

## Consequências da multicolinearidade sobre a análise de trilha em canola

### Multicollinearity consequence on path analysis in canola

Jefferson Luís Meirelles Coimbra<sup>1</sup> Giovani Benin<sup>1</sup> Eduardo Alano Vieira<sup>1</sup>  
Antônio Costa de Oliveira<sup>2</sup> Fernando Iraja Félix Carvalho<sup>2</sup>  
Altamir Frederico Guidolin<sup>3</sup> Adriana Pires Soares<sup>4</sup>

#### RESUMO

A análise estatística do tipo multivariada vem crescendo consideravelmente, motivando a sua ampla utilização por parte dos pesquisadores criando, assim, grande demanda por conhecimentos específicos tanto a respeito da sua aplicação quanto das suas pressuposições ou limitações. Para que a avaliação do grau de associação entre diferentes caracteres de importância agrônoma tenha uma estimativa confiável em termos biológico, é de fundamental importância identificar e quantificar o grau de multicolinearidade entre as variáveis estudadas. Além disso, os tipos de modelos estatísticos e matemáticos utilizados na determinação desta dependência linear entre as variáveis classificatórias ou independentes podem ou não ser adequados a estimativas dos parâmetros biológicos avaliados. O presente trabalho tem como objetivo apresentar uma avaliação crítica sobre o grau de multicolinearidade identificado e avaliado sobre a análise de trilha analisada sobre partes de um experimento de canola. Os resultados permitem inferir que a aplicação da análise de trilha sobre o grau de multicolinearidade severa produz resultados sem nenhuma importância biológica para o melhorista de plantas. No entanto, esta limitação pode ser facilmente identificada e corrigida através da análise de trilha com colinearidade empregando uma constante ( $k$ ) na diagonal da matriz  $X'X$ . O modelo de análise com multicolinearidade severa, entretanto, superestimou, valores de coeficientes de correlação simples, comparativamente com a multicolinearidade fraca. Mesmo assim, pode não ser necessariamente mais precisa, principalmente em virtude da avaliação de um número restrito de variáveis incluídas na análise ou de uma sobreposição destas variáveis explicativas.

**Palavras-chave:** *Brassica spp.*, correlação fenotípica, dependência linear.

#### ABSTRACT

The statistical multivariate analysis has a widespread use by researchers, creating a large demand for specific knowledge regarding its application concerning its assumptions and or limitations. In order to evaluate the degree of association among different characters of agronomic importance with an estimative reliable in biological terms, it is striking to quantify the multicollinearity among the studied variables. In addition, the types of statistical and mathematical models used in determining this linear dependence between classifying or independent variables may or may not be adequate for estimatives of biological parameters evaluated. The present work has as objective to present a critical evaluation on the degree of multicollinearity identified and evaluated on the path analysis performed on parts of a canola experiment. The results allow to postulate that path analysis application on the degree of severe multicollinearity produces results with no biological importance for the plant breeder. However, this limitation can be easily identified and corrected through path analysis with colinearity employing a constant ( $k$ ) on diagonal axis of  $X'X$  matrix. The model of analysis with severe multicollinearity, however overestimated the single correlation coefficient values comparatively with the weak multicollinearity. Even so, it may not be necessarily more precise, mainly regarding the evaluation of a restricted number of variables included in the analysis or an overlapping of the explainable variables.

**Key words:** *Brassica spp.*, phenotypic correlation, linear dependency.

<sup>1</sup>Engenheiro Agrônomo, MSc, Doutorando em Melhoramento, Universidade Federal de Pelotas (UFPEL), Rua Marechal Deodoro, 713/305, Centro, 6020-220, Pelotas, RS. E-mail: coimbrajefferson@pop.com.br. Autor para correspondência.

<sup>2</sup>Engenheiro Agrônomo, PhD, Professor da UFPEL.

<sup>3</sup>Engenheiro Agrônomo, Doutor, Professor, Universidade do Estado de Santa Catarina (UDESC).

<sup>4</sup>Engenheiro Agrônomo, estagiário do Laboratório de Genômica e Fitomelhoramento, UFPEL.

## INTRODUÇÃO

A correlação entre caracteres, basicamente tem duas origens: genética e de ambiente; segundo FALCONER (1987), a correlação genética é ocasionada principalmente pelo pleiotropismo (é a propriedade pela qual um gene condiciona mais de um caráter simultaneamente) e pelo desequilíbrio de ligação (associação não aleatória entre alelos de diferentes locos). Ainda de acordo com este mesmo autor, a correlação fenotípica é definida como a associação entre duas variáveis que pode ser observada diretamente.

Os valores dos coeficientes de correlação permitem ao pesquisador prever, de modo seguro, as alterações em um determinado caráter provocadas pela pressão de seleção exercida sobre outro caráter (COIMBRA et al., 1999). A correlação é apenas uma medida de associação; portanto não permite conclusões sobre causa e o efeito, não possibilitando inferências sobre o tipo de associação que governa o par de caracteres Y/X. Para superar esta limitação, WRIGHT (1921) desenvolveu um método que permite uma partição dos coeficientes de correlações tanto em efeitos diretos quanto indiretos (coeficiente de trilha). De modo contextualizado, um coeficiente de trilha ou análise de causa e efeito pode ser definido como um coeficiente de regressão estandardizado ou padronizado; sendo que a análise de trilha é composta por uma expansão da regressão múltipla quando estão envolvidas inter-relações complexas (CRUZ & CARNEIRO, 2003).

Para que a avaliação da associação entre caracteres tenha uma estimativa e gere uma interpretação biologicamente apropriada e segura, é de fundamental importância que se teste o grau de colinearidade entre as variáveis independentes. Neste caso, a análise dos autovalores da matriz ( $X'X$ ) é utilizada para identificar a natureza aproximada da dependência linear existente entre os caracteres, identificando aquele(s) que contribuem para o aparecimento da multicolinearidade (CARVALHO et al. 2002). É bastante prudente, para obtenção dos efeitos diretos, que a matriz  $X'X$  (não singular das correlações entre as variáveis primárias) esteja bem condicionada, problemas de multicolinearidade podem torná-la não singular e, fazendo, conseqüentemente, com que as estimativas de quadrados mínimos não sejam confiáveis (CRUZ, 1997).

A multicolinearidade se faz presente quando existe algum nível de inter-relação entre as variáveis estudadas (variáveis independentes).

Entretanto, muitas vezes o termo é utilizado erroneamente, como sinônimo de uma correlação muito alta ou perfeito (próxima de +1 ou -1) entre as variáveis independentes (alto grau de multicolinearidade), especialmente quando ocorre uma sobreposição entre as variáveis no modelo de regressão. Seus efeitos danosos não são ocasionados simplesmente pela sua presença, mais sim pelo grau com que se manifesta. Entre os efeitos peculiares de uma elevada multicolinearidade, podem ser citados, estimativa inconsistente do coeficiente de regressão e também uma superestimativa dos efeitos diretos das variáveis explicativas sobre a variável resposta, o que pode levar à interpretação equivocada (CRUZ & CARNEIRO, 2003). O grau de multicolinearidade da matriz pode ser estabelecido com base no número de condições, que é a razão entre o maior e o menor autovalor da matriz. De modo prático, quando o número de condições for menor do que 100, a multicolinearidade é fraca; entre 100 e 1000, a multicolinearidade é moderada a forte e, finalmente, quando maior do que 1000, a multicolinearidade é severa (MONTGOMERY & PECK, 1981). Somente quando o grau de multicolinearidade é considerado fraco, não constitui problema sério para a análise (CARVALHO et al., 1999). Nos demais casos, quando identificado um grau de multicolinearidade tanto moderada quanto forte, deve ser atenuado o problema. Uma maneira prática de contornar este problema é identificar qual (s) variável está inflacionando o grau de multicolinearidade. Esta identificação e seleção das variáveis que maximizam a estimativa do coeficiente de correlação, com interferência de multicolinearidade o menor possível, podem ser realizadas por meio do procedimento *Stepwise* do programa Computacional SAS, por exemplo.

O objetivo deste trabalho foi analisar o efeito da multicolinearidade sobre os coeficientes de correlação simples e análise de trilha avaliada sobre a influência de quatro caracteres primários do rendimento de grãos sobre a produtividade de genótipos fixos de canola sob duas condições de colinearidade (fraca e severa) e inferir a respeito da influência do ambiente na manifestação da multicolinearidade.

## MATERIAL E MÉTODOS

O experimento foi conduzido no município de Lages, localizado no Planalto Sul de Santa Catarina no ano agrícola de 1995/96. As coordenadas geográficas deste município são 27° 52' 30" de latitude sul e 50° 18' 20" de longitude oeste, com altitude média

de 930m e a presença de verões brandos com chuvas bem distribuídas (EMPASC, 1978). Os fatores climáticos ocorridos no período de condução do ensaio como distribuição de chuvas e temperaturas amenas, podem ser considerados normais para a região do Planalto Catarinense.

Foram utilizados doze genótipos de canola (BR-1, LEGENDA, IRIS, BINGO, ICIOLA 42, PFB 2, D1084, B2270, WW-352890, ICIOLA 41, HIOLA 401 (F<sub>n</sub>) e HIOLA 401). Para determinação da correlação fenotípica e posterior análise dos coeficientes de trilha foram avaliados cinco caracteres sendo quatro componentes primários do rendimento de grãos: i) número de síliqua por planta (NSP); ii) número de grãos por planta (NGP); iii) número de grãos por síliqua (NGS); iv) massa de mil grãos em g (MMG) e também o caráter rendimento de grãos em kg ha<sup>-1</sup> (RG).

O delineamento experimental foi o de blocos completamente casualizados, com quatro repetições em duas épocas de semeadura (15/06 e 26/07), as quais constituíram os diferentes ambientes avaliados, respectivamente chamados de ambiente um e dois. O solo da área experimental foi um cambissolo húmico álico, de textura argilosa, pertencente à unidade de mapeamento Lages. O preparo do solo foi convencional, com uma aração e duas gradagens. Cada unidade experimental colhida foi constituída por seis fileiras de cinco metros de comprimento espaçadas de 0,3 metros entre si. A área útil foi composta por quatro fileiras centrais, descartando 0,5m das extremidades, totalizando 4,8m<sup>2</sup>. As duas épocas de semeadura (ambientes) receberam manejo similar. Foi realizado desbaste no estádio de três folhas verdadeiras, para obter uma população de 180 plantas por metro quadrado. Na adubação de base, realizada junto às linhas de semeadura, foram aplicados 20, 80, 60kg ha<sup>-1</sup> de N, P<sub>2</sub>O<sub>5</sub> e K<sub>2</sub>O, respectivamente, segundo as recomendações descritas na Sociedade Brasileira de Ciência do Solo –

Região Sul (1989). A adubação de cobertura foi realizada no estádio de quatro folhas verdadeiras, com uma aplicação de 40kg de N por hectare. O controle de plantas invasoras foi realizado com capinas manuais e o controle de insetos com a aplicação de 500g ha<sup>-1</sup> de Methamidophos (O, S - dimetil-fosforoamidotoado).

Estimativas dos coeficientes de correlação simples ou fenotípicas, obtidas por meio dos dados originais com repetições, entre os caracteres como descrito por STEEL & TORRIE (1980). As correlações foram desdobradas em efeito diretos e indiretos dos caracteres de importância agrônômica para a canola (variáveis independentes do modelo de regressão) sobre o rendimento de grãos (variável básica) por meio da análise de trilha (WRIGHT, 1921).

O grau de multicolinearidade da matriz singular X'X foi estabelecido com base no seu número de condições (MONTGOMERY & PECK, 1981). A análise dos autovalores da matriz de correlação fenotípica foi efetuada para identificar a natureza da dependência linear existente entre os caracteres, detectando aqueles que contribuíram para o aparecimento da multicolinearidade (BELSLEY et al. 1980). Os caracteres que apresentaram os maiores elementos nos autovetores associados aos menores autovalores foram os que mais contribuíram para este aparecimento.

As análises foram realizadas com auxílio do programa computacional GENES (CRUZ, 2001). As correlações fenotípicas foram estimadas pelo método proposto por STEEL & TORRIE (1980) e posteriormente particionadas em efeitos diretos e indiretos pela análise de trilha (CRUZ & CARNEIRO, 2003).

## RESULTADOS E DISCUSSÃO

Na tabela 1, são apresentadas às estimativas de correlação fenotípica (r<sub>p</sub>) entre os caracteres de importância agrônômica avaliados em condição de colinearidade fraca (acima da diagonal) e

Tabela 1 - Coeficientes de correlação fenotípica de Pearson entre cinco caracteres agrônômicos avaliados na cultura de canola, entre o rendimento de grãos (RG) em genótipos de canola, e seus componentes primários (NSP, NGP, NGS e MMG). Acima e abaixo da diagonal os coeficientes de correlações possuem uma colinearidade fraca (ambiente um) e severa (ambiente dois), respectivamente. Pelotas, RS, 2004.

Caracteres	NSP <sup>1</sup>	NGP	NGS	MMG	RG
NSP	1	0,874 <sup>*</sup>	0,737 <sup>*</sup>	0,057	-0,299
NGP	0,982 <sup>*</sup>	1	0,829 <sup>*</sup>	0,177	-0,296
NGS	0,927 <sup>*</sup>	0,978 <sup>*</sup>	1	0,219	0,017
MMG	0,357 <sup>*</sup>	0,438 <sup>*</sup>	0,554 <sup>*</sup>	1	-0,178
RG	0,082	0,029	0,075	0,461 <sup>*</sup>	1

<sup>1</sup>NSP = número de síliqua por planta; NGP = número de grãos por planta; NGS = número de grãos por síliqua; MMG = massa de mil grãos; RG = rendimento de grãos por unidade de área.

\*significativo a 0,01 de probabilidade de erro.

severa (abaixo da diagonal). Os resultados apontaram que um atraso na época de sementeira afetou a manifestação da multicolinearidade, ocasionando um aumento no grau da mesma. Este fato pode ser explicado pelo desempenho contrastante dos cultivares frente às oscilações de ambiente, o que era esperado, uma vez, que a correlação estimada no estudo foi a fenotípica que sofre influência do genótipo, ambiente e da interação dos mesmos. Por exemplo, no ambiente dois (multicolinearidade severa) o caráter MMG apresentou um coeficiente de correlação significativo com todos os caracteres avaliados, comparativamente, no ambiente um (multicolinearidade fraca) o mesmo não revelou correlação significativa com nenhum dos caracteres aferidos. Neste momento, vale a pena ressaltar que este grau de multicolinearidade foi obtido por meio dos dados coletados, sem nenhuma transformação ou indução.

Pode ser observada uma boa concordância na direção das correlações nas condições de multicolinearidade severa e fraca. Entretanto, na condição de multicolinearidade severa, as estimativas de correlações apresentaram magnitude superior em relação à condição de magnitude fraca. Esta maior magnitude pode ser atribuída a efeitos modificadores do ambiente, competição por fatores nutricionais e a diferentes mecanismos fisiológicos controlando a expressão dos caracteres (FALCONER & MACKAY, 1996, CARVALHO et al., 2002), nas duas condições de ambiente avaliadas. Foi observada elevada associação entre os caracteres NSP x NGP (0,874 e 0,982), NSP x NGS (0,737 e 0,927), NGP x NGS (0,829 e 0,978), nas condições de multicolinearidade fraca e severa, respectivamente (Tabela 1).

A ausência de correlação significativa entre os caracteres componentes do rendimento (NSP, NGP, NGS e MMG) com o RG pode ser explicada pela compensação que a canola frequentemente promove, aumentando ou diminuindo o peso dos grãos em função do número de grãos por síliqua e do número de síliqua por planta, em desenvolvimento. A única exceção foi observada para a correlação RG x MMG (0,461) na condição de multicolinearidade severa, possivelmente devido à elevada associação entre as variáveis independentes NGP x NSP (0,982) e NGS x NGP (0,978), o mesmo não foi observado na condição de multicolinearidade fraca, sendo, neste caso, a correlação entre RG x MMG (-0,178) negativa, apesar de não significativa (Tabela 1).

Quando um caráter correlacionar-se positivamente com alguns e negativamente com outros, há a sugestão de ter-se um cuidado adicional

para esse fato, pois ao selecionar um determinado caráter, podem-se provocar mudanças indesejáveis em outros (CRUZ & CARNEIRO, 2003). As correlações fenotípicas obtidas, no presente estudo, indicam que poderá ser difícil reunir em um só genótipo caracteres com maior NSP, NGP, NGS e MMG (Tabela 1). Isso torna a análise de trilha uma estratégia importante para a superação das dificuldades encontradas pelos pesquisadores; permitindo que se possam priorizar alguns caracteres de maior relevância (maior grau de associação) para atingir um maior rendimento de grãos num menor espaço de tempo.

Observadas as tabelas 1 e 2, nas quais os coeficientes de correlações simples entre o par de caracteres MMG x RG sobre o efeito da multicolinearidade fraca (-0,178) e severa (0,461) apontou estimativas bastante contrastantes tanto para magnitude (2,5 vezes maior o coeficiente de correlação estimado sob multicolinearidade severa) dos valores quanto para a direção (sinal oposto).

Para obtenção dos efeitos diretos e indiretos da análise de trilha, é necessário que a matriz  $X'X$  esteja bem condicionada. Problemas da multicolinearidade podem torná-la não-singular, fazendo conseqüentemente, com que as estimativas de mínimos quadrados não sejam confiáveis (CRUZ, 2001); na presença da multicolinearidade, o estimador de mínimos quadrados, obtido  $(X'X)\beta = X'Y$ , pode estar associado a uma variância muito alta e para atenuar esse efeito adverso pode ser utilizada uma constante  $k$  na diagonal da matriz  $X'X$ . De modo geral, procura-se atenuar esse efeito adverso, modificando ligeiramente o sistema de equações normais, pela adição de uma constante  $k$  na diagonal da matriz. CARVALHO (1995) ressalta que deve ser escolhido o menor valor de  $k$ , para o qual a maioria dos coeficientes de trilha, associados aos vários caracteres, estejam estabilizados. Com base em gráficos, pode ser estabelecido o valor de  $k$  facilmente.

Na tabela 2, consta a análise de trilha na qual foram analisados dados de correlação com multicolinearidade "fraca", multicolinearidade "severa", sem a devida correção da multicolinearidade e ainda, com os dados que apresentaram multicolinearidade severa, procedeu-se uma análise de trilha com colinearidade corrigindo este problema (severa com correção), utilizando um valor de  $k$  de 0,278 (determinado de modo idêntico como descrito acima, por meio do Programa Computacional Genes). Pode ser observado que, nas condições de multicolinearidade fraca e de correção da multicolinearidade, os valores diretos e

Tabela 2 - Estimativas dos efeitos diretos ( $P_i$ ) e indiretos ( $P_i \cdot r_{ij}$ ) dos componentes primários do rendimento de grãos (número de síliqua por planta, NSP, número de grãos por planta, NGP, número de grãos por síliqua, NGS e massa de mil grãos em g, MMG) sobre a produtividade de grãos em genótipos de canola obtidas a partir dos dados com diferentes tipos de colinearidade (fraca, severa e severa corrigida). Pelotas, RS, 2004.

Componente secundário	Estimativas sob diferente colinearidade		
	Fraca	Severa	Severa corrigida
Variável ----- NSP			
Efeito direto sobre RG	-0,313	6,237	-0,149
via NGP	-0,644	-10,435	-0,139
via NGS	0,670	4,128	0,165
via MMG	-0,013	0,151	0,119
Correlação fenotípica (Total)	-0,299	0,082	-0,045
Variável ----- NGP			
Efeito direto sobre RG	-0,736	-10,631	-0,143
via NSP	-0,274	6,122	-0,145
via NGS	0,754	4,354	0,246
via MMG	-0,040	0,185	0,177
Correlação fenotípica (Total)	-0,296	0,029	0,095
Variável ----- NGS			
Efeito direto sobre RG	0,909	4,453	0,425
via NSP	-0,231	5,784	-0,058
via NGP	-0,610	-10,395	-0,083
via MMG	-0,050	0,234	0,311
Correlação fenotípica (Total)	0,017	0,075	0,714
Variável ----- MMG			
Efeito direto sobre RG	-0,229	0,423	0,349
via NSP	-0,018	2,225	-0,051
via NGP	-0,130	-4,657	-0,072
via NGS	0,199	2,468	0,379
Correlação fenotípica (Total)	-0,178	0,461	0,702
Coefficiente de determinação	0,369	0,728	0,542
Efeito da variável residual	0,795	0,521	0,676
Número de condições	29,998	3.525,726	11,089
Determinante	0,0419	0,000064	0,448
Tipo de colinearidade	Fraca	Severa	Severa corrigida
Valor de k	-	-	0,278

indiretos dos coeficientes de trilha não foram superiores à unidade, o que não foi observado para a condição de multicolinearidade severa, onde os valores apresentaram magnitude superior a unidade, tanto positiva quanto negativa, evidenciando novamente que uma multicolinearidade severa leva a uma superestimativa dos coeficientes de correlação e de trilha (efeitos diretos e indiretos). Quanto maior for o grau de multicolinearidade, maiores serão os valores das estimativas dos coeficientes de correlações e, conseqüentemente, menor será a precisão em sua estimativa (FREUND & LITTELL, 1981).

O grau de multicolinearidade ocorre quando as observações amostrais das variáveis explicativas, ou suas combinações lineares, são correlacionadas (FERRARI, 1989). Confirmando os resultados obtidos

neste estudo, CARVALHO (1995) e CARVALHO et al. (1999) observaram que, 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 coeficientes de trilha podem assumir valores absurdos ou sem nenhuma coerência com o fenômeno biológico estudado. Neste caso, pode ser observado que o caráter NSP evidenciou efeito direto negativo sobre a variável rendimento de grãos, com magnitude de -0,313 e -0,149 na condição de multicolinearidade fraca e de correção dos efeitos da multicolinearidade, respectivamente. Isto se justifica pelo fato de o número de grãos por planta ter exercido um efeito indireto negativo sobre o número de síliqua por planta.

Entretanto, este caráter deve ser considerado de fundamental importância na seleção de plantas, sendo que o mesmo não foi observado para a condição de multicolinearidade severa sem correção. Neste caso, o efeito direto do NSP sobre o RG foi de 6,237, com elevado efeito direto negativo (-10,345) via NGP e positivo via NGS (4,128). Este resultado é consequência dos efeitos adversos da multicolinearidade (altas variâncias apresentadas pelos estimadores dos coeficientes de trilha), pois não se espera que os caracteres NSP e NGP apresentem efeito direto alto e de sinal positivo sobre o RG simultaneamente.

Ainda na tabela 2, pode ser observado que, na condição de multicolinearidade fraca, o caráter NGS apresentou efeito direto de 0,909 sobre o caráter RG, sendo este valor superior à magnitude da correlação entre os caracteres NGS x RG (0,017). Neste caso, a presença de efeitos indiretos negativos via NSP, NGP e MMG são um indicativo de possíveis dificuldades que os melhoristas de canola se defrontam para selecionar com base nos componentes primários do rendimento de grãos. Dificuldades estas que podem ser mais acentuadas se os dados não forem analisados de forma correta, tal como na condição de multicolinearidade severa. Neste caso, o efeito direto do caráter NGS sobre o RG foi de 4,453, com efeitos indiretos positivo via NSP (5,784) e negativo via NGP (-10,395). Estes valores discrepantes não foram observados na condição onde os dados com multicolinearidade severa foram analisados buscando superar esta dificuldade. Por meio da análise dos dados, inseridos na tabela 2, pode ser observado que os efeitos diretos tiveram o mesmo sinal dos coeficientes de correlações fenotípica e suas magnitudes foram baixas principalmente para os coeficientes de correlações, pois não superam os efeitos residuais. Tal fato evidencia que a variável auxiliar é a principal determinante das variações na variável resposta e, conseqüentemente, prediz que a seleção indireta será eficaz. Por exemplo, os caracteres número de grãos por planta (NGP) e número de grãos por síliqua (NGS) é o par de caracteres mais associados, no entanto evidenciou um efeito direto em sentido contrário; este caráter (NGS) pode ser de grande importância na seleção indireta com RG.

## CONCLUSÕES

Os componentes primários do rendimento de grãos (número de síliqua por planta, número de grãos por planta, número de grãos por síliqua e massa

de mil grãos) têm influência variável no rendimento de grãos. A multicolinearidade severa superestima tanto os valores dos coeficientes de correlações simples quanto os efeitos diretos sobre o caráter rendimento de grãos estimados através da análise de trilha.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BELSLEY, D.A. et al. **Regression diagnostics: identifying data and sources of colinearity.** New York : J. Wiley, 1980. 292p.
- CARVALHO, C.G.P. et al. Análise de trilha sob multicolinearidade em pimentão. **Pesquisa Agropecuária Brasileira**, Brasília, v.34, n.4, p.603-613, 1999.
- CARVALHO, C.G.P. et al. Correlações e análise de trilha em linhagens e soja semeadas em diferentes épocas. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.3, p.311-320, 2002.
- CARVALHO, S.P. de. **Métodos alternativos de estimação de coeficientes de trilha e índices de seleção sob multicolinearidade.** Viçosa : UFV, 1995. 163p.
- COIMBRA, J.L.M. et al. Divergência genética em feijão preto. **Ciência Rural**, Santa Maria, v.29, n.3, p.427-431, 1999.
- CRUZ, C.D. **Programa GENES – Versão Windows**, aplicativo computacional em genética e estatística. Viçosa : UFV, 2001. 648p.
- CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético.** Viçosa : UFV, 2003. 579p.
- CRUZ, C.D. **Aplicativo computacional em genética e estatística.** Viçosa : UFV, 1997. 442p.
- EMPASC. **Zoneamento agroclimático do estado de Santa Catarina.** Florianópolis, 1978. 70p.
- FALCONER, D.S. **Introdução à genética quantitativa.** Viçosa : UFV, 1987. 279p.
- FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics.** 4.ed. England : Longman, 1996. 463p.
- FERRARI, F. **Estimadores viesados para modelos de regressão em presença de multicolinearidade.** Piracicaba: USP-ESALQ, 1989. 127p.
- FREUND, R.J.; LITTELL, R.C. **Sas for linear models.** Cary: SAS Institute, 1981. 229p.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis.** New York : J. Wiley, 1981. 504p.
- STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics: a biometrical approach.** 2.ed. New York : McGraw-Hill, 1980. 631p.
- WRIGHT, S. Correlation and causation. **Journal Agricultural Research**, Washington, v.20, p.557-585, 1921.