

# Análise de trilha na relação entre características morfológicas do milho e sua produtividade de grãos

Tadeu Vilela de Souza <sup>14</sup>

Sílvio de Castro Silveira <sup>2</sup>

João Domingos Scalon <sup>3</sup>

## 1 Introdução

O estudo da relação entre caracteres da planta de milho e sua produtividade é tema de estudo em diversos trabalhos, conduzidos em centros de pesquisa sobre melhoramento genético, objetivando encontrar caracteres que sejam mais relacionados com a produtividade de grãos. Em sua grande parte, é utilizada a análise de correlação nesses estudos, entretanto, deve-se ressaltar que o estudo de correlações entre caracteres não permitem tirar conclusões sobre relações de causa e efeito, ou seja, a correlação é apenas uma medida de associação. Pela importância de se conhecer as relações de causa e efeito entre variáveis, Wright (1921) desenvolveu a análise de trilha (“*path analysis*”), esse método desdobra as correlações entre cada par de variável de um conjunto em efeitos diretos e indiretos. Com esse desdobramento, é possível quantificar a influência direta de uma variável sobre a outra, independente das demais. Sua utilização apropriada depende do conjunto de caracteres estudados, sendo indispensável a elaboração de um diagrama de causa e efeito, que exhibe graficamente o padrão de hipótese das relações de causa e efeito entre um conjunto de variáveis, ou seja, representa a relação dos caracteres explicativos sobre uma característica principal (resposta).

Existem muitos estudos que utilizam a análise de trilha para avaliar qual ou quais características de plantas de milhos híbridos influenciam de maneira direta a sua produtividade de grãos. A maioria desses trabalhos, como Ivanovic e Rosic (1985), Ottaviano e Camussi (1981), Balbinot Jr et al. (2005), estudam os efeitos diretos e indiretos através da relação entre componentes de rendimento do milho e sua produtividade. Porém, observa-se que não é considerado a influência que características morfológicas da planta exercem na sua produtividade.

Este trabalho teve como objetivo identificar caracteres morfológicos da planta que mais contribuem para a produtividade de grãos. Através do desdobramento, em efeitos diretos e indiretos, das correlações entre esses componentes morfológicos avaliados e a produtividade, buscou-se determinar a importância relativa de cada desses componentes na determinação dessa produtividade.

---

<sup>1</sup>DEX - UFLA. e-mail: [tadeu\\_vil\\_souza@yahoo.com.br](mailto:tadeu_vil_souza@yahoo.com.br)

<sup>4</sup>Agradecimento a Fapemig e Capes pelo apoio financeiro.

<sup>2</sup>DEX - UFLA.

<sup>3</sup>DEX - UFLA.

## 2 Material e métodos

O experimento foi conduzido na fazenda experimental da Universidade Federal de Lavras, em Lavras, a 951 m de altitude, nas coordenadas 44°58' longitude Oeste e 21°12' latitude Sul.

Foram escolhidas cinco linhagens de milho do programa de melhoramento da UFLA, obtendo-se 15 tratamentos genéticos. O delineamento utilizado foi o de blocos completos ao acaso com quatro repetições. As parcelas eram constituídas de duas linhas de 2,0 m de comprimento, com espaçamento entre linhas de 0,6 m e quatro plantas por metro linear.

No final do período do florescimento masculino, foram realizadas medições de cinco plantas competitivas no interior da parcela para altura de planta (AP), em metros, considerando a distância do solo ao ponto de inserção da folha bandeira, altura de espiga (AE), em metros, do ponto de inserção da espiga superior formada no colmo e o diâmetro do colmo (DC), em centímetros, a partir do primeiro entrenó do colmo acima do solo, utilizando-se um paquímetro digital.

### 2.1 Escolha do diagrama causal

Determinou-se o diagrama causal combinando duas técnicas: hipótese formulada mediante conhecimento prévio das relações entre as variáveis em estudo e o método da correlação parcial (GOLDSMITH, 1977). Depois de estabelecer a hipótese da relação de causa e efeito entre as variáveis, foi aplicado o método da correlação parcial nas variáveis envolvidas no modelo com o objetivo de verificar a consistência dessa formulação de causa e efeito. De acordo com esse critério, se a diferença entre  $r_{ij}$  (correlação linear simples entre as variáveis  $X_i$  e  $X_j$ ) e  $r_{ij.k}$  (correlação parcial linear entre  $X_i$  e  $X_j$ , controlando para  $X_k$ ) é alta, então pode ser aceito que  $X_k$  participa da trilha causal entre  $X_i$  e  $X_j$ . Essa diferença é escrita como  $|r_{ij} - r_{ij.k}| = \Delta_{ij}$ , e sua significância estatística pode ser avaliada pela transformação de Fisher:

$$z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$$

que é aproximadamente normal com  $\sigma(z) = \sqrt{1/(n-3)}$ , onde  $n$  é o número de observações independentes.

Segundo Vasconcelos, Almeida e Nobre (1998), na prática, se  $\Delta_{ij} < \sigma$  implica que  $X_k$  não influencia a relação entre  $X_i$  e  $X_j$ , enquanto  $\Delta_{ij} > 2\sigma$  implica que  $X_k$  é uma parte importante dessa relação.

### 2.2 Estimação e desdobramento das correlações

Para estimar os coeficientes de trilha é necessário estudar a existência de multicolinearidade entre as variáveis independentes que compõem o modelo. Quando não existe multicolinearidade entre as variáveis independentes, os parâmetros dos modelos são estimados pelo método dos mínimos quadrados resolvendo o seguinte sistema  $(X'X)\beta = X'Y$ , em que  $X'X$  é a matriz de

correlações entre as variáveis independentes;  $\beta$  é o vetor dos estimadores dos coeficientes de trilha; e  $X'Y$  é a matriz de correlações entre a variável dependente com cada variável independente do modelo de regressão. Mas, a presença de multicolinearidade entre as variáveis independentes ocasiona problemas nas estimativas dos coeficientes de trilha, impossibilitando a utilização dos estimadores de mínimos quadrados. Nessa situação, utiliza-se como metodologia a regressão em crista, e os coeficientes de trilha são obtidos pela solução da equação  $(X'X + cI)\beta^* = X'Y$ , em que  $c$  é uma pequena quantidade adicionada aos elementos da diagonal da matriz  $X'X$ ;  $\beta^*$  é o vetor dos estimadores dos coeficientes de trilha; e  $I$  é a matriz identidade.

O grau de multicolinearidade da matriz singular  $X'X$  foi estabelecido com base no seu número de condições (MONTGOMERY e PECK, 1992). Se  $NC < 100$ , a multicolinearidade foi tida como fraca e não constituiu problema sério na análise. Se  $100 \leq NC < 1000$ , a multicolinearidade foi considerada de moderada a forte, e  $NC \geq 1000$  foi indicativo de multicolinearidade severa.

Neste estudo, estimou-se e desdobrou-se a correlação simples usando o estimador da correlação linear de Pearson. Considerou-se como valores nas variáveis a média de cada tratamento dos experimentos. Depois de desdobradas essas correlações, fez-se inferências analisando os efeitos diretos e indiretos das variáveis explicativas sobre a variável básica.

Todas análises e testes estatísticos foram realizados utilizando funções desenvolvidas e existentes no software R (R DEVELOPMENT CORE TEAM, 2014).

### 3 Resultados e discussões

A suposição de normalidade das variáveis foi avaliada pelo teste de Shapiro-Wilk, onde todas foram consideradas normais (PROD, *valor - p* = 0,12), (AP, *valor - p* = 0,07), (AE, *valor - p* = 0,11), (DC, *valor - p* = 0,10). Através da construção de gráficos do tipo *boxplot* para cada uma das variáveis verificou-se não existir *outliers* entre os dados.

Na análise de trilha foi usado um diagrama causal simples, Figura 1. Nessa representação, a variável dependente é a produção de grãos (PROD) e as variáveis independentes são a altura da planta (AP), a altura de espiga (AE) e o diâmetro do colmo da planta (DC). Através do método da correlação parcial foi possível verificar que todas as variáveis que esse diagrama causal desempenham uma parte importante nessa relação.

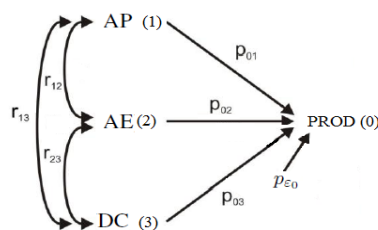


Figura 1: Diagrama causal da análise de trilha.

Avaliando a multicolinearidade entre as variáveis explicativas, usando o teste dos autovalores

e autovetores da matriz de correlação, foi obtido o número de condições igual a 2,5 ( $NC = 2,5$ ), ou seja, uma multicolinearidade fraca. Dessa forma, sem a necessidade de se controlar a multicolinearidade, foi utilizado o método dos mínimos quadrados ordinários para a estimação dos coeficientes do modelo.

Depois de estimados os coeficientes, analisando os resíduos do modelo ajustado, a independência foi confirmada pelo teste de Durbin Watson ( $valor - p = 0,21$ ); a normalidade pôde ser verificada usando o teste de normalidade Shapiro-Wilk ( $valor - p = 0,92$ ); e também se verificou a homocedasticidade pelo teste de Breusch-Pagan ( $valor - p = 0,95$ ). Pela distância de *cook* foi verificada a não existência de pontos influentes.

As correlações estimadas entre as variáveis independentes e a variável dependente foram desdobradas em efeitos diretos e indiretos das variáveis explicativas sobre a produção. Essa correlações e coeficientes de trilha encontram-se na Tabela 1. Pode-se verificar pelo coeficiente de determinação ( $R_{0,123}^2$ ) que estas variáveis explicaram 90,62% da variação da produção de grãos.

Analisando os efeitos diretos e indiretos dos componentes morfológicos sobre a variável principal, verificou-se que as variável AE apresentou o maior efeito direto sobre a produção de grãos (0,456), e efeitos indiretos baixos. Ainda, a variável AE foi a que apresentou a maior correlação com PROD, podendo ser considerada a mais determinante na variação da produção de grãos. Churata e Ayala-Osuna (1996) em seu trabalho, também encontraram um resultado parecido, onde a altura de espiga foi uma das variáveis explicativas mais fortemente relacionada com a produção, com efeito direto de 0,39. A variável AP apresentou efeitos diretos e indiretos baixos, indicando não influenciar de maneira significativa a produtividade. A variável DC apresentou efeito direto alto (-0,428), indicando exercer grande influência na produtividade.

Tabela 1: Estimativas dos efeitos diretos e indiretos das variáveis explicativas sobre a variável básica produção de grãos (PROD).

Variáveis primárias	Vias de associação	Estimador	Estimativa
AP	Efeito direto sobre PROD	$\hat{p}_{01}$	0,287
	Efeito indireto via AE	$\hat{p}_{02}r_{12}$	0,155
	Efeito indireto via DC	$\hat{p}_{03}r_{13}$	0,137
	Total	$r_{01}$	0,58
AE	Efeito direto sobre PROD	$\hat{p}_{02}$	0,456
	Efeito indireto via AP	$\hat{p}_{01}r_{12}$	0,097
	Efeito indireto via DC	$\hat{p}_{03}r_{23}$	0,055
	Total	$r_{02}$	0,61
DC	Efeito direto sobre PROD	$\hat{p}_{03}$	-0,428
	Efeito indireto via AP	$\hat{p}_{01}r_{13}$	-0,092
	Efeito indireto via AE	$\hat{p}_{02}r_{23}$	-0,059
	Total	$r_{03}$	-0,58
$R_{0,123}^2$			0,9062
Efeito residual ( $\hat{p}_\epsilon$ )			0,306

## 4 Conclusões

A variável AE foi a que apresentou a maior correlação com PROD (0,61) e o maior efeito indireto sobre a PROD (0,45), podendo ser considerada a mais determinante na variação da produção de grãos. Sendo assim, é a característica mais indicada para seleção indireta para a PROD. Dessa forma, a seleção de plantas com maior AE possibilita a obtenção de plantas com maior produtividade de grãos.

A variável DC apresentou efeito direto alto (-0,428) e correlação com a PROD negativa (-0,58). Dessa forma, torna-se uma característica influente na produção de grão, indicando que plantas com maior colmo apresenta menor produtividade.

## Referências

- [1] BALBINOT JR, A. A. et al. Contribuição de componentes de rendimento na produtividade de grãos em variedades de polinização aberta de milho. **Revista Brasileira de Agrociência**. Pelotas, v. 11, n. 2, p. 161-166, abr-jun, 2005.
- [2] CHURATA, B. G. M.; AYALA-OZUNA, J. T. Correlações genotípica, fenotípica e de ambiente e análise de trilha em caracteres avaliados no composto de milho (*Zea mays*) arquitetura. **Revista Ceres**, Viçosa, MG, v. 43, n. 249, p. 628-636, 1996.
- [3] GOLDSMITH, J. R. Paths of association in epidemiological analysis: application of confounding factor, since no statistical test or procedure to health effects of environmental exposures. **International Journal of Epidemiology**, Oxford, v. 6, n. 4, p. 391-399, 1977.
- [4] IVANOVIC, M.; ROSIC, K. Path coefficient analysis for three stalk traits and grain yield in maize (*Zea mays* L.). **Maydica**, Bergamo, v. 30, p. 233-239, 1985.
- [5] MONTGOMERY, D. C.; PECK, E. A. **Introduction to linear regression analysis**. 2<sup>nd</sup> ed. New York: J. Wiley, 1992. 544 p.
- [6] OTTAVIANO, E.; CAMUSSI, A. Phenotypic and genetic relationships between yield components in maize. **Euphytica**, Milano, v. 30, n. 3, p. 601-609, dec. 1981.
- [7] R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <http://www.r-project.org>. Acesso em: 20 fev. 2014.
- [8] VASCONCELOS, A. G. G.; ALMEIDA, M. V. A.; NOBRE, F. F. The path analysis approach for the multivariate analysis of infant mortality data. **Annals of Epidemiology**, New York, v. 8, n. 4, p. 262-271, May 1998.
- [9] WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**, Washington, v. 20, n. 7, p. 557-585, Jan. 1921.