

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

Kamylla Rodrigues Leandro

**ANÁLISE DE TRILHA PARA SELEÇÃO DE CARACTERES FENOTÍPICOS EM
LINHAGENS DE MILHO**

Uberlândia

2016

Kamylla Rodrigues Leandro

**ANÁLISE DE TRILHA PARA SELEÇÃO DE CARACTERES FENOTÍPICOS EM
LINHAGENS DE MILHO**

Trabalho de conclusão de curso de graduação apresentado à Faculdade de Matemática da Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de Bacharel(a) em Estatística.

Orientador: Prof. Dr. Marcelo Tavares

Uberlândia

2016

Kamylla Rodrigues Leandro

**ANÁLISE DE TRILHA PARA SELEÇÃO DE CARACTERES FENOTÍPICOS EM
LINHAGENS DE MILHO**

Trabalho de conclusão de curso de graduação apresentado à Faculdade de Matemática da
Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de
Bacharel (a) em Estatística.

Aprovado em: 29 de 06 de 2016


BANCA EXAMINADORA



Prof.º Dr. Lúcio Borges Araújo – Universidade Federal de Uberlândia



Prof.º Dr. Rogério de Melo Costa Pinto – Universidade Federal de Uberlândia



Prof.º Dr. Marcelo Tavares (Orientador) – Universidade Federal de Uberlândia

AGRADECIMENTOS

Primeiramente agradeço a minha amada família, minha mãe Suely Leandro da Silva e irmã Amanda Rodrigues Leandro que com muito apoio, amor e carinho contribuíram para toda essa jornada, que apesar de todas as dificuldades me fortaleceram.

Um agradecimento especial ao Herbert Carvalho Alexandrino, por todo o companheirismo. Obrigada pela paciência, pelo incentivo, e principalmente por depositar em mim confiança e acreditar no meu potencial.

À Karoline Viana, que mesmo seguindo caminhos diferentes, sempre se fez presente em minha vida e foi essência nesse percurso.

Agradeço meu orientador Marcelo Tavares, não somente por me auxiliar no desenvolvimento desta monografia, mas também por ser uma figura importante na minha vida acadêmica e profissional. Obrigada por todas as oportunidades que me proporcionou e por contribuir para meu amadurecimento no processo de formação profissional.

Aos professores participantes da banca examinadora que prontamente aceitaram dividir comigo este momento tão importante e esperado: Lúcio Borges Araújo e Rogério de Melo Costa Pinto.

Aos professores Patrícia Viana da Silva e Thiago Moreira Vargas, por suas vocações inequívocas, por não pouparem esforços como interlocutores dos alunos e por suprirem eventuais falhas e lacunas. Obrigada por contribuírem com tantos ensinamentos e tanto conhecimento. Quero expressar o meu reconhecimento e admiração pela competência profissional de vocês e minha gratidão pela amizade.

À Monsanto e ao time de pesquisa de produção pelo suporte e por ter permitido a realização deste trabalho.

Ao Heyder Silva pela oportunidade e apoio na elaboração da pesquisa para este trabalho. Muito obrigada pela parceria e por compartilhar um pouco do seu vasto conhecimento e expertise.

Ao meu gestor Alexandre Merlin e mentores Renata Dainese e Marcos Bosqueiro. Obrigada por todos os ensinamentos durante minha curta carreira até então. O sucesso de vocês, profissionalismo e carisma são admirados por todos e por mim!

Agradeço a Michelle Silva, grande amiga, obrigada por todas as palavras de encorajamento, por ser companheira para todos os momentos e pelo apoio constante, a Polyane Quinarelli e Mariana Pintar, por me transmitirem tanta alegria e paz e a Gabriela Peres que nunca deixou de me estender a mão nos momentos em que precisei. Agradeço pela amizade de vocês, por fazerem desses anos de graduação, anos mais divertidos e alegres, vocês construíram comigo um sonho que agora se torna realidade.

E enfim, a todos que contribuíram para a realização deste trabalho, seja de forma direta ou indireta, fica registrado aqui, o meu muito obrigado!

RESUMO

A produtividade é um caráter complexo e resultante da expressão e associação de diferentes componentes. As análises de correlações, o conhecimento das relações existentes entre caracteres agronômicos e a produtividade de grãos, tem sido objetivo de estudo no melhoramento genético de plantas para processos de seleção. No entanto, apesar de inferir sobre o entendimento de certo caractere, os coeficientes de correlação, não dão exata importância relativa dos efeitos diretos e indiretos destes fatores. Nesse contexto, a análise de trilha apresenta-se como uma alternativa viável. O objetivo foi avaliar a relação entre a produção de grãos de milho e caracteres agronômicos secundários, visando identificar, por meio da análise de trilha, os melhores caracteres indicados para seleção indireta para produtividade em variedades de milho, tendo como caráter principal a produtividade de grãos, na unidade de sacos de 60 mil sementes vendáveis de milho. Foram analisados experimentos desde o inverno de 2013 até o verão de 2015, em 12 regiões do Brasil. O delineamento foi em blocos casualizados e os tratamentos foram constituídos de 271 linhagens de milho. Avaliaram-se os caracteres: altura de plantas; altura da espiga; plantas dominadas; tombamento de plantas; plantas acamadas; espigas caídas; uniformidade de plantas; plantas sem espigas; peso de espigas colhidas na parcela; peso de mil grãos; enrolamento do estigma; espigas mal formadas; falha de granação; polinização no cartucho; umidade de grãos; produtividade de grãos; sacos de sementes vendáveis; descarte de grãos; grãos ardidos; grãos rachados e viviparidade. Os caracteres: produtividade de grãos, peso de mil grãos, descarte e grãos ardidos foram os mais significativos para a variável resposta produtividade na unidade de sacos de sementes vendáveis. Sendo a produtividade de grãos o caractere mais indicado para seleção indireta, é a variável com maior efeito direto, em sentido favorável a variável resposta em estudo (sacos de 60 mil sementes de milho).

Palavras-Chave: *Linhagens de milho; Correlação genética; Seleção indireta; Análise de Trilha; Produtividade.*

ABSTRACT

Productivity is a complex and resulting character of expression and association of different components. The knowledge of the relationship between agronomic traits and correlations analysis between grain yield and other characters, has been objective of study in plant breeding to selection process. However, despite inferences about understanding certain character, the correlation coefficients do not give exact relative importance of direct and indirect effects of these factors. In this context, the path analysis is presented as a viable alternative. The objective was to evaluate the relationship between the corn grain yield and secondary agronomic traits to identify, through the Path Analysis, the best characters suitable for breeding purposes in maize varieties, the main character yield in salable bags (bags 60 thousand seeds). The experiments were analyzed since the winter of 2013 until the summer of 2015 in 12 locations in Brazil. The design was randomized blocks and treatments consisted of 271 corn inbred lines. The characters were evaluated: Barren Plants; Malformed Ears; Dominated Plants; Dropped Ears; Discard; Ear Height; Failure of Graining; Moisture; Plant Height; Rotten Grains; Root Lodging; Silk balling; Shed in Boot; Shell Weight; Popped Kernels; Saleable Seeds Units; Total Damaged Plants; Thousand Kernels Weight; Plant Uniformity; Viviparity and Yield. The characters: Yield, Thousand Kernels Weight, Discard and Rotten Grains were the most significant for the variable Salable Seed Units. Yield is the character most suitable for indirect selection as it is the variable with the greatest direct effect, on favorable direction, to the response variable.

Key Words: *Corn Inbred Line; Genetic Correlation; Indirect Selection; Path Analysis; Yield.*

LISTA DE TABELAS

Tabela 1 – Produção, Consumo e Exportação de Milho (em mil toneladas).....	15
Tabela 2 – Resumo do método <i>forward</i> para a seleção dos caracteres agronômicos	32
Tabela 3 – Coeficiente de correlação fenotípica entre os caracteres agronômicos em estudo.....	33
Tabela 4 – Efeitos diretos e indiretos das variáveis explicativas produtividade de grãos (YLD), peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR) e residual sobre a variável resposta produtividade na unidade de sacos de sementes vendáveis (SSU).....	34

LISTA DE FIGURAS

Figura 1 – Distribuição da produção de milho nos estados brasileiros.....	13
Figura 2 – Produção de milho em milhões de toneladas por hectare – Brasil.....	14
Figura 3 – Área plantada de milho. Projeções do Agronegócio – 2014/15 a 2024/25	14
Figura 4 – Processo de endogamia e vigor de híbrido (heterose).....	16
Figura 5 – Processo de formação de um híbrido simples.....	17
Figura 6 – Diagrama causal dos efeitos das variáveis explicativas (x_1 , x_2 e x_3) e residual (u) sobre a variável dependente (Y).....	29
Figura 7 – Efeitos diretos e indiretos das variáveis explicativas produtividade de grãos (YLD), peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR) e residual sobre a variável resposta produtividade na unidade de sacos de sementes vendáveis (SSU).....	34

SUMÁRIO

1.	INTRODUÇÃO	10
2.	REFERENCIAL TEÓRICO	13
2.1.	A PRODUÇÃO MILHO NO BRASIL E PROJEÇÕES	13
2.2.	MELHORAMENTO DE MILHO	16
2.3.	O USO DO MODELO MISTO (BLUP) PARA PREDIÇÕES NA OCORRÊNCIA DA INTERAÇÃO GENÓTIPO x AMBIENTE	19
2.4.	MODELOS DE REGRESSÃO LINEAR MISTO	21
2.5.	ANÁLISE DE TRILHA	26
3.	METODOLOGIA.....	31
4.	RESULTADOS E DISCUSSÕES	32
5.	CONSIDERAÇÕES FINAIS.....	36
6.	REFERÊNCIAS	37
7.	ANEXO A – ROTINA R PARA ANÁLISE DE TRILHA.....	41

1. INTRODUÇÃO

A produção de milho a nível mundial, têm aumentado continuamente nas últimas décadas, é uma das culturas mais importantes no mundo em função da sua capacidade produtiva, composição química e valor nutritivo. No que tange à comercialização do milho brasileiro com o mercado externo, o Brasil posiciona-se atrás apenas dos Estados Unidos, no mercado de exportação do grão, com uma participação de 22,5% (FIESP, 2016).

A cultura é de extrema importância para o Brasil, que é um dos maiores produtores do mundo e ocupa a terceira posição no *ranking* mundial, os maiores produtores mundiais de milho são os Estados Unidos, China e Brasil, que, na safra de 2015/2016, produziram aproximadamente: 345,50; 224,60; e 77,50 milhões de toneladas, respectivamente (FIESP, 2016).

O mercado é impulsionado por estratégias baseadas em aumento de produtividade, desenvolvimento contínuo no melhoramento genético para utilização na produção de variedades e híbridos de alta produtividade e adaptados a diferentes sistemas de cultivo e regiões.

A grande demanda por uma produção autossuficiente de milho no Brasil tem estimulado esforços de melhoramento, que são capazes de sustentar o desenvolvimento de plantas de alta performance, influenciando o sucesso dos programas de melhoramento no país. Tal sucesso foi decorrente do esforço realizado no final dos anos 60 e início da década de 70, através da ampliação da base genética, na coleta de variedades locais e obtenção de variedades melhoradas, na mesma época também foi estimulado a introdução de variedades melhoradas desenvolvidos pelo Centro Internacional de Melhoramento de Milho e Trigo (CIMMYT) e de outros programas de desenvolvimento de germoplasma (GAMA et al., 1996).

A pesquisa genética aplicada ao melhoramento de plantas vem passando por inúmeras transformações e desafios ao longo dos anos. Além da elevada demanda, as novas necessidades e competitividades do mercado, o surgimento novas doenças e pragas, a capacidade de adaptação aos ambientes e características específicas desejadas são alguns exemplos. O melhoramento genético, é uma das mais expressivas técnicas que tem contribuído para o aumento de produtividade da cultura de milho. O desenvolvimento e a

utilização do milho híbrido proporcionou, em curto período de tempo, um ganho de produtividade superior a 150% (BUENO et al., 2006).

O desenvolvimento de novas cultivares com maior potencial genético para a produtividade é a principal meta dos programas de melhoramento. Nestes programas, o conhecimento da correlação entre os caracteres (variáveis aleatórias na terminologia estatística) que compõem o rendimento de grãos é de grande importância, pois ao se realizar a seleção com base em certa característica é necessário também avaliar o quanto da alteração de um caráter pode afetar os demais caracteres que estão correlacionados entre si (SANTOS & VENCOVSKI, 1986).

De acordo com SANTOS e VENCOVSKY (1986), as correlações são, em geral, explicadas pelo efeito aditivo dos genes, afetando dois caracteres simultaneamente. Portanto, o conhecimento do grau de associação entre caracteres agronômicos é de grande importância para os melhoristas, principalmente porque a seleção sobre determinado caráter pode alterar o comportamento do outro.

As análises de correlação entre a produtividade de grãos e outros caracteres de importância agronômica têm sido estudadas em diversas culturas e trabalhos de pesquisa na área. Entretanto, apesar de inferir sobre o entendimento de certo caractere, os coeficientes de correlação apesar de serem de grande utilidade na quantificação da magnitude e direção das influências de fatores na determinação de caracteres complexos, não dão exata importância relativa dos efeitos diretos e indiretos destes fatores (CRUZ & REGAZZI, 1997).

Sendo assim, com o intuito de entender a associação entre características fenotípicas de culturas, foi proposto originalmente por WRIGHT (1921) um método denominado análise de trilha (*Path Analysis*), que consiste no desdobramento das correlações estimadas em efeitos diretos e indiretos de uma variável resposta, no caso em questão, a produtividade em grãos. Tais efeitos são obtidos por meio de equações de regressão, onde as variáveis são previamente padronizadas.

Dessa forma, estudos sobre o desdobramento do coeficiente de correlação são feitos pela análise de trilha, que consiste no estudo dos efeitos diretos e indiretos de caracteres sobre uma variável básica, em que as estimativas dos efeitos são obtidos, por meio de equações de regressão, onde as variáveis são previamente padronizadas. Logo, esta análise baseia-se na avaliação do efeito de uma variável independente, explicativas (X) sobre uma variável

dependente (Y), de forma com que as outras variáveis independentes (X_i) não tenham influência sobre os efeitos.

Considerando-se às técnicas e os processos de seleção, tendo em vista a diversidade genética, objetivos e metodologias para achar dentre inúmeras plantas, aquelas que expressam as características mais desejadas, através dos programas de melhoramento e pesquisa de produção, é possível testar as plantas de milho e observar a manifestação de determinada característica para se tomar a decisão de selecioná-la ou não.

O presente trabalho propõe um estudo das características do milho, desejando-se identificar por meio da análise de trilha quais caracteres da planta possuem maior contribuição para a produtividade de grãos na unidade de sacos de sementes vendáveis, objetivando identificar os caracteres mais indicados à seleção indireta para a produtividade, pois este é um caráter complexo e resultante da expressão e associação de diferentes parâmetros agronômicos quantitativos e qualitativos, de forma que este possa servir de ferramenta para o melhorista no momento da seleção. Serão avaliados os componentes das plantas de milho e da produtividade resultante de diferentes linhagens e ambientes.

2. REFERENCIAL TEÓRICO

2.1. A PRODUÇÃO MILHO NO BRASIL E PROJEÇÕES

O milho no país é cultivado em todo território nacional, sendo que 88,2% da produção concentraram-se nas regiões Sul (30,6% da produção), Sudeste (13,7% da produção) e Centro-Oeste (43,9% da produção). Os principais estados produtores são: Mato Grosso, Paraná, Mato Grosso do Sul, Goiás, Minas Gerais, Rio Grande do Sul e São Paulo que respondem por 83,4% da produção nacional (Figura 1).

Atualmente o milho representa 40,5% dentre a produção total de todos os cultivares (Soja, Arroz, Trigo, Feijão, Algodão e demais produtos). O cereal é a segunda maior cultura de importância na produção agrícola no país, superado apenas pela soja que lidera a produção de grãos. (CONAB, 2015).

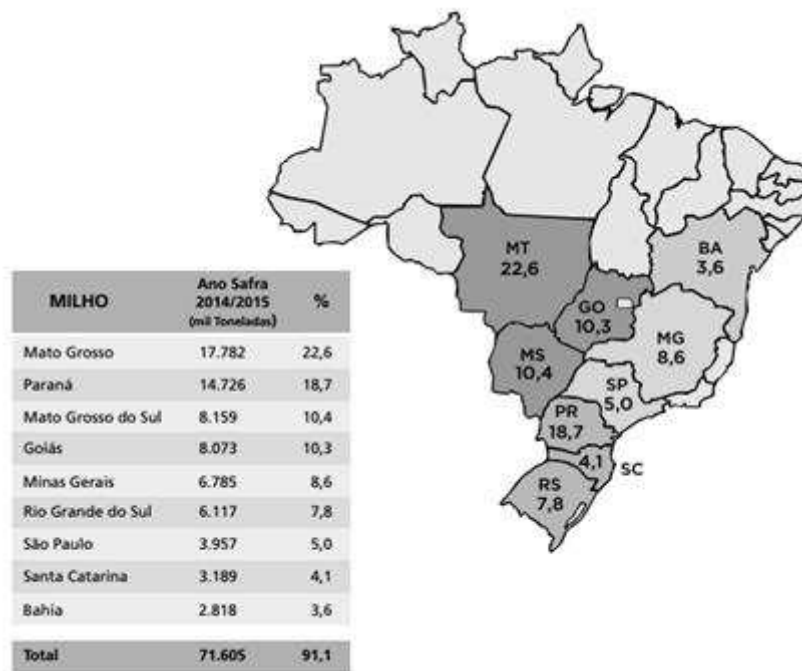


Figura 1 – Distribuição da produção de milho nos estados brasileiros.

Fonte: (MINISTÉRIO DA AGRICULTURA, 2016).

A produção de milho de 2014/2015 no Brasil foi de aproximadamente 84.672,40 milhões de toneladas no total, sendo 30.082,00 na primeira safra e 54.590,50 na segunda safra (Figura 2).

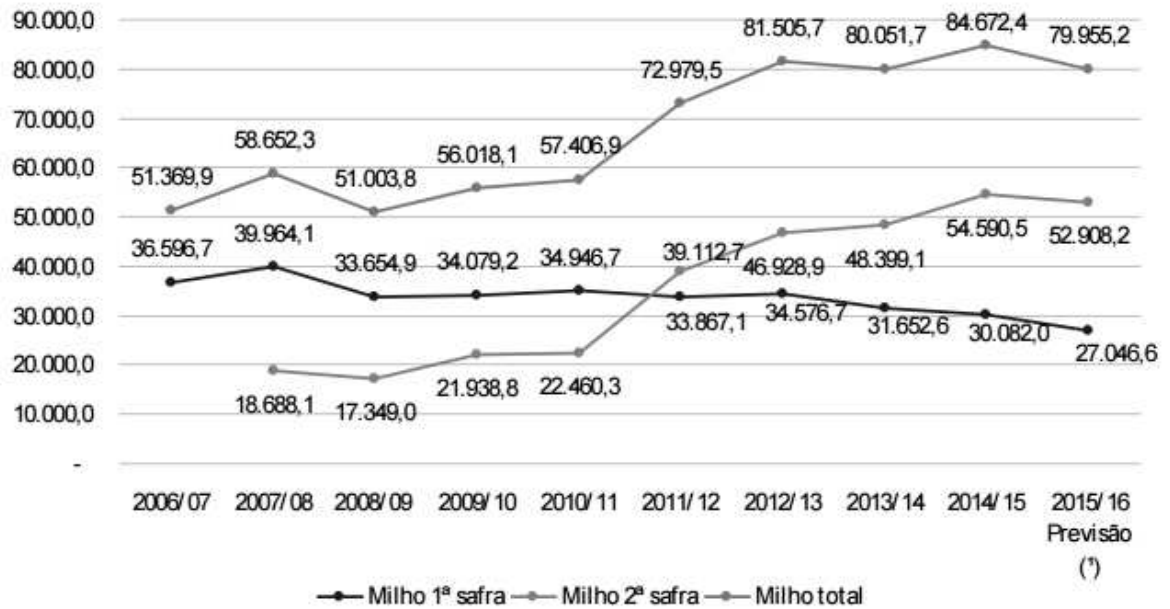


Figura 2 – Produção de milho em milhões de toneladas por hectare – Brasil.

Fonte: (CONAB, 2015)

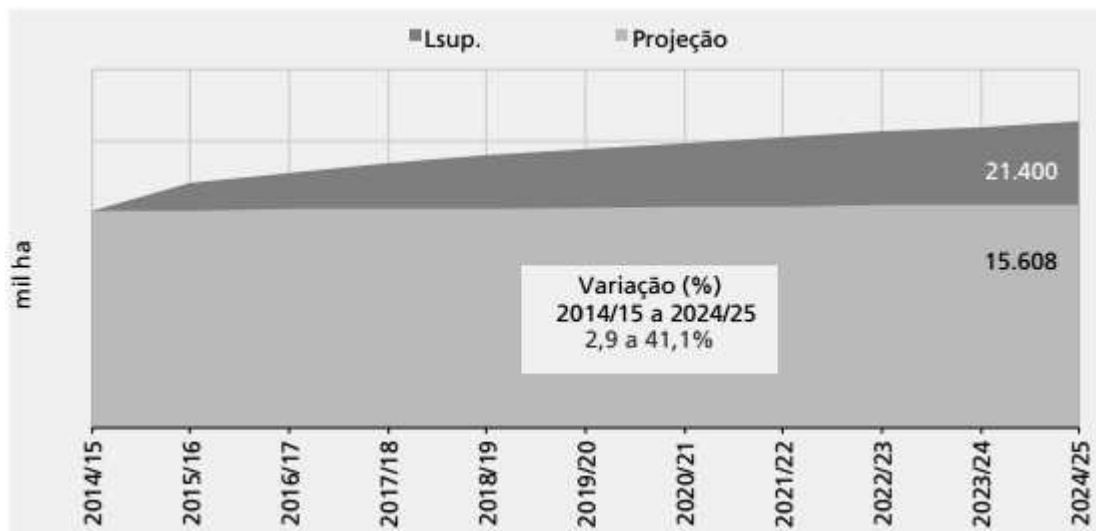


Figura 3 – Área plantada de milho. Projeções do Agronegócio – 2014/15 a 2024/25.

Fonte: (MINISTÉRIO DA AGRICULTURA, 2016).

A área plantada de milho pode chegar a 21,40 milhões de hectares em 2024/25 conforme (Figura 3). O consumo interno de milho que em 2015/16 representa 69,17% da produção deve reduzir-se nos próximos anos para 65,34%. As exportações devem passar de 22,33 milhões de toneladas em 2015/16 para 31,75 milhões de toneladas em 2024/25 (Tabela 1).

Tabela 1 – Produção, Consumo e Exportação de Milho (mil toneladas).

Ano	Produção		Consumo		Exportação	
	Projeção	Limite superior	Projeção	Limite superior	Projeção	Limite superior
2015/16	81,06	95,05	56,07	57,79	22,33	30,47
2016/17	83,14	102,91	57,10	60,05	23,33	33,80
2017/18	85,22	109,44	58,12	62,13	24,40	37,05
2018/19	87,29	115,26	59,13	64,06	25,45	39,94
2019/20	89,37	120,64	60,14	65,88	26,50	42,64
2020/21	91,45	125,70	61,15	67,61	27,55	45,18
2021/22	93,53	130,52	62,16	69,27	28,60	47,61
2022/23	95,60	135,15	63,17	70,88	29,65	49,95
2023/24	97,68	139,63	64,17	72,44	30,70	52,21
2024/25	99,76	143,97	65,18	73,97	31,75	54,41
Variação %	26,30%		18,50%		51,20%	

Fonte: (MINISTÉRIO DA AGRICULTURA, 2016).

Para manter o consumo interno projetado de 65,18 milhões de toneladas e garantir um volume razoável de estoques finais e o nível de exportações projetado, a produção projetada deverá situar-se em pelo menos 99,76 milhões de toneladas em 2024/25, conforme Tabela 1. (MINISTÉRIO DA AGRICULTURA, 2016). Estima-se que a área deve aumentar mais do que está sendo projetado. Entretanto, o milho teve nos últimos anos, elevados ganhos de produtividade resultando em menor necessidade adicional de áreas através dos programas de melhoramento genético de plantas de milho.

2.2. MELHORAMENTO DE MILHO

A descoberta do vigor híbrido aconteceu no século XX, por um botânico e geneticista norte-americano chamado George Harrison Shull. Ele mostrou que, ao fecundar a planta com o próprio pólen (autofecundação), eram produzidos descendentes menos vigorosos. Ao repetir todo o processo nas seis ou oito gerações seguintes, os descendentes fixavam características agrônômicas importantes.

Passaram a ser chamadas por linhas puras, ou linhagens, as plantas que geravam filhos geneticamente semelhantes e também iguais às mães, através do processo de autofecundação (SHULL, 1909). Shull notou que duas linhagens diferentes ao serem cruzadas entre si produziam descendentes com grande vigor, chamado de vigor híbrido ou heterose, dando origem ao milho híbrido.

O cruzamento de duas linhagens endogâmicas (obtida através do processo de autofecundação) restaura o vigor em uma combinação híbrida. Como a semente de um híbrido simples é produzida a partir de uma linhagem endogâmica, a endogamia reduz o potencial de produtividade devido à perda de vigor e, por este fato, tem um custo de produção de sementes relativamente mais alto em consequência da produtividade reduzida, Figura 4.

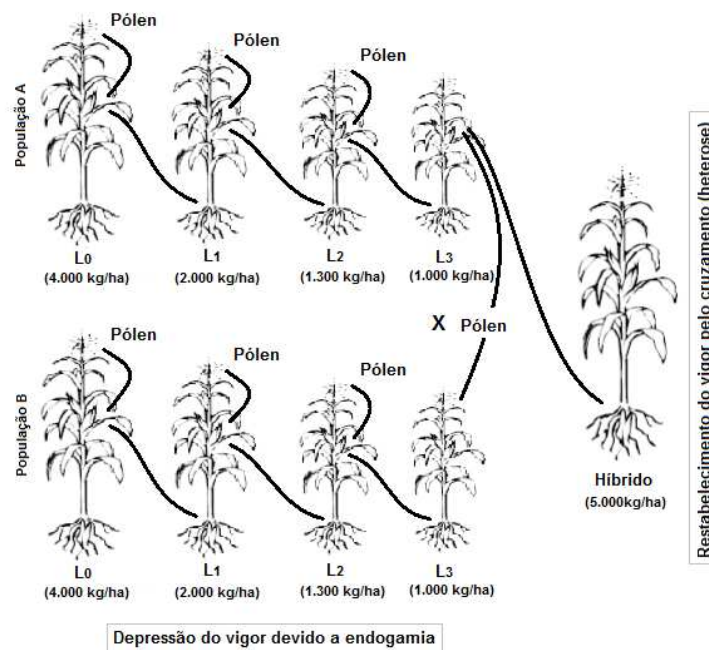


Figura 4 – Processo de endogamia e vigor de híbrido (heterose).

Fonte: Elaborada pelo autor.

O cruzamento entre linhagens e a sua utilização na produção de sementes de milho híbrido impulsionou significativamente o melhoramento genético convencional. Tal descoberta foi uma das maiores contribuições da genética à agricultura mundial, possibilitou ganhos consideráveis, em produtividade, à cultura do milho. As produtividades alcançaram recordes sucessivos e o mercado impulsionou as instituições públicas e privadas de produção de sementes de milho a serem cada vez mais competitivas.

O milho é uma planta monoica (o mesmo indivíduo apresenta órgãos sexuais dos dois sexos) de característica alógama (fecundação cruzada e ao acaso). Assim sendo, possui uma vasta diversidade genética. A seleção, ou seja, a intervenção humana é a responsável pela existência da espécie milho, que deve-se ao fato de que as plantas podem ser colhidas individualmente, contribuindo para o melhoramento da espécie nos métodos de seleção.

A linhagem que produz o pólen é considerada um macho ou polinizador, enquanto que a receptora de pólen, que se colhe apenas a espiga, é chamada de fêmea. (PATERNIANI & MIRANDA FILHO, 1987). A obtenção de linhagens é um dos objetivos básicos num programa de melhoramento genético, pois promove a produção de sementes para origem de híbridos (ALLARD, 1971).

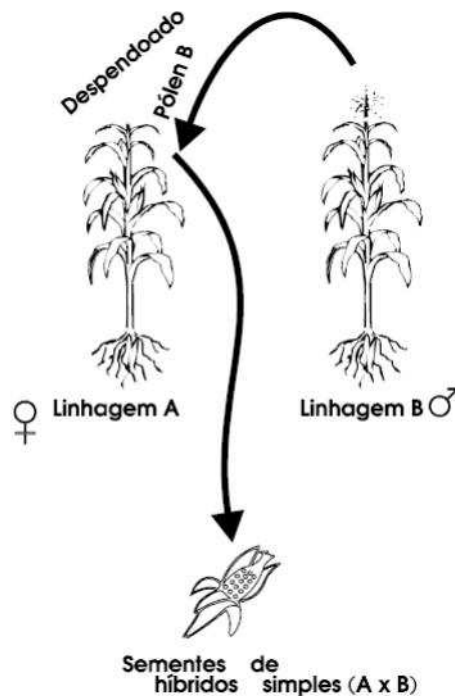


Figura 5 – Processo de formação de um híbrido simples.

Fonte: (DA ROCHA & PEREIRA, 2016).

Um híbrido simples é o produto obtido do cruzamento de duas linhagens endogâmicas: uma com a função parental de macho ou polinizador e outra com função de parental fêmea (KOSHIMA, 2009), exemplificado na Figura 5.

No processo de seleção de linhagens e testes de parentais, tanto o parental fêmea, quanto o parental macho ou polinizador, são avaliados e plantados separadamente. No processo de produção de híbridos, no caso, híbridos simples que vem sendo discutido aqui, é necessário estabelecer o cruzamento entre os parentais.

No plantio, são intercaladas linhas de fêmeas e polinizadores sendo necessário avaliar os ciclos de florescimento, com a finalidade de coincidir a polinização do macho com a emissão dos estilos estigma da fêmea. A proporção da quantidade de linhas de polinizadores e linhas de fêmeas, variam de acordo as características do polinizador e seu potencial de polinização, o parental macho deve ter como característica uma boa produção de pólen e capacidade de polinizar adequadamente as linhas de parentais fêmea (KOSHIMA, 2009).

As plantas das linhas dos parentais fêmea têm os pendões removidos antes de iniciar a polinização (chamado despendoamento, Figura 5), pois a fêmea não pode autopolinizar, para o cruzamento é necessário garantir que a fêmea receba apenas o pólen do polinizador. O parental fêmea, que receberá o pólen da linha de polinizadores, produzirá as espigas com as sementes do híbrido simples final. As características atribuídas a fêmea devem estar relacionadas com uma boa produtividade, bom rendimento em peneiras comerciais (processo que se seleciona os grãos de acordo com comprimento e largura) e sanidade dos grãos.

2.3. O USO DO MODELO MISTO (BLUP) PARA PREDIÇÕES NA OCORRÊNCIA DA INTERAÇÃO GENÓTIPO x AMBIENTE

No melhoramento genético de plantas, as técnicas de avaliação desempenham um papel fundamental, pois permitem a predição dos valores genótipos e fenótipos dos candidatos a seleção, propiciando uma seleção mais acurada.

O uso de delineamentos experimentais repetidos em vários ambientes ou locais é uma prática comum no melhoramento genético de plantas (VENCOVSKY & BARRIGA, 1992). As avaliações e seleções, tornam-se bastante complexas em decorrência do fenômeno da interação GxA (Genótipo x Ambiente), sua importância pode ser evidenciada pelo enorme volume de informação existente, quer seja na forma de artigos científicos (MARIOTTI & CLARIANA, 1994; BACHIREDDY et.al, 1992; GORMAN et al., 1989; HUSSEIN et al., 2000 e SHUKLA, 1983), quer seja como anais de congressos, simpósios, livros e capítulos de livros (GAUCH & ZOBEL, 1996; CRUZ & REGAZZI, 1997; KANG & GAUCH, 1996; DUARTE & VENCOVSKY, 1999). Considerando-se tal fenômeno, nos métodos de seleção é necessária a obtenção de preditores para cada indivíduo considerando-se a variação ambiental.

A metodologia de avaliação genética para os preditores mencionados utilizada atualmente é baseada na teoria dos modelos mistos (HENDERSON, 1963, 1975) conhecida como BLUP - *Best Linear Unbiased Prediction*, que, em português significa “melhor predição linear não viesada”. O procedimento BLUP permite a análise de dados não balanceados, retornando, nestas condições, predições mais confiáveis do que as obtidas pelo método dos quadrados mínimos (MQM) do método de regressão convencional. Ele possibilita a estimativa de valores genéticos mais próximos do verdadeiro valor genotípico (NUNES et al., 2008).

O BLUP é um estimador do tipo “*shrinkage*”. Um estimador do tipo “*shrinkage*” pode ser visto como um coeficiente de confiabilidade que, no caso dos experimentos de desempenho genotípico, é função da herdabilidade ao nível de médias por meio das repetições. Em outras palavras, a média fenotípica é ajustada pelo fator de confiabilidade, que é um estimador do tipo “*shrinkage*”. (FERRAUDO, 2013).

Uma redução na variância genética relativa implica num estreitamento da dispersão das respostas genotípicas médias preditas podendo isto chegar ao limite teórico das médias se igualarem. Trata-se do chamado efeito “*shrinkage*”, relatado na literatura de modelos lineares

mistos. O efeito “*shrinkage*”, nada mais é do que o “encolhimento” da distribuição das médias ajustadas em torno da média geral, quando se passa de uma análise assumindo-os como de efeitos fixos para outra em que tais efeitos são tidos como aleatórios. (DUARTE & VENCOVSKY, 2001)

Pode-se utilizar o BLUP individual dos caracteres agronômicos para cada linhagem estudada em delineamentos experimentais com vários indivíduos por parcela e repetidos em vários locais. As equações de modelo misto multivariado apresentadas a seguir, associadas ao BLUP individual, permitem a obtenção do valor de predição dos caracteres fenotípicos de cada linhagem, considerando de maneira eficiente a ocorrência de interação GxA. As características fenotípicas são resultantes da interação do genótipo com o ambiente, a distinção entre genótipo e fenótipo foi originalmente proposta por JOHANNSEN (1911).

2.4. MODELO DE REGRESSÃO LINEAR MISTO

Um modelo linear misto geral é da seguinte forma (HENDERSON, 1984):

$$y = Xb + Za + e \quad (1),$$

com as seguintes distribuições e estruturas de médias e variâncias:

$$a \sim N(0, G); \quad E(y) = Xb$$

$$e \sim N(0, R); \quad V(y) = V = ZGZ' + R$$

em que:

y : vetor de observações;

b : vetor paramétrico dos efeitos fixos, com matriz de incidência X ;

a : vetor paramétrico dos efeitos aleatórios, com matriz de incidência Z ;

e : vetor de erros aleatórios;

G : matriz de variância covariância dos efeitos aleatórios;

R : matriz de variância covariância dos erros aleatórios;

0 : vetor nulo.

Assumindo como conhecidos G e R , a simultânea estimação dos efeitos fixos e predição dos efeitos aleatórios pode ser obtida pelas equações dadas por:

$$\begin{bmatrix} Z'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1} + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

A solução deste sistema para \hat{b} e \hat{a} conduz a resultados idênticos aos obtidos por:

$\hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y$: estimador de quadrados mínimos generalizados (GLS) ou melhor estimador linear não viciado (BLUE) de b ;

$\hat{a} = GZ'V^{-1}(y - X\hat{b}) = CV^{-1}(y - X\hat{b})$: melhor preditor linear não viciado (BLUP) de a ;
em que $C = GZ' =$ matriz de covariância entre a e y .

Quando G e R não são conhecidas, as componentes de variância a eles associados podem ser estimados empregando-se o procedimento REML (PATTERSON & THOMPSON, 1971; SEARLE, 1992). Exceto por uma constante, a função de verossimilhança restrita a ser maximizada, é dada por:

$$L = -\frac{1}{2} \left(\log|XH^{-1}X| + \log|H| + v \log \sigma_e^2 + \frac{y'Py}{\sigma_e^2} \right) \\ - \frac{1}{2} \left(\log|C| + \log|R| + \log|G| + v \log \sigma_e^2 + \frac{y'Py}{\sigma_e^2} \right)$$

em que:

$$H = R + ZGZ';$$

$$P = H^{-1} - H^{-1}(X'H^{-1}X)^{-1}X'H^{-1};$$

$v = N - r(x) =$ graus de liberdade, em que N é o número total de dados e $r(x)$ é o posto da matriz X ;

$C =$ matriz dos coeficientes das equações de modelo misto.

Sendo geral, o modelo (1) contempla vários modelos inerentes às diferentes situações, tais quais:

- **Modelo univariado, ajustando apenas o vetor de efeitos aditivos (a)**

a : vetor de efeitos genéticos aditivos;

$$G = A\sigma_a^2 ; \quad R = I\sigma_e^2, \text{ em que:}$$

σ_a^2 : variância genética aditiva;

A : matriz de correlação genética aditiva entre os indivíduos em avaliação;

σ_e^2 : variância residual.

- **Modelo univariado com medidas repetidas, ajustando os efeitos aditivos (a^*) e de ambiente permanente (p) (Modelo de Repetibilidade)**

$y = Xb + Za + e$, onde:

$$V(a^*) = A\sigma_a^2; \quad V(p) = I\sigma_p^2; \quad R = I\sigma_e^2, \text{ em que:}$$

$y = Xb + Z_1a^* + Z_2p + e$, sendo:

σ_p^2 : variância dos efeitos permanentes.

- **Modelo multivariado, ajustando os efeitos aditivos**

No caso bivariado tem-se que:

$$Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}; \quad a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix};$$

$$G = A \otimes G_0; \quad R = I \otimes R_0;$$

$$G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{21}} & \sigma_{a_2}^2 \end{bmatrix}; \quad R_0 = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{21}} & \sigma_{e_2}^2 \end{bmatrix} \text{ ou } R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}, \text{ em que:}$$

$\sigma_{a_{12}}$ = covariância genética aditiva entre os caracteres 1 e 2;

$\sigma_{e_{12}}$ = covariância ambiental entre os caracteres 1 e 2.

- **Avaliação em um delineamento em blocos ao acaso, com várias plantas por parcela, várias medições por indivíduo, avaliado em vários experimentos com tratamentos comuns**

Em um Modelo linear misto (modelo aditivo multivariado de repetibilidade), considerando um caso bivariado, tem-se que:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} W_1 & \\ & W_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Estrutura de médias e variâncias:

$$E \begin{bmatrix} y_1 \\ y_2 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} X_1 b_1 \\ X_2 b_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad Var \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ p_1 \\ p_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_{12}} & 0 & 0 & 0 & 0 & 0 & 0 \\ A\sigma_{a_{12}} & A\sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma_{c_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I\sigma_{c_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma_{p_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_{p_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I\sigma_{e_2}^2 \end{bmatrix};$$

em que:

$\sigma_{a_1}^2$ e $\sigma_{a_2}^2$: variâncias genéticas aditivas, nos locais (ou experimentos) 1 e 2, respectivamente;

$\sigma_{p_1}^2$ e $\sigma_{p_2}^2$: variância permanente dentro de parcelas (ambiental + genética não aditiva) nos locais 1 e 2, respectivamente.

$\sigma_{c_1}^2$ e $\sigma_{c_2}^2$: variâncias entre parcelas, nos locais 1 e 2, respectivamente;

$\sigma_{e_1}^2$ e $\sigma_{e_2}^2$: variâncias residuais, nos locais 1 e 2, respectivamente.

Equações do modelo misto:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & XR^{-1}W & X'R^{-1}T \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & ZR^{-1}W & Z'R^{-1}T \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} & W'R^{-1}T \\ T'R^{-1}X & T'R^{-1}Z & T'R^{-1}W & T'R^{-1}T + P^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{c} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \\ T'R^{-1}y \end{bmatrix},$$

em que:

$$X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix}; \quad W = \begin{bmatrix} W_1 & 0 & 0 \\ 0 & W_2 & 0 \\ 0 & 0 & W_3 \end{bmatrix}; \quad T = \begin{bmatrix} T_1 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_3 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}; \quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix}; \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{bmatrix}; \quad \hat{c} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_3 \end{bmatrix}; \quad \hat{p} = \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \end{bmatrix}$$

$$R^{-1} = R_0^{-1} \otimes I; \quad G^{-1} = G_0^{-1} \otimes A^{-1}; \quad C^{-1} = C_0^{-1} \otimes I; \quad P^{-1} = P_0^{-1} \otimes I;$$

$$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}; \quad G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_{12}} \\ \sigma_{a_{12}} & \sigma_{a_2}^2 \end{bmatrix}; \quad C_0 = \begin{bmatrix} \sigma_{c_1}^2 & 0 \\ 0 & \sigma_{c_2}^2 \end{bmatrix}; \quad P_0 = \begin{bmatrix} \sigma_{p_1}^2 & 0 \\ 0 & \sigma_{p_2}^2 \end{bmatrix}$$

Os componentes de variância estão associados aos parâmetros h^2 , c^2 , r e r_a , da seguinte maneira:

$$\sigma_{a_i}^2 = h_i^2 \sigma_{y_i}^2; \quad \sigma_{c_i}^2 = c_i^2 \sigma_{y_i}^2; \quad \sigma_{p_i}^2 = (p_i - h_i^2 - c_i^2) \sigma_{y_i}^2;$$

$$\sigma_{e_i}^2 = (1 - p_i) \sigma_{y_i}^2; \quad \sigma_{a_{ij}} = p_{a_{ij}} \sigma_{a_i} \sigma_{a_j}, \text{ em que } r_i \text{ é a repetibilidade no local } i.$$

$$p_{a_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i} \sigma_{a_j}} = \text{correlação genética entre o desempenho nos locais } i \text{ e } j;$$

$$\sigma_{y_i}^2 = \text{variância fenotípica ao nível de indivíduo no local } i.$$

A variância da interação genótipo x ambiente, é dada por:

$$\sigma_{ae}^2 = \frac{1}{2} (\sigma_{a_i} - \sigma_{a_j})^2 + (1 - p_{a_{ij}}) \sigma_{a_i} \sigma_{a_j}$$

Este tópico objetivou descrever, em termos práticos, os modelos mistos (BLUP) e suas estruturas de médias e variâncias, estimadores e preditores associados ao delineamento experimental deste trabalho. Tal modelo foi exemplificado conforme DE RESENDE (2000) e os principais modelos empregados no melhoramento de plantas podem ser conferidos na referência citada.

2.5. ANÁLISE DE TRILHA

A análise de trilha, nos casos em que se considera um único modelo causal, é simplesmente uma análise de regressão parcial padronizada, sendo útil o desdobramento dos coeficientes de correlação em efeitos diretos e indiretos. Entretanto, a análise de trilha constitui-se numa expansão de regressão múltipla, quando são envolvidas inter-relações complexas e, ou, vários diagramas causais, como é encontrado em estudos de melhoramento envolvendo a produção de grãos (CRUZ & REGAZZI, 1997).

Na análise de trilha as variáveis são previamente padronizadas, e além disso, quando utilizado o BLUP, na maioria das situações práticas, a média será um número entre zero e um, devido ao chamado “*shrinkage*”, relatado anteriormente (DUARTE & VENCOVSKY, 2001). Assim, os BLUP’s dos caracteres agrônômicos também são variáveis previamente padronizadas.

2.5.1. Estimação dos coeficientes de trilha (Efeitos diretos e Indiretos)

A padronização de uma variável é obtida dividindo-se o desvio de cada observação em relação à média pelo desvio padrão da amostra:

$$u_i = \frac{X_i - \bar{X}}{\hat{\sigma}_x}$$

Valem as seguintes propriedades em relação às variáveis padronizadas:

- i. Uma variável padronizada tem média igual a zero e variância igual a um;
- ii. A covariância entre duas variáveis padronizadas é igual à correlação entre estas variáveis (padronizadas ou não).

Serão consideradas, uma variável básica Y (variável resposta) e para exemplificar, três variáveis explicativas (X_1, X_2 e X_3), que se relacionam por meio do modelo:

$$Y - \bar{Y} = b_{o1}(X_1 - \bar{X}_1) + b_{o2}(X_2 - \bar{X}_2) + b_{o3}(X_3 - \bar{X}_3) + e \quad (2)$$

$$\frac{Y - \bar{Y}}{\sigma_y} = b_{o1} \frac{\sigma_{X_1}}{\sigma_y} \frac{(X_1 - \bar{X}_1)}{\sigma_{X_1}} + b_{o2} \frac{\sigma_{X_2}}{\sigma_y} \frac{(X_2 - \bar{X}_2)}{\sigma_{X_2}} + b_{o3} \frac{\sigma_{X_3}}{\sigma_y} \frac{(X_3 - \bar{X}_3)}{\sigma_{X_3}} + \frac{\sigma_e}{\sigma_y} e \quad (3)$$

da qual se obtém:

$$y = p_{01}x_1 + p_{02}x_2 + p_{03}x_3 + p_e u$$

em que:

$$y = \frac{Y - \bar{Y}}{\sigma_y}$$

$$x_i = \frac{(X_i - \bar{X}_i)}{\sigma_{X_i}}$$

$$u = \frac{e}{\sigma_e}$$

$$p_e = \frac{\sigma_e}{\sigma_y}$$

e

$$p_{0i} = \frac{b_{0i}\sigma_{X_i}}{\sigma_y}$$

Neste modelo é verificado que:

$$V(Y) = V(x_i) = V(u) = 1; \quad Cov(y, x_i) = r_{0i}$$

$$Cov(x_i, x_j) = r_{ij}; \quad Cov(u, x_i) = 0$$

Tendo em vista o modelo (3), verificam-se as seguintes relações:

$$\hat{V}(Y) = \hat{p}_{01}^2 + \hat{p}_{02}^2 + \hat{p}_{03}^2 + 2\hat{p}_{01}\hat{p}_{02}r_{12} + 2\hat{p}_{01}\hat{p}_{03}r_{13} + 2\hat{p}_{02}\hat{p}_{03}r_{23} + \hat{p}_e^2 \quad (4)$$

$$C\hat{ov}(y, x_1) = r_{01} = \hat{p}_{01} + \hat{p}_{02}r_{12} + \hat{p}_{03}r_{13};$$

$$C\hat{ov}(y, x_2) = r_{02} = \hat{p}_{01}r_{12} + \hat{p}_{02} + \hat{p}_{03}r_{23}; \quad (5)$$

$$C\hat{ov}(y, x_3) = r_{03} = \hat{p}_{01}r_{13} + \hat{p}_{02}r_{23} + \hat{p}_{03};$$

Em (4) pode ser estimado o coeficiente de determinação do modelo causal ($R_{0.123}^2$), que mede os efeitos das três variáveis explicativas (x_1, x_2 e x_3) sobre y , ou seja:

$$R_{0.123}^2 = \hat{p}_{0_1}^2 + \hat{p}_{0_2}^2 + \dots + 2\hat{p}_{0_2}\hat{p}_{0_3}r_{23}$$

Também em (4) estima-se o efeito da variável residual sobre a variável principal, dado por:

$$\hat{p}_e = \sqrt{1 - R_{0.123}^2}$$

Em (5) é apresentada a decomposição das correlações r_{0_i} em efeitos diretos de x_i sobre a variável básica, expresso por \hat{p}_{0_i} , e os efeitos indiretos de x_i via x_j , expresso por $\hat{p}_{0_j}r_{ij}$.

As seguintes propriedades têm sido associadas aos efeitos diretos da análise de trilha:

- iii. Sendo um coeficiente de regressão, ele tem direção, podendo ser negativo ou positivo e maior ou menor que a unidade;
- iv. Sendo um coeficiente de regressão padronizado, ele pode ser utilizado para comparar efeito de caracteres mensuráveis em diferentes escalas;
- v. Não tendo unidade física, ele se assemelha a um coeficiente de correlação.

A estimação dos efeitos diretos e indiretos é obtida pela solução do sistema de equações descrito em (5), da seguinte maneira:

$$\begin{bmatrix} r_{0_1} \\ r_{0_2} \\ r_{0_3} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} \begin{bmatrix} \hat{p}_{0_1} \\ \hat{p}_{0_2} \\ \hat{p}_{0_3} \end{bmatrix}$$

Os resultados da análise de trilha são de melhor interpretação por meio do diagrama apresentado pela Figura 6.

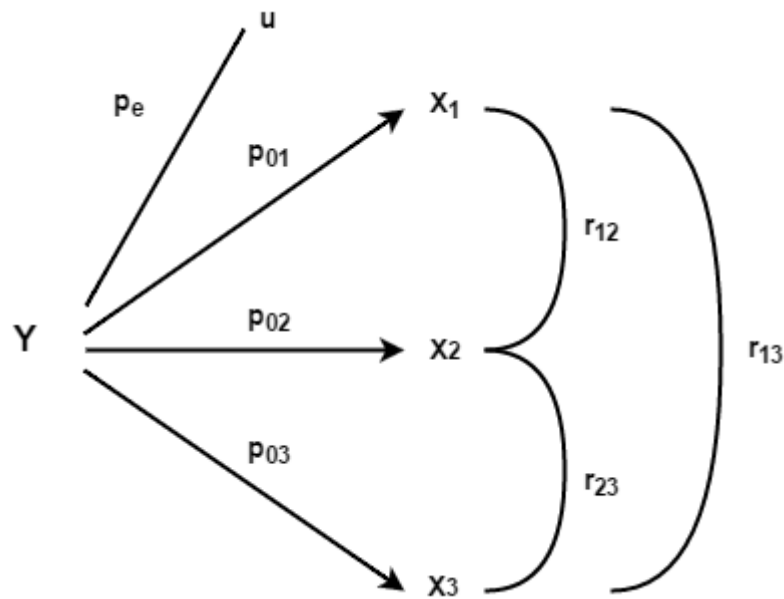


Figura 6 – Diagrama causal dos efeitos das variáveis explicativas (x_1, x_2 e x_3) e residual (u) sobre a variável dependente (Y).

Fonte: (CRUZ & REGAZZI, 1997).

No diagrama de trilha as setas unidirecionais indicam a influência direta de cada variável explicativa sobre a variável básica, e as setas bidirecionais mostram a interdependência dos componentes explicativos e determinam a trilha representativa do efeito indireto sobre a variável básica.

Deve-se ressaltar o fato de que os estimadores dos coeficientes de trilha, obtidos pelo processo descrito anteriormente, são de mínimos quadrados, pois provem, das equações normais $X'X\hat{\beta} = X'Y$, cujas matrizes são derivadas do modelo: $y = p_{01}x_1 + p_{02}x_2 + p_{03}x_3 + p_e u$. Ou seja,

$$X'Y = \begin{bmatrix} \sum x_{1i}y_i \\ \sum x_{2i}y_i \\ \sum x_{3i}y_i \end{bmatrix} = (n-1) \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \end{bmatrix}$$

$$X'X = \begin{bmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \sum x_{1i}x_{3i} \\ \sum x_{1i}x_{2i} & \sum x_{2i}^2 & \sum x_{2i}x_{3i} \\ \sum x_{1i}x_{3i} & \sum x_{2i}x_{3i} & \sum x_{3i}^2 \end{bmatrix} = (n-1) \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{p}_{01} \\ \hat{p}_{02} \\ \hat{p}_{03} \end{bmatrix}$$

As somas de quadrados associadas ao modelo estatístico são dadas por:

$$SQ(total) = Y'Y = \sum y_i^2 = (n - 1)$$

e

$$SQ(regressão) = \hat{\beta}X'Y = (n - 1) (\hat{p}_{01}r_{01} + \hat{p}_{02}r_{02} + \hat{p}_{03}r_{03})$$

Assim, uma maneira alternativa para obtenção do $R_{0.123}^2$ é;

$$R_{0.123}^2 = (\hat{p}_{01}r_{01} + \hat{p}_{02}r_{02} + \hat{p}_{03}r_{03})$$

que corresponde à determinação tanto do diagrama de trilha, quanto do modelo de regressão linear múltipla de Y em função de x_1, x_2 e x_3 .

A finalidade deste tópico foi descrever, em termos práticos, a metodologia da análise de trilha e suas estruturas, associadas ao modelo de regressão múltipla conforme demonstrado. Tal modelo foi exemplificado conforme CRUZ & REGAZZI (1997). Outros modelos de análise de trilha e aplicações, podem ser conferidos na referência citada.

3. METODOLOGIA

Os ensaios foram conduzidos em campos experimentais. Foram avaliadas 271 linhagens de parentais fêmea de polinização aberta, em 12 localidades do Brasil, os dados são desde a safra de inverno de 2013 até o verão de 2015.

Adotou-se o delineamento experimental em blocos casualizados, com três repetições. Cada parcela foi constituída de duas linhas de 5m espaçadas por 0,7m. A população de plantas nos ensaios é 85.000 plantas/ha no verão e 80.000 plantas/ha no inverno.

As plantas foram avaliadas, quanto à altura de plantas (PHT), altura da espiga (EHT), plantas dominadas (DOPLP), tombamento (TDPP), plantas acamadas (RTLP), espigas caídas (DRPP), uniformidade de plantas (UNI) e plantas sem espigas (BARP). Após a maturação dos grãos, foram avaliados o peso de espigas colhidas na parcela (SHW), peso de mil grãos (TKW), enrolamento do estigma (SBLP), espigas mal formadas (DFEAP), falha de granação (KST), polinização no cartucho (SDB), umidade de grãos (MST), produtividade de grãos (YLD) e produtividade final na unidade de sacos de 60 mil sementes vendáveis (SSU) e também avaliadas quanto a qualidade dos grãos em descarte (DSC), grãos ardidos (ROTGR), grãos rachados (SPLK) e viviparidade (VIVPY). Os dados de todos os caracteres foram submetidos à análise de variância. As estimativas de correlação fenotípica entre os caracteres agrônômicos e a produtividade foram obtidas e testadas a 5% de probabilidade pelo teste T-Student.

No modelo misto, os BLUP's por linhagem foram ajustados para efeitos de ano, safra e localidades. Após obtidos os BLUP's individuais por fêmea, realizou-se uma análise de variância para seleção dos principais caracteres e posteriormente feita uma análise das estimativas de correlações fenotípicas entre produtividade em SSU (variável principal) e os caracteres agrônômicos (variáveis explicativas) selecionadas e depois desdobradas em efeitos diretos e indiretos, por meio da análise de trilha (WRIGHT, 1921).

Todas as análises estatísticas foram realizadas através do *freeware* R (TEAM, R. 2013). A implementação computacional da metodologia de modelos mistos é baseada em métodos numéricos, notadamente em álgebra linear numérica visando a obtenção da solução iterativa das equações de modelo misto (obtenção dos BLUP's) e o *script* para a análise de trilha está disponibilizada no Anexo A deste trabalho.

4. RESULTADOS E DISCUSSÕES

A correlação entre caracteres que pode ser diretamente medida é a fenotípica (FALCONER, 1981). Essa correlação tem causas genéticas e ambientais, porém só a genética envolve uma associação de natureza herdável, podendo, por conseguinte, ser utilizada na orientação de programas de melhoramento (CRUZ & REGAZZI, 1997).

Contudo, para entender a contribuição de cada caractere para a resposta produtividade na unidade de sacos de 60 mil sementes vendáveis (SSU), foi realizada uma análise de regressão múltipla com o método de *forward* (passo a frente) para seleção dos caracteres. No método de *forward* cada variável é adicionada ao modelo de acordo com sua estatística F parcial até que não seja incluída mais nenhuma variável explicativa no modelo, tal método foi escolhido pois uma vez que a variável entra, ela não sai mais do modelo, assim, se retêm as variáveis mais correlacionadas com a variável resposta.

As diferenças entre variedades na maioria dos caracteres foram não significativas, dentre os 21 caracteres agronômicos avaliados, apenas quatro foram significativos a 1% de significância (Tabela 2). Os caracteres, cujas diferenças entre variedades foram significativas pela análise, foram selecionados para se estudar a relação de causa e efeito desses sobre a produção de sementes de milho, assim, selecionou-se apenas as variáveis: produtividade de grãos (YLD), peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR).

Tabela 2 – Resumo do método *forward* para a seleção dos caracteres agronômicos.

Variáveis	R ²	n° de variáveis		F-value	p-valor
		adicionadas			
YLD	0,80	1	2573,65	< 0,01	
TKW	0,92	2	925,28	< 0,01	
DSC	0,98	3	2133,01	< 0,01	
ROTGR	0,98	4	81,03	< 0,01	

As estimativas de correlação fenotípica foram avaliadas para as variáveis selecionadas. A magnitude das correlações entre os caracteres agronômicos avaliados variou de -0,31 a 0,82 (Tabela 3).

Tabela 3 – Coeficiente de correlação fenotípica entre os caracteres agrônômicos em estudo.

	ROTGR	DSC	TKW	YLD	SSU
ROTGR	1,00	0,11	-0,09	-0,22	-0,18
DSC	0,80	1,00	-0,49	0,00	0,05
TKW	0,73	0,08	1,00	0,17	-0,31
YLD	0,13	0,66	0,99	1,00	0,82
SSU	0,32	0,96	0,50	0,04 *	1,00

Legenda: Acima da diagonal principal, correlações; Abaixo da diagonal principal, p-valores; () Significativo a 5% de significância.*

Neste estudo, o caractere avaliado em genótipos de milho, que apresentou maior correlação (acima 0,40), em magnitude, com a produtividade final na unidade de sacos de 60 mil sementes vendáveis (SSU) foi apenas a produtividade em grãos (YLD) com 0,82 e significativa testada a 5% significância pelo teste T-Student.

Em CRUZ & REGAZZI (1997) é apontado que o coeficiente de correlação baixo ou igual a zero não implica falta de correlação entre duas variáveis, apenas reflete a ausência de relação linear entre estes caracteres avaliados, logo não é um critério para a análise de trilha.

Para melhor entendimento dos inter-relacionamentos entre os caracteres, foi realizada a análise de trilha. As magnitudes e os sinais dos efeitos diretos e indiretos dos componentes primários (variáveis explicativas) sobre a produtividade (SSU) estão indicados na Figura 7 pelo diagrama causal.

Com um R^2 de aproximadamente 93%, pode-se afirmar que cerca de 93% do efeito sobre a variável SSU consegue e ser explicado pelas variáveis YLD, TKW, DSC e ROTGR. Pode-se notar por meio da Tabela 2, que não há incremento no R^2 com a adição da variável ROTGR.

A partir dos resultados da análise de trilha (Tabela 4), nota-se que o caractere produtividade de grãos (YLD) é o que apresenta maior efeito direto em sentido favorável sobre a produtividade em sacos de sementes vendáveis com 0,92. Com isto, pode-se afirmar que YLD é o melhor caractere secundário para a seleção indireta visando maior produtividade, na unidade de sacos de 60 mil sementes vendáveis (SSU).

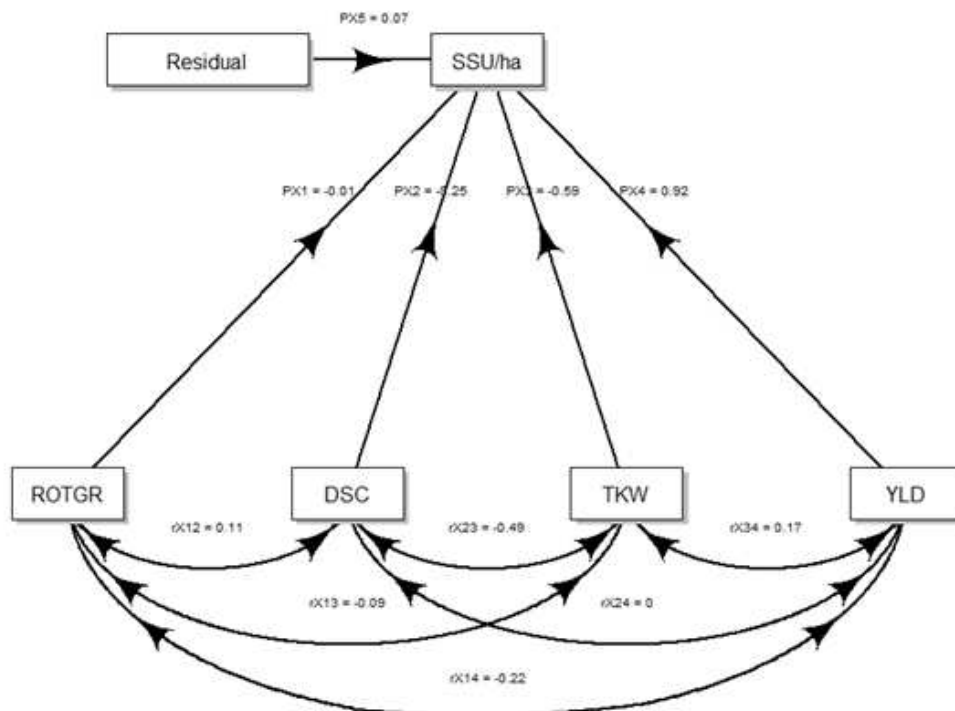


Figura 7 – Diagrama causal dos efeitos das variáveis explicativas produtividade de grãos (YLD), peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR) e residual sobre a variável resposta produtividade na unidade de sacos de sementes vendáveis (SSU).

Tabela 4 – Efeitos diretos e indiretos das variáveis explicativas produtividade de grãos (YLD), peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR) e residual sobre a variável resposta produtividade na unidade de sacos de sementes vendáveis (SSU).

Caractere	Efeito	Via	SSU	Caractere	Efeito	Via	SSU
YLD	Direto		0.92	TKW	Direto		-0.59
	Indireto	TKW	-0.10		Indireto	DSC	0.12
		DSC	-0.00			ROTGR	0.00
		ROTGR	0.00			YLD	0.16
DSC	Direto		-0.23	ROTGR	Direto		-0.01
	Indireto	TKW	0.29		Indireto	TKW	0.05
		ROTGR	-0.00			DSC	-0.03
		YLD	0.00			YLD	-0.20
R²			0,93				

Um outro fator a ser considerado do YLD, é a alta correlação de 0,82. De acordo com FALCONER (1981), uma das causas para altas correlações é o pleiotropismo (um gene responsável por várias características), ou seja, um mesmo gene influencia na expressão de mais de um caráter. Essa informação é útil no melhoramento de plantas, pois favorece a seleção simultânea de dois ou mais caracteres, pela seleção em apenas um destes.

O caractere grãos ardidos (ROTGR) apresenta um efeito direto significativamente baixo, indicando que a sua utilização como auxiliar para seleção pode não proporcionar ganhos satisfatórios na variável básica, produtividade (SSU), conforme Tabela 4. No entanto, existe um efeito indireto de -0,20 com relação ao YLD. Assim, ROTGR pode ser um caractere que a ser considerado simultaneamente com YLD no processo de seleção.

O peso de mil sementes (TKW) e descarte (DSC) apresentaram efeitos diretos contrários sobre a produtividade em sacos de sementes vendáveis (SSU), o que era esperado, pois TKW e DSC são variáveis que impactam negativamente na produtividade. Descarte pois são sementes descartadas em função da sanidade do grão e TKW, pois como é representada pelo peso do grão, quanto mais pesado o grão, menos sementes haverão na espiga, impactado em uma menor quantidade de sementes no rendimento final, uma vez que a produtividade final é medida em sacos de 60 mil sementes vendáveis (SSU).

5. CONSIDERAÇÕES FINAIS

O estudo das influências dos componentes primários sobre a produtividade pode não ter grandes implicações práticas, pois esses caracteres geralmente são igualmente complexos, de baixa herdabilidade (relação entre a variância genotípica e a variância fenotípica) e, portanto, difíceis de serem selecionados (CRUZ & REGAZZI, 1997).

No entanto, apesar de todos os caracteres avaliados serem de grande importância para a seleção em linhagens de milho, no parental fêmea. A seleção proposta pela análise de trilha realizada pode resultar não somente em melhores ganhos genéticos, mas como também permitem direcionar a redução de avaliações em campo nos estágios iniciais de testes, que geram elevado custo e mão de obra, mantendo o foco em um maior rendimento de grãos.

Para fins de melhoramento, é importante identificar, dentre os caracteres aqueles de maior efeito direto em sentido favorável à seleção, de tal forma que a seleção indireta seja eficiente. Desta forma, nas condições experimentais, pode-se concluir que o caractere produtividade de grãos (YLD) é melhor caractere secundário para a seleção indireta. No entanto nas fases iniciais de testes em parentais fêmeas de linhagens de milho, visando a redução de avaliações em campo, é importante que sejam avaliados não somente o YLD, mas também o peso de mil sementes (TKW), descarte (DSC) e grãos ardidos (ROTGR), pois dentre todos os 21 caracteres, estes demonstraram serem os mais significativos para a produtividade final na unidade de sacos de 60 mil sementes de milho.

6. REFERÊNCIAS

ALLARD, R. W. Aspectos gerais da heterozigose e da depressão causada pela endogamia. In: Princípios do melhoramento genético de plantas. São Paulo: Edgard Blucher. p.381, 1971.

BACHIREDDY, V. R.; PAYNE Jr. R.; CHIN, K. L.; KANG, M. S. Conventional selection versus methods that use genotype x environment interaction in sweet corn trials. HortScience, v.27, n.5, p.436-438, 1992.

BUENO, L. C. S.; MENDES, A. N. G.; CARVALHO, S. P. Melhoramento Genético de Plantas: Princípios e Procedimentos. Lavras: UFLA, p.319, 2006.

CONAB. Levantamento de Safras. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1253&t=2>>. Acessos em abril, maio e junho de 2015.

CRUZ, C. D.; REGAZZI, A. J. Modelos biométricos aplicados ao melhoramento genético. 2ª ed. rev. Viçosa: Editora UFV, p.390, 1997.

DA ROCHA, F.; PEREIRA, B. F. Predição de Híbridos e Macho Esterilidade Genético Citoplasmática. Disponível em: <<http://www.genetica.esalq.usp.br/lgn0313/jbp/Predi%E7%E3o%20de%20H%Edbridos%20e%20Macho%20Esterilidade%20Gen%E9tico%20Citoplasm%E1tica.pdf>>. Acesso em junho de 2016.

DE RESENDE, M. D. V.. Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes. Embrapa Florestas, 2000.

DUARTE, J. B.; VENCOVSKY, R. Estimacão e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. Scientia Agricola, v. 58, n. 1, p. 109-117, 2001.

DUARTE, J. B.; VENCOVSKY, R. Interaçãõ genótipos x ambientes: uma introdução à análise AMMI. Ribeirão Preto: Sociedade Brasileira de Genética, p.60, 1999.

FALCONER, D. S.; MACKAY, T. F. C. Introduction to quantitative genetics. 4.ed. England: Longman, p.463, 1996.

FALCONER, D. S. Introdução à genética quantitativa. Viçosa: UFG, Imprensa Universitária, p.279, 1981.

FERRAUDO, G. M. Comparação de modelos mistos, AMMI e Eberhart-Russel via simulação no estudo da interação genótipo x ambiente em cana-de-açúcar. 2013.

FIESP. Safra Mundial de Milho 2016/17 – 2º Levantamento do USDA. Disponível em: <http://www.fiesp.com.br/indices-pesquisas-e-publicacoes/safra-mundial-de-milho2/attachment/boletim_milho_junho2016/>. Acesso em junho de 2016.

GAMA, E. E. G. et al. O programa de melhoramento de milho do Centro Nacional de Pesquisa de Milho e Sorgo CNPMS-EMBRAPA, Brasil. 1996.

GAUCH Jr., H. G.; ZOBEL, R. W. AMMI analysis of yield trials. In.: KANG, M.S.; GAUCH Jr., H.G. (eds.) Genotype-by-environment interaction. Boca Raton: CRC Press, 1996. p. 85-122.

GORMAN, D. P.; KANG, M. S.; MILAM, M. R. Contribution of weather variables to genotype x environment interaction in grain sorghum. Plant Breeding, v.103, n.4, p.299-303, 1989.

GUIA DO MILHO. Tecnologia do campo à mesa. Conselho de informação sobre biotecnologia. 2006. Disponível em: <http://www.cib.org.br/pdf/guia_do_milho_CIB.pdf>. Acesso em junho de 2016.

HENDERSON, C. R. Selection index and expected genetic advance. Statistical genetics and plant breeding, v. 982, p. 141-163, 1963.

HENDERSON, C R. Best linear unbiased estimation and prediction under a selection model. Biometrics, Raleigh, v. 31, n. 2, p. 423-447, June 1975.

HUSSEIN, M. A.; BJORNSTAD, A.; AASTVEIT, A. H. SASG x ESTAB: a SAS program for computing genotype x environment stability statistics. *Agronomy Journal*, v.92, n.3, p.454-459, 2000.

JOHANNSEN W. The genotype conception of heredity. *American Naturalist*, p.129-159, 1911.

KANG, M. S.; GAUCH Jr., H.G. (eds.) *Genotype-by-environment interaction*. Boca Raton: CRC Press, p.416, 1996.

KOSHIMA, F. A. T. Estabilidade e adaptabilidade para caracteres de produção em linhagens de milho. 2009.

MARIOTTI, J. A.; CLARIANA, R. The nature of $G \times E$ interactions in sugarcane hybrid progenies. *Sugar Cane*, 1994.

MINISTÉRIO DA AGRICULTURA. Projeções do Agronegócio - Brasil 2014/15 a 2024/2025. Disponível em: <http://www.agricultura.gov.br/arq_editor/PROJECOES_DO_AGRONEGOCIO_2025_WEB.pdf>. Acesso em junho de 2016.

NUNES, J. A. R.; RAMALHO, M. A. P.; FERREIRA, D. F. Inclusion of genetic relationship information in the pedigree selection method using mixed models. *Genetics and Molecular Biology*, v. 31, n. 1, p. 73-78, 2008.

PATERNIANI, E.; MIRANDA FILHO, J. B. Melhoramento de populações. *Melhoramento e produção do milho*, Campinas, v. 1, p. 217-264, 1987.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, v.58, p.545-554, 1971.

SANTOS, J.; VENCOSKY, R. Correlação fenotípica e genética entre alguns caracteres agronômicos do feijoeiro (*Phaseolus vulgaris*, L.). *Ciência e Prática*, São Paulo, v.10, n.3, p.265-272, 1986.

SEARLE, S. R.; CASELLA, G.; McCULLOCH, C.E. Variance components. New York. J. Wiley, p.528, 1992.

SHUKLA, G. K. Use of regression approach in the analysis of genotypeenvironmental interaction. Journal of the Indian Society of Agricultural Statistics, Interação de genótipos com ambientes 45 v. 35, n.2, p.129-138, 1983.

SHULL, G. H. A pure line method of corn breeding. Report “American Breeders Association”, Washington, v.5, p. 51-59, 1909.

TEAM, R. Core et al. R: A language and environment for statistical computing. 2013.

VENCOVSKY, R.; BARRIGA, P. Genética biométrica no fitomelhoramento. Ribeirão Preto: Sociedade Brasileira de Genética, p.496 p, 1992.

WRIGHT, S. Correlation and causation. Journal Agriculture Research, Islamabad, v.20, p.577-585, 1921.

7. ANEXO A – ROTINA R PARA ANÁLISE DE TRILHA

```
#####
###                               Universidade Federal de Uberlândia          ###
###                               Faculdade de Matemática                    ###
###                               Trabalho de Conclusão de Curso–Bacharelado em Estatística  ###
###                               Autor: Kamylla Rodrigues Leandro              ###
#####
# Bibliotecas
require(pathdiagram)
library(colortools)
library(plspm)
library(semPlot)
library(agricolae)

#####
### Análise de Trilha

cor.x<-data.cor.x # Data frame com as correlações das variáveis explicativas
cor.y<-data.cor.y # Vetor com as correlações da variável resposta

cor.x<-as.matrix(cor.x, mode = "numeric")
cor.y<-as.matrix(cor.y, mode = "numeric")

Path<-path.analysis(cor.x, cor.y)

#Efeitos diretos
diag (Path $ Coef)

Path $ Residual

#Salvando os resultados
resultados<-as.data.frame(Path)
write.table(resultados,file="result.path.txt")

#####
### Diagrama

par(mar = c(1, 1, 1, 1))
openplotmat()

# Coordenadas
coordenadas <- coordinates (c(2, length(cor.y)))
coordenadas[2,1] <- abs((coordenadas[1,1]+coordenadas[1,2])/2) #residual

#Especificando a posição das setas
#1 Residuo para var. dependente
ft1 <- matrix(ncol = 2, byrow = TRUE, data = c(1, 2))
```

```

#2 Var. independente para dependente
ft2 <- matrix(ncol=2, byrow = FALSE,
             data= c(seq((2+length(cor.y)))[3:(length(cor.y)+2)], rep(2, length(cor.y))))
#3 Para corr. x
a <- t(combn(seq((2+length(cor.y)))[3:(length(cor.y)+2)],2))
#4 Distancias
b <- rbind(ft1,ft2)

# Plot distancia
nr <- nrow(b)
arrpos <- matrix(ncol = 2, nrow = nr)
for (i in 1:nr)
  arrpos[i, ] <- straightarrow (to = coordenadas[b[i, 2], ],
                              from = coordenadas[b[i, 1], ],
                              lwd = 2, arr.pos = 0.6, arr.length = 0.5)

# Legendas
text(arrpos[1, 1], arrpos[1, 2] + 0.05,
     paste("P", "X", nrow(cor.x)+1, " = ", round(Path$Residual, 2), sep=""), cex=0.6)

nr <- nrow(arrpos)
for(i in 2:nr){
  text(arrpos[i, 1], arrpos[i, 2] + 0.05,
       paste("P", "X", i-1, " = ", round(diag(Path$Coeff)[i-1], 2), sep=""), cex=0.6)
}

# Plot – flexas de correlação direção 1
nr <- nrow(a)
arrpos <- matrix(ncol = 2, nrow = nr)
for (i in 1:nr)
  arrpos[i, ] <- curvedarrow (to = coordenadas[a[i, 2], ],
                              from = coordenadas[a[i, 1], ],
                              lwd = 2, arr.pos = 0.8, arr.length = 0.5, curve = 0.35)

# Plot – flexas de correlação direção 2
nr <- nrow(a)
arrpos <- matrix(ncol = 2, nrow = nr)
for (i in 1:nr)
  arrpos[i, ] <- curvedarrow (to = coordenadas[a[i, 1], ],
                              from = coordenadas[a[i, 2], ],
                              lwd = 2, arr.pos = 0.8, arr.length = 0.5, curve = -0.35)

# Combinações de corr.x para rxy ...rotulagem em setas de correlação
rcomb <- as.data.frame(t(combn(seq(nrow(cor.x)),2)))
rcomb <- paste(rcomb$V1,rcomb$V2, sep="")

# Lendas
nr <- nrow(a)
arrpos <- matrix(ncol = 2, nrow = nr)
for (i in 1:nr)

```

```

arrpos[i, ] <- curvedarrow (to = coordenadas[a[i, 1], ],
                           from = coordenadas[a[i, 2], ],
                           lwd = 2, arr.pos = 0.5, lcol = "transparent", arr.length = 0.5, curve = -0.35)

nr <- nrow(arrpos)
for(i in 1:nr){
  text(arrpos[i, 1], arrpos[i, 2] + 0.05,
        paste("r", "X", rcomb[i], " = ", round(as.dist(cor.x)[i], 2), sep=""), cex=0.6)
}

# Legenda Residuo
textrect (coordenadas[1,], 0.09, 0.03,lab = "Residual", box.col = "white",
          shadow.col = "grey", shadow.size = 0.005, cex = 1)
# Legenda var dependente
textrect (coordenadas[2,], 0.09, 0.03,lab = attributes(data)$class, box.col = "white",
          shadow.col = "grey", shadow.size = 0.005, cex = 1)
# Legenda var independiente
nr <- nrow(coordenadas)
for (i in 3:nr){
  textrect (coordenadas[i,], 0.09, 0.03,lab = colnames(data)[i-2], box.col = "white",
            shadow.col = "grey", shadow.size = 0.005, cex = 1)
}
#####

```