

**APLICAÇÕES DOS MODELOS LINEARES MISTOS NA PESQUISA
AGROPECUÁRIA**

ÉRICA MIRRE PEREIRA

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO**

CAMPOS DOS GOYTACAZES - RJ

JUNHO – 2011

**APLICAÇÕES DOS MODELOS LINEARES MISTOS NA PESQUISA
AGROPECUÁRIA**

ÉERICA MIRRE PEREIRA

Dissertação apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Produção Vegetal.

Orientador: Profº JOSÉ TARCISIO LIMA THIÉBAUT

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO**

CAMPOS DOS GOYTACAZES - RJ

JUNHO – 2011

FICHA CATALOGRÁFICA

Preparada pela Biblioteca do **CCTA / UENF** 069/2011

Pereira, Érica Mirre

Aplicações dos modelos lineares mistos na pesquisa agropecuária /
Érica Mirre Pereira. – 2011.
91 f.

Orientador: José Tarcisio Lima Thiébaud
Dissertação (Mestrado em Produção Vegetal) – Universidade
Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e
Tecnologias Agropecuárias. Campos dos Goytacazes, RJ, 2011.
Bibliografia: f. 87 – 91.

1. Estrutura de modelos lineares mistos 2. Componente de
variância 3. Delineamento inteiramente casualizado 4. Delineamento
bloco casualizado 5. Programas SAS I. Universidade Estadual do
Norte Fluminense Darcy Ribeiro. Centro de Ciências e Tecnologias
Agropecuárias. II. Título.

CDD – 630.724

APLICAÇÕES DOS MODELOS LINEARES MISTOS NA PESQUISA AGROPECUÁRIA

ÉRICA MIRRE PEREIRA

“Dissertação apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Produção Vegetal.”

Aprovada em 28 de Junho de 2011.

Comissão examinadora:

Prof. Alessandro Coutinho Ramos (Ph.D., Produção Vegetal) – (UVV – ES)

Prof. Geraldo de Amaral Gravina (D.Sc., Fitotecnia) - UENF

Prof. Rogério Figueiredo Daher (D.Sc., Produção Vegetal) - UENF

Prof. José Tarcísio Lima Thiébaud (D.S.c., Produção Animal) – UENF

ORIENTADOR

Aos meus pais, Erasmo(*in memorim*)
e Maria da Conceição;

Aos meus irmãos Eliton e Mariana
Ao meu sobrinho João Vitor e ao
meu Amor ;

Dedico.

AGRADECIMENTO

A Deus, que me colocou onde estou e me faz vencer qualquer obstáculos da Vida.

A Nossa Senhora Aparecida que estar sempre me cobrindo com seu manto SAGRADO.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) e ao Centro de Ciências e Tecnologias Agropecuária (CCTA), pela oportunidade de realizar ao curso de Pós Graduação em Produção Vegetal e por tida a infra-estrutura oferecida.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao técnico Júlio Meirelhes (LEAG), pela atenção, força, serenidade, amizade e disponibilidade de me ajudar a qualquer hora.

Ao profº Gravina por toda orientação, confiança e amizade.

Ao profº Thiébaud, pelo aprendizado.

A minha FAMÍLIA, a qual sempre tem o apoio, incentivo, amor e compreensão em todos os momentos, principalmente na dificuldade que tive durante todo o curso.

A minha Amada Mãe, que sempre me ajuda em tudo.

Ao meu cunhado Kaio e a minha irmã Mariana que abriram os braços e me acolheram em sua residência.

Aos meus tios Décio e Eliane que me acolheram em sua residência e pelas caronas.

Ao Lenilton, por todo carinho, amor, compreensão, ajuda, confiança e apoio durante esses meses finais.

Ao Hildefonso e Eliana de Mimoso pela agradável amizade e convívio durante as disciplinas.

Ao Teólogo e Parapsicólogo PE. Gilberto Roberto Silva pela ajuda e orientação prestada na minha Vida.

SUMÁRIO

RESUMO.....	vi
ABSTRACT.....	viii
1. INTRODUÇÃO.....	01
2. REVISÃO BIBLIOGRÁFICA.....	03
2.1 ESTIMADORES REML.....	04
2.2 Heterogeneidade da Variância.....	15
2.3 Conexidade de dados.....	17
3. MATERIAL E MÉTODOS.....	20
3.1 Estruturação dos Modelos Lineares Mistos.....	20
3.2 Obtenção Derivação das Equações dos Modelos Mistos.....	26
3.3 Componente de Variância.....	27
4.0 RESULTADOS E DISCUSSÃO.....	47

4.1 Fatorial no Delineamento Inteiramente Casualizado – 2 Fatores.....	47
4.2 Fatorial no Delineamento Inteiramente Casualizado – 3 Fatores.....	52
4.3 Experimentos Aninhados e Fatorial – Aninhado.....	58
4.3.1 Experimento com 1 Fator Aninhado.....	58
4.3.2 Experimento Fatorial Aninhado.....	60
4.4 Parcela Sub-Dividida (SPLIT-PLOT).....	64
4.5 Programas Genes.....	69
4.5.1 Delineamento Inteiramente Casualizado.....	69
4.5.2 Delineamento Em Blocos Casualizados Completos.....	69
4.5.3 Esquema Fatorial Completo Em Ambos Delineamentos Mencionados Anteriormente, Quando Os Dados Estão Balanceados.....	70
4.5.4 Esquema Em Parcela Sub-Divididas Em Ambos Os Delineamentos Mencionados Anteriormente, Quando Os Dados Estão Balanceados.....	70
4.6 Programa SAS.....	72
4.6.1 Fatorial Bloco Casualizados I.....	73
4.6.2 Fatorial Bloco Casualizados II.....	77
4.6.3 Fatorial Bloco Casualizados III.....	81
5. RESUMO E CONCLUSÃO.....	86
REFERÊNCIAS BIBLIOGRÁFICAS.....	87

RESUMO

Pereira, Érica Mirre Pereira; M.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro. Junho, 2011. **APLICAÇÕES DOS MODELOS LINEARES MISTOS NA PESQUISA AGROPECUÁRIA**. Orientador: José Tarcísio Lima Thiébaud. Coorientador: Geraldo de Amaral Gravina.

Os modelos mistos são tradicionalmente estudados com o uso do procedimento da análise de variância, quando os componentes de variância são estimados pela solução de um sistema linear de equações. Para a análise dos modelos lineares mistos, alguns pontos são relevantes: a predição dos efeitos aleatórios, testes e estimação dos componentes de variância, estimação e testes de hipóteses para os efeitos definidos como fixos.

Na análise dos modelos lineares mistos desbalanceados, o problema não é tão simples e o uso do pacote estatístico SAS (Statistical Analysis System) pode ser utilizado, demandando um volume maior de conhecimento teórico, em razão da opção pela estrutura da matriz de variância e covariância e da escolha do método de estimação dos componentes. No procedimento MIXED estão disponíveis os métodos ML Maximum Likelihood, REML (Restricted Maximum Likelihood) e MIVQUE (Minimum Variance Quadratic Unbiased), descritos respectivamente por (Hartley & Rao, 1967), (Rao, 1971) e (Patterson & Thompson, 1971). Para uso otimizado do SAS é necessário também que se defina o tipo de soma de quadrados mais adequado: ss1, ss2, ss3 ou ss4, segundo o SAS. Por suas propriedades, o método REML é o mais indicado e, para dados com desbalanceamento, a soma de quadrados do tipo ss3 deve ser escolhida. Quando

dos dados estão balanceados, os quatro tipos de soma de quadrados apresentam os mesmos resultados.

ABSTRACT

The mixed models have been traditionally studied by means of variance analysis in which variance components are estimated by a linear model solution. For mixed model analysis some points are relevant: the prediction of random effects, tests upon and estimation of variance components, and estimation and tests of hypothesis upon fixed effects. In the analysis of unbalanced linear mixed models the problem is not that simple and the use of the SAS program (SAS System Inc., Cary, NC, USA) is preferable, which demand a proper understanding of the theoretical background because of a proper modeling of the structure of the variance-covariance matrix and the choice of the estimation method are required. In the MIXED procedure are available the ML (Maximum Likelihood), REML (Restricted Maximum Likelihood), and MIVQUE (Minimum Variance Quadratic Unbiased) as described respectively by(Hartley & Rao, 1967), (Rao, 1971), and (Patterson & Thompson, 1971). For an optimized use of SAS it is rather necessary to define the appropriate type of sum of squares: SS1, SS2, SS3, or SS4. Because of its properties the use of REML is recommended and for unbalanced data the SS3 type is the appropriate choice. With balanced data the four types of sum of squares yield the same results.

1. INTRODUÇÃO

Ao se estabelecer um modelo estatístico objetiva-se buscar a explicação de uma ou mais variáveis dependentes por meio de efeitos fixos e ou aleatórios. Desta forma, o êxito da modelagem está diretamente ligado às estruturas de variâncias e covariâncias das variáveis aleatórias que, nos modelos lineares mistos, vão ser explicadas por fatores aleatórios, após a eliminação dos efeitos fixos e da caracterização dos efeitos residuais.

Os modelos lineares, nos parâmetros, possuem pelo menos um efeito aleatório designado por erro experimental. Se o modelo apresenta todos os demais efeitos fixos ele é denominado de modelo linear fixo. Quando o modelo linear apresenta, além do erro experimental, outros efeitos aleatórios em comum com outros efeitos fixos, além da média, e denominado de modelo misto.

Nos modelos lineares mistos, a análise de variância apresenta algumas peculiaridades que devem ser mencionadas. A composição das esperanças matemáticas dos quadrados médios permite a estruturação correta dos testes de hipóteses e, caso o interesse seja a estimação dos componentes de variância, os métodos foram amplamente estabelecidos por (Henderson, 1953), (Rao, 1970), (Hartley & Rao, 1967) e (Patterson & Thompson, 1971).

Em 1953, Henderson apresentou os métodos I, II e III de estimação dos componentes de variância. O método da máxima verossimilhança foi apresentado por (Hartley & Rao, 1967). O método de Estimação Quadrática Não-Viesada de Norma Mínima (MINQUE) foi apresentado por (Rao, 1970 e 1971a). Em 1971, Rao

apresentou o Método de Estimação Quadrática Não-Viesada de Variância Mínima (MIVQUE). Também em 1971, Patterson & Thompson apresentaram o método da Máxima Verossimilhança Restrita (REML) de estimação dos componentes de Variância.

Os modelos lineares fixos foram amplamente estudados, em níveis de complexidade diversos, por (Searle, 1971), (Rao, 1972), (Gray Bill & NETER, 1976), dentre outros. Existe um grande número de livros, abordando aspectos teóricos, aplicados a conjunto de dados balanceados e desbalanceados.

Os estudos dos modelos lineares mistos foram implementados de forma intensiva após trabalhos de (Searle, 1982), (Wolfinger, 1993) e (Littel et al., 1996). Com o desenvolvimento do PROC MIXED do SAS (Littel et al., 1996) a metodologia aplicada aos modelos lineares mistos tornou-se mais usual e segura.

Os modelos lineares de melhoramento animal e vegetal, principalmente os modelos de animais, onde há uma estruturação de parentesco, possibilitam a predição de valores genéticos (efeitos aleatórios), por meios dos BLUP's.

Os modelos lineares mistos possibilitam a predição de valores genéticos aleatórios, por meio dos BLUP's (BEST LINEAR UNBIASED PREDICTION), na presença de efeitos fixos. O procedimento é de importância relevante, possibilitando a seleção de novos indivíduos para a composição de rebanhos, considerando os dados de pais e avós, em situações onde o desbalanceamento é a regra que torna o problema muito mais complexo.

2. REVISÃO BIBLIOGRÁFICA

(Smith & Graser, 1986) propuseram a eliminação de construir um conjunto de sub-rotinas, utilizando o conhecimento teórico disponível, e formataram o programa DFREML (Derivate Free Restricted Maximum Likelihood) para a estimação de componentes de variâncias em modelos lineares mistos.

(Boldman & Van Vleck, 1991), adaptando o algoritmo livre de derivadas, proposto por (Smith & Graser, 1986), desenvolveram novo algoritmo acoplando o método de fatoração de Cholesky ao conjunto de matrizes restritas SPARSPAK, desenvolvido por (George et al., 1980), para a obtenção da matriz de (co)variância de única ou múltiplas características e escreveram o conjunto de sub-rotinas que compõem o programa MTDFREML (Multiple Trait Derivate Free Restricted Maximum Likelihood).

Com o objetivo de reduzir o tempo de computação de dados e gerar programas mais eficientes, (Meyer, 1993) incorporou ao programa DFREML a fatoração de Cholesky e o conjunto de sub-rotinas SPARSPAK, para a operacionalização de matrizes esparsas e, conseqüentemente, da matriz dos coeficientes ou matriz C.

Outros métodos de estimação dos componentes de (co) variâncias, com diferentes graus de viés e invariância à translação, segundo (Verneque, 1994), são: o métodos dos estimadores não-viesados de norma mínima MINQUE (Minimum Norm Quadratic Unbiased Estimation), desenvolvido por (RAO, 1971a), o método da estimação não-viesada de mínima variância quadrática MIVQUE

(Minimum Variance Quadratic Unbiased Estimator), desenvolvido por (RAO, 1971a, 1971b), o método da máxima verossimilhança ML (Maximum Likelihood), desenvolvido por (Hatley & RAO, 1967), o método da falsa esperança PEM (Pseudo Expectation Method), desenvolvido por (Schaeffer, 1986), o método R (1994) descrito por (Golden et al., 1994), e o método da estimação de variâncias por meio das verossimilhanças integradas, com base na estatística Bayesiana, VEIL (Variance Estimation Integrated Likelihood), desenvolvido por (Gianola & Foulley, 1990).

A superioridade no emprego dos preditores BLUP em seleção massal foi verificada por (Kuhlers & Kennedy, 1992), que não sofrem os efeitos das amplitudes de variação nos efeitos do ambiente em grupos contemporâneos.

O mesmo BLUP foi, também, observado por (Quinton et al., 1992). Entretanto, para valores baixos de endogamia por geração, a seleção massal foi superior. Assim os autores concluíram que os programas de melhoramento que otimizam ganhos em curtos espaços de tempos, com procedimentos que podem gerar elevada endogamia, podem não ser viáveis a longo prazo.

(Kuhlers & Kennedy, 1982) concluirão pela superioridade do descarte seletivo, com redução do intervalo de gerações, quando associado ao método BLUP. (Sorensen, 1988) concluiu que o método BLUP fornece informações suficientes para comparar indivíduos de gerações diferentes, apesar de apresentar tendências à subestimação dos valores genéticos em animais novos.

2.1 ESTIMADORES REML

Segundo (Searle, 1982); (Searle et al., 1992); (Schaeffer, 1978) e (Patterson & Thompson, 1971), considerando o modelo linear misto para uma característica:

$$Y = Xb + Zg + e, \text{ em que:}$$

y = vetor de dimensão ($n \times 1$) de observações medidas em n animais,

X = matriz de dimensão ($n \times p$) ou de incidência de p efeitos fixos,

b = vetor coluna de dimensão ($p \times 1$) ou de parâmetros desconhecidos de efeitos fixos,

Z = matriz de dimensão ($n \times n$) ou de incidência de valores genéticos, matriz diagonal com 1 ou 0 na diagonal principal, indicando se a característica y foi medida ou não no animal,

g = vetor de dimensões ($n \times 1$) de Valores Genéticos

e = vetor de dimensão ($n \times 1$) de resíduos

As pressuposições da distribuição conjunta dos vetores (y , g , e) são descritas como:

$$\begin{bmatrix} y \\ g \\ e \end{bmatrix} \sim N \left\{ \begin{bmatrix} xb \\ \phi \\ \phi \end{bmatrix}, \begin{bmatrix} ZGZ' + R & ZG & R \\ GZ' & G & \phi \\ R & \phi & R \end{bmatrix} \right\}, \text{ em que:}$$

$$G = A \times G_0 \quad R = I \times R_0$$

A = matriz do número do coeficiente de parentesco entre os indivíduos e de dimensões ($n \times n$),

I = matriz Identidade de ordem n ,

G_0 = matriz de variância genética aditiva.

R_0 = matriz de variância residual.

n = número de animais da amostra.

Se q característica múltiplas são consideradas, as matrizes R_0 e G_0 serão simétricas e de ordem q e denominadas respectivamente de variâncias e (co)variâncias genéticas aditivas e variâncias e (co)variâncias residuais das q características múltiplas. O objetivo, em qualquer situação, é estimar R_0 e G_0 .

Quando os vetores y , g e têm distribuição conjunta normal multivariada, os vetores b e g são estimados e preditos, respectivamente, por:

$$b^0 = [X'R^{-1}X - X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X]^{-1} \\ * [X'R^{-1}y - X'R^{-1}Z(Z'R^{-1}Z + G^{-1})Z'R^{-1}y]$$

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - Xb^0), \text{ em que:}$$

$$G = A \otimes G_0$$

$$R = I \otimes R_0$$

As matrizes já definidas anteriormente e \otimes = operador do produto direto das matrizes.

Para q características múltiplas a função densidade de probabilidade do vetor y é:

$$f(y) = \frac{1}{(2\pi)^{nq/2} |ZGZ' + R|^{0,5}} e^{-0,5[(y-Xb)'(ZGZ' + R)^{-1}(y-Xb)]}$$

A transformação logaritmo Neperiano de $f(y)$ é dada por:

$$LN [f(y)] = (-0,5 n q) LN (2 \pi) - 0,5 LN|ZGZ' + R| - 0,5 [y'(ZGZ' + R)^{-1}y - 2y'(ZGZ' + R)^{-1}Xb + b'X'(ZGZ' + R)^{-1}Xb]$$

A transformação acima pode ser desdobrada em duas parcelas: $LN[f(y)]'$ e $LN[f(y)]''$ denominadas, respectivamente, por logaritmo neperiano da função densidade probabilidade dos contrastes ortogonais entre os efeitos fixos e logaritmo neperiano da função densidade probabilidade dos contrastes ortogonais entre os efeitos aleatórios.

$$LN [f(y)]' = -0,5 p[x] LN(2\pi) - 0,5 LN[X'(ZGZ' + R)^{-1}X] - 0,5 \{y'(ZGZ' + R)^{-1}X[X'(ZGZ' + R)^{-1}X]^{1/2} X'(ZGZ' + R)^{-1}y - 2y'(ZGZ' + R)^{-1}X [X'(ZGZ' + R)^{-1}X]^{1/2} X'(ZGZ' + R)^{-1}Xb + b'X'(ZGZ' + R)^{-1}X [X'(ZGZ' + R)^{-1}X]^{1/2} X'(ZGZ' + R)^{-1}Xb\}$$

$$LN [f(y)]'' = -0,5 p \{k'[k'(ZGZ' + R)k']^{-1}K\} LN (2\pi) - 0,5 LN|K'(ZGZ' + R)K| - 0,5 \{y'K'[K'(ZGZ' + R)k']^{-1}Ky\}, \text{ em que:}$$

$LN [f(y)]'$ = logaritmo neperiano da função densidade probabilidade dos contrastes ortogonais entre os efeitos fixos e, portanto, constante e livre de derivação por não influir nos componentes;

$LN [f(y)]''$ = logaritmo neperiano da função densidade probabilidade dos contrastes ortogonais entre os efeitos aleatórios;

p = indicação de posto da matriz;

K = é a matriz quadrada dos contrastes ortogonais entre os efeitos aleatórios;

Para a estimação dos componentes de (Co) variância, apenas a parcela $LN[p(y)]''$ fornecerá as derivadas em relação aos elementos das matrizes R_0 e G_0 e no ponto extremo da função de verossimilhança, todas as derivadas serão identicamente nulas.

$$\frac{\partial LN[f(y)]''}{\partial \sigma_{ei}^2} = -0,5tr\{K'[K(ZGZ'+R)KR_i]^{-1}\} + 0,5y'K'[K(ZGZ'+R)K']^{-1}$$

$$KR_iK'[K(ZGZ'+R)K']^{-1}Ky \equiv \phi$$

para $i=[1,q]$

$$\frac{\partial LN[f(y)]''}{\partial \sigma_{ei}\partial \sigma_{ei'}} = -0,5tr\{K'[K(ZGZ'+R)K']^{-1}KR_{ii'}\} + 0,5y'K'[K(ZGZ'+R)K']^{-1}$$

$$KR_{ii'}K'[K(ZGZ'+R)K']^{-1}Ky \equiv \phi$$

para $i \neq i'$ e todo $i=[1,q]$

$$\frac{\partial LN[f(y)]''}{\partial \sigma_{gi}^2} = -0,5tr\{K'[K(ZGZ'+R)KG_i]^{-1}\} + 0,5y'K'[K(ZGZ'+R)K']^{-1}$$

$$KG_iK'[K(ZGZ'+R)K']^{-1}Ky \equiv \phi$$

para $i=[1,q]$

$$\frac{\partial LN[f(y)]''}{\partial \sigma_{gi}\partial \sigma_{gi'}} = -0,5tr\{K'[K(ZGZ'+R)K']^{-1}KG_{ii'}\} + 0,5y'K'[K(ZGZ'+R)K']^{-1}$$

$$KG_{ii'}K'[K(ZGZ'+R)K']^{-1}Ky = \phi$$

para $i \neq i'$ e todo $i=[1,q]$, em que:

$$R_i = Z_i \otimes Q_i$$

$$R_{ii'} = Z_{ii'} \otimes Q_{ii'}$$

$$G_i = (Z_iAZ_i') \otimes Q_i$$

$$G_{ij'} = (Z_i A Z_{j'}) \otimes Q_{ij'}$$

Fazendo as simplificações nas $2(C_p^2 + q)$ derivadas parciais e sabendo que $K'[K(ZGZ'+R)K']^{-1}K = P$ é o projetor ortogonal da parte aleatória das observações no espaço coluna da matriz X :

$$Py = y - Xb^0$$

$$tr(PR_i) = y' PR_i Py$$

$$tr(PR_{ii'}) = y' PR_{ii'} Py$$

$$tr(PG_i) = y' PG_i Py$$

$$tr(PG_{ii'}) = y' PG_{ii'} Py$$

em que:

Z_i = matriz de incidência ($n \times n$) da característica $i = [1, q]$;

Q_i = matriz ($q \times q$) que contém 1 na posição (i, i) e 0 nas demais posições;

$Z_{ii'}$ = matriz de incidência simultânea ($n \times n$) das características i e i' ;

$Q_{ii'}$ = matriz ($q \times q$) que contém 1 nas posições simétricas de i e i' ; 0 nas demais posições;

Outra forma da matriz P é:

$$P = (ZGZ' + R)^{-1} - (ZGZ' + R)^{-1} X [X'(ZGZ' + R)^{-1} X]^{-1} X'(ZGZ' + R)^{-1}$$

$$P(ZGZ' + R)P = P$$

Fazendo $ZGZ' + R = V$, teremos:

$$PVP = [V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}] V [V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}]$$

$$PVP = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} + V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}X \\ *(X'V^{-1}X)^{-1}X'V^{-1}$$

Sendo $(X'V^{-1}X)$ matriz simétrica, a escolha apropriada da inversa generalizada nos levará à igualdade:

$$(X'V^{-1}X)^{-1} = (X'V^{-1}X)^{-1} (X'V^{-1}X) (X'V^{-1}X)^{-1}$$

$$PVP = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$$

$PVP = P$, teremos:

$$\text{tr}(PR_i) = \text{tr}[P(ZGZ' + R)PR_i]$$

$$\text{tr}(PR_{i'}) = \text{tr}[P(ZGZ' + R)PR_{i'}]$$

$$\text{tr}(PG_i) = \text{tr}[P(ZGZ' + R)PG_i]$$

$$\text{tr}(PG_{i'}) = \text{tr}[P(ZGZ' + R)PG_{i'}]$$

para todo $i \neq i'$ e $i \in [1, q]$.

Fatorando os seguintes membros das $2(C_p^2 + q)$ equações acima teremos:

$$\text{tr}(PR_i) = t_i e s$$

$$\text{tr}(PR_{i'}) = t_{i'} e s$$

$$\text{tr}(PG_i) = t_i g s$$

$$\text{tr}(PG_{i'}) = t_{i'} g s s = \begin{bmatrix} \sigma_{e1}^2 \\ \sigma_{e1e2}^2 \\ \vdots \\ \sigma_{e1eq}^2 \\ \vdots \\ \sigma_{eq}^2 \\ \sigma_{g1}^2 \\ \sigma_{g1g2}^2 \\ \vdots \\ \sigma_{g1gq} \\ \vdots \\ \sigma_{gq}^2 \end{bmatrix} \quad t_{ie} = \begin{bmatrix} \text{tr}(PR_i PR_{i1}) \\ \text{tr}(PR_i PR_{i2}) \\ \vdots \\ \text{tr}(PR_i PR_{iq}) \\ \text{tr}(PR_i PG_{i1}) \\ \text{tr}(PR_i PG_{i2}) \\ \vdots \\ \text{tr}(PR_i PG_{iq}) \end{bmatrix}$$

todo $i \in [1, q]$, $R_{ij} = R_i$ e $G_{ij} = G_i$

$$t_{ije} = \begin{bmatrix} tr(PR_{ij}PR_1) \\ tr(PR_{ij}PR_{12}) \\ \vdots \\ tr(PR_{ij}PR_{(q-1)q}) \\ tr(PR_{ij}PR_q) \\ tr(PR_{ij}PR_{G1}) \\ tr(PR_{ij}PR_{G12}) \\ \vdots \\ tr(PR_{ij}PR_{(q-1)q}) \\ tr(PR_{ij}PR_q) \end{bmatrix}$$

De forma análoga para t_{ig} e t_{jg} .

Na forma matricial o sistema é:

$Ts = v$, em que:

$$T = \begin{bmatrix} t'_{1e} \\ t'_{12e} \\ \vdots \\ t'_{(q-1)qe} \\ t'_qe \\ \vdots \\ t'_{1g} \\ t'_{12g} \\ \vdots \\ t'_{(q-1)qg} \\ t'_qg \end{bmatrix} \quad v = \begin{bmatrix} y'PR_1Py \\ y'PR_{12}Py \\ \vdots \\ y'PR_{1q}Py \\ \vdots \\ y'PR_{(q-1)q}Py \\ y'PR_qPy \\ y'PG_1Py \\ y'PG_{12}Py \\ \vdots \\ y'PG_{1q}Py \\ \vdots \\ y'PG_{(q-1)q}Py \\ y'PG_qPy \end{bmatrix}$$

As soluções para o vetor s são obtidas iterativamente:

- 1-obter os elementos T e v , com estimativas iniciais de s ;
- 2-Resolver o sistema e obter novas aproximadamente para s ;
- 3-Repetir o procedimento até a obtenção da convergência.

Matricialmente a função de verossimilhança quando o vetor y tem uma distribuição normal multivariada é dada por:

$$f(y) = (2\pi)^{-0,5N} |V|^{-0,5} \exp(-0,5)[(y - Xb)'v^{-1}(y - Xb)]$$

Onde,

$$N = n_1 + n_2 + \dots + n_q$$

$N = nq$ se todos os grupos de heterogeneidade de variância tem o mesmo tamanho.

No REML, quando um grupo de contrastes $K'y$, e quando $K'X = \phi$ e K' tem posto $[N - p(X)]$.

$$f(k'y) = (2\sigma)^{-0,5[N-p(x)]} |K'VK|^{-0,5} \exp(-0,5)[(K'y)'(K'VK)^{-1}(K'y)]$$

A transformação logaritmo neperiano de $f(k'y)$ é dada por:

$$LN[f(y)] = (-0,5)[N - p(X)] LN(2\pi) - (0,5) LN|K'VK| - (0,5)[y'K(K'VK)^{-1}K'y]$$

A constante $(-0,5)[N-p(X)] LN(2\pi)$ não interfere nos componentes de (Co)variância e pode ser ignorado.

Searle (1971) demonstrou que:

$$LN |K'VK| = LN|V| + LN|X'V^{-1}X|$$

$$y'K(K'VK)^{-1}K'y = y'Py = (y - Xb^0)' V^{-1}(y - Xb^0)$$

$$LN[f(K'y)] = -0,5 LN |V| - 0,5 LN|X'V^{-1}X| - 0,5[(y - Xb^0)'V^{-1}(y - Xb^0)]$$

Quando A e D são matrizes de mesma ordem e, portanto, o produto é possível, então: $|AD| = |A| |D|$

$$|V| = |ZGZ' + R|$$

$$|V| = |R(I + R^{-1}ZGZ')|$$

Quando $R^{-1}ZGZ'$ e idempotente:

$$|V| = |R| |I + ZR^{-1}Z'G|$$

$$|V| = |R(G^{-1} + Z'R^{-1}Z)G|$$

$$|V| = |R| |G^{-1} + Z'R^{-1}Z| |G|$$

e, conseqüentemente, tem-se:

$$LN |V| = LN |R| + LN |G^{-1} + Z'R^{-1}Z| + LN |G|$$

Para matrizes quadradas e não singulares:

$$\begin{bmatrix} A & -B \\ Q & D \end{bmatrix} = |A| |D + QA^{-1}B| = |A| |D + A^{-1}BQ| = |AD| + |BQ| = |D| |A + BD^{-1}Q|$$

Se A e D são matrizes identidade:

$$|I + QB| = |I + BQ| = |I + B'Q'| = |I + Q'B'|$$

se θ é uma constante ou escalar:

$$|\theta A| = \theta^m |A|, \text{ quando } m \text{ é a ordem da matriz } A.$$

Da mesma forma, para:

$$C = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}$$

$$|C| = |Z'R^{-1}Z + G^{-1}| |X'(R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1})X|$$

fazendo, $S = (R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1})$

$$|C| = |X'R^{-1}X| |G^{-1} + Z'SZ|$$

$$|C| = |Z'R^{-1}Z + G^{-1}| |X'V^{-1}X|$$

Como $LN |X'V^{-1}X| = LN |C| - LN |Z'R^{-1}Z + G^{-1}|$ e com as simplificações:

$$LN[f(K'y)] = -0,5 LN |R| - (0,5) LN |G| - (0,5) LN |C| - (0,5)y'Py$$

$$LN |R| = LN |I \sigma_0^2| = N LN(\sigma_0^2)$$

$$LN |G| = \sum_{i=1}^q g_i LN(\sigma_i^2)$$

$$LN |C| = LN |X'R^{-1}X| + LN |Z'SZ + G^{-1}|$$

$$\begin{aligned}
LN | X'R^{-1}X | &= LN | X'X\sigma_0^{-2} | = LN[|(\sigma_0^{-2})^{p(x)} | X'X |] \\
LN | X'R^{-1}X | &= LN | X'X | - p(x) LN(\sigma_0^2) \\
Z'SZ + G^{-1} &= \sigma_0^{-2} Z'MZ + G^{-1} = \sigma_0^{-2} (Z'MZ + \sigma_0^2 G^{-1}) \\
LN | C | &= LN | X'X | - p(X) LN(\sigma_0^2) + LN | Z'MZ + G^{-1} \sigma_0^2 | \\
LN[f(K'y)] &= -(0,5)(N - p(x) - q) LN(\sigma_0^2) - 0,5 \sum_{i=1}^q g_i LN(\sigma_i^2) - (0,5) LN | C^* | - (0,5) y'Py
\end{aligned}$$

que é o LN da parcela da função de verossimilhança, sem a constante, e que afeta os componentes de variância.

$$C^* = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\sigma_0^2 \end{bmatrix}$$

A matriz de (Co) variância de y é

$$V = \sum_{i=1}^q Z_i Z_i' \sigma_{gi}^2 + I \sigma_e^2$$

E o projetor da forma quadrática $y'Py$ é $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$

Derivando parcialmente $LN[f(k'y)]$ em relação a cada σ_i^2 ($i = 1, 2, \dots, q$) obtém-se, segundo SEARLE (1971):

$$\frac{\partial LN[f(K'y)]}{\partial \sigma_{gi}^2} = -(0,5) tr(P) + 0,5 y'PPy, \text{ para todo } i = [1, q]$$

$$tr(PZ_i Z_i) = \frac{q_i}{\sigma_i^2} - tr\left(\frac{C_{ii} \sigma_o^2}{\sigma_{gi}^4}\right)$$

$$tr(P) = [N - p(X)] - \sum_{i=1}^q (\hat{\theta}_i' \hat{\theta}_i) / \sigma_{gi}^2$$

Do sistema de equações normais, obtém a solução abaixo.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b^0 \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}Y \end{bmatrix}$$

$$b^0 = \left\{ X' \left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] X \right\}^{IC} *$$

$$* \left\{ X' \left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] y \right\}$$

$$b^0 = (X'V^{-1}X)^{-1} (X'V^{-1}y),$$

Solução múltipla obtida diretamente do sistema de equações normais, e que coincide com a apresentada por (Schaffer, 1993).

Sendo,

$$Py = (V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})y$$

$$b^0 = (X'V^{-1}X)^{-1} (X'V^{-1}y)$$

Em continuidade à solução do sistema de equações normais, obtém o resultado para os valores genéticos preditos, em que:

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1} Z'R^{-1}(y - Xb^0)$$

Henderson, citado por (Schaeffer, 1993), apresenta a solução.

$\hat{g}_i = GiZ_i'Py$, em que:

$$Py = V^{-1}(y - Xb^0)$$

Henderson demonstrou também, que a forma quadrática.

$$y'PPy = y'y - b^0' X'y - \sum_{i=1}^q \hat{g}'_i Z'_i y - \sum_{i=1}^q \hat{g}'_i \hat{g}_i \left(\frac{\alpha_i}{\sigma_e^2} \right)$$

E, no ponto de máximo da função de máxi-verossimilhança, teremos:

$$\hat{\sigma}_{g_i}^2 = \left[\hat{g}'_i \hat{g}_i + \text{tr}(C_{ii}\sigma_e^2) \right] / q_i$$

$$\hat{\sigma}_e^2 = y'Py / (N - p(X))$$

2.2 HETEROGENEIDADE DA VARIÂNCIA

A preocupação com os efeitos da heterogeneidade de variância, em classes de efeitos fixos, fez com que (Lush, 1933) recomendasse a produção de animais nos locais de uso desses animais, quando os genes desejáveis têm maior probabilidade de se expressar. A maior expressão dos genes de interesse, favorece a obtenção de estimadores mais acurados e precisos. Assim, variações genéticas e ou ambientais estão relacionadas com o desempenho dos animais (Dickerson, 1962).

(Falconer, 1952) deu início ao estudo das correlações genéticas entre as características fenotípicas em ambientes distintos, e concluiu pela significância da interação genótipo ambiente: o melhor genótipo em um ambiente pode não ser tão bom em outro ambiente. (Robertson et. al., 1960) recomendam a comparação das herdabilidades, em classes de efeitos fixos, como forma de evitar as alterações na classificação de animais submetidos à seleção.

No melhoramento animal, a seleção de animais tem sido praticada em determinadas condições ambientais. Em características limitadas pelo sexo, a estimativa dos valores dos pais é obtida considerando o desempenho dos filhos e, neste caso, é importante caracterizar o ambiente onde as progênies estão atuando, bem como a distribuição delas nos vários ambientes. Por exemplo, a semelhança entre filhos de um reprodutor, em um determinado rebanho, é o reflexo tanto da covariância ambiental entre meio-irmãs, como da interação genótipo ambiente (Meyer, 1989) e (Short et AL., 1990). A covariância ambiental entre meio-irmãs paternas é função de fatores biológicos, de manejo e, até mesmo estatístico, como a não remoção do efeito de rebanho, de outro fator fixo ou de um modelo inadequado para a estimação dos valores genéticos (Meyer, 1987).

Se o efeito da interação reprodutor-rebanho for considerado no modelo linear de avaliação genética, a avaliação de progenitores é menos acurada e precisa em poucos rebanhos. A interação pode não afetar significativamente a avaliação de reprodutores, com progênies em vários rebanhos, como é o caso de reprodutores utilizados na inseminação artificial. Ignorar a interação pode significar a super estimação da acurácia nas avaliações, principalmente, quando

poucos rebanhos são considerados. Quando as diferenças entre progênes de um mesmo reprodutor não são as mesmas em diferentes rebanhos, a interação reprodutor rebanho, pode não ser significativa se a heterogeneidade de variâncias explicarem parte dessas diferenças.

Variâncias heterogêneas entre rebanhos e aumento da variância associados ao aumento de médias de características produtivas importantes, têm sido verificadas em várias oportunidades e por vários autores (Meyer, 1987), (Boldman & Freeman, 1990), (Dong & Mad, 1990), (Stanton et. al, 1991), (Dodenhoff & Swalve, 1998), (Torres, 1998) .

Segundo (Torres, 1998), quando a heterogeneidade de variância não é levada em conta, embora presente, a produção das filhas de reprodutores deverão ser ponderada pelos desvios padrão dos rebanhos que as contém. As filhas oriundas de rebanhos com variâncias maiores influenciarão mais a avaliação de reprodutores. Os vieses serão maiores, na seleção das mães, quando elas tendem a expressar suas características em um único rebanho.

(Bereskin & Lush, 1965) concluíram que os resultados da avaliação de reprodutores em rebanhos de alto nível, podem não estar de acordo com avaliação de reprodutores, no teste de progênes, em rebanhos de baixo nível de produção.

Mesmo que o número de dados seja insuficiente para a estimação dos componentes de variâncias, rebanhos similares têm sido agrupados e os componentes têm sido obtidos dentro de cada grupo (Mirande & Van Vleck, 1985). Os estudos evidenciaram a existências de variâncias heterogêneas entre os diferentes ambientes.

(Mohammad et. AL.,1982) estimaram os valores genéticos de reprodutores para a produção de leite com modelos reprodutores com ou sem a interação reprodutor-rebanho e interação reprodutor-peso ideal de novilhas. A correlação de SPEARMAN entre os modelos com e sem interação foram significativos, indicando que não houve alteração, também, significativa na ordenação dos valores genéticos preditos nos vários modelos utilizados.

(Van Tassel & Berger, 1994), com dados simulados, estudaram as conseqüências da eliminação da interação reprodutor-rebanho e da matriz de

parentesco na estimação dos componentes de variância em três casos distintos de distribuição de filhas nos rebanhos. Os componentes de variância foram estimados pelo método das estimativas quadráticas não-viesadas de variância mínima (MIVQUE). Quando o parentesco entre os reprodutores foi eliminado, as variâncias do efeito de reprodutor e da interação reprodutor-rebanho tenderam à subestimação, quando os reprodutores eram mais próximos. A variância atribuída aos reprodutores foi superestimada quando, no modelo linear, a interação reprodutor-rebanho não foi considerada. Para os dados resultantes de inseminação artificial, quando não foi considerada a matriz de parentesco, a variância da interação reprodutor-rebanho foi subestimada.

O efeito da interação reprodutor-rebanho quando as variâncias entre rebanhos eram heterogêneas, e seis diferentes combinações de heterogeneidades de variâncias entre rebanho – ano - estação foram simuladas. Quando a característica não é influenciada pela interação reprodutor-rebanho a acurácia das estimativas de valores genéticos pode ser subestimada.

(Spike & Freeman, 1976) verificaram a importância da acurácia na avaliação genética animal. Concluíram que com o aumento da acurácia definida como a eficiência do processo de avaliação na identificação das diferenças observadas nos fenótipos, a mudança genética esperada foi ampliada significativamente.

(Ojala et. al., 1984), trabalhando com dados simulados, avaliaram o efeito do número de amostras nas classes rebanho-reprodutor, sobre a acurácia da avaliação genética. Os resultados demonstraram que a acurácia das estimativas melhora significativamente quando o número de amostras aumenta de 2 para 3 observações por classe. O aumento de número de progênes por reprodutor teve, no entanto, maior efeito sobre acurácia.

2.3 CONEXIDADE DE DADOS

Segundo (Searle, 1971) para modelos fixos de classificação cruzada, a conexão entre os dados existe quando todas as funções lineares entre dois níveis dos fatores considerados são estimáveis, ou seja, quando existem observações em todas as combinações de níveis ou classes de fatores. Nos

modelos mistos, se a pressuposição de esperança dos efeitos aleatórios, dentro de níveis dos efeitos fixos não se verifica, as comparações entre os efeitos aleatórios em diferentes níveis dos efeitos fixos serão viesadas.

A simulação de dados tem sido usada com freqüência em estudo de problemas de conectividade e o efeito na avaliação genética animal, nos diversos tipos de modelos lineares. (Tong et al., 1980), (Kennedy, 1975, 1988, 1991) e (Analla et. al., 1995) foram alguns dos grupos que trabalharam com simulação nos problemas de conectividade dos dados experimentais.

(Foulley et al., 1987) discutiram o impacto da conectividade em relação ao tipo de modelo, fixo ou misto, utilizado e o efeito nas propriedades de um estimador de mérito genético, não-viesado e com quadrado médio do erro mínimo. Segundo os autores, o problema de conectividade restringe o uso do método dos quadrados mínimos regressados, quando há falta de conectividade o processo iterativo é interrompido. Quando a metodologia dos modelos mistos é convenientemente aplicada, o sistema de equações normais sempre apresentará solução, sem levar em conta o grau de conectividade.

(Schenkel, 1991) avaliou as vantagens do uso dos modelos mistos quando o problema de conectividade existe. Nos modelos mistos o processo faz com que as soluções dentro de grupos desconectados somem zero e não provoquem a quebra do processo de convergência.

A teoria do índice de seleção só admite comparações entre animais de um mesmo grupo contemporâneo ou que compartilham o mesmo momento reprodutivo. O método dos quadrados mínimos permite a comparação entre grupos contemporâneos, quando existe a conectividade dos dados.

O efeito de três tipos de falta de conectividade em um modelo linear com 2 fatores classificatórios sobre a estimação dos componentes de variância, usando o método da estimação quadrática não-viesada de norma mínima (MINQUE). O autor concluiu que a falta de conectividade acarreta a falta de associação dos graus de liberdade com as várias formas quadráticas. Ainda segundo (Schenkel, 1991), parece impróprio estimar os componentes de variância sem antes eliminar as subclasses desconectadas. Em pequenos grupos de dados o problema é detectado de imediato, mas, para um grande volume de dados, o problema não é

evidenciado e pode ser a causa de ocorrência de estimativas negativas dos componentes de variância.

3. MATERIAL E MÉTODOS

Os dados que serão utilizados no presente trabalho, em razão de sua formatação teórica e didática, serão simulados em função do modelo linear misto estabelecido.

3.1 Estruturação dos modelos lineares mistos

Matricialmente, um modelo linear misto foi descrito por (Harville, 1977) e por (Laird & Ware, 1982) como:

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + Z \underset{\sim}{v} + \underset{\sim}{e}, \text{ em que:}$$

$\underset{\sim}{y}$ = É um vetor das observações conhecidas, com n linhas;

X = É a matriz de incidência dos efeitos fixos, portanto conhecida, de n linhas e (p+1) colunas;

$\underset{\sim}{\beta}$ = É um vetor desconhecido dos efeitos fixos, com (p+1) linhas;

Z = Matriz de incidência dos efeitos aleatórios, portanto conhecida, com n linha e q colunas;

$\underset{\sim}{v}$ = É um vetor desconhecido dos efeitos aleatórios, com q linhas;

$\underset{\sim}{e}$ = É um vetor dos erros aleatórios, também desconhecido, com n linhas;

n é o número de observações da variável aleatória e dependente;

p é o número de parâmetros dos efeitos fixos.

q é o número de parâmetros dos efeitos aleatórios.

Os q efeitos aleatórios e os n erros são considerados normal e independentemente distribuídos com média zero e matrizes de variâncias e covariâncias dadas, respectivamente, por G e R .

Por hipóteses, as matrizes G e R são positivas definidas e não singulares definidas com:

$$G = Var\left(\underset{\sim}{v}\right) = E\left(\underset{\sim}{v}\underset{\sim}{v}'\right)$$

$$R = Var\left(\underset{\sim}{e}\right) = E\left(\underset{\sim}{e}\underset{\sim}{e}'\right)$$

Assim sendo:

$$Var\left[\begin{array}{c} \underset{\sim}{v} \\ \underset{\sim}{e} \end{array}\right] = \begin{bmatrix} G & \Phi \\ \Phi & R \end{bmatrix}$$

$$Var\left(\underset{\sim}{y}\right) = V = Var\left(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{v} + \underset{\sim}{e}\right)$$

$$V = Var\left(\underset{\sim}{X}\underset{\sim}{\beta}\right) + Var\left(\underset{\sim}{Z}\underset{\sim}{v}\right) + Var\left(\underset{\sim}{e}\right)$$

$$V = \Phi + ZVar\left(\underset{\sim}{v}\right)Z' + I_{(n)}Var\left(\underset{\sim}{e}\right)I'_{(n)}$$

$$V = ZGZ' + R$$

A matriz V é, também por hipótese, positiva definida e não singular.

$$\begin{aligned} E\left(\underset{\sim}{y}\right) &= E\left(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{v} + \underset{\sim}{e}\right) \\ &= E\left(\underset{\sim}{X}\underset{\sim}{\beta}\right) + E\left(\underset{\sim}{Z}\underset{\sim}{v}\right) + E\left(\underset{\sim}{e}\right) \\ &= E\left(\underset{\sim}{X}\underset{\sim}{\beta}\right) + ZE\left(\underset{\sim}{v}\right) + E\left(\underset{\sim}{e}\right) \\ &= X\beta + ZE\left(\underset{\sim}{v}\right) + E\left(\underset{\sim}{e}\right) \end{aligned}$$

$$E\left(\underset{\sim}{v}\right) = \Phi \text{ e } E\left(\underset{\sim}{e}\right) = \Phi$$

$$E\left(\underset{\sim}{y}\right) = X\underset{\sim}{\beta}$$

Portanto, desta forma, o vetor $\underset{\sim}{y}$ é aleatório, com distribuição normal de média $X\underset{\sim}{\beta}$ e variância igual a $(ZGZ' + R)$.

$$\underset{\sim}{y} \sim N\left(\underset{\sim}{X}\underset{\sim}{\beta}, ZGZ' + R\right)$$

$$\underset{\sim}{y} \sim N\left(\underset{\sim}{X}\underset{\sim}{\beta}, V\right)$$

Assumindo que os efeitos aleatórios v_i , com $i = 1, 2, \dots, q$ e e tem distribuição normal com média zero e matriz de variâncias e covariâncias $\sigma_i^2 I_{(n)}$, para $i = 1, 2, \dots, q$ e $\sigma_e^2 I_{(n)}$:

$$V = \sum_{i=1}^q z_i z_i' \sigma_i^2 + \sigma_e^2 I_{(n)} \quad \text{ou} \quad V = \sum_{i=0}^q z_i z_i' \sigma_i^2, \text{ quando } \sigma_0^2 = \sigma_e^2 \text{ e } z_0 z_0' = I$$

Para iniciar a formatação das matrizes de um modelo misto, vamos supor inicialmente um delineamento em blocos casualizado com tratamento fixo e o efeito de bloco aleatório, segundo o quadro teórico de dados dado abaixo.

TRATAMENTO

BLOCO 1 2 3 4

1 $y_{11} y_{21} y_{31} y_{41}$

2 $y_{12} y_{22} y_{32} y_{42}$

3 $y_{13} y_{23} y_{33} y_{43}$

Modelo linear: $y_{ij} = m + t_i + b_j + e_{ij}$

Matricialmente: $\underset{\sim}{y} = X \underset{\sim}{\beta} + Z \underset{\sim}{v} + \underset{\sim}{e}$

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \underset{\sim}{\beta} = \begin{bmatrix} m \\ t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \underset{\sim}{v} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad \underset{\sim}{e} = \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \end{bmatrix}$$

Um outro exemplo, de complexidade maior, vamos supor 3 formatação de adubo e 4 genótipos, segundo um esquema fatorial 3*4, no delineamento inteiramente casualizado com 2 repetições.

$$\text{Modelo linear: } y_{ijk} = m + f_i + g_j + fg_{ij} + e_{ijk}$$

Em que: formulação de adubo é fixo e genótipo é aleatório. Matricialmente,

$$\text{teremos: } \underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{Z} \underset{\sim}{v} + \underset{\sim}{e}$$

$$i=1,2,3 \quad j=1,2,3,4 \quad k=1,2$$

Considerando o quadro teórico de dados, as matrizes serão dadas por:

<i>FR</i>	1				2				3			
<i>G</i>	1	2	3	4	1	2	3	4	1	2	3	4
y_{111}	y_{121}	y_{131}	y_{141}	y_{211}	y_{221}	y_{231}	y_{241}	y_{311}	y_{321}	y_{331}	y_{341}	
y_{112}	y_{122}	y_{132}	y_{142}	y_{212}	y_{222}	y_{232}	y_{242}	y_{312}	y_{322}	y_{332}	y_{342}	

$$\underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{e}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} m \\ f1 \\ f2 \\ f3 \end{bmatrix}$$

$$\underset{\sim}{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{131} \\ e_{132} \\ e_{141} \\ e_{142} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \\ e_{231} \\ e_{232} \\ e_{241} \\ e_{242} \\ e_{311} \\ e_{312} \\ e_{321} \\ e_{322} \\ e_{331} \\ e_{332} \\ e_{341} \\ e_{342} \end{bmatrix}$$

A interação entre um efeito fixo e um efeito aleatório é, também, um efeito aleatório, desta forma teremos:

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Phi \underset{\sim}{v} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ fg_{11} \\ fg_{12} \\ fg_{13} \\ fg_{14} \\ fg_{21} \\ fg_{22} \\ fg_{23} \\ fg_{24} \\ fg_{31} \\ fg_{32} \\ fg_{33} \\ fg_{34} \end{bmatrix}$$

No contexto de modelo misto, as diferenças possuem componentes fixos e aleatórios.

3.2 Obtenção derivação das equações dos modelos mistos

A obtenção do sistema de equações normais dos modelos lineares mistos pode ser feita com a minimização da soma de quadrado dos erros ou com a maximização da função densidade de probabilidade conjunta dos vetores \tilde{y} e \tilde{v} .

Admitindo a distribuição normal do vetor \tilde{y} , a função densidade de probabilidade $g(\tilde{y})$ é dada por:

$$g\left(\begin{matrix} \tilde{y} \\ \tilde{} \end{matrix}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} |ZGZ'+R|^{\frac{1}{2}}} e^{-\frac{1}{2} \left[\begin{matrix} \tilde{y}-X\beta \\ \tilde{} \end{matrix} \right]' (ZGZ'+R)^{-1} \begin{matrix} \tilde{y}-X\beta \\ \tilde{} \end{matrix} \right]}$$

Também o vetor \tilde{v} tem distribuição normal com função densidade de probabilidade:

$$g\left(\begin{matrix} \tilde{v} \\ \tilde{} \end{matrix}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} |G|^{\frac{1}{2}}} e^{-\frac{1}{2} \left[\begin{matrix} \tilde{v}-\Phi \\ \tilde{} \end{matrix} \right]' G^{-1} \begin{matrix} \tilde{v}-\Phi \\ \tilde{} \end{matrix} \right]}$$

A função densidade de probabilidade de conjunta dos vetores \tilde{y} e \tilde{v} , dada por:

$$g\left(\begin{matrix} \tilde{y}, \tilde{v} \\ \tilde{}, \tilde{} \end{matrix}\right) = g\left(\begin{matrix} \tilde{y}/\tilde{v} \\ \tilde{}/\tilde{} \end{matrix}\right) * g\left(\begin{matrix} \tilde{v} \\ \tilde{} \end{matrix}\right)$$

3.3 Componente de variância

Quando X é uma variável aleatória contínua e dp é a sua diferencial de probabilidade, a esperança matemática de X , representada por :

$$E(X) = \int_{-\infty}^{\infty} x \, dp, \text{ em que :}$$

$$\int_{-\infty}^{\infty} dp = 1$$

As propriedades da esperança matemática são:

1) Esperança matemática da constante c

$$E(c) = \int_{-\infty}^{\infty} c \, dp = c \int_{-\infty}^{\infty} dp = c$$

2) Esperança matemática do produto cX .

$$E(cX) = \int_{-\infty}^{\infty} cX \, dp = c \int_{-\infty}^{\infty} X \, dp = cE(X)$$

3) Se X e Y são duas variáveis aleatórias contínuas:

$$E(X + Y) = \int_{-\infty}^{\infty} (X + Y) \, dp = \int_{-\infty}^{\infty} X \, dp + \int_{-\infty}^{\infty} Y \, dp = E(X) + E(Y)$$

4) Se X e Y são independentes:

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XY \, dp = \int_{-\infty}^{\infty} X \, dp \cdot \int_{-\infty}^{\infty} Y \, dp = E(X) \cdot E(Y)$$

A $E(X)$ é o que denominamos de primeiro momento da variável aleatória contínua X em relação à origem. Então, o n -ésimo momento de X em relação à origem é dado por:

$$E(X^n) = \int_{-\infty}^{\infty} X^n \, dp, \text{ para } n = 0, 1, 2, \dots$$

O primeiro momento em relação à origem é a média da distribuição de X .

$$E(X) = \int_{-\infty}^{\infty} X \, dp = \mu_x$$

Com base nas definições dos momentos em relação à origem, estabeleceremos os momentos em relação à média ou em relação à esperança da variável aleatória contínua X . O n -ésimo momento da variável aleatória contínua X em relação à $E(X)$ é dado por:

$$E[X - E(X)]^n = \int_{-\infty}^{\infty} [X - E(X)]^n dp$$

O segundo momento da variável aleatória contínua X , denominado de Variância de X , é dado por:

$$\begin{aligned} E[X - E(X)]^2 &= \int_{-\infty}^{\infty} [X - E(X)]^2 dp \\ &= \int_{-\infty}^{\infty} [X^2 - 2XE(X) + E^2(X)] dp \\ &= \int_{-\infty}^{\infty} X^2 dp - 2\mu_x \int_{-\infty}^{\infty} X dp + \int_{-\infty}^{\infty} \mu_x^2 dp \\ &= E(X^2) - \mu_x^2 = E(X^2) - [E(X)]^2 \end{aligned}$$

A covariância entre duas variáveis aleatórias contínuas X e Y é dada por:

$$\begin{aligned} \text{Cov}(X, Y) &= E\{(X - E(X)) \cdot (Y - E(Y))\} \\ \text{Cov}(X, Y) &= E\{XY - XE(Y) - YE(X) + E(X)E(Y)\} \\ \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ E(XY) &= \text{Cov}(XY) + E(X)E(Y) \end{aligned}$$

Em função da quarta propriedade da esperança matemática de X , se X e Y são linearmente independentes:

$$E(XY) = E(X)E(Y), \text{ então:}$$

$$\text{Cov}(X, Y) = 0, \text{ para } X \text{ e } Y \text{ linearmente independentes.}$$

Com as definições dos momentos em relação à média da distribuição de X , obtém-se as propriedades da $V(X)$.

$$\begin{aligned} 1) \quad V(cX) &= E(cX)^2 - [E(cX)]^2 \\ V(cX) &= c^2 [E(X^2) - E(X)^2] \\ V(cX) &= c^2 V(X) \end{aligned}$$

Quando c é uma constante.

$$\begin{aligned} 2) V(X + Y) &= E[(X + Y) - E(X + Y)]^2 \\ V(X + Y) &= E[X + Y - E(X) - E(Y)]^2 \\ V(X + Y) &= E\{[X - E(X)] + [Y - E(Y)]\}^2 \\ V(X + Y) &= E[X - E(X)]^2 + E[Y - E(Y)]^2 + 2E\{[X - E(X)] \cdot [Y - E(Y)]\} \\ V(X + Y) &= V(X) + V(Y) + 2 \text{Cov}(XY) \end{aligned}$$

De forma análoga, obteremos:

$$\begin{aligned} V(X - Y) &= V(X) + V(Y) - 2\text{Cov}(XY) \\ V(ax + by) &= a^2V(X) + b^2V(Y) + 2ab\text{Cov}(XY) \end{aligned}$$

, quando a e b são duas constantes diferentes de zero. O produto algébrico ab deverá ser sempre considerado.

A variância de uma função linear da variável contínua será então estabelecida como:

$$\begin{aligned} V(a_1X_1 + a_2X_2 + \dots + a_kX_k) &= V\left(\sum_{i=1}^k a_iX_i\right) = a_1^2V(X_1) + a_2^2V(X_2) + \dots + a_k^2V(X_k) \\ &+ 2a_1a_2\text{Cov}(X_1, X_2) + 2a_1a_3\text{Cov}(X_1, X_3) + \dots + 2a_{k-1}a_k\text{Cov}(X_{k-1}, X_k) \end{aligned}$$

Portanto, com c_k^2 expressões de covariância.

Assim sendo, para o modelo estatístico do delineamento inteiramente casualizado:

$$y_{ij} = m + t_i + e_{ij}, \text{ em qualquer situação: } m \text{ é fixo e } e_{ij} \text{ é aleatório, para } i = 1, 2, \dots, v$$

e $j = 1, 2, \dots, r$. Desta forma:

$$E(m) = m \quad \text{e} \quad E(m^2) = m^2$$

$$E(e_{ij}) = 0 \quad \therefore \quad E(e_{ij}^2) = V(e_{ij}) = \sigma^2$$

Os erros são, para efeito de teste, normais e independentemente distribuídos com média 0 e variância σ^2 .

$$e_{ij} \sim \text{NID}(0, \sigma^2)$$

O modelo do delineamento inteiramente casualizado é fixo quando t_i é um efeito fixo e aleatório quando t_i é um efeito aleatório.

Para o modelo do delineamento em blocos casualizados:

$$y_{ij} = m + t_i + b_j + e_{ij}, \text{ em qualquer situação:}$$

m é fixo, e_{ij} é aleatório, para $i = 1, 2, \dots, v$ e $j = 1, 2, \dots, r$.

O modelo do delineamento em blocos casualizados será considerado fixo quando t_i e b_j forem efeitos fixos. Então:

$$E(m) = m \text{ e } E(m^2) = m^2$$

$$E(e_{ij}) = 0 \text{ e } E(e_{ij}^2) = V(e_{ij}) = \sigma^2$$

$$E(t_i) = t_i, E(t_i^2) = t_i^2 \text{ e } \sum_{i=1}^v t_i = 0$$

$$E(b_j) = b_j, E(b_j^2) = b_j^2 \text{ e } \sum_{j=1}^r b_j = 0$$

O modelo do delineamento em blocos casualizados será considerado aleatório quando além de m fixo e e_{ij} aleatório, t_i e b_j forem ambos os efeitos aleatórios.

$$E(m) = m \text{ e } E(m^2) = m^2$$

$$E(e_{ij}) = 0 \text{ e } E(e_{ij}^2) = V(e_{ij}) = \sigma^2$$

$$E(t_i) = 0, E(t_i^2) = V(t_i) = \sigma_t^2$$

$$E(b_j) = 0, E(b_j^2) = V(b_j) = \sigma_b^2$$

Quando t_i é fixo e b_j é aleatório ou quando t_i é aleatório e b_j é fixo o modelo é considerado Misto.

Quando o experimento contém em suas parcelas um número determinado de sub-parcelas ou amostras o modelo para o delineamento inteiramente casualizado é, definido como: $y_{ijk} = m + t_i + e_{ij} + e_{ijk}$, em que:

t_i é o efeito de tratamento.

e_{ij} é o efeito residual entre ou erro experimental.

e_{ijk} é o efeito residual dentro ou erro amostral.

Quando a análise é feita considerando o total ou a média das parcelas, deixa-se de lado uma informação importante que é a variação dentro da parcela. Com a variação dentro da parcela o pesquisador pode efetuar o dimensionamento da parcela que permitira dizer se o número de amostras dentro da parcela foi suficiente ou não.

Supondo: $i = 1, 2, \dots, a$
 $j = 1, 2, \dots, b$
 $k = 1, 2, \dots, c$, teremos

Causas de Variação	G.L
Tratamento	$a - 1$
Resíduo entre	$a(b - 1) = ab - a$
(Parcelas)	$(ab - 1)$
Resíduo dentro	$ab(c-1) = abc - ab$
Total	$abc - 1$

$$SQ_{\text{Trat}} = \frac{1}{bc} \sum_{i=1}^a y^2_{i..} - \frac{y^2 \dots}{abc}$$

$$SQ_{\text{Residuo entre}} = \frac{1}{c} \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{1}{bc} \sum_{i=1}^a y^2_{i..}$$

$$SQ_{\text{Residuo dentro}} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y^2_{ijk} - \frac{1}{c} \sum_{i=1}^a \sum_{j=1}^b y^2_{ij.}$$

Considerando o modelo aleatório, teremos:

$$m - \text{efeito fixo: } E(m) = m, E(m^2) = m^2$$

t_i, e_{ij}, e_{ijk} - efeitos aleatórios.

$$t_i \sim \text{NID}(0, \sigma_t^2) \quad e_{ij} \sim \text{NID}(0, \sigma_{e1}^2) \quad e_{ijk} \sim \text{NID}(0, \sigma_{e2}^2)$$

$$E(t_i) = E(e_{ij}) = E(e_{ijk}) = 0$$

$$E(t_i^2) = V(t_i) = \sigma_t^2 \quad E(e_{ij}^2) = V(e_{ij}) = \sigma_{e1}^2$$

$$E(e_{ijk}^2) = V(e_{ijk}) = \sigma_{e2}^2$$

Para a dedução das esperanças de quadrados médios de Tratamento, Resíduo entre e Resíduo dentro, consideramos os efeitos do modelo da análise não correlacionada e, desta forma, teremos:

$$E \left(\sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c y_{ijk}^2 \right) = \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c E(m + t_i + e_{ij} + e_{ijk})^2$$

$$= \sum_{i=0}^a \sum_{j=0}^b \sum_{k=0}^c [E(m^2) + E(t_i^2) + E(e_{ij}^2) + E(e_{ijk}^2) + E(dp)] \quad \therefore (dp = \text{duplos produtos})$$

$$E\left[\frac{y_{\dots}^2}{abc}\right] = \frac{1}{abc} E\left[abc m + bc \sum_{i=0}^a t_i + c \sum_{i=1}^a \sum_{j=1}^b e_{ij} + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c e_{ijk}\right]$$

$$E\left[\frac{y_{\dots}^2}{abc}\right] = \frac{1}{abc} \left[a^2 b^2 c^2 m^2 + ab^2 c^2 \sigma_t^2 + abc^2 \sigma_{e1}^2 + abc \sigma_{e2}^2 + E(dp) \right]$$

$$E\left[\frac{y_{\dots}^2}{abc}\right] = abcm^2 + bc\sigma_t^2 + c\sigma_{e1}^2 + \sigma_{e2}^2$$

De forma análoga, obtem-se:

$$E\left(\frac{\sum_{i=1}^a y_{i\dots}^2}{bc}\right) = abcm^2 + abc\sigma_t^2 + ac\sigma_{e1}^2 + I\sigma_{e2}^2$$

$$E\left(\frac{\sum_{i=1}^a \sum_{j=1}^b y_{ij\cdot}^2}{c}\right) = abcm^2 + abc\sigma_t^2 + abc\sigma_{e1}^2 + ab\sigma_{e2}^2$$

Então:

$$E(\text{SQ Tratamento}) = E\left[\sum_{i=1}^a \frac{y_{i\dots}^2}{bc} - \frac{y_{\dots}^2}{abc}\right]$$

$$E(\text{SQ Tratamento}) = bc(a-1) \sigma_t^2 + c(a-1) \sigma_{e1}^2 + (a-1) \sigma_{e2}^2$$

$$E(\text{SQ Resíduo entre}) = E\left[\sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij\cdot}^2}{c} - \sum_{i=1}^a \frac{y_{i\dots}^2}{bc}\right]$$

$$E(\text{SQ Resíduo entre}) = ac(b-1) \sigma_{e1}^2 + a(b-1) \sigma_{e2}^2$$

$$E(\text{SQ Resíduo dentro}) = E\left[\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk}^2 - \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij\cdot}^2}{c}\right]$$

$$E(\text{SQ Resíduo dentro}) = ab(c-1) \sigma_{e2}^2$$

Após a divisão pelos respectivos graus de liberdade, obtém-se o sistema linear para o cálculo dos estimadores dos componentes de variância utilizando os quadrados médios da análise estatística do modelo.

$$E(\text{QM Resíduo dentro}) = \sigma_{e2}^2$$

$$E(\text{QM Resíduo entre}) = c\sigma_{e1}^2 + \sigma_{e2}^2$$

$$E(\text{QM Tratamento}) = bc\sigma_t^2 + c\sigma_{e1}^2 + \sigma_{e2}^2$$

A variância da média geral do experimento será:

$$V(\bar{y}...) = V(\hat{m}) = \frac{\sigma_t^2}{a} + \frac{\sigma_{e1}^2}{ab} + \frac{\sigma_{e2}^2}{abc}$$

$$\hat{V}(\hat{m}) = \frac{1}{abc} (bc\hat{\sigma}_t^2 + c\hat{\sigma}_{e1}^2 + \hat{\sigma}_{e2}^2)$$

$$\hat{V}(\hat{m}) = \frac{1}{abc} \text{QM Tratamento} = \frac{1}{abc} \text{VT}$$

Com o quadro da ANOVA

Causas de Variação	G.L	QM	E(QM)
Tratamento(T)	(a -1)	VT	$\sigma_{e2}^2 + c\sigma_{e1}^2 + bc\sigma_t^2$
Resíduo Entre(RE)	a(b -1)	VRE	$\sigma_{e2}^2 + c\sigma_{e1}^2$
Resíduo Dentro(RD)	ab(c -1)	VRD	σ_{e2}^2

1) Observa-se que o resíduo apropriado para o teste da $H_0 : \sigma_t^2 = 0$ é o resíduo entre:

$$F_{\text{cal}} = \frac{\text{VT}}{\text{VRE}}$$

2) Da mesma forma, o resíduo apropriado para o teste do resíduo entre (RE):

$H_0 : \sigma_{e1}^2 = 0$ é o resíduo dentro.

$$F_{\text{cal}} = \frac{\text{VRE}}{\text{VRD}}$$

Para o mesmo modelo estatístico, com o efeito de tratamento fixo (modelo fixo) a alteração do quadro da anova acontece para tratamento.

Causas de Variação	G.L	QM	E(QM)
Tratamento(T)	(a -1)	VT	$\sigma_{e2}^2 + c\sigma_{e1}^2 + \Phi\sigma_t^2$
Resíduo Entre (RE)	a(b -1)	VRE	$\sigma_{e2}^2 + c\sigma_{e1}^2$
Resíduo Dentro (RD)	ab(c -1)	VRD	σ_{e2}^2

Na aplicação do teste F a modificação ocorre na hipótese $H_0 : t_i = 0$ (efeito fixo).

$$F_{\text{cal}} = \frac{\text{VT}}{\text{VRE}}$$

Neste caso:

$$V(\bar{y}...) = V(\hat{m}) = \frac{1}{abc} \sigma_{e2}^2 + \frac{1}{ab} \sigma_{e1}^2 = \frac{1}{abc} E(\text{QM Resíduo entre})$$

$$\hat{V}(\hat{m}) = \frac{1}{abc} \text{QMRE}$$

Como o processo dedutivo é trabalhoso, (Hicks, 1973) criou regras para a obtenção das E(QM) nos modelos balanceados.

Para explicar as regras desenvolvidas por (Hicks, 1973), vamos considerar o modelo do esquema fatorial de dois fatores, no delineamento inteiramente casualizado.

$$y_{ijk} = m + fa_i + fb_j + fab_{ij} + e_{ijk}$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, r$$

Admitindo o modelo aleatório, teremos:

$$E(m) = m$$

$$E(m^2) = m^2$$

$$E(fa_i) = 0$$

$$E(fa_i^2) = \sigma a^2$$

$$E(fb_j) = 0$$

$$E(fb_j^2) = \sigma b^2$$

$$E(fab_{ij}) = 0$$

$$E(fab_{ij}^2) = \sigma ab^2$$

$$E(e_{ijk}) = 0$$

$$E(e_{ijk}^2) = \sigma_e^2$$

Os quadros deverão ser montados na ordem seguinte:

Quadro 1 – Quadro das fontes de variação do modelo.

Fontes

fa_i

fb_j

fab_{ij}

$e_{(ij)k}^*$

(*) os índices i e j são colocados entre parênteses para indicar que a variação é dentro de ij.

Quadro 2 – complementar o quadro inicial, criando a parte superior da tabela com os índices dos efeitos do modelo, classificando-os com as letras F e A para efeitos fixos ou aleatórios e com o número de classes ou níveis de cada índice.

a	b	c
A	A	A
i	j	k

fa_i

fb_j

fab_{ij}

$e_{(ij)k}$

Quadro 3 – Em cada coluna e linha, adicionamos o número de observações na qual o índice da coluna não aparece.

a	b	c
A	A	A
i	j	k

fa_i b c

fb_j a c

fab_{ij} c

$e_{(ij)k}$

Quadro 4 – Nas linhas onde os índices aparecem entre parênteses, adicionar 1 nas colunas referentes a esses índices.

	a	b	c
A	A	A	A
	i	j	k
fa_i		b	c
fb_j	a		c
fab_{ij}			c
$e_{(ij)k}$	1	1	

Quadro 5 – Nas combinações de linhas e colunas em branco do quadro 4, adicionar 1 se o efeito for aleatório ou Φ se o efeito for fixo.

	a	b	c
A	A	A	A
	i	j	k
fa_i	1	b	c
fb_j	a	1	c
fab_{ij}	1	1	c
$e_{(ij)k}$	1	1	1

Quadro 6 – Para a obtenção da E(QM) da linha:

6.1) Cobrir as colunas correspondentes aos índices que aparecem na linha do efeito. Quando o índice aparece entre parênteses, não cobri as colunas.

6.2) Multiplicar os números de observações internos da tabela, em cada linha. Os produtos serão os coeficientes dos componentes da variância do efeito. A soma dos produtos constituem a E(QM) da respectiva linha.

Devem ser eliminadas da soma, todas as parcelas que não contenham o índice do efeito a que se refere a E(QM).

	a	b	c	
	A	A	A	E(QM)
	i	j	k	
fa_i	1	B	c	$\sigma_e^2 + c\sigma_{fab}^2 + e\sigma_{fb}^2 + bc\sigma_{fa}^2$
fb_j	a	1	c	$\sigma_e^2 + e\sigma_{fa}^2 + c\sigma_{fab}^2 + ac\sigma_{fb}^2$
fab_{ij}	1	1	c	$\sigma_e^2 + e\sigma_{fa}^2 + e\sigma_{fb}^2 + c\sigma_{fab}^2$
e_{(ij)k}	1	1	1	σ_e^2

$$E(QMfa) = \sigma_e^2 + c\sigma_{fab}^2 + bc\sigma_{fa}^2$$

$$E(QMfb) = \sigma_e^2 + c\sigma_{fab}^2 + ac\sigma_{fb}^2$$

$$E(QMfab) = \sigma_e^2 + c\sigma_{fab}^2$$

$$E(QM \text{ Resíduo}) = \sigma_e^2$$

Para o mesmo modelo com todos os fatores fixo, o resultado será:

	a	b	c	
	F	F	A	
	i	j	k	
fa_i	σ^2	Bk	Φfa	
fb_j	σ^2	ak	Φfb	
fab_{ij}	σ^2	k	Φfab	
e_{(ij)k}	σ^2			

Em que:

$$\Phi_{fa} = \frac{\sum_{i=1}^a fa_i^2}{(a-1)}$$

$$\Phi_{fb} = \frac{\sum_{j=1}^b fb_j^2}{(b-1)}$$

$$\Phi_{fab} = \frac{\sum_{i=1}^a \sum_{j=1}^b fab_{ij}^2}{(a-1)(b-1)}$$

Φ_{fa} , Φ_{fb} e Φ_{fab} são funções quadráticas de cada um dos efeitos considerados na análise.

Um caso esclarecedor para a formatação do sistema linear que permitira a determinação dos componentes de variância pode ser desenvolvido considerando, por exemplo, quando 20 amostras para a análise de resíduos tóxicos são extraídas de cada uma das 4 localidades, para a análise de eficiência de 5 métodos. Cada amostra é dividida em 5 partes e cada parte é analisada segundo um dos 5 métodos. Três determinações são feitas em cada uma das 5 partes.

L_1 = efeito da localidade, $l = 1, 2, 3, \dots, l = 4$

A_i/L_1 = efeito da amostra dentro de localidade, $i = 1, 2, 3, \dots, a = 20$

M_j = efeito do método, $j = 1, 2, 3, \dots, j = 5$

LM_{lj} = efeito da interação de Local e Método.

AM_{ij}/L_1 = efeito da interação de Amostra e Método dentro de localidade.

D_k/ALM_{ij} = efeito da Determinação ($k = 1, 2, \dots, k = 3$) dentro da classificação cruzada de Amostra, Localidade e Método.

Para o modelo:

$$y = m + L_i + A_i/L_i + M_j + LM_{ij} + AM_{ij}/L_i + D_k/ALM_{ij}$$

O quadro da análise de variância para cada fonte de variação e respectivos graus de liberdade será:

Fonte de Variação	Grau de Liberdade
Localidade (L)	$(l - 1) = 4 - 1 = 3$
Amostra/Localidade (A/L)	$(a - 1)l = (20 - 1)4 = 76$
Método (M)	$(j - 1)l = (5 - 1)4 = 16$
M * L	$(j - 1)(l - 1) = 4 * 3 = 12$
(A * M) / L	$l(a - 1)(j - 1) = 4(20 - 1)(5 - 1) = 304$
Determinação(D) / (A * M / L)	$laj(k - 1) = 4(20)(5)(3 - 1) = 800$
Total Corrigido	$(lajk - 1) = 1200 - 1 = 1199$

Com as regras estabelecidas anteriormente e considerando o modelo aleatório, teremos:

Fonte de Variação	$\sigma^2 D / (AM/L)$	$k\sigma^2 AM/L$	$ak\sigma^2 LM$	$lak\sigma^2 M$	$jk\sigma^2 A/L$	$ja\sigma^2 L$
L	X	X	X	(N)	X	X
A/L	X	X	(N)	(N)	X	(N)
M	X	X	X	X	(N)	(N)
LM	X	X	X	(N)	(N)	(N)
AM / L	X	X	(N)	(N)	(N)	(N)
D / AM/L	X					

O X indica que o componente da coluna irá compor a esperança do quadrado médio da fonte de variação. Da mesma forma, o (N) indica que o componente não participará da esperança do quadrado médio da linha.

Deveremos observar ainda:

- 1) Se um determinado efeito for considerado fixo, haverá a necessidade da eliminação de algum componente na esperança do quadrado médio.
- 2) O componente da menor subclasse aparecerá em todas as esperanças.
- 3) Um componente de interação entre um efeito fixo e um aleatório não aparecerá na esperança de um efeito aleatório. A regra é definir quais os efeitos que são fixos e para qualquer componente, que não seja ao referente à menor subclasse, desconsiderar todos os subscritos à esquerda da primeira barra: Se qualquer um dos subscritos remanescentes representar efeito fixo, o componente não aparecerá na esperança daquele quadrado médio.

Neste exemplo, considerando método (M) como fixo os componentes para a esperança de quadrado médio de Localidade inicialmente definido como:

$$\sigma^2 D / (AM/L) + k\sigma^2 AM/L + ak\sigma^2 LM + jk\sigma^2 A/L + kja\sigma^2 L$$

Considerando os remanescentes, com Localidade é um efeito aleatório, a esperança para o quadrado médio Localidade será:

$$\sigma^2 D(AM/L) + jk\sigma^2 A/L + kja\sigma^2 L$$

Considerando o quadro de esperanças de quadrado médio das fontes, listado anteriormente, teremos duas modificações se o método (M) é fator fixo.

1) $k\sigma^2 AM/L$ será eliminada da esperança matemática da fonte A/L .

2) Em lugar de $lak\sigma^2 M$, aparecerá $lak\Phi M$.

$$\Phi M = \frac{\sum_{j=1}^4 M^2 j}{4-1}$$

ΦM = função quadrática.

Quando consideramos que 10 unidades amostrais são medidas em 4 classes do fator A, por meio de 3 classes do fator B e por meio de 3 classes do fator C, estaremos falando de:

$$10(4)(3)(3) = 360 \text{ dados}$$

Os fatores A, B e C apresentam efeitos cruzados segundo o modelo:

$$y_{ijkl} = m + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + ABC_{ijk} + DI/ABC_{ijk}$$

Em que:

A_i = efeito do fator A, $i = 1, 2, \dots, a = 4$.

B_j = efeito do fator B, $j = 1, 2, \dots, b = 3$.

C_k = efeito do fator C, $k = 1, 2, \dots, c = 3$.

AB_{ij} = efeito cruzado de A com B.

AC_{ik} = efeito cruzado de A com C.

BC_{jk} = efeito cruzado de B com C.

ABC_{ijk} = efeito cruzado de A, com B e C.

$Dl/ABC_{ijk} =$ efeito da amostra dentro das combinações de ABC, $l = 1, 2, \dots, d = 10$.

Fonte de Variação	Grau de Liberdade
A_i	$(a - 1) = 4 - 1 = 3$
B_j	$(b - 1) = 3 - 1 = 2$
AB_{ij}	$(a - 1)(b - 1) = 6$
C_k	$(c - 1) = (3 - 1) = 2$
AC_{ik}	$(a - 1)(c - 1) = 6$
BC_{jk}	$(b - 1)(c - 1) = 4$
ABC_{ijk}	$(a - 1)(b - 1)(c - 1) = 12$
Dl/ABC_{ijk}	$abc(d - 1) = 324$
Total Corrigido	$abcd - 1 = 359$

Para o modelo aleatório.

$$y_{ijkl} = m + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + Dl/ABC_{ijk}$$

ESPERANÇAS QUADRADO MÉDIO

Fonte de Variação	$\sigma^2 D/ABC$	$10\sigma^2 ABC$	$40\sigma^2 BC$	$40\sigma^2 AC$	$120\sigma^2 C$	$40\sigma^2 AB$	$120\sigma^2 B$	$90\sigma^2 A$
A	X	X	(N)	X	(N)	X	(N)	X
B	X	X	X	(N)	(N)	X	X	(N)
AB	X	X	(N)	(N)	(N)	X	(N)	(N)
C	X	X	X	X	X	(N)	(N)	(N)
AC	X	X	(N)	X	(N)	(N)	(N)	(N)

BC	X	X	X	(N)	(N)	(N)	(N)	(N)
ABC	X	X	(N)	(N)	(N)	(N)	(N)	(N)
D/ABC	X							

Considerando B e C como fatores fixos, os componentes que serão eliminados nas esperanças de quadrados médios da várias fontes estão circulados no quadro abaixo:

Fonte de Variação	$\sigma^2 D/ABC$	$10\sigma^2 ABC$	$40\sigma^2 BC$	$40\sigma^2 AC$	$120\sigma^2 C$	$40\sigma^2 AB$	$120\sigma^2 B$	$90\sigma^2 A$
A	X	⊗	(N)	⊗	(N)	⊗	(N)	X
•B	X	⊗	⊗	(N)	(N)	X	X	(N)
AB	X	⊗	(N)	(N)	(N)	X	(N)	(N)
•C	X	⊗	⊗	X	X	(N)	(N)	(N)
AC	X	⊗	(N)	X	(N)	(N)	(N)	(N)
BC	X	X	X	(N)	(N)	(N)	(N)	(N)
ABC	X	X	(N)	(N)	(N)	(N)	(N)	(N)
D/ABC	X							

$$E(QM D/ABC) = \sigma^2 D/ABC$$

$$VE(QM ABC) = \sigma^2 D/ABC + 10\sigma^2 ABC$$

$$VE(QM •BC) = \sigma^2 D/ABC + 10\sigma^2 ABC + 40\sigma^2 BC$$

$$VE(QM AC) = \sigma^2 D/ABC + 40\sigma^2 AC$$

$$E(QM •C) = \sigma^2 D/ABC + 40\sigma^2 AC + 120\sigma^2 C$$

$$VE(QM AB) = \sigma^2 D/ABC + 40\sigma^2 AB$$

$$VE(QM •B) = \sigma^2 D/ABC + 40\sigma^2 AB + 120\sigma^2 B$$

$$VE(QM A) = \sigma^2 D/ABC + 90\sigma^2 A$$

Desta forma, o quadrado médio $\hat{\sigma}^2_{D/ABC}$ vai testar as fontes ABC, AC, AB, e A, O quadrado médio da fonte AB servirá para testar a fonte B. A fonte BC será tratada com o Quadrado Médio da fonte ABC, enquanto a fonte C será testada com o quadrado médio da fonte AC, com os respectivos graus de liberdade.

A obtenção dos componentes de variância, com muitos critérios de classificação e com classes e subclasses apresentando números diferentes de observações (desbalanceamento) é feita utilizando o algoritmo desenvolvido por L.H.Waddell e é um processo difícil e demorado. Desta forma, não comentaremos a obtenção dos componentes de variância, no caso de desbalanceamento, muito comum em análise de dados no melhoramento animal.

4. RESULTADOS E DISCUSSÃO

Os dados para a discussão sobre componentes de variância serão simulados com o objetivo de discutir vários casos e modelos de delineamento. Iniciaremos de um caso mais simples e comum.

4.1 - Fatorial no delineamento inteiramente casualizado – 2 fatores.

Considerando o modelo estatístico:

$$y_{ijk} = m + A_i + B_j + AB_{ij} + e_{ijk}, \text{ em que:}$$

$$i = 1, 2, 3, \dots, a$$

Fixo

$$\sum_{i=1}^a A_i = 0$$

$$\sum_{j=1}^b B_j = 0$$

$$\sum_{i=1}^a \sum_{j=1}^b A_i B_j = 0$$

$$j = 1, 2, 3, \dots, b$$

Aleatório

$$A_i \in \text{NID}(0, \sigma_A^2)$$

$$B_j \in \text{NID}(0, \sigma_B^2)$$

$$AB_{ij} \in \text{NID}(0, \sigma_{AB}^2)$$

$$k = 1, 2, 3, \dots, r$$

Misto

$$\sum_{i=1}^a A_i = 0$$

$$B_j \in \text{NID}(0, \sigma_B^2)$$

$$AB_{ij} \in \text{NID}(0, \sigma_{AB}^2)$$

Fator A com (a - 1) graus de liberdade:

$$SQA = \sum_{i=1}^a \frac{y^2_{i.}}{br} - \frac{y^2_{...}}{abr}$$

Fator B com $(b - 1)$ graus de liberdade:

$$SQB = \sum_{j=1}^b \frac{y^2 \cdot j}{ar} - \frac{y^2 \dots}{abr}$$

Interação AB com $(a - 1)(b - 1) = (ab - a - b + 1)$ graus de liberdade:

$$SQAB = \sum_{i=1}^a \sum_{j=1}^b \frac{y^2 ij}{r} - \sum_{i=1}^a \frac{y^2 \dots}{br} - \sum_{j=1}^b \frac{y^2 \cdot j}{ar} + \frac{y^2 \dots}{abr}$$

Resíduo $(abr - 1) - (a - 1) - (b - 1) = (a - 1)(b - 1) = abr - ab$

$$SQ \text{ Resíduo} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r y^2 ijk - \sum_{i=1}^a \sum_{j=1}^b \frac{y^2 ij}{r}$$

Por dedução ou com as regras de Hicks (1973) para experimentos balanceados obtém-se o quadro da ANOVA:

Causas de Variação	G.L	SQ	E(QM)
A	$(a - 1)$	SQA	$\sigma^2 + r\sigma_{AB}^2 + rb\Phi_A$
B	$(b - 1)$	SQB	$\sigma^2 + r\sigma_{AB}^2 + ar\sigma_B^2$
AB	$(ab - a - b + 1)$	SQAB	$\sigma^2 + r\sigma_{AB}^2$
Resíduo	$(abr - ab)$	SQResíduo	σ^2

Hipóteses Ho: Ho1: $A_i = 0$, para todo $i = 1, 2, \dots, a$

$$Ho2: \sigma_B^2 = 0$$

$$Ho3: \sigma_{AB}^2 = 0$$

i) O resíduo apropriado para as H01 e H02 é o Quadrado Médio da Interação AB com $(ab - a - b + 1)$ g.l.

ii) O resíduo apropriado para a H03 é o Quadrado Médio do Resíduo, com $(abr - ab)$ g.l.

Quadro de Dados

A ₁				A ₂			
B ₁	B ₂	B ₃	B ₄	B ₁	B ₂	B ₃	B ₄
2,3	4,8	5,3	6,2	3,2	6,2	8,2	5,0
2,5	5,2	5,8	6,9	3,8	6,8	9,0	6,2

Quadro ANOVA

Causas de Variação	G.L	SQ	QM	F
A	1	5,52250	5,52250	2,04
B	3	37,48250	12,49417	4,60
AB	3	8,14250	2,71417	11,61
Resíduo	8	1,87000	0,23375	

$$\hat{\sigma}_e^2 = 0,23375$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AB}^2 = 2,71417$$

$$\hat{\sigma}_{AB}^2 = 1,24021$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AB}^2 + 4\hat{\sigma}_B^2 = 12,49417$$

$$\hat{\sigma}_B^2 = 2,44500$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AB}^2 + 8\Phi_A = 5,52250$$

$$\hat{\Phi}_A = 0,254337$$

$$\hat{\Phi}_A = \frac{\sum_{i=1}^a \hat{A}_i^2}{(a-1)}$$

$$\sum_{i=1}^2 \hat{A}_i = (2-1)\hat{\Phi}_A = 0,254337$$

$$\sum_{i=1}^2 \hat{A}_i = 0,254337, \text{ em que } \Phi_A \text{ é uma função quadrática.}$$

Obs: As estimativas de $\sigma_e^2, \sigma_B^2, \sigma_{AB}^2, \Phi_A$ são apresentadas na unidade dos dados ao quadrado.

Os estimadores de componentes de variância são não tendenciosos ($E(\sigma^2) = \sigma^2$) e como podemos obter $\hat{V}(\hat{\sigma}^2)$, podemos obter o intervalo de confiança para σ^2 ao nível α de probabilidade. Também $\hat{V}(\hat{\sigma}^2)$ é estimador não tendencioso:

$$E(\hat{V}(\hat{\sigma}^2)) = V(\hat{\sigma}^2)$$

Um efeito considerado fixo tem as conclusões aplicadas apenas ao material experimental, não constitui uma amostra de uma população. Já o efeito aleatório, como é uma amostra de uma população, as conclusões vão caracterizar a população de trabalho, se a amostra é estatisticamente representativa da população.

O mesmo exemplo foi resolvido com o programa Genes, Aplicativo Computacional em Genética e Estatística, desenvolvido na Universidade Federal de Viçosa. O resumo da saída é apresentado abaixo e podemos listar as alterações que deverão ser editadas:

- 1) O fator A, por se tratar de um programa apropriado ao melhoramento genético vegetal, é denominado Ambiente.
- 2) O fator B é denominado Tratamento e a Interação AB por Trat X AMB.
- 3) Os valores de somas de quadrado estão corretamente calculados, inclusive o teste F apresentando valores calculados com o resíduo correto para os testes Ho1, Ho2 e Ho3.
- 4) $\hat{\sigma}_B^2$ = Componente de variância genotípica = 2,44500.
- 5) $\hat{\sigma}_{AB}^2$ = Componente de variância genótipo x ambiente = 1,24021
- 6) $\hat{\sigma}_e^2$ = Variância residual = 0,23375
- 7) Não apresenta a estimativa $\hat{\Phi}_A = 0,254337$

Resultado da Saída do Programa GENES

Arquivo de dados: C:\Documents and Settings\Administrador\Meus documentos\ERICA EXEMPLOS TESE\ERICA1\GENES\ERICA1.prn

Modelo : $Y_{ijk} = m + G_i + A_j + GA_{ij} + E_{ijk}$

PARÂMETROS GENÉTICOS

Componente de variância genotípica	2.445
Componente de variância genot.xambiente	1.24021
Variância residual	0.23375
Herdabilidade(média) - %	78.2765
Correlação intraclasse (*)	62.389
Coef. variação genético (%)	28.6252
Razão CVg/CVe	3.2342

4.2 - Fatorial no delineamento inteiramente casualizado – 3 fatores.

Considerando o modelo estatístico:

$$y_{ijkl} = m + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + e_{ijkl}$$

$i = 1,2$

$j = 1,2,3$

$k = 1,2,3$

$l = 1,2$

	A ₁			A ₂		
	B ₁	B ₂	B ₃	B ₁	B ₂	B ₃
C ₁	2,2	4,2	6,4	4,5	7,3	5,0
	2,3	4,7	7,0	5,0	7,9	6,0
C ₂	3,6	5,2	8,0	8,2	8,4	7,0
	4,2	5,8	8,4	8,8	8,6	6,8
C ₃	5,0	7,4	7,5	6,4	7,0	6,6
	5,4	7,8	7,2	7,2	7,6	6,8

A_1 e A_2 , bem como B_1, B_2 e B_3 , são selecionados ao acaso e o experimento em mensurar C_1, C_2 e C_3 dentro das combinações AB_{ij} e, portanto, C é fixo.

$$\begin{aligned} A_i &\in \text{NID}(0, \sigma_A^2) & B_j &\in \text{NID}(0, \sigma_B^2) & \sum_{k=1}^3 C_k &= 0 \\ AB_{ij} &\in \text{NID}(0, \sigma_{AB}^2) & AC_{ij} &\in \text{NID}(0, \sigma_{AC}^2) & BC_{jk} &\in \text{NID}(0, \sigma_{BC}^2) \\ ABC_{ijk} &\in \text{NID}(0, \sigma_{ABC}^2) & e_{ijkl} &\in \text{NID}(0, \sigma_e^2) & & \end{aligned}$$

Fator A com 1 grau de liberdade.

$$SQA = \sum_{i=1}^2 \frac{y^2 \cdot i \dots}{18} - \frac{y^2 \dots}{36}$$

Fator B com 2 graus de liberdade.

$$SQB = \sum_{j=1}^3 \frac{y^2 \cdot j \dots}{12} - \frac{y^2 \dots}{36}$$

Fator C com 2 graus de liberdade.

$$SQC = \sum_{k=1}^3 \frac{y^2 \cdot k \dots}{12} - \frac{y^2 \dots}{36}$$

Interação AB_{ij} com 2 graus de liberdade.

$$SQAB = \sum_{i=1}^2 \sum_{j=1}^3 \frac{y^2 \cdot i \cdot k \dots}{6} - \sum_{i=1}^2 \frac{y^2 \cdot i \dots}{18} - \sum_{k=1}^3 \frac{y^2 \cdot k \dots}{12} + \frac{y^2 \dots}{36}$$

Interação BC_{jk} com 4 graus de liberdade.

$$SQBC = \sum_{j=1}^3 \sum_{k=1}^3 \frac{y^2 \cdot j \cdot k \dots}{4} - \sum_{j=1}^3 \frac{y^2 \cdot j \dots}{12} - \sum_{k=1}^3 \frac{y^2 \cdot k \dots}{12} + \frac{y^2 \dots}{36}$$

$$SQTotal = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^2 y^2_{ijkl} - \frac{y^2 \dots}{36}$$

$$SQResiduo = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^2 y^2_{ijkl} - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^3 \frac{y^2 \cdot ijk \dots}{2}$$

Causa de Variação	G.L	SQ	E(QM)
A	1	SQA	$\sigma_e^2 + 6\sigma_{AB}^2 + 18\sigma_A^2$
B	2	SQB	$\sigma_e^2 + 6\sigma_{AB}^2 + 12\sigma_B^2$
AB	2	SQAB	$\sigma_e^2 + 6\sigma_{AB}^2$
C	2	SQC	$\sigma_e^2 + 2\sigma_{ABC}^2 + 6\sigma_{AC}^2 + 4\sigma_{BC}^2 + 12\Phi_e$
AC	2	SQAC	$\sigma_e^2 + 2\sigma_{ABC}^2 + 6\sigma_{AC}^2$
BC	4	SQBC	$\sigma_e^2 + 2\sigma_{ABC}^2 + 4\sigma_{BC}^2$
ABC	4	SQABC	$\sigma_e^2 + 2\sigma_{ABC}^2$
Resíduo	18	SQResíduo	σ_e^2

$$Ho1 : \sigma_A^2 = 0$$

$$Ho2 : \sigma_B^2 = 0$$

$$Ho3 : \sigma_{AB}^2 = 0$$

$$\text{Hipóteses Ho: } Ho4 : \Phi_C = 0$$

$$Ho5 : \sigma_{AC}^2 = 0$$

$$Ho6 : \sigma_{BC}^2 = 0$$

$$Ho7 : \sigma_{ABC}^2 = 0$$

- 1) O resíduo apropriado para os testes de Ho1 e Ho2 é QMAB.
- 2) O resíduo apropriado para o teste de Ho3 é o QMResíduo.
- 3) O resíduo apropriado para o teste de Ho5 e Ho6 é o QMABC.
- 4) O resíduo apropriado para o teste de Ho7 é o QMResíduo.
- 5) Não existe um resíduo apropriado para o teste de σ_C^2 e ele deverá ser composto.

$$H_{04} = E(QMAC + QMBC - QMABC)$$

$$\begin{aligned} \text{Resido ap./} &= \sigma_e^2 + 2\sigma_{ABC}^2 + 6\sigma_{AC}^2 + \sigma_e^2 + 2\sigma_{ABC}^2 + 4\sigma_{BC}^2 - \sigma_e^2 - 2\sigma_{ABC}^2 \\ &= \sigma_e^2 + 2\sigma_{ABC}^2 + 6\sigma_{AC}^2 + 4\sigma_{BC}^2, \end{aligned}$$

$$\text{Com } nH_{04} = \frac{(QMAC + QMBC - QMABC)^2}{(1)^2 \left(\frac{QMAC}{2}\right)^2 + (1)^2 \left(\frac{QMBC}{4}\right)^2 + (-1)^2 \left(\frac{QMABC}{4}\right)^2}$$

Graus de Liberdade.

6) A análise retrata o caso de uma parcela sub-dividida com os fatores A e B aleatórios e em fatorial 2x3 nas parcelas, com o fator C fixo na sub-parcela, no delineamento inteiramente casualizado com duas repetições.

Se todos os fatores forem fixos haverá um erro para testar o que está na parcela (os fatores A, B e AB), portanto com o resíduo 1, e um resíduo para o que está na sub-parcela (o fator c, e as interações AC, BC e ABC), portanto com o resíduo 2.

Quadro ANOVA

Causa de Variação	G.L	SQ	(QM)	F
A	1	14,44000	14,44000	1,113 (ns)
B	2	21,15167	10,57583	0,83 (ns)
AB	2	25,50500	12,76250	91,89 (*)
C	2	22,16167	11,08083	5,21 < F _{0,05} ^(2,2) (ns)

AC	2	5,53167	2,76583	1,57 (ns)
BC	4	4,50167	1,12542	0,64 (ns)
ABC	4	7,0583	1,76458	12,70 (*)
Resíduo	18	2,50000	0,13889	

Composição do Resíduo para teste da H_{o4}

$$\text{Resíduo}/H_{o4} = 2,76583 + 1,12542 - 1,76458$$

$$= 2,12667$$

$$nH_{o4} = \frac{(2,12667)^2}{(1)^2 \left(\frac{2,76583}{2} \right)^2 + (1)^2 \left(\frac{1,12542}{4} \right)^2 + (-1)^2 \left(\frac{1,76458}{4} \right)^2} = \frac{4,522725}{2,186223}$$

$$nH_{o4} = 2,01 \cong 2$$

$$F_{\text{cal}} = \frac{11,08083}{2,12667} = 5,21 < F_{0,05}^{(2,2)}$$

Conclusões: Ao nível de 5% de probabilidade, pelo teste F, são aceitas as hipóteses:

$H_{o1}, H_{o2}, H_{o4}, H_{o5}, H_{o6}$. Da mesma forma, são rejeitadas as hipóteses: H_{o3} e H_{o7} .

$$\hat{\sigma}_e^2 = 0,13889$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{ABC}^2 = 1,76458$$

$$\hat{\sigma}_{ABC}^2 = 0,81284$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{ABC}^2 + 4\hat{\sigma}_{BC}^2 = 1,12542$$

$$\hat{\sigma}_{BC}^2 = \frac{1,12542 - 1,76458}{4}$$

$$\hat{\sigma}_{BC}^2 = -0,15979$$

$\hat{\sigma}_{BC}^2$ como é negativo é uma estimativa de zero e, portanto, não significativa.

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{ABC}^2 + 6\hat{\sigma}_{AC}^2 = 2,76583$$

$$\hat{\sigma}_{AC}^2 = \frac{2,76583 - 1,76458}{6}$$

$$\hat{\sigma}_{AC}^2 = 0,16688$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{ABC}^2 + 6\hat{\sigma}_{AC}^2 + 4\hat{\sigma}_{BC}^2 + 12\hat{\Phi}_C = 11,08083$$

$$\hat{\Phi}_C = \frac{11,08083 - 2,12667}{12}$$

$$\hat{\Phi}_C = 0,74618$$

$$\hat{\sigma}_e^2 + 6\hat{\sigma}_{AB}^2 = 12,76250$$

$$\hat{\sigma}_{AB}^2 = 2,10393$$

$$\hat{\sigma}_e^2 + 6\hat{\sigma}_{AB}^2 + 12\hat{\sigma}_B^2 = 10,57583$$

$$\hat{\sigma}_B^2 = -0,36444$$

$\hat{\sigma}_B^2$ é também uma estimativa de zero.

$$\hat{\sigma}_e^2 + 6\hat{\sigma}_{AB}^2 + 18\hat{\sigma}_A^2 = 14,44000$$

$$\hat{\sigma}_A^2 = 0,09319$$

Uma estimativa negativa de um componente de variância (≥ 0) é indicativo de um problema amostral. O esperado é que o componente seja no mínimo igual a zero.

O programa GENES, voltado para análise de modelos lineares aplicados ao melhoramento animal, não apresenta opção para a solução do modelo conforme o que foi formulado:

3 níveis fatores A e B aleatórios e o fator C fixo, em dentro do fatorial de A com B, fatorial (2x3) no delineamento inteiramente casualizado com 2 repetições.

4.3 – Experimentos aninhados e fatorial – aninhado.

4.3.1 – Experimento com 1 fator aninhado.

A ₁				A ₂							
B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇	B ₈				
2,2	4,3	5,3	7,2	4,0	6,0	8,0	10,2				
2,8	4,9	6,0	7,9	5,0	6,6	9,0	11,0				
Y ₁₍₁₎	Y ₂₍₁₎	Y ₃₍₁₎	Y ₄₍₁₎	Y ₅₍₂₎	Y ₆₍₂₎	Y ₇₍₂₎	Y ₈₍₂₎				
Y _{1..}				Y _{2..}							
i = 1,2				j = 1,2,3,4				k = 1,2			

$$y_{ijk} = m + A_i + B_{j(i)} + e_{k(ij)}$$

Acima temos os dados de um experimento, também no delineamento inteiramente casualizado, onde A é fixo (tipo máquina) e B é aleatório (tipo operador).

$$SQ_{Total} = \sum_{i=1}^2 \sum_{j=1}^4 \sum_{k=1}^2 y_{ijk}^2 - \frac{y_{...}^2}{16} = 94,1100$$

$$SQA = \sum_{i=1}^2 \frac{y_{i..}^2}{8} - \frac{y_{...}^2}{16} = 23,0400$$

Para A = 1

$$SQB_1 = \sum_{j=1}^4 \frac{y_{j(1)}^2}{2} - \frac{y^2 \cdot (1)}{8} = 26,6250$$

Para A = 2

$$SQB_2 = \sum_{j=1}^4 \frac{y_{j(2)}^2}{2} - \frac{y^2 \cdot (2)}{8} = 42,0950$$

$$SQB/A = \frac{\quad}{68,72}$$

$$SQ \text{ Resíduo} = SQ_{\text{Total}} - SQA - SQB(A) = 2,3500$$

Quadro ANOVA

Causa de Variação	G.L	SQ	(QM)	F
Fator A	1	23,0400	23,0400	78,43
Fator B	6	68,7200	11,4533	38,99
Resíduo	8	2,3500	0,29375	
Total	15	94,1100		

Causa de Variação	G.L	E(QM)
A	1	$\sigma_e^2 + 2\sigma_B^2 + 8\Phi_A$
B	2	$\sigma_e^2 + 2\sigma_B^2$
Resíduo	18	σ_e^2
Total	15	

$$\hat{\sigma}_e^2 = 0,29375$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_B^2 = 11,4533$$

$$\hat{\sigma}_B^2 = 5,5798$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_B^2 + 8\hat{\Phi}_A = 23,0400$$

$$\hat{\Phi}_A = 1,4483$$

O programa SAEG resolve este problema quando a ANOVA é feita. O programa foi preparado para resolver problemas como este porque é semelhante a um clássico do melhoramento animal onde o fator A é Touro e o fator B aninhado em A é vaca. A variável é constituída então por uma característica medida na progênie.

Como visto, não existe um programa estatístico, nem o SAS, capaz de liberar os valores de componentes de variância quando solicitado. Temos de aprender a teoria e o método prático estabelecido por (Hicks, 1973) para sim, utilizar o programa.

É bom lembrar que estamos trabalhando com experimentos balanceados. Com o desbalanceamento, as dificuldades são aumentadas.

4.3.2 – Experimento fatorial aninhado.

$$y_{ijkl} = m + A_i + B_j + AB_{ij} + C_{K(j)} + AC_{iK(j)} + e_{l(ijk)}$$

$i = 1,2$ $j = 1,2,3$ $k = 1,2,3$, para cada j $l = 1,2$, para cada i,j,k

A e B são fatores tomados como fixos.

C é fator aleatório, amostra de uma população.

$$\sum_{i=1}^2 A_i = \sum_{j=1}^3 B_j = \sum_{i=1}^2 \sum_{j=1}^3 AB_{ij} = 0$$

$$C_{k(i)} \in \text{NID}(0, \sigma_c^2)$$

$$AC_{ik(j)} \in \text{NID}(0, \sigma_{AC}^2)$$

$$e_{l(ijk)} \in \text{NID}(0, \sigma_e^2)$$

Com o método prático de (Hicks, 1973) obtém – se as esperanças de quadrado médio do quadro abaixo:

Causa de Variação	G.L	SQ	E(QM)
A_i	1	SQA_i	$\sigma_e^2 + 2\sigma_{AC}^2 + 18\Phi_A$
B_j	2	SQB_j	$\sigma_e^2 + 4\sigma_c^2 + 12\Phi_B$
AB_{ij}	2	$SQAB_{ij}$	$\sigma_e^2 + 2\sigma_{AC}^2 + 6\Phi_{AB}$
$C_{k(i)}$	2	$SQC_{k(i)}$	$\sigma_e^2 + 4\sigma_c^2$
$AC_{ik(j)}$	2	$SQAC_{ik(j)}$	$\sigma_e^2 + 2\sigma_{AC}^2$
Resíduo	18	$SQe_{l(ijk)}$	σ_e^2
Total Corrigido	35		

1) O resíduo apropriado para testar o fator A é o quadrado médio da interação

$$AC_{ik(j)}$$

$$H_0 : \Phi_A = 0$$

2) O resíduo apropriado para testar o fator B, também fixo, é o quadrado médio do fator $C_{k(i)}$.

$$H_0 : \Phi_B = 0$$

3) O resíduo apropriado para testar o fator AB_{ij} também é o quadrado médio da interação $AC_{ik(j)}$

$$H_0 : \Phi_{AB} = 0$$

4) O resíduo apropriado para testar $C_{k(i)}$ e $AC_{ik(j)}$ é o quadrado médio do resíduo.

$$H_0 : \sigma_e^2 = 0$$

$$H_0 : \sigma_{AC}^2 = 0$$

	B ₁			B ₂			B ₃		
	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉
A ₁	2,2	4,3	3,6	3,8	4,0	5,0	3,8	3,0	2,8
A ₁	2,5	5,0	3,8	4,0	3,7	3,8	4,2	3,4	2,7
A ₂	1,2	2,3	1,0	1,4	1,3	1,0	1,4	1,0	1,3
A ₂	1,4	2,8	0,8	1,6	1,5	1,3	1,6	1,2	1,4

Quadro ANOVA

Causa de Variação	G.L	SQ	QM	F
A	1	44,6669	44,6669	102,03 *
B	2	0,9172	0,4586	0,35 ns
AB	2	1,0039	0,5019	1,15 ns
C	6	7,9467	1,3244	15,33*
AC	6	2,6267	0,4378	5,07 *
Resíduo	18	1,5550	0,0864	

* Significativo pelo teste F, ao nível de 5% de probabilidade.

$$\hat{\sigma}_e^2 = 0,0864$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AC}^2 = 0,4378$$

$$\hat{\sigma}_{AC}^2 = 0,1757$$

$$\hat{\sigma}_e^2 + 4\hat{\sigma}_C^2 = 1,3244$$

$$\hat{\sigma}_C^2 = 0,3095$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AC}^2 + 6\hat{\Phi}_{AB} = 0,5019$$

$$0,4378 + 6\hat{\Phi}_{AB} = 0,5019$$

$$\hat{\Phi}_{AB} = 0,0107$$

$$\hat{\sigma}_e^2 + 4\hat{\sigma}_C^2 + 12\hat{\Phi}_B = 0,4586$$

$$1,3244 + 12\hat{\Phi}_B = 0,4586$$

$$\hat{\Phi}_B = 0,0722$$

$$\hat{\sigma}_e^2 + 2\hat{\sigma}_{AC}^2 + 18\hat{\Phi}_A = 44,6669$$

$$0,4378 + 18\hat{\Phi}_A = 44,6669$$

$$\hat{\Phi}_A = 2,4572$$

Conclusão: Para $\hat{\Phi}_A, \sigma_e^2$ e σ_{AC}^2 rejeita - se H_0 ao nível de 5% de probabilidade.

4.4 - Parcela sub-dividida (split-plot)

$$y_{ijklm} = m + R_l + A_i + RA_{il} + B_j + RB_{jl} + AB_{ij} + RAB_{ijl} + C_k + RC_{kl} + AC_{ik} + RAC_{ikl} + BC_{jk} + RBC_{jkl} + ABC_{ijk} + RABC_{ijkl} + \sigma_e^2$$

A_i, B_j são fatores fixos: $i = 1, 2, \dots, a$.

$$j = 1, 2, \dots, b.$$

$$\sum_{i=1}^a A_i = \sum_{j=1}^b B_j = \sum_{i=1}^a \sum_{j=1}^b AB_{ij} = 0$$

$C_k, R_l \in E_{m(ijkl)}$ são aleatórios: $k = 1, 2, \dots, c$

$$l = 1, 2, \dots, d$$

$$m = 1, 2, \dots, e$$

$$C_k \in \text{NID}(0, \sigma_e^2)$$

$$R_l \in \text{NID}(0, \sigma_R^2)$$

$$E_{m(ijkl)} \in \text{NID}(0, \sigma_e^2)$$

$$AC_{ik} \in \text{NID}(0, \sigma_{AC}^2)$$

$$BC_{jk} \in \text{NID}(0, \sigma_{BC}^2)$$

$$RC_{lk} \in \text{NID}(0, \sigma_{RC}^2)$$

$$ABC_{ijk} \in \text{NID}(0, \sigma_{ABC}^2)$$

$$RAB \in \text{NID}(0, \sigma_{RAB}^2)$$

$$RAC \in \text{NID}(0, \sigma_{RAC}^2)$$

$$RBC \in \text{NID}(0, \sigma_{RBC}^2)$$

$$RABC \in \text{NID}(0, \sigma_{RABC}^2)$$

$$E_{m(ijkl)} \in \text{NID}(0, \sigma_e^2)$$

Quando $a = b = c = d = 2$ $l = 4$ $n = 2^4 * 4 = 64$ observações.

Quadro ANOVA

Causa de Variação	G.L	SQ	E(QM)
R	1	SQR	$\sigma_e^2 + 16\sigma_{RC}^2 + 32\sigma_R^2$
A	1	SQA	$\sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{AC}^2 + 16\sigma_{RA}^2 + 32\Phi_A$
RA	1	SQRA	$\sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{RA}^2$
B	1	SQB	$\sigma_e^2 + 8\sigma_{RBC}^2 + 16\sigma_{BC}^2 + 16\sigma_{RB}^2 + 32\Phi_B$
RB	1	SQRB	$\sigma_e^2 + 8\sigma_{RBC}^2 + 16\sigma_{RB}^2$
AB	1	SQAB	$\sigma_e^2 + 4\sigma_{RABC}^2 + 8\sigma_{ABC}^2 + 16\Phi_{AB}$
RAB	1	SQRAB	$\sigma_e^2 + 4\sigma_{RABC}^2 + 8\sigma_{RAB}^2$
PARCELA	(7)		
C	1	SQC	$\sigma_e^2 + 16\sigma_{RC}^2 + 32\sigma_C^2$
RC	1	[SQRC	$\sigma_e^2 + 16\sigma_{RC}^2$
AC	1	SQAC	$\sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{AC}^2$
RAC	1	SQRAC	$\sigma_e^2 + 8\sigma_{RAC}^2$
BC	1	SQBC	$\sigma_e^2 + 8\sigma_{RBC}^2 + 16\sigma_{BC}^2$
RBC	1	SQRBC	$\sigma_e^2 + 8\sigma_{RBC}^2$
ABC	1	SQABC	$\sigma_e^2 + 4\sigma_{RABC}^2 + 8\sigma_{ABC}^2$

RABC	1	SQRABC	$\sigma_e^2 + 4\sigma_{RABC}^2$
Em(ijkl)	48	SQRESIDUO	σ_e^2
<hr/>			
Total Corrigido	63		

1) O quadrado médio de RC serve de resíduo para os efeitos aleatórios R_i e C_k .

2) Para testar o quadrado médio do fator fixo A_i o resíduo deverá ser composto.

$$E(QMA) = \sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{AC}^2 + 16\sigma_{RA}^2 + 32\Phi_A$$

$$E(QMRA) = \sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{RA}^2$$

$$E(QMAC) = \sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{AC}^2$$

$$E(QMRAC) = \sigma_e^2 + 8\sigma_{RAC}^2$$

$$E(QMRA) + E(QMAC) - E(QMRAC) = \sigma_e^2 + 8\sigma_{RAC}^2 + 16\sigma_{AC}^2 + 16\sigma_{RA}^2$$

Com

$$n'_A = \frac{[E(QMRA) + E(QMAC) - E(QMRAC)]^2}{(1)^2 \frac{[E(QMRA)]^2}{gl \ RA} + (1)^2 \frac{[E(QMAC)]^2}{gl \ AC} + (-1)^2 \frac{[E(QMRAC)]^2}{gl \ RAC}} \quad g.l.$$

$$F_{cal}(A) = \frac{QMA}{QMRA + QMAC - QMRAC}$$

$$F_{tab}(A) = F_{\alpha}(1, n'_A \quad gl)$$

3) Para testar o quadrado médio da interação RA , o resíduo apropriado é o quadrado médio de RAC .

4) Para testar o quadrado médio do fator fixo B_j , da mesma forma que em 2, o resíduo deverá ser composto.

$$E(QMB) = \sigma_e^2 + 8\sigma_{RBC}^2 + 16\sigma_{BC}^2 + 16\sigma_{RB}^2 + 32\Phi_B$$

$$E(QMRB) + E(QMBC) - E(QMRBC) = \sigma_e^2 + 8\sigma_{BC}^2 + 16\sigma_{RB}^2$$

$$n'_B = \frac{[E(QMRB) + E(QMBC) - E(QMRBC)]^2}{(1)^2 \frac{[E(QMRB)]^2}{gl \ AB} + (1)^2 \frac{[EQMBC]^2}{gl \ BC} + (-1)^2 \frac{[E(QMRBC)]^2}{gl \ RBC}}$$

$$F_{cal}(B) = \frac{QMA}{QMRB + QMBC - QMRBC}$$

$$F_{tab}(B) = F_{\alpha}(1, n'_B, gl)$$

5) O resíduo para testar o QMRB é o quadrado médio de RBC .

6) Para testar a causa de variação AB o resíduo apropriado é o quadrado médio de RAB .

7) Para testar a $H_0 : \sigma_{RAB}^2 = 0$, o resíduo apropriado é o quadrado médio de RABC .

8) Para testar a $H_0 : \sigma_e^2 = 0$, o resíduo é o quadrado médio de RC .

9) $\sigma_{RC}^2 = 0$ e testado com o resíduo puro $\hat{\sigma}_e^2$, com 48 g.l.

10) Também com o resíduo $\hat{\sigma}_e^2$, são testadas as:

$$H_0 : \sigma_{RAC}^2 = 0$$

$$H_0 : \sigma_{RBC}^2 = 0$$

$$H_0 : \sigma_{RABC}^2 = 0$$

11) O resíduo apropriado para testar $H_0 : \sigma_{AC}^2 = 0$ é o QMRAC .

12) O resíduo apropriado para testar $H_0 : \sigma_{BC}^2 = 0$ é o QMRBC .

13) O resíduo apropriado para o teste de $H_0 : \sigma_{ABC}^2 = 0$ é o QMRABC .

Para o modelo:

$$y_{ijkl} = m + R_l + A_i + RA_{il} + B_j + RB_{jl} + AB_{ij} + RAB_{ijl} + C_k + RC_{kl} + AC_{ik} + RAC_{ikl} \\ + BC_{jk} + RBC_{jkl} + ABC_{ijk} + RABC_{ijkl}$$

$$l = 1, 2, 3, 4$$

$$i = 1, 2, 3$$

$$j = 1, 2, 3$$

$$k = 1, 2, 3$$

Se os fatores A, B e C são fixos e o fator R é aleatório, o quadro da ANOVA será dado como:

Quadro ANOVA

Causa de Variação	G.L	E(QM)
R	3	$\sigma_e^2 + 27\sigma_R^2$
A	2	$\sigma_e^2 + 9\sigma_{RI}^2 + 36\Phi_I$
RA	6	$\sigma_e^2 + 9\sigma_{RI}^2$
B	2	$\sigma_e^2 + 9\sigma_{RB}^2 + 36\Phi_t$
RB	6	$\sigma_e^2 + 9\sigma_{RB}^2$
AB	4	$\sigma_e^2 + 3\sigma_{RAB}^2 + 12\Phi_{AB}$
RAB	12	$\sigma_e^2 + 3\sigma_{RAB}^2$
C	2	$\sigma_e^2 + 9\sigma_{RC}^2 + 36\Phi_c$
RC	6	$\sigma_e^2 + 9\sigma_{RC}^2$
AC	4	$\sigma_e^2 + 3\sigma_{RAC}^2 + 12\Phi_{AC}$
RAC	12	$\sigma_e^2 + 3\sigma_{RLM}^2$
BC	4	$\sigma_e^2 + 3\sigma_{RBC}^2 + 12\Phi_{BC}$
RBC	12	$\sigma_e^2 + 3\sigma_{RBC}^2$
ABC	8	$\sigma_e^2 + \sigma_{RABC}^2 + 4\Phi_{ABC}$

RABC	24	$\sigma_e^2 + \sigma_{RABC}^2$
$\hat{\sigma}_e^2$	0	σ_e^2 (não estimável)
<hr/>		
Total Corrigido	107	

Com a formatação de quadro anterior, podemos observar que não existe um resíduo apropriado para testar o efeito de repetição e nem as interações de repetição com os efeitos tomados como fixos.

Já para os efeitos principais e de interação entre os fatores tomados como fixos, sempre vamos poder definir um resíduo apropriado para o teste das estimativas das formas quadráticas.

4.5 – Programas genes

O GENES é um programa apropriado para análise de modelos lineares de experimentos com fatores fixos e aleatórios, inclusive modelos mistos, quando os dados estão balanceados e o planejamento é feito considerando as opções existentes no programa.

4.5.1 – Delineamento inteiramente casualizado

Modelo : $y_{ij} = m + t_i + e_{ij}$

t_i fixo, para $i = 1, 2, \dots, a$

$$\sum_{i=1}^a t_i = 0$$

t_i aleatório : $t_i \in \text{NID}(0, \sigma_t^2)$

Obs: Analisa dados quando o experimento está desbalanceado. Os resultados obtidos são os mesmos da aplicação do método desenvolvido por Hicks.

4.5.2 – Delineamento em blocos casualizados completos

Modelo : $y_{ij} = m + t_i + bl_j + e_{ij}$

bl_j sempre aleatório, para $j = 1, 2, \dots, b$.

$$bl_j \in \text{NID}(0, \sigma_b^2)$$

t_i aleatório ou fixo e portanto:

$$\sum_{i=1}^a t_i = 0 \quad \text{ou} \quad t_i \in \text{NID}(0, \sigma_t^2)$$

4.5.3 – Esquema fatorial completo em ambos os delineamentos mencionados anteriormente, quando os dados estão balanceados.

$$y_{ijk} = m + A_i + B_j + AB_{ij} + e_{ijk}$$

Modelos: $y_{ijk} = m + bl_k + A_i + B_j + AB_{ij} + e_{ijk}$

$$y_{ijk} = m + bl_{k/B_j} + A_i + B_j + AB_{ij} + e_{ijk}$$

Quando bl_k é sempre aleatório.

$$bl_k \in \text{NID}(0, \sigma_{bl}^2)$$

Os fatores A_i e B_j podem ser considerados fixos (modelo fixo), aleatórios (modelo aleatório) ou A fixo e B aleatório ou A aleatório e B fixo (Modelos Mistos).

4.5.4 – Esquema em parcela sub-divididas em ambos os delineamentos mencionados anteriormente, quando os dados estão balanceados.

$$y_{ijk} = m + A_i + e_{ik} + B_j + AB_{ij} + e_{ijk}$$

Modelos: $y_{ijk} = m + A_i + bl_k + e_{ik} + B_j + AB_{ij} + e_{ijk}$

$$y_{ijk} = m + A_i + e_{ik} + B_j + e_{jk} + AB_{ij} + e_{ijk}$$

$$y_{ijk} = m + A_i + bl_k + e_{ik} + B_j + e_{ik} + AB_{ij} + e_{ijk}$$

Os modelos mencionados acima podem ser aleatórios e mistos com apenas A_i fixo, apenas B_j fixo ou A_i, B_j e AB_{ij} fixos. A última opção A_i, B_j e AB_{ij} é um modelo fixo. Em todos os casos a repetição usa sempre aleatória.

Um outro modelo importante analisado com o GENES é o que denominam de experimento em blocos casualizados com informações de procedência / progênie e planta. Claro que o modelo pode ser adaptado para uma outra situação que não seja a análise de dados genéticos. Basta que no arquivo de dados, as quatro primeiras colunas se destinem ordenadamente aos blocos, procedência, progênie e plantas ou indivíduos. As colunas posteriores serão

destinadas às variáveis aleatórias que serão analisadas. O número de progênies pode ser variável. Assim:

Procedência 1 => progênies 1, 2, ... , p1.

Procedência 2 => progênies p1+1, p1 +2, ... , p1+p2.

Procedência 3 => progênies p1+p2+1, p1+p2+2, ... , (p1+p2+p3).

.

.

.

Procedência K => progênies pk-1+2, ... , p1+p2+ ... +pk-1+pk.

O Modelo considerado na análise é o:

$$y_{ijkl} = m + P_i + G/P_{ij} + bl_k + e_{ijk} + e_{ijkl}$$

ou

$$y_{ijkl} = m + A_i + B_j / A_i + bl_k + e_{ijk} + e_{ijkl}$$

K = 1, 2, ... , b blocos.

i = 1, 2, ... , a.

j = conforme especificado para procedência.

Também o número de plantas ou indivíduos(l) pode estar desbalanceado.

$$E(\text{QMBLOCOS}) = \sigma_e^2 + n\sigma_{e1}^2 + (ng = na)\sigma_{b1}^2$$

$$E(\text{QMA}) = \sigma_e^2 + n\sigma_{e1}^2 + nr\Phi_A$$

$r = n^\circ$ blocos

$p = n^\circ$ procedências

$$E(\text{QMB}) = \sigma_e^2 + n\sigma_{e1}^2 + nr\sigma_B^2$$

$$E(\text{QMB/A1}) = \sigma_e^2 + n\sigma_{e1}^2 + nr\sigma_{B1}^2$$

$$E(\text{QMB/A2}) = \sigma_e^2 + n\sigma_{e1}^2 + nr\sigma_{B2}^2$$

.

.

.

$$E(\text{QMB/Ap}) = \sigma_e^2 + n\sigma_{e1}^2 + nr\sigma_{Ba}^2$$

$$E(\text{QMRESÍDUO1}) = \sigma_e^2 + n\sigma_{e1}^2$$

$$E(\text{QMRESIDUO}) = \sigma_e^2$$

4.6 – Programa SAS

Para a análise dos modelos lineares mistos, dois objetivos são importantes:

- 1) Estimação e estabelecimento dos testes de hipóteses para os fatores tomados como fixos.
- 2) Predição dos efeitos aleatórios e estimacão dos componentes de variância.

Vimos que quando os modelos lineares mistos estão balanceados o problema pode ser resolvido sem maiores dificuldades. No entanto, quando o modelo linear misto está desbalanceado, a estimacão dos componentes de variância, de fundamental importância, depende da estruturação da matriz de covariância e do método de estimacão utilizado. Neste caso, um programa como o SAS é de fundamental importância, o que vai exigir para a análise correta um grande volume de conhecimento teórico.

Os modelos lineares mistos são utilizados para a interpretação de dados onde à estrutura de tratamentos é constituída de fatores que são fixos e de fatores que são aleatórios. Então, a análise de um modelo misto consiste em duas fases: a análise de uma parte que é fixa e a análise da outra porção que é aleatória.

A predição dos efeitos aleatórios é feita na presença de efeitos fixos, conforme uma análise seqüencial. A análise da parte fixa do modelo linear consiste na estimação e testes de hipóteses sobre as funções lineares estimáveis dos efeitos fixos.

Os modelos mistos foram estudados inicialmente por (Fischer, 1918), com grande impacto nos estudos genéticos e o modelo foi inicialmente denominado de modelo de componentes de variância.

A partir daí, diversos métodos de estimação dos componentes de variância tem sido proposto.

(Hartley & Rao, 1967) apresentaram o método da máxima verossimilhança (Maximum Likelihood – ML). O método da estimação quadrática não –viesada de variância mínima (Minimum Variance Quadratic Unbiased estimation – MIVQUE) foi descrito em 1971 por Rao. Em 1971, (Patterson & Thompson, 1971) desenvolveram o método da máxima verossimilhança restrita (Restricted Maximum Likelihood – REML). Todos os métodos de estimação do componentes de variância listados acima estão disponíveis no PROC MIXED do SAS.

Segundo (Littel et al., 1996) o PROC MIXED é o procedimento do SAS apropriado para a análise dos modelos mistos com desbalanceamento, por distinguir os efeitos fixos dos efeitos aleatórios.

O SAS permite também a identificação de quatro tipos de somas de quadrados (Tipo I, II, III e IV). Quando os dados são balanceados, todos os tipos de soma de quadrados do tipo ss3, mesmo assim, são as mais apropriadas para os dados desbalanceados.

Aconselha-se também a obtenção dos componentes de variância pelo método REML. A preferência pelo método decorre de suas propriedades estatísticas, superiores as propriedades dos estimadores de quadrados mínimos e máxima verossimilhança (SEARLE et al., 1992)

4.6.1 – Fatorial Bloco Casualizados

Fatores A e B fixo.

Repetição Aleatório.

Balanceado

data ex5;

input rep a b y @@;

datalines;

1 1 1 56 1 1 2 35.3

1 2 1 50 1 2 2 36

1 3 1 39 1 3 2 35

2 1 1 30 2 1 2 25

2 2 1 36 2 2 2 28

2 3 1 33 2 3 2 30

3 1 1 32 3 1 2 24

3 2 1 32.6 3 2 2 27

3 3 1 15 3 3 2 19

4 1 1 30 4 1 2 25

4 2 1 35 4 2 2 30

4 3 1 17 4 3 2 18

;

proc mixed;

class a b rep;

model y=a b a*b;

random rep a*rep;

run;

The Mixed Procedure

Model Information

Data Set	WORK.EX5
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
a	3	1 2 3
b	2	1 2
rep	4	1 2 3 4

Dimensions

Covariance Parameters	3
-----------------------	---

Columns in X	12
Columns in Z	16
Subjects	1
Max Obs Per Subject	24
Observations Used	24
Observations Not Used	0
Total Observations	24

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	138.18502184	
1	1	122.45118792	0.00000000

Convergence criteria met.

Covariance Parameter

Estimates

Cov Parm Estimate

rep	54.8751
a*rep	10.2915

Residual 14.4299
 The SAS System 09:54 Monday, June 21, 2011 2

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood	122.5
AIC (smaller is better)	128.5
AICC (smaller is better)	130.2
BIC (smaller is better)	126.6

Type 3 Tests of Fixed Effects

Effect	Num Den		F Value	Pr > F
	DF	DF		
a	2	6	4.54	0.0629
b	1	9	15.51	0.0034
a*b	2	9	3.35	0.0818

4.6.2 – Fatorial Bloco Casualizados

Fatores A e B fixo.

Repetição Aleatório

data ex5;

```
input rep a b y @@;
datalines;
1 1 1 56 1 1 2 .
1 2 1 50 1 2 2 36
1 3 1 39 1 3 2 35
2 1 1 30 2 1 2 25
2 2 1 36 2 2 2 28
2 3 1 33 2 3 2 30
3 1 1 32 3 1 2 24
3 2 1 . 3 2 2 27
3 3 1 15 3 3 2 19
4 1 1 30 4 1 2 25
4 2 1 35 4 2 2 30
4 3 1 17 4 3 2 18
;
proc mixed;
class a b rep;
model y=a b a*b;
random rep a*rep;
run;
```

The SAS System

10:09 Monday, June 21, 2011 1

The Mixed Procedure

Model Information

Data Set

WORK.EX5

Dependent Variable y
 Covariance Structure Variance Components
 Estimation Method REML
 Residual Variance Method Profile
 Fixed Effects SE Method Model-Based
 Degrees of Freedom Method Containment

Class Level Information

Class Levels Values

a 3 1 2 3

b 2 1 2

rep 4 1 2 3 4

Dimensions

Covariance Parameters	3
Columns in X	12
Columns in Z	16
Subjects	1
Max Obs Per Subject	24
Observations Used	22
Observations Not Used	2
Total Observations	24

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
-----------	-------------	-----------------	-----------

0	1	123.54683201	
1	2	110.01900658	0.02965954
2	1	108.44873153	0.01583636
3	1	107.65668613	0.00529369
4	1	107.40896496	0.00078458
5	1	107.37531094	0.00002305
6	1	107.37439331	0.00000003
7	1	107.37439232	0.00000000

Convergence criteria met.

The SAS System 10:09 Monday, June 21, 2011 2

The Mixed Procedure

Covariance Parameter

Estimates

Cov Parm Estimate

rep	66.3784
a*rep	24.0411
Residual	6.5026

Fit Statistics

-2 Res Log Likelihood	107.4
AIC (smaller is better)	113.4

AICC (smaller is better) 115.4
 BIC (smaller is better) 111.5

Type 3 Tests of Fixed Effects

Effect	Num Den		F Value	Pr > F
	DF	DF		
a	2	6	3.48	0.0993
b	1	7	21.53	0.0024
a*b	2	7	5.14	0.0423

4.6.3 – Fatorial Bloco Casualizados

Fatores A e B fixo.

Repetição Aleatório

data ex5;

input rep a b y @@;

datalines;

1 1 1 56 1 2 1 50

1 2 2 36

1 3 1 39 1 3 2 35

2 1 1 30 2 1 2 25

2 2 1 36 2 2 2 28

2 3 1 33 2 3 2 30

3 1 1 32 3 1 2 24

3 2 2 27

3 3 1 15 3 3 2 19

```
4 1 1 30 4 1 2 25
```

```
4 2 1 35 4 2 2 30
```

```
4 3 1 17 4 3 2 18
```

```
;
```

```
proc mixed;
```

```
class a b rep;
```

```
model y=a b a*b;
```

```
random rep a*rep;
```

```
run;
```

```
The SAS System      10:19 Monday, June 21, 2011  1
```

The Mixed Procedure

Model Information

Data Set	WORK.EX5
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
-------	--------	--------

a 3 1 2 3
 b 2 1 2
 rep 4 1 2 3 4

Dimensions

Covariance Parameters	3
Columns in X	12
Columns in Z	16
Subjects	1
Max Obs Per Subject	22
Observations Used	22
Observations Not Used	0
Total Observations	22

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	123.54683201	
1	2	110.01900658	0.02965954
2	1	108.44873153	0.01583636
3	1	107.65668613	0.00529369
4	1	107.40896496	0.00078458
5	1	107.37531094	0.00002305
6	1	107.37439331	0.00000003
7	1	107.37439232	0.00000000

Convergence criteria met.

The SAS System 10:19 Monday, June 21, 2011 2

The Mixed Procedure

Covariance Parameter

Estimates

Cov Parm Estimate

rep	66.3784
a*rep	24.0411
Residual	6.5026

Fit Statistics

-2 Res Log Likelihood	107.4
AIC (smaller is better)	113.4
AICC (smaller is better)	115.4
BIC (smaller is better)	111.5

Type 3 Tests of Fixed Effects

Effect	Num Den		F Value	Pr > F
	DF	DF		
a	2	6	3.48	0.0993

b	1	7	21.53	0.0024
a*b	2	7	5.14	0.0423

Quando o programa SAS estima a parcela que é equivalente a substituir o dado ponto ou quando é considerado o experimento desbalanceado os resultados são os mesmos. Verificamos esses resultados nos exercícios simulados.

5. RESUMO E CONCLUSÃO

Do estudo realizado sobre modelos mistos e componentes de variância, conclui-se que:

- 1) As regras estabelecidas para a construção do sistema de equações lineares na obtenção dos componentes de variância funcionam corretamente, quando os dados estão balanceados. Conforme as regras desenvolvidas de (Hicks, 1973) nos modelos balanceados.
- 2) Com os experimentos balanceados os tipos SS1, SS2, SS3 e SS4 apresentam os mesmos resultados nos comandos GLM e MIXED do programa SAS.
- 3) O programa GENES esta correto para os modelos formatados no pacote.
- 4) Para dados desbalanceados, sugerimos o uso do programa SAS e o PROC. MIXED, quando as causas de variação deverão ser obtidas com SS3 e os componentes de variância estimados pelo método REML, em razão de suas propriedades.

REFERÊNCIAS BIBLIOGRÁFICAS

- Analla, M. et al. (1995) Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. *Small Rum. Res.*, v. 17, p. 51-5.
- Bereskin, B., Lush, J. L. (1965) Genetic and environmental factors in dairy sire evaluation. III. Influence of environmental and others extraneous correlations among the daughters. *Journal of Dairy Science*, v. 48, n. 3, p. 356-360.
- Boldman. K.G. Freeman. A. E. (1990) Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *Journal of dairy science*. V. 73. N. 2. P. 503-512.
- Boldman, K. G. Van Vleck. L.D. (1991) Derivative-free restricted maximum likelihood estimation in animal models with a sparse matrix solver. *Journal of dairy science*. V. 74. N. 12. P. 4337-4343.
- Dickerson, G.E. (1962) Implications of genetic-environmental interactions in animal breeding. *Animal Production*, v.4, p.47-57.

- Dodenhoff, J., Swalve, H.H. (1998) Heterogeneity of variances across regions of northern Germany and adjustment in genetic evaluation. *Livest. Prod. Sci.*, 53(1):225-236p.
- Dong. M.C. Mad. (1990) i.l. Heterogeneity of (co)variance and heritability in different levels of intra herd milk production and of herd average. *Journal of dairy science*. V. 73. N. 3. P. 843-851.
- Falconer. D.S. (1952) The problem of environment and selection. *Amer. Nat.*. V. 86. N. 830. P. 293-298.
- Fisher, R. A. (1918) The correlation between relatives on the supposition of Medelian inheritance. *Transctions of the Royal Society of Edinburgh*, v.52, p.399-433.
- Foulley, J.L.; IM, S.; Gianola, D. et al. Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet. Sel. Evol.*, v.19, p.197, 1987.
- George, A.; Liu, J.W.H. (1980) User's guide for SPARSPAK: Waterloo Sparse Linear Equations Packages. Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canadá, 151p.
- Gianola. D.; Foulley, j.l. (1990) Variance estimation from integrated likelihood (veil). *Génétique, selection, evolution*. V. 22. P. 403-417.
- Golden. B.I., (1994) Additive genetic groups for animais evaluated in more than one breed association national calhe evaluation. *J. Anim. Sci.*, champaign, v.72, p.2559-67.
- Graybill, F. A. (1976) Theory and application of the linear model. Massachusetts: Duxbury Press, North Scituate, 704p.
- Hartley. H.O. Rao. J.N.K. (1967) Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*. V. 54. N. 1. P. 93-108.
- Harville. D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the american statistical association*. V. 72. N. 358. P. 320-340.
- Henderson, C. R. (1953) Estimation of variance and covariance components. *Biometrika*, washington, v.9, p.226-52.
- Hicks, C. R. (1973) Fundamental concepts in the design of experiments. 2º ed. New york, holt rinehart and winston, 349p.

- Kennedy, B.W., (1975) Comparison of genetic group and relationship methods for mixed model sire evaluation. *J. Dairy sci.*, champaign, v.58, p.1507-14,.
- Kennedy, B.W. (1988) Workshop on estimation of breeding values with the animal model. Guelph, ontario: university of guelph, 134 p.
- Kennedy, B.W. Henderson, O.R. (1988) The unfliniahod legacy. *Journal 081w soisnese*. 74.n. 11, p,4067-4081.
- Kuhlers, D.L, Kennedy, B.W. (1982) Effect of culling on selection response using phenotypic selection or best linear unbiased prediction of breeding values in small, closed herd of swine. *3. Anim. Sei.*, champaign. V.70, p. 2338 - 48.
- Kuhlers, D.L.; Kennedy, B.W. (1992) Effect of culling on selection response using phenotypic selection or best linear unbiased prediction of breeding values in small, closed herds of swine. *J. Anim. Sci.*, v.70, p.2338-2348.
- Little, R.C. et al., (1996) SAS system for mixed models. Cary, NC. Sas Institute, 633p.
- Lush, J.L. (1945) the built index problem in the light of modern genetics. *J. Dairy sci.*, champaign, v.16, p.501- 22.
- Meyer D.J. (1987) The management of cytology specimens. *Comp. Cont. Education - Small Animal* 9(1):10-17
- Meyer. K. (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Génétique, sélection, evolution*. V. 21, n. 2, p. 317-340.
- Meyer. K. (1993) DfremI- version 2.1.09: user notes. Armidate: university of new england. 97p.
- Mirande. S.I. Van Vleck. L.D.(1985) Trends in genetic and phenotypic variances for milk production. *Journal of dairy science*. V. 68. N. 9. P. 2278-2286.
- Mohammad, W.A., Lee, A.J., Grossman, M. (1982) Genotype-environment interaction in sire evaluation. *Journal of Dairy Science*, 65(5):857-860.
- Ojala, M.; Syvajarvi, J.; Hellman, T. (1984) Effect of the size of the comparison on the accuracy of a sire's evaluation. *J. Anim. Breeding*, V. 102, p. 91-94.
- Patterson, H.D., Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, v. 58, n. 2, p. 545-554.
- Quinton, M., Smith, C., Goddard, M.E. (1992) Comparison of selection methods at the sarne level of inbreeding. *J. Anim. Sci.*, champaign, v.70, p. 1060-7.

- RAO, C.R. (1970) Estimation of heteroscedastic variance in linear models. *Journal of the American Statistical Association*, **65**, p.161-72.
- Rao, C.R. (1971) Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis*, v.1,p.445-456.
- Rao. C.R. (1971b) estimation of variance and covariance components minque theory. *The journal of multivariate analysis*. V. 1, p. 257-275.
- Rao. C.R. Minimum variance quadratic unbiased estimation of variance components. *The journal of multivariate analysis*. V. 1. P. 445-456, 1971b.
- Rao, C.R. (1972) Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* v.67, p.112-115.
- Robertson et. al., (1960) Progeny testing dairy bulls at different management levels. *Anim. Prod.*, 2(2):141-152.
- Schaeffer L.R. (1978) Simultaneous estimation of variance and covariances components from multitrait mixed model equations. *Biometrics*. V. 34. N. 2. P. 199-208.
- Schaeffer L.R. (1986) Estimation of variance and covariances within the allowable parameter space. *Journal of dairy science*. V. 69. N. 1, p. 187-194.
- Schaeffer, L.R. (1993) linear models and computing strategies in animal breeding. Guelph, ontario: university of guelph, 213 p.
- Schenkel, F. S. (1991) Aplicação da metodologia de modelos mistos na avaliação genética de suínos. Porto Alegre: UFRGS, 218p. Dissertação Mestrado.
- Searle, S.R. (1971) topics in variance component estimation . *Biometrics*. V. 27, n. 1, p. 1-76,
- Searle. S.R. (1982) Matrix algebra useful to statistics . New york: j. Wiley and sons. 438p. .
- Searle S. et. al., (1992) Variance components . New York : John Willey, 501p.
- Short, R.E.; Bellows, R.A.; Staigmiller, R.B.; Berardinelli, J.G.; Custer, E.E. (1990). Physiological mechanisms controlling anestrus and infertility in postpartum beef cattle. *Journal of Animal Science*, v. 68, p. 799-816.

- Smith, S. P.; Graser, H. V. (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood . *Journal of Dairy Science*, Champaign, v.69, p.1156 – 1165 .
- Sorensen, D.A. (1988) Effect of selection index versus mixed model methods of prediction of breeding value on response to selection in a simulated pig population. *Livest. Prod. Sci.*, Amsterdam, v.20, p. 135-48 .
- Spike, P.L., Freeman, A.E. (1978). Prediction of genetic differences among herds with estimates of breeding values. *J. Dairy Sci.*, 61:1476-1482p.
- Stanton, T.L. et al., (1991) Genotype by environment interaction for Holstein milk yield in Colombia, Mexico, and Porto Rico. *J Dairy Sci*, Champaign, v. 74, n. 5, p. 1700-1714.
- Tong, A.K.W., Kennedy, B.W., Moxley, J.E. (1980). Potential errors in sire evaluation from regional genetic differences. *J. Dairy Sci.*, 63(4):627-633.
- Torres. R. A. (1998) Efeito da heterogeneidade de variância na avaliação genética de bovinos da raça holandesa no brasil. Belo horizonte. MG: UFMG.
- Van Tassell, C.P, Berger, P.J. (1994). Consideration of sire relationships for estimation of variance components with interaction of herd and sire. *J. Dairy Sci.*, 77(1):313 - 324.
- Verneque, R.S. (1994) Procedimentos numéricos e estimação de componentes de covariância em análise multivariada pelo método da máxima verossimilhança restrita - modelos mistos aplicados ao melhoramento animal. Piracicaba: esalq, 1994. 157p. Dissertação (doutorado em estatística) - Escola Superior de Agricultura Luiz de Queiroz.
- Wolfinger, R. D. (1993) Covariance structure selection in general mixed models. *Communcations in statistics simulation and computation*, v.22, n.4, p.1079 - 1106.