TADEU RODRIGUES DA COSTA

MODELOS LINEARES MISTOS: UMA APLICAÇÃO NA PRODUÇÃO DE LEITE DE VACAS DA RAÇA SINDI

RECIFE-PE - JUN/2010

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

MODELOS LINEARES MISTOS: UMA APLICAÇÃO NA PRODUÇÃO DE LEITE DE VACAS DA RAÇA SINDI

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientadora: Profa. Dra. Laélia P. B. Campos dos Santos

Co-orientador: Prof. Dr. Francisco José de Azevedo Cysneiros

RECIFE-PE - JUN/2010.

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

MODELOS LINEARES MISTOS: UMA APLICAÇÃO NA PRODUÇÃO DE LEITE DE VACAS DA RAÇA SINDI

Tadeu Rodrigues da Costa

Dissertação julgada adequada para obtenção do título de mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 04/06/2010 pela Comissão Examinadora.

Orientador:

Prof. Dra. Laélia P. B. Campos dos Santos
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Francisco José de A. Cysneiros
Universidade Federal de Pernambuco
DE-UFPE

Prof. Dr. Juvêncio Santos Nobre
Universidade Federal do Ceará
DEMA-UFC

Prof. Dr. Kleber Régis Santoro
Universidade Federal Rural de Pernambuco
UAG-UFRPE

Dedico às minhas irmãs Suzanne e Nany, meus sobrinhos Rafito e Lela, e minha vida Renata.

Agradecimentos

Agradeço, primeiramente, à minha mãe Luciene por ter me dado a oportunidade de estar aqui hoje. Pela educação, força, carinho e dedicação, praticamente exclusivos. Por seu apoio incondicional. Agradeço eternamente pelos sorrisos e lágrimas de alegria e satisfação que estarão para sempre gravados em um lugar especial dentro do meu ser. Agradeço pelo simples e completo "te amo"e termino também dizendo te amo e muito obrigado por tudo.

Agradeço às minhas irmãs Suzanne e Nany pelo carinho, disposição e pela força dada. Por serem minhas irmãs, amigas e tudo mais que uma pessoa pode ser à outra. Agradeço por ter a oportunidade de fazer de suas felicidades a minha meta e por contribuir significantemente em suas vidas. Sem sombra de dúvidas devotarei o que for preciso para que tudo seja perfeito em suas vidas.

Aos meus sobrinhos, Rafael e Rafaela, por serem crianças tão adoráveis e por me motivarem a continuar caminhando e ignorando obstáculos de qual tamanho forem. Por me terem em seus corações e por me permitirem fazer parte de suas vidas. Sem vocês tudo seria mais difícil e sofrido.

Agradeço à Renata por ser minha vida. Por ser meu alicerce, colunas, paredes e teto. Por me suportar nos momentos mais chatos e por me fazer rir sempre que preciso. Por caminhar ao meu lado em momentos fáceis e difíceis sem questionar o caminho escolhido. Pelos conselhos, sorrisos, lágrimas, sustos e brigas que tanto colaboraram para que eu me tornasse quem sou hoje. Ai shiteiru, vida!

À minha segunda família, Dona Alcione, Leo, Thiago e Aninha. Com certeza vocês tiveram papel fundamental na formação do que hoje é Tadeu. Seu apoio se revelou muitas vezes indispensável, como também seus conselhos. Não usarei a palavra amigo, pois posso substituir sem medo por mãe e irmãos. Sempre terei vocês guardados, cravados em um lugar especial aqui dentro desse ser. Aconteça o que acontecer, sempre contem comigo para o que der e vier.

Não poderia deixar de agradecer aos meus outros irmãos, Diogo, Diego, Teta e Zé, mesmo não sendo irmãos de sangue, mas com certeza, de vida. Nada como rir um pouco

com vocês. Agradeço por todas as madrugadas de jogos, tiração de onda e aquela cervejinha. Tenho um apreço imensurável por todos vocês. A amizade de vocês é um dos grandes presentes que tive a honra de ganhar, conquistar e manter por todo esse tempo. Meu carinho por vocês é enorme e estarei aqui sempre que precisarem.

Aos meus amigos Hemílio, Vanessa, Amanda, Crody, Robinho e Leila, uma nova geração de grandes estatísticos. Acima de tudo, amigos cujas existências em minha vida são significantes com p-valor < 0,0000000000000001. Agradeço pela força e puxadas de orelha. Certamente, sem a intromissão de vocês, eu não teria chegado aqui hoje. Tenho grande orgulho de fazer parte de suas vidas e de vocês fazerem parte da minha. Não só participar, mas agradeço por me permitirem trabalhar aos seus lados. Não deixaria de agradecer aos estudos de última hora, aquele pdf salvador no meio da noite, às dúvidas sanadas no domingão pelo msn enquanto desenvolvíamos nossas atividades curriculares, àquela maminha com fritas em qualquer noite da semana. O que seria de nós sem nós para nos ajudarmos? Obrigado.

Às doutoras Renata e Gaby, pelo acolhimento e paciência. Minha estadia junto a vocês foi curta, mas suficiente para que fossem de extrema importância em minha formação. Agradeço pelas experiências compartilhadas e pelos momentos de estresse. Nada como aquele mar de questionários e os artigos nunca perfeitos. Sem contar na quantidade de exclamações após meu nome na janelinha do gtalk. Muito obrigado por me permitirem estar presente em suas vidas.

Aos professores Francisco José de A. Cysneiros e Audrey Helen Mariz de A. Cysneiros pelo conhecimento disponibilizado, paciência e dedicação. Tenho grande respeito e atribuo meu conhecimento aos senhores. O estatístico que hoje sou é fortemente inspirado em suas posturas e conhecimento. Agradeço por não me deixarem desistir e pela força empregada em minha formação.

À professora Laélia Campos, por me suportar como seu orientando por todo esse tempo e que esteve comigo até o fim. Tenho um carinho especial pela senhora e me sinto muito honrado em participar, mesmo que superficialmente, de sua vida. Ao velho guerreiro e professor Esdras Adriano B. dos Santos, sua contribuição em minha formação foi de grande importância.

Ao professor Moacyr Cunha Filho por sua colaboração significante para o desenvolvimento desta dissertação e disponibilidade do banco de dados. Sua competência, disponibilidade e paciência foram fatores determinantes na conclusão deste processo.

Ao professor Juvêncio Nobre que mesmo distante dispôs-se em ajudar tirando dúvidas

e indicando excelentes materiais. Agradeço a sua ajuda que foram determinísticas para o desenvolvimento desta dissertação.

Ao professor Kleber Santoro por seu apoio e sua presença. Seus ensinamentos e dicas foram fundamentais para o desenvolvimento dessa dissetação.

Agradeço ao André Magalhães e Sônia Fonseca pela compreensão e pelas ajudas durante o desenvolvimento desta dissertação. Pela consideração e pela confiança devotados a minha pessoa como estatístico. A experiência adquirida nesse tempo de trabalho junto a vocês é, e será fortemente aproveitada nos trabalhos que estão por vir.

Ao professor Alexandre Jatobá pela oportunidade concedida e pelo apoio. Com certeza foi uma experiência significante em termos profissionais. Espero ter respondido às expectativas.

À Ana Wilma, minha gerente e amiga. Uma das poucas pessoas em minha vida que conquistaram minha amizade integralmente. Agradeço por permitir participar de sua vida e espero ter conquistado plenamente sua amizade. Tenho um carinho profundo por sua pessoa. Agradeço pela força e pelo apoio, pelos momentos de descontração, pelos cuidados e consideração, por tudo. Fico extremamente honrado e grato em ter uma pessoa de tanta garra, fibra e perseverança presente em minha vida. Espero permanecer perto de você mesmo estando distante. Saiba que estarei aqui para o que precisar.

Às minhas grandes amigas Juliana Holanda, Mirelle Queiroz, Celeste Maia, Gisele Leal e Cristiane Mesquita, principalmente pelas diferenças existentes entre vocês. De cada uma aprendi algo diferente que me fez ser uma pessoa melhor. Devo dizer que vocês foram pessoas-chave e que se alcancei algum sucesso, devo à vocês essa conquista. Muito obrigado por tudo.

Para que eu não cometa o erro de esquecer ninguém e visando não escrever dezenas de páginas de agradecimento, condenso meu profundos agradecimentos aos meus amigos da DINE, alunos e professores do Departamento de Estatística da UFPE e do programa de pós graduação em Biometria e Estatística Aplicada da UFRPE, à Zuleide, a todos colegas e amigos do Ipsep, aos meus amigos e colegas de trabalho da Datamétrica, cunhada, sogros, Nina e ao meu laptop (o que seria de nós sem um computador?).

Fecho os meus agradecimentos salientando novamente que todos foram, são e serão estimativas não-viesadas, consistentes, completas e suficientes dos parâmetros necessários para que se possa convergir na direção de uma pessoa mais responsável e melhor em todos os sentidos. Meus sinceros, verdadeiros e profundos agradecimentos.

"*A mente que se abre a uma nova idéia jamais volta ao seu tamanho original.*"

Albert Einstein

"*Algumas das maiores façanhas do mundo foram feitas por pessoas que não eram suficientemente espertas, para saber que elas eram impossíveis.*"

Doug Larson

"*As idéias geniais são aquelas com as quais nos espantamos por não as ter tido antes.*"

Noel Claraso

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 o entendimento do comportamento da produção daquele determinado rebanho, sendo fundamental na tomada de decisões acerca das condições do rebanho. Dentre as muitas raças leiteiras existentes hoje no Brasil, a raça Sindi tem um papel especial na produção de leite por se adaptar à rigorosidade do clima semi-árido, 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 adequacidade 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: Raça Sindi; Curva de Lactação; Modelo Linear Misto; Análise de Resíduos e Sensibilidade.

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.

Key words: Sindhi breed; lactation curves; linear mixed model; Residue analysis and sensitivity.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

A produção de leite das mais diversas raças (bovinos, caprinos, entre outras) tem sido escopo de estudo de vários pesquisadores e também de instituições bem conceituadas como a EMBRAPA e a EMEPA-PB. Tais estudos, geralmente, têm o intuito de entender o comportamento de um determinado rebanho, a fim de que seja possível a tomada de decisões em diversos níveis, nominalmente, como se deve proceder quanto ao descarte de animais, as circunstâncias adequadas ao aumento do rebanho, e a necessidade de melhorias da alimentação dos rebanhos e das instalações físicas da propriedade, de forma que a eficiência na produção seja alcançada.

A produção 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.

Anterior a qualquer análise que se possa fazer em relação à curva de lactação, torna-se primordial apresentar as características da raça cuja curva de lactação está em estudo neste trabalho. No Brasil, existe uma grande variedade de raças leiteiras, sendo a Sindi uma das raças que melhor se adapta ao clima semi-árido 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. Seguindo a oeste da província de Sind, podem-se encontrar exemplares de maior pureza e de alta produtividade de leite. Produções recordes são freqüentemente encontradas na Índia, chegando um exemplar a produzir 18,12 kg de leite em um único dia (resultado de duas ordenhas), que numa lactação de 300 dias equivale a 5.436 kg de leite. Naquele país, em termos médios, numa lactação são produzidos 1.721,4 kg de leite (LEITE et al., 2001).

Brody et al. (1923) introduziu o estudo da curva de lactação por meio de modelos estatísticos, até hoje utilizados em estudos dessa natureza. Outros modelos foram propostos na tentativa de apresentar maior precisão e adequabilidade às curvas, uma vez que a curva de lactação não apresenta um comportamento padrão, mas sim diferenciado sendo influenciada por diversos fatores, como por exemplo, genética, manejo nutricional entre outros. Com isso, são necessários modelos mais direcionados e mais específicos que se adequem ao formato da curva de lactação da raça que estiver em estudo.

Partindo dos modelos propostos por Brody e colaboradores (1923, 1924) até os dias atuais, uma grande quantidade de modelos tem sido proposta, principalmente modelos que levam em consideração características especificas do fenômeno em estudo, cujas estimativas de parâmetros envolvem métodos iterativos cada vez mais eficientes. Dentre os modelos encontrados na literatura, ajustados às curvas de lactação de várias raças diferentes, é possível citar os modelos lineares e não-lineares múltiplos e modelos quadráticos logarítmicos. Outros mais complexos como o de decaimento exponencial e o modelo gama incompleto, bem como as regressões aleatórias e os modelos polinomiais de Legendre propostos por Pool et al. (2000) e Togashi e Lin (2003).

Estudos vêm surgindo e trazendo novos modelos mais específicos para determinadas curvas de lactação, na tentativa de melhorar a adequação ou de propor novas formas de abordagem de ajuste de modelos a produções de leite. Nessa direção, os modelos lineares mistos surgiram como um novo enfoque em vários casos que estudavam a curva de lactação de algumas raças. Cruz et al. (2008) fazem uso do modelo misto com o intuito de estimar parâmetros genéticos relacionados à produção de leite da raça Sindi. Os autores tomam as quatro primeiras lactações definindo um modelo de repetibilidade. Concluíram que para o estudo genético relacionado à produção de leite, a utilização da produção de leite e gordura no dia do controle pode substituir os valores totais da produção de leite e gordura.

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 a 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 tomas de decisões relacionadas ao modelo. Nesse contexto, os modelos mistos se tor-

nam fundamentais no estudo da curva de lactação de uma raça, pois possibilita 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 visam verificar adequacidade do modelo ajustado. Tais análises têm como objetivo verificar possíveis afastamentos das suposições envolvidas na definição do modelo e observações que possam estar influenciando, 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, predições dos efeitos aleatórios e das previsões individuais e médias.

Considerando a importância da Raça Sindi 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 Sindi proveniente do semi-árido paraibano. Além disso, realizar a análise de resíduos e sensibilidade, tomando como referência a metodologia apresentada por Nobre (2004).

2 Produção de Leite

2.1 A Raça Sindi

A raça Sindi teve sua origem ao norte da província de Sind numa região denominada Kohistan, situada no atual Paquistão, onde a raça é denominada, em inglês, Red Sindhi e é considerada uma das raças mais puras da região, talvez devido à localização isolada da província. Na região, são encontradas as melhores vacas leiteiras dessa raça. Devido à sua adaptação em diversos tipos de solo, clima e resistência a certas doenças, a raça Sindi vem sendo utilizada em toda a Índia por criadores locais ou governamentais. Além disso, grande quantidade de animais tem sido exportada para outros países tais como Coréia, Estados Unidos e Brasil, nesse último se adaptando adequadamente ao clima semi-árido do Nordeste (LEITE et al., 2001).

Na Índia, a raça Sindi tem uma boa produção de leite, chegando alguns casos a aproximadamente 5.500 kg em uma lactação de 300 dias, sendo em média 1.721 kg de leite produzidos. A Sindi tem sido considerada uma raça extremamente econômica devido ao seu baixo custo de manutenção e à sua adaptabilidade. Segundo Joshi e Phillips (1954), a raça faz parte do terceiro grupo básico do gado indiano, no qual também estão englobadas as raças Dangi e Nimari, entre outras.

Tanto na Índia quanto em outros países, assim como o Brasil, a raça é de uma beleza única, apresentando pequeno porte com cabeça pequena e chifres grossos na base crescendo para os lados e curvando-se para cima, orelhas caídas de tamanho mediano, olhos escuros e quartos traseiros caídos e arredondados, membros pequenos e delicados bem aprumados e geralmente apresentam tetas grossas (Figura 2.1). É uma raça resistente a doenças, inclusive à febre aftosa.

Figura 2.1: Vaca da raça Sindi

A Sindi é apropriada para a produção de leite, mas também é boa produtora de carne. Entretanto, devido ao seu pequeno porte não alcança grandes pesos e, portanto, não consegue competir com outras raças de maior porte. Mesmo assim, o gado que não produz ou deixa de produzir leite acaba sendo direcionado para o corte. Na Índia, observa-se que a produção de leite é comparável com a de raças como Sahiwal, Tharpaparkar e Hariana que são consideradas as melhores raças em termos de produção de leite. Numa fazenda em Karachi, observou-se uma produção média de aproximadamente 1.500 kg de leite em 274 dias de lactação. Um determinado grupo, contendo 41 vacas, registrou produção média com um máximo de 2.500 kg de leite, tendo uma média individual de 2.070 kg. Nesta mesma fazenda, foi observado que as melhores lactações atingiram uma média de 3.077 kg em 35 controles.

É evidente a boa produtividade leiteira da raça Sindi, mas vale salientar que a produção de leite depende consideravelmente do local e de outros fatores associados ao clima e condições de vida do animal. Ainda com relação ao rebanho citado no parágrafo anterior, observou-se um período seco médio de 160 dias, intervalo entre partos de aproximadamente 14 meses, idade ao primeiro parto de 41 meses e em média 5 lactações durante a vida (LEITE et al., 2001).

Atualmente o Nordeste vem se destacando na criação de vacas da raça Sindi, principalmente na Paraíba, onde empresas, como a EMEPA, trabalham em estudos sobre a raça constantemente. Sua adaptação no Nordeste brasileiro pode ser explicada pela familiaridade com o clima semi-árido de sua terra original. A Sindi vem como uma solução no que tange à produção de leite para o Nordeste, pois suas características como baixo consumo de alimentos, facilidade de manejo, resistência e adaptabilidade climática, além de ser

um gado de retorno rápido, permitem sua criação em fazendas menores em localidades castigadas pelo clima e com falta de recursos.

Para estudar a produção de leite de gado leiteiro, é utilizada a curva de lactação média que descreve a quantidade de leite produzido por um rebanho durante a sua fase de lactação.

O presente trabalho visa estudar a curva de lactação não apenas do ponto de vista médio, mas considerando a variabilidade entre os animais.

2.2 Curva de Lactação

A produção de leite não só de zebuínos, mas de vários tipos de rebanhos produtores de leite, vem sendo objeto de estudo em todas as partes do mundo por diversos autores cujo interesse básico é obter parâmetros que possibilitem tomadas de decisão acerca do desempenho dos seus rebanhos, otimizando os procedimentos envolvidos com tal produção.

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 configurações semelhantes de produção permitindo a realização de ensaios nutricionais mais eficientes. Dessa forma, podem-se obter respostas mais expressivas quanto à alimentação direcionada para os grupos".

As razões apresentadas por Madsen (1975) refletem a preocupação em ter uma produção de leite eficiente tomando como base a alimentação do rebanho, minimizando a perda das vacas em decorrência de processos relacionados à lactação e à alimentação. A curva de lactação permite ao produtor identificar comportamentos de forma a auxiliar em decisões como descarte do animal e escolha de reprodutores, assim como previsão da produção de leite em um determinado momento da lactação (BIANCHINI SOBRINHO, 1984).

Segundo Cobuci et al. (2003), a curva de lactação apresenta três fases, sendo a primeira a fase ascendente que se inicia no parto e segue até o pico da lactação. A segunda fase é definida ao redor do pico e se apresenta de forma quase constante. A terceira e última fase inicia-se no pico, seguindo até o fim da lactação, apresentando comportamento decrescente. Madsen (1975), entre outros autores, aponta alguns fatores de ambiente como idade da vaca ao parto, estação do parto e ordem do parto como sendo de grande influência na curva de lactação.

Tomando como referência o trabalho de Brody et al. (1923), vários autores iniciaram estudos que se estendem até os dias atuais objetivando apresentar a curva de lactação por meio de uma função, ou melhor, expressar a curva de lactação usando um modelo estatístico que possibilite o entendimento da curva e dos fatores que influenciam na produção, bem como a previsão da produção. Entretanto, a busca pelo melhor modelo é no minímo um caminho árduo, pois há uma diversidade enorme de raças leiteiras bovinas, zebuínas, bubalinas e caprinas, sendo necessário levar em consideração todas as características individuais das raças que geralmente apresentam comportamentos diferentes com relação à produção de leite e lactação.

Na Tabela 2.1 é apresentado um resumo de modelos utilizados por alguns autores. Nos modelos, os β_i, a, b e c são parâmetros desconhecidos, n é o número de dias do parto até o controle leiteiro e t o t-ésimo dia de lactação.

Tabela 2.1: Modelos utilizados por alguns autores no ajuste da curva de lactação

Autor	Raça	Modelo
Singh e Gopal, (1982)	-	$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 lnt + \varepsilon_t$
Cunha Filho et al. (2006)	Sindi	$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 lnt + \varepsilon_t$
Muñoz-Berrocal et al. (2001) e Fraga, et al. (2003)	Murrah	$y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{t} + \varepsilon_t$
Rowlands et al. (1982)	-	$y_t = \beta_0 t^{\beta_1} e^{-\beta_2 t} + \varepsilon_t$
Brody et al. (1923)	-	$y_t = \beta_0 e^{-\beta_2 t}$
Brody et al. (1924)	-	$y_t = \beta_0 e^{-\beta_1 t} - \beta_0 e^{-\beta_2 t}$
Muñoz-Berrocal et al. (2005)	Murrah	$y = \beta_0 + \beta_1 x + \beta_2 x^{-1}$
Muñoz-Berrocal et al. (2005)	Murrah	$y = \beta_0 x^{\beta_1} e^{\beta_2 x}$
Muñoz-Berrocal et al. (2005)	Murrah	$y = x(\beta_0 + \beta_1 x + \beta_2 x^2)^{-1}$
Muñoz-Berrocal et al. (2005)	Murrah	$y = \beta_0 e^{\beta_1 x + \beta_2 x^2}$
Bianchini Sobrinho (1984)	-	$y = a + b^n + \frac{c}{n}$
Wood (1967)	-	$y = an^b e^{-cn}$
Dave (1971)	-	$y = a + bn - cn^2$
Papajcsik e Bodero (1988)	-	$y = \frac{an^b}{cosh(cn)}$
Papajcsik e Bodero (1988)	-	$y = a\, arctan(bn) e^{-cn}$
Papajcsik e Bodero (1988)	-	$y = a\, ne^{-cn}$
Grossman e Koops (1988)	-	$y = \sum_{i=1}^{2} \{a_i b_i [1 - tanh^2(b_i(n - c_i))]\}$
Madalena et al. (1979)	-	$y = a - cn$
Cobuci et al. (2000)	-	$y = a - cn + LN(n)$

Fonte: Elaborada pelo autor tomando como referência de formato, Cobuci et al. (2000)

Rowlands et al. (1982) utilizaram a função gama incompleta para ajustar um modelo que apresentasse uma previsão mais precisa para a produção de leite quando comparado com outros modelos. Uma desvantagem observada nesse modelo é que nas redondezas da lactância média, o modelo perde precisão em relação à previsão. Brody e colaboradores sugeriram dois modelos: um em 1923 e outro em 1924. Como visto, o estudo de 1923 foi o ponto inicial para diversos estudos. O modelo proposto em 1924 tinha como desvantagens o fato de subestimar a produção de leite quando se encontrava na metade da lactância e superestimar quando se encontrava no pico e no final da lactância. Fora esses modelos, Quintero et al. (2007) ainda apresentaram outros modelos e uma breve discussão sobre o assunto mostrando também um modelo misto, mas sem comentários sobre seu ajuste à alguma curva de lactação.

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, tomando como variável explicativa o número de dias entre o parto e o controle leiteiro. Dentre os resultados, os autores expõe que os modelos gama incompleto, quadrático logarítmico e regressão múltipla subestimam a produção no início da lactação, ao contrário do modelo monofásico que superestima a produção no mesmo período. Contudo, os autores sugerem que o modelo difásico é o mais adequado na representação da

curva de lactação dessa raça e, portanto, sendo melhor para estimar as produção de leite.

Estudando a curva de lactação da Holandês-Gir, Oliveira et al. (2007) ajustaram um modelo gama incompleto, utilizando os dias de lactação como variável explicativa. Para tanto o autor delimitou um número mínimo de 4 controles e um máximo de 12 controles, com intervalos de 30 dias. A variável de interesse foi a produção de leite em kg e a variável explicativa, o tempo em dias de lactação. O estudo foi direcionado à questão da época de parição (período de seca e período de águas). Para essa raça, foi observado que a ordem de lactação e a época de parição influenciaram diretamente na forma da curva estimada. Entretanto, os autores concluiram que o modelo não se ajustou adequadamente à produção de leite dessa raça.

Nos modelos ajustados e propostos por Cobuci et al. (2000), y representa a produção de leite, n é o número de dias do parto até o controle leiteiro e a, b, c e d são parâmetros a serem estimados. Uma observação importante é que as letras utilizadas como parâmetros são iguais em todos os modelos, mas seus significados ou interpretação podem ser diferentes. Depois de apresentar seus resultados, os autores sugerem que os modelos $y = ane^{-cn}$, $y = a - cn$, o modelo proposto por Brody et al. (1923) e o modelo proposto pelo autor do estudo são os modelos que melhor se adequam à curva de lactação da raça Guzerá.

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 polimômios de Legendre. Togashi e Lin (2007) propõem 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.

Claramente, observa-se a diversidade de aplicações e modelos para estudos da curva de lactação. Tais estudos mostram a grande variabilidade existente as curvas de lactação, pois nem todas as curvas apresentam pico no decorrer de sua lactação e sim logo no instante do parto. Essas diferenças, entre outras, justificam a utilização de diversos modelos na busca pelo modelo que melhor se ajuste à curva de lactação de uma determinada raça. Os estudos citados aqui e tantos outros que não foram citados, evidenciam uma evolução nos modelos utilizados.

Uma característica marcante na maioria dos ajustes é o fato de considerar as produções médias de leite durante as lactações e, por isso, o uso de modelos para ajuste de valores médios. Além disso, na maioria dos estudos que envolvem um ou mais modelos, a forma de verificação da adequacidade do modelo ou comparação entre modelos se dá pelo coeficiente de determinação ajustado, o que leva a considerar que a adequacidade do modelo pode não ser consistente dado que o coeficiente de determinação na sua forma original ou ajustado, sozinho, não é um bom indicador de qualidade de ajuste, sendo necessário aprofundar-se na verificação da adequacidade para o qual é fortemente indicada a análise de diagnóstico.

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 no estudo da curva de lactação, pois possibilitam incluir na estrutura do modelo um conjunto de variáveis não observáveis, denominados efeitos aleatórios, incorporando a variabilidade do animal, de forma a permitir junto às variáveis observáveis, denominadas de efeitos fixos, o ajuste de um modelo mais consistente, uma vez que esse modelo é capaz de incorporar o comportamento individual de cada animal. Os modelos mistos aparecem na literatura 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).

Com a suspeita de que fatores ambientais e genéticos afetavam a forma da curva de lactação da raça Guzerá, Cobuci et al. (2000) ajustaram um modelo definido por:

$$Y = X\beta + Zg + Wp + \varepsilon \tag{2.1}$$

sendo Y um vetor com as produções de leite da vaca em estudo, X a matriz de efeitos fixos, Z a matriz de efeito genético e W a matriz de efeitos relacionados ao ambiente. Os parâmetros β, g e p são desconhecidos e associados às matrizes descritas, respectivamente. Quanto às variáveis utilizadas como efeito de ambiente, há classe do rebanho, ano de parto, estação do parto e classes de idade ao parto. Os efeitos fixos considerados foram: ano de parto, estação do parto e idade da vaca no momento do parto. Esse modelo foi adotado com o objetivo direcionado para a estimação dos efeitos genéticos, não sendo descritos os resultados detalhadamente. Os autores concluíram que, com base nos resul-

tados encontrados, a escolha de animais, com base na curva de lactação estudada, não é eficiente.

Em 2001, Cobuci et al. apresentaram novo estudo considerando as mesmas diretrizes do estudo por eles realizado um ano antes. Utilizaram um modelo linear misto com os efeitos fixos: rebanho, ano de parto, estação de parto e idade do animal ao parto. Os efeitos aleatórios foram divididos em dois blocos: efeitos diretos genéticos e efeitos permanentes de ambiente. Nesse estudo, a análise do modelo misto é melhor abordada, sendo estudados os resíduos e verificada a presença de correlação serial entre os erros. Os autores indicam que selecionar vacas dessa raça quanto à produção de leite pode provocar acréscimo na produção inicial e alterações no declínio da curva no decorrer da lactação. Os autores concluem que a seleção não altera a forma da curva de lactação, mas que a raça, com maiores produções iniciais, apresenta declínios acentuados no decorrer da lactação. Além disso, indicam que é possível haver alteração na forma da curva se as condições do ambiente em que o animal está alocado passassem por melhorias.

Cruz et al. (2008), trabalhando com o mesmo objetivo de Cobuci et al. (2000, 2001), mas estudando vacas da raça Sindi, objetivaram estimar os parâmetros genéticos para a produção de leite dessa raça. Para tanto, utilizaram 373 lactações e 4.476 controles, dentre os anos de 1986 e 2004, considerando para esse estudo apenas as quatro primeiras lactações. Os autores fazem uma descrição do local de origem do estudo e das condições do manejo durante o período de estudo. Segundo os autores, a ordenha foi manual, duas vezes ao dia, com a lactação dividida em 8 estádios de 35 dias cada. Houve um processo de exclusão de casos em que animais com menos de quatro controles e aqueles com grupos contemporâneos com menos de dois animais, e, após a aplicação deste, o estudo se restringiu a 340 lactações e 2.668 controles. O modelo misto foi utilizado, considerando como efeitos aleatórios os efeitos genéticos aditivos e de ambiente permanentes e como efeitos fixos um grupo de variáveis denominado pelo autor como grupo contemporâneo composto pelas variáveis ano de controle, mês de controle, idade ao parto e duração da lactação. O modelo utilizado é dado por:

$$Y = Xb + Z_1a + Z_2ep + \varepsilon$$

em que Y é um vetor de produção de leite total ou produção de leite no dia do controle, a o vetor de efeitos genéticos aditivos diretos e ep o vetor de efeitos de ambiente permanentes. O método de estimação utilizado foi o de Máxima Verossimilhança Restrita (REML) (PINHEIRO e BATES, 2000), implementado em um programa denominado MTDFREML (BOLDMAN et al., 1995). Como resultados, os autores apresentam que o pico da lactação

foi no início da lactação, descrecendo no decorrer da lactação. Devido ao foco do estudo, o detalhamento dos resultados do modelo foram suprimido, sendo direcionado para os resultados genéticos por ele estimado. Entretanto, os autores concluem que em avaliações genéticas pode ser usada a produção de leite no dia do controle ao invés da produção total. Outro ponto importante é que a produção de leite na fase intermediária da lactação apresenta maior variabilidade genética, sugerindo que essa época é recomendada para avaliações nessa raça sob as condições do estudo.

3 Modelos Lineares Mistos

Grande parte dos estudos práticos nos diversos campos da ciência envolvem o uso de algum tipo de modelo de regressão, com o objetivo de descrever comportamentos, identificar fatores que afetam e/ou explicam certos fenômenos, entre outros objetivos relacionados. Dentre os modelos utilizados mais frequentemente estão os lineares e não-lineares múltiplos, modelos lineares generalizados e variações destes apresentados; mas, quase sempre, utilizando-se a suposição de que as observações são independentes. Entretanto, é bastante comum nas áreas biológicas, agronômicas, saúde e outras, a presença de situações em que os dados apresentam comportamentos dependentes entre si, ou seja, casos em que uma observação depende estocasticamente de outra. De forma geral, são chamados de dados agrupados que compreendem dados longitudinais, medidas repetidas e outros. Além disso, é interesse de alguns estudos a análise da correlação entre-grupos e intra-grupos que está relacionada diretamente com a presença de estruturas de dados agrupados. Nesses casos, os modelos usuais, tanto lineares como não-lineares, acabam por não serem válidos uma vez que a suposição de dados não-correlacionados não pode ser verificada.

Contudo, é possível ajustar modelos na presença de dados agrupados, levando em consideração a possível correlação existente entre as observações, sendo possível modelar as correlações entre e intra-grupos. Tais modelos são denominados *Modelos Mistos* ou também conhecidos como modelos de efeitos mistos, ou modelos de efeitos aleatórios. Eles permitem incorporar e estudar estruturas de covariância que são de fundamental importância no estudo de dados agrupados. Além disso, é possível estudar o comportamento individual e o comportamento médio, ao contrário dos modelos usuais que ajustam o comportamento médio. A classe de modelos mistos, atualmente, tem sido enriquecida com novos métodos e variações. Mesmo com tal riqueza, esse estudo está focado na aplicação de um modelo linear misto, sendo os demais deixados à curiosidade do leitor.

Ao contrário dos modelos lineares múltiplos que levam em consideração um conjunto de covariáveis fixas denominado de efeitos fixos, os modelos lineares mistos levam em

consideração os efeitos fixos e um segundo conjunto de covariáveis denominado de efeitos aleatórios. A incorporação desse conjunto ou bloco de efeitos aleatórios no modelo é a principal característica dos modelos lineares denominados mistos. Com isso, é possível incorporar também uma estrutura de covariâncias que será preponderante na análise da correlação entre e intra-grupos. Dessa forma, tem-se que os modelos lineares mistos são extensões de modelos lineares múltiplos.

3.1 Definição do modelo

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

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, i = 1,\ldots,M \tag{3.1}$$
$$\gamma_i \sim N(0,\Psi), \quad \varepsilon_i \sim N(0,\Sigma)$$

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.

Podemos reescrever o modelo (3.1) na seguinte forma funcional:

$$Y = X\beta + Z\gamma + \varepsilon \tag{3.2}$$

em que

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}; Z = \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_q \end{bmatrix};$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}; \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}.$$

De (3.2) 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 semi-definida. A matriz Ψ pode assumir diversas estruturas de covariâncias que estão relacionadas com particularidades do estudo em questão.

3.2 Estimação dos parâmetros

Atualmente, pode-se encontrar na literatura diversos métodos para estimação de parâmetros em modelos de regressão, além de refinamentos e novos métodos que são apresentados a cada dia. No caso do modelo linear misto, dois métodos são os mais utilizados: o de Máxima Verossimilhança e o de Máxima Verossimilhança Restrita. Nesta seção, são apresentados os dois métodos, apesar de utilizar-se neste estudo o de Máxima Verossimilhança, que será justificado no momento devido.

3.2.1 Estimação por Máxima Verossimilhança

No modelo de regressão usual, com erros seguindo uma distribuição normal, ou em outros, como os modelos lineares generalizados, tem-se que o método de Máxima Verossimilhança é utilizado para estimação dos parâmetros nele envolvidos. Como já é sábido, o método de Máxima Verossimilhança tem como base o comportamento probabilístico do erro associado ao modelo no caso dos modelos lineares usuais ou no comportamento da variável de interesse nos modelos lineares generalizados. Sendo assim, não conhecer o comportamento probabilístico impossibilita o uso do método de Máxima Verossimilhança.

Nos modelos lineares mistos, a estimação dos parâmetros se dá de forma diferenciada quando comparada com outras estimações. Nos modelos usuais, o método de Máxima Verossimilhança é utilizado de forma direta, pois se conhece o comportamento das partes envolvidas no modelo e a matriz de covariáveis é observável e fixa. Como no modelo linear misto, o comportamento probabilístico dos efeitos aleatórios não é conhecido, a utilização do método de Máxima Verossimilhança não é possível.

Dessa forma, a inferência relacionada aos modelos lineares mistos é fundamentada na densidade marginal do vetor de respostas y_i. Sendo assim, segundo Verbeke e Molen-

berghs (2000), condicionando o vetor de respostas aos efeitos aleatórios, assume-se:

$$y_i|\gamma_i \sim N_{n_i}(X_i\beta + Z_i\gamma_i, \Sigma_i)$$

Adicionando a suposição de que $\gamma_i \sim N(0,\Psi)$ e definindo $f(y_i|\gamma_i)$ e $f(\gamma_i)$ como sendo as funções densidades de $y_i|\gamma_i$ e γ_i, respectivamente, tem-se que a densidade marginal de y_i é $y_i \sim N_{n_i}(X_i\beta, Z_i\Psi Z_i^T + \Sigma)$. Sendo assim, $f(y_i)$ é a densidade marginal em que é baseada a estimação dos parâmetros do modelo (2.1).

Segundo Laird e Ware (1982), os modelos lineares mistos são modelos de dois estágios, em outras palavras, modelos com estruturas hierárquicas implícitas. No primeiro estágio, os vetores de coeficientes são considerados fixos assumindo que os ε_i são independentes e, no segundo estágio, assumindo que os γ_i são independentes e ε_i sendo independente de γ_i. Ainda segundo os autores, quando $\Sigma = \sigma^2 I_{n_i}$ o modelo descrito em (2.1) é chamado de *modelo de independência condicional homoscedástico* que será o modelo aplicado nesse estudo.

Dado que é necessário estimar σ^2 conjuntamente com os parâmetros β, a Verossimilhança deve ser construída de forma a permitir tal estimação. Segundo Pinheiro e Bates (2000), denota-se $V_i = Z_i\Psi Z_i^T + \sigma^2 I$ fazendo $\sigma^2 V_i = I + (Z_i\Psi Z_i^T)/\sigma^2$. Dessa forma tem-se que a função de Verossimilhança do modelo marginal para a i-ésima unidade amostral é dada por:

$$l(\gamma_i, \sigma^2|y_i) = -\frac{1}{2}log|V| - \frac{1}{2}(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})$$

Maximizando a Verossimilhança dada em (3.3) com relação a β e σ^2 e definindo $N = \sum_{i=1}^{M} n_i$, obtêm-se os estimadores dos respectivos parâmetros que são dadas por:

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T V^{-1}(y - X\hat{\beta})}{N}$$

3.2.2 Estimação por Máxima Verossimilhança Restrita

O método de Máxima Verossimilhança Restrita, também conhecido como Máxima Verossimilhança Residual, proposto por Patterson and Thompson (1971), visa estimar os componentes da variância. Utilizando o método de Máxima Verossimilhança tem-se que a estimativa de σ^2 é viesada e as estimativas dos parâmetros acabam subestimadas. Segundo Davis (2002), o método de Máxima Verossimilhança Restrita está associado aos

"contraste de erros"e não às observações reais, sendo assim, as estimativas para os componentes da variância são estimados com menos viés.

Segundo Davis (2002), seja Y o vetor referente à variável de interesse, um contraste de erro é uma transformação linear $A'y$ de forma que $E(A'y) = 0$! i. e., que $E(A'X) = 0$! . De posse dessa definição, tem-se que o método de Máxima Verossimilhança Restrita baseia-se na aplicação de método de Máxima Verossimilhança considerando $w = A'y$ em vez de y.

Seja $y \sim N_n(X\beta, V)$, em que $V = Z\Psi Z' + \Sigma$, então, $A'y \sim N_{n-p}(0_{n-p}, A'VA)$. Dessa forma, temos que os parâmetros são estimados baseando a Verossimilhança em $f_w(w, \theta)$, em que θ representa um vetor de parâmetros a serem estimados.

Com base nessas definições, temos que o logaritmo da função de Máxima Verossimilhança restrita é dado por:

$$l(\hat{\beta}, \theta) = -\frac{1}{2}\{log|X^T V^{-1} X| + log|V| + \hat{e}^T \hat{e}\}$$

com $\hat{e} = y - X\tilde{\beta}$ e $\tilde{\beta} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} y$ Maximizando a Verossimilhança com relação aos parâmetros, obtem-se estimadores dados por:

$$\hat{\beta} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y$$

$$\hat{\sigma}^2_{REML} = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}$$

3.2.3 BLUE e BLUP

Fazendo uso da teorema de Gauss-Markov, Harville (1976) obteve o BLUE (Best Linear Unbiased Estimator) para o vetor de efeitos fixos β e o BLUP (Best Linear Unbiased Predictor) para o vetor de efeitos aleatórios γ. O BLUE e o BLUP, $\hat{\beta}$ e $\hat{\gamma}$ respectivamente, são funções lineares de Y, são não-viesados de variância mínima e são os melhores estimadores (preditores) de β e γ, respectivamente.

3.3 Escolha do modelo

Para escolha do modelo, há na literatura, critérios que fornecem estatísticas que auxiliam na decisão de qual modelo escolher. Aqui, faz-se uso de dois critérios: o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion). O AIC e BIC são definidos

como:

$$AIC = -2l + 2p \tag{3.3}$$

$$BIC = -2l + plog(n) \tag{3.4}$$

em que l representa o máximo da log-verossimilhança, p a quantidade de parâmetros e n o quantidade de observações. Como decisão, adota-se como melhor modelo o que apresentar o menor valor em qualquer uma dessas duas estatísticas. É importante ressaltar que o BIC é mais consistente que o AIC e portanto, será o critério de maior peso na escolha do melhor modelo.

3.4 Testes de hipóteses

Indiferente do modelo adotado, os testes de hipóteses configuram uma parte fundamental no processo de ajuste. São responsáveis pela determinação da significância das estimativas dos parâmetros envolvidos e do modelo em si. Para modelos lineares mistos os testes de hipóteses são aproximados. Todavia, vários testes são encontrados na literatura, tais como Wald, Escore e o teste da Razão de Verossimilhanças. Este estudo foca-se no teste da Razão de Verossimilhanças, o qual é apresentado a seguir.

3.4.1 Teste da Razão de Verossimilhanças

Para modelos ajustados pelo método de Máxima Verossimilhança, o teste mais comumente utilizado é o teste da Razão de Verossimilhanças (PINHEIRO e BATES, 2000). Tal teste basea-se na razão entre as Verossimilhanças de dois modelos ajustados. Na prática, é fundamental para verificar se um modelo com menos parâmetros se ajusta tão bem quanto o modelo com a quantidade total de parâmetros. Segundo Pinheiro e Bates (2000), nos casos em que os parâmetros são estimados por Máxima Verossimilhança Restrita, o teste da Razão de Verossimilhança só pode ser aplicado se os dois modelos foram ajustados pelo mesmo método e se os efeitos fixos têm a mesma estrutura.

A estatística do teste da Razão de Verossimilhanças é dada por:

$$\xi_{RV} = 2[log(L_2) - log(L_1)] \tag{3.5}$$

em que L_1 representa o valor maximizado da log-verossimilhança sob o modelo reduzido e L_2, o valor maximizado da log-verossimilhança sob o modelo completo.

O teste da Razão de Verossimilhanças apresenta distribuição χ^2_r, em que r é a diferença entre a quantidade de parâmetros do modelo completo e do modelo restrito. É importante ressaltar que essa aproximação só é válida se o modelo reduzido não se encontrar na fronteira do espaço paramétrico. Além disso, nos casos em que o interesse é testar hipóteses referentes aos efeitos fixos, quando utilizada a Máxima Verossimilhança Restrita, o teste da Razão de Verossimilhanças não é indicado, uma vez que, ao utilizar tal método, os efeitos fixos são desconsiderados. Pinheiro e Bates (2000) não recomendam utilizar a Razão de Verossimilhanças para testar efeitos fixos, uma vez que o teste segue uma distribuição de referência qui-quadrado e assim tende a ser *anticonservativo*.

Visando testar a significância dos efeitos fixos, Pinheiro e Bates (2000) sugerem que uma alternativa é condicionar a especificação desses efeitos às estimativas das variâncias e covariâncias dos efeitos aleatórios. Este teste condicional é dado pelos teste-F e teste-t usuais, como definidos nos modelos lineares, sendo condicionados à

$$\hat{\sigma}^2_R(\theta) = s^2 = \frac{RSS}{N-p} \tag{3.6}$$

em que θ refere-se aos parâmetros envolvidos nos efeitos fixos, RSS é a soma de quadrados do resíduo, N é a soma dos n_i e p a quantidade de parâmetros.

3.4.2 Intervalo de confiança

Da mesma forma que nos testes de hipóteses, os intervalos de confiança para as componentes da matriz de variâncias e covariâncias e para os efeitos fixos são obtidos a partir distribuições aproximadas das estimativas de máxima verossimilhança ou máxima verossimilhança restrita e também do teste-t condicional (PINHEIRO e BATES, 2000). A partir dessas indicações tem-se intervalos de confiança aproximados e que são definidos como se segue.

Denotando $\hat{\beta}_j$ como sendo a estimativa do j-ésimo efeito fixo e gl_j o os graus de liberdade associados ao teste-t condicional referente à estimativa do j-ésimo efeito fixo, então, o intervalo de confiança aproximado com nível (1-α) para β_j é dado por

$$\hat{\beta}_j \pm t_{gl_j}(1-\alpha/2)\hat{\sigma}_R\sqrt{[R^{-1}_{00}R^{-T}_{00}]_{jj}} \tag{3.7}$$

em que $t_{gl_j}(1-\alpha/2)$ representa o quantil de ordem $(1-\alpha/2)$ da distribuição t-student com gl_j graus de liberdade. O termo R_{00} segue da decomposição ortogonal-triangular:

$$\begin{bmatrix} Z_i & X_i & y_i \\ \Delta & 0 & 0 \end{bmatrix} = Q_{(i)} \begin{bmatrix} R_{11(i)} & R_{10(i)} & c_{1(i)} \\ 0 & R_{00(i)} & c_{0(i)} \end{bmatrix};$$

Denotando $[X^{-1}]_{\sigma\sigma}$ como sendo o último elemento da diagonal da inversa da matriz de informação empírica (PINHEIRO e BATES, 2000), o intervalo de confiança aproximado para o desvio padrão intra-grupo σ com nível (1-α) é dado por

$$\left[\hat{\sigma} exp\left(-z(1-\alpha/2)\sqrt{[X^{-1}]_{\sigma\sigma}}\right), \hat{\sigma} exp\left(z(1-\alpha/2)\sqrt{[X^{-1}]_{\sigma\sigma}}\right)\right] \tag{3.8}$$

em que $z_{(1-\alpha/2)}$ representa o quantil de ordem $(1-\alpha/2)$ da distribuição normal padrão.

3.5 Previsão

A previsão é um dos principais interesses no ajuste de um modelo. Os valores ajustados, como define Pinheiro e Bates (2000), são as previsões realizadas para as respostas observadas. As previsões são de grande importância na verificação de adequacidade do modelo, uma vez que deseja-se previsões mais próximas possíveis o que determina menores magnitudes dos resíduos.

Segundo Pinheiro e Bates (2000), a partir do modelo linear misto é possível fazer previsões em dois níveis diferentes: o nível populacional e o nível individual. No nível populacional, os valores ajustados representam os valores preditos para os valores marginais esperados da variável resposta, enquanto que no nível individual, as predições representam os valores esperados condicionados aos efeitos aleatórios estimados daquele indivíduo. Sendo assim temos:

$$E[y_h] = x_h^{'}\beta \tag{3.9}$$

$$E[y_h(i)|\gamma_i] = x_h^{'}\beta + z_h(i)^{'}\gamma_i \tag{3.10}$$

em que x_h representa um vetor de covariáveis associado aos efeitos fixos, $z_h(i)$ é um vetor de covariáveis associadas aos efeitos aleatórios do i-ésimo grupo.

Com isso, os valores preditos são dados por:

$$\hat{y}_h = x_h^{'}\hat{\beta} \tag{3.11}$$

$$\hat{y}_h(i) = x_h^{'}\hat{\beta} + z_h(i)^{'}\hat{\gamma}_i \tag{3.12}$$

em que $\hat{\beta}$ e $\hat{\gamma}$ são o BLUE e o BLUP, respectivamente.

3.6 Análise de resíduos

A análise de resíduos é de importância indiscutível no que tange o estudo da adequacidade do ajuste de modelos de qualquer natureza, sejam modelos lineares em sua forma simples até modelos generalizados lineares e não-lineares e modelos mais complexos. Abordando o estudo da adequabilidadedade de forma geral, esta objetiva verificar se as suposições impostas pelo modelo são atendidas. Contudo, o estudo da adequacidade vai além da verificação de suposições, tendo como preocupação, também, verificar a forma como os casos, observações, influenciam no ajuste do modelo que estiver em questão. Vale ressaltar que cada modelo carrega uma determinada estrutura e apesar da metodologia de ajuste, entenda-se por metodologia de ajuste a estimação dos parâmetros que geralmente trilha pelo caminho da máxima verossimilhança, a abordagem aos resíduos deve ser cuidadosa, uma vez que a estrutura dos resíduos que melhor se encaixa ao estudo da adequacidade varia de modelo para modelo.

Nobre e Singer (2007) apresentam três tipos de erros para os modelos lineares mistos. As três abordagens são necessárias para o estudo da adequabilidade devido às suas características que possibilitam estudar um conjunto de suposições diferentes.

Os três tipos de erros citados por Nobre e Singer (2007) são denominadas e dados por:

Erros Condicionais

$$\varepsilon = Y - X\beta - Z\gamma \tag{3.13}$$

Efeitos aleatórios

$$Z\gamma = E[Y|\gamma] - E[Y] \tag{3.14}$$

Erros marginais

$$\xi = Y - X\beta = Z\gamma + \varepsilon \tag{3.15}$$

Desta forma os valores estimados de ε e ξ denotados por $\hat{\varepsilon}$ e $\hat{\xi}$ são dados por: $\hat{\varepsilon} = Y - X\hat{\beta} - Z\hat{\gamma}$ e $\hat{\xi} = Y - X\hat{\beta}$, em que $\hat{\beta}$ é o BLUE de β e $\hat{\gamma}$ é o BLUP de γ.

Segundo Pinheiros e Bates (2000), antes de quaisquer inferências, duas suposições devem ser verificadas nos modelos lineares mistos: Se os erros intra-grupos são independentes e identicamente distribuídos seguindo uma distribuição normal com média zero e variância σ^2 e se são independentes dos efeitos aleatórios. A outra suposição refere-se à normalidade dos efeitos aleatórios e são independentes para diferentes grupos.

No intuito de verificar o afastamento dessas suposições, Pinheiro e Bates (2000) propõem o uso do gráfico de probabilidade normal dos resíduos condicionais para avaliar a suposição de normalidade e o gráfico dos resíduos condicionais *versus* os valores ajustados para 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 $\hat{\varepsilon}$ apresentarem variâncias diferentes, propõe uma padronização dos resíduos condicionais dada por:

$$\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i}{\sigma\sqrt{q_{ii}}} \tag{3.16}$$

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

O resíduo condicional padronizado, $\hat{\varepsilon}^*$, obtido a partir de (3.10) é mais eficaz na verificação de possível afastamento da suposição de homoscedasticidade, fazendo uso no gráfico de resíduos condicionais padronizados *versus* valores ajustados.

No caso da verificação da normalidade do erro ε, o erro puro, que foi definido por Hilden-Minton (1995) citado por Nobre (2004)como sendo o 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) citado por Nobre (2004) define a fração de confundimento de $\hat{\varepsilon}_i$ como

$$CF_i = 1 - \frac{U_i^TQU_i}{U_i^TQU_i} \tag{3.17}$$

em que U_i é a i-ésima coluna de I_n e $R = I_n$. A medida CF_i é interpretada como sendo a proporção de variabilidade de $\hat{\varepsilon}_i$ 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 $L^T\hat{\varepsilon}$ minimizando o confundimento de $l_i\hat{\varepsilon}$, sendo l_i a i-ésima linha da matrix L. Tomando a restrição de que $Var(l_i^T\hat{\varepsilon}) \propto l_i^TRQRl_i > 0$ e após alguma álgebra, tem-se que o resíduo $(l_i^T\hat{\varepsilon}/\sigma)$ é não correlacionado com a fração de confundimento. Tal resíduo é denominado resíduo com confundimento mínimo e seu uso é mais eficiente na verificação da normalidade dos erros condicionais por intermédio do gráfico de probabilidade normal com envelope.

Dado que a partir do modelo (3.1), é possível estimar o comportamento tanto médio como por indivíduos, é importante estudar não apenas possíveis observações discrepan-

tes, 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

$$\zeta_i = \hat{\gamma}_i^T \hat{Var}(\hat{\gamma}_i^T - \gamma_i)\hat{\gamma}_i^T \tag{3.18}$$

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

Contudo, a análise de resíduo, apesar de fundamental, não é completa necessitando aprofundar ainda mais a consistência do modelo ajustado. Sendo assim, chega-se à análise de sensibilidade, que aborda verificações que completam a análise de resíduo e conseqüentemente a adequacidade do modelo.

3.7 Análise de sensibilidade

A análise de sensibilidade visa entender o comportamento do modelo quando este está sendo sujeitado a algum tipo de perturbação seja nas hipóteses ou nos dados. Dentre as mais utilizadas formas de avaliar a sensibilidade do modelo, encontra-se a eliminação de observações que 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. Fora esse método, outros métodos como identificação de pontos de alavanca, distância de Cook condicional e influência local são utilizados para mensurar a sensibilidade do modelo.

3.7.1 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. Desde então, diversos autores trabalharam e ainda trabalham no sentido de verificar a influência que as observações exercem sobre o ajuste do modelo trazendo novas metodologias baseadas na exclusão de observações e reajuste do modelo. Nesta direção, há o que é denominado de fórmulas de atualização que visam atualizar as

estimativas dos parâmetros quando uma determinada observação for excluída. A utilização dessas fórmulas se torna essencial, pois não é necessário reajustar o modelo. Fei e Pan (2003) propõem utilizar a estrutura de covariância fazendo com que o processo de estimação seja linear e com isso encontrar a relação entre os estimadores do modelo com todas as informações e o modelo com a observação excluída. Com isso, é de extrema importância que a estrutura de covariância seja escolhida adequadamente. Entretanto, nos modelos lineares mistos a eliminação de observações não se apresenta como uma tarefa simples dado que a estimação é realizada iterativamente, sendo necessário aproximações para as fórmulas de atualização.

A preocupação acerca da eliminação de observações, 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 que se trata da análise da influência de um conjunto de observações tomando a idéia da eliminação desse conjunto, como pode ser visto em vários textos em muitas estruturas de modelos distintas, a distância de Cook é a medida mais utilizada nesta tarefa.

A distância de Cook para modelos lineares misto (NOBRE, 2004) é dado por

$$D_I = \frac{(\hat{\beta} - \hat{\beta}_I)^T (X^T V^{-1} X)(\hat{\beta} - \hat{\beta}_I)}{c} = \frac{(\hat{Y} - \hat{Y}_I)^T V^{-1} (\hat{Y} - \hat{Y}_I)}{c} \tag{3.19}$$

em que c é um parâmetro de escala e $\hat{\beta}_I$ e $\hat{Y}_i$ são as estimativas de de β e Y retirando o conjunto de observações I.

Segundo Fung et al. (2002), a equação (3.13) pode ser utilizada para detectar tanto observações como indivíduos influentes, pois, uma vez que I representa um conjunto de observações e fazendo I todas as observações de um determinado indivíduo, tem-se exatamente a influência causada por este. Entretanto, Banerjee (1998) e Tan et al. (2001) citados por Nobre (2004) discutem limitações da distância de Cook dada na forma (3.13) na classe de modelos lineares mistos, pois tal medida pode não detectar observações ou indivíduos influentes quando estes apresentam grande influência em $\hat{\gamma}$.

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

$$D_i^{cond} = \sum_{j=1}^{c} \frac{P_{j(i)}^T P_{j(i)}}{k} \tag{3.20}$$

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 (3.21) é útil para verificação de observações que estejam influenciando as estimativas dos efeitos fixos e aleatórios. Nesse contexto, a equação (3.21) pode ser decomposta em três partes permitindo abordar a influência em partes distintas do modelo ajustado.

A decomposição de (3.14) se dá na forma de

$$D_i^{cond} = D_{1i}^{cond} + D_{2i}^{cond} + D_{3i}^{cond} \tag{3.21}$$

sendo

$$D_{1i}^{cond} = \frac{(\hat{\beta} - \hat{\beta}_i)^T (X^T X)(\hat{\beta} - \hat{\beta}_i)}{K} = \frac{(\hat{Y} - \hat{Y}_{(i)})^T (\hat{Y} - \hat{Y}_{(i)})}{k} \tag{3.22}$$

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

$$D_{2i}^{cond} = \frac{(\hat{\gamma} - \hat{\gamma}_{(i)})^T Z^T Z (\hat{\gamma} - \hat{\gamma}_{(i)})}{k} \tag{3.23}$$

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

e

$$D_{3i}^{cond} = \frac{2(\hat{\beta} - \hat{\beta}_{(i)})^T X^T Z (\hat{\gamma} - \hat{\gamma}_{(i)})}{k} \tag{3.24}$$

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

Contudo, ao excluir o conjunto de observações referentes à um individuo, tem-se que a obtenção de alguns BLUP não é possível (NOBRE, 2004). Com base nesse problema, Nobre (2004) propoe utilizar a médias das distâncias ao invés da distância de Cook condicional definida em (3.21) que é dada por

$$D_{i.}^{cond} = (n_i)^{-1} \sum_{j \in I} D_j^{cond} \tag{3.25}$$

em que n_i é o conjunto de observações do i-ésimo indivíduo.

3.7.2 Pontos de alavanca

Define-se o i-ésimo ponto de alavanca como sendo a i-ésima observação que influencia em sua própria estimativa (PAULA (2004), NOBRE e SINGER (2010)). Nos modelos normais lineares, a i-ésima observação é denominada ponto de alavanca se o i-ésimo elemento da diagonal principal da matriz H, sendo $H = X(X^TX)^{-1}X^T$. Segundo Paula (2004), a matriz H é denominada matriz *hat*, é simétrica e idempotente, sendo $posto(H) = tr(H) = \sum_{i=1}^{n} h_{ii} = p$ em que h_{ii} representa o i-ésimo elemento da diagonal principal de H. Dessa forma, temos que altos valores de h_{ii} indicam influência da observação em sua própria estimativa. Uma outra interpretação para os altos valores de h_{ii} é dada por Paula (2004) que trata h_{ii} como sendo a variação de $\hat{y}_i$ quando y_i é acréscido de um infinitésimo.

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

$$GL(\hat{\beta}) = \frac{\partial \hat{Y}}{\partial Y^T} \tag{3.26}$$

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, $GL(\hat{\beta}) = H$.

Até o momento foram definidos pontos de alavanca como sendo observações que influenciam em suas próprias estimativas. Entretanto, 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

$$H_i = X_i(X^TV^{-1}X)^{-1}X_i^TV_i^{-1} \tag{3.27}$$

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}. Neste sentido, fazendo $h_{ii}^* = GL(\hat{\beta})_{ii}$, tem-se

$$\overline{h}^* = n^{-1}\sum_{i=1}^{n} h_{ii}^* = \frac{p}{n} \tag{3.28}$$

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

$$\frac{tr(H_i)}{n_i} = \frac{\sum_{j\in I}}{n_i} \geq 2p/n \tag{3.29}$$

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

$$GL(\hat{\beta},\hat{\gamma}) = \frac{\partial \hat{Y}^*}{\partial Y^T} = \frac{\hat{Y}}{\partial Y^T} + \frac{\partial Z\hat{\gamma}}{\partial Y^T} = GL(\hat{\beta}) + ZDZ^TQ \tag{3.30}$$

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 ZDZ^TQ que leva em consideração a estrutura da matriz de covariâncias e a matriz de efeitos aleatórios. Sendo assim, $GL(\hat{\beta})_{ii} = h^*_{ii} + (ZDZ^TQ)_{ii}$ 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, 2010).

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 $(n_i)^{-1}\sum_{j\in I} GL(\hat{\beta})_{jj} \geq 2tr(GL(\hat{\beta}))/n$.

No próximo capítulo aplica-se o modelo e as análises aqui definidos na produção de leite das cinquenta e quatro vacas da raça Sindi em estudo. Apresentar-se-á três modelos diferindo quanto à presença dos efeitos aleatórios nos efeitos fixos utilizando duas estruturas da matriz de covariâncias. A escolha do modelo e a análise de resíduo e sensibilidade também são apresentados no próximo capítulo.

4 Aplicação

Neste capítulo são apresentados os resultados obtidos a partir do uso da metologia apresentada anteriormente. Este capítulo está estruturado em quatro partes que consistem na descrição dos dados utilizados, uma análise descritiva, o ajuste do modelo linear misto e a análise de resíduos e sensibilidade. Toda a análise apresentada nesse capítulo foi realizada no software livre R em sua versão 2.9.2. O nível de significância adotado em toda a análise foi de 5% ($\alpha = 0,05$).

4.1 Descrição dos dados

Para o estudo da curva de lactação de vacas da raça Sindi, foram utilizados dados sobre 402 lactações de 54 vacas desta raça em sua primeira ordem de parto. Tais dados foram obtidos a partir de uma base de dados primária na qual haviam dados sobre 1165 lactações de 87 vacas no período de 1987 a 1997 obtida 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 concentrado protéico e energético durante a ordenha. Durante a seca ou época de sol, recebiam alimentação diferenciada baseada em feno de capim *buffel*, capim elefante, palma forrageira picada, raspa de mandioca, bagaço de cana hidrolisado e concentrado protéico. Além disso, sal mineral estava disponível em cochos e uma mistura natural estava disponível o ano todo. 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.

4.2 Análise descritiva

A Tabela 4.1 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 menores médias do período e por conseguinte, 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.

Tabela 4.1: Estatísticas descritivas da produção de leite em kg das 54 vacas por estádio

Estádio	Número de vacas	Mínimo	Máximo	Média	Desvio padrão	Coeficiente de variação	Total produzido
1	54	0,0	9,9	4,9	3,37	0,68	267,0
2	54	0,0	9,7	5,5	3,55	0,65	294,3
3	53	0,0	9,9	6,5	3,06	0,47	344,1
4	52	0,5	10,8	7,0	2,72	0,39	361,6
5	41	0,2	11,0	7,2	2,28	0,31	369,2
6	50	0,2	10,8	6,6	2,02	0,30	331,0
7	46	1,0	9,7	6,1	1,65	0,27	279,6
8	42	3,2	8,4	5,5	1,21	0,22	232,2

Este comportamento observado na Tabela 4.1, sugere uma possível interpretação de que os valores mínimos estão influenciando diretamente na variação da produção de leite nos estádios, apesar das médias estarem se comportando como reflexo dos valores máximos. Contudo, está claro que valores mínimos maiores acarretam em menos variação, como pode ser visto no coeficiente de variação que decai rapidamente com o oitavo estádio, atingindo uma diferença de aproximadamente 40% a menos quando comparado com o primeiro estádio. Vale enfatizar que, analisando o coeficiente de variação separadamente, verifica-se que os três primeiros estádios possuem alta variabilidade e os demais, uma variabilidade moderada, sendo o oitavo estádio o que menos apresentou essa característica.

Analisando o comportamento da produção de leite por vaca indiferente de estádio de lactação, é claramente perceptível, na Tabela 4.2, a diferença existente entre as vacas no que se trata de suas produções totais no período observado da lactação, destacando a vaca V27 que produziu 77,3 kg, enquanto a vaca V41 produziu apenas 9,7 kg, sendo a diferença entre elas de 67,6 kg. Da Figura 4.1 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.

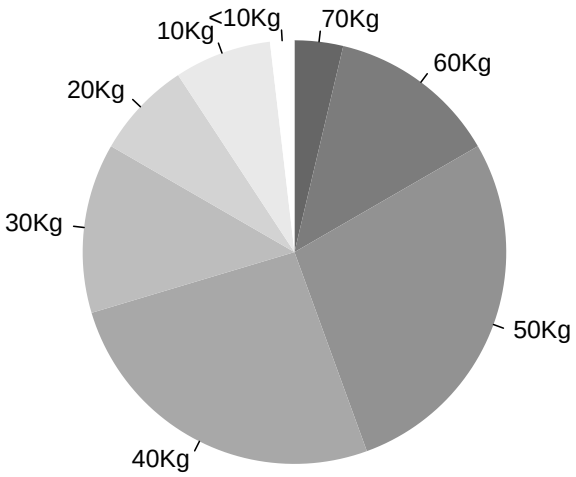

Figura 4.1: Gráfico de setores referente a distribuição da produção de leite total das 54 vacas

Ainda analisando os resultados apresentados na Tabela 4.2, observa-se, com relação à média, que as vacas V27 e V28 apresentaram as maiores produções médias seguidas da vaca V42. Entretanto, ao contrário das vacas V27 e V28, a vaca V42 apresenta uma das menores produções totais. Isso se deve ao fato de que esta vaca só foi observada em dois estádios, tornando os resultados para tal vaca inapropriados. Contudo ao observar os valores mínimo e máximo da produção desta vaca, observa-se que ela apresentou alto valor mínimo e que este retrata a observação referente ao segundo estádio, enquanto que o valor máximo de 9,0 kg representa o primeiro estádio, o que leva a supor, ou melhor, leva a entender que a vaca V42 teria uma boa produção nos demais estádios. No geral, as vacas apresentam produções médias diferenciadas, variando de 9,66 kg a 1,69 kg.

Aprofundando-se na análise da Tabela 4.2, podem-se observar grandes diferenças nos valores do desvio-padrão, que chega a alcançar 5,17 kg, sendo o menor desvio-padrão de 0,33 kg. Analisando os valores mínimos e máximos, observa-se vários valores mínimos abaixo de 1,0 kg, chegando alguns casos a apresentarem produção zero. Mesmo assim, boa parte dos valores mínimos encontra-se acima de 4,0 kg, chegando em três casos aos valores 7,4, 7,4 e 7,9 kg. No que se trata dos valores máximos, o destaque são as vacas

V27, V28, V3, V51 e V54 que apresentaram valores máximos acima de 10 kg. Analisando as medidas conjuntamente, observa-se que as vacas que obtiveram os maiores valores de desvio-padrão foram aquelas que apresentaram valores mínimos menores que 1,0 kg, enquanto os menores desvios-padrão são de vacas que apresentaram altos valores mínimos e, além disso, próximos dos valores máximos.

Tabela 4.2: Estatísticas descritivas da produção de leite em kg das 54 vacas

Vaca	Número de estádios	Mínimo	Máximo	Média	Desvio padrão	Coeficiente de variação	Total produzido
V1	6	3,3	7,8	4,88	1,75	0,36	29,3
V2	8	5,4	9,1	7,43	1,20	0,16	59,4
V3	8	4,4	10,1	7,66	1,91	0,25	61,3
V4	8	0,4	9,3	6,81	2,78	0,41	54,5
V5	8	0,1	9,7	7,45	3,11	0,42	59,6
V6	8	3,2	6,0	4,63	1,10	0,24	37,0
V7	8	4,7	8,2	5,95	1,14	0,19	47,6
V8	8	4,5	8,8	5,86	1,63	0,28	46,9
V9	8	0,4	8,1	5,38	2,99	0,56	43,0
V10	8	3,0	8,3	4,75	1,81	0,38	38,0
V11	8	3,8	7,0	5,78	1,22	0,21	46,2
V12	8	1,0	8,9	5,89	2,98	0,51	47,1
V13	8	1,6	9,0	5,64	2,08	0,37	45,1
V14	8	1,2	8,4	5,38	2,27	0,42	43,0
V15	8	0,2	8,0	3,28	2,84	0,87	26,2
V16	8	0,5	9,8	6,41	3,54	0,55	51,3
V17	8	4,0	8,4	6,39	1,84	0,29	51,1
V18	8	0,4	8,5	6,53	2,63	0,40	52,2
V19	8	5,0	8,7	7,53	1,30	0,17	60,2
V20	8	0,9	9,2	6,06	2,55	0,42	48,5
V21	8	6,3	7,2	6,65	0,33	0,05	53,2
V22	8	5,3	9,8	7,73	1,64	0,21	61,8
V23	8	0,5	9,0	5,36	2,88	0,54	42,9
V24	8	5,1	7,8	6,39	0,91	0,14	51,1
V25	8	0,0	9,0	5,68	3,42	0,60	45,4
V26	8	0,0	9,7	7,04	3,04	0,43	56,3
V27	8	7,9	11,0	9,66	0,92	0,10	77,3
V28	8	7,4	10,8	9,26	1,35	0,15	74,1
V29	8	5,8	8,7	7,60	0,99	0,13	60,8
V30	8	5,9	9,1	7,80	1,00	0,13	62,4
V31	8	4,5	9,0	7,10	1,74	0,24	56,8
V32	8	4,4	7,7	6,14	1,32	0,21	49,1
V33	8	5,0	9,5	7,66	1,84	0,24	61,3
V34	8	5,0	8,5	7,11	1,42	0,20	56,9
V35	8	5,0	8,0	6,68	1,08	0,16	53,4
V36	8	5,0	9,1	7,70	1,28	0,17	61,6
V37	8	4,2	8,7	6,48	1,77	0,27	51,8
V38	8	0,7	9,7	7,20	2,98	0,41	57,6
V39	4	0,3	9,4	6,88	4,40	0,64	27,5
V40	8	0,4	9,7	6,16	2,64	0,43	49,3
V41	3	0,1	9,2	3,23	5,17	1,60	9,7
V42	2	7,4	9,0	8,20	1,13	0,14	16,4
V43	8	1,0	9,5	6,29	3,05	0,49	50,3
V44	7	0,9	2,7	1,69	0,71	0,42	11,8
V45	8	0,2	8,2	4,84	3,14	0,65	38,7
V46	8	0,6	9,8	4,35	3,79	0,87	34,8
V47	5	0,0	9,0	3,90	3,88	1,00	19,5
V48	8	0,8	9,0	4,95	3,25	0,66	39,6
V49	7	0,5	9,4	4,67	3,34	0,72	32,7
V50	7	0,2	8,5	3,83	3,51	0,92	26,8
V51	6	2,3	10,8	6,98	3,70	0,53	41,9
V52	6	0,2	7,9	2,23	2,90	1,30	13,4
V53	7	0,2	9,1	7,04	3,14	0,45	49,3
V54	6	0,1	10,4	6,00	4,64	0,77	36,0

4.3 Modelo Linear Misto

4.3.1 Motivação

A escolha do modelo linear misto foi motivada pela Figura 4.2 que corresponde ao gráfico de perfis das 54 vacas estudadas. Analisando a Figura 4.2 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. Um comportamento padrão pode ser entendido como um comportamento que a maioria das vacas estaria seguindo, como por exemplo, iniciar a lactação num determinado patamar, atingir um pico depois de alguns estádios e descrescer até o fim da lactação ou já iniciar a lactação no pico e decrescer no decorrer dos estádios. O gráfico de perfis mostra exatamente o contrário, pois algumas vacas iniciam a lactação praticamente do zero, alcançando valores altos alguns estádios depois, enquanto que outras vacas iniciam sua lactação já no pico. Além disso, cada vaca possui uma velocidade no aumento da produção de leite bastante particular. Esse comportamento não-padronizado sugere que cada vaca apresenta uma lactação particular que deve ser influenciada por suas próprias condições. Evidente que esse não é o fator determinante, mas pode ser considerado um dos fatores determinísticos e, neste estudo, é o fator motivador da utilização do modelo linear misto.

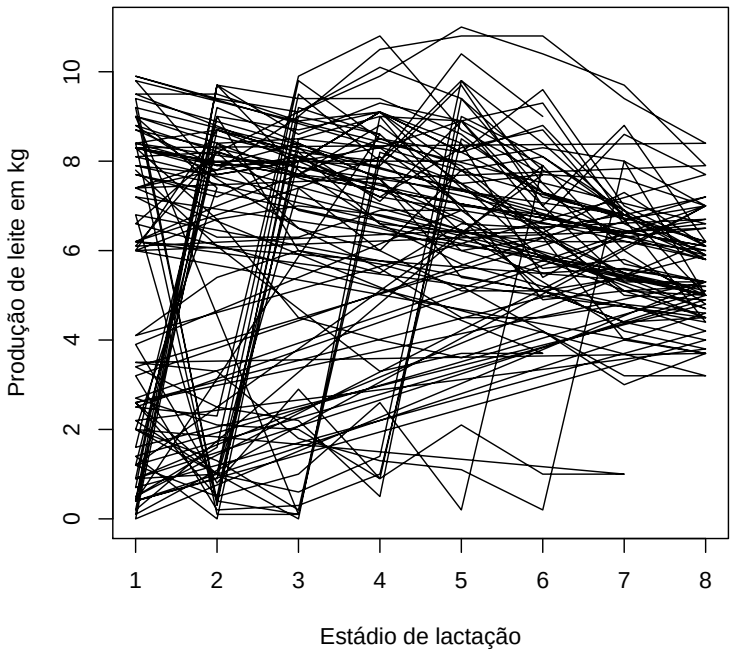

Figura 4.2: Gráfico de perfis das 54 vacas

Ainda analisando o comportamento da produção de leite da raça em estudo, tem-se na Figura 4.3 a curva de lactação média. Pode-se observar dois comportamentos distintos: O primeiro comportamento situa-se entre o primeiro e o quinto estádio e o segundo comportamento segue do quinto ao oitavo estádio. Verifica-se que a curva de lactação média inicia-se com produção bastante baixa no primeiro estádio, crescendo rapidamente até atingir um pico no quinto estádio e logo em seguida, descresce quase que linearmente até o oitavo estádio.

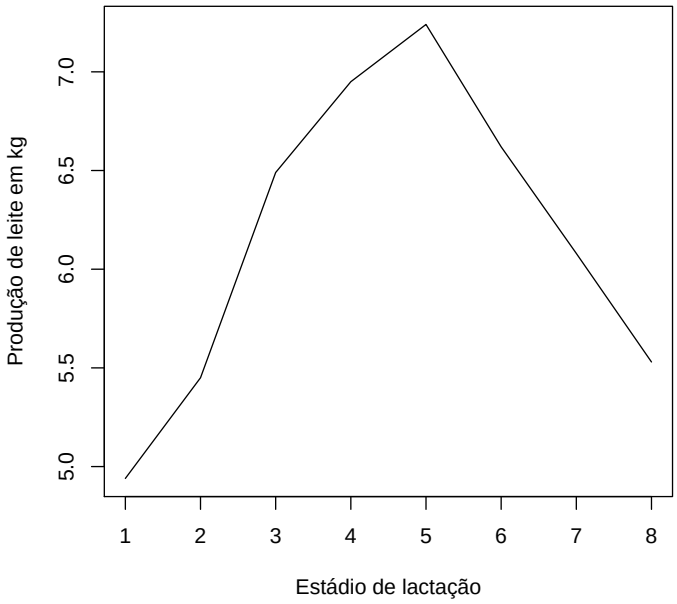

Figura 4.3: Curva de lactação média das 54 vacas

4.3.2 Definição do modelo

O interesse neste estudo é explicar a produção de leite a partir dos estádios da lactação, incorporando no modelo a variabilidade existente de cada vaca. Com isso temos um modelo linear misto com um intercepto e um coeficiente de inclinação, leia-se β_0 e β_1.

Avaliando o modelo a ser ajustado, surge uma questão: A variabilidade causada pela vaca está influenciando no intercepto ou na inclinação da reta? Ou tal varibilidade está influenciando nos dois parâmetros? Tais respostas só serão conhecidas com o ajuste dos modelos e testes. Entretanto, o gráfico de perfis pode indicar uma possível forma de influência. Sendo assim, observa-se, na Figura 4.2, que a curva de lactação inicia-se em vários pontos distintos no eixo da produção de leite. Fora isso, é notório a diferença de

inclinação e posição das curvas no gráfico de perfis. Essas duas observações sugerem o efeito da variabilidade da vaca tanto no intercepto quanto na inclinação. Portanto, tomando como referência o modelo com efeito aleatório no intercepto e na inclinação, optou-se por ajustar três modelos, nos quais a diferença reside na espicificação funcional dos efeitos aleatórios.

Seja o modelo definido em (3.2), dado por

$$Y = X\beta + Z\gamma + \varepsilon$$

Os três modelos a serem ajustados são definidos de forma matricial como se segue

Modelo intercepto aleatório

$$Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix};\ X_i = \begin{bmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix};\ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix};\ Z_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix};\ \gamma = \begin{bmatrix} \gamma_0 \end{bmatrix};\ \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}$$

Modelo inclinação aleatória

$$Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix};\ X_i = \begin{bmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix};\ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix};\ Z_i = \begin{bmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{in_i} \end{bmatrix};\ \gamma = \begin{bmatrix} \gamma_1 \end{bmatrix};\ \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}$$

Modelo intercepto e inclinação aleatórios

$$Y = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix};\ X = \begin{bmatrix} 1 & x_{i12} \\ 1 & x_{i22} \\ \vdots & \vdots \\ 1 & x_{in_i2} \end{bmatrix};\ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix};\ Z = \begin{bmatrix} 1 & z_{i12} \\ 1 & z_{i22} \\ \vdots & \vdots \\ 1 & z_{in_i2} \end{bmatrix};\ \gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix};\ \varepsilon = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}$$

Nos modelos apresentados, tem-se:

- Y_i: Produção de leite em kg da i-ésima vaca;
- X_i: Estádio de lactação da i-ésima vaca;
- Z: Matriz relacionada ao efeito aleatório no intercepto e/ou no estádio;

- β: Efeitos fixos;
- γ: Efeitos aleatórios;
- ε: Erro aleatório associado ao modelo.

Um segundo questionamento gira em torno de qual estrutura de covariância adotar. Para este trabalho foram consideradas duas estruturas 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).

4.3.3 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 estádio como modelo 2 e o modelo com efeito no intercepto e no estádio como sendo o modelo 3.

Da Tabela 4.3, que aporta 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 esse modelo deve ser selecionado. Ou seja, o modelo 3 é o modelo escolhido dentre os três modelos segundo os métodos AIC e BIC.

Continuando a análise dos resultados expostos na Tabela 4.3, 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, indiciando que os erros associados a essas estimativas são pequenos e assim sugerindo boas estimativas. Dentre as estimativas dos três modelos, as estimativas dos parâmetros do modelo 2 foram as que apresentaram intervalos com menor amplitude e, por conseguinte, o erro padrão da estimativa de menor magnitude.

Tabela 4.3: Estimativas $(\pm EP)$ e intervalos de confiança dos parâmetros dos efeitos fixos e intervalos de confiança para os efeitos aleatórios, adotando a matriz de cavariâncias 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	$-0,0070 \pm 0,0005$	$-0,0125 \pm 0,0011$	$-0,0130 \pm 0,0010$
$IC[\beta_0]$	$[7,2044; 7,8722]$	$[8,0103; 8,6593]$	$[7,8619; 8,7934]$
$IC[\beta_1]$	$[-0,0080; -0,0061]$	$[-0,0147; -0,0103]$	$[-0,0150; -0,0110]$
$IC[\sigma_0]$	$[0,4720; 1,0715]$	-	$[1,0447; 1,8704]$
$IC[\sigma_1]$	-	$[0,0033; 0,0069]$	$[0,0035; 0,0070]$
$IC[\sigma]$	-	-	$[1,6275; 1,9175]$
Correlação	–	–	$-0,931$
AIC	$1756,488$	$1710,846$	$1687,595$
BIC	$1772,474$	$1726,832$	$1711,574$
LogLik	$-874,244$	$-851,423$	$-837,798$

Observa-se, ainda na Tabela 4.3, que alguns intervalos de confiança das estimativas dos desvios relacionados aos efeitos aleatórios não existem. Isso se deve ao fato de que os modelos 1 e 2 possuem apenas um efeito aleatório ao contrário do modelo 3 que possui dois efeitos. Esse fato determina que a matriz de variâncias e covariâncias tem apenas um termo e obviamente não possuirá covariâncias. Comparando os intervalos de confiança relativos aos efeitos aleatórios do modelo 3 com os intervalos dos outros dois modelos, pode-se observar que os intervalos do modelo 3 são mais amplos que os dos demais modelos, apesar de que esse "mais amplo"é de uma magnitude não tão distante da magnitude dos demais modelos, sendo possível admitir que os três modelos apresentam intervalos que indiciam boas estimativas desses parâmetros.

No que se trata da estimativa da correlação entre os efeitos aleatórios no intercepto e no parâmetro de inclinação, levando em consideração a definição dos modelos 1 e 2 que possuem somente efeito no intercepto e somente efeito no estádio, obviamente, esses dois modelos não apresentam estimativas da correlação. O modelo 3 apresentou estimativa da correlação entre o efeito aleatório no intercepto e no estádio bastante elevada, sendo esta relação negativa (-93,1) sugerindo a existência de uma forte relação negativa entre as estimativas dos efeitos aleatórios.

Seguindo o mesmo padrão da Tabela 4.3, a Tabela 4.4 mostra os resultados dos ajustes adotando uma estrutura diagonal na matriz de variâncias e covariâncias, isto é, independência entre os efeitos aleatórios. Os ajustes dos modelos 1 e 2 não constam na tabela devido ao fato de que nestes dois modelos só há um efeito aleatório e, por isso, a matriz de variâncias e covariâncias apresenta apenas um termo, não existindo covariâncias. Essa particularidade faz com que a matriz de variâncias e covariâncias tanto no caso da estrutura generalizada como na estrutura diagonal sejam exatamente iguais. Dessa forma, apenas

o modelo 3, que apresenta a estrutura da matriz de variâncias e covariâncias diferente, é apresentado na tabela 4.4.

Da Tabela 4.4 é possível observar que as estimativas dos parâmetros relacionados aos efeitos fixos estão bem próximas 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

Tabela 4.4: Estimativas $(\pm EP)$ e intervalos de confiança dos parâmetros dos efeitos fixos e intervalos de confiança 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	-	-	$-847,0895$

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 melhor 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 estruturada da matriz Ψ, estão melhor definidos e passam a ser os modelos a serem estudados a seguir.

Dando continuidade a análise, uma vez determinada a estrutura de variância e covariância que se adequa melhor, segue-se para a determinação de qual é o modelo mais adequado, dentre aqueles que foram apresentados na Tabela 4.3. Para tanto, são apresentados na Tabela 4.5 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:

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

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 4.5, 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 4.2.

Tabela 4.5: 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

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 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.

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

4.3.4 Análise de resíduos e sensibilidade

Uma vez definido qual o melhor modelo, dentre todos aqueles que foram ajustados, segue-se para a análise de resíduos e sensibilidade. Nesta parte, são verificados os afastamentos das suposições e a sensibilidade do modelo quanto a produções e vacas influentes. 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 4.4 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.

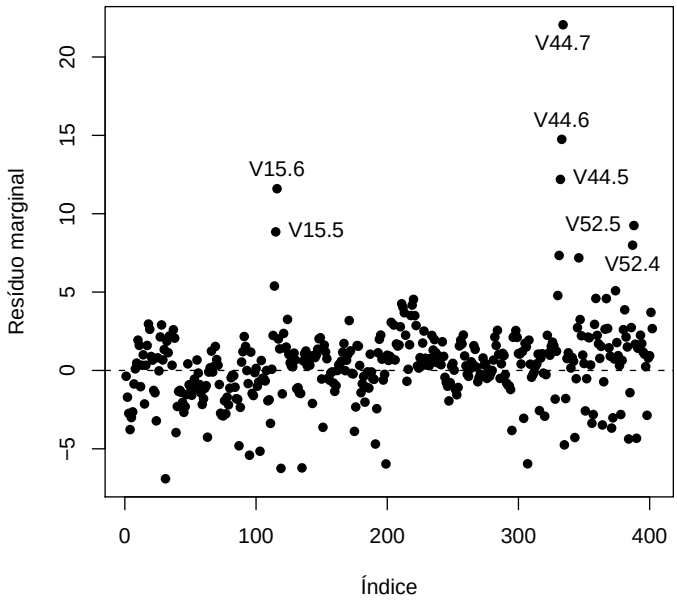

Figura 4.4: Gráfico de dispersão do resíduo marginal versus índice

Estudando os sete pontos que mais se afastam dos demais na figura 4.4, tem-se que as observações V15.5 e V15.6, que representam o quinto e o sexto estádio da vaca V15, são observações que se encontram "dividindo"o comportamento da vaca V15, pois a lactação desta vaca começa num patamar e decresce até o quinto e o sexto estádio, vindo a crescer no sétimo e oitavo estádios. Os pontos V44.5, V44.6 e V44.7 apresentam um comportamento semelhante com as duas observações da vaca V15, só que na direção inversa desta vaca, a observação V44.5 precede uma lactação inferior e antecede duas lactações

inferiores, o que faz da observação V44.5 um ponto incomum tornando os pontos V44.6 e V44.7 também discrepantes. As observações V52.4 e V52.5 apresentam comportamento diferente dos casos aqui apresentados. A observação V52.4 precede uma produção inferior e antecede uma produção inferior, enquanto que a observação V52.5 precede uma produção superior e antecede uma produção superior, indicando uma certa heterogeneidade na produção desta vaca.

O gráfico apresentado na Figura 4.5 expõem o resíduo condicional padronizado para as observações. Na Figura 4.5, pode-se observar 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.

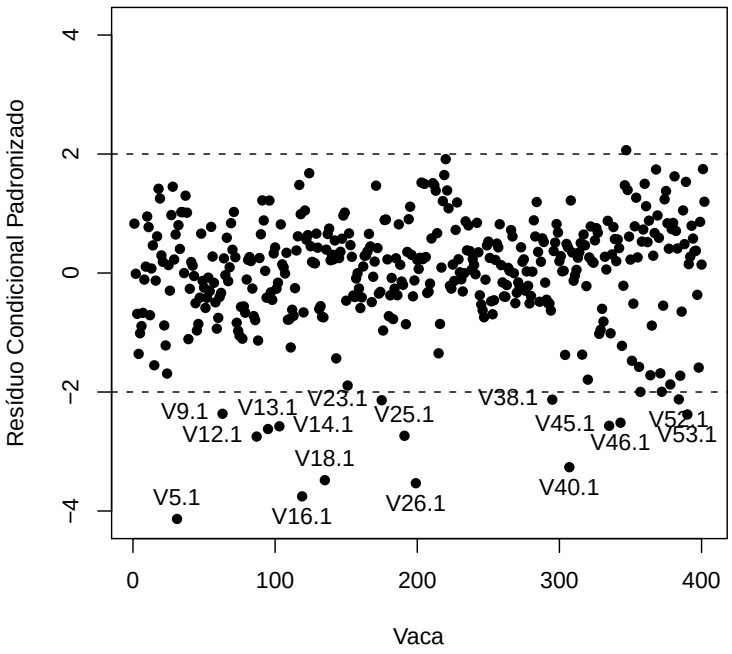

Figura 4.5: Resíduos condicionais padronizados

Quanto à normalidade dos erros, utilizando o resíduo com confundimento mínimo, verifica-se, na Figura 4.6, que não há fortes evidências que impossibilitem indicar o não afastamento da suposição de normalidade, sendo esta verificada. Ainda na Figura 4.6, observa-se uma pequena quebra no ponto zero fazendo 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 superior tendem para a linha central pontilhada

com alguns pontos se afastando. Contudo, mantém-se a indicação de não afastamento da suposição de normalidade.

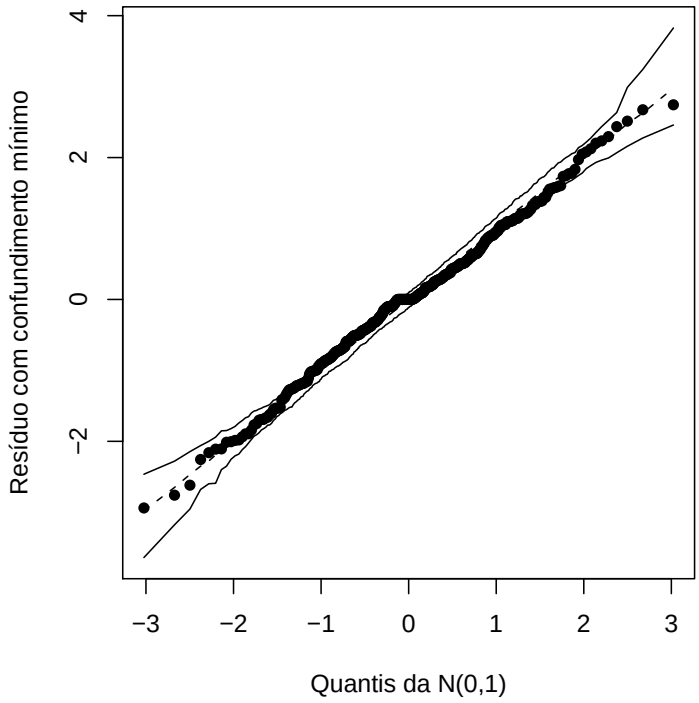

Figura 4.6: Gráfico de probabilidade normal com envelope para o resíduo com confundimento mínimo com grau de confiança de 95%

Analisando a distância de Mahalanobis para identificação de vacas discrepantes, na Figura 4.7, 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, existe 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.

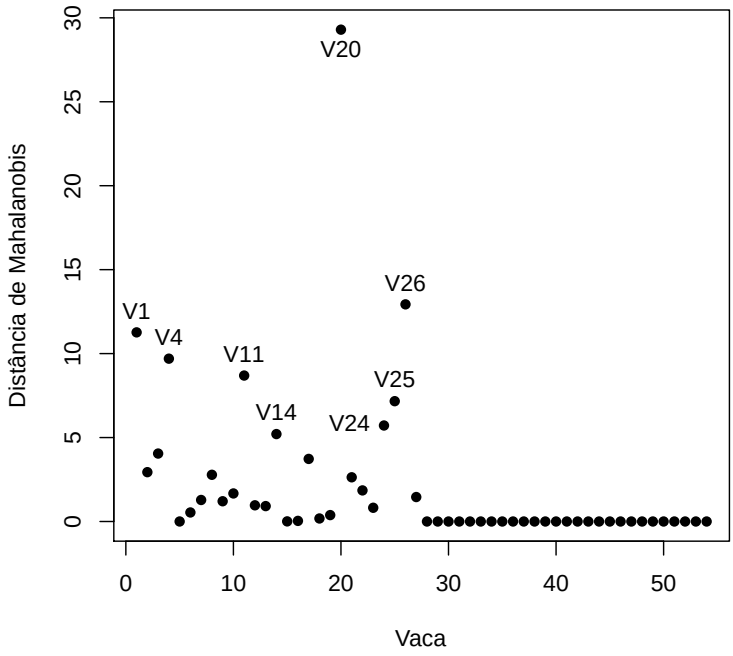

Figura 4.7: Distância de Mahalanobis

Observando o comportamento dessas vacas na Tabela 4.2, verificam-se médias de produção de leite distintas e em patamares diferentes, com exceção das vacas V17 e V24 que apresentam médias idênticas. A vaca V1 apresenta a menor média, enquanto que a vaca V5 apresenta a maior. As demais encontram-se dentro desse intervalo. As oito vacas apresentam valores máximos de produção próximos e valores mínimos variando de 0,1 a 5,1 kg. Analisando a variabilidade por meio do coeficiente de variação, observa-se que, com excessão das vacas V17 e V24, as demais vacas apresentam coeficiente de variação acima de 30%, indicando a existência de variabilidade na produção. Com relação ao total produzido durante os oito estádios da lactação, observa-se que a menor produção, dentre as produções das oito vacas possivelmente discrepantes, é da vaca V1 enquanto que a maior é da vaca V5. Entretanto, tais valores, como também as outras observações feitas sobre estas vacas, não fogem de um comportamento "padrão", uma vez que outras vacas possuem comportamentos semelhantes e não se apresentaram discrepantes.

Seguindo adiante, entrando na análise de sensibilidade propriamente dita, inicia-se a análise pela alavancagem generalizada cujos gráficos são apresentados na Figura 4.8.

Analisando a Figura 4.8(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 as observações com exceção das observações V15.5,

V15.6, V50.6 e V52.5, são relativas aos quatro primeiros estádios da lactação, sendo a maioria de segundo estádio. Outra observação importante é a de 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 estádio da vaca V50 que apresenta produção de 7,3 kg superior à média geral do sexto estádio (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 necessáriamente são classificadas como pontos de alavanca, uma vez que os valores de h_{ii} para essas observações apresentam baixos valores, se aproximando do zero.

Da Figura 4.8(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 sendo pontos de alavanca.

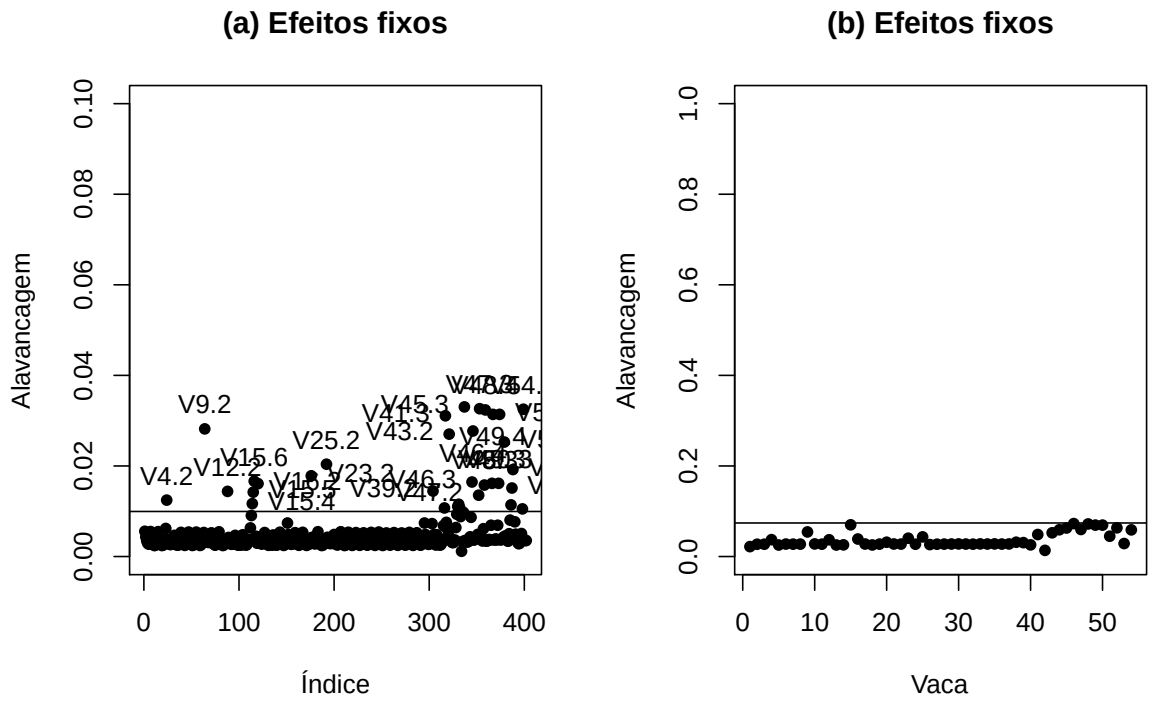

Figura 4.8: Alavancagem generalizada considerando somente os efeitos fixos

Quanto a alavancagem referente aos efeitos fixos e aleatórios, com relação à Figura 4.9(a), observa-se um comportamento semelhante ao da Figura 4.8(a), entretanto, a magnitude do afastamento da linha limite é bem maior do que na Figura 4.8(a). Dessa forma,

essas vinte e uma observações identificadas, são classificadas como pontos de alavanca. Tal indicação sugere que o efeito aleatório influência na estimação da produção de leite, mesmo trazendo pontos de alavanca.

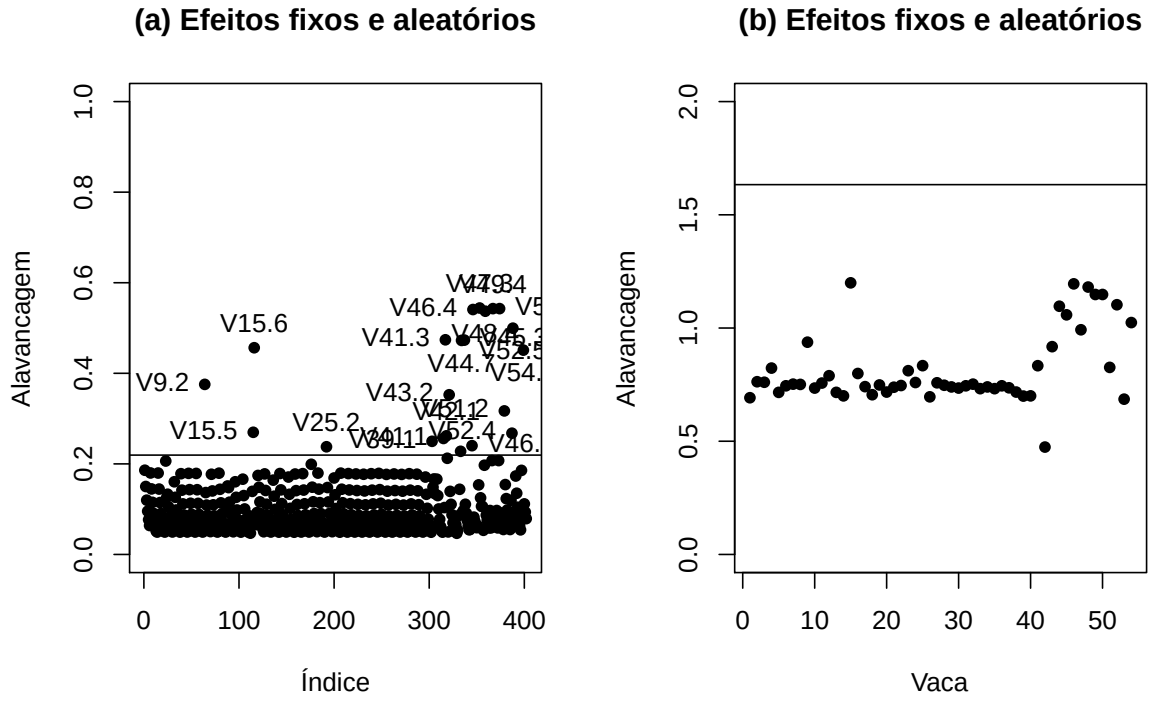

Figura 4.9: Alavancagem generalizada considerando efeitos fixos e aleatórios

Estudando a distância de Cook condicional na sua forma geral, Figura 4.10(a), pode-se verificar que grande parte das observações encontram-se em torno do zero com 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. Destas 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 4.10(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 4.10(a).

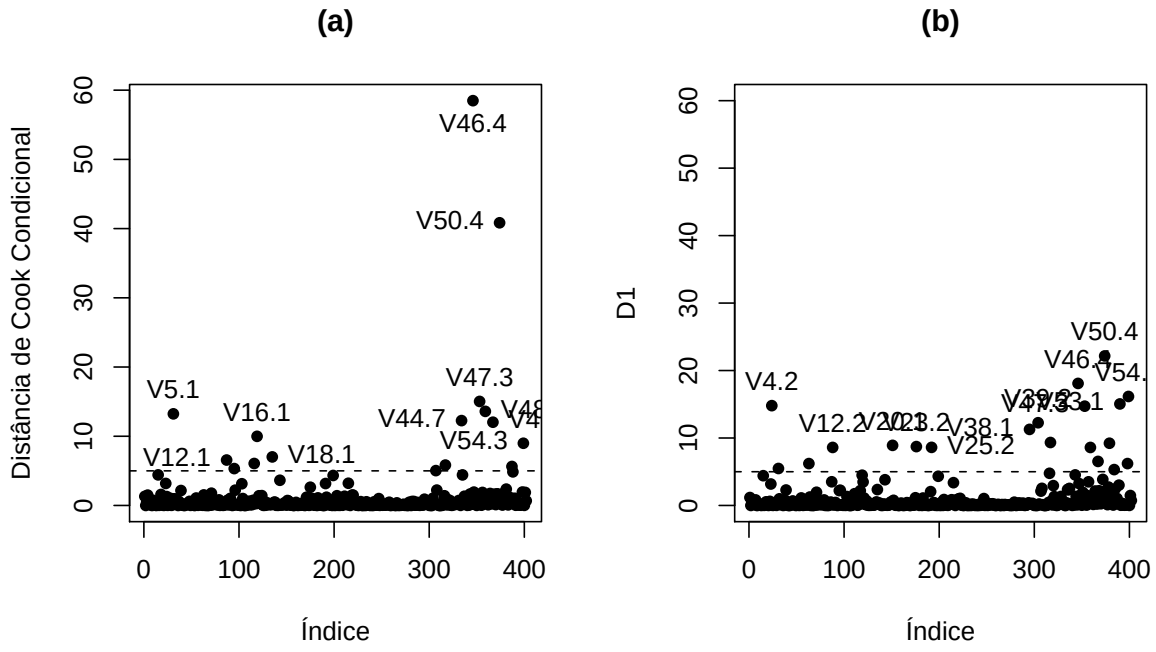

Figura 4.10: Distância de Cook condicional: Geral e primeiro termo da decomposição

Quanto à Figura 4.11(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 4.10(a) se destacam das demais. Dessa forma, existe a possibilidade dessas observações serem consideradas pontos influentes na estimativa dos efeitos aleatórios. Na Figura 4.11(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 4.11(b) é esperado, pois os pontos se concentram em torno do zero, o que indica a possibilidade de não haver problemas.

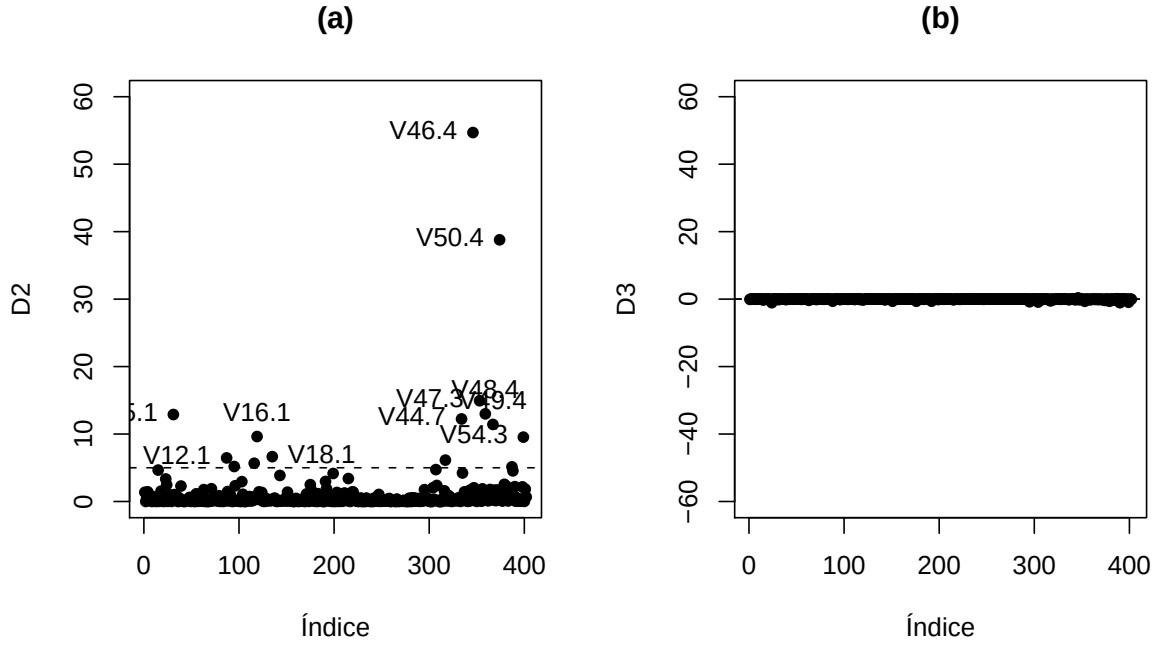

Figura 4.11: Distância de Cook condicional: Segundo e terceiro termos da decomposição

Analisando a variação causada pela retirada das observações identificadas nos gráficos da distância de Cook condicional, pode-se observar na Tabela 4.6, 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 o efeito aleatório referente à inclinação, além de uma influência significante na estimativa de σ.

Tabela 4.6: Comparação das estimativas dos parâmetros do modelo considerando a matriz generalizada ao se retirar as observações possivelmente influentes

Modelo	β_0	β_1	σ_0	σ_1	σ
Completo	8,328	-0,013	1,398	0,005	1,764
V3.1	8,357	-0,013	1,422	0,005	1,755
	(0,003)	(0,000)	(0,017)	(0,000)	(-0,005)
V4.2	8,32	-0,013	1,379	0,005	1,755
	(-0,001)	(0,000)	(-0,014)	(0,000)	(-0,005)
V5.1	8,391	-0,013	1,438	0,005	1,701
	(0,008)	(0,000)	(0,029)	(0,000)	(-0,036)
V9.1	8,325	-0,013	1,395	0,005	1,751
	(0,000)	(0,000)	(-0,002)	(0,000)	(-0,007)
V12.1	8,367	-0,013	1,421	0,005	1,741
	(0,005)	(0,000)	(0,016)	(0,000)	(-0,013)
V12.2	8,304	-0,015	1,394	0,005	1,767
	(-0,003)	(0,154)	(-0,003)	(0,000)	(0,002)
V16.1	8,382	-0,013	1,445	0,005	1,714
	(0,006)	(0,000)	(0,034)	(0,000)	(-0,028)
V16.2	8,32	-0,013	1,388	0,005	1,765
	(-0,001)	(0,000)	(-0,007)	(0,000)	(0,001)
V18.1	8,366	-0,013	1,408	0,005	1,726
	(0,005)	(0,000)	(0,007)	(0,000)	(-0,022)
V20.1	8,298	-0,013	1,399	0,005	1,763
	(-0,004)	(0,000)	(0,001)	(0,000)	(-0,001)
V23.2	8,286	-0,013	1,401	0,005	1,773
	(-0,005)	(0,000)	(0,002)	(0,000)	(0,005)
V25.2	8,286	-0,013	1,394	0,005	1,772
	(-0,005)	(0,000)	(-0,003)	(0,000)	(0,005)
V38.1	8,307	-0,013	1,388	0,005	1,753
	(-0,003)	(0,000)	(-0,007)	v0,000)	(-0,006)
V39.2	8,295	-0,013	1,388	0,005	1,765
	(-0,004)	(0,000)	(-0,007)	(0,000)	(0,001)
V41.2	8,323	-0,013	1,391	0,005	1,761
	(-0,001)	(0,000)	(-0,005)	(0,000)	(-0,002)
V41.3	8,382	-0,013	1,405	0,005	1,754
	(0,006)	(0,000)	(0,005)	(0,000)	(-0,006)
V44.7	8,340	-0,013	1,367	0,005	1,765
	(0,001)	(0,000)	(-0,022)	(0,000)	(0,001)
V46.4	8,337	-0,013	1,403	0,005	1,759
	(0,001)	(0,000)	(0,004)	(0,000)	(-0,003)
V47.3	8,328	-0,013	1,394	0,005	1,765
	(0,000)	(0,000)	(-0,003)	(0,000)	(0,001)
V48.4	8,232	-0,013	1,401	0,005	1,768
	(-0,012)	(0,000)	(0,002)	(0,000)	(0,002)
V49.4	8,331	-0,013	1,402	0,005	1,768
	(0,000)	(0,000)	(0,003)	(0,000)	(0,002)
V50.4	8,375	-0,013	1,385	0,005	1,753
	(0,006)	(0,000)	(-0,009)	(0,000)	(-0,006)
V51.2	8,370	-0,013	1,426	0,005	1,759
	(0,005)	(0,000)	(0,020)	(0,000)	(-0,003)
V52.1	8,368	-0,013	1,349	0,005	1,749
	(0,005)	(0,000)	(-0,035)	(0,000)	(-0,009)
V53.1	8,305	-0,013	1,39	0,005	1,751
	(-0,003)	(0,000)	(-0,006)	(0,000)	(-0,007)
V54.2	8,339	-0,013	1,380	0,005	1,751
	(0,001)	(0,000)	(-0,013)	(0,000)	(-0,007)
V54.3	8,376	-0,013	1,429	0,005	1,757
	(0,006)	(0,000)	(0,022)	(0,000)	(-0,004)
Todas	8,730	-0,015	1,400	0,006	1,496
	(0,048)	(0,154)	(0,001)	(0,200)	(-0,152)

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 indiciou observações influenciando em suas previsões. Entretanto, quando analisado a alavancagem considerando os efeitos fixos e aleatórios, observou-se a existência de observações influenciando 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.

4.3.5 Previsões

Após ajustar o modelo e estudar seus resíduos e sua sensibilidade, apresenta-se a seguir os gráficos dos 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.

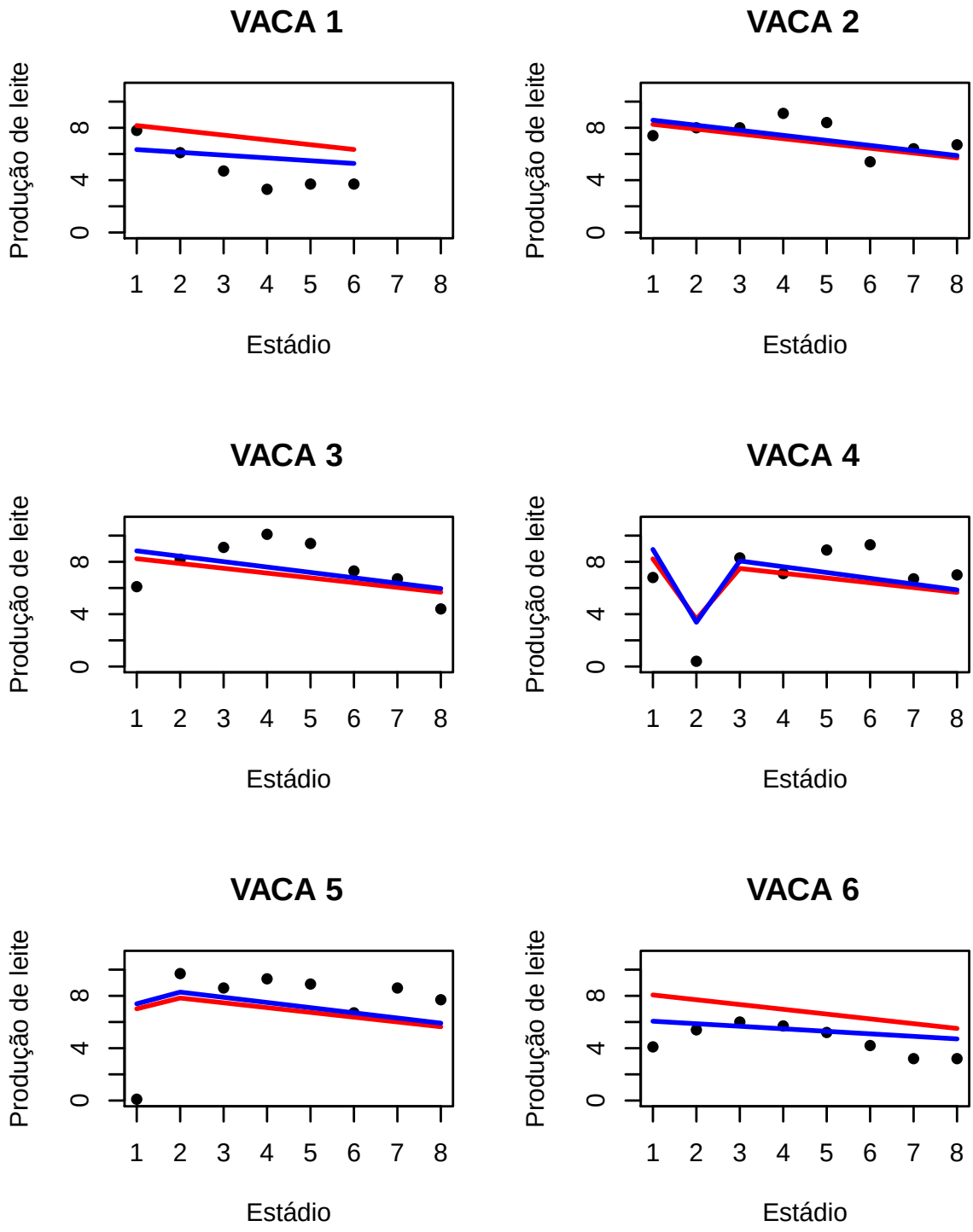

Figura 4.12: Previsão da produção de leite das 54 vacas

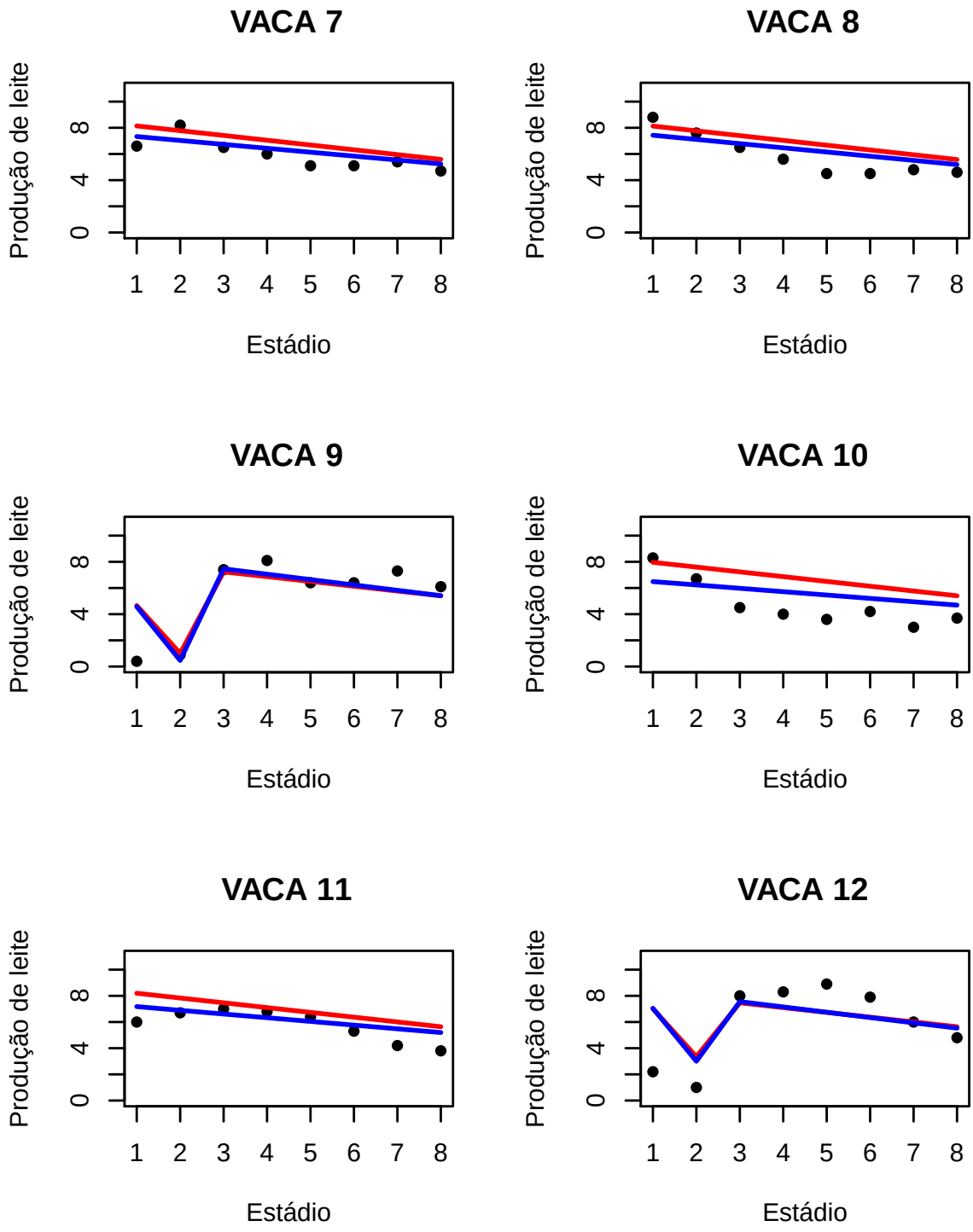

Figura 4.13: Previsão da produção de leite das 54 vacas (cont.)

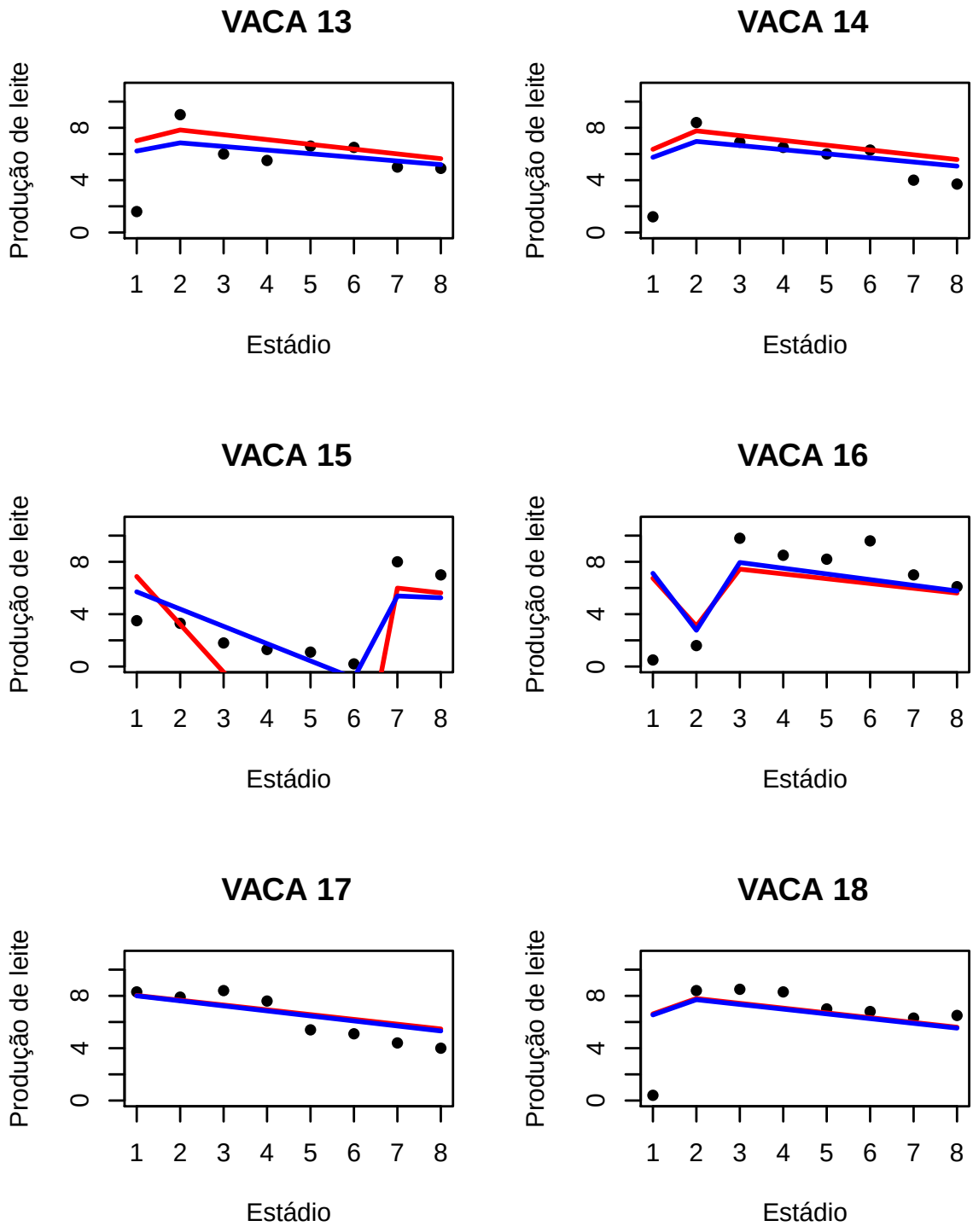

Figura 4.14: Previsão da produção de leite das 54 vacas (cont.)

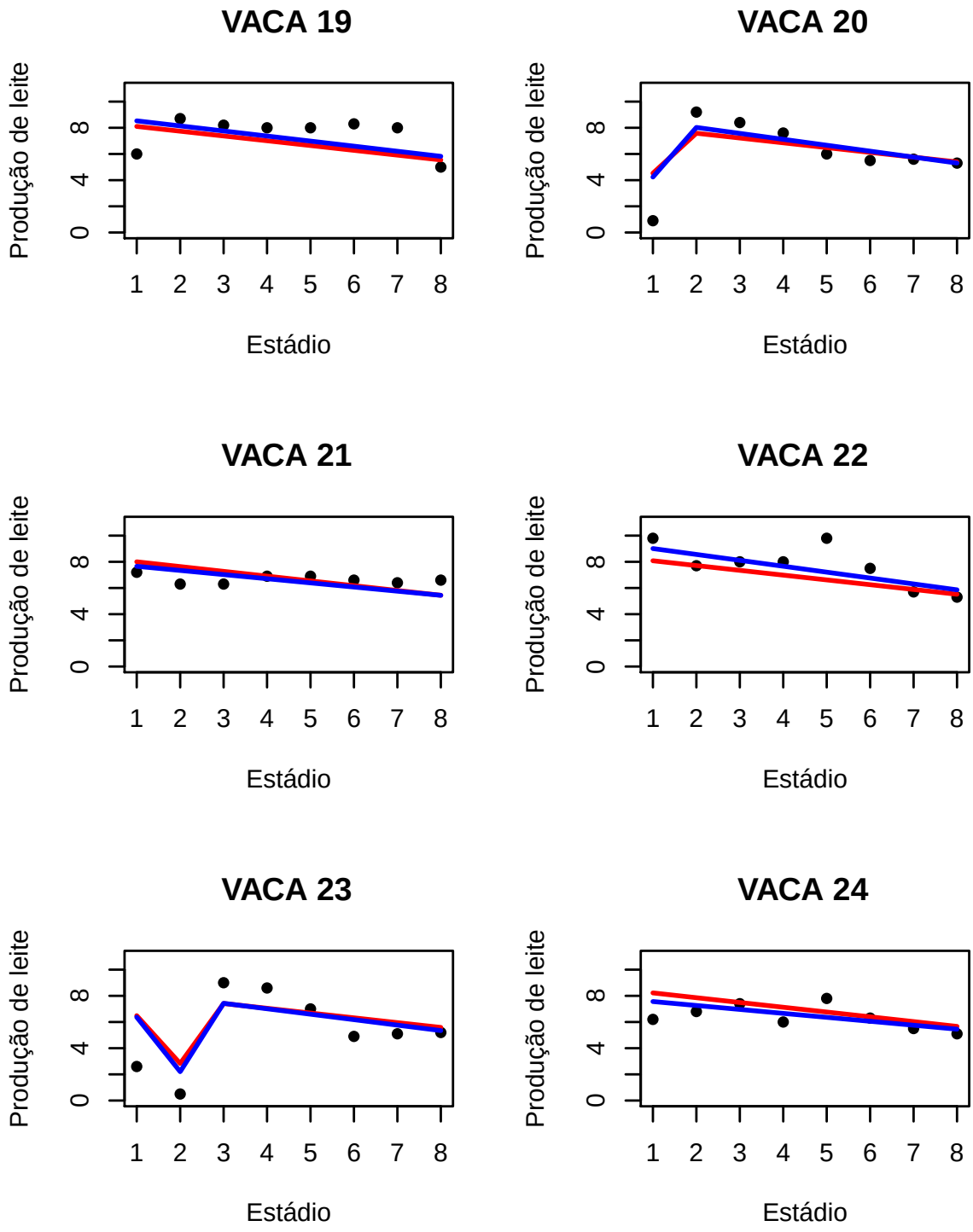

Figura 4.15: Previsão da produção de leite das 54 vacas (cont.)

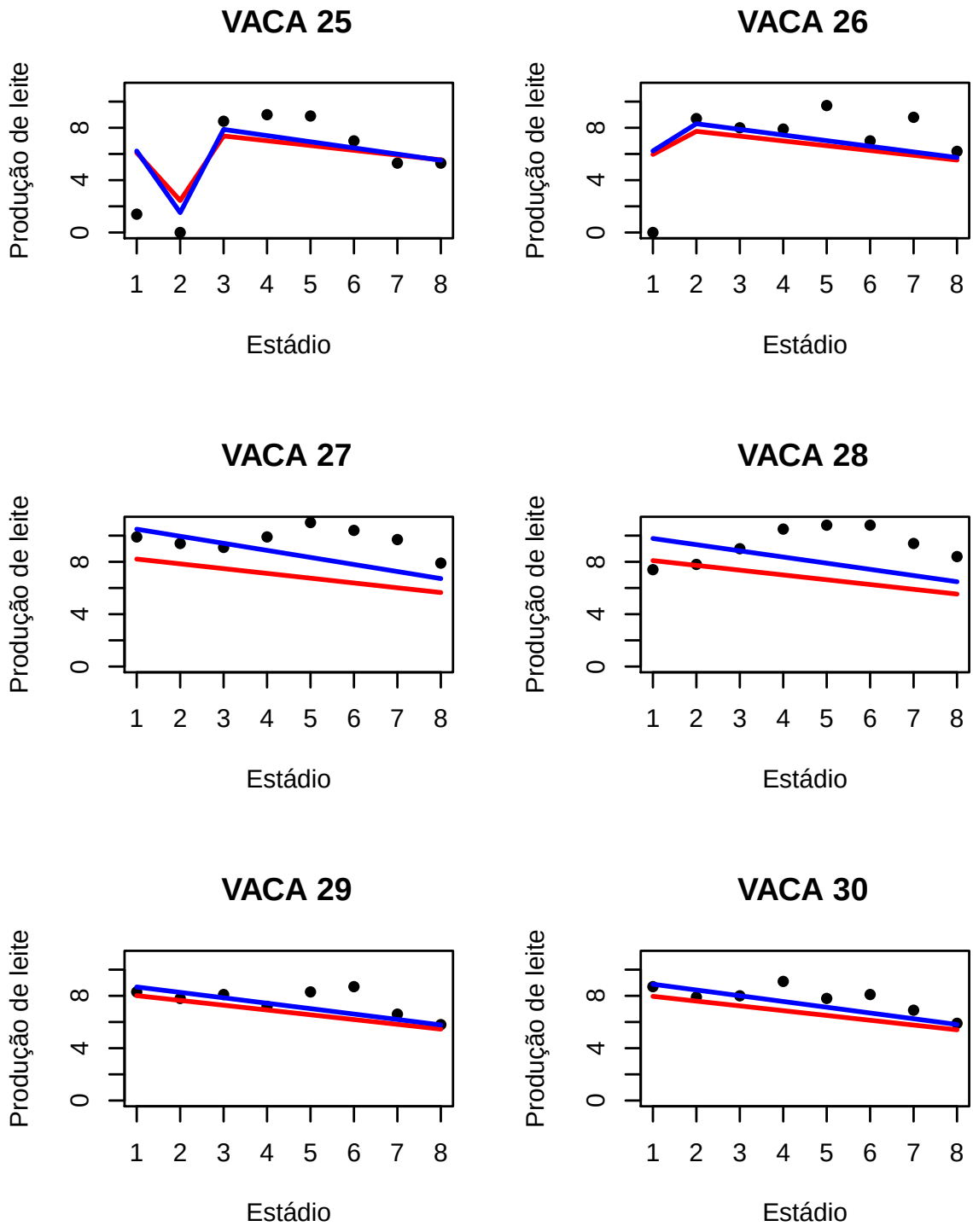

Figura 4.16: Previsão da produção de leite das 54 vacas (cont.)

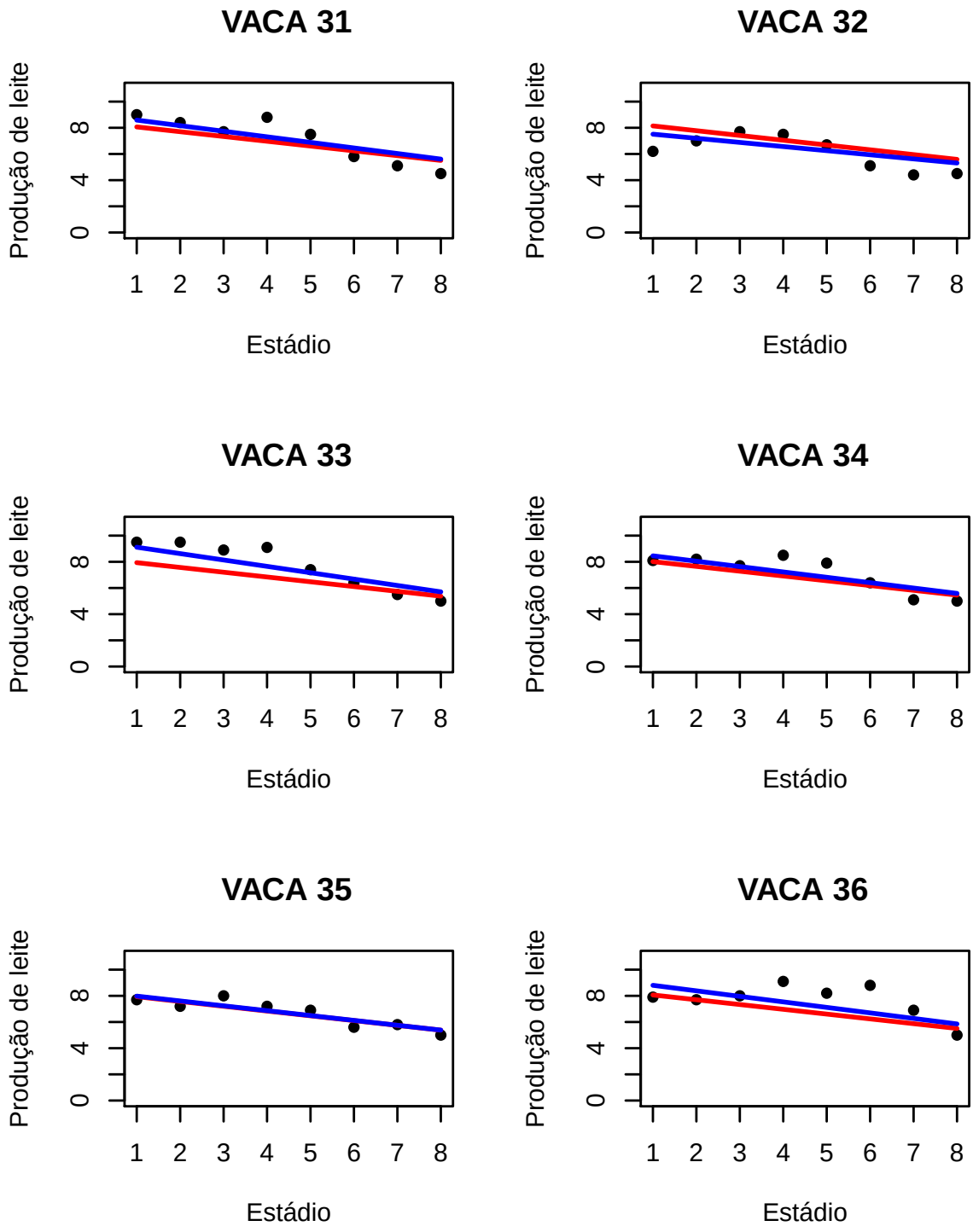

Figura 4.17: Previsão da produção de leite das 54 vacas (cont.)

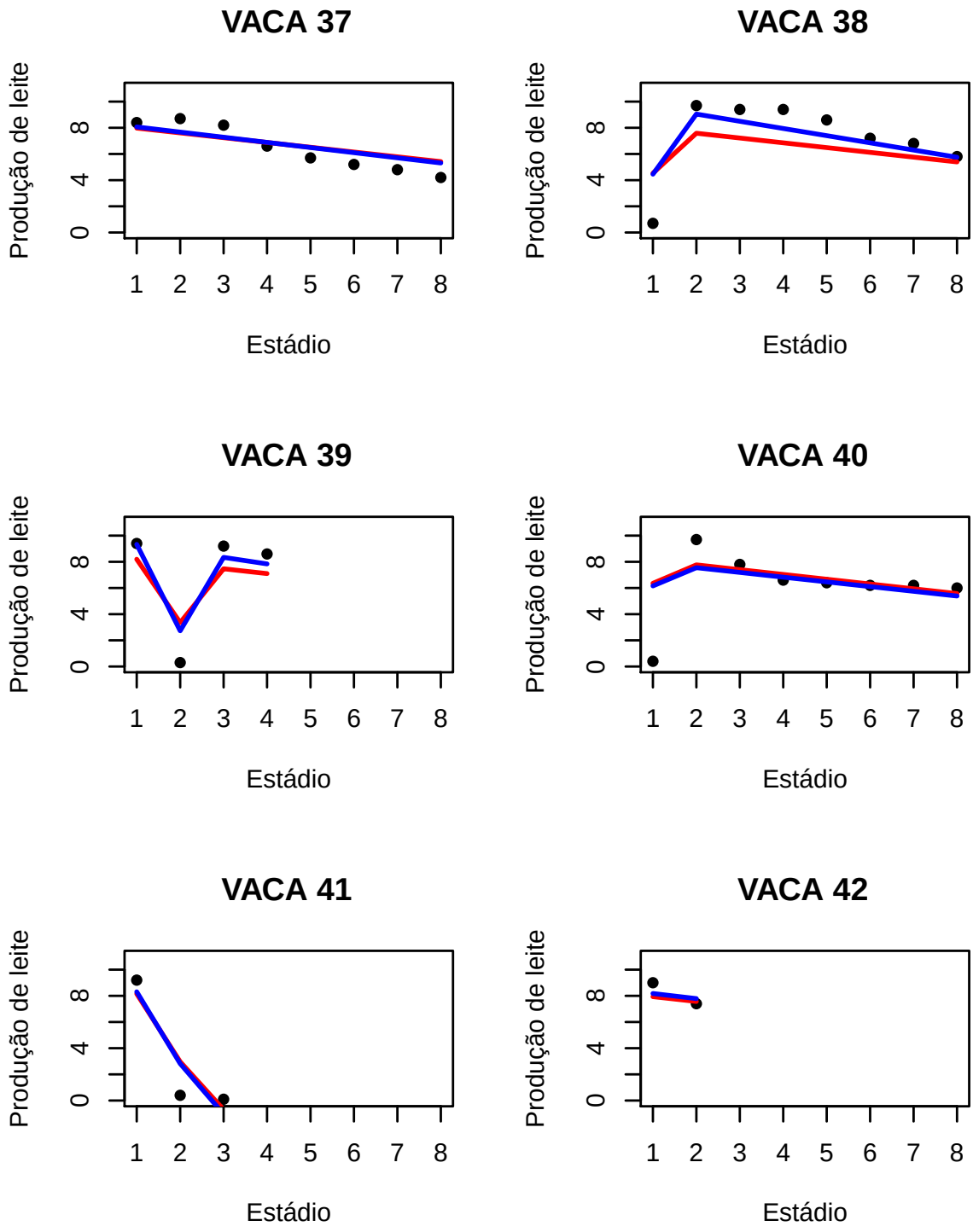

Figura 4.18: Previsão da produção de leite das 54 vacas (cont.)

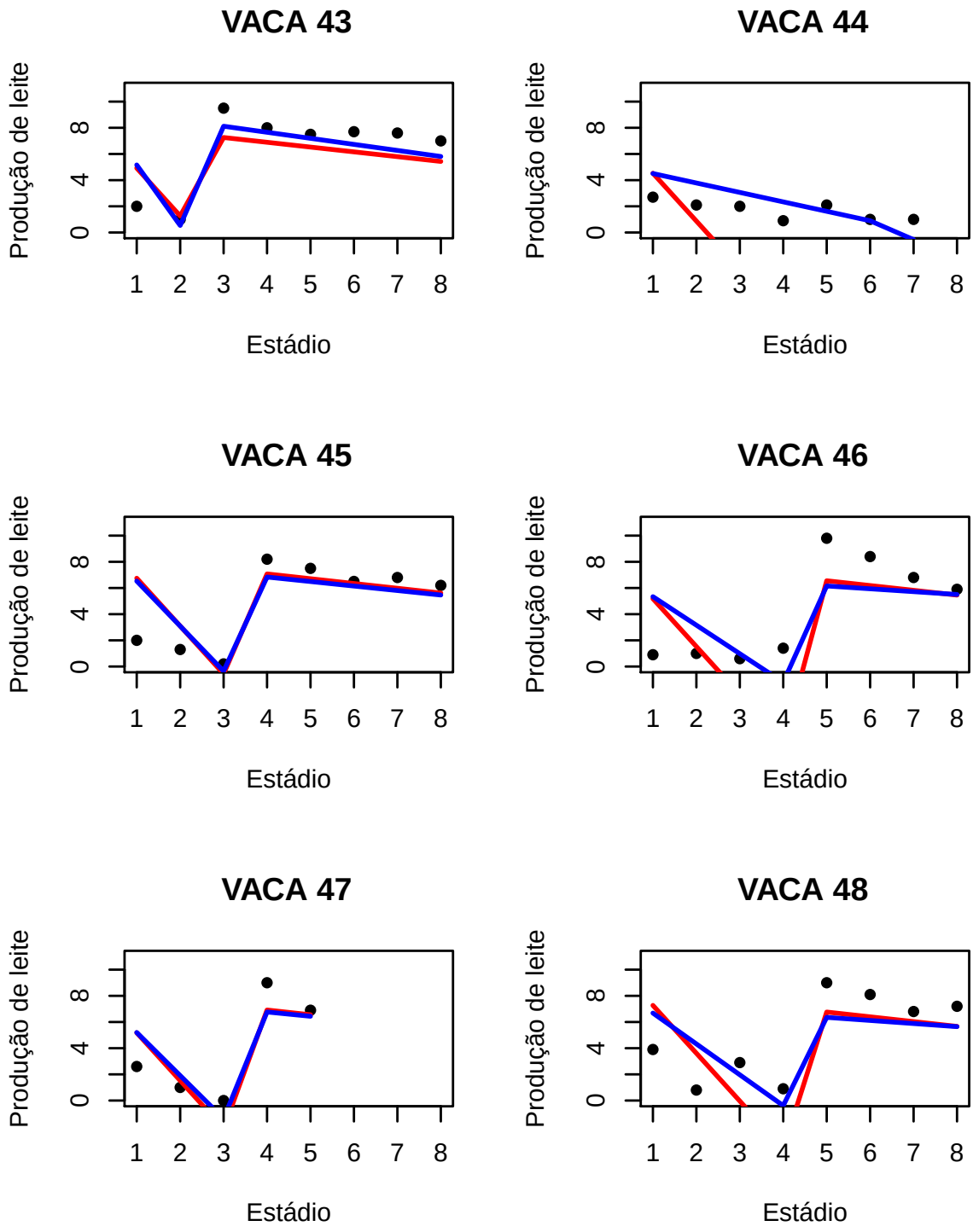

Figura 4.19: Previsão da produção de leite das 54 vacas (cont.)

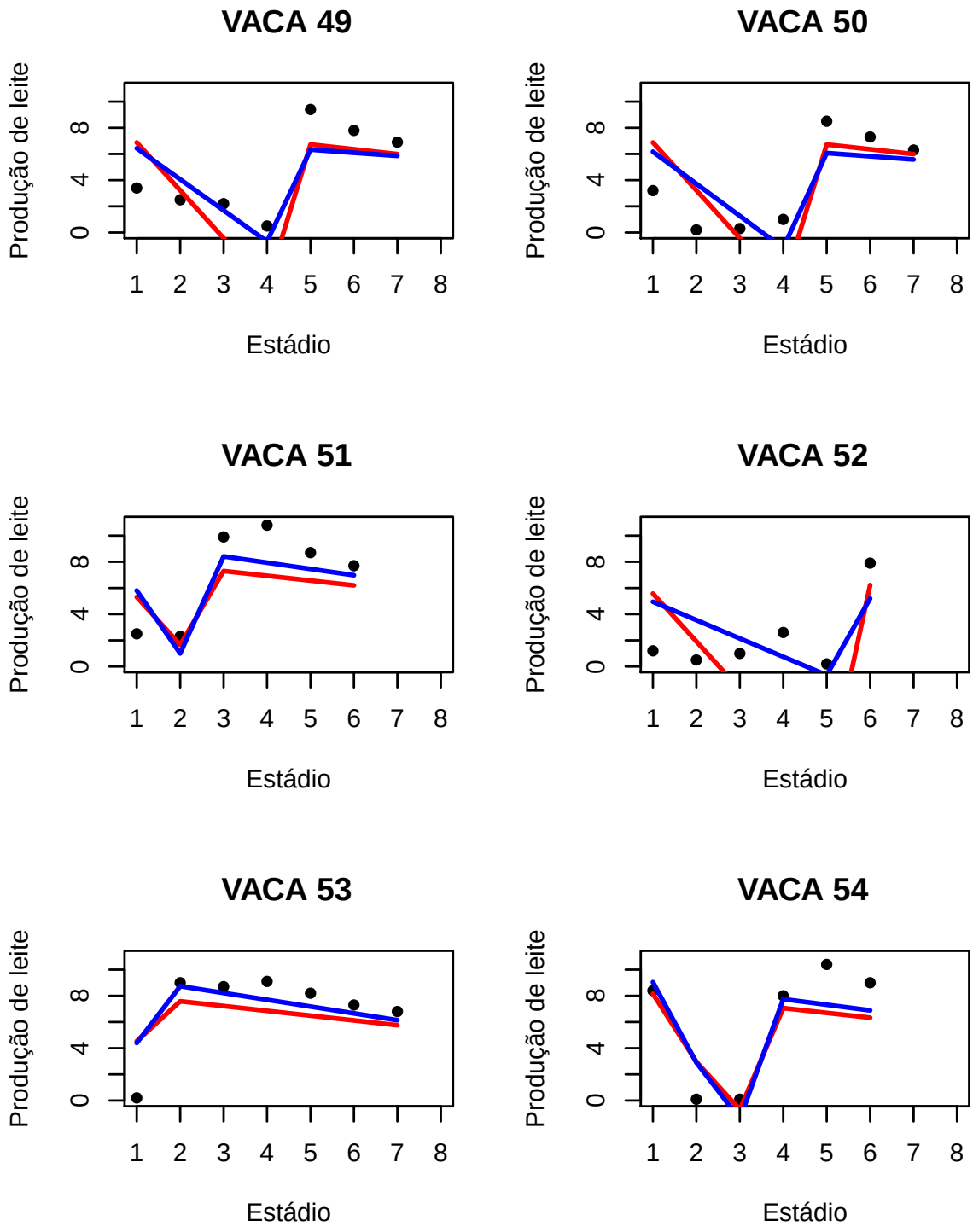

Figura 4.20: Previsão da produção de leite das 54 vacas (cont.)

5 Considerações Finais

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

As vacas estudadas apresentaram boas produções, mas também apresentaram altas variações. A existência de produções tendendo à zero mostrou-se ser uma das maiores causas de variação na produção. Vacas que apresentaram valores altos nos primeiros estádios, possuiram 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 estudado apenas os efeitos fixos, não se observou 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. Tal mudança indica que a inclusão do efeito aleatório pode estar aumentando o efeito das observações em suas próprias previsões.

Quanto à influência, os pontos indicados como possíveis pontos influentes apresenta-

ram 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 estádio. Todavia, de forma geral, as observações possívelmente influentes não causaram variações significativas nas estimativas dos efeitos fixos e da matriz de covariâncias.

Agregando ao ajuste à escolha do modelo, a análise de resíduos e sensibilidade, conclue-se que o modelo escolhido deve ser considerado 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 possíveis observações influentes, em sua maioria, foram observações que representavam produções próximas de zero.

As previsões das produções de leite considerando a presença dos efeitos aleatórios mostraram-se mais próximos da realidade quando comparado 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.

Com base nos resultados obtidos nesse trabalho, seguem algumas sugestões para trabalhos futuros.

1. Identificar e inserir uma variável sazonal;
2. Inserir novas variáveis explicativas na estrutura do modelo e considerando mais ordens de parto;
3. Ajustar um modelo linear generalizado misto com a distribuição de probabilidade da variável resposta positiva, considerando por exemplo, uma distribuição gama e propor a análise de resíduos e sensibilidade para este modelo.

Referências Bibliográficas

Ali, T. E.; Schaeffer, L. R. Accounting for covariances among test day milk in dairy cows. **Canadian Journal of Animal Science**, Vol. 67, No. 3, p. 637-639, 1987.

Banerjee, M.Cook's Distance in Linear Longitudinal Models. **Communications in Statistics, Theory and Methods**, Vol. 27, p. 2973-2983, 1998

Bianchini Sobrinho, E. Estudo da curva de lactação de vacas da raça Gir. Ribeirão Preto,SP: FMVRP/USP, 1984, 88p. Dissertação (Doutorado em Genética) Faculdade de Medicina Veterinária de Ribeirão Preto, Universidade de São Paulo, 1984.

Boldman, K. G.; Kriese, L. A.; Van Vleck, L. D. et al. A manual for use for MTDFREML. A set of programs to obtain estimates of variance and covariance [DRAFT]. Lincoln: Department of Agriculture / Agricultural Research Service, 1995. 120p.

Brody, S. A.; Ragsdale, A. C.; Turner, C. W. The rate of decline of milk secretion with the advance of period of lactation. **Journal of Genetic Physiology**, ; Vol. 5, 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. **Journal of Genetic Physiology**, Vol. 6, p. 541-545, 1924.

Cobuci, J. A.; Euclydes, R. F.; Verneque, R. S.; Teodoro, R. L.; Lopes, P. S.; Silva, M. A. Curva de Lactação na Raça Guzerá. **Revista Brasileira de Zootecnia**, Vol. 29, No. 5, p.1332-1329, 2000.

Cobuci, 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. **Revista Brasileira de Zootecnia**, Vol. 30, No. 4, p.1204-1211, 2001.

Cobuci, J. A.; Euclydes, R. F.; Pereira, C. S.; Ameida Torres, R.; Costa, C. N.; Lopes, P. S. Persistência na lactação - uma revisão. **Archivos Latinoamericanos de Produccion Animal**, Vol. 11, No. 3, p. 163-173, 2003.

Cook, R. D. Detection of Influential Observation in Linear Regression. **Technometrics**, Vol. 19, p. 15-18, 1977.

Cook, R. D.; Weisberg, S. Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression. **Technometrics**, Vol. 22, p. 495-508, 1980.

Cruz, G. R. B.; Ribeiro, M. N.; Filho, E. C. P.; Sarmento, J. L. R. Análise genética de bovinos Sindi utilizando-se as produções de leite e de gordura no dia do controle. **Revista Brasileira de Ciências Agrárias**, Vol. 3, No. 2, p. 179-185, 2008.

Cunha Filho, M. ; Ribeiro, M. N. ; Santos, E. S. ; Oliveira, J. C. V. Estudo da Curva de Lactação em Vacas da Raça Sindi no Estado da Paraíba. **Archivos de Zootecnia**, Espanha, v. III, p. 23-43, 2006.

Dave, B. K. First lactation curve of Indian water buffalo. **Jawaharlal Nehru Krishi Vishwa Vidyalaya Research Journal**, Vol. 5, p. 93, 1971.

Davis, C. S. Statistical methods for the analysis of repeated measurements. New York, Springer - Verlag, p. 442, 2002.

Fei, Y.; Pan, J. Influence Assessments for Longitudinal Data in Linear Mixed Models. In 18th International Woekshop on Statistical Modelling. Eds. G.Verbeke, G. Molenberghs, M. Aerts and S. Fieuws. Leuven: Belgium, p. 143-148.

Fraga LM, Gutiérrez M, Fernández L, Fundora O, González ME. Estudio preliminar de las curvas de lactancia en las búfalas mestizas de Murrah. **Revista Cubana de Ciencia Agrícola**, Vol. 37, p. 151-155, 2003.

Fung, W. K.; Zhu, Z. Y.; Wei, B. C.; He, X. Influence Diagnostics and Outliers test for Semiparametric Mixed Models. **Journal of the Royal Statistical Society B**, Vol. 64, p. 565-579, 2002.

Gonçalves, T. M.; Oliveira, A. I. G.; Freitas, R. T. F.; I. G. Pereira, I. G. Curvas de Lactação em Rebanhos da Raça Holandesa no Estado de Minas Gerais. Escolha do Modelo de Melhor Ajuste. **Revista brasileira de Zootecnia**. Vol. 31, No. 4, p. 1689-1694, 2002.

Grossman, M., Koops, W.J. Multiphasic analysis of lactation curves in dairy cattle. **Journal of Dairy Science**, Vol. 71, No. 6, p. 1598-1608, 1988.

Harville, D. A. Extension of The Gauss-Markov Theorem to Include the Estimation of Random Effects. **The Annals of Statistics**, Vol. 4, p. 384-395, 1976.

Hilden-Minton, J. A. Multilevel Diagnostics for Mixed and Hierarchical Linear Models. PhD Thesis. University of California, Los Angeles, 1995.

Joshi, N. R.; Phillips, R. W.. El ganado cebu de la India y del Pakistan. Roma: Organização das Nações Unidas para a Agricultura e Alimentação, (Publ., 19), 1954.

Laird, N. M.; Ware, J. H. Random-effects models for longitudinal data. **Biometrics**, 1982, 38, p.963-974.

Leite, P. R. M.; Santiago, A. A.; Navarro Filho, H. R.; Albuquerque, R. P. F.; Leite, R. M. Sindi: gado vermelho para o semi-árido. EMEPA-PB. Banco do Nordeste, João Pessoa, 2001, 174p.

Madalena, F. E., Martinez, M. L., Freitas, A.F. Lactation curves of Hostein-Friesian and Holstein-Friesian x Gir cows. **Animal Production**, Vol. 29, p. 101-107, 1979.

Madsen, O. A comparison of some suggested measures of persistency of milk yield in dairy cows. **Animal Production**, Vol. 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. Proceeding of the IV World buffalo congress. Practical Exp 2001; 2:354-420.

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 delactação em Búfalos Murrah e seus mestiços em sistema de criação semi extensivo, no Estado de São Paulo. **Archivos Latinoamericanos de Produccion Animal**, Vol. 13, No. 1, p. 19-23, 2005.

Nobre, J. S. Métodos de diagnostico para modelos lineares mistos. Dissertação de mestrado, IME/USP, São Paulo, 2004.

Nobre, J. S.; Singer, J. M. Residuals Analysis for linear Mixed Models. **Biometrical Journal**, Vol. 49, p. 863-875, 2007.

NOBRE, J. S.; Singer, J. M. Leverage analysis for linear mixed models. **Journal of Applied Statistics**, 2010 (Aceito para publicação).

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. **Arquivos Brasileiros de Medicina Veterinária e Zootecnia**. Vol. 59, No. 1, p. 233-238, 2007.

Papajcsik, I.A., Bodero, J. Modeling lactation curves of Friesian cow in a subtropical climate. **Animal Production**, Vol. 47, p. 201-207, 1988.

Patterson, H. D.; Thompson, R. Recovery of Interblock information when block sizes are unequal. **Biometrika**, Vol. 58, p. 545-554, 1971

Paula, G. A. Modelos de regressão com apoio computacional, São Paulo, IME/USP, http://www.ime.usp.br/ giapaula, 2004

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

Pool, M. H.; Janss, L. L. G.; Meuwissen, T. H. E. Genetic Parameters of Legendre Polynomials for First Parity Lactation Curves. **Journal of Dairy Science**, Vol. 83, 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. **Revista Colombiana de Ciencias Pecuarias**, Vol. 20, No. 2, 2007.

R Development Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2009, ISBN: 3-900051-07-0, http://www.R-project.org.

Rowlands, G. J.; Lucey, S.; Russell, A. M. A comparison of different models of the lactation curve in dairy cattle. **Animal Production**, Vol. 35, p. 135-142, 1982.

Singh RP, Gopal R. lactation curve analysis of buffalo es maintained under village conditions. **Indian Journal Animal Science**, Vol. 52, p. 1157-1163, 1982.

Tan, F. E. S.; Ouwens, M. J. N.; Berger, M. P. F. Detection of Influential Observations in Longitudinal Mixed Effects, Regression Models. **The Statistician**, Vol. 50, p. 271-284, 2001.

Togashi, K.; Lin, C. Y. Modifying the Lactation Curve to Improve Lactation Milk and Persistency. **Journal of Dairy Science**, Vol. 86, 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. **Journal of Dairy Science**, Vol. 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. **Journal of Dairy Science**, Vol. 87, p. 3789-3799, 2004.

Verbeke, G.; Molenberghs, G. Linear mixed models for longitudinal data. New York: Springer - Verlag, 2000, 568p.

Waternaux, C.; Laird, N. M.; Ware, J. H. Methods for Analysis of Longitudinal Data: Blood-lead Concentrations and Cognitive Development. **Journal of the American Statistical Association**, Vol. 84, p. 33-41. 1989.

Wei, B. C.; Hu, Y. Q.; Fung, W. K. Generalizad Leverage and its Applications. **Scandinavian Journal of Statistics**, Vol. 25, p. 25-37, 1998.

Wood, P. D. P. Algebraic model of the lactation curve in cattle. **Nature**, Vol. 216, p. 164-165, 1967.