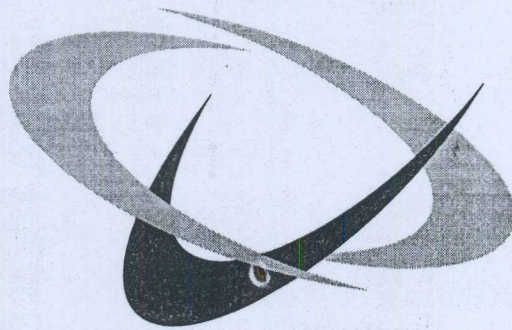


multi  
**multiciência**  
ciência

**PUBLICAÇÃO DO CENTRO UNIVERSITÁRIO  
CENTRAL PAULISTA**



**UNICEP**

**Volume 5 - 2003 - UNICEP - SÃO CARLOS**

PROCI-2003.00020  
FRE  
2003  
SP-2003.00020

# ESTRUTURAS DE COVARIÂNCIAS EM ANÁLISES DE DADOS DE MEDIDAS REPETIDAS

**Alfredo Ribeiro de Freitas**

Pesquisador da Embrapa Pecuária Sudeste, Caixa Postal 339, 13560-970 - São Carlos, SP.  
Bolsista do CNPq.  
ribeiro@cnpse.embrapa.br

**Cátia Valéria Presotti**

Bacharelanda em Estatística da UFSCar-São Carlos - SP.  
catia\_v\_p@hotmail.com

## RESUMO

O objetivo deste estudo foi selecionar as estruturas de covariâncias mais adequadas da variabilidade, quando avaliações sequenciais são feitas em idades diferentes dentro de indivíduos. O estudo foi ilustrado considerando-se dados de nove pesagens obtidas de quatro raças zebuínas: Nelore, Guzerá, Gir e Indubrasil, avaliadas em intervalos trimestrais; do nascimento aos dois anos de idade, obtidos de animais, machos e fêmeas, nascidos de 1972 a 1997, oriundos da Associação Brasileira de Criadores de Zebu - ABCZ. Utilizando-se critérios fornecidos pelo procedimento MIXED do SAS: distribuição de  $c^2$ , AIC ("Akaike's Information Criterion") e SBC ("Schwarz's Bayesian Criterion"), as duas estruturas de covariâncias mais adequadas foram: Fator-Analítico: FA(1) e Autoregressiva de Primeira Ordem com Média Móvel: ARMA(1,1) para Nelore, Não-Estruturada: UN e Simetria Composta Heterogênea: CSH, para Guzerá, UN e FA(1), para Gir e UN e FA(1), para Indubrasil.

**Palavras-chave:** Bovinos de corte; critério de bayesiano de Schwarz; critério de informação de Akaike; procedimento GLM versus MIXED do SAS; seleção de estruturas de covariâncias.

## COVARIANCES STRUCTURES IN REPEATED MEASURES DATA ANALYZES

### ABSTRACT

*The objective of this study was to select the most adequate covariance structures of the variability, when sequential measures are taken at different ages in the same individual. The study was illustrated considering body weight data of nine weighings, from birth to 24 months of age of purebreed Bos Indicus beef cattle in Brazil (Nelore, Guzerá, Gir and Indubrasil), males and females, born from 1972 to 1997. Data were obtained at the National Archive of Brazilian Zebu Breeders Association (ABCZ). Considering the*

*goodness of fit criteria provided by PROC MIXED: ( $\chi^2$  distribution for the null hypothesis, Akaike's Information Criterion and Schwarz's Bayesian Criterion, the best two covariance structures for each category of beef cattle were: Factor-Analytic: FA(1) and First-Order Autoregressive Moving Average: ARMA(1,1), for Nelore, Unstructured: UN and Compound Symmetry Heterogeneous: CSH, for Guzará, UN and FA(1), for Gir and UN and FA(1), for Indubrasil.*

**Key words:** Akaike's Information Criterion; beef cattle; covariance structures selection; GLM versus MIXED procedure of SAS; Schwarz's Bayesian Criterion.

## INTRODUÇÃO

Dados de medidas repetidas, em que as unidades experimentais ou indivíduos são classificados em diferentes subpopulações ou tratamentos, ao longo de diversas condições de avaliação (tempo, doses, etc), têm grande interesse em diversas áreas: indústria (LISKI, 1996), médica, biomédica e veterinária (AZZALINI, 1994; LEE et al. 1996; TROXEL et al. 1998; YAO et al. 1998., REIEZIGEL, 1999), entre outras. Na pesquisa agropecuária (LAMBERT, 1996; MEAD et al. 1996; FREITAS et al., 1998; FREITAS et al., 1999; FREITAS, 2000, FREITAS et al., 2001), dados de medidas repetidas são de fundamental importância nos estudos de crescimento em animais e vegetais, controles de lactação, ensaios de nutrição envolvendo metabolismo e dinâmica do rúmen, cortes sucessivos de forrageiras e principalmente para melhoramento genético animal.

O interesse fundamental do uso dos modelos lineares mistos nesta área (AZZALINI, 1994; TROXEL et al. 1998; FREITAS et al. 1999; REIEZIGEL, 1999) é estimar variâncias e covariâncias a partir dos efeitos aleatórios e produzir testes estatísticos apropriados para efeitos fixos nos vários delineamentos: inteiramente casualizados, blocos, desenho em parcelas subdivididas, desenhos hierárquicos, medidas repetidas, etc.

Para a modelagem de medidas repetidas, é de grande interesse o uso dos modelos lineares mistos (DAVIDIAN e GILTINAN, 1996; LISKI, 1996; LITTEL et al. 1996; LITTEL et al. 1998; REIEZIGEL, 1999)  $y = Xb + Zu + e$ , em que  $y$  é o vetor de observações,  $X$  e  $Z$ , são matrizes de desenho conhecidas, de efeitos fixos e aleatórios;  $b$  e  $u$ , são vetores de efeitos de parâmetros desconhecidos fixos e aleatórios e  $e$ , é o vetor de erros aleatórios não observáveis.

O modelo comumente utilizado é do tipo  $y_{ijk} = m + a_i + d_{ij} + t_k + (at)_{ik} + e_{ijk}$ , em que  $y_{ijk}$  é a resposta no tempo  $k$  do indivíduo  $j$  no grupo de tratamento  $i$ ;  $m$  é o efeito médio geral;  $a_i$  é o efeito fixo do tratamento  $i$ ;  $d_{ij}$  é o efeito aleatório do indivíduo  $j$  no grupo de tratamento  $i$ ;  $t_k$  é o efeito fixo do tempo  $k$ ;  $(at)_{ik}$  é o efeito da interação fixa do tratamento  $i$  com o tempo  $k$  e  $e_{ijk}$  é o erro aleatório (LITTEL et al. 1998). É importante considerar duas fontes de variabilidade: variação aleatória entre e dentro de indivíduos.

Freqüentemente, este tipo de análise tem sido realizada pelo procedimento GLM do SAS sob o enfoque de análise de variância (ANOVA) univariada e delineamento em parcela dividida ("split-plot"), em que os indivíduos representam as parcelas e as avaliações

dentro de indivíduos são as subparcelas. A análise por meio do GLM produz um teste padrão para efeitos entre indivíduos e dois testes para efeitos dentro de indivíduos: univariado e multivariado: o primeiro é válido quando a estrutura de covariância dentro de indivíduos é do tipo HF (condição de Huynh-Feldt). Um teste estatístico é executado pelo GLM conhecido como teste de esfericidade: se é não significativo, a condição de HF é aceita; os testes univariados para efeitos dentro de indivíduos são então implementados; se o teste é significativo o GLM produz dois tipos de ajustamentos: epsilon de Greenhouse-Geisser (G-G) e o HF, o qual é menos conservativo. Para a análise multivariada usa-se quatro diferentes testes: Lâmbda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz Máxima de Roy. Estes testes requerem que a estrutura de variância-covariância dentro de indivíduos seja do tipo não-estruturada (“Unstructured”).

Mesmo considerando-se os ajustes acima, ainda é possível detectar significâncias pelo teste F de efeitos principais e interações destes, que na verdade não existem. Um outro inconveniente sério do procedimento GLM é o fato de ignorar os indivíduos que possuem dados perdidos, o que pode limitar bastante análise em muitas situações.

Nas análises de dados de crescimento de bovinos como medidas repetidas, algumas características inerentes aos dados requerem atenção especial na análise e também inviabilizam o uso do GLM (PRESOTTI e FREITAS, 2002): a) os dados de peso são irregulares no tempo, pois o intervalo entre duas medidas consecutivas quaisquer não é constante; b) possuem estrutura incompleta, isto é, possuem observações perdidas; isto acontece devido a problemas de descartes, mortes, entre outros; c) as avaliações adjacentes são mais estreitamente correlacionadas do que as demais; d) é comum observar variabilidade crescente na resposta dos indivíduos em função do tempo, denominada de “fenômeno da inflação de variância”; a maior dificuldade de se pesar o animal adulto e o acúmulo de influências ambientais contribuem para isto. Assim, um interesse imediato neste tipo de análise é escolher a estrutura de covariância das observações de um mesmo indivíduo que considera o itens de a) a d). O procedimento MIXED do SAS explora adequadamente esse tipo de análise, além de disponibilizar ao usuário cerca de 40 tipos de estruturas de matriz de covariâncias para dados de medidas repetidas, quando os dados tem distribuição normal e seleciona a mais adequada delas (BOZDOGAN, 1987; WOLFINGER, 1993; LITTLE et al. 1996; LITTLE et al., 1998; LINDSEY, 1999; PERRI e IEMMA, 1999).

O objetivo deste estudo foi estimar estruturas de covariâncias por meio do procedimento MIXED do SAS e selecionar, por meio dos critérios AIC (“Akaike’s Information Criterion”) e SBC (“Schwarz’s Bayesian Criterion”), as mais adequadas delas para representar a variabilidade dentro de indivíduos. Foram analisados dados de nove pesagens por animal (medidas repetidas) de quatro raças zebuínas: Nelore, Guzerá, Gir e Indubrasil, avaliadas em intervalos trimestrais; do nascimento aos dois anos de idade, obtidos de animais, machos e fêmeas, nascidos de 1972 a 1997.

## MATERIAL E MÉTODOS

Foram analisados dados de nove pesagens de quatro raças de bovinos zebuínas: Nelore, Guzerá, Gir e Indubrasil, machos e fêmeas, nascidos de 1972 a 1997 e oriundos da Associação Brasileira de Criadores de Zebu - ABCZ; as pesagens foram avaliadas em intervalos trimestrais, do nascimento aos dois anos de idade. O modelo adotado foi o usual para medidas repetidas (LITTELL et al., 1998)  $y_{ijk} = m + a_i + d_{ij} + t_k + (at)_{ik} + e_{ijk}$ , em que  $y_{ijk}$  é a resposta da avaliação  $k$  no indivíduo  $j$  no grupo contemporâneo  $i$ ;  $m$  é a média global,  $a_i$  é o efeito fixo do sexo  $i$ ,  $d_{ij}$  é o efeito aleatório do animal  $j$  no sexo  $i$ ,  $t_k$  é o efeito fixo do tempo  $k$ ,  $(at)_{ik}$  é o efeito de interação e  $e_{ijk}$  é o erro aleatório. O efeito fixo de sexo foi considerado como o fator de variação entre os animais; as múltiplas respostas foram consideradas fator de variação dentro de animais e refletem a estrutura de correlação dentro dos mesmos. As estruturas de covariâncias testadas no presente trabalho (Tabelas 1 e 2) são descritas em vários trabalhos (SEARLE et al. 1992; WOLFINGER, 1993; LITTELL et al. 1996). Foram utilizados três critérios de seleção proporcionados pelo procedimento MIXED: distribuição de  $\chi^2$ , testa a hipótese  $H_0$ : AIC (AIC ("Akaike's Information Criterion",  $AIC = -2L_R + 2q$  e SBC ("Schwarz's Bayesian Criterion",  $SBC = -2L_R(q/y) + q \log(n - p)$ ), em que  $q$  é o número de parâmetros de covariância,  $n$  é o número de observações e  $p$  é o posto da matriz de planejamento  $X$ . A função objetivo para REML (WOLFINGER, 1993) é  $-2L_R((q/y) = \log(V((q) + (y - Xb(q))' V^{-1}(q) [y - Xb(q) + \log|X' V^{-1}(q) X| + (n - p) \log 2p$ . Na análise dos dados de pesos, considerou-se SEXO como tratamento ou fator de variação entre os animais e as condições de avaliação (CA) como medida repetida. Um exemplo do procedimento MIXED (LITTELL et al. 1998) para estimar a estrutura de covariância do tipo VC, é dado abaixo:

```
PROC MIXED DATA=A;
CLASS SEXO CA ANIMAL;
MODEL PESO = SEXO|CA;
REPEATED CA / SUB=ANIMAL(SEXO)TYPE =VC R RCORR;
RUN;
```

A opção CLASS especifica SEXO, CA e ANIMAL como variáveis de classificação; a opção MODEL especifica o modelo para os dados, sendo que SEXO|CA é equivalente a SEXO, CA e interação SEXO\*CA; a opção REPEATED especifica a estrutura de covariância das avaliações dentro de animal, ou seja, considera a dependência entre os  $e_{ijk}$ ; TYPE especifica o tipo de estrutura de covariância a ser determinada (VC, no exemplo); a opção SUB, no presente estudo, define o animal como sendo o efeito de indivíduo, o qual instrui o procedimento MIXED a elaborar  $n$  matrizes ( $n =$  número de animais) bloco diagonal de tamanho  $9 \times 9$  (9 medidas repetidas), sendo que cada bloco tem a estrutura de covariância dada em TYPE. No procedimento MIXED, o REML é o método de estimação padrão para estimar os componentes de variância.

Sob o enfoque matricial dos modelos mistos  $y = Xb + Zu + e$ , o vetor  $b$  contém os

efeitos fixos  $m$ ,  $a_i$ ,  $t_k$  e  $(at)_{ik}$ ; o vetor  $m$  contém os efeitos aleatórios do animal  $j$  no sexo  $i$  (variação entre animal); o vetor  $e$  contém os erros aleatórios (variação dentro do animal). O vetor  $u$  e  $e$  são assumidos terem distribuição normal, ambos com média zero, e variância  $V(u)=G$  e  $V(e) = R$ , respectivamente. O vetor de dados observados  $y$  é então assumido ter distribuição normal com variância  $V=V(y)=V(ZU + e)=ZGZ'+R$ . Na estrutura de medidas repetidas deste trabalho, os dados do animal  $j$  e sexo  $i$  são dados por  $y_{ij1}, y_{ij2}, \dots, y_{ijp}$ , sendo a covariância entre as respostas no tempo  $k$  e  $l$  do mesmo animal dada por  $\sigma^2_d + cov(e_{ijk}, e_{ijl})$ . Para estimar as estruturas de covariâncias, o procedimento MIXED trabalha com a parte  $R = cov(e_{ijk}, e_{ijl})$  do mesmo animal.

**Tabela 1-** Estruturas de covariâncias (VC: Componentes de Variância (*Variance Components*); CS: Simétrica Composta (*Compound Symmetry*); UN: Não Estruturada (*Unstructured*); UN(q): Não Estruturada com q Bandas (*Unstructured With First q Bands*); AR(1): Auto Regressiva de Primeira Ordem (*First-Order Autoregressive*); TOEP; SP(POW): Estrutura de Potência Espacial (*Spatial Power Structure*); ARH1: Auto Regressiva Heterogênea de Primeira Ordem (*Heterogeneous ARH<sub>1</sub>*); CSH: Simetria Composta Heterogênea (*Compound Symmetry Heterogeneous*); FA(1): Fator Analítico de Primeira Ordem (*First-Order Factor Analytic*); HF: Huynh-Feldt e ARMA(1,1): Autoregressiva de Primeira Ordem com Média Móvel (*First-order Autoregressive Moving Average Structure*)).

$VC = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ & \sigma_2^2 & 0 & 0 \\ & & \sigma_3^2 & 0 \\ Sim. & & & \sigma_4^2 \end{bmatrix}$	$CS = \begin{bmatrix} (\sigma^2 + \sigma_1) & \sigma_1 & \sigma_1 & \sigma_1 \\ & (\sigma^2 + \sigma_1) & \sigma_1 & \sigma_1 \\ & & (\sigma^2 + \sigma_1) & \sigma_1 \\ Sim. & & & (\sigma^2 + \sigma_1) \end{bmatrix}$	$UN = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ & & \sigma_{33} & \sigma_{34} \\ Sim. & & & \sigma_{44} \end{bmatrix}$
$UNq = \begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & 0 \\ & \sigma_4 & \sigma_5 & \sigma_6 \\ & & \sigma_7 & \sigma_8 \\ Sim. & & & \sigma_9 \end{bmatrix}$	$AR(1) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & \sigma^2 & 1 & \rho & \rho^2 \\ & & Sim. & 1 & \rho \\ & & & & 1 \end{bmatrix}$	$TOEP = \begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ & \sigma_1 & \sigma_2 & \sigma_3 \\ & & \sigma_1 & \sigma_2 \\ Sim. & & & \sigma_1 \end{bmatrix}$
$SP(POW) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho \\ Sim. & & & \sigma_4^2 \end{bmatrix}$	$ARH1 = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho \\ Sim. & & & \sigma_4^2 \end{bmatrix}$	$CSH = \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ & & 1 & \rho^{d_{34}} \\ Sim. & & & 1 \end{bmatrix}$
$FA(q) = \begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2-\lambda} & \frac{\sigma_1^2 + \sigma_3^2}{2-\lambda} & \frac{\sigma_1^2 + \sigma_4^2}{2-\lambda} \\ & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2-\lambda} & \frac{\sigma_2^2 + \sigma_4^2}{2-\lambda} \\ & & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2-\lambda} \\ Sim. & & & \sigma_4^2 \end{bmatrix}$	$HF = \begin{bmatrix} \lambda_1 + d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ & \lambda_2 + d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ & & \lambda_3 + d & \lambda_4\lambda_3 \\ Sim. & & & \lambda_4 + d \end{bmatrix}$	$ARMA(1,1) = \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ & \sigma^2 & 1 & \gamma & \gamma\rho \\ & & Sim. & 1 & \gamma\rho \\ & & & & 1 \end{bmatrix}$

**Tabela 2-** Estruturas de covariâncias, número de parâmetros da matriz e elemento  $(i,j)$

Estrutura	Nº de parâmetros	Elemento (i,j)
VC	q	$\sigma_k^2 \mathbf{1} (i = j)$ , i corresponde ao k-ésimo efeito
CS	2	$\sigma^2_1 + \sigma^2 (i = j)$
UN	$t(t+1)/2$	$\sigma_{ij}$
UN(q)	$q/2(2t - q+1)$	$\sigma_{ij} \mathbf{1} ( i - j  < q)$
AR(1)	2	$\sigma^2 + \rho^{ i-j }$
TOEP	t	$\sigma  i - j  + 1$
SP(POW)	2	$\sigma^2 \rho^{d^{ij}}$
ARH1	t+1	$\sigma_i \sigma_j \rho^{ i-j }$
CSH	t+1	$\sigma_i \sigma_j [1(i \neq j) + 1 (i = j)]$
FA(1)	$[q/2](2t - q+1) + t$	$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk} + \sigma^2_1 \mathbf{1} (i = j)$
HF	t+1	$(\sigma^2_1 + \sigma^2_j)/2 + \lambda \mathbf{1}(i \neq j)$
ARMA(1,1)	3	$\sigma^2 [\gamma \rho^{ i-j -1} \mathbf{1}(i \neq j) + 1 (i = j)]$

\*q = dimensão da matriz; t = número de parâmetros diferentes

$\gamma$  = componente de média móvel;  $\rho$  = parâmetro autorregressivo

$\sigma^2$  = variância residual;  $\sigma^2_i$  = i-ésimo parâmetro de variância

$d^{ij}$  = Distância Euclidiana entre a i-ésima e j-ésima observação de medidas repetidas.

## RESULTADOS E DISCUSSÃO

A Tabela 3 descreve os resultados das estruturas de covariâncias e três critérios de ajuste proporcionados pelo procedimento MIXED: distribuição de  $\chi^2$ , para o teste da razão de verossimilhança para a hipótese nula de AIC e SBC. Para todas as situações, o teste de esfericidade, o qual testa se a matriz de variância-covariância dentro de indivíduo tem estrutura de covariância do tipo HF (*Huynh-Feldt*), foi altamente significativo ( $\text{Pr} > \chi^2 < 0,0001$ ), indicando para estes conjuntos de dados, que a análise univariada, normalmente feita pelo procedimento GLM, considerando-se medidas repetidas como subparcelas não é apropriada.

Na forma fornecida pelo PROC MIXED (LITTELL et al., 1998), se os valores são negativos para AIC e BIC, quanto maior o valor mais apropriada é a estrutura de covariâncias. Por outro lado, se os valores fornecidos para AIC e BIC, são positivos, os menores são os melhores. De acordo com estes critérios, as duas estruturas de covariâncias mais adequadas (Tabela 3) foram: Fator-Analítico: FA(1) e Autoregressiva de Primeira Ordem com Média Móvel: ARMA(1,1) para Nelore, Não-Estruturada: UN e Simetria Composta Heterogênea: CSH, para Guzerá, UN e FA(1), para Gir e UN e FA(1), para Indubrasil.

Nas Tabelas 4 a 7 são apresentadas as duas estruturas selecionadas para Nelore (Tabela 4), Guzerá (Tabela 5), Gir (Tabela 6) e Indubrasil (Tabela 7). Nas duas estruturas de covariâncias para Nelore, na FA(1), as variâncias tendem a crescer na diagonal e as covariâncias com os últimos pesos tendem a ser maiores; na ARMA(1,1), as variâncias são iguais na diagonal e as pesagens mais próximas são mais estreitamente correlacionadas.

Para Guzerá, em ambas as estruturas, as variâncias são crescentes na diagonal, as covariâncias com os últimos pesos tendem a ser maiores e as pesagens consecutivas são mais estreitamente correlacionadas. As estruturas UN e FA(1) nas raças Gir e Indubrasil, apresentam resultados semelhantes aos obtidos para a raça Guzerá, com as estruturas UN e CSH. De modo geral, exceção à ARMA(1,1) para Nelore, os resultados sugerem que as covariâncias são crescentes na diagonal e as correlações entre duas avaliações diminuem com o aumento do intervalo de tempo (*lag*), isto é, as variações aumentam em função da idade, exibindo o fenômeno da inflação de variância, observado em estudos de curvas de crescimento em animais (DAVIDIAN e GILTINAN, 1996, FREITAS et al, 1998; FREITAS et al, 1999 e FREITAS, 2000).

**Tabela 3** - Estruturas de covariâncias e critérios de ajuste proporcionados pelo procedimento MIXED:  $\chi^2$ , para testar a hipótese nula da razão de verossimilhança, AIC e BIC

Breeds	Covariâncias	Nº de parâmetros	$\chi^2$	Pr> $\chi^2$	AIC	BIC
Nelore	FA(1)	18	7185	<0,0001	362014	362068
	ARMA(1,1)	3	1681	<0,0001	367505	367523
Guzera	UN	45	80215	<0,0001	-206428	-206582
	CSH	10	62511	<0,0001	-215245	-215279
Gir	UN	45	59498	<0,0001	497196	497484
	FA(1)	18	8200	<0,0001	548422	548479
Indubrasil	UN	45	10067	<0,0001	52257	52458
	FA(1)	18	6891	<0,0001	55342	55407

**Tabela 4** - Estruturas de covariâncias para Nelore: FA(1), no triângulo superior e ARMA(1,1), no triângulo inferior. Pi referem-se aos pesos ao nascimento (P0) e até dois anos de idade.

	P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>
P <sub>0</sub>	19342	0	0	0	0	0	0	0	0
P <sub>1</sub>	30847	19347	48	43	76	114	23	540	410
P <sub>2</sub>	7125	30847	19822	428	768	1146	232	5424	4128
P <sub>3</sub>	5406	7125	30847	19722	683	1020	206	4823	3669
P <sub>4</sub>	4102	5406	7125	30847	20569	1833	369	8673	6598
P <sub>5</sub>	3112	4102	5406	7125	30847	22082	552	12957	9859
P <sub>6</sub>	2361	3112	4102	5406	7125	30847	19454	2611	1986
P <sub>7</sub>	1792	2361	3112	4102	5406	7125	30847	80636	46638
P <sub>8</sub>	1359	1792	2361	3112	4102	5406	7125	30847	54827
P <sub>8</sub>	1031	1359	1792	2361	3112	4102	5406	7125	30847

**Tabela 5**- Estruturas de covariâncias para Guzerá: UN no triângulo superior e CHS no triângulo inferior. Pi referem-se aos pesos ao nascimento (P0) e até dois anos de idade.



	P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>
P <sub>0</sub>	10	10	20	26	30	35	43	50	54
P <sub>1</sub>	14	365	349	300	297	365	384	387	413
P <sub>2</sub>	41	429	810	762	745	846	988	1048	1031
P <sub>3</sub>	55	299	783	1292	1287	1355	1556	1788	1771
P <sub>4</sub>	66	362	489	1147	1962	2045	2157	2451	2553
P <sub>5</sub>	81	443	599	725	1717	2951	2980	3173	3292
P <sub>6</sub>	100	545	737	891	1090	2598	3891	4049	3999
P <sub>7</sub>	114	622	841	1017	1244	1531	3384	5039	5013
P <sub>8</sub>	124	787	917	1110	1358	1671	1907	4031	5812
P <sub>8</sub>	140	761	1029	1245	1524	1874	2139	2336	5072

**Tabela 6-** Estruturas de covariâncias para Gir: UN no triângulo superior e FA(1) no triângulo inferior. Pi referem-se aos pesos ao nascimento (P0) e até dois anos de idade.

	P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>
P <sub>0</sub>	9	11	27	-0,9	5	-21	68	-88	-11
P <sub>1</sub>	38573	446	300	525	590	645	558	743	451
P <sub>2</sub>	0	38577	50180	-3785	787	1913	1164	-579	2560
P <sub>3</sub>	0	7	38587	50464	-3844	4887	221	4163	-985
P <sub>4</sub>	0	13	28	38628	49984	-6265	2010	3212	4983
P <sub>5</sub>	0	22	47	92	38728	71521	-3858	361	7249
P <sub>6</sub>	0	82	173	339	572	40688	34736	-17852	-6467
P <sub>7</sub>	0	-76	-160	-315	-532	-1963	40396	33149	71804
P <sub>8</sub>	0	528	1114	2191	3696	13651	-12673	126706	21061
P <sub>8</sub>	0	477	1008	1982	3343	12346	-11462	79711	110667

**Tabela 7-** Estruturas de covariâncias para Indubrasil: FA(1) no triângulo superior e UN no triângulo inferior. Pi referem-se aos pesos ao nascimento (P0) e até dois anos de idade.

	P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>
P <sub>0</sub>	600	2	4	7	10	12	15	17	19
P <sub>1</sub>	19	653	117	207	310	381	462	523	594
P <sub>2</sub>	22	500	857	455	679	837	1014	1148	1302
P <sub>3</sub>	20	432	964	1406	1204	1484	1798	2036	2309
P <sub>4</sub>	13	460	920	1582	2399	2217	2687	3042	3450
P <sub>5</sub>	2	433	917	1572	2546	3331	3310	3748	4250
P <sub>6</sub>	9	448	942	1606	2579	3391	4613	4543	5152
P <sub>7</sub>	18	417	975	1770	2666	3479	4533	5743	5832
P <sub>8</sub>	19	409	964	1864	2815	3540	4656	5758	7214
P <sub>8</sub>	15	486	1092	2038	3171	4005	5038	6089	7525

A Tabela 8 apresenta as estimativas de covariâncias das estruturas FA(1), nas raças Nelore, Gir e Indubrasil; ARMA(1,1), para Nelore, e CSH para Guzerá, utilizadas na elaboração das matrizes das Tabelas 4 a 7. Foram omitidas as estimativas de covariâncias para a estrutura UN, pois as mesmas são usadas diretamente na elaboração das matrizes, não necessitando de cálculos intermediários.

**Tabela 8 - Estimativas de covariâncias**

<b>Nelore</b>		
<b>FA(1) - Covariance Parameter Estimates</b>		
Cov Parm	Subject	Estimate
FA (1)	ANIMAL (SEXO)	19342
FA (1, 1)	ANIMAL (SEXO)	0
FA (2, 1)	ANIMAL (SEXO)	2.1763
FA (3, 1)	ANIMAL (SEXO)	21.9071
FA (4, 1)	ANIMAL (SEXO)	19.4760
FA (5, 1)	ANIMAL (SEXO)	35.0280
FA (6, 1)	ANIMAL (SEXO)	52.3371
FA (7, 1)	ANIMAL (SEXO)	10.5442
FA (8, 1)	ANIMAL (SEXO)	247.58
FA (9, 1)	ANIMAL (SEXO)	188.37
<b>ARIMA (1,1) - Covariance Parameter Estimates</b>		
Cov Parm	Subject	Estimate
Rho	ANIMAL (SEXO)	0.7587
Gamma	ANIMAL (SEXO)	0.2310
Residual		30847
<b>Guzerá</b>		
<b>CSH - Covariance Parameter Estimates</b>		
Cov Parm	Subject	Estimate
Var (1)	ANIMAL (SEXO)	14.4267
Var (2)	ANIMAL (SEXO)	428.62
Var (3)	ANIMAL (SEXO)	783.34
Var (4)	ANIMAL (SEXO)	1147.32
Var (5)	ANIMAL (SEXO)	1716.72
Var (6)	ANIMAL (SEXO)	2598.12
Var (7)	ANIMAL (SEXO)	3384.02
Var (8)	ANIMAL (SEXO)	4031.50
Var (9)	ANIMAL (SEXO)	5072.43
CSH	ANIMAL (SEXO)	0.5163
<b>Gir</b>		
<b>FA(1) - Covariance Parameter Estimates</b>		
Cov Parm	Subject	Estimate
FA (1)	ANIMAL (SEXO)	38573
FA (1, 1)	ANIMAL (SEXO)	0
FA (2, 1)	ANIMAL (SEXO)	1.7771
FA (3, 1)	ANIMAL (SEXO)	3.7534
FA (4, 1)	ANIMAL (SEXO)	7.3799
FA (5, 1)	ANIMAL (SEXO)	12.4497
FA (6, 1)	ANIMAL (SEXO)	45.9822
FA (7, 1)	ANIMAL (SEXO)	-42.6892
FA (8, 1)	ANIMAL (SEXO)	296.87
FA (9, 1)	ANIMAL (SEXO)	268.50

Nas Tabelas 9 a 11, observa-se os resultados do ajuste do procedimento MIXED para a estrutura de covariância ARIMA (1,1) para o rebanho Nelore, utilizando-se o programa:

```

PROC MIXED DATA=NELORE;
TITLE ARMA(1,1);
CLASS SEXO CA ANIMAL;
MODEL PESO = SEXO|CA;
REPEATED CA / SUB=ANIMAL(SEXO) TYPE=ARMA(1,1) R RCORR;
RUN;

```

Na Tabela 9, são fornecidas algumas informações do modelo (*Model Information*): a) parâmetros de covariâncias estimados: 3; b) colunas usadas para a matriz X: 30; c) colunas para matriz Z: 0; d) indivíduos ou animais: 3116; e) avaliações por indivíduo: 9; f) observações usadas: 28044 (3116 x 9 = 28044); g) todas as observações foram usadas e h) total de observações: 28044. A história do processo iterativo (*Iteration History*) na Tabela 9, mostra que cinco iterações foram necessárias para a convergência na obtenção dos parâmetros, sendo produzido o valor máximo do log da verossimilhança, multiplicado por -2 ( $-2 \text{ Res Log Likelihood} = 367499,26136282$ ). Foi utilizado o critério de convergência padrão do procedimento MIXED, com um número de tolerância n:  $\mathbf{g}'_k \mathbf{H}^{-1}_k \mathbf{g}_k / |\mathbf{f}_k| \leq 1\text{E}-8$ , em que,  $\mathbf{g}_k$  é o gradiente (primeira derivada) e  $\mathbf{H}_k$  é a matriz Hessiana (segunda derivada) da função objetivo, ambos na iteração k. De acordo com o resultado, obteve-se na primeira iteração  $\mathbf{g}'_k \mathbf{H}^{-1}_k \mathbf{g}_k / |\mathbf{f}_k| \leq 11,22922528$  e na iteração número 5,  $\mathbf{g}'_k \mathbf{H}^{-1}_k \mathbf{g}_k / |\mathbf{f}_k| \leq 0,00000000$ , tendo-se a convergência.

Duas estimativas parciais de matrizes são apresentadas na Tabela 10: a matriz de covariância e a de correlação. Sob o título **ESTIMATED R COVARIANCE MATRIX FOR ANIMAL(SEXO) 760609 1**, significa uma submatriz de covariância para medidas repetidas do animal de número 760609 no arquivo de dados e pertencente ao tratamento 1 (sexo = 1); por hipótese, todos os outros animais possuem essa mesma estrutura; por exemplo, a covariância entre a pesagem 1 e 2 é igual 7125,37. Com o título **ESTIMATED R CORRELATION MATRIX FOR ANIMAL (SEXO) 760609 1**, é apresentada a submatriz de correlação do mesmo animal, com a correlação entre pesagem 1 e 2 igual a 0,2310. Também, por hipótese, os demais animais do conjunto de dados analisado possuem essa mesma estrutura de correlação.

Na Tabela 11, com o título **COVARIANCE PARAMETER ESTIMATES**, são apresentadas as estimativas dos três componentes de variâncias associados à estrutura ARIMA(1,2), ou seja, residual =  $\sigma^2 = 30847$ ,  $Rho = \rho = 0,7587$  e  $Gamma = \gamma = 0,2310$ . Com estes valores foi construída a matriz triângulo inferior da Tabela 4. Observa-se que essa matriz proporciona a mesma variância entre as diversas medidas repetidas, o que difere do esperado em curvas de crescimento de animal, conforme o item "d" de Introdução: "é comum observar variabilidade crescente na resposta dos indivíduos em função do tempo, denominada de fenômeno da inflação de variância"; contudo, atende o item c: "as avaliações adjacentes são mais estreitamente correlacionadas do que as demais". Possivelmente, esta seja a razão desta estrutura ter sido a segunda opção de escolha na raça Nelore.

Com o **FIT STATISTICS** são fornecidos três critérios pelo procedimento MIXED para comparar modelos:  $-2 \text{ Res Log Likelihood}$ , AIC ("Akaike's Information Criterion") e SBC ou BIC ("Schwarz's Bayesian Criterion"). Por exemplo,  $-2 \text{ Res Log Likelihood} = 367499,3$ , serve para construir testes de razão de verossimilhança restrito e comparar duas estruturas de variâncias. O resultado é uma distribuição de *Chi-Square*,

com graus de liberdade igual à diferença do número de parâmetros. Na forma fornecida pelo PROC MIXED (LITTELL et al., 1998), como os valores para AIC e BIC são positivos, quanto menor o valor mais apropriada é a estrutura de covariâncias; assim, critério de AIC (367505,3) é levemente superior ao de BIC (367523,4).

A saída **NULL MODEL LIKELIHOOD RATIO TEST** mostra a estatística Qui-quadrado com dois graus de liberdade e  $Pr > ChiSq$ ; este resultado equivale ao teste de razão de verossimilhança restrito comparando o modelo associado à matriz ARIMA(1,1), em uso, *versus* o modelo nulo ajustado pelo procedimento GLM do SAS ( $R = I\sigma^2$ ). O resultado ( $Pr > \chi^2 = 1681,90 < 0,0001$ ), mostra que o modelo que estima a matriz a ARIMA(1,1) é significativamente ( $P < ,0001$ ) melhor do que o modelo simples nulo. Finalmente, sob o título **TYPE 3 TESTS OF FIXED EFFECTS**, tem-se os testes de efeitos fixos tipo 3 como produzido pelo SAS (1996). Houve significância ( $P < 0,0001$ ) do efeito global de tratamento (SEXO), considerando-se a média das nove avaliações do efeito das avaliações dentro do animal, independente de sexo; esta significância entre as avaliações ou pesagens é uma constatação de que os pesos avaliados no mesmo animal têm respostas crescentes com a idade; finalmente, o último valor mostra interação das medidas repetidas com o sexo. Os resultados das Tabelas de 9 a 11, mostram que o procedimento MIXED atua em duas etapas; na primeira, estima as estruturas de variância e covariâncias, as quais são colocadas no lugar dos parâmetros verdadeiros; a segunda e última etapa consiste de cálculos usando as distribuições t e F, conforme apresentada em **TYPE 3 TESTS OF FIXED EFFECTS**.

**Tabela 9 - Model Information e Iteration History do procedimento MIXED**

Model Information			
Covariance Parameters			3
Columns in X		30	
Columns in Z		0	
Subjects	3116		
Max Obs Per Subject		9	
Observations Used		28044	
Observations Not Used		0	
Total Observations		28044	
Iteration History			
Iteration	Evaluations	-2 Res Log Likelihood	Criterion
0	1	369181.16485045	
1	3	368245.58127071	11.22922528
2	2	367671.46109101	0.00181140
3	2	367518.62021255	0.00012667
4	1	367499.32009889	0.00000037
5	1	367499.26136282	0.00000000
<b>Convergence criteria met</b>			

**Tabela 10 - Estimated R Matrix e R Correlation Matrix for ANIMAL(SEXO) 760609 1**

<b>Estimated R Covariance Matrix for ANIMAL(SEXO) 760609 1</b>									
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	30847	7125.37	5406.25	4101.89	3112.24	2361.36	1791.64	1359.37	1031.40
2	7125.37	30847	7125.37	5406.25	4101.89	3112.24	2361.36	1791.64	1359.37
3	5406.25	7125.37	30847	7125.37	5406.25	4101.89	3112.24	2361.36	1791.64
4	4101.89	5406.25	7125.37	30847	7125.37	5406.25	4101.89	3112.24	2361.36
5	3112.24	4101.89	5406.25	7125.37	30847	7125.37	5406.25	4101.89	3112.24
6	2361.36	3112.24	4101.89	5406.25	7125.37	30847	7125.37	5406.25	4101.89
7	1791.64	2361.36	3112.24	4101.89	5406.25	7125.37	30847	7125.37	5406.25
8	1359.37	1791.64	2361.36	3112.24	4101.89	5406.25	7125.37	30847	7125.37
9	1031.40	1359.37	1791.64	2361.36	3112.24	4101.89	5406.25	7125.37	30847

<b>Estimated R Correlation Matrix for ANIMAL(SEXO) 760609 1</b>									
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.0000	0.2310	0.1753	0.1330	0.1009	0.07655	0.05808	0.04407	0.03344
2	0.2310	1.0000	0.2310	0.1753	0.1330	0.1009	0.07655	0.05808	0.04407
3	0.1753	0.2310	1.0000	0.2310	0.1753	0.1330	0.1009	0.07655	0.05808
4	0.1330	0.1753	0.2310	1.0000	0.2310	0.1753	0.1330	0.1009	0.07655
5	0.1009	0.1330	0.1753	0.2310	1.0000	0.2310	0.1753	0.1330	0.1009
6	0.07655	0.1009	0.1330	0.1753	0.2310	1.0000	0.2310	0.1753	0.1330
7	0.05808	0.07655	0.1009	0.1330	0.1753	0.2310	1.0000	0.2310	0.1753
8	0.04407	0.05808	0.07655	0.1009	0.1330	0.1753	0.2310	1.0000	0.2310
9	0.03344	0.04407	0.05808	0.07655	0.1009	0.1330	0.1753	0.2310	1.0000

**Tabela 11 - Covariance Parameter Estimates, Fit Statistics, Null Model Likelihood Ratio Test e Type 3 Tests of Fixed Effects**

<b>Covariance Parameter Estimates</b>		
Cov Parm	Subject	Estimate
Rho	ANIMAL (SEXO)	0.7587
Gamma	ANIMAL (SEXO)	0.2310
Residual		30847

<b>Fit Statistics</b>	
-2 Res Log Likelihood	367499.3
AIC (smaller is better)	367505.3
BIC (smaller is better)	367523.4

<b>Null Model Likelihood Ratio Test</b>		
DF	Chi-Square	Pr > ChiSq
2	1681.90	<.0001

<b>Type 3 Tests of Fixed Effects</b>					
Num	Den	DF	DF	F Value	Pr > F
	Effect				
	SEXO	1	3114	115.10	<.0001
	ca	8	25E3	1876.66	<.0001
	SEXO*ca	8	25E3	33.22	<.0001

## CONCLUSÕES

1. As estruturas de covariâncias selecionadas de dados de peso corporais de bovinos zebuínos e analisadas como medidas repetidas indicaram que as correlações associadas às medidas repetidas avaliadas no mesmo animal tendem a diminuir com o intervalo do tempo ("lag") entre duas medidas consecutivas.
2. As variâncias tendem a aumentar em função da idade ou tempo, exibindo o fenômeno da inflação de variância, observado normalmente em estudos de curvas de crescimento de animais.
3. Todos os ajustes modelos ajustaram significativamente melhor do que o modelo simples nulo ajustado pelo procedimento GLM do SAS ( $R = I\sigma^2$ ), indicando, para estes conjuntos de dados, que a análise univariada, normalmente feita pelo procedimento GLM, considerando-se medidas repetidas como subparcela, não é apropriada.
4. As duas melhores estruturas de covariâncias para cada raça foram: Fator-Analítico: FA(1) e Autoregressiva de Primeira Ordem com Média Móvel: ARMA(1,1) para Nelore; Não-estruturada: UN e a Simetria Composta Heterogênea: CSH para Guzerá, UN e FA(1) para Gir e FA(1) e UN para a raça Indubrasil.

## REFERÊNCIAS BIBLIOGRÁFICAS

AZZALINI, A. 1994. Logistic regression for auto correlated data with application to repeated measures. *Biometrics*. 8,n., p.767-775.

BOZDOGAN, H. 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, v.52, n.3, p.345-370, 1987

DAVIDIAN, M.; GILTINAN, D. M. Nonlinear models for repeated measurement data. 2.ed. London: Chapman Hall, 1996. 359p.

FREITAS, A. R. ALENCAR, M.M; SILVAL, A.M. Ajuste de modelos não lineares em bovinos de corte II. influência do mês de nascimento In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA., 34, 1997, Botucatu. Anais... Botucatu: SBZ, 1998. p.190-192

FREITAS, A.R.; FELICIANO SILVA, A.E.D.; UNANIAN, M.M. 1999. Estimativa do tamanho amostral em medidas repetidas de circunferência escrotal de bovinos Nelore. *Revista da Sociedade Brasileira Zootecnia*, v.28, n.2, p.278-286.

FREITAS, A. R. 2000. Método para estimar o tamanho amostral em análises de medidas repetidas Pesquisa . *Agropecuária Brasileira*. v.35, n.4, p.697-702.

FREITAS, A. R. de., PRIMAVESI, O, CORREA, L.A., PRIMAVESI, OTTEB., MASCIOLLI, A.S. 2001. Repeated measurement analyses of forages in cropping systems. In: XIX INTERNATIONAL GRASSLAND CONGRESS., 19, 2001, Piracicaba, Anais... Piracicaba:, 2001. p.1046-1047.

LAMBERT, P. 1996. Modelling of repeated series of count data measured at unequally spaced times. *Applied Statistics*, v.45, n.1, p.31-38.

LEE, S.J., KIM, K., TSIATIS, A.A. 1996. Repeated significance testing in longitudinal chemical trials. *Biometrika*, v.83, n.4, p.779-789.

LINDSEY, J. K. 1999. Multivariate elliptically contoured distributions for repeated measurement. *Biometris*. 55: 1277-1280.

LISKI, E.P., NUMMI, T. 1996. Prediction in repeated-measures models with engineering application. *Technometrics*, v.38, n.1, p.25-36.

LITTELL, R.C., HENRY, P.R., AMMERMAN, C.B. 1998. Statistical analysis of repeated measures data using SAS procedures. *J. Anim. Sci.* 76:1216-1231.

LITTELL, R.C., MILLIKEN, G.A., STROUP, W.W., et al. 1996. *SAS System for Mixed Models*. Cary: Statistical Analysis System Institute, 633p.

MEAD, R., CURNOW, R.N., HASTED, A.M. 1996. *Statistical methods in agriculture and experimental biology*. 2.ed; Chapman & Hall. 415p.

PERRI, S.H.V. ; IEMMA, A.F. 1999. Procedimento MIXED do sistema estatístico SAS para análise de modelos mistos. *Sci. Agric.* 56(4): 959-967.

PRESOTTI, C. V.; FREITAS, A. R. Procedimentos do SAS: GLM VS MIXED em medidas repetidas In: JORNADA REGIONAL DE ESTATÍSTICA, 3., SEMANA DA ESTATÍSTICA, 2., 2002, Maringá, PR, Maringá: SBE, 2002. p.142-149.

RAO, C.R. 1972. Estimation of variance and covariance components in linear models. *Jour. Amer. Statist. Ass.* 67: 112-115.

REIEZIGEL, J. 1999. Analysis of experimental data with repeated measurement. *Biometrics*. 55: 1059-1063.

SAS INSTITUTE Inc. 1996. *Course notes*. Cary: Statistical Analysis System institute. 614p.

SAS INSTITUTE. 1996. Advanced general linear models with an emphasis on mixed models. Cary: Statistical Analysis System Institute. 614p.

SAS INSTITUTE. 1997. Univariate & multivariate General Linear Models: Theory and applications using SAS Software . Cary: Statistical Analysis System Institute. 619p.

SAS Institute. (1993b). SAS/STAT User's guide: statistics. Release 6.4. Cary, SAS Inst. 1686p.

SEARLE, S.R. 1987. Linear Models for Unbalanced Data. Wiley. 536 p.

TROXEL, A.B., LIPSITZ, S.R., HARRINGTON,D.P. 1998. Marginal models for the analysis of longitudinal measurements. *Biometrika*, v.85,n.3,p.661-672.

WOLFINGER, R. 1993. Covariance structure selection in general mixed models. *Commun.Statist. Simula.*, 22(4):1079-1106.

YAO,Q.,WEI,L.J.,HOEGAN,J.W. ,1998. Analysis of incomplete repeated-measurements with dependent censoring times. *Biometrika*,v.85,n.1,p.139-149.