ANÁLISE COMPARATIVA DE MODELOS PARA DADOS LONGITUDINAIS NO ESTUDO DA CONTAGEM DO NÚMERO DE BACTÉRIAS PRESENTES NO LEITE DE VACA

Idemauro Antonio Rodrigues de LARA[1]
Maria Helena Constantino SPYRIDES[2]
Mirela Gurgel GUERRA[3]
Adriano Henrique do Nascimento RANGEL[3]

▪ RESUMO: Nesse trabalho é descrito um estudo desenvolvido no estado do Rio Grande do Norte, Brasil, sobre a contagem bacteriana total no leite de vaca a fim de avaliar os impactos dos procedimentos de manejo, na ordenha, limpeza de equipamentos e do tanque de resfriamento sob essa variável durante o processo de produção do leite. O objetivo foi comparar duas abordagens metodológicas para análise de dados longitudinais: os modelos marginais e os modelos de efeitos mistos. Uma discussão comparativa no tocante a utilização desses modelos, estimação e interpretação dos resultados é apresentada. Em termos práticos, os resultados mostraram que os procedimentos adequados no manejo na ordenha e na limpeza do tanque de resfriamento são os que mais contribuem para uma maior redução nos níveis de contaminação do leite.

▪ PALAVRAS-CHAVE: CBT; dados longitudinais; modelos lineares generalizados; EEG's; efeitos aleatórios.

1 Introdução

A contagem bacteriana total (CBT) é um indicador da qualidade do leite produzido, refletindo o grau de higiene e limpeza, tanto na produção quanto no

[1]Universidade de São Paulo – USP, Escola Superior de Agricultura "Luiz de Queiroz"-- ESALQ, Departamento de Ciências Exatas, CEP:13418-900, Piracicaba, SP, Brasil. E-mail: *idemauro@usp.br*
[2]Universidade Federal do Rio Grande do Norte – UFRN, Departamento de Estatística, CEP: 59078-970, Natal, RN, Brasil. E-mail: *spyrides@ccet.ufrn.br*
[3]Unidade Acadêmica Especializada em Ciências Agrárias. E-mail: *mirelaguerra@yahoo.com.br* / *adrianohrangel@yahoo.com.br*

manuseio, além da temperatura de armazenamento. O leite recém ordenhado, em geral, apresenta baixo número de microorganismos, no entanto, condições inadequadas de higiene, assim como, de refrigeração e armazenamento do leite podem favorecer a proliferação do número de bactérias. Outros fatores de igual importância que favorecem a contaminação microbiana do leite cru são a saúde da glândula mamária, os cuidados com a ordenha, o ambiente em que o animal fica alojado, além da qualidade da água utilizada (SANTOS e FONSECA, 2001). No Brasil, novas normas de sanidade animal, higiene, refrigeração e nutrição animal foram implementadas com vistas a atender às exigências do mercado internacional, promover a melhoria da qualidade do leite e derivados, garantir a segurança alimentar e aumentar a competitividade dos produtos lácteos (SANTOS et al., 2009). O Ministério da Agricultura, Pecuária e Abastecimento (Mapa) criou o Programa Nacional de Melhoria da Qualidade do Leite - PNQL e publicou a Instrução Normativa N.^{o} 51/2002 (IN51/2002), estabelecendo limites legais (gordura, proteína, sólidos totais, células somáticas e contagem bacteriana total) dos diversos tipos de leite no país, bem como a coleta e o transporte a granel de leite cru refrigerado (DÜRR, 2006; ZANELA et al., 2006). Quanto à CBT, nas regiões Norte e Nordeste do Brasil, o limite de contagem padrão em placas para o leite refrigerado cru deve ser até 750.000 ufc/mL a partir de julho de 2010, passando para o máximo de 100.000 ufc/mL a partir de julho de 2012 (BRASIL, 2002). Diversos estudos mostram que é possível produzir leite com CBT menor ou igual a 100.000 ufc/mL, e esta contagem é regularmente alcançada por produtores em diversos países. Na União Européia e na Nova Zelândia o limite legal de CBT é 100.000 ufc/mL, e no Canadá o limite máximo é 50.000 ufc/mL (SOUTO et al., 2009). Apesar dos novos esforços para a melhoria da qualidade do leite cru, os produtores brasileiros ainda estão longe de atender, na sua totalidade, os parâmetros propostos pela normativa e às necessidades contemporâneas de segurança alimentar. Após dez meses de monitoramento oficial realizado pela Rede Brasileira de Laboratórios de Qualidade do Leite (RBQL), mais de 50% das amostras de leite cru analisadas apresentaram CBT acima do índice permitido pela legislação, dando sinal das precárias condições de higiene e conservação na maioria das fazendas leiteiras do país (DÜRR, 2006).

Sendo assim, é pertinente propor modelos estatísticos que possam descrever de forma adequada o relacionamento funcional entre a CBT e um conjunto de covariáveis que descrevam as condições de produção do leite. Diversos experimentos têm sido planejados com objetivo de identificar covariáveis ou fatores que mais contribuem para o aumento da CBT, com vistas a estabelecer metas e medidas corretivas. Como a CBT refere-se a uma variável discreta, no caso contagem, um dos modelos mais apropriados para a análise é o modelo de regressão Poisson (PAULA, 2004). Como o estudo envolve dados longitudinais, os modelos marginais e de efeitos aleatórios são duas opções interessantes e de objetivos específicos apropriados para descrever estruturas de dependência para dados correlacionados (DIGGLE et al., 2002). Há de se ressaltar que a escolha do modelo para a análise de dados longitudinais deve levar em conta não somente a natureza da variável resposta, mas também, as hipóteses do estudo. Zeger, Liang e Albert (1988),

trabalhando com dados binários, chamaram a atenção para distinção entre essas duas abordagens. Quando o conjunto de dados envolve apenas efeitos fixos e o interesse principal está em modelar a resposta média populacional, os modelos marginais apresentam-se como os mais apropriados. Nesse caso, está se descrevendo uma "curva média" de respostas em função das covariáveis. No entanto, em algumas situações experimentais, assume-se que a variável resposta é uma função de covariáveis cujos coeficientes de regressão variam de um indivíduo para o outro. Essa variabilidade, em geral, é atribuída a fatores não mensuráveis e pode ser representada por uma distribuição de probabilidade. A correlação entre as observações é atribuída à presença dessa variável aleatória (DIGGLE et al., 2002). Assim, na presença de efeitos aleatórios o foco passa a ser na descrição de "curvas individuais" de respostas, ou ainda, em parâmetros individuais. Nesse contexto, o presente trabalho tem como objetivo estabelecer um estudo comparativo entre essas duas modelagens estatísticas, tendo como base de dados os resultados de um estudo longitudinal, envolvendo dados de CBT de leite de vaca, conduzido no interior do Rio Grande do Norte.

2 Material e métodos

2.1 Material

O estudo foi conduzido em oito propriedades localizadas na região agreste do Estado do Rio Grande do Norte. Nessas localidades foram coletadas mensalmente quatro medidas repetidas, uma por semana, de amostras de leite cru de tanques de resfriamento, durante o período de janeiro de 2010 a julho de 2011, totalizando 74 amostras. Primeiramente, o leite foi homogeneizado acionando-se o agitador dos tanques durante cinco minutos. Após esse período, o leite foi transferido com auxílio de uma concha de aço inoxidável para frascos plásticos esterilizados com volume de 40 mL, onde foram adicionadas quatro gotas do conservante azidiol®. Posteriormente, foram enviadas em caixas isotérmicas com gelo reciclável ao Laboratório da Clínica do Leite do Departamento de Zootecnia da Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. O estudo foi desenvolvido em três períodos. O período 1 compreende a fase de diagnóstico, no qual foi aplicado um questionário para avaliar a atual situação dos procedimentos de ordenha em cada sistema de produção. Este período compreendeu os meses de janeiro a abril de 2010. No segundo momento (período 2) foi realizada a capacitação dos colaboradores em boas práticas agropecuárias na produção de leite seguida de uma reavaliação dos diversos fatores relacionados aos procedimentos de ordenha. Esta fase durou 8 meses, referentes a maio a dezembro de 2010. E o terceiro momento (período 3) correspondeu à fase de acompanhamento das medidas da CBT, realizados durante o primeiro semestre de 2011. Dentre os diversos fatores relacionados ao manejo de ordenha utilizado nas propriedades, considerou-se a realização (ou não) das seguintes práticas: 1) teste da caneca; 2) pré-dipping; 3) secagem com papel toalha; 4) sanitização adequada da ordenha; 5) limpeza alcalina adequada da ordenha; 6) limpeza ácida adequada da ordenha;

7) resfriamento do tanque; 8) sanitização adequada do tanque; 9) limpeza adequada do tanque; 10) limpeza mecânica adequada do tanque. Para efeito de análise, agruparam-se estes fatores em três categorias: manejo da ordenha compreendendo os itens de 1 a 3; limpeza 1 correspondendo aos itens 4 a 6 e limpeza 2 englobando os itens de 7 a 10. Estas três covariáveis correspondem ao número de práticas realizadas. Sabe-se que estes dez itens são importantes na produção, o interesse ao especificar estas três covariáveis é avaliar se um número maior de procedimentos recomendados associados à ordenha, limpeza 1 e limpeza 2 auxilia na qualidade do leite. Considerando a IN 51, criou-se também uma variável resposta categorizada, que assume valor 1 se CBT < 750.000 ufc/mL e 0 se CBT ≥ 750.000 ufc/mL.

2.2 Modelagem estatística

Apresenta-se, a seguir, um resumo das duas metodologias para dados longitudinais abordadas nesse trabalho. Uma discussão comparativa mais aprofundada entre esses modelos pode ser vista em Zeger e Liang (1992), Zeger, Liang e Albert (1998), Diggle et al. (2002), dentre outros. Para estabelecer a notação, considere $\boldsymbol{y}_i$ $(i = 1, 2, \ldots, N)$ o vetor de variáveis respostas da i-ésima unidade (no presente trabalho as unidades correspondem às propriedades), de dimensões $n_i \times 1$, isto é, $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$, de tal forma que a cada ocasião t $(t = 1, 2, \ldots, n_i)$, uma observação y_{it} esteja associada a um vetor, $\boldsymbol{x}_i$, de variáveis explicativas ou covariáveis. Aqui, $\boldsymbol{y}_i$ representa um perfil individual de respostas, ou seja, um vetor em que cada componente é CBT da i-ésima propriedade na t-ésima ocasião, a qual estão associadas às covariáveis ordenha, limpeza 1 e limpeza 2, conforme descrito na seção 2.1. Como a variável resposta refere-se a uma contagem, é usual, supor que $Y_{it} = \text{CBT}_{it} \sim \text{Poisson}(\mu_{it})$.

Sabe-se que a distribuição de Poisson pertence à família exponencial canônica, ou seja:

$$f(y_{it}, \theta_{it}, \phi) = \exp\left\{ y_{it} \ln \mu_{it} - \mu_{it} - \ln(y_{it}) \right\} \tag{1}$$

Assume-se que $\theta_{it} = \ln(\mu_{it})$ é o parâmetro canônico, que não é de interesse prático, pois há um para cada observação e $b(\theta_{it}) = \exp(\theta_{it}) = \mu_{it}$ é uma função monótona e derivável. Assim, verifica-se que $\text{E}(Y_{it}) = b'(\theta_{it}) = \mu_{it}$ e $\text{Var}(Y_{it}) = a_i(\phi)b''(\theta_{it}) = a_i(\phi)V(\mu_{it}) = \mu_{it}$, em que $V(\mu_{it})$ é a função de variância. Nos modelos com dados de contagem, exceto em casos com superdispersão, deve-se observar que $\text{E}(Y_{it}) = \text{Var}(Y_{it})$ e $a_i(\phi) = \phi = 1$ é um parâmetro de escala suposto conhecido.

Para cada ocasião t tem-se, por definição, um modelo linear generalizado. O princípio do modelo marginal é estabelecer uma relação funcional para $\text{E}(Y_{it}) = \mu_{it}$. Para tanto, acrescenta-se uma função que faz a ligação entre a parte aleatória (1) e a parte sistemática que inclui as covariáveis do estudo:

$$\eta_{it} = \ln(\mu_{it}) = \beta_0 + \beta_{1j}\text{período}_j + \beta_2\text{ordenha} + \beta_3\text{limpeza 1} + \beta_4\text{limpeza 2}, \tag{2}$$

em que $\boldsymbol{\eta}_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_0, \beta_{1j}, \beta_2, \beta_3, \beta_4)'$ é o vetor de parâmetros desconhecidos e de interesse, em que $j = 1, 2, 3$ refere-se aos períodos de diagnóstico, capacitação e acompanhamento, respectivamente. Como se tratam de medidas repetidas no tempo, é de se esperar que não haja independência entre as observações (y_{it}) e isso deve ser levado em conta no processo de estimação dos parâmetros do modelo.

Para tratar deste tipo de problema, Liang e Zeger (1986) apresentaram um método de estimação intitulado de Equações de Estimação Generalizadas (EEG's), que requer a especificação de uma distribuição marginal e de uma estrutura de correlação. Esse procedimento, por sua vez, pode ser visto como uma extensão das equações de estimação de quase-verossimilhança para dados independentes. Proposto por Wedderburn (1974), o método da quase-verossimilhança requer apenas a especificação de uma função de variância para a variável resposta e uma relação funcional entre a resposta média e os parâmetros $\boldsymbol{\beta}$. McCullagh e Nelder (1983) mostram que para uma amostra aleatória de dimensão n a função escore para $\boldsymbol{\beta}$ é expressa por:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{1}{\phi}\boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \tag{3}$$

em que $\boldsymbol{V} = \text{diag}[V(\mu_i)]$, $\boldsymbol{D} = \dfrac{\partial \mu}{\partial \beta} = W^{1/2}V^{1/2}X$, $\boldsymbol{W} = \text{diag}\left[\dfrac{1}{V(\mu_i)}\left(\dfrac{\partial \mu_i}{\partial \eta_i}\right)^2\right]$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ e $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$. Deste modo, a solução $\hat{\boldsymbol{\beta}}$ é resultante da resolução do sistema $\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$, que é resolvido via processo iterativo de mínimos quadrados ponderados. Usando o método escore de Fisher, tem-se:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\boldsymbol{D}^{(m)}\boldsymbol{V}^{-1(m)}\boldsymbol{D}^{(m)}]^{-1}\boldsymbol{D}'^{(m)}\boldsymbol{V}^{-1(m)}[\boldsymbol{y} - \boldsymbol{\mu}^{(m)}], \tag{4}$$

que independe do parâmetro de dispersão, ϕ. Registra-se que o logaritmo da função de verossimilhança para distribuições da família exponencial aparece como um caso particular do logaritmo da função de quase-verossimilhança. Nesse contexto, as EEG's consistem em uma modificação no procedimento anterior a fim de obter estimativas confiáveis para os parâmetros, levando em conta uma estrutura de correlação. A ideia desse procedimento é introduzir, no processo de estimação dos parâmetros, uma matriz $\boldsymbol{R}(\boldsymbol{\alpha})$, simétrica e de dimensão t_i, que satisfaça as condições de uma matriz de correlação, sendo $\boldsymbol{\alpha}$ um vetor de dimensões $s \times 1$ que caracteriza completamente a matriz $\boldsymbol{R}(\boldsymbol{\alpha})$. Dessa forma, tem-se que:

$$\boldsymbol{\Omega}_i = Var(Y_i) = \phi\boldsymbol{V}_i^{1/2}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{V}_i^{1/2}, \tag{5}$$

em que Ω_i é a matriz de variâncias e covariâncias de Y_i se a verdadeira correlação entre as observações repetidas for dada por $\boldsymbol{R}(\boldsymbol{\alpha})$, em que ϕ denota um parâmetro de dispersão, que no presente estudo é suposto igual a 1. Quando não conhecido deve ser estimado assim como os demais parâmetros. Para se estimar o vetor de parâmetros $\boldsymbol{\beta}$, referentes aos efeitos das covariáveis, deve-se resolver o sistema:

$$\boldsymbol{U}_G(\boldsymbol{\beta}) = \boldsymbol{0}, \tag{6}$$

em que o índice G significa "generalizada" e é utilizado para se diferenciar da função escore dos modelos de quase-verossimilhança (equação 3). A equação 6 depende simultaneamente dos parâmetros ϕ (quando desconhecido) e $\boldsymbol{\alpha}$, que devem ser estimados separadamente. Para implementação computacional, deve-se atribuir estimativas inciais para $\boldsymbol{\beta}$, cujos valores serão posteriormente atualizados. Com isso, estimativas para os parâmetros $\boldsymbol{\alpha}$ e ϕ são obtidas como função dos resíduos de Pearson. A seguir, troca-se ϕ por $\hat{\phi}$ quando $\boldsymbol{\beta}$ é conhecido e $\boldsymbol{\alpha}$ por $\hat{\boldsymbol{\alpha}}$, quando $\boldsymbol{\beta}$ e ϕ são conhecidos. Portanto, as EEG's (7) tem a forma:

$$\sum_{i=1}^{N} \boldsymbol{U}_i \left[\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}} \left(\hat{\boldsymbol{\beta}}, \hat{\phi}(\boldsymbol{\beta}) \right) \right] = \mathbf{0}, \tag{7}$$

em que N corresponde ao número de unidades amostrais. A solução $\hat{\boldsymbol{\beta}}_G$ é obtida a partir do método escore de Fisher modificado (LIANG e ZEGER, 1986):

$$\boldsymbol{\beta}_G^{(m+1)} = \boldsymbol{\beta}_G^{(m)} + \left\{ \sum_{i=1}^{N} (\boldsymbol{D}_i')^{(m)} (\boldsymbol{\Omega}_i^{-1})^{(m)} (\boldsymbol{D}_i)^{(m)} \right\}^{-1} \left\{ \sum_{i=1}^{N} (\boldsymbol{D}_i')^{(m)} (\boldsymbol{\Omega}_i^{-1})^{(m)} (\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(m)}) \right\}. \tag{8}$$

Note que a equação 8 é uma generalização da equação 4 para uma estrutura com medidas repetidas no tempo. Liang e Zeger (1986) registram que as estimativas pontuais dos parâmetros do vetor $\boldsymbol{\beta}$, em geral, são invariantes para diversas estruturas de correlação, porém as estatísticas de precisão, como os erros-padrão, por exemplo, costumam ser afetados pela escolha da estrutura de correlação. Sob condições moderadas de regularidade, $\hat{\boldsymbol{\beta}}_G$ é consistente e converge em distribuição para uma normal multivariada. Maiores detalhes sobre esse processo podem ser vistos em Liang e Zeger (1986) e Zeger e Liang (1986).

Nos modelos marginais, os parâmetros $\boldsymbol{\beta}$ referem-se essencialmente a efeitos fixos e em decorrência disso têm interpretações para a média populacional, isto é, são modelos da classe PA (*population-averaged*) (ZEGER, LIANG e ALBERT, 1988). Da relação estabelecida em (2) pode-se estimar as médias por $\hat{\mu}_{it} = g^{-1}(\hat{\eta}) = \exp(\hat{\eta})$.

De modo similar, por meio das EEG's, pode-se ajustar um modelo marginal para descrever a probabilidade (π_{it}) de que as fazendas estejam com a CBT dentro dos limites aceitáveis pela Instrução Normativa 51 (IN 51), como função das variáveis explicativas incluídas no estudo. Para tanto, assume-se uma variável resposta categorizada em 1 se CBT $<$ 750.000 ufc/mL e 0 se CBT $\geq$ 750.000 ufc/mL e, considera-se que esta "nova" variável resposta tem distribuição Bernoulli com parâmetro π_{it}. O modelo marginal adotando-se função de ligação canônica (logística) é:

$$\eta_{it} = \ln\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = \beta_0 + \beta_{1j}\text{ordenha}_j + \beta_2\text{limpeza1} + \beta_2\text{limpeza2}, \tag{9}$$

em que $j = 1, 2, 3$ categorias para a covariável período. Assumindo o preditor linear (9), pode-se modelar π_{it} pela expressão:

$$\mu_{it} = \pi_{it} = \frac{\exp(\eta_{it})}{1+\exp(\eta_{it})}. \tag{10}$$

O procedimento básico para se ajustar modelos marginais para respostas correlacionadas, por meio das EEG's, atualmente, encontra-se implementado em vários *softwares*, como no SAS (proc *genmod*) e no R (pacotes *gee* e *geepack*). Em particular, pacote *geepack* do *software* R permite, além da técnica básica de ajuste de modelos, algumas implementações adicionais, como por exemplo, a definição de uma estrutura de correlação do "usuário". Halekoh e Hojsgaard (2006) fazem uma descrição bem aplicada dos principais aspectos adicionais deste pacote em relação aos demais aplicativos computacionais.

Por outro lado, num modelo de efeitos aleatórios assume-se que a variável resposta é uma função de uma ou mais variáveis explanatórias com coeficientes de regressão que variam de um indivíduo para o outro, devido a fatores não mensuráveis. Assim, nesses modelos os coeficientes de regressão são resultantes de uma contribuição média mais um efeito específico inerente a cada indivíduo, razão pela qual são classificados como modelos da classe SS (*subject-specific*)(ZEGER, LIANG e ALBERT, 1988).

Na definição do modelo linear generalizado misto (MLGM), assume-se que dado um ou mais efeitos aleatórios as variáveis respostas são variáveis aleatórias independentes, seguindo um modelo linear generalizado (DIGGLE et.al., 2002; MOLENBERGHS e VERBEKE, 2005). Assim, neste estudo para a CBT, tem-se:

$$Y_{it}|\boldsymbol{b}_i \sim \text{Poisson}(\mu_{it})$$

$$f_{Y_{it}|\boldsymbol{b}_i}(y_{it}|\boldsymbol{b}_i) = \exp\left\{ y_{it}\ln\mu_{it} - \mu_{it} - \ln(y_{it}) \right\}$$

Porém, em comparação ao modelo marginal, deve-se observar que:

$$\mathbf{E}(Y_{it}|\boldsymbol{b}_i) = \mu_{it}$$

é a média condicional da variável aleatória Y_{it}, funcionalmente ligada à parte sistemática por uma função de ligação:

$$g(\mathbf{E}(Y_{it}|\boldsymbol{b}_i)) = g(\mu_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{b}_i, \tag{11}$$

em que $\boldsymbol{x}_{it}$ é a i-ésima linha da matriz de delineamento associada aos efeitos fixos, $\boldsymbol{\beta}$ é o vetor de parâmetros dos efeitos fixos, $\boldsymbol{z}_{it}$ é a i-ésima linha da matriz do modelo associada aos efeitos aleatórios e $\boldsymbol{b}_i$ é o vetor dos efeitos aleatórios. A parte sistemática do modelo (11) inclui tanto os efeitos fixos como os aleatórios. Assume-se que o conjunto de todos os coeficientes $\boldsymbol{b}_i$ constitui-se numa amostra de uma variável aleatória q-dimensional:

$$\boldsymbol{b}_i \sim f_b(\boldsymbol{b}),$$

sendo usual (mas não restrita) a escolha da distribuição normal multivariada para esses efeitos aleatórios, ou seja, $\boldsymbol{b}_i \sim N_q(\mathbf{0}, \boldsymbol{G})$. O objetivo da análise é estimar os

coeficientes dos efeitos fixos, $\boldsymbol{\beta}$, os parâmetros de f_b e, em alguns casos, o parâmetro de escala, ϕ. Por exemplo, no presente estudo, se considerarmos fazenda e período como de efeitos aleatórios, o preditor linear terá a seguinte estrutura:

$$\eta_{it} = (\beta_0 + b_{i0}) + (\beta_{1j} + b_{i1j})\text{periodo}_j + \beta_2\text{ordenha} + \beta_3 \text{ limpeza 1} + \beta_4\text{limpeza 2}, \quad (12)$$

em que $\boldsymbol{b} = (b_{i0}, b_{i1j})'$, $j = 1, 2, 3$ períodos, denota o vetor referente aos efeitos aleatórios de fazenda e período, respectivamente. A escolha de fazenda e período como efeitos aleatórios neste caso, pode ser justificada pela heterogeneidade entre as localidades e períodos e também pelo fato de se poder representar esses efeitos como uma amostra de uma distribuição de probabilidade. Nesse sentido, a correlação entre as observações é resultante dessa variável aleatória latente. Registra-se, adicionalmente, que a significância estatística para esses efeitos aleatórios pode ser verificada por meio de um teste de razão de verossimilhanças.

A literatura com relação à estimação dos parâmetros do modelo linear generalizado misto (11) é ampla, sendo usual o ajuste por meio de máxima verossimilhança (BRESLOW, CLAYTON, 1993; PINHEIRO, BATES, 2000; DIGGLE et al., 2002). A função de verossimilhança para o vetor de parâmetros desconhecidos $\boldsymbol{\psi}$, que inclui ambos os elementos de $\boldsymbol{\beta}$ e $\boldsymbol{G}$ é:

$$L(\boldsymbol{\psi}, \phi, \boldsymbol{y}) = \prod_{i=1}^{N} \int \prod_{t=1}^{n_i} f(y_{it} \mid \boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i, \boldsymbol{G}) d\boldsymbol{b}_i, \quad (13)$$

que é obtida integrando a distribuição conjunta de $(\boldsymbol{Y}, \boldsymbol{b})$ em relação a $\boldsymbol{b}$. Na equação 13, N representa o total de propriedades e n_i é o número de observações repetidas em cada uma delas.

O problema para maximizar a função (13) é a presença de N integrais duplas sob os efeitos aleatórios, $\boldsymbol{b}_i$, pois conforme descreve a equação 12, no estudo em questão, o vetor $\boldsymbol{b}_i$ tem dimensão dois. No caso do modelo linear com distribuição normal e função de ligação identidade, essas integrais podem ser resolvidas analiticamente. Entretanto, em grande parte dos casos, a integral dada em (13) não tem solução analítica exata, sendo necessários métodos numéricos para obter uma solução aproximada. Molenberhgs e Verbeke (2005) salientam que, em geral, as aproximações numéricas podem ser divididas em três grupos: a aproximação dos integrandos, a aproximação dos dados e a aproximação da integral.

Em particular, a técnica de aproximação dos dados baseia-se na sua decomposição nos termos média e erro, com uma expansão em séries de Taylor em torno da média da decomposição, que é uma função não linear do preditor linear. Sem perda de generalidade, considere a decomposição:

$$Y_{it} = \mu_{it} + \varepsilon_{it} = g^{-1}(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{b}_i) + \varepsilon_{it}, \quad (14)$$

em que $g^{-1}(.)$ é a inversa da função de ligação e o termo ε_{it} tem distribuição apropriada com variância igual a $\text{Var}(Y_{it} \mid \boldsymbol{b}_i) = \phi \text{V}(\mu_{it})$. As técnicas conhecidas como quase-verossimilhança penalizada (QVP) e quase-verossimilhança marginal

(QVM) podem ser implementadas para a aproximação de μ_{it} na expressão (14). Embora esses métodos sejam distintos, a ideia básica dos processos é permitir uma aproximação para a contribuição individual na função de verossimilhança. A implementação computacional para o ajuste dos modelos de efeitos mistos requer que a distribuição da variável resposta, da função de ligação e da distribuição dos efeitos aleatórios sejam corretamente especificados. Os programas desenvolvidos nos *softwares* estatísticos para o ajuste de modelos lineares generalizados mistos permitem estimar os parâmetros referentes aos efeitos fixos e aleatórios, bem como a variância desses efeitos. Como opções computacionais disponíveis, tem-se, por exemplo, o procedimento *GLIMMIX* no SAS e o pacote *glmmPQL* no *software* R. Nesses pacotes, a integral da função de verossimilhança é feita a partir da aproximação dos dados. Nos procedimentos *NLMIXED* do SAS e no *glmmML* do R, a solução da integral se dá por quadratura gaussiana. Maiores detalhes sobre esses procedimentos podem ser encontrados em Breslow e Clayton (1993), Pinheiro e Bates (2000) dentre outros.

No presente trabalho, para o ajuste dos modelos marginal (2) e misto (12) utilizou-se o *software* R, versão 2.13. Os modelos marginais foram ajustados com auxílio dos pacotes *gee* e *geepack*, enquanto que o modelo misto foi ajustado por meio do pacote *glmmPQL* .

3 Resultados e discussão

Os dados descritos na seção 2.1 foram analisados segundo as duas metodologias descritas, afim de subsidiar uma discussão comparativa dos resultados. Inicialmente realizou-se uma análise exploratória da CBT, por período e propriedade.

Tabela 1 - Estatísticas descritivas da contagem bacteriana total (em ufc/mL) em amostras de leite produzido no agreste do estado do Rio Grande do Norte nos anos de 2010 e 2011, por período

Período	Amostras	Média	Erro-padrão	Coeficiente de Variação (%)
Diagnóstico	59	2.049	425,45	$159,5$
Capacitação	179	440	113,08	$343,8$
Acompanhamento	203	416	25,75	$88,4$

A média, o erro-padrão e o coeficiente de variação dos dados da CBT estão apresentados na Tabela 1. Observou-se que, no período de diagnóstico, a média da CBT estava bastante elevada, indicando que o leite estava sendo produzido fora dos padrões aceitáveis de higiene. Ocorreu uma redução nos níveis médios de CBT nas fases de capacitação e acompanhamento, indicando que as propriedades passaram a atender, em média, o limite máximo de 750.000 ufc/mL estabelecido pela IN 51. Após a capacitação, além da redução dos níveis médios de CBT, houve também uma redução na variabilidade do processo de produção, revelando maior

homogeneidade dos resultados entre as propriedades. Observa-se, também, que a redução da variação acompanha a redução da média, o que indica uma relação linear entre média e variância, sugerindo a utilização da distribuição Poisson.

Considerando-se o percentual de visitas às propriedades, verificou-se que na fase de diagnóstico, 52,5% das visitas às fazendas apresentaram CBT adequadas à IN 51, passando nas fases de capacitação e acompanhamento para aproximadamente 90% (Tabela 2). Dessa forma, pode-se ressaltar não só a importância do uso dos procedimentos adequados, mas também das recomendações em boas práticas agropecuárias por meio da capacitação dos profissionais que lidam diretamente com a produção.

Tabela 2 - Adequação das propriedades à IN51 quanto à contagem bacteriana total, nos três períodos, no agreste do Estado do Rio Grande do Norte, 2010 a 2011

	$\geq$ 750.000 ufc/mL		$<$ 750.000 ufc/mL	
Período	Amostras	%	Amostras	%
Diagnóstico	28	47,5	31	52,5
Capacitação	21	11,7	158	88,3
Acompanhamento	21	10,3	182	89,6

A Figura 1 mostra o gráfico dos perfis individuais de resposta, para cada localidade. Nota-se que a CBT tem distribuição heterogênea entre as propriedades com valores variando entre 2 ufc/mL (mínimo) e 3.250 ufc/mL (máximo). Observa-se que na fase de diagnóstico, os menores valores médios são observados nas propriedades 6, 4, 2 e 3, respectivamente. Na fase de capacitação, os menores valores médios são observados nas propriedades 6, 4, 3 e 2, respectivamente. Já na fase de acompanhamento, as memores médias estão associadas às propriedades 3, 2, 5, 4 e 6. Este gráfico sugere um comportamento não linear, com heterogeneidade entre as propriedades, sinalizando a possibilidade de inclusão de efeitos aleatórios no modelo estatístico.

A fim de explicar a estrutura funcional e de dependência desses dados, ajustaram-se modelos marginais para CBT, usando função de ligação logarítmica e a distribuição Poisson para as contagens de bactérias, conforme apresentado em (2). Algumas estruturas de correlação foram testadas para descrever a dependência entre as observações, entre elas, a simétrica ou permutável (*exchangeable*), *m*-dependente, não estruturada (*unstructured*) e auto-regressiva (AR). Observou-se que as estimativas dos parâmetros dos modelos ajustados com estas quatro estruturas de correlação foram muito similares, porém optou-se pela escolha do modelo com a estrutura de correlação AR1, tendo em vista que o seu ajuste levou à menor função desvio (*deviance* residual).

A Tabela 3 apresenta as estimativas dos parâmetros, os erros padrões e as estatísticas do teste de Wald para cada um dos coeficientes. A estimativa do parâmetro de correlação foi $\hat{\alpha} = 0,80$ com erro-padrão $0,07$. Os resultados mostram

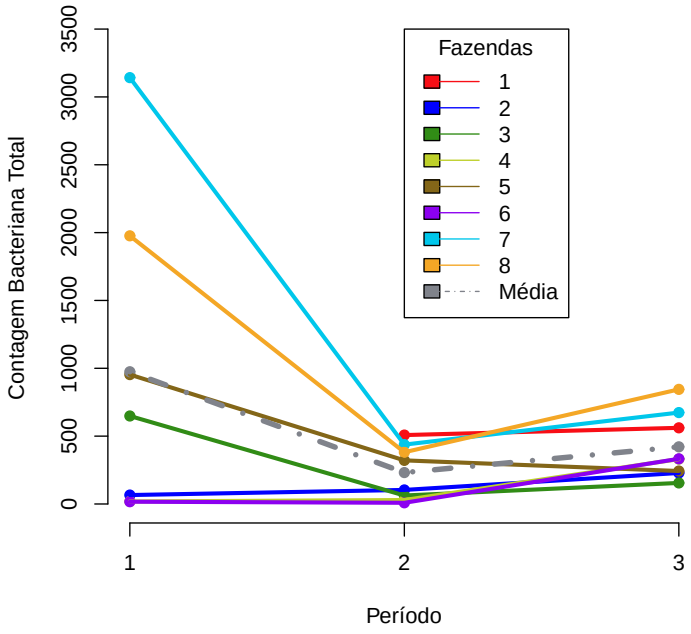

Figura 1 - Perfis individuais de contagem bacteriana total.

Tabela 3 - Estimativas dos parâmetros do modelo marginal referente à contagem bacteriana total com estrutura de correlação AR-1

Parâmetro	Estimativa	Erro-padrão	Estatística de Wald	Nível descritivo
intercepto	6,98	0,52	181,46	$< 0,001$
período 1	-0,37	0,48	0,59	0,441
período 2	-0,06	0,47	0,02	0,886
ordenha	0,53	0,21	6,35	0,012
limpeza 1	-0,26	0,05	24,77	$< 0,001$
limpeza 2	-0,06	0,07	0,66	0,417

que os efeitos referentes à ordenha (teste da caneca e secagem das tetas) e limpeza 1 (limpeza dos equipamentos de ordenha) são significativos a 5%. Os sinais negativos dos coeficientes associados às covariáveis limpeza 1 e limpeza 2 indicam que quanto maior o número de procedimentos realizados, menores serão os níveis médios de contaminação do leite. O efeito de período não foi estatisticamente significativo

($p > 0,10$), sendo que a fase de acompanhamento (período 3) foi tomada como categoria de referência no processo de estimação.

Na ótica dos modelos lineares generalizados de efeitos mistos ajustaram-se dois modelos, um considerando apenas fazenda como efeito aleatório e outro considerando fazenda e período como efeitos aleatórios, sendo que o melhor ajuste foi o do modelo com efeito aleatório em fazenda. Constatou-se que a variância do efeito aleatório em fazenda foi estatisticamente significativa, justificando a inclusão desse efeito para explicar a contagem de bactérias presentes no leite de vaca. Para ambos os modelos foram testadas algumas intererações de interesse prático como, por exemplo, entre limpeza 1 e limpeza 2, período e limpeza, ordenha e limpeza, porém não foram significativas. A estrutura funcional do modelo a ser considerado é dada pela equação 15 e as estimativas dos seus parâmetros são apresentadas na Tabela 4.

$$\eta_{it} = (\beta_0 + b_{i0}) + \beta_{1j}\text{periodo}_j + \beta_2\text{ordenha} + \beta_3 \text{ limpeza 1} + \beta_4\text{limpeza 2}, \qquad (15)$$

em que $j = 1, 2, 3$ períodos. No processo de estimação, a categoria 3 de período é tomada como referência.

Tabela 4 - Estimativas dos efeitos fixos do modelo misto referente à contagem bacteriana total, com efeitos aleatórios em fazenda e período

Parâmetro	Estimativa	Erro-padrão	Estatística t	Nível descritivo
intercepto	9,10	0,49	18,57	$< 0,001$
período 1	0,02	0,35	0,07	$0,946$
período 2	0,66	0,34	1,90	$0,058$
ordenha	-0,25	0,14	-1,72	$0,085$
limpeza 1	-0,17	0,04	-4,00	$< 0,001$
limpeza 2	-0,24	0,07	-3,39	$< 0,001$

De acordo com os resultados obtidos, nota-se que diferentemente do modelo marginal, o efeito de limpeza 2 (que representa as questões de higiene e resfriamento do tanque) foi estatisticamente significativo, assim como o efeito de período 2 (a 6%) captando dessa forma, o efeito do treinamento e capacitação dos colaboradores. Adicionalmente, dada a presença do efeito aleatório em fazenda, pode-se pensar em níveis de contaminação individuais, ou seja, a CBT também pode ser explicada por um efeito adicional intrínseco às propriedades. Desta forma, temos um parâmetro a mais estimado, correspondendo à variância do efeito aleatório de intercepto (fazenda), cuja estimativa do erro-padrão foi $0,64$. Esse termo capta a variabilidade não mensurável entre as fazendas.

Adicionalmente, ajustou-se um modelo marginal de regressão logística para explicar a probabilidade (π_{it}) de que as fazendas estejam com a CBT dentro dos limites aceitáveis pela Instrução Normativa 51 (IN 51), conforme descrito na seção 2.2. Assumindo o preditor linear dado pela equação (9), algumas estruturas de correlação foram avaliadas e, com base na precisão dos erros-padrões e a função

desvio, selecionou-se a estrutura de correlação AR-1. A Tabela 5 apresenta as estimativas dos parâmetros e os respectivos resultados dos testes de Wald para cada um dos coeficientes.

Tabela 5 - Estimativas para os parâmetros do modelo marginal de regressão logística com estrutura de correlação AR-1

Parâmetro	Estimativa	Erro-padrão	Estatística de Wald	Nível descritivo
Intercepto	-4,29	1,29	11,04	$< 0,001$
Período 1	-1,40	0,51	7,50	$0,006$
Período 2	-1,07	1,06	1,02	$0,312$
Ordenha	1,75	0,56	9,84	$0,001$
Limpeza 1	-0,05	0,03	2,18	$0,139$
Limpeza 2	0,38	0,13	8,45	$0,003$

A estimativa do parâmetro de correlação foi $\hat{\alpha} = 0,75$ com erro padrão $0,19$. Os resultados mostram significância estatística dos efeitos de ordenha, período 1 e limpeza 2, mostrando a importância com os cuidados na ordenha, bem como a limpeza e resfriamento do tanque em que o leite fica armazenado, sendo que a covariável ordenha é a que mais contribui para se ter uma probabilidade maior de que a propriedade esteja com CBT abaixo do limite de 750.000 ufc/mL.

A seguir, ainda considerando a estrutura de regressão logística, foram ajustados modelos de efeitos mistos, considerando efeitos aleatórios nos termos fazenda e período. Porém, constatou-se a pertinência do efeito aleatório apenas no termo fazenda. O efeito de período também não foi significativo. Assim, a estrutura funcional do modelo de regressão logística com intercepto aleatório ajustado foi:

$$\eta_{it} = \ln\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = (\beta_0 + b_{i0}) + \beta_1 \text{ limpeza 1} + \beta_2 \text{limpeza 2}, \tag{16}$$

em que $b_{i0} \sim \mathrm{N}(0, \sigma^2)$. As estimativas dos parâmetros para o modelo (16) bem como os testes de significância para cada coeficiente individual estão dispostos na Tabela 6.

Tabela 6 - Estimativas dos efeitos fixos do modelo misto de regressão logística (16), considerando efeito aleatório em fazenda

Parâmetro	Estimativa	Erro-padrão	Estatística t	Nível descritivo
Intercepto	-7,70	2,51	-3,06	$0,002$
Ordenha	1,09	0,37	2,94	$0,003$
Limpeza 1	0,30	0,36	0,81	$0,418$
Limpeza 2	0,78	0,20	3,82	$< 0,001$

Embora neste caso não se tenha o efeito de período, os resultados da Tabela 6 concordam com aqueles obtidos com o uso das EGG's, uma vez que foram significativos os efeitos de ordenha e limpeza 2. Adicionalmente, esse modelo estima um parâmetro a mais referente à variância do efeito aleatório de fazenda, cujo erro-padrão estimado foi $2,48$. Esse termo de variabilidade capta a heterogeneidade entre as propriedades leiteiras.

Verificou-se que a capacitação dos profissionais do manejo permitiu que nas etapas seguintes houvesse um maior uso dos procedimentos recomendados pela IN 51. Constatou-se que quanto maior o número de procedimentos adequados maior é a probabilidade das fazendas em atenderem às metas recomendadas. O estudo mostra que uma parcela considerável das propriedades atingiu os limites ideais da CBT estabelecidos pela IN 51 quando houve uma adoção de procedimentos de higiene no processo de obtenção de leite, bem como a refrigeração imediata do mesmo. A maioria das propriedades necessita, portanto, canalizar esforços na melhoria das condições higiênicas durante a ordenha e o armazenamento do leite, e na sua refrigeração rápida na temperatura de 4^oC, para reduzir os níveis de contaminação microbiana e atender ao padrão definitivo para 2011.

Considerações finais

Nesse estudo, tanto o modelo marginal como o modelo misto possibilitaram destacar a importância dos procedimentos adequados no manejo da ordenha e na limpeza do tanque de resfriamento para a redução do número de bactérias no leite de vaca. Nos dois casos, assume-se que a variável resposta tem distribuição Poisson. Há de se ressaltar que uma pressuposição usual em modelos com esta variável aleatória é de que a média seja aproximadamente igual à variância. Porém, quando é assumida a distribuição Poisson para a variável resposta, pode ocorrer, em alguns casos, uma variância muito maior do que aquela esperada pelo modelo. Esse fato pode ser indicativo de superdispersão e uma provável causa dessa violação é a existência de heterogeneidade entre as unidades amostrais. Algumas alternativas para o problema da superdispersão são o emprego da distribuição binomial negativa ou a inclusão de efeitos aleatórios pertinentes para acomodar a heterogeneidade excessiva, a qual pode comprometer as estatísticas de precisão do modelo. Neste trabalho, optou-se pela segunda abordagem, ou seja, assume-se que a distribuição condicional da variável resposta dado um efeito aleatório (variável aleatória latente) tem distribuição Poisson. Neste caso, a variância dos efeitos aleatórios pode ser usada para explicar a variabilidade inter-unidades amostrais, responsável, em grande parte dos casos, pelo fenômeno da superdispersão. Esta é, portanto, uma das vantagens do uso dos modelos mistos para dados de contagem.

Comparando os resultados expressos pelas Tabelas 3 e 4 nota-se, especificamente, diferenças em relação aos efeitos de período e limpeza 2. Como explicar os resultados distintos obtidos para o mesmo conjunto de dados sobre contagem de bactérias? A resposta a essa questão está na fundamentação metodológica que responde pelas questões do estudo, as quais, por sua vez, devem ser estabelecidas

no planejamento do experimento. As duas abordagens metodológicas ilustradas nesse trabalho, embora sejam apropriadas para dados correlacionados, tratam de hipóteses distintas e os processos estimação dos parâmetros, embora centrados na teoria da máxima verossimilhança, não são idênticos. Na abordagem dos modelos marginais, a resposta média populacional é uma função das covariáveis consideradas nesse estudo. Isso é pertinente quando as hipóteses visam testar os efeitos dos fatores sobre a resposta média populacional. Além disso, em um modelo marginal a correlação é tratada como um parâmetro *nuisance*, haja vista que a estrutura de correlação de trabalho, mesmo que mal especificada, é usada a fim de se obterem estimativas mais confiáveis para os parâmetros, resultando em menores erros padrões o que certamente permite a construção de intervalos de confiança mais precisos para os mesmos. Na prática, podemos dizer que se o objetivo do pesquisador é simplesmente inferir sobre a contagem do número médio de bactérias, em termos da população de fazendas, período não é relevante, porém as covariáveis limpeza 1 e ordenha tem implicação direta no aumento ou diminuição desse índice, ou seja, é pertinente a importância das práticas do teste da caneca, pré-*dipping* e secagem das tetas para uma melhor qualidade na produção do leite. Ao passo que, se o interesse do pesquisador está focado nos indivíduos, nesse estudo representado pelas propriedades, o modelo de efeitos mistos é mais apropriado. Note que ao considerarmos a presença de efeito aleatório no intercepto, isto é, adotarmos a estrutura funcional de um modelo linear generalizado misto, têm-se hipóteses envolvendo parâmetros individuais. Isso pode justificar as diferenças encontradas com relação aos modelos marginais para os efeitos de algumas covariáveis, tanto para o modelo de dados de contagem de CBT quanto para o modelo de regressão logística. Por fim, quando se ajustam modelos marginais e de efeitos mistos para dados correlacionados, não é raro, obterem-se resultados não concordantes por essas duas metodologias. Nesse contexto, deve-se estar atento às hipóteses do estudo e à necessidade, ou não, de se considerar efeitos aleatórios pertinentes.

LARA, I. A. R., SPYRIDES, M. H. C.; GUERRA, G. G.; RANGEL, A. H. N. Comparative analysis of models for longitudinal data in the study of counting the number of bacteria present in cow's milk. *Rev. Bras. Biom.,* São Paulo, v.30, n.4, p.492-508, 2012.

▪ *ABSTRACT: This work describes a study conducted in the state of Rio Grande do Norte - Brazil, on the total bacterial count in cows' milk in order to assess the impacts of management procedures for milking, cleaning of equipment and the cold water tank in this variable during the process of milk production. The objective was to compare two methodological approaches for analyzing longitudinal data: the marginal models and mixed effects models. A comparative discussion regarding the use of these models, estimation and interpretation of the results is presented. In practical terms, the results showed that the handling procedures for milking and the cleanliness of the cooling tank are the ones that contribute to a further reduction in the levels of contamination of milk.*

▪ *KEYWORDS: TBC; longitudinal data; generalized linear models; GEE's; random effects.*

Referências

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Instrução Normativa nº* 51, de 18/09/2002. Diário Oficial da União, Brasília, 20 set. 2002. Seção I, p. 13 – 22, set-2002.

BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, Boston, v. 88, n. 421, p. 9-25, 1993.

DIGGLE, P.J.; HEAGERTY, P.J.; LIANG, K.Y.; ZEGER, S.L. *Analysis of longitudinal data*. New York: Oxford University Press, 2002, 379 p.

DURR, J.W. Controle de qualidade e aumento da competitividade da indústria láctea. *Congresso Pan-Americano do Leite* - Tendências e avanços do agronegócio do leite nas Américas: mais leite = mais saúde. Ed. Carlos Eugênio Martins et al., Porto Alegre-RS, 2006. (CD ROM)

HALEKOH, U.; HOJSGAARD, S. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, v.15, p.1-10, 2006.

LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. *Biometrika*, Cambridge, v. 73, n. 1, p. 13-22, 1986.

McCULLAGH, P.; NELDER, J.A. Quasi-likelihood functions. **Annals of Statistics**, Hayward, v. 11, p. 59 – 67, 1983.

MOLENBERGHS, G. VERBEKE, G. *Models for discrete longitudinal data*. New York: Springer-Verlag, 2005, 683 p.

PAULA, G.A. **Modelos de Regressão com Apoio Computacional**. São Paulo: IME, 2004, 245 p.

PINHEIRO, J.C.; BATES, D.M. *Mixed-effects in S and S-PLUS*, New York: Springer-Verlag, 2000, 552 p.

R Development Core Team. R: A language and environment for statistical computing 2.13.1. Vienna, Austria, 2011. Disponível em: http://www.R-project.org. Acesso em 22 out. 2011.

SANTOS, L.G.C.; NADAI, F, E. A.; BACCHI, M. A.; SARRIÉS, G. A.; BLUMER, L; BARBOSA J. F. Chemical composition of bovine milk from Minas Gerais State, Brazil. *Journal of Radioanalytical and Nuclear Chemistry*, v.282, 117-123, 2009.

SANTOS, M.V.; FONSECA, L.F.L. Importância e efeito de bactérias psicotróficas sobre a qualidade do leite. *Revista Higiene Alimentar*, v.15, p.13-19, 2001.

SAS Institute Inc. *The GLIMMIX Procedure*. Carry NC, 2004.

SOUTO, L.I.M.; SAKATA, S.T.; MINIGAWA, C.Y.; TELLES, E.O.; GARBUGLIO, M.A.; BENITES, N.R. Qualidade higiênico-sanitária do leite cru produzido no estado de São Paulo, Brasil. *Veterinária e Zootecnia*, Botucatu, v. 16, n. 3, p. 491-499, 2009.

WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, Cambridge, v. 61, p. 439 – 447, 1974.

ZANELA, M.; FISCHER, V.; RIBEIRO, M.; STUMPF JUNIOR, W; ZANELA, C.; MARQUES, L.; MARTINS, P. Qualidade do leite em sistemas de produção na região Sul do Rio Grande do Sul. Brasília, *Pesquisa Agropecuária Brasileira*, jan., v. 41, n. 1, 2006.

ZEGER, S.L.; LIANG, K.Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, Washington, v. 42, p. 121-130, 1986.

ZEGER, S.L.; LIANG, K.Y. An overview of methods for the analysis of longitudinal data. **Statistics in Medicine**, Chichester, v. 11, p. 1825 – 1839, 1992.

ZEGER, S. L.; LIANG, K. Y.; ALBERT, P. S. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. v. 44, p. 1049-1060, 1988.

Recebido em 04.12.2012.

Aprovado após revisão em 03.06.2013.