

ANÁLISE DE TRILHA SOB MULTICOLINEARIDADE EM PIMENTÃO¹

CLAUDIO GUILHERME PORTELA DE CARVALHO², VALTER RODRIGUES OLIVEIRA³,
COSME DAMIÃO CRUZ⁴ e VICENTE WAGNER DIAS CASALI⁵

RESUMO - Os objetivos deste trabalho foram: a) desdobrar correlações genótípicas em efeitos diretos e indiretos de componentes primários e secundários sobre a produção de frutos, em genótipos de pimentão (*Capsicum annuum* L.), por meio de análises de trilha; b) comparar métodos alternativos de estimação dos coeficientes de trilha para contornar os efeitos adversos da multicolinearidade. Com base nas magnitudes dos efeitos diretos e indiretos, nenhum dos componentes primários ou secundários mostrou ser importante na seleção dos genótipos mais produtivos. Em populações segregantes, o caráter altura da planta poderá ser útil em resposta correlacionada, por meio de seleção indireta. Na ocorrência de multicolinearidade de moderada a severa entre os caracteres, a eliminação de variáveis do modelo de regressão e a análise de trilha em crista mostraram-se eficientes para reduzir as altas variâncias dos coeficientes de trilha encontradas. Contudo, quando várias variáveis necessitam ser eliminadas, o processo de descarte torna-se difícil e muitas análises podem ser necessárias. A análise de trilha em crista, por sua vez, deve contornar os efeitos adversos da multicolinearidade, sem que haja grande redução nos valores dos coeficientes de determinação dos modelos de regressão.

Termos para indexação: *Capsicum annuum*, estudo de correlações, melhoramento genético.

PATH ANALYSIS UNDER MULTICOLLINEARITY IN GREEN PEPPER

ABSTRACT - The main goals of this work were: a) to display genotype correlations in direct and indirect effects of primary and secondary components on the production of fruits, in green pepper (*Capsicum annuum* L.) genotypes, by mean of path analysis; b) to compare alternative methods of estimation of the path coefficients to avoid the adverse effects of multicollinearity. Based on the magnitude of direct and indirect effects, none of the primary or secondary components showed to be more important in the selection of more productive genotypes. In segregating populations, the character height of the plant might be useful in correlated response, by means of indirect selection. In the occurrence of moderate to severe multicollinearity between the characters, the elimination of variables from the regression model and the ridge path analysis showed to be efficient in reducing the high variances of the path coefficients found. However, when several variables need to be eliminated, the disposal process becomes difficult and many analysis might be necessary. Ridge path analysis, although, should by-pass the adverse effects of multicollinearity, without having great reduction on the coefficient determination values of the regression models.

Index terms: *Capsicum annuum*, correlation studies, genetic improvement.

INTRODUÇÃO

Em programas de melhoramento genético, o conhecimento da correlação entre caracteres é importante quando se deseja fazer seleção simultânea de

caracteres, ou quando um caráter de interesse apresenta baixa herdabilidade, problemas de medição ou de identificação. Neste caso, ao selecionar outro caráter de alta herdabilidade, de fácil medição, de fácil identificação e que apresenta alta correlação com o caráter desejado, o melhorista poderá obter progressos mais rápidos em relação ao uso de seleção direta.

A quantificação e a interpretação da magnitude de uma correlação podem, contudo, resultar em equívocos na estratégia de seleção, pois correlação alta entre dois caracteres pode ser resultado do efeito,

¹ Aceito para publicação em 2 de julho de 1998.

² Eng. Agr., M.Sc., Dep. de Biol. Geral, UFV, CEP 36571-000 Viçosa, MG.

³ Eng. Agr., D.Sc., Dep. de Fitotecnia, UFV.

⁴ Eng. Agr., D.Sc., Dep. de Biol. Geral, UFV.

⁵ Eng. Agr., Ph.D., Dep. de Fitotecnia, UFV.

sobre estes, de um terceiro ou de um grupo de caracteres (Cruz & Regazzi, 1994). Quando se deseja, por exemplo, incremento na produção de frutos por meio da seleção de seus componentes primários e/ou secundários, o estudo de correlação não indica a importância relativa dos efeitos diretos e indiretos desses componentes na produção e, assim, não é possível constatar se a correlação foi estabelecida por verdadeiras relações de causa e efeito.

Com o intuito de entender melhor as causas envolvidas nas associações entre caracteres, Wright (1921) propôs um método denominado de análise de trilha ("path analysis") que desdobra as correlações estimadas em efeitos diretos e indiretos de caracteres sobre uma variável básica. Esse método foi inicialmente utilizado em plantas por Dewey & Lu (1959), sendo posteriormente aplicado em diversas culturas. Em *Capsicum annuum*, Gill et al. (1977), Gupta & Yadav (1984), Cruz et al. (1988), Khurana et al. (1993) e Nandi et al. (1994) relatam sua utilização.

O sucesso da análise de trilha reside basicamente na formulação do relacionamento causa-efeito entre as variáveis (Schuster, 1996). Além disso, o desdobramento de correlações é dependente do conjunto de caracteres estudados, que normalmente é estabelecido pelo conhecimento prévio do pesquisador de sua importância e de possíveis inter-relações expressas em "diagramas de trilha" (Cruz & Regazzi, 1994).

Na mensuração dos efeitos diretos e indiretos de um conjunto de caracteres sobre uma variável básica, faz-se necessário estimar coeficientes de trilha, obtidos por meio de equações de regressão em que as variáveis são previamente padronizadas. A estimação destes coeficientes, contudo, pode ser adversamente afetada pelos efeitos de multicolinearidade entre os caracteres envolvidos. A multicolinearidade ocorre quando as observações amostrais das variáveis explicativas, ou suas combinações lineares, são correlacionadas (Matsuo, 1986; Ferrari, 1989). Segundo Carvalho (1995), em presença de multicolinearidade, as variâncias associadas aos estimadores dos coeficientes de trilha podem atingir valores demasiadamente elevados, tornando-os pouco confiáveis. Além disso, as estimativas dos parâmetros podem assumir valores absurdos ou sem nenhuma coerência com o fenômeno biológico estudado. Shrivastava & Sharma (1976), por exemplo,

encontraram efeito direto negativo na produtividade de arroz com relação aos caracteres número de perfilhos, número de grãos, peso de grãos e comprimento da panícula.

Para contornar os efeitos adversos da multicolinearidade, pode-se realizar a eliminação de variáveis do modelo de regressão. Shrivastava & Sharma (1976), em seu estudo, verificaram que eliminando-se o comprimento da panícula da análise, todas as outras contribuíram positivamente. Os autores propõem, portanto, que se faça uma seleção cuidadosa das variáveis para o estudo e que se tenha cautela no uso dessa técnica.

Além da eliminação de variáveis, encontra-se também na literatura a possibilidade de empregar uma metodologia alternativa ao dos quadrados mínimos, proposta por Carvalho (1995) e denominada análise de trilha em crista, para estimação dos parâmetros.

Este trabalho teve como objetivo avaliar as correlações genotípicas e seus desdobramentos em efeitos diretos e indiretos, pela análise de trilha, de componentes primários e secundários sobre a produção de frutos em genótipos de pimentão (*C. annuum* L.) e comparar métodos alternativos de contornar os efeitos adversos da multicolinearidade na estimação dos coeficientes de trilha.

MATERIAL E MÉTODOS

Foram analisados os dados obtidos de um ensaio com 128 linhagens, um acesso e quatro cultivares de pimentão da coleção de *Capsicum* do Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH-UFV), realizado em condições de campo, no município de Viçosa, MG.

O experimento constituiu-se na produção de mudas em sementeira, sendo transplantadas quando apresentavam cinco folhas definitivas, em média, para um solo adubado com 270 kg de N (60 kg no transplantio + 210 kg em cobertura), 30 kg de P₂O₅ (30+0), 170 kg de K₂O (120+50), 72 kg de Ca (72+0), 30 kg de Mg (30+0), 10 kg de Zn (10+0) e 3,5 kg de B (3,5+0), por hectare, nas formas de nitrocálcio, superfosfato simples, cloreto de potássio, nitrocálcio, sulfato de magnésio, sulfato de zinco e bórax, respectivamente. Irrigações suplementares e demais tratamentos culturais, como capinas, tutoramento das plantas e controle fitossanitário foram efetuados quando se fizeram necessários, de modo a manter as plantas em condições ótimas de crescimento e desenvolvimento.

O delineamento experimental utilizado foi o de blocos completos casualizados, com três repetições. Cada parcela foi composta por uma fileira com seis plantas, a espaços de 0,40 m, mantendo-se um espaçamento entre linhas de 0,90 m.

Os caracteres avaliados foram: a) produção total de frutos em g/planta, considerando-se cinco colheitas realizadas a intervalos médios de dez dias (PTF); b) número total de frutos/planta, considerando-se cinco colheitas realizadas a intervalos médios de dez dias (NTF); c) peso médio de fruto em grama, considerando-se cinco colheitas realizadas a intervalos médios de dez dias (PMF); d) diâmetro do pedúnculo, em mm (DP); e) altura da planta, em cm (AP); f) altura da primeira bifurcação, em cm (AB); g) diâmetro do caule, em cm (DC); h) número de ramos na primeira bifurcação (NR); i) comprimento do fruto, em cm (CF); j) largura do fruto, em cm (LF); k) razão comprimento/largura do fruto (RCL); l) espessura da polpa, em mm (EP); m) produção precoce em g/planta, considerando-se as três primeiras colheitas (PPP); n) produção precoce em número/planta, considerando-se as três primeiras colheitas (PPN); o) peso médio dos frutos precoces em g/fruto, considerando-se as três primeiras colheitas (PMP); p) índice de precocidade com base no peso (IPP); q) índice de precocidade com base no número (IPN); r) índice da concentração da produção com base no peso (ICPP); s) índice da concentração da produção com base no número (ICPN).

Estimativas de correlação genotípica e de coeficiente de determinação genotípico foram obtidas conforme sugerem Mode & Robinson (1959) e Vencovsky & Barriga (1992), respectivamente. Desdobramentos destas correlações em efeitos diretos e indiretos foram realizados por meio da análise de trilha desenvolvida por Wright (1921).

Diagramas causais mostrando o inter-relacionamento das variáveis analisadas são apresentados na Fig. 1. Nesses diagramas, a seta unidirecional indica efeito direto (coeficiente de trilha) de cada variável explicativa, enquanto a seta bidirecional representa a interdependência de duas variáveis explicativas, cuja magnitude é quantificada pela correlação genotípica. A Fig. 1a mostra o diagrama correspondente às relações causa-efeito entre as variáveis primárias, NTF e PMF, e a variável básica PTF. Como estas variáveis são inter-relacionadas por um efeito multiplicativo, a transformação logarítmica das observações correspondentes foi usada, resultando em LN_{NTF}, L_{PMF} e L_{PTF}. As Figs. 1b, 1c, 1d e 1e apresentam os diagramas em cadeia envolvendo variáveis secundárias, primárias e básica. As variáveis primárias e básica são as mesmas apresentadas na Fig. 1a. Além disso, foram considerados quatro grupos de variáveis secundárias: a) DP, AP, AB, DC e NR (Fig. 1b); b) CF, LF, RCL e EP (Fig. 1c); c) PPP, PPN e PMP (Fig. 1d); d) IPP, IPN, ICPP e ICPN (Fig. 1e).

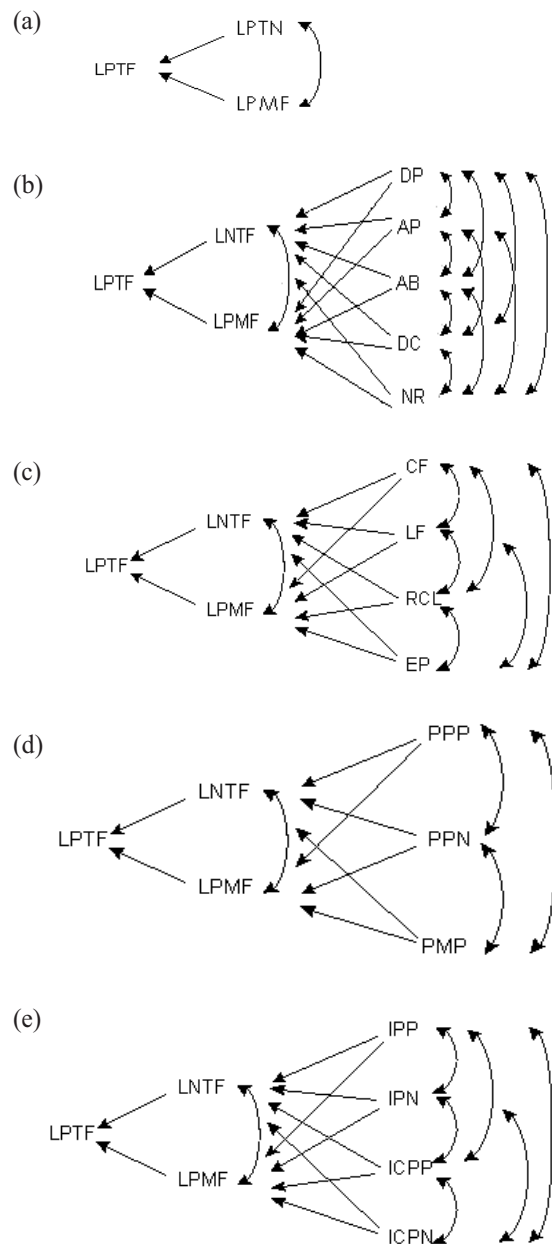


FIG. 1. Diagramas causais indicando o inter-relacionamento das variáveis analisadas: a) diagrama mostrando as relações causa-efeito entre os componentes primários e a produção total do fruto de pimentão; b-e) diagramas em cadeia apresentando as relações causa-efeito entre os componentes primários e secundários e a produção total de pimentões.

Na ocorrência de multicolinearidade moderada a severa entre as variáveis primárias ou secundárias de um diagrama causal específico, adotou-se o procedimento da eliminação de variáveis que contribuíram para o aparecimento dessa multicolinearidade, e empregou-se a metodologia alternativa aos dos quadrados mínimos, idealizada por Carvalho (1995) para estimação dos coeficientes de trilha. Carvalho (1995) propôs a modificação do sistema de equações normais, pela introdução de uma constante K à diagonal da matriz $X'X$, à semelhança do método de regressão em crista ("ridge regression") proposto por Hoerl & Kennard (1970a). Assim, utilizando-se esta metodologia, os coeficientes de trilha foram obtidos pela solução da equação

$$(X'X + KI_p) \theta^* = X'Y$$

em que $X'X$ é a matriz de correlações entre as variáveis independentes do modelo de regressão; K é uma pequena quantidade adicionada aos elementos da diagonal da matriz $X'X$; I_p é a matriz identidade; θ^* 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. O valor adequado referente à constante K foi determinado, neste ensaio, pelo exame do traço da crista (Hoerl & Kennard, 1970b). O traço da crista foi obtido plotando os parâmetros estimados (coeficientes de trilha) em função dos valores de K no intervalo de $0 < K < 1$. O menor valor de K capaz de estabilizar a maioria dos estimadores dos coeficientes de trilha foi empregado.

O grau de multicolinearidade da matriz $X'X$ foi estabelecido de acordo com os critérios indicados por Montgomery & Peck (1981), que se baseiam nos valores do determinante e do número de condição (NC = razão entre o maior e o menor autovalor) dessas matrizes. Para detectar os caracteres que contribuíram para o aparecimento da multicolinearidade, foi efetuada a análise dos elementos dos autovetores associados aos autovalores, descrita por Belsley et al. (1980). A diagnose da multicolinearidade, e todas as outras análises desse ensaio foram efetuadas pelo programa computacional GENES (Cruz, 1997).

RESULTADOS E DISCUSSÃO

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, 1994). Neste estudo, os caracteres

avaliados em genótipos de pimentão, que apresentaram maior correlação genotípica (acima de 0,4), em magnitude, com a produção total de frutos (LPTF) foram: AP (0,543), AB (0,485), LNTF (0,478), PMP (0,468), ICPP (-0,451), ICPN (-0,413), LPMF (0,406) e LF (0,401). Contudo, para melhor entendimento dos inter-relacionamentos entre os caracteres, realizaram-se análises de trilha, cujos diagramas causais são indicados na Fig. 1.

As magnitudes e os sinais dos efeitos diretos e indiretos dos componentes primários LNTF e LPMF sobre LPTF, de acordo com o diagrama causal da Fig. 1a, encontram-se na Tabela 1. Os caracteres LNTF e LPMF apresentaram magnitude de correlações acima de 0,4 com LPTF, e os de efeitos diretos e indiretos, acima de 0,6 sobre essa variável básica. Assim, no processo de seleção, devem ser considerados, simultaneamente, NTF e PMF.

O estudo das influências dos componentes primários sobre a produção pode, contudo, não ter grandes implicações práticas, pois esses caracteres geralmente são igualmente complexos, de baixa herdabilidade e, portanto, difíceis de serem selecionados (Cruz & Regazzi, 1994). Neste ensaio, os coeficientes de determinação genotípicos para LPMF, LNTF e LPTF foram 84,70%, 66,05% e 55,75%, respectivamente. Como os dois caracteres primários

TABELA 1. Estimativas dos efeitos diretos e indiretos de variáveis primárias sobre a variável básica (produção total de frutos transformada para a escala logarítmica, LPTF), na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96.

| Caráter ¹ | Efeito | Estimativa |
|-----------------------------|-------------------|------------|
| LNTF | Direto sobre LPTF | 1,151 |
| | Indireto via LPMF | -0,673 |
| | Total | 0,478 |
| LPMF | Direto sobre LPTF | 1,107 |
| | Indireto via LNTF | -0,701 |
| | Total | 0,406 |
| Coeficiente de determinação | | 1,000 |

¹ LNTF: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto.

devem ser considerados simultaneamente e LNTF mostrou coeficiente de determinação genotípico similar a LPTF, a resposta correlacionada, utilizando-se os componentes primários da produção, poderá não ser uma estratégia de seleção eficiente.

O coeficiente de determinação do modelo da análise de trilha (R^2), em questão, foi igual à unidade, pois as variações da variável básica são totalmente explicadas por esse esquema causal. Os caracteres primários mostraram multicolinearidade fraca (determinante da matriz $X'X = 0,6298$ e $NC = 4,107$), já que, segundo Montgomery & Peck (1981), à medida que o determinante de matriz de correlação entre os caracteres se aproxima de zero, a multicolinearidade torna-se mais intensa. Além disso, se $NC < 100$, a multicolinearidade não constitui problema sério (multicolinearidade fraca). Se $100 < NC < 1000$, a multicolinearidade é de moderada a forte, e se $NC > 1000$ há indício de multicolinearidade severa.

A mensuração dos efeitos diretos e indiretos dos caracteres botânicos sobre LPTF, de acordo com a Fig. 1b, encontra-se na Tabela 2. Altura da planta que apresentou a mais alta correlação com LPTF (0,543) teve o maior efeito direto (0,571) e efeitos indiretos baixos (próximos de zero) sobre esta variável e coeficiente de determinação genotípico de 88,8%. Por outro lado, apesar do caráter AB ter mostrado a segunda maior correlação com LPTF (0,485), seu efeito direto sobre este caráter foi praticamente nulo (0,073). Os coeficientes de determinação genotípicos para DP, AB, DC e NR foram 72,51%, 93,69%, 71,33% e 82,27%, respectivamente.

O modelo de regressão incluindo somente os caracteres botânicos como variáveis independentes explicou 27,89% da variação de LNTF, e 53,44% de LPMF. Assim como no esquema causal anterior, esses caracteres apresentaram multicolinearidade fraca (determinante da matriz $X'X = 0,2492$ e $NC = 13,146$). Além disso, nenhum desses, quando selecionados isoladamente ou em conjunto com os outros caracteres botânicos, poderá ser importante na seleção dos genótipos mais produtivos deste ensaio, uma vez que individualmente apresentam baixa correlação com os componentes primários da produção, e em conjunto pouco explicam a variação destes componentes.

O caráter AP, entre os caracteres botânicos analisados, foi o que mais explicou LPTF e, por conseguinte, em populações segregantes oriundas do cruzamento dos genótipos avaliados, esse caráter poderá ser útil em resposta correlacionada, por meio da seleção indireta.

Os caracteres CF, LF, RCL e EP mostraram multicolinearidade de moderada a forte (determinante da matriz $X'X = 0,0068$ e $NC = 481,093$). O desdobramento das correlações estimadas em efeitos diretos e indiretos dos caracteres do fruto sobre a variável básica, conforme a Fig. 1c, encontra-se na Tabela 3. Apesar de a correlação entre CF e LPTF ter sido de 0,276, o efeito direto deste componente secundário sobre a variável básica foi de -1,219. Este resultado pode ser consequência dos efeitos adversos da multicolinearidade (altas variâncias apresentadas pelos estimadores dos coeficientes de trilha), pois não se espera que o caráter CF apresente efeito direto alto e de sinal negativo com LPTF. Além disso, por meio dessa análise, o caráter RCL, que mostrou correlação com LPTF de 0,033, também teve efeito direto alto sobre a variável básica (1,958).

O modelo de regressão incluindo somente os caracteres do fruto como variáveis independentes explicou 21,62% da variação de LNTF e 80,13% de LPMF. Os coeficientes de determinação genotípicos para CF, LF, RCL e EP foram 87,06%, 78,96%, 91,63% e 81,46%, respectivamente.

Para contornar os efeitos adversos da multicolinearidade, empregou-se a metodologia proposta por Carvalho (1995) e da exclusão de variáveis do modelo de regressão, para estimação dos parâmetros.

O resultado da análise de trilha, de acordo com o diagrama causal em cadeia, apresentado na Fig. 1c, e empregando-se a metodologia proposta por Carvalho (1995), encontra-se na Tabela 4. Nesta análise, o caráter CF teve efeito direto positivo (0,157) e RCL apresentou efeito direto com magnitude de apenas 0,202. Largura do fruto foi a variável, entre os caracteres do fruto analisados, que apresentou maior efeito direto (0,454) com LPTF. O modelo de regressão explicou 20,00% da variação de LNTF e 76,95% de LPMF. O valor $K = 0,05$ foi usado para a estimação dos coeficientes de trilha. Devido à presença de multicolinearidade de moderada a forte, estes mostraram-se instáveis à medida que K aumen-

tou e tenderam a se estabilizar para algum valor de K ($K = 0,05$), como pode ser visto no exame do traço da crista (Figs. 2 e 3). Os valores da constante K foram considerados dentro do intervalo $0 < K < 1$, uma vez que $X'X$ encontra-se no formato de correlações.

Segundo Kalil (1977), como no método baseado no traço da crista K é escolhido segundo critérios subjetivos, este pode, conseqüentemente, variar de acordo com o pesquisador. Valores maiores que $K = 0,05$ também estabilizaram os coeficientes de regressão e poderiam, assim, ser utilizados na análise.

Contudo, segundo Hoerl & Kennard (1970b), o estimador em crista (θ^*) é um estimador viesado. Apesar de a variância do estimador em crista ser uma função decrescente de K, o quadrado do seu viés é uma função crescente de K. Desta forma, Carvalho (1995) concluiu que o valor de K a ser escolhido deve ser o menor valor capaz de estabilizar a maioria dos estimadores dos coeficientes de trilha.

A exclusão de variáveis do modelo de regressão para reduzir a variância dos estimadores dos coeficientes de trilha mostrou resultados semelhantes à

TABELA 2. Estimativas dos efeitos diretos e indiretos de caracteres botânicos sobre o caráter básico (produção total de frutos transformada para a escala logarítmica, LPTF), na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96¹.

| Descrição dos efeitos | Componentes primários | | LPTF |
|--------------------------------|-----------------------|--------|--------|
| | LNTF | LPMF | |
| Efeito direto secundário de DP | -0,443 | 0,715 | 0,272 |
| Efeito indireto de DP via AP | -0,009 | -0,003 | -0,012 |
| Via AB | 0,010 | -0,015 | -0,005 |
| Via DC | -0,069 | 0,026 | -0,043 |
| Via NR | 0,004 | -0,004 | -0,000 |
| Total | -0,507 | 0,719 | 0,212 |
| Efeito direto secundário de AP | 0,462 | 0,109 | 0,571 |
| Efeito indireto de AP via DP | 0,009 | -0,015 | -0,006 |
| Via AB | -0,111 | 0,164 | 0,053 |
| Via DC | -0,120 | 0,045 | -0,075 |
| Via NR | -0,011 | 0,011 | 0,000 |
| Total | 0,229 | 0,314 | 0,543 |
| Efeito direto secundário de AB | -0,152 | 0,225 | 0,073 |
| Efeito indireto de AB via DP | 0,029 | -0,047 | -0,018 |
| Via AP | 0,337 | 0,079 | 0,416 |
| Via DC | 0,022 | -0,008 | 0,014 |
| Via NR | -0,001 | 0,001 | 0,000 |
| Total | 0,235 | 0,250 | 0,485 |
| Efeito direto secundário de DC | -0,311 | 0,117 | -0,194 |
| Efeito indireto de DC via DP | -0,098 | 0,158 | 0,060 |
| Via AP | 0,178 | 0,041 | 0,219 |
| Via AB | 0,011 | -0,016 | -0,005 |
| Via NR | 0,004 | -0,004 | 0,000 |
| Total | -0,216 | 0,296 | 0,080 |
| Efeito direto secundário de NR | 0,129 | -0,133 | -0,004 |
| Efeito indireto de NR via DP | -0,014 | 0,023 | 0,009 |
| Via AP | -0,039 | -0,009 | -0,048 |
| Via AB | 0,001 | -0,002 | -0,001 |
| Via DC | -0,009 | 0,003 | -0,006 |
| Total | 0,068 | -0,118 | -0,050 |

¹ LNTF: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto; DP: diâmetro do pedúnculo; AP: altura da planta; AB: altura da primeira bifurcação; DC: diâmetro do caule; NR: número de ramos na primeira bifurcação.

análise de trilha em crista. Neste caso, dois grupos de variáveis secundárias foram formados: (1) CF, LF e EP; (2) RCL e EP. Na análise de trilha, utilizando-se o primeiro grupo de componentes secundários, a variável LF teve efeito direto com magnitude de 0,378 e efeitos indiretos baixos (próximos de zero). Nenhum dos caracteres do fruto avaliados, quando selecionado isoladamente ou em conjunto com outros desta análise, poderá ser importante em resposta correlacionada, por meio de seleção indireta.

Comparando-se as duas metodologias empregadas para contornar os efeitos adversos de multicolinearidade, pode-se mencionar que no descarte de variáveis não necessariamente a que tem maior importância econômica é a que mais explica a variável básica (por exemplo, em pimentão, a variável RCL é mais importante que LF, mas, para os dados em questão, esta última explicou mais LPTF), o

que torna difícil o processo de exclusão, principalmente quando várias variáveis necessitam ser eliminadas. Para reduzir essa dificuldade, muitas análises devem ser realizadas. O método proposto por Carvalho (1995), por sua vez, além de incluir e analisar todas as variáveis, possibilita a realização de apenas uma única análise.

As magnitudes e os sinais dos efeitos diretos e indiretos dos caracteres relacionados com produção precoce, de acordo com a Fig. 1d, são indicados na Tabela 5. Peso médio dos frutos precoces, que apresentou correlação de 0,468 com LPTF, teve efeito direto com magnitude de 0,456 e efeitos indiretos baixos (próximos de zero) sobre esta variável e coeficiente de determinação genotípico de 83,43%. Para os caracteres PPN e PPP, este coeficiente foi 73,19% e 68,24%, respectivamente. O modelo de regressão, incluindo somente os caracteres relacionados com

TABELA 3. Estimativas dos efeitos diretos e indiretos de caracteres do fruto sobre o caráter básico (produção total de frutos transformada para a escala logarítmica, LPTF), na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96¹.

| Descrição dos efeitos | Componentes primários | | LPTF |
|---------------------------------|-----------------------|--------|--------|
| | LPTN | LPMF | |
| Efeito direto secundário de CF | -1,055 | -0,164 | -1,219 |
| Efeito indireto de CF via LF | -0,057 | -0,138 | -0,195 |
| Via RCL | 1,206 | 0,482 | 1,688 |
| Via EP | -0,004 | 0,006 | 0,002 |
| Total | 0,090 | 0,186 | 0,276 |
| Efeito direto secundário de LF | 0,395 | 0,966 | 1,361 |
| Efeito indireto de LF via CF | 0,151 | 0,024 | 0,175 |
| Via RCL | -0,862 | -0,345 | -1,207 |
| Via EP | -0,129 | 0,201 | 0,072 |
| Total | -0,445 | 0,846 | 0,401 |
| Efeito direto secundário de RCL | 1,398 | 0,560 | 1,958 |
| Efeito indireto de RCL via CF | -0,909 | -0,141 | -1,050 |
| Via LF | -0,244 | -0,596 | -0,840 |
| Via EP | 0,062 | -0,097 | -0,035 |
| Total | 0,307 | -0,274 | 0,033 |
| Efeito direto secundário de EP | -0,294 | 0,458 | 0,164 |
| Efeito indireto de EP via CF | -0,016 | -0,002 | -0,018 |
| Via LF | 0,174 | 0,425 | 0,599 |
| Via RCL | -0,297 | -0,118 | -0,415 |
| Total | -0,433 | 0,763 | 0,330 |

¹ LNTF: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto; CF: comprimento do fruto; LF: largura do fruto; RCL: razão comprimento/largura do fruto; EP: espessura da polpa.

TABELA 4. Estimativas dos efeitos diretos e indiretos de caracteres do fruto sobre o caráter básico (produção total de frutos transformada para a escala logarítmica, LPTF), obtidas por meio da análise de trilha em crista, na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96¹.

| Descrição dos efeitos | Componentes primários | | LPTF |
|---------------------------------|-----------------------|--------|--------|
| | LPTN | LPMF | |
| Efeito direto secundário de CF | -0,113 | 0,270 | 0,157 |
| Efeito indireto de CF via LF | 0,028 | -0,093 | -0,065 |
| Via RCL | 0,185 | -0,011 | 0,174 |
| Via EP | -0,004 | 0,006 | 0,002 |
| Total | 0,096 | 0,172 | 0,268 |
| Efeito direto secundário de LF | -0,194 | 0,648 | 0,454 |
| Efeito indireto de LF via CF | .0,016 | -0,039 | -0,023 |
| Via RCL | -0,132 | 0,008 | -0,124 |
| Via EP | -0,126 | 0,197 | 0,071 |
| Total | -0,436 | 0,814 | 0,378 |
| Efeito direto secundário de RCL | 0,215 | -0,013 | 0,202 |
| Efeito indireto de RCL via CF | -0,098 | 0,234 | 0,136 |
| Via LF | 0,119 | -0,399 | -0,280 |
| Via EP | 0,061 | -0,095 | -0,034 |
| Total | 0,297 | -0,273 | 0,024 |
| Efeito direto secundário de EP | -0,286 | 0,448 | 0,162 |
| Efeito indireto de EP via CF | -0,002 | 0,004 | 0,002 |
| Via LF | -0,085 | 0,285 | 0,200 |
| Via RCL | -0,046 | 0,003 | -0,043 |
| Total | -0,419 | 0,740 | 0,321 |

¹ LNTF: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto; CF: comprimento do fruto; LF: largura do fruto; RCL: razão comprimento/largura do fruto; EP: espessura da polpa.

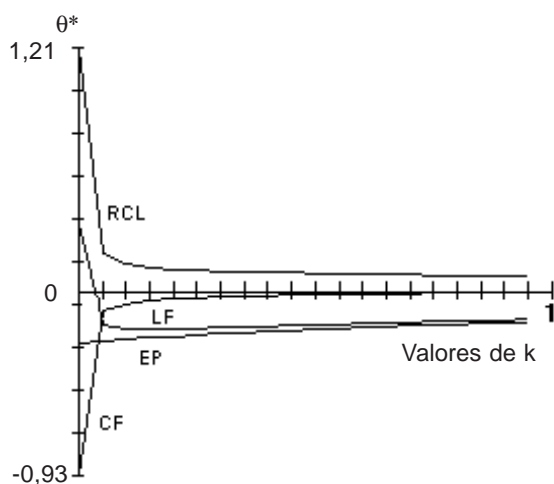


FIG. 2. Estimativas dos coeficientes de trilha (θ^*) em função dos valores de k e obtidas da análise utilizando-se como variável básica o logaritmo do número total de frutos.

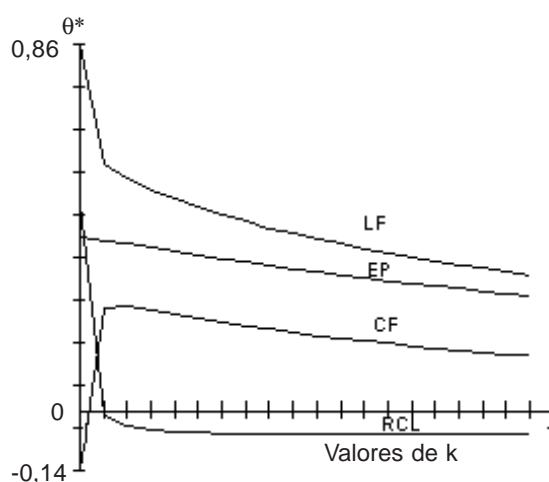


FIG. 3. Estimativas dos coeficientes de trilha (θ^*) em função dos valores de k e obtidas da análise utilizando-se como variável básica o logaritmo do peso médio de frutos.

produção precoce como variáveis independentes, explicou 40,00% de LNTF e 97,00% de LPMF. Essas variáveis secundárias apresentaram multicolinearidade fraca (determinante da matriz $X'X = 0,1284$ e $NC = 29,251$). Além disso, nenhuma dessas, quando selecionada isoladamente ou em conjunto com outras desta análise, poderá ser importante em resposta correlacionada, por meio de seleção indireta. O caráter PMP, entre os caracteres relacionados com produção precoce, foi o que mais explicou LPTF.

Os caracteres IPP, IPN, ICPP e ICPN mostraram multicolinearidade de moderada a forte (determinante da matriz $X'X = 0,0003$ e $NC = 214,458$). Com o desdobramento das correlações estimadas em efeitos diretos e indiretos dos índices mensurados com a variável básica, de acordo com a Fig. 1e, foram observadas altas variâncias dos estimadores dos coeficientes de trilha, à semelhança do ocorrido com os caracteres do fruto. O modelo de regressão incluindo esses caracteres como variáveis independentes explicou 90% de LNTF e 78,00% de LPMF. Os coeficientes de determinação genotípico para IPP, IPN, ICPP e ICPN foram 57,17%, 62,02%, 69,31% e 72,42%,

respectivamente. Para contornar os efeitos adversos da multicolinearidade, foi feita a exclusão de variáveis do modelo de regressão e uso da metodologia proposta por Carvalho (1995). A mensuração dos efeitos diretos e indiretos dos caracteres IPP, IPN e ICPN sobre LPTF encontra-se na Tabela 6. O modelo de regressão contendo esses caracteres como variáveis independentes explicou 41,00% de LNTF e 68,88% de LPMF. Os caracteres IPP, IPN e ICPN mostraram efeitos diretos de 0,738; -0,732 e -0,362, respectivamente. Com base nas magnitudes dos efeitos diretos e indiretos, poder-se-ia utilizar seleção simultânea desses caracteres visando ganhos na produção. Contudo, pelo fato de as magnitudes dos coeficientes de determinação genotípica para IPP e IPN terem sido similares à do LPTF, esta estratégia de seleção poderá não ser eficiente. Ao ser aplicada a análise de trilha em crista, não se conseguiu encontrar um valor de k que reduzisse a variância dos coeficientes de trilha sem que houvesse uma grande redução nos valores dos coeficientes de determinação dos modelos de regressão analisados.

TABELA 5. Estimativas dos efeitos diretos e indiretos de caracteres relacionados com a produção precoce sobre o caráter básico (produção total de frutos transformada para a escala logarítmica, LPTF), na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96¹.

| Descrição dos efeitos | Componentes primários | | LPTF |
|---------------------------------|-----------------------|--------|--------|
| | LPTN | LPMF | |
| Efeito direto secundário de PPP | 1,218 | -1,072 | 0,146 |
| Efeito indireto de PPP via PPN | -1,382 | 1,322 | -0,060 |
| Via PMP | 0,029 | -0,045 | -0,016 |
| Total | -0,135 | 0,205 | 0,070 |
| Efeito direto secundário de PPN | -1,522 | 1,456 | -0,066 |
| Efeito indireto de PPN via PPP | 1,106 | -0,973 | 0,133 |
| Via PMP | 0,201 | -0,315 | -0,114 |
| Total | -0,215 | 0,168 | -0,047 |
| Efeito direto secundário de PMP | -0,801 | 1,257 | 0,456 |
| Efeito indireto de PMP via PPP | -0,044 | 0,039 | -0,005 |
| Via PPN | 0,382 | -0,365 | 0,017 |
| Total | -0,463 | 0,931 | 0,468 |

¹ LPTN: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto; PPP: produção precoce em g/planta; PPN: produção precoce em número/planta; PMP: peso médio dos frutos precoces.

TABELA 6. Estimativas dos efeitos diretos e indiretos de caracteres com base em índices sobre o caráter básico (produção total de frutos transformada para a escala logarítmica, LPTF), na cultura do pimentão. Viçosa, Minas Gerais. Ano agrícola 1995/96¹.

| Descrição dos efeitos | Componentes primários | | LPTF |
|----------------------------------|-----------------------|--------|--------|
| | LPTN | LPMF | |
| Efeito direto secundário de IPP | -2,413 | 3,151 | 0,738 |
| Efeito indireto de IPP via IPN | 2,046 | -2,740 | -0,694 |
| Via ICPN | 0,618 | -0,285 | 0,333 |
| Total | 0,251 | 0,126 | 0,377 |
| Efeito direto secundário de IPN | 2,158 | -2,890 | -0,732 |
| Efeito indireto de IPN via IPP | -2,288 | 2,987 | 0,699 |
| Via ICPN | 0,576 | -0,266 | 0,310 |
| Total | 0,446 | -0,109 | 0,277 |
| Efeito direto secundário de ICPN | -0,671 | 0,309 | -0,362 |
| Efeito indireto de ICPN via IPP | 2,223 | -2,902 | -0,679 |
| Via IPN | -1,850 | 2,478 | 0,628 |
| Total | -0,298 | -0,115 | -0,413 |

¹ LNTF: logaritmo do número total de frutos; LPMF: logaritmo do peso médio de fruto; IPP: índice de precocidade com base no peso; IPN: índice de precocidade com base no número; ICPN: índice da concentração da produção com base no número.

CONCLUSÕES

1. Os caracteres que, isoladamente, têm maior importância em explicar as variações na produção de frutos são altura de planta (entre os caracteres botânicos analisados), largura do fruto (entre os caracteres do fruto), e peso médio dos frutos precoces (entre os caracteres relacionados com produção precoce).

2. Nenhum dos índices avaliados (índice de precocidade com base no peso; de precocidade com base no número; da concentração da produção com base no peso; e da concentração da produção com base no número) mostra ser superior para explicar a produção de frutos.

3. A exclusão de variáveis e o método proposto por Carvalho (1995) mostram ser eficientes para reduzir a variância dos coeficientes de trilha.

REFERÊNCIAS

- BELSLEY, D.A.; KUH, E.; WELCH, R.E. **Regression diagnostics: identifying data and sources of collinearity**. New York: John Wiley & Sons, 1980. 292p.
- CARVALHO, S.P. de. **Métodos alternativos de estimação de coeficientes de trilha e índices de seleção, sob multicolinearidade**. Viçosa: UFV, 1995. 163p.
- CRUZ, C.D. **Programa Genes: aplicativo computacional em genética e estatística**. Viçosa: UFV, 1997. 442p.
- CRUZ, C.D.; MIRANDA, J.E.C.; COSTA, C.P. Correlações, efeitos diretos e indiretos de caracteres agrônômicos sobre a produção de pimentão (*Capsicum annuum* L.). **Revista Brasileira de Genética**, v.11, n.4, p.921-928, 1988.
- CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 1994. 390p.
- DEWEY, D.R.; LU, K.H. A correlation and path coefficient analysis of components of crested wheatgrass seed production. **Agronomy Journal**, v.51, p.515-518, 1959.
- FALCONER, D.S. **Introdução à genética quantitativa**. Viçosa: UFV, 1981. 279p.
- FERRARI, F. **Estimadores viesados para modelos de regressão em presença de multicolinearidade**. Piracicaba: USP-ESALQ, 1989. 127p.
- GILL, H.S.; ASSAWA, B.M.; THAKUR, P.C.; THAKUR, T.C. Correlation, path coefficient and multiple

- regression analysis in sweet-pepper. **Indian Journal of Agricultural Science**, v.47, p.408-410, 1977.
- GUPTA, C.R.; YADAV, R.D.S. Genetic variability and path analysis in chili (*Capsicum annuum* Linn.). **Genetica Agraria**, v.38, p.425-432, 1984.
- HOERL, A.E.; KENNARD, R.W. Ridge regression: applications to nonorthogonal problems. **Technometrics**, v.12, n.1, p.69-82, 1970a.
- HOERL, A.E.; KENNARD, R.W. Ridge regression: biased estimation for nonorthogonal problems. **Technometrics**, v.12, n.1, p.55-68, 1970b.
- KALIL, M.N. **Aplicação do método de regressão de cumeeira ("ridge regression") na estimação de funções de demanda e de produção**. Piracicaba: USP-ESALQ, 1977. 135p.
- KHURANA, S.C.; PANDITA, M.L.; THAKRAL, K.K.; SINGH, C.B. Correlation and path analysis studies in chilli. **Maharashtra Journal of Horticulture**, v.7, p.76-80, 1993.
- MATSUO, T. **O uso da regressão de cumeeira em experimentos agrônomicos**. Piracicaba: USP-ESALQ, 1986. 89p.
- MODE, J.C.; ROBINSON, H.F. Pleiotropism and the genetic variance and covariance. **Biometrics**, v.15, p.518-537, 1959.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York: John Wiley & Sons, 1981. 504p.
- NANDI, A.; SENAPATI, N.; TRIPATHY, P. Path coefficient analysis in chilli *Capsicum annuum* L. **Environment and Ecology**, v.12, p.943-944, 1994.
- SCHUSTER, I. **Correlações, coeficientes de trilha, composição de gluteninas e qualidade do trigo para panificação**. Viçosa: UFV, 1996. 98p.
- SHRIVASTAVA, M.N.; SHARMA, K.K. Analysis of path coefficients in rice. **Zeitschrift fuer Pflanzenzuechtung**, v.77, p.174-177, 1976.
- VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496p.
- WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**, v.20, p.557-585, 1921.