

MODELOS LINEARES MISTOS: UMA APLICAÇÃO NA CURVA DE LACTAÇÃO DE VACAS DA RAÇA SINDI

Tadeu Rodrigues da COSTA¹

Laélia CAMPOS²

Francisco José de Azevedo CYSNEIROS³

Moacyr CUNHA FILHO⁴

- **RESUMO:** Curvas de lactação representam de forma gráfica a produção de leite individual ou de um rebanho durante seu período de lactação e carregam uma importância indiscutível no que tange ao entendimento do comportamento da produção daquele determinado rebanho, sendo fundamental na tomada de decisões acerca das condições do mesmo. Dentre as muitas raças leiteiras existentes hoje no Brasil, a raça Sindi tem um papel especial na produção de leite por se adaptar à rigurosidade do clima semiárido, tornando-se uma alternativa viável para a produção de leite no Nordeste. Nesse sentido, o objetivo desse trabalho foi o de aplicar um modelo linear misto em um banco de dados de um rebanho da raça Sindi com o intuito de verificar a produção de leite e a previsão individual dos animais desse rebanho. Além disso, foi feita a análise de resíduos e sensibilidade para verificação da adequabilidade do modelo. Como resultado principal, o modelo linear misto foi considerado adequado para estudar o comportamento individual de cada animal e a previsão da produção de leite.
- **PALAVRAS-CHAVE:** Análise de resíduos e sensibilidade; curva de lactação; modelo linear misto.

¹Universidade Federal Rural de Pernambuco – UFRPE, Departamento de Estatística e Informática, CEP: 52.171-900, Recife, PE, Brasil. E-mail: *tadeudrigues@gmail.com*

²Universidade Federal de Sergipe, Departamento de Física – CCET, CEP: 49.100-000, São Cristóvão, SE, Brasil. E-mail: *lpcbcampos@gmail.com*

³Universidade Federal de Pernambuco – UFPE, Departamento de Estatística – CCEN, CEP: 50.740-540, Recife, PE, Brasil. E-mail: *cysneiros@de.ufpe.br*

⁴Universidade Federal de Pernambuco – UFPE, Departamento de Economia, CEP: 50.740-590, Recife, PE, Brasil. E-mail: *moacyr@deinfo.ufrpe.br*

1 Introdução

No Brasil, existe uma grande variedade de raças leiteiras, sendo a Sindi uma das raças que melhor se adapta ao clima semiárido da região Nordeste. Essa raça teve sua origem onde hoje situa-se o Paquistão, na província do Sind, local que determinou seu nome. Também é conhecida como Red Sindhi e muitas vezes chamada de gado vermelho ou Sindi vermelho. Na Índia, é considerada a raça mais pura dentre as raças existentes. A raça apresenta pequeno porte com cabeça pequena, chifres grossos na base crescendo para os lados e curvando-se para cima, orelhas caídas de tamanho mediano, olhos escuros, quartos traseiros caídos e arredondados, membros pequenos e delicados bem apumados e geralmente apresentam tetas grossas. É uma raça resistente a doenças, inclusive à febre aftosa.

A produção de leite pode ser estudada por meio da curva de lactação. Essa curva expressa o comportamento da produção de leite durante a lactação do rebanho. O estudo dessa curva, como apresenta Madsen (1975), permite, por exemplo, observar a velocidade em que a produção de um rebanho ou animal decai com o decorrer do período correspondente a esse processo. Dessa forma, o produtor pode interferir no processo produtivo, fazendo ajustes na alimentação, adequando-a às necessidades observadas nesse período.

A classe de modelos denominada modelos mistos, em sua forma linear, não-linear e generalizada, é bastante utilizada em ajustes que envolvem dados repetidos ou quando o interesse é adicionar ao modelo em estudo o componente “indivíduo”, ou seja, incorporar ao modelo a variabilidade causada pelo indivíduo em si. Nesta classe de modelos, além de incorporar essa informação, é possível também realizar estimação para cada indivíduo (entenda-se indivíduo como sendo a unidade experimental), além da estimação do comportamento médio como geralmente é feito nos modelos usuais. A possibilidade de estudar e estimar individualmente estabelece parâmetros de grande importância nas tomadas de decisões relacionadas ao modelo. Nesse contexto, os modelos mistos se tornam fundamentais no estudo da curva de lactação de uma raça, pois possibilitam estudar o comportamento de cada animal amostrado e a previsão de sua produção, permitindo ao produtor tomar decisões quanto ao seu rebanho tanto de forma média como individual.

Como parte integrante e fundamental do ajuste de um modelo, seja ele misto, generalizado, em suas formas lineares ou não-lineares e em outras tantas estruturas de modelos existentes na literatura, a análise de resíduos e sensibilidade visa verificar a adequabilidade do modelo ajustado. Tal análise tem como objetivo averiguar possíveis afastamentos das suposições envolvidas na definição do modelo e observações que possam exercer influência, de forma desproporcional, nas estimativas dos parâmetros e previsões. Na classe dos modelos lineares mistos, a análise de resíduos e sensibilidade tem o papel determinante de verificar se as suposições de linearidade, normalidade e homoscedasticidade são atendidas, além de identificar possíveis observações e unidades experimentais que estejam influenciando, de forma desproporcional, nas estimativas dos efeitos fixos, previsões dos efeitos aleatórios e das previsões individuais e médias.

Considerando a importância da Raça Sindí no Nordeste e a difusão e aplicação de metodologias que visem melhor adaptabilidade ao estudo de determinadas curvas de lactação, esse estudo vem propor o ajuste de um modelo linear misto à curva de lactação de um rebanho da raça Sindí proveniente do semiárido paraibano. Além disso, realiza a análise de resíduos e sensibilidade, tomando como referência a metodologia apresentada por Nobre (2004).

2 Curva de lactação

Brody *et al.* (1923) introduziu o estudo da curva de lactação por meio de modelos matemáticos, apresentando uma forma de estudar a produção de leite utilizando um gráfico denominado curva de lactação que, segundo Ali e Schaeffer (1987), é a representação gráfica da produção de leite de uma vaca ou rebanho, tomando como referência o tempo ou, mais precisamente, o período de lactação. Segundo Quintero *et al.* (2007), a curva de lactação é a representação de um processo biológico que pode ser afetado por certas características como nível de produção inicial e persistência de lactação que, segundo Cobuci *et al.* (2003), é definida como a capacidade da vaca manter sua produção de leite após atingir a produção máxima na lactação. Além disso, Madsen (1975) retrata a importância do estudo da curva de lactação considerando três fatores:

1. O conhecimento prévio do comportamento da produção por meio da curva de lactação permite ministrar a alimentação do rebanho de forma que vacas com curvas de lactação com menores declínios necessitem de menos alimento que vacas com curvas cujo declínio é mais acentuado;
2. Altas produções de leite, no início da lactação, podem causar desordem reprodutiva e enfermidades metabólicas devido à alta atividade fisiológica demandada pela vaca;
3. A curva de lactação permite identificar grupos com características semelhantes de produção, possibilitando a realização de ensaios nutricionais mais eficientes. Dessa forma, podem-se obter respostas mais expressivas quanto à alimentação direcionada para os grupos.

Com base nos estudos de Brody *et al.* (1923, 1924), vários autores têm direcionado esforços para a modelagem das curvas de lactação. Nos estudos encontrados na literatura, são apresentados ajustes das curvas de lactação de várias raças diferentes utilizando-se uma gama de modelos. Dentre os modelos encontrados na literatura, podem-se citar os modelos não-lineares utilizados por Muñoz-Berrocal *et al.* (2001, 2005) e Fraga, *et al.* (2003) e o modelo quadrático utilizado por Cunha Filho *et al.* (2006). Gonçalves *et al.* (2002) ajustaram modelos como gama incompleto, monofásico, difásico, quadrático logarítmico e modelos de regressão múltipla para a curva de lactação da raça Holandesa. Na mesma direção, Oliveira *et al.* (2007) trabalhou um modelo gama incompleto estudando a Holandês-Gir.

Sem especificação da raça, Val-Arreola *et al.* (2004) estudaram a curva de lactação de gado leiteiro na região central do México, iniciando a modelagem a partir de um modelo de decaimento exponencial que não leva em consideração o pico da lactação. Entre os demais quatro modelos também ajustados, encontra-se um modelo considerando a equação gama. Pool *et al.* (2000) e Togashi e Lin (2003) estudaram curvas de lactação por meio de regressão aleatória e polinômios de Legendre. Togashi e Lin (2007) propuseram modificar a curva de lactação através dos autovalores da matriz de covariância da regressão aleatória, mostrando que estes autovalores são úteis na estimativa dos parâmetros, na avaliação genética, seleção e desenvolvimento de critérios relacionados ao melhoramento genético.

Pode-se observar pelos estudos de Cobuci *et al.* (2000), Val-Arreola *et al.* (2004), Pool *et al.* (2000), Togashi e Lin (2003) e Togashi e Lin (2007) que outros fatores influenciam na produção de leite, tais como fatores genéticos e ambientais, indicando que as características do animal e do local onde ele é tratado interferem no seu desempenho quanto à produção de leite. É nesse sentido que os modelos lineares mistos surgem como uma ferramenta valiosa. Os modelos mistos aparecem em estudos de produção de leite e de gordura, geralmente, com o intuito de analisar fatores ambientais e/ou genéticos relacionados ou que se supõem serem fatores determinantes na produção de leite ou gordura (COBUCI *et al.*, 2000; COBUCI *et al.*, 2001; CRUZ *et al.*, 2008), principalmente em estudos de melhoramento genético, como pode ser visto no trabalho de Holanda *et al.* (2004) no qual foram avaliados os ganhos genéticos dos pesos em bovinos da raça Nelore. Já Cunha *et al.* (2009) analisaram, por meio do modelo misto, o impacto de se ignorar efeitos genéticos não aditivos de dominância na avaliação genética animal. Utilizando o modelo em questão, Facó *et al.* (2008) estimaram efeitos genéticos aditivos e não aditivos para caracterização reprodutiva e produtiva de vacas mestiças (Holandês-Gir). De forma geral, percebe-se a importância do uso de modelos mistos na avaliação animal, seja geneticamente ou em estudos de produtividade.

3 Material e métodos

3.1 Modelo linear misto

O modelo linear misto proposto por Laird e Ware (1982) para o i -ésimo grupo, abrangendo apenas um nível de agrupamento é dado por:

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i, i = 1, \dots, c \quad (1)$$

$$\gamma_i \sim N(0, \Psi), \quad \epsilon_i \sim N(0, \Sigma);$$

em que Y_i é um vetor $n_i \times 1$ referente à variável resposta, β é um vetor $p \times 1$ de parâmetros desconhecidos denominado de efeitos fixos, X_i é uma matriz $n_i \times p$ de covariáveis observadas relacionada com os efeitos fixos, γ_i é um vetor $q \times 1$ de parâmetros denominado de efeitos aleatórios, Z_i é uma matriz $n_i \times q$ associada aos efeitos aleatórios e ϵ_i é um vetor $n_i \times 1$ de erros aleatórios. Assume-se que os vetores γ e ϵ são independentes.

De (1) temos que:

$$E(Y) = X\beta$$
$$Var(Y) = V = Z\Psi Z^T + \Sigma;$$

em que Ψ é denominada de matriz de variâncias e covariâncias do vetor de efeitos aleatórios sendo uma matriz simétrica positiva semidefinida. A matriz Ψ pode assumir diversas estruturas de covariâncias que estão relacionadas com particularidades do estudo em questão.

3.2 Análise de resíduos

No contexto da análise de resíduos, esse estudo aborda os resíduos condicionais padronizados, resíduo com confundimento mínimo, distância de Mahalanobis, distância de Cook e os pontos de alavanca que são definidos a seguir.

3.2.1 Resíduo condicional padronizado

Segundo Pinheiro e Bates (2000), a partir dos resíduos condicionais pode-se avaliar a suposição de homocedasticidade. Além disso, os resíduos condicionais também podem ser utilizados para identificação de pontos discrepantes. Entretanto, Nobre (2004), com base na possibilidade dos elementos de $\hat{\epsilon}$ apresentarem variâncias diferentes, propõe uma padronização dos resíduos condicionais dada por:

$$\hat{\epsilon}_i^* = \frac{\hat{\epsilon}_i}{\sigma\sqrt{q_{ii}}}; \quad (2)$$

em que $\hat{\epsilon}_i$ é o i -ésimo elemento de $\hat{\epsilon}$ e q_{ii} o i -ésimo elemento da matriz Q , sendo esta definida como $Q = M - MX(X^T MX)^{-1}X^T M$ e $M = \sigma^2 V^{-1}$.

O resíduo condicional padronizado, $\hat{\epsilon}^*$, obtido a partir de (2), é mais eficaz na verificação de possível afastamento da suposição de homoscedasticidade.

3.2.2 Resíduo com confundimento mínimo

Para verificação da normalidade do erro ϵ , Hilden-Minton (1995) propõe utilizar o resíduo com confundimento mínimo, pois o erro puro, erro que depende apenas das componentes fixas do modelo, acaba por não ser uma boa opção devido ao confundimento causado por γ . Ou seja, o resíduo confundido, resíduo que depende de mais de uma fonte de erro, não produz um parâmetro confiável para verificação da suposição de normalidade.

Sendo assim, Hilden-Minton (1995) define a fração de confundimento de $\hat{\epsilon}_i$ como

$$CF_i = 1 - \frac{U_i^T Q U_i}{U_i^T Q U_i}; \quad (3)$$

em que U_i é a i -ésima coluna de I_n . A medida CF_i é interpretada como sendo a proporção de variabilidade de $\hat{\epsilon}_i$ quando confundida por $\hat{\gamma}$; é um valor entre zero e um e quanto mais próximo de um for o CF_i maior é o confundimento. Dessa forma,

o autor propõe utilizar a transformação linear $L^T \hat{\epsilon}$ minimizando o confundimento de $l_i \hat{\epsilon}$. O resíduo com confundimento mínimo é mais eficiente na verificação da normalidade dos erros condicionais por intermédio do gráfico de probabilidade normal com envelope.

3.2.3 Distância de Mahalanobis

Dado que a partir do modelo (1) pode-se estimar o comportamento tanto médio como por indivíduos, é importante estudar não apenas possíveis observações discrepantes, mas também possíveis indivíduos discrepantes. Sendo assim, segundo Pinheiro e Bates (2000), o gráfico de $\hat{\gamma}_i$ versus o índice dos indivíduos pode ser usado na identificação de indivíduos discrepantes. Entretanto, Waternaux *et al.* (1989) propõem utilizar a distância de Mahalanobis em vez de usar $\hat{\gamma}_i$ ou $Z_i \hat{\gamma}_i$ para a identificação de indivíduos discrepantes.

A distância de Mahalanobis é dada por

$$\zeta_i = \hat{\gamma}_i^T \hat{Var}(\hat{\gamma}_i^T - \gamma_i) \hat{\gamma}_i^T; \quad (4)$$

que, sob a validade do modelo, segue uma distribuição qui-quadrado com n_i graus de liberdade, sendo n_i suficientemente grande.

3.3 Análise de sensibilidade

A análise de sensibilidade visa entender o comportamento do modelo quando este está sujeito a algum tipo de perturbação seja nas hipóteses ou nos dados. Dentre as formas mais utilizadas para avaliar a sensibilidade do modelo, encontra-se a eliminação de observações. Essa forma consiste em excluir uma determinada observação ou um determinado conjunto de observações e avaliar o quanto a exclusão dessa observação altera o ajuste do modelo. Há, ainda, outros métodos para mensurar a sensibilidade do modelo, a saber: identificação de pontos de alavanca, distância de Cook condicional e influência local.

3.3.1 Eliminação de observações

Cook (1977) apresentou o método de eliminação de observações com o intuito de identificar possíveis pontos que estivessem influenciando, de forma desproporcional, na estimativa dos parâmetros.

A eliminação de observações gera certa preocupação que, segundo Cook e Weisberg (1980), é fundamentada na possibilidade de que, ao se excluir uma determinada observação e analisar sua influência nas estimativas, a influência causada por um conjunto de observações seja mascarada e, com isso, observações que conjuntamente com outras estejam influenciando as estimativas não sejam identificadas.

No sentido de evitar a não detecção de possíveis conjuntos de observações influentes, Tan *et al.* (2001) propuseram a Distância de Cook Condicional que tem

como fundamento o uso do BLUE e do BLUP não condicionais, sendo esta medida dada por

$$D_i^{cond} = \sum_{j=1}^c \frac{P_{j(i)}^T P_{j(i)}}{k}; \quad (5)$$

em que $P_{j(i)} = (X_j \hat{\beta} + Z_j \hat{\gamma}_j) - (X_j \hat{\beta}_i + Z_j \hat{\gamma}_{j(i)})$, sendo $\hat{\beta}_i$ e $\hat{\gamma}_{j(i)}$ as estimativas dos parâmetros β e γ_j do modelo cuja i -ésima observação foi excluída e $k = \sigma^2([n - 1]c + p)$.

A distância de Cook condicional definida em (5) é útil para verificação de observações que estejam influenciando as estimativas dos efeitos fixos e aleatórios. Nesse contexto, a equação (5) pode ser decomposta em três partes, permitindo abordar a influência em partes distintas do modelo ajustado.

A decomposição de (5) se dá na forma de

$$D_i^{cond} = D_{1i}^{cond} + D_{2i}^{cond} + D_{3i}^{cond}; \quad (6)$$

sendo

$$D_{1i}^{cond} = \frac{(\hat{\beta} - \hat{\beta}_i)^T (X^T X) (\hat{\beta} - \hat{\beta}_i)}{K} = \frac{(\hat{Y} - \hat{Y}_{(i)})^T (\hat{Y} - \hat{Y}_{(i)})}{k}; \quad (7)$$

que serve para verificar se a i -ésima observação está influenciando na estimativa $\hat{\beta}$.

$$D_{2i}^{cond} = \frac{(\hat{\gamma} - \hat{\gamma}_{(i)})^T Z^T Z (\hat{\gamma} - \hat{\gamma}_{(i)})}{k}; \quad (8)$$

que serve para verificar se a i -ésima observação está influenciando na estimativa $\hat{\gamma}$, e

$$D_{3i}^{cond} = \frac{2(\hat{\beta} - \hat{\beta}_{(i)})^T X^T Z (\hat{\gamma} - \hat{\gamma}_{(i)})}{k}; \quad (9)$$

que serve como medida de covariância entre a mudança nas estimativas do BLUE e do BLUP ao excluir a i -ésima observação (NOBRE, 2004).

3.3.2 Pontos de alavanca

Define-se o i -ésimo ponto de alavanca como sendo a i -ésima observação que influencia em sua própria estimativa (NOBRE e SINGER, 2011). Seja H dada por $H = X(X^T X)^{-1} X^T$, h_{ii} representa o i -ésimo elemento da diagonal principal de H , então altos valores de h_{ii} indicam influência da observação em sua própria estimativa.

No intuito de generalizar a alavancagem para vários modelos, Wei *et al.* (1998) propuseram uma matriz denominada matriz de alavancagem generalizada definida por

$$GL(\hat{\beta}) = \frac{\partial \hat{Y}}{\partial Y^T}. \quad (10)$$

A alavancagem generalizada de uma observação é determinada por $GL(\hat{\beta})_{ii}$ que representa a taxa de mudança instantânea da estimativa de y_i quando y_i é acrescido de um infinitésimo. Pode-se observar que, nos modelos normais lineares, $GL(\hat{\beta}) = H$. No caso de medidas repetidas, também é fundamental identificar um indivíduo discrepante ou um *indivíduo alavanca*. Nesse contexto, Nobre (2004) propõe o uso de uma matriz de alavancagem para cada indivíduo, sendo esta definida como

$$H_i = X_i(X^T V^{-1} X)^{-1} X_i^T V_i^{-1}; \quad (11)$$

em que V_i é a matriz de covariâncias do i -ésimo indivíduo.

Contudo, é importante definir o quão “grande” é um valor de h_{ii} . Nesse sentido, fazendo $h_{ii}^* = GL(\hat{\beta})_{ii}$, tem-se

$$\bar{h}^* = n^{-1} \sum_{i=1}^n h_{ii}^* h_{jj}^* = \frac{p}{n}. \quad (12)$$

Portanto, uma observação é dita ser alavanca se $h_{ii}^* \geq 2p/n$. Para o caso em que o interesse é verificar se um indivíduo é considerado alavanca, tem-se que esse será considerado alavanca se

$$\frac{tr(H_i)}{n_i} = \frac{\sum_{j \in I} h_{jj}^*}{n_i} \geq 2p/n; \quad (13)$$

em que I indica o conjunto de observações referentes ao i -ésimo indivíduo.

Entretanto, segundo Nobre (2004), as definições de alavancagem expostas até o momento são voltadas diretamente para a verificação de alavancagem nos efeitos fixos. Em se tratando dos modelos lineares mistos, é fundamental observar a alavancagem nos efeitos aleatórios também. Sendo assim, Nobre (2004) propõe incorporar as informações relativas aos efeitos aleatórios na matriz de alavancagem generalizada. Dessa forma, tem-se

$$GL(\hat{\beta}, \hat{\gamma}) = \frac{\partial \hat{Y}^*}{\partial Y^T} = \frac{\hat{Y}}{\partial Y^T} + \frac{\partial Z \hat{\gamma}}{\partial Y^T} = GL(\hat{\beta}) + Z D Z^T Q. \quad (14)$$

Portanto, tem-se uma matriz de alavancagem que permite avaliar a alavancagem tanto nos efeitos fixos como nos efeitos aleatórios, uma vez que o termo $GL(\hat{\beta})$ considera os pontos de alavanca que influenciam as estimativas dos efeitos fixos e o termo $Z D Z^T Q$ que leva em consideração a estrutura da matriz de covariâncias e a matriz de efeitos aleatórios. Sendo assim, $GL(\hat{\beta})_{ii} = h_{ii}^* + (Z D Z^T Q)_{ii}$ é o i -ésimo elemento da diagonal principal e define a alavancagem generalizada da i -ésima observação nas estimativas dos parâmetros fixos e aleatórios (NOBRE e SINGER, 2011).

Para determinar o quão grande é o valor de $GL(\hat{\beta})_{ii}$, define-se que uma observação é dita ser um ponto de alavanca se $GL(\hat{\beta})_{ii} \geq 2tr(GL(\hat{\beta}))/n$ e um indivíduo diz-se ser um indivíduo alavanca se $(n_i)^{-1} \sum_{j \in I} GL(\hat{\beta})_{jj} \geq 2tr(GL(\hat{\beta}))/n$.

4 Aplicação

A análise apresentada nesta seção foi realizada no software livre R em sua versão 2.9.2. O método de estimação dos parâmetros do modelo foi o Método de Máxima Verossimilhança. O nível de significância adotado em toda a análise foi de $\alpha = 0,05$.

4.1 Descrição dos dados

Para o estudo da curva de lactação de vacas da raça Sindi, foram utilizados dados de 402 controles referentes às lactações de 54 vacas em sua primeira ordem de parto. Tais dados foram obtidos a partir de uma base primária na qual havia registros sobre 1165 controles referentes às lactações de 87 vacas, no período de 1987 a 1997, conseguidos por meio da pesquisa de Cunha Filho *et al.* (2006). As vacas são de propriedade da fazenda Carnaúba, pertencente à AMDA (Agropecuária Manoel Dantas Ltda.), situada no município de Taperoá, microrregião do Cariri Ocidental do Estado da Paraíba. Os animais foram criados em sistema semi-intensivo e a alimentação variava de acordo com a situação climática. Em épocas de chuva, os animais pastavam em campo aberto durante o dia, retornando ao curral para passarem a noite. Sua alimentação era baseada em ração concentrada durante a ordenha. Durante a seca, recebiam alimentação diferenciada baseada em feno de capim *buffel*, capim elefante, palma forrageira picada, raspa de mandioca, bagaço de cana hidrolisado e concentrado proteico. Além disso, sal mineral estava disponível em cochos, assim como uma mistura natural. O controle leiteiro foi realizado a cada 28 dias classificados como estádios de lactação, totalizando 8 estádios de lactação com duas ordenhas diárias de 12 horas de intervalo (CUNHA FILHO *et al.*, 2006). Devido a fatores não especificados na pesquisa de Cunha Filho *et al.* (2006), doze vacas não totalizaram os oito estádios, tornando o estudo desbalanceado.

A Tabela 1 traz algumas medidas descritivas referentes à produção de leite por estádio. Analisando a média, pode-se observar que os estádios extremos, estádios 1, 2, 7 e 8, apresentam as médias do período e os estádios 3, 4, 5 e 6 apresentam as maiores, sendo a maior média atribuída ao quinto estádio. O mesmo comportamento é observado para os valores máximos de produção e o valor total produzido. Entretanto, pode-se observar que os valores mínimos e o desvio padrão não possuem o mesmo comportamento. Os valores mínimos crescem timidamente até o sétimo estádio, dando um salto ao chegar no oitavo estádio, enquanto o desvio padrão decresce com o aumento do estádio, indicando que os últimos estádios apresentam produções mais homogêneas.

Na Figura 1, observa-se que 54% das vacas produziram entre 40 e 50 kg de leite no período observado. Logo em seguida, 13% produziram 60 kg e outras 13% produziram 40 kg. Apenas 4% das vacas produziram em torno dos 70 kg de leite. Os 17% restante, produziram menos de 20 kg. Com essas informações, tem-se que a maioria das vacas produziu entre 30 e 60 kg (80% das vacas). Pode-se entender, a partir de tal comportamento, que a maioria das vacas está num patamar de produção mediano, enquanto uma pequena quantidade apresenta produções extremas, sejam produções altas ou muito baixas.

Tabela 1 - Estatísticas descritivas da produção de leite em kg das 54 vacas por estádio

Estádio	1	2	3	4	5	6	7	8
Número de vacas	54	54	53	52	41	50	46	42
Mínimo	0	0	0	0,5	0,2	0,2	1	3,2
Máximo	9,9	9,7	9,9	10,8	11	10,8	9,7	8,4
Média	4,9	5,5	6,5	7	7,2	6,6	6,1	5,5
Desvio-padrão	3,37	3,55	3,06	2,72	2,28	2,02	1,65	1,21
Coefficiente de variação	0,68	0,65	0,47	0,39	0,31	0,3	0,27	0,22
Total produzido	267	294,3	344,1	361,6	369,2	331	279,6	232,2

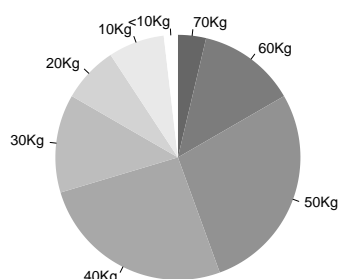


Figura 1 - Gráfico de setores referente a distribuição das 54 vacas de acordo com sua produção de leite total.

4.2 Motivação

A escolha do modelo linear misto foi motivada pela Figura 2 que corresponde ao gráfico de perfis das 54 vacas estudadas. Analisando a Figura 2, observa-se a inexistência de um comportamento padrão, ou seja, cada vaca tem seu próprio comportamento no que diz respeito à sua produção de leite durante a lactação. Esse comportamento não-padroneado sugere que cada vaca apresenta uma lactação particular que deve ser influenciada por suas próprias condições.

4.3 Definição do modelo

Da Figura 2, pode-se observar que a curva de lactação inicia-se em vários pontos distintos no eixo da produção de leite e que existe uma diferença de inclinação e posição das curvas. Essas duas observações sugerem o efeito da variabilidade da

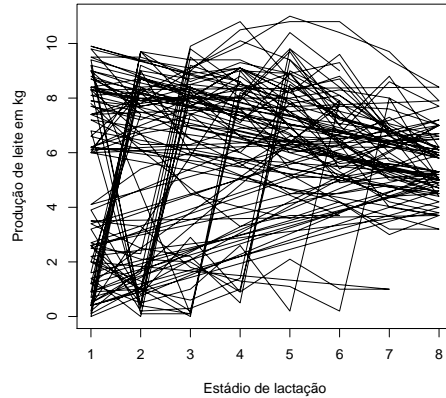


Figura 2 - Gráfico de perfis das 54 vacas.

vaca tanto no intercepto quanto na inclinação. Portanto, optou-se por ajustar três modelos, tomando como resposta a produção de leite e como variável explicativa, os estádios de lactação. Os três modelos são: modelo com efeito aleatório somente no intercepto, um modelo com efeito aleatório somente no estágio e um modelo com efeito aleatório tanto no intercepto como no estágio.

Foram consideradas duas estruturas de covariâncias definidas por

$$\Psi_1 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}; \Psi_2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

A estrutura definida por Ψ_1 é denominada matriz de covariâncias não estruturada positiva-definida e admite a existência de correlação entre os efeitos aleatórios. A estrutura Ψ_2 é denominada matriz de covariâncias diagonal e transcreve a suposição de que os efeitos aleatórios são não correlacionados (independentes sob normalidade).

4.4 Ajuste e escolha do modelo

Para efeito de análise, definimos o modelo com efeito no intercepto como sendo o modelo 1, o modelo com efeito no estágio como modelo 2 e o modelo com efeito no intercepto e no estágio como sendo o modelo 3.

Da Tabela 2, que aponta os resultados dos ajustes dos três modelos por meio do método de Máxima Verossimilhança considerando a matriz de variâncias e covariâncias dos efeitos aleatórios não estruturada, tomando como referência os métodos de seleção já definidos, tem-se que o modelo 3 apresentou menores valores tanto do AIC como do BIC, indicando que o modelo 3 é o modelo que deve ser escolhido dentre os três modelos. Estudando as estimativas dos efeitos fixos, pode-se

observar que as estimativas encontram-se próximas, sendo as estimativas tanto de β_0 como de β_1 para o modelo 1, as que apresentaram menor magnitude. Além disso, os intervalos de confiança referentes às estimativas destes parâmetros indicam baixa variabilidade das estimativas, indicando que os erros associados a essas estimativas são pequenos e assim sugerindo boas estimativas.

Tabela 2 - Estimativas ($\pm EP$) e intervalos de confiança (IC) dos parâmetros dos efeitos fixos (β 's) e intervalos de confiança (IC) para os efeitos aleatórios, adotando a matriz de covariâncias não estruturada

Parâmetro	Modelo 1	Modelo 2	Modelo 3
β_0	7,5383 \pm 0,1702	8,3348 \pm 0,1654	8,3276 \pm 0,2374
β_1	-0,0070 \pm 0,0005	-0,0125 \pm 0,0011	-0,0130 \pm 0,0010
IC[β_0]	[7,2044; 7,8722]	[8,0103; 8,6593]	[7,8619; 8,7934]
IC[β_1]	[-0,0080; -0,0061]	[-0,0147; -0,0103]	[-0,0150; -0,0110]
IC[σ_0]	[0,4720; 1,0715]	-	[1,0447; 1,8704]
IC[σ_1]	-	[0,0033; 0,0069]	[0,0035; 0,0070]
IC[σ]	-	-	[1,6275; 1,9175]
Correlação	-	-	-0,931
AIC	1756,488	1710,846	1687,595
BIC	1772,474	1726,832	1711,574
LogLik	-874,244	-851,423	-837,798

No que se trata da estimativa da correlação, o modelo 3 apresentou estimativa da correlação entre o efeito aleatório no intercepto e no estágio bastante elevada (-0,931), sugerindo a existência de uma forte relação negativa entre as estimativas dos efeitos aleatórios.

A Tabela 3 mostra os resultados dos ajustes adotando uma estrutura diagonal na matriz de variâncias e covariâncias, isto é, independência entre os efeitos aleatórios. Da Tabela 3, é possível observar que as estimativas dos parâmetros relacionados aos efeitos fixos se aproximam das estimativas quando é adotada a estrutura generalizada da matriz de variâncias e covariâncias. Os intervalos de confiança para as estimativas dos efeitos fixos apresentam menor amplitude quando comparados com os intervalos de confiança para os mesmos efeitos fixos estimados adotando a matriz generalizada de variâncias e covariâncias. O mesmo comportamento é observado nos intervalos de confiança relativos aos efeitos aleatórios.

Comparando os dois conjuntos de ajustes, considerando a matriz não estruturada e diagonal de Ψ , fica evidente que os ajustes para os quais foi adotada a matriz não estruturada, a matriz Ψ_1 , aparentam estar mais bem determinados. Tomando os métodos de seleção AIC e BIC, tem-se que o modelo 3, considerando a estrutura diagonal, apresenta valores de AIC e BIC superiores à essas mesmas medidas quando considerado a matriz estruturada. Sendo assim, pode-se considerar que os ajustes, levando em conta a estrutura de covariâncias estruturada da matriz Ψ , estão mais bem definidos e passam a ser os modelos a serem estudados a seguir.

Dando continuidade a análise, uma vez determinada a estrutura de variância e

Tabela 3 - Estimativas ($\pm EP$) e intervalos de confiança (IC) dos parâmetros dos efeitos fixos (β 's) e intervalos de confiança (IC) para os efeitos aleatórios, adotando uma estrutura de covariâncias diagonal

Parâmetro	Modelo 1	Modelo 2	Modelo 3
β_0	-	-	$8,3097 \pm 0,1887$
β_1	-	-	$-0,0125 \pm 0,001$
$IC[\beta_0]$	-	-	[7,9394; 8,6799]
$IC[\beta_1]$	-	-	[-0,0145; -0,0105]
$IC[\sigma_0]$	-	-	[0,4754; 1,1954]
$IC[\sigma_1]$	-	-	[0,0029; 0,0062]
$IC[\sigma]$	-	-	[1,6491; 1,9367]
Correlação	-	-	-
AIC	-	-	1704,179
BIC	-	-	1724,161
LogLik	-	-	-847,0895

covariância que se adequa melhor, segue-se para a determinação de qual é o modelo mais adequado, dentre aqueles que foram apresentados na Tabela 2. Para tanto são apresentados, na Tabela 4, os testes da Razão de Verossimilhanças para indicação de quais efeitos aleatórios devem permanecer no modelo.

Lembrando que o modelo 3 é aquele com efeito aleatório tanto no intercepto quanto no estádio e que os modelos 1 e 2 são os modelos com efeito aleatório no intercepto e no estádio, respectivamente. Sendo β_0 o efeito fixo relativo ao intercepto, β_1 o efeito fixo relativo ao estádio e γ_0 e γ_1 os efeitos aleatórios no intercepto e no estádio. Tem-se duas hipóteses a serem testadas que são definidas como:

$$T_1 = \begin{cases} H_0 : \sigma_2^2 = 0 \\ H_1 : \sigma_2^2 \neq 0 \end{cases} ; T_2 = \begin{cases} H_0 : \sigma_0^1 = 0 \\ H_1 : \sigma_0^1 \neq 0 \end{cases} .$$

O primeiro teste visa a verificação da hipótese relacionada com a presença do efeito aleatório do estádio no modelo, enquanto que o segundo teste tem a finalidade de verificar a hipótese da presença do efeito aleatório do intercepto no modelo. Com base nessas definições, verifica-se na Tabela 4, segundo o teste da Razão de Verossimilhanças, que as hipóteses de que o efeito no intercepto e o efeito no estádio são nulos, são rejeitadas a um nível de significância de 1%. Esses resultados indicam que o modelo com efeitos aleatórios tanto no intercepto quanto no estádio é o modelo a ser escolhido, ou seja, o modelo mais adequado. Tal indicação corrobora com a suposição feita a partir do gráfico de perfis, apresentado na Figura 2.

Fundamentando-se nos resultados até o momento apresentados, pode-se concluir que o modelo com efeito aleatório tanto no intercepto quanto no estádio adotando uma estrutura generalizada positiva definida para a matriz de variâncias e covariâncias, é o melhor modelo para o estudo em questão.

Para finalizar essa etapa de ajuste, segue-se para a estimação de Ψ_1 que é apresentada logo abaixo. Observa-se que a variância relacionada ao efeito aleatório

Tabela 4 - Teste da razão de verossimilhanças

Hipótese	Modelo	Graus de liberdade	LogLik	Razão de Verossimilhanças	Valor - P
T_1	Modelo 3	6	-837,7976		
	Modelo 2	4	-851,4229	27,25061	0,0001
T_2	Modelo 3	6	-837,7976		
	Modelo 1	4	-874,2441	72,89302	0,0001

no estádio é extremamente pequena, enquanto que a variância relacionada ao efeito no intercepto é bem maior. A matriz Ψ_1 evidencia um grau de variabilidade relativamente baixo, que aglomerado ao que já foi visto sobre o ajuste desse modelo, indica um bom ajuste do mesmo.

$$\hat{\Psi}_1 = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 1,953934 & -0,006450 \\ -0,006450 & 0,000025 \end{bmatrix}.$$

4.5 Análise de resíduos e sensibilidade

Na análise que se segue, pode-se observar por exemplo, a identificação “V1.8” que representa a observação referente ao oitavo estádio da vaca 1.

Iniciando a análise de resíduos e sensibilidade observa-se, na Figura 3, que os resíduos marginais apresentam comportamento aleatório em torno do zero com exceção de alguns pontos que se afastam muito dos demais. Entretanto, é uma quantidade pequena, se comparada com a amostra, e não interfere fortemente da indicação da presença de linearidade.

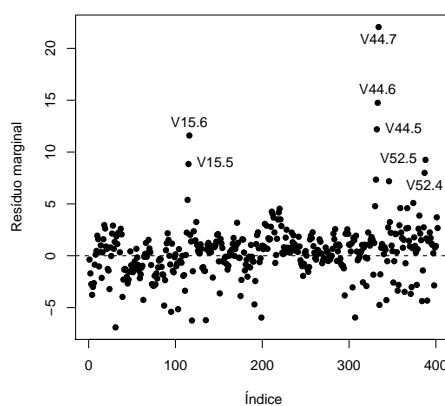


Figura 3 - Gráfico de dispersão do resíduo marginal versus índice.

O gráfico apresentado na Figura 4 expõe o resíduo condicional padronizado para as observações. Na Figura 4, pode-se verificar que não há afastamento da suposição de homogeneidade em relação às observações, pois nenhum padrão sistemático é observado. Algumas observações encontram-se fora dos limites, mas essas não comprometem a verificação da suposição.

Quanto à normalidade dos erros, utilizando o resíduo com confundimento mínimo, verifica-se, na Figura 5, que não há fortes evidências que impossibilitem indicar o não afastamento da suposição de normalidade, sendo esta verificada. Ainda na Figura 5, observa-se uma pequena quebra no ponto zero fazendo com que os pontos da metade superior tendam para a banda inferior, enquanto a metade inferior apresenta comportamento contrário. Fora isso, tanto a cauda inferior como a superior tendem para a linha central pontilhada com alguns pontos se afastando. Contudo, mantém-se a indicação de não afastamento da suposição de normalidade.

Analisando a distância de Mahalanobis para identificação de vacas discrepantes, na Figura 6, observa-se que há um certo distúrbio até a vigésima oitava vaca do estudo, passando para um comportamento linear e de distância mínima. Sendo assim, existem dois comportamentos presentes na distância de Mahalanobis. Focando na parte em que os pontos são mais espalhados, observa-se que as vacas V1, V4, V5, V10, V13, V17, V18 e V24 se distanciam em maior escala das demais, sendo consideradas possíveis vacas discrepantes.

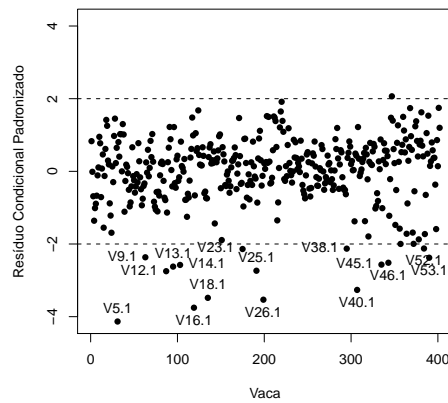


Figura 4 - Resíduos condicionais padronizados.

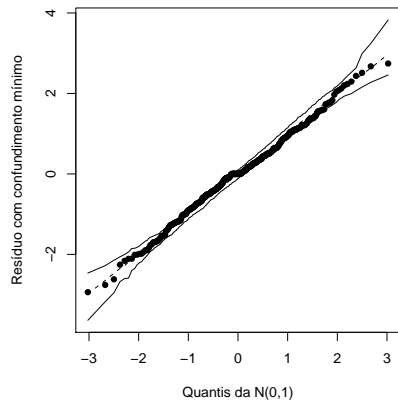


Figura 5 - Gráfico de probabilidade normal com envelope para o resíduo com confundimento mínimo com grau de confiança de 95%.

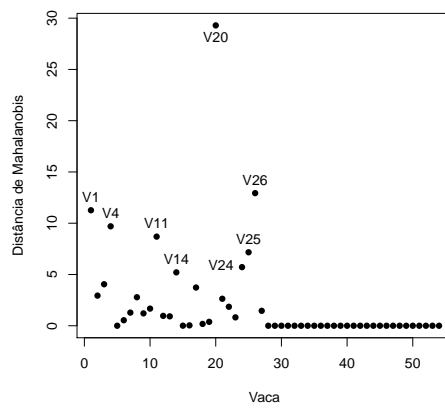


Figura 6 - Distância de Mahalanobis.

Seguindo adiante, entrando na análise de sensibilidade propriamente dita, inicia-se a análise pela alavancagem generalizada cujos gráficos são apresentados na Figura 7.

Analisando a Figura 7(a), que traz o gráfico da alavancagem generalizada por observação, percebe-se vinte e seis pontos que se distanciam da linha limite. Estudando tais observações, percebe-se que excetuando as observações V15.5, V15.6, V50.6 e V52.5, as demais são relativas aos quatro primeiros estádios da lactação, sendo a maioria de segundo estágio. Destacamos, ainda, que dezenove das vinte e seis observações (o equivalente a 73%) são produções menores ou iguais a 1,0 kg, seis observações (23%) apresentam produções entre 1,1 kg e 2,9 kg e uma observação (4%), referente ao sexto estágio da vaca V50, que apresenta produção de 7,3 kg superior à média geral do sexto estágio (6,6 kg). As seis observações são de segundo, terceiro, quarto e quinto estádios cujas médias gerais são bem superiores. Contudo, pode-se dizer que as observações identificadas não necessariamente são classificadas como pontos de alavanca, uma vez que os valores de h_{ii} para essas observações apresentam-se baixos, aproximando-se do zero.

Da Figura 7(b), pode-se observar que algumas vacas encontram-se em cima da linha limite, mas as demais encontram-se abaixo. Dessa forma, não há indícios de vacas que possam ser classificadas como pontos de alavanca.

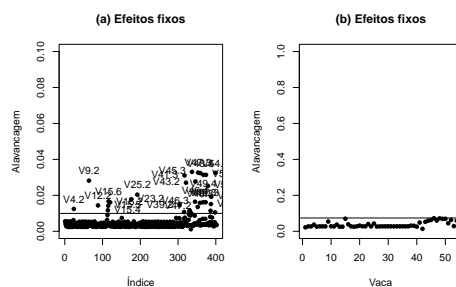


Figura 7 - Alavancagem generalizada considerando somente os efeitos fixos.

Quanto à alavancagem referente aos efeitos fixos e aleatórios, com relação à Figura 8(a), observa-se um comportamento semelhante ao da Figura 7(a), entretanto, a magnitude do afastamento da linha limite é bem maior do que na Figura 7(a). Dessa forma, essas vinte e uma observações identificadas, são classificadas como pontos de alavanca. Tal indicação sugere que o efeito aleatório influencia na estimação da produção de leite, mesmo trazendo pontos de alavanca.

Estudando a distância de Cook condicional na sua forma geral, Figura 9(a), pode-se verificar que grande parte das observações encontra-se em torno do zero com

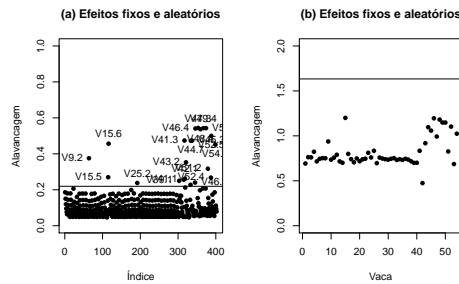


Figura 8 - Alavancagem generalizada considerando efeitos fixos e aleatórios.

poucos pontos se aproximando da linha limite. Onze observações apresentam-se afastadas das demais, podendo ser classificadas como possíveis pontos influentes nas estimativas dos efeitos fixos e aleatórios. Dessas onze observações, duas se distanciam consideravelmente das demais, sendo uma das observações referente à produção de leite do quarto estádio da vaca V46 e a outra, referente à produção do quarto estádio da vaca V50, sendo essas produções inferiores a 1,5 kg. As outras observações encontram-se afastadas das demais, mas não tão distantes da linha de limite.

Já, na Figura 9(b), referente ao primeiro termo da decomposição da distância de Cook condicional, os pontos se apresentam mais espalhados, mas apenas doze observações ultrapassam a linha limite, podendo ser classificadas como possíveis observações influentes na estimativa dos efeitos fixos. Dessas doze observações, quatro também se destacam na Figura 9(a).

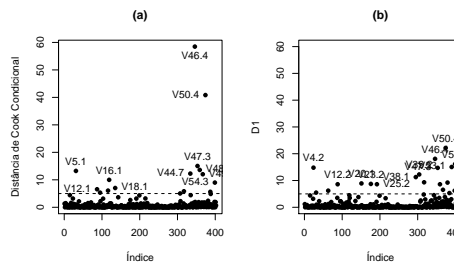


Figura 9 - Distância de Cook condicional: Geral e primeiro termo da decomposição.

Quanto à Figura 10(a), referente ao segundo termo da decomposição da distância de Cook condicional e que serve para identificação de observações que estão influenciando na estimativa dos efeitos aleatórios, pode-se verificar a partir dela que as mesmas observações identificadas na Figura 9(a) se destacam das demais.

Dessa forma, existe a possibilidade dessas observações serem consideradas pontos influentes na estimativa dos efeitos aleatórios. Na Figura 10(b), é apresentado o terceiro termo da decomposição da distância de Cook condicional, sendo esse tido como uma medida de covariância nas estimativas dos efeitos fixos e aleatórios. O comportamento apresentado na Figura 10(b) é esperado, pois os pontos se concentram em torno do zero, o que indica a possibilidade de não haver problemas.

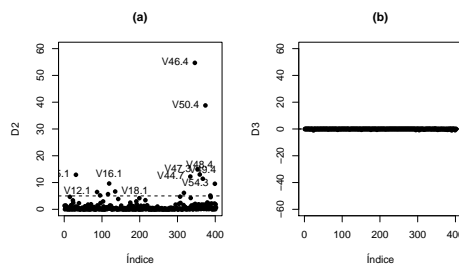


Figura 10 - Distância de Cook condicional: Segundo e terceiro termos da decomposição.

Analisando a variação causada pela retirada das observações identificadas nos gráficos da distância de Cook condicional, pode-se observar que a exclusão individual, de forma geral, não influencia as estimativas dos parâmetros dos efeitos fixos e aleatórios. As maiores influências são da observação V12.2 que provocou variação de 15,4% na estimativa de β_1 e variação de -3,5% na estimativa de σ_0 quando a observação V52.1 é excluída. Com relação à exclusão conjunta, novamente, observam-se alterações consideráveis nas estimativas dos parâmetros, sendo as estimativas de β_1 e σ_1 as que foram mais influenciadas. Uma observação interessante é a variação da estimativa de σ que ficou em -15,2%. Tais resultados indicam alta influência das observações, quando retiradas conjuntamente, nos parâmetros relacionados à inclinação e ao efeito aleatório referente à inclinação, além de uma influência significativa na estimativa de σ .

Em resumo, pode-se dizer que o modelo apresentou consistência, pois as suposições de normalidade e homogeneidade foram satisfeitas. As observações que foram identificadas como possíveis pontos discrepantes mostraram-se pouco influentes nas estimativas do modelo, salvo alguns pontos específicos. A análise relacionada aos pontos de alavanca, no que diz respeito aos efeitos fixos, não indicou observações influenciando em suas previsões. Entretanto, ao avaliar o gráfico de alavancagem, considerando os efeitos fixos e aleatórios, verificou-se que existem algumas observações que influenciam em suas respectivas previsões. Somando a análise de resíduos e sensibilidade à análise descritiva, pode-se considerar que produções em pequenas quantidades, precisamente tendendo a zero, acabam sendo

classificadas como possíveis observações influentes. Contudo, ao se observar a vaca em si, a variação existente na produção individual pode estar induzindo a um comportamento padrão, mesmo existindo produções que tendam a zero.

4.6 Previsões

Após ajustar o modelo e estudar seus resíduos e sua sensibilidade, serão apresentados os gráficos (Figuras 11 a 13) contendo os valores da produção de leite juntamente com suas respectivas estimativas. Nos gráficos a seguir, a linha vermelha representa a estimativa considerando apenas os efeitos fixos e a linha azul representa as estimativas considerando os efeitos aleatórios.

Observa-se que, nos casos em que a produção se aproxima do zero, o modelo estimou valores negativos, mas foi capaz de acompanhar a variabilidade causada por esses casos. Também é possível verificar que na maioria das vacas o modelo foi capaz de acompanhar o comportamento da produção de leite, mostrando sua eficiência no que tange à previsão.

Considerações finais

Esse trabalho teve como foco a produção de leite individual de cada vaca no que tange ao ajuste de um modelo que possibilitasse estudar o comportamento médio e individual e sua predição. Nesse contexto, o problema foi abordado por meio de um modelo linear misto estudando sua adequacidade através da análise de resíduos e sensibilidade.

As vacas estudadas apresentaram boas produções, mas também altas variações. A existência de produções tendendo a zero mostrou-se ser uma das maiores causas de variação na produção. Vacas que apresentaram valores altos nos primeiros estádios, possuíram menores variações no decorrer da lactação.

O modelo linear misto mostrou-se bastante eficaz no ajuste à produção de leite da amostra de vacas da raça Sindi em estudo. Dentre os modelos estudados, o modelo no qual estavam presentes os efeitos aleatórios no intercepto e no estádio, considerando a estrutura de covariância generalizada, ou seja, supondo dependência entre os efeitos aleatórios, foi o que se apresentou mais adequado. Tal modelo apresentou alta correlação negativa entre os efeitos aleatórios, indicando uma relação de dependência de ordem inversamente proporcional.

A análise de resíduos não evidenciou afastamento da suposição de linearidade e homoscedasticidade. Em se tratando da normalidade, o gráfico de probabilidade normal com envelope não acusou sérios afastamentos, sendo possível considerar a normalidade, entretanto, indicando a possibilidade de inclusão de uma variável sazonal.

No que se refere à alavancagem, quando estudados apenas os efeitos fixos, não se verificou a presença de observações ou vacas com essa característica, mas ao se incluir os efeitos aleatórios, algumas observações passaram a ser consideradas de alavanca. Quanto às vacas, não houve indicação de alavancagem por meio destas.

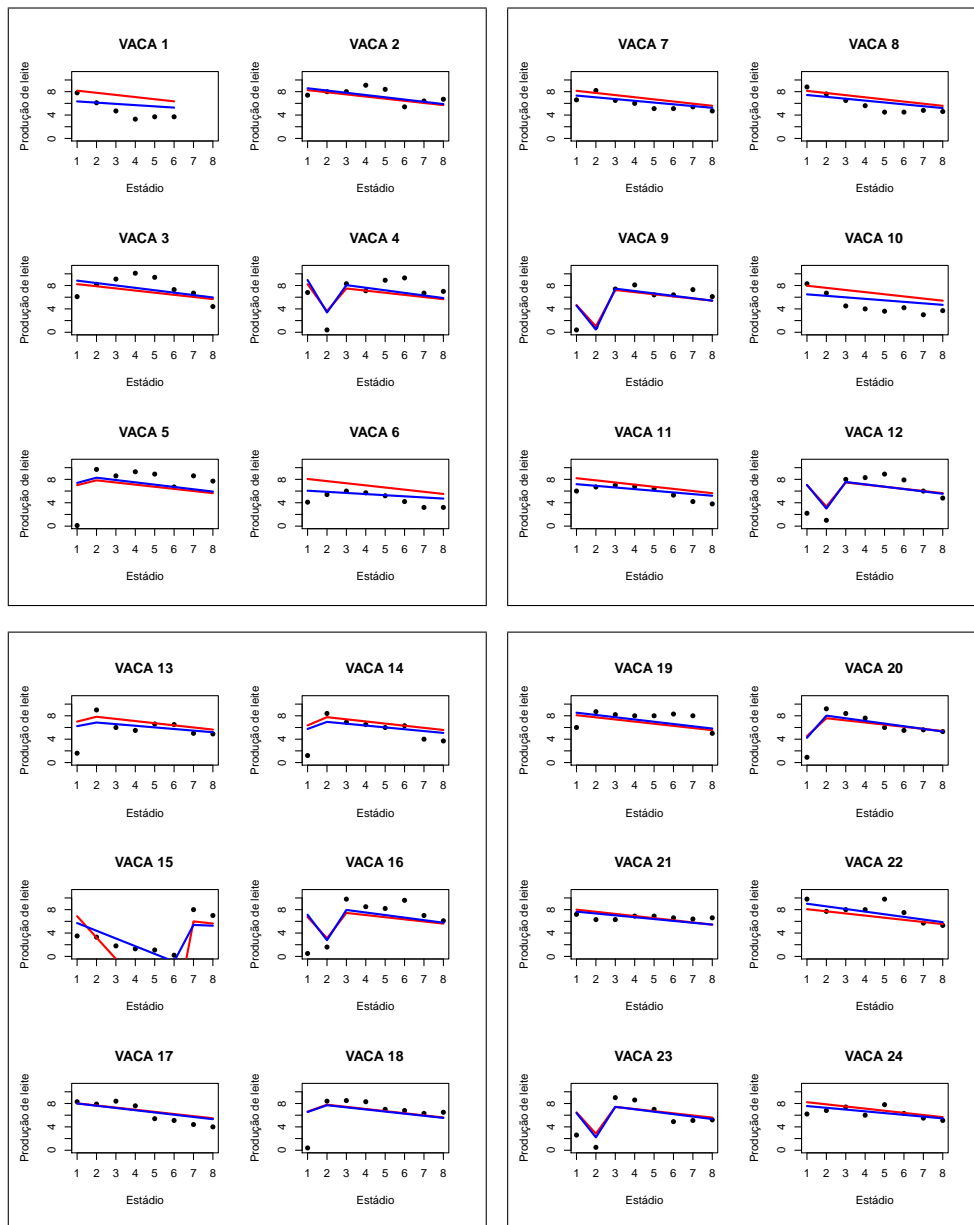


Figura 11 - Previsão da produção de leite das 54 vacas.

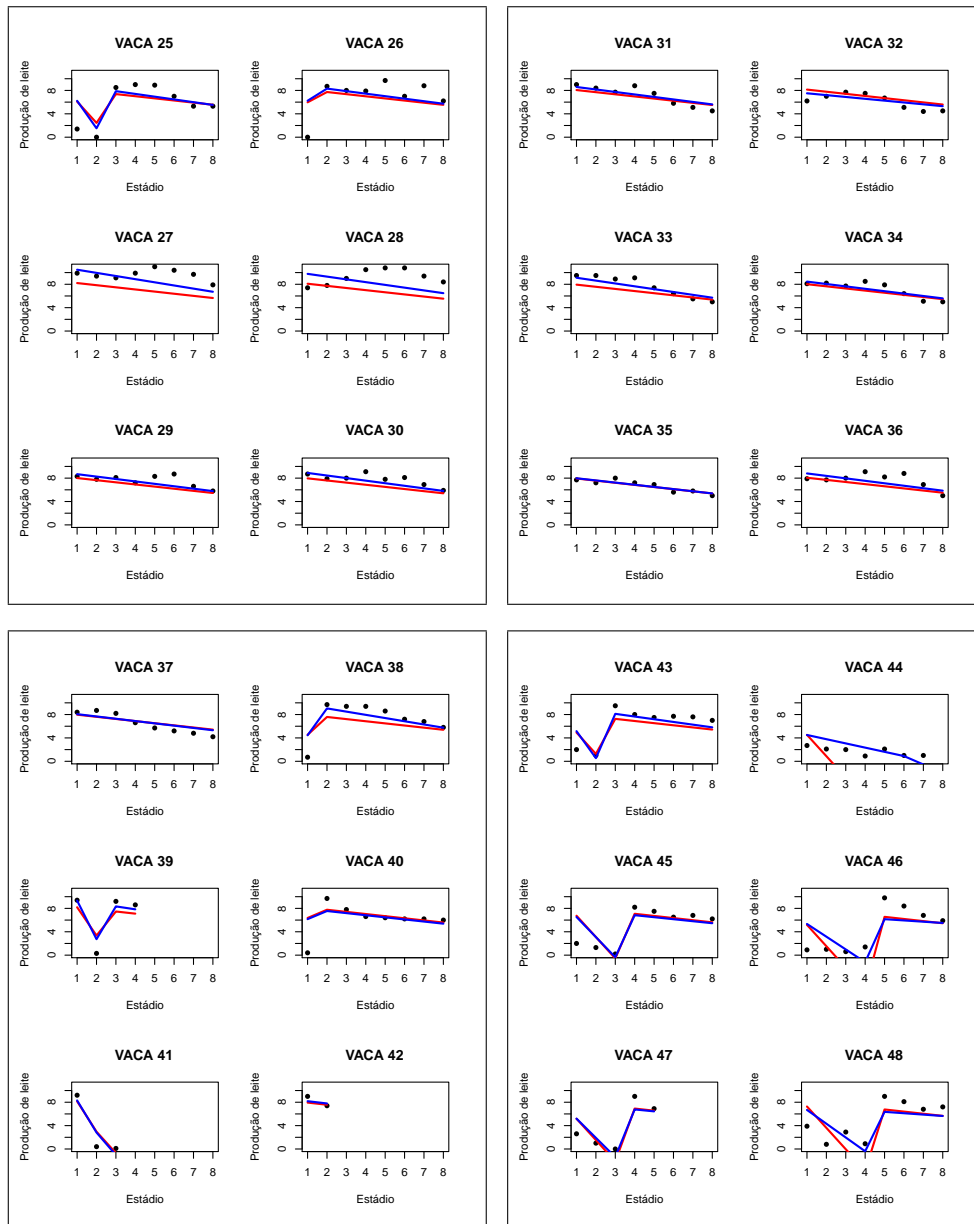


Figura 12 - Previsão da produção de leite das 54 vacas.

Tal mudança indica que a inclusão do efeito aleatório pode estar aumentando o efeito das observações em suas próprias previsões.

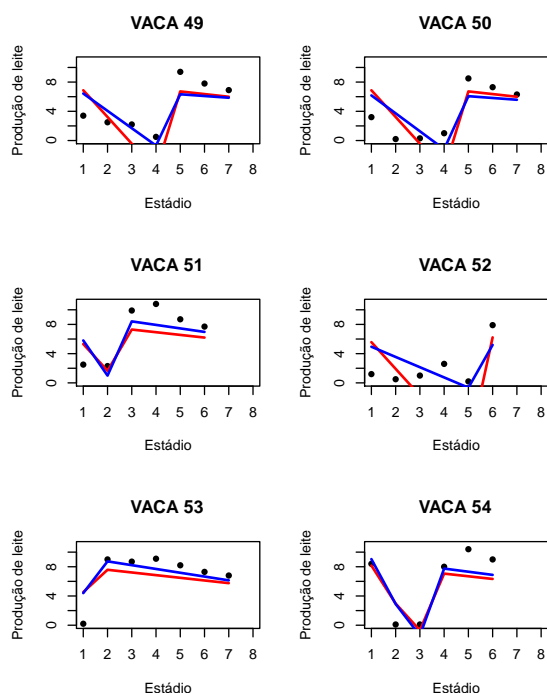


Figura 13 - Previsão da produção de leite das 54 vacas.

Quanto à influência, os pontos indicados como possíveis pontos influentes apresentaram variações nas estimativas dos parâmetros de ordem inferior a 3% em termos absolutos. Apenas uma observação causou variação superior a 15% na estimativa do efeito fixo referente ao estágio. Todavia, de forma geral, as observações possivelmente influentes não causaram variações significativas nas estimativas dos efeitos fixos nem na matriz de covariâncias.

Agregando ao ajuste à escolha do modelo, além da análise de resíduos e sensibilidade, conclui-se que o modelo escolhido deve ser considerado como modelo final, cujo ajuste pode ser definido como pouco sensível a observações influentes. Todavia, produções de leite próximas do zero devem ser consideradas prejudiciais ao ajuste, pois, mesmo considerando baixa influência nas estimativas, as observações classificadas como possivelmente influentes, em sua maioria, foram observações que representavam produções próximas de zero.

As previsões das produções de leite considerando os efeitos aleatórios mostraram-se mais próximas da realidade quando comparadas às previsões considerando apenas os efeitos fixos, ou seja, o comportamento médio. Contudo, a previsão para produções cujo valor observado estava próximo do zero, em alguns casos, foi negativa. Tal comportamento é esperado, uma vez que o modelo linear

misto utilizado levava em consideração a suposição de normalidade dos erros.

Agradecimentos

Agradecemos ao CNPq, pelo suporte financeiro; e também ao Dr. Juvêncio Nobre, pelas suas valiosas contribuições que enriqueceram o trabalho.

COSTA, T. R.; CAMPOS, L.; CISNEIROS, F. J. A.; CUNHA FILHO, M. Mixed linear models: an application to the lactation curve lactation sindhi's cows. *Rev. Bras. Biom.*, São Paulo, v.30, n.1, p.23-49, 2012.

■ **ABSTRACT:** *Lactation curves graphically represent individual milk or dairy herd production during their lactation period and they carry an unquestionable importance in terms of understanding the behavior of that particular herd production, which is fundamental to take decisions over conditions of the herd. Among many Brazilian dairy breeds that exist nowadays, the Sindhi breed has a special role in milk production because of its adaptation to the hard semi-arid climate, turning it into a feasible alternative for milk production in Brazil's Northeast. Therefore, the deal of this work was to use a linear mixed model in a database of a Sindhi breed herd, in order to verify milk production and animals individual forecast of this herd. Furthermore, the analysis of the waste and the sensitivity to verify model adaptability were done. The main result was that mixed linear model was suitable to study the behavior of each animal and the prediction of milk production.*

■ **KEYWORDS:** *Residue analysis and sensitivity; lactation curves; linear mixed model.*

Referências

ALI, T. E.; SCHAEFFER, L. R. Accounting for covariances among test day milk in dairy cows. *Can. J. Anim. Sci.*, Ottawa, v.67, n.3, p.637-639, 1987.

BRODY, S. A.; RAGSDALE, A. C.; TURNER, C. W. The rate of decline of milk secretion with the advance of period of lactation. *J. Gen. Physiol.*, New York, v.5, n.4, p.441-444, 1923.

BRODY, S. A.; RAGSDALE, A. C.; TURNER, C. W. The relation between the initial rise and the subsequent decline of milk secretion following parturition. *J. Gen. Physiol.*, New York, v.6, n.5, p.541-545, 1924.

COBUCCI, J. A.; EUCLYDES, R. F.; VERNEQUE, R. S.; TEODORO, R. L.; LOPES, P. S.; SILVA, M. A. Curva de lactação na raça Guzerá. *Rev. Bras. Zoot.*, Viçosa, v.29, n.5, p.1332-1329, 2000.

COBUCCI, J. A.; EUCLYDES, R. F.; TEODORO, R. L.; VERNEQUE, R. S.; LOPES, P. S.; SILVA, M. A. Aspectos genéticos e ambientais da curva de lactação de vacas da raça Guzerá. *Rev. Bras. Zoot.*, Viçosa, v.30, n.4, p.1204-1211, 2001.

- COBUCCI, J. A.; EUCLYDES, R. F.; PEREIRA, C. S.; AMEIDA TORRES, R.; COSTA, C. N.; LOPES, P. S. Persistência na lactação - uma revisão. *Arch. Latinoam. Prod. Anim.*, Maracaibo, v.11, n.3, p.163-173, 2003.
- COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, Baltimore, v.19, n.1, p.15-18, 1977.
- COOK, R. D.; WEISBERG, S. Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, Baltimore, v.22, n.4, p.495-508, 1980.
- CUNHA, E. E.; EUCLYDES, R. F.; TORRES, R. A.; SARMENTO, J. L. R.; CARNEIRO, P. L. S.; CARNEIRO, A. P. S. Impactos de se ignorarem os efeitos genéticos não-aditivos de dominância na avaliação genética animal. *Rev. Bras. Zoot.* Viçosa, v.38, n.12, p.2354-2361, 2009.
- CUNHA FILHO, M.; RIBEIRO, M. N. ; SANTOS, E. S. ; OLIVEIRA, J. C. V. Estudo da curva de lactação em vacas da raça Sindí no Estado da Paraíba. *Arch. Zoot.*, Maracaibo, v.3, p.23-43, 2006.
- CRUZ, G. R. B.; RIBEIRO, M. N.; FILHO, E. C. P.; SARMENTO, J. L. R. Análise genética de bovinos Sindí utilizando-se as produções de leite e de gordura no dia do controle. *Rev. Bras. Ciênc. Agrár.*, Recife, v.3, n.2, p.179-185, 2008.
- FACÓ, O.; LÔBO, R. N. B.; MARTINS FILHO, R.; MARTINS, G. A.; OLIVEIRA, S. M. P.; AZEVÊDO, D. M. M. R. Efeitos genéticos aditivos e não-aditivos para características produtivas e reprodutivas em vacas mestiças Holandês x Gir. *Rev. Bras. Zoot.* Viçosa, v.37, n.1, p.48-53, 2008.
- FRAGA L. M.; GUTIÉRREZ M.; FERNÁNDEZ L.; FUNDORA O.; GONZÁLEZ M. E. Estudio preliminar de las curvas de lactancia en las búfalas mestizas de Murrah. *Rev. Cubana Cienc. Agríc.*, Havana, v.37, p.151-155, 2003.
- GONÇALVES, T. M.; OLIVEIRA, A. I. G.; FREITAS, R. T. F.; PEREIRA, I. G. Curvas de lactação em rebanhos da raça Holandesa no Estado de Minas Gerais. Escolha do modelo de melhor ajuste. *Rev. Brasil. Zootec.*, Viçosa, v.31, n.4, p.1689-1694, 2002.
- HILDEN-MINTON, J. A. *Multilevel diagnostics for mixed and hierarchical linear models*. Thesis (Ph.D.) - University of California, Los Angeles, 1995.
- HOLANDA, M. C. R.; BARBOSA, S. B. P.; RIBEIRO, A. C.; SANTORO, K. R. Tendências genéticas para crescimento em bovinos Nelore em Pernambuco, Brasil. *Arch. Zootec.*, Córdoba, v.53, p.185-194, 2004.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, Oxford, v.38, n.4, p.963-974, 1982.
- MADSEN, O. A comparison of some suggested measures of persistency of milk yield in dairy cows. *Anim. Prod.*, Cambridge, v.20, p.191-197, 1975.
- MUÑOZ-BERROCAL, M.; THOLON, P.; PELICION, L. C.; TONHATI, H. Uso de polinomios ordinarios y segmentados en el ajuste de curvas de lactancia de búfalas

Murrah y sus mestizas en Brasil. The buffalo an alternative for animal agricultural in the third Millenium. In: WORLD BUFFALO CONGRESS, 4., 2001, Maracaibo. *Proceeding...* v.2, p.354-420.

MUÑOZ-BERROCAL, M.; TONHATI, H.; CERÓN-MUÑOZ, M.; DUARTE, J. M. C.; CHABARIBERI, R. L. Uso de modelos lineares e não lineares para o estudo da curva de lactação em búfalos Murrah e seus mestiços em sistema de criação semi extensivo, no Estado de São Paulo. *Arch. Latinoam. Prod. Anim.*, Maracaibo, v.13, n.1, p.19-23, 2005.

NOBRE, J. S. *Métodos de diagnóstico para modelos lineares mistos*. 2004. Dissertação (Mestrado em Estatística) – Universidade de São Paulo, Instituto de Matemática e Estatística, São Paulo, 2004.

NOBRE, J. S.; SINGER, J. M. Leverage analysis for linear mixed models. *J. Appl. Stat.*, Oxon, v.38, n.5, p.1063-1072, 2011.

OLIVEIRA, H. T. V.; REIS, R. B.; GLÓRIA, J. R.; QUIRINO, C. R.; PEREIRA, J. C. C. Curvas de lactação de vacas F1 Holandês-Gir ajustadas pela função gama incompleta. *Arq. Bras. Med. Vet. Zoot.*, Belo Horizonte, v.59, n.1, p.233-238, 2007.

PINHEIRO, J. C.; BATES, D. M. *Mixed-effects models in S and S-PLUS*. New York: Springer - Verlag, 2000. 528p.

POOL, M. H.; JANSSE, L. L. G.; MEUWISSEN, T. H. E. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.*, Champaign, v.83, n.11, p.2640-2649, 2000.

QUINTERO, J. C.; SERNA, J. I.; HURTADO, N. A.; NOGUERA, R. R.; CERÓN-MUÑOZ, M. F. Modelos matemáticos para curvas de lactancia en ganado lechero. *Rev. Colomb. Cienc. Pec.*, Medellin, v.20, n.2, p.174-156, 2007.

R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. Disponível em: <<http://www.R-project.org>>.

TAN, F. E. S.; OUWENS, M. J. N.; BERGER, M. P. F. Detection of influential observations in longitudinal mixed effects, regression models. *The Statistician*, London, v.50, n.2, p.271-284, 2001.

TOGASHI, K.; LIN, C. Y. Modifying the lactation curve to improve lactation milk and persistency. *J. Dairy Sci.*, Champaign, v.86, n.4, p.1487-1493, 2003.

TOGASHI, K.; LIN, C. Y. Genetic modification of the lactation curve by bending the eigenvectors of the additive genetic random regression coefficient matrix. *J. Dairy Sci.*, Champaign, v.90, p.5753-5758, 2007.

VAL-ARREOLA, D.; KEBREAD, E.; DIJKSTRA, J.; FRANCE, J. Study of the lactation curve in dairy cattle on farms in Central Mexico. *J. Dairy Sci.*, Champaign, v.87, n.11, p.3789-3799, 2004.

WATERNAUX, C.; LAIRD, N. M.; WARE, J. H. Methods for analysis of longitudinal data: Blood-lead concentrations and cognitive development. *J. Am. Stat. Assoc.*, Schaumburg, v.84, p.33-41. 1989.

WEI, B. C.; HU, Y. Q.; FUNG, W. K. Generalized leverage and its applications. *Scand. J. Stat.*, Oxford, v.25, n.1, p.25-37, 1998.

Recebido em 26.09.2011.

Aprovado após revisão em 03.05.2012.