

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CÂMPUS DE JABOTICABAL**

**ALTERNATIVAS PARA REDUÇÃO DE EFEITOS DE MULTICOLINEARIDADE EM  
MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS DE CORTE**

**Eduardo da Cruz Gouveia Pimentel**

**Orientador: Prof. Dr. Luiz Alberto Fries**

**Co-orientadora: Profa. Dra. Sandra Aidar de Queiroz**

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Zootecnia (Genética e Melhoramento Animal).

JABOTICABAL – SÃO PAULO – BRASIL

Fevereiro – 2004

## **DADOS CURRICULARES DO AUTOR**

**EDUARDO DA CRUZ GOUVEIA PIMENTEL** – nasceu em 17 de janeiro de 1975, no município de Recife – PE. Iniciou o curso de Zootecnia na Universidade Federal Rural de Pernambuco em 1993 e concluiu em 1999. Iniciou o curso de mestrado em Zootecnia, área de Genética e Melhoramento Animal, na FCAV/Unesp – Jaboticabal em março de 2002 e concluiu em fevereiro de 2004.

A meu pai  
e minha mãe  
Dedico este trabalho

## AGRADECIMENTOS

A Deus.

A meu pai, Clóvis, pelo amor, amizade, apoio, incentivo, torcida, e por tudo o que sempre fez pelos filhos.

A minha saudosa mãe, Laélia, que sempre nos deu muito amor, e sei que torce por nós em todos os momentos de nossas vidas.

A meus irmãos Marcelo e Ana Lúcia, minha cunhada Verônica, meus sobrinhos Laélia e Marcelinho, e meu tio Zé, pelo amor, apoio, incentivo e torcida.

A meus maiores amigos Fabão e Mará.

A Luiz Fries e Sandra Queiroz pela orientação e amizade.

A meus colegas e amigos Roberto e Vânia, que sempre me ajudaram e contribuíram muito no desenvolvimento deste trabalho.

Ao programa de melhoramento Conexão Delta G e à empresa GenSys - Consultores Associados S/C Ltda., pelo fornecimento e preparação dos dados usados neste trabalho.

Ao prof. José Reis, com quem aprendi a gostar de Melhoramento Animal.

A meus amigos e companheiros de república: Beto, Frank, Rafael e Zinaldo.

A Toinho, Léo, Márcio, Sarah, Auriclécia, Fabiana, Helena, Márcio (doido), Fábio, Carlo e todos os amigos que tive a felicidade de conhecer em Jaboticabal.

Aos membros das bancas examinadoras de qualificação e defesa, prof. Euclides Malheiros, prof. Antonio Ferraudó, prof. Maurício Alencar e prof. Danísio Munari, pelas discussões, correções e sugestões.

A todos os professores do departamento de Zootecnia.

Ao programa de Pós-Graduação em Zootecnia (Genética e Melhoramento Animal), pela oportunidade de cursar o mestrado nesta instituição.

À CAPES, pelo apoio financeiro.

## SUMÁRIO

Página

Resumo .....	iii
Summary .....	iv

### CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

1.1 Introdução.....	01
1.2 O problema da multicolinearidade em análises de regressão .....	01
1.3 Estimação viesada de coeficientes de regressão (regressão de cumeeira).....	04
1.4 Efeitos genéticos envolvidos em características de populações cruzadas.....	07
1.5 Objetivos.....	11
Referências .....	11

### CAPÍTULO 2 – UM PROGRAMA PARA IMPLEMENTAÇÃO DA REGRESSÃO DE CUMEEIRA USANDO O MÉTODO GIVENS DE DECOMPOSIÇÃO ORTOGONAL PARA OBTENÇÃO DAS SOLUÇÕES

Resumo .....	18
2.1 Introdução.....	19
2.2 Teoria .....	21
2.2.1 Regressão de cumeeira.....	21
2.2.2 Rotações de Givens .....	23
2.3 Estrutura do programa.....	25
2.3.1 Entrada de dados .....	25
2.3.2 Operações executadas pelo programa.....	26
2.3.3 Arquivos de saída .....	27
2.4 Aplicações em problemas de melhoramento animal .....	29
Referências .....	29

### CAPÍTULO 3 – USO DA REGRESSÃO DE CUMEEIRA PARA REDUÇÃO DOS EFEITOS DA MULTICOLINEARIDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS DE CORTE

Resumo .....	31
Introdução .....	32
Material e métodos .....	34
Resultados e discussão.....	39
Conclusões.....	56
Referências .....	56

### CAPÍTULO 4 – EFEITOS DA INCLUSÃO DE PARÂMETROS PARA EPISTASIA E COMPLEMENTARIEDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS CRUZADOS

Resumo .....	59
Introdução .....	60
Material e métodos .....	62
Resultados e discussão.....	66
Conclusões.....	75
Referências .....	75

CAPÍTULO 5 – IMPLICAÇÕES .....	79
APÊNDICE .....	80

## ALTERNATIVAS PARA REDUÇÃO DE EFEITOS DE MULTICOLINEARIDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS DE CORTE

**RESUMO** – O problema da multicolinearidade em análises de regressão foi abordado. A técnica da regressão de cumeieira foi empregada na estimação de parâmetros de efeitos genéticos sobre o desempenho de animais cruzados, por meio de dois modelos: um contendo apenas efeitos de ação aditiva e dominância (AD); e outro incluindo, além desses, efeitos de epistasia e complementariedade (ADEC). Um programa foi desenvolvido, em linguagem Fortran 90, para implementação de cinco versões da regressão de cumeieira: o método proposto originalmente; o implementado pelo SAS; e três formas de ponderação do coeficiente  $\lambda$ . Três critérios matemáticos para escolha de  $\lambda$  foram testados: a soma e a média harmônica dos valores absolutos da estatística t de Student, e o valor de  $\lambda$  a partir do qual os valores dos fatores de inflação de variância passavam a ser todos menores que trezentos. As comparações entre os cinco métodos e os três critérios foram feitas, usando-se o modelo ADEC, pelo exame de superfícies de predição obtidas a partir dos coeficientes estimados. Superfícies de predição também foram usadas para comparação entre os dois modelos, para cada método. Com o conjunto de dados utilizado, superfícies de predição biologicamente coerentes puderam ser obtidas em todos os métodos de implementação, usando-se o critério com base nos valores de FIV para determinação de  $\lambda$ . Recomenda-se que um critério matemático seja usado como ferramenta auxiliar para escolha de  $\lambda$ , não dispensando o exame dos sinais e valores das estimativas e um bom conhecimento do fenômeno em estudo. A inclusão de parâmetros para efeitos de epistasia e complementariedade em modelos de avaliação de efeitos genéticos em animais cruzados pôde representar um ganho tanto em termos de ajuste do modelo quanto de capacidade de predição de desempenho de genótipos não testados.

**Palavras-chave:** complementariedade, cruzamento, epistasia, estimação viesada, heterose, regressão de cumeieira

## ALTERNATIVES TO REDUCE THE EFFECTS OF MULTICOLLINEARITY IN BEEF CATTLE GENETIC EFFECTS EVALUATION MODELS

**SUMMARY** – The problem of multicollinearity in regression analysis was studied. Ridge regression techniques were used to estimate genetic parameters affecting performance of crossbred animals, using two models: the additive-dominance model; and an alternative model including additive, dominance, complementarity and epistatic effects. A software was developed, in Fortran 90, to perform five variant types of ridge regression: the originally proposed method; the one implemented by SAS; and three forms of weighting the ridge coefficient  $\lambda$ . Three mathematical criteria were tested with the aim of choosing a value for the  $\lambda$  coefficient: the sum and the harmonic mean of absolute Student t-values, and the value of  $\lambda$  from which all variance inflation factors (VIFs) became lower than 300. Prediction surfaces, obtained from estimated coefficients, were used to compare the five methods and three criteria, using the alternative model. Prediction surfaces were also used to compare the two models, for each method. In this study (and this particular data structure), prediction surfaces showed quite acceptable biological interpretation, for all five methods, when criterion based on VIF values was used to choose the  $\lambda$  coefficient. A mathematical criterion to choose  $\lambda$  is recommended as an indicator tool, without excluding an exam of signs and values of estimated coefficients, and a good understanding of the phenomenon under study. Inclusion of complementarity and epistatic effects, in models for genetic effects evaluation in crossbred animals, represented a better fit of the model, and an improvement in its ability to predict performance of untested genotypes.

**Keywords:** biased estimation, complementarity, crossbreeding, epistasis, heterosis, ridge regression



## CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

### 1.1 Introdução

Os modelos mais usados para avaliação de efeitos genéticos em características de crescimento de bovinos cruzados levam em consideração apenas efeitos de ação aditiva e de dominância. Contudo, a partir da década de 80, têm sido relatadas evidências de que a contribuição de efeitos epistáticos para a expressão de tais características nem sempre é negligenciável.

O melhor conhecimento dos efeitos genéticos envolvidos em características de bovinos cruzados pode melhorar a qualidade das predições de valor genético em avaliações de populações multirraciais, bem como as predições de desempenho de animais cruzados para planejamento de programas de cruzamentos.

As estimativas de parâmetros de efeitos genéticos, usando-se modelos que contemplam efeitos epistáticos, obtidas por meio de análises de regressão por quadrados mínimos ordinários, têm se mostrado de difícil interpretação biológica. A presença de correlações entre as variáveis incluídas no modelo (multicolinearidade) resultam na obtenção de estimativas incoerentes, muitas vezes com sinal trocado, e pode ser o motivo para o uso corrente de modelos mais simples.

Um procedimento alternativo para contornar os problemas causados pela multicolinearidade é o uso de métodos viesados de estimação dos coeficientes de regressão. A regressão de cumeeira é um desses métodos.

### 1.2 O problema da multicolinearidade em análises de regressão

Supondo-se um modelo  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ .

Em que:

$\mathbf{y}$  é o vetor de observações da variável resposta, de ordem  $\mathbf{n} \times \mathbf{1}$ ;

$\mathbf{X}$  é a matriz de observações das variáveis explicativas, de ordem  $\mathbf{n} \times \mathbf{p}$ ;

$\boldsymbol{\beta}$  é o vetor dos parâmetros a serem estimados, de ordem  $\mathbf{p} \times \mathbf{1}$ ;

$\boldsymbol{\epsilon}$  é o vetor dos resíduos, de ordem  $\mathbf{n} \times \mathbf{1}$ ;

As soluções para  $\beta$ , por quadrados mínimos ordinários, são obtidas pela expressão  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . Contudo, se  $\mathbf{X}'\mathbf{X}$  é singular, sua inversão não pode ser feita e as equações normais não têm solução única. Em vez disso, uma infinidade de soluções existem. Quando isso acontece, advém do fato de que existe pelo menos uma combinação linear entre as colunas da matriz  $\mathbf{X}$  que é igual a zero. Ou, para colocar de outra maneira, pelo menos uma coluna de  $\mathbf{X}$  é linearmente dependente das outras colunas. Diria-se então que existe colinearidade (ou multicolinearidade) entre as colunas de  $\mathbf{X}$  (DRAPER & SMITH, 1998).

Sendo  $n$  o número de observações e  $p$  o número de parâmetros, existe multicolinearidade perfeita quando  $\text{posto}(\mathbf{X}) < p$ . Isso implica a existência de uma ou mais relações lineares exatas entre as colunas da matriz  $\mathbf{X}$ . Como  $\text{posto}(\mathbf{X}) = \text{posto}(\mathbf{X}'\mathbf{X})$ , a multicolinearidade perfeita é sempre caracterizada por  $\det(\mathbf{X}'\mathbf{X}) = 0$ . Isso implica que  $\mathbf{X}'\mathbf{X}$  não possui inversa verdadeira e, com isso, estimativas por quadrados mínimos ordinários não podem ser obtidas. Um problema potencialmente maior é o da multicolinearidade imperfeita, associada a correlações altas, mas não iguais a 1 ou -1, entre as colunas de  $\mathbf{X}'\mathbf{X}$ . Sua inversão introduz instabilidade numérica às estimativas de  $\beta$  e traz sérias conseqüências à sua precisão (SPANOS & MCGUIRK, 2002). Em situações como essas, as estimativas dos coeficientes obtidas por quadrados mínimos ordinários terão altos erros-padrão (apesar de algumas apresentarem valores de  $t$  estatisticamente significativos, para a hipótese de nulidade). Em particular, esses coeficientes podem ser muito altos e inapropriados para predição (BJÖRKSTRÖM & SUNDBERG, 1999). A multicolinearidade também torna difícil a determinação da contribuição de cada variável explicativa, pois seus efeitos são confundidos e as estimativas dos coeficientes de regressão são altamente instáveis (RIKARDBSEN & JOHANSEN, 2003).

Indicações de multicolinearidade aparecem quando os estimadores sofrem grandes mudanças em virtude da adição ou retirada de uma variável, quando os sinais algébricos dos coeficientes não correspondem à expectativa, ou quando coeficientes considerados importantes têm grande erro padrão. A presença de multicolinearidade

também é indicada pela magnitude dos coeficientes de correlação existentes entre as variáveis explicativas. Uma alta correlação entre um par de variáveis explicativas indica forte relação linear entre essas duas variáveis (CHATTERJEE & PRICE, 1991).

O diagnóstico mais simples de colinearidade é feito pelo exame da matriz de correlações entre as variáveis explicativas (HAIR JR. *et al.*, 1992; BAYHAN & BAYHAN, 1998; FRASER, 2000). Contudo, segundo MOORE *et al.* (1984), a matriz de correlações pode falhar no diagnóstico de multicolinearidade se as dependências lineares correspondem a inter-relações entre grupos de variáveis, e não a correlações simples entre pares delas. Uma melhor investigação da multicolinearidade envolve o exame do valor do coeficiente de determinação ( $R^2$ ), que resulta da regressão de cada variável explicativa contra todas as outras. A relação entre as variáveis explicativas pode ser julgada pelo exame do valor do *fator de inflação de variância* (FIV). O *fator de inflação de variância* é uma ferramenta de diagnóstico bastante útil pois indica diretamente o prejuízo causado pela colinearidade à precisão das estimativas (FOX & MONETTE, 1992). Se  $R_i^2$  é o quadrado do coeficiente de correlação múltipla de  $x_i$  contra as outras variáveis explicativas, a *inflação da variância* de  $x_i$  é dada por:

$$FIV(x_i) = \frac{1}{1 - R_i^2}$$

Fica claro que se  $x_i$  tiver forte relação linear com as outras variáveis explicativas,  $R_i^2$  ficará próximo de 1 e o valor do FIV tenderá a ser muito alto. Na ausência de qualquer relação linear entre as variáveis explicativas (se elas são ortogonais, por exemplo),  $R_i^2$  seria igual a zero e o FIV seria igual a 1. À medida que o valor do FIV aumenta, caracteriza-se a saída da ortogonalidade para a tendência à colinearidade (CHATTERJEE & PRICE, 1991). Segundo STINE (1995), não existe um valor crítico bem definido que caracterize um alto valor de FIV. LECLERC & PIREAUX (1995) sugerem que *fatores de inflação de variância* maiores que 300 denotam problema de multicolinearidade.

De acordo com ÖZTÜRK & AKDENIZ (2000), a melhor maneira de se verificar a existência de multicolinearidade é o exame dos autovalores da matriz  $X'X$ . Segundo

MOORE *et al.* (1984), à medida que a matriz  $X$  se torna menos ortogonal em razão da presença de colinearidade, os autovalores da matriz de correlações decrescem em magnitude. Um autovalor igual a zero existe para cada combinação linear exata, enquanto que as dependências são caracterizadas por baixos autovalores. KNIGHT *et al.* (1992) sugerem o exame do índice de condição, que se define como sendo a raiz quadrada da razão entre o maior e o menor autovalor. Índices de condição superiores a 30 podem ser tomados como indicativos de severa multicolinearidade.

A importância da multicolinearidade deve ser enfatizada do ponto de vista do inerente obstáculo que ela introduz: estimativas confiáveis dos coeficientes de regressão não podem ser obtidas da amostra em si. A amostra é pobre em termos de variação independente dos regressores, e não se pode esperar que se obtenham da amostra estimativas dos efeitos independentes dos regressores. Nesses casos, o uso de informação *a priori* pode ajudar (GOLDBERGER, 1964).

### **1.3 Estimação viesada de coeficientes de regressão (regressão de cumeeira)**

Várias alternativas têm sido propostas no sentido de superar o problema que a multicolinearidade traz à estimação por quadrados mínimos. Segundo FREUND & WILSON (1998), essas alternativas podem ser divididas em três categorias: seleção de variáveis, redefinição de variáveis e estimação viesada. Uma quarta alternativa a ser considerada é o uso de informação *a priori*.

Na seleção de variáveis, propõe-se um modelo inicial, sempre contendo uma grande quantidade de variáveis explicativas, e usam-se análises estatísticas para que um modelo final, contendo uma subsérie de variáveis explicativas que adequadamente expliquem a variação na resposta, seja revelado (FREUND & LITTELL, 2000).

Mesmo na presença de multicolinearidade, pode ser interessante manter todas as variáveis possíveis no modelo e estimar sua importância comparativa com relação à variável dependente. Segundo CUADRAS (1993), a afirmação de que variáveis correlacionadas contêm informação redundante sobre a resposta nem sempre é correta, e a seleção de variáveis pode incorrer no descarte de algumas que apresentam efeito combinado altamente significativo.

A redefinição consiste em transformações lineares das variáveis do modelo em novas variáveis, não correlacionadas. Sua aplicação é pertinente apenas em casos em que essas novas variáveis têm um significado prático útil na discriminação do fenômeno em estudo.

Métodos viesados de estimação de coeficientes de regressão têm sido amplamente empregados como uma alternativa à regressão por quadrados mínimos ordinários quando as variáveis preditoras são altamente correlacionadas. Segundo ÖZÇELİK *et al.* (2002) e WAN (2002), quando se usa um procedimento viesado de estimação, pretende-se essencialmente introduzir um certo viés às estimativas com o intuito de reduzir suas variâncias. Alguns desses procedimentos apresentam propriedades reducionais, e o benefício de sua aplicação resulta do fato de que a redução em variância mais do que compensa o acréscimo em viés, de maneira que o quadrado médio do erro de predição também é reduzido (WEIGEL *et al.*, 1991; SIMIANER, 1991; BUTLER & DENHAM, 2000), lembrando que  $MSE = PEV + viés^2$ , em que MSE é o quadrado médio do erro de predição e PEV a variância do erro de predição. Segundo ROOK *et al.* (1990), alguns métodos de estimação viesada podem ser eficientes na remoção do efeito da colinearidade, fornecendo interpretações mais aceitáveis das relações da variável resposta com as explicativas, e funcionando melhor como ferramenta da predição.

FRANK & FRIEDMAN (1993) apresentam um estudo de simulação comparando, em diversas situações, os três métodos reducionais mais usados para estimação de coeficientes de regressão: regressão de cumeeira, regressão de componentes principais e quadrados mínimos parciais. Os resultados desse estudo mostraram que, por possuírem propriedades semelhantes, as soluções obtidas pelos três métodos, com os mesmos dados, são geralmente bastante similares. Pequena vantagem, em termos de acurácia de estimação dos coeficientes de regressão individuais, foi observada usando-se o método da regressão de cumeeira (HOERL & KENNARD, 1970).

HOERL (1962) sugeriu que para controlar a inflação e instabilidade geral associada às estimativas de quadrados mínimos na presença de multicolinearidade, pequenas quantidades positivas ( $\lambda$ ) fossem adicionadas aos elementos da diagonal

principal da matriz  $X'X$ . O vetor de estimativas passa então a ser obtido por:  $\beta^* = (X'X + \lambda I)^{-1} X'Y$ , em que  $I$  é uma matriz identidade de ordem igual à de  $X'X$  (HOERL & KENNARD, 1970). Ao incorporar à análise uma informação que não está presente nos dados, a regressão de cumeeira pode ser considerada um método de estimação viesada combinado ao uso de informação *a priori*.

Desde que o valor da constante  $\lambda$  não seja muito alto, a estimação por regressão de cumeeira apresenta menor quadrado médio do erro que a análise por quadrados mínimos ordinários (CHIPMAN, 1999; HAWKINS & YIN, 2002). O problema chave na estimação de cumeeira é a escolha do coeficiente  $\lambda$  (SHI & WANG, 1999). Em seus primeiros trabalhos sobre regressão de cumeeira, HOERL & KENNARD (1970) sugeriram um método gráfico para adoção do coeficiente  $\lambda$  obedecendo aos seguintes critérios: o valor de  $\lambda$  deve coincidir com o ponto a partir do qual as estimativas dos parâmetros tornam-se relativamente estáveis; os coeficientes com sinais aparentemente incorretos para  $\lambda = 0$  já terão sido alterados para o sinal apropriado; o valor da soma dos quadrados dos resíduos da regressão não deverá ter sofrido aumento excessivo nesse nível, em relação ao nível  $\lambda = 0$ .

Diversas metodologias estatísticas têm sido propostas no intuito de determinar critérios menos subjetivos para escolha do coeficiente  $\lambda$ . Algumas delas são descritas por VINOD (1976), HOERL *et al.* (1986) e VAN GAANS & VRIEND (1990). Segundo MONTGOMERY & PECK (1982), citados por VAN GAANS & VRIEND (1990), nenhum dos critérios matemáticos propostos para seleção de  $\lambda$  provaram alguma superioridade. O método gráfico tem sido o mais utilizado pela facilidade na identificação de mudanças nos coeficientes.

Segundo MOORE *et al.* (1984), a comparação dos resultados obtidos por quadrados mínimos ordinários com os obtidos por regressão de cumeeira, só pode ser feita tendo-se como referência os sinais dos coeficientes e seus níveis de significância estatística. FRIES *et al.* (2000a) sugerem que diversos valores de  $\lambda$  sejam testados até o ponto em que a redução nos valores absolutos das estimativas não seja mais compensada por reduções no FIV e nos erros-padrão. A divisão dessas estimativas

pelos respectivos erros-padrão resulta nos valores de  $t$  de Student, cuja soma fornece a estatística “soma(  $| t |$  )”. Essa função de  $t$  não é monotônica em  $\lambda$ , e pode ser um referencial mais pragmático e informativo para definição de seu valor do que o método gráfico. Segundo PICCOLI *et al.* (2002), a função “média harmônica dos valores absolutos de  $t$  (HARM\_AT)” pode ajudar a identificar o menor valor possível de  $\lambda$  que produza a maior uniformidade em termos de resultados estatisticamente significativos.

Algumas implementações alternativas do método da regressão de cumeieira têm sido sugeridas. ALDRIN (1997) apresenta a regressão de cumeieira de comprimento modificado, pela qual, estimativas preliminares obtidas por regressão de cumeieira são multiplicadas por fatores de ajuste, num procedimento similar ao estimador James-Stein. FOUCART (1999), pelo que definiu como regressão de cumeieira parcial, propõe a adição do coeficiente  $\lambda$  somente aos elementos da diagonal principal correspondentes a covariáveis envolvidas em multicolinearidade, produzindo estimativas satisfatórias dos coeficientes de regressão superestimados por quadrados mínimos ordinários, sem perturbar os outros coeficientes, em contraste com a regressão de cumeieira clássica, em que  $\lambda$  é adicionado a todos os elementos da diagonal. LIPOVETSKY & CONKLIN (2001) sugerem a adição de valores diferentes de  $\lambda$  para cada elemento da diagonal principal da matriz de coeficientes.

#### **1.4 Efeitos genéticos envolvidos em características de populações cruzadas**

Características de importância econômica apresentam considerável variação entre raças de gado de corte. Diferenças entre as raças são um importante recurso genético a ser usado em cruzamentos para a exploração da heterose e da complementariedade. O conhecimento dos efeitos genéticos das raças e de seus cruzamentos em diversas condições climáticas pode ser usado na identificação das melhores combinações de raças e sistemas de cruzamento para os diversos mercados existentes (FRANKE *et al.*, 2001).

Em rebanhos cruzados, o melhoramento genético pode ser feito pela combinação de dois métodos: (1) maximização da heterose e da retenção de heterose, pelo melhor uso possível de combinações de raças e sistemas de cruzamento; e (2)

seleção para aumentar o valor genético aditivo das raças componentes. No primeiro método, o desafio é identificar as causas da heterose, para que a retenção da heterose em vários cruzamentos e gerações possa ser predita e maximizada (MACKINNON *et al.*, 1996).

Uma descrição de caso de vigor híbrido foi feita por MENDEL (1865) em seus experimentos com hibridização de plantas. O autor relatou que o cruzamento de plantas com caules de 1 e 6 pés de comprimento produziu, sem exceção, híbridos mais vigorosos, com comprimento de caule variando de 6 a 7,5 pés. SHULL (1914), citado por SHERIDAN (1981), empregou o termo heterose, para descrever o incremento em vigor apresentado por indivíduos cruzados, em relação a geração parental, independentemente da causa. De acordo com DICKERSON (1993), esse incremento resulta do aumento em heterozigose (redução de endogamia) para alelos dominantes, efeitos de complementariedade materna/paterna, e mudanças em interações não alélicas.

O modelo mais utilizado para estimação de parâmetros de efeitos genéticos e planejamento de programas de cruzamento contém apenas efeitos aditivos e de dominância, e foi proposto por FISHER (1918). Segundo WILLHAM (1970), a definição de heterose seria, de fato, uma definição de interação. Interação é a falha de um esquema aditivo em descrever os fatos. Logo, o uso da dominância serviria para ilustrar o que acontece geneticamente com uma população cruzada. A inclusão de vários loci, em vez de um ou dois, seria um processo simplesmente aditivo, a menos que a epistasia fosse considerada.

Coeficientes de heterozigose medem a proporção esperada de loci em que os dois alelos derivam de raças ancestrais diferentes, e espera-se que estejam relacionados com a contribuição da dominância para a heterose (MACKINNON *et al.*, 1996). Contudo, segundo SHERIDAN (1981), a limitada evidência experimental disponível indica que, em muitos casos, o nível de heterose em populações cruzadas, exceto para populações F1, é menor do que seria predito com base na porcentagem de heterozigose. De acordo com DICKERSON (1993), os desvios da associação linear



com heterozigose, entre as gerações parentais, F1, F2 e retrocruzadas, fornecem evidência de efeitos de interação gênica não alélica.

WRIGHT (1978), em seus estudos de combinações gênicas, reconhece que um organismo vivo nunca deve ser visto como um mosaico de características unitárias, cada uma determinada por um único gene, mas como uma vasta rede de sistemas que interagem. O autor vislumbrava o valor adaptativo de uma população na forma de um mapa topográfico com picos e vales. Cada ponto dessa superfície adaptativa representando um genótipo em particular, e a altura desse ponto sendo determinada pelo valor adaptativo desse genótipo em seu ambiente. Num sistema puramente aditivo, haveria um pico na superfície adaptativa, correspondendo ao valor aditivo máximo. Com interações não aditivas e efeitos epistáticos contribuindo para o fenótipo, e/ou efeitos pleiotrópicos entre características contribuindo para o valor adaptativo geral, deveriam existir diversos picos de diferentes alturas correspondendo a diferentes genótipos (SCHAMBER & MUIR, 2001) . De acordo com WRIGHT (1978), ao longo do processo de formação das raças, a seleção natural deve então, de algum modo, operar sobre sistemas interativos de genes, como blocos, produzindo combinações que precisam ser mais eficientes.

Segundo DICKERSON (1972), citado por WILLHAM & POLLAK (1985), na produção de uma população F2 ou de sintéticos, existe perda por recombinação, em consequência de quebras de combinações epistáticas favoráveis nas populações parentais. Essa perda por recombinação foi definida por DICKERSON (1973) como a fração média de pares de genes, com segregação independente, presentes nos gametas de ambos os pais, que espera-se que sejam combinações não-parentais.

Se efeitos de epistasia contribuem para o desempenho, e se a meiose destrói combinações favoráveis de genes, os valores observados para um cruzamento devem ser menores (com heterose positiva) do que predições feitas usando-se um modelo que considera apenas efeitos aditivos e de dominância (RUTLEDGE, 2001). Segundo KINGHORN & VERCOE (1989), a heterose em animais cruzados deve conter um componente negativo causado pela quebra de interações favoráveis entre loci, fixadas no processo de seleção natural e artificial. KINGHORN (1980) definiu essa perda por

epistasia como sendo proporcional à probabilidade de que dois genes não alelos, em um indivíduo diplóide, sejam de raças diferentes.

Pelo fato da hipótese da dominância ter se mostrado insuficiente para explicar o desempenho de animais cruzados, diversas pesquisas vêm sendo conduzidas, estendendo-se o modelo estatístico para avaliação de efeitos genéticos, de maneira a incluir epistasia. Contudo, segundo KINGHORN & VERCOE (1989), o grau de expressão de efeitos epistáticos em um genótipo cruzado tem uma base teórica mais complexa. Uma estratégia alternativa é fazer uma pressuposição sobre a natureza biológica das interações epistáticas, e então adicionar um único parâmetro na construção do modelo. Em muitos casos relatados na literatura, a derivação dos coeficientes tem sido feita, com base nas teorias de perda por recombinação e de perda por epistasia (BAAS *et al.*, 1992; HIROOKA *et al.*, 1998; ARTHUR *et al.*, 1999; KAHN *et al.*, 2000; SHARMA *et al.*, 2000; EL FADILI & LEROY, 2001; CASSADY *et al.*, 2002; DEMEKE *et al.*, 2003a; DEMEKE *et al.*, 2003b).

FRIES *et al.* (2000b) examinaram a hipótese, numa série de experimentos com peso à desmama de bezerros cruzados, de que uma maior heterozigose em uma geração é concomitante com maiores efeitos epistáticos na geração seguinte. Logo, perdas por recombinação ou epistasia serão proporcionais à epistazigose, definida como a heterozigose média observada nos pais de um indivíduo. A heterozigose dá a probabilidade de que genes de um determinado par tenham se originado de diferentes raças. Essa probabilidade é válida para quaisquer dois genes, alelos ou não. A heterozigose também informa a fração esperada de gametas recombinantes produzidos por um indivíduo cruzado, para qualquer número de loci, formas e configurações espaciais, com respeito às raças originais. Essa interpretação da heterozigose é muito próxima da perda por epistasia de KINGHORN (1980) e complementa a definição de DICKERSON (1973) de perda por recombinação.

Quando se consideram características de crescimento de animais cruzados em ambiente tropical, pode-se imaginar que dois conjuntos de genes estejam atuando: um relacionado a potencial de crescimento e outro a adaptação. Esses dois conjuntos podem ser tomados como características que se complementam e cujo produto

(interação) pode fornecer uma base teórica para a heterose que não depende de dominância ou epistasia. Essa relação multiplicativa entre as ações aditivas para as duas características foi definida por KINGHORN (1993) como “profit heterosis”, e pode ser interpretada como um efeito de complementariedade entre as características que determinam crescimento.

Uma maneira de tentar identificar composições raciais mais adequadas às diversas condições climáticas presentes em diferentes regiões, é a consideração, no modelo, de efeitos de latitude. A inclusão de parâmetros para epistasia e complementariedade, e para interações de efeitos genéticos com efeitos de latitude, no modelo de avaliação de efeitos genéticos tradicional (ação aditiva e dominância), incorre em problemas de multicolinearidade, tornando necessário o uso de métodos alternativos de estimação dos coeficientes (BRITO *et al.*, 2002; PICCOLI *et al.*, 2002).

### 1.5 Objetivo

O objetivo do presente trabalho foi desenvolver um programa computacional que permitisse: testar diferentes implementações da regressão de cumeeira; testar diferentes critérios para determinação do coeficiente  $\lambda$ ; aplicar a técnica da regressão de cumeeira para estimação de parâmetros em modelos de avaliação de efeitos genéticos em populações cruzadas de bovinos de corte; e comparar diferentes modelos utilizados para avaliação de efeitos genéticos, quanto a capacidade de predição.

### Referências

- ALDRIN, M. Length modified ridge regression. **Computational Statistics & Data Analysis**, v.25, p.377-398. 1997.
- ARTHUR, P.F.; HEARNshaw, H.; STEPHENSON, P.D. Direct and maternal additive and heterosis effects from crossing *Bos indicus* and *Bos taurus* cattle: cow and calf performance in two environments. **Livestock Production Science**, v.57, p.231-241. 1999.

- BAAS, T.J.; CHRISTIAN, L.L.; ROTHSCHILD, M.F. Heterosis and Recombination Effects in Hampshire and Landrace Swine: I. Maternal Traits. **Journal of Animal Science**, v.70, p.89-98. 1992.
- BAYHAN, G.M.; BAYHAN, M. Forecasting using autocorrelated errors and multicollinear predictor variables. **Computers and Industrial Engineering**, v.34, n.2, p.413-421. 1998.
- BJÖRKSTRÖM, A.; SUNDBERG, R. A Generalized view on Continuum Regression. **Scandinavian Journal of Statistics**, v. 26, p.17-30. 1999.
- BRITO, F.V.; PICCOLI, M.L.; SEVERO, J.L.P.; SCHENKEL, F.S.; ROSO, V.M.; FRIES, L.A. Estimating environmental and genotypic effects on preweaning weight gain of Angus x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** p.773-776.
- BUTLER, N.A.; DENHAM, M.C. The peculiar shrinkage properties of partial least squares regression. **Journal of The Royal Statistical Society**, Ser. B, v.62, part 3, p.585-593. 2000.
- CASSADY, J.P.; YOUNG, L.D.; LEYMASTER, K.A. Heterosis and recombination effects on pig growth and carcass traits. **Journal of Animal Science**, v.80, p.2286-2302. 2002.
- CHATTERJEE, S.; PRICE, B. **Regression analysis by example**. 2nd ed. New York: Wiley series in probability and mathematical statistics, 1991. 278p.
- CHIPMAN, J.S. Linear restrictions, rank reduction, and biased estimation in linear regression. **Linear Algebra and its Applications**, v. 289, p.55-74. 1999.
- CUADRAS, C.M. Interpreting an inequality in multiple regression. **The American Statistician**, v.47, n.4, p.256-259. 1993.
- DEMEKE, S.; NESER, F.W.C.; SCHOEMAN, S.J. Early growth performance of *Bos taurus* x *Bos indicus* cattle crosses in Ethiopia: estimation of individual crossbreeding effects. **Journal of Animal Breeding and Genetics**, v.120, p.245-257. 2003a.
- DEMEKE, S.; NESER, F.W.C.; SCHOEMAN, S.J. Variance components and genetic parameters for early growth traits in a mixed population of purebred *Bos indicus* and crossbred cattle. **Livestock Production Science**, v.84, p.11-21. 2003b.

- DICKERSON, G.E. Inbreeding and Heterosis in Animals. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF DR. JAY L. LUSH. 1973, Champaign, **Proceedings...** p.54-77.
- DICKERSON, G.E. **Evaluation of breeds and crosses of domestic animals.** Roma: FAO, 1993. p.1-49. (FAO Animal Production and Health Paper, 108).
- DRAPER, N.R.; SMITH, H. **Applied regression analysis.** 3rd ed. New York: Wiley series in probability and statistics, 1998. 706p.
- EL FADILI, M.; LEROY, P.L. Estimation of additive and non-additive genetic parameters for reproduction, growth and survival traits in crosses between the Moroccan D'man and Timahdite sheep breeds. **Journal of Animal Breeding and Genetics**, v.118, p.341-353. 2001.
- FISHER, R.A. The correlation between relatives on the supposition of Mendelian inheritance. **Transactions of the Royal Society of Edinburgh**, v. 52, p. 399-433. 1918.
- FOUCART, T. Stability of the inverse correlation matrix. Partial ridge regression. **Journal of Statistical Planning and Inference**, v.77, p.141-154. 1999.
- FOX, J.; MONETTE, G. Generalized collinearity diagnostics. **Journal of the American Statistical Association**, v.87, n.417, p.178-185. 1992.
- FRANK, I.E.; FRIEDMAN, J.H. A Statistical View of Some Chemometrics Regression Tools. **Technometrics**, v.35, n.2, p.109-135. 1993.
- FRANKE, D.E.; HABET, O.; TAWAH, L.C.; WILLIAMS, A.R.; DEROUEN, S.M. Direct and maternal genetic effects on birth and weaning traits in multibreed cattle data and predicted performance of breed crosses. **Journal of Animal Science**, v.79, p.1713-1722. 2001.
- FRASER, I. An application of maximum entropy estimation: the demand for meat in the United Kingdom. **Applied Economics**, v.32, i1, p.45-63. 2000.
- FREUND, R.J.; LITTELL, R.C. **SAS System for Regression**, 3rd ed. Cary: SAS Institute Inc., 2000. 235p.
- FREUND, R.J.; WILSON, W.J. **Regression Analysis:** statistical modeling of a response variable. San Diego: Academic Press, 1998. 444p.

- FRIES, L.A.; GRASER, H.U.; JOHNSTON, D.J.; HEARNSHAW, H. Using Ridge Regression to Estimate Genetic Effects in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, suppl. B, p. 241, 2000a.
- FRIES, L.A.; JOHNSTON, D.J.; HEARNSHAW, H.; GRASER, H.U. Evidence of Epistatic Effects on Weaning Weight in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, suppl. B, p. 242, 2000b.
- GOLDBERGER, A.S. **Econometric Theory**. New York: Wiley Series in Probability and Mathematical Statistics, 1964. 399p.
- HAIR JR., J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Multivariate Data Analysis**, 3rd ed. New York: Macmillan Publishing Company, 1992. 544p.
- HAWKINS, D.M.; YIN, X. A faster algorithm for ridge regression of reduced rank data. **Computational Statistics & Data Analysis**, v.40, p.253-262. 2002.
- HIROOKA, H.; GROEN, A.F.; VAN DER WERF, J.H.J. Estimation of additive and non-additive genetic parameters for carcass traits on bulls in dairy, dual purpose and beef cattle breeds. **Livestock Production Science**, v.54, p.99-105. 1998.
- HOERL, A.E. Application of Ridge Analysis to Regression Problems. **Chemical Engineering Progress**, v.58, p.54-59. 1962.
- HOERL, A.E.; KENNARD, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v.12, n.1, p.55-67. 1970.
- HOERL, R.W.; SCHUENEMEYER, J.H.; HOERL, A.E. A Simulation of Biased Estimation and Subset Selection Regression Techniques. **Technometrics**, v.28, n.4, p.369-380. 1986.
- KAHI, A.K.; THORPE, W.; NITTER, G.; BAKER, R.L. Crossbreeding for dairy production in the lowland tropics of Kenya. I. Estimation of individual crossbreeding effects on milk production and reproductive traits and on cow live weight. **Livestock Production Science**, v.63, p.39-54. 2000.
- KINGHORN, B.P. The expression of 'recombination loss' in quantitative traits. **Journal of Animal Breeding and Genetics**, v.97, p.138-143. 1980.

- KINGHORN, B.P. Theory of Breed Utilisation. In: ANIMAL GENETICS AND BREEDING UNIT. AGBU. **Design of Livestock Breeding Programs**: Short course in animal Breeding. Armidale, 1993. part.4, chap.19, p.187-204.
- KINGHORN, B.P.; VERCOE, P.E. The effects of using the wrong genetic model to predict the merit of crossbred genotypes. **Animal Production**, v.49, p.209-216. 1989.
- KNIGHT, J.R.; HILL, R.C.; SIRMANS, C.F. Biased Prediction of Housing Values. **Journal of the American Real Estate and Urban Economics Association**, v. 20, n.3, p.427-456. 1992.
- LECLERC, G.; PIREAUX, J.J. The use of least squares for XPS peak parameters estimation. Part 3. Multicollinearity, ill-conditioning and constraint-induced bias. **Journal of Electron Spectroscopy and Related Phenomena**, v. 71, p.179-190. 1995.
- LIPOVETSKY, S.; CONKLIN, W.M. Multiobjective regression modifications for collinearity. **Computers & Operations Research**, v.28, p.1333-1345. 2001.
- MACKINNON, M.J.; THORPE, W.; BAKER, R.L. Sources of genetic variation for milk production in a crossbred herd in the tropics. **Animal Science**, v.62, p.5-16. 1996.
- MENDEL, J.G. **Experiments in Plant Hybridization**. 1865. Disponível em: <<http://www.mendelweb.org/Mendel.plain.html>>. Acesso em: 23 de agosto de 2003.
- MOORE, J.S.; REICHERT, A.K.; CHO, C.C. Analysing the Temporal Stability of Appraisal Model Coefficients: An Application of Ridge Regression Techniques. **Journal of the American Real Estate and Urban Economics Association**, v. 12, n.1, p.50-71. 1984.
- ÖZÇELİK, Y.; KULAKSIZ, S.; ÇETİN, M.C. Assessment of the wear of diamond beads in the cutting of different rock types by the ridge regression. **Journal of Materials Processing Technology**, v.127, p.392-400. 2002.
- ÖZTÜRK, F.; AKDENİZ, F. Ill-conditioning and multicollinearity. **Linear Algebra and its Applications**, v.321, p.295-305. 2000.
- PICCOLI, M.L.; ROSO, V.M.; BRITO, F.V.; SEVERO, J.L.P.; SCHENKEL, F.S.; FRIES, L.A. Additive, complementarity (additive\*additive), dominance, and epistatic effects on preweaning weight gain of Hereford x Nelore calves. In: WORLD CONGRESS ON

- GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** CD-ROM.
- RIKARDBSEN, A.H.; JOHANSEN, M. A morphometric method for estimation of total lipid level in live Arctic charr: a case study of its application on wild fish. **Journal of Fish Biology**, v.62, p.724-734. 2003.
- ROOK, A.J.; DHANOA, M.S.; GILL, M. Prediction of the voluntary intake of grass silages by beef cattle. **Animal Production**, v.50, p.439-454. 1990.
- RUTLEDGE, J.J. Greek temples, tropical kine and recombination load. **Livestock Production Science**, v.68, p.171-179. 2001.
- SCHAMBER, E.M.; MUIR, W.M. Wright's shifting balance theory of evolution in artificial breeding programmes: empirical testing using the model organism *Tribolium castaneum*. **Journal of Animal Breeding and Genetics**, v.118, p.181-188. 2001.
- SHARMA, B.S.; PRABHAKARAN, V.T.; PIRCHNER, F. Gene action and heterosis in lifetime traits of Friesian x Sahiwal crosses. **Journal of Animal Breeding and Genetics**, v.117, p.319-330. 2000.
- SHERIDAN, A.K. Crossbreeding and Heterosis. **Animal Breeding Abstracts**, v.49, n.3, p.131-144. 1981.
- SHI, L.; WANG, X. Local influence in ridge regression. **Computational Statistics & Data Analysis**, v.31, p.341-353. 1999.
- SIMIANER, H. Prospects for Third Generation Methods of Genetic Evaluation. In: ANNUAL MEETING OF THE EUROPEAN ASSOCIATION FOR ANIMAL PRODUCTION, 42., 1991, Berlin, **Proceedings...**
- SPANOS, A.; MCGUIRK, A. The problem of near-multicollinearity revisited: erratic vs systematic volatility. **Journal of Econometrics**, v.108, p.365-393. 2002.
- STINE, R.A. Graphical Interpretation of Variance Inflation Factors. **The American Statistician**, v.49, n.1, p.53-56. 1995.
- VAN GAANS, P.F.M.; VRIEND, S.P. Multiple linear regression with correlations among the predictor variables. Theory and computer algorithm RIDGE (Fortran 77). **Computers & Geosciences**, v.16, n.7, p.933-952. 1990.



- VINOD, H.D. Application of New Ridge Regression Methods to a Study of Bell System Scale Economies. **Journal of the American Statistical Association**, v.71, n.356, p.835-841. 1976.
- WAN, A.T.K. On generalized ridge regression estimators under collinearity and balanced loss. **Applied Mathematics and Computation**, v.129, p.455-467. 2002.
- WEIGEL, K.A.; GIANOLA, D.; TEMPELMAN, R.J.; MATOS, C.A.; CHEN, I.H.C.; WANG, T.; BUNGE, R.; LO, L.L. Improving Estimates of Fixed Effects in a Mixed Linear Model. **Journal of Dairy Science**, v.74, p.3174-3182. 1991.
- WILLHAM, R.L. Genetic consequences of crossbreeding. **Journal of Animal Science**, v.30, p.690-693. 1970.
- WILLHAM, R.L.; POLLAK, E. Symposium: heterosis and crossbreeding. **Journal of Dairy Science**, v.68, p.2411-2417. 1985.
- WRIGHT, S. The relation of livestock breeding to theories of evolution. **Journal of Animal Science**, v.46, n.5, p.1192-1200. 1978.

## CAPÍTULO 2 – UM PROGRAMA PARA IMPLEMENTAÇÃO DA REGRESSÃO DE CUMEEIRA USANDO O MÉTODO GIVENS DE DECOMPOSIÇÃO ORTOGONAL PARA OBTENÇÃO DAS SOLUÇÕES

**RESUMO** – A regressão de cumeeira é um dos métodos de estimação viesada mais utilizados para contornar os efeitos da multicolinearidade em análises de regressão. Neste trabalho, apresenta-se um programa para implementação da regressão de cumeeira, usando o método Givens de decomposição ortogonal para obtenção das soluções do sistema. A aplicação das rotações planares de Givens sobre a matriz de dados constituem um método capaz de conferir maior precisão às estimativas, e maior praticidade computacional quando se trabalha com atualizações sucessivas dos arquivos de dados, caso que se observa na implementação da técnica da regressão de cumeeira. O programa permite quatro implementações alternativas da regressão de cumeeira: uma utilizando os valores de  $\lambda$  definidos pelo usuário, e as outras aplicando três formas de ponderação de  $\lambda$ . Podem ser utilizados para essa ponderação os valores dos fatores de inflação de variância (FIVs), os valores da razão da soma de quadrados da correspondente covariável pela menor soma de quadrados, e o produto desses dois ponderadores. Nos quatro métodos, os valores são adicionados somente aos elementos da diagonal principal da matriz de coeficientes, correspondentes às covariáveis que apresentam valor de FIV superior ao definido pelo usuário como crítico. O programa gera quatro arquivos que podem auxiliar o usuário na escolha de  $\lambda$ : um contendo as estimativas dos parâmetros obtidas para cada valor de  $\lambda$  empregado (podendo ser utilizadas para construção do “traço ridge”); um com os valores da estatística t de Student; um com os valores de FIV; e um contendo os valores da soma e da média harmônica dos valores absolutos de t.

**Palavras-chave:** estimação viesada, multicolinearidade, rotações de Givens

## 2.1 Introdução

Análises de regressão são delineadas para situações em que uma variável está supostamente relacionada a uma ou mais medidas feitas, geralmente, no mesmo objeto. O propósito da análise é usar os dados (valores observados das variáveis) para estimar a forma dessa relação (SEARLE, 1971).

Assumindo-se que a variável resposta seja função da combinação linear de uma série de variáveis explicativas, essa relação pode ser representada, em notação matricial, por um modelo linear do tipo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{com} \quad \mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{e} \quad \mathbf{V}(\boldsymbol{\epsilon}) = \mathbf{I}\sigma_{\boldsymbol{\epsilon}}^2$$

Em que:

$\mathbf{y}$  é o vetor de observações da variável resposta, de ordem  $\mathbf{n} \times \mathbf{1}$ ;

$\mathbf{X}$  é a matriz de observações das variáveis explicativas, de ordem  $\mathbf{n} \times \mathbf{p}$ ;

$\boldsymbol{\beta}$  é o vetor dos parâmetros a serem estimados, de ordem  $\mathbf{p} \times \mathbf{1}$ ;

$\boldsymbol{\epsilon}$  é o vetor dos resíduos, de ordem  $\mathbf{n} \times \mathbf{1}$ ;

$\mathbf{n}$  é o número de observações;

$\mathbf{p}$  é o número de parâmetros a serem estimados.

Segundo VAN GAANS & VRIEND (1990), a regressão por quadrados mínimos ordinários é a técnica predominantemente aplicada para solução desse tipo de problema. Os autores atribuem isso à objetividade e relativa simplicidade do método, e à ampla disponibilidade de pacotes computacionais para esse fim.

O método de estimação por quadrados mínimos ordinários consiste em escolher  $\hat{\boldsymbol{\beta}}$  como o valor de  $\boldsymbol{\beta}$  que minimiza a soma dos quadrados dos desvios das observações de seus valores esperados. Esse processo envolve a derivação de  $\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$  em relação a  $\boldsymbol{\beta}$ . Igualando-se o resultado a zero, e escrevendo-se as equações resultantes em termos de  $\hat{\boldsymbol{\beta}}$ , tem-se:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Essas equações são mais conhecidas como as equações normais, e desde que  $(\mathbf{X}'\mathbf{X})^{-1}$  exista, elas têm solução única, dada por:

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

Formar e resolver as equações normais numericamente constituem geralmente uma maneira insatisfatória de solucionar problemas de quadrados mínimos lineares, em parte por causa da perda em precisão que resulta quando se forma a matriz de produtos cruzados  $(\mathbf{X}'\mathbf{X})$ , e em parte porque pode haver necessidade de adicionar ou remover observações, e é desejável que se faça proveito de cálculos já feitos (GENTLEMAN, 1985).

Certos métodos numéricos, como a decomposição da matriz de dados no produto de uma matriz ortogonal por uma matriz triangular superior, podem ser usados para minimizar a imprecisão das estimativas, e permitir maior praticidade de implementação computacional. Uma descrição detalhada de alguns desses métodos pode ser encontrada em LAWSON & HANSON (1974) e MAINDONALD (1984).

Métodos de decomposição ortogonal baseados em transformações de Givens (GIVENS, 1954) apresentam duas grandes vantagens: zeros na matriz de delineamento representam ganhos substanciais na solução do problema, e a matriz de delineamento pode ser naturalmente processada por linhas (GENTLEMAN, 1985).

Uma pressuposição do modelo de regressão linear clássico é a de que não existam relações lineares exatas entre os valores observados das variáveis explicativas. Se existirem, o vetor de parâmetros  $\hat{\boldsymbol{\beta}}$  não é estimável. Na prática, uma relação linear exata é altamente improvável, mas é comum uma interdependência geral entre as variáveis que explicam determinado fenômeno. À existência de correlações entre as variáveis explicativas, dá-se o nome de multicolinearidade (GOLDBERGER, 1964).

A presença de multicolinearidade torna difícil a determinação e a interpretação da contribuição individual da cada variável explicativa, leva a estimativas com altas variâncias e erros-padrão, e pode produzir equações de predição inapropriadas. Nesses casos, métodos alternativos à estimação por quadrados mínimos ordinários são

indicados para contornar os efeitos da multicolinearidade. A regressão de cumeeira é um desses métodos.

Este trabalho apresenta um programa para implementação da regressão de cumeeira, usando o método das rotações planares de Givens para obtenção das soluções do sistema.

## 2.2 Teoria

### 2.2.1 Regressão de cumeeira

O método da regressão de cumeeira (HOERL, 1962) consiste na adição de um escalar ( $\lambda$ ) aos elementos da diagonal principal da matriz de coeficientes, visando quebrar as dependências lineares verificadas entre suas colunas. O estimador torna-se:

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

Originalmente, HOERL & KENNARD (1970) propuseram um método gráfico, denominado “traço ridge”, para escolha do valor de  $\lambda$ . Por esse método, as estimativas dos parâmetros são plotadas contra diversos valores de  $\lambda$ , partindo de zero. Escolhe-se, então, o valor de  $\lambda$  a partir do qual se pode verificar uma estabilização nos valores das estimativas.

Em virtude da subjetividade desse método, diversos critérios matemáticos têm sido propostos para uma determinação mais objetiva do valor de  $\lambda$ . Porém, segundo MONTGOMERY & PECK (1982), citados por VAN GAANS & VRIEND (1990), nenhum desses critérios apresentou, efetivamente, alguma superioridade.

A identificação do problema de multicolinearidade em análises de regressão geralmente é feita pelo exame das correlações entre pares de variáveis explicativas. Porém, em alguns casos, associações entre três ou mais variáveis podem não ser detectadas pelo exame de correlações aos pares. Uma forma mais eficiente de

diagnóstico de multicolinearidade é o exame dos fatores de inflação de variância (FIVs), definidos por:

$$FIV(x_i) = \frac{1}{1 - R_i^2}$$

Em que  $R_i^2$  é o quadrado do coeficiente de correlação múltipla que resulta da regressão da variável  $x_i$  contra todas as outras variáveis explicativas incluídas no modelo.

As reduções nos valores dos FIVs, observadas para os diversos valores de  $\lambda$  testados, podem servir de referencial para que se escolha o valor mínimo de  $\lambda$  para os quais se verificam valores aceitáveis de FIV.

FRIES *et al.* (2000) sugerem que diversos valores de  $\lambda$  sejam testados até o ponto em que a redução nos valores absolutos das estimativas não seja mais compensada por reduções no FIV e nos erros-padrão. A divisão dessas estimativas pelos respectivos erros-padrão resulta nos valores de t de Student, cuja soma dos valores absolutos fornece a estatística “soma( | t | )”. Essa função de t não é monotônica em  $\lambda$ , e pode ser um referencial mais pragmático e informativo para definição de seu valor do que o método gráfico. Segundo PICCOLI *et al.* (2002), a função “média harmônica dos valores absolutos de t (HARM\_AT)” pode ajudar a identificar o menor valor possível de  $\lambda$  que produza a maior uniformidade em termos de resultados estatisticamente significativos.

Segundo HARVILLE (1983), citado por FRIES *et al.* (2000), uma forma de fazer com que programas de análise de regressão implementem a regressão de cumeieira, consiste em adicionar pseudo-observações. Se a matriz de dados original é  $[X \ y]$  com p variáveis explicativas, então adicionam-se p pseudo-observações  $[S \ 0]$  para formar:

$$\begin{bmatrix} X_{n \times p} & y_{n \times 1} \\ S_{p \times p} & 0_{p \times 1} \end{bmatrix}$$

De maneira que:

$$\begin{bmatrix} X'_{p \times n} & S'_{p \times p} \\ y'_{1 \times n} & 0'_{1 \times p} \end{bmatrix} \begin{bmatrix} X_{n \times p} & y_{n \times 1} \\ S_{p \times p} & 0_{p \times 1} \end{bmatrix} = \begin{bmatrix} X'X + S'S & X'y \\ y'X & y'y \end{bmatrix}$$

Fazendo-se  $S'S = \lambda I$ , chega-se ao estimador da regressão de cumeeira. Quando se implementa a regressão de cumeeira pela adição de pseudo-observações, torna-se interessante o uso de um método numérico capaz de processar os dados por linha, como é o caso das rotações de Givens.

### 2.2.2 Rotações de Givens

Um dos principais métodos para solução de problemas de quadrados mínimos é o das rotações de Givens, usam o fato de que  $\| X\hat{\beta} - y \| = \| QX\hat{\beta} - Qy \|$ , em que  $Q$  é uma matriz ortogonal, e por meio de uma série de pré-multiplicações (cujo produto denota-se por  $Q'$ ) reduz  $X_{n \times p}$  a:

$$Q'X = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

Em que  $R$  é uma matriz triangular superior de ordem  $p \times p$ . A esse processo dá-se o nome de decomposição QR de  $X$ . Uma vez que a redução para a forma triangular superior se completa,  $\hat{\beta}$  pode ser obtido por retro-substituição resolvendo-se  $R\hat{\beta} = c$ , em que:

$$Q'y = \begin{bmatrix} c \\ d \end{bmatrix} \quad (\text{EĞECIOĞLU \& SRINIVASAN, 1995}).$$

Transformações de Givens são indicadas para os casos em que se deseja precisão numérica ou transformar apenas um elemento de um vetor  $v$  em zero. Como essa transformação modifica apenas dois componentes de um vetor, uma descrição do método pode ser feita pela operação de um vetor de ordem 2.

*Dado um vetor de ordem 2:  $v = (v_1, v_2)'$ , com  $v_1 \neq 0$  ou  $v_2 \neq 0$ . Existe uma matriz ortogonal  $2 \times 2$ :*

$$G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

$$\text{com } c^2 + s^2 = 1$$

*tal que:*

$$Gv = \begin{bmatrix} (v_1^2 + v_2^2)^{1/2} \\ 0 \end{bmatrix}.$$

*Para tanto, definem-se  $c$  e  $s$  como:*

$$c = \frac{v_1}{(v_1^2 + v_2^2)^{1/2}} \quad \text{e} \quad s = \frac{v_2}{(v_1^2 + v_2^2)^{1/2}} \quad (\text{LAWSON \& HANSON, 1974}).$$

Aplicando-se essas rotações em todos os elementos da matriz de dados, linha por linha, zeram-se todos os elementos necessários para que se chegue a uma matriz triangular superior. Dada uma matriz  $X$ , de ordem  $n \times p$ , se  $n > p$ , soluções de quadrados mínimos podem, então, ser obtidas do sistema triangular superior por retro-substituição.



Uma das grandes vantagens desse método reside na possibilidade de armazenagem da triangular superior em arquivo, para análises posteriores. Como os dados são operados por linha, a adição de novas linhas de dados ao problema requer que apenas os novos registros sejam rotados contra a triangular superior já armazenada.

## **2.3 Estrutura do programa**

O programa apresentado neste trabalho é uma adaptação do sistema GIVENS (FRIES, 1987) para solução de modelos lineares fixos ou mistos por rotações planares de Givens aplicadas sobre a matriz de dados. O sistema GIVENS retorna soluções de quadrados mínimos para o problema apresentado.

Trechos de algoritmos em linguagem Fortran 90 foram incorporados ao programa fonte do sistema GIVENS, de modo a gerar soluções de regressão de cumeieira. A adaptação foi feita com o objetivo de aliar a implementação da técnica de regressão de cumeieira à praticidade computacional apresentada pelo método Givens de decomposição ortogonal.

### **2.3.1 Entrada de dados**

O arquivo de entrada requerido pelo programa consiste da matriz de valores observados para todas as variáveis explicativas e dependentes, constantes do modelo em estudo. Na tela, o programa pede que sejam informados: os formatos de entrada dos dados; o nome do arquivo de dados; e os nomes das variáveis explicativas e dependentes.

A informação do formato de entrada dos dados deve seguir as convenções da linguagem Fortran 77. Deve-se abrir e fechar o formato com um par de parênteses, usando-se no máximo 127 colunas para explicitar a informação. O arquivo de dados

deve ser informado com o diretório e o nome, seguido da extensão. Os nomes das variáveis explicativas e dependentes devem ser fornecidos na mesma ordem em que aparecem no arquivo de dados.

O sistema permite a análise de até cinco respostas (variáveis dependentes), conforme modelos contendo até 40 covariáveis, e até cinco variáveis classificatórias. A primeira variável classificatória pode conter até 5000 níveis, a segunda até 200 e da terceira à quinta, até 20 níveis.

Na tela, o programa também pede que sejam informados o valor inicial e o valor final de  $\lambda$  a serem testados, bem como o valor de FIV a partir do qual se deseja que os coeficientes  $\lambda$  sejam aplicados. Essa opção permite que só sejam adicionados valores aos elementos da matriz de coeficientes cujas correspondentes variáveis estão envolvidas em colinearidade, sendo esta uma vantagem em relação a outros programas que implementam a regressão de cumeeira (ex.: SAS – Statistical Analysis System).

### **2.3.2 Operações executadas pelo programa**

Ao final do fornecimento de todas as informações requeridas pelo programa, o arquivo de dados é lido e rotações de Givens são aplicadas na matriz de dados, linha por linha.

Lidas e rotadas todas as linhas de dados, a matriz triangular superior formada pelas rotações é armazenada em arquivo.

A técnica da regressão de cumeeira, implementada pelo programa, testa 101 valores de  $\lambda$ , do valor inicial ao valor final determinados pelo usuário, acrescidos a cada rodada da centésima parte da diferença entre os dois valores. Independente do valor inicial de  $\lambda$  informado pelo usuário, o programa faz uma primeira análise com  $\lambda = 0$ , para que se conheçam os valores das estimativas dos parâmetros e dos fatores de inflação de variância, obtidos por quadrados mínimos ordinários. Os valores de FIV para cada covariável são calculados como o produto do elemento da diagonal de  $X'X$  da

respectiva covariável pelo elemento da diagonal da inversa de  $X'X$  correspondente (MAINDONALD, 1984).

Para cada valor de  $\lambda$  empregado, uma nova análise de regressão é conduzida. A adição dos valores de  $\lambda$  aos elementos da diagonal principal da matriz de coeficientes é feita pela inclusão de uma matriz de pseudo-observações  $[S \ 0]$ , em que:  $S$  é uma matriz diagonal, de ordem igual à de  $X'X$ , cujos valores da diagonal correspondem à raiz quadrada de  $\lambda$ ; e  $0$  é um vetor de zeros, de ordem igual à de  $S$ .

A cada novo valor de  $\lambda$  testado, apenas as matrizes de pseudo-observações são lidas e rotadas contra a triangular superior armazenada em arquivo.

O programa oferece quatro métodos alternativos de implementação da regressão de cumeira: (1) adição dos valores de  $\lambda$  definidos para o intervalo informado pelo usuário; (2) ponderação dos valores de  $\lambda$  pela razão da soma de quadrados da correspondente covariável pela menor soma de quadrados – ajustando para a magnitude dos valores dos elementos da matriz de coeficientes; (3) ponderação de  $\lambda$  pelos valores de FIV apresentados para cada covariável – ajustando para o grau de envolvimento de cada covariável no problema de multicolinearidade; e (4) ponderação de  $\lambda$  pelo produto da razão de soma de quadrados pelo FIV – ajustando para magnitude e grau de envolvimento em multicolinearidade, simultaneamente. Nos quatro métodos, os valores de  $\lambda$  são adicionados somente aos elementos da matriz de coeficientes correspondentes às covariáveis que apresentam FIV superior ao valor crítico definido pelo usuário.

### 2.3.3 Arquivos de saída

Além dos arquivos de saída gerados originalmente pelo sistema GIVENS, quatro arquivos são criados no intuito de auxiliar o usuário na escolha de  $\lambda$ : três contendo os valores das estimativas dos parâmetros (beta.out), dos FIVs (fiv.out) e dos valores de  $t$

de Student (tes.out), para cada valor de  $\lambda$  testado; e um contendo a soma e a média harmônica dos valores absolutos de t (estt.out), para cada valor de  $\lambda$ .

O formato de saída dos três primeiros arquivos citados pode ser visto na Figura 1, em que var\_1, var\_2 e var\_3 correspondem aos nomes das variáveis, fornecidos via teclado.

lambda	var_1	var_2	var_3
.00000	44.01738	-38.72578	37.98986
.00000	44.01738	-38.72578	37.98986
.01000	44.23532	-38.95546	38.18827
.02000	44.24528	-38.93964	38.18006
.03000	44.25742	-38.93110	38.17506
.04000	44.26813	-38.92521	38.17075
.05000	44.27739	-38.92081	38.16700
.06000	44.28547	-38.91742	38.16380
.07000	44.29259	-38.91474	38.16109
.08000	44.29892	-38.91259	38.15882
.09000	44.30461	-38.91084	38.15690
.10000	44.30975	-38.90941	38.15528

Figura 1. Formato de saída dos arquivos que contêm os valores das estimativas dos parâmetros, dos FIVs e dos valores de t de Student, de acordo com o valor de  $\lambda$  empregado.

O formato de saída do arquivo “estt.out”, que contém os valores das estatísticas baseadas em funções dos valores de t de Student, é apresentado na Figura 2.

lambda	sum(abs)t	harm(abs)t
.00000	50.73994	.54233
.00000	50.73994	.54233
.01000	60.63769	.71276
.02000	62.81973	.54440
.03000	63.20676	.33784
.04000	62.52417	.65621
.05000	61.36139	.67865
.06000	60.00588	.62207
.07000	58.59187	.53422
.08000	57.18511	.43072
.09000	55.81808	.31742
.10000	54.50618	.19632

Figura 2. Formato de saída do arquivo que contém os valores da soma e da média harmônica dos valores absolutos da estatística t de Student, de acordo com o valor de  $\lambda$  empregado.

## 2.4 Aplicações em problemas de melhoramento animal

Incorporações de novos registros aos arquivos de dados, analisados a cada safra, são rotinas freqüentes em programas de melhoramento animal. A possibilidade de obtenção de novas soluções, com necessidade de que apenas as novas linhas de dados sejam lidas e rotadas contra uma matriz triangular superior armazenada em arquivo, pode representar economia em tempo de processamento e demanda por memória convencional. Essa economia pode ser particularmente relevante em problemas que envolvem grande volume de informação, como os que se observam em tais programas.

Análises de regressão são procedimentos amplamente empregados na área de melhoramento animal. Em muitas situações, a presença de multicolinearidade torna difícil a interpretação dos coeficientes estimados. A regressão de cumeieira é uma ferramenta que possibilita a obtenção de estimativas interpretáveis dos coeficientes de regressão, mesmo na presença de forte multicolinearidade.

## Referências

- EĞECIOĞLU, Ö.; SRINIVASAN, A. Givens and Householder Reductions for Linear Least Squares on a Cluster of Workstations. In: INTERNATIONAL CONFERENCE ON HIGH PERFORMANCE COMPUTING (HiPC), 1995, New Delhi, **Proceedings...** p.734-739.
- FRIES, L.A. Mixed Model Methodology for within-herd beef cattle selection, using Givens Rotations and a microcomputer. In: INTERNATIONAL CONFERENCE ON QUANTITATIVE GENETICS, 2., 1987, Raleigh, **Proceedings...** p.38.
- FRIES, L.A.; GRASER, H.U.; JOHNSTON, D.J.; HEARNshaw, H. Using Ridge Regression to Estimate Genetic Effects in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, supl. B, p. 241, 2000.

- GENTLEMAN, W.M. Basic Procedures for Large, Sparse or Weighted Linear Least Squares Problems. In: GRIFFITHS, P.; HILL, I.D. (Ed.). **Applied Statistics Algorithms**. Chichester: Ellis Horwood series in mathematics and its applications, 1985. p.130-144.
- GIVENS, W. Numerical Computation of the Characteristic Values of a Real Symmetric Matrix. **Oak Ridge National Laboratory Report**, Oak Ridge, ORNL – 1574, 1954.
- GOLDBERGER, A.S. **Econometric Theory**. New York: Wiley Series in Probability and Mathematical Statistics, 1964. 399p.
- HOERL, A.E. Application of Ridge Analysis to Regression Problems. **Chemical Engineering Progress**, v.58, p.54-59. 1962.
- HOERL, A.E.; KENNARD, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v.12, n.1, p.55-67. 1970.
- LAWSON, C.L.; HANSON, R.J. **Solving Least Squares Problems**. New Jersey: Prentice-Hall series in automatic computation, 1974. 340p.
- MAINDONALD, J.H. **Statistical computation**. New York: Wiley series in probability and mathematical statistics, 1984. 370p.
- PICCOLI, M.L.; ROSO, V.M.; BRITO, F.V.; SEVERO, J.L.P.; SCHENKEL, F.S.; FRIES, L.A. Additive, complementarity (additive\*additive), dominance, and epistatic effects on preweaning weight gain of Hereford x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** CD-ROM.
- SEARLE, S.R. **Linear Models**. New York: Wiley series in probability and mathematical statistics, 1971. 532p.
- VAN GAANS, P.F.M.; VRIEND, S.P. Multiple linear regression with correlations among the predictor variables. Theory and computer algorithm RIDGE (Fortran 77). **Computers & Geosciences**, v.16, n.7, p.933-952. 1990.

### **CAPÍTULO 3 – USO DA REGRESSÃO DE CUMEEIRA PARA REDUÇÃO DOS EFEITOS DA MULTICOLINEARIDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS DE CORTE**

**RESUMO** – Uma alternativa para redução dos efeitos da multicolinearidade em análises de regressão foi abordada por meio da utilização da regressão de cumeeira para estimação de parâmetros de efeitos genéticos em bovinos de corte. O modelo utilizado relacionava o ganho médio diário pré-desmama de bezerros Hereford x Nelore a efeitos diretos e maternos de ação aditiva, dominância, epistasia e complementariedade. Cinco métodos de implementação da regressão de cumeeira foram testados: o método proposto originalmente; o implementado pelo SAS; e mais três formas diferentes de ponderação do coeficiente  $\lambda$ . Testaram-se, como critérios matemáticos para escolha de  $\lambda$ , a soma e a média harmônica dos valores absolutos da estatística t de Student, e o valor de  $\lambda$  a partir do qual os valores dos fatores de inflação de variância passavam a ser todos menores que 300. Superfícies de predição foram construídas para comparações entre os cinco métodos e os três critérios. A regressão de cumeeira pôde fornecer superfícies de predição bem diferentes das obtidas por quadrados mínimos ordinários, e com interpretações biológicas bem mais aceitáveis. Para o modelo e o conjunto de dados deste estudo, a ponderação de  $\lambda$  pelo FIV se mostrou um método mais robusto, quanto ao critério para escolha de  $\lambda$ . Superfícies de predição biologicamente coerentes puderam ser obtidas em todos os métodos, usando-se o critério baseado no FIV. Recomenda-se que critérios matemáticos sejam usados como ferramenta auxiliar para escolha de  $\lambda$ , não dispensando o exame dos sinais e valores das estimativas (ou valores preditos) e um bom entendimento do fenômeno em estudo.

**Palavras-chave:** complementariedade, epistasia, estimador viesado, heterose

## Introdução

Muitos problemas em melhoramento animal requerem a predição de uma variável, a variável dependente, como uma função de diversas outras, os preditores, regressores ou variáveis explicativas. O método mais popular de se ajustar as equações de predição é a regressão por quadrados mínimos. Contudo, quando algumas variáveis preditoras são altamente correlacionadas, os coeficientes de regressão por quadrados mínimos, embora não viesados, podem ter altas variâncias (BERGMANN & HOHENBOKEN, 1995).

O termo colinearidade se refere à associação, medida como uma correlação, entre duas variáveis explicativas. Multicolinearidade se refere à correlação entre três ou mais variáveis, embora os dois termos sejam usados indistintamente. Seus efeitos nos resultados de uma análise de regressão são substanciais. A multicolinearidade impõe um limite ao coeficiente de determinação e torna difícil o incremento em qualidade de predição pela adição de variáveis. Torna-se difícil também a determinação da contribuição individual de cada variável explicativa, porque seus efeitos são confundidos em consequência da colinearidade. Os coeficientes de regressão podem ser estimados incorretamente e apresentarem sinais trocados (HAIR JR. *et al.*, 1992).

Diversos critérios têm sido usados no diagnóstico de multicolinearidade. DRAPER & SMITH (1998) sugerem cinco passos: (1) verificar se alguns coeficientes estão com sinal trocado, com base em conhecimento prévio; (2) verificar se regressores importantes, com base em conhecimento prévio, apresentam pequenos valores da estatística *t* de Student, que testa a hipótese de nulidade das estimativas dos coeficientes; (3) verificar se a exclusão de uma linha ou coluna da matriz de dados provoca mudanças bruscas no modelo ajustado; (4) verificar as correlações entre todos os pares de variáveis explicativas; e (5) examinar os fatores de inflação de variância (FIVs).

Quando há suspeita de multicolinearidade, recomendam-se métodos alternativos de análise, como a regressão de cumeira (HOERL & KENNARD, 1970). Os coeficientes de regressão estimados resultantes podem sugerir uma nova interpretação



dos dados e, assim, levar a um melhor entendimento do processo em estudo (CHATTERJEE & PRICE, 1991).

O procedimento de análise por regressão de cumeeira consiste na adição de uma pequena quantidade positiva ( $\lambda$ ) aos elementos da diagonal principal da matriz de coeficientes, reduzindo assim as dependências lineares entre suas colunas. O vetor de estimativas passa então a ser obtido por:  $\beta^* = (X'X + \lambda I)^{-1} X'Y$ , em que  $I$  é uma matriz identidade de ordem igual à de  $X'X$  (HOERL & KENNARD, 1970). O problema na regressão de cumeeira é que o valor de  $\lambda$  é definido por meio de argumentos relativamente subjetivos (FREUND & LITTELL, 2000).

Originalmente, um método gráfico, denominado “traço ridge” (HOERL & KENNARD, 1970), foi proposto como critério para adoção do coeficiente  $\lambda$ . Valores das estimativas dos regressores são plotados contra  $\lambda$ , até que se identifique o ponto a partir do qual as estimativas se estabilizem, e coeficientes com sinal trocado já tenham mudado de sinal.

Diversos trabalhos têm sido conduzidos no intuito de determinar critérios claros e objetivos para escolha do coeficiente  $\lambda$ . Segundo MONTGOMERY & PECK (1982), citados por VAN GAANS & VRIEND (1990), nenhum dos critérios matemáticos propostos para seleção de  $\lambda$  apresentaram alguma superioridade. O método gráfico tem sido o mais utilizado pela facilidade na identificação de mudanças nos coeficientes.

Algumas implementações alternativas da regressão de cumeeira também têm sido propostas, como a adição de  $\lambda$  somente aos elementos da diagonal correspondentes a variáveis envolvidas em colinearidade (FOUCART, 1999), e a aplicação de diferentes valores de  $\lambda$  (LIPOVETSKY & CONKLIN, 2001).

O objetivo do presente trabalho foi testar diferentes implementações da regressão de cumeeira, bem como diferentes critérios para escolha do coeficiente  $\lambda$ , para reduzir os efeitos da multicolinearidade na estimação de parâmetros de efeitos genéticos em bovinos de corte.

## Material e métodos

Foram utilizados dados do programa de melhoramento de gado de corte Conexão Delta G, preparados pela empresa GenSys Consultores Associados S/C Ltda. O banco de dados da Conexão Delta G contém registros de 109.614 bezerros Hereford x Nelore nascidos entre 1974 e 1999, distribuídos em 4.665 grupos de contemporâneos, e criados em 29 fazendas localizadas nos estados de Mato Grosso, Goiás, Mato Grosso do Sul, São Paulo, Paraná e Rio Grande do Sul, entre as latitudes 9°S e 31°S.

As estimativas dos parâmetros de efeitos de ação aditiva direta (**aa**) e materna (**am**), heterozigose direta (**ha**) e materna (**hm**), epistasia direta (**ea**) e materna (**em**), e complementariedade direta (**aaz**) e materna (**amz**), no ganho de peso médio diário do nascimento à desmama (**gmdnd**), foram obtidas pelo seguinte modelo:

$$y = X\beta + Wg + \epsilon$$

onde:

**y** é o vetor de observações do **gmdnd**;

**X** é a matriz de incidência dos efeitos fixos ambientais;

**$\beta$**  é o vetor dos parâmetros dos efeitos fixos ambientais;

**W** é a matriz de incidência dos efeitos genéticos fixos e interações;

**g** é o vetor dos parâmetros dos efeitos genéticos fixos e interações;

**$\epsilon$**  é o vetor de efeitos residuais aleatórios.

Covariáveis para **aa** e **am** foram definidas pela contribuição dos genes Nelore na composição genética de cada indivíduo. Covariáveis para **aaz** e **amz** entre Nelore e Hereford foram calculadas como sendo **aaz=aa\*(1-aa)** e **amz=am\*(1-am)**, respectivamente (BRITO *et al.*, 2002; PICCOLI *et al.*, 2002). As fórmulas usadas para calcular e decompor heterozigose direta e materna foram as mesmas usadas por BÉRTOLI (1991) e SCHENKEL (1993). Para o cálculo, foram necessárias as

composições genéticas dos pais e avôs maternos dos indivíduos. Quando não se dispunha de informação sobre a composição genética dos ancestrais maternos, foi assumido que o animal foi produzido por acasalamento inter se. Os efeitos de **ea** e **em** foram modelados por epistazigose como em FRIES *et al.* (2000a) e calculados como a heterozigose média presente nos gametas que geraram cada indivíduo ou como a heterozigose média nos pais de um indivíduo.

Os efeitos ambientais incluídos no modelo foram: grupo de contemporâneos, efeito linear e quadrático de idade da vaca de acordo com o sexo do bezerro, efeito linear e quadrático de idade média à desmama, e polinômio segmentado quadrático-quadrático de data juliana de nascimento.

Em virtude do número de classes do efeito de grupo de contemporâneos, e do longo tempo de processamento que a sua inclusão no modelo impunha à análise, optou-se por pré-absorver esse efeito. A inclusão de efeitos principais de latitude (linear e quadrático) no modelo, resultou em estimativas nulas desses coeficientes. Como todos os indivíduos do mesmo grupo de contemporâneos estão na mesma latitude, o efeito de latitude foi removido pela absorção do efeito de grupo de contemporâneos. Optou-se, então, por incluir o efeito de latitude como um agente modificador, na forma de interações dos efeitos genéticos com efeitos lineares (**L1**) e quadráticos (**L2**) de latitude. Assim, o modelo final foi definido sem intercepto, e com todas as variáveis explicativas, totalizando trinta e três, na forma de covariáveis, expressas como desvio da média do grupo de contemporâneos.

O diagnóstico de multicolinearidade foi feito pelo exame dos fatores de inflação de variância (FIVs), dos autovalores da matriz de coeficientes e do índice de condição.

Em muitos casos, a multicolinearidade advém de correlações entre três ou mais variáveis, e o exame das correlações aos pares pode não revelar tais dependências. Por ser uma função do coeficiente de correlação múltipla entre uma variável explicativa com todas as outras incluídas no modelo, o FIV pode ser considerado uma ferramenta bem mais informativa do que o exame das correlações aos pares das variáveis explicativas. Os valores de FIV para cada covariável foram calculados como o produto

do elemento da diagonal de  $X'X$  da respectiva covariável pelo elemento da diagonal da inversa de  $X'X$  correspondente (MAINDONALD, 1984).

A condição de colinearidade também pode ser indicada pela presença de um ou mais autovalores com valor muito próximo de zero. De fato, se o determinante de uma matriz é dado pelo produto de seus autovalores, a presença de um ou mais autovalores com valor próximo de zero implica em uma matriz próxima da singularidade (HOCKING, 2003). O índice de condição, que se define como a raiz quadrada da razão do maior pelo menor autovalor da matriz de correlações, é uma estatística amplamente usada para verificação de autovalores muito baixos. Índices de condição superiores a 30 podem ser considerados como indicativo de forte multicolinearidade (KNIGHT *et al.*, 1992). Tanto os autovalores da matriz de correlações quanto o índice de condição foram obtidos usando-se a opção COLLINOINT do comando PROC REG implementado pelo SAS (Statistical Analysis System).

Um programa para solução de regressão linear múltipla por regressão de cumeeira foi desenvolvido em linguagem Fortran 90. Testaram-se cinco métodos de implementação da regressão de cumeeira:

$$(1) \hat{\beta}_{RC} = (X'X + \lambda I)^{-1} X'y$$

Implementação original do método da regressão de cumeeira, proposto por HOERL & KENNARD (1970);

$$(2) \hat{\beta}_{SAS} = [R_{XX}(I + D_\lambda)]^{-1} R_{Xy}$$

Procedimento implementado pelo SAS, conforme descrito por FREUND & LITTELL (2000), em que  $R_{XX}$  é a matriz de correlações das variáveis explicativas,  $R_{Xy}$  é o vetor de correlações das variáveis explicativas com a variável resposta, e  $D_\lambda$  é uma matriz diagonal com valores de  $\lambda$  na diagonal.

$$(3) \hat{\beta}_{RCW} = (X'X + \lambda W)^{-1} X'y$$

Em que  $W$  é uma matriz diagonal cujos elementos da diagonal são iguais aos de  $X'X$  divididos pelo menor elemento da diagonal de  $X'X$ . Nesse método, os valores de  $\lambda$  foram ponderados pelas razões da soma de quadrados da covariável pela menor soma de quadrados na matriz de coeficientes, visando ajustar o valor adicionado em função da magnitude do elemento de  $X'X$ .

$$(4) \hat{\beta}_{FIV} = (X'X + \lambda F)^{-1} X'y$$

Em que  $F$  é uma matriz diagonal, tal que:

$$F = \{FIV_i\} \text{ se } FIV_i \geq 300;$$

$$F = \{0\} \text{ caso contrário.}$$

Nesse método, os valores de  $\lambda$  foram ponderados pelos valores de FIV das correspondentes covariáveis. Foram usados para essa ponderação os valores de FIV encontrados com  $\lambda$  igual a zero.

$$(5) \hat{\beta}_{CFW} = (X'X + \lambda C)^{-1} X'y$$

Em que  $C$  é uma matriz diagonal cujos elementos da diagonal são iguais aos produtos dos correspondentes elementos de  $W$  pelos de  $F$ . Nesse método, os valores de  $\lambda$  foram ponderados pelo produto da razão de soma de quadrados descrita em (3) pelo FIV das correspondentes covariáveis. Por meio dessa combinação dos métodos (3) e (4), buscou-se ajustar o valor de  $\lambda$  para a magnitude do elemento de  $X'X$  e o grau de envolvimento da variável em colinearidade, simultaneamente.

Nas implementações (1), (3), (4) e (5) o coeficiente  $\lambda$  foi aplicado somente às covariáveis genéticas com valor de FIV superior a 300 (LECLERC & PIREAUX, 1995). Essa opção pode evitar a imposição de um viés desnecessário às estimativas de parâmetros não envolvidos em colinearidade.

Os critérios estudados para determinação do valor de  $\lambda$  foram baseados nos valores de FIV e em funções dos valores da estatística t de Student observados para as variáveis envolvidas em multicolinearidade. Essas funções foram a soma (**soma (abs)t**) e a média harmônica (**harm (abs)t**) dos valores absolutos de t (FRIES *et al.*, 2000b; PICCOLI *et al.*, 2002). Um valor máximo da soma dos valores absolutos de t pode indicar o máximo que se pode conseguir em redução nos erros padrão das estimativas, com o menor valor de  $\lambda$  possível. Um maior valor de média harmônica pode indicar maior uniformidade em significância estatística das estimativas dos parâmetros. O critério baseado nos valores de FIV constou da adoção do valor de  $\lambda$  a partir do qual todos os valores de FIV eram menores que trezentos.

A presença, no modelo considerado, de parâmetros para os efeitos de interação entre os efeitos genéticos e os de latitude, impossibilitou uma interpretação mais clara dos sinais e valores das estimativas dos coeficientes individuais. As comparações entre os cinco tipos de implementação e os três critérios para adoção de  $\lambda$  foram, então, feitas tomando-se como referencial as superfícies de predições obtidas usando-se as soluções estimadas por cada tipo e para cada valor de  $\lambda$ . Predições de **gmdnd** foram feitas para nove genótipos de gerações avançadas, do puro Hereford ao puro Nelore, de oitavo em oitavo, e em seis latitudes diferentes: 16, 19, 22, 25, 28 e 31°S. Um valor correspondente aos efeitos ambientais médios (0,61kg/dia) foi fixado para todas as predições.

É importante salientar que as covariáveis para os efeitos genéticos aditivos foram definidas como a proporção de genes Nelore no genótipo de cada indivíduo. Logo, todos os coeficientes para as covariáveis genéticas para indivíduos 100% Hereford são iguais a zero. Isso explica porque todas as predições para indivíduos 100% Hereford são iguais ao valor fixado para os efeitos ambientais.

## Resultados e discussão

Na Tabela 1 são apresentados os valores das estatísticas utilizadas para o diagnóstico de multicolinearidade. Constan na tabela os valores observados para as covariáveis genéticas e suas interações com latitude, efeitos para os quais o problema de colinearidade se mostrou mais evidente.

Tabela 1. Autovalores da matriz de coeficientes, correspondentes às respectivas covariáveis, números de condição e fatores de inflação de variância (FIV).

Efeito*	Autovalor	Número de condição	FIV
aa	0,1916200	6,5	1980
am	0,1488400	7,3	2277
ha	0,0825100	9,8	17494
hm	0,0808100	9,9	33738
aaz	0,0535600	12,2	27136
amz	0,0350800	15,1	76792
ea	0,0315600	15,9	5705
em	0,0311700	16,0	30429
aal1	0,0240400	18,2	7901
aml1	0,0198300	20,1	11888
aazl1	0,0111200	26,8	114879
amzl1	0,0081800	31,2	371906
hal1	0,0042400	43,4	77754
hml1	0,0026200	55,2	201750
eal1	0,0013400	77,1	38916
eml1	0,0006179	113,6	169839
aal2	0,0002692	172,1	2471
aml2	0,0002039	197,7	4275
aazl2	0,0001241	253,4	35792
amzl2	0,0000734	329,6	125496
hal2	0,0000429	430,9	24763
hml2	0,0000350	477,6	76658
eal2	0,0000031	1608,0	15620
eml2	0,0000009	<b>2969,0</b>	61884

\* aa, am, ha, hm, aaz, amz, ea, em, aal1, aml1, hal1, hml1, aazl1, amzl1, eal1, eml1, aal2, aml2, hal2, hml2, aazl2, amzl2, eal2 e eml2 correspondem aos efeitos genéticos diretos e maternos de ação aditiva, heterose, complementariedade e epistasia, e suas interações com efeitos linear (l1) e quadrático (l2) de latitude, respectivamente.

Na segunda coluna verificam-se 13 autovalores menores que 0,01, sendo nove menores que 0,001, revelando o quanto a matriz de correlações se aproxima da singularidade.

Na terceira coluna são apresentados os números de condição, que se definem como a raiz quadrada da razão entre o maior autovalor pelo autovalor da respectiva covariável. O último número da coluna corresponde ao índice de condição, obtido pela raiz quadrada da razão entre maior e o menor autovalor. Observou-se um índice de condição igual a 2969, ou seja, quase 100 vezes maior do que o considerado crítico (30), indicando a presença de severa multicolinearidade.

Os valores dos fatores de inflação de variância (FIV) são bastante elevados, alguns chegando à casa das centenas de milhares, revelando a intensa associação existente entre as variáveis explicativas.

Na Figura 1 são apresentadas as mudanças observadas nas estimativas dos parâmetros e nos respectivos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (1) de implementação da regressão de cumeieira. Os valores de  $t$  sofrem aumento gradativo, em módulo, até certo ponto, passando a regredir a partir de então (Figura 1b).



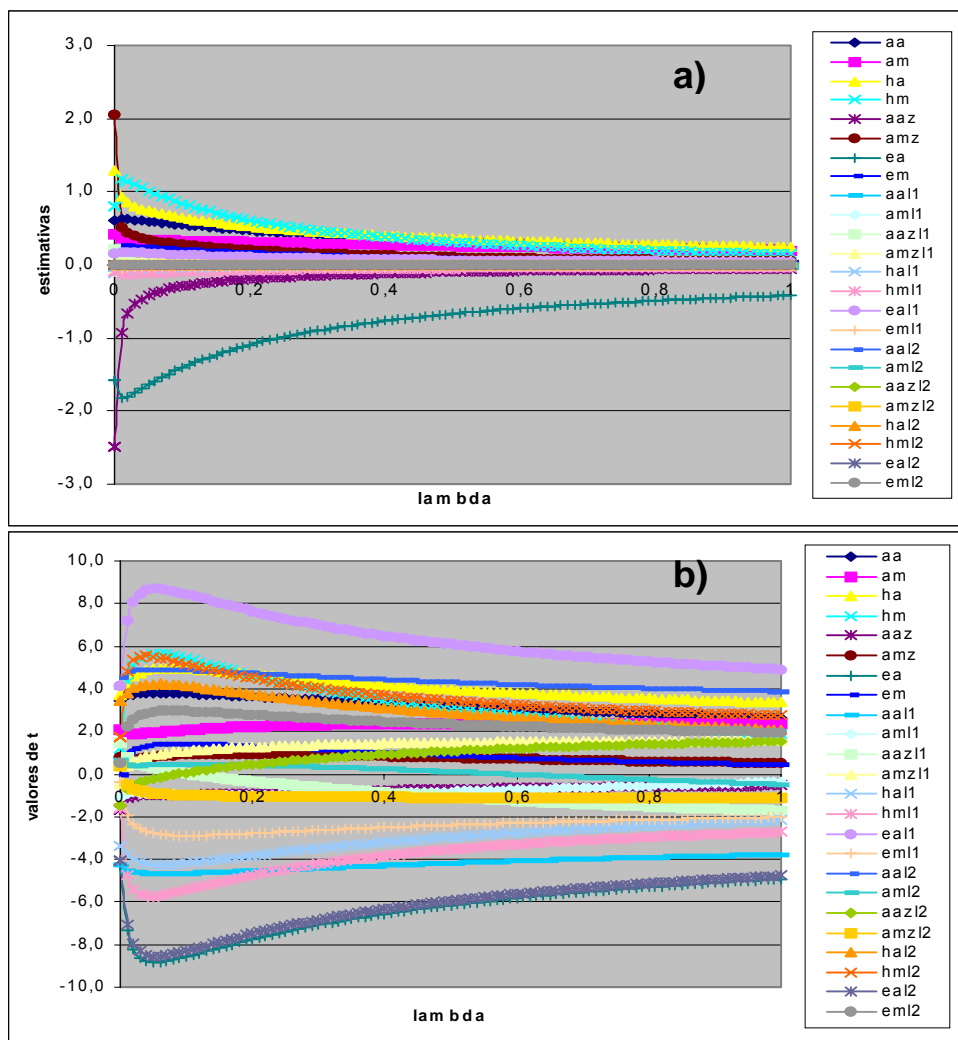


Figura 1. Mudanças nas estimativas dos coeficientes (“traço ridge”) e nos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (1) de implementação da regressão de cumeeira.

Na Figura 2, são apresentados os gráficos dos valores de **soma(abs)t**, **harm(abs)t** e FIVs de acordo com o valor do coeficiente  $\lambda$ , e as superfícies de predição obtidas a partir dos coeficientes estimados por cada critério. O valor máximo (76,7886) para **soma(abs)t** é atingido com  $\lambda = 0,08$  (Figura 2a), e o valor máximo (4,7762) para **harm(abs)t** é atingido com  $\lambda = 0,06$  (Figura 2c).

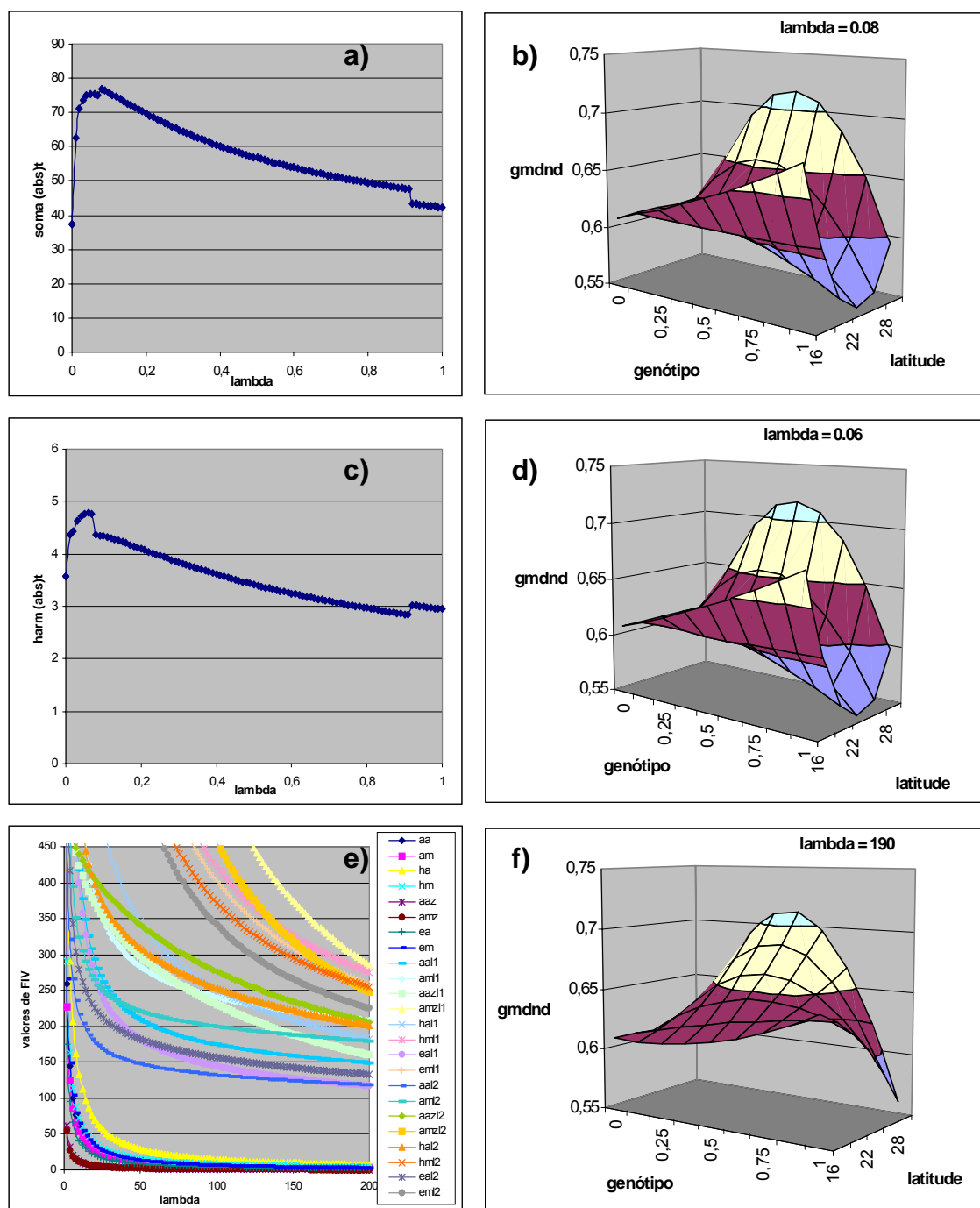


Figura 2. Valores de **soma(abs)t**, **harm(abs)t** e FIVs de acordo com o valor do coeficiente  $\lambda$ , e superfícies de predição obtidas a partir dos coeficientes estimados por cada critério, pelo método (1) de implementação da regressão de cumeeira.

Os valores preditos de **gmdnd**, calculados a partir dos coeficientes estimados com  $\lambda = 0,08$  e  $\lambda = 0,06$ , são apresentados nas Figuras 2b e 2d, respectivamente. Em latitudes intermediárias, observam-se valores preditos de **gmdnd** inferiores aos preditos para latitudes extremas, para todas as composições raciais. Esse comportamento é particularmente mais evidente em animais 100% Nelore. Espera-se que o desempenho de animais zebuínos, melhor adaptados a ambientes tropicais, seja inferior em estados como o Rio Grande do Sul (latitude 31°S) do que, por exemplo, em São Paulo (latitude 23°S).

Pela observação do “traço ridge”, percebe-se que os valores das estimativas são regredidos em direção a zero à medida que se aumentam os valores de  $\lambda$ , verificando-se mudança mais acentuada até valores de  $\lambda$  próximos a 0,8 (Figura 1a). Isso pode sugerir que os pontos onde **soma(abs)t** e **harm(abs)t** apresentaram maior valor, não coincidiram com valores de  $\lambda$  que provocassem alterações suficientes nas estimativas, para fornecer predições com interpretações biológicas mais condizentes com as expectativas. De fato, uma redução mais efetiva nos valores de FIV é observada com a utilização de valores de  $\lambda$  mais elevados. Somente a partir de  $\lambda = 190$ , os valores de FIV passam a ser, todos, menores que 300 (Figura 2e).

As predições feitas a partir dos coeficientes estimados com  $\lambda = 190$  são apresentadas na Figura 2f. Nessa superfície já se pode observar um declínio constante nas predições de **gmdnd** de indivíduos 100% Nelore, à medida que as latitudes vão se tornando mais elevadas.

Verifica-se também que o desempenho predito dos animais puros Nelore é sempre inferior ao de todas as outras composições raciais, exceto nas menores latitudes, onde genes para maior adaptação parecem assumir maior importância na determinação de desempenho, do que genes para alto metabolismo, incorporados ao genótipo dos animais pela proporção Hereford.

As mudanças nas estimativas dos parâmetros e nos valores de  $t$ , obtidas pelo método (2) de implementação, são ilustradas nas Figuras 3a e 3b, respectivamente.

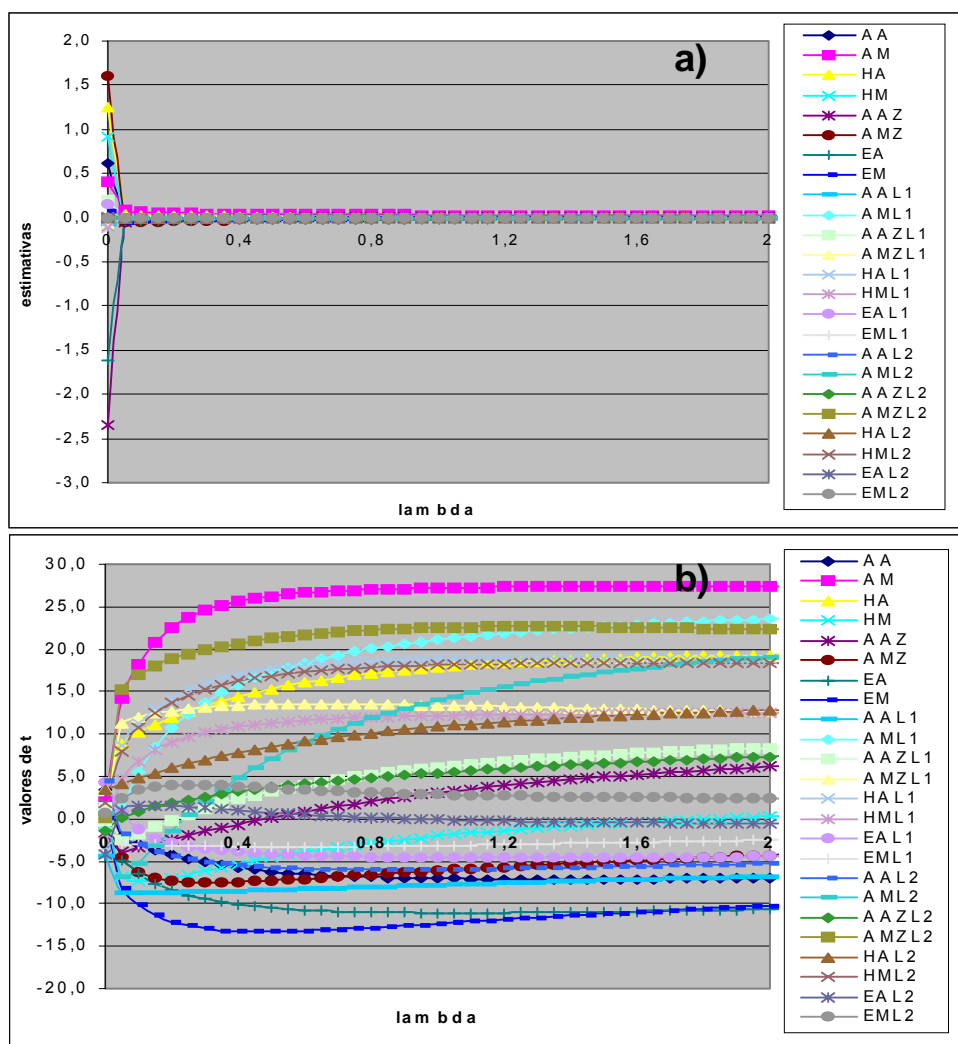


Figura 3. Mudanças nas estimativas dos coeficientes (“traço ridge”) e nos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (2) de implementação da regressão de cumeeira.

A observação das Figura 3a e 3b revela uma situação diferente da encontrada no método (1). As maiores mudanças nas estimativas ocorrem com valores bem baixos do coeficiente  $\lambda$  (até 0,1). Já nos valores de  $t$ , ocorre um aumento, porém de maneira mais lenta, atingindo-se um valor máximo (4,3937) para **harm(abs)t** com  $\lambda = 0,35$  (Figura 4c), e um valor máximo (264,04) para **soma(abs)t** com  $\lambda = 2,65$  (Figura 4a).

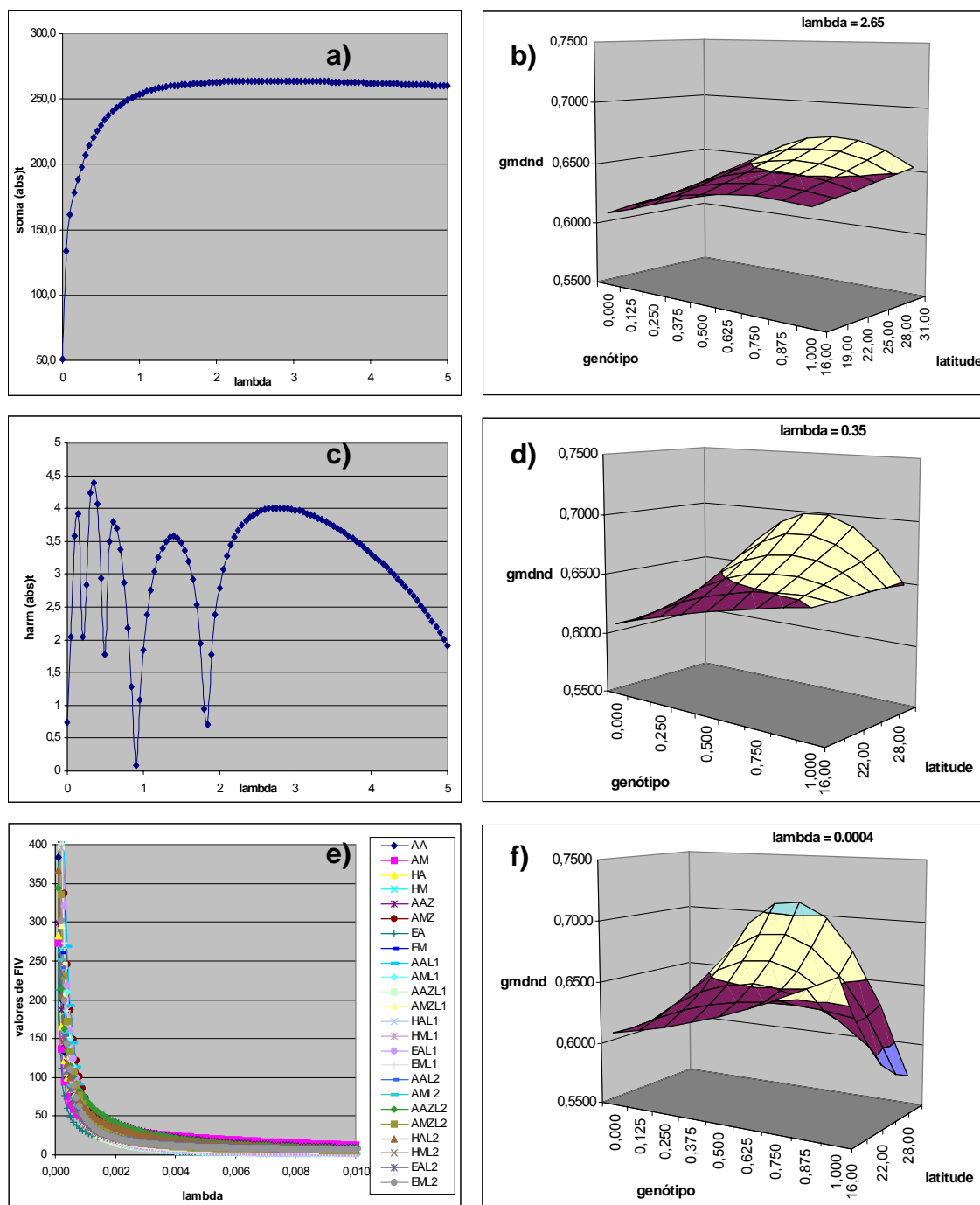


Figura 4. Valores de  $\text{soma}(\text{abs})t$ ,  $\text{harm}(\text{abs})t$  e FIVs de acordo com o valor do coeficiente  $\lambda$ , e superfícies de predição obtidas a partir dos coeficientes estimados por cada critério, pelo método (2) de implementação da regressão de cumeeira.

Uma drástica redução nos valores de FIV também pode ser observada com valores de  $\lambda$  bastante baixos (Figura 4e). Todos os valores de FIV já são menores que 300 com  $\lambda = 0,0004$ .

Esse maior impacto da regressão de cumeeira, implementada pelo SAS, nas estimativas dos parâmetros e nos respectivos valores de FIV, pode ser explicado pelo fato de que, nesse método, adicionam-se valores a todos os elementos da matriz de correlações, e não só à sua diagonal principal. Com isso, uma maior quantidade de informação externa é incorporada à análise, impondo, possivelmente, um maior viés às estimativas. Um outro ponto a ser considerado é a padronização das variáveis. Como o SAS usa a matriz de correlações na implementação da regressão de cumeeira, pequenos coeficientes  $\lambda$  adicionados a essa matriz produzem um efeito maior do que seria conseguido ao se utilizar a matriz de variâncias e covariâncias. Também deve-se ressaltar que, nesse método, o coeficiente  $\lambda$  é aplicado a todas as covariáveis incluídas no modelo, inclusive as ambientais, sem distinção quanto ao envolvimento de cada uma delas em colinearidade.

As superfícies de predição construídas a partir dos coeficientes estimados, tomando-se **soma(abs)t** e **harm(abs)t** como critérios para adoção de  $\lambda$ , são bastante atenuadas (Figuras 4b e 4d), quando comparadas à superfície obtida com  $\lambda = 0,0004$  (Figura 4f). Esse pode ser um indicativo de que, com  $\lambda = 0,35$  e  $\lambda = 2,65$ , há uma redução maior do que a necessária nas estimativas dos parâmetros.

A superfície de predição obtida com  $\lambda = 0,0004$  se assemelha à obtida pelo método (1) com  $\lambda = 190$ , porém uma vantagem mais pronunciada do Nelore, nas menores latitudes, pode ser observada.

No método (3) de implementação da regressão de cumeeira, os valores de  $\lambda$  adicionados a cada elemento da matriz de coeficientes são ponderados pela razão entre o respectivo elemento de  $X'X$  e o menor deles. O objetivo, com isso, foi ajustar os valores a serem adicionados, para a ordem de grandeza dos elementos de  $X'X$ . Essa ponderação dos coeficientes  $\lambda$  procura eliminar a influência das unidades de medida de cada variável explicativa, promovendo uma redução (ou uma penalização) mais

uniforme nas estimativas. As mudanças nas estimativas e nos valores de  $t$ , de acordo com os valores de  $\lambda$  utilizados, são ilustradas na Figura 5.

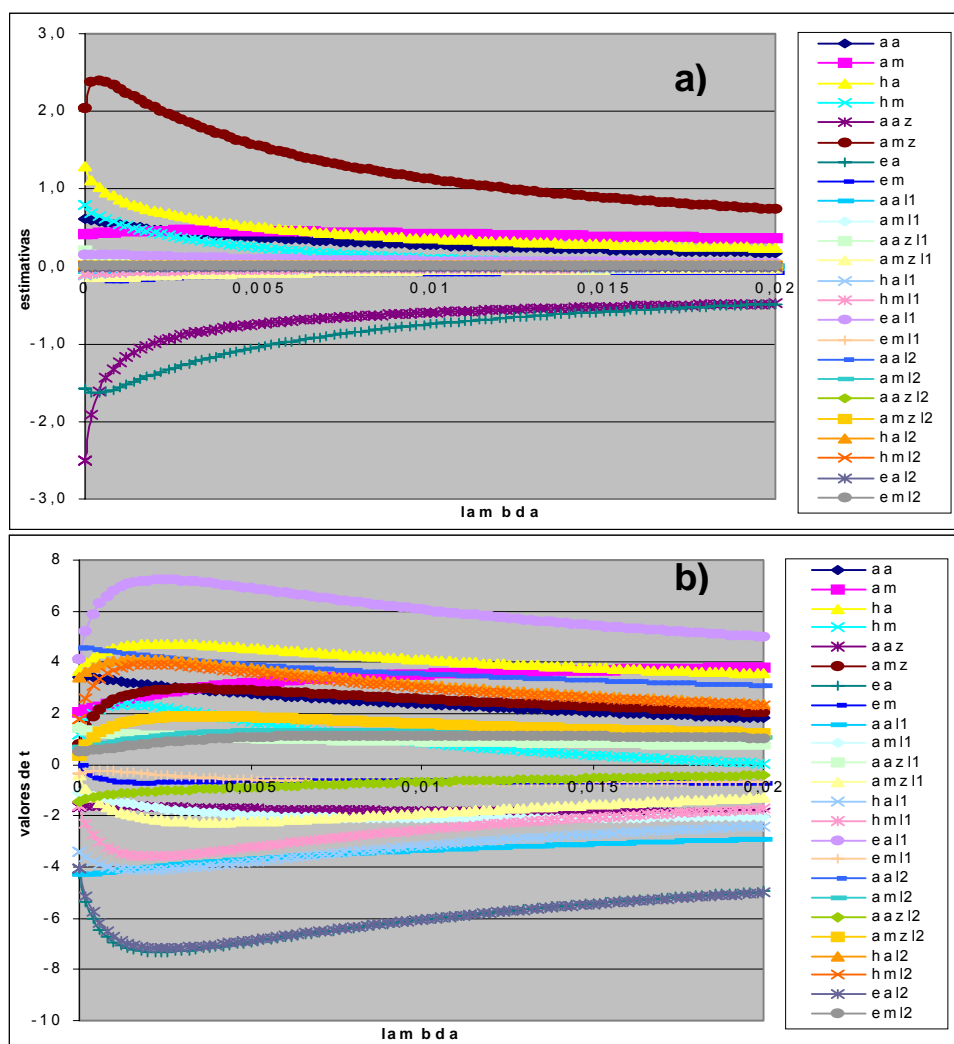


Figura 5. Mudanças nas estimativas dos coeficientes (“traço ridge”) e nos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (3) de implementação da regressão de cumeeira.

Uma situação semelhante à obtida pelo método (1) pode ser observada. Há um aumento nos valores de  $t$  até um certo ponto, a partir do qual eles passam a regredir. O valor máximo (63,5543) para **soma(abs)t** foi alcançado com  $\lambda = 0,0022$ , e o valor máximo (3,8477) para **harm(abs)t** foi alcançado com  $\lambda = 0,0014$  (Figuras 6a e 6c).

As predições feitas a partir dos coeficientes estimados com  $\lambda = 0,0022$  e com  $\lambda = 0,0014$  são apresentadas nas Figuras 6b e 6d, respectivamente. As superfícies obtidas são bem próximas às obtidas pelo método (1), usando **soma(abs)t** e **harm(abs)t** como critérios para adoção de  $\lambda$ . Uma explicação para essa similaridade pode advir do fato de que as covariáveis genéticas são todas expressas em termos de porcentagem, excetuando-se os efeitos de complementariedade direta e materna, cujo valor máximo é 0,25. Por isso, para esse conjunto de dados, o ajuste para as unidades de medida das variáveis explicativas não surtiu grande efeito, ou, pelo menos, o efeito que poderia se esperar num conjunto de dados com variáveis explicativas expressas em unidades mais discrepantes.

Tomando-se como critério o ponto a partir do qual todos os valores de FIV são menores que 300, chega-se a  $\lambda = 0,26$  (Figura 6e). A superfície de predição construída a partir dos coeficientes estimados, usando-se esse valor de  $\lambda$ , apresentada na Figura 6f, assemelha-se às obtidas pelos métodos (1) e (2), quando esse mesmo critério foi empregado para a determinação do valor de  $\lambda$ .



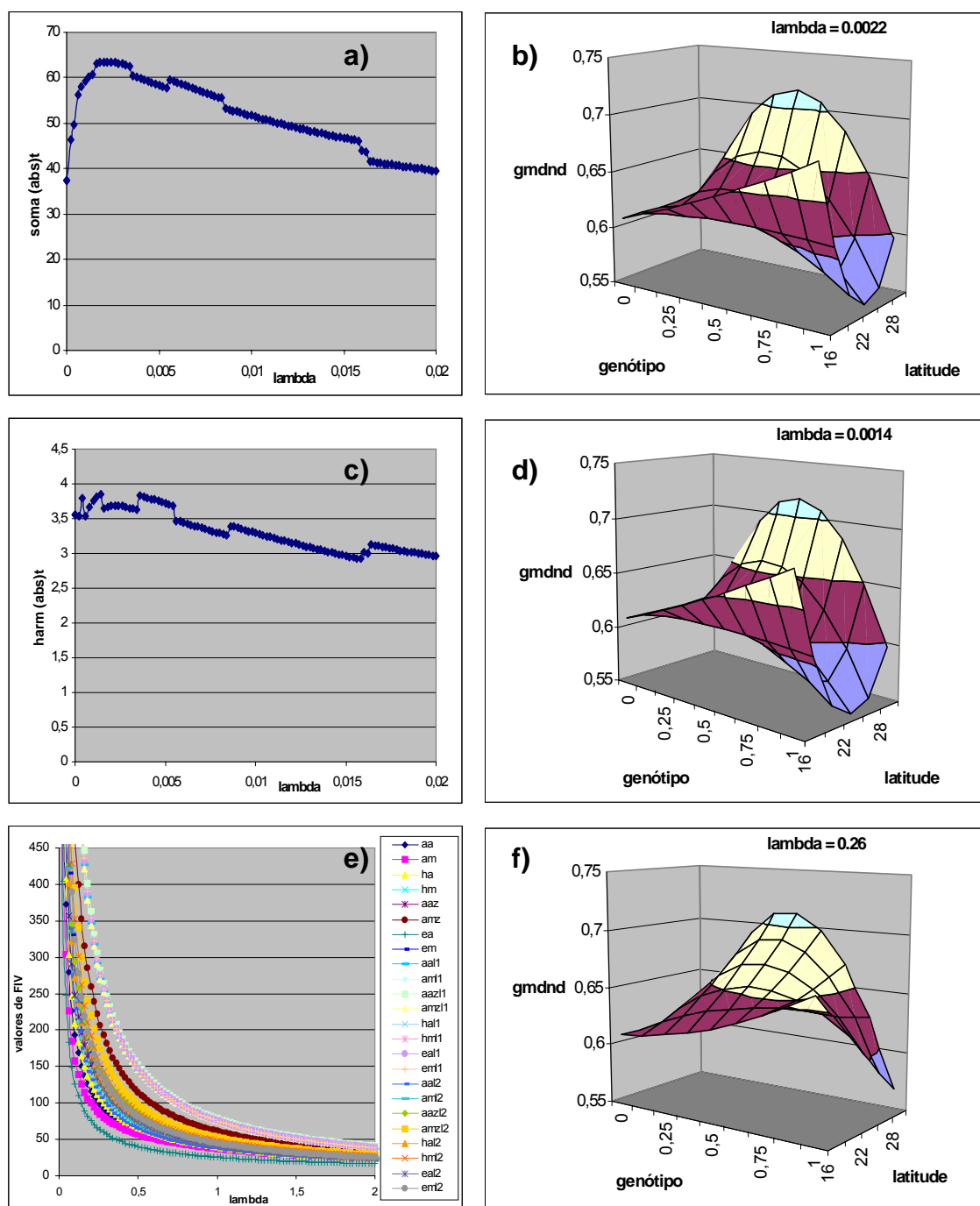


Figura 6. Valores de **soma(abs)t**, **harm(abs)t** e FIVs de acordo com o valor do coeficiente  $\lambda$ , e superfícies de predição obtidas a partir dos coeficientes estimados por cada critério, pelo método (3) de implementação da regressão de cumeeira.

No método (4) de implementação, os valores de  $\lambda$  são ponderados pelo valor de FIV da correspondente covariável. Esse procedimento visa adicionar coeficientes aos elementos de  $X'X$  de acordo com o grau de envolvimento de cada variável no problema de multicolinearidade. As mudanças nas estimativas e nos valores de  $t$  são apresentadas na Figura 7.

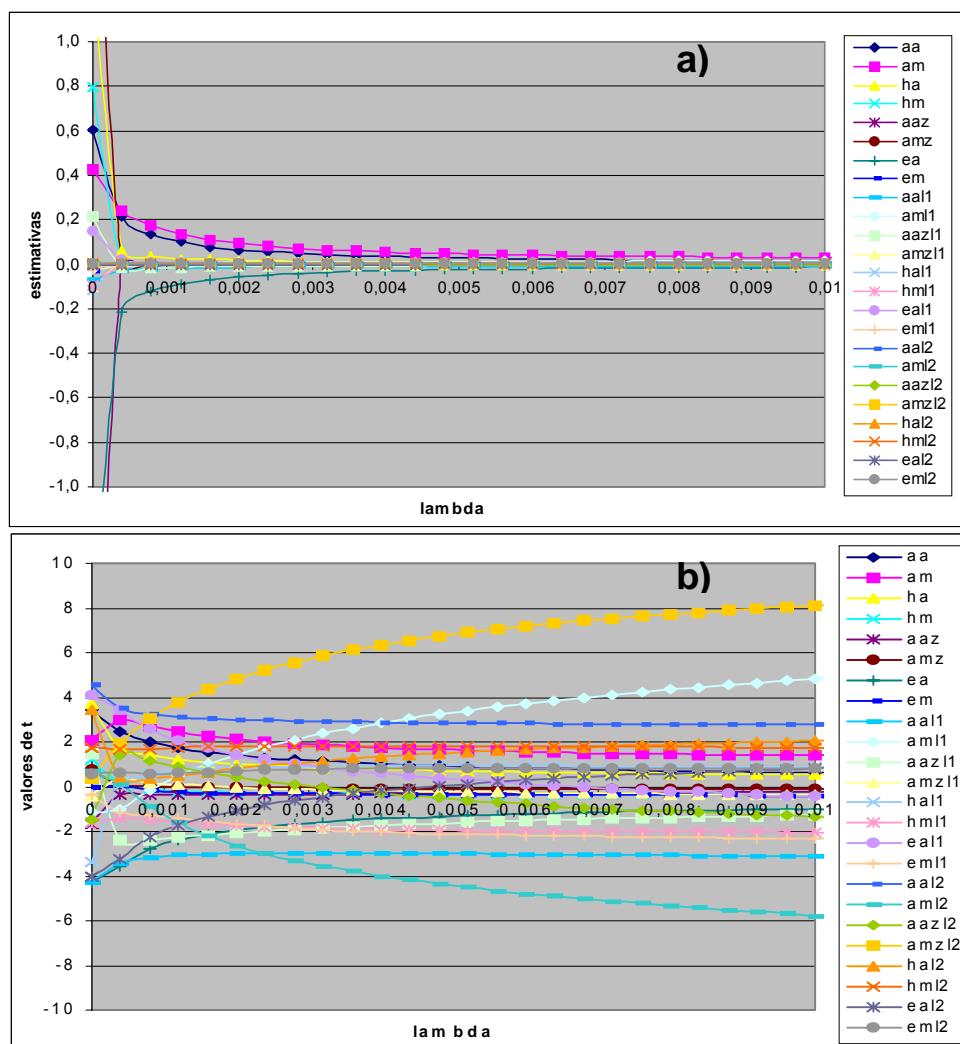


Figura 7. Mudanças nas estimativas dos coeficientes (“traço ridge”) e nos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (4) de implementação da regressão de cumeieira.

Percebe-se que as maiores mudanças nas estimativas dos parâmetros se verifica com valores de  $\lambda$  até 0,01 (Figura 7a). O valor máximo (63,6018) de **soma(abs)t** é atingido com  $\lambda = 0,29$ , e o valor máximo (0,6786) de **harm(abs)t** é atingido com  $\lambda = 0,016$  (Figuras 8a e 8c). Todos os valores de FIV passaram a ser menores que 300, a partir de  $\lambda = 0,01$ .

As previsões de **gmdnd** feitas a partir das estimativas obtidas segundo os três critérios são ilustradas nas Figuras 8b, 8d e 8f. Pode-se perceber uma similaridade de tendência entre as três superfícies, havendo uma variação apenas no grau de inclinação, notadamente com  $\lambda = 0,29$ . Com esse método de implementação da regressão de cumeeira, parece ter havido uma concordância entre os critérios para adoção de  $\lambda$ , no que diz respeito à capacidade de previsão, principalmente entre os critérios baseados nos valores de FIV e na média harmônica dos valores absolutos de t.

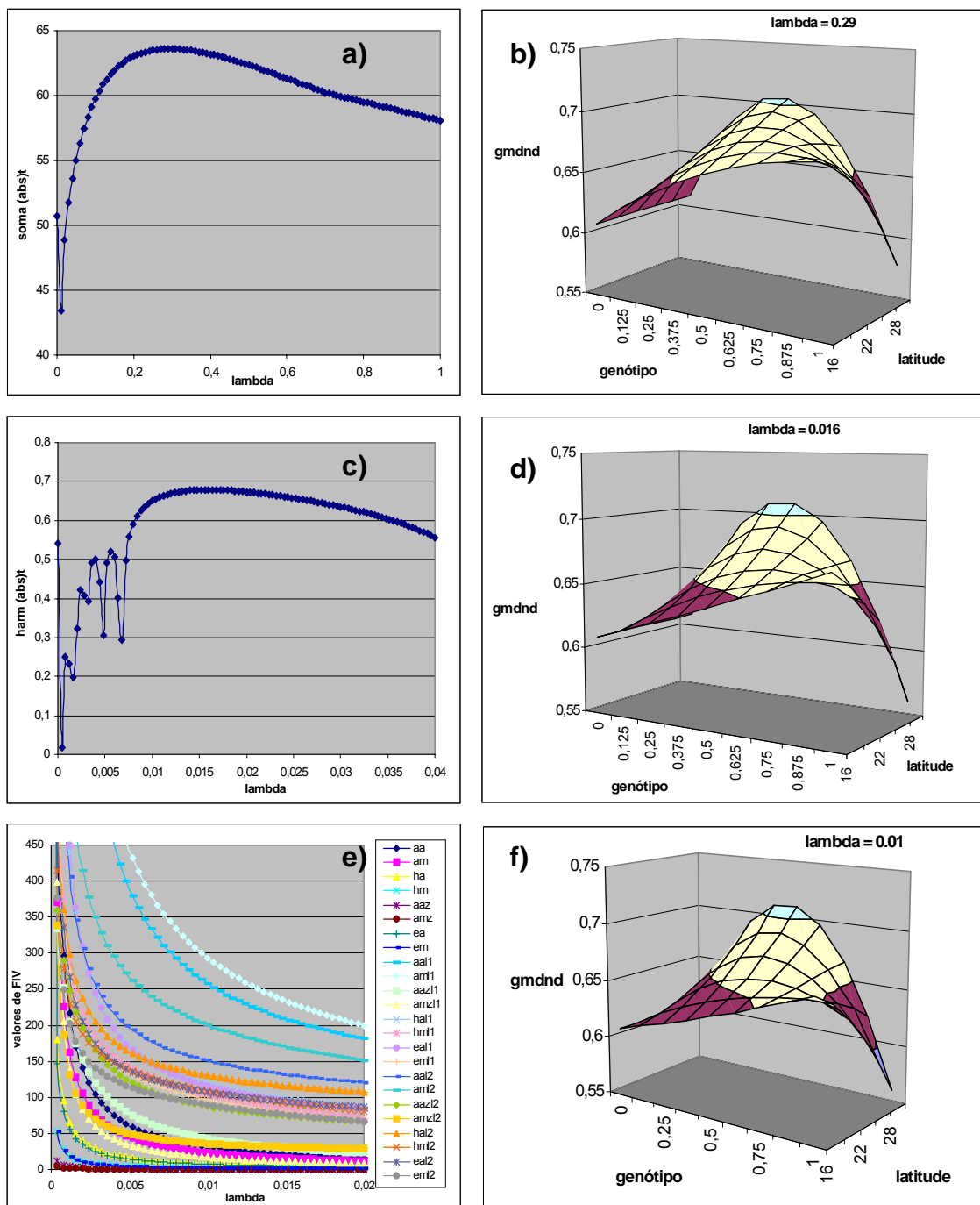


Figura 8. Valores de **soma(abs)t**, **harm(abs)t** e FIVs de acordo com o valor do coeficiente  $\lambda$ , e superfícies de predição obtidas a partir dos coeficientes estimados por cada critério, pelo método (4) de implementação da regressão de cumeeira.

Nas Figuras 9a e 9b são apresentadas as modificações nos valores das estimativas dos parâmetros e das respectivos valores de  $t$ , de acordo com os valores de  $\lambda$  empregados pelo método (5). Observa-se que as maiores mudanças ocorrem até valores de  $\lambda$  em torno de 0,0006, tendendo a haver uma estabilização nos valores das estimativas e de  $t$ , a partir de então.

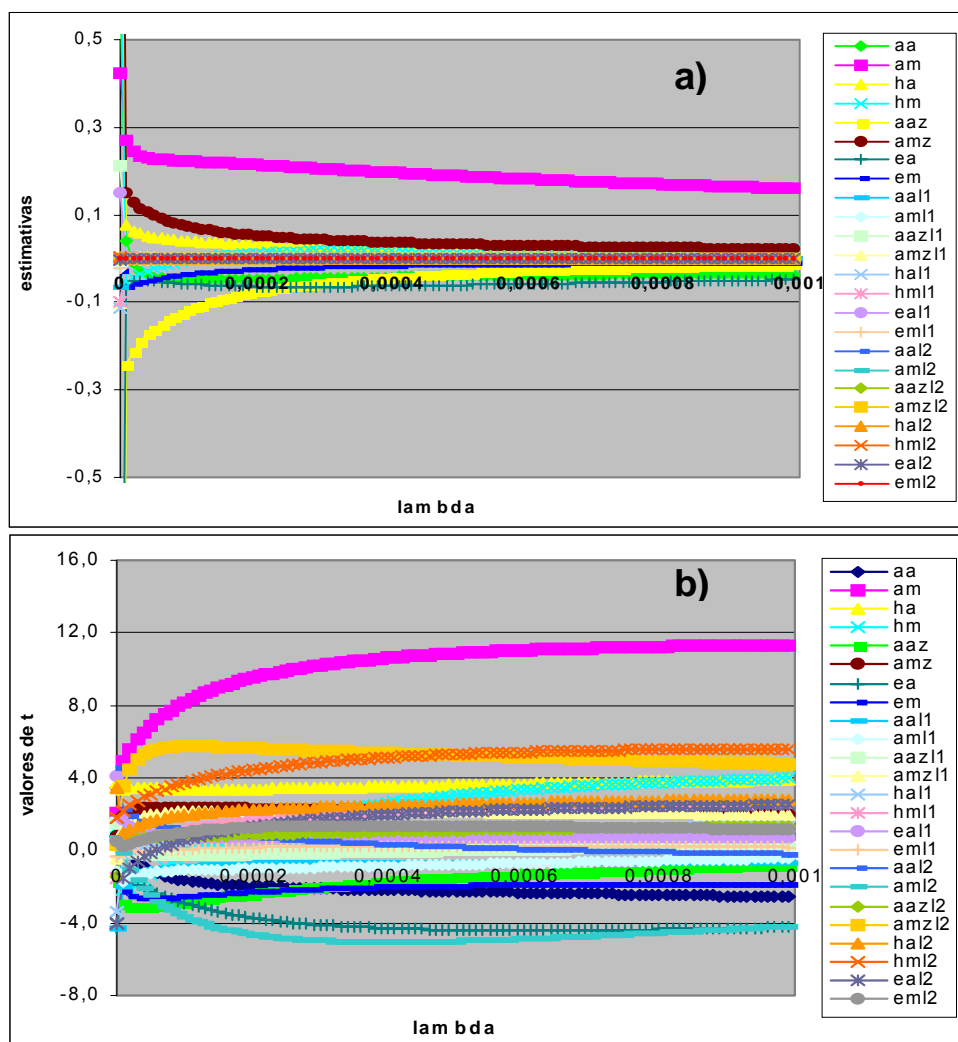


Figura 9. Mudanças nas estimativas dos coeficientes (“traço ridge”) e nos valores de  $t$ , de acordo com os valores do coeficiente  $\lambda$  adotado, pelo método (5) de implementação da regressão de cumeeira.

Na Figura 10 são apresentadas as superfícies de predição construídas a partir dos coeficientes estimados para cada  $\lambda$ . O valor máximo (63,2383) de **soma(abs)t** foi atingido com  $\lambda = 0,0275$ , e o valor máximo (0,9306) de **harm(abs)t** foi atingido com  $\lambda = 0,00022$ . Os valores de FIV passaram a ser menores que 300 a partir de  $\lambda = 0,00004$ . Como, nesse método,  $\lambda$  é ponderado pelo produto de FIV pela razão de soma de quadrados, uma grande redução nos FIVs já pôde ser obtida com valores de  $\lambda$  bem baixos.

A superfície obtida com  $\lambda = 0,00004$  se assemelha às obtidas pelo método (3) com  $\lambda = 0,26$  e pelo método (4) com  $\lambda = 0,01$ . Ou seja, superfícies bem próximas puderam ser conseguidas pelos métodos (3), (4) e (5) usando-se o critério baseado nos FIVs. A diferença nos valores de  $\lambda$  utilizados pode, então, ser decorrente da ordem de grandeza do ponderador empregado em cada método.

Fazendo-se uma comparação entre os cinco métodos de implementação estudados neste trabalho, verifica-se que, com o presente conjunto de dados e modelo considerado, o método (4) mostrou maior robustez quanto ao critério empregado na determinação de  $\lambda$ . Quanto ao critério para adoção do coeficiente  $\lambda$ , o ponto, a partir do qual os valores de FIV são menores que 300, produziu, em todos os métodos de implementação, estimativas razoáveis, se avaliadas via capacidade de predição.

Na literatura, é comum a consideração de que, quando uma análise de regressão é feita com finalidade de predição, pouca diferença se pode esperar ao usar o método da regressão de cumeira ou dos quadrados mínimos ordinários, em caso de presença de multicolinearidade. A recomendação do uso da regressão de cumeira geralmente é feita para casos em que o interesse, na análise de regressão, é a interpretação dos coeficientes de regressão individuais (BERGMANN & HOHENBOKEN, 1995; FREUND & WILSON, 1998). Os resultados deste trabalho indicam que, pelo menos para este conjunto de dados e este modelo, a regressão de cumeira pôde fornecer superfícies de predição bem diferentes, cujas interpretações biológicas são bem mais aceitáveis, em relação ao método dos quadrados mínimos ordinários.

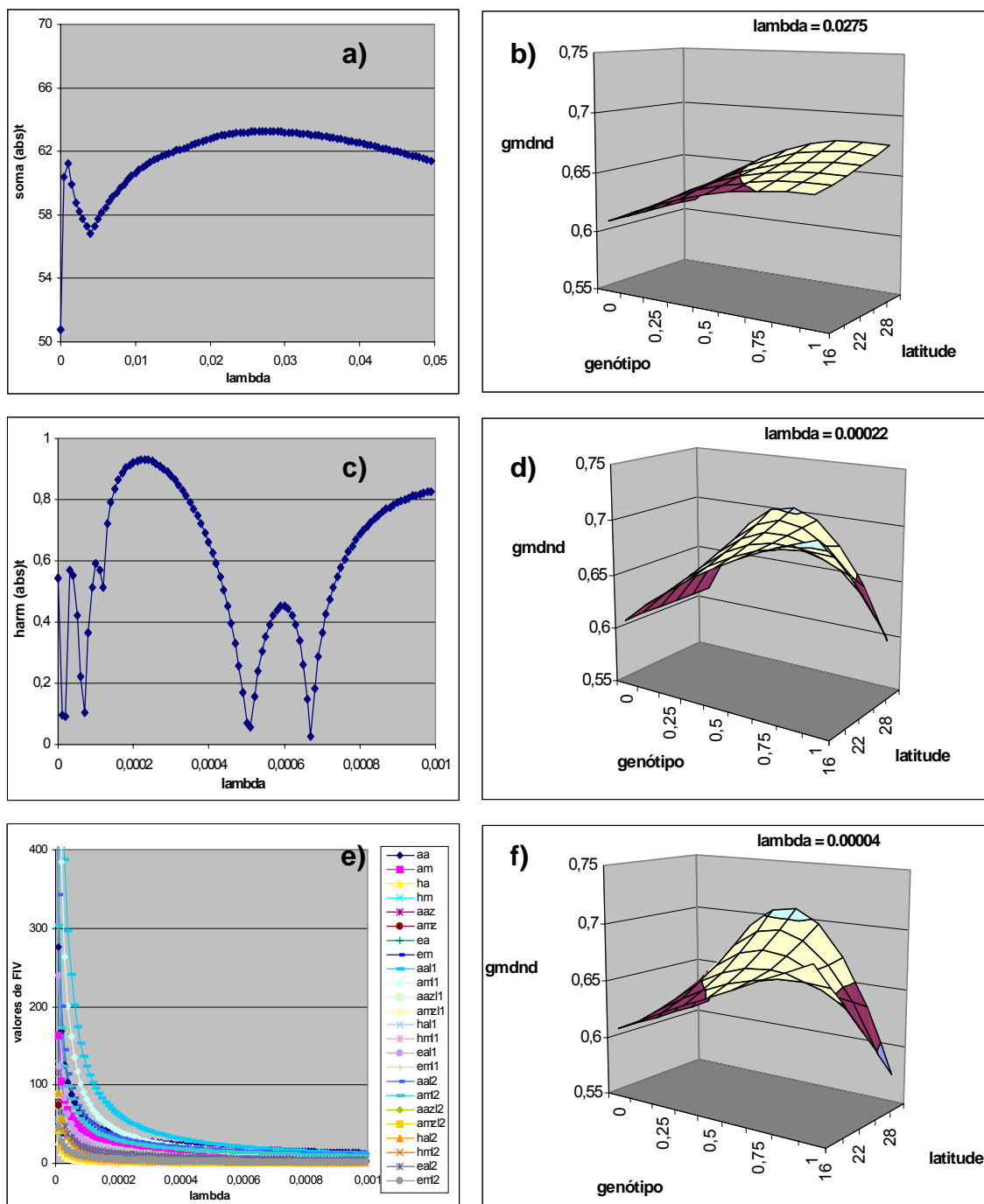


Figura 10. Valores de **soma(abs)t**, **harm(abs)t** e FIVs de acordo com o valor do coeficiente  $\lambda$ , e superfícies de predição obtidas a partir dos coeficientes estimados por cada critério, pelo método (5) de implementação da regressão de cumeeira.

## Conclusões

Para o modelo e o conjunto de dados estudados, o método (4) de implementação da regressão de cumeeira se mostrou mais robusto, quanto ao critério para escolha de  $\lambda$ . Superfícies de predição biologicamente coerentes puderam ser obtidas em todos os métodos de implementação, usando-se o critério baseado no valor de FIV para escolha de  $\lambda$ . Um critério matemático deve ser usado como uma ferramenta auxiliar para escolha de  $\lambda$ , porém não excluindo a necessidade da observação do “traço ridge”, e principalmente um bom conhecimento do fenômeno em estudo.

O método da regressão de cumeeira pode ser uma alternativa eficaz para contornar o problema da multicolinearidade, não só quando o objetivo da análise é a interpretação dos sinais e valores dos coeficientes, mas também quando a análise é feita com finalidade de predição.

## Referências

- BERGMANN, J.A.G.; HOHENBOKEN, W.D. Alternatives to least squares in multiple linear regression to predict production traits. **Journal of Animal Breeding and Genetics**, v.112, p.1-16. 1995.
- BÉRTOLI, C.D. **Sistema Cruza - Controle de Produção e Avaliação dos Valores Genéticos dentro de uma População Bovina Sintética**. Dissertação (Mestrado em Agronomia) Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 1991. 233 p.
- BRITO, F.V.; PICCOLI, M.L.; SEVERO, J.L.P.; SCHENKEL, F.S.; ROSO, V.M.; FRIES, L.A. Estimating environmental and genotypic effects on preweaning weight gain of Angus x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** p.773-776.



- CHATTERJEE, S.; PRICE, B. **Regression analysis by example**. 2nd ed. New York: Wiley series in probability and mathematical statistics, 1991. 278p.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: Wiley series in probability and statistics, 1998. 706p.
- FOUCART, T. Stability of the inverse correlation matrix. Partial ridge regression. **Journal of Statistical Planning and Inference**, v.77, p.141-154. 1999.
- FREUND, R.J.; LITTELL, R.C. **SAS System for Regression**, 3rd ed. Cary: SAS Institute Inc., 2000. 235p.
- FREUND, R.J.; WILSON, W.J. **Regression Analysis**: statistical modeling of a response variable. San Diego: Academic Press, 1998. 444p.
- FRIES, L.A.; JOHNSTON, D.J.; HEARNSHAW, H.; GRASER, H.U. Evidence of Epistatic Effects on Weaning Weight in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, suppl. B, p. 242, 2000a.
- FRIES, L.A.; GRASER, H.U.; JOHNSTON, D.J.; HEARNSHAW, H. Using Ridge Regression to Estimate Genetic Effects in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, suppl. B, p. 241, 2000b.
- HAIR JR., J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Multivariate Data Analysis**, 3rd ed. New York: Macmillan Publishing Company, 1992. 544p.
- HOCKING, R.R. **Methods and Applications of Linear Models**: regression and the analysis of variance, 2nd ed. New Jersey: Wiley series in probability and statistics, 2003. 741p.
- HOERL, A.E.; KENNARD, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v.12, n.1, p.55-67. 1970.
- KNIGHT, J.R.; HILL, R.C.; SIRMANS, C.F. Biased Prediction of Housing Values. **Journal of the American Real Estate and Urban Economics Association**, v. 20, n.3, p.427-456. 1992.
- LECLERC, G.; PIREAUX, J.J. The use of least squares for XPS peak parameters estimation. Part 3. Multicollinearity, ill-conditioning and constraint-induced bias. **Journal of Electron Spectroscopy and Related Phenomena**, v. 71, p.179-190. 1995.

- LIPOVETSKY, S.; CONKLIN, W.M. Multiobjective regression modifications for collinearity. **Computers & Operations Research**, v.28, p.1333-1345. 2001.
- MAINDONALD, J.H. **Statistical computation**. New York: Wiley series in probability and mathematical statistics, 1984. 370p.
- PICCOLI, M.L.; ROSO, V.M.; BRITO, F.V.; SEVERO, J.L.P.; SCHENKEL, F.S.; FRIES, L.A. Additive, complementarity (additive\*additive), dominance, and epistatic effects on preweaning weight gain of Hereford x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** CD-ROM.
- SCHENKEL, F.S. **Cálculo das heterozigoses**. Porto Alegre, RS: Gensys Consultores Associados. 1993.
- VAN GAANS, P.F.M.; VRIEND, S.P. Multiple linear regression with correlations among the predictor variables. Theory and computer algorithm RIDGE (Fortran 77). **Computers & Geosciences**, v.16, n.7, p.933-952. 1990.

## **CAPÍTULO 4 – EFEITOS DA INCLUSÃO DE PARÂMETROS PARA EPISTASIA E COMPLEMENTARIEDADE EM MODELOS DE AVALIAÇÃO DE EFEITOS GENÉTICOS EM BOVINOS CRUZADOS**

**RESUMO** – Modelos utilizados para avaliação de efeitos genéticos envolvidos em características de crescimento em animais cruzados foram estudados. Compararam-se dois modelos: um contendo parâmetros para efeitos diretos e maternos de ação aditiva e dominância; e outro incluindo, além desses, efeitos diretos e maternos de epistasia e complementariedade. Para contornar os efeitos da multicolinearidade presente na estrutura dos dados, a técnica da regressão de cumeeira foi empregada na estimação dos parâmetros. Quatro métodos de implementação da regressão de cumeeira foram testados: o método proposto originalmente; o implementado pelo SAS; e mais duas formas de ponderação do coeficiente  $\lambda$ . As comparações entre os dois modelos, para cada um dos quatro métodos de implementação, foram feitas por meio de: superfícies de predição obtidas usando-se as soluções estimadas para cada modelo; valores de  $R^2$  apresentados por cada modelo; e um teste F para verificar se a adição de mais parâmetros representa um ganho efetivo, em termos de ajuste do modelo. Concluiu-se que a inclusão de parâmetros para efeitos de epistasia e complementariedade em modelos de avaliação de efeitos genéticos em animais cruzados pode representar um ganho tanto em termos de ajuste do modelo quanto de capacidade de predição de desempenho de genótipos não testados. O método da regressão de cumeeira se mostrou uma ferramenta que possibilita a obtenção de estimativas de efeitos de ação aditiva, dominância, epistasia e complementariedade, mesmo na presença de forte multicolinearidade.

**Palavras-chave:** bovino de corte, cruzamento, estimação viesada, heterose, multicolinearidade, regressão de cumeeira

## Introdução

Sistemas de cruzamentos são fundamentados na exploração de efeitos raciais e de heterose. O uso eficiente dos efeitos raciais se baseia no conhecimento das características raciais diretas e maternas. Adicionalmente, o uso efetivo da heterose requer a caracterização dos efeitos de heterose nas vacas e bezerros cruzados (ROBERSON *et al.*, 1986).

A heterose, e seus efeitos geralmente benéficos ao desempenho, tem sido atribuída aos efeitos de dominância de genes de ação melhoradora. Em muitas investigações, essa hipótese tem sido satisfatória na explicação da superioridade observada em indivíduos cruzados, como reportado por GREGORY *et al.* (1991). Contudo, segundo SHARMA *et al.* (2000), em alguns experimentos e investigações em que indivíduos  $F_2$  ou sintéticos são estudados, a hipótese da dominância tem se provado insuficiente para ajustar os resultados. Com isso, vários pesquisadores têm estendido os modelos estatísticos de maneira a incluir epistasia (HIROOKA *et al.*, 1998; ARTHUR *et al.*, 1999; KAHN *et al.*, 2000; MOHAMED *et al.*, 2001; CASSADY *et al.*, 2002; DEMEKE *et al.*, 2003).

Segundo MACKAY (2001), qualquer processo bioquímico que culmine na expressão de uma característica quantitativa, é determinada por redes de loci que interagem em nível genético e molecular. De acordo com WRIGHT (1978), ao longo do processo de formação das raças, a seleção natural deve, de algum modo, operar sobre sistemas interativos de genes, como blocos, produzindo combinações que precisam ser mais eficientes.

A quebra de associações epistáticas positivas entre genes de origem parental parece ser a principal causa do declínio no desempenho de crescimento em gerações avançadas de cruzamentos *Bos taurus* x *Bos indicus*. Logo, a epistasia parece ser um efeito genético importante, que não deve ser ignorado em avaliações de desempenho de cruzamentos, particularmente quando a finalidade do exercício é a predição de desempenho de genótipos não testados, para decisões sobre futuros programas de cruzamentos (DEMEKE *et al.*, 2003).

Segundo BURKE & ARNOLD (2001), as evidências disponíveis sugerem que a inferioridade dos híbridos resulta de um amplo efeito epistático negativo no histórico genético desses indivíduos. Se a seleção natural retém combinações de interações gênicas favoráveis, então a hibridização produz um vasto arranjo de genótipos recombinantes nunca antes sujeitos a seleção. Na média, esses genótipos serão menos adaptados que seus ascendentes, dando margem a algum nível de seleção contra os híbridos. Em experimento com peso à desmama de bezerros cruzados, FRIES *et al.* (2000) examinaram a hipótese de que uma maior heterozigose em uma geração é concomitante com maiores efeitos epistáticos na geração seguinte; e sugerem a modelagem de efeitos epistáticos por epistazigose, definida como a heterozigose média observada nos pais de um indivíduo.

Quando se consideram características de crescimento de animais cruzados em ambiente tropical, pode-se imaginar que dois conjuntos de genes estejam atuando: um relacionado a potencial de crescimento e outro a adaptação. Esses dois conjuntos podem ser tomados como características que se complementam e cujo produto (interação) pode fornecer uma base teórica para a heterose que não depende de dominância ou epistasia. Essa relação multiplicativa entre as ações aditivas para as duas características foi definida por KINGHORN (1993) como “profit heterosis”, e pode ser interpretada como um efeito de complementariedade entre as características que determinam crescimento.

Segundo BRITO *et al.* (2002), numa tomada de decisões acerca do delineamento de programas de cruzamento para regiões e mercados específicos, devem-se considerar: as contribuições aditivas direta e materna de cada raça; interações entre as contribuições aditivas de cada raça (complementariedade); efeitos diretos e maternos de dominância; efeitos epistáticos direto e materno; e interações entre componentes genéticos e variáveis ambientais.

O objetivo do presente trabalho foi comparar dois modelos de avaliação de efeitos genéticos no desempenho de bezerros Hereford x Nelore: um contendo efeitos diretos e maternos de ação aditiva e dominância; e outro incluindo, além desses quatro, efeitos diretos e maternos de epistasia e complementariedade.

## Material e métodos

Foram utilizados dados do programa de melhoramento de gado de corte Conexão Delta G, preparados pela empresa GenSys Consultores Associados S/C Ltda. O banco de dados da Conexão Delta G contém registros de 109.614 bezerros Hereford x Nelore nascidos entre 1974 e 1999, distribuídos em 4.665 grupos de contemporâneos, e criados em 29 fazendas localizadas nos estados de Mato Grosso, Goiás, Mato Grosso do Sul, São Paulo, Paraná e Rio Grande do Sul, entre as latitudes 9°S e 31°S.

Dois modelos foram empregados para obtenção de estimativas dos parâmetros de efeitos genéticos no ganho de peso médio diário do nascimento à desmama (**gmdnd**): o modelo **AD** contendo parâmetros para efeitos de ação aditiva direta (**aa**) e materna (**am**), e heterozigose direta (**ha**) e materna (**hm**); e o modelo **ADEC** incluindo, além desses quatro efeitos, parâmetros para efeitos de epistasia direta (**ea**) e materna (**em**), e complementariedade direta (**aaz**) e materna (**amz**).

Covariáveis para **aa** e **am** foram definidas pela contribuição dos genes Nelore na composição genética de cada indivíduo. As fórmulas usadas para calcular e decompor heterozigose direta e materna foram as mesmas usadas por BÉRTOLI (1991) e SCHENKEL (1993). Para o cálculo, foram necessárias as composições genéticas dos pais e avôs maternos dos indivíduos. Quando não se dispunha de informação sobre a composição genética dos ancestrais maternos, foi assumido que o animal foi produzido por acasalamento inter se.

Covariáveis para **aaz** e **amz** entre Nelore e Hereford foram calculadas como sendo **aaz=aa\*(1-aa)** e **amz=am\*(1-am)**, respectivamente (BRITO *et al.*, 2002; PICCOLI *et al.*, 2002). Os efeitos de **ea** e **em** foram modelados por epistazigose como em FRIES *et al.* (2000) e calculados como a heterozigose média presente nos gametas que geraram cada indivíduo ou como a heterozigose média nos pais de um indivíduo.

Os efeitos ambientais incluídos em ambos os modelos foram: grupo de contemporâneos, efeito linear e quadrático de idade da vaca de acordo com o sexo do bezerro, efeito linear e quadrático de idade média à desmama, e polinômio segmentado

quadrático-quadrático de data juliana de nascimento. Em virtude do número de classes do efeito de grupo de contemporâneos, e do longo tempo de processamento que a sua inclusão nos modelos impunha às análises, optou-se por pré-absorver esse efeito.

A inclusão de efeitos principais de latitude (linear e quadrático) nos modelos, resultou em estimativas nulas desses coeficientes. Como todos os indivíduos do mesmo grupo de contemporâneos estão na mesma latitude, o efeito de latitude foi removido pela absorção do efeito de grupo de contemporâneos. Optou-se, então, por incluir o efeito de latitude como um agente modificador, na forma de interações dos efeitos genéticos com efeitos lineares (**L1**) e quadráticos (**L2**) de latitude. Segundo FACÓ *et al.* (2002), deve-se sempre investigar a ocorrência de interação genótipo-ambiente, sob risco de se incorrer em graves erros na condução de programas de cruzamento.

Devido à presença de forte multicolinearidade, constatada pelos altos fatores de inflação de variância (FIVs) observados nas análises por quadrados mínimos ordinários, as estimativas dos coeficientes de regressão foram obtidas pelo método da regressão de cuneeira (HOERL & KENNARD, 1970). Esse método consiste na adição de um escalar ( $\lambda$ ) aos elementos da diagonal principal da matriz de coeficientes, visando quebrar as dependências lineares existentes entre suas colunas. Estimativas dos coeficientes de regressão foram obtidas, para os dois modelos, empregando-se quatro métodos distintos de implementação da regressão de cuneeira:

$$(1) \hat{\beta}_{RC} = (X'X + \lambda I)^{-1} X'Y$$

Implementação original do método da regressão de cuneeira, proposto por HOERL & KENNARD (1970);

$$(2) \hat{\beta}_{SAS} = [R_{XX}(I + D_{\lambda})]^{-1} R_{XY}$$

Procedimento implementado pelo SAS, conforme descrito por FREUND & LITTELL (2000), em que  $R_{XX}$  é a matriz de correlações das variáveis explanatórias,  $R_{XY}$  é o vetor de correlações das variáveis explanatórias com a variável resposta, e  $D_{\lambda}$  é uma matriz diagonal com valores de  $\lambda$  na diagonal.

$$(3) \hat{\beta}_{RCW} = (X'X + \lambda W)^{-1} X'Y$$

Em que  $W$  é uma matriz diagonal cujos elementos da diagonal são iguais aos de  $X'X$  divididos pelo menor elemento da diagonal de  $X'X$ . Nesse método, os valores de  $\lambda$  foram ponderados pelas razões da soma de quadrados da covariável pela menor soma de quadrados na matriz de coeficientes, visando ajustar o valor adicionado em função da magnitude do elemento de  $X'X$ .

$$(4) \hat{\beta}_{FIV} = (X'X + \lambda F)^{-1} X'Y$$

Em que  $F$  é uma matriz diagonal, tal que:

$$F = \{FIV_i\} \text{ se } FIV_i \geq 300;$$

$$F = \{0\} \text{ caso contrário.}$$

Nesse método, os valores de  $\lambda$  foram ponderados pelos valores de FIV das correspondentes covariáveis. Foram usados para essa ponderação os valores de FIV encontrados com  $\lambda = 0$ .

Nas implementações (1), (3) e (4) o coeficiente  $\lambda$  foi aplicado somente às covariáveis genéticas com valor de FIV superior a 300 (LECLERC & PIREAUX, 1995). Essa opção pode evitar a imposição de um viés desnecessário às estimativas de parâmetros não envolvidos em colinearidade.

Testaram-se três critérios matemáticos para escolha do coeficiente  $\lambda$ : a soma e a média harmônica dos valores absolutos da estatística  $t$  de Student (que testa a hipótese de nulidade dos coeficientes estimados), e o valor de  $\lambda$  a partir do qual todos os valores de FIV se tornam menores que trezentos.

A presença, nos modelos considerados, de parâmetros para os efeitos de interação entre os efeitos genéticos e os de latitude, impossibilitou uma interpretação mais clara dos sinais e valores das estimativas dos coeficientes individuais. As comparações entre os dois modelos, para cada um dos quatro tipos de implementação, foram, então, feitas tomando-se como referencial: as superfícies de predições obtidas usando-se as soluções estimadas para cada modelo; os valores dos coeficientes de



determinação apresentados por cada modelo; e um teste F para verificar se a inclusão de covariáveis para epistasia e complementariedade são realmente importantes, após o ajuste do modelo para os efeitos de ação aditiva e dominância. Uma comparação entre os dois modelos em termos de interpretação dos sinais e valores das estimativas dos parâmetros pode ser encontrada em PIMENTEL *et al.* (2003), onde as interações dos efeitos genéticos com latitude não foram incluídas nos modelos.

Predições de **gmdnd** foram feitas para nove genótipos de gerações avançadas, do puro Hereford ao puro Nelore, de oitavo em oitavo, e em seis latitudes diferentes: 16, 19, 22, 25, 28 e 31°S. Um valor correspondente aos efeitos ambientais médios (0,61kg/dia) foi fixado para todas as predições. É importante salientar que as covariáveis para os efeitos genéticos aditivos foram definidas como a proporção de genes Nelore no genótipo de cada indivíduo. Logo, todos os coeficientes para as covariáveis genéticas para indivíduos 100% Hereford são iguais a zero. Isso explica porque todas as predições para indivíduos 100% Hereford são iguais ao valor fixado para os efeitos ambientais.

Para testar se a inclusão de parâmetros para os efeitos de epistasia e complementariedade efetivamente melhorou o modelo, em termos de ajuste, um teste F foi computado. O valor de F, conforme descrito por WEISBERG (1980), foi calculado pela seguinte fórmula:

$$F = \frac{(SQR_{AD} - SQR_{ADEC}) / (gl_{AD} - gl_{ADEC})}{SQR_{ADEC} / gl_{ADEC}}$$

Em que:  $SQR_{AD}$  e  $SQR_{ADEC}$  são as somas de quadrados dos resíduos do modelo **AD** e **ADEC**, respectivamente; e  $gl_{AD}$  e  $gl_{ADEC}$  são os graus de liberdade do resíduo do modelo **AD** e **ADEC**, respectivamente.

O teste F fornece evidência contra o modelo mais simples (**AD**) se o valor de F é maior que o valor tabelado para uma distribuição  $F(gl_{AD} - gl_{ADEC}, gl_{ADEC}, \alpha)$ .

## Resultados e discussão

Com o modelo **ADEC**, nos quatro métodos de implementação, puderam se verificar valores máximos (com  $\lambda \neq 0$ ) tanto para a soma quanto para a média harmônica dos valores absolutos de  $t$ , permitindo a escolha do valor de  $\lambda$  com base nesses critérios. Já com o modelo **AD**, tanto a soma quanto a média harmônica apresentaram valor máximo com  $\lambda = 0$ , ou seja, segundo esses critérios, as análises por quadrados mínimos ordinários seriam as indicadas. Como os FIVs observados nas análises por quadrados mínimos chegaram à casa dos milhares, indicando severa multicolinearidade, os critérios baseados nos valores absolutos de  $t$ , para esse modelo, não foram considerados apropriados para escolha de  $\lambda$ . Optou-se, então, por empregar apenas o critério baseado nos valores de FIV, para os dois modelos, permitindo-se uma comparação mais adequada entre os dois.

Os valores de  $\lambda$  empregados, segundo o critério baseado nos valores de FIV, foram: 26 para o modelo **AD** e 190 para o modelo **ADEC**, pelo método (1) de implementação da regressão de cumeeira; 0,0004 para o modelo **AD** e 0,0004 para o modelo **ADEC**, pelo método (2); 1,16 para o modelo **AD** e 0,26 para o modelo **ADEC**, pelo método (3); 0,019 para o modelo **AD** e 0,01 para o modelo **ADEC**, pelo método (4).

As análises de regressão conduzidas usando-se o modelo **AD** resultaram em valores de coeficientes de determinação de 0,0463 pelo método (1) de implementação da regressão de cumeeira, 0,0469 pelo método (2), 0,0464 pelo método (3), e 0,0462 pelo método (4). Usando-se o modelo **ADEC**, os valores encontrados foram de 0,0479 pelo método (1), 0,0485 pelo método (2), 0,0481 pelo método (3), e 0,0479 pelo método (4). Vale ressaltar que a baixa magnitude dos valores de  $R^2$  apresentados nessas análises deve-se a absorção do efeito de grupo de contemporâneos, pela qual foram removidos do modelo os efeitos da média e do grupo de contemporâneos. Em média, nos quatro tipos de implementação da regressão de cumeeira, a inclusão de efeitos de epistasia e complementariedade no modelo de avaliação de efeitos genéticos representou um aumento de 3,55% no valor de  $R^2$ .

ALBUQUERQUE & FRIES (2000) compararam um modelo contendo apenas o efeito de grupo de contemporâneos (GC) com outro que, além do GC, incluía efeitos de idade à desmama e data juliana de nascimento no ganho médio diário do nascimento à desmama em gado Nelore. Os autores encontraram valores de  $R^2$  de 0,478 e 0,485 para o primeiro e segundo modelos, respectivamente, indicando, num primeiro momento, uma diferença bastante pequena (0,007) entre os dois modelos, em termos de ajuste. Porém, uma análise de regressão dos efeitos de idade à desmama e data juliana de nascimento no ganho médio diário do nascimento à desmama, após a correção do ganho para o efeito de grupo de contemporâneos, resultou em  $R^2$  de 0,322, bem superior à diferença de 0,007 entre os valores de  $R^2$  encontrados para os dois modelos. Os autores sugerem que parte dos efeitos de idade à desmama e data juliana de nascimento, quando não incluídos explicitamente no modelo, aparece nas estimativas de GC. Se as covariáveis não estão no modelo, quase toda a soma de quadrados correspondente a elas é incluída/confundida no efeito de grupo de contemporâneos.

PIMENTEL *et al.* (2003) compararam dois modelos similares aos modelos **AD** e **ADEC**, porém não incluindo interações dos efeitos genéticos com latitude. As estimativas dos parâmetros de efeitos genéticos obtidas nesse trabalho são apresentadas na Tabela 1.

Tabela 1. Estimativas dos parâmetros obtidas usando-se o modelo 1 e o modelo 2, por quadrados mínimos ordinários (QM) e por regressão de cumeeira (RC).

		AA	AM	AAZ	AMZ	HA	HM	EA	EM
<b>Mod. 1</b>	<b>QM</b>	-0,03513	0,12719			0,03631	0,04738		
<b>Mod. 2</b>	<b>QM</b>	-0,03694	0,11699	-0,29742	0,34847	0,11266	-0,03102	0,03300	-0,07061
	<b>RC</b>	-0,01349	0,07687	0,02755	0,03245	0,02366	0,01939	-0,01370	-0,00997

Os autores observaram que as estimativas para heterose obtidas através do modelo 1 foram superiores às encontradas quando se consideraram efeitos de epistasia e complementariedade. Os coeficientes estimados para HA (0,0363) e HM (0,0473) pelo

modelo 1 se aproximaram do que se obtinha pela adição dos coeficientes estimados pelo modelo 2, por regressão de cumeieira, para HA, AAZ e EA (0,0376), e HM, AMZ e EM (0,0418), respectivamente. Os autores concluíram que pelo uso de modelos mais completos para avaliação de efeitos genéticos em bovinos de corte, pode-se identificar a contribuição de efeitos de epistasia e complementariedade, que encontram-se embutidos no efeito de heterose estimado por modelos mais simples.

KINGHORN & VERCOE (1989) compararam o modelo de dominância com seis modelos alternativos que incluíam parâmetros para epistasia, diferindo entre si quanto à interpretação biológica das interações epistáticas entre dois loci. Os modelos que incluíam epistasia apresentaram melhor ajuste que o modelo de dominância para as treze características estudadas, porém essas diferenças só foram significativas ( $p < 0,05$ ) para duas delas, indicando que a epistasia foi de pouca ou nenhuma importância. Para testar essa conclusão, uma análise similar foi conduzida comparando o grau de ajuste desses seis modelos com um modelo contendo apenas o efeito de epistasia (nenhum parâmetro para efeito de dominância incluído). Os resultados dessa segunda análise revelaram que os modelos que incluíam efeitos de epistasia e dominância só apresentaram ajuste significativamente melhor para quatro das treze características, sugerindo que a dominância tem relativamente pequena importância para a heterose. Os autores atribuem a impossibilidade de uma comparação confiável entre a importância dos efeitos de dominância e epistasia à forte colinearidade apresentada pelos coeficientes para esses dois efeitos. Essas considerações sugerem que sejam testados métodos alternativos de estimação que visem contornar os efeitos da multicolinearidade.

FRIES et al. (2000), em experimento com peso à desmama de bezerros cruzados, reportam redução de 60% na falta de ajuste do modelo aditivo-dominância, pela adoção de um modelo que incluía efeitos de epistasia e complementariedade, quando comparados com um modelo contendo todos os genótipos presentes nos dados como níveis de um efeito classificatório (100% do ajuste possível). Segundo os autores, a estimação de todos esses componentes genéticos, apresentando severa

multicolinearidade na estrutura dos dados, só foi possível graças ao emprego da regressão de cumeeira.

O teste F conduzido para verificar a significância da diferença de ajuste entre os dois modelos resultou em valores de F altamente significativos ( $p < 0,0001$ ), indicando superioridade do modelo **ADEC** em relação ao **AD**, em termos de ajuste. Os valores de F encontrados foram de 16,59 pelo método (1) de implementação da regressão de cumeeira, 17,13 pelo método (2), 16,71 pelo método (3), e 16,51 pelo método (4).

A discordância entre os resultados obtidos neste estudo e os apresentados por KINGHORN & VERCOE (1989), em termos de diferença significativa de ajuste entre os modelos, pode se dever ao emprego da regressão de cumeeira na estimação dos parâmetros, pela qual contornaram-se os efeitos da colinearidade presente na estrutura dos dados.

Nas Figuras 1 a 4 são ilustradas as superfícies construídas com as estimativas obtidas pelos dois modelos, pelos quatro métodos de implementação da regressão de cumeeira.

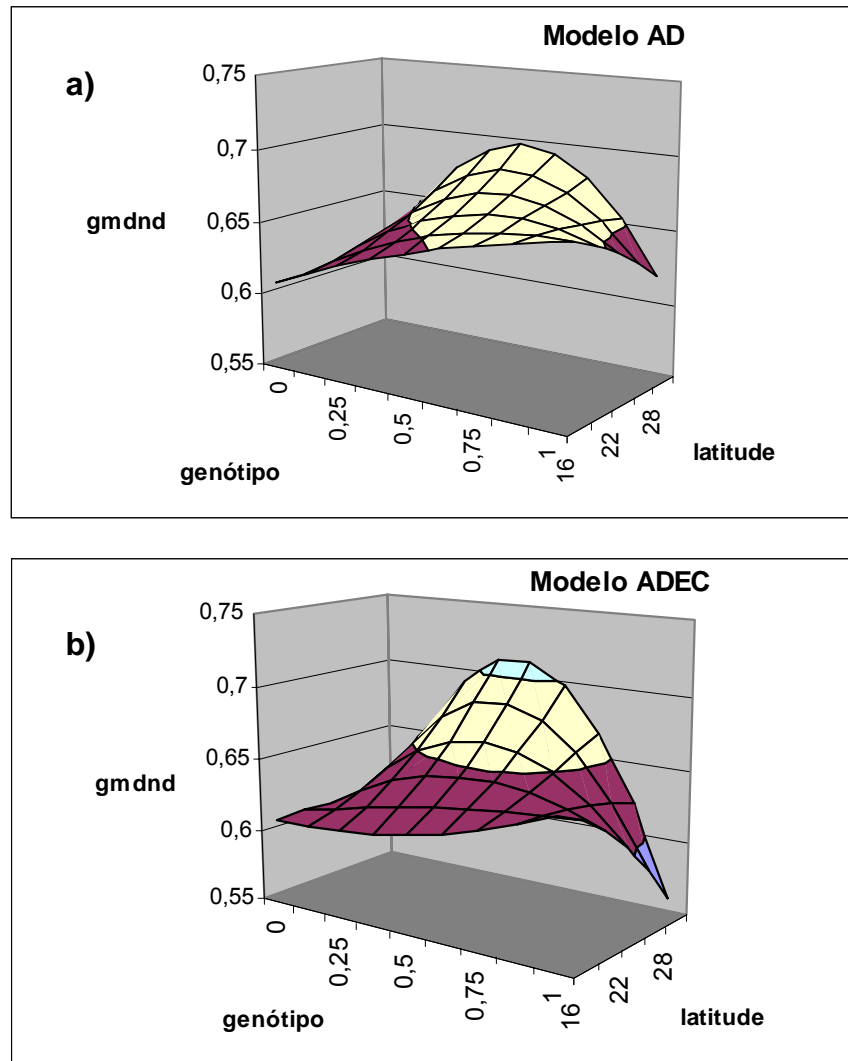


Figura 1. Predições de **gmdnd** obtidas a partir dos coeficientes estimados usando-se o modelo **AD** e **ADEC** pelo método (1) de implementação da regressão de cumeeira.

$$\hat{\beta}_{RC} = (X'X + \lambda I)^{-1} X'Y$$

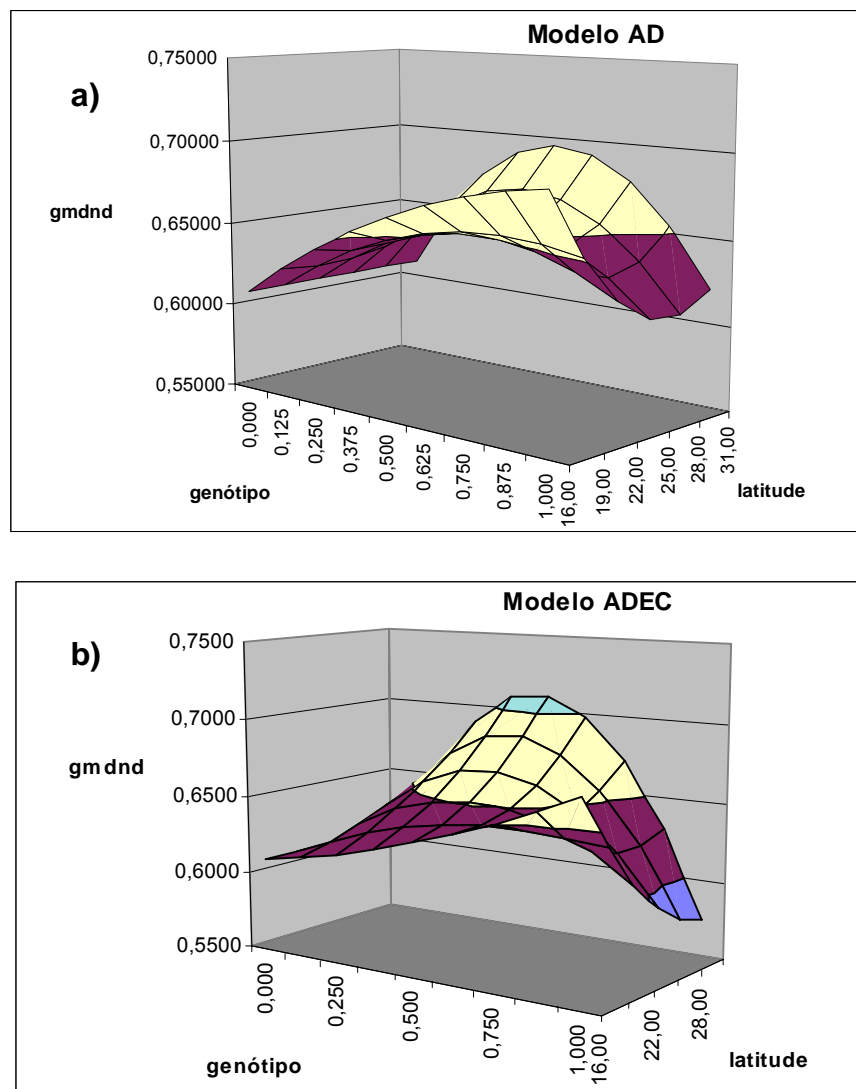


Figura 2. Predições de **gmdnd** obtidas a partir dos coeficientes estimados usando-se o modelo **AD** e **ADEC** pelo método (2) de implementação da regressão de cumeeira.

$$\hat{\beta}_{SAS} = [R_{XX}(I + D_{\lambda})]^{-1}R_{XY}$$

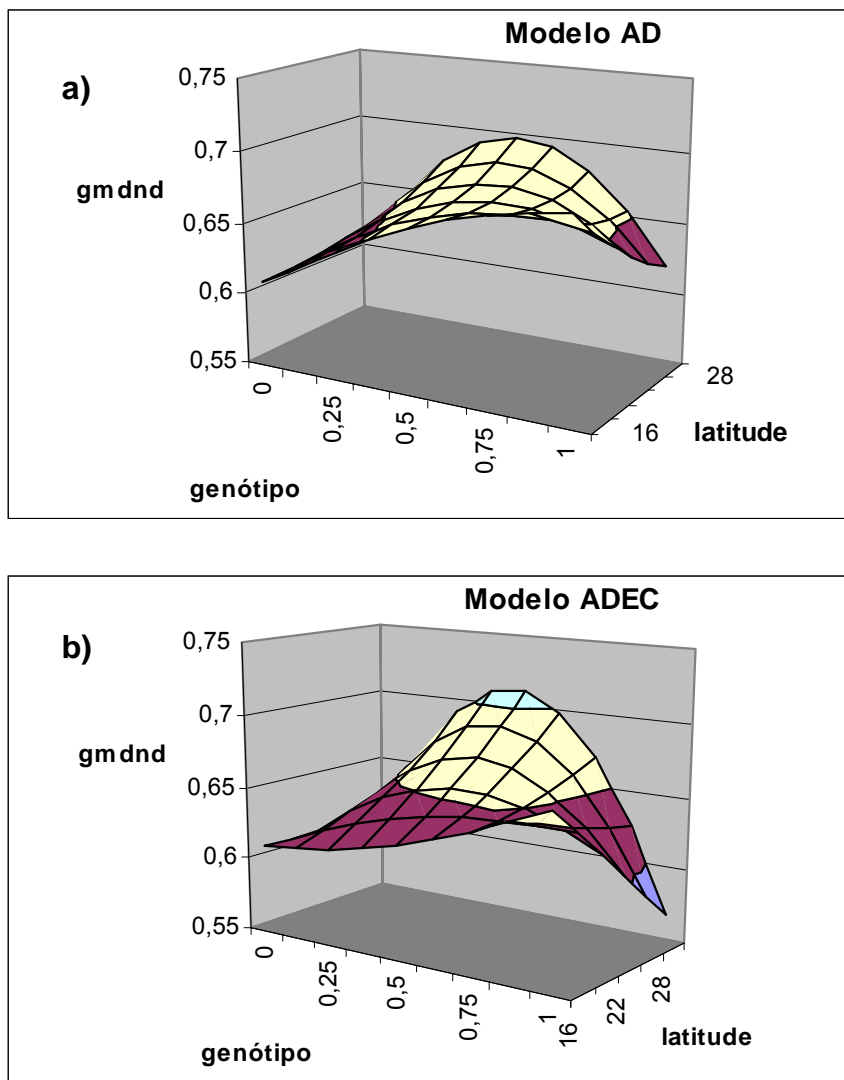


Figura 3. Predições de **gmdnd** obtidas a partir dos coeficientes estimados usando-se o modelo **AD** e **ADEC** pelo método (3) de implementação da regressão de cumeieira.

$$\hat{\beta}_{RCW} = (X'X + \lambda W)^{-1} X'Y$$



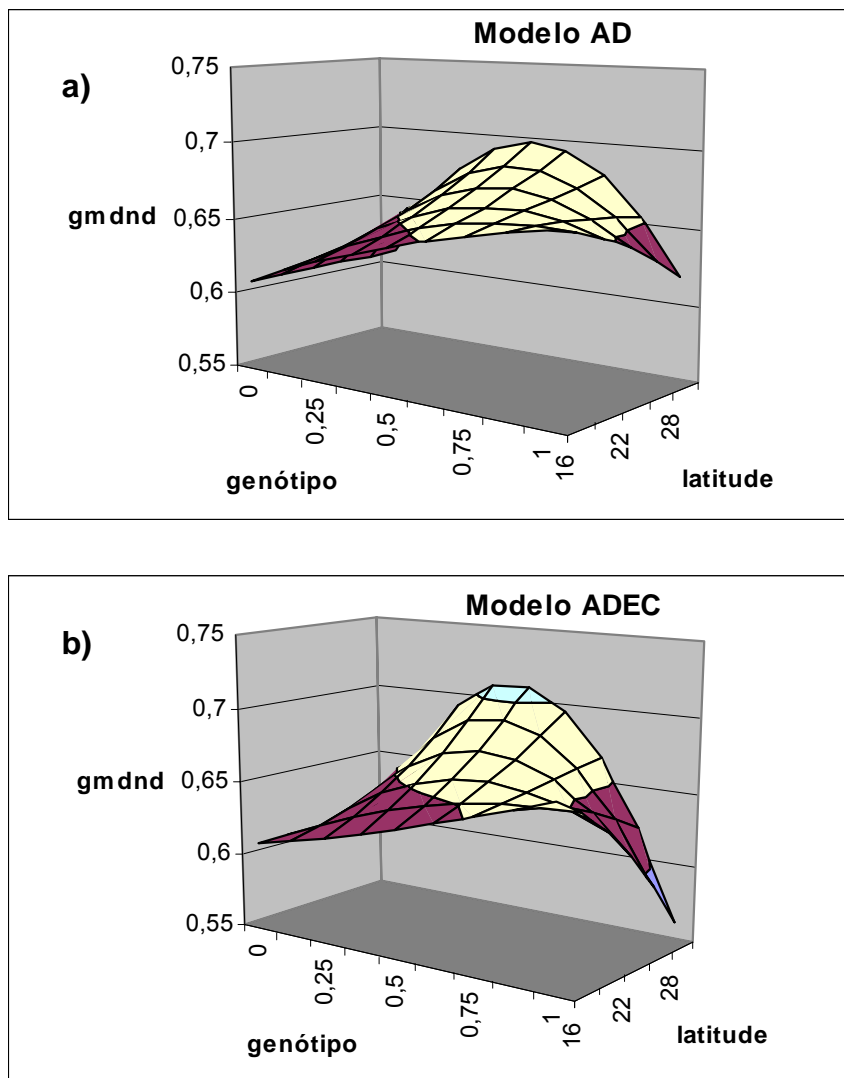


Figura 4. Predições de **gmdnd** obtidas a partir dos coeficientes estimados usando-se o modelo **AD** e **ADEC** pelo método (4) de implementação da regressão de cumeeira.

$$\hat{\beta}_{FIV} = (X'X + \lambda F)^{-1} X'Y$$

Nas Figuras 1a e 1b podem ser observadas as superfícies de predição construídas a partir dos coeficientes estimados com os dois modelos pelo método (1) de implementação da regressão de cumeieira. As duas superfícies apresentam uma similaridade em tendência, ambas mostrando: para a latitude 16°S, aumento em **gmdnd** à medida que cresce a proporção de Nelore na composição genética dos animais; decréscimo nas predições de **gmdnd** de animais Nelore à medida em que as latitudes se tornam mais elevadas; e superioridade dos animais de composição raciais intermediárias, exceto nas latitudes mais baixas, onde o desempenho predito dos indivíduos 100% Nelore superam todas as outras composições raciais, revelando que nessas latitudes a adaptação a clima tropical do zebu parece assumir um papel mais determinante do desempenho do que o potencial genético de crescimento do taurino somado aos efeitos de heterose.

FACÓ *et al.* (2002), estudando diversos grupos genéticos Holandês x Gir no Brasil, não verificaram qualquer benefício em elevar a proporção de genes da raça Holandesa sob condições de ambiente hostis.

O que se percebe na comparação das duas superfícies é que a correspondente ao modelo **ADEC** se apresenta mais acentuada do que a correspondente ao modelo **AD**. A inclusão de parâmetros para epistasia e complementariedade parece ter promovido um incremento no poder discriminatório do modelo, quanto à capacidade de predição de desempenho de diferentes composições raciais.

As superfícies obtidas com os métodos de implementação (3) e (4) confirmam os resultados observados para o método (1), como se pode verificar pela comparação entre as Figuras 1, 3 e 4.

Na Figura 2 são apresentados os resultados obtidos com o método de implementação (2). A superfície correspondente ao modelo **AD** (fig. 2a) apresenta um comportamento bem diferente de todas as outras superfícies construídas a partir dos dois modelos, pelos quatro métodos. Em latitudes intermediárias, observam-se predições de **gmdnd** inferiores às preditas em latitudes extremas, para todas as composições raciais. Os valores preditos de **gmdnd** dos animais Nelore nas latitudes

mais altas superam os preditos para latitudes intermediárias, indo contra o que se poderia esperar, ou ao que seria biologicamente mais coerente.

## Conclusões

A inclusão de parâmetros para efeitos de epistasia e complementariedade em modelos de avaliação de efeitos genéticos em animais cruzados pode representar um ganho tanto em termos de ajuste do modelo quanto de capacidade de predição de desempenho de genótipos não testados.

O método da regressão de cumeieira é uma ferramenta que possibilita a obtenção de estimativas de efeitos de ação aditiva, dominância, epistasia e complementariedade, mesmo na presença de forte colinearidade na estrutura dos dados.

## Referências

- ALBUQUERQUE, L.G.; FRIES, L.A. Consequences of Ignoring some Fixed Effects on Preweaning Average Daily Gain in Nelore Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, supl. B, p. 216, 2000.
- ARTHUR, P.F.; HEARNshaw, H.; STEPHENSON, P.D. Direct and maternal additive and heterosis effects from crossing *Bos indicus* and *Bos taurus* cattle: cow and calf performance in two environments. **Livestock Production Science**, v.57, p.231-241. 1999.
- BÉRTOLI, C.D. **Sistema Cruza - Controle de Produção e Avaliação dos Valores Genéticos dentro de uma População Bovina Sintética**. Dissertação (Mestrado em Agronomia) Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 1991. 233p.

- BRITO, F.V.; PICCOLI, M.L.; SEVERO, J.L.P.; SCHENKEL, F.S.; ROSO, V.M.; FRIES, L.A. Estimating environmental and genotypic effects on preweaning weight gain of Angus x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** p.773-776.
- BURKE, J.M.; ARNOLD, M.L. Genetics and the Fitness of Hybrids. **Annual Review of Genetics**, v.35, p.31-52. 2001.
- CASSADY, J.P.; YOUNG, L.D.; LEYMASTER, K.A. Heterosis and recombination effects on pig growth and carcass traits. **Journal of Animal Science**, v.80, p.2286-2302. 2002.
- DEMEKE, S.; NESER, F.W.C.; SCHOEMAN, S.J. Early growth performance of *Bos taurus* x *Bos indicus* cattle crosses in Ethiopia: Evaluation of different crossbreeding models. **Journal of Animal Breeding and Genetics**, v.120, p.39-50. 2003.
- FACÓ, O.; LÔBO, R.N.B.; MARTINS FILHO, R.; MOURA, A.A.A. Análise do Desempenho Produtivo de Diversos Grupos Genéticos Holandês x Gir no Brasil. **Revista Brasileira de Zootecnia**, v.31, n.5, p.1944-1952. 2002.
- FREUND, R.J.; LITTELL, R.C. **SAS System for Regression**, 3rd ed. Cary: SAS Institute Inc., 2000. 235p.
- FRIES, L.A.; JOHNSTON, D.J.; HEARNshaw, H.; GRASER, H.U. Evidence of Epistatic Effects on Weaning Weight in Crossbred Beef Cattle. **Asian-Australasian Journal of Animal Sciences**, Marrickville, v.13, supl. B, p. 242, 2000.
- GREGORY, K.E.; CUNDIFF, L.V.; KOCH, R.M. Breed effects and heterosis in advanced generations of composite populations for growth traits in both sexes of beef cattle. **Journal of Animal Science**, v.69, p.3202-3212. 1991.
- HIROOKA, H.; GROEN, A.F.; VAN DER WERF, J.H.J. Estimation of additive and non-additive genetic parameters for carcass traits on bulls in dairy, dual purpose and beef cattle breeds. **Livestock Production Science**, v.54, p.99-105. 1998.
- HOERL, A.E.; KENNARD, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v.12, n.1, p.55-67. 1970.
- KAHI, A.K.; THORPE, W.; NITTER, G.; BAKER, R.L. Crossbreeding for dairy production in the lowland tropics of Kenya. I. Estimation of individual crossbreeding effects on

- milk production and reproductive traits and on cow live weight. **Livestock Production Science**, v.63, p.39-54. 2000.
- KINGHORN, B.P. Theory of Breed Utilisation. In: ANIMAL GENETICS AND BREEDING UNIT. AGBU. **Design of Livestock Breeding Programs**: Short course in animal Breeding. Armidale, 1993. part.4, chap.19, p.187-204.
- KINGHORN, B.P.; VERCOE, P.E. The effects of using the wrong genetic model to predict the merit of crossbred genotypes. **Animal Production**, v.49, p.209-216. 1989.
- LECLERC, G.; PIREAUX, J.J. The use of least squares for XPS peak parameters estimation. Part 3. Multicollinearity, ill-conditioning and constraint-induced bias. **Journal of Electron Spectroscopy and Related Phenomena**, v. 71, p.179-190. 1995.
- MACKAY, T.F.C. The Genetic Architecture of Quantitative Traits. **Annual Review of Genetics**, v.35, p.303-339. 2001.
- MOHAMED, S.A.; ROTTMANN, O.; PIRCHNER, F. Components of heterosis for growth traits and litter size in line crosses of mice after long-term selection. **Journal of Animal Breeding and Genetics**, v.118, p.263-270. 2001.
- PICCOLI, M.L.; ROSO, V.M.; BRITO, F.V.; SEVERO, J.L.P.; SCHENKEL, F.S.; FRIES, L.A. Additive, complementarity (additive\*additive), dominance, and epistatic effects on preweaning weight gain of Hereford x Nelore calves. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier, **Proceedings...** CD-ROM.
- PIMENTEL, E.C.G.; QUEIROZ, S.A.; CARVALHEIRO, R.; FRIES, L.A. Efeitos da inclusão de epistasia e complementariedade em modelos de avaliação genética em bovinos de corte. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 40., 2003, Santa Maria, **Anais...** CD-ROM.
- ROBERSON, R.L.; SANDERS, J.O.; CARTWRIGHT, T.C. Direct and Maternal Genetic Effects on Preweaning Characters of Brahman, Hereford and Brahman-Hereford Crossbred Cattle. **Journal of Animal Science**, v.63, p.438-446. 1986.
- SCHENKEL, F.S. **Cálculo das heterozigoses**. Porto Alegre, RS: Gensys Consultores Associados. 1993.

SHARMA, B.S.; PRABHAKARAN, V.T.; PIRCHNER, F. Gene action and heterosis in lifetime traits of Friesian x Sahiwal crosses. **Journal of Animal Breeding and Genetics**, v.117, p.319-330. 2000.

WEISBERG, S. **Applied Linear Regression**. 1st ed. New York: Wiley Series in probability and mathematical statistics, 1980. 283p.

WRIGHT, S. The relation of livestock breeding to theories of evolution. **Journal of Animal Science**, v.46, n.5, p.1192-1200. 1978.

## CAPÍTULO 5 – IMPLICAÇÕES

O programa computacional desenvolvido neste trabalho permite que quatro métodos alternativos de implementação da regressão de cumeieira sejam testados, fornecendo três critérios numéricos para escolha do valor de  $\lambda$ , além de gerar saídas com os valores das estimativas dos parâmetros e dos respectivos valores de  $t$ . Essas opções oferecem ao usuário uma maior quantidade de informação para auxiliar uma tomada de decisão do que outros pacotes computacionais disponíveis para análises de regressão, como o SAS, por exemplo.

Outro aspecto interessante do programa decorre do uso das rotações de Givens para obtenção das soluções, que permite a armazenagem do arquivo de dados na forma de uma matriz triangular superior. Dessa forma, atualizações das soluções do sistema podem ser feitas com menor tempo de processamento e demanda por memória convencional.

Pelo uso da regressão de cumeieira, puderam-se contornar os efeitos da multicolinearidade verificada ao se usarem modelos para avaliação de efeitos genéticos, em animais cruzados, que incluem efeitos de epistasia e complementariedade, além dos aditivos e heteróticos usualmente estimados. Com isso, pôde-se decompor o efeito de heterose, estimado por modelos mais simples, em efeitos de dominância, epistasia e complementariedade. Uma identificação mais adequada dos efeitos genéticos envolvidos na expressão de características de animais cruzados pode melhorar a qualidade das predições de valor genético em avaliações de populações multirraciais, e predições de desempenho de genótipos não testados para planejamento de programas de cruzamentos.

## APÊNDICE

Programa fonte do sistema GIVENS, em linguagem Fortran 90, adaptado para implementação da regressão de cumeeira:



```

$DEBUG
c$STORAGE:2
c      interface to subroutine time (n,str)
c      character*10 str [near,reference]
c      integer*2 n [value]
c      end
CHARACTER*8 NOMEVC(20),NOMEVCV(40),NOMEY(20)
c      CHARACTER*10 TSTR
CHARACTER*15 INAME
CHARACTER*127 INFMT
CHARACTER*8 REG1VC(20)
CHARACTER*1 VC,CV,T1,CE
CHARACTER*3 FIM
CHARACTER*1 RC
INTEGER NBVC,NBCV,NBY,IPOS(20),intvcd(20,2),nbivcd,ivcd
INTEGER IRC,ICOV,calc(40)
REAL REG1CV(40),REG1Y(20),elem(40)
REAL LBEG,LEND,LBY,LAMB,lcom,vlfiv
      noint=0
      WRITE(*,997)
      WRITE(*,997)
      WRITE(*,9001)
      WRITE(*,9002)
      WRITE(*,9003)
      WRITE(*,9004)
      WRITE(*,9005)
      WRITE(*,9006)
      WRITE(*,9007)
      WRITE(*,9008)
      WRITE(*,9009)
      WRITE(*,9010)
      WRITE(*,9011)
      WRITE(*,9012)
      WRITE(*,9013)
      WRITE(*,9014)
      WRITE(*,9015)
      WRITE(*,997)
      WRITE(*,999)' VERSÃO GR1 - 23 DE SETEMBRO DE 1987.'
      WRITE(*,997)
      WRITE(*,999)' DOMINIO PUBLICO, para fins n,º-comerciais.'
      PAUSE 'Tecla CR para continuar, por favor.'
write(*,997)
write(*,997)
write(*,999)'      SOLUÇÃO DE MODELOS LINEARES FIXOS OU MISTOS'
WRITE(*,999)'      ATRAV S DAS ROTAS™ES PLANARES DE GIVENS'
WRITE(*,999)'      APLICADAS SOBRE A MATRIZ DE DADOS.'
      WRITE(*, '(/,A,/))' '      Vers,º Rapida 1.'
WRITE(*,999)' O sistema GIVENS-GR1 permite a an lise de at,:'
WRITE(*,999)' 5 respostas(vari veis dependentes), conforme'
WRITE(*,999)'      mod'los contendo at,'
WRITE(*,999)' 5 vari veis classificat'rias, sendo que a'
WRITE(*,999)' primeira e a segunda informadas podem conter
*ate 200 niveis e'
write(*,999)'      da terceira a quinta at, 20 n,veis cada; e a
8te '
write(*,999)' 5 vari veis cont;nuas (covari veis ou regressores)
*.'
write(*,997)
write(*,999)' Estes limites foram impostos prevendo um equipamento
* com 640Kb de RAM e'
write(*,999)' dois acionadores de discos.'

```

```

write(*,997)
      pause ' Tecla CR para continuar, por favor.'
write(*,997)
write(*,997)
write(*,997)
write(*,997)
write(*,999)' Este sistema foi desenvolvido com o apoio do FINEC -
* FUNDAÇÃO BANCO DO BRASIL'
WRITE(*,999)' e CNPq - CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTI
*FICO E TECNOLÓGICO.'
write(*,997)
write(*,997)
write(*,997)
write(*,999)' Solicitam-se informa'ões, cr'íticas e coment rios sob
*re o desempenho do sistema.'
write(*,997)
write(*,999)' Para obter manuais, fltimas vers'ões e enviar sugest
*ões, escreva para:'
write(*,999)'      Projeto Givens'
write(*,999)'      DZ-AGRO-UFRGS'
write(*,999)'      Caixa Postal 776'
write(*,999)'      (0512)365011 - Ramal 46'
write(*,999)'      90.001-PORTO ALEGRE-RS'
      PAUSE 'Tecla CR para continuar, por favor.'
9001 FORMAT(13X,4('G'),7X,'I',6X,'V',5X,'V',5X,4('E'),4X,'N',6X,'N',4X,
*4('S'))
9002 FORMAT(12X,'G',4X,'G',6X,'I',6X,'V      V',4X,'E      E      NN',5X,
*'N      S      S')
9003 FORMAT(11X,'G      G      I      V      V      E      E      NN      N      S',
*6X,'S')
9004 FORMAT(11X,'G',12X,'I',6X,'V      V      E',9X,'N      N      N      S')
9005 FORMAT(11X,'G',12X,'I',7X,'V      V      E',9X,'N      N      N      S')
9006 FORMAT(11X,'G',12X,'I',7X,'V      V      E',9X,'N      N      N      S')
9007 FORMAT(11X,'G',12X,'I',7X,'V      V      E',9X,'N      N      N      S')
9008 FORMAT(11X,'G',12X,'I',7X,'V      V      E',9X,'N      N      N      S')
9009 FORMAT(11X,'G',12X,'I',8X,'V      V      E',9X,'N      N      N      S')
9010 FORMAT(11X,'G      GGGGG',5X,'I',8X,'V      V      EEEEE      N      N      N',4X,
*4('S'))
9011 FORMAT(11X,'G      G      I      V      V      E',9X,'N      N      N',8X,'S')
9012 FORMAT(11X,'G      G      I      V      V      E',9X,'N      N      N',9X,'S')
9013 FORMAT(11X,'G      G      I      V      E      E      N      N      N',
*9X,'S')
9014 FORMAT(12X,'G      G      I      V      E      E      N      NN      S',
*5X,'S')
9015 FORMAT(13X,'GGGG      I      V      EEEE      N',6X,'N      ',
*'SSSS')
write(*,997)
999 FORMAT(A)
997 FORMAT('0')
995 FORMAT(A,\)
899 FORMAT(A1)
10 CONTINUE
20 CONTINUE
25 CONTINUE
write(*,997)
write(*,999)' Como seu primeiro passo, , necess rio que voce indi
2que qual o '
write(*,999)' formato de entrada dos seus dados. Informe este for
2mato seguindo as '
write(*,999)' conven'ões da linguagem Fortran77. Veja exemplos no
2 seu manual.'

```

```

WRITE(*,999)' RESTRIÇÃO: voce dispoe de 127 colunas, no maximo, para
2 para explicitar '
write(*,999)' o formato dos seus dados de entrada.'
WRITE(*,997)
WRITE(*,999)'   NOTA: use as mesmas formas de leitura (T) para que,
write(*,999)'   nos formatos de entrada, voce indique:'
write(*,999)'   PRIMEIRO, os formatos das variaveis clas-
write(*,999)'   sificatçrias;'
WRITE(*,999)'   SEGUNDO, os formatos das covariaveis; e'
write(*,999)'   TERCEIRO, os formatos das respostas.'
write(*,997)
WRITE(*,999)'   P.S.: Nao esqueça de abrir e fechar o formato com
2 um par de parenteses.'
write(*,997)
write(*,997)
WRITE(*,975)
975 FORMAT(' Entre com o formato dos seus dados:')
READ(*,999,ERR=25) INFMT
WRITE(*,999)' O seu formato de leitura ,:'
WRITE(*,972) INFMT
972 FORMAT(1X,A)
write(*,999)' Confirma e responde (S/N) se este formato de entrada
2 de dados esta
WRITE(*,970)
970 format(' correto:')\
READ(*,899) T1
IF ((T1.EQ.'N').OR.(T1.EQ.'n')) GOTO 25
31 CONTINUE
36 CONTINUE
write(*,997)
WRITE(*,999)' Nosso proximo passo sera definir: '
write(*,999)'   PRIMEIRO, os nomes das variaveis classifica-
write(*,999)'   çrias (se houverem); '
write(*,999)'   SEGUNDO, os nomes das covariaveis do seu mo-
write(*,999)'   delo (se houverem); e '
write(*,999)'   TERCEIRO, os nomes das respostas ou vari-
write(*,999)'   veis dependentes.'
write(*,997)
WRITE(*,999)' NOTA 1: A indicaçao destes nomes deve seguir exa-
write(*,999)'   tamente a mesma ordem daquela fornecida
write(*,999)'   para o formato de leitura dos dados.'
write(*,997)
WRITE(*,999)' NOTA 2: Os nomes das (co-)variaveis podem ocupar um
write(*,999)'   maximo de 6(seis) colunas.'
WRITE(*,997)
38 continue
WRITE(*,983)
983 FORMAT(' Existem variaveis classificatçrias no seu modelo ou dados
2 (S/N) :'\
READ(*,899) VC
IF ((VC.EQ.'S').OR.(VC.EQ.'s')) THEN
FIM=' '
I=0
40 CONTINUE
I=I+1
WRITE(*,997)
WRITE(*,950) I
950 FORMAT(' Entre com o nome da variavel classificatçria nª ',i2,
2' :'\
READ(*,948) NOMEVC(I)
948 FORMAT(A6)
42 CONTINUE
write(*,999)' Se esta , a ultima variavel classificatçria, digite'
write(*,999)' FIM , ou qualquer outra letra, e tecle RETURN.'
write(*,999)' Caso contrario, dê um espaco e apçs tecle '
write(*,963)
963 format(1x,'RETURN :'\
READ(*,'(A3)',ERR=42) FIM
IF(FIM.EQ.' ') GOTO 40
WRITE(*,997)
WRITE(*,999)' Os nomes das variaveis classificatçrias s,o:'
WRITE(*,999)' Nª ORDEM '
WRITE(*,999)' DE LEITURA NOME '
WRITE(*,999)' _____ '
DO 45 J=1,I
45 WRITE(*,962) J,NOMEVC(J)
962 FORMAT(4X,I2,11X,A6,/)
44 CONTINUE
write(*,997)
write(*,999)' Voce aceita estes nomes e esta a relaçao completa?'
write(*,995)' Responda aqui (S/N):'
read(*,'(a1)',err=44) ce
IF((CE.EQ.'N').OR.(CE.EQ.'n')) GOTO 38
NBVC=I
ELSE
NBVC=0
ENDIF
WRITE(*,997)
WRITE(*,997)
nbivcd=0
if(nbvc.le.1) goto 46
43 continue
write(*,999)' Existem interaçoes duplas entre as variaveis classif
2 icatçrias?:'
write(*,995)' Responda aqui(S/N):'
read(*,'(a1)') ce
if(ce.eq.'n') ce='N'
if(ce.eq.'N') goto 46
i=0
47 continue
write(*,999)' Indique, no formato 2I2, quais os numeros de ordem
2 das variaveis classificatç-
write(*,999)' rias para as quais voce deseja averiguar se existe
2 interaçao dupla.'
48 continue
write(*,999)' Indique o par de numeros de ordem a seguir:'
write(*,999)' 1234'
i=i+1
read(*,'(2i2)',err=48) intvcd(i,1),intvcd(i,2)
write(*,999)' As variaveis classificatçrias entre as quais
2 existe interaçao dupla s,o:'
i1=intvcd(i,1)
i2=intvcd(i,2)
write(*,'(4x,a6,a3,a6,a1)') nomevc(i1),' e ',nomevc(i2),'.'
write(*,995)' Voce concorda com esta informaçao(S/N)?'
read(*,'(a1)') ce
if(ce.eq.'n') ce='N'
if((ce.eq.'N').and.(i.eq.1)) goto 43
if((ce.eq.'N').and.(i.gt.1)) then
i=i-1
goto 47
endif

```

```

49 continue
write(*,999)' Se esta , a fltima intera#o dupla entre as vari vei
2s classificat#rias, digite'
write(*,999)' FIM, ou qualquer outra letra, e tecler RETURN.'
write(*,999)' Caso contr rio, de um espao e ap#s tecler
write(*,963)
read(*,'(a3)',err=49) fim
if(fim.eq.' ') go to 48
nbivcd=i
WRITE(*,997)
WRITE(*,997)
46 CONTINUE
WRITE(*,980)
980 FORMAT(' Existem covari veis no seu modelo ou dados (S/N) :'\)
READ(*,899) CV
IF ((CV.EQ.'S').OR.(CV.EQ.'s')) THEN
C
50 CONTINUE
FIM=' '
I=0
52 CONTINUE
I=I+1
WRITE(*,997)
WRITE(*,945) I
945 FORMAT(' Entre com o nome da covari vel n# ,i2, ' :'\)
read(*,948) nomecv(i)
write(*,999)'
C WRITE(*,999)' Digite l para que essa covari vel seja inclu;da ou'
C WRITE(*,999)' 0 para que ela n#o seja inclu;da ou no c lculo da'
C WRITE(*,999)' soma e m,dia harm#nica dos valores absolutos de t.'
C read(*,949) calc(i)
C 949 FORMAT(i1)
55 CONTINUE
WRITE(*,999)' Se esta , a ultima covari vel, digite FIM, ou qualqu
2er outra letra,'
write(*,999)' e tecler RETURN.'
write(*,999)' Caso contr rio, d^ um espao e ap#s tecler
write(*,963)
READ(*,'(A3)',ERR=55) FIM
IF (FIM.EQ.' ') GOTO 52
WRITE(*,997)
WRITE(*,999)' Os nomes das covari veis s,o:'
WRITE(*,999)' N# ORDEM '
WRITE(*,999)' DE LEITURA NOME '
WRITE(*,999)' _____ '
do 60 j=1,i
60 write(*,965) j,nomecv(j)
965 FORMAT(4X,I2,11X,A6,/)
write(*,997)
62 continue
write(*,999)' Voce aceita estes nomes e est a rela#o completa?'
write(*,995)' Responda aqui (S/N):'
READ(*,'(A1)',ERR=62) CE
IF ((CE.EQ.'N').OR.(CE.EQ.'n')) GOTO 46
NBCV=I
if((vc.eq.'n').or.(vc.eq.'N')) then
write(*,999)' Voce deseja a inclus#o de uma constante no modelo?
2'
write(*,995)' Responda aqui (S/N):'
read(*,'(a1)') ce
if((ce.eq.'N').or.(ce.eq.'n')) noint=1

```

```

endif
ELSE
NBCV=0
IF((VC.EQ.'N').OR.(VC.EQ.'n')) THEN
WRITE(*,999)' Existe algo de estranho no seu problema. Como varia
2veis independentes'
write(*,999)' voce deve ter ou vari veis classificat#rias ou covar
2 veis.'
write(*,999)' Volte a informar novamente o problema. '
write(*,999)' Caso queira abortar a execut#o e repensar seu proble
2ma, pressione CONTROL-C'
GOTO 10
ENDIF
ENDIF
WRITE(*,997)
WRITE(*,997)
65 CONTINUE
FIM=' '
I=0
67 CONTINUE
I=I+1
WRITE(*,997)
write(*,942) i
942 format(' Entre com o nome da resposta (vari vel dependente) de n#
2 ',i2,' :'\)
read(*,948) NOMEY(I)
70 CONTINUE
WRITE(*,999)' Se esta , a fltima resposta, digite FIM, ou qualquer
2 outra letra,'
write(*,999)' e tecler RETURN.'
WRITE(*,999)' Caso contr rio, d^ um espao em branco e ap#s tecler'
write(*,963)
READ(*,'(A3)',ERR=70) FIM
IF(FIM.EQ.' ') GOTO 67
WRITE(*,997)
WRITE(*,999)' Os nomes das respostas s,o:'
WRITE(*,999)' N# ORDEM '
WRITE(*,999)' DE LEITURA NOME '
WRITE(*,999)' _____ '
do 75 J=1,I
75 WRITE(*,962) J,NOMEY(J)
WRITE(*,997)
77 CONTINUE
write(*,999)' Voce aceita estes nomes e est a rela#o completa?'
write(*,995)' Responda aqui (S/N):'
READ(*,'(A1)',ERR=77) CE
IF ((CE.EQ.'N').OR.(CE.EQ.'n')) GOTO 65
NBY=I
80 CONTINUE
WRITE(*,997)
WRITE(*,999)' O pr#ximo passo ser voce indicar o nome do arquivo
2onde est#o os dados.'
write(*,999)' O arquivo de dados dever estar no acionador B:'
write(*,999)' Ao fornecer o nome do arquivo, n#o esque#a de inicia
*r por B:'
write(*,940)
format(1x,' O nome do arquivo de dados ,:'\)
read(*,999,err=80) iname
OPEN(5,FILE=INAME)
WRITE(*,997)
WRITE(*,999)' O primeiro registro ser lido e impresso para que v

```

```

2oçe possa conferir'
write(*,999)' e decidir qual o próximo passo a tomar.'
write(*,997)
write(*,999)' O primeiro registro cont,m, dentro das seguintes var
2i veis, os valores:'
write(*,997)
if((vc.eq.'S').or.(vc.eq.'s')) then
IF((CV.EQ.'S').OR.(CV.EQ.'s')) THEN
IOS=1
READ(5,INFMT,ERR=69) (REG1VC(I),I=1,NBVC),(REG1VC(J),J=1,NBCV),
2 (REG1Y(K),K=1,NBY)
WRITE(*,999)' -----VARI+VEIS CLASSIFICAT+RIAS'
DO 85 I=1,NBVC
85 WRITE(*,935) NOMEVC(I),REG1VC(I)
935 FORMAT(1X,A6,' : ',A20)
WRITE(*,999)' -----COVARI+VEIS-----'
DO 86 J=1,NBCV
86 WRITE(*,933) NOMEVC(J),REG1VC(J)
933 FORMAT(1X,A6,' : ',F20.6)
ELSE
C
C S• TEM VARI+VEIS CLASSIFICAT+RIAS
C
IOS=2
READ(5,INFMT,ERR=69) (REG1VC(I),I=1,NBVC),(REG1Y(K),K=1,NBY)
WRITE(*,999)' -----VARI+VEIS CLASSIFICAT+RIAS'
DO 90 I=1,NBVC
90 WRITE(*,935) NOMEVC(I),REG1VC(I)
ENDIF
ELSE
C
C S• TEM COVARI+VEIS
C
IOS=3
READ(5,INFMT,ERR=69) (REG1VC(J),J=1,NBCV),(REG1Y(K),K=1,NBY)
WRITE(*,999)' -----COVARI+VEIS-----'
DO 94 J=1,NBCV
94 WRITE(*,933) NOMEVC(J),REG1VC(J)
ENDIF
C
WRITE(*,999)' -----RESPOSTAS--'
DO 95 K=1,NBY
95 WRITE(*,933) NOMEY(K),REG1Y(K)
GOTO 100
69 CONTINUE
WRITE(*,999)' Verifique o formato fornecido ou o arquivo de dados
2. N,º foi possível'
write(*,999)' ler o primeiro registro.'
GOTO 25
100 CONTINUE
WRITE(*,999)' Voçe concorda com estes dados e com a sua formata+o
2 ?'
write(*,995)' Responda aqui (S/N):'
READ(*,'(A1)') CE
IF((CE.EQ.'S').OR.(CE.EQ.'s')) THEN
WRITE(*,999)' A sua resposta indica que a leitura do arquivo compl
2eto pode continuar.'
write(*,999)' Nesta primeira leitura ser pesquisado no seu arquiv
2o os seguintes'
write(*,999)' parâmetros : '
write(*,999)' 1. Se existirem vari veis classificat+rias,para cada

```

```

2 uma destas ser '
write(*,999)' calculado:'
write(*,999)' - número de classes e'
write(*,999)' - número de observat+es por classe.'
write(*,999)' 2. Se existirem covari veis, para cada uma destas se
2r calculada:'
write(*,999)' - a m,dia e'
write(*,999)' - a varifncia.'
write(*,999)' 3. Para cada resposta, ser calculada:'
write(*,999)' - a m,dia e'
write(*,999)' - a varifncia.'
write(*,999)' '
write(*,999)' Al,m disto, o tamanho total do problema ser deter
2minado, permitindo'
write(*,999)' conhecer o número de parâmetros a ser estimado e a
2 necessidade de '
write(*,999)' memória dinfmica(RAM).'
REWIND 5
123 CONTINUE
C L^ os dados conforme os valores de IOS
C
C 1= VC,CV,Y
C 2= VC,Y
C 3= CV,Y
C
c call time(10,tstr)
c write(*,*) ' Inicio da primeira leitura do arquivo ...s ',tstr
C
CALL LE1(NBVC,NBCV,NBY,NOMEVC,NOMEVC,NOMEY,INAME,INFMT,ios,
2NTOTVC,IREG,IPOS,NTAMU,NRS,nbivcd,ivcd,intvcd,noint,RC,LBEG,
2LEND,LBY,IRC,lcom,vlfiv,calc)
C
c call time(10,tstr)
c write(*,*) ' Final da primeira leitura do arquivo ...s ',tstr
REWIND 5
132 CONTINUE
WRITE(*,997)
WRITE(*,999)' Rota+es Givens ser,º aplicadas na Matriz de Dados.'
c call time(10,tstr)
CALL LE2(NBVC,NBCV,NBY,NOMEVC,NOMEVC,NOMEY,INAME,INFMT,ios,
2NTOTVC,IREG,IPOS,NTAMU,NRS,nbivcd,ivcd,intvcd,noint,IRC,LBEG,
2LEND,LBY,LAMB,elem,ICOV,lcom,vlfiv,RC,calc)
c write(*,*) ' Inicio da segunda leitura e das rota+es ...s ',tstr
c call time(10,tstr)
c write(*,*) ' Final da segunda leitura e das rota+es ...s ',tstr
WRITE(*,997)
write(*,999)' Use o comando-DOS: A>TYPE GR1.OUT >PRN'
write(*,999)' A>TYPE GR2.OUT >PRN'
write(*,999)' para obter um relatorio impresso.'
write(*,999)' Use o comando-DOS: A>TYPE GR1.OUT |MORE'
write(*,999)' A>TYPE GR2.OUT |MORE'
write(*,999)' para obter o relatorio na tela.'
ELSE
200 CONTINUE
WRITE(*,997)
WRITE(*,999)' Indique a seguir qual o seu problema nos dados, atra
2v,s da letra correspondente'
WRITE(*,999)' a op+o desejada:'
write(*,999)' (F) Formato'
write(*,999)' (V) Vari veis classificat+rias'
write(*,999)' (C) Covari veis'
write(*,999)' (R) Respostas'
write(*,999)' Ou voçe deseja : '

```

```

write(*,999) ' (I) Iniciar novamente o problema'
write(*,999) ' (A) Abandonar a execu#o'
write(*,995) ' -----> Indique sua op#o : '
READ(*,'(A1)') T1
IF((T1.EQ.'F').OR.(T1.EQ.'f')) GOTO 25
IF((T1.EQ.'V').OR.(T1.EQ.'v')) GOTO 36
IF((T1.EQ.'C').OR.(T1.EQ.'c')) GOTO 46
IF((T1.EQ.'R').OR.(T1.EQ.'r')) GOTO 65
IF((T1.EQ.'I').OR.(T1.EQ.'i')) GOTO 10
IF((T1.EQ.'A').OR.(T1.EQ.'a')) GOTO 9999
ENDIF
9999 CONTINUE
STOP
end
SUBROUTINE LE1(NBVC,NBCV,NBY,NOMEVC,NOMEV,NOMEY,INAME,INFMT,
2IOS,NTOTVC,IREG,IPOS,NTAMU,NRS,nbivcd,ivcd,intvcd,NOINT,RC,
2LBEG,LEND,LBY,IRC,lcom,vlfiv,calc)

CHARACTER*8 NOMEVC(NBVC),NOMEV(NBCV),NOMEY(NBY)
CHARACTER*127 INFMT
CHARACTER*15 INAME
CHARACTER*1 T, RC
c$LARGE
CHARACTER*8 REGVC(20),LCLASS(5000),M1CLAS(200),M2CLAS(20),
2 CLASS(2,20),NAM
INTEGER IPOS(20),IVC(20,2),ICLASS(2,20),ICV(40),IY(20),nbivcd,IRC
integer iclasl(5000),iclam1(200),iclam2(20),intvcd(20,2),ivcd
integer calc(40)
REAL REGCV(40),REGY(20),MCV(40),VCV(40),MY(20),VY(20)
REAL LBEG,LEND,LBY,lcom,vlfiv
COMMON NAM(5260)
OPEN(5,FILE=INAME)
OPEN(10,FILE='GR1.OUT',STATUS='unknown')
NBCLASS=1
IREG=0
DO 5 J=1,NBCV
MCV(J)=0
VCV(J)=0
5 ICV(J)=0
DO 6 K=1,NBY
MY(K)=0
VY(K)=0
6 IY(K)=0
WRITE(*,9997)
WRITE(*,9997)
WRITE(*,9997)
WRITE(*,9997)
WRITE(*,999) ' Voc^ deseja fazer esta an lise'
WRITE(*,999) ' por Regress#o de Cumeeira ?'
WRITE(*,995) ' Responda aqui (s/n):'
READ(*,'(A1)') RC
IF((RC.EQ.'n').or.(RC.EQ.'N')) GOTO 10
IRC=2
GOTO 8
7 CONTINUE
WRITE(*,9997)
WRITE(*,999) ' Erro. Verifique se o formato est correto e se'
WRITE(*,999) ' o valor final de lambda , maior que o inicial.'
8 CONTINUE
WRITE(*,9997)
WRITE(*,999) ' Defina, no formato F20.10,'

```

```

WRITE(*,995) ' o valor inicial de lambda:'
READ(*,'(F20.10)',ERR=7) lcom
WRITE(*,9997)
WRITE(*,999) ' Defina, no formato F20.10,'
WRITE(*,995) ' o valor final de lambda:'
READ(*,'(F20.10)',ERR=7) LEND
IF(LEND.LE.LCOM) GOTO 7
LBY=(LEND-LCOM)/100.0
WRITE(*,9997)
WRITE(*,999) ' Defina, no formato F20.10, o valor de FIV a partir'
WRITE(*,995) ' do qual se aplicara lambda:'
READ(*,'(F20.10)',ERR=7) vlfiv
10 CONTINUE
IF(IOS.EQ.1) THEN
READ(5,INFMT,END=300) (REGVC(I),I=1,NBVC),(REGCV(J),J=1,NBCV),
2 (REGY(K),K=1,NBY)
ELSE
IF(IOS.EQ.2) READ(5,INFMT,END=300) (REGVC(I),I=1,NBVC),
2 (REGY(K),K=1,NBY)
IF(IOS.EQ.3) READ(5,INFMT,END=300) (REGCV(J),J=1,NBCV),
2 (REGY(K),K=1,NBY)
ENDIF
IREG=IREG+1
IF(IOS.EQ.3) GOTO 55
IF(IREG.EQ.1) THEN
if(regvc(1).ne.' ') LCLASS(1)=REGVC(1)
if(regvc(2).ne.' '.and.NBVC.GE.2) M1CLAS(1)=REGVC(2)
IF(REGVC(3).NE.' '.AND.NBVC.GE.3) M2CLAS(1)=REGVC(3)
IF(NBVC.GE.4) THEN
DO 12 I=4,NBVC
IF(REGVC(I).NE.' ') THEN
CLASS(I-3,1)=REGVC(I)
ENDIF
12 CONTINUE
ENDIF
DO 13 I=1,NBVC
IF(REGVC(I).NE.' ') THEN
IVC(I,2)=1
ELSE
IVC(I,2)=0
ENDIF
if(i.gt.3) then
IF(REGVC(I).NE.' ') THEN
ICLASS(I-3,1)=1
ELSE
ICLASS(I-3,1)=0
ENDIF
else
if(i.eq.1) iclasl(1)=1
if(i.eq.2) iclam1(1)=1
if(i.eq.3) iclam2(1)=1
IF(REGVC(I).EQ.' ') THEN
IF(I.EQ.1) ICLASL(1)=0
IF(I.EQ.2) ICLAM1(1)=0
IF(I.EQ.3) ICLAM2(1)=0
ENDIF
endif
13 IVC(I,1)=I
ENDIF
IF(IREG.GE.2) THEN
DO 50 I=1,NBVC

```

```

      IF(REGVC(I).EQ.' ') GOTO 49
IFLAG=0
II=IVC(I,2)
DO 20 J=1,II
  IF((I.EQ.1).AND.(LCLASS(J).EQ.REGVC(1))) GOTO 15
  IF((I.EQ.2).AND.(M1CLAS(J).EQ.REGVC(2))) GOTO 15
  IF((I.EQ.3).AND.(M2CLAS(J).EQ.REGVC(3))) GOTO 15
  IF((I.GE.4).AND.(CLASS(I-3,J).EQ.REGVC(I))) GOTO 15
GOTO 17
15 CONTINUE
  if(i.eq.1) iclas1(j)=iclas1(j)+1
  if(i.eq.2) iclam1(j)=iclam1(j)+1
  if(i.eq.3) iclam2(j)=iclam2(j)+1
  if(i.gt.3) then
    ICLASS(I-3,J)=ICLASS(I-3,J)+1
  endif
IFLAG=1
17 CONTINUE
20 CONTINUE
IF(IFLAG.EQ.0) THEN
  IVC(I,2)=IVC(I,2)+1
  II=IVC(I,2)
  if(i.eq.1) iclas1(ii)=1
  if(i.eq.2) iclam1(ii)=1
  if(i.eq.3) iclam2(ii)=1
  if(i.gt.3) iclass(i-3,ii)=1
  IF(I.EQ.1) LCLASS(II)=REGVC(1)
  IF(I.EQ.2) M1CLAS(II)=REGVC(2)
  IF(I.EQ.3) M2CLAS(II)=REGVC(3)
  IF(I.GE.4) CLASS(I-3,II)=REGVC(I)
ENDIF
49 CONTINUE
50 CONTINUE
ENDIF
55 CONTINUE
IF(IOS.EQ.2) GOTO 105
DO 100 J=1,NBCV
  IF(REGCV(J).NE.0.0) ICV(J)=ICV(J)+1
  A=REGCV(J)-MCV(J)
  B=A*A
  VCV(J)=VCV(J)+B-B/IREG
  MCV(J)=MCV(J)+A/IREG
C   IF(REGCV(J).EQ.0.00) WRITE(10,910) IREG,NOMEVC(J)
100 CONTINUE
105 CONTINUE
C 910 FORMAT(1X,' 0 registro ',I8,' apresenta o valor de 0.0 para a cov
C      2ariavel ',A6)
C 920 FORMAT(1X,' 0 registro ',I8,' apresenta o valor de 0.0 para a res
C      2posta ',A6)
DO 200 K=1,NBY
  IF(REGY(K).NE.0.0) IY(K)=IY(K)+1
  A=REGY(K)-MY(K)
  B=A*A
  VY(K)=VY(K)+B-B/IREG
  MY(K)=MY(K)+A/IREG
C   IF(REGY(K).EQ.0.00) WRITE(10,920) IREG,NOMEY(K)
200 CONTINUE
GOTO 10
300 CONTINUE
WRITE(*,999)' Todos os registros j foram lidos.'
901 format(A,\)

```

```

IF(IOS.EQ.3) GOTO 410
IPOS(1)=1
DO 350 I=2,NBVC
350 IPOS(I)=IPOS(I-1)+IVC(I-1,2)
  ivcd=0
  if(nbivcd.eq.0) goto 395
  do 380 i=1,nbivcd
    ii=intvcd(i,1)
    ij=intvcd(i,2)
    ik=ivc(ii,2)
    il=ivc(ij,2)
    ivcd=ivcd+ik*il
380 continue
395 continue
  NTOTVC=IPOS(NBVC)+IVC(NBVC,2)-1
  DO 400 I=1,NBVC
    IBEG=IPOS(I)
    IF(I.EQ.NBVC) THEN
      IEND=NTOTVC
    ELSE
      IEND=IPOS(I+1)-1
    ENDIF
    DO 400 II=IBEG,IEND
      IK=II-IBEG+1
      IF(I.EQ.1) NAM(II)=LCLASS(IK)
      IF(I.EQ.2) NAM(II)=M1CLAS(IK)
      IF(I.EQ.3) NAM(II)=M2CLAS(IK)
      IF(I.GE.4) NAM(II)=CLASS(I-3,IK)
400 CONTINUE
410 CONTINUE
  ILARGE=0
  IF(IOS.NE.3) ILARGE=IVC(1,2)
  IF(IOS.EQ.3) NTOTVC=0
  NPARAM=1+NTOTVC+NBCV+ivcd-noint
  NCOLU=NPARAM+NBY
  NRS=NCOLU
  A=NCOLU/2.0
  B=NCOLU+1
  NTAMU=A*B
  IBYTP=8*NCOLU
  IBYTU=4*NTAMU
997 FORMAT(1X,///)
  IF((T.EQ.'S').OR.(T.EQ.'s')) T='S'
  WRITE(10,998) IREG,' registros foram lidos.'
  WRITE(*,998) IREG,' registros foram lidos.'
  WRITE(10,940) NPARAM,NBY,NCOLU,NTAMU,IBYTU,IBYTP
  WRITE(*,940) NPARAM,NBY,NCOLU,NTAMU,IBYTU,IBYTP
940 FORMAT(1X,' Existem ',I4,' parametros a serem estimados.',/, ' Junt
2o com as ',I2,' respostas, resultam em ',I4,' colunas de U.',/, ' E
3sta matriz U ocupara ',I8,' posicoes ou ',I8,' bytes de RAM.',/,
4      ' Para conter um registro e uma linha de U serao necess
5arios ',I7,' bytes de RAM.')
  WRITE(*,997)
450 CONTINUE
IF(IOS.EQ.3) GOTO 470
WRITE(10,999) ' LISTAGEM DAS VARIAVEIS CLASSIFICATORIAS '
WRITE(10,999) ' _____ '
WRITE(10,999) ' Nome da variavel, Nome das classes e Numero de Obs
2ervacoes por classe.'
DO 460 I=1,NBVC
  WRITE(10,950) NOMEVC(I),NOMEVC(I)

```

```

950 FORMAT(1X,A6,66X,A6)
IBEG=IPOS(I)
IF(I.EQ.NBVC) IEND=NTOTVC
IF(I.NE.NBVC) IEND=IPOS(I+1)-1
DO 460 II=IBEG,IEND
IK=II-IBEG+1
if(i.eq.1) WRITE(10,955) NAM(II),ICLAsL(ik)
if(i.eq.2) WRITE(10,955) NAM(II),ICLAm1(ik)
if(i.eq.3) WRITE(10,955) NAM(II),ICLAm2(ik)
if(i.gt.3) WRITE(10,955) NAM(II),ICLASS(I-3,IK)
460 continue
955 FORMAT(8X,A20,5X,I10)
470 CONTINUE
IF(10S.EQ.2) GOTO 490
WRITE(10,999) ' LISTAGEM DAS COVARIABEIS'
WRITE(10,999) ' _____'
WRITE(10,999) ' Nome da Covariavel          Media          Variancia
2'
write(10,999) ' _____          _____          _____
2'
DO 480 J=1,NBCV
VCV(J)=VCV(J)/(IREG-1)
WRITE(10,960) NOMEVC(J),MCV(J),VCV(J)
IF(ICV(J).NE.IREG) THEN
IM=IREG-ICV(J)
WRITE(10,970) IM
ELSE
WRITE(10,'(1X,/)')
ENDIF
480 CONTINUE
960 FORMAT(6X,A6,3X ,F20.8,4X,F20.8\ )
970 FORMAT(I20,' obs. c/cov=0.0',/)
490 CONTINUE
WRITE(10,999) ' LISTAGEM DAS RESPOSTAS '
WRITE(10,999) ' _____'
WRITE(10,999) ' Nome da Resposta          Media          Variancia
2'
write(10,999) ' _____          _____          _____
2'
DO 500 K=1,NBY
VY(K)=VY(K)/(IREG-1)
WRITE(10,960) NOMEY(K),MY(K),VY(K)
IF(IY(K).NE.IREG) THEN
IM=IREG-IY(K)
WRITE(10,976) IM
ELSE
WRITE(10,'(1X,/)')
ENDIF
500 CONTINUE
976 FORMAT(I20,' obs. c/res=0.0',/)
995 FORMAT(A,\ )
998 FORMAT(I10,A)
999 FORMAT(A)
9997 FORMAT('0')
RETURN
END
C *****
SUBROUTINE LE2(NBVC,NBCV,NBY,NOMEVC,NOMEYC,NOMEY,INAME,INFMT,
2IOS,NTOTVC,IREG,IPOS,NTAMU,NRS,nbivcd,ivcd,intvcd,noint,IRC,
2LBEG,LEND,LBY,LAMB,elem,ICOV,lcom,vlfiv,RC,calc)
CHARACTER*8 NOMEVC(NBVC),NOMEYC(NBCV),NOMEY(NBY)

```

```

CHARACTER*127 INFMT
CHARACTER*10 IMP
CHARACTER*1 AF(5),T1,T2,RC
CHARACTER*15 INAME
c$ LARGE:NAM
CHARACTER*8 NAM,REGVC(20)
real lby,lamb,lend,lbeg,lcom,vlfiv
real lat,g,gmd,l1,l2
real*8 sqx(40),fiv(40),sdsqx(40),pond(40),minsqx,ponder(40)
REAL R(5301),S(5301),REGCV(40),REGY(20),MY(20),SE(5301),MSE(40)
REAL SUMAT,HARMT,SINV,TEE(40),ATEE(40),INVT(40),qme(40),razt(40)
REAL RATIO(40),BPB(40,40),RREG(40,40),E(40,40),rre(40,40),srt
real URAM(14052951)
INTEGER IPOS(NBVC),IU(5301,2),intvcd(20,2),glid(40),glvc(40)
INTEGER ivcd,nbivcd,IRC,DIAG,ICOV,nn,nofiv,desc,calc(40)
integer denovo,ncgr
integer aplv(40)
real ELEM(40)
COMMON NAM(5260)
OPEN(5,FILE=INAME)
OPEN(10,FILE='GR2.OUT',STATUS='unknown')
OPEN(15,FILE='URIDG',STATUS='unknown',FORM='UNFORMATTED')
OPEN(20,FILE='beta.OUT',STATUS='unknown')
open(30,file='elem.out',status='unknown')
open(40,file='pobs.out',status='unknown')
open(50,file='fiv.out',status='unknown')
open(60,file='tes.out',status='unknown')
open(70,file='estt.out',status='unknown')
open(80,file='pred.out',status='unknown')
open(90,file='pond.out',status='unknown')
999 FORMAT(A)
ZERO=0.0000000
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(20,'(A,\)')' lambda'
if(noint.eq.0) write(20,'(A,\)')' interc.'
do 2 i=1,nbcv
write(20,'(14x,A6,\)')nomecv(i)
2 continue
WRITE(20,'(A)')' '
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(50,'(A,\)')' lambda'
do 3 i=1,nbcv
write(50,'(14x,A6,\)')nomecv(i)
3 continue
WRITE(50,'(A)')' '
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(60,'(A,\)')' lambda'
do 4 i=1,nbcv
write(60,'(14x,A6,\)')nomecv(i)
4 continue
WRITE(60,'(A)')' '
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(70,'(A,\)')' lambda'
write(70,'(A)')' sum(abs)t harm(abs)t sum(t/
2fiv)'
write(80,'(A)')' 0.000 0.125 0.250 0.375
2 0.500 0.625 0.750 0.875 1.000'
DO 5 I=1,NTAMU
5 URAM(I)=ZERO
minsqx=0.d00
do 6 j=1,nbcv
sqx(j)=0.d00
sdsqx(j)=0.d00
pond(j)=0.d00
6 fiv(j)=0.d00

```

```

        if(ivcd.ne.0) then
            do 8 i=1,ivcd
2 se(1+ntotvc+i)=zero
        endif
        NP=1+NTOTVC+NBCV+ivcd-noint
        NCU=NP+NBY
        DO 30 I=1,NCU
        IU(I,1)=(NCU-I/2.0)*(I-1)+I
30 IU(I,2)=(NCU-(I+1)/2.0)*I+I
        ITEST=1
        iler=0
50 CONTINUE
        iler=iler+1
        IF(IOS.EQ.1) THEN
            READ(5,INFMT,END=300) (REGVC(I),I=1,NBVC), (REGCV(J),J=1,NBCV),
2 (REGY(K),K=1,NBY)
        ELSE
            IF(IOS.EQ.2) READ(5,INFMT,END=300) (REGVC(I),I=1,NBVC),
2 (REGY(K),K=1,NBY)
            IF(IOS.EQ.3) READ(5,INFMT,END=300) (REGCV(J),J=1,NBCV),
2 (REGY(K),K=1,NBY)
        ENDIF
        NBLU=1
        IF(IOS.NE.3) THEN
            DO 70 I=1,NBVC
            IB=IPOS(I)
            IF(I.LT.NBVC) IE=IPOS(I+1)-1
            IF(I.EQ.NBVC) IE=NTOTVC
            DO 70 K=IB,IE
            IF(REGVC(I).EQ.NAM(K)) THEN
                NBLU=K
                S(K+1)=1.00
            ELSE
                S(K+1)=0.0
            ENDIF
70 CONTINUE
        ENDIF
        if(nbivcd.ne.0) then
            ic=0
            do 78 i=1,nbivcd
            i1=intvcd(i,1)
            i2=intvcd(i,2)
            ib1=ipos(i1)
            iel=ntotvc
            if(i1.lt.nbvc) iel=ipos(i1+1)-1
            ib2=ipos(i2)
            ie2=ntotvc
            if(i2.lt.nbvc) ie2=ipos(i2+1)-1
            do 75 j=ib1,iel
            jj=j-ib1
            do 75 k=ib2,ie2
            kk=k-ib2+1
            kj=ie2-ib2+1
            jk=jj*kj+kk
            ijk=1+ntotvc+jk
            if((s(j+1).eq.1.00).and.(s(k+1).eq.1.00)) then
                s(ijk+ic)=1.00
                se(ijk+ic)=1.00
            endif
75 continue
            ic=ic+jk

```

```

78 continue
        endif
        JJ=NTOTVC+1+ivcd-noint
        if(noint.eq.0) s(1)=1.00
C ^ ESTE 1 PORQUE MU EST+ INCLUIDO NO MODELO *****
        IF(IOS.NE.2) THEN
            DO 80 J=1,NBCV
            sqx(j)=sqx(j)+regcv(j)*regcv(j)
80 S(JJ+J)=REGCV(J)
        ENDIF
        DO 90 K=1,NBY
80 S(NP+K)=REGY(K)
            call opera(s,1,ncu,iu,r,URAM)
            itest=mod(iler,1000)
            IF(ITEST.EQ.0) WRITE(*,900) ILER
            GOTO 50
C VOLTA A LER NOVO REGISTRO *****
300 WRITE(15)URAM
            IF((IOS.NE.2).and.(IRC.EQ.2))THEN
C write(*,'(a,f20.10)')' lbeg eh:',lbeg
C write(*,'(a,f20.10)')' lby eh:',lby
                ilam=1
                denovo=1
            DO 302 J=1,NBCV
302 aply(J)=0
                GOTO 310
            else
                GOTO 390
            endif
305 CONTINUE
            rewind 15
            READ(15)URAM
            ilam=ilam+1
            if(ilam.eq.2)then
                lbeg=lcom
            else
                if(ilam.gt.2)then
                    LBEG=LBEG+LBY
                endif
            endif
            IF(LBEG.GT.LEND) GOTO 8100
            WRITE(30,997)
            WRITE(80,997)
            write(80,9000)lbeg
            write(80,'(A)')' '
            write(80,'(A)')' ' 0.000 0.125 0.250 0.375
2 0.500 0.625 0.750 0.875 1.000'
            WRITE(40,997)
310 CONTINUE
            LAMB=0.0
            if(ilam.eq.1)then
                lbeg=0.0
            endif
            LAMB=SQRT(LBEG)
            ILER=1
            GOTO 320
315 CONTINUE
            ILER=ILER+1
            IF(ILER.GT.NBCV) GOTO 350
            CONTINUE
320 DO 321 J=1,NBCV

```



```

321 ELEM(J)=0.0
    DIAG=ILER
    if (aply(diag).eq.1) then
      ELEM(DIAG)=LAMB*ponder(diag)*pond(diag)
    endif
    DO 322 N=1,NBCV
322 write(30,9102)elem(N)
    write(30,'(a)') ' '
    do 323 i=1,ncu
323 S(i)=0.0
    LL=NTOTVC+1+ivcd-noint
    DO 325 n=1,NBCV
325 S(LL+n)=ELEM(n)
    do 331 j=1,ncu
331 write(40,9102)S(j)
    write(40,'(a)') ' '
    CALL OPERA(S,1,ncu,IU,R,URAM)
    GOTO 315
350 CONTINUE
390 CONTINUE
800 FORMAT(I4,F10.3)
C   RESTRIÇÕES, MM *****
    WRITE(*,997)
899 FORMAT(A1)
995 FORMAT(A,\)
    IF(NBVC.NE.0) THEN
      WRITE(*,999) ' O modelo proposto , Fixo (F) ou Mixto (M)?'
      write(*,995) ' Indique (F/M):'
      READ(*,899) T1
      IF(T1.EQ.'f') T1='F'
      IF(T1.EQ.'m') T1='M'
      IF(T1.EQ.'F') THEN
        WRITE(*,*) ' Como o modelo , fixo, ser necess rio impor restrições
2'
        write(*,999) ' sobre as soluções para que o problema se torne ident
2ific vel.'
        write(*,999) ' As restrições ser, o do tipo:'
        write(*,999) '
        WRITE(*,999) '          SOMAT•RIO [ EFEITOS ] = 0.0 '
        WRITE(*,997)
        DO 500 I=1,NBVC
        DO 405 J=1,NRS
405 S(J)=0.00
        KBEG=IPOS(I)+1
        IF(I.LT.NBVC) KEND=IPOS(I+1)
        IF(I.EQ.NBVC) KEND=NTOTVC+1
        glvc(i)=kend-kbeg
        DO 442 J=KBEG,KEND
442 S(J)=1.00
        call opera(s,kbeg,ncu,iu,r,URAM)
500 CONTINUE
        if(nbivcd.ne.0) then
          intbeg=0
          do 660 i=1,nbivcd
            do 601 j=1,ncu
601 s(j)=0.00
          il=intvcd(i,1)
          i2=intvcd(i,2)
          ib1=ipos(il)+1
          iel=ntotvc+1
          if(il.lt.nbvc) iel=ipos(il+1)

```

```

          ib2=ipos(i2)+1
          ie2=ntotvc+1
          if(i2.lt.nbvc) ie2=ipos(i2+1)
          nbeg=1+ntotvc+intbeg
          ic1=iel-ib1+1
          ic2=ie2-ib2+1
          intbeg=intbeg+ic1*ic2
          do 620 j=ib1,iel
            kbeg=nbeg+(j-ib1)*ic2+1
            do 610 k=ib2,ie2
610 s(kbeg+k-ib2)=1.00
            call opera(s,kbeg,ncu,iu,r,uram)
620 continue
            do 640 k=ib2,ie2
              do 630 j=ib1,iel
                jbeg=nbeg+(j-ib1)*ic2+1+k-ib2
                if(j.eq.ib1) kbeg=jbeg
630 s(jbeg)=1.00
            call opera(s,kbeg,ncu,iu,r,uram)
640 continue
            nbeg=nbeg+1
            nend=nbeg+ic1*ic2-1
            ljk=0
            do 650 l=nbeg,nend
              IF(SE(L).EQ.0.00) THEN
                DO 645 J=L,NCU
645 S(J)=0.00
                S(L)=1.00
                CALL OPERA(S,L,NCU,IU,R,uram)
                ENDIF
650 ljk=ljk+se(l)
                glid(i)=ljk-ic1-ic2+1
660 continue
            endif
            ELSE
C   ***          MODELO          MIXTO          *****
          IFLAG=0
          WRITE(*,997)
          write(*,999) ' Como o modelo que voce est utilizando , mixto, se
2r necess rio informar quais'
          write(*,999) ' vari veis classificatórias s, o aleatórias (A) e qua
2is s, o fixas (F).'
          write(*,999) ' Indique a seguir, com um A ou F : '
          write(*,999) ' VARIÁVEL : A/F '
          WRITE(*,999) ' _____ '
          DO 505 I=1,NBVC
501 CONTINUE
960 FORMAT(2X,A6,9X,\)
          WRITE(*,960) NOMEVC(I)
          READ(*,899) AF(I)
          IF(AF(I).EQ.'A') GOTO 502
          IF(AF(I).EQ.'F') THEN
            GOTO 502
          ELSE
            GOTO 501
          ENDIF
502 CONTINUE
          IF(AF(I).NE.'A') IFLAG=1
505 CONTINUE
          WRITE(*,997)
          IF(IFLAG.EQ.1) THEN

```

```

WRITE(*,999) ' Como pelo menos uma variavel , fixa, ser necessario
2 impor restrições sobre as'
write(*,999) ' soluções, para que o problema se torne identificavel
2.'
WRITE(*,999) ' As restrições ser, o do tipo:'
write(*,999) ' '
write(*,999) '          SOMAT•RIO [ EFEITOS ] = 0.0 '
WRITE(*,997)
                ENDIF
DO 600 I=1,NBVC
IF(AF(I).EQ.'F') THEN
DO 541 J=1,NRS
541 S(J)=0.00
KBEG=IPOS(I)+1
IF(I.LT.NBVC) KEND=IPOS(I+1)
IF(I.EQ.NBVC) KEND=NTOTVC+1
glvc(i)=kend-kbeg
DO 542 J=KBEG,KEND
542 S(J)=1.00
call opera(s,kbeg,ncu,iu,r,uram)
ENDIF
C ^ FECHA IF VC(I)   FIXA.   *** *** ***   *****
600 CONTINUE
C AGORA COMEÇA SE VC   ALEATORIA   *****
DO 2050 I=1,NBVC
IF(AF(I).EQ.'A') THEN
2001 CONTINUE
WRITE(*,997)
WRITE(*,977) ' Voce informou que a variavel ',nomevc(i),
2 ' ', aleatoria.'
977 format(a,a6,a)
write(*,999) ' Ser necessario que voce informe, com formato F10.0,
2, baseado em informaç, o'
write(*,976) ' a priori, qual a relação, o entre V(E)/V(' ,NOMEVC(I),
3 ' ):'
976 FORMAT(A,A6,A\ )
READ(*,'(F10.0)',ERR=2001) RATIO(I)
WRITE(10,997)
WRITE(10,*) NOMEVC(I)
WRITE(10,'(A,F10.5)') ' Relação informada foi :', ratio(I)
RATIO(I)=SQRT(RATIO(I))
KBEG=IPOS(I)+1
IF(I.LT.NBVC) KEND=IPOS(I+1)
IF(I.EQ.NBVC) KEND=NTOTVC+1
glvc(i)=kend-kbeg
DO 2049 M=KBEG,KEND
DO 2005 J=1,NRS
2005 S(J)=0.00
S(M)=RATIO(I)
call opera(s,m,ncu,iu,r,uram)
2049 CONTINUE
ENDIF
2050 CONTINUE
ENDIF
C ^ FECHA MODELO MIXTO ?   *****
ENDIF
C ^ FECHA NBVC > 0 ?   *****
C NEXT STEP          BACKSOLVING
C R(.) conter as respostas
C S(.) ser utilizado para ler cada linha de S
DO 1500 K=1,NBY

```

```

LK=NP+K
DO 1100 I=1,NP
L=NP-I+1
IBEG=IU(L,1)
IEND=IU(L,2)
DO 1010 J=IBEG,IEND
M=L+J-IBEG
1010 s(m)=uram(j)
IF(L.EQ.NP) THEN
R(L)=S(LK)/S(L)
ELSE
RHS=0.00
LL=L+1
DO 1020 J=LL,NP
1020 RHS=RHS+R(J)*S(J)
RHS=S(LK)-RHS
R(L)=RHS/S(L)
ENDIF
1100 CONTINUE
WRITE(10,997)
WRITE(10,997)
write(10,999) ' ***** '
write(10,'(a,f8.4)') ' RESULTADOS PARA LAMBDA =',LBEG
write(10,999) ' ***** '
WRITE(10,997)
WRITE(10,'(A,A6)') ' TABELA DA ANALISE DA VARIANCIA DE ',NOMEY(K)
WRITE(10,'(A)') ' _____
2 _____
WRITE(10,999) ' FONTE DE VARIACAO   GL   SQ SEQUENCIAIS   QUADRA
2DO MEDIO   F   R**2 '
WRITE(10,999) ' _____
2 _____
IF(NBVC.NE.0) THEN
DO 1102 I=1,NBVC
S(I)=0.00
I1=IPOS(I)+1
IF(I.EQ.NBVC) I2=NTOTVC+1
IF(I.LT.NBVC) I2=IPOS(I+1)
DO 1101 J=I1,I2
JJ=IU(J,2)-NBY+K
S(I)=S(I)+uram(jj)*uram(jj)
1101 CONTINUE
1102 CONTINUE
if(nbivcd.ne.0) then
intbeg=0
npl=ntotvc+1
do 1118 i=1,nbivcd
j=np1+i
s(j)=0.00
i1=intvcd(i,1)
i2=intvcd(i,2)
i1l=ipos(i1)+1
i1e=npl
if(i1.lt.nbvc) i1e=ipos(i1+1)
i2l=ipos(i2)+1
i2e=npl
if(i2.lt.nbvc) i2e=ipos(i2+1)
nbeg=np1+intbeg+1
ic1=i1e-i1l+1
ic2=i2e-i2l+1
intbeg=ic1*ic2+intbeg

```

```

nend=nbeg+ic1*ic2-1
do 1110 m=nbeg,nend
kk=iu(m,2)-nby+k
1110 s(j)=s(j)+uram(kk)*uram(kk)
1118 continue
endif
ENDIF
NP1=NP+1
SSS=0.00
DO 1125 LL=NP1,LK
LLL=IU(LL,1)+LK-LL
1125 SSS=SSS+uram(LLL)*uram(LLL)
IDFF=IREG-1+noint
if(nbvc.ne.0) then
ij=0
do 1126 i=1,nbvc
1126 ij=ij+glvc(i)
idff=idff-ij
endif
if(nbivcd.ne.0) then
ij=0
do 1127 i=1,nbivcd
1127 ij=ij+glid(i)
idff=idff-ij
endif
if(nbvc.ne.0) idff=idff-nbvc
IF(T1.EQ.'M') THEN
DO 1130 I=1,NBVC
IF(AF(I).EQ.'A') THEN
I1=IPOS(I)
IF(I.EQ.NBVC) THEN
I2=NTOTVC
ELSE
I2=IPOS(I+1)-1
ENDIF
I3=I2-I1
IDFF=IDFF+I3
DO 1128 J=I1,I2
A=R(J+1)
B=A*A
1128 BPB(I,K)=BPB(I,K)+B
ENDIF
1130 CONTINUE
ENDIF
MSE(K)=SSS/IDFF
if(lbeg.eq.0.0)then
qme(k)=mse(k)
endif
AII=0.00
IF(NBVC.NE.0) THEN
DO 1140 J=1,NBVC
AII=AII+S(J)
QM=S(J)/GLVC(J)
F=QM/MSE(K)
WRITE(10,661) NOMEVC(J),GLVC(J),S(J),QM,F
661 format(1x,a6,T18,I5,2F20.8,F10.3)
1140 CONTINUE
ENDIF
IF(nbivcd.NE.0) THEN
IC=0
DO 1160 J=1,nbivcd
JJ=NTOTVC+1+J
AII=AII+S(JJ)
I1=INTVCD(J,1)
I2=INTVCD(J,2)
QM=S(JJ)/GLID(J)
F=QM/MSE(K)
WRITE(10,967) NOMEVC(I1),'*',NOMEVC(I2),GLID(J),S(JJ),QM,F
967 FORMAT(1X,A6,A1,A6,T18,I5,2F20.8,F10.3)
1160 CONTINUE
endif
IF(NBCV.NE.0) THEN
NP1=NTOTVC+1-noint
DO 1180 J=1,NBCV
NN=NP1+J+IVCD
RREG(J,K)=R(NN)
JJ=IU(NN,2)-NBY+K
S(NN)=uram(jj)*uram(jj)
F=S(NN)/MSE(K)
AII=AII+S(NN)
WRITE(10,961) NOMEVC(J),' 1',S(NN),S(NN),F
961 FORMAT(1X,A6,T18,A5,2F20.8,F10.3)
1180 CONTINUE
ENDIF
AIJ=AII/(AII+SSS)
WRITE(10,964) IDFF,SSS,MSE(K),AIJ
964 FORMAT(1X,'RESIDUO',T18,I5,2F20.8,T74,F7.4)
WRITE(10,999)'
2
WRITE(10,999)' OBS.:SE OS DADOS NAO FOREM BALANCEADOS, AS HIPOTESE
2S DE NULIDADE TESTADAS'
WRITE(10,999)' ACIMA PODERAO CONTER FORMAS QUADRATICAS DE OUT
2ROS EFEITOS.'
WRITE(10,999)' CONSIDERE O USO DAS SQ PARCIAIS.'
WRITE(10,997)
WRITE(10,940) NOMEY(K)
940 FORMAT(1X,'SOLUCOES PARA ',A6,/,42('_',),/, ' VARIAVEL',T25,
2'ESTIMATIVA',/,20('_',),2X,20('_',))
if(noint.eq.0) WRITE(10,941) R(1)
941 FORMAT(/,' CONSTANTE',T22,F20.8)
IF(NBVC.NE.0) THEN
DO 1200 I=1,NBVC
WRITE(10,942) NOMEVC(I)
IBEG=IPOS(I)
IEND=NTOTVC
IF(I.LT.NBVC) IEND=IPOS(I+1)-1
DO 1190 J=IBEG,IEND
WRITE(10,943) NAM(J),R(J+1)
1190 CONTINUE
943 FORMAT(1X,A8,T22,F20.8)
942 FORMAT(1X,/,1X,A6)
1200 CONTINUE
ENDIF
IF(NBIVCD.NE.0) THEN
IC=0
DO 1220 J=1,NBIVCD
J1=INTVCD(J,1)
J2=INTVCD(J,2)
WRITE(10,944) NOMEVC(J1),'*',NOMEVC(J2)
944 FORMAT(1X,/,1X,A6,A1,A6)
IB1=IPOS(J1)
IE1=NTOTVC

```

```

IF(J1.LT.NBVC) IE1=IPOS(J1+1)-1
IB2=IPOS(J2)
IE2=NTOTVC
IF(J2.LT.NBVC) IE2=IPOS(J2+1)-1
DO 1210 JJ=IB1,IE1
DO 1210 JK=IB2,IE2
IJK=1+NTOTVC+(JJ-IB1)*(IE2-IB2+1)+JK+1-IB2+IC
WRITE(10,945) NAM(JJ),'*',NAM(JK),R(IJK)
1210 continue
IC=IC+(IE1-IB1+1)*(IE2-IB2+1)
945 FORMAT(1X,A8,A1,A8,T22,F20.8)
1220 CONTINUE
ENDIF
IF(NBCV.NE.0) THEN
NP1=1+NTOTVC+IVCD-noint
DO 1230 J=1,NBCV
NN=NP1+J
WRITE(10,946) NOMEVC(J),R(NN)
1230 continue
946 FORMAT(1X,/,1X,A6,T22,F20.8)
ENDIF
WRITE(10,947)
947 FORMAT(1X,42('_'))
WRITE(10,999)' OBS.:ESTIMATIVAS DOS REGRESSORES SAO BLUE PARA OS P
2ARAMETROS.'
WRITE(10,999)' SOLUCOES PARA AS VARIAVEIS CLASSIFICATORIAS NA
20 SAO BLUE PARA OS PARAMETROS,'
WRITE(10,999)' MAS SAO BLUE PARA ALGUMA COMBINACAO LINEAR DES
2TES (E.G.:CONSTRASTES).'
WRITE(10,997)
C ERROS PADROES *****
IF((RC.EQ.'s').or.(RC.EQ.'S')) WRITE(20,9000)LBEG
ncgr=1+nbcv-noint
DO 1232 J=1,ncgr
WRITE(20,9101)R(J)
1232 CONTINUE
9000 FORMAT(f10.5,\)
9100 FORMAT(f20.5,\)
9101 FORMAT(E20.8,\)
9102 FORMAT(f15.5,\)
WRITE(20,'(A)')' '
1500 CONTINUE
DO 1502 I=1,NP
II=IU(I,1)
uram(ii)=1.0/uram(ii)
SE(I)=uram(ii)*uram(ii)
1502 CONTINUE
NPM1=NP-1
DO 1510 I=1,NPM1
N1L1=I+1
DO 1508 J=N1L1,NP
SSS=0.00
JM1=J-1
DO 1503 L=I,JM1
AIL=uram(i11)
ILJ=IU(L,1)+J-L
ALJ=uram(ilj)
SSS=SSS+AIL*ALJ
1503 CONTINUE
IJJ=IU(J,1)
AJJ=uram(ijj)
IIJ=IU(I,1)+J-I
AIJ=-SSS*AJJ
uram(iij)=aij
SE(I)=SE(I)+AIJ*AIJ
1508 CONTINUE
1510 CONTINUE
C *****
DO 1550 K=1,NBY
IF((NBVC.GT.0).AND.(T1.EQ.'M')) THEN
WRITE(10,'(A20,A6)')' Para a resposta : ',nomey(k)
DO 1530 I=1,NBVC
IF(AF(I).EQ.'A') THEN
TRACO=0.00
I1=IPOS(I)
IF(I.EQ.NBVC) THEN
I2=NTOTVC
ELSE
I2=IPOS(I+1)-1
ENDIF
I3=I2-I1+1
DO 1520 J=I1,I2
TRACO=TRACO+SE(J+1)
BPB(I,K)=BPB(I,K)/(I3-(RATIO(I)*RATIO(I)*TRACO))
ratio(i)=mse(k)/bpb(i,k)
WRITE(10,'(A10,F18.8)')' TRACO = ', TRACO
WRITE(10,'(A12,A5,A,F18.8)')' Varifncia('',nomevc(i),')='',BPB(I,K)
write(10,'(a,f18.8)')' Nova relacão , = ',ratio(i)
endif
1530 continue
endif
WRITE(10,997)
WRITE(10,948) NOMEY(K)
948 FORMAT(1X,'ERROS PADROES PARA ',A6,/,42('_'),/, ' VARIAVEL',
2T25,'ERRO PADRAO',/,20('_'),2X,20('_'))
SEP=SQRT(SE(1)*MSE(K))
IF(NOINT.EQ.0) WRITE(10,941) SEP
IF(NBVC.NE.0) THEN
DO 1240 I=1,NBVC
WRITE(10,942) NOMEVC(I)
IBEG=IPOS(I)
IEND=NTOTVC
IF(I.LT.NBVC) IEND=IPOS(I+1)-1
DO 1235 J=IBEG,IEND
SEP=SQRT(SE(J+1)*MSE(K))
WRITE(10,943) NAM(J),SEP
1235 CONTINUE
1240 CONTINUE
ENDIF
IF(NBIVCD.NE.0) THEN
IC=0
DO 1260 J=1,NBIVCD
J1=INTVCD(J,1)
J2=INTVCD(J,2)
WRITE(10,944) NOMEVC(J1),'*',NOMEVC(J2)
IB1=IPOS(J1)
IE1=NTOTVC
IF(J1.LT.NBVC) IE1=IPOS(J1+1)-1
IB2=IPOS(J2)
IE2=NTOTVC
IF(J2.LT.NBVC) IE2=IPOS(J2+1)-1

```

```

DO 1250 JJ=IB1,IE1
DO 1250 JK=IB2,IE2
IJK=1+NTOTVC+(JJ-IB1)*(IE2-IB2+1)+JK+1-IB2+IC
SEP=SQRT(SE(IJK)*MSE(K))
WRITE(10,945) NAM(JJ),'*',NAM(JK),SEP
1250 continue
IC=IC+(IE1-IB1+1)*(IE2-IB2+1)
1260 CONTINUE
ENDIF
IF(NBCV.NE.0) THEN
NP1=1+NTOTVC+IVCD-noint
sumat=0.d00
srt=0.d00
sinv=0.d00
nofiv=0
DO 1270 J=1,NBCV
NN=NP1+J
fiv(j)=sqx(j)*se(nn)
if((fiv(j).gt.vlfiv).and.(lbeg.eq.0.0))then
aply(j)=1
endif
if(lbeg.eq.0.0)then
ponder(j)=dsqrt(fiv(j))
endif
SEP=SQRT(SE(NN)*qme(K))
tee(j)=R(NN)/SEP
atee(j)=abs(tee(j))
invt(j)=1/atee(j)
if(aply(j).eq.0.or.atee(j).lt.0.0000000000000001)then
atee(j)=0.0
endif
razt(j)=atee(j)/fiv(j)
srt=srt+razt(j)
sumat=sumat+atee(j)
if(aply(j).eq.0.or.atee(j).lt.0.0000000000000001)then
invt(j)=0.0
nofiv=nofiv+1
endif
sinv=sinv+invt(j)
WRITE(10,946) NOMEVCV(J),SEP
1270 continue
desc=nbvcv-nofiv
rnbcv=REAL(desc)
harmt=rnbcv/sinv
ENDIF
WRITE(10,947)
1550 CONTINUE
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(50,9000)lbeg
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(60,9000)lbeg
IF((RC.EQ.'s').or.(RC.EQ.'S')) write(70,9000)lbeg
do 1555 j=1,nbcv
write(50,9100)fiv(j)
write(60,9100)tee(j)
1555 continue
write(50,'(a)')' '
write(60,'(a)')' '
do 1556 k=1,nby
write(70,'(4X,f12.5,T20,f10.5,T40,f15.10)')sumat,harmt,srt
*****
*INVERSA*
1556 continue

do 1558 lat = 16,31,3
write(80,9000)lat
ll=lat
l2=lat*lat
do 1557 g = 0,1,0.125
gmd=0.61+
2R(10)*g+
2R(11)*g+
2R(12)*(1-g*g-(1-g)*(1-g))+
2R(13)*(1-g*g-(1-g)*(1-g))+
2R(14)*g*(1-g)+
2R(15)*g*(1-g)+
2R(16)*(1-g*g-(1-g)*(1-g))+
2R(17)*(1-g*g-(1-g)*(1-g))+
2R(18)*g*ll+
2R(19)*g*ll+
2R(20)*g*(1-g)*ll+
2R(21)*g*(1-g)*ll+
2R(22)*(1-g*g-(1-g)*(1-g))*ll+
2R(23)*(1-g*g-(1-g)*(1-g))*ll+
2R(24)*(1-g*g-(1-g)*(1-g))*ll+
2R(25)*(1-g*g-(1-g)*(1-g))*ll+
2R(26)*g*ll+
2R(27)*g*ll+
2R(28)*g*(1-g)*ll+
2R(29)*g*(1-g)*ll+
2R(30)*(1-g*g-(1-g)*(1-g))*ll+
2R(31)*(1-g*g-(1-g)*(1-g))*ll+
2R(32)*(1-g*g-(1-g)*(1-g))*ll+
2R(33)*(1-g*g-(1-g)*(1-g))*ll
write(80,9000)gmd
1557 continue
write(80,'(A)')' '
1558 continue
if(ios.ne.2.and.irc.eq.2.and.lbeg.eq.0.0)then
minsqx=9999999999999999.999999999
do 1559 i=1,nbcv
if(sqx(i).lt.minsqx.and.sqx(i).ne.0.d00.and.aply(i).eq.1)then
minsqx=sqx(i)
endif
write(90,9901) i,sqx(i),minsqx
9901 format(4X,i2,4X,f30.5,4X,f30.5)
1559 continue
write(90,997)
do 1560 j=1,nbcv
sdsqx(j)=sqx(j)/minsqx
pond(j)=dsqrt(sdsqx(j))
write(90,9901) j,sdsqx(j),pond(j)
1560 continue
endif
C WRITE(*,997)
C WRITE(*,999) ' O c lculo da matriz de VARCOV (*1/QME) requer'
C WRITE(*,999) ' '
C WRITE(*,999) ' [(NP**3)/6 + O(NP**2)] opera#es.'
C write(*,999) ' '
C WRITE(*,995) ' Voce deseja este c lculo e impress#o (S/N) ? : '
C READ(*,899) T2
C IF(T2.EQ.'s') T2='S'
C IF(T2.EQ.'S') THEN
C WRITE(10,'(/)')
C WRITE(10,999) ' MATRIZ DE VARIANCIAS E COVARIANCIAS [*1/QME] '

```

```

WRITE(10,999) '
WRITE(10,999) ' LINHA      COLUNA      VARCOV '
WRITE(10,999) '
DO 1600 I=1,NP
DO 1600 J=I,NP
IF(I.EQ.J) THEN
                AIJ=SE(I)
                GOTO 1590
            ENDIF

SS=0.00
DO 1580 L=J,NP
    IIL=IU(I,1)+L-I
        IJL=IU(J,1)+L-J
SS=SS+uram(iil)*uram(ijl)
1580 CONTINUE
AIJ=SS
1590 CONTINUE
WRITE(10,933) I,J,AIJ
c  if ((i.gt.jj).and.(i.eq.j)) fiv(i-jj)=sqx(i-jj)*aij

1600 CONTINUE
C  ENDIF
C  WRITE(*,*)' Calculando SQP PARCIAIS...'
C  E E E E E E E E E E E E E E E E E E E E E E E E E E E E E E
    IF(NBY.GT.1) THEN
        WRITE(10,'(/,1X,A)')
        2'Matriz E (Soma de quadrados e produtos dos residuos)'
        write(10,'(A)')
        2'
            DO 7350 K=1,NBY
            DO 7350 J=1,NBY
                RRE(K,J)=ZERO
7350 BPB(K,J)=ZERO
            DO 7400 K=1,NBY
            DO 7400 J=K,NBY
                NL=NP+K
                LLL=IU(NL,1)+J-K
7400 RRE(K,J)=uram(111)
            DO 7450 K=1,NBY
            DO 7450 J=1,NBY
                IF(K.GT.J) THEN
                    BPB(K,J)=BPB(J,K)
                    GOTO 7430
                ELSE
                    DO 7420 L=1,K
7420 BPB(K,J)=BPB(K,J)+RRE(L,K)*RRE(L,J)
                ENDIF
                WRITE(10,'(1X,A8,1X,A8,1X,F20.10)') NOMEY(K),NOMEY(J),BPB(K,J)
7430 CONTINUE
                E(K,J)=BPB(K,J)
7450 CONTINUE
                CALL INVLUG(E,NBY)
                WRITE(10,'(/,1X,A)') ' MATRIZ E**-1 (E-INVERSA) '
                DO 7460 K=1,NBY
                DO 7460 J=K,NBY
7460 WRITE(10,'(1X,A8,1X,A8,1X,F20.10)') NOMEY(K),NOMEY(J),E(K,J)
                WRITE(10,997)
                ENDIF
C SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4
C SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4 SS4
    IF(NBVC.NE.0) THEN

```

```

write(10,*)' SQP PARCIAIS PARA VARIABEIS CLASSIFICATORIAS'
write(10,*)'
DO 7200 I=1,NBVC
DO 7002 K=1,NBY
DO 7002 J=1,NBY
7002 BPB(K,J)=ZERO
    I1=IPOS(I)+1
    IF(I.LT.NBVC) THEN
        I2=IPOS(I+1)
        ELSE
        I2=NTOTVC+1
        ENDIF
    I3=I2-I1+1
    I4=I3+1
    I5=I3+NBY
    NTAMR=I5*(I5+1)/2.0
    OPEN(9,FILE='R',STATUS='unknown',ACCESS='DIRECT',RECL=4)
    DO 7005 J=1,NTAMR
7005 WRITE(9,REC=J) ZERO
    DO 7006 J=1,I5
        R(J)=ZERO
7006 S(J)=ZERO
    DO 7008 J=I4,I5
        KU=IU(1,2)-NBY+1+J-I4
        R(j)=uram(ku)
7008 WRITE(9,REC=J) R(J)
    DO 7050 J=2,NP
    DO 7010 K=I1,I2
        KK=K-I1+1
        IF(j.GE.k) THEN
            IK=IU(K,1)+J-K
                S(KK)=uram(ik)
            ELSE
                S(KK)=ZERO
            ENDIF
7010 CONTINUE
    DO 7012 L=I4,I5
        JL=IU(J,2)+L-I4-NBY+1
7012 S(L)=uram(jl)
    DO 7030 K=1,I5
        IF(S(K).EQ.ZERO) GOTO 7030
        KBEG=(I5-K/2.0)*(K-1.0)+K
        KEND=(I5-(K+1.0)/2.0)*K+K
        DO 7014 L=KBEG,KEND
            LK=L-KBEG+K
7014 READ(9,REC=L) R(LK)
        IF(R(K).EQ.ZERO) THEN
            ISS=1
            IF(S(K).LT.ZERO) ISS=-1
            DO 7016 M=K,I5
                SWAP=-1*R(M)
                R(M)=ISS*S(M)
                S(M)=SWAP
            ELSE
                CALL G1(R(K),S(K),COS,SIN,R(K))
                S(K)=ZERO
                KM1=K+1
                IF(KM1.GT.I5) GOTO 7019
                DO 7018 M=KM1,I5
                CALL G2(COS,SIN,R(M),S(M))
7018 CONTINUE
7019

```

```

                                ENDIF
                                DO 7020 M=KBEG,KEND
                                MK=M-KBEG+K
7020 WRITE(9,REC=M) R(MK)
7030 CONTINUE
7050 CONTINUE
                                WRITE(10, '(A8)') NOMEVC(I)
                                DO 7080 M=1,I3
                                MEND=(I5-(M+1)/2.0)*M+M
                                DO 7070 K=1,NBY
                                MK=MEND-NBY+K
7070 READ(9,REC=MK) R(K)
                                DO 7060 J=1,NBY
                                DO 7060 K=J,NBY
7060 BPB(J,K)=BPB(J,K)+R(K)*R(J)
7080 CONTINUE
                                DO 7090 J=1,NBY
                                DO 7090 K=J,NBY
                                WRITE(10,7561) NOMEY(J),NOMEY(K),BPB(J,K)
7090 continue
                                CLOSE(9,STATUS='DELETE')
                                IF(NBY.GT.1) THEN
                                DO 7095 J=1,NBY
                                DO 7095 K=1,J
                                    IF(K.LT.J) BPB(J,K)=BPB(K,J)
7095 CONTINUE
                                CALL MABB(E,BPB,NBY)
                                TR=0.00
                                DO 7100 J=1,NBY
7100 TR=TR+BPB(J,J)
                                WRITE(10,'(/,1X,A,A6,/,A,F20.10)')
                                2' VALOR DO TRACO DE HOTELLING-LAWLEY PARA TESTAR A HIPOTESE DE NUL
                                3IDADE DOS EFEITOS CONJUNTOS DE ',NOMEVC(I),' SOBRE AS RESPOSTAS: '
                                4,TR
                                JQ=GLVC(I)
                                JS=MIN(JQ,NBY)
                                JM=ABS(NBY-JQ)-1
                                GN=0.5*(IDFF-NBY-1.0)
                                GL1=JS*(JM+JS+1.0)
                                GL2=2.0*(JS*GN+1.0)
                                FF=GL2*TR/(JS*GL1)
                                IGL1=INT(GL1)
                                IGL2=INT(GL2)
                                WRITE(10,7199) 'APROXIMACAO A DISTRIBUICAO DE F COM ',IGL1,
                                2' E ',IGL2,' G.L. : ',FF
7199 FORMAT(1X,A,I4,A3,I4,A,F8.2)
                                WRITE(10,997)
                                ENDIF
7200 CONTINUE
                                ENDIF
C*****
                                IF(NBIVCD.NE.0) THEN
                                IC=0
                                DO 7700 I=1,NBIVCD
                                DO 7501 K=1,NBY
                                DO 7501 J=1,NBY
7501 BPB(K,J)=ZERO
                                I1=INTVCD(I,1)
                                I2=INTVCD(I,2)
                                WRITE(10,7660) NOMEVC(I1),NOMEVC(I2)
7660 FORMAT(1X,A6,'*',A6)
                                IB1=IPOS(I1)+1
                                IE1=NTOTVC+1
                                IF(I1.LT.NBVC) IE1=IPOS(I1+1)
                                IB2=IPOS(I2)+1
                                IE2=NTOTVC+1
                                IF(I2.LT.NBVC) IE2=IPOS(I2+1)
                                I3=(IE1-IB1+1)*(IE2-IB2+1)
                                II1=IC+NTOTVC+2
                                IC=IC+I3
                                II2=IC+NTOTVC+1
                                I4=I3+1
                                I5=I3+NBY
                                NTAMR=I5*(I5+1)/2.0
                                OPEN(9,FILE='R',STATUS='unknown',ACCESS='DIRECT',RECL=4)
                                DO 7505 J=1,NTAMR
7505 WRITE(9,REC=J) ZERO
                                DO 7506 J=1,I5
                                R(J)=ZERO
7506 S(J)=ZERO
                                DO 7508 J=I4,I5
                                KU=IU(1,2)-NBY+1+J-I4
                                r(j)=uram(ku)
7508 WRITE(9,REC=J) R(J)
                                DO 7550 J=2,NP
                                DO 7510 K=II1,II2
                                KK=K-II1+1
                                IF(J.GE.K) THEN
                                IK=IU(K,1)+J-K
                                s(kk)=uram(ik)
                                    ELSE
                                    S(KK)=ZERO
                                ENDIF
7510 CONTINUE
                                DO 7512 L=I4,I5
                                JL=IU(J,2)+L-I4-NBY+1
7512 s(l)=uram(jl)
                                DO 7530 K=1,I5
                                IF(S(K).EQ.ZERO) GOTO 7530
                                KBEG=(I5-K/2.0)*(K-1.0)+K
                                KEND=(I5-(K+1.0)/2.0)*K+K
                                DO 7514 L=KBEG,KEND
                                LK=L-KBEG+K
7514 READ(9,REC=L) R(LK)
                                IF(R(K).EQ.ZERO) THEN
                                ISS=1
                                IF(S(K).LT.ZERO) ISS=-1
                                DO 7516 M=K,I5
                                SWAP=-1*R(M)
                                R(M)=ISS*S(M)
                                S(M)=SWAP
7516 ELSE
                                CALL G1(R(K),S(K),COS,SIN,R(K))
                                S(K)=ZERO
                                KM1=K+1
                                IF(KM1.GT.I5) GOTO 7519
                                DO 7518 M=KM1,I5
7518 CALL G2(COS,SIN,R(M),S(M))
7519 CONTINUE
                                    ENDIF
                                DO 7520 M=KBEG,KEND
                                MK=M-KBEG+K

```

```

7520 WRITE(9,REC=M) R(MK)
7530 CONTINUE
7550 CONTINUE
      DO 7580 M=1, I3
      MEND=(I5-(M+1)/2.0)*M+M
      DO 7570 K=1,NBY
      MK=MEND-NBY+K
7570  READ(9,REC=MK) R(K)
      DO 7560 J=1,NBY
      DO 7560 K=J,NBY
7560  BPB(J,K)=BPB(J,K)+R(K)*R(J)
7580  CONTINUE
      DO 7590 J=1,NBY
      DO 7590 K=J,NBY
      WRITE(10,7561) NOMEY(J),NOMEY(K),BPB(J,K)
7590  continue
      CLOSE(9,STATUS='DELETE')
7561  FORMAT(1X,A8,1X,A8,1X,F20.8)
      IF(NBY.GT.1) THEN
      DO 7592 J=1,NBY
      DO 7592 K=1,J
      IF(K.LT.J) BPB(J,K)=BPB(K,J)
7592  CONTINUE
      CALL MABB(E,BPB,NBY)
      TR=0.00
      DO 7595 J=1,NBY
7595  TR=TR+BPB(J,J)
      WRITE(10,'(//,1X,A,A6,A1,A6,/,A,F20.10)')
2'  VALOR DO TRACO DE HOTELLING-LAWLEY PARA TESTAR A HIPOTESE DE NUL
3IDADE DOS EFEITOS CONJUNTOS DA INTERACAO ',NOMEVC(I1),'*',
4NOMEVC(I2),' SOBRE AS RESPOSTAS: ',TR
      JQ=GLID(I)
      JS=MIN(JQ,NBY)
      JM=ABS(NBY-JQ)-1
      GN=0.5*(IDFF-NBY-1.0)
      GL1=JS*(JM+JS+1.0)
      GL2=2.0*(JS*GN+1.0)
      FF=GL2*TR/(JS*GL1)
      IGL1=INT(GL1)
      IGL2=INT(GL2)
      WRITE(10,7199) 'APROXIMACAO A DISTRIBUICAO DE F COM ',IGL1,
2'  E ',IGL2,' G.L. : ',FF
      WRITE(10,997)
      ENDIF
7300  CONTINUE
      ENDIF
933  FORMAT(I6,4X,I6,2X,F12.8)
900  FORMAT(T5,I7,' a. observat,o j rodada.')
997  FORMAT(1X,///)
8000  CONTINUE
      IF((IOS.NE.2).and.(IRC.EQ.2))THEN
      GOTO 305
      else
      goto 8100
      endif
8100  CONTINUE
      denovo=denovo+1
      if(denovo.eq.2)then
      lbeg=lcom
      WRITE(20,997)
      write(30,997)
      write(40,997)
      write(50,997)
      write(60,997)
      write(70,997)
      write(80,997)
      write(80,997)
      goto 310
      else
      goto 8101
      endif
8101  continue
      CLOSE (15, STATUS='DELETE')
      RETURN

```



```

END
C*****
SUBROUTINE OPERA(S,KBEG,NCU,IU,R,URAM)
REAL S(5301),R(5301),URAM(14052951)
INTEGER KBEG,ncu,IU(5301,2)
DO 20 J=KBEG,NCU
IF(S(J).EQ.0.00) GOTO 20
IBEG=IU(J,1)
IEND=IU(J,2)
DO 4 L=IBEG,IEND
LJ=L-IBEG+J
4 R(LJ)=URAM(L)
IF(R(J).EQ.0.00) THEN
ISS=1
IF(S(J).LT.0.00) ISS=-1
DO 6 ITR=J,NCU
SWAP=-1*R(ITR)
R(ITR)=ISS*S(ITR)
6 S(ITR)=SWAP
ELSE
CALL G1(R(J),S(J),COS,SIN,R(J))
S(J)=0.00
IC=J+1
IF(IC.GT.NCU) GOTO 10
DO 8 KK=IC,NCU
8 CALL G2(COS,SIN,R(KK),S(KK))
10 CONTINUE
ENDIF
DO 12 L=IBEG,IEND
LJ=L-IBEG+J
12 URAM(L)=R(LJ)
20 CONTINUE
RETURN
END
C *****
SUBROUTINE G1(A,B,COS,SIN,SIG)
ZERO=0.0
ONE=1.0
IF(A.LE.B) GOTO 10
XR=B/A
YR=SQRT(ONE+XR**2)
COS=SIGN(ONE/YR,A)
SIN=COS*XR
SIG=ABS(A)*YR
RETURN
10 IF(B) 20,30,20
20 XR=A/B
YR=SQRT(ONE+XR**2)
SIN=SIGN(ONE/YR,B)
COS=SIN*XR
SIG=ABS(B)*YR
RETURN
30 SIG=ZERO
COS=ZERO
SIN=ONE
RETURN
END
C *****
SUBROUTINE G2(COS,SIN,X,Y)
XR=COS*X+SIN*Y
Y=-SIN*X+COS*Y

```

```

X=XR
RETURN
END
SUBROUTINE MABB(A,B,N)
PERFORMA B=A*B
C REAL A(5,5),B(5,5),V(5)
DO 20 J=1,N
DO 16 I=1,N
V(I)=0.00
DO 16 K=1,N
16 V(I)=V(I)+A(I,K)*B(K,J)
DO 20 I=1,N
20 B(I,J)=V(I)
RETURN
END
SUBROUTINE INVLUG(A,N)
PROGRAMA PARA INVERTER NO LOCAL UMA MATRIZ "A"
C REAL A(5,5)
DO 250 I=1,N
IF (ABS(A(I,I)).LT.0.0000001) GOTO 300
A(I,I)=1.0/A(I,I)
DO 50 J=1,N
IF(I.EQ.J)GOTO 50
A(I,J)=A(I,J)*A(I,I)
50 CONTINUE
DO 150 K=1,N
IF(I.EQ.K)GOTO 150
DO 100 J=1,N
IF(I.EQ.J)GOTO 100
A(K,J)=A(K,J)-A(I,J)*A(K,I)
100 CONTINUE
150 CONTINUE
DO 200 K=1,N
IF(I.EQ.K)GOTO 200
A(K,I)=-1.00*A(K,I)*A(I,I)
200 CONTINUE
250 CONTINUE
RETURN
300 WRITE(*,310)I
310 FORMAT(' **** SINGULARIDADE NA LINHA ',I5)
RETURN
END

```