrência de doenças foliares, como a ferrugem da folha (MARTINELLI, 2003), que exerce influência no desempenho quantitativo e qualitativo dos genótipos e, pode resultar em reduções superiores a 50% na produtividade de grãos, principalmente em condições de ambiente desfavorável à cultura (BENIN et al., 2003a), aumentando a magnitude da interação entre os materiais genéticos e o ambiente (BENIN et al., 2005). O controle desta doença demanda aplicações frequentes de fungicidas, característica que onera consideravelmente os custos de produção (MARTINELLI, 2003). A aplicação de fungicida afeta os parâmetros de adaptabilidade, responsividade e estabilidade dos materiais genéticos, sobretudo para materiais suscetíveis, sugerindo que para fins de melhoramento genético e recomendações assertivas o desempenho produtivo deve ser estudado considerando as características de ambiente com e sem aplicação de fungicida de forma individualizada (BENIN et al., 2005; LORENCETTI et al., 2004).

1.2 JUSTIFICATIVA

Para que as respostas obtidas por meio da análise de trilha sejam válidas é necessário que as pressuposições estatísticas sejam atendidas como: a normalidade dos resíduos, a ausência de multicolinearidade severa e a homocedasticidade das variâncias (HAIR et al., 2009). No entanto, em partes dos estudos já realizados, estas informações não são encontradas, isto é, os pesquisadores não divulgam se os pressupostos foram testados, se utilizaram de técnicas para contornar a violação dos pressupostos estatísticos ou simplesmente desprezam os pressupostos estatísticos. Toebe & Cargnelutti Filho (2013) ao procederem a comparação dos resultados obtidos por análise de trilha na qual os dados não apresentavam normalidade e ao corrigirem a não normalidade por meio da transformação dos dados, verificaram que violar o pressuposto gera viés nos resultados, isto é, as respostas verificadas não apresentavam aplicação biológica. De maneira semelhante, Couto et al. (2009) trabalhando com a cultura da abobrinha, verificaram a necessidade de transformar os dados para proceder a análise de trilha, pois o não atendimento dos pressupostos mascaram os resultados.

No entanto, em algumas situações a transformação dos dados não é suficiente para superar a multicolinearidade severa entre as variáveis explicativas. Sendo assim, algumas estratégias podem e devem ser empregadas a fim de contornar esse viés, como por exemplo eliminação de variáveis correlacionadas (MONTGOMERY; PECK; VINING, 2012) ou a realização da análise de trilha sob multicolinearidade (análise de trilha em crista) (CRUZ; REGAZZI; CARNEIRO, 2012). Na literatura podem ser encontradas diversas pesquisas que utilizaram de maneira eficientes estas estratégias (CARVALHO et al., 2017; FERRARI et al., 2018; NARDINO et al., 2016; OLIVOTO et al., 2017a, 2017b; TOEBE; CARGNELUTTI FILHO, 2013; TOEBE; CARGNELUTTI FILHO, 2013)

Ao proceder análises estatísticas multivariadas, como a análise de trilha, os parâmetros do modelo matemático relativos ao delineamento experimental e tratamentos não são considerados no emprego das técnicas. Geralmente utiliza-se os valores médios de cada tratamento ou repetição, resultantes de uma soma de efeitos, sem estratificar a influência dos parâmetros, aspectos que podem condicionar a ocorrência de possíveis viés nos resultados. Sendo assim, a realização deste estudo se justifica devido a importância da análise de trilha aos programas de melhoramento genético e a falta de informações sobre o efeito da violação dos pressupostos estatísticos associados as implicações da remoção dos parâmetros do modelo matemático nas relações lineares dos componentes de rendimento da cultura da aveia branca.

1.3 HIPÓTESES

- A violação dos pressupostos estatísticos gera viés nos resultados da análise de trilha.
- A remoção dos parâmetros do modelo matemático gera implicações na magnitude e direção dos coeficientes de trilha.
- Ao estratificar ambientes e cenários de cultivo a remoção dos parâmetros do modelo matemático gera implicações na magnitude e direção das relações.

1.4 OBJETIVOS

Avaliar e caracterizar as implicações da violação dos pressupostos estatísticos e da remoção dos parâmetros do modelo matemático sobre as relações lineares, em ensaios com a cultura da aveia branca.

2 ARTIGO I – MULTIVARIATE ASSUMPTIONS AND EFFECT OF MODEL PARAMETERS IN PATH ANALYSIS IN OAT CROP

Submetido para o periódico: *International Journal of Plant Production*
Situação: Sob Revisão

Multivariate assumptions and effect of model parameters in path analysis in oat crop

Jaqueline Sgarbossa[1*], Alessandro Dal'Col Lúcio[1], José Antônio Gonzalez da Silva[2], Braulio Otomar Caron[3], Maria Inês Diel[4], Tiago Olivoto[5], Claiton Nardini[3], and Odenis Alessi[2]

[1]Department of Plant Science, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil. E-mail: sgarbossajs@yahoo.com; adlucio@ufsm.br. Orcid: https://orcid.org/0000-0001-7541-090X; https://orcid.org/0000-0003-0761-4200
[2]Department of Agronomy, Regional University of Northwestern Rio Grande do Sul State, Ijuí, Rio Grande do Sul, Brazil. E-mail: jose.gonzales@unijui.edu.br; odenisalessi@hotmail.com. Orcid: https://orcid.org/0000-0002-9335-2421; https://orcid.org/0000-0002-3509-6984
[3]Department of Agronomy and Natural Sciences, Federal University of Santa Maria *Campus* Frederico Westphalen, Frederico Westphalen, Rio Grande do Sul, Brazil. E-mail: otomarcaron@yahoo.com.br; claitonnardini@live.com; Orcid: https://orcid.org/0000-0002-6557-3294; https://orcid.org/0000-0001-5791-6720
[4]Federal University of Pampa (UNIPAMPA), Itaqui, Rio Grande do Sul, Brazil. E-mail: mariaines.diel@hotmail.com. Orcid: https://orcid.org/0000-0002-7905-2166
[5]Department of Plant Science, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil. E-mail: tiagoolivoto@gmail.com. Orcid: https://orcid.org/0000-0002-0241-9636

Correspondent author*: Jaqueline Sgarbossa; e-mail: sgarbossajs@yahoo.com;

2.1 ABSTRACT

Path analysis is a multivariate statistical technique that evaluates direct and indirect effects of explanatory variables on the interest variable. When performing path analysis, the effects of mathematical model parameters are disregarded, working only with the average effects of the treatments. Thus, this study aimed to analyze the effect of statistical assumptions and the removal of mathematical model parameters on the path analysis results in oat. The experimental design employed was a two-factor 22×5 randomized complete block design, characterized by twenty-two cultivars and five fungicide applications, with three repetitions. Six explanatory variables, namely panicle length, panicle dry mass, panicle spikelet number, panicle grain number, panicle grain dry mass, and harvest index, and the primary variable yield were measured. Correlation coefficients were calculated, and diagnoses of multicollinearity were performed. Moreover, the path analysis was performed in three ways: traditional, under multicollinearity, and traditional with the elimination of variables. The occurrence of multicollinearity resulted in obtaining path coefficients without biological application. Removing the model's parameters modifies the path coefficients, with average changes of 10.5% and 13.3% in the direction of the associations and 24.7% and 23.0% in the magnitude of the direct and indirect effects, respectively, regardless of the type of path analysis performed.

Keywords: *Avena sativa*, multicollinearity, parameter removal, simple correlation, variable elimination.

2.2 INTRODUCTION

Oats are one of the main cereals grown in the world. In Brazil, this cereal stands out as the second most significant winter crop, showing an increase of 5.4% in the cultivation area between the 2020/2021 and 2021/2022 seasons (Conab, 2021). In this regard, the state of Rio Grande do Sul accounts for about 70% of this production area (Conab, 2019), and a considerable fraction of this area is cultivated with black oat (*Avena strigosa L.*), intended for soil cover and forage production (Cassol et al., 2011; F. V. Moreira et al., 2008). Another fraction is occupied with oats (*Avena sativa L.*), intended for grain production for human food or animal feed, forage, silage, hay, soil cover, and raw material for industry (Caierão et al., 2001; Kaziu et al., 2019).

Numerous studies have been conducted to enhance the oat production system, whether the breeding program is directed to the development of new cultivars of oats or black oats (Caierão et al. 2006, Mantai et al. 2017, Alessi et al. 2018, Meira et al. 2019a, b, c), as well as the improvement of cultivation management (Dornelles et al., 2020; Kraisig et al., 2020; Mantai et al., 2020, 2020; Silva et al., 2020). To obtain superior genotypes, it is essential to proceed with efficient selection, which can often be laborious and time-consuming when performed directly on the trait. This difficulty can be overcome by selecting populations based on their yield components or other adaptive traits that indirectly increase grain yield. However, indirect selection requires a high correlation between the variable under selection and the objective variable (Falconer & Mackay, 1997).

Simple correlation studies allow analyzing the direction and intensity of the linear association between two variables, but do not indicate cause-and-effect relationship between the variables (Cruz et al., 2012; Ferreira, 2009; Sari et al., 2018; Vencovski & Barriga, 1992). Path analysis (PA) is more appropriate for situations where the study objective encompasses more than two traits, because it provides information about the interrelationships of the traits. In this analysis, correlation coefficients are broken down into direct and indirect effects, allowing the influence of one variable over the other to be quantified (Cruz et al., 2012). Thus, statistical techniques that identify the cause-and-effect relationships between variables are fundamental because they allow the identification of characteristics that can be used in the indirect selection of plant cultivars (Cargnelutti Filho et al., 2015).

For the responses obtained through PA to be valid, statistical assumptions must be met, such as multivariate normality of the residuals and the absence of multicollinearity (Hair et al., 2009). Toebe & Cargnelutti Filho (2013), when comparing results obtained through traditional PA, in which the residuals were not normal and when correcting the non-normality through data transformation, found that violating the assumption generates a bias in the results, i.e., the obtained responses had no biological application. Similarly, Couto et al. (2009), working with the zucchini crop, observed the need for data transformation to proceed with PA because not meeting the assumptions masks the results.

Furthermore, it should be considered that when performing multivariate analysis, such as PA, the parameters of the mathematical model are not considered, only the mean values of each treatment or each repetition are worked with, without stratifying the effects of factors, for factorial, block, for randomized block design, and interaction, for factorial. These aspects can lead to the occurrence of bias in the results obtained. Accordingly, the following hypotheses were generated: i) the violation of statistical assumptions generates a bias in the analysis results; ii) removing parameters from the mathematical model modifies the analysis results. To meet these hypotheses, the present study aimed to analyze the effect of statistical assumptions and the removal of parameters from the mathematical model on the results of the path analysis in the oat crop.

2.3 MATERIALS AND METHODS

2.3.1 Study area and experimental design

The research was carried out with results from experiments conducted during the agricultural years 2015, 2016, 2017, 2018, and 2019, in the city of Augusto Pestana, in southern Brazil, under geographical coordinates of 28º 26' 30'' S and 54º 00' 58'' W, with an altitude of 400 meters (m) above sea level. According to Köppen’s climate classification, the region's climate is Cfa, characterized by an average air temperature of 19.1 °C, ranging from 0 to 38 °C, and an accumulated rainfall of 2040 mm (Alvares et al., 2013). The soil of the experimental area is classified as *Latossolo Vermelho distrófico típico* (Oxisol) (Tedesco et al., 1995).

The experimental design employed was a two-factor 22×5 randomized complete block design, characterized by twenty-two oat cultivars: URS Altiva, URS Brava, URS Guará, URS Estampa, URS Corona, URS Torena, URS Charrua, URS Guria, URS Tarimba, URS Taura, URS 21, FAEM 007, FAEM 006, FAEM 5 Chiarasul, FAEM 4 Carlasul, Brisasul, Barbarasul, Fapa Slava, IPR Afrodite, UPFPS Farroupilha, UPFA Ouro, and UPFA Gaudéria, and five fungicide applications: 0 (no fungicide application), 1 (application performed at 60 days after emergence [DAE]), 2 (applications performed at 60 DAE and 75 DAE), 3 (applications performed at 60, 75 and 90 DAE), and 4 (applications performed at 60, 75, 90 and 105 DAE), with three repetitions.

Oat sowing in all agricultural years was performed between May 15 and June 15, following the technical recommendations for the crop. Harvest was performed from late October to early November in all agricultural years. The production performance was analyzed by collecting the plants from 3 central rows, 5 meters long, selected on the day of harvest, and quantifying the following variables: panicle length (PL - cm); panicle dry mass (PDM - g), by weighing on a precision balance; panicle spikelet number (PSN), by counting; panicle grain number (PGN), by counting; panicle grain dry mass (GDM), by weighing on a precision balance; harvest index (HI), determined by the ratio between panicle grain dry mass and panicle dry mass; and grain yield (yield - kg ha^{-1}), determined by weighing the grains in the plot and later extrapolated to kg ha^{-1}.

2.3.2 Statistical Analysis

In the statistical analyses, five agricultural years were addressed as five environments: environment 1 refers to the year 2015, environment 2 to the year 2016, environment 3 to the year 2017, environment 4 to the year 2018, and environment 5 to the year 2019. Thus, univariate statistical assumptions, normality of residuals, and homoscedasticity of residual variances were initially tested for all measured variables. Subsequently, multivariate statistical assumptions, multivariate normality, and multicollinearity were also tested. For the multivariate normality diagnosis, the multivariate Shapiro-Wilk test was applied (Royston, 1983). Shapiro-Wilk test was applied (p-value ≤ 0.05) to analyze the normality of the residuals (Shapiro & Wilk, 1965) and the homogeneity of variances by Bartlett's test (p-value ≤ 0.05) (Steel et al., 1997). The multicollinearity diagnosis was performed considering the variance inflation factor (VIF) and the condition number (CN).

Pearson correlation analysis was performed, generating the matrix of correlation coefficients. Subsequently, the CN was obtained by the ratio between the largest and smallest eigenvalue of the correlation matrix X'X. CN $\leq$ 100 indicates the occurrence of weak multicollinearity, 100 > CN < 1,000, moderate to severe multicollinearity, and CN $\geq$ 1,000, severe multicollinearity (Montgomery & Peck, 1982). The VIF, on the other hand, was obtained for each variable in the inverse diagonal of the correlation matrix X'X. When a VIF value > 10 is obtained, it indicates the occurrence of severe multicollinearity (Hair et al., 2009). The occurrence of multicollinearity between the explanatory variables was defined by obtaining values of CN $\geq$ 1,000 and VIF >10. When violation of assumptions was verified, data transformation was performed employing the Box-Cox family of transformations (Box & Cox, 1964). After the data transformation, the multivariate normality and multicollinearity diagnosis was performed again to check the transformation's efficiency. For situations in which the data transformation was not efficient in circumventing the violation of assumptions, we opted for the exclusion of variables and/or grouping of highly correlated variables, as recommended by Cruz et al. (2012).

In each environment (before and after data transformation), the path analysis was performed, considering the primary variable grain yield as a function of the explanatory variables (PL, PDM, PSN, PGN, GDM, and HI) (Cruz et al., 2012). The direct and indirect effects of the explanatory variables on yield were estimated by means of three methods of path analysis: traditional, traditional with the elimination of variables, and multicollinearity conditions (ridge path analysis). In all methods, it was considered that each explanatory variable has a direct effect on yield and acts indirectly through its effects on the other explanatory variables.

In the traditional analysis, the direct and indirect effects were estimated, disregarding possible multicollinearity biases. To this end, the variables were standardized, and the model of the path analysis was established: yield = $\hat{\beta}_1$PL+ $\hat{\beta}_2$PDM+ $\hat{\beta}_3$PSN+ $\hat{\beta}_4$PGN $\hat{\beta}_5$GDM+ $\hat{\beta}_6$HI + residual, where $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$, $\hat{\beta}_5$ and $\hat{\beta}_6$ are the estimators of the direct effects of the variables PL, PDM, PSN, PGN, GDM and HI, respectively. Following this, by means of a system of normal equations $X'X\hat{\beta} = X'Y$, the direct and indirect effects of each explanatory variable on yield were obtained, as described by Cruz et al. (2012).

In the traditional path analysis with variable elimination, a high linear association was found between the variables PDM, GDM and HI. In all five environments, the elimination of the PDM variable was efficient in reducing multicollinearity to satisfactory levels. Thus, this explanatory variable was eliminated from the path analysis in each environment and, subsequently, the direct and indirect effects of the five explanatory variables on yield were calculated, following the path analysis model: yield = $\hat{\beta}_1$PL+ $\hat{\beta}_2$PSN+ $\hat{\beta}_3$PGN $\hat{\beta}_4$GDM+ $\hat{\beta}_5$HI + residual, where $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ and $\hat{\beta}_5$ are the estimators of the direct variables PL, PSN, PGN, GDM and HI. As in the traditional path analysis, the normal equation system $X'X\ \hat{\beta} = X'Y$ was used to obtain the indirect effect of each variable on the yield.

In the ridge path analysis, the six explanatory variables were kept (PL, PDM, PSN, PGN, GDM and HI), for estimating the direct and indirect effects on grain yield. However, a constant "k" was added to the diagonal of the correlation matrix X'X, in order to reduce the variance associated with the least-squares estimator of the path analysis. Thus, the system of normal equations $X'X\ \hat{\beta} = X'Y$ became $(X'X + k)\ \hat{\beta} = X'Y$. The addition of values of the constant "k" was tested to choose the smallest value at which the path coefficients stabilized (Cruz et al., 2012).

Additionally, the effects of mathematical model parameters (cultivar, application, and block) were isolated and removed, proceeding again with the traditional path analysis, path analysis with the elimination of variables, and ridge path analysis. These results were compared to the ones observed in the "traditional" path analysis to identify whether the removal of the model factors generates changes in the results of the analysis.

The removal of the effects of the model parameters refers to the uniformity trials (without applying treatments), considering that each observation is composed of the overall mean plus the random effect of the error. The group of data in which the removal of model parameter effects was performed was called predicted, and the group of data with the maintenance of parameter effects was called original.

The mathematical model for two-factor experiments (fixed effect) under randomized block design is presented in equation 1:

$Y_{ijk} = m + a_i + d_j + (ad)_{ij} + b_k + e_{ijk}$ Equation 1

Where: Y_{ijk} is an observation in block k (k = 1, 2, and 3) referring to the treatment level i (22 levels) of factor A (cultivar) with level j (5 levels) of factor D (applications): m is the overall mean of the experiment; ai is the effect of level i (i = 22) of factor A; dj is the effect of level j (j = 5) of factor D; $(ad)_{ij}$ is the effect of the interaction of level i of factor A with level j of factor D; b_k is the random effect of block k; e_{ijk} is the random effect of experimental error.

The estimates of the model parameters m, a_i, d_j, $(ad)_{ij}$, and b_k were obtained based on the method of least squares, according to the equations presented below:

$\widehat{m} = \frac{Y...}{IJK}$ Equation 2

Where: $\widehat{m}$ is the overall mean of the experiment; Y... is the sum of all the observations of the experiment; I is the levels of factor A; J is the levels of factor D; K is the number of blocks.

$\hat{a}_i = \overline{Y}_{i..} - \widehat{m}$ Equation 3

Where: $\hat{a}_i$ is the effect of factor A; $\overline{Y}_{i..}$ is the mean of level i of factor A; $\widehat{m}$ is the overall mean of the experiment.

$\hat{d}_j = \overline{Y}_{.j.} - \widehat{m}$ Equation 4

Where: $\hat{d}_j$ is the effect of factor D; $\overline{Y}_{.j.}$ is the mean of level j of factor D; $\widehat{m}$ is the overall mean of the experiment.

$\hat{b}_k = \overline{Y}_{..k} - \widehat{m}$ Equation 5

Where: $\hat{b}_k$ is the effect of block; $\overline{Y}_{..k}$ is the mean of level k of factor block; $\widehat{m}$ is the overall mean of the experiment.

$(\widehat{ad})_{ij} = \overline{Y}_{ij.} - \widehat{m} - \hat{a}_i - \hat{d}_j$ Equation 6

Where: $(\widehat{ad})_{ij}$ is the effect of the interaction; $\overline{Y}_{ij.}$ is the mean of the combination of level i of factor A and level j of Factor D; $\hat{a}_i$ is the effect of factor A; $\hat{d}_j$ is the mean of factor D.

A 5% probability level of error was adopted in all statistical analyses, and all analyses were performed using Excel software and R software (R Core Team, 2021). The analyses were performed using the packages stats (R Core Team, 2021), car (Fox & Weisberg, 2019), MVN (Korkmaz et al., 2014), pracma (Borchers, 2021), faraway (Faraway, 2016), Hmisc (Harrell, 2021), biotools (A. R. da Silva et al., 2017), rpanel (Bowman et al., 2007), and tkrplot (Tierney, 2021).

2.3.3 Meteorological conditions

The data of air temperature (minimum, average, and maximum) and accumulated rainfall were obtained from the mobile automatic weather station, located approximately 200 meters away from the experimental area. During the oat season in 2015, an average air temperature of 17.4 °C, ranging from -0.16 °C to 33.0 °C, and an accumulated rainfall of 798.3 mm were verified (Fig. 1). For the 2016 growing period, the average air temperature was 16.5 °C, ranging from -0.3 °C to 34.7 °C, with an accumulated rainfall of 711.2 mm. During the 2017 season, the average air temperature was 19.1 °C, ranging from -4.2 °C to 34.3 °C, with an accumulated rainfall of 715.5 mm. For the 2018 growing season, the average air temperature was 16.5 °C, ranging from -1.60 °C to 32.30 °C, with an accumulated rainfall of 687.4 mm. During 2019, the average air temperature was 17.0 °C, ranging from -4.2 °C to 36.1 °C, with an accumulated rainfall of 645.5 mm. Oats require temperatures between 0 °C and 35 °C (Leite et al., 2012) for growth and development to occur. Therefore, during all crop years, the minimum and maximum temperatures exceeded or were very close to the limits established for the cultivation. In the year 2015, a period with no rainfall was observed right after crop fertilization. This aspect associated with the occurrence of high temperatures during the anthesis period, when the development of the reproductive system is particularly sensitive to water stress and high temperatures, may have impacted the performance of the crop.

2.4 RESULTS

2.4.1 Statistical assumptions

The diagnosis of univariate normality and homoscedasticity of the residual variances allowed the identification of the violation of statistical assumptions in all growing environments, indicating the need for data transformation (Table 1). When performing the data transformation, it was observed that this technique was not always efficient in bypassing the violation of assumptions. However, it helped to improve the coefficients. In general, the highest rates of assumption violation were verified in Environment 1 and the lowest rates in Environment 2.

The multivariate normality and multicollinearity diagnosis demonstrated a violation of statistical assumptions for all environments under study (Table 2). Furthermore, the variance inflation factor (VIF) and condition number (CN) statistics allowed identifying the variables that caused noise in the data, which were PDM, GDM and HI. Thus, as an initial strategy to circumvent the tendency of the analyses, the variables that did not show univariate normality were transformed. However, the data transformation was not efficient in circumventing the violation of the assumptions. An existing alternative to circumvent the violation of assumptions is the exclusion of variables that generate tendencies in the analyses. Thus, the variables that showed multicollinearity (PDM, GDM and HI) were individually excluded, and again, the diagnosis of multicollinearity and multivariate normality was performed. The elimination of the PDM variable proved efficient in circumventing multicollinearity (Table 3). However, this alternative was not sufficient to obtain multivariate normality.

In general, the same trend is observed in the results of the multicollinearity and multivariate normality diagnosis between the data groups, but with divergent absolute values. In environment 1, the absolute values of the VIF and CN statistics were higher for the original data pool in contrast to the predicted ones. However, for the other environments, the highest absolute values of all variables were verified for the predicted database, except for HI in environment 2. Similar responses were observed when PDM exclusion was performed, and the highest absolute values of the VIF and CN statistics for all variables were obtained in the predicted data group, except for environment 1 and the variable PSN in environment 3.

2.4.2 Simple Correlation

When analyzing the linear relationship between the yield components of oats, considering the original and predicted database, high rates of significant correlation between the variables, regardless of the cultivation environment, were observed (Table 4). In general, all traits under study showed a correlation with yield, but with r values of low magnitude for PL. The PDM showed high correlations with PGN, GDM and PSN, which is justified because these variables are panicle components. PGN and GDM also showed a good correlation, suggesting that panicles with higher grain numbers result in higher values of grain mass.

When reviewing individually the 105 values of correlation coefficients obtained from the combination of each pair of variables in the 5 environments between data groups (original and predicted), several situations are observed. Generally, the response pattern was maintained in 79% of combinations, with reverse direction in the response in 4.8% of combinations and alteration in 50% of the absolute value of the coefficient, with the maintenance of sign in 16.2% of combinations. Furthermore, it should be noted that the response pattern was maintained in the environments 2, 3, 4, and 5, with minor changes in the magnitude of the coefficients. In environment 1, the response pattern with the inversion of sign was verified in 19.0% of the combinations and change in 50% of the absolute value of the coefficient, with maintenance of the sign in 66.75%, coincidentally this environment had the highest data of the experimental sites.

2.4.3 Path analysis *vs* removal of model parameters

The values of correlation coefficient between PL and yield of the oats in the five environments were of low magnitude (≤0.19), indicating a low association between these variables. The other variables showed correlation coefficients with higher magnitudes, suggesting a possible cause-effect relationship, regardless of the environment (Table 5). It is also important to note that these responses are maintained even after the transformation of the variables.

When analyzing the correlation coefficients between the explanatory variables and the yield, for the group of predicted data, a low linear association between PL and the yield is also observed, as well as specific changes in the direction of the correlation (+ or -) for PSN and in the magnitude of the correlation for PDM, PGN, GDM, and HI, in environment 1. In addition, when performing a new Pearson correlation between the five coefficients obtained for the group of original data and the five coefficients obtained for the group of predicted data, we can observe associations of low magnitude, especially for the variables PGN, HI, PSN, PDM, and GDM.

Considering the existence of a linear relationship between the yield and most variables and that, in some environments, the PL showed a significant correlation with the main variable, it is justifiable to carry out the path analysis to stratify the direct and indirect effects of the explanatory variables on the main variable. Thus, it is necessary to carry out a new diagnosis of multicollinearity, contemplating only the explanatory variables and checking the possibility of bias in the path analyses.

The diagnosis of multicollinearity based on VIF indicated that the variables PL, PSN, and PGN do not have a high correlation with the other explanatory variables (FIV<10), regardless of the environment, the data group (original or predicted), the use of data transformation techniques, or exclusion of variables (Table 6). For the variables PDM, GDM, and HI, the VIF statistic (VIF>10) indicated the existence of multicollinearity, regardless of the data group or data transformation. However, the technique of eliminating the PDM variable circumvented the violation of the assumption, and absence of severe multicollinearity between the variables was verified.

Similar results were observed when performing the diagnosis of multicollinearity using the condition number statistic, which indicated the existence of severe multicollinearity (CN>1,000) among the explanatory variables, regardless of the data group, or the transformation of the original data (Table 7). As observed for the VIF, the elimination of the PDM variable circumvented the problem of multicollinearity of the variables, resulting in CN values < 100.

The original data transformation technique was not efficient in circumventing the multicollinearity between the explanatory variables, and in some situations it inflated the coefficients, indicating the existence of a high correlation between the explanatory variables. Thus, when performing the path analysis, we chose to use the original data, without transforming variables. Therefore, first, the path analysis was performed considering the original and predicted data groups, and later the PDM variable was eliminated in each data group, and a new path analysis was performed.

The variables PL, PDM, PSN, PGN, GDM, and HI showed explanatory power of 54.3%, 30.2%, 42.4%, 43.2% and 35.7% of the variance in oat yield, for environments 1, 2, 3, 4, and 5, respectively, for the original data group (Table 8). For the predicted data group, the variables showed explanatory power of 9.9%, 35.9%, 46.5%, 49.0% and 42.3% of the variance in yield, for environments 1, 2, 3, 4, and 5, respectively. Thus, for environments 2, 3, 4, and 5 removing the effects of the model parameters increased the explanatory capacity of the variables by 15%. In contrast, for environment 1, there was a reduction of 81.8% in the explanatory power of the variance in yield.

In general, the removal of the effects of the model parameters resulted in a change in the direction of the direct effects in 16.7% of the combinations and promoted changes greater than 50% in the magnitude of the direct effects in 26.7%, with maintenance in the pattern of the direct effects in 56.6% of the combinations. Another aspect to be highlighted is that once again most of the changes, whether in direction or in magnitude, were verified in environment 1, coinciding with the highest rates of violation of univariate assumptions.

When analyzing the impacts of removing the model parameters on the indirect effects, we can observe, in general, a change of 18.67% in the direction of the response of the path coefficients, a change of 26.67% in the magnitude (>50% of the absolute value) of the coefficients and maintenance of the response pattern in 54.66% of the situations. In environments 3, and 4 the pattern of responses was maintained above 90%, while in environments 1, 2, and 5 the pattern was maintained in only 3.3%, 33.3%, and 46.7% of the situations, respectively. On the other hand, when observing the direct and indirect effects of the explanatory variables on the yield of oats, values for path coefficients that exceed the unit (1) were obtained, regardless of the data group.

When carrying out a path analysis with the elimination of PDM, it was identified that the other variables showed the ability to explain the variance in the yield of oats in 54.10%, 30.20%, 41.80%, 42.20%, and 34.20% for environments 1, 2, 3, 4, and 5, respectively, for the original data group (Table 9). Considering the group of predicted data, the variables showed explanatory power of 9.9%, 35.9%, 46.1%, 48.2%, and 39.4% of the variance in the yield for environments 1, 2, 3, 4, and 5, respectively. Thus, it can be inferred that removing the effects of the model parameters resulted in an increase of 14.6% in the explanation capacity of the variables, for environments 2, 3, 4, and 5. However, for environment 1, it resulted in a reduction of 81.5% in the ability to explain variance in the yield.

In general, removing the effects of the model parameters resulted in an 8% change in the direction of the direct effects of the explanatory variables on yield, a 50% variation in the magnitude of the direct effects in 24% of situations, and maintenance of 68% in the pattern of path coefficients. Regarding indirect effects, alterations of 12% were observed in the direction of the coefficients, with alterations greater than 50% in the magnitude of the effects in 25% of the combinations and maintenance of the pattern of responses in 63% of the combinations. Another aspect to be highlighted is that most of the changes, whether in direction or in magnitude, were verified in environment 1, coinciding with the highest rates of violation of assumptions.

Another strategy that can be used by researchers to overcome the problems with multicollinearity, it's to carry out ridge path analysis. In the ridge path analysis, the variables PL, PDM, PSN, PGN, GDM, and HI explained 52.8%, 29.2%, 41.0%, 41.0%, and 33.6% of the variance in yield, for environments 1, 2, 3, 4 and 5, respectively, for the original data group (Table 10). For the predicted data group, the variables studied were able to explain 9%, 34.8%, 45.2%, 46.8%, and 38.8% of the variance in yield, for environments 1, 2, 3, 4, and 5, respectively. Thus, removing the effects of the model parameters resulted in an average increase of 14.8% in the ability to explain the variables of variance in yield, for environments 2, 3, 4, and 5. In contrast, there was a reduction of 82.9% in the explanatory power of variance in environment 1.

In general, the removal of the effects of the model parameters resulted in a change in the direction of the direct effects in 6.67% of the combinations and promoted changes greater than 50% in the magnitude of these effects in 23.33%, with maintenance in the response pattern in 70.0% of the combinations. Regarding the indirect effects, the removal of the effects of the model parameters resulted in a change in the direction of the coefficients in 9.33% of the combinations, and promoted changes greater than 50% in the magnitude of the coefficients in 19.33%, with the maintenance of the pattern of indirect effects in 71.34% of the combinations. Another aspect to be highlighted is that, again, most of the changes, whether in direction or in magnitude, were verified in environment 1, coinciding with the highest rates of violation of assumptions.

In order to analyze in a more precise way, the interference of the removal of the model parameters in the magnitude and direction of the coefficients for traditional path analysis, traditional path analysis with the elimination of variables, and path analysis under multicollinearity (ridge), a Pearson's correlation was performed between the values of the direct and indirect effects of each variable, between the data groups (original and predicted). For the variables PL, PGN, and HI, correlations ranging from moderate to strong degree (0.59 to 0.80) were verified, regardless of the type of path analysis performed (Table 11). For the variables GDM and PDM, correlations from strong to high degree (≥0.71 to ≤0.96) were obtained, regardless of the type of path analysis. However, for PSN, there was no significant correlation between the path coefficients considering the original and predicted data groups, regardless of the type of path analysis performed.

2.4.4 Interpretation of path analysis in different data groups

When analyzing the cause-effect relationships for the original data group vs the predicted data group, it was identified for the PL variable in environments 1, 2, 3, and 5 that the direct effects are in the opposite direction to the correlation coefficients, with low magnitude values, regardless of the data group, suggesting that the correlation is caused by indirect effects (Table 8). Only for environment 4, do the correlation coefficient and the direct effects show similar sign and magnitude, that is, in this environment the direct effects explain the true association between the traits. Regarding the indirect effects, there were very low magnitudes in the coefficients, confirming the absence of a cause and an effect of this variable on the yield.

For the PDM variable, in the original data group, positive direct effects of high magnitude were obtained, indicating that the direct effects explain the true association between the variables, regardless of the environment. Similar responses were observed for the predicted data group, except for environments 1 and 2. For environment 1, the correlation coefficient and direct effect values were negligible. For environment 2, the association between variables can be attributed to indirect effects. When analyzing the indirect effects in environment 2, we can observe a negative association of greater magnitude via GDM, PGN, and PSN.

For PSN, correlation coefficients and direct effects of the same sign were observed for environments 1 and 5, regardless of the data group, but with negligible direct effects. In environments 2, 3, and 4, correlation coefficients of low magnitude and negative direct effects of low magnitude were verified, indicating that the true association with yield can be explained by indirect effects. When analyzing the indirect effects, it was identified that they are negligible. For PGN, the direct effects were of the same sign as the correlation coefficients, but with magnitudes lower than 0.26, regardless of the data group or environment, indicating that the true association of this variable with yield may be related to indirect effects. The coefficients referring to indirect effects are also of low magnitude (<0.19), with the highest values observed via PSN, PDM, and GDM.

For GDM, in the original data group, the direct effects were in the opposite direction and of high magnitude compared to the correlation coefficients, indicating that the direct effects must be considered in the analysis. When analyzing the indirect effects, there was a high negative association via PDM, PGN, PSN, and HI. Similar results are observed for the predicted data group, in environments 1, 3, 4, and 5. On the other hand, environment 2 suggests a cause-effect relationship between the variables, via direct effect. For the HI, correlation coefficients and direct effects with similar direction and magnitude are observed, regardless of the data group or environment, suggesting that the true association between the variables is explained by the direct effects. In addition, there was the contribution of GDM to obtain these correlation coefficients.

When interpreting the cause-effect relationships for the original and predicted data groups with variable elimination, it is observed for the PL variable that the direct effects and correlation coefficients were of low magnitude or showed a reverse direction, regardless of the group of variables, data or environment, a condition that suggests that the true association between the variables must be explained by indirect effects (Table 9). In turn, the indirect association effects were of low magnitude, indicating the absence of a cause-effect relationship between PL and yield. For PSN, we can observe an inverse and low magnitude relationship between the correlation coefficients and the direct effects, regardless of the fact that the true association between the variables is explained by the indirect effects, of the data group or environment. When analyzing the indirect effects, in general, the coefficients are of low magnitude, indicating that there is no association between PSN and yield.

For the PGN variable, the direct effects are positive as well as the correlation coefficients, but of low magnitude (<0.27), regardless of the data group and environment, so the indirect effects must be considered to explain the true association between the variables. For the indirect effects, most of the PGN *vs* yield relationship is explained via PSN and PGN, but with path coefficients that do not exceed 0.20, indicating low magnitude. When considering the direct effects and the correlation coefficient for GDM, there were similar positive directions and magnitude, regardless of data group or environment. However, the values of direct effects are 44% and 36.7% lower than the correlation coefficients, for the original and predicted data groups, respectively. Thus, a fraction of the association between the variables can be attributed to indirect effects, via PGN and HI.

Similar responses are observed for the HI, that is, regardless of the environment or data group, the correlation coefficients and direct effects are of the same sign and magnitude, so the direct effects explain a good part of these relationships. However, the indirect effects verified for GDM also contribute to obtaining these relationships, a situation that is justified because the HI is obtained by the ratio between GDM and PDM.

When interpreting the ridge path analysis for the original *vs* predicted data groups, correlation coefficients and direct effects of reverse direction or very low magnitudes were verified for PL, regardless of the data group or environment, suggesting that indirect effects explain the true association between the variables (Table 10). When analyzing the indirect effects, we can verify that coefficients of low magnitudes were obtained, indicating once again a low association between PL and yield. For PDM, correlation coefficients and direct effects of the same sign and magnitudes that can be considered similar were observed, suggesting a direct relationship between the variables. However, when considering the indirect effects, the correlation can also be explained via GDM and PGN, situations that are justified, since both variables are components of the panicle and contribute to obtaining the absolute values of this trait.

In general, for PSN, the direct effects showed values with reverse direction or very low magnitudes, in relation to the correlation coefficients, regardless of the data group or environment. This condition suggests that the true association between the variables is explained by indirect effects. When analyzing the indirect effects, values with very low magnitude were identified (<0.11), suggesting the absence of a cause-effect relationship between these variables. For PGN, we can verify direct effects in the same direction as the correlation coefficients, regardless of the data group or environment. However, only in environments 3, 4, and 5 the magnitudes of the direct effects can be considered somewhat interesting. In these environments, we can also verify indirect effects via PSN, GDM, and PDM.

For GDM, the correlation coefficients and direct effects were in the same direction, but with lower magnitudes, regardless of the data group or environment, suggesting that the association can be explained by indirect effects. For the indirect effects, the path coefficients also did not exceed values of 0.24, but suggest that part of the correlation can be explained via PDM, GDM, and PSN. When analyzing the correlation coefficients and the direct effects for HI, similar directions and values were identified, evidencing that the direct effects explain the association between the variables. However, when analyzing the path coefficients for the indirect effects, a contribution of PGN and PDM was verified to obtain these correlations, a situation that is justified due to the relationship between these variables for the calculation of HI.

2.5 DISCUSSION

2.5.1 Statistical assumptions and simple correlation

The diagnosis of univariate normality and homoscedasticity of the residual variances allowed the identification of the violation of statistical assumptions in all growing environments. The use of the data transformation technique helped to improve the coefficients. However, it was not always efficient in bypassing the violation of assumptions. The improvement in the coefficients obtained for the statistics of the assumption tests is associated with the characterization of the data transformation technique, which is based on adding a lambda value that maximizes the maximum likelihood estimator and minimizes the residual (Kutner et al., 2004).

When analyzing the linear relationships between the yield components, considering the data groups studied, high rates of significant correlations were verified between the variables, regardless of the growth environment. The high number of significant linear combinations is related to the number of observations, a situation that conditions the achievement of significance even in situations with correlation coefficient values of low magnitude (Lúcio et al., 2013).

In general, all traits under study showed a correlation with yield, but with r values of low magnitude for PL, indicating that individually this variable has a lower influence on yield (Sari et al., 2017). The PDM showed high correlations with PGN, GDM and PSN, which is justified because these variables are panicle components. PGN and GDM also showed a good correlation, suggesting that panicles with higher grain numbers result in higher values of grain mass. Kaziu et al. (2019), studying the linear relationships of yield components of oats, found a strong to very strong association of yield with panicle mass, harvest index, and panicle grain number and a weak association of panicle length ($r = 0.31$), similar to the results found in this study. On the other hand, Dumlupinar et al. (2011), analyzing the correlation between the yield components of different oat genotypes, verified that PDM and PGN are highly correlated with each other, as well as the thousand grain mass (TGM) *vs* PDM, PDM *vs* yield and yield *vs* TGM, showing that grain mass plays a determining role for yield.

Benin et al. (2003), analyzing the linear relationship between the yield characteristics of oats, found that panicle mass, the number of panicles per plant, and the average mass of grains have a positive correlation with the yield of the crop. Similar responses were observed by Caierão et al. (2001). The authors found a positive correlation between the number of grains per panicle and grain mass and yield, demonstrating that they can be used in indirect selection for yield. In a study with the black oat crop, Meira et al. (2019a) found a strong positive correlation between PGN and HI and yield, thus demonstrating that increases in yield occur as the number and mass of grains per plant increase.

When reviewing individually the correlation coefficients obtained from the combination of each pair of variables in the 5 environments between the data groups, it was verified that the response pattern was maintained in environments 2, 3, 4, and 5, with minor changes in the magnitude of the coefficients. In contrast, the most significant divergences between the coefficients were obtained in environment 1. These responses can be attributed to the extreme weather conditions in critical periods of the crop, which impacted the alteration of its development and yield.

2.5.2 Path analysis

Pearson's correlation indicated the existence of a linear relationship between the explanatory variables and yield, regardless of the growth environment and data group. When carrying out a new diagnosis of multicollinearity, considering only the group of explanatory variables, a violation of the assumption was observed in all the scenarios studied. The variable elimination technique was an efficient alternative to overcome the problems related to multicollinearity. Meira et al. (2019a) when working with the crop of black oats also verified the occurrence of severe multicollinearity in the correlation matrix between the explanatory traits, and the authors chose to eliminate the variables to circumvent the multicollinearity. In the literature, studies are found that demonstrate the need to eliminate some explanatory variables to properly carry out the path analysis with corn (Cargnelutti Filho 2013, Olivoto et al. 2017, Toebe et al. 2017a, b), tomato (Rodrigues et al., 2010; Sari et al., 2017, 2018), pepper (Carvalho et al., 1999; S. O. Moreira et al., 2013) and jabuticaba (Salla et al., 2015) crops.

In general, the removal of the model parameters promoted changes in the capacity to explain the variance in yield, by the independent variables. Furthermore, changes in the direction and magnitude of path coefficients (direct and indirect) were observed in all environments and types of path analysis performed, especially for environment 1. This is a response resulting from the variation in the meteorological conditions of this environment compared to the others. In addition, when path analysis was performed without using techniques to circumvent the multicollinearity between the explanatory variables, values for path coefficients that exceeded unity (1) were observed, regardless of the data group. This situation is indicative of the existence of bias in the analyses, such as multicollinearity, aspects that can impact the biological application of the results (Olivoto et al. 2017, Toebe et al. 2017b). These observations corroborate what was verified in the multicollinearity diagnosis, which identified the assumption violation.

Considering that to overcome the problems with multicollinearity, it was necessary to eliminate the variable PDM and that studies suggest that this trait has a high correlation with yield and a good prospect of indirect selection via PDM (Caierão et al. 2001, 2006, Mantai et al. 2020a) and that the grain mass corresponds to about 80% to 85% of the panicle mass (Caierão et al., 2001), ridge path analysis was performed. However, for ridge path analysis, the removal of model parameters also resulted in changes in the ability to explain variance in yield, as well as in the magnitude and direction of path coefficients.

In general, when analyzing the cause-effect relationships, considering the types of path analysis performed, the environments and the data groups, there was no cause-effect relationship of the PL and PSN variables with yield. The results verified in this study are similar to those found in the literature, regardless of the data group or type of path analysis performed. Benin et al. (2003) indicate that the variables panicle mass, number of panicles per plant, and average grain mass have the greatest direct and indirect effects on yield, thus being the main characteristics to be considered for the selection of superior genotypes for white oat. Meira et al. (2019a) found no direct effect of panicle length on black oat yield, but a moderate indirect positive effect via the number of grains per panicle. The authors also observed that the number of grains per panicle exerts a direct positive influence of high magnitude on the yield. Moradi et al. (2005) and Bibi et al. (2012) found that the number of grains per panicle has a greater direct effect on oat yield.

2.6 CONCLUSIONS

The occurrence of multicollinearity among the explanatory variables resulted in obtaining path coefficients with magnitudes that exceed the unit and without biological application.

Removing the model's parameters modified the path coefficients, with average changes of 10.5% and 13.3% in the direction of the associations and 24.7% and 23.0% in the magnitude of the direct and indirect effects, respectively, regardless of the type of path analysis performed.

The indirect selection for the white oat yield from the harvest index, considering the grain and panicle dry mass, is an interesting alternative for selecting cultivars with higher production potential.

2.7 COMPETING INTERESTS

There is no conflict of interest among the authors of this paper.

2.8 FUNDING

No funding was obtained for this study.

2.9 ACKNOWLEDGEMENTS

The authors are grateful to the Coordination for Improvement of Higher Education Personnel (Capes-Brazil) for their financial support of the first author (Finance code 001, Process Nº.88887.499817/2020-00). Also, I would like to thank the members of the research groups on Technical Systems of Agricultural Production at the Regional University of Northwestern Rio Grande do Sul State and the research group on Agricultural Experimentation at the Federal University of Santa Maria for help in this project.

2.10 REFERENCES

Alessi, O., Dornelles, E. F., Mamann, Â. T. W. de, Kraisig, A. R., Henrichsen, L., Marolli, A., Pansera, V., & Silva, J. A. G. da. (2018). Aplicação de modelos de regressão e de adaptabilidade e estabilidade na identificação de cultivares de aveia branca com maior resistência genética a doenças foliares. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, *6*(2), 7.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. de M., & Sparovek, G. (2013). Köppen’s climate classification map for Brazil. *Meteorologische Zeitschrift*, *22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Benin, G., Carvalho, F. I. F. de, Oliveira, A. C. de, Marchioro, V. S., Lorencetti, C., Kurek, A. J., Silva, J. A. G., Cargnin, A., & Simoni, D. (2003). Estimativas de correlações e coeficientes de trilha como critérios de seleção para rendimento de grãos em aveia. *Revista Brasileira de Agrociência*, *9*(1), 09–16.

Bibi, A., Shahzad, A., Sadaqat, H., Tahir, M., & Fatima, B. (2012). *Genetic characterization and inheritance studies of oats (Avena sativa L.) for green fodder yield*. *1*, 450–460.

Borchers, H. W. (2021). *pracma: Practital numerical math functions*. CRAR.R-Project. https://cran.r-project.org/package=pracma

Bowman, R., Alexander, G., Crawford, E., & Bowman, A. (2007). rpanel: Simple Interactive Controls for R Functions Using the tcltk Package. *Journal of Statistical Software*, *17*, 1–18. https://doi.org/10.18637/jss.v017.i09

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Caierão, E., Carvalho, F. I. F. de, & Floss, E. L. (2006). Seleção indireta para o incremento do rendimento de grãos em aveia. *Ciência Rural*, *36*(4), 1126–1131. https://doi.org/10.1590/S0103-84782006000400013

Caierão, E., Carvalho, F. I. F. de, Pacheco, M. T., Lonrecetti, C., Marchioro, V. S., & Silva, J. G. (2001). Seleção indireta em aveia para o incremento no rendimento de grãos. *Ciência Rural*, *31*(2), 231–236. https://doi.org/10.1590/S0103-84782001000200007

Cargnelutti Filho, A., Toebe, M., Alves, B. M., Burin, C., Santos, G. O. dos, Facco, G., & Neu, I. M. M. (2015). Relações lineares entre caracteres de aveia preta. *Ciência Rural*, *45*(6), 985–992. https://doi.org/10.1590/0103-8478cr20140500

Carvalho, C. G. P. de, Oliveira, V. R., Cruz, C. D., & Casali, V. W. (1999). Análise de trilha sob multicolinearidade em pimentão. *Pesquisa Agropecuaria Brasileira*, *34*(4), 603–613.

Cassol, L. C., Piva, J. T., Soares, A. B., & Assmann, A. L. (2011). Desempenho de forrageiras hibernais sob distintos níveis de luminosidade. *Revista Ceres*, *58*, 438–443. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-737X2011000400006&nrm=iso

Conab. (2019). *Acompanhamento da safra brasileira de grãos - Terceiro levantamento* (No. 3; 3rd ed.).

Conab. (2021). *Acompanhamento da safra brasileira de grãos-Safra 2021/22 2º levantamento* (v. 9; 2).

Couto, M. R. M., Lúcio, A. D., Lopes, S. J., & Carpes, R. H. (2009). Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. *Ciência Rural*, *39*(6), 1701–1707. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782009000600011&nrm=iso

Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético* (UFV, Ed.; 4ª). 2012.

Dornelles, E. F., Silva, J. A. G. da, Carvalho, I. R., Alessi, O., Pansera, V., Lautenchleger, F., Stumm, E. M. F., Carbonera, R., Bárta, R. L., & Tisott, J. v. (2020). Research article resistance of oat cultivars to reduction in fungicide use and a longer interval from application to harvest to promote food security. *Genetics and Molecular Research*, *19*, 1–12.

Dumlupinar, Z., Maral, H., Kara, R., Dokuyucu, T., & Akkaya, A. (2011). Evaluation of Turkish oat landraces based on grain yield, yield components and some quality traits. *Turkish Journal of Field Crops*, *16*(2), 190–196.

Falconer, D. S., & Mackay, T. F. (1997). *Introduction to quantitative genetics* (4th ed.).

Faraway, J. (2016). *faraway: Functions and datasets for books by Julian Faraway*. Cran.R-Project. https://cran.r-project.org/package=faraway

Ferreira, D. F. (2009). *Estatística Básica* (UFLA, Ed.; 2nd ed.).

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Hair, J. F., Black, W. C., Babin, B., Anderson, R., & Tatham, R. L. (2009). *Análise multivariada de dados*.

Harrell, F. E. (2021). *Frank E Harrell*. CRAR.R-Project. https://cran.r-project.org/package=Hmisc

Kaziu, I., Kashta, F., & Celami, A. (2019). Estimation of Grain Yield, Grain Components and Correlations between Them in some Oat Cultivars. *Albanian Journal of Agricultural Sciences*, *18*(1), 13–19. https://www.proquest.com/scholarly-journals/estimation-grain-yield-components-correlations/docview/2317872862/se-2

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*(2), 151–162.

Kraisig, A. R., da Silva, J. A. G., Carvalho, I. R., Mamann, Â. T. W. de, Corso, J. S., & Norbert, L. (2020). Time of nitrogen supply in yield, industrial and chemical quality of oats grains. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *24*, 700–706.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models* (McGraw-Hill, Ed.; 4th ed.).

Leite, J. D. B., Federizzi, L., & Bergamaschi, H. (2012). Mudanças climáticas e seus possíveis impactos aos sistemas agrícolas no Sul do Brasil. *Agrária*, *7*, 337. https://doi.org/10.5039/agraria.v5i3a1239

Lúcio, A. D., Storck, L., Krause, W., Gonçalves, R. Q., & Nied, A. H. (2013). Relações entre os caracteres de maracujazeiro-azedo. *Ciência Rural*, *43*(2), 225–232. https://doi.org/10.1590/S0103-84782013000200006

Mantai, R. D., Silva, J. A. G. da, Binelo, M. O., Sausen, A. T. Z. R., Rossi, D. S., & Corso, J. S. (2020). Nitrogen management in the relationships between oat inflorescence components and productivity. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *24*(6), 385–393. https://doi.org/10.1590/1807-1929/agriambi.v24n6p385-393

Mantai, R. D., Silva, J. A. G. da, Marolli, A., Mamann, Â. T. W. de, Sawicki, S., & Krüger, C. A. M. B. (2017). Simulation of oat development cycle by photoperiod and temperature. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *21*(1), 3–8. https://doi.org/10.1590/1807-1929/agriambi.v21n1p3-8

Mantai, R. D., Silva, J. A. G. da, Scremin, O. B., Carvalho, I. R., Magano, D. A., Fachinetto, J. M., Lautenchleger, F., Rosa, J. A. da, Peter, C. L., Berlezi, J. D., & Babeski, C. M. (2020). Nitrogen levels in oat grains and its relation to productivity. *Genetics and Molecular Research*, *19*(4), 1–13.

Meira, D., Meier, C., Olivoto, T., Follmann, D. N., Rigatti, A., Lunkes, A., Marchioro, V. S., & Souza, V. Q. de. (2019). Multivariate analysis revealed genetic divergence and promising traits for indirect selection in black oat. *Revista Brasileira de Ciências Agrárias*, *14*, 1–7. https://doi.org/10.5039/agraria.v14i4a6514

Meira, D., Meier, C., Olivoto, T., Nardino, M., Klein, L. A., Moro, E. D., Fassini, F., Marchioro, V. S., & Souza, V. Q. de. (2019). Estimates of genetic parameters between and within black oat populations. *Bragantia*, *78*, 43–51. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0006-87052019000100043&nrm=iso

Meira, D., Meier, C., Olivoto, T., Nardino, M., Rigatti, A., Klein, L. A., Caron, B. O., Marchioro, V. S., & Souza, V. Q. de. (2019a). Phenotypic variance of black oat growing in crop seasons reveals genetic effects predominance. *Anais Da Academia Brasileira de Ciências*, *91*. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652019000500864&nrm=iso

Meira, D., Meier, C., Olivoto, T., Nardino, M., Rigatti, A., Klein, L. A., Caron, B. O., Marchioro, V. S., & Souza, V. Q. de. (2019b). Phenotypic variance of black oat growing in crop seasons reveals genetic effects predominance. *Anais Da Academia Brasileira de Ciências*, *91*(3). https://doi.org/10.1590/0001-3765201920180036

Montgomery, D. C., & Peck, E. A. (1982). *Introduction to linear regression analysis*.

Moradi, M., A, R., & Arzani, A. (2005). Path Analysis for Yield and Related Traits in Oats. *Journal of Science and Technology of Agriculture and Natural Resources*.

Moreira, F. V., Cecato, U., Prado, I. N. do, Wada, F. Y., Rego, F. C. de A., & Nascimento, W. G. do. (2008). Avaliação de aveia preta cv Iapar 61 submetida a níveis crescentes de nitrogênio em área proveniente de cultura de soja. *Acta Scientiarum Animal Sciences*, *23*. https://doi.org/10.4025/actascianimsci.v23i0.2608

Moreira, S. O., Gonçalves, L. S. A., Rodrigues, R., Sudré, C. P., Júnior, A. T. do A., & Medeiros, A. M. (2013). Correlações e análise de trilha sob multicolinearidade em linhas recombinadas de pimenta (*Capsicum annuum* L.). *Revista Brasileira de Ciências Agrárias*, *8*(1), 15–20.

Olivoto, T., de Souza, V. Q., Nardino, M., Carvalho, I. R., Ferrari, M., de Pelegrin, A. J., Szareski, V. J., & Schmidt, D. (2017). Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. *Agronomy Journal*, *109*(1), 131–142. https://doi.org/https://doi.org/10.2134/agronj2016.04.0196

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Rodrigues, G. B., Marim, B. G., Silva, D. J. H. da, Mattedi, A. P., & Almeida, V. de S. (2010). Análise de trilha de componentes de produção primários e secundários em tomateiro do grupo salada. *Pesquisa Agropecuaria Brasileira*, *45*(2), 155–162.

Royston, J. P. (1983). Some Techniques for Assessing Multivariate Normality Based on the Shapiro- Wilk W. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *32*(2), 121–133. https://doi.org/10.2307/2347291

Salla, V. P., Danner, M. A., Citadin, I., Sasso, S. A. Z., Donazzolo, J., & Gil, B. V. (2015). Análise de trilha em caracteres de frutos de jabuticabeira. *Pesquisa Agropecuaria Brasileira*, *50*(3), 218–223.

Sari, B. G., Lúcio, A. D., Olivoto, T., Krysczun, D. K., Tischler, A. L., & Drebes, L. (2018). Interference of sample size on multicollinearity diagnosis in path analysis. *Pesquisa Agropecuaria Brasileira*, *53*(6), 769–773.

Sari, B. G., Lúcio, A. D., Santana, C. S., & Lopes, S. J. (2017). Linear relationships between cherry tomato traits. *Ciência Rural*, *47*(3), 1–7.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Silva, A. R. da, Malafaia, G., & Menezes, I. P. P. (2017). biotools: an R function to predict spatial gen diversity via an individual-based approach. *Genetics and Molecular Research*, *16*(2), 1–6.

Silva, J. A. G. da, Mamann, Â. T. de, Scremin, O. B., Carvalho, I. R., Pereira, L. M., Lima, A. R. C. de, Lautenchlegere, F., Basso, N. C. F., Argenta, C. V., Berlezi, J. D., Porazzi, F. U., Matter, E. M., & Norbert, L. (2020). Biostimulants in the Indicators of Yield and Industrial and Chemical Quality of Oat Grains. *Journal of Agricultural Studies*, *8*, 68. https://doi.org/10.5296/jas.v8i2.15728

Steel, R. G. D., Torrie, J. H., & Dickey, D. A. (1997). *Principles and procedures of statistic: a biometrical approach* (M. H. Book, Ed.; 3rd ed.).

Tedesco, M. J., Gianello, C., Bissani, C. A., Bohnen, H., & Volkweiss, S. J. (1995). *Análise de solo, plantas e outros materiais* (UFRGS, Ed.; 2nd ed.).

Tierney, L. (2021). *tkrplot: TK Rplot*. CRAR.R-Project. https://cran.r-project.org/package=tkrplot

Toebe, M., & Cargnelutti Filho, A. (2013). Multicollinearity in path analysis of maize (*Zea mays* L.). *Journal of Cereal Science, 57*, 453–462. https://doi.org/10.1016/j.jcs.2013.01.014

Toebe, M., & Cargnelutti Filho, A. (2013). Não normalidade multivariada e multicolinearidade na análise de trilha em milho. *Pesquisa Agropecuária Brasileira*, *48*(5), 466–477. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2013000500002&nrm=iso

Toebe, M., Cargnelutti Filho, A., Lindolfo, S., & Lúcio, A. D. (2017). Direct effects on scenarios and types of path analyses in corn hybrids. *Genetics and Molecular Research*, *16*(1), 1–15.

Toebe, M., Cargnelutti Filho, A., Lindolfo, S., & Lúcio, A. D. (2017). Sample size for estimation of direct effects in path analysis of corn. *Genetics and Molecular Research*, *16*(2), 1–23.

Vencovski, R., & Barriga, P. (1992). *Genética biométrica no fitomelhoramento*.

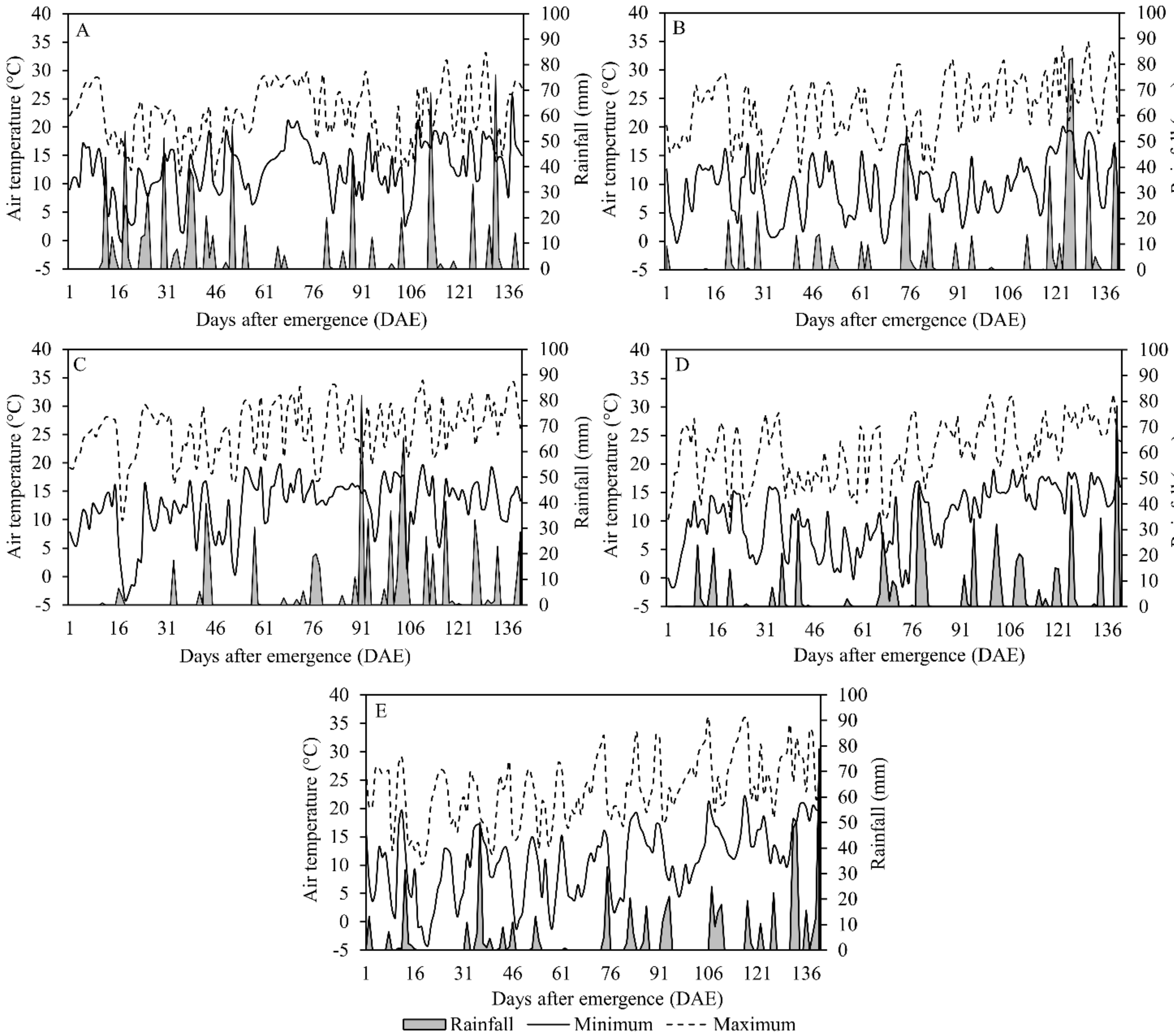


Fig. 1 Air temperature (minimum and maximum) and accumulated rainfall during the oat growing season (May to October) in five agricultural years: 2015 (A), 2016 (B), 2017 (C), 2018 (D), and 2019 (E).

Table 1 Shapiro-Wilk (SW) univariate normality test p-value and Bartlett (B) univariate homoscedasticity test of variances (B) for 7 variables, from 22 white oat cultivars in 5 environments, for original data. Without transformation (B and SW) and with transformation (Bt and SWt) of data, of variables that did not fit in the adherence tests, using the Box-Cox methodology and the respective λ value used.

Statistics	Variables						
	Yield	PL	PDM	NSP	NGP	WGP	HI
				Environment 1			
VIF	2.1873	1.1743	480.0791	2.7202	3.3897	578.0092	16.4623
CN				4467.852			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.433	1.3479	1113.2523	4.6444	6.5444	6.5444	1205.0463
CN				9426.906			
Shapiro-Wilk				0.0000			
				Environment 3			
VIF	1.7348	1.2697	384.8615	2.39778	3.4635	469.147	18.464
CN				3388.211			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7614	1.2459	512.164	3.4139	4.1872	587.614	14.1485
CN				4045.003			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5557	1.3288	744.883	3.9402	4.4292	814.6135	18.4659
CN				6189.635			
Shapiro-Wilk				0.0000			
				——————————Transformed data——————————			
				Environment 1			
VIF	2.1873	1.1743	480.0791	2.7202	3.3897	578.0093	16.4623
CN				4467.852			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.4329	1.3479	1113.2523	4.6444	6.5444	1205.0463	19.0268
CN				9426.906			
Shapiro-Wilk				0.0000			

				Environment 3			
VIF	1.7448	1.2697	384.8615	2.3978	3.4634	469.147	18.464
CN				3388.211			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7614	1.2459	512.164	3.4139	4.1872	587.6139	14.1485
CN				4045.003			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5557	1.3288	744.883	3.9402	4.4292	814.6135	18.4659
CN				6189.635			
Shapiro-Wilk				0.0000			
				−−−−−−−−−−−−−Variable elimination−−−−−−−−−−−−−			
				Environment 1			
VIF	2.1804 (0.32%)	1.1536 (1.76%)	−	2.7163 (0.14%)	3.3888 (0.03%)	3.2965 (99.43%)	2.2336 (86+43%)
CN				18.3585 (99.59%)			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.4316 (0.10%)	1.3395 (0.62%)	−	4.6412 (0.07%)	6.5208 (0.36%)	3.5315 (46.04%)	1.2716 (99.89%)
CN				33.564 (99.64%)			
Shapiro-Wilk				0.0000			
				Environment 3			
VIF	1.7187 (0.93%)	1.2526 (1.35%)	−	2.39 (0.32%)	3.4089 (1.58%)	3.136 (99.33%)	2.0588 (88.85%)
CN				15.6728 (99.54%)			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7311 (1.72%)	1.2069 (3.13%)	−	3.3951 (0.55%)	4.1764 (0.26%)	2.2595 (99.62%)	1.9725 (86.06%)
CN				19.9597 (99.51%)			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5198 (2.31%)	1.3263 (0.19%)	−	3.9399 (0.01%)	4.3947 (0.78%)	2.5447 (99.69%)	1.3945 (92.45%)
CN				23.3466 (99.62%)			
Shapiro-Wilk				0.0000			

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 2 Diagnosis of multicollinearity, based on the variance inflation factor (VIF) and condition number (CN), and diagnosis of multivariate normality, using the Shapiro-Wilk test modified by Royston, for the oat yield components, considering the original data from 5 environments, data transformation and exclusion of variables to meet the assumptions.

Statistics	Variables						
	Yield	PL	PDM	NSP	NGP	WGP	HI
				Environment 1			
VIF	2.1873	1.1743	480.0791	2.7202	3.3897	578.0092	16.4623
CN				4467.852			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.433	1.3479	1113.2523	4.6444	6.5444	6.5444	1205.0463
CN				9426.906			
Shapiro-Wilk				0.0000			
				Environment 3			
VIF	1.7348	1.2697	384.8615	2.39778	3.4635	469.147	18.464
CN				3388.211			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7614	1.2459	512.164	3.4139	4.1872	587.614	14.1485
CN				4045.003			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5557	1.3288	744.883	3.9402	4.4292	814.6135	18.4659
CN				6189.635			
Shapiro-Wilk				0.0000			
				————————Transformed data————————			
				Environment 1			
VIF	2.1873	1.1743	480.0791	2.7202	3.3897	578.0093	16.4623
CN				4467.852			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.4329	1.3479	1113.2523	4.6444	6.5444	1205.0463	19.0268
CN				9426.906			
Shapiro-Wilk				0.0000			
				Environment 3			

VIF	1.7448	1.2697	384.8615	2.3978	3.4634	469.147	18.464
CN				3388.211			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7614	1.2459	512.164	3.4139	4.1872	587.6139	14.1485
CN				4045.003			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5557	1.3288	744.883	3.9402	4.4292	814.6135	18.4659
CN				6189.635			
Shapiro-Wilk				0.0000			
				−−−−−−−−−−−−−Variable elimination−−−−−−−−−−−−−−			
				Environment 1			
VIF	2.1804 (0.32%)	1.1536 (1.76%)	−	2.7163 (0.14%)	3.3888 (0.03%)	3.2965 (99.43%)	2.2336 (86+43%)
CN				18.3585 (99.59%)			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.4316 (0.10%)	1.3395 (0.62%)	−	4.6412 (0.07%)	6.5208 (0.36%)	3.5315 (46.04%)	1.2716 (99.89%)
CN				33.564 (99.64%)			
Shapiro-Wilk				0.0000			
				Environment 3			
VIF	1.7187 (0.93%)	1.2526 (1.35%)	−	2.39 (0.32%)	3.4089 (1.58%)	3.136 (99.33%)	2.0588 (88.85%)
CN				15.6728 (99.54%)			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.7311 (1.72%)	1.2069 (3.13%)	−	3.3951 (0.55%)	4.1764 (0.26%)	2.2595 (99.62%)	1.9725 (86.06%)
CN				19.9597 (99.51%)			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.5198 (2.31%)	1.3263 (0.19%)	−	3.9399 (0.01%)	4.3947 (0.78%)	2.5447 (99.69%)	1.3945 (92.45%)
CN				23.3466 (99.62%)			
Shapiro-Wilk				0.0000			

Between parentheses were shown in percentage, the effects of eliminating the variable in the VIF and NC statistics.

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 3 Diagnosis of multicollinearity, based on the variance inflation factor (VIF) and condition number (CN), and diagnosis of multivariate normality, using the Shapiro-Wilk test modified by Royston, for the oat yield components, considering the predicted data from 5 environments, with the exclusion of variables to meet the assumptions.

Statistics	Variáveis						
	Yield	PL	PDM	NSP	NGP	WGP	HI
				Environment 1			
VIF	1.1104	1.0133	244.2612	2.1068	2.1175	253.0662	8.9986
CN				1346.755			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.5609	1.3799	1313.4052	5.2035	7.7337	1441.3268	24.1068
CN				11618.97			
Shapiro-Wilk				0.0003			
				Environment 3			
VIF	1.8688	1.2539	485.7508	2.5454	3.8773	608.9549	25.2638
CN				4470.507			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.959	1.2886	666.6767	3.6986	4.5494	777.1991	18.4365
CN				5409.635			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.7342	1.3999	907.9075	4.2447	4.8462	1000.1063	22.9544
CN				7802.556			
Shapiro-Wilk				0.0000			
				——————————Variable elimination——————————			
				Environment 1			
VIF	1.1104 (0.00%)	1.0074 (0.58%)	–	2.0885 (0.87%)	2.1166 (0.04%)	1.3525 (99.47%)	1.1439 (87.29%)
CN				7.0499 (99.48%)			
Shapiro-Wilk				0.0000			
				Environment 2			
VIF	1.5607 (0.01%)	1.3625 (1.26%)	–	5.1924 (0.21%)	7.6784 (0.72%)	3.8165 (99.74%)	1.3111 (94.56%)
CN				41.1201 (99.65%)			
Shapiro-Wilk				0.0000			

				Environment 3			
VIF	1.8536 (0.81%)	1.2514 (0.20%)	–	2.5415 (0.15%)	3.8101 (1.73%)	3.5568 (99.42%)	2.2764 (90.99%)
CN				17.9893 (99.60%)			
Shapiro-Wilk				0.0000			
				Environment 4			
VIF	1.9289 (1.54%)	1.2648 (1.85%)	–	3.6943 (0.12%)	4.5492 (0.00%)	2.479 (99.68%)	2.2476 (87.81%)
CN				22.465 (99.58%)			
Shapiro-Wilk				0.0000			
				Environment 5			
VIF	1.6494 (4.89%)	1.3997 (0.01%)	–	4.2433 (0.03%)	4.8104 (0.74%)	2.8159 (99.72%)	1.4402 (93.73%)
CN				26.4658 (99.66%)			
Shapiro-Wilk				0.0000			

Between parentheses were shown in percentage, the effects of eliminating the variable in the VIF and NC statistics.
Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 4 Pearson's correlation coefficients for the yield components of white oat grown in 5 environments, with original data (below the diagonal) and predicted data (above the diagonal), with n=330.

Variables	Yield	PL	PDM	PSN	PGN	GDM	HI
				Environment 1			
Yield	**1**	0.03^{ns}	0.07^{ns}	−0.16*	0.04^{ns}	0.10^{ns}	0.18*
PL	0.04^{ns}	**1**	0.00^{ns}	0.01^{ns}	0.06^{ns}	0.00^{ns}	−0.02^{ns}
PDM	0.58*	0.25*	**1**	0.40*	0.45*	0.98*	0.02^{ns}
PSN	0.42*	0.33*	0.63*	**1**	0.69*	0.36*	−0.17*
PGN	0.48*	0.25*	0.73*	0.77*	**1**	0.42*	−0.11*
GDM	0.63*	0.24*	0.99*	0.61*	0.73*	**1**	0.20*
HI	0.68*	0.07^{ns}	0.54*	0.31*	0.41*	0.64*	**1**
				Environment 2			
Yield	**1**	0.18*	0.50*	0.31*	0.42*	0.53*	0.44*
PL	0.12*	**1**	0.50*	0.41*	0.42*	0.49*	0.14*
PDM	0.44*	0.48*	**1**	0.70*	0.80*	0.99*	0.28*
PSN	0.26*	0.40*	0.69*	**1**	0.89*	0.69*	0.23*
PGN	0.35*	0.39*	0.79*	0.88*	**1**	0.81*	0.31*
GDM	0.47*	0.47*	0.99*	0.69*	0.80*	**1**	0.40*
HI	0.41*	0.07^{ns}	0.25*	0.21*	0.29*	0.36*	**1**
				Environment 3			
Yield	**1**	−0.01*	0.59*	0.26*	0.55*	0.63*	0.56*
PL	0.00^{ns}	**1**	0.25*	0.35*	0.16*	0.20*	−0.11*
PDM	0.55*	0.26*	**1**	0.55*	0.74*	0.99*	0.52*
PSN	0.24*	0.37*	0.54*	**1**	0.70*	0.51*	0.12^{ns}
PGN	0.52*	0.17*	0.73*	0.69*	**1**	0.75*	0.50*
GDM	0.59*	0.21*	0.99*	0.51*	0.74*	**1**	0.65*
HI	0.54*	−0.12*	0.48*	0.11^{ns}	0.47*	0.61*	**1**
				Environment 4			
Yield	**1**	−0.17*	0.45*	0.23*	0.39*	0.51*	0.65*
PL	−0.10^{ns}	**1**	0.28*	0.18*	0.15*	0.24*	−0.14*
PDM	0.42*	0.30*	**1**	0.49*	0.61*	0.99*	0.49*
PSN	0.21*	0.22*	0.49*	**1**	0.84*	0.47*	0.19*
PGN	0.37*	0.21*	0.61*	0.83*	**1**	0.61*	0.37*
GDM	0.48*	0.25*	0.99*	0.48*	0.61*	**1**	0.59*
HI	0.61*	−0.10^{ns}	0.43*	0.17*	0.34*	0.55*	1
				Environment 5			

Yield	**1**	0.19*	0.53*	0.42*	0.48*	0.56*	0.45*
PL	0.19*	**1**	0.51*	0.37*	0.34*	0.48*	0.03ns
PDM	0.48*	0.47*	**1**	0.63*	0.69*	0.99*	0.31*
PSN	0.39*	0.36*	0.62*	**1**	0.87*	0.62*	0.17*
PGN	0.43*	0.33*	0.67*	0.86*	**1**	0.68*	0.26*
GDM	0.52*	0.45*	0.99*	0.61*	0.67*	**1**	0.44*
HI	0.43*	0.02ns	0.29*	0.16*	0.24*	0.41*	**1**

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.
nsNot significant.
*Significant at 5% error probability.

Table 5 Pearson's correlation coefficients of the explanatory variables with the grain yield of oats, in 5 environments, Pearson's correlation coefficient between the 5 coefficients obtained with the original values of the variables and after data transformation (between groups) and Pearson's correlation coefficient between the 5 coefficients obtained with the original and predicted values of the variables (between groups), with n=5.

	Variables					
Environments	PL	PDM	PSN	PGN	GDM	HI
			Original data			
1	0.04^{ns}	0.58^{*}	0.42^{*}	0.48^{*}	0.63^{*}	0.68^{*}
2	0.12^{*}	0.44^{*}	0.26^{*}	0.35^{*}	0.47^{*}	0.41^{*}
3	0.00^{ns}	0.55^{*}	0.24^{*}	0.52^{*}	0.59^{*}	0.54^{*}
4	-0.10^{ns}	0.42^{*}	0.21^{*}	0.37^{*}	0.48^{*}	0.61^{*}
5	0.19^{*}	0.48^{*}	0.39^{*}	0.43^{*}	0.52^{*}	0.43^{*}
Mean	0.05	0.49	0.30	0.43	0.54	0.53
			Transformed data			
1	0.04^{ns}	0.59^{*}	0.43^{*}	0.51^{*}	0.65^{*}	0.68^{*}
2	0.12^{*}	0.44^{*}	0.26^{*}	-0.34^{*}	0.47^{*}	0.40^{*}
3	0.01^{ns}	0.55^{*}	0.26^{*}	0.52^{*}	0.59^{*}	0.56^{*}
4	-0.10^{ns}	0.43^{*}	0.24^{*}	0.38^{*}	0.49^{*}	0.62^{*}
5	-0.18^{*}	0.48^{*}	0.38^{*}	0.43^{*}	0.51^{*}	0.44^{*}
Mean	-0.02	0.50	0.31	0.30	0.54	0.54
			Predicted data			
1	0.03^{ns}	0.07^{ns}	-0.16^{*}	0.04^{ns}	0.10^{ns}	0.18^{*}
2	0.18^{*}	0.50^{*}	0.31^{*}	0.42^{*}	0.53^{*}	0.44^{*}
3	-0.01^{ns}	0.59^{*}	0.26^{*}	0.55^{*}	0.63^{*}	0.56^{*}
4	-0.17^{*}	0.45^{*}	0.23^{*}	0.39^{*}	0.51^{*}	0.65^{*}
5	0.19^{*}	0.53^{*}	0.42^{*}	0.48^{*}	0.56^{*}	0.45^{*}
Mean	0.04	0.43	0.21	0.38	0.47	0.46
$r^{(1)}$	-0.05^{ns}	1.00^{*}	0.99^{*}	0.74^{ns}	0.99^{*}	1.00^{*}
$r^{(2)}$	0.98^{*}	-0.50^{ns}	-0.41^{ns}	-0.10^{ns}	-0.58^{ns}	-0.27^{ns}

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index. $r^{(1)}$ Pearson's correlation coefficient between the original and transformed data groups; $r^{(2)}$Pearson's correlation coefficient between the original and modified data sets.

^{ns}Not significant.

*Significant at 5% error probability.

Table 6 Variance Inflation Factor (VIF) for the explanatory variables of the yield of oats cultivated in 5 environments, with original data, transformed data, original data excluding variables, predicted data and modified data excluding variables.

Environments	Variables					
	PL	PDM	PSN	PGN	GDM	HI
			Original data			
1	1.151	478.581	2.683	3.389	577.082	15.642
2	1.337	1112.270	4.624	6.543	1204.839	18.838
3	1.265	381.302	2.382	3.368	466.407	17.996
4	1.212	503.389	3.397	4.116	579.004	13.077
5	1.328	727.694	3.934	4.414	798.742	17.539
			Transformed data			
1	1.160	1086.342	3.035	3.875	1351.405	40.629
2	1.222	1079.775	3.878	4.865	1164.683	18.470
3	1.259	468.478	2.382	3.323	578.117	22.314
4	1.213	504.312	3.326	4.032	580.074	13.080
5	1.310	627.500	3.938	4.416	687.984	15.131
			Original data with elimination of variables			
1	1.132	−	2.678	3.388	3.154	1.749
2	1.329	−	4.621	6.520	3.202	1.756
3	1.249	−	2.372	3.298	2.957	1.971
4	1.181	−	3.383	4.111	2.200	1.634
5	1.326	−	3.934	4.371	2.423	1.279
			Predicted data			
1	1.013	244.261	1.978	2.045	253.060	8.976
2	1.369	1313.202	5.160	7.709	1440.177	24.035
3	1.244	481.814	2.522	3.764	606.054	24.790
4	1.228	656.422	3.689	4.492	767.099	17.150
5	1.395	863.520	4.236	4.827	957.012	21.125
			Predicted data with elimination of variables			
1	1.007	−	1.960	2.044	1.344	1.122
2	1.352	−	5.149	7.653	3.455	1.204
3	1.242	−	2.517	3.680	3.293	2.210
4	1.212	−	3.686	4.491	2.403	1.838
5	1.395	−	4.235	4.777	2.632	1.328

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 7 Condition number (CN) for explanatory variables of the yield of oats cultivated in 5 environments, with original data (Orig.), transformed data (Orig.t), original data excluding variables (Orig.e), modified data (Pred.), and predicted data with exclusion of variables (Pred.e).

Environments	Orig.	Orig.t	Orig.e	Pred.	Pred.e
1	3952.276	9272.453 (134.61%)	15.574 (99.61%)	1346.278 (65.94%)	6.4941 (58.30%)
2	8889.842	8522 (4.14%)	31.161 (99.65%)	10795.51 (21.44%)	37.4764 (20.27%)
3	3011.417	3717.6 (23.45%)	13.471 (99.55%)	3943.674 (30.96%)	15.3252 (13.76%)
4	3657.622	3663.841 (0.17%)	17.693 (99.52%)	4846.142 (32.49%)	19.6587 (11.11%)
5	5532.334	4756.321 (14.03%)	20.711 (99.63%)	6725.201 (21.56%)	23.1911 (11.97%)

Between parentheses were shown in percentage, the effects of eliminating the variable and removing the parameters of the mathematical model on the CN statistics.

Table 8 Direct and indirect effects of explanatory variables on yield of oats cultivated in 5 environments, for original (Orig.) and predicted (Pred.) data, without considering multicollinearity (n=330).

	Env 1		Env 2		Env 3		Env 4		Env 5	
Effects	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.
					PL					
Direct on Yield	-0.104	0.021	-0.088	-0.082	-0.053	-0.072	-0.139	-0.176	-0.018	-0.056
Indirect via PDM	0.210	0.000	0.400	-0.181	0.372	0.366	0.662	0.644	1.573	2.560
Indirect via PSN	0.043	-0.004	-0.047	-0.069	-0.035	-0.039	-0.021	-0.012	0.023	0.027
Indirect via PGN	0.005	0.016	0.012	0.054	0.041	0.040	0.041	0.026	0.033	0.036
Indirect via GDM	-0.154	0.000	-0.179	0.425	-0.261	-0.250	-0.562	-0.537	-1.442	-2.412
Indirect via HI	0.044	-0.002	0.025	0.030	-0.062	-0.058	-0.080	-0.111	0.016	0.034
r	0.044	0.032	0.122	0.177	0.003	-0.013	-0.098	-0.166	0.186	0.189
					PDM					
Direct on Yield	0.828	0.010	0.828	-0.361	1.432	1.451	2.238	2.288	3.324	5.059
Indirect via PL	-0.026	0.000	-0.042	-0.041	-0.014	-0.018	-0.041	-0.050	-0.009	-0.028
Indirect via PSN	0.082	-0.137	-0.082	-0.116	-0.051	-0.061	-0.047	-0.034	0.040	0.046
Indirect via PGN	0.014	0.116	0.024	0.102	0.170	0.183	0.122	0.104	0.067	0.071
Indirect via GDM	-0.645	0.074	-0.378	0.852	-1.238	-1.228	-2.190	-2.251	-3.162	-4.935
Indirect via HI	0.328	0.003	0.090	0.061	0.248	0.262	0.339	0.394	0.221	0.318
r	0.580	0.065	0.440	0.496	0.547	0.589	0.421	0.452	0.482	0.531
					PSN					
Direct on Yield	0.130	-0.340	-0.119	-0.166	-0.095	-0.111	-0.097	-0.069	0.065	0.073
Indirect via PL	-0.034	0.000	-0.035	-0.034	-0.019	-0.025	-0.030	-0.031	-0.006	-0.021
Indirect via PDM	0.519	0.004	0.571	-0.251	0.780	0.802	1.098	1.113	2.067	3.198
Indirect via PGN	0.015	0.175	0.027	0.113	0.161	0.173	0.167	0.145	0.086	0.090
Indirect via GDM	-0.400	0.027	-0.263	0.597	-0.638	-0.638	-1.056	-1.077	-1.950	-3.091
Indirect via HI	0.189	-0.023	0.075	0.048	0.055	0.062	0.131	0.154	0.124	0.175
r	0.418	-0.157	0.256	0.306	0.244	0.263	0.213	0.234	0.385	0.424
					PGN					
Direct on Yield	0.020	0.256	0.030	0.126	0.235	0.246	0.201	0.172	0.100	0.104
Indirect via PL	-0.026	0.001	-0.034	-0.035	-0.009	-0.012	-0.028	-0.027	-0.006	-0.019
Indirect via PDM	0.608	0.004	0.654	-0.290	1.039	1.080	1.359	1.390	2.233	3.470
Indirect via PSN	0.101	-0.233	-0.105	-0.149	-0.065	-0.078	-0.080	-0.058	0.056	0.063
Indirect via GDM	-0.476	0.032	-0.303	0.696	-0.924	-0.938	-1.355	-1.390	-2.136	-3.406
Indirect via HI	0.253	-0.016	0.104	0.067	0.244	0.254	0.269	0.301	0.186	0.268
r	0.480	0.044	0.346	0.416	0.520	0.552	0.365	0.387	0.433	0.481
					GDM					
Direct on Yield	-0.651	0.076	-0.381	0.858	-1.257	-1.246	-2.211	-2.271	-3.194	-4.985
Indirect via PL	-0.025	0.000	-0.041	-0.041	-0.011	-0.015	-0.035	-0.042	-0.008	-0.027
Indirect via PDM	0.820	0.010	0.822	-0.358	1.412	1.430	2.217	2.268	3.290	5.009
Indirect via PSN	0.080	-0.123	-0.082	-0.116	-0.048	-0.057	-0.046	-0.033	0.040	0.045
Indirect via PGN	0.014	0.108	0.024	0.102	0.173	0.185	0.123	0.105	0.067	0.071

Indirect via HI	0.393	0.028	0.131	0.085	0.319	0.329	0.427	0.480	0.320	0.448
r	0.631	0.098	0.473	0.532	0.587	0.628	0.475	0.508	0.515	0.562
						HI				
Direct on Yield	0.612	0.142	0.363	0.214	0.520	0.503	0.780	0.810	0.772	1.027
Indirect via PL	-0.007	0.000	-0.006	-0.012	0.006	0.008	0.014	0.024	0.000	-0.002
Indirect via PDM	0.443	0.000	0.206	-0.102	0.682	0.756	0.972	1.114	0.952	1.565
Indirect via PSN	0.040	0.056	-0.025	-0.038	-0.010	-0.014	-0.016	-0.013	0.010	0.012
Indirect via PGN	0.008	-0.029	0.009	0.040	0.110	0.124	0.069	0.064	0.024	0.027
Indirect via GDM	-0.418	0.015	-0.138	0.341	-0.771	-0.815	-1.211	-1.345	-1.325	-2.176
r	0.677	0.183	0.410	0.444	0.538	0.564	0.609	0.653	0.433	0.454
R^2	0.543	0.099	0.302	0.359	0.424	0.465	0.432	0.490	0.357	0.423
Residual	0.676	0.949	0.835	0.800	0.759	0.732	0.754	0.715	0.802	0.759

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 9 Direct and indirect effects of explanatory variables on yield of oats cultivated in 5 environments, for original (Orig.) and predicted (Pred.) data, excluding variables (n=330).

	Env 1		Env 2		Env 3		Env 4		Env 5	
Effects	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.
					PL					
Direct on Yield	-0.099	0.021	-0.085	-0.083	-0.043	-0.070	-0.121	-0.165	-0.012	-0.056
Indirect via PSN	0.044	-0.004	-0.047	-0.069	-0.037	-0.040	-0.018	-0.011	0.023	0.026
Indirect via PGN	0.005	0.016	0.010	0.055	0.044	0.043	0.040	0.026	0.042	0.049
Indirect via GDM	0.060	0.000	0.226	0.238	0.067	0.076	0.047	0.047	0.128	0.162
Indirect via HI	0.034	-0.002	0.017	0.037	-0.027	-0.022	-0.045	-0.063	0.006	0.009
r	0.044	0.032	0.122	0.177	0.003	-0.013	-0.098	-0.166	0.186	0.189
					PSN					
Direct on Yield	0.133	-0.340	-0.118	-0.168	-0.102	-0.116	-0.085	-0.064	0.064	0.070
Indirect via PL	-0.032	0.000	-0.034	-0.035	-0.016	-0.024	-0.026	-0.029	-0.004	-0.021
Indirect via PGN	0.016	0.175	0.023	0.115	0.175	0.186	0.161	0.146	0.108	0.124
Indirect via GDM	0.157	0.031	0.331	0.334	0.164	0.193	0.089	0.094	0.173	0.207
Indirect via HI	0.145	-0.023	0.053	0.059	0.024	0.023	0.074	0.088	0.044	0.044
r	0.418	-0.157	0.256	0.306	0.244	0.263	0.213	0.234	0.385	0.424
					PGN					
Direct on Yield	0.021	0.256	0.027	0.129	0.254	0.265	0.194	0.173	0.126	0.143
Indirect via PL	-0.024	0.001	-0.033	-0.035	-0.008	-0.011	-0.025	-0.025	-0.004	-0.019
Indirect via PSN	0.103	-0.233	-0.104	-0.150	-0.070	-0.082	-0.070	-0.054	0.055	0.061
Indirect via GDM	0.186	0.036	0.382	0.390	0.237	0.284	0.114	0.122	0.189	0.228
Indirect via HI	0.195	-0.016	0.074	0.082	0.106	0.095	0.152	0.171	0.066	0.068
r	0.480	0.044	0.346	0.416	0.520	0.552	0.365	0.387	0.433	0.481
					GDM					
Direct on Yield	0.255	0.086	0.480	0.481	0.323	0.377	0.185	0.199	0.283	0.334
Indirect via PL	-0.023	0.000	-0.040	-0.041	-0.009	-0.014	-0.031	-0.039	-0.006	-0.027
Indirect via PSN	0.082	-0.123	-0.081	-0.116	-0.052	-0.059	-0.040	-0.031	0.039	0.043
Indirect via PGN	0.015	0.108	0.021	0.104	0.187	0.200	0.119	0.106	0.084	0.098
Indirect via HI	0.303	0.028	0.094	0.104	0.139	0.124	0.242	0.273	0.114	0.114
r	0.631	0.098	0.473	0.532	0.587	0.628	0.475	0.508	0.515	0.562
					HI					
Direct on Yield	0.471	0.140	0.259	0.262	0.226	0.189	0.442	0.461	0.275	0.261
Indirect via PL	-0.007	0.000	-0.006	-0.012	0.005	0.008	0.012	0.023	0.000	-0.002
Indirect via PSN	0.041	0.056	-0.024	-0.038	-0.011	-0.014	-0.014	-0.012	0.010	0.012
Indirect via PGN	0.009	-0.029	0.008	0.040	0.120	0.134	0.067	0.064	0.030	0.037
Indirect via GDM	0.164	0.017	0.173	0.191	0.198	0.247	0.101	0.118	0.117	0.146
r	0.677	0.183	0.410	0.444	0.538	0.564	0.609	0.653	0.433	0.454
R^2	0.541	0.099	0.302	0.359	0.418	0.461	0.422	0.482	0.342	0.394
Residual	0.677	0.949	0.836	0.800	0.763	0.735	0.760	0.720	0.811	0.779

Yield; PL: panicle length; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 10 Direct and indirect effects of the explanatory variables on yield of oats, with the addition of k value (0.05) in the diagonal of the X'X matrix and considering the original and predicted data =(n=330).

	Env 1		Env 2		Env 3		Env 4		Env 5	
Effects	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.
					PL					
Direct on Yield	-0.093	0.022	-0.077	-0.075	-0.047	-0.071	-0.122	-0.165	-0.012	-0.052
Indirect via PDM	0.028	0.000	0.103	0.104	0.045	0.048	0.032	0.030	0.078	0.102
Indirect via PSN	0.038	-0.003	-0.038	-0.051	-0.030	-0.031	-0.013	-0.008	0.024	0.028
Indirect via PGN	0.009	0.014	0.011	0.042	0.039	0.038	0.034	0.023	0.039	0.046
Indirect via GDM	0.034	0.000	0.108	0.120	0.029	0.033	0.022	0.024	0.051	0.060
Indirect via HI	0.033	-0.002	0.019	0.040	-0.030	-0.026	-0.045	-0.062	0.006	0.009
r	0.049	0.031	0.126	0.181	0.005	-0.010	-0.092	-0.158	0.187	0.192
					PDM					
Direct on Yield	0.111	0.028	0.214	0.208	0.172	0.192	0.108	0.106	0.164	0.201
Indirect via PL	-0.024	0.000	-0.037	-0.038	-0.012	-0.018	-0.036	-0.046	-0.005	-0.026
Indirect via PSN	0.073	-0.121	-0.067	-0.085	-0.044	-0.050	-0.030	-0.022	0.042	0.047
Indirect via PGN	0.025	0.098	0.022	0.080	0.163	0.173	0.100	0.090	0.079	0.091
Indirect via GDM	0.144	0.056	0.228	0.242	0.138	0.164	0.084	0.100	0.113	0.122
Indirect via HI	0.245	0.003	0.069	0.079	0.121	0.120	0.189	0.219	0.082	0.086
r	0.574	0.064	0.429	0.486	0.538	0.580	0.415	0.447	0.474	0.521
					PSN					
Direct on Yield	0.117	-0.299	-0.097	-0.122	-0.080	-0.090	-0.061	-0.044	0.068	0.074
Indirect via PL	-0.031	0.000	-0.031	-0.031	-0.017	-0.025	-0.026	-0.029	-0.004	-0.019
Indirect via PDM	0.069	0.011	0.148	0.144	0.094	0.106	0.053	0.051	0.102	0.127
Indirect via PGN	0.027	0.148	0.025	0.089	0.154	0.163	0.137	0.125	0.101	0.115
Indirect via GDM	0.089	0.021	0.158	0.169	0.071	0.085	0.041	0.048	0.070	0.077
Indirect via HI	0.141	-0.023	0.057	0.064	0.027	0.028	0.073	0.086	0.046	0.047
r	0.413	-0.142	0.261	0.312	0.248	0.267	0.216	0.237	0.382	0.421
					PGN					
Direct on Yield	0.034	0.216	0.028	0.099	0.225	0.232	0.165	0.149	0.117	0.132
Indirect via PL	-0.023	0.001	-0.030	-0.032	-0.008	-0.011	-0.025	-0.025	-0.004	-0.018
Indirect via PDM	0.081	0.013	0.169	0.167	0.125	0.143	0.066	0.064	0.110	0.138
Indirect via PSN	0.090	-0.205	-0.085	-0.109	-0.055	-0.063	-0.051	-0.037	0.058	0.064
Indirect via GDM	0.106	0.024	0.183	0.197	0.103	0.125	0.052	0.062	0.076	0.084
Indirect via HI	0.190	-0.016	0.079	0.088	0.120	0.116	0.150	0.167	0.069	0.073
r	0.479	0.033	0.344	0.411	0.509	0.541	0.357	0.379	0.427	0.474
					GDM					
Direct on Yield	0.145	0.057	0.229	0.243	0.140	0.166	0.085	0.101	0.114	0.123
Indirect via PL	-0.022	0.000	-0.036	-0.037	-0.010	-0.014	-0.031	-0.039	-0.005	-0.025
Indirect via PDM	0.109	0.028	0.212	0.206	0.169	0.189	0.107	0.105	0.163	0.199
Indirect via PSN	0.072	-0.108	-0.067	-0.085	-0.041	-0.046	-0.029	-0.021	0.041	0.046
Indirect via PGN	0.025	0.091	0.023	0.081	0.165	0.175	0.101	0.091	0.078	0.091

Indirect via HI	0.295	0.028	0.100	0.112	0.156	0.150	0.238	0.267	0.118	0.122
r	0.624	0.095	0.461	0.519	0.580	0.619	0.471	0.503	0.510	0.555
					HI					
Direct on Yield	0.459	0.140	0.277	0.281	0.254	0.230	0.435	0.451	0.286	0.279
Indirect via PL	-0.007	0.000	-0.005	-0.011	0.006	0.008	0.012	0.022	0.000	-0.002
Indirect via PDM	0.059	0.001	0.053	0.059	0.082	0.100	0.047	0.051	0.047	0.062
Indirect via PSN	0.036	0.049	-0.020	-0.028	-0.009	-0.011	-0.010	-0.008	0.011	0.013
Indirect via PGN	0.014	-0.025	0.008	0.031	0.106	0.117	0.057	0.055	0.028	0.035
Indirect via GDM	0.093	0.011	0.083	0.097	0.086	0.109	0.046	0.060	0.047	0.054
r	0.655	0.176	0.396	0.429	0.525	0.552	0.587	0.631	0.419	0.440
R^2	0.528	0.090	0.292	0.348	0.410	0.452	0.410	0.468	0.336	0.388
Residual	0.687	0.954	0.842	0.808	0.768	0.740	0.768	0.730	0.815	0.782
k	0.050									

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index.

Table 11 Pearson's correlation coefficients between path coefficients (direct and indirect effects) obtained between data groups (original and modified), for traditional path analysis (without considering multicollinearity), traditional path analysis with elimination of variables and path analysis under multicollinearity (ridge).

Explanatory variables	Traditional[(1)]	With elimination[(2)]	Under multicollinearity[(1)]
PL	0.68*	0.64*	0.65*
PDM	0.95*	−	0.88*
PSN	0.01^{ns}	−0.01^{ns}	−0.04^{ns}
PGN	0.59*	0.64*	0.72*
GDM	0.96*	0.86*	0.82*
HI	0.80*	0.71*	0.68*

Yield; PL: panicle length; PDM: panicle dry mass; PSN: panicle spikelet number; PGN: panicle grain number; GDM: panicle grain dry mass; and HI: harvest index; [1] n = 30; [2] n = 25.

3 ARTIGO II – LINEAR RELATIONSHIPS AND IMPLICATIONS OF REMOVING MODEL PARAMETERS IN TRIALS WITH OAT

Submetido para o periódico: *European Journal of Agronomy*
Situação: Sob Revisão

Linear relationships and implications of removing model parameters in trials with oat

Jaqueline Sgarbossa*[1], Alessandro Dal'Col Lúcio[1], José Antonio Gonzalez da Silva[2], Maria Inês Diel[3], Mariane Peripolli[1], Darlei Michalski Lambrecht[1], Lana Bruna de Oliveira Engers[1], Claiton Nardini[4], Odenis Alessi[2] and Cibele Luisa Peter[2]

[1] Department of Plant Science, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil. E-mail: sgarbossajs@yahoo.com; adlucio@ufsm.br; mperipolli@gmail.com; darleilambrecht@gmail.com; engers.lana@gmail.com

[2] Department of Agronomy, Regional University of Northwestern Rio Grande do Sul State, Ijuí, Rio Grande do Sul, Brazil. E-mail: jose.gonzales@unijui.edu.br; odenisalessi@hotmail.com; cibele.peter2017@gmail.com

[3]Federal University of Pampa, Itaqui, Rio Grande do Sul, Brazil. E-mail: mariaines.diel@hotmail.com

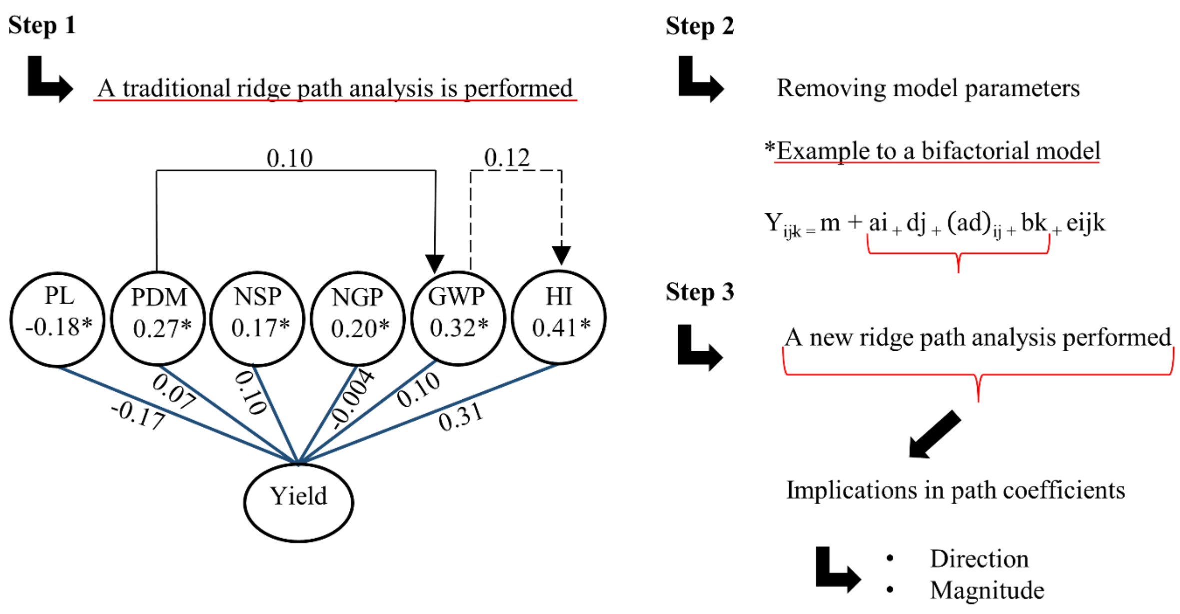
[4]Department of Agronomy and Natural Sciences, Federal University of Santa Maria *Campus* Frederico Westphalen, Frederico Westphalen, Rio Grande do Sul, Brazil. E-mail: claitonnardini@live.com

Corresponding Author*: sgarbossajs@yahoo.com

3.1 HIGHLIGHTS

- Ridge path analysis with model parameters removal is proposed;
- Linear relationships of oat yield traits vary between cropping scenarios.
- Removing parameters from the model generates implications in the path coefficients;

3.2 GRAFICAL ABSTRACT



3.3 ABSTRACT

1. CONTEXT: Path analysis (PA) is a multivariate statistical technique used to analyze cause-and-effect relationships between traits. When carrying out this analysis, the parameters of the mathematical model referring to the experimental design are disregarded, and average observations are used, without stratification of effects. 2. OBJECTIVE: Thus, this study was conducted with the aim of analyzing the implications of removing the parameters from the mathematical model on the results of path analysis, with oats grown in different years and agricultural scenarios (with and without fungicide). 3. METHODS: For this, two field trials were conducted in southern Brazil, in five years of growth. The experimental design used in trial I (with fungicide application) was randomized complete blocks (RCB), in a 22 × 4 bifactorial arrangement, characterized by twenty-two oat cultivars and four fungicide applications. For trial II (without fungicide application) the RCB design was used, and the treatments were characterized by twenty-two oat cultivars, with three replications. The traits

measured were panicle length, panicle mass, number of spikelets, number of grains, grain mass, and grain yield. For each year, data group, and scenario, the correlation coefficients between the explanatory variables and grain yield were calculated. In addition, the diagnosis of multicollinearity in the correlation matrix was made, through the methods of variance inflation factor and condition number, and later, a path analysis was carried out, considering grain yield as the main trait. 4. RESULTS AND CONCLUSIONS: The diagnosis of multicollinearity indicated the violation of the statistical assumption, so it was necessary to proceed with a path analysis under multicollinearity (ridge). The removal of parameters from the mathematical model promoted changes in the linear relationships between the oat yield traits, with the maintenance of the linear correlation coefficients in 3.30% and 20% of the situations, for the scenarios with and without fungicide application, respectively. Regarding the path coefficients, it was observed that the direct effects were maintained in 3.30% and 30% and indirect effects in 7.33% and 24.67% of the situations, for the scenarios with and without fungicide application, respectively. 5. SIGNIFICANCE: The study revealed that the removal of the model parameters promotes changes in the path coefficients and, consequently, has implications for the results and information generated. This new approach is proposed in order to expand the scope of the information generated, as it does not have effects arising from the design and treatments used in the experiment.

Keywords: *Avena sativa*; multicollinearity; path analysis; simple correlation.

3.4 INTRODUCTION

Oat (*Avena sativa* L.) is one of the main winter cereals grown in the world, with about 9,442,749 hectares (ha) of cultivated area, production of 23,132,209 tons of grains and average yield of 2,450 kg ha^{-1} (FAO, 2022). Brazil stands out as the second most interesting winter crop, with an 8% increase in the planted area between the 2021 and 2022 harvests, production of 1,262.6 tons of grain, and yield of 2,321 kg ha^{-1}. In turn, Rio Grande do Sul accounts for 71.2% of this area destined for oat exploitation (387,600 ha), production of 937,200 tons of grains, and yield of 2,418 kg ha^{-1}, which is higher than the national average (CONAB, 2022) .

The increase in the area destined for the cultivation of oats is directly related to its use in human food and animal feed, vegetation cover, and straw for the no-tillage system (Achleitner et al., 2008; Bortolini et al., 2005; Buerstmayr et al., 2007; Castro et al., 2012; Floss et al., 2007; Fontaneli et al., 2009; Marchioro et al., 2001). The demand for oats for human consumption has increased considerably over time, due to the benefits generated for the diet, such as whole grains, source of soluble fiber, balanced energy, and nutritional supply, having in its chemical constitution amino acids, fatty acids, vitamins and essential mineral salts for the human body (Achleitner et al., 2008; Duda et al., 2021; Gutkoski et al., 2009) . Additionally, the consumption of oats in the diet has been related to reduction of risks and improvement in the conditions of numerous diseases such as diabetes, hyperglycemic complications, dyslipidemia, and hypercholesterolemia, reducing the risk of cardiovascular diseases and gastrointestinal disorders, improving immune functions, and aiding in the control overweight and obesity (Guimarães et al., 2021).

Studies have been carried out to enhance oat production systems through either genetic improvement programs aimed at the development of new materials (Alessi et al., 2018; Caierão et al., 2006; Klein et al., 2019; Mantai et al., 2017; Meira et al., 2019a) or the improvement of cultural management and treatments (Dornelles et al., 2020; Kraisig et al., 2020; Mantai et al., 2020a, 2020b). To obtain superior genotypes it is essential to carry out efficient selection, which can often be laborious and time-consuming when performed directly on the trait of interest. This difficulty can be overcome by selecting materials based on their yield components and/or other adaptive traits that indirectly favor the main trait of interest. However, indirect selection requires a high correlation between the trait under selection and the object trait (Falconer and Mackay, 1997).

Simple correlation analyses make it possible to identify the direction (+ or -) and magnitude of the linear association between two traits. However, they do not indicate a cause-and-effect relationship between the traits (Cruz et al., 2012; Ferreira, 2009; Sari et al., 2018; Vencovski and Barriga, 1992) . When the object of study involves more than two traits of interest, path analysis (PA) becomes more suitable, as it provides information about the interrelationships of the traits. In PA, the simple linear correlation coefficients are broken down into direct and indirect effects, allowing the measurement of the influence of one trait over the other (Cruz et al., 2012). Therefore, it is essential to know and use adequate statistical techniques that allow the identification of cause-and-effect relationships between traits, as they generate reliable information, even indicating the traits that can be used in the indirect selection of genotypes (Bello et al., 2010; Cargnelutti Filho et al., 2015; Nardino et al., 2016) .

On the other hand, for the results obtained through the PA to be reliable, it is essential that the statistical assumptions of the model are met (Hair et al., 2009), especially the absence of multicollinearity between the explanatory traits (Olivoto et al., 2017; Sari et al., 2018). Olivoto et al. (2017) , when comparing the results of estimated traditional PA with plot means and estimated traditional PA considering each observation in the plot, found that violating the assumption generates biased path coefficients, with little biological interpretation. Similarly, Toebe et al. (2017), working with corn, found that carrying out traditional path analysis with severe multicollinearity between the explanatory traits can result in inaccurate estimates of path coefficients, indicated by the obtaining of direct effects above |1|.

Additionally, studies have shown that the production performance of the oat crop is influenced by the genetic constitution of the materials used and the cultivation environment (Holland et al. 2000; Benin et al. 2003a). Another limiting factor for oat production is leaf rust disease (Martinelli, 2003), which influences the quantitative and qualitative performance of the genotypes and can generate reductions of more than 50% in grain yield, especially under unfavorable environmental conditions (Benin et al. 2003a), increasing the magnitude of the interaction between genetic materials and the environment (Benin et al., 2005a). The control of this disease requires frequent applications of fungicides, which considerably increases production costs (Martinelli, 2003). In addition, the application of fungicide affects the parameters of adaptability, responsiveness, and stability of genetic materials, indicating that for genetic improvement and correct recommendations, production performance should be studied considering the characteristics of the environment with and without fungicide application (scenarios) in an individualized way (Benin et al., 2005a; Lorencetti et al., 2004).

When carrying out multivariate analyses, such as PA, the parameters of the mathematical model related to the design and the treatment characteristics are not considered in the use of the technique. Generally, one works with the average values of each treatment or repetition, which arise from a sum of effects, without stratifying the influence of factors (for factorials), block (random block design), and interaction (factorial) aspects that can condition the occurrence of bias in the results obtained. Considering the scarcity of information on the subject and the importance of the PA technique for genetic improvement programs, as well as the influence of contrasting scenarios on oat performance, the following hypotheses were generated: i) the removal of parameters from the mathematical model generates divergent results compared to traditional path analysis; ii) the results obtained by the traditional path methods and with the removal of the model parameters are influenced by the different agricultural scenarios. To meet these hypotheses, the study aimed to analyze the effect of removing the parameters from the mathematical model on the results of path analysis, with the oat crop, grown in different years and agricultural scenarios with and without fungicide applications.

3.5 MATERIAL AND METHODS

3.5.1 Study area and experimental design

The study was carried out with results of experiments conducted during the agricultural years 2015, 2016, 2017, 2018, and 2019, in the city of Augusto Pestana, in southern Brazil, under geographic coordinates of 28º 26' 30” S and 54º 00' 58 ” W, with an altitude of 400 meters (m) above sea level. According to the Köppen climate classification, the climate of the region is Cfa, characterized by an average air temperature of 19.1 °C, ranging from 0 to 38 °C, and accumulated rainfall of 2,040 mm (Alvares et al., 2013). The soil of the experimental area is classified as *Latossolo Vermelho distrófico típico* (Oxisol) (Tedesco et al., 1995).

3.5.1.1 Trial I - with fungicide application

The experimental design used was randomized complete blocks, in a 22×4 bifactorial arrangement, characterized by twenty-two oat cultivars: URS Altiva, URS Brava, URS Guará, URS Estampa, URS Corona, URS Torena, URS Charrua, URS Guria, URS Tarimba, URS Taura, URS 21, FAEM 007, FAEM 006, FAEM 5 Chiarasul, FAEM 4 Carlasul, Brisasul, Barbarasul, Fapa Slava, IPR Afrodite, UPFPS Farroupilha, UPFA Ouro and UPFA Gaudéria, and four fungicide applications: 1 (application performed at 60 days after emergence [DAE]), 2 (applications performed at 60 DAE and 75 DAE), 3 (applications performed at 60, 75, and 90 DAE), and 4 (applications performed at 60, 75, 90 and 105 DAE), with three repetitions.

3.5.1.2 Trial II - without fungicide application

The experimental design used was randomized complete blocks, in a unifactorial arrangement, characterized by twenty-two oat cultivars: URS Altiva, URS Brava, URS Guará, URS Estampa, URS Corona, URS Torena, URS Charrua, URS Guria, URS Tarimba, URS Taura, URS 21, FAEM 007, FAEM 006, FAEM 5 Chiarasul, FAEM 4 Carlasul, Brisasul, Barbarasul, Fapa Slava, IPR Afrodite, UPFPS Farroupilha, UPFA Ouro and UPFA Gaudéria, with three replications.

3.5.2 Crop management

The sowing of oats in all agricultural years was carried out from May 15 to June 15, following the technical recommendations for the crop. Harvesting was carried out from late October to early November in all agricultural years. Production performance was analyzed by collecting plants from three central rows, 5 m long, selected on the day of harvest. The number of spikelets per panicle (NSP) and the number of grains per panicle (NGP) were determined by counting. Panicle length (PL) was measured with a graduated ruler. Panicle dry mass (PDM) and grain weight per panicle (GWP) were determined by weighing on a precision scale. The panicle harvest index (HI) was determined by the ratio between grain weight and panicle dry mass. Grain yield (yield) was determined by weighing grains from the usable plot, with moisture adjusted to 13% and later converted to kg ha^{-1}.

3.5.3 Statistical analysis

To carry out the statistical analyses, the five agricultural years were approached as being five environments, in which environment 1 refers to the year 2015, environment 2 to the year 2016, environment 3 to the year 2017, environment 4 to the year 2018, and environment 5 to the year 2019. Initially, the statistical assumption of multicollinearity between the explanatory traits was tested. The diagnosis of multicollinearity was made considering the variance inflation factor (VIF) and the condition number (CN).

Pearson's correlation analysis was performed, generating the matrix of correlation coefficients; posteriorly, the CN was obtained by the ratio between the highest and the lowest eigenvalue of the X'X correlation matrix. CN ≤ 100 indicates weak multicollinearity, 100 < CN > 1,000, moderate to severe multicollinearity, and CN ≥ 1,000, severe multicollinearity (Montgomery and Peck, 1982). The VIF was obtained for each variable, on the inverse diagonal of the X'X correlation matrix; when the VIF value > 10 is obtained, the occurrence of severe multicollinearity is considered (Hair et al., 2009). The occurrence of multicollinearity between the explanatory variables was defined by obtaining values of CN ≥ 1,000 and VIF >10.

In each environment, path analysis was performed, considering yield as the main trait, depending on the explanatory traits (PL, PDM, NSP, NGP, GWP, and HI) (Cruz et al., 2012). The direct and indirect effects of the explanatory traits on yield were estimated by means of path analysis under multicollinearity conditions (ridge path analysis). Therefore, it was considered that each explanatory trait has a direct effect on yield and acts indirectly through its effects on the other explanatory traits.

In the ridge path analysis, the six explanatory traits (PL, PDM, NSP, NGP, GWP, and HI) were considered to estimate the direct and indirect effects on yield. However, a constant “k” was added to the diagonal of the correlation matrix X'X, to reduce the variance associated with the least squares estimator of the path analysis. Thus, the system of normal equations X'X "β"ˆ= X'Y became (X'X + k) "β"ˆ= X'Y. The addition of values of the constant “k” was tested and its lowest value was chosen, from which the path coefficients stabilized (Cruz et al., 2012).

Additionally, the effects of the parameters of the mathematical model (cultivar, application, and block for trial I and cultivar and block for trial II) and ridge path analysis were also isolated and removed. These results were compared to those observed in the “traditional” path analysis under multicollinearity, to identify whether the removal of model effects generates changes in the results of the path analysis.

Removing the effects of the model parameters is referred to as the uniformity trials (without the application of treatments), considering that each observation is composed of the overall mean plus the random effect of the error. The data group from which the effects of the parameters of the model were removed was designated as predicted, and the data group with the maintenance of the effects of the parameters was designated as original.

The mathematical model for trial I, which is characterized as a two-factor experiment (fixed effect) under a randomized block design, is shown in equation 1:

$$Y_{ijk} = m + a_i + d_j + (ad)_{ij} + b_k + e_{ijk} \quad (1)$$

Where: Y_{ijk} is an observation in block k (k = 1, 2, and 3) referring to treatment level i (22 levels) of factor A (cultivar) with level j (4 levels) of factor D (applications): m is the overall mean of the experiment; a_i is the effect of level i (i = 22) of factor A; d_j is the effect of level j (j = 4) of factor D; $(ad)_{ij}$ is the effect of the interaction of level i of factor A with level j of factor D; b_k is the random effect of block k; e_{ijk} is the random effect of experimental error.

The mathematical model for trial II, which is characterized as a one-factor experiment (fixed effect) under a randomized block design, is shown in equation 2:

$$Y_{ij}=m+ t_i+ b_j+ e_{ij} \quad (2)$$

Where: Y_{ij} is the observed value of the variable Y, in the experimental unit that received level i (22 levels) of treatment t, in block k (k = 1, 2 and 3); m is the overall mean of the experiment; t_i is the effect of level i (i=22) of the treatment; b_j is the random effect of block j; e_{ij} is the experimental error effect.

In all statistical analyses, the level of 5% probability of error was adopted, and all analyses were performed using Excel software and R software (R Core Team, 2021). The analyses were performed using the following packages: stats (R Core Team, 2021), car (Fox and Weisberg, 2019), MVN (Korkmaz et al., 2014), pracma (Borchers, 2021), faraway (Faraway, 2016), Hmisc (Harrell, 2021), biotools (Silva et al., 2017), rpanel (Bowman et al., 2007) and tkrplot (Tierney, 2021).

3.5.4 Weather conditions

The data of air temperature (minimum, average, and maximum) and accumulated rainfall were obtained from a mobile automatic weather station, located approximately 200 meters away from the experimental area. During the period of oat cultivation in 2015, the average air temperature was 17.4 °C, ranging from -0.16 °C to 33.0 °C, and the accumulated rainfall was 798.3 mm (Figure 1). In 2016, the average air temperature was 16.5 °C, ranging from -0.3 °C to 34.7 °C, and the accumulated rainfall was 711.2 mm. In 2017, the average air temperature was 19.1 °C, ranging from -4.2 °C to 34.3 °C, and the accumulated rainfall was

715.5 mm. For 2018, the average air temperature was 16.5 °C, ranging from -1.60 °C to 32.30 °C, and the accumulated rainfall was 687.4 mm. During cultivation in 2019, the average air temperature was 17.0 °C, ranging from -4.2 °C to 36.1 °C, and the accumulated rainfall was 645.5 mm. For some growth and development to occur, oats require temperatures between 0 °C and 35 °C (Leite et al., 2012). Therefore, during all the years of cultivation, the minimum and maximum temperatures exceeded or were very close to the limits established for the crop. Also, in 2015, a period without rainfall was observed soon after crop fertilization, and this aspect was associated with the occurrence of high temperatures during the anthesis period when the development of the reproductive system is particularly sensitive to water stress, and high temperatures may have impacted crop performance.

3.6 RESULTS

3.6.1 Simple correlation

When analyzing the linear relationships between yield components and yield of oat, high levels of significant correlations (76.7%) were observed for the original data group with fungicide application, with r values ranging from |0.02| to |0.41|. On the other hand, for the group of original data without fungicide application, statistical significance was observed in 23.3% of the coefficients, with values ranging from |0.33| to |0.00|. For the predicted data group, the lowest rates of statistical significance (3.30%) were observed, regardless of the scenario (with and without fungicide application), with r values ranging from |0.23| to |0.01| (Table 1).

When analyzing the group of original data with fungicide application, a positive and significant linear relationship of GWP, HI, PDM, and NGP with grain yield was observed, in most environments studied (95%). PL and NSP, on the other hand, showed coefficients of low magnitude in most environments and without statistical significance (60%). For the group of original data without fungicide application, PL and HI showed no significant correlation with yield, in all cultivation environments. PDM was significantly and positively correlated with yield in environments 2 and 3. NSP was negatively and significantly associated with yield only in environment 4. NGP was significantly correlated with yield in environments 1 and 3, while GWP had a significant influence in environments 2 and 3. However, for the group of predicted data, significance was observed only for NSP in environment 5, for the scenario with fungicide application. For the scenario without fungicide application, a significant correlation was observed only for PL, in environment 2.

When individually analyzing the estimates of the correlation coefficients obtained from the combination of each pair of variables with yield in 5 environments, considering the different scenarios and data groups, it is possible to observe different situations. In general, the response pattern was maintained in 11.70% of the combinations, with an inverse direction (signal) in 45% of the combinations and a change greater than 50% in the absolute value of the coefficient, with signal maintenance in 43.30% of the combinations. When analyzing the effect of removing the model parameters on the correlation coefficients, for the scenario with fungicide application, a change in the direction of the associations was observed in 56.70% of the combinations, a change greater than 50% in the magnitude of the coefficients, with the maintenance of signal in 40% of the combinations and maintenance of the response pattern in 3.30% of the situations. For the scenario without fungicide application, there was an inverse direction in the associations in 33.30% of the combinations, >50% change in the magnitude of the coefficients in 46.70% of the combinations with the maintenance of 20% in the response pattern.

3.6.2 Multicollinearity

The multicollinearity diagnoses indicated a violation of the statistical assumption for all environments under study and the data group evaluated (Table 2). The diagnosis of multicollinearity based on VIF indicated that the variables PL, NSP, and NGP do not have a high correlation with the other explanatory variables (VIF<10), regardless of the environment, scenario, and data group. For the variables PDM, GWP and HI, the VIF statistic indicated the existence of multicollinearity, regardless of the data group, scenario, or cultivation environment. Similar responses were observed when making the diagnosis of multicollinearity of the CN statistic, which indicated the existence of moderate to severe multicollinearity between the explanatory variables, regardless of the data group (Table 3).

3.6.3 Traditional path analysis under multicollinearity *vs* Modified path analysis under multicollinearity

The traits PL, PDM, NSP, NGP, GWP, and HI showed explanatory capacity of 22.7%, 18.2%, 18.7%, 12.7%, and 11.5% of the variance in oat yield for environments 1, 2, 3, 4 and 5, respectively, for the original data group with fungicide application (Table 4). For the group of predicted data with fungicide application, the traits could explain 3.0%, 2.9%, 2.0%, 2.0%, and 4.9% of the variance in yield for environments 1, 2, 3, 4, and 5, respectively. Thus, removing parameters from the mathematical model resulted in an average change of 82.10% in the explanatory capacity of the traits.

In general, the removal of the parameters of the mathematical model caused a change in the direction of the path coefficients of the direct effects in 63.30% of the combinations, with a change greater than 50% in the magnitude of the direct effects in 43.30%, with the maintenance of the pattern of response in 3.30% of situations. For the indirect effects, a change of 48% in the direction of the path coefficients, a change of 44.67% in the magnitude (>50% of the absolute value) of the coefficients, and maintenance of the response pattern in 7.33% of the combinations were observed.

For the condition without fungicide application, the traits were able to explain 14.2%, 8.0%, 17.8%, 15.7%, and 6.0% of the accumulated variance, in environments 1, 2, 3, 4, and 5, considering the original data group. For the predicted data group, the traits showed explanatory power of 6.10%, 27.20%, 16.7%, 5.3%, and 10.5% of the accumulated variance. Thus, the removal of model parameters reduced by 61.6% the explanatory capacity of variance in environments 1 and 4, and increased by 157.50% the explanatory capacity in environments 2 and 5, with maintenance in environment 3 (Table 5).

When analyzing the direct effects, the removal of the model parameters resulted in a 13.33% change in the direction of the path coefficients, a change in the magnitude of the coefficients (>50%) in 56.67% of the combinations and maintenance of the response pattern in 30% of the situations. For the indirect effects, a change of 28% in the direction of the path coefficients, a change greater than 50% in the absolute values of the coefficients in 57.30%, and maintenance of the response pattern in 24.67% of the situations were observed.

In order to analyze more specifically the impacts of removing parameters from the mathematical model on the direction and magnitude of the path coefficients, Pearson's correlation was estimated between the values of the direct and indirect effects of each trait, between the groups of data (original and predicted) in each scenario. After estimating Pearson's correlation between the original and predicted data, in the scenario with fungicide application, positive and significant correlations were found for the traits PDM, NSP, NGP, and GWP, ranging from moderate to strong (0.48 to 0.62) (Table 6). For HI there was a strong negative and significant correlation, while for PL a weak correlation was obtained. In the scenario without fungicide application, positive correlations were observed for all traits under study. However, only PL and NSP showed statistical significance, with moderate correlations (0.40 to 0.51).

3.7 DISCUSSION

3.7.1 Simple correlation and multicollinearity

When analyzing the linear relationships between panicle components and oat yield, high levels of significant correlation were found for the scenario with fungicide application, regardless of the data group (original and predicted). Obtaining correlation coefficients with low magnitude values, but with statistical significance, is related to the sensitivity of the correlation coefficient to the number of observations (n) used in the estimates (Lúcio et al., 2013). This result is a consequence that the sample n is a parameter in the equations for estimating the correlation coefficient and the minimum absolute value for the correlation coefficient to show significance, respectively (Filho and Júnior, 2009). Thus, when n is reduced, the value of the correlation coefficient needs to show a high magnitude (|1|), to be statistically significant (Cargnelutti Filho et al., 2010; Hair et al., 2009; Stevenson, 2001). Thus, situations may occur, such as those in the present study, in which the magnitudes of the correlations are relatively low (<0.40) and, even so, statistical significance is identified, especially for the scenario with fungicide application in which n is higher (n=264).

For the scenario with fungicide application, positive and significant correlations were found between the yield components GWP, HI, PDM, and NGP with yield, regardless of the data group (Table 1). There was no response pattern between the data groups for the scenario without fungicide application. However, a tendency to obtain a non-significant correlation between traits and yield was observed. Caierão et al. (2001), when studying the linear relationships of the traits NGP, PDM, and thousand-grain weight (TGW), observed a trend of positive associations with yield, with r of 0.62, 0.72, and 0.36, respectively, values with greater magnitudes than those observed in the present study. In addition, the authors point out that the linear correlation coefficient between panicle weight and yield (0.72) provides good perspectives for indirect selection via panicle weight, especially if one considers that the grains represent about 80 to 85% of the panicle weight, unlike other cereals, in which the percentage is lower.

Benin et al. (2005a), when analyzing different plant selection methods, found correlations of 0.15, 0.47, and 0.27 of PDM with yield for the methods carried out based on individual plant yield, selection carried out based on the average weight of grains and combined selection, respectively. The same authors observed similar results for the NGP trait, obtaining correlation coefficients of 0.12, 0.58, and 0.23, for methods based on individual plant yield, average grain weight, and the use of combined methods, respectively. Mantai et al. (2016), when studying the performance of oats subjected to different doses of N, found low correlations of the traits PDM, NSP, NGP, and PL with yield, at the lowest doses of N applied (30 and 60 kg N ha^{-1}). GWP (mean r = 0.66) and HI (mean r = 0.86) showed high correlations.

Similar responses were observed by Mantai et al. (2020a) , who analyzed the linear relationships between the panicle components and the yield of oats, grown in different succession systems and applied N doses, and observed weak correlations for the soybean/oat system with PL, NSP, NGP, and PDM, with values of $r \leq 0.35$. For the variables GWP and HI, correlations of greater magnitude were found ($0.37 \geq r \leq 0.53$). In the corn/oat succession system, associations of low magnitude were obtained for NSP, NGP, and PDM with yield, regardless of the N rate applied (0, 30, 60, and 120 kg ha^{-1}). However, PL, GWP, and HI were significantly associated with yield. Also, the authors observed that PL negatively influences yield, that is, panicles with greater length result in less productive plants, regardless of the dose of N supplied (Mantai et al., 2020a). The existence of significant correlations indicates the viability of indirect selection to obtain gains in the most important trait, which also directly depends on the heritability of the considered trait (Cruz et al., 2012).

Considering the existence of a linear relationship between oat grain yield and most of the analyzed yield components, mainly for the group of original data with fungicide application, and that the main purpose of this research was to investigate the implications of removing the parameters of the model on the cause-and-effect relationships, and that obtaining statistical significance of the correlation coefficient is sensitive to the number of observations, it was decided to maintain all explanatory traits in the path analysis. Therefore, it was necessary to make a diagnosis of multicollinearity between the explanatory traits, to avoid obtaining biased results.

The multicollinearity diagnoses indicated a violation of the statistical assumption in all environments under study, data group, and scenarios evaluated (Table 2 and Table 3), and the traits PDM, GWP, and HI caused severe multicollinearity. The occurrence of severe multicollinearity between the explanatory traits is a recurrent result in the literature, being found for tomato (Rodrigues et al., 2010; Sari et al., 2017), corn (Entringer et al., 2014; Olivoto et al., 2017; Toebe et al., 2017), soybean (Carvalho et al., 2002; Nogueira et al., 2012; Zuffo et al., 2018), jabuticaba (Salla et al., 2015), black oat (Meira et al., 2019b), wheat (Gondim et al., 2008), canola (Amorim et al., 2008), among others. Among the negative aspects caused by multicollinearity, the inflation of the variance of the estimates of the path coefficients can be highlighted, leading to values that are very high, imprecise, and without biological interpretation (Sari et al., 2017; Toebe et al., 2017).

To overcome the problems related to the occurrence of multicollinearity between the explanatory traits, some strategies can be adopted, such as excluding non-additive traits from the model (they generate multicollinearity) or carrying out a path analysis under multicollinearity (ridge) (Cruz et al., 2012; Montgomery et al., 2012). Considering the first strategy, it would be necessary to investigate the removal of PDM, GWP, and HI. Studies suggest that these traits have a high correlation with yield and a good perspective for indirect selection via PDM (Caierão et al., 2006, 2001; Mantai et al., 2020b). The same studies describe that the grain weight corresponds to 80% to 85% of the panicle mass and that the panicle harvest index is obtained by the ratio between PDM and GWP. Thus, it was decided to carry out the ridge path analysis without removing any variable from the database.

3.7.2 Path analysis under multicollinearity *vs* Path analysis under multicollinearity with removal of model parameters

Removing parameters from the mathematical model resulted in changes in the ability to explain the variance in oat yield, especially for the scenario with fungicide application. In general, the traits made it possible to explain an average of 12.2% and 3.0% of the variance, for the original and predicted data groups, for the scenario with fungicide application, and 12.3% and 13.2% for the original and predicted data groups, for the scenario without fungicide application. Caierão et al. (2001), when analyzing the cause-and-effect relationships of the thousand-grain weight (TGW), NGP, PDM, plant height (PH), days from emergence to flowering (DEF), days from emergence to maturation (DEM) and days from flowering to maturation (DFM) with the yield of oat genotypes, found determination coefficients of 0.59, indicating that about 60% of the observed yield comes from the effects of the analyzed traits. Furthermore, it is interesting to point out that the coefficient of determination is restricted to these levels, because the main trait is quantitative, with a large number of genes with little effect on the trait, showing considerable environmental variance, reducing its heritability (Vesohoski et al., 2011).

In the literature, there are studies that obtained determination coefficients of low magnitude and, consequently, a high residual effect, for example in the industrial grain yield of oats grown in succession to corn, in which the residual effects ranged from 0.62 to 0.86 for the traits panicle length, number of spikelets per panicle, number of grains per panicle, panicle mass, panicle grain mass, and panicle harvest index (Mantai et al., 2020a). Benin et al. (2003) analyzed the cause-and-effect relationships of the traits number of panicles per plant, panicle weight, number of grains per panicle, average grain weight, vegetative cycle and plant height in relation to grain production per oat plant and observed residual effect of 0.50. In the black oat crop, the study of the cause-and-effect relationships of the traits plant height, number of leaves per plant, and number of tillers per plant on the fresh mass and dry mass produced indicated residual effects between 0.40 and 0.72 .(Cargnelutti Filho et al., 2015) For soybean and sweet potato crops, estimates of the direct and indirect effects of secondary traits on primary traits indicated high residual effects, ranging from 0.69 to 0.97 and from 0.82 to 0.92, respectively (Cavalcante et al., 2006; Nogueira et al., 2012).

The determination or explanation coefficient is an indicator of the goodness of fit of the adopted model. In situations where the determination coefficient values are close to or equal to the unit (1), it is accepted that variations in the dependent trait are explained by variations in the explanatory traits (Borges et al., 2011; Kavalco et al., 2014). The coefficient of determination values of the path analysis model, observed in the present study, were of low magnitude and the residual effects were high, indicating that the independent traits considered as predictors of the model explain a small fraction of the variation observed for the dependent trait. This aspect shows that, for the conditions of the related study, the independent traits do not interfere with the yield variance, so there are other traits that may provide a greater impact in terms of selection (Cruz et al., 2012) and should be included in path diagrams (Nogueira et al., 2012). Thus, it was decided not to discuss the results of the path analyses, considering only the implications of removing parameters from the mathematical model on the path coefficients, in each scenario and environment.

The removal of parameters from the mathematical model resulted in changes in the direction and magnitude (>50%) of the path coefficients, for all environments and scenarios studied. In general, maintenance of the response pattern of direct effects was observed in 3.30% and 30% of the combinations, for the scenarios with fungicide application and without fungicide application, respectively (Table 4 and Table 5). For the indirect effects, maintenance of the response pattern was observed in 7.33% and 24.67% of the combinations, for the scenarios with fungicide application and without fungicide application, respectively. Furthermore, the Pearson correlation performed for the path coefficients in each scenario and for each analyzed trait indicated the influence of the data group, confirming the initially indicated results.

The removal of parameters from the mathematical models and the stratification within each scenario are strategies that must be considered by the researcher during the planning of the experiment, in order to avoid possible results that may not show the real relationship between the measured variables. Thus, for situations in which one seeks to expand the scope of the information generated about the cause-and-effect relationships between the different variables measured in agricultural trials, the use of the proposed new approach is strongly suggested, as it allows for removing the influences of treatments and design on the observations and, consequently, on the path coefficients and their interpretations. Thus, it is possible to reduce the possible bias in the estimates of the coefficients, highlighting the real relationship between those variables.

3.8 CONCLUSION

The removal of parameters from the mathematical model promoted changes of 56.7% and 33.3% in the direction and 40.0% and 46.7% in the magnitude of the linear associations between the oat yield traits, for the scenarios with fungicide application and without fungicide application.

The removal of parameters from the mathematical model implies changes in the path coefficients. For the scenario with fungicide application, average changes of 63.3% and 48.0% were obtained in the direction of associations and 43.3% and 44.7% in the magnitude of direct and indirect effects, respectively. For the scenario without fungicide application, average changes of 13.3% and 28.0% were found in the direction and 56.7% and 24.7% in the magnitude of the direct and indirect coefficients, respectively.

3.9 DECLARATION OF COMPETING INTEREST

The authors declare no conflicts of interest.

3.10 DATA AVAILABLE

Data will be made available on request.

3.11 ACKNOWLEDGEMENTS

The authors are grateful to the Coordination for Improvement of Higher Education Personnel (Capes-Brazil) for their financial support of the author Jaqueline Sgarbossa (Process Nº.88887.499817/2020-00). Also, I would like to thank the members of the research groups on Technical Systems of Agricultural Production at the Regional University of Northwestern Rio Grande do Sul State and the research group on Agricultural Experimentation at the Federal University of Santa Maria for help in this project.

3.12 REFERENCES

Achleitner, A., Tinker, N.A., Zechner, E., Buerstmayr, H., 2008. Genetic diversity among oat varieties of worldwide origin and associations of AFLP markes with quantitative traits. Theoretical and Applied Genetics 117, 1041–1053.

Alessi, O., Dornelles, E.F., Mamann, Â.T.W. de, Kraisig, A.R., Henrichsen, L., Marolli, A., Pansera, V., Silva, J.A.G. da, 2018. Aplicação de modelos de regressão e de adaptabilidade e estabilidade na identificação de cultivares de aveia branca com maior resistência genética a doenças foliares. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics 6, 7.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J. de M., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Amorim, E.P., Ramos, N.P., Ungaro, M.R.G., Kiihl, T.A.M., 2008. Correlações e análise de trilha em girassol. Bragantia 67, 307–316. https://doi.org/10.1590/S0006-87052008000200006

Bello, O.B., Abdulmaliq, S.Y., Afolabi, M.S., Ige, S.A., 2010. Correlation and path coefficient analysis of yield and agronomic characters among open pollinated maize varieties and their F1 hybrids in a diallel cross. Afr J Biotechnol 9, 2633–2639.

Benin, G., Carvalho, F.I.F. de, Oliveira, A.C., Lorencetti, C., Vieira, E.A., Coimbra, J.L.M., Valério, I.P., Floss, E.L., Bertan, I., Silva, G.O. da, 2005a. Adaptabilidade e estabilidade em aveia em ambientes estratificados. Ciência Rural 35, 295–302.

Benin, G., Carvalho, F.I.F. de, Oliveira, A.C. de, Lorencetti, C., Valério, I.P., Schmidt, D.A.M., Hartwig, I., Ribeiro, G., Vieira, E.A., Silva, J.A.G. da, 2005b. Early generation selection strategy for yield and yield components in white oat. Sci Agric 62, 357–365. https://doi.org/10.1590/S0103-90162005000400009

Benin, G., Carvalho, F., Oliveira, A., Assmann, I., Floss, E., Lorencetti, C., Marchioro, V.S., Silva, J.A.G. da, 2003. Implicações do ambiente sobre o rendimento de grãos em aveia e suas influências sobre estimativas de parâmetros genéticos. Revista Brasileira de Agrociência 9, 207–214.

Borchers, H.W., 2021. pracma: Practital numerical math functions [WWW Document]. CRAR.R-project. URL https://cran.r-project.org/package=pracma

Borges, V., Sobrinho, F.S., Lédo, F.J. da S., Kopp, M.M., 2011. Associação entre caracteres e análise de trilha na seleção de progênies de meios-irmãos de Brachiaria ruziziensis. Revista Ceres 58, 765–772. https://doi.org/10.1590/S0034-737X2011000600013

Bortolini, P.C., Moraes, A. de, Carvalho, P.C. de F., 2005. Produção de forragem e de grãos de aveia branca sob pastejo. Revista Brasileira de Zootecnia 34, 2192–2199. https://doi.org/10.1590/S1516-35982005000700005

Bowman, R., Alexander, G., Crawford, E., Bowman, A., 2007. rpanel: Simple Interactive Controls for R Functions Using the tcltk Package. J Stat Softw 17, 1–18. https://doi.org/10.18637/jss.v017.i09

Buerstmayr, H., Krenn, N., Stephan, U., Grausgruber, H., Zechner, E., 2007. Agronomic performance and quality of oat (Avena sativa L.) genotypes of worldwide origin produced under Central European growing conditions. Field Crops Res 101, 343–351. https://doi.org/https://doi.org/10.1016/j.fcr.2006.12.011

Caierão, E., Carvalho, F.I.F. de, Floss, E.L., 2006. Seleção indireta para o incremento do rendimento de grãos em aveia. Ciência Rural 36, 1126–1131. https://doi.org/10.1590/S0103-84782006000400013

Caierão, E., Carvalho, F.I.F. de, Pacheco, M.T., Lonrecetti, C., Marchioro, V.S., Silva, J.G., 2001. Seleção indireta em aveia para o incremento no rendimento de grãos. Ciência Rural 31, 231–236. https://doi.org/10.1590/S0103-84782001000200007

Cargnelutti Filho, A., Toebe, M., Alves, B.M., Burin, C., Santos, G.O. dos, Facco, G., Neu, I.M.M., 2015. Relações lineares entre caracteres de aveia preta. Ciência Rural 45, 985–992. https://doi.org/10.1590/0103-8478cr20140500

Cargnelutti Filho, A., Toebe, M., Burin, C., Silveira, T.R. da, Casarotto, G., 2010. Tamanho de amostra para estimação do coeficiente de correlação linear de Pearson entre caracteres de milho. Pesqui Agropecu Bras 45, 1363–1371. https://doi.org/10.1590/S0100-204X2010001200005

Carvalho, C.G.P. de, Arias, C.A.A., Toledo, J.F.F. de, Oliveira, M.F. de, Vello, N.A., 2002. Correlações e análise de trilha em linhagens de soja semeadas em diferentes épocas. Pesqui Agropecu Bras 37, 311–320. https://doi.org/10.1590/S0100-204X2002000300012

Castro, G.S.A., Costa, C.H.M. da, Neto, J.F., 2012. Ecofisiologia da aveia branca. Scientia Agraria Paranaensis 11, 1–15.

Cavalcante, J.T., Ferreira, P.V., Soares, L., Borges, V., Silva, P.P. da, Silva, J.W. da, 2006. Análise de trilha em caracteres de rendimento de clones de batata-doce (Impomea batatas (L.) Lam). Acta Sci Agron 28, 261–266.

CONAB, C.N. de A., 2022. Acompanhamento da safra brasileira de grãos - Safra 2021/2022, 11th ed. file:///C:/Users/sgarb/Downloads/E-book_Boletim-de-Safras-11o-levantamento-compactado.pdf, Brasília.

Cruz, C.D., Regazzi, A.J., Carneiro, P.C.S., 2012. Modelos biométricos aplicados ao melhoramento genético, 4ª. ed. 2012, Viçosa.

Dornelles, E.F., Silva, J.A.G. da, Carvalho, I.R., Alessi, O., Pansera, V., Lautenchleger, F., Stumm, E.M.F., Carbonera, R., Bárta, R.L., Tisott, J. v, 2020. Research article resistance of oat cultivars to reduction in fungicide use and a longer interval from application to harvest to promote food security. Genetics and Molecular Research 19, 1–12.

Duda, M., Tritean, N., Racz, I., Kadar, R., Russu, F., Fițiu, A., Muntean, E., Vâtcă, A., 2021. Yield performance of spring oats varieties as a response to fertilization and sowing distance. Agronomy 11. https://doi.org/10.3390/agronomy11050815

Entringer, G.C., Santos, P.H.A.D., Vettorazzi, J.C.F., Cunha, K.S. da, Pereira, M.G., 2014. Correlação e análise de trilha para componentes de produção de milho superdoce. Revista Ceres 61, 356–361. https://doi.org/10.1590/S0034-737X2014000300009

Falconer, D.S., Mackay, T.F., 1997. Introduction to quantitative genetics, 4th ed. Londres.

FAO, F. and A.O. of the U.N., 2022. Production/Yield quantities of oats in world [WWW Document]. https://www.fao.org/faostat/en/#data/QCL/visualize.

Faraway, J., 2016. faraway: Functions and datasets for books by Julian Faraway [WWW Document]. Cran.R-project. URL https://cran.r-project.org/package=faraway

Ferreira, D.F., 2009. Estatística Básica, 2nd ed. Lavras.

Filho, D.B.F., Júnior, J.A. da S., 2009. Desvendando os mistérios do coeficiente de correlação de Pearson (r). Revista Política Hoje 18, 115–146.

Floss, E.L., Palhano, A.L., Filho, C.V.S., Premazzi, L.M., 2007. Crescimento, produtividade, caracterização e composição química da aveia branca. Acta Sci 29, 1–7.

Fontaneli, Renato Serena, Fontaneli, Roberto Serena, Santos, H.P. dos, Junior, A. do N., Minella, E., Caierão, E., 2009. Rendimento e valor nutritivo de cereais de inverno de duplo propósito: forragem verde e silagem ou grãos. Revista Brasileira de Zootecnica 38, 2126–2120.

Fox, J., Weisberg, S., 2019. An R companion to applied regression, 3rd ed. Sage.

Gondim, T.C. de O., Rocha, V.S., Sediyama, C.S., Miranda, G.V., 2008. Análise de trilha para componentes do rendimento e caracteres agronômicos de trigo sob desfolha. Pesqui Agropecu Bras 43, 487–493. https://doi.org/10.1590/S0100-204X2008000400007

Guimarães, G.N.H.G., Dadalto, J.O., Figueiredo, L., Prado, L.F., Morais, M.F.S., Gaspar, R., Coimbra, C.N., Quinones, E.M., Diniz, R., Maccagnan, P., 2021. Aveia e saúde humana: uma revisão bibliográfica. Revista Higeia@-Revista Científica de Saúde 3, 1–8.

Gutkoski, L.C., Teixeira, D.M. de F., Durigon, A., Ganzer, A.G., Bertolin, T.E., Colla, L.M., 2009. Influência dos teores de aveia e de gordura nas características tecnológicas e funcionais de bolos. Ciência e Tecnologia de Alimentos 29, 254–261. https://doi.org/10.1590/S0101-20612009000200003

Hair, J.F., Black, W.C., Babin, B., Anderson, R., Tatham, R.L., 2009. Análise multivariada de dados.

Harrell, F.E., 2021. Frank E Harrell [WWW Document]. CRAR.R-project. URL https://cran.r-project.org/package=Hmisc

Holland, J.B., Bjornstad, A., Frey, K.J., Gullord, M., Wesenberg, D.M., Buraas, T., 2000. Recurrent selection in oat for adaptation to diverse environments. Euphytica 113, 195–205.

Kavalco, S.A.F., Figueiredo, R., Groli, E.L., Zimmer, C.M., Baretta, D., Tessmann, E.W., Magalhães Júnior, A.M. de, Oliveira, A.C. de, 2014. Análise de trilha em genótipos de trigo submetidos ao estresse por encharcamento. Semin Cienc Agrar 35, 1683–1696.

Klein, L.A., Marchioro, V.S., Souza, V.Q. de, Meira, D., Meier, C., 2019. Dissimilaridade genética entre genótipos de aveia preta. Revista Brasileira de Iniciação Científica 6, 114–115.

Korkmaz, S., Goksuluk, D., Zararsiz, G., 2014. MVN: An R package for assessing multivariate normality. R J 6, 151–162.

Kraisig, A.R., da Silva, J.A.G., Carvalho, I.R., Mamann, Â.T.W. de, Corso, J.S., Norbert, L., 2020. Time of nitrogen supply in yield, industrial and chemical quality of oats grains. Revista Brasileira de Engenharia Agrícola e Ambiental 24, 700–706.

Leite, J.D.B., Federizzi, L., Bergamaschi, H., 2012. Mudanças climáticas e seus possíveis impactos aos sistemas agrícolas no Sul do Brasil. Agrária 7, 337. https://doi.org/10.5039/agraria.v5i3a1239

Lorencetti, C., Carvalho, F.I.F. de, Marchioro, V.S., Benin, G., Oliveira, A.C. de, Foss, E.L., 2004. Implicações da aplicação de fungicida na adaptabilidade e estabilidade de rendimento de grãos em aveia branca. Ciência Rural 34, 693–700. https://doi.org/10.1590/S0103-84782004000300007

Lúcio, A.D., Storck, L., Krause, W., Gonçalves, R.Q., Nied, A.H., 2013. Relações entre os caracteres de maracujazeiro-azedo. Ciência Rural 43, 225–232. https://doi.org/10.1590/S0103-84782013000200006

Mantai, R.D., Silva, J.A.G. da, Arenhardt, E.G., Heck, T.G., Sausen, A.T.Z.R., Krüger, C.A.M.B., Bandeira, L.M., 2016. The Dynamics of Relation Oat Panicle with Grain Yield by Nitrogen. Am J Plant Sci 7, 17–27.

Mantai, R.D., Silva, J.A.G. da, Binelo, M.O., Sausen, A.T.Z.R., Rossi, D.S., Corso, J.S., 2020a. Nitrogen management in the relationships between oat inflorescence components and productivity. Revista Brasileira de Engenharia Agrícola e Ambiental 24, 385–393. https://doi.org/10.1590/1807-1929/agriambi.v24n6p385-393

Mantai, R.D., Silva, J.A.G. da, Marolli, A., Mamann, Â.T.W. de, Sawicki, S., Krüger, C.A.M.B., 2017. Simulation of oat development cycle by photoperiod and temperature. Revista Brasileira de Engenharia Agrícola e Ambiental 21, 3–8. https://doi.org/10.1590/1807-1929/agriambi.v21n1p3-8

Mantai, R.D., Silva, J.A.G. da, Scremin, O.B., Carvalho, I.R., Magano, D.A., Fachinetto, J.M., Lautenchleger, F., Rosa, J.A. da, Peter, C.L., Berlezi, J.D., Babeski, C.M., 2020b. Nitrogen levels in oat grains and its relation to productivity. Genetics and Molecular Research 19, 1–13.

Marchioro, V.S., Carvalho, F.I.F., Oliveira, A.C. de, Kurek, A.J., Lorencetti, C., Silva, J.A.G. da, Cargnin, A., 2001. Estratégias para a modificação do potencial de rendimento de grãos em genótipos de aveia: época de semeadura e aplicação de fungicida. Revista Brasileira de Agrociência 7, 33–36.

Martinelli, J.A., 2003. Manejo integrado de doenças da aveia. Fitopatol Bras 28, 98–101.

Meira, D., Meier, C., Olivoto, T., Follmann, D.N., Rigatti, A., Lunkes, A., Marchioro, V.S., Souza, V.Q. de, 2019a. Multivariate analysis revealed genetic divergence and promising traits for indirect selection in black oat. Revista Brasileira de Ciências Agrárias 14, 1–7. https://doi.org/10.5039/agraria.v14i4a6514

Meira, D., Meier, C., Olivoto, T., Nardino, M., Rigatti, A., Klein, L.A., Caron, B.O., Marchioro, V.S., Souza, V.Q. de, 2019b. Phenotypic variance of black oat growing in crop seasons reveals genetic effects predominance. An Acad Bras Cienc 91. https://doi.org/10.1590/0001-3765201920180036

Montgomery, D.C., Peck, E.A., 1982. Introduction to linear regression analysis. New York.

Montgomery, D.C., Peck, E.A., Vining, G., 2012. Introduction to linear regression analysis, 5th ed. Hoboken.

Nardino, M., Souza, V.Q. de, Baretta, D., Konflanz, V.A., Carvalho, I.R., Follmann, D.N., Caron, B.O., 2016. Association of secondary traits with yield in maize F 1 's. Ciência Rural 46, 776–782. https://doi.org/10.1590/0103-8478cr20150253

Nogueira, A.P.O., Sediyama, T., Souza, L.B. de, Hamawaki, O.T., Cruz, C.D., Pereira, D.G., Matsuo, É., 2012. Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. Bioscience Journal 28, 877–888.

Olivoto, T., de Souza, V.Q., Nardino, M., Carvalho, I.R., Ferrari, M., de Pelegrin, A.J., Szareski, V.J., Schmidt, D., 2017. Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. Agron J 109, 131–142. https://doi.org/https://doi.org/10.2134/agronj2016.04.0196

R Core Team, 2021. R: A Laguage and Environment for Statistical Computing.

Rodrigues, G.B., Marim, B.G., Silva, D.J.H. da, Mattedi, A.P., Almeida, V. de S., 2010. Análise de trilha de componentes de produção primários e secundários em tomateiro do grupo salada. Pesqui Agropecu Bras 45, 155–162.

Salla, V.P., Danner, M.A., Citadin, I., Sasso, S.A.Z., Donazzolo, J., Gil, B.V., 2015. Análise de trilha em caracteres de frutos de jabuticabeira. Pesqui Agropecu Bras 50, 218–223.

Sari, B.G., Lúcio, A.D., Olivoto, T., Krysczun, D.K., Tischler, A.L., Drebes, L., 2018. Interference of sample size on multicollinearity diagnosis in path analysis. Pesqui Agropecu Bras 53, 769–773.

Sari, B.G., Lúcio, A.D., Santana, C.S., Lopes, S.J., 2017. Linear relationships between cherry tomato traits. Ciência Rural 47, 1–7.

Silva, A.R. da, Malafaia, G., Menezes, I.P.P., 2017. biotools: an R function to predict spatial gen diversity via an individual-based approach. Genetics and Molecular Research 16, 1–6.

Stevenson, W., 2001. Estatística aplicada a administração, 1ª. ed. São Paulo.

Tedesco, M.J., Gianello, C., Bissani, C.A., Bohnen, H., Volkweiss, S.J., 1995. Análise de solo, plantas e outros materiais, 2nd ed. Porto Alegre.

Tierney, L., 2021. tkrplot: TK Rplot [WWW Document]. CRAR.R-project. URL https://cran.r-project.org/package=tkrplot

Toebe, M., Cargnelutti Filho, A., Lindolfo, S., Lúcio, A.D., 2017. Direct effects on scenarios and types of path analyses in corn hybrids. Genetics and Molecular Research 16, 1–15.

Vencovski, R., Barriga, P., 1992. Genética biométrica no fitomelhoramento. Ribeirão Preto - SP.

Vesohoski, F., Marchioro, V.S., Franco, F. de A., Cantelle, A., 2011. Componentes do rendimento de grãos em trigo e seus efeitos diretos e indiretos na produtividade. Revista Ceres 58, 337–341. https://doi.org/10.1590/S0034-737X2011000300014

Zuffo, A.M., Ribeiro, A.B.M., Bruzi, A.T., Zambiazzi, E.V., Fonseca, W.L., 2018. Correlação e análise de trilha em cultivares de soa cultivadas em diferentes densidades de plantas. Cultura Agronômica 27, 78–90.

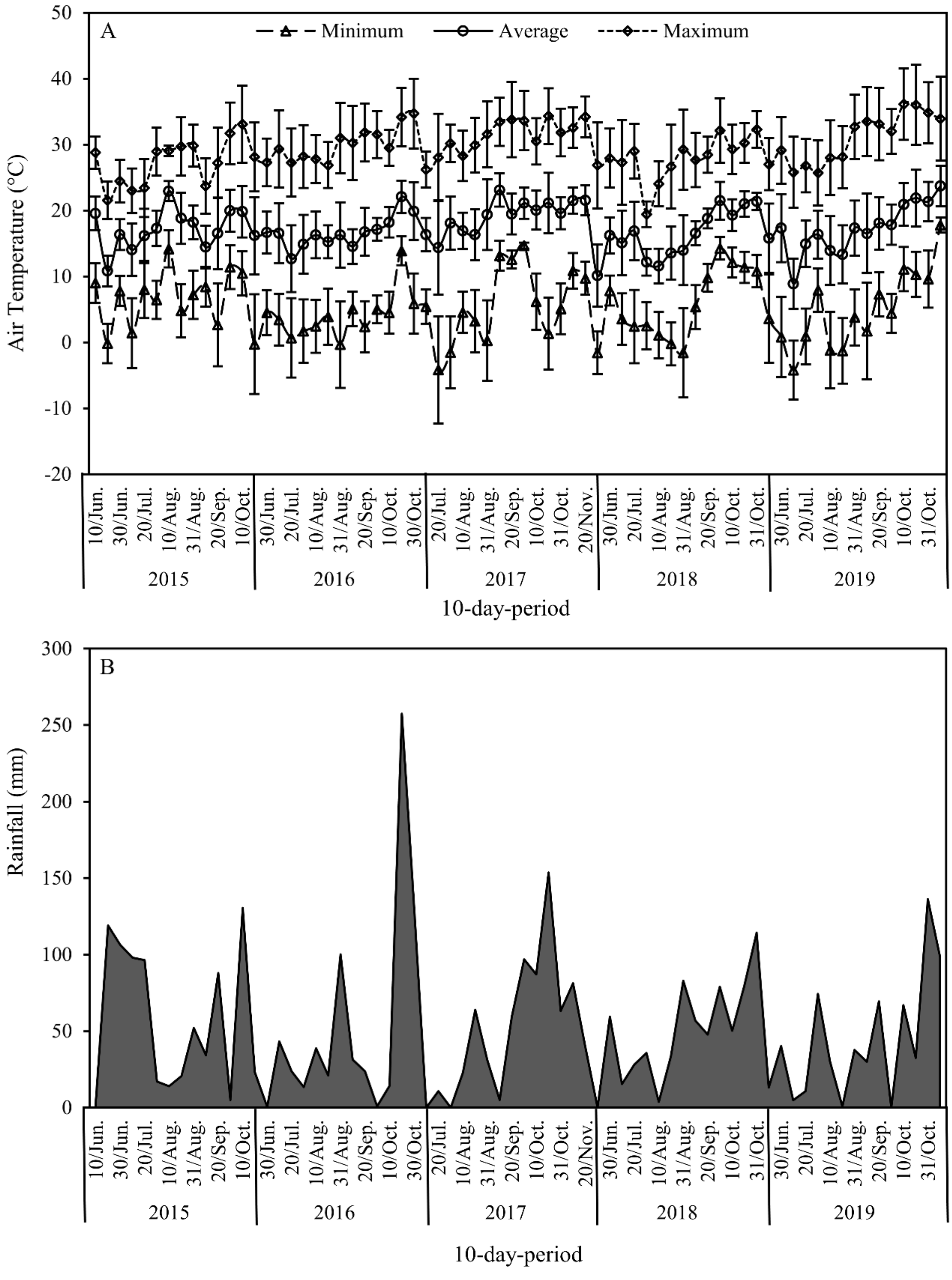


Figure 1 – Air temperature (minimum, average and maximum - A) and accumulated rainfall (B) (10-day periods), during the oat cultivation period (June to October), in five agricultural years: 2015, 2016, 2017, 2018 and 2019.

Table 1 – Pearson correlation coefficients of the explanatory variables with the yield of oats, cultivated with and without fungicide application and in five environments, considering the original and predicted data groups.

	Variables					
Environments	PL	PDM	NSP	NGP	WGP	HI
			Originals with fungicide[(1)]			
1	−0.18*	0.27*	0.17*	0.20*	0.32*	0.41*
2	−0.04^{ns}	0.33*	0.09^{ns}	0.16*	0.35*	0.20*
3	−0.02^{ns}	0.29*	0.04^{ns}	0.34*	0.33*	0.30^{ns}
4	−0.05^{ns}	0.16*	0.02^{ns}	0.10^{ns}	0.20*	0.35*
5	0.13*	0.27*	0.20*	0.16*	0.30*	0.22*
			Originals without fungicide[(2)]			
1	−0.05^{ns}	0.14^{ns}	0.19^{ns}	0.27*	0.17^{ns}	0.18^{ns}
2	0.13^{ns}	0.25*	0.04^{ns}	0.17^{ns}	0.25*	−0.05^{ns}
3	−0.18^{ns}	0.30*	0.06^{ns}	0.27*	0.33*	0.24^{ns}
4	0.08^{ns}	0.07^{ns}	−0.30*	−0.12^{ns}	0.11^{ns}	0.22^{ns}
5	0.04^{ns}	−0.21^{ns}	−0.06^{ns}	−0.07^{ns}	−0.20^{ns}	0.00^{ns}
			Predicted with fungicide[(1)]			
1	−0.03^{ns}	−0.04^{ns}	0.02^{ns}	−0.10^{ns}	−0.06^{ns}	−0.09^{ns}
2	0.05^{ns}	0.06^{ns}	−0.10^{ns}	−0.02^{ns}	0.06^{ns}	0.02^{ns}
3	0.07^{ns}	−0.03^{ns}	−0.07^{ns}	0.01^{ns}	−0.04^{ns}	−0.03^{ns}
4	−0.02^{ns}	−0.02^{ns}	−0.02^{ns}	−0.05^{ns}	−0.03^{ns}	−0.06^{ns}
5	0.05^{ns}	0.09^{ns}	0.19^{ns}	0.07^{ns}	0.08^{ns}	0.008^{ns}
			Predicted without fungicide[(2)]			
1	−0.06^{ns}	0.15^{ns}	0.13^{ns}	0.07^{ns}	0.20^{ns}	0.07^{ns}
2	0.37*	−0.01^{ns}	−0.20^{ns}	−0.14^{ns}	−0.06^{ns}	−0.19^{ns}
3	−0.20^{ns}	0.07^{ns}	0.23^{ns}	0.19^{ns}	0.07^{ns}	−0.01^{ns}
4	0.01^{ns}	0.01^{ns}	−0.17^{ns}	−0.19^{ns}	0.00^{ns}	−0.04^{ns}
5	0.13^{ns}	0.07^{ns}	0.16^{ns}	0.07^{ns}	0.09^{ns}	0.14^{ns}

Yield: PL: panicle length; PDM: panicle dry mass; NSP: number of spikelets per panicle; NGP: number of grains per panicle; GWP: grain weight per panicle; HI: Harvest index. [(1)] number of observations n = 264; [(2)] number of observations n=66.

^{ns}Not significant.

*Significant at 5% error probability.

Table 2 – Variance inflation factor (VIF) for the explanatory variables of yield of oats cultivated with and without fungicide application and in five environments (Env.), considering the original and predicted data groups.

Env.	PL	PDM	NSP	NGP	GWP	HI
			Originals with fungicide			
1	1.090	1038.635	2.127	2.471	1150.754	31.698
2	1.281	1572.106	4.419	5.884	1636.786	19.098
3	1.255	623.835	2.270	2.915	717.227	29.318
4	1.165	831.029	3.188	3.545	916.523	17.253
5	1.358	936.714	3.673	3.940	985.557	23.880
			Originals without fungicide			
1	1.327	504.888	2.656	3.907	553.991	16.958
2	1.329	1232.192	2.627	4.073	1243.946	32.411
3	1.379	247.226	1.959	2.336	307.199	17.931
4	1.928	530.541	2.807	3.084	540.075	16.755
5	1.156	601.320	2.811	2.265	636.528	19.131
			Predicted with fungicide			
1	1.014 (7.02%)	617.944 (40.50%)	2.146 (0.89%)	2.199 (11.02%)	664.356(42.27%)	22.279 (29.72%)
2	1.092 (14.76%)	580.773 (63.06%)	2.864 (35.18%)	3.419 (41.89%)	624.157 (61.87%)	18.570 (2.77%)
3	1.102 (12.20%)	416.381 (33.25%)	2.159 (4.88%)	2.406 (14.48%)	441.674 (38.42%)	26.470 (9.71%)
4	1.098 (5.80%)	613.775 (26.14%)	2.558 (19.78%)	2.737 (22.79%)	646.921 (29.42%)	14.651 (15.08%)
5	1.271 (6.44%)	790.085 (15.65%)	3.210 (12.61%)	3.146 (20.15%)	795.149 (19.32%)	24.049 (0.71%)
			Predicted without fungicide			
1	1.091 (17.84%)	64.861 (87.15%)	1.723 (35.14%)	1.855 (52.52%)	76.580 (86.18%)	11.054 (34.82%)
2	1.190 (10.48%)	419.856 (65.93%)	2.559 (2.58%)	3.339 (18.00%)	453.759 (63.52%)	33.447 (3.20%)
3	1.702 (23.44%)	168.478 (31.85%)	1.998 (1.99%)	2.577 (10.32%)	218.259 (28.95%)	17.549 (2.13%)
4	1.370 (28.98%)	412.581 (22.23%)	2.467 (12.11%)	3.225 (4.57%)	427.842 (20.78%)	15.111 (9.81%)
5	1.060 (8.30%)	416.128 (30.80%)	2.864 (1.87%)	2.052 (9.39%)	442.242 (30.52%)	23.867 (24.75%)

Values in parentheses represent, in percentage, the effects of removing parameters from the mathematical model on the VIF statistics.
Yield: PL: panicle length; PDM: panicle dry mass; NSP: number of spikelets per panicle; NGP: number of grains per panicle; GWP: grain weight per panicle; HI: Harvest index.

Table 3 – Condition number (CN) for the explanatory variables of yield of oats cultivated with and without fungicide application and in five environments, considering the original and predicted data groups.

Environments	With fungicide		Without fungicide	
	Originals	Predicted	Originals	Predicted
1	6845.192	3368.011 (50.80%)	3683.204	322.507 (91.24%)
2	11630.570	3787.780 (67.43%)	8604.037	2639.418 (69.32%)
3	4411.413	2594.084 (41.20%)	1723.691	1227.982 (28.76%)
4	5470.483	3533.938 (35.40%)	3489.785	2591.458 (25.74%)
5	6574.209	5171.948 (21.33%)	3602.747	2456.653 (31.81%)

Values in parentheses represent, in percentage, the effects of removing parameters from the mathematical model on the CN statistics.

Table 4 – Direct and indirect effects of the explanatory variables on the yield of oats cultivated with fungicide application and in five environments (Env.), considering the original (Orig.) and predicted (Pred.) data groups, with the addition of a k value on the diagonal of the X'X matrix of correlation.

	Env 1		Env 2		Env 3		Env 4		Env 5	
Effects	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.
PL										
Direct on Yield	-0.169	-0.025	-0.173	0.060	-0.007	0.082	-0.070	-0.013	0.018	-0.006
Indirect via PDM	0.009	0.001	0.107	0.010	0.018	-0.001	0.010	-0.003	0.056	0.012
Indirect via NSP	0.023	0.006	-0.042	-0.050	-0.085	-0.030	-0.021	0.004	0.055	0.101
Indirect via NGP	0.000	-0.012	-0.028	0.015	0.056	0.022	0.025	-0.008	-0.038	-0.055
Indirect via GWP	0.010	-0.002	0.116	0.013	0.020	-0.011	0.017	0.004	0.049	-0.005
Indirect via HI	-0.043	0.001	-0.007	0.000	-0.018	0.001	-0.003	-0.002	-0.009	0.000
r	-0.18^{*}	-0.03^{ns}	-0.04^{ns}	0.05^{ns}	0.02^{ns}	0.07^{ns}	-0.05^{ns}	-0.02^{ns}	0.13^{*}	0.05^{ns}
PDM										
Direct on Yield	0.066	0.012	0.246	0.047	0.065	-0.005	0.030	-0.010	0.115	0.028
Indirect via PL	-0.022	-0.002	-0.075	0.013	-0.002	0.022	-0.023	-0.004	0.009	-0.003
Indirect via NSP	0.051	0.050	-0.078	-0.098	-0.117	-0.069	-0.033	0.011	0.083	0.170
Indirect via NGP	-0.002	-0.074	-0.064	0.035	0.220	0.065	0.048	-0.029	-0.067	-0.101
Indirect via GWP	0.100	-0.022	0.274	0.058	0.086	-0.042	0.054	0.016	0.106	-0.012
Indirect via HI	0.070	-0.008	0.012	0.000	0.035	-0.001	0.086	-0.006	0.020	0.000
r	0.27^{*}	-0.04^{ns}	0.33^{*}	0.06^{ns}	0.29^{*}	-0.03^{ns}	0.16^{*}	-0.02^{ns}	0.27^{*}	0.09^{ns}
NSP										
Direct on Yield	0.101	0.148	-0.118	-0.203	-0.233	-0.150	-0.080	0.033	0.149	0.319
Indirect via PL	-0.038	-0.001	-0.062	0.015	-0.002	0.016	-0.018	-0.001	0.007	-0.002
Indirect via PDM	0.034	0.004	0.162	0.022	0.033	-0.002	0.012	-0.004	0.064	0.015
Indirect via NGP	-0.003	-0.140	-0.074	0.047	0.223	0.086	0.076	-0.053	-0.095	-0.154
Indirect via GWP	0.050	-0.007	0.180	0.028	0.040	-0.018	0.022	0.005	0.058	-0.007
Indirect via HI	0.016	0.006	0.005	0.000	-0.002	0.003	0.009	0.002	0.010	0.000
r	0.17^{*}	0.02^{ns}	0.09^{ns}	-0.10^{ns}	0.04^{ns}	-0.07^{ns}	0.02ns	-0.02^{ns}	0.20^{*}	0.19^{ns}
NGP										
Direct on Yield	-0.004	-0.194	-0.085	0.058	0.335	0.122	0.092	-0.069	-0.112	-0.189
Indirect via PL	-0.018	-0.001	-0.058	0.015	-0.001	0.015	-0.019	-0.002	0.006	-0.002
Indirect via PDM	0.041	0.004	0.187	0.028	0.043	-0.003	0.015	-0.004	0.069	0.015
Indirect via NSP	0.071	0.107	-0.103	-0.163	-0.155	-0.106	-0.065	0.026	0.127	0.261
Indirect via GWP	0.062	-0.008	0.210	0.035	0.058	-0.023	0.028	0.007	0.063	-0.006
Indirect via HI	0.050	0.001	0.012	0.000	0.040	-0.004	0.045	-0.002	0.014	0.000
r	0.20^{*}	-0.10^{ns}	0.16^{*}	-0.02^{ns}	0.34^{*}	0.01^{ns}	0.10^{ns}	-0.05^{ns}	0.16^{*}	0.07^{ns}
GWP										
Direct on Yield	0.102	-0.022	0.275	0.059	0.088	-0.043	0.055	0.016	0.107	-0.012
Indirect via PL	-0.017	-0.002	-0.073	0.013	-0.002	0.021	-0.022	-0.004	0.008	-0.002
Indirect via PDM	0.066	0.012	0.244	0.046	0.064	-0.005	0.029	-0.010	0.113	0.028
Indirect via NSP	0.049	0.045	-0.077	-0.095	-0.107	-0.062	-0.032	0.011	0.081	0.170
Indirect via NGP	-0.002	-0.070	-0.064	0.034	0.223	0.064	0.047	-0.029	-0.066	-0.099

Indirect via HI	0.118	-0.019	0.024	0.000	0.056	-0.010	0.120	-0.014	0.045	-0.001
r	0.32*	−0.06ns	0.35*	0.06ns	0.33*	−0.04ns	0.20*	−0.03ns	0.30*	0.08ns
					HI					
Direct on Yield	0.314	-0.066	0.112	-0.001	0.120	-0.037	0.287	-0.053	0.170	-0.006
Indirect via PL	0.023	0.001	0.011	0.001	0.001	-0.003	0.001	-0.001	-0.001	0.000
Indirect via PDM	0.015	0.001	0.027	0.005	0.019	0.000	0.009	-0.001	0.013	-0.001
Indirect via NSP	0.005	-0.014	-0.005	-0.002	0.005	0.014	-0.003	-0.001	0.009	0.019
Indirect via NGP	-0.001	0.003	-0.009	0.004	0.111	0.014	0.014	-0.002	-0.009	-0.002
Indirect via GWP	0.038	-0.007	0.059	0.016	0.041	-0.012	0.023	0.004	0.028	-0.002
r	0.41*	−0.09ns	0.20*	0.02ns	0.30ns	−0.03ns	0.35*	−0.06ns	0.22*	0.01ns
R^2	0.227	0.030	0.182	0.029	0.187	0.020	0.127	0.020	0.115	0.049
Residual	0.879	0.985	0.904	0.985	0.901	0.990	0.934	0.990	0.941	0.975
k	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060

Yield: PL: panicle length; PDM: panicle dry mass; NSP: number of spikelets per panicle; NGP: number of grains per panicle; GWP: grain weight per panicle; HI: Harvest index.

Table 5 – Direct and indirect effects of the explanatory variables on the yield of oats cultivated without fungicide application and in five environments, considering the original (Orig.) and predicted (Pred.) data groups, with the addition of a k value on the diagonal of the X'X correlation matrix.

	Env 1		Env 2		Env 3		Env 4		Env 5	
Effects	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.	Orig.	Pred.
					PL					
Direct on Yield	-0.209	-0.088	0.005	0.446	-0.265	-0.383	0.093	0.028	0.117	0.144
Indirect via PDM	-0.048	-0.026	0.068	0.021	0.029	0.035	0.027	0.024	-0.063	-0.003
Indirect via NSP	0.034	0.006	-0.050	-0.034	0.009	0.137	-0.109	-0.015	0.010	-0.016
Indirect via NGP	0.172	-0.009	0.037	-0.012	0.025	0.009	0.020	-0.038	-0.001	0.004
Indirect via GWP	0.003	0.067	0.058	-0.045	0.041	0.003	0.059	0.012	-0.033	-0.010
Indirect via HI	0.012	-0.002	0.007	-0.030	-0.003	0.024	-0.020	-0.005	-0.001	0.000
r	-0.05^{ns}	-0.06^{ns}	0.13^{ns}	0.37^{*}	-0.18^{ns}	-0.20^{ns}	0.08^{*}	0.01^{ns}	0.04^{ns}	0.13^{ns}
					PDM					
Direct on Yield	-0.146	-0.161	0.139	0.065	0.098	0.096	0.044	0.087	-0.186	-0.026
Indirect via PL	-0.068	-0.014	0.002	0.142	-0.079	-0.138	0.058	0.008	0.040	0.018
Indirect via NSP	0.040	0.015	-0.102	-0.078	0.007	0.101	-0.163	-0.037	0.040	0.198
Indirect via NGP	0.273	-0.001	0.081	-0.025	0.076	0.028	0.029	-0.112	-0.003	-0.039
Indirect via GWP	0.008	0.324	0.122	-0.124	0.185	0.012	0.101	0.055	-0.099	-0.083
Indirect via HI	0.041	0.000	0.004	0.006	0.005	-0.032	0.003	0.002	0.006	0.005
r	0.14^{ns}	0.15^{ns}	0.25^{*}	-0.01^{ns}	0.30^{*}	0.07^{ns}	0.07^{*}	0.01^{ns}	-0.21^{ns}	0.07^{ns}
					NSP					
Direct on Yield	0.073	0.069	-0.186	-0.240	0.022	0.312	-0.379	-0.082	0.078	0.380
Indirect via PL	-0.096	-0.008	0.001	0.063	-0.111	-0.168	0.027	0.005	0.016	-0.006
Indirect via PDM	-0.079	-0.034	0.076	0.021	0.033	0.031	0.019	0.039	-0.094	-0.013
Indirect via NGP	0.291	0.032	0.086	-0.032	0.067	0.025	0.044	-0.158	-0.005	-0.054
Indirect via GWP	0.004	0.064	0.066	-0.035	0.051	0.003	0.040	0.025	-0.044	-0.034
Indirect via HI	-0.004	0.002	0.005	0.033	-0.003	0.011	-0.024	0.004	-0.019	-0.132
r	0.19^{ns}	0.13^{ns}	0.04^{ns}	-0.20^{ns}	0.06^{ns}	0.23^{ns}	-0.30^{*}	-0.17^{ns}	-0.06^{ns}	0.16^{ns}
					NGP					
Direct on Yield	0.376	0.058	0.110	-0.043	0.123	0.045	0.057	-0.209	-0.007	-0.079
Indirect via PL	-0.096	0.014	0.002	0.124	-0.054	-0.075	0.033	0.005	0.010	-0.007
Indirect via PDM	-0.106	0.004	0.102	0.038	0.061	0.061	0.023	0.047	-0.080	-0.013
Indirect via NSP	0.057	0.038	-0.145	-0.175	0.012	0.177	-0.293	-0.062	0.058	0.262
Indirect via GWP	0.006	-0.056	0.090	-0.073	0.120	0.008	0.053	0.032	-0.040	-0.038
Indirect via HI	0.015	0.007	0.002	-0.003	0.005	-0.033	0.002	0.011	-0.008	-0.049
r	0.27^{*}	0.07^{ns}	0.17^{ns}	-0.14^{ns}	0.27^{*}	0.19^{ns}	-0.12^{ns}	-0.19^{ns}	-0.07^{ns}	0.07^{ns}
					GWP					
Direct on Yield	0.008	0.351	0.124	-0.129	0.190	0.012	0.103	0.056	-0.100	-0.085
Indirect via PL	-0.067	-0.017	0.002	0.157	-0.057	-0.090	0.054	0.006	0.038	0.016
Indirect via PDM	-0.144	-0.149	0.137	0.063	0.096	0.093	0.043	0.086	-0.183	-0.025
Indirect via NSP	0.037	0.013	-0.100	-0.064	0.006	0.084	-0.148	-0.037	0.035	0.155
Indirect via NGP	0.263	-0.009	0.081	-0.025	0.077	0.029	0.029	-0.121	-0.003	-0.035

Indirect via HI	0.070	-0.007	-0.004	-0.057	0.008	-0.058	0.027	0.008	0.017	0.072
r	0.17^{ns}	0.20^{ns}	0.25*	−0.06^{ns}	0.33*	0.07^{ns}	0.11^{ns}	0.00^{ns}	−0.20^{ns}	0.09^{ns}
					HI					
Direct on Yield	0.184	-0.020	-0.054	-0.232	0.015	-0.107	0.137	0.031	0.064	0.296
Indirect via PL	-0.014	-0.007	-0.001	0.059	0.053	0.085	-0.013	-0.005	-0.001	0.000
Indirect via PDM	-0.032	0.000	-0.011	-0.002	0.034	0.029	0.001	0.005	-0.017	0.000
Indirect via NSP	-0.002	-0.006	0.017	0.034	-0.004	-0.032	0.066	-0.010	-0.023	-0.170
Indirect via NGP	0.031	-0.021	-0.004	-0.001	0.039	0.014	0.001	-0.074	0.001	0.013
Indirect via GWP	0.003	0.129	0.010	-0.032	0.102	0.007	0.020	0.014	-0.026	-0.020
r	0.18^{ns}	0.07^{ns}	−0.05^{ns}	−0.19^{ns}	0.24^{ns}	−0.01^{ns}	0.22^{ns}	−0.04^{ns}	0.00^{ns}	0.14^{ns}
R^2	0.142	0.061	0.080	0.272	0.178	0.167	0.157	0.053	0.060	0.105
Residual	0.926	0.969	0.959	0.854	0.907	0.913	0.918	0.973	0.969	0.946
k	0.060	0.050	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060

Yield: PL: panicle length; PDM: panicle dry mass; NSP: number of spikelets per panicle; NGP: number of grains per panicle; GWP: grain weight per panicle; HI: Harvest index.

Table 6 – Pearson correlation coefficients between the path coefficients (direct and indirect effects) obtained between the data groups (original and predicted), for the scenarios with fungicide application and without fungicide application (n=30).

Variables	With fungicide	Without fungicide
PL	0.18^{ns}	0.51^{*}
PDM	0.51^{*}	0.28^{ns}
NSP	0.62^{*}	0.40^{*}
NGP	0.55^{*}	0.29^{ns}
GWP	0.48*	0.23^{ns}
HI	−0.78*	0.33^{ns}

PL: panicle length; PDM: panicle dry mass; NSP: number of spikelets per panicle; NGP: number of grains per panicle; GWP: grain weight per panicle; HI: Harvest index.

4 CONSIDERAÇÕES FINAIS

Nos programas de melhoramento genético são amplamente utilizadas técnicas estatísticas visando compreender as relações lineares de caracteres de interesse, entre as quais destacam-se a correlação linear de Pearson e análise de trilha. Em um primeiro momento a correlação linear é empregada a fim avaliar a existência ou não de relação entre um par de caracteres. Porém, não permite determinar como os outros caracteres podem estar interferindo nesta relação (SARI et al., 2018). Assim, quando são mensurados mais de dois caracteres de interesse, as possíveis relações diretas e indiretas de caracteres correlacionados pode ser estudada por meio da análise de trilha (RODRIGUES et al., 2010; SARI et al., 2017).

Neste estudo foram utilizadas duas abordagens referentes ao estudo das relações lineares e organizadas em dois artigos. No artigo I foram analisados os efeitos do não atendimento dos pressupostos estatísticos e da remoção dos parâmetros do modelo matemático nos resultados de análise de trilha, análise de trilha com eliminação de variável e análise de trilha sob multicolinearidade, em ensaios com a aveia cultivada em cinco ambientes. O grupo de dados com remoção dos parâmetros do modelo matemático foi designado de preditos e o grupo de dados com manutenção dos parâmetros foi designado de original.

Ao analisar inicialmente os dados quanto a normalidade multivariada e multicolinearidade, foi verificado violação dos pressupostos estatísticos, em todos os ambientes e grupo de dados. Assim, como estratégia inicial, foi utilizado a transformação dos dados por meio da família Box-Cox (BOX; COX, 1964), a qual não foi eficiente para contornar estes problemas. Como estratégias alternativas para superar a violação dos pressupostos deve-se considerar a exclusão das variáveis que geram viés na análise. A eliminação da variável massa seca de panícula foi eficiente para contornar os problemas relacionados a multicolinearidade.

Ao analisar a correlação linear simples entre os componentes de rendimento da aveia, nos cinco ambientes de cultivo e em cada grupo de dados (originais e preditos), foram observados elevados índices de correlação significativa. De modo geral todas as variáveis apresentaram correlação com a produtividade, porém com valores de baixa magnitude, sobretudo para o comprimento de panícula. A massa de panícula apresentou elevada correlação com o número de grãos por panícula, massa de grãos por panícula e número de espiguetas por panícula, aspecto que se justifica, pois, estas variáveis são componentes de panícula. A massa de grãos por panícula apresentou correlação com o número de grãos por panícula, indicando que panículas com elevado número de grãos resultam em grãos mais pesados. Esses resultados corroboram aos verificados a literatura (BENIN et al., 2003b; CAIERÃO et al., 2001; DUMLUPINAR et al., 2011; KAZIU; KASHTA; CELAMI, 2019; KLEIN et al., 2019).

Ao analisar a influência da remoção dos parâmetros do modelo matemático sob a correlação linear foram verificadas diversas situações. No geral foi verificado manutenção no padrão de resposta em 79% das combinações, com direção inversa na resposta em 4,8% das combinações e alteração superior a 50% na magnitude do valor dos coeficientes em 16,2% das combinações. Além disso, foi observado manutenção no padrão de resposta nos ambientes 2, 3, 4 e 5, com pequenas alterações na magnitude dos coeficientes. Em contraponto, as maiores alterações foram obtidas para o ambiente 1, sendo verificado 19,05% de inversão na direção das combinações e alteração na magnitude dos coeficientes (>50%), com manutenção no sinal em 66,77% das combinações. Esses resultados estão relacionados a ocorrência de condições meteorológicas extremas em períodos críticos da cultura, que impactaram no crescimento, desenvolvimento e expressão da produtividade.

Considerando que 87,8% das variáveis apresentaram correlações significativas com a produtividade de grãos em pelo menos um ambiente e que a transformação dos dados não foi eficiente em contornar a violação dos pressupostos, optou-se por utilizar os dados sem transformação e proceder diferentes tipos de análise de trilha, conforme interesse e necessidade. Desse modo, inicialmente foi realizado novo diagnóstico de multicolinearidade, somente entre o conjunto de variáveis explicativas, o qual evidenciou a ocorrência de multicolinearidade e indicou que a eliminação da variável massa de panícula é suficiente para contornar a violação do pressuposto.

A ocorrência multicolinearidade e a necessidade de eliminar variáveis para contornar o viés relacionado a essa situação são recorrentes na literatura (MEIRA et al., 2019a; OLIVOTO et al., 2017a; SALLA et al., 2015; SARI et al., 2017, 2018; TOEBE et al., 2017a, 2017b). Por outro lado, estudos sugerem que a variável massa de panícula tem elevada correlação com a produtividade de grãos e boa perspectiva par seleção indireta (CAIERÃO et al., 2001; CAIERÃO; CARVALHO; FLOSS, 2006; MANTAI et al., 2020). Assim, optou-se também por proceder a análise de trilha sob multicolinearidade (sem eliminação de variáveis).

A nova abordagem proposta para a análise de trilha promove alterações na capacidade explicativa da variância na produtividade de grãos pelas variáveis preditoras, independentemente do ambiente, do cenário analisado e da característica da técnica (com eliminação de variável, sob multicolinearidade e sem considerar multicolinearidade). Estes resultados mantiveram-se quando analisadas as interrelações entre caracteres, isto é, houve influência sob a direção e magnitude dos coeficientes de trilha obtidos.

Ao interpretar os resultados das análises de trilha foi observado ausência de relação direta ou indireta das variáveis comprimento de panícula e número de espiguetas por panícula, com a produtividade de grãos. Esses resultados se assemelham aos verificados na literatura, os quais indicam que massa de panícula, massa de grãos por panícula e número de grãos por panícula exercem influência sob a produtividade de grãos da aveia (BENIN et al., 2003b; BIBI et al., 2012; MORADI; A; ARZANI, 2005).

Considerando as informações geradas até o momento e que pesquisas evidenciaram que a ocorrência de doenças foliares, como a ferrugem da folha, exercem influências sob o desempenho produtivo dos materiais genéticos (BENIN et al., 2003a, 2005; MARTINELLI, 2003), no artigo II, estudou-se se as implicações da remoção dos parâmetros do modelo matemático nas relações lineares dos componentes de rendimento da aveia se mantêm em cenários contrastantes, com e sem aplicação de fungicida.

Ao analisar as correlações entre os componentes de rendimento e a produtividade foi observado significância estatística para 76,7% dos coeficientes, para o grupo de dados originais com aplicação de fungicida. Para o grupo de dados originais sem aplicação de fungicida, foi obtido significância estatística em 23,3% dos coeficientes. Em contraponto, para o grupo de dados preditos foram observados os menores índices de significância estatística (3,3%), independentemente do cenário.

Ao comparar as estimativas dos coeficientes de correlação obtidos de cada par de variável com a produtividade de grãos nos cinco ambientes de cultivo, considerando os diferentes cenários e grupos de dados, foram verificadas alterações na direção e magnitude dos coeficientes. De modo geral o padrão de resposta foi mantido em 11,70% das combinações, sendo verificado direção inversa em 45% e alteração superior a 50% do valor absoluto de coeficiente, com manutenção do sinal em 43,3% das combinações.

Os diagnósticos de multicolinearidade indicaram violação do pressuposto estatístico para todos os ambientes de cultivo e grupo de dados, sendo as variáveis correlacionadas massa de panícula, massa de grãos por panícula e índice de colheita da panícula. Quando identificada a ocorrência de multicolinearidade no grupo de variáveis explicativas, duas estratégias podem ser adotadas par contornar viés nas análises, sendo: a eliminação de variáveis correlacionadas ou a utilização da análise de trilha sob multicolinearidade. Considerando a importância dessas variáveis para predição do rendimento da cultura, optou-se em proceder a análise de trilha em crista.

Para o cenário com aplicação de fungicida, a remoção dos parâmetros do modelo resultou em alteração média de 82,10% na capacidade explicativa dos caracteres, em relação a variância na produtividade. Além disso foram verificadas alterações na direção dos efeitos diretos em 63,3% das combinações, magnitude (>50%) em 43,30% e manutenção no padrão de resposta em 3,30% das situações. Para os efeitos indiretos, foi verificado alteração de 48% na direção dos coeficientes, magnitude (>50%) 44,67% dos coeficientes e manutenção no padrão de resposta em 7,33% das combinações.

Para o cenário sem aplicação de fungicida, a remoção dos parâmetros do modelo também ocasionou alterações na capacidade explicativa dos caracteres preditores. Ainda foram observadas alterações de13,33% na direção dos efeitos diretos, alteração na magnitude dos coeficientes (>50%) em 56,67% das combinações e manutenção no padrão de resposta em 30% das situações. Para os efeitos indiretos foi verificado de 28% de mudança na direção dos coeficientes de trilha, alteração superior a 50% no valor absoluto dos coeficientes em 57,30% e manutenção no padrão de resposta em 24,67% das combinações.

Com bases nas informações geradas neste estudo, a remoção dos parâmetros do modelo matemático e a estratificação de ambientes e cenários são estratégias que devem ser consideradas durante o uso da análise de trilha, pois resulta na obtenção de coeficientes de trilha que evidenciam a real relação entre as variáveis estudas e livres de efeitos oriundos de tratamento e delineamento experimental.

Recomenda-se o uso da nova abordagem proposta para análise de trilha, para situações em que as variáveis foram mensuradas em experimentos que contenham tratamentos e ou redes de ensaios em que o delineamento experimental não é mantido nos multiambientes. Situações que demandariam ao pesquisador proceder para cada ambiente e tratamento uma nova análise de trilha e interpretá-la de forma isolada.

Além disso, novas pesquisas devem ser desenvolvidas com o intuito de estudar as implicações e a viabilidade da remoção dos parâmetros do modelo matemático de outras técnicas estatísticas multivariadas, a fim de tornar os resultados fidedignos e que ampliem a abrangência das informações geradas.

5 CONCLUSÕES GERAIS

A ocorrência de multicolinearidade no conjunto de variáveis explicativas resulta na obtenção de coeficientes de trilha com magnitudes que ultrapassam a unidade e sem interpretação biológica, independentemente do grupo de dados analisados, com manutenção ou remoção dos parâmetros do modelo matemático.

As magnitudes dos valores absolutos das estatísticas número de condição e fator de inflação da variância são influenciadas pela remoção dos parâmetros do modelo matemático, independentemente do ambiente de cultivo e do cenário agrícola (com e sem aplicação e fungicida).

As técnicas eliminação de variáveis e análise de trilha em crista são alternativas eficientes para contornar a ocorrência de multicolinearidade entre as variáveis explicativas, inclusive para situações em que a análise de trilha é procedida com remoção dos parâmetros do modelo matemático.

A remoção dos parâmetros do modelo matemático modifica os coeficientes de trilha, com alterações médias de 10,5% e 13,3% na direção das associações e 24,7% e 23,0% na magnitude dos efeitos diretos e indiretos, respectivamente, independentemente do tipo de análise de trilha procedida.

As relações lineares entre os caracteres de rendimento da aveia são influenciadas pela remoção dos parâmetros do modelo matemático, com alterações de 56,7% e 33,3% na direção e 40,0% e 46,7% na magnitude dos coeficientes de correlação, para os cenários com aplicação de fungicida e sem aplicação de fungicida.

A remoção dos parâmetros do modelo matemático implica em alterações na direção e na magnitude dos coeficientes de trilha, independentemente do cenário com aplicação de fungicida e sem aplicação de fungicida.

No cenário com aplicação de fungicida, a remoção dos parâmetros do modelo matemático resultou em alterações médias de 53,3% e 48,0% na direção e 43,3% e 44,7% na magnitude dos efeitos diretos e indiretos, respectivamente.

Para o cenário sem aplicação de fungicida, a remoção dos parâmetros do modelo matemático implicou em alterações médias de 13.3% e 28.0% na direção e 56.7% e 24.7% na magnitude dos coeficientes diretos e indiretos, respectivamente.

REFERÊNCIAS

ACHLEITNER, Andreas *et al*. Genetic diversity among oat varieties of worldwide origin and associations of AFLP markes with quantitative traits. **Theoretical and Applied Genetics**, [*S. l*], v. 117, p. 1041–1053, 2008. Disponível em: https://link.springer.com/article/10.1007/s00122-008-0843-y. Acesso em: 15 jul. 2022.

ALIYU, L.; AHMED, M. K.; MAGAJI, M. D. Correlation and multiple regression analysis between morphological characters and components of yield in peper (*Capsicum annuum* L.). **Crop Research**, [*S. l.*], v. 19, p. 318–323, 2000. Disponível em: https://www.cabdirect.org/cabdirect/abstract/20001612157. Acesso em: 07 maio. 2022.

AZEVEDO, Camila Ferreira *et al*. Updated knowledge in the estimation of genetics parameters: a Bayesian approach in white oat (*Avena sativa* L.). **Euphytica**, [*S. l.*], v. 218, n. 4, p. 43, []p, 2022. Disponível em: https://link.springer.com/article/10.1007/s10681-022-02995-0. Acesso em: 05 jul. 2022.

BELLO, Basgir Omolaran *et al*. Correlation and path coefficient analysis of yield and agronomic characters among open pollinated maize varieties and their F1 hybrids in a diallel cross. **African Journal of Biotechnology**, [*S. l.*], v. 9, n. 18, p. 2633–2639, 3 maio 2010. Disponível em: http://www.academicjournals.org/AJB. Acesso em: 04 jul. 2022.

BENIN, Giovani *et al*. Implicações do ambiente sobre o rendimento de grãos em aveia e suas influências sobre estimativas de parâmetros genéticos. **Revista Brasileira de Agrociência**, [*S. l.*], v. 9, n. 3, p. 207–214, 2003a. Disponível em: https://www.researchgate.net/publication/263696548. Acesso em: 04 jul. 2022.

BENIN, Giovani *et al*. Estimativas de correlações e coeficientes de trilha como critérios de seleção para rendimento de grãos em aveia. **Revista Brasileira de Agrociência**, [*S. l.*], v. 9, n. 1, p. 09–16, 2003b.

BENIN, Giovani *et al*. Adaptabilidade e estabilidade em aveia em ambientes estratificados. **Ciência Rural**, [*S. l.*], v. 35, n. 2, p. 295–302, 2005. Disponível em: https://doi.org/10.1590/S0103-84782005000200008. Acesso em: 15 jul. 2022.

BIBI, Amir *et al*. Genetic characterization and inheritance studies of oats (*Avena sativa* L.) for green fodder yield. **International Journal of Biology, Pharmacy and Allied Sciences**, [*S. l.*], v. 1, p. 450–460, 1 abr. 2012. Disponível em: https://www.researchgate.net/publication/279651387. Acesso em: 15 jul. 2022.

BORTOLINI, Patrícia Cambrussi; MORAES, Aníbal de; CARVALHO, Paulo César de Faccio. Produção de forragem e de grãos de aveia branca sob pastejo. **Revista Brasileira de Zootecnia**, [*S. l.*], v. 34, n. 6, p. 2192–2199, dez. 2005. Disponível em: https://doi.org/10.1590/S1516-35982005000700005. Acesso em: 30 jun. 2022.

BOX, George Edward Pelham; COX, Davis Roxbee. An Analysis of Transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, [*S. l.*], v. 26, n. 2, p. 211–243, 1964. Disponível em: https://www.jstor.org/stable/2984418. Acesso em: 10 maio. 2022.

CAIERÃO, Eduardo *et al*. Seleção indireta em aveia para o incremento no rendimento de grãos. **Ciência Rural**, [*S. l.*], v. 31, n. 2, p. 231–236, abr. 2001. Disponível em: https://doi.org/10.1590/S0103-84782001000200007. Acesso em: 30 maio. 2022.

CAIERÃO, Eduardo; CARVALHO, Fernando Irajá Félix de; FLOSS, Elmar Luis. Seleção indireta para o incremento do rendimento de grãos em aveia. **Ciência Rural**, [*S. l.*], v. 36, n. 4, p. 1126–1131, ago. 2006. Disponível em: https://doi.org/10.1590/S0103-84782006000400013. Acesso em: 29 maio. 2022.

CARVALHO, Ivan Ricardo *et al*. Path analysis of grain yield associated characters in Brazilians wheat genotypes (*Triticum aestivum* L.). **Australian Journal of Crop Science**, [*S. l.*], v. 11, n. 11, p. 1406–1410, 2017. Disponível em: https://www.researchgate.net/publication/321672385. Acesso em: 01 ago. 2022.

CASTRO, Gustavo Spadotti Amaral; COSTA, Claudio Hideo Martins da; NETO, Jayme Ferrari. Ecofisiologia da aveia branca. **Scientia Agraria Paranaensis**, [*S. l.*], v. 11, n. 3, p. 1–15, 2012. Disponível em; https://doi.org/10.18188/sap.v11i3.4808. Acesso em: 10 maio, 2022.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira de grãos - Safra 2021/2022**. 11. ed. Brasília, v. 9, n.11, 2022. Disponível em: https://www.conab.gov.br/info-agro/safras/graos. Acesso em: 10 ago. 2022.

COUTO, Maria Rubia Machado *et al*. Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, [*S. l.*], v. 39, n. 6, p. 1701–1707, 2009. Disponível em: https://doi.org/10.1590/S0103-84782009005000110. Acesso em: 10 maio. 2022.

CRUZ, Cosme Damião; REGAZZI, Adair José; CARNEIRO, Pedro Crescêncio Souza **Modelos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa, MG: UFV, 2012. 514 p. ISBN 978-85-7269-433-9.

DUMLUPINAR, Ziya *et al*. Evaluation of Turkish oat landraces based on grain yield, yield components and some quality traits. **Turkish Journal of Field Crops**, [*S. l.*], v. 16, n. 2, p. 190–196, 2011. Disponível em: https://www.researchgate.net/publication/257930629. Acesso em: 01 ago. 2022.

FALCONER, Douglas Scott; MACKAY, Trudy Frances Charlene. **Introduction to quantitative genetics**. 4. Ed. EN:Londres, 1997. 464 p.

FAO – Food and Agriculture Organization of the United Nations. **Production/Yield quantities of oats in world**. 2022 [*S. l.*]. Disponível em: https://www.fao.org/faostat/en/#data/QCL/visualize. Acesso em: 12 set. 2022.

FERRARI, Mauricio *et al*. Path analysis and phenotypic correlation among yield components of soybean using environmental stratification methods. **Australian Journal of Crop Science**, [*S. l.*], v. 12, n. 02, p. 193–202, 2018. Disponível em: https://www.researchgate.net/publication/323666303. Acesso em: 10 ago. 2022.

FLOSS, Elmar Luis *et al*. Crescimento, produtividade, caracterização e composição química da aveia branca. **Acta Scientiarum. Animal Sciences**, [*S. l.*], v. 29, n. 1, p. 1–7, 2007. Disponível: https://doi.org/10.4025/actascianimsci.v29i1.241. Acesso em: 30 jun. 2022.

FONTANELI, Renato Serena *et al*. Rendimento e valor nutritivo de cereais de inverno de duplo propósito: forragem verde e silagem ou grãos. **Revista Brasileira de Zootecnica**, [*S. l.*], v. 38, n. 11, p. 2126–2120, 2009. Disponível em: https://doi.org/10.1590/S1516-35982009001100007. Acesso em: 15 jun. 2022.

GUTKOSKI, Luiz Carlos *et al*. Influência dos teores de aveia e de gordura nas características tecnológicas e funcionais de bolos. **Ciência e Tecnologia de Alimentos**, Campinas, v. 29, n. 2, p. 254–261, jun. 2009. Disponível em: https://doi.org/10.1590/S0101-20612009000200003. Acesso em: 20 jun. 2022.

HAIR, Joseph F Jr *et al*. **Análise multivariada de dados**. 5. ed. Porto Alegre, RS: Bookman, 2005. 593 p. ISBN: 0-13-017706-7.

HOLLAND, James B. *et al*. Recurrent selection in oat for adaptation to diverse environments. **Euphytica**, *[S. l]*, v. 113, p. 195–205, 2000. Disponível em: https://doi.org/10.1023/A:1003933421378. Acesso em: 01 mar. 2023.

KAZIU, Isuf; KASHTA, Foto; CELAMI, Asllan. Estimation of Grain Yield, Grain Components and Correlations between Them in some Oat Cultivars. **Albanian Journal of Agricultural Sciences**, [*S. l.*], v. 18, n. 1, p. 13–19, 2019. Disponível em: https://ajas.ubt.edu.al/wp-content/uploads/2019/04/3_-ISUF-KAZIU-pp-13-19.pdf. Acesso em: 15 jun. 2022.

KLEIN, Luis Antônio *et al*. Dissimilaridade genética entre genótipos de aveia preta. **Revista Brasileira de Iniciação Científica**, Itapetinga, v. 6, n. 6, p. 114–115, 2019. Disponível em: https://periodicos.itp.ifsp.edu.br/index.php/IC/article/view/1634. Acesso em: 10 jun. 2022.

LORENCETTI, Claudir *et al*. Implicações da aplicação de fungicida na adaptabilidade e estabilidade de rendimento de grãos em aveia branca. **Ciência Rural**, [*S. l*], v. 34, n. 3, p. 693–700, jun. 2004. Disponível em: https://doi.org/10.1590/S0103-84782004000300007. Acesso em: 01 mar. 2023.

MANTAI, Rubia Diana *et al*. The dynamics of relation oat panicle with grain yield by nitrogen. **American Journal of Plant Sciences**, [*S. l*], v. 7, p. 17–27, 2016. Disponível em; http://dx.doi.org/10.4236/ajps.2016.71003. Acesso em: 01 mar. 2023.

MANTAI, Rubia Diana *et al*. Nitrogen management in the relationships between oat inflorescence components and productivity. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 24, n. 6, p. 385–393, 2020. Disponível em: https://doi.org/10.1590/1807-1929/agriambi.v24n6p385-393. Acesso em: 12 maio. 2022.

MANTAI, Rubia Diana *et al*. Contribution of nitrogen on industrial quality of oat grain components and the dynamics of relations with yield. **Australian Journal of Crop Science**, [*S. l.*], v. 15, n. 3, p. 334–342, 2021a. Disponível em: https://www.cropj.com/march2021.html. Acesso em: 10 jun. 2022.

MANTAI, Rubia Diana *et al*. Technical and agronomic efficiency of nitrogen use on the yield and quality of oat grains. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 25, n. 8, p. 529–537, 2021b. Disponível em: https://doi.org/10.1590/1807-1929/agriambi.v25n8p529-537. Acesso em: 10 jun. 2022.

MARTINELLI, José Antônio. Manejo integrado de doenças da aveia. **Fitopatologia Brasileira**, [*S. l*], v. 28, n. 1, p. 98–101, 2003.

MEINERZ, Gilmar Roberto *et al*. Valor nutritivo da forragem de genótipos de cereais de inverno de duplo propósito. **Revista Brasileira de Zootecnia**, [*S. l.*], v. 40, n. 6, p. 1173–1180, 2011. Disponível em: https://doi.org/10.1590/S1516-35982011000600003. Acesso em: 12 jun. 2021.

MEIRA, Daniela *et al*. Multivariate analysis revealed genetic divergence and promising traits for indirect selection in black oat. **Revista Brasileira de Ciências Agrárias**, Recife, v. 14, n.4, p. 1–7, 2019a. Disponível em: https://doi.org/10.5039/agraria.v14i4a6514. Acesso em: 10 jun. 2012.

MEIRA, Daniela *et al*. Phenotypic variance of black oat growing in crop seasons reveals genetic effects predominance. **Anais da Academia Brasileira de Ciências**, [*S. l.*], v. 91, 2019b. Disponível em: https://doi.org/10.1590/0001-3765201920180036. Acesso em: 12 abr. 2022.

MEIRA, Daniela *et al*. Estimates of genetic parameters between and within black oat populations. **Bragantia**, Campinas, v. 78, n. 1, p. 43–51, 11 mar. 2019c. Disponível em: https://doi.org/10.1590/1678-4499.2018116. Acesso em: 01 jun. 2022.

MONTGOMERY, Douglas C; PECK, Elizabeth A; VINING, Geoffrey. **Introduction to linear regression analysis**. 5. ed. Hoboken, 2012. 821 p. ISBN 978-0-470-54281-1.

MORADI, Mohammad; REZAI, Abdolmajid; ARZANI, Ahmad. Path Analysis for Yield and Related Traits in Oats. **Journal of Science and Technology of Agriculture and Natural Resources**, [*S. l.*], v. 9, p. 173-180, 2005. Disponível em: https://www.researchgate.net/publication/26622524_Path_Analysis_for_Yield_and_Related_Traits_in_Oats. Acesso em: 20 fev. 2022.

MOURA, Natã Balssan *et al*. Akaike criteria and selection of physiological multi-character indexes for the production of black oat seeds. **Comunications in Plant Sciences**, [*S. l.*], v. 11, p. 22–29, 2021. Disponível em: https://pdfs.semanticscholar.org/6d14/0d2a0165f48b99e16ce79a0f7cd6dc39185b.pdf. Acesso em: 05 jul. 2022.

NARDINO, Maicon *et al*. Association of secondary traits with yield in maize F 1 ’s. **Ciência Rural**, [*S. l.*], v. 46, n. 5, p. 776–782, 2, 2016a. Disponível em: https://doi.org/10.1590/0103-8478cr20150253. Acesso em: 21 ago. 2022.

OLIVOTO, Tiago *et al*. Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. **Agronomy Journal**, Madison, v. 109, n. 1, p. 131–142, 2017a. Disponível em: https://doi.org/10.2134/agronj2016.04.0196. Acesso em: 18 abr. 2022.

OLIVOTO, Tiago *et al*. REML/BLUP and sequential path analysis in estimating genotypic values and interrelationships among simple maize grain yield-related traits. **Genetics and Molecular Research**, [*S. l.*], v. 16, n. 1, p. 1–19, 2017b. Disponível em: http://www.funpecrp.com.br/gmr/year2017/vol16-1/pdf/gmr-16-01-gmr.16019525.pdf. Acesso em: 10 ago. 2022.

REGINATTO, Douglas Cezar *et al*. Sustainable optimization of nitrogen uses in oat at sowing and trop-dressing stages. **Australian Journal of Crop Science**, [*S. l.*], v. 15, n. 1, p. 23–31, 2021a. Disponível em: https://www.cropj.com/reginatto_15_1_2020_23_31.pdf. Acesso em: 17 fev. 2022.

REGINATTO, Douglas Cezar *et al*. Nitrogen management at sowing and topdressing with the time of supply in the main biotype of oats grown in southern Brazil. **Australian Journal of Crop Science**, [*S. l.*], v. 15, n. 04, p. 524–530, 2021b. Disponível em: https://www.cropj.com/carvalho_15_4_2021_524_530.pdf. Acesso em: 17 fev. 2022.

RODRIGUES, Gabriel Belfort *et al*. Análise de trilha de componentes de produção primários e secundários em tomateiro do grupo salada. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 45, n. 2, p. 155–162, 2010. Disponível em: https://doi.org/10.1590/S0100-204X2010000200006. Acesso em: 25 jul. 2022.

SALLA, Vanessa Padilha *et al*. Análise de trilha em caracteres de frutos de jabuticabeira. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 50, n. 3, p. 218–223, 2015. Disponível em: https://doi.org/10.1590/S0100-204X2015000300005. Acesso em: 05 maio. 2022.

SARI, Bruno Giacomini *et al*. Linear relationships between cherry tomato traits. **Ciência Rural**, Santa Maria, v. 47, n. 3, p. 1–7, 2017. Disponível em: https://doi.org/10.1590/0103-8478cr20160666. Acesso: 20 jan. 2021.

SARI, Bruno Giacomini *et al*. Intereference of sample size on multicollinearity diagnosis in path analyses. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 53, n. 6, p. 769–773. 2018. Disponível em: https://doi.org/10.1590/S0100-204X2018000600014. Acesso em: 20 jan. 2021.

TOEBE, Marcos *et al*. Direct effects on scenarios and types of path analyses in corn hybrids. **Genetics and Molecular Research**, [*S. l.*], v. 16, n. 1, p. 1–15, 2017a. Disponível em: https://www.geneticsmr.com/sites/default/files/articles/year2017/vol16-1/pdf/gmr-16-01-gmr.16019529_0.pdf. Acesso em: 03 jan. 2021.

TOEBE, Marcos *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, 2017b. Disponível em: https://www.geneticsmr.com/sites/default/files/articles/year2017/vol16-2/pdf/gmr-16-02-gmr.16029523_0.pdf. Acesso em: 03 jan. 2021.

TOEBE, Marcos; CARGNELUTTI FILHO, Alberto. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, Brasília, v. 48, n. 5, p. 466–477, 2013. Acesso em: 03 jan. 2021.

TOEBE, Marcos; CARGNELUTTI FILHO, Alberto. Multicollinearity in path analysis of maize (*Zea mays* L.). **Journal of Cereal Science**, [*S. l.*], v. 57, p. 453–462, 2013. Disponível em: https://doi.org/10.1016/j.jcs.2013.01.014. Acesso em: 23 abr. 2022.