



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

ANDERSON PAULO SCORSATO

VINÍCIUS MENARIN

**MODELOS LINEARES E NÃO-LINEARES APLICADOS À PRODUÇÃO DE LEITE DE
CASTRO/PR**

CURITIBA/PR

2013

ANDERSON PAULO SCORSATO

VINÍCIUS MENARIN

MODELOS LINEARES E NÃO-LINEARES APLICADOS À PRODUÇÃO DE LEITE DE
CASTRO/PR

Trabalho de Conclusão de Curso
apresentado à disciplina Laboratório de
Estatística, curso de bacharelado em
Estatística do Setor de Ciências Exatas da
Universidade Federal do Paraná.

Orientadora: Prof.^a Dr.^a Suely Ruiz Giolo

CURITIBA/PR

2013

AGRADECIMENTOS

Primeiramente a Deus, que iluminando nossos caminhos nos permitiu alcançar essa meta dando-nos força e sabedoria indispensáveis.

Agradecemos também aos nossos pais, familiares, amigos, namoradas e colegas de faculdade por acreditarem em nós, pelo apoio e compreensão nos momentos em que não pudemos estar presentes.

À nossa orientadora, Prof.^a Dr.^a Suely Ruiz Giolo, que aceitou nosso projeto, pelos valiosos ensinamentos dados e que sempre se mostrou disposta a nos ajudar nas adversidades encontradas no decorrer do trabalho.

Ao membro da banca, Prof. Dr. Cesar Augusto Taconeli, pela avaliação do trabalho e pelas contribuições dadas.

Aos professores do Departamento de Estatística da Universidade Federal do Paraná, pela contribuição em nossa formação profissional.

Aos funcionários do Departamento de Estatística da Universidade Federal do Paraná, pelo suporte dado no decorrer da nossa graduação.

À Cooperativa Castrolanda, pelo primeiro contato que viabilizou a realização deste trabalho.

À Associação Paranaense dos Criadores de Bovinos da Raça Holandesa, que disponibilizou o banco de dados e prestou suporte no entendimento do mesmo.

SUMÁRIO

1.	INTRODUÇÃO.....	12
2.	REVISÃO DE LITERATURA.....	13
2.1.	RELEVÂNCIA ECONÔMICA.....	13
2.2.	O MUNICÍPIO DE CASTRO.....	14
2.3.	COOPERATIVA CASTROLANDA.....	15
2.4.	ASSOCIAÇÃO DOS CRIADORES DE BOVINOS DA RAÇA HOLANDESA.....	15
2.5.	RAÇAS.....	16
2.5.1.	Holandesa.....	16
2.5.2.	Pardo suíço.....	17
2.5.3.	Jersey.....	18
2.6.	MODELOS PROPOSTOS NA LITERATURA.....	19
3.	MATERIAL E MÉTODOS.....	19
3.1.	MATERIAL.....	19
3.2.	MÉTODOS.....	21
3.2.1.	Modelos Lineares Generalizados (MLG) para Dados Longitudinais.....	21
3.2.2.	Modelos Não-Lineares.....	24
3.2.3.	Métodos Estatísticos para Escolha dos Modelos.....	25
3.2.4.	Seleção de Modelos Lineares.....	27
3.2.5.	Seleção de Modelos Não-Lineares.....	28
3.2.6.	Obtenção de Medidas de Interesse.....	29
4.	RESULTADOS E DISCUSSÃO.....	29
4.1.	ANÁLISE DESCRITIVA.....	29
4.1.1.	Raça Holandesa.....	33
4.1.2.	Raça Pardo Suíço.....	36
4.1.3.	Raça Jersey.....	38
4.1.4.	Multicolinearidade.....	41
4.2.	MODELOS LINEARES AJUSTADOS.....	41

4.2.1.	Raça Holandesa	41
4.2.2.	Raça Pardo Suíço	47
4.2.3.	Raça Jersey.....	52
4.3.	MODELOS NÃO-LINEARES AJUSTADOS	55
4.3.1.	Raça Holandesa	55
4.3.2.	Raça Pardo Suíço	59
4.3.3.	Raça Jersey.....	62
5.	CONSIDERAÇÕES FINAIS	65
6.	BIBLIOGRAFIA	71

LISTA DE TABELAS

Tabela 1 Modelos matemáticos utilizados no ajuste das curvas de lactação.	19
Tabela 2 Total de animais de cada raça utilizados no trabalho.	20
Tabela 3 Funções dos parâmetros tempo de pico e persistência.	29
Tabela 4 Estatísticas da produção diária (em kg) de acordo com a raça.	32
Tabela 5 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o número de ordenhas.	33
Tabela 6 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o grau de pureza do sangue.	34
Tabela 7 Classificação dos produtores da raça Holandesa quanto ao porte.	35
Tabela 8 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o tamanho do produtor.	35
Tabela 9 Estatísticas da produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o número de ordenhas.	37
Tabela 10 Estatísticas da produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o grau de pureza de sangue.	37
Tabela 11 Produção diária (em kg) dos animais da raça Jersey, de acordo com o número de ordenhas.	39
Tabela 12 Produção diária (em kg) dos animais da raça Jersey, de acordo com o grau de pureza de sangue.	39
Tabela 13 Correlação entre as covariáveis ordem de parto e idade no parto.	41
Tabela 14 Correlações lineares entre as curvas observadas e estimadas, para a raça Holandesa.	41
Tabela 15 Valores de QIC para a raça Holandesa.	42
Tabela 16 Modelos para a raça Holandesa com menores AIC	43
Tabela 17 Estimativas dos parâmetros do modelo linear ajustado para a raça Holandesa.	43
Tabela 18 Estimativas dos parâmetros do modelo linear ajustado para a raça Holandesa, com o grau de sangue agrupado.	44
Tabela 19 Interpretação das estimativas dos parâmetros das covariáveis do modelo linear para a raça Holandesa.	45
Tabela 20 Tempo e produção de leite no pico estimados pelo modelo linear ajustado para raça Holandesa.	47
Tabela 21 Correlações lineares entre as curvas observadas e estimadas, para a raça Pardo Suíço.	47
Tabela 22 Valores de QIC para a raça Pardo Suíço.	48
Tabela 23 Resultados de $QICu$ para a raça Pardo Suíço.	48

Tabela 24 Estimativas dos parâmetros do modelo linear ajustado para a raça Pardo Suíço.	49
Tabela 25 Tempo e produção de leite no pico estimados pelo modelo linear ajustado pela raça Pardo Suíço.	51
Tabela 26 Correlações lineares entre as curvas observadas e estimadas, para a raça Jersey.....	52
Tabela 27 Valores de <i>QIC</i> para a raça Jersey.....	53
Tabela 28 Estimativas dos parâmetros do modelo linear ajustado para a raça Jersey.....	53
Tabela 29 Interpretação da estimativa do parâmetro da idade no modelo linear para a raça Jersey.....	53
Tabela 30 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Holandesa.	55
Tabela 31 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Holandesa.....	56
Tabela 32 Avaliação do impacto da idade nas curvas estimadas de dois modelos não-lineares da raça Holandesa.	56
Tabela 33 Estimativas dos parâmetros do modelo não-linear da raça Holandesa.	57
Tabela 34 Interpretação das estimativas dos parâmetros das covariáveis do modelo não-linear para a raça Holandesa.	58
Tabela 35 Tempo e produção de leite no pico estimados pelo modelo não-linear ajustado para raça Holandesa.....	58
Tabela 36 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Pardo Suíço.....	59
Tabela 37 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Pardo Suíço.	59
Tabela 38 Avaliação do impacto da idade nas curvas estimadas de dois modelos não-lineares da raça Pardo Suíço.....	60
Tabela 39 Estimativas dos parâmetros do modelo não-linear da raça Pardo Suíço.....	60
Tabela 40 Interpretação da estimativa do parâmetro da idade no modelo não-linear para a raça Pardo Suíço.	61
Tabela 41 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Jersey.....	62
Tabela 42 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Jersey.	62
Tabela 43 Estimativas dos parâmetros do modelo não-linear da raça Jersey.....	63
Tabela 44 Interpretação da estimativa do parâmetro da idade no modelo não-linear para a raça Jersey.	64

Tabela 45 Tempo de pico estimado pelos modelos lineares e não-lineares para a raça Holandesa	67
Tabela 46 Produções no pico estimadas pelos modelos lineares e não-lineares para a raça Holandesa.	68
Tabela 47 Tempo e produção de pico estimados pelos modelos lineares e não-lineares para a raça Pardo Suíço.	69

LISTA DE FIGURAS

Figura 1 Produção de leite anual de Castro, comparada com a produção estadual, de 1974 a 2010.	14
Figura 2 Usina de Beneficiamento de Leite da Cooperativa Castrolanda, em Castro/PR.....	15
Figura 3 Animal da raça Holandesa – variedade HPB.	17
Figura 4 Animal da raça Holandesa – variedade HVB.	17
Figura 5 Animal da raça Pardo Suíço.	18
Figura 6 Animal da raça Jersey.	18
Figura 7 Produção diária média (em kg) das raças Holandesa, Jersey e Pardo Suíço ao longo do tempo.	30
Figura 8 Dispersão dos valores de produção diária (em kg) em torno da média de cada raça.	31
Figura 9 Distribuição da produção de leite (em kg) por raça.	31
Figura 10 Perfis individuais de produção diária (em kg) em torno da média de cada raça. ...	32
Figura 11 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o número de ordenhas.....	33
Figura 12 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o grau de pureza do sangue.	34
Figura 13 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o tamanho do produtor.....	35
Figura 14 Produção diária média (em kg) dos animais da raça Holandesa de acordo com a idade do animal no parto, com o número de animais em cada faixa.	36
Figura 15 Produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o número de ordenhas.....	36
Figura 16 Produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o grau de pureza de sangue.	37
Figura 17 Produção diária média (em kg) dos animais da raça Pardo Suíço de acordo com a idade do animal no parto, com o número de animais em cada faixa.	38
Figura 18 Produção diária de leite (em kg) dos animais da raça Jersey, de acordo com o número de ordenhas.....	39
Figura 19 Produção diária (em kg) dos animais da raça Jersey, de acordo com o grau de pureza de sangue.	40
Figura 20 Produção diária média (em kg) dos animais da raça Jersey de acordo com a idade do animal no parto, com o número de animais em cada faixa.....	40
Figura 21 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Holandesa.....	42

Figura 22 Modelo linear ajustado para a raça Holandesa, com as correlações entre as curvas observadas e estimadas.....	44
Figura 23 Análise dos resíduos <i>deviance</i> associados ao modelo linear ajustado para a raça Holandesa.	45
Figura 24 Modelo linear ajustado para a raça Holandesa com o grau de sangue agrupado, com as correlações entre as curvas observadas e estimadas.	46
Figura 25 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Pardo Suíço.	48
Figura 26 Resíduos brutos do modelo linear da raça Pardo Suíço versus o tempo.	49
Figura 27 Resíduos brutos do modelo linear da raça Pardo Suíço, para cada animal em cada tempo.	50
Figura 28 Modelo linear ajustado para a raça Pardo Suíço com as correlações entre as curvas observadas e estimadas.....	51
Figura 29 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Jersey.	52
Figura 30 Resíduos brutos do modelo linear da raça Jersey versus o tempo.	54
Figura 31 Resíduos brutos do modelo linear da raça Jersey, para cada animal em cada tempo.	54
Figura 32 Modelo linear ajustado para a raça Jersey, com a correlação entre a curva observada e estimada.....	55
Figura 33 Análise de resíduos do modelo não-linear para a raça Holandesa.....	57
Figura 34 Modelo não-linear ajustado para a raça Holandesa com o grau de sangue agrupado, com as correlações entre as curvas observadas e estimadas.....	58
Figura 35 Análise de resíduos do modelo não-linear para a raça Pardo Suíço.	60
Figura 36 Modelo não-linear ajustado para a raça Pardo Suíço, com as correlações entre as curvas observadas e estimadas.....	61
Figura 37 Curvas estimadas e observadas de modelos não-lineares rejeitados da raça Jersey.	63
Figura 38 Análise de resíduos do modelo não-linear para a raça Jersey.	64
Figura 39 Modelo não-linear ajustado para a raça Jersey, com a correlação entre a curva observada e estimada.....	65
Figura 40 Modelos lineares e não-lineares ajustados à raça Holandesa.....	67
Figura 41 Modelos lineares e não-lineares ajustados à raça Pardo Suíço.	68
Figura 42 Modelos lineares e não-lineares ajustados à raça Jersey.	69

RESUMO

A indústria de leite e de seus derivados desempenha um papel importante na economia. O estudo da produção leiteira de um rebanho, usualmente realizado por meio do ajuste de curvas de lactação, possibilita ao produtor um melhor entendimento do comportamento do rebanho, auxiliando-o na tomada de decisões a fim de aumentar a produtividade do rebanho e obter melhores resultados financeiros.

No âmbito nacional, o município de Castro/PR destaca-se por ser o maior produtor de leite. Dada essa importância, o objetivo deste trabalho foi estudar as curvas de lactação de três raças de bovinos de Castro (Holandesa, Pardo Suíço e Jersey) e quais fatores influenciam na produção de leite; além disso, foram obtidas medidas tais como: tempo e produção de pico e persistência. Para isso, foram utilizados modelos lineares e não-lineares, em uma abordagem de dados longitudinais. Os principais critérios de seleção de modelos utilizados, dependendo do tipo de modelo, foram: o *AIC*, o *QIC* e o *QICu*.

Dentre as covariáveis consideradas para cada raça, basicamente as mesmas mostraram-se significativas nos dois tipos de modelos considerados (lineares e não-lineares). Para expressar a correlação entre as observações provenientes de um mesmo animal foi selecionada a matriz permutável nos modelos lineares ajustados para as raças Pardo Suíço e Jersey, enquanto que para a raça Holandesa foi selecionada a matriz independente. Dos modelos não-lineares propostos na literatura, o modelo de Wood (1967) foi o que melhor se ajustou para as três raças.

A raça Holandesa é a raça predominante em Castro e, nesse trabalho, foi a que apresentou os melhores indicadores de produção de leite, principalmente para os animais puros que foram ordenhados três vezes por dia. Os modelos não-lineares também se apresentaram como a alternativa mais adequada para o estudo das curvas de lactação, tanto em relação ao ajuste dos modelos quanto em relação às estimativas do tempo e produção de pico.

Palavras-chave: dados longitudinais, curva de lactação, modelos lineares, modelos não-lineares, Castro.

1. INTRODUÇÃO

A indústria de alimentos sempre desempenhou um importante papel na economia brasileira, sendo que, dentre os diversos setores da indústria alimentícia o setor de laticínios destaca-se entre os quatro principais. Estima-se que a participação dos laticínios no faturamento total da indústria de alimentos seja de aproximadamente 10% (CARVALHO, 2010).

No cenário nacional de produção leiteira destaca-se o município de Castro/PR, que é considerado uma das principais bacias leiteiras do Brasil, tanto em produtividade quanto em qualidade genética. De acordo com dados do Instituto de Pesquisa Econômica Aplicada (IPEA), Castro ocupa a primeira posição no ranking nacional de produção de leite desde 1992.

O estudo da produção leiteira é, usualmente, realizado por meio de gráficos de curvas de lactação, que representam a produção de uma vaca no decorrer da lactação (OLIVEIRA, REIS, *et al.*, 2007). A análise dessas curvas possibilita um melhor entendimento acerca do sistema de produção, trazendo informações sobre a produção de leite, o que pode auxiliar o produtor na previsão da produção de suas vacas em determinado estágio de lactação, bem como na tomada de decisões quanto a descarte ou manejo das vacas (COBUCI, EUCLYDES, *et al.*, 2001).

O ajuste de curvas de lactação pode ser feito por meio de funções lineares, intrinsecamente lineares ou não-lineares. As lineares são as mais utilizadas devido à fácil interpretação dos parâmetros obtidos e à maior flexibilidade na estimação dos mesmos (FARO e ALBUQUERQUE, 2002). Muitas das funções propostas na literatura para o ajuste de curvas de lactação são modificações da função de Wood (WOOD, 1967). Dentre elas, a mais conhecida é a gama incompleta, seja na sua forma não-linear ou linearizada. No que diz respeito aos modelos não-lineares, algumas informações de interesse que podem ser obtidas a partir dos mesmos são: a produção inicial, o tempo de pico, a produção no pico e a persistência, definida como a constância na produção de leite durante a lactação (CKROH e COIMBRA, 2005).

Neste contexto, a utilização de técnicas e ferramentas estatísticas no estudo de curvas de lactação pode contribuir para um melhor entendimento de fatores associados à produção de leite. O objetivo deste trabalho foi analisar, a partir do banco de dados disponibilizado pela Associação dos Criadores de Bovinos da Raça Holandesa (APCBRH), o comportamento da produção leiteira de três raças do município de Castro/PR por meio de modelos estatísticos, no contexto de dados longitudinais (em que várias medições são feitas ao longo do tempo dentro da mesma unidade experimental). Foram ajustados modelos

lineares e não-lineares para explicar a curva média de lactação dos animais, investigando a influência de covariáveis ambientais na produção. Além disso, foram obtidas algumas medidas de interesse, como persistência e tempo de pico de lactação, com os parâmetros estimados por esses modelos.

2. REVISÃO DE LITERATURA

2.1. RELEVÂNCIA ECONÔMICA

A produção de leite e seus derivados desempenha um papel importante no cenário econômico nacional e internacional. A produção brasileira vem crescendo nos últimos anos sendo responsável por mudanças tanto no aspecto econômico quanto social. No primeiro quesito é um peso positivo na balança comercial e na geração de renda. Já no aspecto social, auxilia na geração de empregos, manutenção da população do meio rural ou melhoria da qualidade de vida em locais distantes dos aglomerados urbanos (CAPUCHO, 2010) (NERO, VIÇOSA e PEREIRA, 2009).

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil produziu no ano de 2010, 30,7 bilhões de litros de leite. Foi o 5º maior produtor mundial naquele ano, segundo avaliação do USDA. Contudo, a grande maioria da produção provém de animais de baixa produtividade. Tais valores de produção têm aumentado de maneira modesta desde 2001 (MEZZADRI, 2012). Uma das suposições que busca justificar este baixo desempenho na produtividade é que 90% dos sistemas de produção nacional são extensivos. Assim sendo, problemas nutricionais e de manejo são os maiores responsáveis pelo baixo desempenho de produção (MEZZADRI, 2012).

O Estado do Paraná tem papel de destaque na produção leiteira. Segundo levantamento do IBGE, dos 20 primeiros municípios produtores do nosso país, quatro são paranaenses, sendo o primeiro lugar pertencente a Castro com 180 milhões de litros produzidos em 2010 (MEZZADRI, 2012).

No ano de 2010, o Estado do Paraná produziu 3.595.775.000 litros de leite, participando com 11,7% da produção brasileira. A produtividade média das vacas é superior à média nacional, atingindo 10,9 litros. Entretanto existe uma grande diferenciação conforme o porte e nível tecnológico dos produtores, variando de 7,1 litros/vaca/dia a 18,5 litros/vaca/dia, não sendo incomum encontrar rebanhos nas regiões mais tecnificadas onde os animais ultrapassam 30 litros/vaca/dia. As principais bacias leiteiras do Estado

encontram-se no Centro Oriental (Campos Gerais), Oeste e Sudoeste, as quais concentram 48,5% dos produtores e 53,0% da produção estadual de leite (MEZZADRI, 2012).

A Figura 1 mostra a importância de Castro no cenário estadual de produção leiteira. Observa-se que a produção de leite é crescente ao longo do tempo (linha cheia). A linha tracejada indica o quanto Castro representa na produção do estado do Paraná, sendo que nos últimos anos esse valor ficou em torno de 5%.

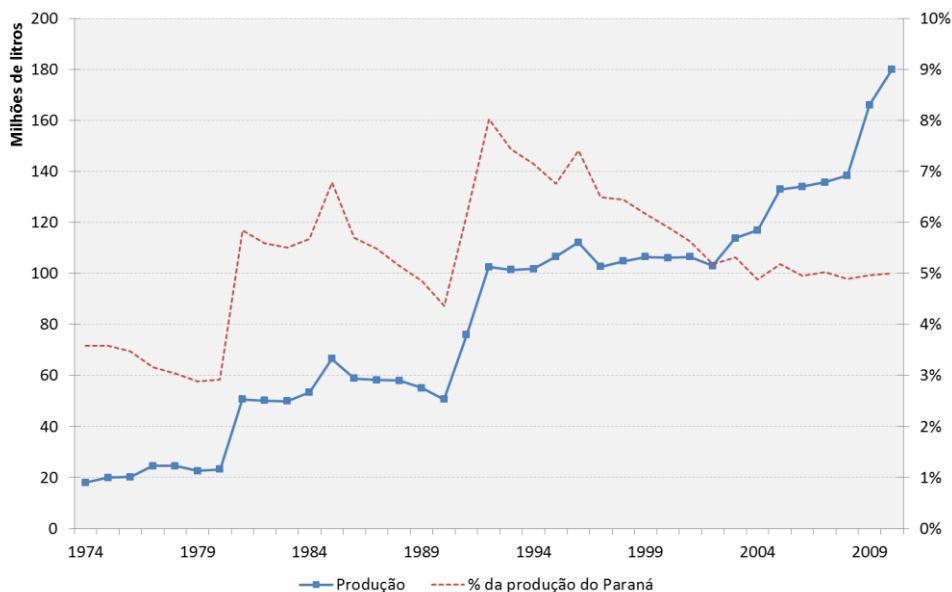


Figura 1 Produção de leite anual de Castro, comparada com a produção estadual, de 1974 a 2010.

2.2. O MUNICÍPIO DE CASTRO

O município de Castro está localizado na região Centro-Sul do Estado do Paraná (região dos Campos Gerais), tendo uma área de 2.532 km² e os biomas Cerrado e Mata Atlântica. De acordo com o censo de 2010 do IBGE, a população de Castro é de 67.084 habitantes, sendo 73,4% considerados população urbana e 26,6% população rural.

O município tem como base de sua economia a atividade agropecuária, sendo considerado um dos maiores produtores do Estado do Paraná. Na agricultura destacam-se como principais produtos o milho, a soja, o feijão, a batata, o trigo e a aveia. Também merecem destaque a pecuária de corte, a criação de gado, de suínos e de aves. Em relação à pecuária leiteira, o município é considerado uma das principais bacias leiteiras do Brasil, tanto em produtividade quanto qualidade genética. Castro também recebeu o título de maior produtor de calcário agrícola da América Latina, devido à exploração e extração mineral. Seu comércio é amplo, reunindo prestadoras de serviços, indústrias de móveis, pincéis e

alimentos, além de contar com um Parque Industrial para implantação de novas indústrias, em franco desenvolvimento (CASTRO).

2.3. COOPERATIVA CASTROLANDA

A Colônia Castrolanda está localizada a 6 km do centro da cidade de Castro. Ela foi fundada pelos primeiros imigrantes vindos da Holanda entre 1951 e 1954 que trouxeram consigo equipamentos agrícolas e gado, fundando assim a Cooperativa Castrolanda (CASTRO). Atualmente a Cooperativa está presente em vários municípios, sendo eles: Castro (PR), Ponta Grossa (PR), Piraí do Sul (PR), Curiúva (PR), Ventania (PR) e Itaberá (SP).

A atividade leiteira foi a primeira exploração agropecuária da Cooperativa Castrolanda. Com as famílias holandesas vieram animais com genética diferenciada comparada ao plantel brasileiro da época. Esses imigrantes holandeses e produtores, com foco na produtividade e na qualidade da produção, posicionaram o município de Castro, no primeiro lugar na produção nacional de leite (CASTROLANDA).

A Cooperativa conta hoje com a Usina de Beneficiamento de Leite (Figura 2), localizada no Parque Industrial de Castro, de onde saem produtos como leite do tipo longa vida e creme de leite, entre outros. A capacidade atual máxima da usina é de um milhão de litros de leite por dia (CASTROLANDA).



Figura 2 Usina de Beneficiamento de Leite da Cooperativa Castrolanda, em Castro/PR.

2.4. ASSOCIAÇÃO DOS CRIADORES DE BOVINOS DA RAÇA HOLANDESA

A Associação Paranaense de Criadores de Bovinos (APCB), fundada em 27 de março de 1953, considerada de utilidade pública em julho de 1955, inicialmente reunia em

seu quadro associativo criadores de bovinos de diferentes raças. Todavia, com o passar do tempo, os criadores passaram a buscar identidade própria e promoção de suas raças. Assim nasceram várias outras Associações que se desvincularam da APCB, como a Associação Paranaense dos Criadores de Bovinos da Raça Holandesa (APCBRH). A Entidade acompanhou a evolução da tecnologia e informatizou todos os seus serviços, como por exemplo: Controle Leiteiro, Programa de Classificação Linear para Tipo, Programa de Atualização Automática de Provas de Touros, e Programa para Cadastro e Envio de Coberturas. Com a informatização, as informações chegam até a entidade com mais facilidade e o retorno para o criador também ficou mais rápido. Os serviços da entidade ficaram mais atualizados e aprimorados. O objetivo principal é otimizar o atendimento ao criador (APCBRH).

Apesar de o foco da APCBRH ser a raça Holandesa foram encontrados animais das raças Jersey e Pardo Suíço no banco de dados fornecido.

2.5. RAÇAS

2.5.1. Holandesa

Raça de origem europeia com habitat natural localizado na região Frísia (norte da Holanda). Os maiores rebanhos brasileiros encontram-se nas regiões sul e sudeste. A raça holandesa apresenta três variedades: Frísia, Crominga e, por último, Mosa, Reno e Yessel (NEIVA, 2000).

A variedade Frísia, como o nome sugere, é originária da província de Frísia e é conhecida como holandês preto e branco (HPB) ou simplesmente holandês. Tal variedade é constituída de animais de grande porte. As vacas desta variedade apresentam, em geral, um peso de 500 a 700 kg e os touros 900 kg a 1.000 kg. Os bezerros nascem com peso médio entre 38 e 40 kg. A estatura dos animais é de 1,30 a 1,40 m e de 1,40 a 1,42 m para vacas e touros, respectivamente. A pelagem destes animais é malhada, com malhas brancas e pretas distribuídas pelo corpo desuniforme e desordenadamente, variando de animal para animal (Figura 3). As vacas apresentam úbere bem inserido, dotado de boa vascularização, quartos simétricos entre si e bem desenvolvidos, tetos de tamanho médio e bem espaçados. A HPB tem uma aptidão específica para a produção leiteira e apresenta uma média que pode variar de 4000 a 7.000 kg de leite por lactação, com 305 dias de persistência. No entanto, animais superiores podem com facilidade ultrapassar esta média (NEIVA, 2000).

A variedade de Mosa, Reno e Yessel (MRY) é originária dos vales que dão nome à variedade, localizados na região oriental da Holanda, também conhecida com holandesa vermelha e branca (HVB). No Brasil, os principais rebanhos estão também localizados na região Sul e Sudeste. As vacas desta variedade apresentam, em geral, um peso de 500 a 600 kg e os touros 800 kg a 900 kg. A estatura dos animais é 1,28 m e 1,38 m para vacas e touros, respectivamente. A pelagem destes animais é vermelha e branca, com predominância da cor vermelha no pescoço e na cauda, e da branca mais frequente nas partes baixas do corpo, ou seja, ventre, membros e úbere (Figura 4). As vacas apresentam úbere volumoso e desenvolvido, bem irrigado, tetos de tamanhos regulares, grossos e bem delimitados. A MRY é um gado especializado na produção de leite, mais rústico que a variedade frísia, e pode ser bem manejado em semi-estabulação. Produz, em média, de 4.000 a 7.000 kg de leite por lactação, com persistência média de 305 dias (NEIVA, 2000).



Figura 3 Animal da raça Holandesa – variedade HPB.



Figura 4 Animal da raça Holandesa – variedade HVB.

2.5.2. Pardo suíço

Raça originária da região dos lagos, na Suíça. É a raça mais antiga selecionada pelo homem que se tem notícia. No Brasil, os maiores rebanhos estão concentrados nas regiões sul e sudeste. No Brasil, durante anos foram feitos vários cruzamentos dessa raça na busca de se obter animais mais produtivos tanto para carne quanto para leite. As vacas desta variedade apresentam, em geral, um peso de 550 a 700 kg e os touros 800 kg a 1.200 kg. Os bezerros pesam, ao nascer, de 35 a 40 kg. A estatura dos animais é de 1,38 a 1,48 m e de 1,55 a 1,65 m para vacas e touros, respectivamente. A pelagem destes animais varia do cinza claro ao cinza escuro (Figura 5). No macho, a tonalidade de cinza é sempre mais escura que na fêmea. No focinho e ao redor dos olhos nota-se uma auréola mais clara, as partes internas dos membros anteriores e posteriores são claras e finalmente apresenta uma faixa clara na região dorsal que vai do pescoço até a inserção da cauda. As vacas apresentam úbere bem implantado, volumoso e com bastante irrigação, tetos médios e quartos bem definidos (NEIVA, 2000).

Tanto os animais puros quanto os seus mestiços são de dupla aptidão (carne e leite). A produção média das vacas está entre 4.500 e 6.200 kg por lactação e o teor médio de gordura varia de 3,5 a 4% (NEIVA, 2000).



Figura 5 Animal da raça Pardo Suíço.

2.5.3. Jersey

Raça originária da ilha de Jersey, localizada no canal da mancha, na Inglaterra. Nos EUA, o gado Jersey foi selecionado e aperfeiçoado para alcançar melhor performance leiteira com alto teor de gordura. Atualmente é bastante difundida em todo o continente americano, e no Brasil está distribuída em todo o território (NEIVA, 2000).

Sendo uma raça de porte pequeno a vaca pesa em média 350 kg e tem estatura de, em média, 1,15 m. Os touros da Jersey pesam 600 a 700 kg e têm em média 1,20 m de estatura. Os bezerros nascem com 25 kg em média. A pelagem destes animais varia do pardo claro ao pardo escuro (Figura 6). No macho, há predominância da cor pardo escura, podendo chegar ao negro. A presença de malhas brancas no peito, ventre e úbere são aceitáveis. O úbere das vacas é glanduloso, bem inserido, com irrigação sanguínea abundante, quartos bem definidos e tetos simétricos e pequenos (NEIVA, 2000).

Animais desta raça tem aptidão exclusiva para leite, alcançam médias de produção entre 3.500 e 5.500 kg por lactação, com persistência de 305 dias e com teor de gordura variando entre 5,5 a 6,0%. Por este motivo é considerada uma raça manteigueira (NEIVA, 2000).



Figura 6 Animal da raça Jersey.

2.6. MODELOS PROPOSTOS NA LITERATURA

Vários modelos algébricos têm sido propostos com o objetivo de descrever as curvas de lactação com maior precisão (OLIVEIRA, REIS, *et al.*, 2007). Tais modelos são usualmente expressos em termos de funções lineares, intrinsecamente lineares e não-lineares. Funções lineares são aquelas em que a variável resposta pode ser expressa como uma combinação linear dos parâmetros associados às covariáveis, enquanto as intrinsecamente lineares são as que podem ser linearizadas por meio de alguma transformação. Em contrapartida, as não-lineares são as que não podem sofrer transformações de modo a produzirem funções lineares dos parâmetros (SANTOS, 2012).

Nos modelos apresentados na Tabela 1, $y(t)$ denota a produção total de leite no instante t , sendo t o tempo decorrido do parto até a data do controle leiteiro. As notações a, b, c e d indexam os parâmetros específicos de cada modelo, os quais descrevem características da forma da curva de lactação (SANTOS, 2012).

Tabela 1 Modelos matemáticos utilizados no ajuste das curvas de lactação.

Modelo	Tipo	Autor	Função
1	LIN	(DAVE, 1971)	$y(t) = a + bt + ct^2$
2	LIN	(SANTOS, 2012)	$y(t) = a + bt + ct^2 + dt^3$
3	LIN	(SINGH e GOPAL, 1982)	$y(t) = a + bt + c \ln t$
4	LIN	(BIANCHINI SOBRINHO, 1984)	$y(t) = a + bt + ct^2 + d \ln t$
5	LIN	(BIANCHINI SOBRINHO, 1984)	$y(t) = a + bt + c/t$
6	INLIN	(BRODY, RAGSDALE e TURNER, 1924)	$y(t) = a \exp(-ct)$
7	INLIN	(WOOD, 1967)	$y(t) = at^b \exp(-ct)$
8	NLIN	(SIKKA, 1950)	$y(t) = a \exp(bt - ct^2)$
9	NLIN	(DHANOA, 1981)	$y(t) = at^{bc} \exp(-ct)$
10	NLIN	(JENKINS e FERRELL, 1984)	$y(t) = \frac{a}{t} \exp(-ct)$
11	NLIN	(PAPAJCSIK e BODERO, 1988)	$y(t) = at \exp(-ct)$
12	NLIN	(CAPPIO-BORLINO, PULINA e ROSSI, 1995)	$y(t) = at^{bc} \exp(-ct)$

* Tipo: LIN (linear), INLIN (intrinsecamente linear), NLIN (não-linear)

3. MATERIAL E MÉTODOS

3.1. MATERIAL

A base de dados original fornecida pela APCBRH contém registros da produção leiteira de Castro provenientes de 27.716 animais, abrangendo principalmente as raças

Holandesa, Pardo Suíço e Jersey (aproximadamente 98% dos animais), entre outras raças (em pequena quantidade). O período compreendido é de janeiro de 2008 a maio de 2012.

Para as análises realizadas foi considerada uma amostra de animais das raças Holandesa, Jersey ou Pardo Suíço que satisfizeram algumas condições, totalizando 753 animais. As condições impostas foram:

- informação disponível da data do último parto;
- informação disponível da primeira medida de produção de leite no mês ou no mês seguinte ao parto;
- idade no parto superior a 1 ano;
- informação disponível de todas as medidas de produção de leite após o primeiro parto sem interrupções. Foram considerados na amostra apenas os registros relativos à produção de leite do primeiro parto de cada animal. Dessa forma a correlação entre diferentes partos de um mesmo animal não precisa ser considerada.

Cada raça foi modelada separadamente, dadas as grandes diferenças percebidas entre o comportamento da produção de leite entre elas. Ainda, em função do tempo de lactação de cada raça, foi considerado nas análises um período de acompanhamento de 18 meses após o parto para a raça Holandesa e de 12 meses para as outras duas raças. A Tabela 2 apresenta informações das bases de dados original e final de cada uma das três raças analisadas neste trabalho.

Tabela 2 Total de animais de cada raça utilizados no trabalho.

Raça	Período de tempo analisado	Animais na base original	Animais na base final	% da base original
Holandesa	18 meses	24.511	557	2,3%
Pardo Suíço	12 meses	825	120	14,5%
Jersey	12 meses	1.121	76	6,8%

As observações da produção de leite foram feitas uma vez por mês em cada animal, e as variáveis que foram utilizadas na modelagem da produção de leite dos animais são:

- Tempo após o parto (em meses);
- Idade do animal no parto (dias);
- Grau sanguíneo (puro, meio-puro ou outros);
- Número de ordenhas (duas ou três);
- Porte do produtor de acordo com o número de animais (pequeno, médio ou grande);
- Ordem de parto (primeiro, segundo, terceiro, etc.).

Para a análise dos dados e ajuste dos modelos foi utilizado o *software* R 2.14.1 (R DEVELOPMENT CORE TEAM, 2011), e para o gerenciamento do banco de dados foram utilizados os *softwares* Microsoft Access® 2010 e Microsoft Excel® 2010. Para estimação dos parâmetros dos modelos foram utilizados os pacotes *nlme* e *geepack*.

3.2. MÉTODOS

Foram utilizadas duas abordagens de modelos estatísticos: modelos lineares e não-lineares. Ambas as abordagens foram aplicadas no contexto de dados longitudinais, no qual várias medidas são feitas ao longo do tempo em cada animal.

3.2.1. Modelos Lineares Generalizados (MLG) para Dados Longitudinais

Em dados longitudinais, como é o caso dos dados analisados neste trabalho, tem-se $\underline{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ uma sequência de n_i medidas observadas para o i -ésimo animal ($i = 1, \dots, m$) e t_j ($j = 1, \dots, n_i$) os tempos em que estas medidas são tomadas para cada animal. Associado a cada y_{ij} tem-se, também, valores x_{ijk} ($k = 1, \dots, p$) de p covariáveis. Devido à estrutura longitudinal dos dados é esperado que exista correlação entre as observações registradas de um mesmo animal, sendo que isso foi testado.

Assume-se que os y_{ij} são realizações de variáveis aleatórias Y_{ij} com distribuição de probabilidade pertencentes à família exponencial, de modo que a esperança para o modelo marginal, $E[Y_{ij}]$, seja escrita como função das covariáveis. Além disso, têm-se as seguintes suposições:

- i. A média da produção de leite de cada animal em cada tempo depende das covariáveis X_{ij} pela função de ligação $g(\mu_{ij}) = X_{ij}^T \underline{\beta}$. Neste trabalho, a função de ligação utilizada foi a inversa:

$$g(\mu_{ij}) = \frac{1}{\mu_{ij}}$$

- ii. A variância depende da média de modo que $V[Y_{ij}] = V(\mu_{ij})\phi^{-1}$, em que $V(\mu_{ij}) = \mu^2$ é uma função de variância conhecida e ϕ é um parâmetro de dispersão que pode ser necessário estimar.

- iii. A covariância entre duas medidas de um mesmo animal Y_{ij} e Y_{ik} é uma função de parâmetros desconhecidos, α , isto é, $Cov[Y_{ij}, Y_{ik}] = V_i(\alpha)$, onde

$$V_i(\alpha) = A_i^{1/2} R(\alpha) A_i^{1/2},$$

em que A_i é a matriz diagonal da função de variância para cada indivíduo e $R(\alpha)$ é a chamada matriz de trabalho, que descreve como as medições dentro de cada animal se relacionam ao longo do tempo (ZEGGER e LIANG, 1986).

Algumas matrizes de trabalho comumente utilizadas, e que serão estudadas neste trabalho, são mostradas a seguir (AGRANONIK, 2009):

- Independente:

Considera que as respostas obtidas em cada medida são independentes entre si. Dessa forma, a matriz R é a matriz identidade:

$$R(\alpha) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n_i \times n_i}$$

em que $n_i \times n_i$ indica que cada animal tem uma matriz dessa forma.

- Auto-regressiva:

A estrutura de correlação entre as respostas é dada por $cor(Y_{ij}, Y_{ij=k}) = \alpha^k$, para $j = 1, \dots, n_i$ e $k = 1, \dots, n_i$, e igual a zero se $j + k > n_i$. Por exemplo, no modelo autoregressivo de ordem 1, com 4 medidas dentro do animal, tem-se:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix}_{4 \times 4}$$

- Permutável:

Essa estrutura de correlação indica que a correlação entre quaisquer medidas é a mesma:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{bmatrix}_{n_i \times n_i}$$

O método de estimação dos parâmetros dos modelos se deu de acordo com a matriz de trabalho escolhida. Se esta foi a matriz independente, a estimação foi feita via Método da Máxima Verossimilhança assumindo-se a distribuição Gama para a variável resposta (produção diária de leite do animal), cuja função de densidade é dada por (PAULA, 2010):

$$f(y|\alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\beta)} \exp(-\alpha y) y^{\beta-1}, \text{ onde } \alpha > 0, \beta > 0 \text{ e } y > 0.$$

A distribuição Gama e duas variações são amplamente utilizadas na literatura para a modelagem de dados não-negativos de natureza contínua, quando estes não são simetricamente distribuídos em relação aos seus valores (PRUDENTE, 2009).

Nos demais casos, a estimação foi realizada via Equações de Estimação Generalizadas (EEG). As EEG foram desenvolvidas para produzir estimativas mais eficientes e não-viciadas para os parâmetros do modelo de regressão quando se lida com dados correlacionados, pois considera a estrutura de correlação entre as observações. EEG são uma extensão dos MLG, sendo que não é necessário assumir que a variável resposta seja da família exponencial. Assume-se, entretanto, que a média e a variância estão caracterizadas como em um MLG (AGRANONIK, 2009). Ao contrário do Método de Máxima Verossimilhança não é assumida, nesses casos, distribuição de probabilidade para a variável resposta, sendo necessário o conhecimento apenas do primeiro e segundo momentos.

A estimativa de $\underline{\beta}$, onde $\underline{\beta}$ é o vetor de z parâmetros, é a solução do sistema de equações diferenciais quasi-score definidas por:

$$U_k(\underline{\beta}) = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta_k} V_i^{-1}(y_i - \mu_i) = 0 \rightarrow k = 1, \dots, z.$$

Para obter $\underline{\beta}$, é sugerido um processo iterativo baseado no escore de Fisher modificado e nas estimativas de α e ϕ , obtidas por meio do método dos momentos (LIANG e ZEGGER, 1986). Neste processo, altera-se entre estimar $\underline{\beta}$ para valores de $\hat{\phi}$ e $\hat{\alpha}$ e estimar (ϕ, α) para valores fixos de $\underline{\beta}$. Maiores detalhes acerca da estimação de $\underline{\beta}$, da variância de $\underline{\beta}$, da estimação da matriz de trabalho e do parâmetro de dispersão podem ser encontrados em (AGRANONIK, 2009).

3.2.2. Modelos Não-Lineares

A forma geral do modelo não-linear é apresentada da seguinte forma:

$$Y_i = g(X_i, \beta) + \varepsilon_i,$$

em que:

- \underline{X}_i é o vetor de observações das covariáveis para o i-ésimo animal:

$$\underline{X}_i = [1 \quad X_{i,1} \quad \dots \quad X_{i,p-1}]^T$$

- $\underline{\beta}$ é o vetor de parâmetros desconhecidos a serem estimados:

$$\underline{\beta} = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_p]^T$$

- Os erros ε_i têm distribuição Normal com média zero, variância constante σ^2 , são independentes e não correlacionados, ou seja:

$$\varepsilon_i \sim N(0, I\sigma^2)$$

- $g(X_i, \beta)$ é uma função diferenciável.

Nos modelos de regressão não-linear o número de parâmetros p não é, necessariamente, diretamente relacionado ao número de variáveis explicativas no modelo. A principal característica desses modelos é que a parte fixa, $g(X, \beta)$, em geral, decorre de um processo determinístico deduzido a partir de suposições teóricas, sendo os parâmetros resultantes interpretáveis e a parte aleatória composta de erros homogêneos.

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes com a estrutura especificada de forma não-linear, supondo normalidade para elas. A estimativa do vetor de parâmetros β é calculada pelo método dos mínimos quadrados que, nesse caso, coincide com o método de máxima verossimilhança. A estimativa de β é, então, calculada, minimizando-se a função quadrática definida por:

$$S(\beta) = \sum_{i=1}^n (Y_i - g(X_i, \beta))^2$$

Derivando-se $S(\beta)$ em relação à β_r , obtém-se

$$\frac{\partial S(\beta)}{\partial \beta_r} = 2 \sum_{i=1}^n (Y_i - g(X_i, \beta)) \left[\frac{\partial g(X_i, \beta)}{\partial \beta_r} \right].$$

A estimativa $\hat{\beta}$ do vetor de parâmetros β é obtida igualando-se $\frac{\partial S(\beta)}{\partial \beta_r}$ a zero para $r = 1, \dots, p$. Em geral, as equações $\frac{\partial S(\beta)}{\partial \beta_r} = 0$ não são lineares e têm que ser resolvidas numericamente, o que requer cálculos computacionais intensivos (MOUTINHO, PRUDENTE e DEMÉTRIO, 2009).

3.2.3. Métodos Estatísticos para Escolha dos Modelos

Para a escolha do melhor modelo para cada raça foram utilizadas algumas técnicas, descritas a seguir.

3.2.3.1. Significância das Covariáveis

A estatística de teste a ser utilizada para testar a hipótese relativa à permanência ou não das covariáveis no modelo será a estatística de Wald, que é um teste baseado na distribuição assintótica normal de $\hat{\beta}_i$, e é uma generalização da estatística t de Student, dada por:

$$W = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{V}[\hat{\beta}_i]}}$$

A hipótese testada por ela é

$$\begin{cases} H_0: \beta_i = 0 \\ H_a: \beta_i \neq 0. \end{cases}$$

O nível de significância adotado neste trabalho será de 5% ou 10%. Logo, quando o p-valor relativo a uma covariável for inferior a esse nível, esta permanecerá no modelo.

3.2.3.2. AIC

O *AIC* (*Akaike's Information Criterion*) é o valor do logaritmo neperiano da função de verossimilhança, penalizado pelo número de parâmetros estimados no modelo. Isto é,

$$AIC = -2l + 2p,$$

sendo l o máximo do logaritmo da função de verossimilhança e p o número de parâmetros estimados no modelo.

O *AIC* auxilia na escolha do modelo que melhor se ajusta aos dados. Tal critério pode ser usado nas classes de modelos lineares, modelos lineares generalizados e modelos não-lineares. O modelo que apresentar o menor valor para o *AIC* é considerado o mais adequado (AKAIKE, 1974).

3.2.3.3. QIC

Quando se utilizam as EEG no ajuste do modelo não é assumida nenhuma distribuição para a variável resposta e, portanto, não há uma função de verossimilhança. Com isso o uso do *AIC* na seleção do modelo fica inviável, e faz-se necessário outro critério. O *QIC* (*Quasi-Likelihood Information Criterion*), baseado na quase-verossimilhança $Q(\cdot)$, é definido por $QIC(R) = -2Q(\hat{\beta}(R); I; D) + 2 \text{tr}(\hat{\Omega}_I \hat{V}_r)$ (PAN, 2001), em que:

- R é a matriz de trabalho assumida;
- $\hat{\beta}(R)$ são as estimativas dos parâmetros utilizando a matriz R ;
- I é a matriz de trabalho independente;
- D representa os valores observados nos dados de cada animal, de forma que $D = \{(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)\}$;
- $\hat{\Omega}_I = -\frac{\partial^2 Q(\beta; I; D)}{\partial \beta \partial \beta'}$
- \hat{V}_r é o estimador de $\text{cov}(\hat{\beta})$.

De acordo com Pan (2001), o uso do $QIC_u(R) = -2Q(\hat{\beta}(R); I; D) + 2p$ é sugerido como uma aproximação para o $QIC(R)$. O $QIC_u(R)$ deve ser utilizado na seleção das variáveis, enquanto que o $QIC(R)$ deve ser utilizado na escolha da matriz de trabalho adequada. Caso o $QIC(R)$ indique que a matriz de trabalho adequada aos dados seja a independente, o modelo recai em um MLG conforme descrito na seção 3.2.1.

3.2.3.4. Gráficos de Resíduos

Quando um MLG é ajustado pode-se utilizar os resíduos *deviance*, que são assumidos apresentarem distribuição aproximadamente Normal. Foram analisados os resíduos brutos e padronizados quanto à homocedasticidade e normalidade através de um gráfico quantil-quantil (QQ-Plot). Além disso, avaliou-se a presença de pontos influentes com a distância de Cook, que não deve ser superior a 1 (PAULA, 2010).

Para os ajustes via EEG ainda não há muitos métodos de diagnóstico implementados no *software* utilizado. Sendo assim, foram obtidos gráficos dos resíduos brutos ao longo do tempo e os gráficos dos resíduos em cada tempo, para cada animal (BALLINGER, 2004). Dessa forma consegue-se avaliar a homocedasticidade e a presença de pontos influentes no modelo (PREISSER e QAQISH, 1996). Nesse caso, como o ajuste é feito via EEG, não há suposição de normalidade dos resíduos.

Por último, para os modelos não-lineares, o pacote *nlme* disponibiliza o gráfico dos resíduos padronizados *versus* os valores ajustados, permitindo a avaliação da homocedasticidade. Em modelos não-lineares há novamente a suposição de normalidade dos resíduos, que foi verificada graficamente por meio do QQ-Plot (PINHEIRO e BATES, 2000).

3.2.3.5. Curvas Observadas e Estimadas

O passo final na escolha dos modelos foram os gráficos das curvas médias observadas e estimadas, verificando se as curvas estavam bem ajustadas. A correlação linear de Pearson foi utilizada para quantificar quão próximas essas curvas se apresentaram uma da outra.

3.2.4. Seleção de Modelos Lineares

Decidiu-se modelar as raças separadamente devido às diferenças observadas na análise descritiva. Por exemplo, uma covariável que seja significativa para a raça Holandesa pode não ser significativa para a Jersey. Além disso, como se trata de dados longitudinais, é esperado que haja uma estrutura de correlação entre as observações de cada animal. Um modelo que considere as três raças conjuntamente assumiria a mesma estrutura para as três raças, o que é uma suposição forte. Modelando as raças separadamente se ganha, assim, mais liberdade na escolha de um modelo adequado para cada uma delas.

A escolha do modelo linear de cada raça foi baseada em quatro passos, resumidos a seguir:

1. Escolha do modelo considerando apenas o tempo;
2. Escolha da matriz de trabalho adequada;
3. Escolha das variáveis significativas;

4. Análise de resíduos e qualidade do modelo ajustado.

O primeiro passo consiste em escolher, dentre os modelos lineares descritos na literatura (Tabela 1), qual modelo (considerando apenas o tempo) é mais adequado aos dados. Para realizar essa escolha foram comparadas as correlações de Pearson entre as curvas observadas e estimadas, bem como os gráficos dessas curvas. Escolhida a forma como o tempo entra no modelo, precisa-se determinar a estrutura de correlação mais adequada entre as observações de cada animal e também as variáveis significativas.

Os segundo e terceiro passos têm por objetivo selecionar a matriz de trabalho e as covariáveis significativas, respectivamente. Para isso foi utilizado o procedimento adotado por Cui (CUI, 2007). Na escolha da matriz de trabalho foram ajustados três modelos, cada um com uma das matrizes de trabalho propostas na metodologia (independente, AR1 e permutável) e todas as covariáveis de interesse. Neste passo foram deixadas as covariáveis significativas ao nível de 10% de significância por ainda não se tratar da escolha final das covariáveis, e sim da matriz de trabalho. A matriz de trabalho escolhida será aquela que apresentar o menor *QIC*.

O terceiro passo consiste na escolha das covariáveis finais do modelo (além do tempo). Fixando-se a matriz de trabalho escolhida no passo anterior, ajustaram-se modelos com várias combinações entre as covariáveis. Dentre esses modelos, o que apresentar menor *QICu* será o escolhido. No quarto e último passo foi feita a análise dos resíduos e do modelo ajustado, comparando-se as curvas estimadas e observadas.

3.2.5. Seleção de Modelos Não-Lineares

A metodologia para a escolha dos modelos não-lineares adequados a cada uma das raças foi mais simples do que aquela utilizada nos modelos lineares. Os passos adotados foram análogos àqueles descritos na seção anterior, com algumas diferenças.

O pacote *nlme* possibilita a inclusão de uma estrutura de correlação entre as observações no modelo com a função *gnls*, porém sua aplicação na base de dados utilizada não foi possível devido a problemas computacionais na estimação. A função utilizada foi a *nls*, desconsiderando a matriz de trabalho, o que descarta o segundo passo; com isso, na escolha das covariáveis significativas foi utilizado o *AIC*, ao invés do *QICu*.

3.2.6. Obtenção de Medidas de Interesse

Há um interesse especial por algumas medidas quando se trabalha com curvas de lactação, tais como: tempo de pico (t_p), produção no pico (p_p) e persistência (P), embora essa última só seja obtida nos modelos não-lineares.

Nos modelos lineares o tempo de pico foi obtido derivando-se a equação do modelo em relação ao tempo. Já a produção de leite nesse pico foi obtida substituindo-se a estimativa do tempo de pico e os valores das estimativas dos parâmetros das covariáveis significativas na equação do modelo ajustado.

Nos modelos não-lineares, em particular no modelo de Wood, essas medidas são obtidas como mostrado na Tabela 3, de acordo com Ckroh e Coimbra (2005). Como os modelos dependem de covariáveis, a produção no pico foi obtida da mesma forma que nos modelos lineares, aplicando-se o modelo estimado.

Tabela 3 Funções dos parâmetros tempo de pico e persistência.

Medida	Expressão
Tempo de pico	$t_p = \frac{b}{c}$
Persistência	$P = -(b + 1) \ln c$

4. RESULTADOS E DISCUSSÃO

4.1. ANÁLISE DESCRITIVA

Para um melhor entendimento dos dados disponíveis no banco de dados é apresentada, a seguir, uma análise descritiva dos mesmos.

Comparando as três raças quanto à produção de leite de cada uma delas (Figura 7), observa-se que a raça Holandesa apresenta a maior produção média de leite, seguida pela Pardo Suíço e depois pela Jersey, sendo que as três raças apresentam pico de produção no terceiro mês após o parto.

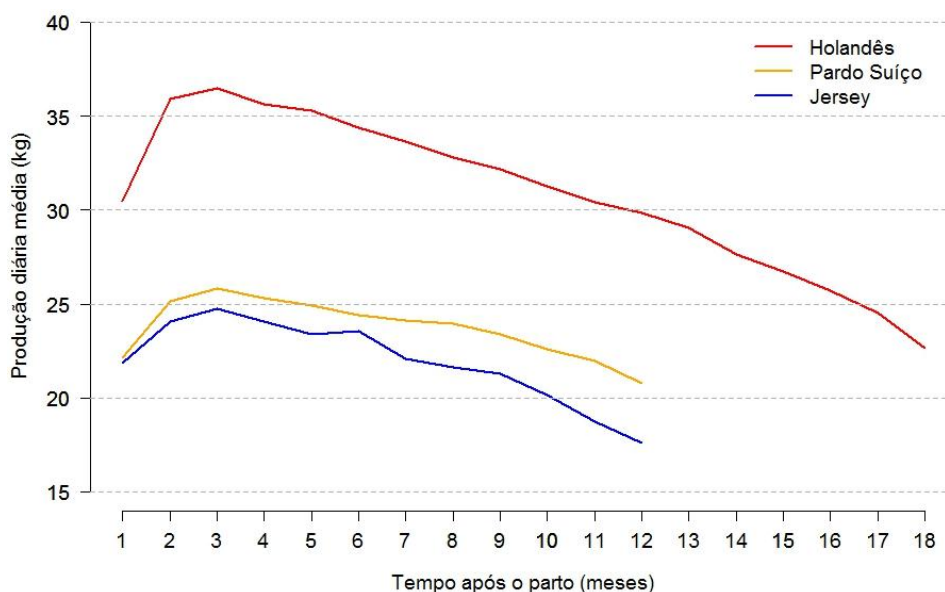


Figura 7 Produção diária média (em kg) das raças Holandesa, Jersey e Pardo Suíço ao longo do tempo.

No que diz respeito à variabilidade da produção de leite ao longo do tempo, observa-se na Figura 8 que as três raças têm variabilidade constante ao longo do tempo após o parto, embora a raça Holandesa apresente uma amplitude maior do que as demais. Isso mostra que a variância pode ser considerada constante e que não precisa ser modelada, e também que as três raças têm comportamentos distintos. Logo, sugere-se que sejam modeladas separadamente.

Na Figura 9, que também mostra a variabilidade da produção de leite em torno da média, para cada raça, nota-se que a média é bastante próxima da mediana, indicando uma simetria e que valores extremos não estão influenciando a média.

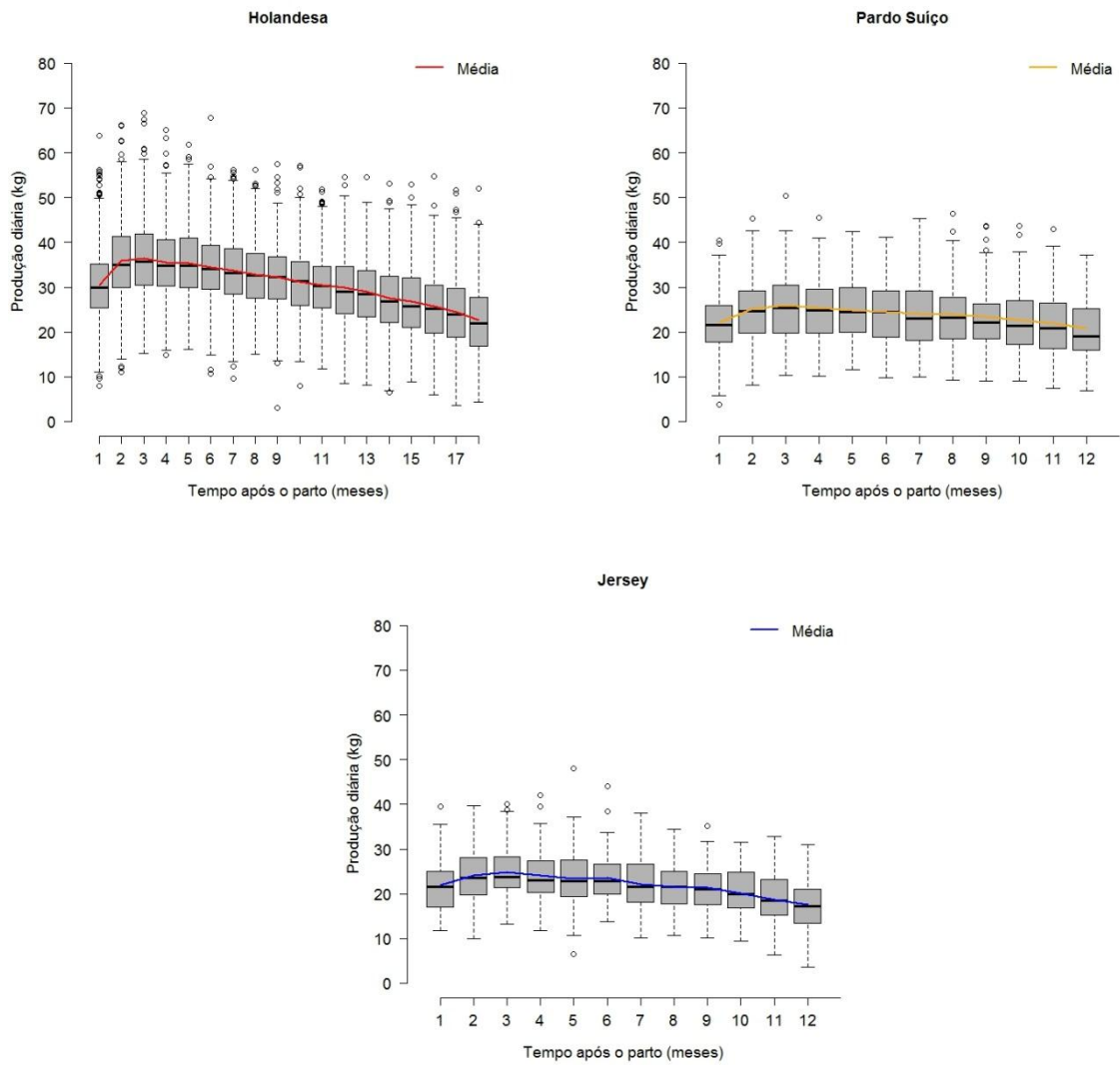


Figura 8 Dispersão dos valores de produção diária (em kg) em torno da média de cada raça.

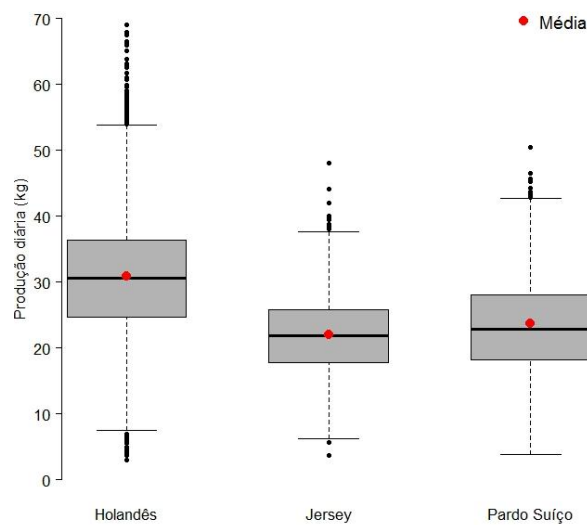


Figura 9 Distribuição da produção de leite (em kg) por raça.

A Tabela 4 apresenta as principais medidas resumo da produção diária de acordo com as raças.

Tabela 4 Estatísticas da produção diária (em kg) de acordo com a raça.

RAÇA	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
Holandesa	557	3,0	30,8	30,6	69,0	8,9
Pardo Suíço	120	3,8	23,7	22,9	50,4	7,3
Jersey	76	3,6	21,9	21,8	48,0	6,2

Na Figura 10 são mostrados alguns perfis individuais de produção de leite. Devido à grande quantidade de animais optou-se por mostrar alguns animais com altas e baixas produções em torno da média. Nota-se uma grande variabilidade entre os animais.

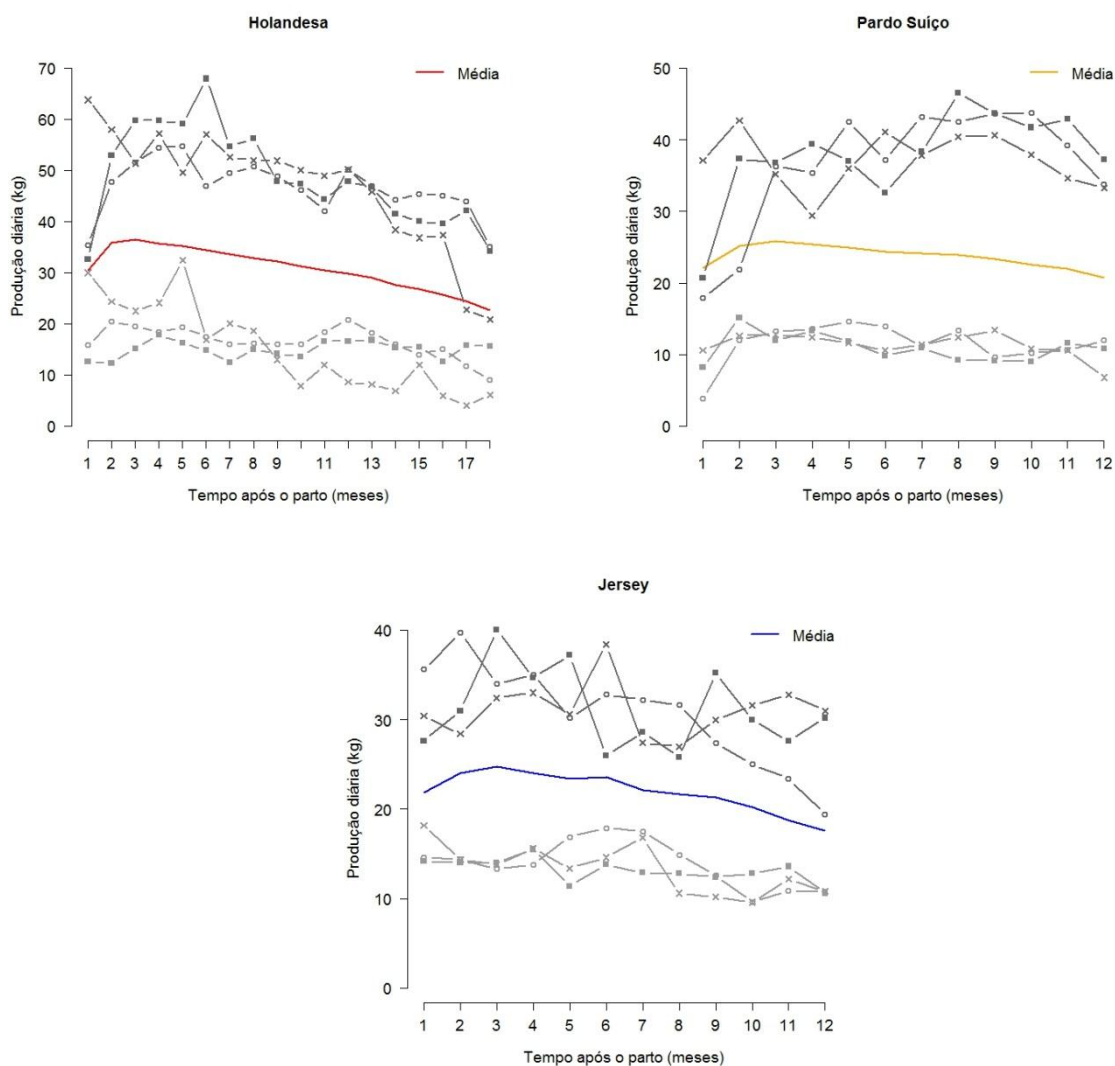


Figura 10 Perfis individuais de produção diária (em kg) em torno da média de cada raça.

Nas seções a seguir serão apresentadas análises descritivas de acordo com cada raça.

4.1.1. Raça Holandesa

Uma possível covariável de interesse no modelo para explicar a variabilidade na produção de leite é o número de ordenhas (duas ou três). Do total de 557 animais dessa raça, têm-se duas ordenhas para 62% e três ordenhas para 38%. A Figura 11 mostra que tanto a produção média quanto a produção total são maiores para os animais que tiveram três ordenhas, e que a variabilidade é a mesma tanto para duas quanto para três ordenhas.

A Tabela 5 mostra um resumo da produção de leite de acordo com o número de ordenhas, pela qual se observa que a diferença entre as duas formas de manejo é de 5 kg na média, e que o desvio padrão é ligeiramente maior para três ordenhas.

Tabela 5 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o número de ordenhas.

NÚMERO DE ORDENHAS	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
2	387	3,0	28,5	28,3	65,1	8,1
3	233	6,5	34,6	34,5	69,0	8,7

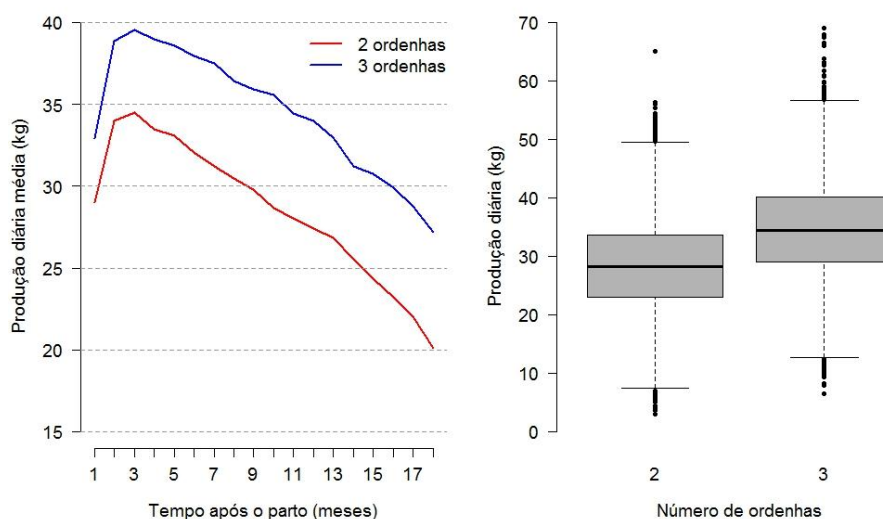


Figura 11 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o número de ordenhas.

Outra covariável de interesse é o grau de sangue (pureza do animal). Do total de animais, 68% são puros, 8% são meio-puros e os demais 24% representam outras classificações (resultantes de misturas de raças). A Tabela 6 e a Figura 12 mostram que a maior produtividade média é dos animais puros e meio-puros, embora estes últimos representem apenas 8% do total de animais. Isso pode explicar por que a curva média não apresenta o mesmo comportamento dos demais animais. Observa-se também que os animais puros apresentam pontos discrepantes (*outliers*). Ainda, os animais da raça Holandesa que foram cruzados com outras raças (curva cinza) apresentam uma produção de leite menor.

Tabela 6 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o grau de pureza do sangue.

GRAU DE SANGUE	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
1/2 Puro	45	12,5	33,6	33,8	54,7	7,7
Puro	377	3,7	31,0	30,6	69,0	8,8
Outros	135	3,0	29,4	29,4	58,6	9,1

A partir desta análise, a covariável grau de pureza aparentemente não apresenta relevância para explicar a produção média de leite dos animais, mas isso deve ser verificado na modelagem.

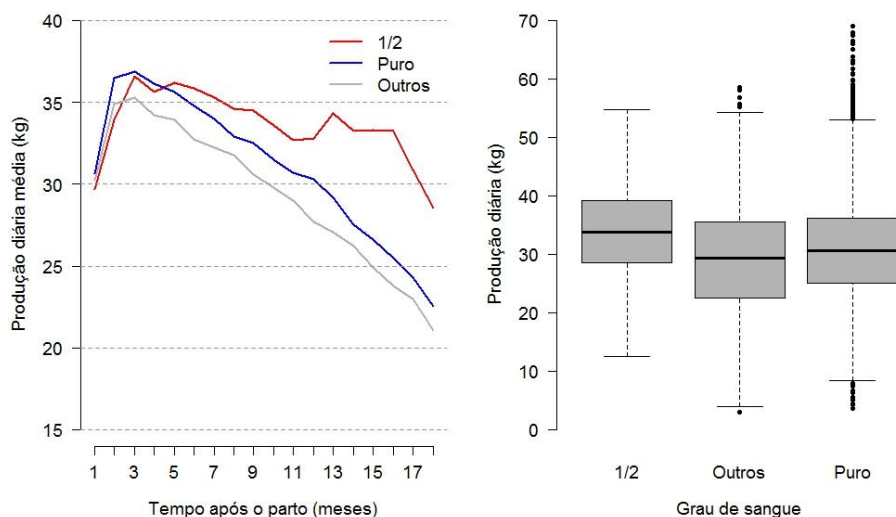


Figura 12 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o grau de pureza do sangue.

Devido à grande quantidade de animais da raça holandesa e às diferenças detectadas quanto ao tamanho dos rebanhos (variando de 1 a 60 animais), os 46 produtores foram classificados em pequenos, médios ou grandes, como mostrado na Tabela 7.

Tabela 7 Classificação dos produtores da raça Holandesa quanto ao porte.

TAMANHO	ANIMAIS	PRODUTORES	%
Grande	30 ou mais	4	8,7%
Médio	Entre 9 e 30	22	47,8%
Pequeno	Menos que 9	20	43,5%

Pela Figura 13 conclui-se que os produtores dos diferentes portes não apresentam grande diferença na média de produção de leite ao longo do tempo após o parto. Como esperado, grandes produtores produzem mais do que os médios, que por sua vez produzem mais do que os pequenos.

As estatísticas da produção de todo o período analisado, para os três portes de produtores, é mostrada na Tabela 8.

Tabela 8 Estatísticas da produção diária (em kg) dos animais da raça Holandesa, de acordo com o tamanho do produtor.

TAMANHO	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
Pequeno	93	3,0	29,4	29,2	69,0	8,5
Médio	271	6,5	30,5	30,0	67,9	9,1
Grande	193	3,7	31,9	32,1	65,1	8,5

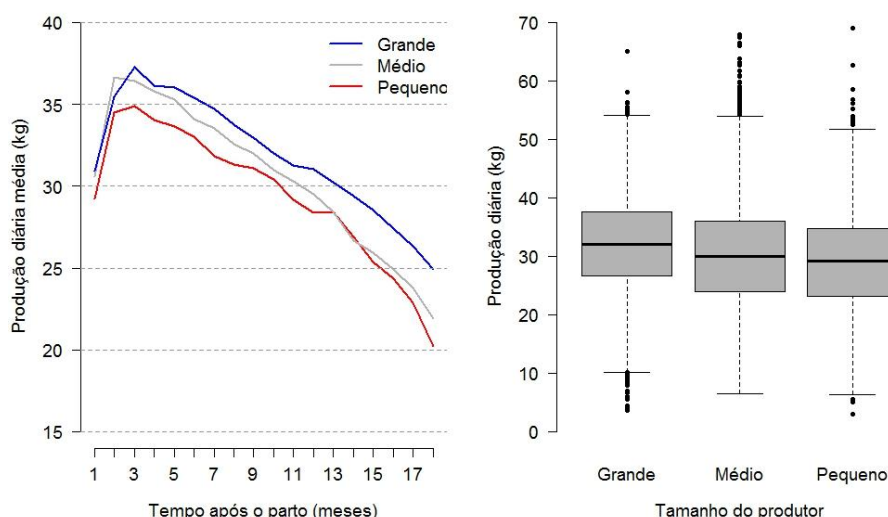


Figura 13 Produção diária (em kg) dos animais da raça Holandesa, de acordo com o tamanho do produtor.

A idade do animal no parto é outra covariável disponível no banco de dados e que deve ser estudada. Na Figura 14 tem-se a média de produção, por mês, de cada faixa de idade (em meses). Observa-se uma produção maior para as vacas holandesas mais novas, sendo que essa média se estabiliza ao longo dos anos de vida do animal. Ressalta-se que o

número de animais nas idades mais avançadas é menor, portanto a média é baseada em poucos valores.

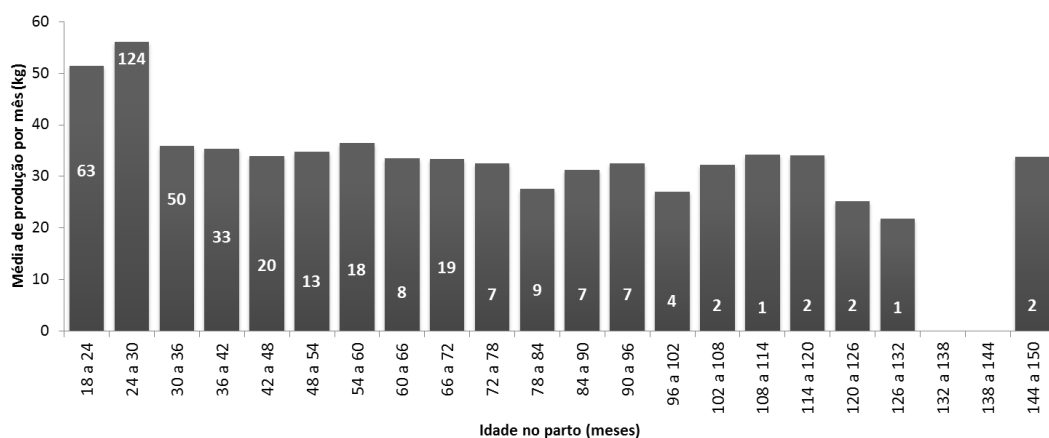


Figura 14 Produção diária média (em kg) dos animais da raça Holandesa de acordo com a idade do animal no parto, com o número de animais em cada faixa.

4.1.2. Raça Pardo Suíço

Analisando-se a produção de leite de acordo com o número de ordenhas para os animais da raça Pardo Suíço têm-se os resultados mostrados na Figura 15 e Tabela 9.

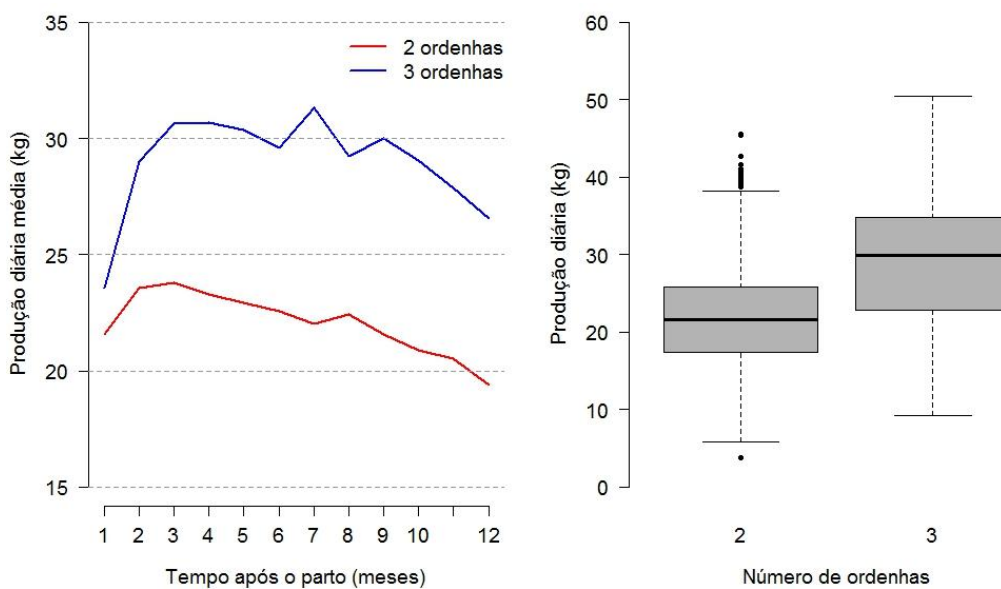


Figura 15 Produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o número de ordenhas.

Tabela 9 Estatísticas da produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o número de ordenhas.

NÚMERO DE ORDENHAS	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
2	101	3,8	22,0	21,6	45,6	6,3
3	39	9,2	29,0	30,0	50,4	7,6

Da Figura 15 conclui-se que os animais com três ordenhas têm média de produção maior que os demais. Ainda, na Tabela 9 observa-se que a diferença entre as duas formas de manejo é de 7 kg, na média.

Na Figura 16 nota-se que os animais da raça Pardo Suíço que têm sangue diferente de puro ou meio-puro têm uma produção média bastante superior. Na média, esse grupo produz cerca de 4,5 kg de leite a mais, de acordo com a Tabela 10.

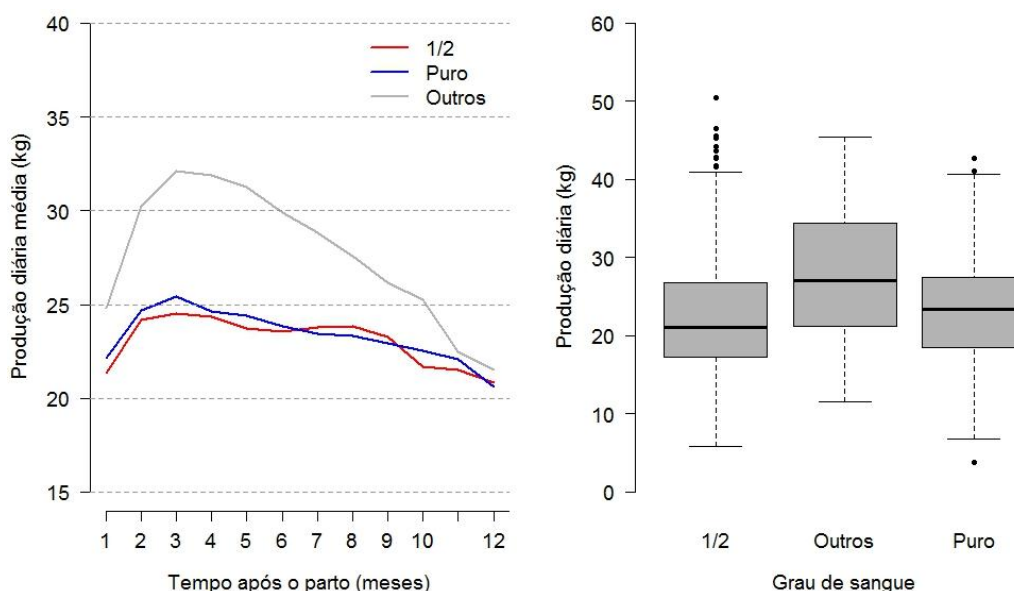


Figura 16 Produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o grau de pureza de sangue.

Tabela 10 Estatísticas da produção diária (em kg) dos animais da raça Pardo Suíço, de acordo com o grau de pureza de sangue.

GRAU DE SANGUE	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
1/2 Puro	40	5,8	23,1	21,1	50,4	7,8
Puro	67	3,8	23,3	23,5	42,7	6,6
Outros	13	11,5	27,7	27,0	45,4	8,0

Para essa raça os animais não foram classificados de acordo com o tamanho do rebanho devido ao pequeno número de produtores (apenas 6).

Com relação à idade do animal no parto, conforme a Figura 17, as diferentes faixas de idade parecem não diferir quanto à produção de leite.

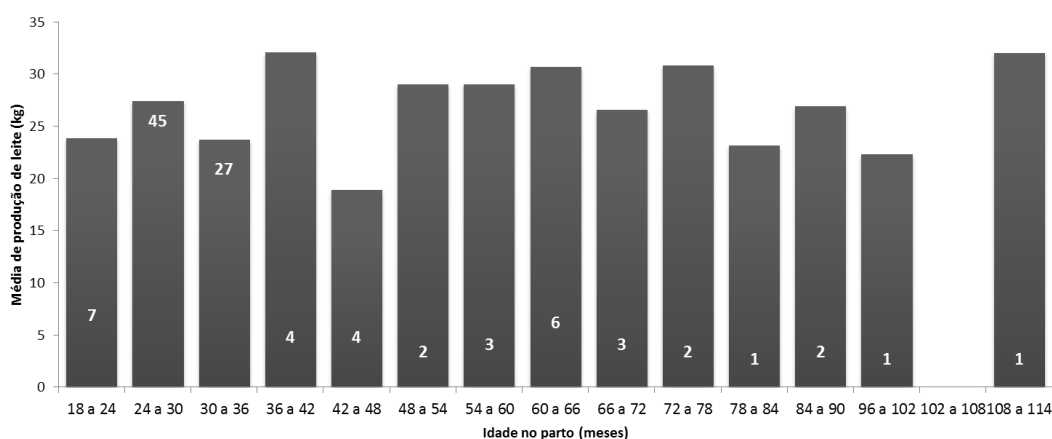


Figura 17 Produção diária média (em kg) dos animais da raça Pardo Suíço de acordo com a idade do animal no parto, com o número de animais em cada faixa.

4.1.3. Raça Jersey

A Figura 18 mostra a produção média de leite de acordo com o número de ordenhas, para a raça Jersey. Observa-se, ao contrário das outras duas raças, que não há grande diferença entre as duas formas de manejo.

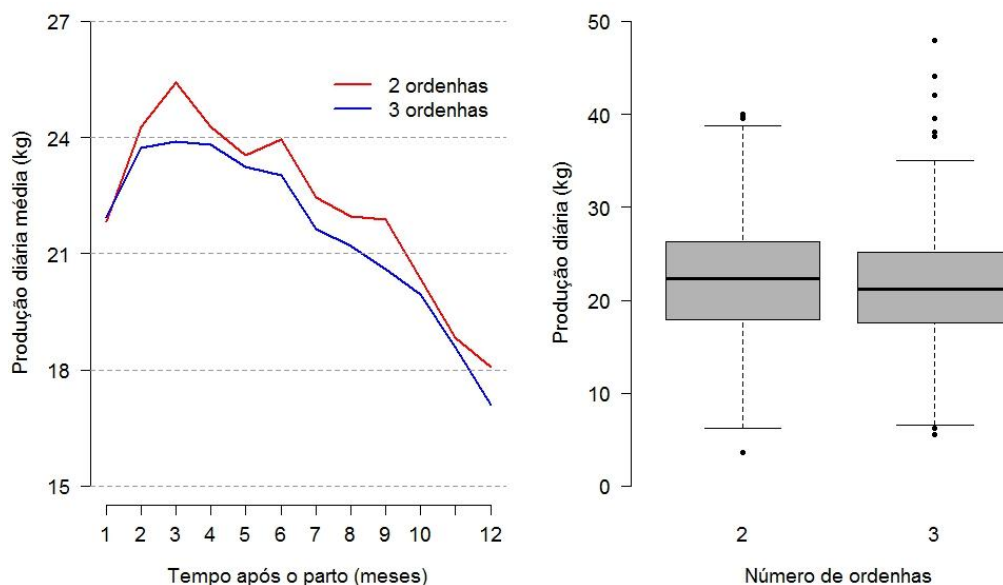


Figura 18 Produção diária de leite (em kg) dos animais da raça Jersey, de acordo com o número de ordenhas.

Tabela 11 Produção diária (em kg) dos animais da raça Jersey, de acordo com o número de ordenhas.

NÚMERO DE ORDENHAS	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
2	47	3,6	22,3	22,3	40,0	6,3
3	34	5,6	21,5	21,2	48,0	5,9

Conforme a Tabela 11, a diferença entre duas ou três ordenhas dos animais da raça Jersey não chega a ser de 1 kg de leite na média.

Os resultados da produção de leite em função do grau de pureza de sangue dos animais da raça Jersey são apresentados na Tabela 12 e na Figura 19. Observa-se que não há grande diferença entre os três níveis de pureza.

Tabela 12 Produção diária (em kg) dos animais da raça Jersey, de acordo com o grau de pureza de sangue.

GRAU DE SANGUE	ANIMAIS	MÍNIMO	MÉDIA	MEDIANA	MÁXIMO	DESVIO PADRÃO
1/2 Puro	25	3,6	21,9	22,0	39,7	6,3
Puro	40	5,6	22,0	21,6	48,0	5,9
Outros	11	6,2	21,9	21,9	40,0	7,0

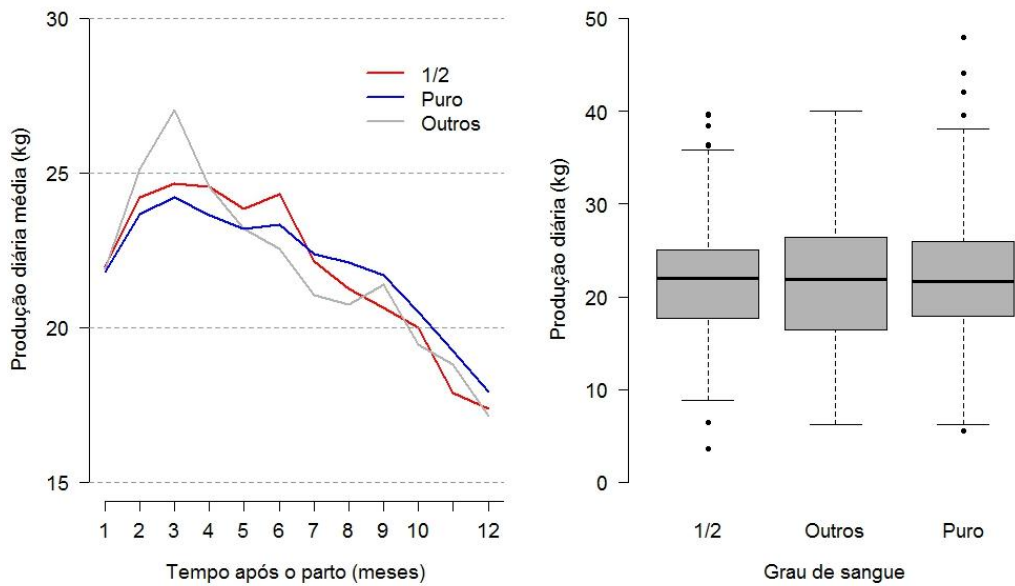


Figura 19 Produção diária (em kg) dos animais da raça Jersey, de acordo com o grau de pureza de sangue.

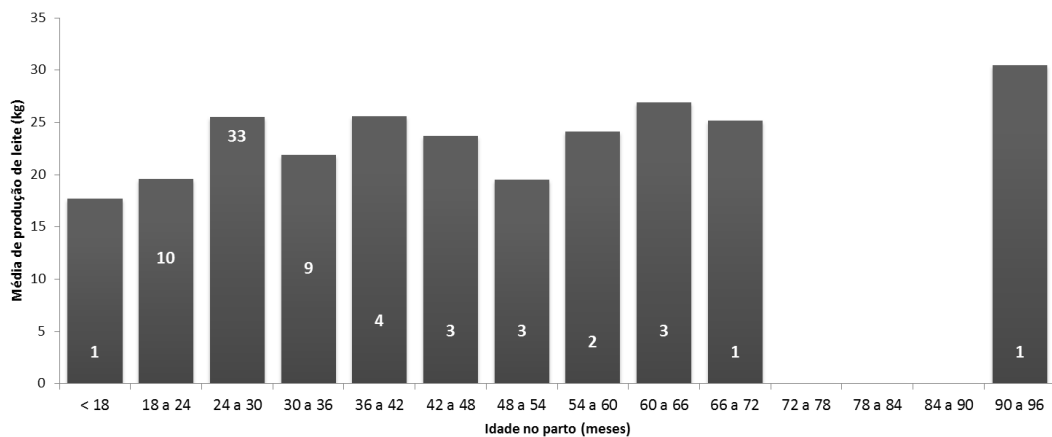


Figura 20 Produção diária média (em kg) dos animais da raça Jersey de acordo com a idade do animal no parto, com o número de animais em cada faixa.

Na Figura 20 é estudado o efeito da idade do animal na produção de leite, para os animais da raça Jersey. Não se observa influência acentuada, mas isso deverá ser testado na modelagem.

4.1.4. Multicolinearidade

Para as três raças foi detectada uma correlação alta entre duas covariáveis descritas na seção 3.1: ordem de parto e idade do animal no parto. Tal fato era esperado, pois quanto mais velho o animal, mais partos teve. As correlações de Pearson obtidas em cada raça são mostradas Tabela 13.

Tabela 13 Correlação entre as covariáveis ordem de parto e idade no parto.

RAÇA	CORRELAÇÃO
Holandesa	0,8733
Pardo Suíço	0,8466
Jersey	0,7176

Como os modelos não podem ter covariáveis que apresentam correlação acentuada entre si, decidiu-se manter apenas a idade do animal no parto nas análises.

4.2. MODELOS LINEARES AJUSTADOS

4.2.1. Raça Holandesa

Como descrito na metodologia, o primeiro passo consistiu na escolha do modelo mais adequado considerando apenas a covariável tempo. Desse modo, para os modelos mostrados na Tabela 1, são mostradas na Tabela 14 as correlações de Pearson entre as curvas observadas e estimadas. Observa-se que a estrutura que apresenta maior correlação é a do modelo 5, para qualquer uma das matrizes de trabalho possíveis. O termo quadrático do tempo no modelo 4 não foi significativo, o que o deixa igual ao modelo 3. As curvas observadas e estimadas para esse modelo são mostradas na Figura 21, na qual se observa que o ajuste é satisfatório.

Tabela 14 Correlações lineares entre as curvas observadas e estimadas, para a raça Holandesa.

Modelo	Independente	AR1	Permutável
1 $y(t) = a + bt + ct^2$	0,9644	0,9330	0,9644
2 $y(t) = a + bt + ct^2 + dt^3$	0,9700	0,9559	0,9700
3 $y(t) = a + bt + c \ln t$	0,9868	0,9839	0,9868
4 $y(t) = a + bt + ct^2 + d \ln t$	0,9865	0,9842	0,9865
5 $y(t) = a + bt + c/t$	0,9874	0,9853	0,9874

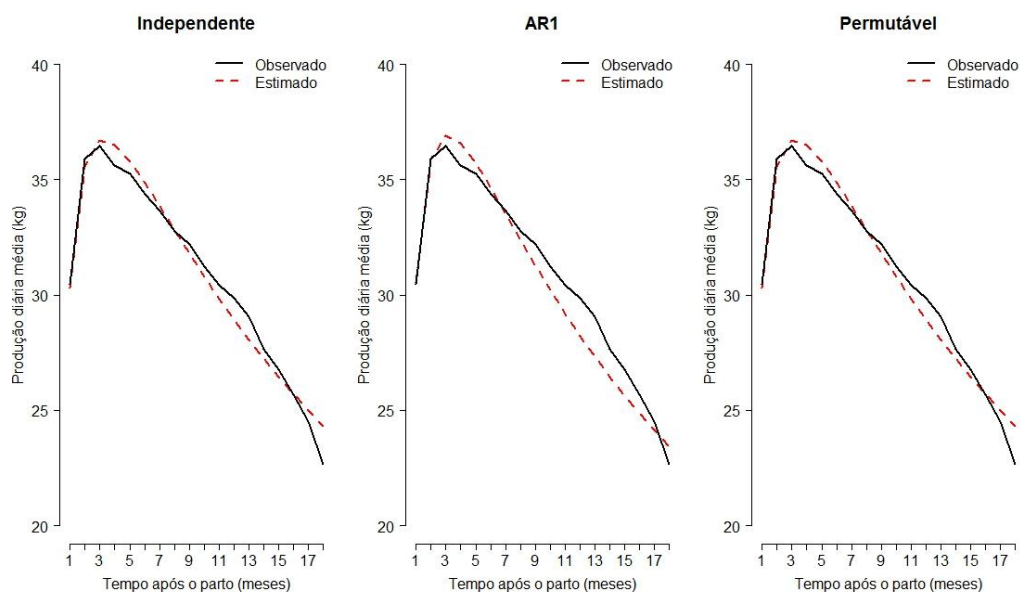


Figura 21 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Holandesa.

Com esse modelo escolhido, parte-se agora para a escolha da matriz de trabalho e das covariáveis significativas. De acordo com o método adotado por Cui (2007), é necessário avaliar o *QIC* nos modelos com todas as covariáveis. Como já descrito, foram avaliadas todas as covariáveis significativas ao nível de 10% de significância, e os resultados são mostrados na Tabela 15.

Tabela 15 Valores de *QIC* para a raça Holandesa.

Matriz de trabalho	Covariáveis ($\alpha = 0, 10$)	<i>QIC</i>
Independente	Número de ordenhas, idade, grau de sangue	22.680,5
AR1	Número de ordenhas, idade, grau de sangue	23.049,9
Permutável	Número de ordenhas, idade, grau de sangue, tamanho do produtor	22.820,1

A matriz de trabalho que apresentou o menor valor de *QIC* foi a independente. Com isso pode-se ajustar o modelo utilizando a abordagem de MLG, que considera ausência de relação entre as observações de cada animal. A seguir, foram ajustados vários modelos com essa estrutura, considerando as combinações possíveis entre as variáveis. Entre as diversas combinações possíveis, as cinco com menores *AICs* são apresentadas na Tabela 16.

Tabela 16 Modelos para a raça Holandesa com menores AIC.

Modelo	Covariáveis ($\alpha = 0,05$)	AIC
1	Número de ordenhas, grau de sangue, tamanho do produtor, idade	68.897
2	Número de ordenhas, grau de sangue, idade	68.907
3	Número de ordenhas, grau de sangue, tamanho do produtor	68.925
4	Número de ordenhas, grau de sangue	68.936
5	Número de ordenhas, tamanho do produtor, idade	69.017

A partir da Tabela 16, é possível observar que os modelos 1 a 4 apresentaram valores de *AIC* bem próximos, não justificando, assim, a escolha dos modelos com maior número de variáveis como os modelos 1, 2 e 3. Em análises adicionais verificou-se também que a idade não exerce grande influência nas curvas estimadas, justificando sua retirada do modelo. Portanto, escolheu-se o modelo 4 (com número de ordenhas e grau de sangue), cujas estimativas dos parâmetros são mostradas na Tabela 17.

Tabela 17 Estimativas dos parâmetros do modelo linear ajustado para a raça Holandesa.

Variável	Estimativa	Erro padrão	Valor t	P-valor
Intercepto	2,14e-02	3,95e-04	54,22	< 2,00e-16
Tempo	1,12e-03	2,29e-05	49,00	< 2,00e-16
1/Tempo	1,17e-02	5,08e-04	23,07	< 2,00e-16
Ordenhas - 3	-5,90e-03	1,59e-04	-37,13	< 2,00e-16
Sangue - outros	2,82e-03	3,05e-04	9,26	< 2,00e-16
Sangue - puro	7,83e-04	2,73e-04	2,87	0,0042
Parâmetro de dispersão (ϕ)	0,0599	-	-	-

Representação gráfica do modelo ajustado é mostrada na Figura 22.

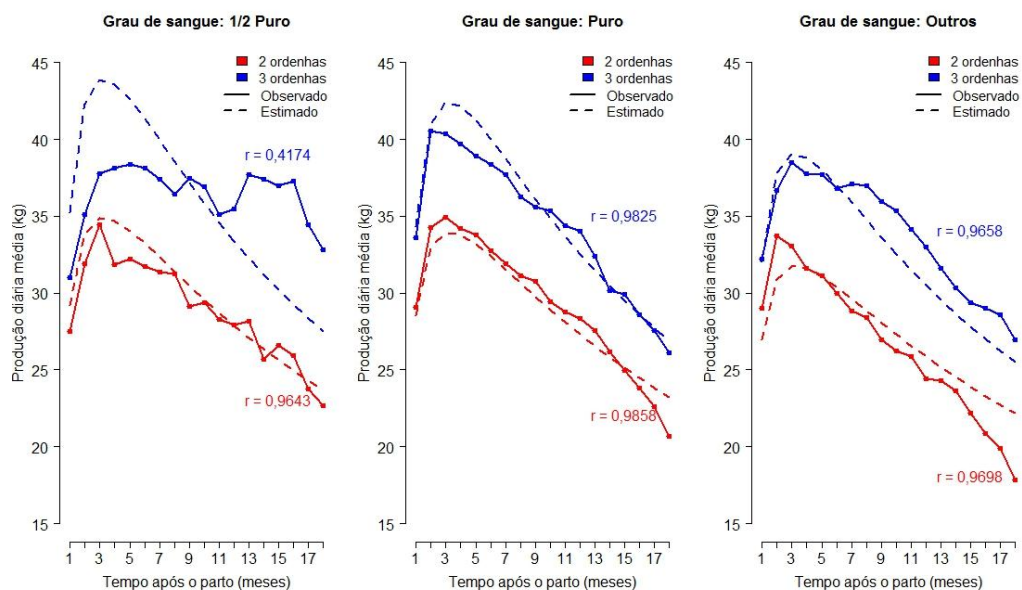


Figura 22 Modelo linear ajustado para a raça Holandesa, com as correlações entre as curvas observadas e estimadas.

A partir da Figura 22 observa-se que o ajuste do modelo para os animais meio-puros (gráfico à esquerda) não ficou muito bom, o que pode ser percebido pela baixa correlação entre os valores observados e estimados dos animais com três ordenhas. Já a correlação para as demais curvas foi alta (acima de 0,96). Optou-se, assim, por agrupar os animais meio-puros com os animais de outros graus de sangue. Estimativas associadas a esse modelo são mostradas na Tabela 18 e sua representação gráfica na Figura 24.

Tabela 18 Estimativas dos parâmetros do modelo linear ajustado para a raça Holandesa, com o grau de sangue agrupado.

Variável	Estimativa	Erro padrão	Valor t	P-valor
Intercepto	2,35e-02	3,25e-04	71,87	< 2,00e-16
Tempo	1,12e-03	2,29e-05	48,82	< 2,00e-16
1/Tempo	1,17e-02	5,09e-04	22,87	< 2,00e-16
Ordenhas – 3	-6,07e-03	1,58e-04	-38,39	< 2,00e-16
Sangue – puro	-1,25e-03	1,70e-04	-7,37	1,77e-13
Parâmetro de dispersão (ϕ)	0,0603	-	-	-

A análise de resíduos deste modelo é mostrada na Figura 23:

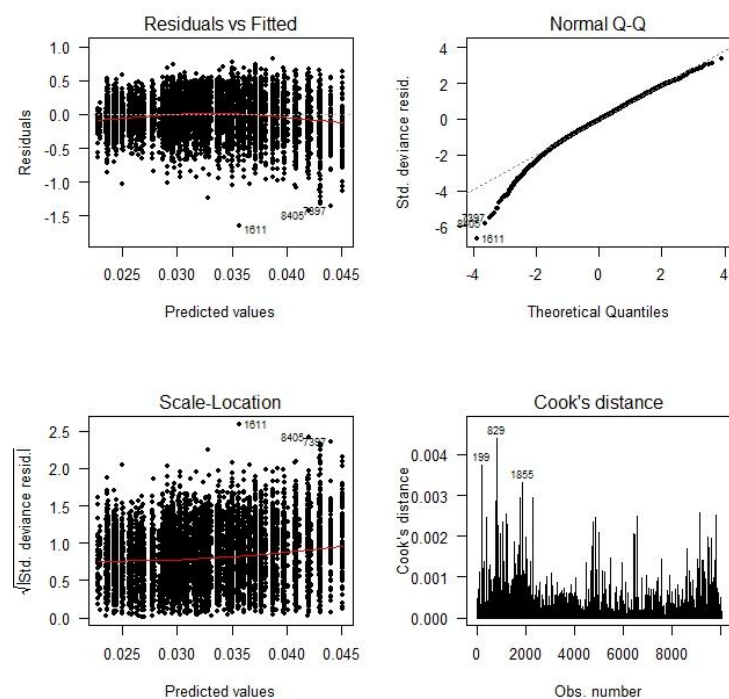


Figura 23 Análise dos resíduos *deviance* associados ao modelo linear ajustado para a raça Holandesa.

Embora seja possível observar uma pequena fuga na cauda inferior do gráfico de normalidade dos resíduos *deviance* (QQ-Plot), o modelo ajustado não apresentou sérias distorções quanto às suposições usualmente assumidas para um modelo linear (homocedasticidade e normalidade dos resíduos *deviance*). Também não são observados pontos influentes no ajuste do modelo, de acordo com a distância de Cook.

A interpretação das estimativas dos parâmetros das covariáveis pode ser feita por meio da razão entre a produção em um dado mês após o parto variando-se os níveis das covariáveis *dummy*. Os resultados dessas razões, utilizando-se o tempo de pico estimado, são mostrados na Tabela 19.

Tabela 19 Interpretação das estimativas dos parâmetros das covariáveis do modelo linear para a raça Holandesa.

Grau de sangue	Ordenhas	Efeito	Ordenhas	Grau de sangue	Efeito
Não puro	3 vs 2	1,2458	2	Puro vs Não puro	1,0424
Puro	3 vs 2	1,2589	3	Puro vs Não puro	1,0534

* Valores obtidos para o tempo de pico estimado (3,23).

Estima-se que quando são realizadas três ordenhas em animais não-puros, há um ganho na produção média no tempo de pico igual a 24,6% em relação ao manejo com duas

ordenhas. Este ganho é um pouco maior para os animais puros, com a produção média no pico aumentando aproximadamente 26%.

Os animais puros que foram ordenhados duas vezes tiveram produção máxima estimada em 4,2% maior do que os animais não-puros. Esse comportamento foi semelhante para os animais que tiveram três ordenhas, com ganho, nesse caso, pouco superior a 5%.

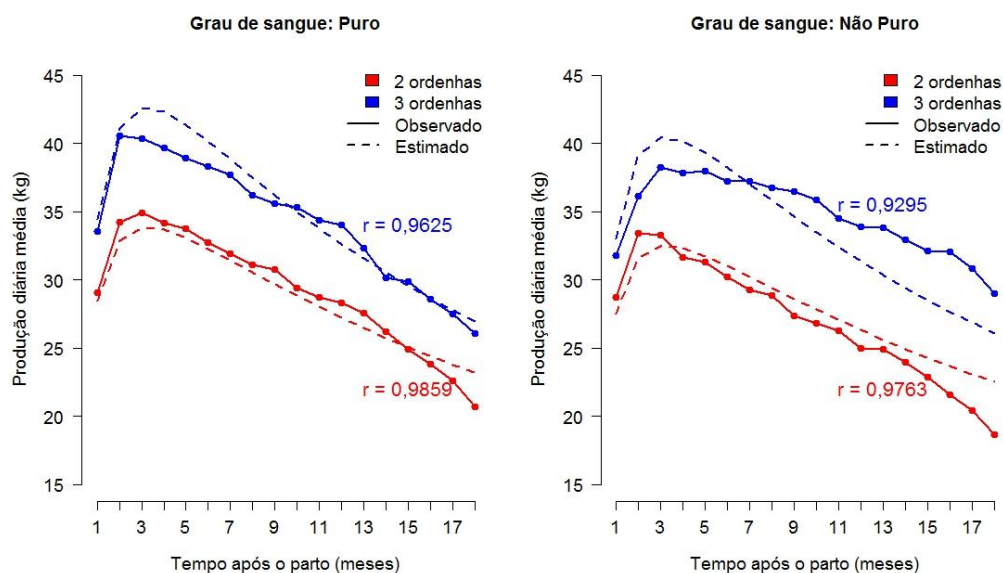


Figura 24 Modelo linear ajustado para a raça Holandesa com o grau de sangue agrupado, com as correlações entre as curvas observadas e estimadas.

Observa-se na Figura 24 que o agrupamento dos animais meio-puro e outros proporcionou um melhor ajuste do modelo, observado principalmente na correlação dos animais não-puros e de três ordenhas (0,93). Portanto esse modelo linear foi considerado como o mais adequado para a raça Holandesa, e com isso pode-se estimar o tempo e a produção no pico, derivando-se a equação do modelo em relação ao tempo. Os resultados obtidos são mostrados na Tabela 20, de onde se conclui que para os animais com duas ordenhas o modelo está subestimando a produção no pico, enquanto que para três ordenhas a produção está sendo superestimada, com um erro maior do que para duas ordenhas.

Tabela 20 Tempo e produção de leite no pico estimados pelo modelo linear ajustado para raça Holandesa.

Número de ordenhas	Grau de sangue	Tempo de pico estimado	Produção estimada	Produção observada	Diferença
2	Puro	3,23	33,86	34,95	-1,09
2	Não puro	3,23	32,48	33,41	-0,93
3	Puro	3,23	42,62	40,57	2,05
3	Não puro	3,23	40,46	38,26	2,20

Portanto a estimativa da produção de leite de um animal da raça Holandesa em um dado mês t após o parto é dada por:

$$\left\{ \begin{array}{l} \hat{y}(t \mid \text{puro, 2 ordenhas}) = \frac{1}{0,0235 + 0,0011t + 0,0117/t - 0,0013} \\ \hat{y}(t \mid \text{puro, 3 ordenhas}) = \frac{1}{0,0235 + 0,0011t + 0,0117/t - 0,0013 - 0,0061} \\ \hat{y}(t \mid \text{não puro, 2 ordenhas}) = \frac{1}{0,0235 + 0,0011t + 0,0117/t} \\ \hat{y}(t \mid \text{não puro, 3 ordenhas}) = \frac{1}{0,0235 + 0,0011t + 0,0117/t - 0,0061} \end{array} \right.$$

4.2.2. Raça Pardo Suíço

Nesta seção são mostrados os resultados do modelo linear ajustado para a raça Pardo Suíço, seguindo os mesmos passos da raça Holandesa.

Na Tabela 21 observa-se que o modelo considerando somente o tempo que tem maior correlação é o modelo 5. O modelo 4 foi desconsiderado porque o termo quadrático do tempo não foi significativo ao nível de 5%. Portanto, este modelo recai no modelo 3. As curvas observadas e estimadas para o modelo 5 são mostradas na Figura 25.

Tabela 21 Correlações lineares entre as curvas observadas e estimadas, para a raça Pardo Suíço.

Modelo	Independente	AR1	Permutável
1) $y(t) = a + bt + ct^2$	0,9037	0,8473	0,9037
2) $y(t) = a + bt + ct^2 + dt^3$	0,9344	0,9260	0,9344
3) $y(t) = a + bt + c \ln t$	0,9720	0,9699	0,9720
4) $y(t) = a + bt + ct^2 + d \ln t$	0,9748	0,9741	0,9748
5) $y(t) = a + bt + c/t$	0,9828	0,9803	0,9828

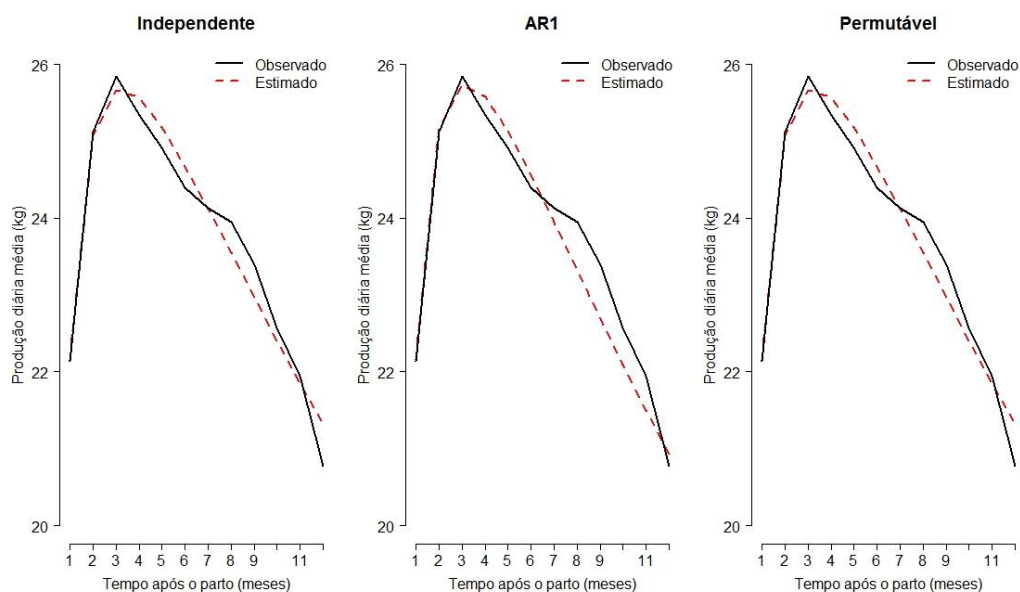


Figura 25 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Pardo Suíço.

O próximo passo é a escolha da matriz de correlação adequada para essa raça. Os resultados são mostrados na Tabela 22, de onde se conclui que estrutura permutável é a mais adequada, pois apresentou o menor *QIC*.

Tabela 22 Valores de *QIC* para a raça Pardo Suíço.

Matriz de trabalho	Variáveis ($\alpha = 0,10$)	<i>QIC</i>
Independente	Número de ordenhas, idade	3.396,1
AR1	Número de ordenhas, idade	3.425,2
Permutável	Número de ordenhas, idade	3.395,0

Diferentemente do que foi feito para a raça Holandesa, não foi possível ajustar um MLG, pois a matriz de trabalho escolhida não foi a independente. Ajustou-se, assim, um modelo utilizando Equações de Estimação Generalizadas.

A seleção de variáveis foi feita utilizando o *QICu*, e os resultados são mostrados na Tabela 23. Desta tabela se conclui que o modelo final deve considerar apenas o número de ordenhas.

Tabela 23 Resultados de *QICu* para a raça Pardo Suíço.

Modelo	Variáveis ($\alpha = 0,05$)	<i>QICu</i>
1	Número de ordenhas	6.205,8
2	Idade	6.208,4
3	Número de ordenhas, idade	6.208,4

As estimativas dos parâmetros do modelo selecionado são mostradas na Tabela 24.

Tabela 24 Estimativas dos parâmetros do modelo linear ajustado para a raça Pardo Suíço.

Variável	Estimativa	Erro padrão	Teste Wald	P-valor
Intercepto	0,032696	0,001613	410,68	< 2,0e-16
Tempo	0,001185	0,000158	56,30	6,2e-14
1/Tempo	0,012863	0,001486	74,96	< 2,0e-16
Ordenhas - 3	-0,004659	0,001676	7,73	0,0054
Parâmetro – permutável (α)	0,618	0,0516		
Parâmetro de dispersão (ϕ)	0,0815	0,00678		

A interpretação do parâmetro associado ao número de ordenhas é realizada utilizando a razão entre as produções no tempo de pico (estimado em 3,29). Essa razão foi igual a 1,13, o que significa que os animais da raça Pardo Suíço ordenhados três vezes têm produção no pico 13% maior do que aqueles ordenhados duas vezes.

O parâmetro da matriz de trabalho permutável significa que a correlação entre quaisquer tempos, em cada animal, é constante e, nesse caso, estimado ser igual a 0,618. A baixa estimativa do parâmetro de dispersão confirma o que foi constatado na análise descritiva, mostrando que há uma grande dispersão na variável resposta.

Os resíduos mostrados para o modelo da raça Holandesa (Figura 23) não podem ser obtidos para este modelo, uma vez que não se trata de um MLG. Contudo, a Figura 26 mostra os resíduos brutos distribuídos ao longo do tempo, podendo-se observar homogeneidade dos mesmos em torno do zero. Na Figura 27 são mostrados, para cada tempo, os resíduos brutos de cada animal, na qual não se verifica nenhum ponto discrepante e uma distribuição aleatória em torno de zero.

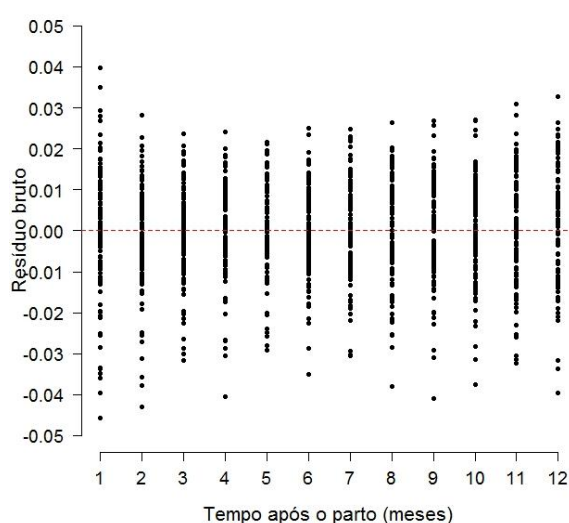


Figura 26 Resíduos brutos do modelo linear da raça Pardo Suíço versus o tempo.

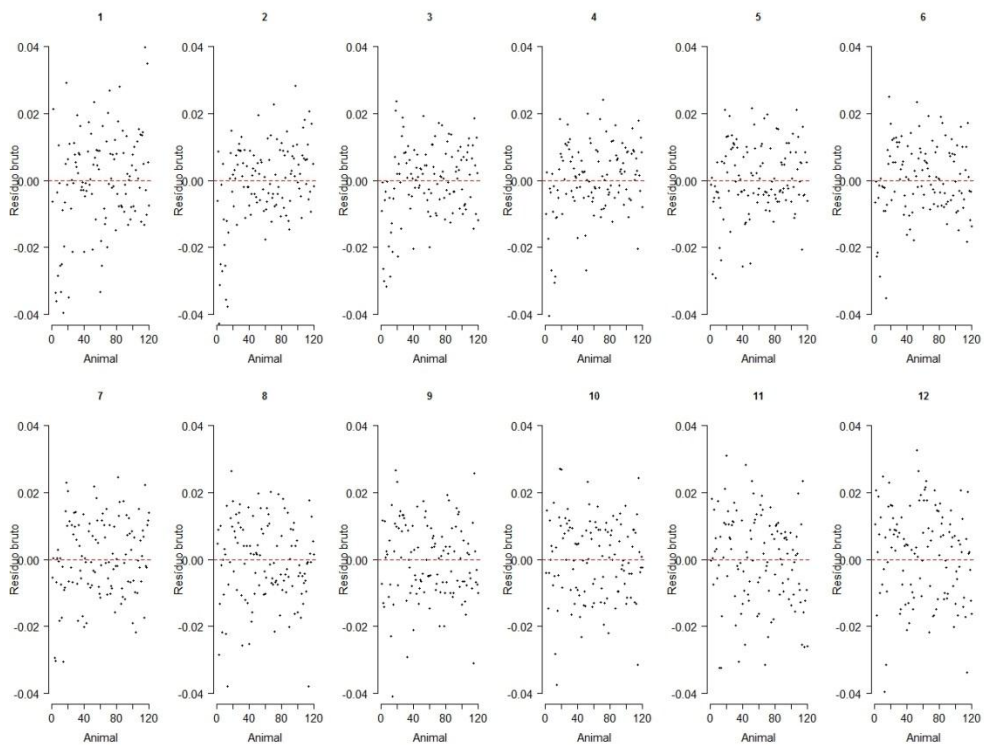


Figura 27 Resíduos brutos do modelo linear da raça Pardo Suíço, para cada animal em cada tempo.

A Figura 28 mostra o ajuste do modelo aos dados dos animais, comparando as curvas observadas e estimadas. Observa-se um ajuste melhor para os animais que tiveram duas ordenhas, o qual teve uma correlação de 0,94 entre a curva observada e a estimada, enquanto que essa correlação foi de apenas 0,72 para os animais com três ordenhas, mostrando uma subestimação. Isso é justificado pelo número menor de animais nessa categoria (39, de acordo com a Tabela 9).

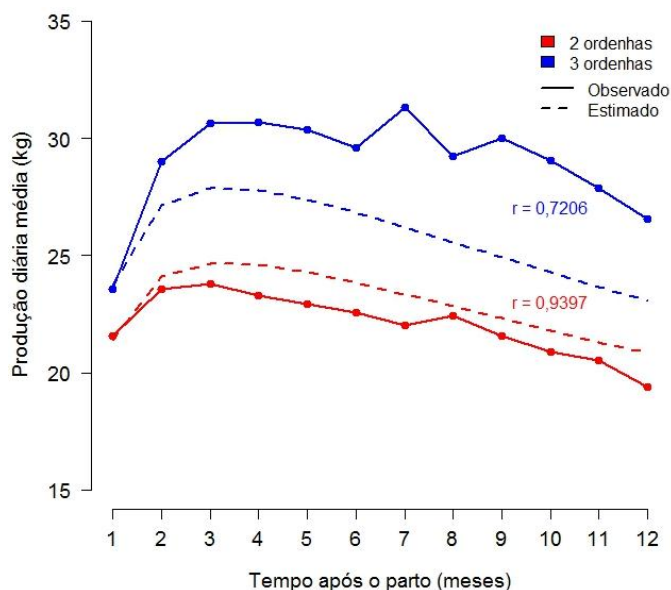


Figura 28 Modelo linear ajustado para a raça Pardo Suíço com as correlações entre as curvas observadas e estimadas.

Foi estimado também o tempo de pico e a produção neste tempo (Tabela 25). O erro foi maior para os animais com três ordenhas, em que a produção no pico (estimado) foi subestimada em 2,7 kg. Já os animais com duas ordenhas tiveram essa produção superestimada em menos de 1 kg.

Tabela 25 Tempo e produção de leite no pico estimados pelo modelo linear ajustado pela raça Pardo Suíço.

Número de ordenhas	Tempo de pico estimado	Produção estimada	Produção observada	Diferença
2	3,29	24,7	23,8	0,90
3	3,29	27,9	30,6	-2,72

Portanto, a estimativa da produção de leite de um animal da raça Pardo Suíço em um dado mês t após o parto é dada por:

$$\begin{cases} \hat{y}(t|2 \text{ ordenhas}) = \frac{1}{0,0327 + 0,0012t + 0,0129/t} \\ \hat{y}(t|3 \text{ ordenhas}) = \frac{1}{0,0327 + 0,0012t + 0,0129/t - 0,0047} \end{cases}$$

4.2.3. Raça Jersey

Nesta seção são mostrados os resultados do modelo linear ajustado para a raça Jersey, seguindo os mesmos passos das raças Holandesa e Pardo Suíço.

Na Tabela 26 observa-se que o modelo considerando somente o tempo que tem maior correlação é a do modelo 4. Contudo, assim como nas raças anteriores, o termo quadrático não foi significativo. Logo a estrutura selecionada para o tempo foi a do modelo 3, cujas curvas observadas e estimadas para esse modelo são mostradas na Figura 29.

Tabela 26 Correlações lineares entre as curvas observadas e estimadas, para a raça Jersey.

Modelo	Independente	AR1	Permutável
1) $y(t) = a + bt + ct^2$	0,9708	0,9592	0,9708
2) $y(t) = a + bt + ct^2 + dt^3$	0,9766	0,9729	0,9766
3) $y(t) = a + bt + c \ln t$	0,9850	0,9847	0,9849
4) $y(t) = a + bt + ct^2 + d \ln t$	0,9865	0,9861	0,9865
5) $y(t) = a + bt + c/t$	0,9759	0,9707	0,9759

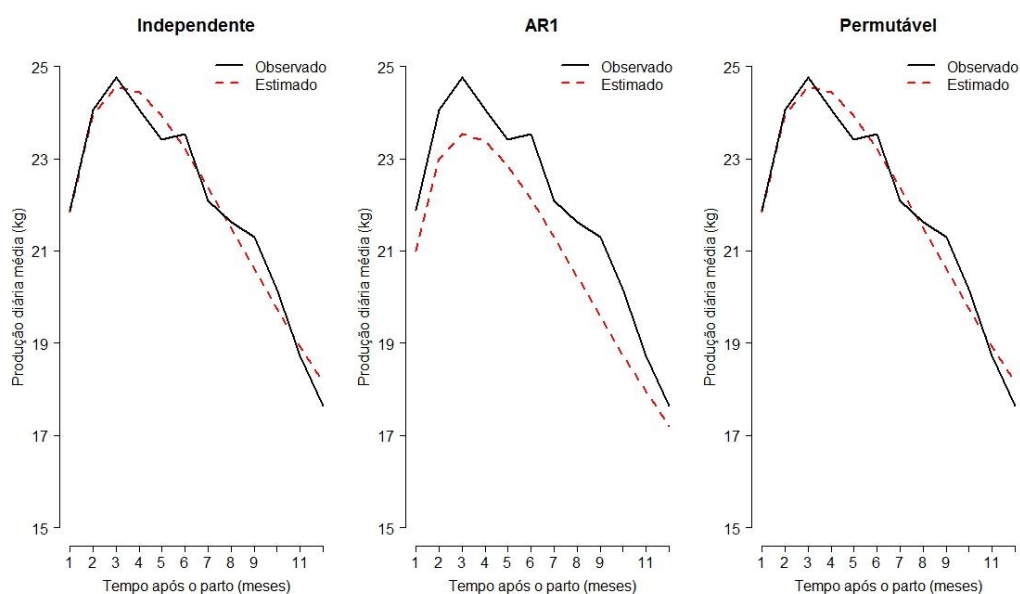


Figura 29 Curvas observadas e ajustadas do modelo considerando apenas o tempo, para a raça Jersey.

Observa-se que o modelo adotado ajusta-se bem quando se utiliza a matriz de trabalho permutável ou a independente, mas não para a estrutura AR1.

Os resultados do *QIC* são mostrados na Tabela 27, de onde se conclui que a permutável deve ser a matriz de trabalho escolhida. Como apenas a variável idade foi significativa nessa estrutura, não é necessário avaliar o *QICu* para a seleção de variáveis.

Tabela 27 Valores de *QIC* para a raça Jersey.

Matriz de trabalho	Variáveis ($\alpha = 0, 10$)	<i>QIC</i>
Independente	Número de ordenhas, idade, grau de sangue	2.195,0
AR1	Número de ordenhas, idade	2.185,1
Permutável	Idade	2.152,3

As estimativas dos parâmetros para o modelo escolhido são mostradas na Tabela 28. A interpretação da estimativa do parâmetro da idade é mostrada na Tabela 29, na qual foram comparadas a produções no pico de animais com diferentes idades. Observa-se que quanto maior a diferença de idade maior é o efeito na produção de leite no pico (estimado), e que animais mais velhos têm um efeito maior quando se compara a mesma diferença de idade.

Estima-se que a correlação entre quaisquer meses é de 0,53, o que é uma correlação intermediária. Para esses animais também há uma grande dispersão na produção de leite, de acordo com a estimativa do parâmetro (ϕ).

Tabela 28 Estimativas dos parâmetros do modelo linear ajustado para a raça Jersey.

Variável	Estimativa	Erro padrão	Teste Wald	P-valor
Intercepto	0,052450	0,001980	701,48	<2e-16
Tempo	0,003187	0,000372	73,51	<2e-16
ln(Tempo)	-0,010370	0,001415	53,68	2,36e-13
Idade	-9,812e-06	1,471e-06	44,51	2,53e-11
Parâmetro – permutável (α)	0,5347	0,050980	-	-
Parâmetro de dispersão (ϕ)	0,06635	0,006519	-	-

Tabela 29 Interpretação da estimativa do parâmetro da idade no modelo linear para a raça Jersey.

Diferença	Idades comparadas	Efeito na produção no pico
6 meses	2 anos vs 1,5 ano	4,1%
	5 anos vs 4,5 anos	5,5%
1 ano	3 anos vs 2 anos	9,0%
	6 anos vs 5 anos	12,3%
5 anos	6,5 anos vs 1,5 ano	65,6%
	7 anos vs 2 anos	70,0%

Na Figura 30 tem-se a distribuição dos resíduos brutos ao longo do tempo, onde se observa que o erro é maior para tempos maiores, embora esse erro não seja grande a ponto de afetar a estimação. Análise de resíduos apresentados na Figura 31 mostra que à medida

que o tempo avança há um aumento na dispersão dos resíduos, mas sem perda da homocedasticidade.

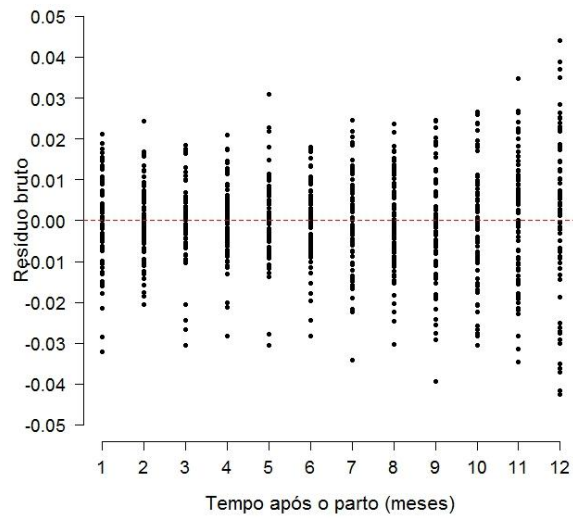


Figura 30 Resíduos brutos do modelo linear da raça Jersey versus o tempo.

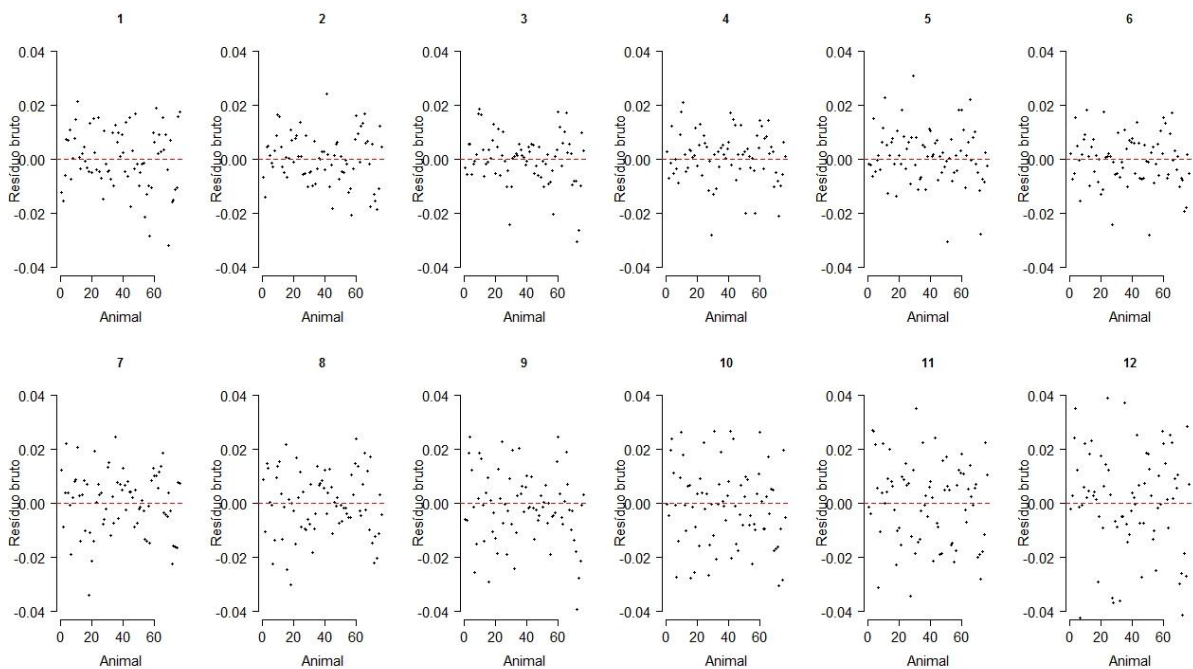


Figura 31 Resíduos brutos do modelo linear da raça Jersey, para cada animal em cada tempo.

O modelo ajustado, considerando a idade média dos animais, é mostrado na Figura 32. O tempo de pico foi estimado em 3,25 meses, e a produção nesse pico (considerando-se a idade média) foi estimada em 24,38 kg.

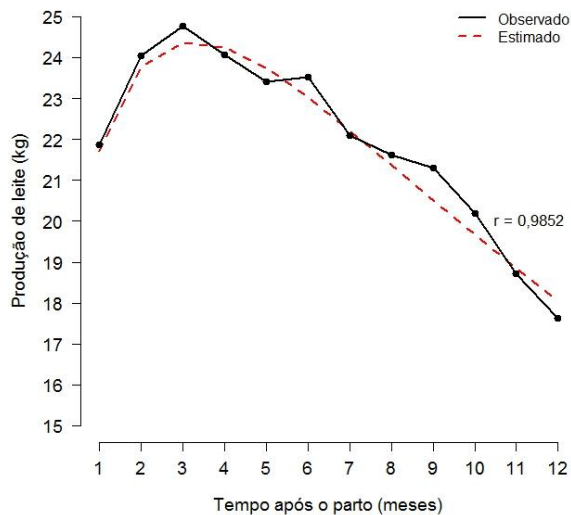


Figura 32 Modelo linear ajustado para a raça Jersey, com a correlação entre a curva observada e estimada.

Para um animal da raça Jersey, estimativa de sua produção de leite em um dado mês t após o parto pode, portanto, ser obtida por:

$$\hat{y}(t|idade) = \frac{1}{0,0524 + 0,00319t - 0,0104 \ln t - 9,81 \times 10^{-6} \cdot idade}$$

4.3. MODELOS NÃO-LINEARES AJUSTADOS

4.3.1. Raça Holandesa

Conforme descrito na metodologia utilizada para modelos não-lineares, o primeiro passo foi a escolha do modelo considerando o tempo. Ou seja, dentre os modelos intrinsecamente lineares e não-lineares mostrados na Tabela 30, escolheu-se aquele que melhor se ajustou aos dados, utilizando-se, para isso, os três critérios citados.

Tabela 30 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Holandesa.

Modelo	Tipo	Função	AIC	l	r
1	INLIN	$y(t) = a \exp(-ct)$	70.442,9	-35.218,4	0,8801
2	INLIN	$y(t) = at^b \exp(-ct)$	69.935,1	-34.963,6	0,9882
3	NLIN	$y(t) = a \exp(bt - ct^2)$	70.071,2	-35.031,6	0,9610
4	NLIN	$y(t) = at^{bc} \exp(-ct)$	69.935,1	-34.963,6	0,9882
5	NLIN	$y(t) = \frac{a}{t} \exp(-ct)$	81.761,1	-40.877,5	-0,3555
6	NLIN	$y(t) = at \exp(-ct)$	74.275,6	-37.134,8	0,7097
7	NLIN	$y(t) = at^{bc} \exp(-ct)$	70.369,8	-35.180,9	0,8970

A partir da Tabela 30 conclui-se que as formulações que têm menor *AIC*, maior *l* e *r* são os modelos 2 e 4, de Wood (1967) e Dhanoa (1981), respectivamente. Por ser mais utilizado na literatura e ter uma expressão mais simples o modelo de Wood foi adotado.

O passo seguinte foi a seleção das covariáveis significativas, utilizando novamente o *AIC* e a *l*. Assim como feito no modelo linear da raça Holandesa, os animais meio-puros e outros foram agrupados como não-puros; também foi desconsiderado o tamanho do produtor devido ao tamanho de amostra, que foi verificado apresentar poucos animais em alguns grupos. Os resultados são mostrados na Tabela 31.

Tabela 31 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Holandesa.

Modelo	Covariáveis ($\alpha = 0,05$)	<i>AIC</i>	<i>l</i>
1	Sem covariáveis	69.935,1	-34.963,6
2	Número de ordenhas	68.491,4	-34.240,7
3	Número de ordenhas, idade	68.476,6	-34.232,3
4	Número de ordenhas, idade, grau de sangue	68.416,9	-34.201,4
5	Número de ordenhas, grau de sangue	68.434,8	-34.211,4

Nota-se que o modelo 4 foi o que apresentou o menor *AIC* e a maior *l*, mas, assim como feito no modelo linear dessa raça, decidiu-se estudar se a covariável idade é realmente importante no modelo, ou se sua significância é decorrente do grande tamanho de amostra. Comparando as curvas estimadas de dois modelos, com e sem a covariável idade, tem-se os resultados da Tabela 32. Observa-se que a diferença entre as curvas estimadas pelos dois modelos é, em média, próxima de zero para todas as combinações das covariáveis *dummy*. Ainda, na soma de todos os tempos, a diferença não chega a 0,5 kg entre os dois modelos. Logo, conclui-se que a covariável idade não exerce grande influência no modelo e pode ser retirada.

Tabela 32 Avaliação do impacto da idade nas curvas estimadas de dois modelos não-lineares da raça Holandesa.

2 ordenhas	Soma	Média	3 ordenhas	Soma	Média
Puro	0,2371	0,0132	Puro	0,1036	0,0058
Não-puro	-0,3275	-0,0182	Não-puro	-0,4611	-0,0256

* Utilizou-se a idade média no modelo que continha a idade.

Tabela 33 Estimativas dos parâmetros do modelo não-linear da raça Holandesa.

Covariável	Estimativa	Erro padrão	Valor t	P-valor
<i>a</i>	24,01951	0,34888	68,847	< 2e-16
<i>b</i>	0,28275	0,01254	22,548	< 2e-16
<i>c</i>	0,07322	0,00212	34,523	< 2e-16
Ordenhas – 3	6,08523	0,15171	40,110	< 2e-16
Grau de sangue – puro	1,20756	0,15752	7,666	1,94e-14

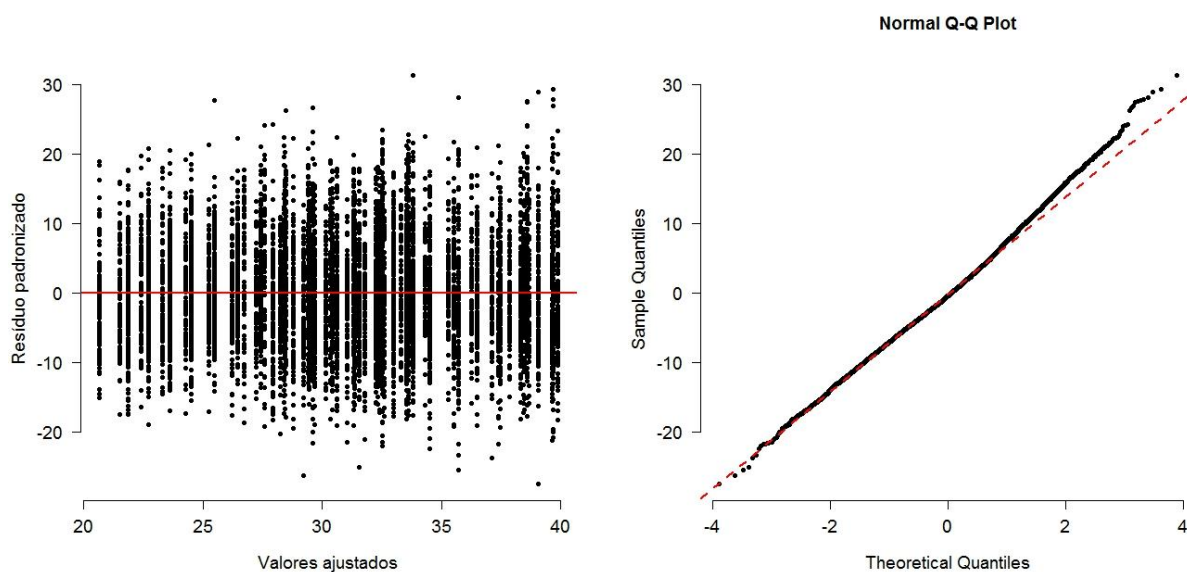


Figura 33 Análise de resíduos do modelo não-linear para a raça Holandesa

A análise de resíduos (Figura 33) mostra que estes estão distribuídos aleatoriamente em torno de zero, sem padrões de comportamento e sem sérios desvios da normalidade.

Conforme descrito na metodologia, o tempo de pico é estimado em 3,86 meses. As produções estimadas para animais puros de duas e três ordenhas são 33,82 e 39,90, respectivamente. Analogamente, as produções estimadas para animais não-puros de duas e três ordenhas são 32,61 e 38,70, respectivamente. Além disso, a persistência é estimada em 3,35.

A interpretação das estimativas dos parâmetros segue a mesma linha daquela feita para o modelo linear, isto é, em função da razão entre as produções no pico para cada nível das covariáveis *dummy* (Tabela 34). Deste modo, o modelo estima que, em média, animais não-puros que foram ordenhados três vezes produzem 18,7% a mais no pico do que aqueles ordenhados duas vezes, e que animais puros com três ordenhas produzem 18% a mais do que aqueles com duas ordenhas. Esse efeito é menor quando se comparam os animais puros versus não-puros, para duas ou três ordenhas: de 3% a 4% a mais no pico de produção.

Tabela 34 Interpretação das estimativas dos parâmetros das covariáveis do modelo não-linear para a raça Holandesa.

Grau de sangue	Ordenhas	Efeito	Ordenhas	Grau de sangue	Efeito
Não-puro	3 vs 2	1,1866	2	Puro vs Não-puro	1,0370
Puro	3 vs 2	1,1799	3	Puro vs Não-puro	1,0312

* Valores obtidos para o tempo de pico estimado (3,86).

A Figura 34 mostra as curvas estimadas pelo modelo não-linear e as curvas observadas para os animais da raça Holandesa. Observa-se um bom ajuste com altas correlações entre as curvas, todas acima de 0,90.

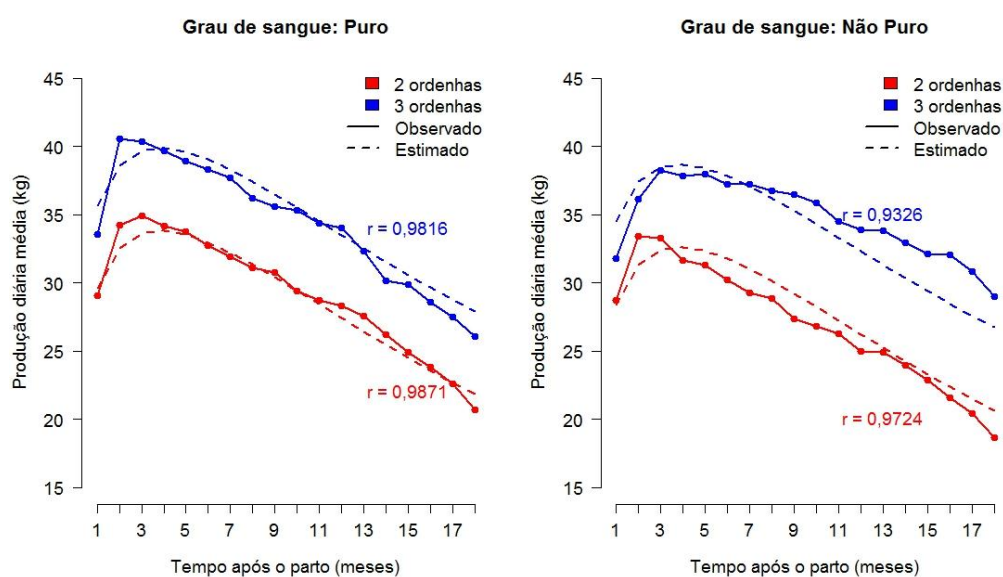


Figura 34 Modelo não-linear ajustado para a raça Holandesa com o grau de sangue agrupado, com as correlações entre as curvas observadas e estimadas.

Na Tabela 35 têm-se as produções estimadas e observadas para cada grupo de animais. Observa-se que as diferenças são menores do que aquelas obtidas no modelo linear da raça Holandesa (Tabela 20).

Tabela 35 Tempo e produção de leite no pico estimados pelo modelo não-linear ajustado para raça Holandesa.

Número de ordenhas	Grau de sangue	Tempo de pico estimado	Produção estimada	Produção observada	Diferença
2	Puro	3,86	33,82	34,95	1,13
2	Não puro	3,86	32,61	33,41	0,80
3	Puro	3,86	39,90	40,57	0,67
3	Não puro	3,86	38,70	38,26	-0,44

Concluindo, o modelo não-linear ajustado para a raça Holandesa é dado por:

$$\begin{cases} \hat{y}(t| \text{puro, 2 ordenhas}) = 24,01951 \cdot t^{0,28275} e^{-0,07322 \cdot t} + 1,20756 \\ \hat{y}(t| \text{puro, 3 ordenhas}) = 24,01951 \cdot t^{0,28275} e^{-0,07322 \cdot t} + 1,20756 + 6,08523 \\ \hat{y}(t| \text{não puro, 2 ordenhas}) = 24,01951 \cdot t^{0,28275} e^{-0,07322 \cdot t} \\ \hat{y}(t| \text{não puro, 3 ordenhas}) = 24,01951 \cdot t^{0,28275} e^{-0,07322 \cdot t} + 6,08523 \end{cases}$$

4.3.2. Raça Pardo Suíço

Como feito na raça Holandesa, foram testados os modelos intrinsecamente lineares e não-lineares nos dados da raça Pardo Suíço e, novamente, foi selecionado o modelo de Wood conforme os resultados da Tabela 36.

Tabela 36 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Pardo Suíço.

Modelo	Tipo	Função	AIC	<i>l</i>	<i>r</i>
1	INLIN	$y(t) = a \exp(-ct)$	9.799,6	-4.896,8	0,6282
2	INLIN	$y(t) = at^b \exp(-ct)$	9.767,3	-4.879,6	0,9734
3	NLIN	$y(t) = a \exp(bt - ct^2)$	9.775,7	-4.883,9	0,9013
4	NLIN	$y(t) = at^{bc} \exp(-ct)$	9.767,3	-4.879,6	0,9734
5	NLIN	$y(t) = \frac{a}{t} \exp(-ct)$	10.763,0	-5.378,5	-0,7659
6	NLIN	$y(t) = at \exp(-ct)$	10.124,0	-5.059,0	0,7743
7	NLIN	$y(t) = at^{bc} \exp(-ct)$	9.771,7	-4.881,8	0,9365

Testando os efeitos das covariáveis têm-se os resultados mostrados na Tabela 37. Assim como na raça Holandesa, o grau de sangue foi agrupado em puro e não-puro, mas observa-se que essa covariável não traz um grande ganho no AIC e na *l*, o que justifica a sua retirada do modelo. Logo, o modelo escolhido tem as covariáveis número de ordenhas e idade.

Tabela 37 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Pardo Suíço.

Modelo	Covariáveis ($\alpha = 0,05$)	AIC	<i>l</i>
1	Sem covariáveis	9.505,9	-4.748,0
2	Número de ordenhas	9.440,6	-4.714,3
3	Número de ordenhas, idade	9.438,1	-4.712,0
4	Número de ordenhas, idade, grau de sangue	9.479,1	-4.733,6
5	Número de ordenhas, grau de sangue	9.505,9	-4.748,0

De modo similar ao que foi feito para a raça Holandesa, estudou-se o impacto da inclusão da covariável idade, fazendo-se as diferenças entre um modelo com e sem essa covariável em cada tempo. De acordo com a Tabela 38, observa-se que há uma diferença relativamente grande entre os modelos, tanto para duas quanto três ordenhas, principalmente na soma das produções estimadas. Portanto, a idade é realmente importante na predição da produção de leite.

Tabela 38 Avaliação do impacto da idade nas curvas estimadas de dois modelos não-lineares da raça Pardo Suíço.

Ordenhas	Soma	Média
2	2,31	0,13
3	-4,67	-0,26

** Utilizou-se a idade média no modelo que continha a idade.*

As estimativas dos parâmetros do modelo selecionado são mostradas na Tabela 39.

Tabela 39 Estimativas dos parâmetros do modelo não-linear da raça Pardo Suíço.

Covariável	Estimativa	Erro padrão	Valor t	P-valor
<i>a</i>	12,31000	0,76060	16,185	< 2e-16
<i>b</i>	0,36390	0,06247	5,826	7,00e-09
<i>c</i>	0,08931	0,01315	6,790	1,63e-11
Ordenhas – 3	6,45500	0,39670	16,270	< 2e-16
Idade	0,00272	0,00033	8,287	2,64e-16

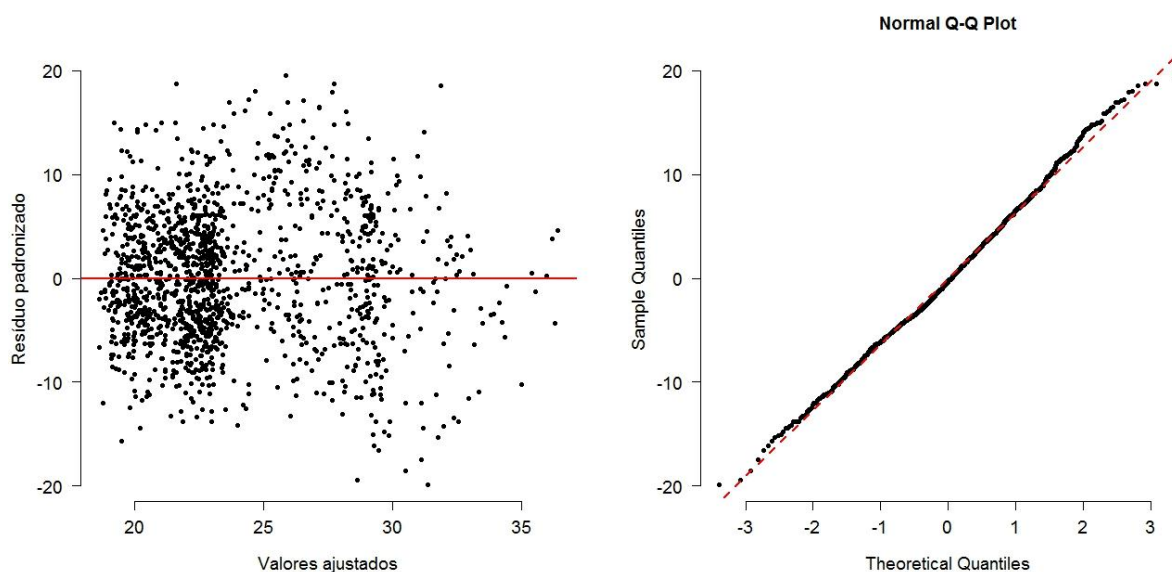


Figura 35 Análise de resíduos do modelo não-linear para a raça Pardo Suíço.

Na análise de resíduos apresentada na Figura 35 não se observam violações no que se refere às suposições de homocedasticidade e normalidade dos resíduos.

O tempo de pico foi estimado em 4,1 meses, como descrito na metodologia. A produção no pico para duas ordenhas, fixando-se a idade média, foi estimada em 23,7 kg, enquanto que para três ordenhas foi estimada em 30,2 kg. Isso significa que animais com a idade média (1.095 dias) que foram ordenhados três vezes têm produção máxima 27% maior do que aqueles ordenhados duas vezes. A persistência foi estimada em 3,29.

Tabela 40 Interpretação da estimativa do parâmetro da idade no modelo não-linear para a raça Pardo Suíço.

Diferença	Idades comparadas	Efeito na produção no pico	
		2 ordenhas	3 ordenhas
6 meses	2,5 anos vs 2 anos	2,2%	1,7%
	5 anos vs 4,5 anos	2,0%	1,6%
1 ano	3 anos vs 2 anos	4,4%	3,4%
	6 anos vs 5 anos	3,9%	3,1%
5 anos	7 anos vs 2 anos	21,9%	17,0%
	8 anos vs 3 anos	21,0%	16,5%

Na Tabela 40 observa-se que a idade tem um efeito maior em animais de duas ordenhas, mas esse efeito é menor em animais mais velhos. Ainda, quanto maior a diferença de idade entre os animais, maior é o efeito na produção de leite.

O ajuste do modelo é mostrado na Figura 36, considerando a idade média dos animais:

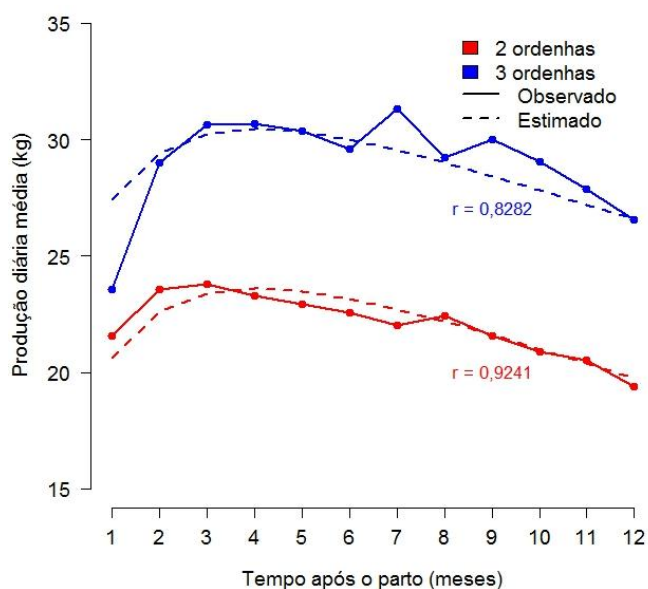


Figura 36 Modelo não-linear ajustado para a raça Pardo Suíço, com as correlações entre as curvas observadas e estimadas.

Estimativa da produção de leite de um animal da raça Pardo Suíço em um dado mês t após o parto pode, portanto, ser obtida por:

$$\begin{cases} \hat{y}(t| \text{idade}, 2 \text{ ordenhas}) = 12,31 \cdot t^{0,3639} e^{-0,08931 \cdot t} + 0,00272 \cdot \text{idade} \\ \hat{y}(t| \text{idade}, 3 \text{ ordenhas}) = 12,31 \cdot t^{0,3639} e^{-0,08931 \cdot t} + 0,00272 \cdot \text{idade} + 6,455 \end{cases}$$

4.3.3. Raça Jersey

Na Tabela 41 são mostrados os resultados do ajuste dos modelos propostos na literatura aos dados da raça Jersey. Dela se conclui, novamente, que o modelo mais bem ajustado é o de Wood (modelo 2).

Tabela 41 Estatísticas de ajuste dos modelos intrinsecamente lineares e não-lineares aos dados de animais da raça Jersey.

Modelo	Tipo	Função	AIC	l	r
1	INLIN	$y(t) = a \exp(-ct)$	5.835,5	-2.914,8	0,8197
2	INLIN	$y(t) = at^b \exp(-ct)$	5.800,9	-2.896,5	0,9879
3	NLIN	$y(t) = a \exp(bt - ct^2)$	5.805,7	-2.898,8	0,9682
4	NLIN	$y(t) = at^{bc} \exp(-ct)$	5.800,9	-2.896,5	0,9879
5	NLIN	$y(t) = \frac{a}{t} \exp(-ct)$	6.598,6	-3.296,3	-0,4395
6	NLIN	$y(t) = at \exp(-ct)$	6.083,9	-3.038,9	0,7494
7	NLIN	$y(t) = at^{bc} \exp(-ct)$	5.831,9	-2.911,9	0,8481

Na Tabela 42 são mostrados os valores de AIC e l para o modelo de Wood com as covariáveis. O modelo 3 foi o selecionado, uma vez que os modelos 5 e 6 tiveram ajustes ruins quando foram comparadas as curvas observadas e estimadas (Figura 37). As estimativas dos parâmetros são mostradas na Tabela 43.

Tabela 42 Estatísticas de ajuste do modelo de Wood com covariáveis para os animais da raça Jersey.

Modelo	Covariáveis ($\alpha = 0,05$)	AIC	l
1	Sem covariáveis	5.800,9	-2.896,5
2	Número de ordenhas	5.799,8	-2.894,9
3	Idade	5.724,4	-2.857,2
4	Número de ordenhas, idade	5.726,4	-2.857,2
5	Idade, grau de sangue	5.715,1	-2.851,6
6	Número de ordenhas, idade, grau de sangue	5.703,0	-2.844,5

Apesar de os modelos 5 e 6 terem apresentado menor *AIC* e maior *l*, observa-se que estes apresentam um ajuste pobre em relação às curvas observadas. As curvas estimadas do modelo 5, para animais puros versus não-puros, estão muito próximas; no modelo 6, para animais não-puros, há uma falta de ajuste devido ao fato de haver apenas um animal com três ordenhas.

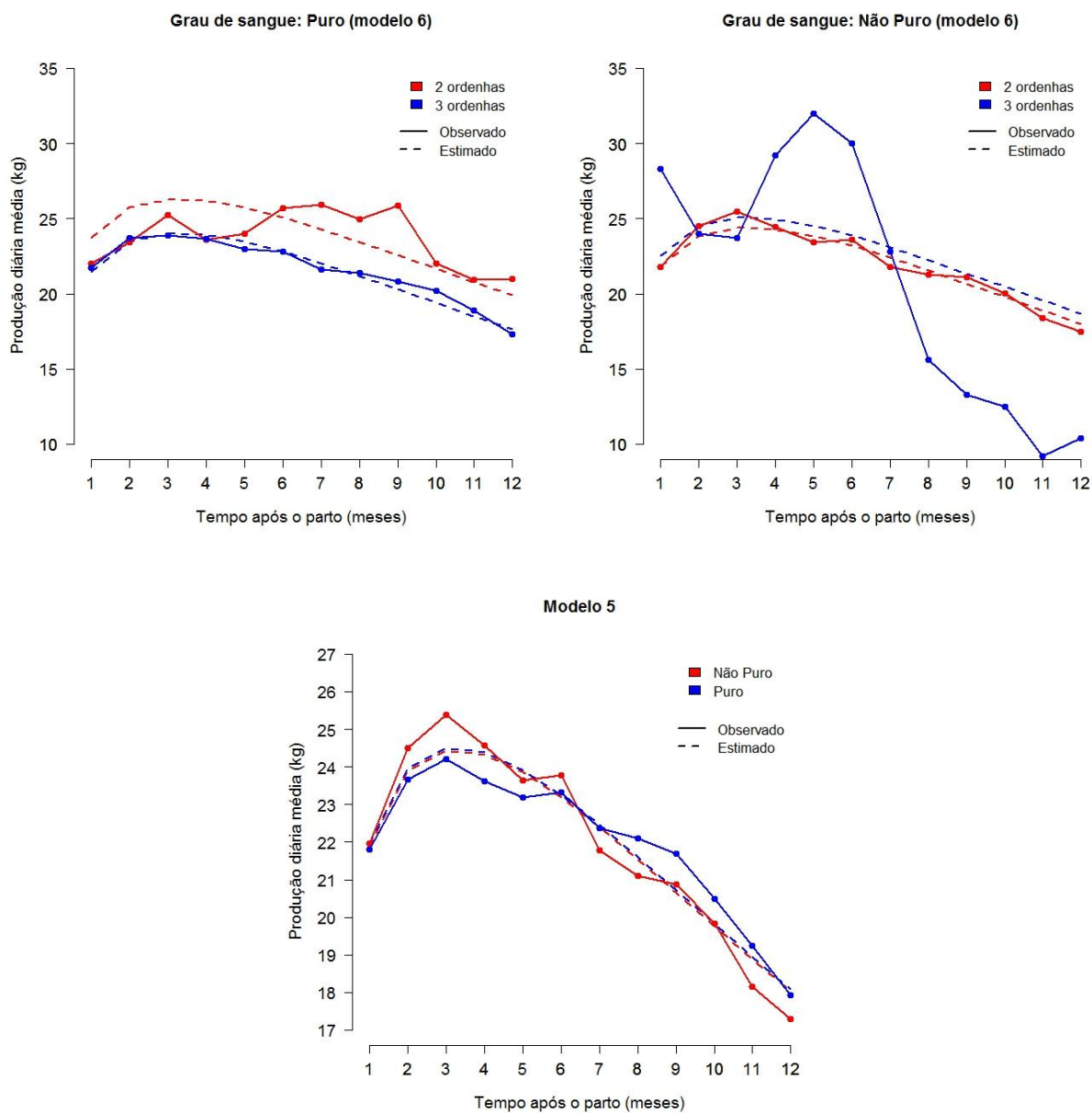


Figura 37 Curvas estimadas e observadas de modelos não-lineares rejeitados da raça Jersey.

Tabela 43 Estimativas dos parâmetros do modelo não-linear da raça Jersey.

Covariável	Estimativa	Erro padrão	Valor t	P-valor
<i>a</i>	19,47000	0,68580	28,393	< 2e-16
<i>b</i>	0,27290	0,04477	6,096	1,6e-09
<i>c</i>	0,08398	0,00981	8,560	< 2e-16
Idade	0,00414	0,00046	9,034	< 2e-16

Na Figura 38 observa-se que não há grandes desvios nos resíduos, quanto à homocedasticidade e normalidade.

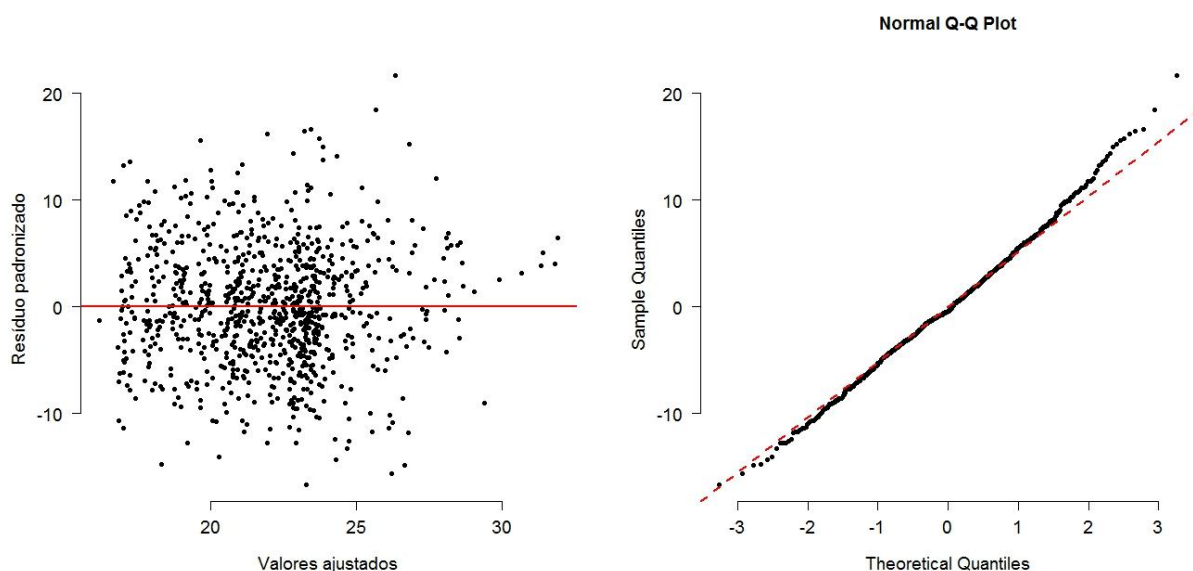


Figura 38 Análise de resíduos do modelo não-linear para a raça Jersey.

O tempo de pico foi estimado em 3,25 meses, e a persistência em 3,15. Para um animal com a idade média (975 dias) a produção nesse pico foi estimada em 24,5 kg.

A interpretação da estimativa do parâmetro idade é feita da mesma forma como no modelo linear para a raça Jersey (Tabela 44). Quanto maior a diferença de idade, maior é o efeito na produção de leite.

Tabela 44 Interpretação da estimativa do parâmetro da idade no modelo não-linear para a raça Jersey.

Diferença	Idades comparadas	Efeito na produção no pico
6 meses	2 anos vs 1,5 ano	3,3%
	5 anos vs 4,5 anos	2,8%
1 ano	3 anos vs 2 anos	6,4%
	6 anos vs 5 anos	5,4%
5 anos	6,5 anos vs 1,5 ano	33,3%
	7 anos vs 2 anos	32,2%

O ajuste do modelo é mostrado na Figura 39, considerando a idade média dos animais.

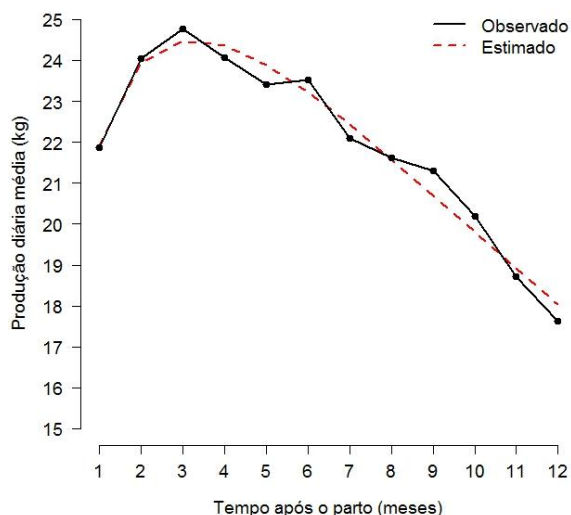


Figura 39 Modelo não-linear ajustado para a raça Jersey, com a correlação entre a curva observada e estimada.

Portanto o modelo não-linear ajustado para os animais da raça Jersey fica, desse modo, expresso por:

$$\hat{y}(t| \text{idade}) = 19,47 \cdot t^{0,2729} e^{-0,08398 \cdot t} + 0,00414 \cdot \text{idade}$$

5. CONSIDERAÇÕES FINAIS

Os modelos lineares utilizados apresentaram, de forma geral, um bom ajuste à produção de leite das três raças utilizadas neste trabalho. Em tais modelos estruturas de correlação (matrizes de trabalho) diferentes foram indicadas pelos procedimentos de seleção utilizados. Para a raça Holandesa, que representa a maioria do rebanho de Castro, foi indicada a estrutura de correlação independente, o que significa que as observações dos animais ao longo do tempo foram consideradas não correlacionadas. Isso pode ser devido ao fato de a amostra conter muitos animais; assim, a precisão das estimativas com essa estrutura de correlação já é satisfatória, o que justifica a modelagem utilizando Modelos Lineares Generalizados. A distribuição considerada para a produção de leite diária foi a Gama, sendo que no modelo foi utilizada a função de ligação inversa.

As raças Pardo Suíço e Jersey tiveram a matriz Permutável incluída no modelo linear, que foi estimado via Equação de Estimação Generalizadas. Nesse caso não foi considerada uma distribuição para a produção de leite diária, mas a função de ligação inversa também foi utilizada.

Nessas duas metodologias utilizadas nos modelos lineares foi possível a inclusão de covariáveis para explicar a produção diária média. Entretanto, as covariáveis não foram as mesmas para as três raças.

Dentre os modelos não-lineares sugeridos na literatura, aquele que proporcionou melhor ajuste foi o de Wood (1967), independentemente da raça. Para esses modelos foi assumida distribuição Normal para a produção diária de leite. Uma diferença importante nos modelos não-lineares é a ausência de uma matriz de trabalho, devido a restrições do pacote *nlme* do R. Assim como nos modelos lineares foi possível a inclusão de covariáveis nos modelos, sendo que as covariáveis selecionadas só foram diferentes para a raça Pardo Suíço, que incluiu a idade do animal no parto no modelo não-linear.

Na seleção dos modelos não-lineares foram utilizadas as metodologias mais empregadas na literatura, como *AIC* e valor da verossimilhança, tornando-a mais simples do que a seleção utilizada nos modelos lineares. Nestes, os critérios adotados, como *QIC* e *QICu*, são menos abordados na literatura, o que dificultou a seleção destes modelos. Além disso, quando se utiliza estimação via EEG não há uma literatura formal para análise de resíduos.

Na Figura 40 são mostrados os ajustes dos modelos lineares e não-lineares da raça Holandesa, sendo que os modelos não-lineares tiveram um ajuste melhor.

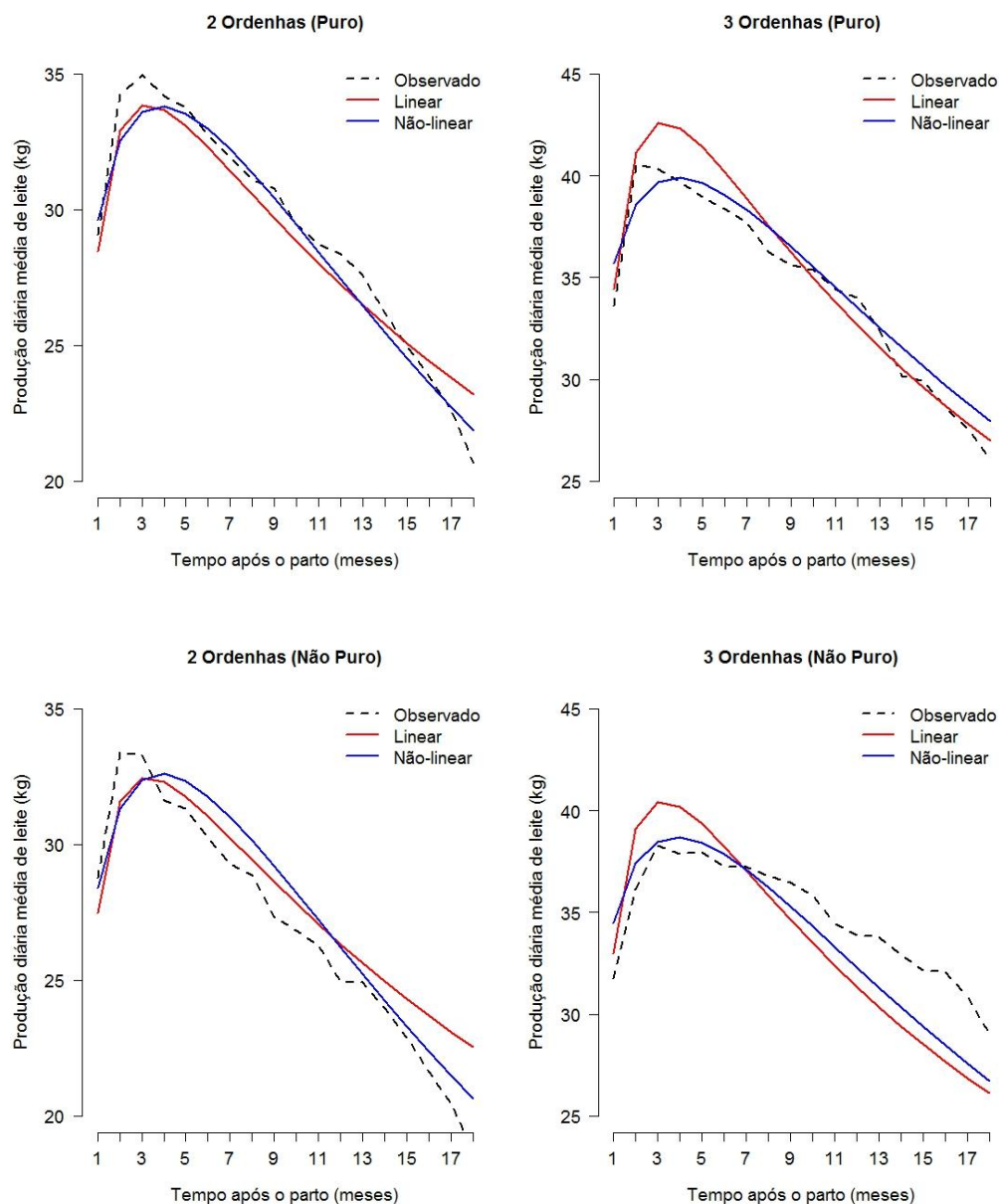


Figura 40 Modelos lineares e não-lineares ajustados à raça Holandesa.

Na Tabela 45 são mostradas as estimativas do tempo de pico (t_p) nos modelos lineares e não-lineares. Como o pico observado foi no 3º mês, observa-se que os modelos lineares têm um erro menor quanto a essa medida.

Tabela 45 Tempo de pico estimado pelos modelos lineares e não-lineares para a raça Holandesa

Número de ordenhas	Grau de sangue	\hat{t}_p (Linear)	\hat{t}_p (Não-linear)
2	Puro	3,23	3,86
2	Não puro	3,23	3,86
3	Puro	3,23	3,86
3	Não puro	3,23	3,86

As produções no pico estimadas pelos modelos lineares e não-lineares são mostradas na Tabela 46, na qual se observa que os modelos não-lineares se aproximaram mais da produção máxima real, principalmente para três ordenhas.

Tabela 46 Produções no pico estimadas pelos modelos lineares e não-lineares para a raça Holandesa.

Número de ordenhas	Grau de sangue	Produção observada	Produção estimada (Linear)	Produção estimada (Não-linear)
2	Puro	34,95	33,86	33,82
2	Não puro	33,41	32,48	32,61
3	Puro	40,57	42,62	39,90
3	Não puro	38,26	40,46	38,70

Os modelos ajustados para a raça Pardo Suíço são mostrados na Figura 41. Observa-se novamente que os modelos não-lineares tiveram um ajuste melhor do que os lineares.

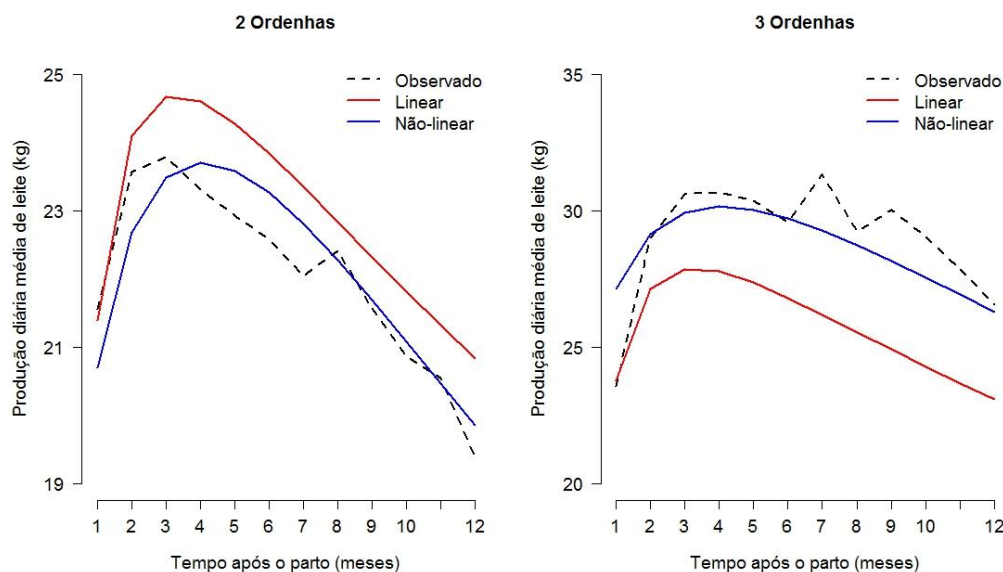


Figura 41 Modelos lineares e não-lineares ajustados à raça Pardo Suíço.

A Tabela 47 mostra as estimativas de tempo e produção de pico dos modelos lineares e não-lineares. Enquanto os modelos lineares têm uma assertividade maior no tempo de pico, os modelos não-lineares têm produção de pico estimada mais próxima da produção real.

Tabela 47 Tempo e produção de pico estimados pelos modelos lineares e não-lineares para a raça Pardo Suíço.

Número de ordenhas	\hat{t}_p (Linear)	\hat{t}_p (Não-linear)	Produção observada	Produção estimada (Linear)	Produção estimada (Não-linear)
2	3,29	4,10	23,8	24,7	23,7
3	3,29	4,10	30,6	27,9	30,2

Para a raça Jersey, os modelos ajustados são mostrados na Figura 42. Observa-se que não há grandes diferenças entre o modelo linear e o modelo não-linear. Quanto ao tempo de pico, os dois modelos o estimaram em 3,25 meses. Pela Figura 42 observa-se que a estimativa da produção no pico dos dois modelos é bastante parecida.

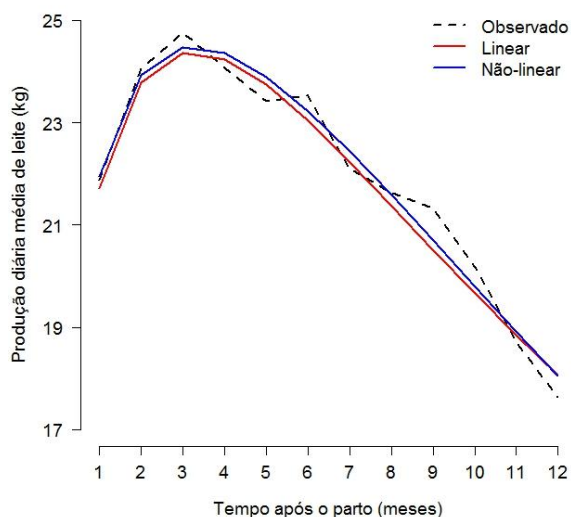


Figura 42 Modelos lineares e não-lineares ajustados à raça Jersey.

Dados os resultados anteriores conclui-se que os modelos não-lineares são mais adequados para o estudo de curvas de lactação das três raças estudadas, para o município de Castro/PR. Comparando-se as raças conclui-se que a raça Holandesa é a que tem maior produção de leite e, entre os animais desta raça, aqueles que são ordenhados três vezes são os que têm melhores índices de produção. Em relação à persistência os animais da raça Pardo Suíço têm uma pequena vantagem, o que significa que eles mantêm sua produção máxima por mais tempo.

Estudos subsequentes a este seriam os que consideram a inclusão de covariáveis relacionadas ao manejo dos animais, pois efeitos de ambiente não foram analisados neste

trabalho por não estarem disponíveis na base de dados. Exemplos: alimentação, condicionamento, tipo de ordenha, estação do ano no parto, entre outras.

Uma continuidade, no âmbito de modelos não-lineares, é a inclusão de uma estrutura de correlação entre as observações. Com isso melhora-se a precisão das estimativas dos parâmetros dos modelos.

Modelos como esses podem ser ajustados para rebanhos específicos de cada produtor. Nesse caso a variabilidade dos dados é menor, o que possibilita um ajuste mais refinado.

6. BIBLIOGRAFIA

AGRANONIK, M. **Equações de Estimação Generalizadas (GEE): Aplicação em estudo sobre mortalidade neonatal em gemelares de Porto Alegre, RS (1995-2007)**. Universidade Federal do Rio Grande do Sul. Porto Alegre. 2009.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.

APCBRH. A Associação. **APCBRH - Associação Paranaense de Criadores de Bovinos da Raça Holandesa**. Disponível em: <<http://www.holandeparana.com.br/associacao.html>>. Acesso em: 21 Novembro 2012.

BALLINGER, G. Using Generalized Estimating Equations for Longitudinal Data Analysis. **Organizational Research Methods**, v. 7, n. 2, p. 127-150, Abril 2004.

BIANCHINI SOBRINHO, E. **Estudo da curva de lactação de vacas da raça Gir**. Universidade de São Paulo. Ribeirão Preto. 1984.

BRODY, S.; RAGSDALE, A. C.; TURNER, C. W. The relation between the initial rise and the subsequent decline of milk secretion following parturition. **The Journal of General Physiology**, v. 6, p. 541-545, 1924.

CAPPIO-BORLINO, A.; PULINA, G.; ROSSI, G. A non-linear modification of Wood's equation fitted to lactation curves of Sardinian dairy ewes. **Small Ruminant Research**, v. 18, p. 75-79, 1995.

CAPUCHO, T. O. **Produção leiteira no Paraná: um estudo considerando os efeitos espaciais**. Universidade Estadual de Maringá. Maringá. 2010.

CARVALHO, G. R. A Indústria de laticínios no Brasil: passado, presente e futuro. **Circular Técnica 102**, Dezembro 2010. 1-12.

CASTRO, P. M. D. Dados Gerais. **Site da Prefeitura Municipal de Castro**. Disponível em: <http://castro.pr.gov.br/site/index.php?option=com_content&view=article&id=3&Itemid=4>. Acesso em: 21 Novembro 2012.

CASTRO, P. M. D. História. **Site da Prefeitura Municipal de Castro**. Disponível em: <http://castro.pr.gov.br/site/index.php?option=com_content&view=article&id=4&Itemid=5>. Acesso em: 21 Novembro 2012.

CASTROLANDA, C. Áreas - Leite. **Cooperativa Castrolanda**. Disponível em: <<http://www.castrolanda.coop.br/areas/leite>>. Acesso em: 21 Novembro 2012.

- CASTROLANDA, C. Processo Industrial. **Leite Castrolanda**. Disponível em: <<http://leite.castrolanda.coop.br/producao>>. Acesso em: 21 Novembro 2012.
- CKROH, A. P.; COIMBRA, E. D. F. **MÉTODOS ESTATÍSTICOS NO AJUSTE DE CURVAS DE LACTAÇÃO**. Universidade Federal do Paraná. Curitiba. 2005.
- COBUCCI, J. A. et al. Aspectos Genéticos e Ambientais da Curva de Lactação de Vacas da Raça Guzerá. **Rev. bras. zootec.**, 30, n. 4, 2001. 1204-1211.
- CUI, J. QIC Program and Model Selection in Gee Analyses. **The Stata Journal**, 7, 2007. 209-220.
- DAVE, B. K. First lactation curve of Indian water buffalo. **Jawaharlal Nehru Krishi Vishwa Vidyalaya Research Journal**, v. 5, p. 93, 1971.
- DHANOVA, M. S. A note on an alternative form of the lactation model of Wood. **Animal Production**, v. 32, p. 342, 1981.
- FARO, L. E.; ALBUQUERQUE, L. G. Comparação de alguns modelos matemáticos para o ajuste às curvas de lactação individuais de vacas da raça Caracu. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, 54, 2002. 295-302.
- JENKINS, T. G.; FERRELL, C. L. A note on lactation curves of crossbred cows. **Animal Production**, v. 39, p. 479-482, 1984.
- LIANG, K.-Y.; ZEGER, S. L. Longitudinal Data Analysis Using Generalized Linear Models. **Biometrika**, v. 73, n. 1, p. 13-22, Abril 1986.
- MEZZADRI, F. P. **Análise da Conjuntura Agropecuária**. Secretaria da Agricultura e do Abastecimento - Departamento de Economia Rural. [S.l.]. 2012.
- MOUTINHO, G. C.; PRUDENTE, A. A.; DEMÉTRIO, C. G. B. Uma Revisão dos Modelos Normais Não-Lineares. **Revista Brasileira de Biometria**, São Paulo, v. 27, n. 3, p. 360-393, 2009.
- NEIVA, R. S. **Produção de Bovinos Leiteiros**. 2ª. ed. Lavras: UFLA/FAEPE, 2000.
- NERO, L. A.; VIÇOSA, G. N.; PEREIRA, F. E. V. Qualidade microbiológica do leite determinada por características de produção. **Ciência e Tecnologia de Alimentos**, Campinas, v. 29, n. 2, p. 386-390, Abril 2009. ISSN 0101-2061.
- OLIVEIRA, H. T. V. et al. Curvas de lactação de vacas F1 Holandês-Gir ajustadas pela função gama incompleta. **Arq. Bras. Med. Vet. Zootec.**, 59, 2007. 233-238.
- PAN, W. Akaike's Information Criterion in Generalized Estimating Equations. **Biometrics**, v. 57, p. 120-125, Março 2001.

PAPAJCSIK, I. A.; BODERO, J. Modeling lactation curves of Friesian cow in a subtropical climate. **Animal Production**, v. 47, n. 2, 1988.

PAULA, G. A. **Modelos de Regressão com Apoio Computacional**. São Paulo: USP, 2010.

PINHEIRO, J. C.; BATES, D. M. Mixed-Effects Models in S and S-PLUS. In: _____ **Nonlinear Mixed-Effects Models: Basic Concepts and Motivating Examples**. [S.l.]: Springer, 2000. Cap. 6, p. 273-304.

PREISSER, J. S.; QAQISH, B. F. Deletion Diagnostics for Generalised Estimating Equations. **Biometrika**, v. 83, n. 3, p. 551-562, Setembro 1996.

PRUDENTE, A. A. **Modelos Não-Lineares de Regressão: Alguns Aspectos de Teoria Assintótica**. Universidade Federal Rural de Pernambuco. Recife. 2009.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. **The R Project for Statistical Computing**, 2011. ISSN 3-900051-07-0. Disponível em: <www.R-project.org>.

SANTOS, E. P. B. **Ajustes de Curvas de Lactação de Vacas Mestiças das Raças Holandesa e Gir**. Universidade Federal do Ceará. Fortaleza. 2012.

SIKKA, L. C. A study of lactation as affected by breeding and environment. **Journal Dairy Research**, v. 17, n. 3, p. 231-252, 1950.

SINGH, R. P.; GOPAL, R. Lactation curves analysis of buffaloes maintained under village conditions. **Indian Journal of Animal Science**, v. 52, n. 12, p. 1157-1160, 1982.

WOOD, P. D. P. Algebraic Model of the Lactation Curve in Cattle. **Nature**, 216, 1967. 164-165.

ZEGER, S. L.; LIANG, K.-Y. Longitudinal Data Analysis for Discrete and Continuous Outcomes. **Biometrics**, v. 42, p. 121-130, 1986.