

SAMUEL PEREIRA DE CARVALHO

Métodos Alternativos de Estimação de
Coeficientes de Trilha e Índices de Seleção,
sob Multicolinearidade


VIÇOSA

MINAS GERAIS - BRASIL

FEVEREIRO - 1995

15195

SAMUEL PEREIRA DE CARVALHO

UFV	BIBLIOTECA BBT	OBRA RG000510850
	CLASSIFICACAO T 575.1015195 / C331m	
TITULO: Metodos alternativos de estimacao de coefi		
		
102770		BBT

MÉTODOS ALTERNATIVOS DE ESTIMAÇÃO DE COEFICIENTES DE TRILHA E ÍNDICES DE SELEÇÃO, SOB MULTICOLINEARIDADE

Tese Apresentada à Universidade Federal de Viçosa, como Parte das Exigências do Curso de Genética e Melhoramento, para Obtenção do Título de "Doctor Scientiae".

BIBLIOTECA CENTRAL - U F V - 102-770 03-01-95
--

T
575.1015195
C331m
1995
ex-2

DOAÇÃO

VIÇOSA
MINAS GERAIS - BRASIL
FEVEREIRO - 1995

Ficha catalográfica preparada pela Seção de Catalogação
e Classificação da Biblioteca Central da UFV

T

C33lm
1995
Carvalho, Samuel Pereira de, 1947 -
Métodos alternativos de estimação de coeficien-
tes de trilha e índices de seleção, sob multicoli-
nearidade/ Samuel Pereira de Carvalho. - Viçosa:
UFV, 1995.
163p. : il.

Orientador: Cosme Damião Cruz.
Tese (doutorado) - Universidade Federal de Viço-
sa.

1. Melhoramento genético - Métodos estatísticos.
2. Biometria. 3. Genética quantitativa. I. Univer-
sidade Federal de Viçosa. II. Título.

CDD. 18 ed. 575.1015195

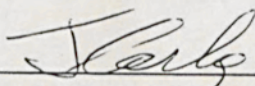
CDD. 19 ed. 575.1015195

SAMUEL PEREIRA DE CARVALHO

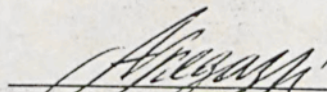
MÉTODOS ALTERNATIVOS DE ESTIMAÇÃO DE COEFICIENTES DE
TRILHA E ÍNDICES DE SELEÇÃO, SOB MULTICOLINEARIDADE

Tese Apresentada à Universidade
Federal de Viçosa, como Parte das
Exigências do Curso de Genética e
Melhoramento, para Obtenção do
Título de "Doctor Scientiae".

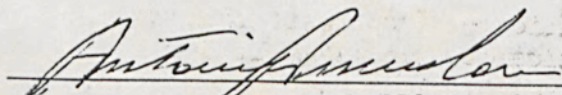
APROVADA: 07 de julho de 1994.



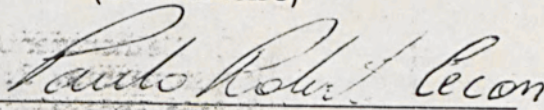
Prof. José Carlos Silva
(Conselheiro)



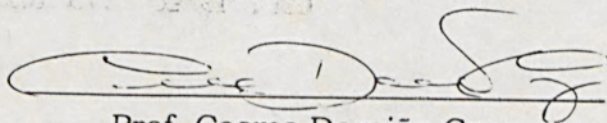
Prof. Adair José Regazzi
(Conselheiro)



Prof. Antônio Américo Cardoso



Prof. Paulo Roberto Cecon



Prof. Cosme Damião Cruz
(Orientador)

AGRADECIMENTOS

A Deus, Autor da vida.

Aos meus familiares.

A quem este trabalho possa ser útil.

Ao Departamento de Ciências Agrárias da Universidade Federal de Lavras - ESAL, especialmente aos professores Dr. A. Patto Ramalho, João Bosco dos Santos e César Brasil P. Pinto, pelo estímulo.

A Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior - CAPES e ao Programa Institucional de Capacitação de Docentes - PICD, pela bolsa de estudos concedida.

A Universidade Federal de Viçosa - UFV, pela oportunidade concedida.

Ao professor Cosme Damião Cruz, pela orientação segura e constante, pelo estímulo à busca do aperfeiçoamento, pelos esforços no sentido do aprimoramento do trabalho e pela amizade.

Ao professor Cibas Vieira, pelo dedicado apoio e pela orientação na fase inicial do curso.

Aos professores Adair José Regazzi e José Carlos Silva, pelo aconselhamento decisivo ao êxito do trabalho de tese.

Aos professores Antonio Americo Cardoso e Paulo Roberto Ceccon, pelas sugestões apresentadas.

Aos professores do curso, pelos ensinamentos transmitidos.

A todos os funcionários da UFV que de algum modo contribuíram para a realização desse trabalho.

Aos colegas de curso, pela amizade e pelo companheirismo, especialmente aos colegas Antonio Alves Pereira, Antonio Oliveira Júnior, Lailton Soares, Eduardo Shiguera, José Marcelo, José Tarciso e José Reinaldo, com quem trabalhei e estudei.

AGRADECIMENTOS

A Igreja Presbiteriana de Viçosa, onde eu e meus familiares encontramos um segundo lar.

Ao Departamento de Biologia da Escola Superior de Agricultura de Lavras - ESAL, especialmente aos professores Magno A. Patto Ramalho, João Bosco dos Santos e César Brasil P. Pinto, pelo estímulo.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior - CAPES e ao Programa Institucional de Capacitação de Docentes - PICD, pela bolsa de estudos concedida.

À Universidade Federal de Viçosa - UFV, pela oportunidade concedida.

Ao professor Cosme Damião Cruz, pela orientação segura e constante, pelo estímulo à busca do aperfeiçoamento, pelos esforços no sentido do aprimoramento do trabalho e pela amizade.

Ao professor Clibas Vieira, pelo dedicado apoio e pela orientação na fase inicial do curso.

Aos professores Adair José Regazzi e José Carlos Silva, pelo aconselhamento decisivo ao êxito do trabalho de tese.

Aos professores Antonio Américo Cardoso e Paulo Roberto Cecon, pelas sugestões apresentadas.

Aos professores do curso, pelos ensinamentos transmitidos.

A todos os funcionários da UFV que de algum modo contribuíram para a realização desse trabalho.

Aos colegas de curso, pela amizade e pelo companheirismo, especialmente aos colegas Antonio Alves Pereira, Antonio Oliveira Júnior, Lailton Soares, Eduardo Shiguera, José Marcelo, José Tarciso e José Reinaldo, com quem compartilhamos trabalhos e estudos.

À Igreja Presbiteriana de Viçosa, onde eu e meus familiares encontramos um segundo lar.

Samuel Pereira de Carvalho, filho de José Teixeira de Carvalho e Lúcia Pereira de Carvalho, nasceu a 10 de agosto de 1947 em Macaia, município de Bom Sucesso, Minas Gerais. cursou o primário e o ginásio em Ribeirão Vermelho, MG, realizando em seguida o curso científico no Colégio Estadual Dr. João Batista Hermeto, em Lavras, MG. Em 1969 ingressou na Escola Superior de Agricultura de Lavras, obtendo o diploma de engenheiro-agrônomo em 1972. Admitido em 1973 no extinto Instituto Brasileiro do Café - IBC, por meio de concurso público, atuou no serviço de assistência técnica daquele órgão no estado do Paraná, ocupando o cargo de chefe do Serviço Local de Assistência à Caficultura em Londrina, PR. Em 1986 passou ao setor de pesquisas da Divisão de Assistência à Caficultura do IBC em Varginha, MG. Em 1987, no cargo de engenheiro-agrônomo, foi transferido por redistribuição para a Escola Superior de Agricultura de Lavras, onde realizou o curso de Mestrado em Genética e Melhoramento de Plantas. No segundo semestre de 1990 iniciou o curso de Doutorado em Genética e Melhoramento, na Universidade Federal de Viçosa - UFV.

BIOGRAFIA

Samuel Pereira de Carvalho, filho de José Teixeira de Carvalho e Luzia Pereira de Carvalho, nasceu a 10 de agosto de 1947 em Macaia, município de Bom Sucesso, Minas Gerais. Coursou o primário e o ginásial em Ribeirão Vermelho, MG, realizando em seguida o curso científico no Colégio Estadual Dr. João Batista Hermeto, em Lavras, MG. Em 1969 ingressou na Escola Superior de Agricultura de Lavras, obtendo o diploma de engenheiro-agrônomo em 1972. Admitido em 1973 no extinto Instituto Brasileiro do Café - IBC, por meio de concurso público, atuou no serviço de assistência técnica daquele órgão no estado do Paraná, ocupando o cargo de chefe do Serviço Local de Assistência à Cafeicultura em Londrina, PR. Em 1986 passou ao setor de pesquisas da Divisão de Assistência à Cafeicultura do IBC em Varginha, MG. Em 1987, no cargo de engenheiro-agrônomo, foi transferido por redistribuição para a Escola Superior de Agricultura de Lavras, onde realizou o curso de Mestrado em Genética e Melhoramento de Plantas. No segundo semestre de 1990 iniciou o curso de Doutorado em Genética e Melhoramento, na Universidade Federal de Viçosa - UFV.

2.3.5.1. Regressão em Componentes Principais	28
2.3.5.2. Regressão em Cristas ou em Cuneiras	29
2.3.5.2.1. Propriedades dos Estimadores sobre Cristas ou em Cuneiras.....	32
2.3.5.2.2. Métodos para a Determinação de k	33
3. MATERIAL E MÉTODOS	37
3.1. Caracterização dos Experimentos.....	37
3.1.1. Materiais.....	37
3.1.2. Local e Épocas de Experimentos.....	38
3.1.3. Delineamento Experimental	38
3.1.4. Caracteres Avaliados.....	38
3.2. Metodologias.....	41
3.2.1. Análise de Variância.....	41
3.2.2. Estimação de Componentes de Covariâncias e Análise de Fenótipos, Genótipos e de	41
EXTRATO	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	5
2.1. Uso da Análise de Trilha no Melhoramento de Plantas...	5
2.2. Uso do Índice de Seleção no Melhoramento de Plantas..	10
2.3 Efeitos da Multicolinearidade sobre os Parâmetros Genéticos.....	18
2.3.1. Conceito e Causas da Multicolinearidade.....	18
2.3.2. Efeitos da Multicolinearidade sobre a Análise de Trilha.....	20
2.3.3. Efeitos da Multicolinearidade sobre os Índices de Seleção	21
2.3.4. Diagnóstico da Multicolinearidade.....	22
2.3.5. Métodos Alternativos de Análise de Regressão em Presença de Multicolinearidade.....	28
Utilizando o Método de Cristas	29

Página

2.3.5.1. Regressão em Componentes Principais	28
2.3.5.2. Regressão em Cristas ou em Cumeeira	29
2.3.5.2.1. Propriedades dos Estimadores sobre Cristas ou em Cumeeira.....	32
2.3.5.2.2. Métodos para a Determinação de k	33
3. MATERIAL E MÉTODOS	37
3.1. Caracterização dos Experimentos.....	37
3.1.1. Materiais.....	37
3.1.2. Local e Épocas de Instalação dos Experimentos.....	38
3.1.3. Delineamento Experimental	38
3.1.4. Caracteres Avaliados.....	38
3.2. Metodologias.....	41
3.2.1. Análise de Variância.....	41
3.2.2. Estimação de Componentes de Covariâncias e Correlações Fenotípicas, Genotípicas e de Ambientes	43
3.2.2.1. Estimação das Covariâncias	43
3.2.2.2. Estimação das Correlações	46
3.2.3. Diagnóstico de Multicolinearidade.....	47
3.2.3.1. Análise Descritiva das Correlações.....	47
3.2.3.2. Fatores de Inflação da Variância	47
3.2.3.3. Análise dos Autovalores e Autovetores de $X'X$	48
3.2.3.4. Decomposição em Valores Singulares e Decomposição da Variância.....	49
3.2.4. Análise de Trilha	51
3.2.4.1. Estimação dos Coeficientes de Trilha pelo Método dos Mínimos Quadrados.....	52
3.2.4.2. Estimação dos Coeficientes de Trilha Utilizando o Método de Cristas	59

3.3. Estimação de Ganhos por Seleção.....	63
3.3.1. Resposta Direta à Seleção	63
3.3.2. Resposta Correlacionada.....	66
3.3.3. Seleções Simultâneas com Base em Vários Caracteres.....	69
3.3.3.1. Estimação do Índice Clássico sem Considerar a Ocorrência de Multicolinearidade.....	69
3.3.3.2. Estimação do Índice Clássico Considerando a Ocorrência de Multicolinearidade	72
4. RESULTADOS E DISCUSSÃO	77
4.1. Análise da Variância	77
4.2. Estimativas de Parâmetros Genéticos.....	78
4.3. Correlações Genotípicas, Fenotípicas e de Ambiente.....	81
4.4. Análise de Trilha.....	87
4.4.1. Análise de Trilha sem Considerar a Possibilidade de Ocorrência de Multicolinearidade entre as Observações	88
4.4.1.1. Efeitos das Variáveis Primárias sobre a Variável Básica.....	88
4.4.1.2. Efeitos das Variáveis Secundárias sobre as Variáveis Primárias	90
4.4.1.3. Efeitos Indiretos das Variáveis Secundárias sobre a Variável Básica.....	97
4.4.2. Análise de Trilha com Correção para os Efeitos da Multicolinearidade entre as Observações.....	98
4.4.2.1. Efeitos das Variáveis Primárias sobre a Variável Básica, Estimados com Correção para os Efeitos da Multicolinearidade	98

4.4.2.2. Efeitos das Variáveis Secundárias sobre as Variáveis Primárias, Estimados com Correção para os Efeitos da Multicolinearidade	108
4.4.2.3. Efeitos das Variáveis Secundárias sobre a Variável Básica, Estimados com Correção para os Efeitos da Multicolinearidade	117
4.4.2.4. Aplicação do Método de Estimação em Cristas sob Multicolinearidade Intensa	117
4.5. Ganho de Seleção	133
4.5.1. Seleção Direta e Seleção Indireta	136
4.5.2. Seleção com Base no Índice Clássico, de SMITH (1936) e HAZEL (1943)	140
5. RESUMO E CONCLUSÕES	148
BIBLIOGRAFIA	152
APÊNDICE	161

Informações obtidas de amostras de genótipos de feijão, avaliadas em dois ensaios conduzidos em Viçosa-MG, durante o ano agrícola 1991/92, foram utilizadas para estudo de correlações e ganhos de seleção. Para isso, procedeu-se à estimação de coeficientes de tripla e índices de seleção. Considerando os efeitos adversos da multicolinearidade, quanto à eficiência e à exequibilidade dessas metodologias, foram propostos métodos alternativos de estimação, com base na solução em cristas ou em cuneca. Verificou-se que, sob boa precisão experimental, a amostra em estudo foi eficiente para discriminar os genótipos favoráveis. As correlações entre os caracteres foram determinadas, preponderantemente, pelos componentes genotípicos. Componentes primários da produção de grãos no feijoeiro;

como número médio de vagens por planta e número médio de sementes por vagem, apresentaram-se fortemente correlacionados com esse caráter. O índice de colheita, que é um componente secundário, também mostrou-se altamente correlacionado com a produção de grãos. Pela análise de trilha, observou-se que a seleção indireta para aumento da produtividade de grãos é mais viável por meio do número médio de vagens por planta. Observa-se, ainda, que a seleção com esse objetivo é também viável por meio de caracteres secundários de menor complexidade genética. Assim, para aumentar o número médio de vagens por planta, o qual apresenta alta correlação positiva com a produtividade de grãos pela seleção com base em índice de colheita e

EXTRATO

CARVALHO, Samuel Pereira de, D. S., Universidade Federal de Viçosa, fevereiro de 1995. **Métodos Alternativos de Estimação de Coeficientes de Trilha e Índices de Seleção, sob Multicolinearidade.** Professor Orientador: Cosme Damião Cruz. Professores Conselheiros: Adair José Regazzi e José Carlos Silva.

Informações obtidas de amostras de genótipos de feijão, avaliadas em dois ensaios conduzidos em Viçosa-MG, durante o ano agrícola 1991/92, foram utilizadas para estudo de correlações e ganhos de seleção. Para isso, procedeu-se à estimação de coeficientes de trilha e índices de seleção. Considerando os efeitos adversos da multicolinearidade, quanto à eficiência e à exequibilidade dessas metodologias, foram propostos métodos alternativos de estimação, com base na solução em cristas ou em cumeeira. Verificou-se que, sob boa precisão experimental, a amostra em estudo foi eficiente para discriminar os genótipos favoráveis. As correlações entre os caracteres foram determinadas, preponderantemente, pelos componentes genotípicos. Componentes primários da produção de grãos no feijoeiro,

como número médio de vagens por planta e número médio de sementes por vagem, apresentaram-se fortemente correlacionados com esse caráter. O índice de colheita, que é um componente secundário, também mostrou-se altamente correlacionado com a produção de grãos. Pela análise de trilha, observou-se que a seleção indireta para aumento da produtividade de grãos é mais viável por meio do número médio de vagens por planta. Observa-se, ainda, que a seleção com esse objetivo é também viável por meio de caracteres secundários de menor complexidade gênica. Assim, pode-se aumentar o número médio de vagens por planta, o qual apresenta alta correlação positiva com a produtividade de grãos pela seleção com base em índice de colheita e número de dias até a maturação. A estimação dos coeficientes de trilha, pelo método em cristas, proporcionou a redução da magnitude dos fatores de inflação da variância associados aos efeitos diretos e indiretos, especialmente quando sob multicolinearidade intensa. Não foi possível obter ganhos para aumento do número de vagens e sementes por vagem sem reduzir o tamanho das sementes, em razão de forte associação negativa entre esses caracteres. Contudo, pela otimização dos fatores de ponderação desses caracteres para a seleção, o índice de seleção clássico possibilita a obtenção desses ganhos, minimizando as perdas no tamanho das sementes. O método proposto para estimação de índices de seleção em cristas mostrou-se eficiente. Os resultados evidenciam a relevância dos efeitos da multicolinearidade.

1. INTRODUÇÃO

A análise de trilha e o índice de seleção são técnicas que ocupam posição de destaque no melhoramento de plantas.

A análise de trilha tem sido utilizada no estudo dos efeitos diretos e indiretos de um conjunto de variáveis sobre um determinado caráter, considerado como principal, caracterizado por grande complexidade, porém, de considerável importância econômica. Consiste no desdobramento dos coeficientes de correlação entre duas variáveis em seus efeitos diretos e indiretos, uma vez que esta correlação é constituída pela soma dos valores de todos os efeitos associados às duas variáveis (LI, 1975). Assim é obtido o coeficiente de trilha, definido como o produto do coeficiente de regressão parcial, pela relação entre o desvio-padrão de um efeito determinado por uma dada causa e o desvio-padrão total do efeito. A causa em questão é constituída pela variável explicativa, enquanto a variável principal ou básica constitui o efeito. Sendo Y o efeito e X a causa, o coeficiente de trilha da causa X para o efeito Y é $\beta_{yx} \sigma_x / \sigma_y$, conforme SINGH e CHAUDHARY (1979).

Esse método, é considerado como sendo uma generalização da análise de regressão parcial padronizada, e consiste no estudo de um sistema de variáveis múltiplas, linearmente relacionadas, que envolve os fatores explicativos ou causas, e suas variáveis resultantes, ou efeitos. Na aplicação do método, é indispensável a construção de um diagrama de causa e efeito, indicando as inter-relações entre as variáveis, em conformidade com a hipótese que se deseja estudar.

O método da análise de trilha tem sido eficiente para revelar a verdadeira natureza das inter-relações de causa e efeito entre um caráter de importância econômica e seus componentes primários e secundários, em várias espécies vegetais.

O índice de seleção, por outro lado, tem sido utilizado na identificação de genótipos superiores, com base em um conjunto de caracteres e relaciona-se, também, com a correlação entre estes. O índice de seleção caracteriza-se por maximizar a probabilidade de se obter a ordenação correta dos valores genéticos dos indivíduos submetidos ao processo seletivo. É potencialmente útil, quando os melhoristas têm necessidade de avaliar vários caracteres de naturezas diversas simultaneamente, para melhor inferir sobre a superioridade relativa dos mesmos, conforme CRUZ (1990). A seleção, com base em vários caracteres, tem-se mostrado mais adequada por possibilitar a obtenção de um produto final superior em todos os aspectos de interesse, qualquer que seja a natureza dos mesmos. Isso evita que o material melhorado seja rejeitado por agricultores, consumidores, industriais ou outros segmentos aos quais o produto é destinado.

Sendo uma modalidade de análise multivariada, o índice de seleção possibilita a combinação das informações múltiplas contidas na unidade experimental, tornando possível a seleção com base num complexo de variáveis. Assim, torna-se possível a identificação dos

genótipos mais promissores, principalmente nos contextos genético e evolutivo.

Em relação a outros métodos alternativos, utilizados para seleção com base em caracteres múltiplos, a superioridade do índice de seleção torna-se maior à medida que cresce o número de caracteres envolvidos na seleção. Essa superioridade evidencia-se também quando as importâncias relativas dos caracteres tornam-se mais semelhantes entre si, conforme HAZEL e LUSH (1942).

Apesar de serem procedimentos distintos e com finalidades também distintas no melhoramento, as técnicas de análise de trilha e cálculo de índice de seleção apresentam dificuldades comuns, pois seus parâmetros são estimados a partir de matrizes de correlações ou covariâncias fenotípicas ou genotípicas, as quais podem ser mal condicionadas por efeito de multicolinearidade entre as variáveis envolvidas. Conseqüentemente, não poderiam proporcionar estimativas precisas dos parâmetros genéticos. Porém, a presença de multicolinearidade nas matrizes de trabalho pode ser diagnosticada mediante técnicas capazes de prover informações sobre sua intensidade e ainda identificar as variáveis envolvidas, possibilitando a adoção de medidas capazes de atenuar possíveis efeitos adversos.

O presente trabalho tem por objetivos:

- Avaliar a adequação de matrizes de correlações para uso em análise de trilha e na estimação de índices de seleção;
- apresentar procedimentos estatísticos para obtenção de parâmetros genéticos em análise de trilha e estimação de índices de seleção, em presença de multicolinearidade; e

- aplicar os procedimentos tradicionais, apresentados na literatura, e também os procedimentos propostos no presente trabalho, para análise de trilha e estimação de índices de seleção em populações de feijão, na geração F_3 .

2. REVISÃO DE LITERATURA

2.1. *Uso da Análise de Trilha no Melhoramento de Plantas*

O método da análise de trilha ou coeficiente de trilha foi desenvolvido por WRIGHT (1921a). Conforme o próprio autor, destina-se a fornecer meios de se calcular o grau com que um dado efeito é determinado por uma das causas em um sistema complexo de variáveis correlacionadas. Assumindo que estas variáveis estejam relacionadas de forma aproximadamente linear e que as influências das várias causas combinam-se, também aproximadamente, por aditividade, o coeficiente de trilha estará medindo a importância de um dado relacionamento, ou o grau de influência, a partir de uma causa até seu efeito. Desse modo, o coeficiente de trilha difere do coeficiente de correlação por apresentar uma direção, e é ainda definido como sendo a proporção do desvio padrão do efeito em relação ao desvio padrão total, quando todas as causas são constantes, à exceção da única em questão, conforme WRIGHT (1921b). Nesse mesmo trabalho, o autor empregou o método

2. REVISÃO DE LITERATURA

2.1. *Uso da Análise de Trilha no Melhoramento de Plantas*

O método da análise de trilha ou coeficiente de trilha foi desenvolvido por WRIGHT (1921a). Conforme o próprio autor, destina-se a fornecer meios de se calcular o grau com que um dado efeito é determinado por uma das causas em um sistema complexo de variáveis correlacionadas. Assumindo que estas variáveis estejam relacionadas de forma aproximadamente linear e que as influências das várias causas combinam-se, também aproximadamente, por aditividade, o coeficiente de trilha estará medindo a importância de um dado relacionamento, ou passo de influência, a partir de uma causa até seu efeito. Desse modo, o coeficiente de trilha difere do coeficiente de correlação por apresentar uma direção, e é ainda definido como sendo a proporção do desvio padrão do efeito em relação ao desvio padrão total, quando todas as causas são constantes, à exceção da única em questão, conforme WRIGHT (1921b). Nesse mesmo trabalho, o autor empregou o método

do coeficiente de trilha para estudo do relacionamento biométrico entre um pai e sua descendência, considerando acasalamentos de indivíduos endógamos e não-endógamos.

Como descrito por LI (1956), coeficiente de trilha é um método de análise multivariada apropriado para lidar com um sistema "fechado" de variáveis linearmente relacionadas. Caso estas não se relacionem, faz-se necessária uma transformação de escalas para que o método possa ser aplicado. Por sistema linear fechado, entende-se que cada variável pode ser tanto uma combinação linear de algumas outras variáveis, como pode ser um dos fatores básicos, e nesse caso pode ainda ser correlacionada ou independente de outros fatores do sistema. O autor ressalta que o princípio fundamental da teoria de análise de trilha é que a correlação entre duas variáveis é a soma de todos os passos que as une. Aplicando-se esse princípio à genética, obtém-se a correlação entre parentes. Indivíduos aparentados dividem um ou mais ancestrais comuns, os quais constituem as causas para a herança dos parentes. Consegue-se, assim, não só calcular a correlação genética, mas também dividi-la em seus componentes, de modo que cada componente represente a contribuição de um ancestral comum, através de uma linha particular de descendentes. LI (1956) mostra ainda que, dentre outras aplicações, o coeficiente de trilha pode ser usado para avaliar a não-aditividade de efeitos gênicos. Também o equilíbrio gênico pode ser avaliado por esse método naquelas situações em que ocorrem autofecundações sucessivas. Neste último caso as consequências da endogamia podem ser avaliadas a um custo inferior ao de qualquer outro método.

Conforme JOHNSON e WICHERN (1988), a análise de trilha consiste em duas partes:

a) estabelecimento de um modelo de relacionamento de causa e efeito, ou diagrama causal, entre as variáveis; e

b) decomposição das correlações observadas em um conjunto de termos denominado coeficientes de trilha, os quais representam os caminhamentos simples e complexos.

Assim é possível medir os efeitos diretos e indiretos de uma variável sobre a outra.

O coeficiente de trilha mede o efeito de um caráter sobre o outro. É um coeficiente de regressão padronizado, podendo ser negativo ou positivo e ainda maior ou menor que a unidade. O diagrama causal deve ser estabelecido dispondo-se as variáveis de modo a representar a natureza de suas interrelações. LI (1975) salienta que esse esquema causal, ou diagrama, é praticamente indispensável para especificar a exata natureza da estrutura proposta, de acordo com a qual serão feitas as análises subseqüentes. A construção desse esquema causal deve estar baseada em relações causa-efeito, determinadas em estudos prévios, ou por meio de hipóteses, as quais se deseja testar. A escolha de uma forma de relacionamento particular fica a critério do pesquisador, mas uma vez escolhida deve ser mantida em todo o diagrama. Ainda segundo esse autor, a análise de trilha não é um meio para se conhecer formas de relacionamento a partir de correlações observadas entre variáveis inter-relacionadas. Ao contrário, quanto mais se conhece a relação entre as variáveis, melhores resultados podem ser obtidos com a análise de trilha.

A construção gráfica do esquema causal possibilita a obtenção de um conjunto de equações simultâneas. A solução destas fornece informações sobre a contribuição direta e indireta das causas sobre o efeito, conforme SINGH e CHAUDHARY (1979).

Para a análise de trilha, o grau de influência de uma variável sobre a outra deve ser expresso em termos quantitativos. Estabelecido o diagrama causal, efetua-se a análise, a qual consiste em subdividir essa influência em efeitos direcionais diretos e indiretos. Desse modo, são determinados os valores correspondentes a esses efeitos, os quais são denominados coeficientes de caminamento ou coeficientes de trilha.

Quando é considerado um único diagrama causal, a análise de trilha torna-se simplesmente uma análise de regressão parcial padronizada. Porém, quando são envolvidos vários diagramas causais, o método corresponde a uma expansão da regressão múltipla, como geralmente ocorre em estudos de melhoramento, envolvendo, por exemplo, os componentes primários e secundários da produção de grãos ou frutos, conforme CRUZ e REGAZZI (1992).

O uso do método de análise de trilha para estudo do relacionamento entre caracteres em plantas teve início com DEWEY e LU (1959). BRATT (1973), trabalhando com a cultura do trigo, demonstrou a importância do desdobramento da correlação entre altura da planta e produção de grãos, via análise de trilha.

Trabalhos com a cultura da seringueira, PAIVA et al. (1982) e VASCONCELOS (1982), e com o feijoeiro, DUARTE e ADAMS (1972), evidenciaram a importância da análise de trilha no melhoramento de plantas. Esses estudos possibilitaram, por exemplo, o conhecimento da influência de um caráter qualquer sobre o caráter principal, mesmo quando a seleção era efetuada sobre um terceiro caráter.

VIANNA et al. (1980) utilizaram a análise de trilha para estudar as inter-relações entre sete caracteres em progênies de milho, obtendo uma visualização e interpretação mais adequadas do real inter-relacionamento entre os caracteres estudados, quando em comparação com a análise de correlação simples. Trabalhando também com a

cultura do milho, LENG (1963) já havia analisado a influência do número de fileiras de grãos na espiga, número de grãos por fileira, peso de grãos e número de espigas por planta, sobre a produtividade de grãos.

O método da análise de trilha tem sido utilizado para estudo de inter-relacionamento de caracteres em várias outras culturas, como no algodão, por BUTANY et al. (1968); no amendoim, por KHANGURA e SANDHU (1972); no arroz, por LENKA e MISRA (1973); na cevada, por SHARMA et al. (1973); no fumo, por RAO et al. (1973); na linhaça, por SINGH (1980); na soja, por LAL e HAQUE (1971), e PANDEY e TORRIE (1973); no sorgo, por NAPHADE (1972); no trigo, por SHRIVASTAVA et al. (1980) e SOUZA (1985). NASCIMENTO FILHO (1988) utilizou coeficientes de caminhamento para estudar as correlações entre peso de raízes, comprimento e diâmetro basal do ramo, número de folhas, área foliar, peso seco da parte aérea e peso seco total, no guaranazeiro.

Trabalhando com a cultura do feijão, CASTOLDI (1991) utilizou o método dos coeficientes de trilha para desdobrar correlações genotípicas em componentes de efeito direto e indireto. Nesta análise, as variáveis correspondentes ao número de vagens por parcela (X_1), número de sementes por vagem (X_2) e peso médio de 100 sementes (X_3), foram consideradas como fatores primários da produção (Y). Os valores observados referentes a estas variáveis foram transformados em logaritmos. As variáveis correspondentes à área foliar (X_5), número de folíolos (X_6), altura média de plantas (X_7) e peso da palhada (X_8), foram considerados como sendo fatores secundários. Observou-se que as variáveis utilizadas explicaram totalmente a variação da variável básica com base no esquema causal proposto, proporcionando um coeficiente de determinação equivalente a 100%. Evidenciou-se, ainda,

a vantagem de se adotar a seleção indireta para produção de grãos via número de vagens por parcela.

2.2. Uso do Índice de Seleção no Melhoramento de Plantas

Sabe-se que grande parte dos caracteres de interesse econômico, tanto em vegetais como em animais, têm controle genético quantitativo, o que os torna dependentes das condições ambientais. Assim, em programas de melhoramento, surgem dificuldades advindas da influência das variações não herdáveis sobre as variações hereditárias. Ao se estabelecer métodos de seleção, é desejável que a probabilidade de obtenção de genótipos superiores seja a mais alta possível. Vários métodos têm sido propostos com essa finalidade, como níveis independentes de eliminação, seleção em "tandem" e índices de seleção. Esses métodos, por sua vez, envolvem geralmente um grande número de variações.

Quando a seleção tem por base vários caracteres, estes podem ser considerados simultaneamente em níveis independentes de eliminação ou podem ser selecionados um de cada vez, no método em "tandem", conforme SILVA (1982). Estes dois processos apresentam, no entanto, algumas limitações. Quando os níveis de eliminação são considerados independentemente, indivíduos e progênies que apresentam quase todos os caracteres desejáveis podem ser eliminados indevidamente, caso um dos caracteres considerados não atinja o limite máximo, ou mínimo, estabelecido. Por outro lado, o método em "tandem" baseia-se na pressuposição de que ao se proceder a seleção para um caráter, os outros permanecem inalterados. Havendo, porém, correlações entre os caracteres essa condição nem sempre é satisfeita. Assim, a utilização desse método pode tornar-se complexa e ineficiente.

Os índices de seleção são estabelecidos com base em um conjunto de caracteres ou em um caráter isoladamente, como é o caso do índice com base em informações de parentes, e têm por finalidade classificar, em conformidade com determinado critério, indivíduos e progênes submetidos ao processo seletivo. PIRCHNER (1983) define o índice de seleção como sendo uma equação de regressão múltipla, na qual os valores genéticos dos indivíduos submetidos ao processo seletivo constituem as variáveis dependentes e as informações correspondentes ao desempenho desses indivíduos, ou de seus parentes, constituem as variáveis independentes. O índice de seleção pode ser usado para determinar a melhor combinação das informações obtidas do indivíduo, ou do indivíduo e de seus parentes, em relação a um determinado caráter, ou então para combinar em um índice agregado as informações obtidas de um modo semelhante, relativas a um conjunto de caracteres, conforme SILVA (1980). O índice de seleção foi proposto inicialmente para características múltiplas em plantas, por SMITH (1936). Posteriormente, foi utilizado por HAZEL (1943) no melhoramento animal. HENDERSON (1963) ampliou seu uso para seleção de linhas e cruzamentos entre linhas.

Estudos comparativos evidenciam a superioridade do índice de seleção em relação a outros métodos, quando são considerados conjuntos de caracteres, conforme HAZEL e LUSH (1942), YOUNG (1961) e ELGIN et al. (1970).

O índice de seleção é estabelecido de modo a maximizar a correlação entre o próprio índice (I) e um valor genético verdadeiro ou agregado genotípico (H), o qual é desconhecido na prática. Conforme LIN (1978), tem-se

$$H = a_1g_1 + \dots + a_mg_m = a'g$$

$$I = b_1x_1 + \dots + b_mx_m = b'x,$$

em que

$g' = (g_1 \ g_2 \ \dots \ g_m)$ = vetor de m valores genotípicos desconhecidos

$x' = (x_1 \ x_2 \ \dots \ x_m)$ = vetor de m valores fenotípicos conhecidos

$a' = (a_1 \ a_2 \ \dots \ a_m)$ = vetor de pesos econômicos relativos aos m

caracteres avaliados

$b' = (b_1 \ b_2 \ \dots \ b_m)$ = vetor de coeficientes de ponderação dos

caracteres no índice a ser estimado.

Observa-se ainda as relações:

$$\sigma_I^2 = b'Pb; \quad \sigma_H^2 = a'Ga; \quad \sigma_{IH} = b'Ga.$$

P e G são as matrizes de variância e covariância fenotípica e genotípica, respectivamente.

De acordo com CRUZ (1990), o índice de seleção constitui um caráter adicional, estabelecido pela combinação ótima de vários caracteres. Isso torna possível a seleção simultânea de vários caracteres, com eficiência. Sendo

$$\rho_{IH} = b'Ga / [(b'Pb)(a'Ga)]^{1/2},$$

a equação $P\hat{b} = Ga$ é obtida pela maximização de ρ_{IH} . Assim, matrizes de variâncias e covariâncias bem estimadas, e pesos econômicos, relativos aos caracteres, bem estabelecidos, constituem condições básicas para que estimativas confiáveis dos índices possam ser obtidas. PESEK e BAKER (1969), substituíram os pesos econômicos relativos por

ganhos desejados. Utilizando esta metodologia, CRUZ (1990) substituiu o vetor de ganhos esperados (Δg) na expressão

$$\Delta g = G \cdot \hat{b} \cdot i_1 / [V(I)]^{1/2}$$

pelo vetor de ganhos desejados (Δg_d), estimando \mathbf{b} pela expressão

$$\hat{b} = G^{-1} \cdot \Delta g_d \frac{[V(I)]^{1/2}}{i_1}.$$

Como $[V(I)]^{1/2}/i_1$ não afeta a proporcionalidade dos \hat{b}_i 's, pode-se estimar \mathbf{b} simplesmente por

$$\hat{b} = G^{-1} \cdot \Delta g_d.$$

Assim, especificando-se os ganhos desejados, podem ser maximizados os ganhos em cada caráter, conforme a preferência do melhorista. No trabalho citado, o autor estabeleceu como ganho desejado o equivalente a um desvio-padrão genético do caráter sob seleção, com sinal apropriado. O método dispensa o estabelecimento prévio dos pesos econômicos dos caracteres, mas estes podem ser estimados "a posteriori", através do vetor $\hat{\mathbf{b}}$, pela expressão

$$\hat{\mathbf{a}} = G^{-1} \cdot P \cdot \hat{\mathbf{b}}.$$

Pode-se, então, estimar os ganhos indiretos esperados para cada caráter $[\Delta g_i(I)]$, e também o ganho esperado no agregado genotípico.

TAI (1977) sugeriu uma variação no método de PESEK e BAKER ao propor a construção de um índice de seleção com base na especificação de ganhos genéticos desejados para características de

valor econômico. Esse autor sugere, ainda, que caracteres secundários também podem ser incluídos no índice auxiliando o processo de seleção.

Um índice que não tem como base pesos econômicos foi proposto por ELSTON (1963). BINET (1965) propôs um índice para seleção indireta, com base em dois caracteres, objetivando o melhoramento de um terceiro.

KEMPTHORNE e NORDSKOG (1959) propuseram o uso de índices de seleção restritos, pela fixação de limites para algumas variáveis, de acordo com a preferência do melhorista. Modificações desse processo foram introduzidas por TALLIS (1962) e por Moem (1968). Neste último trabalho, o autor, citado por SILVA (1980), baseia-se na introdução de uma "variável muda" ao índice, para cada restrição imposta. JAMES (1968) obteve um índice de seleção sob restrições lineares nos coeficientes e nas alterações das médias, apropriado para seleção indireta de caracteres não avaliáveis, com a garantia de que as médias correspondentes mantenham-se inalteradas.

A avaliação de um índice de seleção foi discutida por WILLIAMS (1962), enquanto SUBANDI et al. (1973) compararam a eficiência de índices de seleção para três caracteres em cruzamentos de duas variedades de milho.

Algumas questões enumeradas por LIN (1978), como dificuldade de obtenção de estimativas precisas de variâncias e covariâncias, possibilidade de mudança nos parâmetros genéticos ao longo do processo seletivo, e dificuldade na atribuição de pesos econômicos aos caracteres, têm limitado o uso dos índices de seleção no melhoramento de plantas. Apesar disso, muitos trabalhos têm sido feitos com índices de seleção.

HANSON e JOHNSON (1957) apresentaram um método para cálculo de um índice pelo agrupamento de dados obtidos em dois ou mais experimentos. BRIM et al. (1959) utilizaram duas populações de linhagens **F₃** de soja, avaliadas em locais diferentes, para construir um Índice de seleção com base nos teores de óleo e proteína, produção de grãos, acamamento, peso da semente e período de frutificação. Considerando o valor relativo entre óleo e proteína equivalente a 1:6, o ganho esperado para teor de óleo e peso das sementes chegou a superar em 18% o ganho no esperado com base no teor de óleo e proteína e na produção.

Verifica-se que o índice de seleção tem sido utilizado com mais freqüência no melhoramento animal. No melhoramento vegetal seu uso é quase restrito a espécies perenes ou semi-perenes. CRUZ (1990) considera que os índices não são de uso rotineiro como critério de seleção em espécies de ciclo curto em consequência da atuação de fatores limitantes, como a falta de uma manipulação mais precisa dos dados experimentais, juntamente com a escassez de recursos computacionais, incluindo-se aí a falta de programas apropriados. Esses fatores quase sempre impossibilitam a rápida identificação das unidades superiores. Em se tratando de espécies de ciclo curto, essa rapidez é essencial. Há ainda escassez de comparações entre os ganhos a serem obtidos por diferentes critérios de seleção. Um fator que também tem contribuído para a não-generalização do uso do índice de seleção é a falta de evidência de sua superioridade em relação a outras técnicas. Para comparar a eficiência de diferentes índices de seleção em relação à seleção direta, CRUZ (1990) adotou duas estratégias:

a) Comparar os ganhos totais, proporcionados pela seleção direta em determinado caráter, com os ganhos obtidos ao utilizar como critério de seleção o índice clássico proposto por SMITH (1936) e HAZEL

(1943), e ainda o índice por ganhos desejados proposto por PESEK e BAKER (1969), considerando um caráter como principal e os demais como secundários.

b) Comparar os ganhos totais, proporcionados pela seleção direta, com os ganhos obtidos por diferentes índices de seleção, com vários caracteres considerados como principais. Para esse estudo foram utilizadas 100 progênies de irmãs completas de milho, avaliando-se oito caracteres de importância econômica.

Nesse trabalho, CRUZ utilizou ainda os seguintes índices: i) Índice base, proposto por WILLIAMS (1962); ii) Índice "livre de pesos" proposto por ELSTON (1963); iii) Índice com base em somas de postos ou "ranks" proposto por MULAMBA e MOCK (1978).

As comparações entre os ganhos proporcionados pelo uso de índices e pela seleção direta permitiram concluir que os ganhos pela seleção direta são em geral maiores no caráter principal, porém, inferiores em termos totais. Assim, os índices são vantajosos por proporcionar maiores ganhos totais.

O autor verificou, ainda, que os diversos índices de seleção proporcionaram ganhos totais superiores aos obtidos pela seleção em um só caráter. O índice mais eficiente foi o de SMITH (1936) e HAZEL (1943). O menos eficiente foi o índice proposto por PESEK e BAKER (1969). Os demais índices proporcionaram ganhos intermediários.

Outros trabalhos também têm demonstrado a eficiência dos índices de seleção. MATZINGER et al. (1976) utilizaram um índice para obter aumento do número de folhas em plantas de fumo e, simultaneamente, reduzir o porte da planta. MARTIN et al. (1982) citam que ROSIELLE et al. (1977) também conseguiram aumentar a produção de aveia sem causar mudanças indesejáveis em outros caracteres. KAUFFMANN e DUDLEY (1979), utilizando essa mesma técnica,

obtiveram aumento na produção de grãos e no teor de proteína em milho.

O uso de índices de seleção, em plantas alógamas de ciclo curto, ocorre mais freqüentemente na cultura de milho. Em trabalhos de ROBINSON et al. (1951), LAIBLE e DIRKS (1968), WOLFF (1972) e GUTIERREZ (1974), constata-se a superioridade do método de seleção através de índices, quando em comparação com outros métodos, como níveis independentes, seleção em "tandem" etc. Nesses exemplos, os índices foram calculados com base em médias de progênies. REGAZZI (1978), aplicando o método de PESEK e BAKER (1969), ao estabelecer índices de seleção para progênies derivadas de um composto de milho, utilizando oito caracteres considerados de valor econômico, verificou que índices com base em variâncias e covariâncias genotípicas podem apresentar estimativas de ganhos genéticos irreais. O autor observou também que os índices de seleção podem ser sensíveis a possíveis erros de estimação das variâncias e covariâncias.

Trabalhando com a cultura da cenoura, VIEIRA (1988) estimou um índice de seleção por meio do ganho genético desejado, segundo a metodologia de PESEK e BAKER (1969), com base em sete caracteres considerados de importância econômica. Para isso, estabeleceu-se que o ganho genético desejado seria equivalente a um desvio-padrão genético do caráter correspondente. Dentre os sete valores estimados para b_i , dois apresentaram sinal negativo, indicando que durante o processo seletivo deveria ser dada preferência àquelas progênies que apresentassem os menores valores correspondentes a esses dois \hat{b}_i 's. Verificou-se ainda que, para todos os caracteres considerados no índice, seriam necessárias, no máximo, duas gerações de seleção para que o ganho genético desejado fosse atingido. Para um dos caracteres avaliados, a taxa de ganho esperado foi excepcionalmente alta. O autor

atribuiu esse problema à existência de erros inerentes às estimativas dos componentes de variância utilizados.

2.3. Efeitos da Multicolinearidade sobre os Parâmetros Genéticos

2.3.1. Conceito e Causas da Multicolinearidade

Se em uma análise de regressão múltipla as variáveis independentes, ou explicativas, forem correlacionadas entre si, diz-se que há intercorrelação ou multicolinearidade entre elas. Algumas vezes o segundo termo é reservado para os casos em que a correlação entre as variáveis é muito alta ou mesmo perfeita conforme NETER e WASSERMAN (1974). Os problemas causados pela multicolinearidade não são devidos simplesmente à sua presença, mas sim ao grau com que se manifesta. KMENTA (1971) salienta que ao estar relacionada às variáveis explicativas, as quais assume-se não serem estocásticas, a multicolinearidade é uma característica da amostra e não da população. Assim, não se aplica um teste para multicolinearidade, mas pode-se avaliar sua intensidade em uma amostra particular.

Ocorrendo correlação perfeita entre as variáveis, verifica-se que infinitos modelos de regressão podem ser ajustados perfeitamente aos dados. Em uma situação como esta não se pode interpretar qualquer conjunto de coeficientes de regressão como se esses refletissem os efeitos das variáveis. Ainda segundo esses autores, na prática, raramente são encontradas variáveis perfeitamente relacionadas ou dados que não contenham algum componente aleatório de erro, que normalmente contribuem para evitar a multicolinearidade perfeita. Mesmo assim, as implicações citadas permanecem relevantes.

Segundo NETER et al. (1983), em citação de FERRARI (1989), as variáveis independentes tendem a ser correlacionadas entre si em muitas situações nas várias áreas das ciências, incluindo-se entre estas as ciências biológicas.

A multicolinearidade é devida a alguns fatores como forma de obtenção dos dados, restrições impostas e também a especificação e super-parametrização do modelo, conforme MONTGOMERY e PECK (1981). Sabe-se que as variáveis multicolineares fornecem informações não muito diferentes daquelas proporcionadas pelas outras variáveis. BELSLEY et al. (1980) enfatizam que isso torna difícil avaliar as influências dessas variáveis sobre a resposta na variável principal.

Quanto aos seus efeitos adversos, a multicolinearidade pode afetar as estimativas de mínimos quadrados dos coeficientes de regressão. Essas estimativas são mais afetadas pela multicolinearidade que pelo relacionamento entre a resposta e as variáveis independentes, conforme MASON et al. (1975).

HOERL e KENNARD (1970b) demonstram que a multicolinearidade pode afetar o quadrado da distância entre o estimador de mínimos quadrados $\hat{\beta}$ e o parâmetro β . Representando essa distância por L , tem-se:

$$L = \hat{\beta} - \beta$$

$$L^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \|\hat{\beta} - \beta\|^2$$

$$E(L^2) = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2$$

$$E(L^2) = \sum_{j=1}^p V(\hat{\beta}_j) = \sigma^2 \text{Tr}(X'X)^{-1}. \text{ Assim, tem-se:}$$

$$E(L^2) = \sigma^2 \sum_{j=1}^p \lambda_j^{-1}.$$

Nessa expressão tem-se que $\lambda_j > 0$ ($j = 1, 2, \dots, p$) corresponde aos autovalores da matriz $\mathbf{X}'\mathbf{X}$, considerada na forma de correlações. Se $\mathbf{X}'\mathbf{X}$ é mal condicionada por efeito da multicolinearidade, pelo menos um λ_j será muito pequeno, provocando a ocorrência de valores altos para as distâncias entre $\hat{\beta}$ e β , segundo SILVEY (1969). WEBSTER et al. (1974) chamam a atenção para a estimativa da variância de $\hat{\beta}_j$ expressa por

$$V(\hat{\beta}_j) = (1 - R_j^2)^{-1} \sigma^2,$$

na qual R_j^2 é o coeficiente de determinação múltiplo da regressão de X_j sobre as variáveis independentes restantes. Ocorrendo multicolinearidade forte entre X_j e algum subconjunto das demais variáveis independentes, R_j^2 atingirá um valor próximo da unidade, tornando a variância de $\hat{\beta}_j$ muito grande.

MANDEL (1982) observa que o método dos mínimos quadrados pode produzir estimativas satisfatórias dos parâmetros em presença de multicolinearidade apenas nas situações em que as predições estiverem restritas às regiões do espaço das variáveis independentes onde existe a multicolinearidade.

2.3.2. Efeitos da Multicolinearidade sobre a Análise de Trilha

Sendo a análise de trilha uma forma de estudo de regressão com as variáveis padronizadas, com base em matrizes de correlações, pode-se deduzir, por analogia, o virtual dano que a multicolinearidade causa ao processo de estimação dos coeficientes. Assim é de se esperar que, em presença de multicolinearidade, as variâncias associadas aos

estimadores dos coeficientes de trilha atinjam valores demasiadamente elevados tornando os mesmos pouco confiáveis.

A multicolinearidade pode afetar, também, o coeficiente de determinação total (R^2) na análise de trilha, o qual corresponde à determinação da variável tomada como base, em função das variáveis explicativas, tanto no diagrama de trilha como no modelo de regressão linear múltipla, conforme CRUZ e REGAZZI (1992). LI (1956) define esse coeficiente como sendo a soma dos quadrados dos efeitos diretos mais o duplo produto entre esses efeitos e o coeficiente de correlação entre as variáveis explicativas. Se a variável tomada como base é completamente explicada por estas, tem-se $R^2=1$. Caso contrário, R^2 será sempre inferior à unidade. No entanto, freqüentemente encontra-se $R^2 > 1$.

2.3.4. Diagnóstico de Multicolinearidade

2.3.3. Efeitos da Multicolinearidade sobre os Índices de Seleção

Um índice de seleção, como o índice clássico ou o índice com base em ganhos desejados, é função da inversa da matriz de variâncias e covariância fenotípicas (P) ou genotípicas (G), respectivamente, como visto. Sabe-se, também, que o grau de condicionamento dessas matrizes é determinado pelo relacionamento entre as variáveis consideradas. Se ocorrer multicolinearidade perfeita entre algumas dessas variáveis, as matrizes P e G serão singulares, isto é, não terão inversa única. Desse modo, infinitos índices poderiam ser estabelecidos a partir de inversas generalizadas de P e G, mas nenhum desses índices teria sentido prático. A condição hipotética de multicolinearidade perfeita constitui um caso extremo. Pode-se, porém, entender que à medida que P e G aproximam-se dessa condição, o índice correspondente torna-se menos confiável. Isso pode ser explicado pela variância associada aos coeficientes de ponderação do índice (b's), os

quais tornam-se maiores à medida que P e G aproximam-se da singularidade.

Taxas de ganho por seleção, com base em índices, excessivamente altas e não condizentes com observações anteriores, podem ser devidas a erros associados às estimativas dos componentes de variância, constituindo, assim, indícios de multicolinearidade forte. Cunningham (1969), citado por VIEIRA (1988), aponta esses erros de estimativas como sendo a fonte que mais contribui para reduzir a eficiência do índice. O próprio VIEIRA (1988), ao encontrar alta taxa de ganho esperado para um caráter, atribuiu o problema à "existência de erros inerentes às estimativas dos componentes de variância utilizados".

2.3.4. Diagnóstico da Multicolinearidade

Um procedimento de diagnóstico eficiente para multicolinearidade deve refletir diretamente o grau com que esta se manifesta e ainda possibilitar a identificação das variáveis independentes envolvidas no problema, como ressaltam MONTGOMERY e PECK (1981). FERRARI (1989) destaca os seguintes procedimentos mais comumente utilizados para detectar multicolinearidade:

a) Métodos informais

Alguns métodos informais podem ser usados para detectar multicolinearidade, apesar de não fornecerem informações sobre sua intensidade e natureza. NETER et al. (1983) destacam as seguintes evidências de multicolinearidade, conforme citação de FERRARI (1989):

- efeito da exclusão ou inclusão de uma variável, ao provocar grandes alterações nos coeficientes de regressão estimados, ou efeito semelhante, quando uma observação é alterada ou retirada;

- coeficientes de regressão que apresentam resultados não-significativos em testes individuais para variáveis independentes importantes;

- coeficientes de regressão que não apresentam o sinal algébrico esperado em consonância com experiências anteriores; e

- grandes amplitudes para intervalos de confiança correspondentes a coeficientes de regressão que representam importantes variáveis independentes.

b) Exame do determinante de $X'X$

Desde que $X'X$ esteja na forma de correlação, seu determinante, $|X'X|$, pode variar de zero a um. Se $|X'X| = 1$, as variáveis independentes são ortogonais. Se $|X'X| = 0$, existe uma dependência linear completa entre estas variáveis. À medida em que $|X'X|$ aproxima-se de zero a multicolinearidade torna-se mais intensa. Porém, como observam MONTGOMERY e PECK (1981), esse método não possibilita a identificação das variáveis causadoras de multicolinearidade.

c) Análise da matriz de correlação

Consiste na análise dos elementos não-diagonais r_{ij} na matriz $X'X$. Se as variáveis independentes, ou explicativas, X_i e X_j são aproximadamente linearmente dependentes, então $|r_{ij}|$ será próximo de 1. Verifica-se que um coeficiente de correlação alto indica multicolinearidade. Porém, quando o número total de variáveis

independentes é maior que dois, esta condição torna-se apenas suficiente, mas não necessária. Nesse caso, a ausência de correlação alta entre duas variáveis não evidencia ausência de multicolinearidade, conforme KMENTA (1971).

d) Análise dos autovalores e autovetores de $\mathbf{X}'\mathbf{X}$

Existindo uma ou mais dependências lineares aproximadas entre as variáveis, um ou mais autovalores $(\lambda_1, \lambda_2, \dots, \lambda_p)$ de $\mathbf{X}'\mathbf{X}$ serão pequenos, conforme BELSLEY et al. (1980), KENDALL (1957) e SILVEY (1969). MONTGOMERY e PECK (1981) propõem a avaliação do número de condição (**NC**) de $\mathbf{X}'\mathbf{X}$. Dada à simetria desta matriz o número de condição consiste na razão entre o maior e o menor autovalor, definindo-se

$$NC = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}$$

Salientam os autores que se **NC** < **100**, a multicolinearidade não constitui problema sério. Se **100** < **NC** < **1000**, a multicolinearidade é de moderada a forte, e **NC** > **1000** constitui indicio de multicolinearidade severa.

A análise dos autovalores pode identificar a natureza aproximada da dependência linear existente entre as variáveis segundo BELSLEY et al. (1980). Para isso, faz-se $\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, em que $\mathbf{\Lambda}$ é a matriz diagonal de dimensões $\mathbf{p} \times \mathbf{p}$, cujos elementos da diagonal são os autovalores λ_j , ($j = 1, 2, \dots, p$) de $\mathbf{X}'\mathbf{X}$, e \mathbf{V} é a matriz ortogonal de dimensões $\mathbf{p} \times \mathbf{p}$, cujas colunas são os autovetores normalizados de $\mathbf{X}'\mathbf{X}$. Sendo $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$, as colunas de \mathbf{V} , um autovalor (λ_j) próximo a zero

é indício de dependência linear entre as observações. Os elementos do autovetor (v_j) associado a esse autovalor descrevem a natureza dessa dependência.

e) Fatores de inflação da variância

MARQUARDT (1970) designa os elementos diagonais $v_{jj} = (1 - R_j^2)^{-1}$, da matriz $R_{xx}^{-1} = (X'X)^{-1}$, como sendo os fatores de inflação da variância (VIF) quando a matriz $X'X$ é tomada na forma de correlações. Estes fatores podem ser usados para detectar multicolinearidade, conforme MONTGOMERY e PECK (1981). A variância do j -ésimo coeficiente de regressão de mínimos quadrados é dada por $v_{jj}\sigma^2$. Assim, v_{jj} pode ser interpretado como sendo o fator que aumenta a variância de $\hat{\beta}$ quando ocorre dependência linear entre as variáveis. Para NETER et al. (1983), se qualquer um dos VIF's assume valor maior que 10, é possível que os coeficientes de regressão associados a esses valores tenham estimativas de mínimos quadrados demasiadamente influenciadas pela multicolinearidade.

f) Decomposição em valores singulares

Qualquer matriz ${}_nX_p$, sendo n observações e p variáveis, pode ser decomposta em $X=UDV'$, conforme LAWSON e HANSON (1974) e BELSLEY et al. (1980). Verifica-se que $X'X = (UDV')'(UDV') = VD^2V' = VAV'$, em que U tem dimensões $n \times p$, constituindo-se em uma matriz cujas colunas são os autovetores associados aos autovalores de $X'X$. V tem dimensões $p \times p$ e é a matriz dos autovetores normalizados de $X'X$. Tem-se que $U'U = V'V = I_p$. D é uma matriz diagonal de dimensões $p \times p$, com elementos diagonais não

negativos μ_j ($j=1, 2, \dots, p$), representando os valores singulares de \mathbf{X} . Então, $\mathbf{X}=\mathbf{UDV}'$ é a forma de decomposição de \mathbf{X} em seus valores singulares. Segundo MONTGOMERY e PECK (1981), o grau do mal-condicionamento em \mathbf{X} afeta o tamanho dos valores singulares e será tanto maior quanto maior for o valor singular para cada dependência linear aproximada. Essa relação é denominada índice de condição (η_k) da matriz \mathbf{X} , por BELSLEY et al. (1980) e IEMMA (1988), definindo-se,

$$\eta_k = \frac{\mu_{\text{máx}}}{\mu_k}, k=1, 2, \dots, p.$$

Assim tem-se que $\eta_k \geq 1$ para todo k .

A variância do estimador de mínimos quadrados de β pode ser escrita como

$$V(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{V}\Lambda^{-1}\mathbf{V},$$

ou, para o k -ésimo componente de $\hat{\beta}$:

$$V(\hat{\beta}_k) = \sigma^2 \sum_{j=1}^p \frac{v_{kj}^2}{\mu_j^2} = \sigma^2 \sum_{j=1}^p \frac{v_{kj}^2}{\lambda_j},$$

conforme FERRARI (1989), citando MONTGOMERY e PECK (1981) e IEMMA (1987). À exceção de σ^2 , o k -ésimo elemento diagonal de $\mathbf{V}\Lambda^{-1}\mathbf{V}$ é o k -ésimo fator de inflação da variância, isto é

$$(\text{VIF})_k = \sum_{j=1}^p \frac{v_{kj}^2}{\mu_j^2} = \sum_{j=1}^p \frac{v_{kj}^2}{\lambda_j}.$$

Torna-se então evidente que a ocorrência de um ou mais valores singulares, ou autovalores pequenos, provocará inflação nas variâncias de $\hat{\beta}_j$.

Pode-se também efetuar a decomposição da variância para medir o grau de multicolinearidade, como sugerem BELSLEY et al. (1980):

$$\Pi_{jk} = \frac{\left(\frac{v_{kj}^2}{\mu_j^2} \right)}{(VIF)_k}, j=1, 2, \dots, p.$$

Os elementos Π_{jk} são ordenados em uma matriz Π de dimensões $p \times p$, como no Quadro 1. Nesta matriz, os elementos de cada coluna de Π são as proporções da variância de cada $\hat{\beta}_k$, que é também cada fator de inflação da variância $(VIF)_k$, associado ao i -ésimo valor singular.

Proporções altas de Π_{jk} associadas a μ_j baixo constituem indicativos de que este valor singular está associado à multicolinearidade e esta, por sua vez, está inflacionando as variâncias de $\hat{\beta}_p$.

QUADRO 1 - Proporções da Decomposição da Variância

Valor singular associado	$\text{Var}(\hat{\beta}_1)$	$\text{Var}(\hat{\beta}_2)$...	$\text{Var}(\hat{\beta}_p)$
μ_1	Π_{11}	Π_{12}	...	Π_{1p}
μ_2	Π_{21}	Π_{22}	...	Π_{2p}
...
μ_p	Π_{p1}	Π_{p2}	...	Π_{pp}

FERRARI (1989).

2.3.5. Métodos Alternativos de Análise de Regressão em Presença de Multicolinearidade

Em presença de multicolinearidade, pode-se optar pela obtenção de dados adicionais ou pela reespecificação do modelo. A reespecificação pode ser feita eliminando-se as variáveis problemáticas ou empregando-se métodos alternativos ao método dos quadrados mínimos para estimação dos parâmetros. Entretanto, como salientam MONTGOMERY e PECK (1981), citados por FERRARI (1989), a obtenção de dados adicionais nem sempre é viável e eficiente, pois os novos dados podem também apresentar multicolinearidade. Quanto à eliminação de variáveis, esses autores ressaltam que apesar de ser uma técnica potencialmente eficiente, ela pode não ser satisfatória se as variáveis a serem retiradas do modelo estiverem explicando proporções consideráveis da variável resposta. A multicolinearidade pode ainda afetar alguns dos procedimentos empregados para selecionar as variáveis, tornando o processo ineficiente.

Os métodos alternativos ao de mínimos quadrados podem ser vantajosos na estimação de coeficientes de regressão, apesar de fornecerem estimadores tendenciosos, como salientam FERRARI (1989) e GUNST e MASON (1977). Dentre os métodos alternativos, destacam-se:

- a) Regressão em componentes principais;
- b) Regressão em cristas ou em cumeeira.

2.3.5.1. Regressão em Componentes Principais

De acordo com JOHNSON et al. (1973), a regressão em componentes principais pode ser vista como um estimador de mínimos quadrados restrito.

Os componentes principais podem ser obtidos, segundo HOCKING et al. (1976), considerando o modelo na forma canônica

$$y = Z\alpha + \varepsilon,$$

no qual as variáveis independentes são consideradas em ordem decrescente de seus autovalores, isto é, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Os "s" últimos desses autovalores são considerados como sendo aproximadamente iguais a zero. As colunas da matriz **Z** a eles correspondentes são excluídas da análise. A regressão em componentes principais é então obtida aplicando-se o método dos quadrados mínimos aos componentes restantes.

2.3.5.2. Regressão em Cristas ou em Cumeeira

A utilização da regressão em cristas foi proposta por HOERL e KENNARD (1970a,b), para as situações de quase singularidade da matriz **X'X** por efeito de multicolinearidade, originando estimativas instáveis dos parâmetros. Essa denominação foi dada por HOERL (1962) em decorrência da analogia com o método de análise em cristas ou em cumeeira, proposto anteriormente por ele mesmo, para análises de superfícies de respostas multidimensionais. HOERL (1985) descreve o método de análise em crista proposto por HOERL (1959) e discute a conexão existente entre análise em cristas e regressão em cristas. Conforme FERRARI (1989), os coeficientes de regressão em crista são obtidos utilizando-se uma versão parcialmente modificada das equações normais. Para isso, considera-se o modelo de regressão múltipla na notação matricial:

$$y = X\beta + \varepsilon,$$

em que

y : vetor com os valores da variável dependente;

X : matriz de dimensões $n \times p$, dos valores das variáveis independentes;

β : vetor dos parâmetros a serem estimados;

ε : vetor dos erros.

Admite-se que:

$$E(\varepsilon) = \Phi$$

$$E(\varepsilon'\varepsilon) = I\sigma_e^2,$$

sendo σ_e^2 a variância dos erros.

Para aplicação do método, as variáveis devem ser primeiramente padronizadas, evitando-se, assim, a influência das unidades de medida, como destaca KALIL (1977). Desse modo, $X'X$ e $X'y$ passam a ser matrizes de correlações simples. Assim faz-se

$$y_j = \frac{Z_j - \bar{Z}}{\sqrt{\sum (Z_j - \bar{Z})^2}}.$$

Nessa expressão, tem-se:

y_j : valor padronizado do j -ésimo elemento do vetor y ;

Z_j : valor original da j -ésima observação da variável dependente;

\bar{Z} : média aritmética dos valores de Z_j , para $j = 1, 2, \dots, n$.

De igual modo, tem-se a expressão

$$x_{ij} = \frac{V_{ij} - \bar{V}_i}{\sqrt{\sum (V_{ij} - \bar{V}_i)^2}},$$

em que

x_{ij} : valor padronizado da j -ésima observação da i -ésima variável que constituem a matriz X ;

V_{ij} : original da j -ésima observação da i -ésima variável independente, para $j = 1, 2, \dots, p$; $i = 1, 2, \dots, n$.

\bar{V}_i : média aritmética dos valores de V_{ij} para $i = 1, 2, \dots, p$.

Sabe-se que o estimador de mínimos quadrados ordinários de β é

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Como o método de regressão em cristas tem por objetivo reduzir a multicolinearidade pelo afastamento da solução de mínimos quadrados, o estimador em cristas (β^*) é obtido adicionando-se pequenas quantidades positivas (\mathbf{k}) aos elementos da diagonal da matriz $X'X$, na forma de matriz de correlações. Assim tem-se

$$\beta^* = (X'X + kI_p)^{-1}X'y.$$

São considerados os valores da constante \mathbf{k} , geralmente dentro do intervalo $0 \leq \mathbf{k} \leq 1$, uma vez que $X'X$ encontra-se na forma de matriz de correlações. A determinação correta dessa constante constitui a essência da regressão sobre cristas, conforme HOCKING et al. (1976).

DRAPER e SMITH (1981) salientam que existem situações para as quais o método da regressão em cristas constitui uma técnica teoricamente apropriada. No entanto, seu uso indiscriminado deve ser evitado.

2.3.5.2.1. *Propriedades dos Estimadores sobre Cristas ou em Cumeeira*

MARQUARDT e SNEE (1975) apresentam uma interpretação geométrica da regressão sobre cristas. Observam esses autores que o estimador de mínimos quadrados ($\hat{\beta}$) constitui a solução para a qual a soma de quadrados do resíduo (**SQR**) é mínima. O estimador sobre cristas (β^*) é o vetor solução de menor norma que proporciona um valor de **SQR** maior que o mínimo em proporção especificada.

HOERL e KENNARD (1970b) e KALIL (1977) destacam as seguintes propriedades apresentadas pelo estimador em cristas (β^*):

$$a) \mathbf{E}(\beta^*) = -k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta .$$

Logo, β^* é um estimador tendencioso de β ;

$$b) \beta^* = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X})\hat{\beta} ,$$

isto é, β^* é uma transformação linear de $\hat{\beta}$;

$$c) \mathbf{VAR} = \mathbf{COV}(\beta^*) = \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} .$$

Conseqüentemente, a variância de β^* é uma função decrescente de k ;

d) Designando $QME(\beta^*)$ o erro quadrático médio para o estimador β^* , tem-se

$$E[QME(\beta^*)] = E[(\beta^* - \beta)'(\beta^* - \beta)]$$

$$E[QME(\beta^*)] = k^2 \beta' (X'X + kI)^{-2} \beta + \sigma^2 \text{Tr}(X'X + kI)^{-1} (X'X) (X'X + kI)^{-1}$$

Assim tem-se que

$$E[QME(\beta^*)] = \sum [\text{VIÉS}(\beta^*)]^2 + \sum \text{VAR}(\beta_I^*) .$$

Constata-se que o quadrado do viés é uma função crescente de k .

e) Se $\beta'\beta$ é limitado, então existe $k > 0$, tal que

$$E[QME(\beta^*) | E(\hat{\beta})] = \sigma^2 \sum_{i=1}^p \lambda_i^{-1} .$$

HOERL e KENNARD (1970a) demonstram esta propriedade, à qual denominam teorema da existência.

2.3.5.2.2. Métodos para a Determinação de k

A determinação correta da constante k constitui a essência da regressão em cristas conforme HOCKING et al. (1976). Vários métodos são descritos na literatura, sendo que alguns são relacionados por FERRARI (1989). Dentre estes, destacam-se os seguintes:

a) Método com base no teorema da existência

Um valor apropriado para \mathbf{k} , com base no teorema da existência de HOERL e KENNARD (1970a) e descrito por HOERL et al. (1975), é dado por

$$\mathbf{k} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} .$$

Nessa expressão tem-se que $p = \text{posto}(\mathbf{X}'\mathbf{X})$. Os estimadores $\hat{\beta}$ e $\hat{\sigma}^2$ são obtidos pela solução de mínimos quadrados. Em consequência da multicolinearidade, $\hat{\beta}'\hat{\beta}$ pode assumir valores excessivamente elevados, tornando \mathbf{k} muito pequeno como ressalta FERRARI (1989). Conseqüentemente, por esse método, o estimador sobre cristas (β^*) estaria assumindo um valor muito próximo do estimador de mínimos quadrados ($\hat{\beta}$).

b) Método iterativo

Esse procedimento, proposto por HOERL e KENNARD (1976), baseia-se no método anterior. Consiste em seqüências de estimações de β e \mathbf{k} . Partindo-se de um estimador de mínimos quadrados $\hat{\beta}$, estima-se o primeiro valor de \mathbf{k} , pela expressão

$$\mathbf{k}_0 = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} .$$

Com base em \mathbf{k}_0 obtém-se o primeiro estimador em crista β^* :

$$\beta^*(\mathbf{k}_0) = (\mathbf{X}'\mathbf{X} + \mathbf{k}_0\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} .$$

O processo continua com a estimação de k_1 , com base em $\beta^*(k_0)$, e assim sucessivamente, até que se verifique

$$\frac{k_{j+1} - k_j}{k_j} < \rho^{-1,3}.$$

Nessa expressão tem-se que $\rho = \text{traço}(X'X)^{-1}$.

c) Método com base no traço de crista

O traço de crista é constituído por um gráfico dos elementos β_i^* , $i = 1, 2, \dots, p$, tomados em relação a \mathbf{k} . Em se tratando de variáveis padronizadas, os valores de \mathbf{k} encontram-se no intervalo $0 < \mathbf{k} < 1$. Uma ilustração desse gráfico, apresentado por DRAPER e SMITH (1981), página 316, é aqui reproduzido na Figura 1. Nesse gráfico, os valores de β_i^* , $i = 1, 2, \dots, p$, $i = 1, 2, 3, 4$ representam os coeficientes de regressão em crista, em função dos valores assumidos por \mathbf{k} .

Em presença de multicolinearidade alta, as estimativas dos coeficientes de regressão em crista mostram-se instáveis à medida que \mathbf{k} aumenta. Essas estimativas tendem a se estabilizar para algum valor de \mathbf{k} , como pode ser visto no gráfico da Figura 1.

A análise de crista consiste então em se estimar o valor de \mathbf{k} , para o qual as estimativas de β_i^* se estabilizam, conforme MONTGOMERY e PECK (1981) e HOERL e KENNARD (1970b). Esse valor de \mathbf{k} deve ser o menor possível, observando-se as condições:

- coeficientes já estabilizados e apresentando comportamento semelhante ao de um sistema ortogonal;

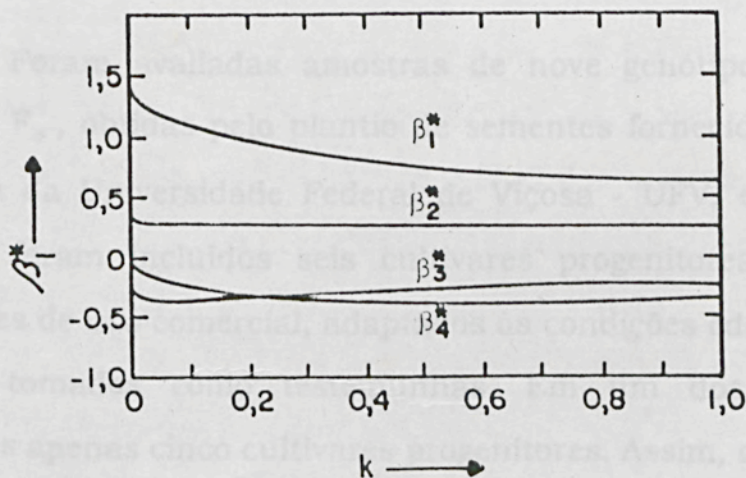
- coeficientes que apresentavam sinais não esperados em $\mathbf{k} = \mathbf{0}$, já apresentando sinais coerentes com a natureza do caráter correspondente; e

- resíduos não muito elevados, em relação ao resíduo obtido pela solução de mínimos quadrados ($\mathbf{k} = \mathbf{0}$).

3. MATERIAL E MÉTODOS

3.1. Caracterização dos Experimentos

3.1.1. Materiais



FONTE: DRAPER e SMITH (1981).

FIGURA 1 - Gráfico dos Elementos β_i^* , $i = 1, 2, \dots, p$, Tomados em Relação \mathbf{k} .

No Quadro 2 é apresentada a relação dos genótipos utilizados no experimento.

3.1.2. Local e Épocas de Instalação dos Experimentos

Os ensaios foram instalados em Viçosa, MG, em local situado a 649m de altitude, 20° 45' 20" S de latitude e 42° 52' 40" W de longitude.

O primeiro ensaio, aqui denominado Experimento 1, foi instalado no dia 10/09/1991. O segundo, denominado Experimento 2, foi instalado no dia 31/10/1991.

3. MATERIAL E MÉTODOS

3.1.3. Definição Experimental

3.1. Caracterização dos Experimentos

Utilizou-se, nos dois ensaios, o delineamento em blocos

3.1.1. Materiais

completamente aleatorizados, com quatro repetições. Cada parcela experimental foi constituída por duas fileiras de 5,00m de comprimento, espaçadas

Foram avaliadas amostras de nove genótipos de feijão, em geração F₃, obtidas pelo plantio de sementes fornecidas pelo setor de genética da Universidade Federal de Viçosa - UFV, em dois ensaios. Nestes, foram incluídos seis cultivares progenitores e ainda cinco cultivares de uso comercial, adaptados às condições edafo-climáticas da região, tomados como testemunhas. Em um dos ensaios foram incluídos apenas cinco cultivares progenitores. Assim, o primeiro ensaio foi constituído por 20 e o segundo por 19 tratamentos. Os cultivares progenitores dividem-se em dois grupos. Um grupo é constituído por cultivares portadores de genes para tolerância ao frio (Diacol Andino, Ica Llanogrande, Ica Tundama e Rojo 70). O outro grupo (Rio Tibagi, Carioca e Roxo) é constituído por cultivares de uso comercial e também adaptados à região onde o ensaio foi instalado.

No Quadro 2 é apresentada a relação dos genótipos utilizados no experimento.

3.1.2. Local e Épocas de Instalação dos Experimentos

Os ensaios foram instalados em Viçosa, MG, em local situado a 649m de altitude, 20° 45' 20" S de latitude e 42° 52' 40" W de longitude.

O primeiro ensaio, aqui denominado Experimento 1, foi instalado no dia 10/09/1991. O segundo, denominado Experimento 2, foi instalado no dia 31/10/1991.

3.1.3. Delineamento Experimental

Utilizou-se, nos dois ensaios, o delineamento em blocos completos casualizados, com quatro repetições. Cada parcela experimental foi constituída por duas fileiras de 5,00m de comprimento, espaçadas de 0,60m, perfazendo uma área de 6,00m². A densidade de sementeira foi de 15 sementes por metro linear. Uma bordadura constituída por um dos cultivares adaptados à região foi mantida ao redor de todo o ensaio. Práticas culturais preconizadas por VIEIRA (1983), para o feijoeiro, foram aplicadas ao ensaio.

3.1.4. Caracteres Avaliados

Foram avaliados os seguintes caracteres:

a) **Número de dias até o florescimento (FLR)**: Número de dias contados a partir da sementeira até que aproximadamente 50% das plantas apresentassem uma flor;

QUADRO 2 - Material Experimental de Feijão em Geração F₃, Cultivares Progenitores e Testemunhas, Utilizados no Experimento. Viçosa, MG, 1991

Item	Identificação de controle na UFV	Identificação descritiva
Materiais em geração F ₃	00 x 19	Rojo 70 x Rio Tibagi
	00 x 98	Rojo 70 x Ica Tundama
	00 x 94	Rojo 70 x Diacol Andino
	19 x 94	Rio Tibagi x Diacol Andino
	98 x 94	Ica Tundama x Diacol Andino
	97 x 19	Ica Llanogrande x Rio Tibagi
	97 x 98	Ica Llanogrande x Ica Tundama
	00 x 30	Rojo 70 x Carioca
Cultivares progenitores	54 x 30	Roxo x Carioca
	1994	Diacol Andino
	1998	Ica Tundama
	2000	Rojo 70
	1719	Rio Tibagi
	3030	Carioca
Cultivares testemunhas	2054	Roxo
	1919	Ouro
	1992	Ouro Negro
	1735	Rico
	2157	RA 94
	2164	CNF 5547

b) Número de dias até a maturação (MAT): Número de dias contados a partir da semeadura até que aproximadamente 90% das vagens estivessem secas;

c) Estande final (STD): Número de plantas existentes na parcela experimental por ocasião da colheita;

d) Número de vagens (NVG): Número médio de vagens por planta da parcela;

e) Número de sementes por vagem (NSV): Número médio de sementes por vagem;

f) Número de sementes (NSM): Número médio de sementes por planta da parcela;

g) Peso médio de 100 sementes (PCS): Tomado em nível de parcela, foi determinado pela expressão.

$$PCS = \frac{\text{Peso de sementes (gramas)}}{\text{Número de sementes}} \times 100;$$

h) Produção de grãos (PRD): Peso da produção de grãos, expresso em gramas, em nível de planta da parcela;

i) Peso da palhada (PPL): Peso médio dos restos de colheita (hastes, parte do sistema radicular e vagens vazias) em nível de planta da parcela, expresso em gramas; e

j) Índice de colheita (INC): Tomado em nível de parcela, foi obtido pela expressão.

$$INC = \frac{\text{Peso de grãos}}{\text{Peso da palhada + peso de grãos}}$$

3.2. Metodologias

3.2.1. Análise de Variância

Efetuuou-se a análise de variância para cada um dos caracteres avaliados, considerando-se o modelo estatístico

$$Y_{ij} = m + g_i + b_j + e_{ij}$$

em que

Y_{ij} : valor do i -ésimo tratamento (materiais segregantes ou cultivares) no j -ésimo bloco;

m : média geral;

g_i : efeito do i -ésimo tratamento (materiais segregantes ou cultivares), considerado aleatório, sendo $g_i \sim \text{NID}(0, \sigma_G^2)$. Para isso, assumiu-se que os tratamentos constituem amostra do germoplasma da UFV, em uso no programa de melhoramento, visando tolerância ao frio e adequados às modalidades de análises estatísticas e genéticas propostas para os dados. ($i=1, 2, \dots, g$).

b_j : efeito aleatório do bloco j , para $j=1, 2, \dots, r$; e

e_{ij} : efeito aleatório do erro entre parcelas, sendo $e_{ij} \sim \text{NID}(0, \sigma_e^2)$.

O esquema da análise de variância para cada um dos caracteres avaliados e esperanças dos quadrados médios são apresentados a seguir:

FV	GL	QM	E(QM)
Blocos	($r-1$)		
Tratamentos	($g-1$)	QMT	$\sigma_e^2 + r\sigma_G^2$
Erro	($r-1$) ($g-1$)	QMR	σ_e^2

Com base no esquema apresentado, foram obtidos os seguintes estimadores de parâmetros:

- $\hat{\sigma}_{P(média)}^2$: estimador da variância fenotípica entre as médias dos tratamentos;
- $\hat{\sigma}_{G(média)}^2$: estimador da variância genotípica entre as médias dos tratamentos;
- $\hat{\sigma}_{e(média)}^2$: estimador da variância residual média; e
- $\hat{\sigma}_e^2$: estimador da variância residual.

Esses estimadores foram obtidos pelas expressões correspondentes:

$$\hat{\sigma}_{P(média)}^2 = \frac{QMT}{r}$$

$$\hat{\sigma}_{G(média)}^2 = \frac{QMT - QMR}{r}$$

$$\hat{\sigma}_{e(média)}^2 = \frac{QMR}{r}$$

$$\hat{\sigma}_e^2 = QMR$$

Utilizando-se esses estimadores, foi obtida a herdabilidade em nível de média de tratamento ($h_{média}^2$), a herdabilidade no sentido amplo, em nível de parcela experimental ($h_{indiv.}^2$), e o coeficiente de variação genético [$CV_{genético} (\%)$], pelas expressões:

$$h^2_{(m\u00e9dia)} = \frac{\hat{\sigma}^2_{G(m\u00e9dia)}}{\hat{\sigma}^2_{P(m\u00e9dia)}}$$

$$h^2_{(indiv.)} = \frac{\hat{\sigma}^2_{G(m\u00e9dia)}}{\hat{\sigma}^2_{G(m\u00e9dia)} + \hat{\sigma}^2_e}$$

$$CV_{(gen\u00e9tico)}(\%) = \frac{\sqrt{\hat{\sigma}^2_{G(m\u00e9dia)}}}{\hat{m}} \cdot 100$$

Na \u00faltima express\u00e3o tem-se que

\hat{m} : m\u00e9dia geral do car\u00e1ter em estudo.

Obteve-se, ainda, a raz\u00e3o $CV_{(gen\u00e9tico)}(\%)/CV_{(experimental)}(\%)$, a qual constitui ind\u00edcios de situa\u00e7\u00e3o favor\u00e1vel ao melhoramento gen\u00e9tico, quando assume valor igual ou superior \u00e0 unidade, conforme VENCOSKY (1987).

3.2.2. Estima\u00e7\u00e3o de Componentes de Covari\u00e2ncias e Correla\u00e7\u00f5es Fenot\u00edpicas, Genot\u00edpicas e de Ambientes

3.2.2.1. Estima\u00e7\u00e3o das Covari\u00e2ncias

Para obten\u00e7\u00e3o dessas estimativas, foram realizadas an\u00e1lises de vari\u00e2ncia e de covari\u00e2ncia para cada par de caracteres, as quais s\u00e3o apresentadas esquematicamente no Quadro 3.

Sabe-se que a an\u00e1lise de vari\u00e2ncia de duas observa\u00e7\u00f5es (**X** e **Y**) de um mesmo indiv\u00edduo pode ser feita considerando-se as observa\u00e7\u00f5es compostas (**X + Y**) e analisando-as como se fossem observa\u00e7\u00f5es isoladas conforme KEMPTHORNE (1966).

QUADRO 3 - Esquema de Análises para Estimação das Covariâncias entre cada Par de Caracteres

F.V.	G.L.	Q. M.			P. M.	E(PM)
		X	Y	X+Y		
Blocos	(r-1)					
Tratamentos	(g-1)	$\hat{V}_f(X)$	$\hat{V}_f(Y)$	$\hat{V}_f(X+Y)$	$\text{côv}_f(X, Y) = [\hat{V}_f(X) - \hat{V}_f(Y)]/2$	$\sigma_e(X, Y) + r\sigma_G(X, Y)$
Erro	(r-1)(g-1)	$\hat{V}_a(X)$	$\hat{V}_a(Y)$	$\hat{V}_a(X+Y)$	$\text{côv}_a(X, Y) = [\hat{V}_a(X) - \hat{V}_a(Y)]/2$	$\sigma_e(X, Y)$

A variância da soma de duas variáveis **X** e **Y** pode ser obtida somando-se as variâncias das duas variáveis individuais com o dobro da covariância entre elas, ou seja,

$$V(X+Y) = V(X) + V(Y) + 2COV(X,Y). \text{ Logo,}$$

$$COV(X,Y) = \frac{V(X+Y) - V(X) - V(Y)}{2}$$

O estimador da covariância, equivalente ao produto médio entre **X** e **Y**, foi obtido pelos quadrados médios de **X**, **Y** e **X+Y** nas respectivas análises de variância. Desse modo, foram obtidos também os produtos médios associados a tratamentos (PMT_{XY}) e resíduo (PMR_{XY}), pelas expressões:

$$PMT_{XY} = (QMT_{X+Y} - QMT_X - QMT_Y)/2 = C\hat{O}V_F(X,Y) = \hat{\sigma}_{F(X,Y)}$$

e

$$PMR_{XY} = (QMR_{X+Y} - QMR_X - QMR_Y)/2 = C\hat{O}V_e(X,Y) = \hat{\sigma}_{e(X,Y)},$$

nas quais tem-se que:

$\hat{\sigma}_{F(X,Y)}$: estimador da covariância fenotípica entre dois caracteres (X e Y);

$\hat{\sigma}_{e(X,Y)}$: estimador da covariância de ambiente entre dois caracteres (X e Y).

Os componentes de covariância são estimados, conhecendo-se as esperanças do produto médio das fontes de variação, as quais são obtidas de modo equivalente às esperanças dos quadrados médios da análise de variância, apenas substituindo-se pelas covariâncias as

expressões correspondentes às variâncias. Assim, utilizando-se as expressões de $\mathbf{E}(\mathbf{PM})$, são obtidos os estimadores da covariância genotípica ($\hat{\sigma}_{G,XY}$) e de ambiente ($\hat{\sigma}_{e,XY}$), entre dois caracteres (\mathbf{X} e \mathbf{Y}), pelas expressões:

$$\hat{\sigma}_{G(X,Y)} = \frac{\text{PMT}_{XY} - \text{PMR}_{XY}}{r}$$

$$\hat{\sigma}_{e(X,Y)} = \text{PMR}_{XY}.$$

3.2.2.2. Estimação das Correlações

Os coeficientes de correlação fenotípica (\mathbf{r}_f), de ambiente (\mathbf{r}_a) e genotípica (\mathbf{r}_g), são obtidos pelas expressões correspondentes, como ilustram CRUZ e REGAZZI (1992), com base em produtos médios e componentes de variância e covariância, já definidos:

a) Correlação fenotípica

$$r_f = \frac{\text{PMT}_{XY}}{\sqrt{\text{QMT}_X \cdot \text{QMT}_Y}}.$$

b) Correlação de ambiente

$$r_a = \frac{\text{PMR}_{XY}}{\sqrt{\text{QMR}_X \cdot \text{QMR}_Y}}.$$

c) Correlação genotípica

$$r_g = \frac{\hat{\sigma}_{G(X,Y)}}{\sqrt{\hat{\sigma}_G^2(X) \cdot \hat{\sigma}_G^2(Y)}}.$$

3.2.3. Diagnóstico de Multicolinearidade

Para avaliar possíveis ocorrências de multicolinearidade nas matrizes de trabalho, foram utilizados quatro procedimentos descritos na literatura:

- análise descritiva das correlações;
- fatores de inflação da variância (VIF 's);
- análise dos autovalores e autovetores de $X'X$; e
- decomposição em valores singulares e decomposição da variância.

Esses métodos foram utilizados simultaneamente porque alguns deles são complementares entre si. Aproveitou-se também a oportunidade para comparar essas metodologias quanto à exeqüibilidade e eficiência.

3.2.3.1. Análise Descritiva das Correlações

Foram analisados os elementos não-diagonais (r_{ij}) das matrizes de trabalho obtidas a partir das variáveis padronizadas previamente.

Se as variáveis independentes (x_i , x_j) apresentarem dependência linear aproximada entre si, então r_{ij} , tomado em valor modular $|r_{ij}|$ será próximo à unidade (1). Contudo, a ausência de correlação alta entre duas variáveis não evidencia ausência de multicolinearidade, como já discutido no item 2.3.4.

3.2.3.2. Fatores de Inflação da Variância

Por esse método foi considerada a magnitude dos elementos diagonais v_{ij} da matriz $R_{xx}^{-1} = (X'X)^{-1}$, denominados VIF's ou fatores de

inflação da variância. A ocorrência de qualquer VIF com valor superior a 10 constitui indicativo de possíveis efeitos adversos provocados pela multicolinearidade, sobre os estimadores em uso, conforme NETER et al. (1983).

3.2.3.3. *Análise dos Autovalores e Autovetores de $X'X$*

Para cada matriz de trabalho $(X'X)$, foram obtidos os autovalores, ou raízes características, identificados por λ_j , para $j=1, 2, \dots, p$, sendo p o número de caracteres em estudo. Fazendo-se o determinante de $(X'X - \lambda I_p)$ igual a zero, isto é, $|(X'X - \lambda I_p)| = 0$, determinou-se λ_j . A seguir foi estabelecido o número de condição (**NC**), referente a cada matriz, pela expressão

$$NC = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}$$

Nos casos em que foi constatada multicolinearidade moderada a severa, isto é, **NC** > 100, conforme classificação de MONTGOMERY e PECK (1981), procurou-se reduzir sua influência pelo emprego de estimadores com base em traço de crista.

Tomando-se $(X'X - \lambda_j I_p)v$ e fazendo esta expressão identicamente nula, isto é, $(X'X - \lambda_j I_p)v = \Phi$, obtém-se o autovetor v , associado a cada autovalor λ_j .

Fazendo -se $X'X = V\Lambda V'$, sendo Λ uma matriz diagonal de dimensões $p \times p$, cujos elementos da diagonal são os autovalores λ_j da matriz $X'X$, e V uma matriz ortogonal, também de dimensões $p \times p$, cujas colunas são os autovetores normalizados de $X'X$, tem-se que um

autovalor (λ_j) com valor próximo a zero constitui indício de dependência linear aproximada entre os dados, enquanto os maiores elementos do autovetor (v_j), associado a esse autovalor, correspondem às variáveis que apresentam multicolinearidade entre si. Assim foi possível não só conhecer a intensidade da multicolinearidade em cada caso, como também identificar as variáveis envolvidas. O autovetor (\mathbf{v}) normalizado foi obtido dividindo-se seus elementos pela norma, dada por:

$$\|\mathbf{v}\| = \sqrt{\sum_{j=1}^p v_j^2},$$

em que v_j representa os elementos do autovetor \mathbf{v} .

3.2.3.4. Decomposição em Valores Singulares e Decomposição da Variância

Considerando que uma matriz ${}_nX_p$, sendo n observações e p variáveis, pode ser decomposta em ${}_nX_p = {}_nU_{pp}D_{pp}V'_p$ tem-se que $\mathbf{X}'\mathbf{X} = (\mathbf{U}\mathbf{D}\mathbf{V}')'(\mathbf{U}\mathbf{D}\mathbf{V}') = \mathbf{V}\mathbf{D}^2\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$. Observa-se, ainda, que $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_p$ e \mathbf{D} é uma matriz diagonal de dimensões $p \times p$, com elementos diagonais não negativos μ_j , $j = 1, 2, \dots, p$, sendo μ_j os valores singulares de \mathbf{X} . Assim, para cada matriz, foram calculados os valores de μ_j , determinando-se, também, o índice de condição (η_k) definido como

$$\eta_k = \frac{\mu_{\text{máx}}}{\mu_{\text{mín}}},$$

sendo $k = 1, 2, \dots, p$, e $\eta_k \geq 1$, para todo k .

A variância de um estimador de mínimos quadrados ($\hat{\theta}$) pode ser expressa por

$$V(\hat{\theta}) = \sigma^2(X'X)^{-1} = \sigma^2V\Lambda^{-1}V',$$

como já foi visto anteriormente. Para o k -ésimo componente de $\hat{\theta}$, tem-se também

$$V(\hat{\theta}_k) = \sigma^2 \sum \frac{v_{kj}^2}{\mu_j^2} = \sigma^2 \sum \frac{v_{kj}^2}{\lambda_j},$$

como visto em 2.3.4. Desse modo, calculou-se, para cada matriz de trabalho, o k -ésimo fator de inflação da variância $(VIF)_k$, definido como

$$(VIF)_k = \frac{\sum_{j=1}^p \frac{v_{kj}^2}{\mu_j^2}}{\sum_{j=1}^p \frac{v_{kj}^2}{\lambda_j}}.$$

Isso possibilitou também a identificação das variáveis envolvidas em multicolinearidade.

Aplicou-se, ainda, o método da decomposição da variância em seus elementos constitutivos (Π_{jk}) para avaliar a intensidade da multicolinearidade, utilizando a expressão

$$\Pi_{jk} = \frac{v_{kj}^2/\mu_j^2}{(VIF)_k}, \quad j = 1, 2, \dots, p,$$

como foi definida anteriormente. Os elementos Π_{jk} foram então ordenados em matrizes (Π) , de dimensões $p \times p$, nas quais os elementos de cada coluna constituem as proporções da variância de cada estimador ($\hat{\theta}_k$). Esses elementos são também os fatores de inflação da variância $(VIF)_k$ associados ao j -ésimo valor singular ou autovalor.

Assim, a associação entre proporções altas da decomposição da variância (Π_{jk}) e um valor singular (μ_j) baixo, foi tomada como indício de que a variável correspondente está, associada a um determinado grau de multicolinearidade, e esta por sua vez, está inflacionando as variâncias dos estimadores ($\hat{\theta}_k$) envolvidos.

3.2.4. Análise de Trilha

Utilizou-se o método dos coeficientes de trilha, desenvolvido por WRIGHT (1921a), para desdobrar as correlações genotípicas em componentes de efeitos diretos e indiretos.

Como preconizado na literatura, foram apresentados os diagramas ou esquemas causais, estabelecendo-se as relações entre cada variável básica e as variáveis explicativas, e também suas inter-relações. Nesses diagramas as setas unidirecionais indicam os efeitos diretos de cada variável explicativa, enquanto as setas bidirecionais simbolizam a interdependência das variáveis explicativas, cuja magnitude é quantificada pelo coeficiente de correlação genotípica (r_g). Os dados correspondentes aos caracteres em estudo foram padronizados previamente, dividindo-se o desvio de cada observação em relação à média, pelo respectivo desvio-padrão. As observações assim padronizadas passam a ter média igual a zero e variância igual à unidade.

Tomou-se por variável básica (Y) a produção de grãos e como fatores primários as variáveis:

a) número de vagens (X_1); b) número de sementes por vagem (X_2), c) peso médio de 100 sementes (X_3). Como fatores secundários foram consideradas as variáveis: número de dias para o florescimento (X_4); número de dias para a maturação (X_5) e peso da palhada (X_6); d) estande final (X_7); e) índice de colheita (X_8).

3.2.4.1. *Estimação dos Coeficientes de Trilha pelo Método dos Mínimos Quadrados*

a) **Efeitos de componentes primários sobre a variável básica**

Na Figura 2 é apresentado o diagrama correspondente às relações causa-efeito entre as variáveis primárias X_1 , X_2 e X_3 e a variável básica Y . O modelo matemático proposto para explicar o relacionamento entre essas variáveis é $Y = X_1 \cdot X_2 \cdot X_3$. Trata-se, portanto, de um modelo de efeitos multiplicativos entre as variáveis independentes. Porém, em um esquema causal, a variável básica é descrita por um modelo aditivo dos efeitos das variáveis explicativas. Assim, tornou-se necessário transformar o modelo multiplicativo em aditivo, adotando-se, para isso, a transformação logarítmica na base 10, obtendo-se

$$\log Y = \log X_1 + \log X_2 + \log X_3.$$

Conseqüentemente, essas variáveis passam a ser identificadas como logaritmo da produção de grãos ($\log Y$), logaritmo do número de

FIGURA 2 - Diagrama Representativo das Relações Causa-efeito entre as Variáveis Primárias e a Variável Básica.

vagens ($\log X_1$), logaritmo do número de sementes por vagem ($\log X_2$) e logaritmo do peso médio de 100 sementes ($\log X_3$).

Para o diagrama proposto, estabeleceu-se o seguinte sistema de equações normais padronizadas, cuja solução possibilitou a obtenção dos coeficientes de trilha:

$$\begin{cases} r_{y1} = p_{y1} + p_{y2}r_{12} + p_{y3}r_{13} \\ r_{y2} = p_{y1}r_{12} + p_{y2} + p_{y3}r_{23} \\ r_{y3} = p_{y1}r_{13} + p_{y2}r_{23} + p_{y3} \end{cases}$$

Matricialmente, na forma $(X'X)\hat{\beta} = X'Y$, esse sistema torna-se

$$\begin{bmatrix} r_{y1} \\ r_{y2} \\ r_{y3} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix} \begin{bmatrix} p_{y1} \\ p_{y2} \\ p_{y3} \end{bmatrix}$$

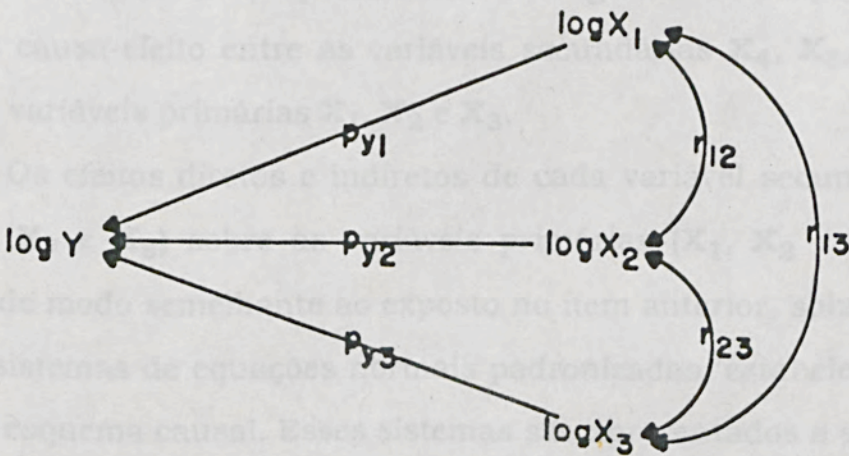


FIGURA 2 - Diagrama Representativo das Relações Causa-efeito entre as Variáveis Primárias e a Variável Básica.

A solução desse sistema para $(X'X)\hat{\beta} = X'Y$ é apresentada na forma

$$\begin{bmatrix} P_{y1} \\ P_{y2} \\ P_{y3} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{y1} \\ r_{y2} \\ r_{y3} \end{bmatrix}$$

Assim, podem ser conhecidos os efeitos diretos das variáveis explicativas primárias sobre a variável básica, representados por P_{y1} , P_{y2} , e P_{y3} e também os efeitos indiretos, representados por $P_{y2}r_{12}$, $P_{y3}r_{13}$, $P_{y1}r_{13}$, $P_{y3}r_{23}$, $P_{y1}r_{13}$ e $P_{y2}r_{23}$.

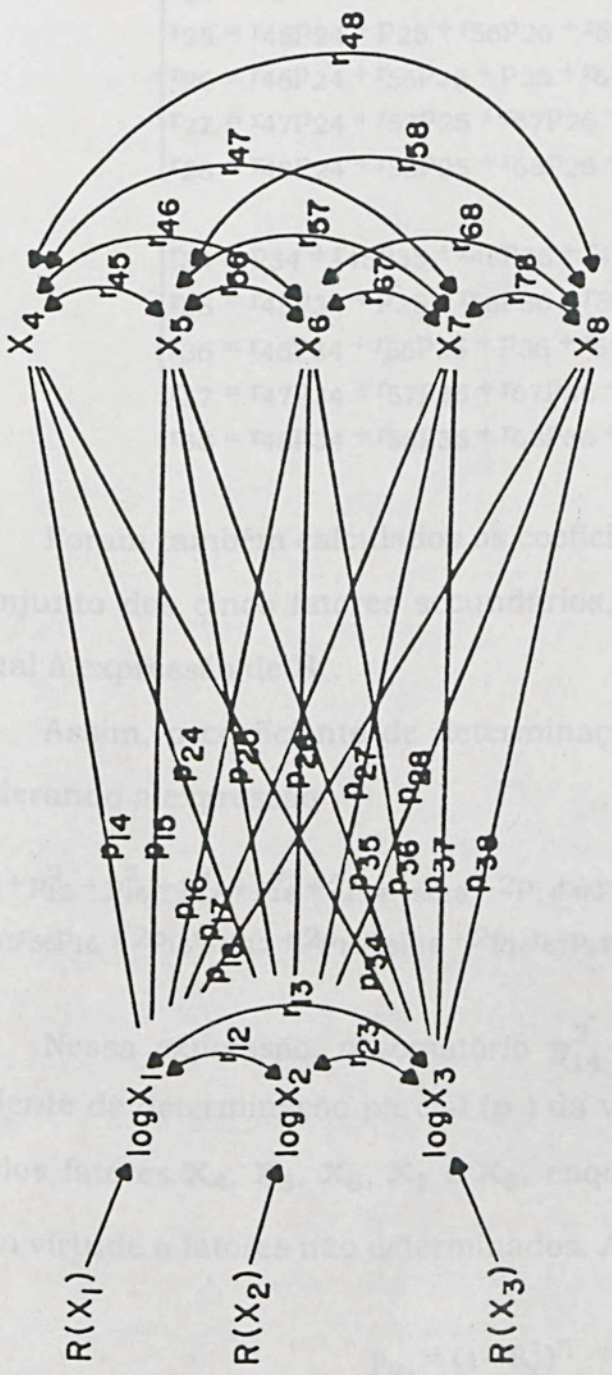
As variações da variável básica são totalmente explicadas por esse esquema causal. Logo, o coeficiente de determinação é igual à unidade. Matematicamente, esse coeficiente é expresso sob a forma

$$R_{y(1,2,3)}^2 = 1 = P_{y1}^2 + P_{y2}^2 + P_{y3}^2 + 2P_{y1}r_{12}P_{y2} + 2P_{y1}r_{13}P_{y3} + 2P_{y3}r_{23}P_{y2}.$$

b) Efeitos de Componentes Secundários sobre os Componentes Primários

Na Figura 3 é apresentado o diagrama correspondente às relações causa-efeito entre as variáveis secundárias X_4 , X_5 , X_6 , X_7 e X_8 e as variáveis primárias X_1 , X_2 e X_3 .

Os efeitos diretos e indiretos de cada variável secundária (X_4 , X_5 , X_6 , X_7 e X_8) sobre as variáveis primárias (X_1 , X_2 e X_3) foram obtidos de modo semelhante ao exposto no item anterior, solucionando-se três sistemas de equações normais padronizadas, estabelecidos com base no esquema causal. Esses sistemas são apresentados a seguir.



$R(X_1)$: Variável residual associada a X_1
 $R(X_2)$: Variável residual associada a X_2
 $R(X_3)$: Variável residual associada a X_3

FIGURA 3 - Diagrama Mostrando as Relações Causa-Efeito entre as Variáveis Secundárias (X_4, X_5, X_6, X_7 e X_8) e as Variáveis Primárias (X_1, X_2 e X_3), e Também as Variáveis Residuais Associadas.

$$\begin{cases} r_{14} = p_{14} + r_{45}p_{15} + r_{46}p_{16} + r_{47}p_{17} + r_{48}p_{18} \\ r_{15} = r_{45}p_{14} + p_{15} + r_{56}p_{16} + r_{57}p_{17} + r_{58}p_{18} \\ r_{16} = r_{46}p_{14} + r_{56}p_{15} + p_{16} + r_{67}p_{17} + r_{68}p_{18} \\ r_{17} = r_{47}p_{14} + r_{57}p_{15} + r_{67}p_{16} + p_{17} + r_{78}p_{18} \\ r_{18} = r_{48}p_{14} + r_{58}p_{15} + r_{68}p_{16} + r_{78}p_{17} + p_{18} \end{cases}$$

$$\begin{cases} r_{24} = p_{24} + r_{45}p_{25} + r_{46}p_{26} + r_{47}p_{27} + r_{48}p_{28} \\ r_{25} = r_{45}p_{24} + p_{25} + r_{56}p_{26} + r_{57}p_{27} + r_{58}p_{28} \\ r_{26} = r_{46}p_{24} + r_{56}p_{25} + p_{26} + r_{67}p_{27} + r_{68}p_{28} \\ r_{27} = r_{47}p_{24} + r_{57}p_{25} + r_{67}p_{26} + p_{27} + r_{78}p_{28} \\ r_{28} = r_{48}p_{24} + r_{58}p_{25} + r_{68}p_{26} + r_{78}p_{27} + p_{28} \end{cases}$$

$$\begin{cases} r_{34} = p_{34} + r_{45}p_{35} + r_{46}p_{36} + r_{47}p_{37} + r_{48}p_{38} \\ r_{35} = r_{45}p_{34} + p_{35} + r_{56}p_{36} + r_{57}p_{37} + r_{58}p_{38} \\ r_{36} = r_{46}p_{34} + r_{56}p_{35} + p_{36} + r_{67}p_{37} + r_{68}p_{38} \\ r_{37} = r_{47}p_{34} + r_{57}p_{35} + r_{67}p_{36} + p_{37} + r_{78}p_{38} \\ r_{38} = r_{48}p_{34} + r_{58}p_{35} + r_{68}p_{36} + r_{78}p_{37} + p_{38} \end{cases}$$

Foram também calculados os coeficientes de determinação total do conjunto dos cinco fatores secundários, incluindo-se uma variável residual à expressão de R^2 .

Assim, o coeficiente de determinação total para X_1 foi obtido considerando a expressão

$$1 = p_{14}^2 + p_{15}^2 + p_{16}^2 + p_{17}^2 + p_{18}^2 + 2p_{14}r_{45}p_{15} + 2p_{14}r_{46}p_{16} + 2p_{14}r_{47}p_{17} + 2p_{14}r_{48}p_{18} + 2p_{15}r_{56}p_{16} + 2p_{15}r_{57}p_{17} + 2p_{15}r_{58}p_{18} + 2p_{16}r_{67}p_{17} + 2p_{16}r_{68}p_{18} + 2p_{17}r_{78}p_{18} + p_{1R_1}^2$$

Nessa expressão, o somatório $p_{14}^2 + \dots + 2p_{17}r_{78}p_{18}$ constitui o coeficiente de determinação parcial (p) da variável explicativa primária X_1 pelos fatores X_4 , X_5 , X_6 , X_7 e X_8 , enquanto $p_{1R_1}^2$ é a variação em X_1 , em virtude a fatores não determinados. Assim,

$$p_{1R_1} = (1 - R_1^2)^{1/2}$$

De igual modo foram estudadas as determinações para X_2 e X_3 .

c) Efeitos de Componentes Secundários sobre a Variável

Básica

Na Figura 4 é apresentado um diagrama em cadeia envolvendo fatores secundários, fatores primários e a variável básica. O cálculo dos coeficientes de trilha relativos a esse diagrama foi realizado a partir de cinco equações normais padronizadas.

Considerando primeiramente a variável secundária X_4 , tem-se que a correlação (r_{4Y}) entre X_4 e a variável básica Y pode ser expressa por

$$\begin{aligned} r_{4Y} = \text{COV}(X_4, Y) &= PY_1r_{14} + PY_2r_{24} + PY_3r_{34} \\ &= PY_1(p_{14} + r_{45}p_{15} + r_{46}p_{16} + r_{47}p_{17} + r_{48}p_{18}) + \\ &\quad + PY_2(p_{24} + r_{45}p_{25} + r_{46}p_{26} + r_{47}p_{27} + r_{48}p_{28}) + \\ &\quad + PY_3(p_{34} + r_{45}p_{35} + r_{46}p_{36} + r_{47}p_{37} + r_{48}p_{38}) \end{aligned}$$

$$\begin{aligned} r_{4Y} = PY_1p_{14}^{(a)} + PY_1p_{15}r_{45}^{(b)} + PY_1p_{16}r_{46}^{(c)} + PY_1p_{17}r_{47}^{(d)} + PY_1p_{18}r_{48}^{(e)} + \\ + PY_2p_{24}^{(f)} + PY_2p_{25}r_{45}^{(g)} + PY_2p_{26}r_{46}^{(h)} + PY_2p_{27}r_{47}^{(i)} + PY_2p_{28}r_{48}^{(j)} + \\ + PY_3p_{34}^{(k)} + PY_3p_{35}r_{45}^{(l)} + PY_3p_{36}r_{46}^{(m)} + PY_3p_{37}r_{47}^{(n)} + PY_3p_{38}r_{48}^{(o)} \end{aligned}$$

em que

- a: efeito indireto primário de X_4 sobre Y , via X_1 ;
- b: efeito indireto secundário de X_4 sobre Y , via X_5 , via X_1 ;
- c: efeito indireto secundário de X_4 sobre Y , via X_6 , via X_1 ;
- d: efeito indireto secundário de X_4 sobre Y , via X_7 , via X_1 ;
- e: efeito indireto secundário de X_4 sobre Y , via X_8 , via X_1 ;
- f: efeito indireto primário de X_4 sobre Y , via X_2 ;
- g: efeito indireto secundário de X_4 sobre Y , via X_5 , via X_2 ;
- h: efeito indireto secundário de X_4 sobre Y , via X_6 , via X_2 ;
- i: efeito indireto secundário de X_4 sobre Y , via X_7 , via X_2 ;
- j: efeito indireto secundário de X_4 sobre Y , via X_8 , via X_2 ;
- k: efeito indireto primário de X_4 sobre Y , via X_3 ;
- l: efeito indireto secundário de X_4 sobre Y , via X_5 , via X_3 ;
- m: efeito indireto secundário de X_4 sobre Y , via X_6 , via X_3 ;
- n: efeito indireto secundário de X_4 sobre Y , via X_7 , via X_3 ;
- o: efeito indireto secundário de X_4 sobre Y , via X_8 , via X_3 .

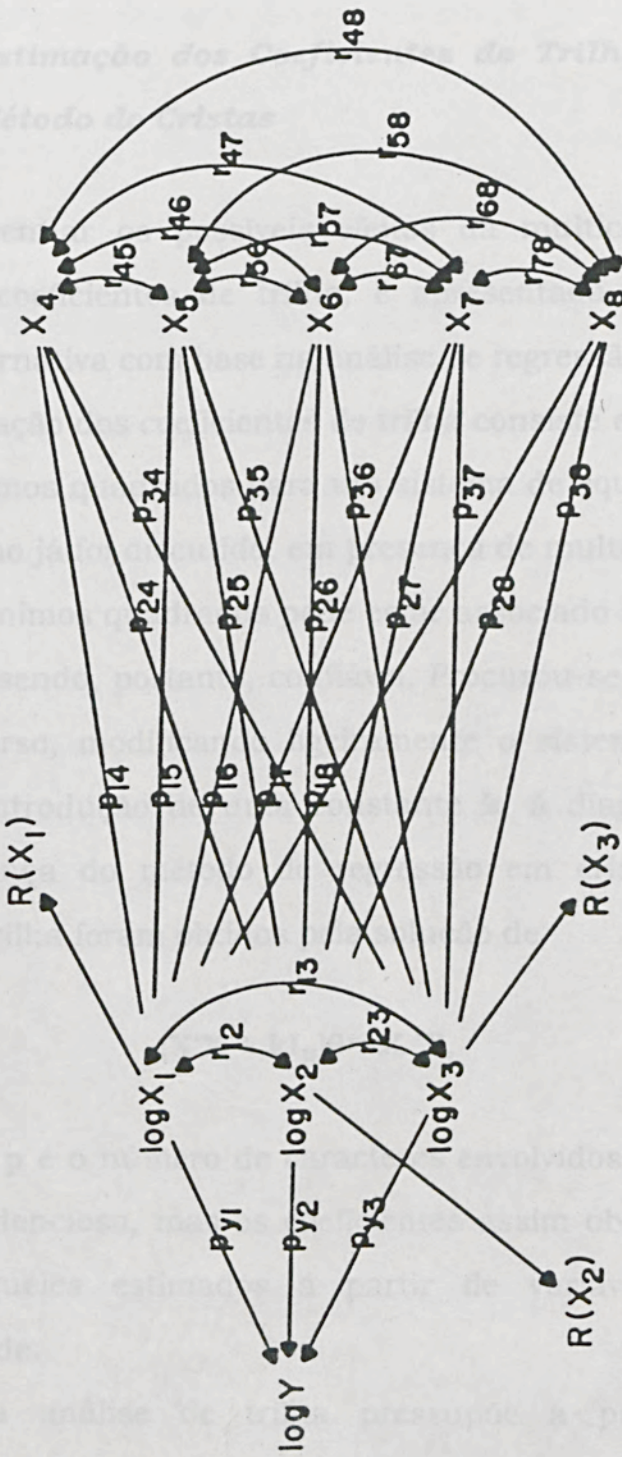


FIGURA 4 - Diagrama Mostrando as Relações Causa-efeito entre as Variáveis Primárias (X_1, X_2 e X_3), Variáveis Secundárias (X_4, X_5, X_6, X_7 e X_8) e a Variável Básica (Y), e Também as Variáveis Residuais Associadas.

De igual modo foram estimados os efeitos dos demais componentes secundários sobre a variável básica.

3.2.4.2. *Estimação dos Coeficientes de Trilha Utilizando o Método de Cristas*

Para atenuar os possíveis efeitos da multicolinearidade na estimação dos coeficientes de trilha, é apresentada, a seguir, uma metodologia alternativa com base na análise de regressão em cristas.

A estimação dos coeficientes de trilha consiste em se obter uma solução de mínimos quadrados para um sistema de equações na forma $(\mathbf{X}'\mathbf{X})\beta = \mathbf{X}'\mathbf{Y}$. Como já foi discutido, em presença de multicolinearidade, o estimador de mínimos quadrados pode estar associado a uma variância muito alta, não sendo, portanto, confiável. Procurou-se, então, atenuar esse efeito adverso, modificando ligeiramente o sistema de equações normais, pela introdução de uma constante \mathbf{k} , à diagonal da matriz $\mathbf{X}'\mathbf{X}$, à semelhança do método de regressão em cristas. Assim, os coeficientes de trilha foram obtidos pela solução de

$$(\mathbf{X}'\mathbf{X} + \mathbf{kI}_p)\theta^* = \mathbf{X}'\mathbf{Y},$$

em que $\mathbf{k} \geq 0$ e p é o número de caracteres envolvidos. Essa forma de estimação é tendenciosa, mas os coeficientes assim obtidos são muito semelhantes àqueles estimados a partir de variáveis isentas de multicolinearidade.

Como a análise de trilha pressupõe a padronização de variáveis, a matriz $\mathbf{X}'\mathbf{X}$ encontra-se na forma de correlações. Por isso, foram tomados valores de \mathbf{k} no intervalo, $0 \leq \mathbf{k} \leq 1$. Dentro desse intervalo, determinou-se o valor adequado para \mathbf{k} pelo exame de

gráficos de cristas, nos quais os coeficientes de trilha foram plotados em relação aos valores de k . Assim, foram escolhidos os menores valores de k , para os quais a maioria dos coeficientes já se encontravam estabilizados. Ressalta-se que, na literatura disponível, não foram encontradas referências sobre a aplicação dessa adaptação à análise de trilha.

a) Efeitos de componentes primários sobre a variável básica

a.1) Efeitos do logaritmo do número de vagens ($\log X_1$), logaritmo do número de sementes por vagem ($\log X_2$) e logaritmo do peso médio de 100 sementes ($\log X_3$), sobre logaritmo da produção de grãos $\log(Y)$

Considerando o mesmo diagrama apresentado na Figura 2, estimou-se novamente os coeficientes de trilha pelo método de cristas, solucionando-se o sistema $(X'X + kI_p)\theta^* = X'Y$. O valor adequado de k foi determinado pelo exame do gráfico de cristas.

Deve ser ressaltado que, pela metodologia proposta, a decomposição da correlação entre as variáveis explicativas (componentes primários) e a variável básica é dada por:

$$r_{y1} = (1+k)p_{y1} + p_{y2}r_{12} + p_{y3}r_{13}$$

$$r_{y2} = p_{y1}r_{12} + (1+k)p_{y2} + p_{y3}r_{23}$$

$$r_{y3} = p_{y1}r_{13} + p_{y2}r_{23} + (1+k)p_{y3}$$

Assim, tem-se:

$(1+k)p_{yj}$: efeito direto da variável j sobre a variável básica;

$p_{rj}r_{jk}$: efeito indireto da variável k , através da variável j ,

sobre a variável básica.

O fator de inflação da variância e covariância dos elementos de θ^* é dado por:

$$\omega = (X'X + kI)^{-1} (X'X) (X'X + kI)^{-1}$$

Assim, define-se:

$(1+k)^2 \omega_{jj}$: fator de inflação da variância do efeito direto da j-ésima variável;

$r_{jk}^2 \omega_{kk}$: fator de inflação da variância do efeito indireto da variável k através da variável j.

a.2) Efeitos do logaritmo do número de vagens ($\log X_1$), logaritmo do número de sementes por vagem ($\log X_2$), logaritmo do peso médio de 100 sementes ($\log X_3$) e logaritmo do número de sementes ($\log X_9$), sobre logaritmo da produção ($\log Y$)

Com o objetivo de verificar a eficiência do método de cristas para a análise de trilha, quanto à capacidade de atenuar os efeitos adversos da multicolinearidade, estimaram-se os efeitos diretos e indiretos dos componentes primários sobre a variável básica a partir do esquema causal apresentado na Figura 5. Nesse diagrama, está incluída a variável logaritmo do número de sementes ($\log X_9$), a qual é perfeitamente correlacionada com logaritmo do número de vagens ($\log X_1$) e logaritmo do número de sementes por vagem ($\log X_2$), uma vez que $\log X_9 = \log X_1 + \log X_2$. Utilizou-se o procedimento de análise em crista, determinando-se um valor adequado para k pelo exame do traço de crista. Assim, foi possível comparar os efeitos diretos das variáveis $\log X_1$, $\log X_2$ e $\log X_3$, sobre $\log Y$, obtidos pela estimação em cristas, com os efeitos diretos dessas mesmas variáveis, estimados pelo método

dos quadrados mínimos, quando em presença de multicolinearidade proporcionada pela inclusão de $\log X_9$ no diagrama de causa e efeito.

b) Efeitos dos componentes secundários sobre os primários

b.1) Efeitos dos componentes secundários sobre $\log X_1$, $\log X_2$ e $\log X_3$

Considerando o diagrama apresentado na Figura 3, foram estimados os efeitos diretos e indiretos das variáveis secundárias sobre as primárias, utilizando o método de cristas, como já descrito para outros esquemas causais.

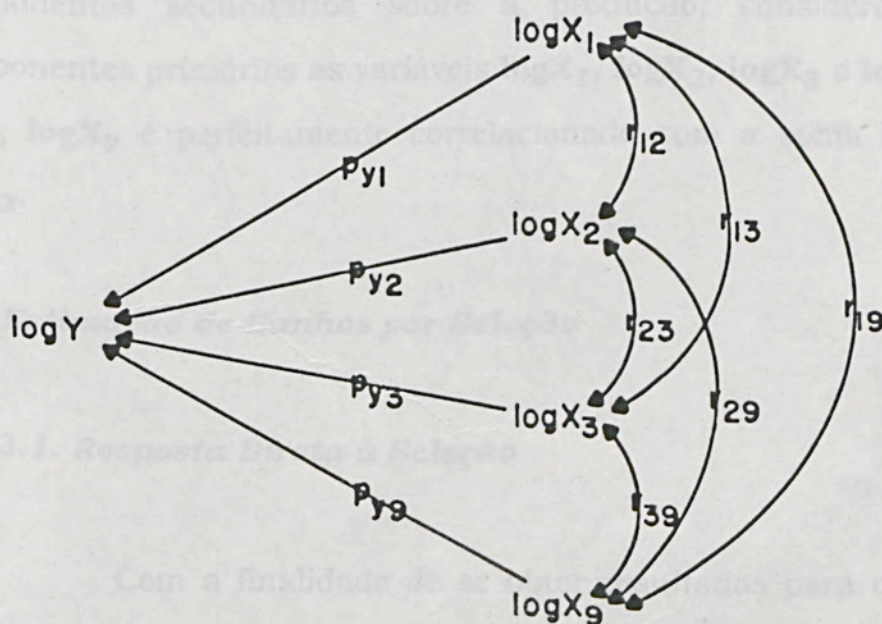


FIGURA 5 - Diagrama Apresentando as Relações Causa-efeito entre as Variáveis Primárias ($\log X_1$, $\log X_2$, $\log X_3$ e $\log X_9$) e a Variável Básica ($\log Y$).

b.2) Efeitos dos componentes secundários sobre os componentes primários, sob multicolinearidade

Na Figura 6, é apresentado o diagrama correspondente às relações causa-efeito entre as variáveis secundárias X_4 , X_5 , X_6 , X_7 e X_8 , sobre as variáveis primárias $\log X_1$, $\log X_2$, $\log X_3$ e $\log X_9$, sendo $\log X_9$ perfeitamente correlacionada com a soma de $\log X_1$ e $\log X_2$.

Os efeitos diretos e indiretos das variáveis secundárias sobre as primárias foram estimados de modo semelhante ao já descrito, sendo também o valor adequado de k estabelecido com base no traço de crista.

c) Efeitos dos componentes secundários sobre a variável básica

Tomando-se por base o diagrama apresentado na Figura 7, estimou-se, pelo método de cristas, os efeitos diretos e indiretos dos componentes secundários sobre a produção, considerando como componentes primários as variáveis $\log X_1$, $\log X_2$, $\log X_3$ e $\log X_9$. Como visto, $\log X_9$ é perfeitamente correlacionado com a soma de $\log X_1$ e $\log X_2$.

3.3. Estimação de Ganhos por Seleção

3.3.1. Resposta Direta à Seleção

Com a finalidade de se obter resultados para comparação de metodologias, considerou-se um esquema seletivo com seleção de

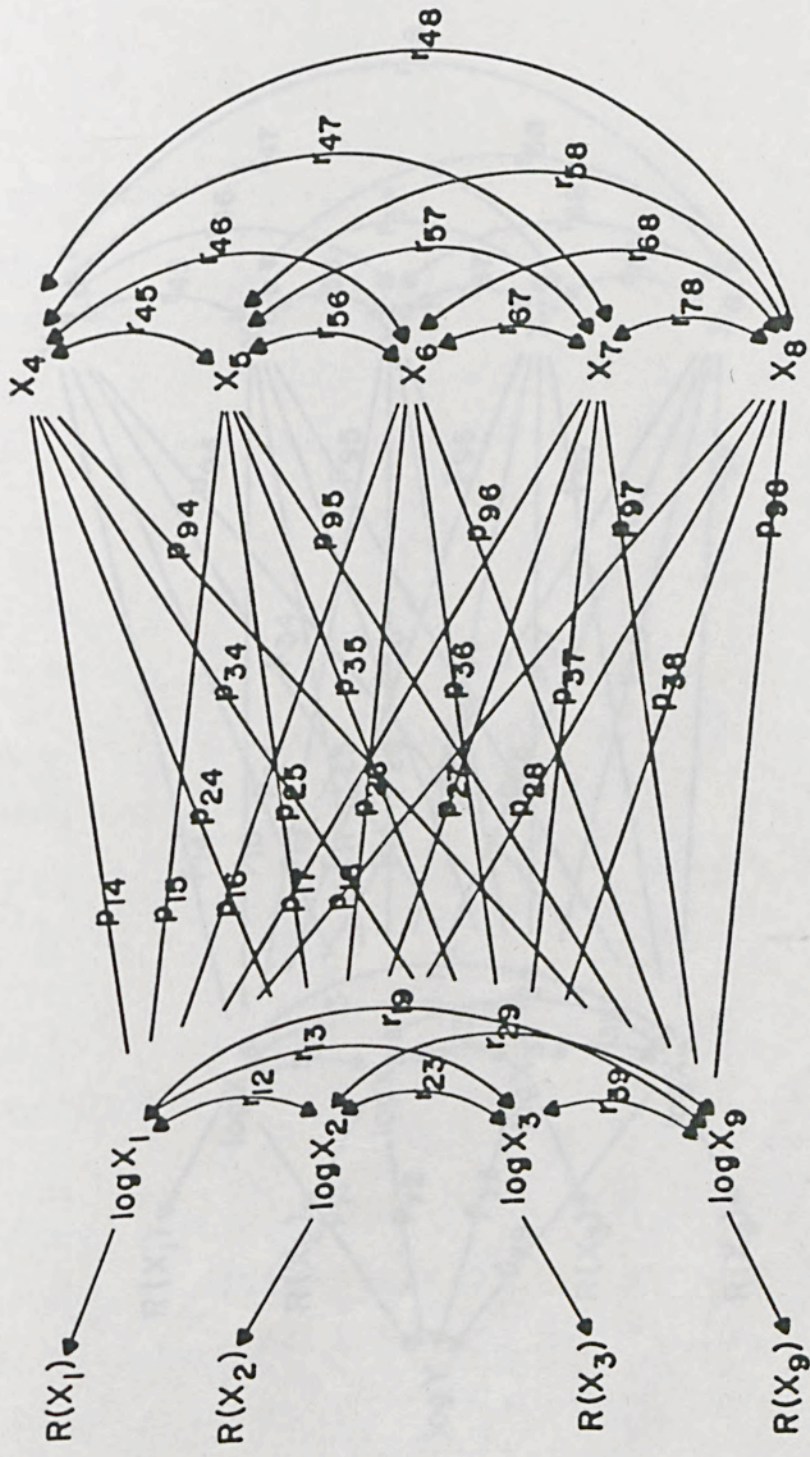


FIGURA 6 - Diagrama Apresentando as Relações Causa-efeito entre Variáveis Primárias e Variáveis Secundárias.

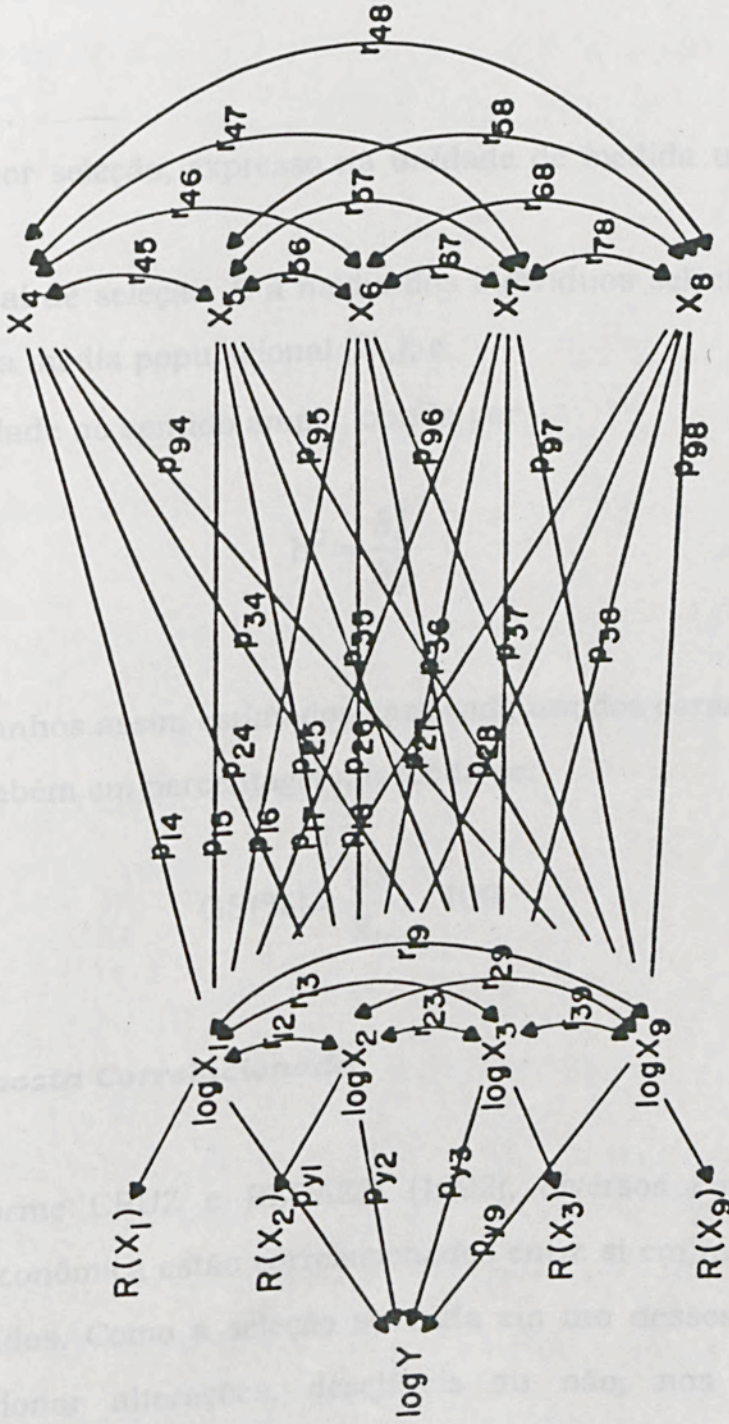


FIGURA 7 - Diagrama Apresentando as Relações Causa-efeito entre Variáveis Primárias, Variáveis Secundárias e a Variável Básica.

20% dos materiais em estudo e a recombinação das unidades selecionadas. Assim, o ganho foi estimado pela expressão

$$GS = DS \cdot h^2,$$

em que

GS: Ganho por seleção, expresso na unidade de medida usada para a variável;

DS: Diferencial de seleção. É a média dos indivíduos selecionados (\bar{x}_s) menos a média populacional (\bar{x}_0); e

h^2 : Herdabilidade no sentido amplo, obtida por

$$h^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_F^2}$$

Os ganhos assim estimados para cada um dos caracteres foram expressos também em percentagem, fazendo-se:

$$GS(\%) = \frac{GS}{\bar{x}_0} \cdot 100$$

3.3.2. Resposta Correlacionada

Conforme CRUZ e REGAZZI (1992), diversos caracteres de importância econômica estão correlacionados entre si em magnitudes e sentidos variados. Como a seleção aplicada em um desses caracteres pode proporcionar alterações, desejáveis ou não, nos demais, a quantificação dos efeitos indiretos da seleção aplicada a um ou mais caracteres torna-se essencial em um programa de melhoramento. O

conhecimento desses efeitos indiretos é útil também quando se deseja melhorar caracteres de difícil mensurabilidade.

A resposta esperada em um caráter **Y** em consequência de seleção aplicada no caráter **X** é definida como sendo o produto da resposta direta em **X** pelo coeficiente de regressão dos valores genéticos de **Y**, em função dos valores genéticos de **X**, isto é,

$$RY_{(x)} = \hat{\beta}RX$$

em que

RY_(x): Resposta indireta em **Y**, resultante da seleção em **X**;

RX: Resposta direta em **X**, expressa por

$$RX = DS_x p \frac{\hat{\sigma}_{Gx}^2}{\hat{\sigma}_{Fx}^2} = \frac{k_x p \hat{\sigma}_{Gx}^2}{\hat{\sigma}_{Fx}} = k_x p h_x \hat{\sigma}_{Gx} ;$$

p: Controle parental. Se as unidades recombinadas são as próprias unidades selecionadas, tem-se **p = 1**.

k: Diferencial de seleção em unidades de desvio-padrão fenotípico. Isto é,

$$k = \frac{DS}{\hat{\sigma}_F}$$

β̂: Coeficiente de regressão que mede a variação nos valores genéticos de **Y** quando ocorrem alterações de uma unidade nos valores genéticos de **X**. Tem-se que

$$\hat{\beta} = \frac{CÔV(X, Y)}{\hat{\sigma}_{Gx}^2} = \frac{r_G \hat{\sigma}_{Gy}}{\hat{\sigma}_{Gx}} ;$$

r_G: Coeficiente de correlação genotípica entre **X** e **Y**.

Tem-se então:

$$RY_{(x)} = k_x p h_x r_G \hat{\sigma}_{Gy} .$$

Sendo **RY** a resposta à seleção direta em **Y**, tem-se

$$RY = k_y p h_y \hat{\sigma}_{Gy} .$$

Assim, pode ser obtida a relação:

$$\frac{RY_{(x)}}{RY} = \frac{k_x h_x r_G}{k_y h_y} .$$

Para uma mesma intensidade de seleção aplicada a **X** e a **Y**, tem-se:

$$\frac{RY_{(x)}}{RY} = \frac{h_x r_G}{h_y} .$$

Logo, $RY_{(x)} > RY$, se $r_G h_x > h_y$.

Desse modo, conforme CRUZ e REGAZZI (1992), a resposta indireta proporciona ganhos superiores à resposta direta se forem satisfeitas as seguintes condições:

- a) o caráter **Y** apresentar baixa herdabilidade;
- b) o caráter **X** apresentar alta herdabilidade;
- c) ocorrer alta correlação genética entre os caracteres **X** e **Y**.

Empregando a metodologia descrita, foram então estimadas as respostas indiretas à seleção, considerando-se a seleção direta em cada um dos caracteres em estudo. Para isso, foram utilizados os coeficientes de correlação genética entre os caracteres, obtidos por meio de análises de variâncias e covariâncias, descritas no item 3.2.2.

3.3.3. Seleções Simultâneas com Base em Vários Caracteres

Como salientam CRUZ e REGAZZI (1992), o material selecionado deve reunir em si vários atributos favoráveis, simultaneamente. Assim, a possibilidade de êxito em um programa de melhoramento torna-se maior à medida que um maior número de caracteres de importância econômica é considerado no processo seletivo.

Dentre os processos propostos para seleção, com base em vários caracteres, destaca-se o método dos índices de seleção por possibilitar a combinação ideal das informações múltiplas contidas na unidade experimental.

No presente trabalho foi utilizado o índice clássico, proposto por SMITH (1936) e HAZEL (1943). Procurou-se, principalmente, avaliar possíveis efeitos adversos provocados pela multicolinearidade sobre o índice.

3.3.3.1. Estimação do Índice Clássico sem Considerar a Ocorrência de Multicolinearidade

Como proposto por SMITH (1936) e HAZEL (1943), o índice de seleção consiste em uma combinação linear dos caracteres de importância econômica, ponderados por coeficientes escolhidos de

modo a maximizar a correlação entre o índice (**I**) e o agregado genotípico (**H**), descritos a seguir. Este agregado consiste em uma combinação linear dos valores genéticos, desconhecidos, ponderados por seus respectivos valores econômicos. Tem-se então:

$$H = a_1g_1 + a_2g_2 + \dots + a_n g_n = \sum_{i=1}^n a_i g_i = g'a$$

$$I = b_1y_1 + b_2y_2 + \dots + b_n y_n = \sum_{i=1}^n b_i y_i = y'b$$

em que

n: número de caracteres avaliados;

g': vetor linha de dimensão **n**, de valores genéticos desconhecidos, dos **n** caracteres considerados;

y': vetor linha de dimensão **n**, de médias de um determinado genótipo;

a: vetor coluna de dimensão **n**, de pesos econômicos previamente estabelecidos;

b: vetor coluna, de dimensão **n** dos coeficientes de ponderação do índice a ser estimado.

Considera-se também:

P: matriz de covariâncias fenotípicas, de dimensões **n x n**;

G: matriz de covariâncias genéticas entre os caracteres, de dimensões **n x n**.

As seguintes estatísticas podem ser estabelecidas para o índice e o agregado genotípico:

i. Variância do índice:

$$\sigma_I^2 = b'Pb ;$$

ii. Variância do agregado genotípico:

$$\sigma_H^2 = a'Ga ;$$

iii. Covariância entre **I** e **H**:

$$\sigma_{IH} = \mathbf{b}' \mathbf{G} \mathbf{a} ;$$

iv. Correlação entre **I** e **H**:

$$\rho_{IH} = \rho = \frac{\mathbf{b}' \mathbf{G} \mathbf{a}}{\sqrt{(\mathbf{b}' \mathbf{P} \mathbf{b})(\mathbf{a}' \mathbf{G} \mathbf{a})}} .$$

O vetor **b** é estimado pela maximização da correlação entre **I** e **H**. Conforme CRUZ e REGAZZI (1992), o estimador **b** pode ser obtido mais facilmente pela maximização da função

$$\omega = \log_e(\rho_{IH}) = \log_e(\mathbf{b}' \mathbf{G} \mathbf{a}) - \frac{1}{2} [\log_e(\mathbf{b}' \mathbf{P} \mathbf{b}) + \log_e(\mathbf{a}' \mathbf{G} \mathbf{a})] .$$

Assim, é obtida a relação

$$\hat{\mathbf{P}} \hat{\mathbf{b}} = \mathbf{G} \mathbf{a} \frac{\hat{\mathbf{b}}' \hat{\mathbf{P}} \hat{\mathbf{b}}}{\hat{\mathbf{b}}' \mathbf{G} \mathbf{a}} .$$

Como o escalar $\frac{\hat{\mathbf{b}}' \hat{\mathbf{P}} \hat{\mathbf{b}}}{\hat{\mathbf{b}}' \mathbf{G} \mathbf{a}}$ pode ser desconsiderado sem afetar a proporcionalidade dos $\hat{\mathbf{b}}$'s, tem-se a expressão

$$\hat{\mathbf{b}} = \mathbf{P}^{-1} \mathbf{G} \mathbf{a} ,$$

a qual foi utilizada para estimação dos índices.

Considerou-se como peso econômico inicial para cada caráter o respectivo coeficiente de variação genética. Posteriormente, foram usados valores simulados como pesos econômicos, de modo a se ter combinações de ganhos possivelmente mais adequados ao melhoramento. Os índices, assim estabelecidos, foram posteriormente

comparados entre si, considerando o ganho proporcionado por cada um.

3.3.3.2. *Estimação do Índice Clássico Considerando a Ocorrência de Multicolinearidade*

Como visto anteriormente, à medida que a matriz $\mathbf{X}'\mathbf{X}$, ou matriz de variâncias e covariâncias fenotípicas, aproxima-se da condição de quase singularidade por efeito de multicolinearidade entre as variáveis envolvidas na estimação é de se esperar que o índice de seleção a ela correspondente torne-se menos confiável por efeito de um possível aumento da variância associada aos coeficientes de ponderação do índice (\mathbf{b} 's). Procurou-se, então, reduzir o grau de quase singularidade da matriz de variâncias e covariâncias fenotípicas pela adição de uma constante (\mathbf{k}) à diagonal dessa matriz. Para tornar possível a utilização de valores de \mathbf{k} , no intervalo $0 \leq \mathbf{k} \leq 1$, foram obtidas transformações das expressões correspondentes ao índice (\mathbf{I}) e ao agregado genotípico (\mathbf{H}). Essas transformações são apresentadas a seguir.

Sejam \mathbf{I} e \mathbf{H} , dados por

$$\begin{cases} \mathbf{I} = b_1 Y_1 + b_2 Y_2 + \dots + b_p Y_p & (1) \\ \mathbf{H} = a_1 G_1 + a_2 G_2 + \dots + a_p G_p & (2) \end{cases}$$

Então, pela maximização da correlação entre \mathbf{I} e \mathbf{H} , obtém-se

$$P\hat{\mathbf{b}} = G\mathbf{a}.$$

Sejam \mathbf{I}_p e \mathbf{H}_p , dados por

$$\begin{cases} \mathbf{I}_p = p_1 Y_1 + p_2 Y_2 + \dots + p_p Y_p & (3) \\ \mathbf{H}_p = q_1 g_1 + q_2 g_2 + \dots + q_p g_p & (4), \end{cases}$$

em que

$$y_j = \frac{Y_j}{\hat{\sigma}_{Fj}} ; \quad g_j = \frac{G_j}{\hat{\sigma}_{Gj}} ,$$

sendo $\hat{\sigma}_{Fj}$ e $\hat{\sigma}_{Gj}$ os desvios-padrão fenotípico e genotípico do caráter j.

De (1) obtém-se:

$$I = b_1 \sigma_{f1} \frac{Y_1}{\sigma_{f1}} + b_2 \sigma_{f2} \frac{Y_2}{\sigma_{f2}} + \dots + b_p \sigma_{fp} \frac{Y_p}{\sigma_{fp}}$$

$$I = p_1 Y_1 + p_2 Y_2 + \dots + p_p Y_p = p' y = I_p, \text{ sendo}$$

$$p_j = b_j \sigma_{fj},$$

ou, de forma matricial, tem-se

$$\underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \dots \\ p_p \end{bmatrix} = \begin{bmatrix} \sigma_{f1} & 0 & \dots & 0 \\ & \sigma_{f2} & \dots & 0 \\ & & \dots & 0 \\ & & & \sigma_{fp} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{bmatrix}$$

$$\underline{p} = D_F \underline{b},$$

em que D_F é a matriz diagonal cujos elementos não-nulos são os desvios-padrão fenotípicos dos caracteres estudados.

Da mesma forma, a partir de (2) obtém-se:

$$H = a_1 \sigma_{g1} \frac{G_1}{\sigma_{g1}} + a_2 \sigma_{g2} \frac{G_2}{\sigma_{g2}} + \dots + a_p \sigma_{gp} \frac{G_p}{\sigma_{gp}}$$

$$H = q_1 g_1 + q_2 g_2 + \dots + q_p g_p = q' g = H_p$$

Tem-se que:

$$q_j = a_j \sigma_{g_j}$$

ou, na forma matricial:

$$\underline{q} = \begin{bmatrix} q_1 \\ q_2 \\ \dots \\ q_p \end{bmatrix} = \begin{bmatrix} \sigma_{g1} & 0 & \dots & 0 \\ & \sigma_{g2} & \dots & 0 \\ & & \dots & 0 \\ & & & \sigma_{gp} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix}$$

$$\underline{q} = D_G a$$

Nesta expressão, D_G é a matriz diagonal, cujos elementos não-nulos são os desvios-padrão genotípicos dos caracteres estudados.

Verificam-se, também, as seguintes relações:

$$a) V(I_p) = V(p'y) = p'V(y)p = p'R_F p$$

$$b) V(H_p) = V(q'g) = q'V(g)q = q'R_G q$$

$$c) \text{Cov}(I_p, H_p) = p' D_F^{-1} G D_G^{-1} q$$

$$d) r_{HI} = \frac{\text{Cov}(I_p, H_p)}{\sqrt{V(I) V(H)}} = \frac{p' D_F^{-1} G D_G^{-1} q}{\sqrt{(p'R_F p) (q'R_G q)}}$$

Estima-se o índice de seleção pela maximização de r_{HI} , obtendo-se a expressão:

$$R_F \hat{p} = D_F^{-1} G D_G^{-1} q \cdot \frac{\hat{p}' R_F q}{\hat{p}' D_F^{-1} G D_G^{-1} q}$$

Para efeitos práticos, a constante $\hat{p}' R_F q / \hat{p}' D_F^{-1} G D_G^{-1} q$ não afeta a proporcionalidade do índice. Assim, tem-se:

$$R_F \hat{p} = D_F^{-1} G D_G^{-1} q$$

$$\hat{p} = R_F^{-1} D_F^{-1} G D_G^{-1} q$$

Se houver multicolinearidade perfeita entre as variáveis estudadas a matriz R_F será singular, e o processo de estimação não proporcionará valores confiáveis para o vetor de coeficientes do índice de seleção. Uma alternativa para a obtenção desse vetor de coeficientes é por meio da expressão

$$(R_F + kI)\hat{p} = D_F^{-1} G D_G^{-1} q.$$

O vetor \hat{p} assim estimado proporciona uma correlação entre o índice (I) e o agregado genotípico (H) inferior ao obtido pelo processo clássico proposto por SMITH e HAZEL. Entretanto, para alguns valores de k , é possível obter estimativas adequadas de \hat{p} , com níveis satisfatórios de correlação entre I e H.

Sendo $M = D_F^{-1} G D_G^{-1}$,

tem-se

$$r_{HI} = \frac{p' M q}{\sqrt{(p' R_F p) (q' R_G q)}}.$$

Sob colinearidade, tem-se

$$r_{HI}^* = \frac{p^* M q}{\sqrt{(p^* R_F p^*) (q' R_G q)}},$$

em que

$$r_{HI}^* \leq r_{HI}.$$

A determinação dos valores adequados de k foi feita pelo exame de gráficos de cristas, no qual os coeficientes de ponderação (p 's) foram plotados em relação aos valores de k . Considerou-se apropriado o valor para o qual a maioria dos coeficientes se estabilizaram.

4. RESULTADOS E DISCUSSÃO

4.1. Análise de Variância

Os dados obtidos em cada um dos experimentos foram submetidos à análise de variância separadamente. Foram analisadas as observações correspondentes à número de dias até o florescimento (FLR), número de dias até a maturação (MAT), peso da palhada (PPL), estande final (STO), índice de colheita (INC), número de sementes (NSM), número de vagens (NVO), número de sementes por vagem (NSV), peso médio de 100 sementes (PCS) e produção de grãos (PRD). Optou-se pela prévia transformação logarítmica das observações correspondentes a NSM, NVO, NSV, PCS, e PRD, visando sua utilização posterior em análise de tripla em que se prevê uma relação linear e aditiva entre a produção de grãos e seus componentes primários. Assim, obteve-se, respectivamente, LNSM, LNVO, LNSV, LPCS e LPRD. Os resultados da análise de variância de cada caráter quanto aos quadrados médios e

coeficientes de variação, para cada experimento, são apresentados no Quadro 4.

Ocorreram diferenças significativas entre as populações em relação a todos os caracteres estudados. Observa-se, também, que houve boa precisão experimental, conforme critério de classificação de GOMES (1987), pois, a maioria dos caracteres apresentaram coeficiente de variação experimental (CV_e) inferior a 20% nos dois experimentos.

4. RESULTADOS E DISCUSSÃO

4.1. Análise da Variância

Os dados obtidos em cada um dos experimentos foram submetidos à análise da variância separadamente. Foram analisadas as observações correspondentes a número de dias até o florescimento (FLR), número de dias até a maturação (MAT), peso da palhada (PPL), estande final (STD), índice de colheita (INC), número de sementes (NSM), número de vagens (NVG), número de sementes por vagem (NSV), peso médio de 100 sementes (PCS) e produção de grãos (PRD). Optou-se pela prévia transformação logarítmica das observações correspondentes a NSM, NVG, NSV, PCS, e PRD, visando sua utilização posterior em análise de trilha em que se prevê uma relação linear e aditiva entre a produção de grãos e seus componentes primários. Assim, obteve-se, respectivamente, LNSM, LNVG, LNSV, LPCS e LPRD. Os resultados da análise de variância de cada caráter quanto aos quadrados médios e

coeficientes de variação, para cada experimento, são apresentados no Quadro 4.

Ocorreram diferenças significativas entre as populações em relação a todos os caracteres estudados. Observa-se, também, que houve boa precisão experimental, conforme critério de classificação de GOMES (1987), pois, a maioria dos caracteres apresentaram coeficiente de variação experimental (CV_e) inferior a 20% nos dois experimentos.

4.2. Estimativas de Parâmetros Genéticos

Com base nos resultados das análise de variância, foram obtidas, para cada um dos caracteres em estudo, as estimativas da variância genética ($\hat{\sigma}_G^2$), do coeficiente de herdabilidade no sentido amplo em nível de parcela experimental (h^2) e do coeficiente de variação genética (CV_G). Essas estimativas são apresentadas no Quadro 5, juntamente com a estimativa da razão entre o coeficiente de variação genética (CV_G) e o coeficiente de variação experimental (CV_e).

Observa-se que alguns caracteres apresentam os coeficientes de herdabilidade em nível de parcela experimental (h^2) relativamente mais elevados e também apresentam valores mais altos para a proporção CV_G/CV_e . Esta associação constitui situação favorável para a seleção, como ressalta VENCOVSKY (1987). Pode-se verificar no Quadro 5 que, para o Experimento 1, os caracteres peso médio de 100 sementes (LPCS), número de sementes por vagem (LNSV), número de sementes (LNSM), índice de colheita (INC), número de vagens (LNVG), produção de grãos (LPRD) e número de dias para a floração (FLR) podem ser considerados como os mais favoráveis para a seleção na cultura do feijão, por apresentarem h^2 mais elevado, acima de 60%, e

QUADRO 4 - Resumo das Análises de Variância e Estimativas dos Coeficientes de Variação Experimental (CV_e) e das Médias, Relativas a 10 Caracteres de Feijão, para o Experimento 1 (Exp. 1) e para o Experimento 2 (Exp. 2). Viçosa, MG, 1991/92

FV	GL	FLR	MAT	PPL	STD	INC	LMCS	LNVG	LNSV	LPCS	LPRD
Tratamentos											
Exp. 1	19	26,2105**	104,7500**	11,3202**	756,1414**	0,0321**	0,3169**	0,1880**	0,0221**	0,0590**	0,1512**
Exp. 2	18	71,9262**	443,4514**	18,5632**	733,7917**	0,0961**	0,4059**	0,2474**	0,0419**	0,0248**	0,2913**
Blocos											
Exp. 1	3	4,2114	7,9792	1,0317	750,0625	0,0078	0,0153	0,0198	0,0007	0,0013	0,0223
Exp. 2	3	75,3687	9,7039	4,8965	65,5658	0,0104	0,0296	0,0272	0,0066	0,0090	0,0344
Resíduo											
Exp. 1	57	3,2095	15,7478	3,0289	223,9879	0,0019	0,0159	0,0132	0,0010	0,0006	0,0172
Exp. 2	54	12,6517	20,9331	2,5388	36,6075	0,0064	0,0157	0,0162	0,0059	0,0181	0,0295
Médias											
Exp. 1		47,8409	89,3185	6,5544	82,3500	0,5452	1,4913	0,8602	0,6311	1,3769	0,8682
Exp. 2		49,2692	94,1289	7,8570	77,2368	0,4804	1,4497	0,8890	0,5606	1,3726	0,8224
CV_e (%)											
Exp. 1		3,7447	4,4429	26,6339	18,1739	8,0953	8,4595	13,3431	5,0143	1,7816	15,1220
Exp. 2		7,2194	4,8666	20,2795	7,8336	16,7010	8,6385	14,3377	13,6491	9,8123	20,8806

**Significativo em nível de 1% de probabilidade, pelo teste F.

QUADRO 5 - Estimativas da Variância Genética ($\hat{\sigma}_G^2$), Coeficiente de Herdabilidade no Sentido Amplo em Nível de Parcela Experimental Expresso em Percentagem ($h^2\%$), Coeficiente de Variação Genético (CV_G), Coeficiente de Variação Experimental (CV_e) e Razão CV_G/CV_e , Obtidos para os Experimentos 1 e 2, Relativos a 10 Caracteres de Feijão. Viçosa, MG, 1991/92

Caráter	$\hat{\sigma}_G^2$		$(h^2\%)$		CV_G		CV_e		CV_G/CV_e	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2
	FLR	5,750	139,819	64,178	91,702	5,012	24,000	3,745	7,219	1,338
MAT	22,250	105,629	58,556	83,460	5,281	10,919	4,443	4,861	1,189	2,246
PPL	2,073	4,006	40,630	61,210	22,033	25,474	26,634	20,279	0,827	1,256
STD	133,038	74,296	37,263	82,642	14,006	17,093	18,174	7,834	0,771	2,182
INC	0,007	0,022	79,494	77,682	15,939	31,158	8,095	16,701	1,969	1,866
LNSM	0,075	0,097	82,543	86,149	18,395	21,544	8,459	8,639	2,174	2,494
LNVG	0,044	0,058	76,841	78,057	24,305	27,042	13,343	14,338	1,821	1,886
LNSV	0,005	0,009	84,039	60,626	11,506	16,937	5,014	13,649	2,295	1,241
LPCS	0,015	0,002	96,040	8,365	8,773	2,965	1,782	9,812	4,923	0,302
LPRD	0,033	0,065	66,026	68,945	21,081	31,112	15,122	20,881	1,394	1,490

por apresentarem também a razão CV_G/CV_e acima da unidade. O caráter LPCS, quando avaliado no Experimento 1, destacou-se por apresentar $CV_G/CV_e=4,923$ e $h^2=96,040\%$. Esses valores foram os mais altos observados. Contudo, essa performance foi conseqüência da alta precisão experimental com que o caráter foi avaliado, pois na respectiva análise da variância observou-se $CV_e=1,782\%$.

Para o Experimento 2, os caracteres apresentaram a razão CV_G/CV_e sempre superior à unidade e h^2 superior a 60%, à exceção de LPCS. Observa-se que, sob boa precisão experimental, a amostra em estudo mostrou-se eficiente para discriminar genótipos favoráveis.

4.3. Correlações Genotípicas, Fenotípicas e de Ambiente

As estimativas das correlações genotípicas, fenotípicas e de ambiente, entre os caracteres em estudo, para o Experimento 1 e para o Experimento 2, são apresentadas nos Quadros 6 e 7, respectivamente.

Observa-se que, em geral, houve boa concordância de sinais nas correlações fenotípicas e genotípicas, para os dois experimentos. Quanto às magnitudes, houve ligeira tendência de as correlações genotípicas superarem as fenotípicas. Assim, pode-se deduzir que os componentes genotípicos têm maior influência na determinação das correlações que os componentes de ambiente. Pode-se observar também que, em alguns casos, as correlações de ambiente diferiram das genotípicas em magnitude e sinal. Essa constatação constitui indícios de que a seleção indireta, com base nos coeficientes de correlação genotípica, pode ser prejudicada pela ação diferencial do ambiente sobre as variáveis envolvidas, favorecendo uma em detrimento da outra.

QUADRO 6 - Estimativas dos Coeficientes de Correlação Genotípica (r_G), Fenotípica (r_F) e de Ambiente (r_A), entre os Caracteres em Estudo, com Amostras de Feijão. Experimento 1. Viçosa, MG, 1991/92

	I	FLR	MAT	PPL	STD	INC	LNSM	LNVG	LNSV	LPCS	LPRD
FLR	G		0,1895	-0,3070	-0,6309	-0,1069	-0,0425	-0,1318	0,2232	-0,2990	-0,2592
	F		0,2318	-0,2488	-0,5855	-0,0919	-0,0474	-0,1300	0,2022	-0,2731	-0,2379
	A		0,5027	-0,0146	-0,4710	0,0608	-0,1104	-0,1175	-0,0282	0,1566	-0,0794
MAT	G			0,0358	0,1871	-0,8968	-0,7218	-0,7276	-0,6293	0,6275	-0,6654
	F			0,0800	0,0689	-0,8252	-0,6507	-0,6498	-0,5662	0,5857	-0,5748
	A			0,2583	-0,3590	-0,2516	-0,0255	-0,0296	0,0065	0,2605	0,0188
PPL	G				0,6099	0,0003	0,4858	0,5044	0,3827	-0,2261	0,5792
	F				0,4773	-0,0393	0,4938	0,5201	0,3527	-0,1931	0,5945
	A				0,1400	-0,3105	0,7643	0,7581	0,2975	-0,0109	0,7323
STD	G					-0,0625	-0,0062	0,0415	-0,1441	0,3996	0,2553
	F					-0,0603	0,0247	0,0647	-0,0953	0,3217	0,2372
	A					-0,0710	0,2446	0,2159	0,1974	-0,2150	0,1939
INC	G						0,8068	0,8126	0,7085	-0,6002	0,8130
	F						0,7770	0,7769	0,6764	-0,5672	0,7705
	A						0,2700	0,2678	0,1048	0,4655	0,3470

Continua...

QUADRO 7 - Estimativas dos Coeficientes de Correlação Genotípica (r_G), Fenotípica (r_F) e de Ambiente (r_A), entre os Caracteres em Estudo, com Amostras de Feijão, para o Experimento 2.Viçosa, MG, 1991/92

	r	FLR	MAT	PPL	STD	INC	LNSM	LNVG	LNSV	LPCS	LPRD
FLR	G		0,9595	0,2224	-0,7342	-0,8875	-0,8929	-0,8655	-0,7460	1,0370	-0,9247
	F		0,9282	0,1806	-0,7038	-0,8560	-0,8808	-0,8463	-0,6849	0,5123	-0,8900
	A		0,0618	-0,4310	0,1150	-0,2157	-0,5170	-0,4981	-0,0110	-0,1414	-0,4871
MAT	G			0,2558	-0,5795	-0,9588	-0,8998	-0,8275	-0,8666	0,8236	-0,9669
	F			0,2377	-0,5530	-0,9037	-0,8775	-0,7916	-0,8082	0,4772	-0,8964
	A			0,0718	-0,0326	0,0054	-0,3816	-0,1955	-0,2901	0,3302	-0,0233
PPL	G				0,1160	-0,3994	-0,0534	0,0488	-0,2990	0,2893	-0,0189
	F				0,0907	-0,3879	-0,0073	0,1080	-0,2849	0,0784	0,0145
	A				-0,1743	-0,3079	0,5697	0,6771	-0,1973	-0,1915	0,2651
STD	G					0,5715	0,7146	0,8025	0,3196	-0,9667	0,7186
	F					0,5322	0,6730	0,7461	0,2820	-0,4908	0,6512
	A					-0,1016	-0,2254	-0,1746	-0,0836	-0,0181	-0,1812
INC	G						0,9364	0,8764	0,8611	-1,3882	0,9222
	F						0,9002	0,8172	0,8154	-0,5198	0,9108
	A						0,2619	-0,0157	0,4542	0,7836	0,8058

Continua...

Constituem exemplos dessa situação, as correlações entre logaritmo do peso médio de 100 sementes (LPCS) e índice de colheita (INC), dentre outros. SANTOS e VENCOVSKY (1986) e CASTOLDI (1991) encontraram correlações negativas entre o número de sementes por vagem e peso médio de 100 sementes. Isso foi também constatado no presente trabalho.

Dentre os caracteres mais correlacionados com a produção de grãos (PRD), destacaram-se alguns de seus componentes primários como o número de sementes (NSM), número de vagens (NVG) e número de sementes por vagem (NSV). Outro caráter bem correlacionado com PRD foi o índice de colheita (INC). ABRAHÃO (1960) verificou que no feijoeiro esse índice está correlacionado positivamente com a produtividade de grãos. SHARMA et al. (1987) afirmam que esta medida da proporção entre peso de grãos e peso da palhada somado ao peso de grãos é pouco sensível às flutuações de ambiente. Assim, esse caráter é eficiente para discriminar genótipos.

VIEIRA e GOMES (1969) constataram associação entre o peso da palhada e a produção de grãos, no feijoeiro. Contudo, no presente trabalho, essa correlação não se mostrou consistente nos dois experimentos.

O estudo das correlações apresenta algumas limitações, apesar de proporcionar muitas informações úteis ao melhoramento. Assim, é que no estudo de um caráter de natureza complexa como a produção de grãos, a qual é determinada em geral pelo inter-relacionamento de vários fatores, as estimativas dos coeficientes de correlação não fornecem informações precisas sobre as influências diretas e indiretas de cada componente primário sobre esta variável, como ressalta ALMEIDA

(1991). Para um entendimento mais aprofundado desses inter-relacionamentos, as variáveis foram submetidas também à análise trilha.

4.4. Análise de Trilha

O estudo dos coeficientes de trilha mostra-se eficiente ao revelar a verdadeira natureza das inter-relações de causa e efeito entre a produção de grãos e seus componentes primários. Consiste rotineiramente no desdobramento das correlações genótípicas entre essas variáveis em efeitos diretos e indiretos. No presente estudo, optou-se por trabalhar com correlações fenotípicas, pois, pretende-se analisar efeitos de multicolinearidade intensa para evidenciar seus efeitos. Para se conseguir multicolinearidade intensa, torna-se necessário o estabelecimento de combinações quase perfeitas entre as variáveis. Estas combinações são mais facilmente obtidas a partir dos efeitos fenotípicos. Sem perda de generalização, os resultados aqui obtidos certamente poderão ser aplicados às análises de trilha rotineiras, com base nas correlações genótípicas entre as variáveis.

Tomando-se por base os diagramas causais estabelecidos previamente, foi feita inicialmente a análise de trilha sem levar em conta os possíveis efeitos adversos da multicolinearidade. Em seguida, procurou-se atenuar esses efeitos, estimando-se os coeficientes de trilha por procedimento análogo ao método de regressão em cristas ou em cumeeira. Os resultados assim obtidos foram depois comparados entre si.

4.4.1. Análise de Trilha sem Considerar a Possibilidade de Ocorrência de Multicolinearidade entre as Observações

4.4.1.1. Efeitos das Variáveis Primárias sobre a Variável Básica

Com base no esquema causal mostrado na Figura 2, foram estimados os efeitos diretos e indiretos das variáveis primárias sobre a produção de grãos para os Experimentos 1 e 2. Esses resultados são apresentados no Quadro 8. São também apresentados os fatores de inflação da variância (VIF's) associados a cada efeito e calculados conforme MARQUARDT (1970). Esses VIF's foram posteriormente comparados com aqueles associados aos efeitos diretos e indiretos, estimados de modo a atenuar os efeitos da multicolinearidade. Observa-se que, em ambos os experimentos, as variáveis explicaram totalmente a variação da variável básica, como era de se esperar. Os efeitos diretos das três variáveis primárias sobre a variável básica foram positivos, apesar de ter sido negativa a correlação entre a variável explicativa primária (logaritmo do peso médio de 100 sementes - LPCS) e a variável básica (logaritmo do peso de grãos - LPRD). Logaritmo do número de vagens (LNVG) é o caráter mais correlacionado com a produção e também com maior efeito direto sobre esta. Verifica-se, assim, que a seleção indireta para maior produção de grãos é mais viável por meio do número médio de vagens por planta do que por meio de qualquer outro componente primário, o que está em harmonia com os resultados obtidos por CASTOLDI (1991), ao verificar que a seleção indireta para

QUADRO 8 - Efeitos Diretos e Indiretos das Variáveis Primárias sobre a Variável Básica e Fatores de Inflação da Variância (VIF's) para os Experimento 1 e 2

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNVG	Direto sobre LPRD	1,1138	3,2829	0,9215	2,0087
	Indireto via LNSV	0,3184	3,8687	0,2171	0,5261
	" LPCS	-0,4800	2,5363	-0,1937	0,8522
	Total	0,9522		0,9449	
LNSV	Direto sobre LPRD	0,3842	5,6320	0,3796	1,6085
	Indireto via LNVG	0,9231	2,2551	0,5270	0,6570
	" LPCS	-0,5452	3,7711	-0,1602	0,5829
	Total	0,7621		0,7464	
LPCS	Direto sobre LPRD	0,6250	4,2990	0,2919	1,9346
	Indireto via LNVG	-0,8554	1,9363	-0,6116	0,8846
	" LNSV	-0,3351	4,2854	-0,2083	0,4846
	Total	-0,5655		-0,5281	
Coeficiente de determinação		0,9999		0,9999	

LNVG: logaritmo do número médio de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes;

LPRD: logaritmo da produção de grãos.

produção de grãos no feijoeiro pode ser feita através do número de vagens por parcela.

4.4.1.2. Efeitos das Variáveis Secundárias sobre as Variáveis Primárias

Nos Quadros 9 e 10, em relação aos experimentos 1 e 2 respectivamente, são apresentados os valores estimados para os efeitos diretos e indiretos das variáveis explicativas secundárias sobre as variáveis explicativas primárias, com base no esquema causal mostrado na Figura 3.

Observa-se que peso da palhada (PPL) e índice de colheita (INC) apresentaram efeitos diretos positivos e relativamente altos sobre logaritmo do número de vagens (LNVG) para os dois experimentos. INC foi também a variável mais consistentemente correlacionada com LNVG. Número de dias até a maturação (MAT) apresentou alto efeito indireto negativo sobre LNVG via INC. Número de dias até o florescimento (FLR), juntamente com INC e MAT, apresentaram os maiores efeitos diretos sobre logaritmo do número de sementes (LNSV). Dentre esses, o efeito de MAT foi negativo e os demais positivos. MAT apresentou, também, correlação negativa e relativamente alta com LNSV, enquanto INC apresentou a maior correlação positiva com essa variável primária. Estande final (STD) demonstrou alguma influência sobre LNSV apenas indiretamente, via FLR, sendo esse efeito negativo.

O caráter primário que apresentou o maior coeficiente de correlação com a produção de grãos (PRD), e também o maior efeito

QUADRO 9 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variáveis Primárias, e Estimativas dos Fatores de Inflação da Variância (VIF's) para o Experimento 1

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNVG	0,0132	1,8083
	Indireto via MAT	-0,0277	0,1963
	" " PPL	-0,1617	0,0807
	" " STD	0,1098	0,0648
	" " INC	-0,0638	0,0280
	Total	-0,1297	
MAT	Direto sobre LNVG	-0,1194	3,6527
	Indireto via FLR	0,0031	0,0972
	" " PPL	0,0520	0,0083
	" " STD	-0,0129	0,0093
	" " INC	-0,5725	2,2552
	Total	-0,6497	
PPL	Direto sobre LNVG	0,6498	1,3033
	Indireto via FLR	-0,0033	0,1119
	" " MAT	-0,0095	0,0234
	" " STD	-0,0895	0,4485
	" " INC	-0,0273	0,0051
	Total	0,5202	
STD	Direto sobre LNVG	-0,1876	1,9685
	Indireto via FLR	-0,0077	0,6199
	" " MAT	-0,0082	0,0173
	" " PPL	0,3101	0,2969
	" " INC	-0,0418	0,0120
	Total	0,0648	
INC	Direto sobre LNVG	0,6938	3,3118
	Indireto via FLR	-0,0012	0,0153
	" " MAT	0,0985	2,4873
	" " PPL	-0,0255	0,0020
	" " STD	0,0113	0,0072
	Total	0,7769	
Coeficiente de determinação		0,9407	

Continua...

QUADRO 9, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSV	0,4175	1,8083
	Indireto via MAT	-0,0705	0,1963
	" " PPL	-0,1304	0,0807
	" " STD	0,0299	0,6748
	" " INC	-0,0442	0,0280
	Total	0,2023	
MAT	Direto sobre LNSV	-0,3044	3,6527
	Indireto via FLR	0,0967	0,0972
	" " PPL	0,0419	0,0083
	" " STD	-0,0035	0,0093
	" " INC	-0,3970	2,2552
	Total	-0,5663	
PPL	Direto sobre LNSV	0,5242	1,3033
	Indireto via FLR	-0,1039	0,1119
	" " MAT	-0,0243	0,0234
	" " STD	-0,0244	0,4485
	" " INC	-0,0189	0,0051
	Total	0,3527	
STD	Direto sobre LNSV	-0,0511	1,9685
	Indireto via FLR	-0,2444	0,6199
	" " MAT	-0,0210	0,0173
	" " PPL	0,2502	0,2969
	" " INC	-0,0290	0,0120
	Total	-0,0953	
INC	Direto sobre LNSV	0,4811	3,3118
	Indireto via FLR	-0,0384	0,0153
	" " MAT	0,2512	2,4873
	" " PPL	-0,0206	0,0020
	" " STD	0,0031	0,0072
	Total	0,6764	
Coeficiente de determinação		0,7719	

Continua...

QUADRO 9, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LPCS	-0,3783	1,8083
	Indireto via MAT	0,1408	0,1963
	" " PPL	0,1175	0,0807
	" " STD	-0,1625	0,6748
	" " INC	0,0094	0,0280
	Total	-0,2731	
MAT	Direto sobre LPCS	0,6075	3,6527
	Indireto via FLR	-0,0877	0,0772
	" " PPL	-0,0378	0,0083
	" " STD	0,0191	0,0093
	" " INC	0,0846	2,2552
	Total	0,5857	
PPL	Direto sobre LPCS	-0,4723	1,3033
	Indireto via FLR	0,0941	0,1119
	" " MAT	0,0486	0,0234
	" " STD	0,1325	0,4485
	" " INC	0,0040	0,0051
	Total	-0,1931	
STD	Direto sobre LPCS	0,2776	1,9685
	Indireto via FLR	0,2215	0,6199
	" " MAT	0,0419	0,0173
	" " PPL	-0,2255	0,2969
	" " INC	0,0062	0,0120
	Total	0,3217	
INC	Direto sobre LPCS	-0,1025	3,3118
	Indireto via FLR	0,0348	0,0153
	" " MAT	-0,5013	2,4873
	" " PPL	0,0186	0,0020
	" " STD	-0,0167	0,0072
	Total	-0,5671	
Coeficiente de determinação		0,6978	

Continua...

QUADRO 10 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variáveis Primárias, e Estimativas dos Fatores de Inflação da Variância (VIF's) para o Experimento 2

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNVG	-0,3292	11,6880
	Indireto via MAT	0,2945	10,7616
	" " PPL	0,0762	0,0476
	" " STD	-0,1247	1,2674
	" " INC	-0,7631	5,1223
	Total	-0,8463	
MAT	Direto sobre LNVG	0,3173	12,4909
	Indireto via FLR	-0,3056	10,0699
	" " PPL	0,1002	0,0824
	" " STD	-0,0980	0,7825
	" " INC	-0,8056	5,7091
	Total	-0,7917	
PPL	Direto sobre LNVG	0,4217	1,4590
	Indireto via FLR	-0,0594	0,3812
	" " MAT	0,0754	0,7057
	" " STD	0,0161	0,0211
	" " INC	-0,3458	1,0519
	Total	0,1080	
STD	Direto sobre LNVG	0,1772	2,5587
	Indireto via FLR	0,2317	5,7895
	" " MAT	-0,1755	3,8198
	" " PPL	0,0382	0,0120
	" " INC	0,4744	1,9800
	Total	0,7460	
INC	Direto sobre LNVG	0,8914	6,9907
	Indireto via FLR	0,2818	8,5642
	" " MAT	-0,2867	10,2001
	" " PPL	-0,1636	0,2195
	" " STD	0,0943	0,7247
	Total	0,8172	
Coeficiente de determinação		0,9337	

Continua...

QUADRO 10, Cont.

Variável	Efeito	Estimativa	VIF .
FLR	Direto sobre LNSV	0,2718	11,6880
	Indireto via MAT	-0,6103	10,7616
	" " PPL	0,0132	0,0476
	" " STD	0,1508	1,2674
	" " INC	-0,5103	5,1223
	Total	-0,6848	
MAT	Direto sobre LNSV	-0,6575	12,4909
	Indireto via FLR	0,2522	10,0699
	" " PPL	0,0173	0,0824
	" " STD	0,1185	0,7825
	" " INC	-0,5387	5,7091
	Total	-0,8082	
PPL	Direto sobre LNSV	0,0730	1,4590
	Indireto via FLR	0,0491	0,3812
	" " MAT	-0,1563	0,7057
	" " STD	-0,0194	0,0211
	" " INC	-0,2312	2,0519
	Total	-0,2848	
STD	Direto sobre LNSV	-0,2142	2,5587
	Indireto via FLR	-0,1913	5,7895
	" " MAT	0,3636	3,8198
	" " PPL	0,0066	0,0120
	" " INC	0,3173	1,9800
	Total	0,2820	
INC	Direto sobre LNSV	0,5961	6,9907
	Indireto via FLR	-0,2326	8,5642
	" " MAT	0,5942	10,2010
	" " PPL	-0,0283	0,2195
	" " STD	-0,1140	0,7247
	Total	0,8154	
Coeficiente de determinação		0,7502	

Continua...

QUADRO 10, Cont.

Variável	Efeito	Estimativa	VIF .
FLR	Direto sobre LPCS	0,0502	11,6880
	Indireto via MAT	-0,1350	10,7616
	" " PPL	-0,0110	0,0476
	" " STD	0,1906	1,2674
	" " INC	0,4175	5,1223
	Total	0,5123	
MAT	Direto sobre LPCS	-0,1455	12,4909
	Indireto via FLR	0,0466	10,0699
	" " PPL	-0,0144	0,0824
	" " STD	0,1498	0,7825
	" " INC	0,4407	5,7091
	Total	0,4772	
PPL	Direto sobre LPCS	-0,0607	1,4590
	Indireto via FLR	0,0091	0,3812
	" " MAT	-0,0346	0,7057
	" " STD	-0,0246	0,0211
	" " INC	0,1892	1,0519
	Total	0,0784	
STD	Direto sobre LPCS	-0,2709	2,5587
	Indireto via FLR	-0,0353	5,7895
	" " MAT	0,0804	3,8198
	" " PPL	-0,0055	0,0120
	" " INC	-0,2595	1,9800
	Total	-0,4908	
INC	Direto sobre LPCS	-0,4877	6,6907
	Indireto via FLR	-0,0430	8,5642
	" " MAT	0,1315	10,2010
	" " PPL	0,0235	0,2153
	" " STD	-0,1441	0,7247
	Total	-0,5198	
Coeficiente de determinação		0,3380	

direto, nos dois experimentos, foi LNVG. Assim, é desejável a resposta indireta sobre este pela seleção em caracteres secundários de menor complexidade gênica. Considerando a consistência das correlações genotípicas nos experimentos 1 e 2, verifica-se que é possível a obtenção de ganhos em LNVG pela seleção em INC e MAT. Por meio da análise de trilha, pode-se verificar, porém, que a seleção indireta via INC é mais eficiente que via MAT, pois o efeito direto de INC mostrou-se consistentemente positivo para os dados dos dois experimentos.

4.4.1.3. Efeitos Indiretos das Variáveis Secundárias sobre a Variável Básica

No Quadro 11, são apresentadas as estimativas dos efeitos indiretos das variáveis secundárias sobre a variável básica para os experimentos 1 e 2.

Observa-se que o índice de colheita (INC) apresentou a mais alta correlação positiva e número de dias até a maturação (MAT) e a mais alta correlação negativa com a produção de grãos (PRD) consistentes para os dados dos dois experimentos. O INC foi, ainda, o caráter que apresentou o maior efeito indireto primário e positivo sobre PRD, ressaltando mais uma vez a importância desse caráter, cuja ação concentra-se, principalmente, sobre o número de vagens (LNVG).

O peso da palhada (PPL) também apresentou efeito indireto primário positivo sobre PRD, com ação concentrada sobre LNVG.

4.4.2. Análise de Trilha com Correção para os Efeitos da Multicolinearidade entre as Observações

Nesse estudo, procurou-se verificar o efeito da multicolinearidade sobre os coeficientes de trilha, comparando-se os fatores de inflação da variância (VIF's) associados a esses coeficientes. Para isso, efetuou-se novamente a análise de trilha com base nos diagramas causais apresentados nas Figuras 2, 3 e 4, empregando-se um procedimento análogo ao método de regressão em cristas ou em cumeeira, proposto por HOERL e KENNARD (1970a,b). Esse procedimento consiste na solução do sistema

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

em que θ^* representa os coeficientes de trilha; $X'X$ é a matriz de correlações entre as variáveis explicativas e $X'Y$ corresponde às correlações entre as variáveis explicativas e a variável tomada como base. Os valores de k foram determinados pelo exame dos gráficos de cristas, como já descrito.

4.4.2.1. Efeitos das Variáveis Primárias sobre a Variável Básica, Estimados com Correção para os Efeitos da Multicolinearidade

Com base no esquema causal mostrado na Figura 2, foram estimados os efeitos diretos e indiretos das variáveis primárias sobre a produção de grãos, para os Experimentos 1 e 2, com correção para os efeitos da multicolinearidade. Primeiramente, foram aplicados os métodos para diagnóstico de multicolinearidade nas matrizes de

QUADRO 11 - Efeitos Indiretos Primários e Secundários das Variáveis Secundárias sobre a Variável Básica, para os Experimentos 1 e 2

Variável	Efeito	Exp. 1	Exp. 2
FLR	Indireto primário via LNVG	0,0147	-0,3034
	" " " LNSV	0,1604	0,1032
	" " " LPCS	-0,2365	0,1047
	Total primário	-0,0614	-0,1855
	Ind. secundário via MAT via LNVG	-0,0308	0,2714
	" " " MAT " LNSV	-0,0271	-0,2317
	" " " MAT " LPCS	0,0880	-0,0394
	Total secundário via MAT	0,0301	0,0003
	Ind. secundário via PPL via LNVG	-0,1801	0,0702
	" " " PPL " LNSV	-0,0501	0,0050
	" " " PPL " LPCS	0,0735	-0,0032
	Total secundário via PPL	-0,1567	0,0720
	Ind. secundário via STD via LNVG	0,1223	-0,1149
	" " " STD " LNSV	0,0115	-0,0572
	" " " STD " LPCS	-0,1016	0,0556
	Total secundário via STD	0,0322	-0,1165
	Ind. secundário via INC via LNVG	-0,0710	-0,7032
	" " " INC " LNSV	-0,0170	-0,1937
	" " " INC " LPCS	0,0059	0,1218
	Total secundário via INC	-0,0821	-0,7751
Total (r_{y4})	-0,2379	-1,0048	
MAT	Indireto primário via LNVG	-0,1330	0,2924
	" " " LNSV	-0,1169	-0,2496
	" " " LPCS	0,3797	-0,0425
	Total primário	0,1298	0,0003
	Ind. secundário via FLR via LNVG	0,0034	-0,2816
	" " " FLR " LNSV	0,0372	0,0957
	" " " FLR " LPCS	-0,0548	0,0136
	Total secundário via FLR	-0,0142	-0,1723
	Ind. secundário via PPL via LNVG	0,0579	0,0924
	" " " PPL " LNSV	0,0161	0,0066
	" " " PPL " LPCS	-0,0236	-0,0042
	Total secundário via PPL	0,0504	0,0948

Continua...

QUADRO 11, Cont.

Variável	Efeito	Exp. 1	Exp. 2
MAT	Ind. secundário via STD via LNVG	-0,0144	-0,0903
	" " " STD " LNSV	-0,0013	0,0450
	" " " STD " LPCS	0,0119	0,0437
	Total secundário via STD	-0,0038	-0,0016
	Ind. secundário via INC via LNVG	-0,6377	-0,7424
	" " " INC " LNSV	-0,1525	-0,2045
	" " " INC " LPCS	-0,0529	0,1286
	Total secundário via INC	-0,8431	-0,8183
	Total (r_{Y5})	-0,6809	-0,8971
	PPL	Indireto primário via LNVG	0,7237
" " " LNSV		0,2014	0,0277
" " " LPCS		-0,2952	-0,0177
Total primário		0,6299	0,3987
Ind. secundário via FLR via LNVG		-0,0037	-0,0548
" " " FLR " LNSV		-0,0399	0,0186
" " " FLR " LPCS		0,0588	0,0026
Total secundário via FLR		0,0152	-0,0336
Ind. secundário via MAT via LNVG		-0,0106	0,0695
" " " MAT " LNSV		-0,0093	-0,0593
" " " MAT " LPCS		0,0304	-0,0101
Total secundário via MAT		0,0105	0,0001
Ind. secundário via STD via LNVG		-0,0997	0,0148
" " " STD " LNSV		-0,0094	-0,0074
" " " STD " LPCS		0,0828	-0,0072
Total secundário via STD		-0,0263	0,0002
Ind. secundário via INC via LNVG		-0,0304	-0,3187
" " " INC " LNSV		-0,0073	-0,0878
" " " INC " LPCS	0,0025	0,0552	
Total secundário via INC	-0,0352	-0,3513	
Total (r_{Y6})	0,5941	0,0141	
STD	Indireto primário via LNVG	-0,2090	0,1633
	" " " LNSV	-0,0196	-0,0813
	" " " LPCS	0,1735	-0,0791
	Total primário	-0,0551	0,0029

Continua...

QUADRO 11, Cont.

Variável	Efeito	Exp. 1	Exp. 2	
STD	Ind. secundário via FLR via LNVG	-0,0086	0,2135	
	" " " FLR " LNSV	-0,0939	-0,0726	
	" " " FLR " LPCS	-0,1384	-0,0103	
	Total secundário via FLR	-0,2409	0,1306	
	Ind. secundário via MAT via LNVG	-0,0092	-0,1617	
	" " " MAT " LNSV	-0,0081	-0,1380	
	" " " MAT " LPCS	0,0262	0,0235	
	Total secundário via MAT	0,0089	-0,2762	
	Ind. secundário via PPL via LNVG	0,3454	0,0353	
	" " " PPL " LNSV	0,0961	0,0025	
	" " " PPL " LPCS	-0,1409	-0,0016	
	Total secundário via PPL	0,3006	0,0362	
	Ind. secundário via INC via LNVG	-0,0466	0,4372	
	" " " INC " LNSV	-0,0111	0,1204	
	" " " INC " LPCS	0,0039	-0,0757	
	Total secundário via INC	-0,0538	0,4819	
	Total (r_{Y7})	0,0177	0,3174	
	INC	Indireto primário via LNVG	0,7728	0,8215
		" " " LNSV	0,1849	0,2263
" " " LPCS		-0,0641	-0,1423	
Total primário		0,8936	0,9055	
Ind. secundário via FLR via LNVG		-0,0013	0,2597	
" " " FLR " LNSV		-0,0147	-0,0883	
" " " FLR " LPCS		0,0217	-0,0125	
Total secundário via FLR		0,0057	0,1589	
Ind. secundário via MAT via LNVG		0,1097	-0,2642	
" " " MAT " LNSV		0,0965	0,2256	
" " " MAT " LPCS		-0,3133	0,0384	
Total secundário via MAT		-0,1071	-0,0002	
Ind. secundário via PPL via LNVG		-0,0284	-0,1507	
" " " PPL " LNSV		-0,0079	-0,0107	
" " " PPL " LPCS		0,0116	0,0069	
Total secundário via PPL		-0,0247	-0,1545	
Ind. secundário via STD via LNVG		0,0126	0,0869	
" " " STD " LNSV		0,0012	-0,0433	
" " " STD " LPCS		-0,0105	-0,0421	
Total secundário via STD		0,0033	0,0015	
Total (r_{Y8})	0,7708	0,9112		

correlações fenotípicas correspondentes às três variáveis explicativas primárias. Os resultados desse diagnóstico são apresentados no Quadro 12.

É desejável que um método de diagnóstico de multicolinearidade forneça informações sobre o grau com que esta se manifesta e ainda possibilite a identificação das variáveis envolvidas no problema. O exame da matriz de correlações, como proposto por NETER et al. (1983), constitui-se na simples verificação da magnitude dos seus elementos não-diagonais. No Quadro 12, observa-se que para as duas matrizes as maiores correlações não se aproximam demasiadamente da unidade. Logo, é possível que não esteja ocorrendo alto grau de colinearidade. Contudo, faz-se necessária a observação dos demais métodos para que essa hipótese seja confirmada.

As variáveis envolvidas em correlações simples são identificadas pelo exame da matriz de correlações. Porém, é possível que três ou mais variáveis independentes estejam envolvidas em multicolinearidade sem que quaisquer pares dessas variáveis sejam altamente correlacionadas. Uma situação como esta não é diagnosticada pelo exame da matriz de correlações, como salienta FERRARI (1989).

Observa-se também, no Quadro 12, que os fatores de inflação da variância (VIF's) não atingiram valor superior a 10, o que é também um indicativo de que não está ocorrendo multicolinearidade em grau elevado. Outro método de diagnóstico consiste no exame do determinante de cada matriz de correlação. Nessas matrizes, o determinante varia de zero a um, caso as variáveis sejam perfeitamente correlacionadas ou ortogonais entre si, respectivamente. Constata-se

QUADRO 12 - Resultados dos Testes para Diagnóstico de Multicolinearidade Aplicados às Matrizes de Correlações Fenotípicas Correspondentes às Variáveis Explicativas Primárias, para os Experimentos 1 e 2

Item observado	Experimento 1	Experimento 2
Maior Correlação	0,8288 (LNVG x LNSV)	0,5719 (LNVG x LNSV)
Menor Correlação	-0,8723 (LNSV x LPCS)	-0,6637 (LNVG x LPCS)
Número de VIF's ≥ 10	0	0
Determinante	0,0728	0,3478
Autovalores	[2,6468 0,2372 0,1160]	[2,1911 0,4740 0,3349]
Número de Condição	22,8155	6,5420
Valor Singular	[2,6468 0,2372 0,1160]	[2,1911 0,4740 0,3349]
Índice de Condição	[1,0000 11,1585 22,8172]	[1,0000 4,6226 6,5425]
VIF _k	[4,4505 3,4469 5,3165]	[1,8301 2,1243 1,5975]
Proporções de Decomposição da Variância dos Efeitos Diretos	[0,0103 0,0254 0,0018] 1,3839 0,1402 2,0890 5,5442 7,7540 4,3079	[0,0399 0,0101 0,0712] 0,7464 1,4375 0,0197 1,6684 0,8854 2,4917

LNVG: logaritmo do número médio de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes.

que o determinante correspondente à primeira matriz em estudo foi menor que o determinante da segunda matriz, sem que ambos tendessem para a nulidade, o que é também indício de multicolinearidade fraca.

O exame dos autovalores e do número de condição, dado pela razão entre o maior e o menor dentre eles, possibilita identificar o grau de multicolinearidade presente. Assim, nos dois casos, a multicolinearidade pode ser classificada como fraca, conforme critério de MONTGOMERY e PECK (1981), pois os números de condição encontrados foram inferiores a 100.

Como visto, esses primeiros métodos discutidos são pouco informativos quanto à multicolinearidade, pois, apesar de detectarem sua presença e até sua intensidade, geralmente são incapazes de identificar as variáveis envolvidas no problema.

A decomposição da matriz de correlações em valores singulares e a determinação dos índices de condição, dados pela relação entre o maior e os demais valores singulares, fornecem também informações sobre o grau de multicolinearidade presente. Constata-se que a primeira matriz apresentou índices de condição mais elevados que a segunda matriz. Pode-se, então, concluir que no primeiro caso a multicolinearidade é um pouco maior que no segundo. Os fatores de inflação da variância associada a cada coeficiente de trilha (VIF_k) constituem indicativos de efeitos adversos da multicolinearidade. Pode-se verificar que, na primeira matriz, os valores de VIF foram relativamente mais altos que na segunda, evidenciando novamente a presença de multicolinearidade mais intensa na primeira.

A combinação dos métodos de decomposição em valores singulares com o método de decomposição da variância dos efeitos

diretos (p) da análise de trilha, em analogia ao proposto por BELSLEY et al. (1980), possibilita a identificação das variáveis envolvidas em multicolinearidade. Assim, para a primeira matriz, correspondente ao Experimento 1, as maiores proporções da decomposição da variância correspondem às variáveis 1 (logaritmo do número de vagens - LNVG) e 2 (logaritmo do número de sementes por vagem - LNSV), sendo estas as variáveis mais correlacionadas entre si. Para a segunda matriz, correspondente ao Experimento 2, verifica-se, pelo mesmo processo, que estão mais correlacionadas às variáveis LNVG e logaritmo do peso médio de 100 sementes - LPCS.

Os valores de k foram determinados pelo exame do gráfico de cristas correspondente a cada experimento apresentado na Figura 8. Procurou-se determinar, através desses gráficos, o menor valor de k para o qual a maioria dos coeficientes de trilha se estabilizam, em analogia ao que preconizam HOERL e KENNARD (1970b) e MONTGOMERY e PECK (1981), para regressão em cristas. Pelo exame dos gráficos em questão, verifica-se que essa estabilização pode ser considerada aceitável para $k=0,0808$, sendo esse o valor utilizado na estimação dos coeficientes, os quais são apresentados no Quadro 13. Comparando-se esses resultados com aqueles obtidos pelo método dos mínimos quadrados apresentados no Quadro 8 pode-se observar que a análise de trilha pelo método de cristas proporcionou redução dos fatores de inflação das variâncias (VIF's) associadas aos efeitos diretos e indiretos das variáveis primárias sobre a variável básica, mesmo em presença de multicolinearidade fraca. As estimativas dos efeitos diretos e indiretos foram pouco afetadas, sem alteração de sinais. No entanto, verificou-se redução nos coeficientes de determinação.

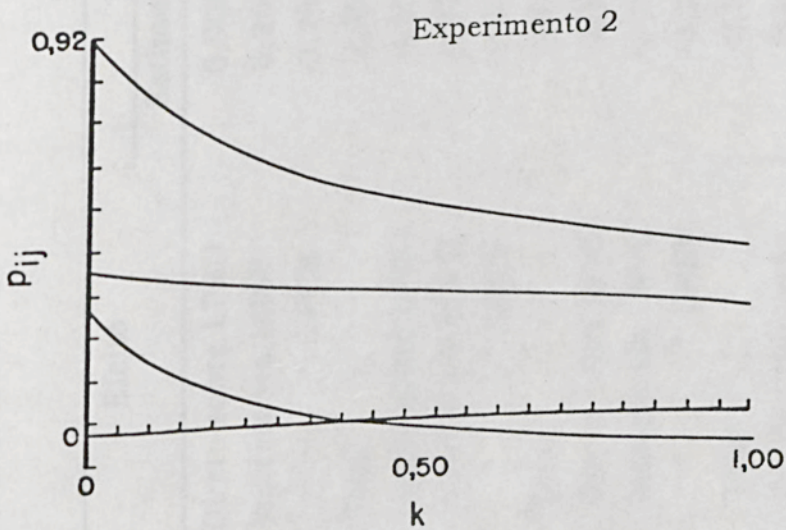
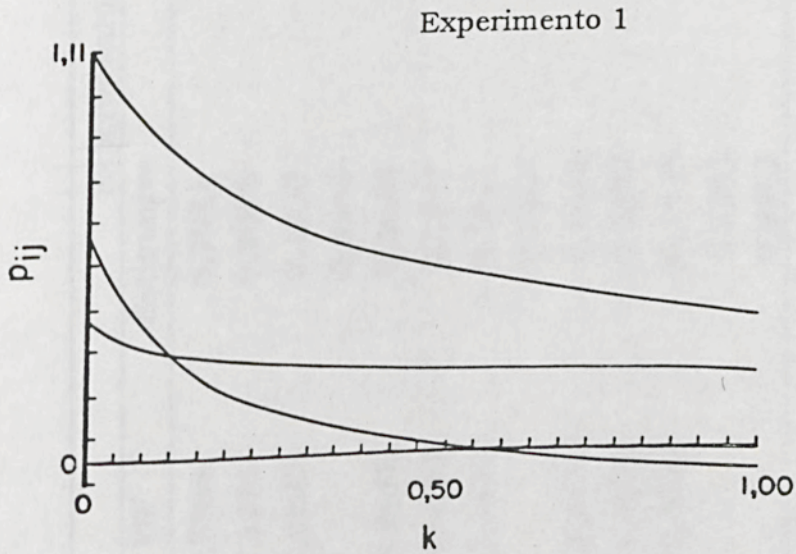


FIGURA 8 - Gráficos de Cristas para os Coeficientes de Trilha, entre as Variáveis Primárias e a Variável Básica, Correspondentes aos Valores de k no Intervalo $0 < k < 1$, para os Experimentos 1 e 2.

QUADRO 13 - Efeitos Diretos e Indiretos das Variáveis Primárias sobre a Variável Básica, Fatores de Inflação da Variância (VIF's) e Coeficiente de Determinação, Obtidos pelo Método de Cristas, para os Experimentos 1 e 2

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNVG	Direto sobre LPRD	0,9086	2,7868	0,7951	1,9325
	Indireto via LNSV	0,2549	2,3275	0,2079	0,4542
	" " LPCS	-0,2846	1,6630	-0,1223	0,7071
	Total	0,9522		0,9449	
LNSV	Direto sobre LPRD	0,3075	3,9578	0,3635	1,6221
	Indireto via LPRD	0,7530	1,6388	0,4547	0,5411
	" " LPCS	-0,3233	2,1453	-0,1011	0,4836
	Total	0,7621		0,7464	
LPCS	Direto sobre LPRD	0,3706	3,2933	0,1842	1,8750
	Indireto via LNVG	-0,6978	1,4072	-0,5277	0,7288
	" " LNSV	-0,2682	2,5782	-0,1995	0,4184
	Total	-0,5655		0,5281	
	Coeficiente de determinação	0,8899		0,9253	

LNVG: logaritmo do número médio de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes.

4.4.2.2. Efeitos das Variáveis Secundárias sobre as Variáveis Primárias, Estimados com Correção para os Efeitos da Multicolinearidade

Os efeitos diretos e indiretos das cinco variáveis explicativas secundárias sobre as três variáveis explicativas primárias, conforme esquema causal mostrado na Figura 3, foram estimados com correção, para os efeitos da multicolinearidade, para os Experimentos 1 e 2. A correção foi aplicada sabendo-se, porém, que ambas as matrizes de correlações entre as variáveis explicativas apresentaram multicolinearidade fraca. Isso pode ser constatado pelos resultados dos respectivos diagnósticos apresentados no Quadro 14.

Os coeficientes de trilha foram estimados pelo método de cristas determinando-se o valor de k pelo exame do respectivo gráfico. Dois desses gráficos correspondentes aos efeitos das variáveis secundárias sobre a variável primária logaritmo do número de vagens - LN_{VG}, para os Experimentos 1 e 2, são apresentados na Figura 9. Percebe-se, pelo exame desses gráficos, que a maioria dos coeficientes de trilha estabilizam-se em $k=0,0750$. Tomando-se esse valor para k , foram, então, estimados os efeitos diretos e indiretos. Observou-se, ainda, que ao se estimar os efeitos das variáveis explicativas secundárias sobre as demais variáveis explicativas primárias, esse mesmo valor de k mostrou-se adequado. Assim, tomou-se $k=0,0750$ para a estimação dos efeitos das variáveis secundárias sobre todas as primárias.

Nos Quadros 15 e 16, são apresentados os efeitos diretos e indiretos das cinco variáveis explicativas secundárias sobre as três primárias e também os respectivos fatores de inflação da variância

QUADRO 14 - Resultados dos Testes para Diagnóstico de Multicolinearidade, Aplicados às Matrizes de Correlações Fenotípicas Correspondentes às Variáveis Explicativas Secundárias, para os Experimentos 1 e 2

Item observado	Experimento 1	Experimento 2
Maior Correlação	0,4773 (PPL x STD)	0,9282 (FLR x MAT)
Menor Correlação	-0,8252 (MAT x IND)	-0,9037 (MAT x IND)
Número de VIF's ≥ 10	0	2
Determinante	0,1347	0,0074
Autovalores	[1,91 1,85 0,73 0,35 0,15]	[3,33 1,11 0,41 0,11 0,04]
Número de Condição	12,9255	74,9877
Valor Singular	[1,91 1,85 0,73 0,35 0,15]	[3,33 1,11 0,41 0,11 0,04]
Índice de Condição	[1,00 1,03 2,62 5,46 12,73]	[1,00 3,00 8,12 30,27 83,25]
VIF _k	[2,23 2,96 1,37 2,30 3,18]	[7,77 4,40 6,88 13,89 2,25]
Proporções de Decomposição da Variância dos Efeitos Diretos	[0,03 0,01 0,06 0,03 0,01] 0,03 0,03 0,00 0,00 0,05 0,09 0,10 0,84 0,09 0,00 0,76 0,60 0,09 1,89 0,08	[0,00 0,00 0,00 0,00 0,01] 0,03 0,00 0,02 0,00 0,19 0,02 1,01 0,18 0,01 0,01 1,63 4,00 6,84 0,06 1,27
	3,71 4,43 1,57 1,41 5,94	17,30 3,17 3,31 22,05 13,47

FLR: Número de dias até o florescimento;

MAT: Número de dias até a maturação;

PPL: Peso da palhada;

STD: Estande final;

IND: Índice de colheita.

QUADRO 15

Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variáveis Primárias em Função da Direção da Variação da Variável Primária LNVG, para os Experimentos 1 e 2.

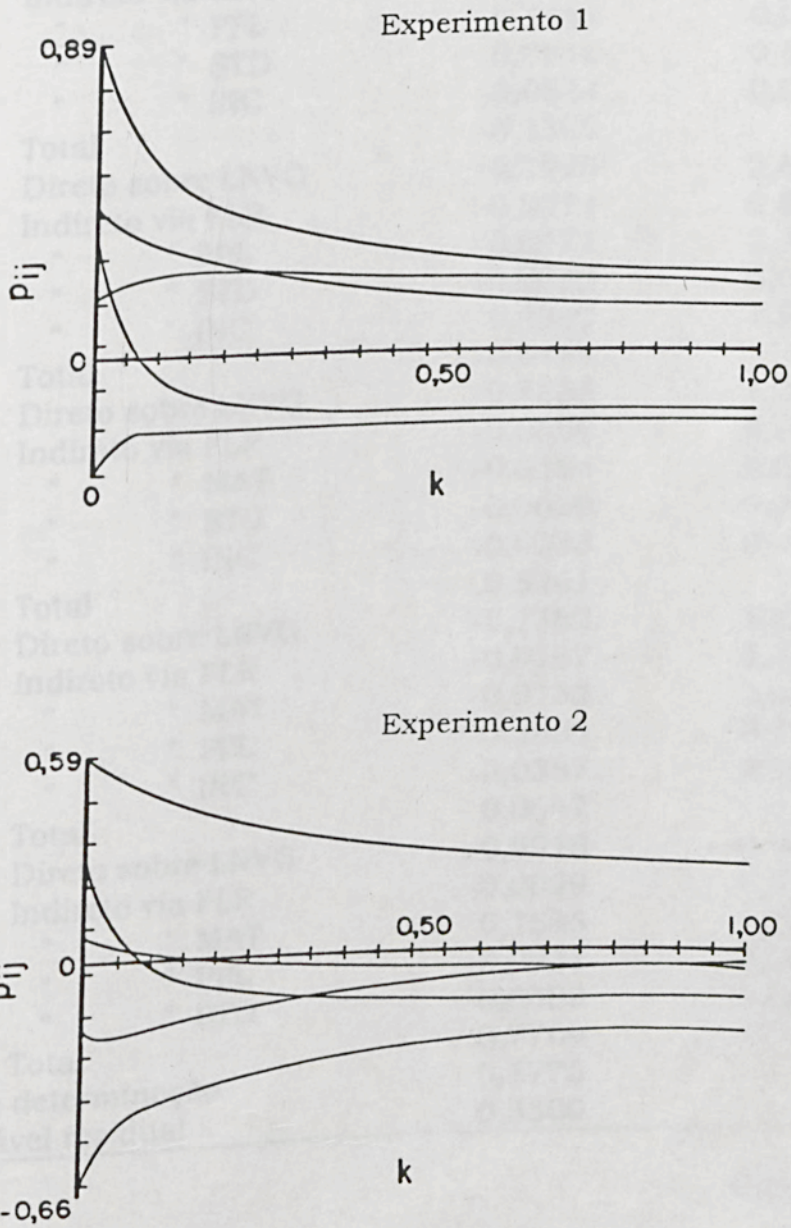


FIGURA 9 - Gráficos de Cristas para os Coeficientes de Trilha entre as Variáveis Secundárias e a Variável Primária LNVG, Correspondentes aos Valores de k no Intervalo $0 < k < 1$, para os Experimentos 1 e 2.

QUADRO 15 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variáveis Primárias Estimados com Correção para Multicolinearidade, Fatores de Inflação da Variância (VIF's), Coeficiente de Determinação e Efeito da Variável Residual, Para o Experimento 1

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNMG	0,0320	1,7168
	Indireto via MAT	-0,0445	0,1347
MAT	" " PPL	-0,1464	0,0720
	" " STD	0,0810	0,5556
MAT	" " INC	-0,0544	0,0196
	Total	-0,1300	
	Direto sobre LNMG	-0,1920	2,8981
	Indireto via FLR	0,0074	0,0798
	" " PPL	0,0471	0,0075
	" " STD	-0,0095	0,0077
PPL	" " INC	0,4883	1,5799
	Total	-0,6498	
	Direto sobre LNMG	0,5885	1,3449
	Indireto via FLR	-0,0080	0,0920
	" " MAT	-0,0154	0,0161
	" " STD	-0,0659	0,3693
STD	" " INC	-0,0233	0,0036
	Total	0,5201	
	Direto sobre LNMG	-0,1382	1,8731
	Indireto via FLR	-0,0187	0,5093
	" " MAT	-0,0132	0,0119
	" " PPL	0,2809	0,2651
INC	" " INC	-0,0357	0,0084
	Total	0,0647	
	Direto sobre LNMG	0,5918	2,6812
	Indireto via FLR	-0,0029	0,0125
	" " MAT	0,1585	1,7077
	" " PPL	-0,0231	0,0018
INC	" " STD	0,0083	0,0059
	Total	0,7769	
	Coeficiente de determinação	0,8775	
Efeito da variável residual		0,3500	

Continua...

QUADRO 15, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSV	0,3703	1,7168
	Indireto via MAT	-0,0678	0,1347
	" " PPL	-0,1185	0,0720
	" " STD	0,0318	0,5557
	" " INC	-0,0414	0,0196
	Total	0,0222	
MAT	Direto sobre LNSV	-0,2926	2,8981
	Indireto via FLR	0,0858	0,0798
	" " PPL	0,0381	0,0075
	" " STD	-0,0037	0,0077
	" " INC	-0,3719	1,5799
	Total	-0,5662	
PPL	Direto sobre LNSV	0,4762	1,3449
	Indireto via FLR	-0,0921	0,0920
	" " MAT	-0,0234	0,0161
	" " STD	-0,0259	0,3693
	" " INC	-0,0177	0,0036
	Total	0,3527	
STD	Direto sobre LNSV	-0,0544	1,8732
	Indireto via FLR	-0,2168	0,5093
	" " MAT	-0,0202	0,0119
	" " PPL	0,2273	0,2651
	" " INC	-0,0272	0,0084
	Total	0,0953	
INC	Direto sobre LNSV	0,4506	2,6812
	Indireto via FLR	-0,0340	0,0125
	" " MAT	0,2414	1,7077
	" " PPL	-0,0187	0,0018
	" " STD	0,0033	0,0059
	Total	0,6764	
Coeficiente de determinação		0,7185	
Efeito da variável residual		0,5306	

Continua...

QUADRO 15, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LPCS	-0,3284	1,7168
	Indireto via MAT	0,1145	0,1347
	" " PPL	0,1035	0,0720
	" " STD	-0,1542	0,5557
	" " INC	0,0163	0,0196
	Total	0,2731	
MAT	Direto sobre LPCS	0,4938	2,8981
	Indireto via FLR	-0,0761	0,0798
	" " PPL	-0,0333	0,0075
	" " STD	0,0181	0,0077
	" " INC	0,1461	1,5799
	Total	0,5857	
PPL	Direto sobre LPCS	-0,4158	1,3449
	Indireto via FLR	0,0817	0,0920
	" " MAT	0,0395	0,0161
	" " STD	0,1257	0,3693
	" " INC	0,0070	0,0036
	Total	-0,1931	
STD	Direto sobre LPCS	0,2634	1,8731
	Indireto via FLR	0,1923	0,5093
	" " MAT	0,0340	0,0120
	" " PPL	-0,1985	0,2651
	" " INC	0,0107	0,0084
	Total	0,3217	
INC	Direto sobre LPCS	-0,1771	2,6812
	Indireto via FLR	0,0302	0,0125
	" " MAT	-0,4075	1,7077
	" " PPL	0,0163	0,0018
	" " STD	-0,0159	0,0059
	Total	0,5672	
		0,6444	
	Coefficiente de determinação	0,5963	
	Efeito da variável residual		

Continua...

QUADRO 16 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variáveis Primárias com Correção para os Efeitos da Multicolinearidade, Fatores de Inflação da Variância, Coeficiente de Determinação e Efeito da Variável Residual, para o Experimento 2

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNVG	-0,2165	5,8226
	Indireto via MAT	-0,0181	4,3010
	" " PPL	0,0596	0,0394
	" " STD	-0,1609	0,8841
	" " INC	-0,4943	2,8838
	Total	-0,8463	
MAT	Direto sobre LNVG	-0,0195	5,7691
	Indireto via FLR	-0,2009	4,3409
	" " PPL	0,0785	0,0683
	" " STD	-0,1264	0,5459
	" " INC	-0,5218	3,2142
	Total	-0,7916	
PPL	Direto sobre LNVG	0,3302	1,3962
	Indireto via FLR	-0,0391	0,1643
	" " MAT	-0,0046	0,2821
	" " STD	0,0207	0,0147
	" " INC	-0,2240	0,5922
	Total	0,1080	
STD	Direto sobre LNVG	0,2286	2,0627
	Indireto via FLR	0,1523	2,4957
	" " MAT	0,0108	1,5267
	" " PPL	0,0299	0,0094
	" " INC	0,3073	1,1147
	Total	0,7461	
INC	Direto sobre LNVG	0,5774	4,5482
	Indireto via FLR	0,1853	3,6919
	" " MAT	0,0176	4,0770
	" " PPL	-0,1281	0,1818
	" " STD	0,1217	0,5056
	Total	0,8172	
	Coeficiente de determinação	0,8767	
	Efeito da variável residual	0,3511	

Continua...

QUADRO 16, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSV	0,0078	5,8226
	Indireto via MAT	-0,4253	4,3010
	" " PPL	0,0061	0,0394
	" " STD	0,1539	0,8841
	" " INC	-0,4280	2,8838
	Total	-0,6849	
MAT	Direto sobre LNSV	-0,4582	5,7691
	Indireto via FLR	0,0073	4,3409
	" " PPL	0,0081	0,0683
	" " STD	0,1209	0,5459
	" " INC	-0,4519	3,2142
	Total	-0,8082	
PPL	Direto sobre LNSV	0,0339	1,3962
	Indireto via FLR	0,0014	0,1643
	" " MAT	-0,1089	0,2821
	" " STD	-0,0198	0,0147
	" " INC	-0,1940	0,5922
	Total	-0,2849	
STD	Direto sobre LNSV	-0,2187	2,0627
	Indireto via FLR	-0,0055	2,4957
	" " MAT	0,2534	1,5267
	" " PPL	0,0031	0,0099
	" " INC	0,2661	1,1147
	Total	0,2820	
INC	Direto sobre LNSV	0,5000	4,5482
	Indireto via FLR	-0,0067	3,6919
	" " MAT	0,4141	4,0770
	" " PPL	-0,0131	0,1818
	" " STD	-0,1164	0,5056
	Total	0,8154	
		0,7014	
	Coefficiente de determinação	0,5465	
	Efeito da variável residual		

Continua...

QUADRO 16, Cont.

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LPCS	0,0613	5,8226
	Indireto via MAT	-0,0072	4,3010
	" " PPL	-0,0057	0,0394
	" " STD	0,1806	0,8841
	" " INC	0,2787	2,8838
	Total	0,5123	
MAT	Direto sobre LPCS	-0,0077	5,7691
	Indireto via FLR	0,0569	4,3409
	" " PPL	-0,0075	0,0683
	" " STD	0,1419	0,5459
	" " INC	0,2942	3,2142
	Total	0,4772	
PPL	Direto sobre LPCS	-0,0315	1,3962
	Indireto via FLR	0,0111	0,1643
	" " MAT	-0,0018	0,2821
	" " STD	-0,0233	0,0147
	" " INC	0,1263	0,5922
	Total	0,0784	
STD	Direto sobre LPCS	-0,2566	2,0627
	Indireto via FLR	-0,0432	2,4957
	" " MAT	0,0043	1,5267
	" " PPL	-0,0029	0,0099
	" " INC	-0,1733	1,1147
	Total	-0,4908	
INC	Direto sobre LPCS	-0,3255	4,5482
	Indireto via FLR	-0,0525	3,6919
	" " MAT	0,0070	4,0770
	" " PPL	0,0122	0,1818
	" " STD	-0,1365	0,5056
	Total	-0,5198	
		0,3204	
	Coefficiente de determinação	0,8244	
	Efeito da variável residual		

(VIF's) estimados com correção para os efeitos da multicolinearidade. Comparando-se esses resultados com as estimativas obtidas pelo método dos mínimos quadrados, apresentadas nos Quadros 9 e 10, pode-se observar, também, que em geral, a estimação em cristas proporcionou redução dos fatores de inflação da variância (VIF's) associados aos efeitos. Estes, por sua vez, foram ligeiramente afetados; mas por estarem sujeitos a uma variância menor, podem ser mais confiáveis. Os coeficientes de determinação (r^2) sofreram redução inferior a 8%, em média.

4.4.2.3. Efeitos das Variáveis Secundárias sobre a Variável Básica, Estimados com Correção para os Efeitos da Multicolinearidade

Os efeitos diretos e indiretos das variáveis explicativas secundárias sobre a variável básica, estimados com correção para os efeitos da multicolinearidade, são apresentados no Quadro 17, para os Experimentos 1 e 2. Como nos casos anteriores de estimação em cristas, verifica-se ligeira alteração nos efeitos em relação à estimação por mínimos quadrados, cujos resultados estão apresentados no Quadro 11.

4.4.2.4. Aplicação do Método de Estimação em Cristas sob Multicolinearidade Intensa

Considerou-se o diagrama causal com quatro variáveis explicativas primárias, apresentado na Figura 5, em que a variável logaritmo do número de sementes ($\log X_9$ ou LNSM) é perfeitamente

QUADRO 17 - Efeitos Indiretos Primários e Secundários das Variáveis Secundárias sobre a Variável Básica, com Correção para os Efeitos da Multicolinearidade, para os Experimentos 1 (Exp. 1) e 2 (Exp. 2)

Variável	Efeito	Exp. 1	Exp. 2
FLR	Indireto primário via LNVG	0,0291	-0,1721
	" " " LNSV	0,1139	0,0028
	" " " LPCS	-0,1217	0,0113
	Total primário	0,0213	-0,1580
	Ind. secundário via MAT via LNVG	-0,0404	-0,0144
	" " " MAT " LNSV	-0,0209	-0,1546
	" " " MAT " LPCS	0,0424	-0,0013
	Total secundário via MAT	-0,0189	-0,1703
	Ind. secundário via PPL via LNVG	-0,1330	0,0474
	" " " PPL " LNSV	-0,0364	0,0022
	" " " PPL " LPCS	0,0383	-0,0010
	Total secundário via PPL	-0,1311	0,0486
	Ind. secundário via STD via LNVG	0,0735	-0,1279
	" " " STD " LNSV	0,0098	0,0559
	" " " STD " LPCS	-0,0571	0,0333
	Total secundário via STD	0,0262	-0,0387
	Ind. secundário via INC via LNVG	-0,0494	-0,3930
	" " " INC " LNSV	-0,0127	-0,1556
	" " " INC " LPCS	0,0060	0,0513
Total secundário via INC	-0,0565	-0,4973	
Total (r_{Y4})	-0,1590	-0,8157	
MAT	Indireto primário via LNVG	-0,1745	-0,0155
	" " " LNSV	-0,0900	-0,1665
	" " " LPCS	0,1830	-0,0014
	Total primário	-0,0815	-0,1834
	Ind. secundário via FLR via LNVG	0,0067	-0,1597
	" " " FLR " LNSV	0,0264	0,0026
	" " " FLR " LPCS	-0,0282	0,0105
	Total secundário via FLR	0,0049	-0,1466
	Ind. secundário via PPL via LNVG	0,0428	0,0624
	" " " PPL " LNSV	0,0117	0,0029
	" " " PPL " LPCS	-0,0123	-0,0014
	Total secundário via PPL	0,0422	0,0639

Continua...

QUADRO 17, Cont.

Variável	Efeito	Exp. 1	Exp. 2
MAT	Ind. secundário via STD via LNMG	-0,0086	-0,1005
	" " " STD " LNSV	-0,0011	-0,0439
	" " " STD " LPCS	0,0067	0,0261
	Total secundário via STD	-0,0030	-0,1183
	Ind. secundário via INC via LNMG	-0,4437	-0,4149
	" " " INC " LNSV	-0,1143	-0,1642
	" " " INC " LPCS	0,0541	0,0542
	Total secundário via INC	-0,5039	-0,5299
	Total (r_{Y5})	-0,5413	-0,9143
	PPL	Indireto primário via LNMG	-0,5347
" " " LNSV		0,1464	0,0123
" " " LPCS		-0,1541	-0,0058
Total primário		0,5270	0,2690
Ind. secundário via FLR via LNMG		-0,0072	-0,0311
" " " FLR " LNSV		-0,0283	0,0005
" " " FLR " LPCS		0,0303	0,0020
Total secundário via FLR		-0,0052	-0,0286
Ind. secundário via MAT via LNMG		-0,0139	-0,0037
" " " MAT " LNSV		-0,0072	-0,0396
" " " MAT " LPCS		0,0146	-0,0003
Total secundário via MAT		-0,0065	-0,0436
Ind. secundário via STD via LNMG		-0,0599	0,0165
" " " STD " LNSV		-0,0080	-0,0072
" " " STD " LPCS		0,0466	-0,0043
Total secundário via STD		-0,0213	0,0050
Ind. secundário via INC via LNMG		-0,0211	-0,1781
" " " INC " LNSV		-0,0054	-0,0705
" " " INC " LPCS		0,0026	0,0233
Total secundário via INC		-0,0239	-0,2253
Total (r_{Y6})	0,4701	-0,0235	
STD	Indireto primário via LNMG	-0,1255	0,1817
	" " " LNSV	-0,0167	-0,0795
	" " " LPCS	0,0976	-0,0473
	Total primário	-0,0446	0,0549

Continua...

QUADRO 17, Cont.

Variável	Efeito	Exp. 1	Exp. 2	
STD	Ind. secundário via FLR via LNVG	-0,0170	0,1211	
	" " " FLR " LNSV	-0,0667	-0,0020	
	" " " FLR " LPCS	0,0713	-0,0079	
	Total secundário via FLR	-0,0124	0,1112	
	Ind. secundário via MAT via LNVG	-0,0120	0,0086	
	" " " MAT " LNSV	-0,0062	0,0921	
	" " " MAT " LPCS	0,0126	0,0008	
	Total secundário via MAT	-0,0056	0,1015	
	Ind. secundário via PPL via LNVG	0,2552	0,0238	
	" " " PPL " LNSV	0,0699	0,0011	
	" " " PPL " LPCS	0,0237	-0,0005	
	Total secundário via PPL	0,3488	0,0244	
	Ind. secundário via INC via LNVG	-0,0324	0,2443	
	" " " INC " LNSV	-0,0083	0,0967	
	" " " INC " LPCS	0,0039	-0,0319	
	Total secundário via INC	-0,0368	0,3091	
	Total (r_{y7})	0,2494	0,6011	
	INC	Indireto primário via LNVG	0,5377	0,4591
		" " " LNSV	0,1386	0,1817
		" " " LPCS	-0,0656	-0,0600
Total primário		0,6107	0,5808	
Ind. secundário via FLR via LNVG		-0,0027	0,1473	
" " " FLR " LNSV		-0,0105	-0,0024	
" " " FLR " LPCS		0,0112	-0,0097	
Total secundário via FLR		-0,0020	0,1352	
Ind. secundário via MAT via LNVG		0,1440	0,0140	
" " " MAT " LNSV		0,0742	0,1505	
" " " MAT " LPCS		-0,1510	0,0013	
Total secundário via MAT		0,0672	0,1658	
Ind. secundário via PPL via LNVG		-0,0210	-0,1018	
" " " PPL " LNSV		-0,0057	-0,0048	
" " " PPL " LPCS		0,0061	0,0023	
Total secundário via PPL		-0,0206	-0,1043	
Ind. secundário via STD via LNVG		0,0076	0,0967	
" " " STD " LNSV		0,0010	-0,0423	
" " " STD " LPCS		-0,0059	-0,0251	
Total secundário via STD		0,0027	0,0293	
Total (r_{y8})	0,6580	0,8068		

correlacionada com a soma do logaritmo de número de vagens ($\log X_1$ ou LNVG) e logaritmo de número de sementes por vagem ($\log X_2$ ou LNSV), gerando multicolinearidade perfeita. Considerando-se também os diagramas apresentados nas Figuras 6 e 7 foram estimados:

a) os efeitos diretos e indiretos das variáveis primárias LNSM, LNVG, LNSV e LPCS, respectivamente, logaritmo do número de sementes, logaritmo do número de vagens, logaritmo do número de sementes por vagem e logaritmo do peso médio de 100 sementes, sobre a variável básica LPRD (logaritmo da produção de grãos);

b) os efeitos diretos e indiretos das variáveis secundárias - número de dias até o florescimento (FLR), número de dias até a maturação (MAT), peso da palhada (PPL), estande final (STD), e índice de colheita (INC), sobre a variável explicativa primária LNSM; e

c) os efeitos indiretos primários e secundários das cinco variáveis explicativas secundárias sobre a variável básica (LPRD) através da variável explicativa primária LNSM.

As duas matrizes de correlações fenotípicas entre as variáveis explicativas correspondentes a cada um dos experimentos foram submetidas ao diagnóstico de multicolinearidade através dos métodos já descritos. Os resultados desse exame são apresentados no Quadro 18. Para ambos os experimentos, pode-se verificar que a introdução da quarta variável, gerada a partir da combinação de outras duas, provocou multicolinearidade intensa nas matrizes de correlações. Conforme classificação de MONTGOMERY e PECK (1981), estas matrizes estão sendo afetadas por multicolinearidade severa, pois ambas apresentam número de condição (NC) maior que 1000. Observa-se que a multicolinearidade presente na matriz correspondente ao Experimento 2 mostra-se mais intensa apesar das variáveis tomadas no Experimento 1

QUADRO 18 - Resultados dos Testes para Diagnóstico de Multicolinearidade Aplicados às Matrizes de Correlações Fenotípicas Correspondentes a Quatro Variáveis Explicativas Primárias, para os Experimentos 1 e 2

Item observado	Experimento 1	Experimento 2
Maior Correlação	0,9890 (LNVG x LNSV)	0,9645 (LNVG x LNSV)
Menor Correlação	-0,8723 (LPCS x LNSM)	-0,6948 (LNVG x LNSM)
Número de VIF's ≥ 10	3	3
Determinante	$5,7141 \times 10^{-6}$	$6,8987 \times 10^{-6}$
Autovalores		
Número de Condição	[3,59342 0,29052 0,11601 0,000005]	[3,12369 0,47413 0,40216 0,000001]
Valor Singular	76164,2800	269691,6000
Índice de Condição	[3,59342 0,29052 0,11601 0,000005]	[3,12369 0,47413 0,40216 0,000001]
VIF _k	[1,0000 12,3689 30,9751 71868,4000]	[1,0000 6,5883 7,7673 312369,0000]
Proporções de Decomposição da Variância dos Efeitos Diretos	[4899,92 9800,59 6505,56 1,76]	[18426,64 20415,93 47498,70 1,09]
	[0,00 0,00 0,00 0,03]	[0,00 0,00 0,00 0,05]
	[0,00 0,00 0,00 2,41]	[0,00 0,00 0,00 1,45]
	[0,00 0,00 0,01 1,77]	[0,00 0,00 0,00 0,34]
	[21182,07 21191,45 21177,19 2,19]	[86331,87 86329,58 86336,05 344,00]

LNVG: logaritmo do número de vagens por planta;

LNSV: " " " sementes por vagem;

LPCS: " " " peso médio de 100 sementes;

LNSM: " " " número de sementes por planta.

apresentarem correlações simples mais elevadas. Evidencia-se, assim, que a ocorrência de correlações altas não constitui condição necessária para a existência de multicolinearidade conforme KMENTA (1971).

Para estabelecer padrões comparativos, os coeficientes de trilha foram estimados, primeiramente, pelo método dos mínimos quadrados e, a seguir, pelo método de cristas. Nos Quadros 19 e 20, são apresentadas as estimativas dos efeitos diretos e indiretos das variáveis primárias sobre a variável básica, para os dois experimentos, estimados por mínimos quadrados e por cristas, respectivamente. Observa-se, no Quadro 19, que os VIF's atingiram valores extremamente altos, o que já era de se esperar em decorrência da intensa multicolinearidade presente nas matrizes de correlações entre as variáveis primárias, nos dois experimentos, como visto no Quadro 18.

Verifica-se que a multicolinearidade afeta também as estimativas dos próprios efeitos diretos e indiretos das variáveis. Isso pode ser observado comparando-se, por exemplo, as estimativas dos efeitos das variáveis LNVG e LNSV em relação ao Experimento 2, obtidas sob colinearidade fraca, como apresentadas no Quadro 8, com as estimativas desses mesmos efeitos obtidas sob multicolinearidade intensa, como apresentadas no Quadro 19. Como visto, sob multicolinearidade severa, essas estimativas sofreram alterações tanto em magnitude como em sinal.

Nos Quadros 21, 22, 23 e 24, são apresentadas as estimativas dos efeitos das variáveis secundárias sobre a variável LNSM, para os dois experimentos, obtidas por mínimos quadrados e por cristas. São juntamente apresentadas as estimativas dos fatores de inflação da variância (VIF's) associados aos efeitos.

QUADRO 19 - Estimativas dos Efeitos Diretos e Indiretos das Variáveis Explicativas Primárias sobre a Variável Básica, Estimados pelo Método dos Mínimos Quadrados, Fatores de Inflação da Variância (VIF's) Associados, Coeficiente de Determinação e Efeito da Variável Residual, Para os Experimentos 1 e 2

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNSM	Direto sobre LPRD	0,5130	12746,02	3,9853	50427,21
	Indireto via LNVG	0,7107	7400,38	-2,1103	28561,11
	" LNSV	0,2246	726,59	-0,6936	3082,67
	" LPCS	-0,5137	2,90	-0,2036	0,94
	Total	0,9346		0,9777	
LNVG	Direto sobre LPRD	0,7186	7565,92	-2,1880	30702,28
	Indireto via LNSM	0,5074	12467,15	3,8438	46910,43
	" LNSV	0,2063	613,04	-0,5164	1708,51
	" LPCS	-0,4801	2,53	-0,1945	0,85
	Total	0,9522		0,9449	

Continua...

QUADRO 19, Cont.

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNSV	Direto sobre LPRD	0,2489	892,46	-0,9029	5223,70
	Indireto via LNSM	0,4629	10377,11	3,0615	29758,70
	" LNVG	0,5956	5197,10	-1,2513	10041,78
	" LPCS	-0,5453	3,27	-0,1609	0,58
	Total	0,7621		0,7464	
LPCS	Direto sobre LPRD	0,6251	4,30	0,2931	1,94
	Indireto via LNSM	-0,4216	8608,09	-2,7690	24343,59
	" LNVG	-0,5519	4462,56	1,4522	13524,28
	" LNSV	-0,2171	679,08	0,4956	1573,85
	Total	-0,5655		-0,5281	
Coefficiente de determinação		0,9999			
Efeito da variável residual		0,0086	1,0002		

LNSM: logaritmo do número de sementes;

LNVG: logaritmo do número médio de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes.

QUADRO 20 - Estimativas dos Efeitos Diretos e Indiretos das Variáveis Explicativas Primárias sobre a Variável Básica Estimados pelo Método de Cristas, Fatores de Inflação da Variância Associados (VIF's), Coeficiente de Determinação e Efeito da Variável Residual, Para os Experimentos 1 e 2

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNSM	Direto sobre LPRD	0,5011	9,76	0,4649	13,25
	Indireto via LNVG	0,5847	5,94	0,4850	7,44
	" " LNSV	0,1565	3,35	0,1701	1,58
	" " LPCS	-0,3453	1,98	-0,1656	0,84
	Total	0,9346		0,9777	
LNVG	Direto sobre LPRD	0,5912	7,02	0,5029	8,81
	Indireto via LNSM	0,4956	8,26	0,4484	11,18
	" " LNSV	0,1438	2,83	0,1266	0,88
	" " LPCS	-0,3227	1,73	-0,1582	0,76
	Total	0,9522		0,9449	

Continua...

QUADRO 20, Cont.

Variáveis	Efeito	Experimento 1		Experimento 2	
		Estimativa	VIF	Estimativa	VIF
LNSV	Direto sobre LPRD	0,1735	4,76	0,2214	2,96
	Indireto via LNSM	0,4521	6,87	0,3572	7,09
	" " LNVG	0,4900	4,17	0,2876	2,61
	" " LPCS	-0,3665	2,23	-0,1308	0,52
	Total	0,7621		0,7464	
LPCS	Direto sobre LPRD	0,4201	3,39	0,2383	1,91
	Indireto via LNSM	-0,4118	5,70	-0,3230	5,80
	" " LNVG	-0,4540	3,58	-0,3338	3,52
	" " LNSV	-0,1513	3,13	-0,1215	0,81
	Total	-0,5655		-0,5281	
Coeficiente de determinação		0,9258		0,9691	
Efeito da variável residual		0,2723		0,1757	

LNSM: logaritmo do número de sementes;

LNVG: logaritmo do número médio de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes.

QUADRO 21 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre as Variável Primária LNSM, Estimados pelo Método dos Mínimos Quadrados e Fatores de Inflação da Variância (VIF's) Associados, Para o Experimento 1

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSM	0,1199	1,8083
	Indireto via MAT	-0,0404	0,1963
	" " PPL	-0,1590	0,0807
	" " STD	0,0927	0,6748
	" " INC	-0,0606	0,0280
	Total	-0,0474	
MAT	Direto sobre LNSM	-0,1743	3,6527
	Indireto via FLR	0,0278	0,0972
	" " PPL	0,0511	0,0083
	" " STD	-0,0109	0,0093
	" " INC	-0,5445	2,2552
	Total	-0,6507	
PPL	Direto sobre LNSM	0,6391	1,3033
	Indireto via FLR	-0,0298	0,1119
	" " MAT	-0,0139	0,0234
	" " STD	-0,0756	0,4485
	" " INC	-0,0259	0,0051
	Total	0,4938	
STD	Direto sobre LNSM	-0,1583	1,9685
	Indireto via FLR	-0,0702	0,6199
	" " MAT	-0,0120	0,0173
	" " PPL	0,3050	0,2969
	" " INC	-0,0398	0,0120
	Total	0,0247	
INC	Direto sobre LNSM	0,6598	3,3118
	Indireto via FLR	-0,0110	0,0153
	" " MAT	0,1438	2,4873
	" " PPL	-0,0251	0,0020
	" " STD	0,0095	0,0072
	Total	0,7770	
	Coefficiente de determinação	0,9320	
	Efeito da variável residual	0,2607	

QUADRO 22 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre a Variável Primária LNSM, Estimados pelo Método dos Mínimos Quadrados e Fatores de Inflação da Variância (VIF's) Associados, Para o Experimento 2

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSM	-0,1718	11,6880
	Indireto via MAT	0,0374	10,7616
	" " PPL	0,0638	0,0476
	" " STD	-0,0483	1,2674
	" " INC	-0,7618	5,1223
	Total	-0,8808	
MAT	Direto sobre LNSM	0,0403	12,4909
	Indireto via FLR	-0,1595	10,0699
	" " PPL	0,0839	0,0824
	" " STD	-0,0380	0,7825
	" " INC	-0,8042	5,7091
	Total	-0,8775	
PPL	Direto sobre LNSM	0,3531	1,4590
	Indireto via FLR	-0,0310	0,3812
	" " MAT	0,0096	0,7057
	" " STD	0,0062	0,0211
	" " INC	-0,3452	1,0519
	Total	-0,0073	
STD	Direto sobre LNSM	0,0687	2,5587
	Indireto via FLR	0,1209	5,7895
	" " MAT	-0,0223	3,8198
	" " PPL	0,0320	0,0120
	" " INC	0,4736	1,9800
	Total	0,6730	
INC	Direto sobre LNSM	0,8899	6,9907
	Indireto via FLR	0,1471	8,5642
	" " MAT	-0,0364	10,2010
	" " PPL	-0,1370	0,2195
	" " STD	0,0365	0,7247
	Total	0,9002	
	Coefficiente de determinação	0,9608	
	Efeito da variável residual	0,1981	

QUADRO 23 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre a Variável Primária LNSM, Estimados pelo Método de Cristas e Fatores de Inflação da Variância (VIF's) Associados, para o Experimento 1

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSM	0,1218	1,7156
	Indireto via MAT	-0,0527	0,1337
	" " PPL	-0,1437	0,0718
	" " STD	0,0705	0,5532
	" " INC	-0,0526	0,0195
	Total	-0,0474	
MAT	Direto sobre LNSM	-0,2273	2,8860
	Indireto via FLR	0,0282	0,0795
	" " PPL	0,0462	0,0074
	" " STD	-0,0083	0,0077
	" " INC	-0,4721	1,5683
	Total	-0,6507	
PPL	Direto sobre LNSM	0,5778	1,3461
	Indireto via FLR	-0,0303	0,0916
	" " MAT	-0,0182	0,0159
	" " STD	-0,0574	0,3677
	" " INC	-0,0225	0,0036
	Total	0,4938	
STD	Direto sobre LNSM	-0,1203	1,8717
	Indireto via FLR	-0,0713	0,5071
	" " MAT	-0,0157	0,0118
	" " PPL	0,2758	0,2644
	" " INC	-0,0345	0,0084
	Total	0,0247	
INC	Direto sobre LNSM	0,5721	2,6711
	Indireto via FLR	-0,0112	0,0125
	" " MAT	0,1876	1,6945
	" " PPL	-0,0227	0,0018
	" " STD	0,0073	0,0059
	Total	0,7770	
Coeficiente Efeito da	de determinação	0,8689	
	variável residual	0,3620	

QUADRO 24 - Efeitos Diretos e Indiretos das Variáveis Secundárias sobre a Variável Primária LNSM, Estimados pelo Método de Cristas e Fatores de Inflação da Variância (VIF's) Associados, para o Experimento 2

Variável	Efeito	Estimativa	VIF
FLR	Direto sobre LNSM	-0,1684	5,7062
	Indireto via MAT	-0,1529	4,1784
	" " PPL	0,0480	0,0391
	" " STD	-0,0766	0,8740
	" " INC	-0,5177	2,8236
	Total	-0,8808	
MAT	Direto sobre LNSM	-0,1647	5,6447
	Indireto via FLR	-0,1563	4,2239
	" " PPL	0,0632	0,0678
	" " STD	-0,0602	0,5396
	" " INC	-0,5465	3,1471
	Total	-0,8775	
PPL	Direto sobre LNSM	0,2660	1,3960
	Indireto via FLR	-0,0304	0,1599
	" " MAT	-0,0392	0,2740
	" " STD	0,0099	0,0145
	" " INC	-0,2346	0,5798
	Total	-0,0073	
STD	Direto sobre LNSM	0,1088	2,0536
	Indireto via FLR	0,1185	2,4284
	" " MAT	0,0911	1,4831
	" " PPL	0,0241	0,0099
	" " INC	0,3219	1,0915
	Total	0,6730	
INC	Direto sobre LNSM	0,6048	4,4852
	Indireto via FLR	0,1441	3,5923
	" " MAT	0,1489	3,9607
	" " PPL	-0,1032	0,1805
	" " STD	0,0579	0,4997
	Total	0,9002	
Coefficiente de determinação		0,9086	
Efeito da variável residual		0,3023	

Ao se estimar os coeficientes de trilha, pelo método de cristas, procurou-se determinar, como preconizado, um valor mínimo para a constante k , através do exame do gráfico de cristas estabelecido em cada caso. A estimação dos efeitos das variáveis primárias sobre a variável básica foi feita utilizando-se $k=0,0500$ e $k=0,0750$ para as análises referentes aos Experimentos 1 e 2, respectivamente. Comparando-se esses resultados, os quais são apresentados no Quadro 20, com aqueles obtidos pelo método dos mínimos quadrados apresentados no Quadro 19, constata-se que a estimação em cristas proporcionou redução dos VIF's em níveis aceitáveis.

As estimativas dos coeficientes de trilha obtidas pelo método de cristas foram mais semelhantes às obtidas sob multicolinearidade fraca, quando comparadas com as estimativas proporcionadas pelo método dos mínimos quadrados. Isso pode ser constatado comparando-se os efeitos diretos e indiretos das variáveis LNVG e LNSV obtidos sob multicolinearidade fraca e severa, como apresentados nos Quadros 8 e nos Quadros 19 e 20, respectivamente. Em relação aos dados do Experimento 2, por exemplo, verifica-se que o efeito direto de LNVG sobre LPRD, sob multicolinearidade fraca, foi 0,9215. Sob multicolinearidade severa, essa estimativa caiu para -2,1880, quando obtida por mínimos quadrados, enquanto o método de cristas proporcionou o valor 0,5029, que é bem mais próximo ao valor inicial, e o seu sinal é concordante com aquele primeiro valor, estando dentro da expectativa biológica.

Para a estimação dos efeitos diretos e indiretos das variáveis secundárias sobre LNSM, pelo método em cristas, tomou-se $k=0,07692$ e $k=0,07885$, para os Experimentos 1 e 2, respectivamente, pelo exame dos gráficos de cristas. As estimativas desses efeitos, juntamente com os respectivos VIF's, são apresentadas nos Quadros 23 e 24.

Comparando-se esses resultados com aqueles obtidos pelo método dos mínimos quadrados, apresentados nos Quadros 21 e 22, constata-se mais uma vez a eficiência do método em cristas em proporcionar sensível redução nos fatores de inflação da variância (VIF's) associados às estimativas dos efeitos.

Finalmente, foram estimados os efeitos indiretos primários e secundários das variáveis explicativas secundárias sobre a variável básica (LPRD), através da variável explicativa primária LNSM, utilizando-se o método de cristas. Esses resultados são apresentados no Quadro 25. Ao se comparar estas estimativas com aquelas proporcionadas por mínimos quadrados apresentadas no Quadro 26, constata-se novamente que o método de cristas proporciona estimativas de coeficientes de trilha não muito diferentes daquelas proporcionadas pelo método de mínimos quadrados.

Conclui-se, assim, que a teoria desenvolvida para análise de regressão em cristas, recomendável para situações de multicolinearidade forte, pode ser aplicada perfeitamente à análise de trilha. Como visto, a multicolinearidade severa torna menos confiáveis os efeitos diretos e indiretos em que são desdobradas as correlações entre as variáveis na análise de trilha. O problema pode ser atenuado pela aplicação do método proposto.

4.5. Ganho de Seleção

Diferentes critérios de seleção podem ser utilizados para maximizar o êxito de um programa de melhoramento. Informações de vários caracteres são normalmente utilizadas quando se deseja obter ganhos em todos eles, simultaneamente. Nesse contexto, algumas

QUADRO 25 - Efeitos Indiretos Primários e Secundários das Variáveis Secundárias sobre a Variável Básica, Via LNSM, Estimados pelo Método de Cristas, para os Experimentos 1e 2

Variável	Efeito	Exp. 1	Exp. 2
FLR	Indireto primário	0,0120	-0,0783
	Ind. secundário via MAT, via LNSM	-0,0264	-0,0711
	" " " PPL, via LNSM	-0,0720	0,0223
	" " " STD, via LNSM	0,0353	-0,0356
	" " " INC, via LNSM	-0,0263	-0,2407
MAT	Indireto primário	-0,1139	-0,0766
	Ind. secundário via FLR, via LNSM	0,0141	-0,0727
	" " " PPL, via LNSM	0,0232	0,0294
	" " " STD, via LNSM	-0,0042	-0,0280
	" " " INC, via LNSM	-0,2365	-0,2541
PPL	Indireto primário	0,2895	0,1237
	Ind. secundário via FLR, via LNSM	-0,0152	-0,0141
	" " " MAT, via LNSM	-0,0091	-0,0182
	" " " STD, via LNSM	-0,0288	0,0046
	" " " INC, via LNSM	-0,0173	0,1497
STD	Indireto primário	-0,0603	0,0506
	Ind. secundário via FLR, via LNSM	-0,0357	0,0551
	" " " MAT, via LNSM	-0,0078	0,0423
	" " " PPL, via LNSM	0,1382	0,0112
	" " " INC, via LNSM	-0,0113	-0,1091
INC	Indireto primário	0,2866	0,2812
	Ind. secundário via FLR, via LNSM	-0,0056	0,0670
	" " " MAT, via LNSM	0,0940	0,0692
	" " " PPL, via LNSM	-0,0114	-0,0480
	" " " STD, via LNSM	0,0036	0,0269

QUADRO 26 - Efeitos Indiretos Primários e Secundários das Variáveis Secundárias sobre a Variável Básica, Via LNSM, Estimados pelo Método dos Mínimos Quadrados, para os Experimentos 1 e 2

Variável	Efeito	Exp. 1	Exp. 2
FLR	Indireto primário	0,0615	-0,6848
	Ind. secundário via MAT, via LNSM	0,0207	0,1489
	" " " PPL, via LNSM	-0,0816	0,2542
	" " " STD, via LNSM	0,0475	-0,1926
	" " " INC, via LNSM	-0,0311	3,0359
MAT	Indireto primário	-0,0894	0,1604
	Ind. secundário via FLR, via LNSM	0,0143	-0,6356
	" " " PPL, via LNSM	0,0262	0,3345
	" " " STD, via LNSM	-0,0056	-0,1514
	" " " INC, via LNSM	-0,2793	3,2050
PPL	Indireto primário	0,3279	1,4073
	Ind. secundário via FLR, via LNSM	-0,0153	-0,1237
	" " " MAT, via LNSM	-0,0071	0,0381
	" " " STD, via LNSM	-0,0388	0,0248
	" " " INC, via LNSM	-0,0204	1,8875
STD	Indireto primário	-0,0812	0,2737
	Ind. secundário via FLR, via LNSM	-0,0360	0,4819
	" " " MAT, via LNSM	-0,0061	-0,0887
	" " " PPL, via LNSM	0,1565	0,1276
	" " " INC, via LNSM	-0,0133	1,3757
INC	Indireto primário	0,3385	3,5466
	Ind. secundário via FLR, via LNSM	-0,0057	0,5862
	" " " MAT, via LNSM	0,0738	-0,1450
	" " " PPL, via LNSM	-0,0129	-0,5459
	" " " STD, via LNSM	0,0049	0,1457

modalidades de seleção sobressaem-se por serem de fácil execução e eficientes, como a seleção direta no próprio caráter considerado, e a seleção indireta, com base em resposta correlacionada. No entanto, no caso do feijoeiro, o ganho na produção de grãos através da resposta correlacionada não deve superar o ganho proporcionado pela seleção direta nesse caráter, conforme resultados obtidos por RAMALHO et al. (1979).

Em algumas situações, a seleção direta e indireta pode não atender ao objetivo geral do melhoramento. É o caso, por exemplo, de se trabalhar com componentes primários da produção quando alguns deles são inversamente correlacionados entre si. Em situações como esta, pode-se não obter os ganhos em alguns caracteres considerados como principais sem que ocorra perda em outros, também importantes, fazendo-se necessária a utilização de outro critério, como a seleção por meio de índices.

Assim, foram estimados primeiramente os ganhos proporcionados pela seleção direta e indireta. A seguir, procurou-se obter as estimativas dos ganhos pela seleção com base no índice clássico de SMITH (1936) e HAZEL (1943), introduzindo-se correção para os efeitos da multicolinearidade presente. Os resultados foram então comparados.

4.5.1. Seleção Direta e Seleção Indireta

Considerou-se a seleção com base na produção de grãos e em seus três componentes primários. Como no presente trabalho, as observações correspondentes a esses caracteres foram previamente transformadas em logaritmos, optou-se por utilizar os dados assim transformados, tendo-se então logaritmo do número de vagens (LNVG),

logaritmo do número de sementes por vagem (LNSV), logaritmo do peso médio de 100 sementes (LPCS) e logaritmo da produção de grãos (LPRD). As estimativas dos ganhos pela seleção direta e indireta de 40% dos materiais em estudo, com base nesses quatro caracteres, são apresentadas no Quadro 27. Observa-se que LNVG apresentou a maior estimativa de ganho esperado pela seleção direta nos dois experimentos. A segunda maior estimativa de ganho esperado no Experimento 1 foi apresentado pela produção de grãos. No Experimento 2, esse caráter apresentou o maior ganho esperado pela seleção direta. O peso médio de 100 sementes apresentou a menor estimativa de ganho esperado pela seleção direta nos dois experimentos, apesar de estar associado à herdabilidade mais alta no Experimento 1. No Experimento 2, esse caráter apresentou herdabilidade baixa. Sua alta herdabilidade, no Experimento 1, pode ter sido condicionada apenas pela alta precisão experimental, como discutido no item 4.2. Assim, sob precisão experimental mais baixa, como aquela verificada no Experimento 2, cujo coeficiente de variação (CV%) foi relativamente superior ao apresentado pelo Experimento 1, evidencia-se que o peso médio de 100 sementes não é um caráter favorável à seleção. O estudo de ganhos de seleção vem confirmar essa particularidade.

A seleção para produção de grãos com base em número de vagens mostrou-se eficiente nos dois experimentos, igualando à seleção direta, sem suplantá-la. A seleção direta para produção de grãos também mostrou-se eficiente nos dois experimentos. Verifica-se que, em nenhum dos casos, a resposta correlacionada superou o ganho obtido pela seleção direta, como era de se esperar, pois, para estimação da resposta correlacionada, foram utilizadas as informações do diferencial de seleção indireta.

QUADRO 27 - Estimativas da Herdabilidade (h^2) e Ganhos de Seleção (GS) Obtidos pela Seleção Direta e Indireta sobre Quatro Caracteres, Aplicada ao Material Experimental em Estudo, para os Experimentos 1 e 2. Viçosa, MG. 1991/92

Caráter sob seleção	h^2		LNVG		LNSV		LPCS		LPRD		Total		
	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	Exp. 1	Exp. 2	
LNVG	93,00	93,44	GS	0,2045	0,2062	0,0601	0,0756	-0,0801	-0,0100	0,1789	0,2439		
			GS %	23,78	23,19	9,53	13,49	-5,82	-0,73	20,61	29,66		
LNSV	95,47	86,07	GS	0,1683	0,1763	0,0656	0,0773	-0,0913	-0,0122	0,1394	0,2104	48,10	65,61
			GS %	19,56	19,95	10,39	13,79	-6,63	-0,89	16,06	25,59		
LPCS	98,98	26,82	GS	-0,1629	-0,1462	-0,0630	-0,0405	0,1250	0,0199	-0,1018	-0,1162		
			GS%	-18,93	-0,1645	-9,99	-7,22	9,08	1,45	-11,73	-14,13		
LPRD	88,60	89,87	GS	0,2045	0,2062	0,0601	0,0756	-0,0801	-0,0100	0,1789	0,2439	-31,57	-36,35
			GS%	23,78	0,2062	9,53	13,49	-5,82	-0,73	20,61	29,66	48,10	65,61

h^2 : herdabilidade do caráter, no sentido amplo;

LNVG: logaritmo do número de vagens por planta da parcela;

LNSV: logaritmo do número médio de sementes por vagem;

LPCS: logaritmo do peso médio de 100 sementes;

LPRD: logaritmo da produção de grãos.

Número de vagens e número de sementes por vagem são caracteres fortemente correlacionados com a produção de grãos. Assim, é desejável que sejam selecionados materiais genotípicos que apresentem maior número de vagens por planta e maior número de sementes por vagem. Porém, esses caracteres são correlacionados negativamente com o peso médio de 100 sementes. Assim, ao se aplicar a seleção no sentido de aumentar o número de vagens e o número de sementes por vagem, obtém-se, também, uma redução no tamanho das sementes, o que pode ser indesejável. Essa dificuldade pode ser constatada pelo exame dos resultados apresentados no Quadro 27. ADAMS (1967) apresenta uma explicação para o problema afirmando que este ocorre porque componentes primários da produção, como número de vagens, número de sementes por vagem e peso médio de 100 sementes, competem por uma quantidade limitada de nutrientes inorgânicos e materiais sintetizados, liberados por uma mesma unidade nutricional, à medida em que as fases de desenvolvimento se sucedem na planta. Define-se, resumidamente, unidade nutricional como sendo uma estrutura composta por algumas folhas e vasos, responsável pelo armazenamento, síntese e distribuição de nutrientes, relacionada a uma determinada parte da planta.

A questão discutida a respeito do inter-relacionamento entre os componentes primários da produção de grãos evidencia a limitação da seleção direta e indireta via resposta correlacionada. Como visto, utilizando-se estas técnicas para os dados em questão, não se tem como obter ganhos para aumento do número de vagens e sementes por vagem, sem provocar redução no tamanho das sementes. Contudo, esse problema pode ser evitado, ao menos parcialmente, com a utilização de índices de seleção, como será visto a seguir.

4.5.2. Seleção com Base no Índice Clássico de SMITH (1936) e HAZEL (1943)

Como citado anteriormente o índice clássico consiste em um vetor de ponderação (b) dos caracteres de importância econômica obtidos pela relação

$$\hat{b} = P^{-1} G a,$$

em que

P: matriz de variâncias e covariâncias fenotípicas;

G: matriz de variâncias e covariâncias genotípicas; e

a: vetor de pesos econômicos.

Considerando então a produção de grãos (LPRD) e seus componentes primários (LNVG, LNSV e LPCS), procurou-se estimar um índice que proporcionasse a maximização dos ganhos para esses caracteres, selecionando-se 40% dos materiais em estudo. Verifica-se, contudo, que as observações correspondentes ao número de vagens, número de sementes por vagem e peso de sementes são perfeitamente correlacionadas com a produção de grãos. Conseqüentemente, a matriz de variâncias e covariâncias fenotípicas (P) entre essas variáveis é singular, não possuindo inversa única. Logo, o índice não pode ser obtido. As matrizes P_1 e P_2 correspondentes aos Experimentos 1 e 2, respectivamente, são apresentadas a seguir:

$$P_1 = \begin{bmatrix} 0,04700 & 0,01335 & -0,02022 & 0,04014 \\ 0,01335 & 0,00552 & -0,00787 & 0,01101 \\ -0,02022 & -0,00787 & 0,01474 & -0,01335 \\ 0,04014 & 0,01101 & -0,01335 & 0,03780 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 0,06186 & 0,01456 & -0,01299 & 0,06342 \\ 0,01456 & 0,01048 & -0,00442 & 0,02062 \\ -0,01299 & -0,00442 & 0,00619 & -0,01122 \\ 0,06342 & 0,02062 & -0,01122 & 0,07284 \end{bmatrix}$$

Para ser em singulares, as matrizes P , relativas aos dois experimentos, devem ter determinantes nulos, isto é, devem apresentar $\det(P_1) = \det(P_2) = 0$. Utilizando o programa para análises genéticas e estatísticas - GENES, obteve-se $\det(P_1) = -5,5849 \times 10^{-12}$ e $\det(P_2) = 4,1868 \times 10^{-11}$. Esses valores tendem para zero, como era de se esperar.

Apesar de os determinantes serem da ordem de 10^{-12} a 10^{-11} , verifica-se que por uma questão técnica com base em aproximações numéricas, o programa GENES não reconhece a singularidade das matrizes P_1 e P_2 , conseguindo assim invertê-las e possibilitando a estimação dos respectivos índices de seleção. Contudo, por razões já expostas, em consequência da multicolinearidade, os índices assim estimados não são confiáveis.

Desta forma, optou-se por estimar um novo índice, corrigindo-se os efeitos adversos da multicolinearidade a partir do sistema

$$(R_F + kI)\hat{p} = D_F^{-1} G D_G^{-1} q$$

obtido pela decomposição de $\hat{P}b = Ga$, como visto.

As matrizes R_{F1} e R_{F2} , correspondentes aos Experimentos 1 e 2, respectivamente, são apresentadas a seguir. Os resultados do diagnóstico de multicolinearidade dessas matrizes são apresentados no

Quadro 28. Por esses resultados, pode-se constatar que essas matrizes apresentam multicolinearidade severa. Como seus determinantes aproximam-se de zero, ambas tendem à singularidade.

$$R_{F1} = \begin{bmatrix} 1,0000 & 0,8288 & -0,7680 & 0,9522 \\ 0,8288 & 1,0000 & -0,8723 & 0,7621 \\ -0,7680 & -0,8723 & 1,0000 & -0,5655 \\ 0,9522 & 0,7621 & -0,5655 & 1,0000 \end{bmatrix}$$

$$R_{F2} = \begin{bmatrix} 1,0000 & 0,5719 & -0,6637 & 0,9449 \\ 0,5719 & 1,0000 & -0,5489 & 0,7464 \\ -0,6637 & -0,5489 & 1,0000 & -0,5281 \\ 0,9449 & 0,7464 & -0,5281 & 1,0000 \end{bmatrix}$$

O processo proposto para atenuar os efeitos da multicolinearidade consiste em somar uma quantidade constante (k) à diagonal da matriz de correlações fenotípicas R_F . Assim, esta matriz afasta-se da condição de quase singularidade, possibilitando a estimação do índice.

Os valores de k foram escolhidos pelo exame de gráficos de cristas, os quais foram construídos plotando-se os valores de \hat{p} em função dos valores de k no intervalo $0 \leq k \leq 1$. Procurou-se encontrar um valor de k para o qual os coeficientes \hat{p} 's se mostrassem estabilizados. Procurou-se, também, evitar que o valor de k viesse a provocar redução substancial na correlação entre o índice e o agregado genotípico. Os valores dessa correlação, em função de k , são apresentados no Quadro 29, para os dois experimentos.

Optou-se por tomar $k=0,0500$ para as estimações correspondentes aos dois experimentos. Pode-se observar, no Quadro 29, que tanto para o Experimento 1 como para o Experimento 2, esse valor não implica em redução substancial no coeficiente de correlação entre o índice e o agregado genotípico.

QUADRO 28 - Resultados dos Testes para Diagnóstico de Multicolinearidade, Aplicados às Matrizes de Correlações Fenotípicas entre a Produção de Grãos e seus Componentes Primários

Item observado	R_{F1}	R_{F2}
Maior Correlação	0,9522 (LNVG x LPRD)	0,9449 (LNVG x LPRD)
Menor Correlação	-0,5655 (LPCS x LPRD)	-0,6637 (LNVG x LPCS)
Número de VIF's ≥ 10	4	4
Determinante	$6,9381 \times 10^{-6}$	$2,0984 \times 10^{-5}$
Autovalores	[3,3815 0,50008 0,11974 0,000003]	[3,01845 0,52707 0,45445 0,000003]
Número de Condição	98606,5200	103998,3000
Valor Singular	[3,3815 0,50008 0,11974 0,000003]	[3,01845 0,52707 0,45445 0,000003]
Índice de Condição	[1,0000 6,7619 28,2403 112716,6600]	[1,0000 5,7268 6,6420 100615,0000]
VIF _k	[6913,65 11549,47 220,42 10499,32]	[10095,83 6134,47 1651,19 16577,08]
Proporções de Decomposição	0,00 0,00 0,00 0,00	0,00 0,00 0,00 0,00
da Variância dos Efeitos Diretos	0,00 0,00 0,01 0,00	0,00 0,00 0,00 0,00
	0,00 0,00 0,006 0,00	0,00 0,00 0,00 0,00
	29161,78 29162,49 28788,33 29168,31	34450,88 34444,16 34426,22 34453,41

LNVG: logaritmo do número de vagens por planta;

LPCS: " do peso médio de 100 grãos;

LPRD: " da produção de grãos.

QUADRO 29 - Valores da Correlação entre o Índice e o Agregado Genotípico (r_{IH}), em Função dos Valores Assumidos pela Constante k , no Intervalo $0 \leq k \leq 1$, para os Experimentos 1 e 2

k	rIH		k	rIH	
	Exp. 1	Exp. 2		Exp. 1	Exp. 2
0,0000	0,8993	0,9792	0,4000	0,5318	0,7944
0,0500	0,8202	0,9461	0,5000	0,4875	0,7623
0,1000	0,7568	0,9178	0,6000	0,4514	0,7338
0,1500	0,7041	0,8927	0,7000	0,4213	0,7073
0,2000	0,6593	0,8701	0,8000	0,3958	0,6828
0,2500	0,6208	0,8492	0,9000	0,3739	0,6600
0,3000	0,5873	0,8298	1,0000	0,3547	0,6388

Uma das dificuldades mais comuns no processo de estimação de índices de seleção consiste na atribuição de pesos econômicos aos caracteres em estudo, pois, freqüentemente, não há como se obter os valores relativos dos mesmos. Ressalta-se, ainda, que o interesse primordial do melhorista é obter o maior ganho nos caracteres avaliados, considerados conjuntamente. Um índice ideal seria aquele capaz de proporcionar ganhos em todos os caracteres de interesse, ou que ao menos proporcionasse ganhos para a maioria dos caracteres. Em situações extremas, sob fortes correlações negativas, poder-se-ia tolerar a redução total do ganho, ou mesmo um mínimo de perdas em um ou mais caracteres. O que se deseja, enfim, é encontrar uma combinação ideal para os coeficientes de ponderação econômica dos caracteres, de modo a proporcionar um índice com as características desejadas. No presente trabalho, tomou-se o coeficiente de variação genético (CV_G) como referencial para o estabelecimento dos pesos econômicos, como utilizado por CRUZ (1990). Procurou-se identificar a combinação ideal dos coeficientes, variando-se os valores desses

coeficientes e verificando os ganhos de seleção obtidos em cada simulação. Assim, para os dados do Experimento 1, partindo-se da proporção aproximada 2:1:3, entre os CV_G 's correspondentes a LNVG, LNSV e LPCS, identificou-se a proporção 4:3:8 como sendo a que proporcionou os ganhos mais satisfatórios. Os resultados obtidos pela aplicação desse índice aos dados são apresentados no Quadro 30. Comparando-se esses resultados com aqueles apresentados no Quadro 27, proporcionados pela resposta correlacionada, torna-se evidente o potencial do índice de seleção em proporcionar ganhos totais satisfatórios, sem perda em nenhum dos caracteres considerados. Observa-se que, enquanto a seleção via resposta correlacionada proporcionou ganhos em número de vagens e número de sementes por vagem com redução considerável e simultânea do peso médio de 100 sementes, o índice estimado proporcionou ganhos nos dois primeiros caracteres, enquanto o tamanho das sementes era mantido praticamente intacto. Por esse índice, seriam selecionados os seguintes genótipos, em ordem decrescente de importância relativa: Ouro Negro, CNF 5547, RA 94, Ica Tundama, Ouro, Rio Tibagi, Diacol Andino, Ica Llanogrande x Ica Tundama e Rico.

Para a estimação do índice correspondente ao Experimento 2, utilizou-se também inicialmente a proporção entre os CV_G 's correspondentes a LNVG, LNSV e LPCS, que neste caso foi 9:6:1. Vários índices foram simulados e estimados com base em variações dessa proporção inicial, buscando-se a combinação ótima. Contudo, não se obteve um índice capaz de proporcionar ganhos em número de vagens e número de sementes por vagem, sem provocar redução no peso médio de 100 sementes. Procurou-se, então, identificar um índice que possibilitasse a obtenção de ganhos satisfatórios nos dois primeiros caracteres, com um mínimo de perdas no tamanho das sementes.

Constatou-se que o índice estimado com base em coeficientes genotípicos de ponderação equivalentes à proporção 2:2:1 atende a essas exigências. No Quadro 31, são apresentados os ganhos de seleção com base nesse índice. Comparando-se esses resultados com aqueles apresentados no Quadro 28, observa-se que a performance do índice assim estimado foi exatamente igual ao efeito da seleção indireta aplicada a número de vagens ou a produção de grãos.

QUADRO 30 - Resultados Proporcionados pelo Índice de Seleção sob Multicolinearidade, Considerando $k=0,0500$, para o Experimento 1

Caráter	a	q	b	p	GS	GS %
LNVG	4	0,4181	0,3387	0,0734	0,1171	13,6100
LNSV	3	0,1089	0,0078	0,0006	0,0269	4,2700
LPCS	8	0,4831	1,8061	0,2193	0,0001	0,0100
LPRD	0	0,0000	1,1322	0,2202	0,1367	15,2400
Ganho Total					0,2750	33,6700

a: valor econômico;

q: valor econômico padronizado;

b: coeficiente dos caracteres no índice de seleção;

p: coeficiente padronizado dos caracteres no índice de seleção;

GS : ganho de seleção.

Com base nesse índice, seriam selecionados os seguintes genótipos, em ordem decrescente de importância relativa: Ouro, Rico, Rio Tibagi, RA 94, CNF 5547, Roxo x Carioca, Ouro Negro e Carioca.

Observa-se que, neste caso, a utilização do índice de seleção não trouxe vantagens ao melhoramento. Contudo, esta situação foi condicionada por correlações negativas desfavoráveis entre as variáveis

consideradas, o que pode ser constatado observando-se os resultados apresentados no Quadro 7. Sob condições menos drásticas, a eficiência do índice de seleção ficou evidenciada pelos resultados obtidos em relação ao Experimento 1.

QUADRO 31 - Resultados Proporcionados pelo Índice de Seleção sob Multicolinearidade, Considerando $k=0,0500$, para o Experimento 2

Caráter	a	q	b	p	GS	GS %
LNVG	2	0,4808	0,8694	0,2162	0,2062	23,1900
LNSV	2	0,1899	0,9145	0,0936	0,0756	13,4900
LPCS	1	0,0407	-0,2649	-0,0995	-0,0100	-0,7300
LPRD	0	0,0000	0,7579	0,2045	0,2439	29,6600
Ganho Total					0,5157	65,6100

a: valor econômico

q: valor econômico padronizado

b: coeficiente dos caracteres no índice de seleção

p: coeficiente padronizado dos caracteres no índice de seleção

GS : ganho de seleção.

A multicolinearidade, como visto, pode interferir no processo de estimação do índice de seleção, tornando-o menos confiável. Porém, seus efeitos adversos podem ser identificados e removidos, ampliando assim o potencial de uso dessa metodologia no melhoramento genético.

variáveis foi feita pela utilização dos métodos descritos na literatura pertinente. Avaliaram-se os efeitos adversos da multicolinearidade sobre os coeficientes de trilha estimando-se os respectivos fatores de inflação da variância (VIF's) associados. Para reduzir esses efeitos os coeficientes de trilha foram também estimados por procedimento análogo ao método de regressão em cristas ou em cumeeira.

Propôs-se verificar a eficiência do índice de seleção em comparação com a seleção direta e indireta visando o aumento da produtividade de grãos. Também se avaliou a viabilidade de utilização do índice de seleção sob multicolinearidade intensa, utilizando-se um procedimento baseado na estimação em cristas ou em cumeeira.

5. RESUMO E CONCLUSÕES

Com o objetivo de verificar a adequação de matrizes de correlações para uso em análise de trilha e na estimação de índices de seleção, em presença de multicolinearidade, amostras de genótipos de feijão foram avaliadas em dois ensaios conduzidos em Viçosa-MG, durante o ano agrícola 1991/92, em relação a dez caracteres considerados de importância agronômica. As observações correspondentes a esses caracteres foram submetidas primeiramente à análise da variância. Assim foram também obtidas as estimativas dos parâmetros genéticos.

Realizou-se o estudo do inter-relacionamento entre os caracteres pela estimação dos respectivos coeficientes de correlação fenotípica, genotípica e de ambiente. Em seguida as correlações fenotípicas foram desdobradas em componentes de efeitos diretos e indiretos, em conformidade com diagramas causais previamente estabelecidos. A determinação do grau de condicionamento das matrizes de trabalho em consequência de multicolinearidade entre as

variáveis foi feita pela utilização dos métodos descritos na literatura pertinente. Avaliaram-se os efeitos adversos da multicolinearidade sobre os coeficientes de trilha estimando-se os respectivos fatores de inflação da variância (VIF's) associados. Para atenuar esses efeitos os coeficientes de trilha foram também estimados por procedimento análogo ao método de regressão em cristas ou em cumeeira.

Procurou-se verificar a eficiência do índice de seleção em comparação com a seleção direta e indireta, visando o aumento da produtividade de grãos. Avaliou-se também a viabilidade de utilização do índice de seleção sob multicolinearidade intensa, utilizando-se um procedimento baseado na estimação em cristas ou em cumeeira.

Os resultados das análises dos dados relativos aos dois experimentos possibilitaram as seguintes conclusões:

- a) sob precisão experimental moderada a alta, a amostra em estudo foi eficiente para discriminar genótipos favoráveis, com base nos caracteres avaliados, em programas de melhoramento do feijoeiro;
- b) as correlações entre os caracteres estudados foram determinadas principalmente pelos componentes genotípicos, pois, observou-se ligeira tendência de predominância das correlações genotípicas sobre as fenotípicas;
- c) alguns componentes primários da produção de grãos no feijoeiro, como o número médio de vagens por planta e número médio de sementes por vagem, destacaram-se por apresentar alta correlação com esse caráter. O mesmo pode ser verificado em relação ao índice de colheita, que é um caráter secundário. A princípio, esses caracteres poderiam ser utilizados em processos de seleção indireta para aumento na produção de grãos;

d) a decomposição dos efeitos dos componentes primários sobre a produção de grãos possibilitou constatar que a seleção indireta para incremento desse caráter é mais viável através do número médio de vagens por planta;

e) é possível a obtenção de ganhos em número médio de vagens por planta da parcela pela seleção com base em índice de colheita e número de dias até à maturação. Essa constatação pode ser de muita utilidade no melhoramento do feijoeiro, pois, sendo o número médio de vagens por planta o caráter a apresentar o maior coeficiente de correlação e o maior efeito direto sobre a produção de grãos, é desejável a resposta indireta sobre este, por meio da seleção em caracteres secundários de menor complexidade gênica;

f) a estimação dos coeficientes de trilha pelo método em cristas afetou os mesmos apenas ligeiramente quanto à magnitude, quando em comparação com os resultados obtidos pelo método dos mínimos quadrados. Contudo, em geral, a estimação em cristas proporcionou a redução dos fatores de inflação da variância, associados aos efeitos;

g) sob multicolinearidade intensa, na análise de trilha, os fatores de inflação da variância associados aos efeitos diretos e indiretos das variáveis atingiram valores extremamente altos. Especialmente, sob essa condição, constatou-se a eficiência do método de estimação em cristas, o qual proporcionou sensível redução na magnitude desses fatores;

h) pela estimação dos ganhos, constatou-se a viabilidade da seleção para a produção de grãos com base no número médio de vagens por planta, embora o ganho assim estimado não tenha ultrapassado aquele estimado através da seleção direta;

i) dada à forte associação negativa entre número de vagens e número de sementes por vagem em relação ao peso médio de 100 sementes, através da seleção direta e indireta, não foi possível obter ganhos para aumento do número de vagens e sementes por vagem sem provocar redução no tamanho das sementes;

j) o índice de seleção clássico torna possível a obtenção de ganhos em número de vagens e número de sementes por vagem, minimizando ou mesmo eliminando perdas no tamanho das sementes

l) o processo proposto para estimação de índices de seleção em cristas mostrou-se eficiente sob multicolinearidade; e

m) os resultados evidenciam a relevância dos problemas advindos da multicolinearidade. Quanto ao seu diagnóstico, observa-se que os métodos informais fornecem indicações apenas gerais, enquanto a combinação de alguns outros métodos possibilita a quantificação de sua intensidade e a identificação das variáveis envolvidas.

BIBLIOGRAFIA

ABRAHÃO, I.O. Melhoramento do feijoeiro. *Eragantia*, 19(10):156-61, 1960.

BIBLIOGRAFIA

ALMEIDA, C.M.V. C. Caracteres no estágio adulto e possibilidade de seleção precoce em híbridos de cacau [*Theobroma cacao* L.]. Piracicaba, USP/ESALQ, 1991. 194p. (Tese D.S.).

ADAMS, M.W. Basis of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*. *Crop. Sci.*, 7(5):505-10, 1967.

BELSLEY, D.A.; KUH, E.; WELCH, R.E. *Regression Diagnostics: Identifying data and sources of collinearity*. New York, John Wiley & Sons, 1980. 292p.

BINET, F.E. On the construction of an index for indirect selection. *Biometrics*, 21: 291-9, 1965.

BRATT, G.M. Significance of the path coefficient analysis in determining the nature of character association. *Euphytica*, 22(2):338-43, 1973.

BRIM, C.A.; JOHNSON, H. W.; COCKHAM, C. C. Multiple selection criteria in soybeans. *Agronomy Journal*, 51: 42-6, 1959.

BUTANY, W.T.; SINGH, M.; MENRA, R.B. Path analysis of yield components in cotton. *Ind. J. Genet. Plant Breed.*, 28(1):40-3, 1968.

BIBLIOGRAFIA

- ABRAHÃO, I.O. Melhoramento do feijoeiro. **Bragantia**, **19**(10):156-61, 1960.
- ALMEIDA, C.M.V. C. **Correlações entre caracteres no estágio adulto e possibilidade de seleção precoce em híbridos de cacau (*Theobroma cacao* L.)**. Piracicaba, USP/ESALQ, 1991. 194p. (Tese D.S.).
- ADAMS, M.W. Basis of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*. **Crop. Sci.**, **7**(5):505-10, 1967.
- BELSLEY, D.A.; KUH, E.; WELCH, R.E. **Regression Diagnostics: Identifying data and sources of collinearity**. New York, John Wiley & Sons, 1980. 292p.
- BINET, F.E. On the construction of an index for indirect selection. **Biometrics**, **21**: 291-9, 1965.
- BRATT, G.M. Significance of the path coefficient analysis in determining the nature of character association. **Euphytica**, **22**(2):338-43, 1973.
- BRIM, C.A.; JOHNSON, H. W.; COCKHAM, C. C. Multiple selection criteria in soybeans. **Agronomy Journal**, **51**: 42-6, 1959.
- BUTANY, W.T.; SINGH, M.; MEHRA, R.B. Path analysis of yield components in cotton. **Ind. J. Genet. Plant Breed.**, **28**(1):40-3, 1968.

- CASTOLDI, F.L. **Análises das interrelações entre o rendimento e diversas características agrônômicas do feijoeiro (*Phaseolus vulgaris* L.)**. Viçosa, MG, UFV, 1991.73p. (Tese D.S.)
- CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. Piracicaba, USP/ESALQ, 1990. 188p. (Tese D.S.)
- CRUZ, C.D. & REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, MG, UFV, Impr.Univ.,1992. (no prelo).
- CUNNINGHAM, C.P. **Animal breeding theory**. Oslo, Londbruksbokhandlen Universitetes for Aget. 1969. 272p.
- DEWEY, D.R. & LU, K.H. A correlation and path coefficient analysis of components of crested wheatgrass seed production. **Agronomy Journal**, **51**(9):515-8, 1959.
- DRAPPER, N.R. & SMITH, H. **Applied regression analysis**. 2.ed. New York, John Wiley & Sons, 1981. 709p.
- DUARTE, R.A. & ADAMS, M.W. A path coefficient analysis of some yield component interrelations in field beans (*Phaseolus vulgaris* L.). **Crop Sci.**, **12**(5):579-82, 1972.
- ELSTON, R. C. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. **Biometrics**, **19**: 85-97, 1963.
- ELGIN, J.H.; HILL, R.R.; ZEIDERS, D. E. Comparison of four methods of multiple trait selection for fine traits in alfalfa. **Crop Sci.**, **10**: 190-3, 1970.
- FERRARI, F. **Estimadores viesados para modelos de regressão em presença de multicolinearidade**. Piracicaba, USP/ESALQ,1989. 127p. (Tese D.S.).
- GOMES, F. P. **Curso de estatística experimental**. 12. ed. São Paulo, Nobel, 1987. 467p.
- GUNST, R. F. & MASON, R. L. Biased estimation in regression: An evaluation using mean squared error. **J. Amer. Statist. Assoc.**, **72**(3539):616-28, 1977.
- GUTIERREZ, A.P. **Estimação de um índice de seleção para um composto de milho**. Piracicaba, USP/ESALQ, 1974.50p. (Tese M.S.).

- HAZEL, L. N. The genetics bases for constructing selection indexes. **Genetics**, **28**:476-90, 1943.
- HAZEL, L.N. & LUSH, J.L. The efficiency of three methods of selection. **J. Hered.**, **33**:393-9, 1942.
- HANSON, W.D. & JOHNSON, H.W. Methods for calculating and evaluating a general selection index obtained by pooling information from two or more experiments. **Genetics**, **42**:421-32, 1957.
- HENDERSON, C.R. Selection index and expected genetic advance. In: HANSON, W.D. & ROBINSON, H.F. **Statistical genetics and plant breeding**. Washington, D.C., National Academy of Sciences, National Research Council, 1963. p.141-163. (Publication, 982).
- HOERL, A.E. Applications of ridge analysis to regression problems. **Chem. Eng. Progr.**, **58**:54-9, 1962
- HOERL, A.E. Optimum solution of many variables equations. **Chem. Eng. Progr.**, **55**:69-78, 1959.
- HOERL, A.E. Ridge analysis 25 years later. **The Amer. Statist.**, **39**(3):186-92, 1985.
- HOERL, A.E. & KENNARD, R.W. Ridge regression: Applications to nonorthogonal problems. **Technometrics**, **12**(1):69-82, 1970a.
- HOERL, A.E. & KENNARD, R.W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, **12**(1):55-68, 1970b.
- HOERL, A.E. & KENNARD, R.W. Ridge regression: Iterative estimation of the biasing parameters. **Comm. Statist.**, **A5**:77-8, 1976.
- HOERL, A.E.; KENNARD, R.W.; BALDWIN, K.F. Ridge regression: Some simulations. **Comm. Statist.**, **A4**:105-23, 1975.
- HOCKING, R.R.; SPEED, F.M.; LYNN, M. J. A class of biased estimators in linear regression. **Technometrics**, **18** (4): 426-37, 1976.
- IEMMA, A.F. **Matrizes para estatística**. Um texto para profissionais de ciências aplicadas. Piracicaba, USP/ESALQ, 1988. 339p.
- IEMMA, A.F. **Modelos lineares. Uma introdução para profissionais da pesquisa agropecuária**. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 32: REUNIÃO DA SOCIEDADE BRASILEIRA DE BIOMETRIA, Londrina, 1987. Londrina, SBB, 1987. 263p.

- JAMES, J.W. Index selection with restrictions. **Biometrics**, **24**: 1015-15, 1968.
- JOHNSON, R.A. & WICHERN, D.W. **Applied multivariate statistical analysis**. 2.ed. New Jersey, Prentice Hall, 1988. 607p.
- JOHNSON, S.R.; REIMER, S.C.; ROTHROCK, T.P. Principal components and the problem of multicollinearity. **Metroeconomía**, **25**:306-17, 1973.
- KALIL, M.N. **Aplicação do método de regressão de cumeeira ("ridge regression") na estimação de funções de demanda e produção**. Piracicaba, USP/ESALQ, 1977. 135p. (Tese M.S.).
- KAUFFMANN, K.D. & DUDLEY, J.W. Selection indices for grain yield, percent protein, and kernel weight. **Crop Sci.**, **19**:583-8, 1979.
- KEMPTHORNE, O. **An introduction to the genetic statistics**. New York, John Willey, 1966. 545p.
- KEMPTHORNE, O. & NORDSKOG, A.W. Restricted selection indices. **Biometrics**, **15**:10-19, 1959.
- KENDALL, M.G. **A course in multivariate analysis**. London, Griffing, 1957. 185p.
- KHANGURA, B.S. & SANDHU, R.S. Path analysis in groundnut (*Arachis hypogaea*). **Ind. J. Agric. Sci.**, **42**(9):792-5, 1972.
- KMENTA, J. **Elements of Econometrics**. New York, Macmillan Publishing, 1971. 655p.
- LAIBLE, C.A. & DIRKS, V.A. Genetic variance and selective value of ear number in corn (*Zea mays* L). **Crop Sci.**, **8**:540-3, 1968.
- LAL, V.S. & HAQUE, M.D.F. Path analysis of yield components in soybean. **Ind. J. Genet. Plant Breed.**, **31**(2):357-62, 1971.
- LAWSON, C.L. & HANSON, R. J. **Solving least square problems**. Englewood Cliffs, Prentice Hall, 1974. 340p.
- LENG, E. R. Component analysis in inheritance studies of grain yield in maize. **Crop Sci.**, **3**(3):187-90, 1963.
- LENKA, D. & MISRA, B. Path-coefficient analysis of yield in rice varieties. **Ind. J. Agric. Sci.**, **43**(4): 376-79, 1973.

- LI, C.C. The concept of path coefficient and its impact on population genetics. **Biometrics**, **12**(2):190-210, 1956.
- LI, C.C. **Path analysis: a primer**. Pacific Grove, Boxwood Press, 1975. 346p.
- LIN, C.Y. Index selection for genetic improvement of quantitative characters. **Theor. Appl. Gen.**, **52**:49-56, 1978.
- MANDEL, J. Use of the singular value decomposition in the regression analysis. **Amer. Statist.**, **36**(1):15-24, 1982.
- MARQUARDT, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. **Technometrics**, **12**(3):591-612, 1970.
- MARQUARDT, D.W. & SNEE, R.D. Ridge regression in practice. **Amer. Statist.**, **29**:3-20, 1975.
- MARTIN St., S.K.; LOESCH JR., P.J., DEMOPULOS-RODRIGUES, J.T.; WIZER, W.J. Selection indexes for the improvement of opaque-2 maize. **Crop sci.**, **22**: 478-95, 1982.
- MASON, R.L.; GUNST, R.F.; WEBSTER, J.T. Regression analysis and problems of multicollinearity. **Comm. Statist.**, **A3**: 277-92, 1975.
- MATZINGER, D.F., COCKERHAM, C.C.; WERNSMAN, E.A. Single character and index selection with random mating in a naturally self-fertilizing species. In: POLLAK, E., KEMPTHORNE, O.; BAILY JR., B. (ed). **Proceedings international conference on quantitative genetics**. Ames, Iowa State University. Press, 1976. p.311-20.
- MONTGOMERY, D.C. & PECK, E.A. **Introduction to linear regression analysis**. New York, John Wiley & Sons, 1981. 504p.
- MULAMBA, N.N. & MOCK, J.J. Improvement of yield potencial of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. **Egypt J. Gen. Citol.**, **7**: 40-51, 1978.
- NAPHADE, D.S. Correlation and path analysis for some characters contributing to fodder yield in sorghum. **Ind. J. A. Sci.**, **42**(9):790-1, 1972.
- NASCIMENTO FILHO, F.J. **Coeficientes de caminhamento entre caracteres da parte aérea e do sistema radicular em guaraná (*Paullinea cupana* var. *sorbilis*)**. Piracicaba, USP/ESALQ, 1988. 101p. (Tese M. S.).

- NETER, J. & WASSERMAN, W. **Applied linear statistical models**. Homewood, Richard D. Irwin, 1974. 842p.
- NETER, J.; WASSERMAN, W.; KUTNER, M.W. **Applied linear regression models**. Homewood, Richard D. Irwin, 1983. 547p.
- PAIVA, J.R.; ROSSETTI, A.G.; GONÇALVES, P.S. Uso do coeficiente de caminhamento no melhoramento da seringueira. **Pesq. Agropec. Bras.**, **17**(3):433-40, 1982.
- PANDEY, J.P. & TORRIE, J.H. Path coefficient analysis of seed components in soybeans (*Glicine max L. Morr.*). **Crop Sci.** **13**(5):505-7, 1973.
- PESEK, J. & BAKER, R.J. Desired improvement in relation to selection indices. **Can. J. Plant Sci.**, **49**:803-4, 1969.
- PIRCHNER, F. **Population genetics in animal breeding**. 2. ed. New York, Plenum Press, 1983. 414p.
- RAMALHO, M.A.P.; ANDRADE, L.A.B.; TEIXEIRA, N.C.S. Correlações genéticas e fenotípicas entre caracteres do feijão (*Phaseolus vulgaris L.*). **Ci. Prát.**, **3**(1):63-70, 1979.
- RAO, C.V.; NARASIMHYA, G.; APPARAO, K. Path analysis in flue-cured tobacco (*Nicotiana tabacum L.*). **Ind. J. Agric. Sci.**, **43**(2):198-201, 1973.
- REGAZZI, A.J. **Variâncias, covariâncias genéticas e índices de seleção num composto de milho (*Zea mays L.*)**. Viçosa, MG, UFV, 1978. 57p. (Tese M.S.).
- ROBINSON, H.F.; COMSTOCK, R.E. HARVEY, P.H. Genotypic and phenotypic correlations in corn and their implications in selection. **Agron. J.**, **43**:283-7, 1951.
- ROSIELLE, A.A.; EAGLES, H.A.; FREY, K.J. Application of restricted selection indexes for improvement of economic value in oats. **Crop Sci.**, **17**:359-61, 1977.
- SANTOS, J.B. & VENCOVSKY, R. Correlação fenotípica e genética entre alguns caracteres agrônômicos do feijoeiro (*Phaseolus vulgaris L.*). **Ci. Prát.**, **10**(3):265-72, 1986.
- SHARMA, R.C.; BHATNAGAR, S.M.; BHATNAGAR, V.K.; BHARGAVA, B. D. Path coefficient analysis of grain and fodder yields, and selection indices in 6-row barley. **Ind. J. Agric. Sci.**, **43**(4):380-5, 1973.

- SHARMA, S.K.; SINGH, V.P.; SING, R.K. Harvest indexes as a criterion for selection in wheat. **Ind. J. Genet. Plant Breed.**, **47**(2):119-23, 1987.
- SHRIVASTAVA, S.N.; SARKAR, D.K.; MALLIC, M.H. Association analysis in rainfed wheat. **Ind. J. Genet. Plant Breed.**, **40**(3):512-14, 1980.
- SILVA, M.A. **Melhoramento Animal** (índice de seleção). Viçosa, MG, UFV, Impr. Univ. 1980. 65p.
- SILVA, M.A. **Melhoramento Animal** (métodos de seleção). Viçosa, MG, UFV, Impr. Univ. 1982. 51p.
- SILVEY, S.D. Multicollinearity and imprecise estimation. **J. Royal Statist. Soc.**, Serie B. **31**(3):589-52, 1969.
- SINGH, K.N. Path analysis in linseed under sodic soil conditions. **Ind. J. Genet. Plant Breed.**, **40**(2):385-7, 1980.
- SINGH, R.K. & CHAUDHARY, B.D. **Biometrical methods in quantitative genetic analysis**. New Delhi, Kajyani Publishers, 1979. 340p.
- SMITH, H.F.A. A discriminant function for plant selection. **Ann. Eugen.**, **7**:240-50, 1936.
- SOUZA, M.A. **Adaptabilidade, estabilidade, correlações e coeficientes de trilha em genótipos de trigo (*T. aestivum* L.) em doze ambientes de Minas Gerais**. Viçosa, MG, UFV, 1985. 118p. (Tese M.S.)
- SUBANDI, W.; COMPTON, A.; EMPIG, L.T. Comparison of the efficiencies of selection indexes for three traits in two variety crosses of corn. **Crop Sci.**, **13**:184-6, 1973.
- TALLIS, G.W. A selection index for optimum genotype. **Biometrics**, **18**:120-2, 1962.
- TAI, G.C.G. Index selection with desired gains. **Crop Sci.**, **17**:182-3, 1977.
- VASCONCELOS, M.E.C. **Análise do coeficiente de caminhamento ("Path coefficient") (*Hevea* spp.)**. Piracicaba, USP/ESALQ, 1982. 77p. (Tese M.S.).
- VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E. & VIÉGAS, G. L. **Melhoramento e produção do milho** Campinas, Fundação Cargill, 1987. v.1, cap. 5, p.137-214.

- VIANNA, R.T.; GAMA, E.E.G.; NASPOLINI FILHO, V.; MORO, J.R. Correlações e análises do coeficiente vetor (path coefficient) em linhagens endogâmicas de milho (*Zea mays* L.). **Ci. Cult.**, **32**(9):1235-41, 1980.
- VIEIRA, C. **Cultura do feijão**. 2.ed. Viçosa, UFV, MG, Impr. Univ. 1983. 146p.
- VIEIRA, C. & GOMES, F.R. Correlação entre o peso da palha e o das sementes, em variedades de feijão (*Phaseolus vulgaris* L.). **R. Ceres**, **16**(88):81-7, 1969.
- VIEIRA, J. V. **Herdabilidade, correlações e índice de seleção em populações de cenoura** (*Daucus carota* L.). Viçosa, MG, UFV, 1988. 86p. (Tese D.S.).
- WEBSTER, J.T.; GUNST, R.F.; MASON, R.L. Latent root regression analysis. **Technometrics**, **16**(4):513-22, 1974.
- WILLIAMS, J.S. The evaluation of a selection index. **Biometrics**, **18**:375-93, 1962.
- WOLFF, F. Mass selection in maize composites by means of selection indexes. **Meded. Landbouwhogsch.**, Wageningen, **79**:80, 1972.
- WRIGHT, S. Correlation and causation. **J. Agric. Res.**, **20**:557-85, 1921(a).
- WRIGHT, S. Systems of mating. I. The biometric relations between parent and offspring. **Genetics**, **6**:111-23, 1921(b).
- YOUNG, S.S.Y. A further examination of the relative efficiency of three methods of selection for genetic gain under less restricted conditions. **Gene. Res.** **2**:106-21, 1961.

APÊNDICE A

QUADRO 1 - Produtividade Média de Grãos (PRG) nos Genótipos de Soja, em Gramas por Planta, Para o Experimento 1. Uruçuva, MG, 1991/92

Genótipo	Produtividade Média
Rojo	7,20
Rojo	5,71
Rojo	5,01
Rio Tibagi	9,35
Ica Tundama	4,89
Ica Lis	5,08
Rio Tibagi	5,45
Roxo	5,25
Diacol	8,91
Ica Tundama	4,55
Rojo	6,82
Rio Tibagi	3,77
Caro	9,40
Roxo	6,33
Duro	7,52
Our	13,39
Rico	15,77
RA	12,86
CM	13,83
	14,13

APÊNDICE

APÊNDICE A

QUADRO 1A - Produtividade Média de Grãos (PRD) dos Genótipos de Feijão, em Gramas por Planta, Para o Experimento 1. Viçosa, MG, 1991/92

Genótipo	Produtividade Média
Rojo 70 x Rio Tibagi	7,20
Rojo 70 x Ica Tundama	5,71
Rojo 70 x Diacol Andino	5,01
Rio Tibagi x Diacol Andino	9,35
Ica Tundama x Diacol Andino	4,89
Ica Llanogrande x Rio Tibagi	5,08
Ica Llanogrande x Ica Tundama	5,46
Rojo 70 x Carioca	5,26
Roxo x Carioca	8,91
Diacol Andino	4,55
Ica Tundama	6,82
Rojo 70	3,77
Rio Tibagi	9,40
Carioca	8,33
Roxo	7,52
Ouro	13,39
Ouro Negro	15,77
Rico	12,86
RA 94	13,83
CNF 5547	14,13

QUADRO 2A - Produtividade Média de Grãos (PRD) dos Genótipos de Feijão, em Gramas por Planta, Para o Experimento 2. Viçosa, MG, 1991/92

Genótipo	Produtividade Média
Rojo 70 x Rio Tibagi	5,80
Rojo 70 x Ica Tundama	5,44
Rojo 70 x Diacol Andino	4,80
Rio Tibagi x Diacol Andino	6,16
Ica Tundama x Diacol Andino	3,78
Ica Llanogrande x Rio Tibagi	5,08
Ica Llanogrande x Ica Tundama	3,06
Rojo 70 x Carioca	6,19
Roxo x Carioca	9,75
Diacol Andino	2,78
Ica Tundama	3,30
Rio Tibagi	9,99
Carioca	10,33
Roxo	7,13
Ouro	17,05
Ouro Negro	14,65
Rico	11,54
RA 94	14,35
CNF 5547	14,06