



DIANA CAMPOS DE OLIVEIRA

**TESTE PARA REGRESSÃO SPLINE LINEAR:
SIMULAÇÃO E APLICAÇÃO A DADOS DE
CRESCIMENTO DE OVINOS**

LAVRAS – MG

2015

DIANA CAMPOS DE OLIVEIRA

**TESTE PARA REGRESSÃO SPLINE LINEAR: SIMULAÇÃO E
APLICAÇÃO A DADOS DE CRESCIMENTO DE OVINOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. Joel Augusto Muniz

LAVRAS - MG

2014

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Oliveira, Diana Campos de.

Teste para regressão spline linear : simulação e aplicação a dados de crescimento de ovinos / Diana Campos de Oliveira. – Lavras : UFLA, 2015.

108 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2014.

Orientador(a): Joel Augusto Muniz.

Bibliografia.

1. Regressão segmentada linear e não linear. 2. Curva de crescimento. 3. Ovino. 4. Propriedades dos estimadores. 5. Simulação. I. Universidade Federal de Lavras. II. Título.

DIANA CAMPOS DE OLIVEIRA

**TESTE PARA REGRESSÃO SPLINE LINEAR: SIMULAÇÃO E
APLICAÇÃO A DADOS DE CRESCIMENTO DE OVINOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 18 de dezembro de 2014.

Dr. Antônio Policarpo Sousa Carneiro	UFV
Dr. Luiz Alberto Beijo	UNIFAL
Dr. Augusto Ramalho de Moraes	UFLA
Dr. Lucas Monteiro Chaves	UFLA

Dr. Joel Augusto Muniz
Orientador

LAVRAS – MG

2014

AGRADECIMENTOS

Primeiramente a DEUS, por me ter dado força para prosseguir ao longo dessa caminhada.

Aos meus pais, pelo exemplo de vida, pelo amor incondicional e incentivo, por sempre estarem ao meu lado contribuindo para minhas conquistas.

Às minhas irmãs pela amizade, por estarem ao meu lado sempre, apoiando-me e incentivando-me nas horas mais difíceis.

Ao meu esposo, Tiago, pelo constante incentivo, pelo companheirismo e carinho, pelo apoio em todos os momentos.

Aos amigos, pelos momentos de estudos e de diversão, pelo companheirismo e presença em todos os momentos difíceis dessa jornada, pessoas que, por mais que a distância nos separe, jamais as esquecerei.

Ao Professor Joel Augusto Muniz, pela orientação e colaboração, por me receber prontamente e sempre tão disposto.

Ao Professor Antônio Policarpo Sousa Carneiro, pela coorientação, sendo sua atuação de grande importância para a conclusão deste trabalho.

À Universidade Federal de Lavras, pela oportunidade no curso de Doutorado em Estatística e Experimentação Agropecuária.

A todos os professores do Departamento de Ciência Exatas, pelos conhecimentos transmitidos.

Aos membros da banca examinadora, Dr. Luiz Alberto Beijo, Dr. Augusto Ramalho de Moraes, Dr. Lucas Monteiro Chaves, Dr. Devanil Jaques de Souza e Dr. Denismar Alves Nogueira, pela disponibilidade e sugestões que contribuíram para o aprimoramento deste trabalho.

Aos funcionários do Departamento de Ciências Exatas por sempre serem tão gentis e prestativos.

Ao Departamento de Estatística e à Universidade Federal de Ouro Preto por terem me concedido afastamento para conclusão deste trabalho.

À Empresa Baiana de Desenvolvimento Agrícola S.A. pela concessão dos dados utilizados no trabalho.

À CAPES pelo apoio financeiro para a realização deste trabalho.

RESUMO

No Brasil, o efetivo de ovinos vem crescendo, contudo, o rebanho é insuficiente para abastecer o mercado interno de forma que o país continua importando carne ovina de outros países o que demonstra uma oportunidade para os produtores brasileiros desenvolverem mecanismos com intuito de produzir carne com qualidade, gerando renda. Uma forma consistente e prática de analisar a eficiência produtiva de animais é por meio do estudo de curvas de crescimento, que descrevem uma relação funcional entre peso e idade. Geralmente as curvas de crescimento são representadas por modelos não lineares que apresentam parâmetros com interpretações biológicas. Os objetivos deste trabalho foram utilizar a regressão linear e não linear, com até três segmentos, para descrever curvas de crescimento em ovinos cruzados, e avaliar, por meio de estudos de simulação, as propriedades dos estimadores dos parâmetros das funções segmentadas lineares, como consistência e tendência, além de avaliar a soma de quadrado extra como uma ferramenta estatística para determinar o número de segmentos significativos das regressões segmentadas lineares. Foram utilizadas informações de pesagens de animais dos seguintes grupos genéticos de ovinos de corte: Dorper x Morada Nova (DMN), Dorper x Rabo Largo (DRL), Dorper x Santa Inês (DSI). Os estimadores dos parâmetros das funções segmentadas lineares apresentaram pequeno viés e são consistentes quando os erros apresentam variâncias homocedásticas e são independentes. O teste estatístico, baseado nas somas de quadrado extra, mostrou-se como excelente ferramenta para determinar o número de segmentos significativos nas regressões segmentadas lineares, quando o tamanho da amostra for grande.

Palavras-chave: Regressão segmentada linear e não linear. Curva de crescimento. Ovinos. Propriedades dos estimadores. Modelos não lineares. Simulação.

ABSTRACT

The sheep flock has been increasing in Brazil but, is not yet sufficient to supply the domestic market, so that Brazil continues importing sheep meat from other countries so far. This shows to Brazil graziers the opportunity of developing mechanisms seeking to obtain high production, productivity, and high quality. A consistent and practical way of analysing the livestock productive efficiency is by studying growth curves, which describe a functional relationship between weight and age. Growth curves, however, are generally described using nonlinear models, which represent parameters for biological interpretation. The purpose of this study was to use linear and nonlinear regressions, with up to three segments, to describe growth curves on breed sheeps. We also assessed, by means of simulation studies, properties of parameters estimators of piecewise linear function, such as consistency and trend. Besides, we assessed the extra sum of squares as a statistical tool to estimate the number of significant segments of these segmented linear regressions. Weight data of the following genetic beef sheep groups were used: Dorper \times Morada Nova, Dorper \times Rabo Largo, and Dorper \times Santa Inês. Parameters estimators of piecewise linear functions showed a slight bias, were found to be independent, and were consistent when errors had homoscedastic variance. Therefore, based on the extra sum of squares, the statistical test was found to be an excellent tool in estimating the number of significant segmented linear regressions.

Key-words: Linear and nonlinear regressions. Growth curve. Sheep flock. Properties of estimators. Nonlinear models. Computer simulation.

LISTA DE FIGURAS

- Figura 1 Modelo linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DMN 71
- Figura 2 Modelo linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DSI 72
- Figura 3 Modelos linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DRL..... 73
- Figura 4 Estimativas dos parâmetros β_{00} e β_{01} em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear com erros correlacionados e variâncias heterocedásticas..... 83
- Figura 5 Estimativas dos parâmetros β_{00} , β_{01} e β_1 , em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear-linear com erros correlacionados e variâncias heterocedásticas..... 84
- Figura 6 Estimativas dos parâmetros β_{00} , β_{01} , β_1 e β_2 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear-linear-linear com erros correlacionados e variâncias heterocedásticas..... 85
- Figura 7 Estimativas dos parâmetros β_{00} e β_{01} em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear com erros normais e homocedásticos 86
- Figura 8 Estimativas dos parâmetros β_{00} , β_{01} e β_1 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear-linear com erros normais e homocedásticos 87

Figura 9 Estimativas dos parâmetros β_{00} , β_{01} , β_1 e β_2 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear-linear-linear com erros normais e homocedásticos 88

LISTA DE TABELAS

Tabela 1	Modelos de regressão segmentada para descrever curvas de crescimento de ovinos cruzados, y representa o peso e x a idade do animal.....	32
Tabela 2	Esquema para análise da variância com a decomposição da soma de quadrados da regressão para três variáveis explanatórias.....	48
Tabela 3	Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DMN, com os respectivos intervalos de confiança.....	55
Tabela 4	Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DSI, com os respectivos intervalos de confiança.....	56
Tabela 5	Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DRL, com os respectivos intervalos de confiança.....	57
Tabela 6	Resultados dos avaliadores da qualidade de ajuste, critério de informação de Akaike (AIC), critério de informação bayesiano (BIC) e coeficiente de determinação ajustado (R^2_{aj}) para as funções segmentadas não lineares, considerando os três grupos genéticos (DMN, DSI e DRL).....	61

Tabela 7	Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DMN ..	63
Tabela 8	Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DSI.....	64
Tabela 9	Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DRL ...	64
Tabela 10	Análise da variância para o peso de ovinos do grupo genético DMN com a decomposição da soma de quadrados da regressão, considerando 1.000 repetições	66
Tabela 11	Análise da variância para os pesos de ovinos do grupo genético DRL com a decomposição da soma de quadrados da regressão considerando, 1.000 repetições	68
Tabela 12	Análise da variância para os pesos de ovinos do grupo genético DSI com a decomposição da soma de quadrados da regressão, considerando 1.000 repetições	69
Tabela 13	Avaliadores da qualidade do ajuste para estimativas dos parâmetros das funções segmentadas linear nos parâmetros para os três grupos genéticos.....	70
Tabela 14	Valores médios do AIC, BIC e Raj2 para os 1.000 modelos simulados.....	74
Tabela 15	ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 50 observações	75
Tabela 16	ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 100 observações	76

Tabela 17 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 150 observações	77
Tabela 18 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 195 observações	78
Tabela 19 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 50 observações	79
Tabela 20 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 100 observações	80
Tabela 21 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1000 repetições e 150 observações	81
Tabela 22 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 195 observações	82

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO DE LITERATURA	17
2.1	Ovinocultura	17
2.2	Curvas de Crescimento	19
2.3	Modelos segmentados	22
2.4	Avaliadores da qualidade do ajuste de modelos	24
2.5	Simulação pelo método de Monte Carlo	25
2.6	Propriedades dos Estimadores dos Parâmetros	27
3	MATERIAL E MÉTODOS	30
3.1	Origem dos dados	30
3.2	Modelos segmentados	30
3.3	Estimação dos parâmetros das regressões segmentadas não lineares	33
3.3.1	Processo Iterativo	36
3.3.2	Derivadas parciais para funções segmentadas não lineares	38
3.4	Estimação dos parâmetros da regressão segmentada linear nos parâmetros	41
3.5	Soma de quadrados extra	45
3.6	Testes de hipóteses sobre os coeficientes de regressão usando as somas de quadrados extras	48
3.7	Avaliadores da qualidade do ajuste	51
3.8	Simulação de dados	52
4	RESULTADOS E DISCUSSÃO	55
4.1	Regressão segmentada ajustada a dados de ovinos	55
4.2	Regressão segmentada ajustada a dados simulados	74
5	CONCLUSÕES	90
	REFERÊNCIAS	91
	APÊNDICE	98
	ANEXOS	102

1 INTRODUÇÃO

No Brasil, dentre os animais de médio porte, o efetivo nacional de ovinos, chegou a 16,8 milhões de cabeças em 2012 (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA- IBGE, 2013).

Contudo, o rebanho ovino é insuficiente para abastecer o mercado interno brasileiro, de forma que o país continua importando em torno de 50% da carne ovina consumida no país (HOLANDA JÚNIOR; AS; ARAUJO, 2003), o que reforça a necessidade de estudos relacionados às curvas de crescimento, que ajudam na definição do manejo adequado colaborando para o aumento da produtividade dos animais.

A escassez desses animais demonstra que existem oportunidades para os produtores brasileiros desenvolverem a cadeia produtiva de ovinos e assim ocupar de forma mais racional as imensas áreas de pasto disponíveis no país (SORIO, 2010). De acordo com Medeiros (2001), o crescimento do consumo de carne ovina no Brasil está direcionado para nichos de mercado existentes nas grandes cidades, nas quais o poder aquisitivo da população é maior, exigindo maior qualidade, cortes especiais e continuidade de abastecimento no mercado.

Segundo Malhado et al. (2009), um dos grandes problemas para a prática desta atividade é que grande parte dos produtores não está consciente das vantagens de se produzir carne de boa qualidade, e abastecem o mercado com carcaças sem qualidade, derivadas do abate de animais com idade avançada comprometendo a qualidade da carne.

Uma forma consistente e prática de analisar a eficiência produtiva de animais é por meio do estudo de curvas de crescimento, as quais relacionam o peso do animal com a sua idade. Os modelos não lineares, usados para descrever tais curvas, sintetizam um grande número de medidas em apenas alguns

parâmetros que possuem interpretações práticas (SILVA; AQUINO; OLIVEIRA, 2001).

As curvas de crescimento têm se mostrado úteis na obtenção de animais que apresentam carne de qualidade, uma vez que possuem parâmetros com fácil interpretação biológica, como velocidade de crescimento, peso na idade adulta e o ponto em que o animal passa de uma fase de crescimento acelerada para uma fase de crescimento menor (SILVA; AQUINO; OLIVEIRA, 2001).

Modelos não lineares usuais têm sido muito utilizados para descrever curvas de crescimento. Entre os modelos de regressão utilizados na modelagem do crescimento, merece destaque a regressão segmentada conectada por nós (pontos de junção), as quais são denominadas funções segmentadas ou splines, que podem ser lineares ou não lineares nos parâmetros.

As funções segmentadas apresentam ajustes definidos por partes, em que para cada intervalo da variável independente são estimados modelos distintos. Essas funções apresentam facilidade de convergência, são fáceis de manipular e avaliar, são úteis quando um polinômio de grau menor não se ajusta bem aos dados e o fenômeno em estudo apresenta grandes alterações.

Na literatura, existem poucos trabalhos envolvendo estudo de crescimento com funções segmentadas aplicadas ao estudo de curvas de crescimento. De acordo com Hernandez (2010), regressão segmentada é uma ferramenta estatística útil na modelagem de fenômenos em transição e com grandes mudanças no comportamento.

Nas regressões segmentadas, o ponto na variável explicativa no qual ocorre alteração no comportamento dos dados é denominado de nó, geralmente é de algum interesse do pesquisador. As regiões delimitadas pelos nós são referidas como fases ou etapas.

Os nós das regressões segmentadas podem ser estimados ou determinados pelo pesquisador por meio do diagrama de dispersão ou pelo

conhecimento do pesquisador. Tem-se um modelo não linear quando os nós são parâmetros do modelo, sendo, portanto estimados por meio de métodos iterativos e um modelo linear se os nós forem determinados previamente pelo pesquisador. Neste caso, o nó torna-se uma constante e as derivadas parciais em relação aos parâmetros não são funções de nenhum parâmetro desconhecido, caracterizando a linearidade do modelo.

Objetivou - se estimar os parâmetros dos modelos segmentados lineares e não lineares, com um, dois e três segmentos, utilizando dados reais, e apresentar os métodos de estimação para o caso linear e não linear.

Objetivou-se também avaliar a soma de quadrado extra como uma ferramenta estatística para determinar o número de segmentos significativos das regressões segmentadas lineares, além de apresentar e avaliar, por meio de estudos de simulação, as propriedades dos estimadores dos parâmetros das funções segmentadas lineares, como consistência e tendência.

2 REVISÃO DE LITERATURA

2.1 Ovinocultura

Explorada em todos os continentes, a ovinocultura é uma atividade econômica que está presente em áreas sob as mais diversas características climáticas, edáficas e botânicas (VIDAL et al., 2006).

De acordo com Fernandes (1998), a criação de ovinos está presente na história da humanidade como sendo a atividade que proporciona a maior fonte de alternativas para subsistência, oferecendo lã e pele para vestuário além de carne e leite para alimentação.

Segundo Viana (2008), a produção de lã foi o principal objetivo da exploração econômica da ovinocultura no século XX, enquanto a produção de carne, produto considerado secundário, apenas supria o consumo dos estabelecimentos rurais.

A crise neste setor, proporcionada pelos elevados estoques australianos de lã e o início da comercialização de tecidos sintéticos no mercado têxtil internacional, fez com que a atividade deixasse de ser economicamente vantajosa, com isso muitos produtores abandonaram o setor, reduzindo significativamente o rebanho comercial e gerando a desestruturação de toda a cadeia produtiva (NOCCHI, 2001).

O aumento do poder aquisitivo da população e o incremento do abate de animais jovens trouxe um novo mercado para a ovinocultura. A carne ovina começou a ser apreciada e isso levou a uma maior demanda de consumo, o que indica um excelente potencial para se tornar um produto substituto à lã no mercado (VIANA, 2008).

Vidal et al. (2006) lembram que, na região nordeste do Brasil, os ovinos desempenham papel primordial no suprimento alimentar e geração de renda. No

entanto, o baixo potencial genético dos rebanhos, associado à escassez de pastos na estação seca, às práticas de manejo inadequadas, às precárias condições sanitárias, às graves limitações zootécnicas e de assistência técnica, além das imperfeições de mercado, concorrem para os baixos índices de produtividade e de rentabilidade.

Uma solução para este problema seria a utilização de métodos de criação de cordeiros mais eficientes, que forneçam carne de qualidade e a baixo custo, aumentando a lucratividade, as curvas de crescimento podem ser uma alternativa para contribuir com isso.

Atualmente, poucas instituições públicas de pesquisa do Brasil mantêm rebanhos de ovinos de raças naturalizadas do Nordeste brasileiro, com objetivo de utilizar estes animais em cruzamentos com raças exóticas especializadas para a produção de carne (MALHADO et al., 2009). Além disso, as informações disponíveis sobre a eficiência dos resultados desses cruzamentos são insuficientes, fazendo-se necessário o desenvolvimento de pesquisas com a finalidade de identificar cruzamentos com melhor desempenho.

A raça Santa Inês é encontrada em todas as regiões do Brasil, as fêmeas apresentam boa habilidade materna e conseguem facilmente parir cordeiros vigorosos (Ovinos Santa Inês: Estado de Arte e Perspectivas).

Uma das raças de ovinos com grande utilização para produção de carne com qualidade é a Dorper, originária da África do Sul, que surgiu do cruzamento da ovelha Blackhead Persian com o Dorset Horn (SOUZA; LEITE, 2000).

A raça Rabo Largo apresenta um efetivo reduzido, é considerada umas das raças ovinas mais rústicas do Nordeste. Já a raça Morada Nova compõe uma das principais raças de ovinos deslanados explorada para carne, representando fonte de proteína na alimentação rural (MALHADO et al., 2009).

2.2 Curvas de Crescimento

De acordo com Silva, Aquino e Oliveira (2001), uma forma prática e eficiente de se analisar o crescimento de animais é por meio do estudo de suas curvas de crescimento, que descrevem uma relação funcional entre peso e idade, geralmente representada por modelos não lineares.

Uma das principais vantagens na utilização desses modelos é o agrupamento de várias informações de pesagens associadas à idade, durante o crescimento, em poucos parâmetros biologicamente interpretáveis (SARMENTO et al., 2006). Os parâmetros, nos modelos não lineares, podem ser úteis para prever taxas de crescimento, necessidades alimentares, peso à maturidade e graus de maturidade.

O conhecimento do processo de ganho de massa corporal do animal possibilita que se faça um controle da produção de carne. Este incremento pode ser influenciado pela alimentação, por condições climáticas, pelo estado sanitário e pelas características genéticas associadas aos animais (GOTTSCHALL, 1999).

Souza e Bianchini Sobrinho (1994) argumentam que o ganho de peso, associado às características produtivas dos animais, pode ser uma alternativa para programas de seleção, visando à precocidade com maior peso e melhor qualidade de carcaça.

Os modelos não lineares têm sido utilizados para descrever o crescimento do animal ao longo do tempo, possibilitando avaliar os fatores genéticos e ambientais que influenciam a forma da curva de crescimento, permitindo a seleção de animais com maior velocidade de crescimento, sem alterar o peso adulto, em vez de selecionar animais cada vez maiores (SARMENTO et al., 2006).

Dessa forma, as curvas de crescimento permitem conhecer o processo de ganho de massa corporal do animal, informações que irão contribuir na eficiência da produção de carne gerando redução dos custos e maior rentabilidade ao produtor.

Segundo Fitzhugh Junior (1976), alguns requisitos devem ser atendidos para que um modelo de regressão não linear descreva adequadamente a relação peso-idade: interpretação biológica dos parâmetros, alta qualidade de ajuste e facilidade de convergência, uma vez que requer métodos iterativos para sua estimação. De forma geral, para que um modelo não linear possa ser usado para descrever curvas de crescimento, é necessário que os parâmetros apresentem interpretação relevante sob o ponto de vista biológico, ajustes com pequenos desvios além de proporcionar altas taxas de convergência, uma vez que modelos não lineares requerem métodos iterativos de estimação.

Com relação aos estudos de curvas de crescimento de ovinos, Lobo et al. (2006) ajustaram os modelos von Bertalanffy, Brody, Gompertz, Richardse Logístico a dados de peso-idade de ovinos da raça Santa Inês e o modelo Richards foi o que melhor se ajustou aos dados (BERTALANFFY, 1957; BRODY, 1945; RICHARDS, 1959). Entretanto, Guedes et al. (2004) ajustaram estes mesmos modelos aos dados de peso-idade de cordeiros da raça Bergamácia e Santa Inês e concluíram que os modelos von Bertalanffy e Gompertz proporcionaram o melhor ajuste.

Já Silveira et al. (2012) concluiu que os modelos Richards e von Bertalanffy foram, respectivamente, os que apresentaram os melhores ajustes para os conjuntos de dados médios e individuais de ovinos dos mesmos grupos genéticos utilizados neste trabalho.

Carneiro et al. (2014) avaliaram o uso de teste de identidade de modelos não lineares, na comparação de curvas de crescimento de bovinos da raça Tabapuã, de cinco regiões de produção do Nordeste do Brasil. Os autores

concluíram que o modelo reduzido, com igualdade da taxa de maturidade, com quatorze parâmetros, foi o mais adequado para descrever o crescimento dos animais.

Mazzini et al. (2003) utilizaram os modelos não lineares no ajuste de curvas de crescimento de machos Hereford. Nesse estudo, o modelo Gompertz apresentou a melhor qualidade de ajuste, sendo esta verificada por meio do coeficiente de determinação, erro de predição médio e porcentagem de convergência.

Mazzini et al. (2005) também realizaram estudos com curvas de crescimento de bovinos e apontaram que os modelos com estrutura de erros autorregressivos apresentaram estimativas dos coeficientes mais próximas da realidade biológica e melhor qualidade de ajuste, sendo, portanto, mais eficientes para descrever a curva de crescimento do rebanho.

Mendes et al. (2008) compararam modelos logísticos difásicos ponderados aplicados ao estudo de curvas de crescimento de fêmeas Hereford com três diferentes estruturas de erros: erros independentes, autorregressivos de primeira ordem e autorregressivo de segunda ordem a dados de peso-idade de 55 fêmeas da raça Hereford obtiveram ajustes mais eficientes do que no ajuste dos modelos sem autocorrelação.

Silveira et al. (2011) utilizaram a análise de agrupamento para classificar modelos de regressão não lineares usados para descrever a curva de crescimento de ovinos cruzados, tendo em vista os resultados de diferentes avaliadores de qualidade de ajuste.

Guedes et al. (2005) utilizaram curvas de crescimento e ajustaram, por meio da metodologia Bayesiana, o modelo logístico a dados de peso-idade de cordeiros da raça Santa Inês e realizaram estimação por intervalo para as taxas de crescimento absolutas em tempo distintos considerados no estudo.

São vários os estudos com modelos não lineares aplicados a curvas de crescimento, entretanto, ainda são raros, na literatura, trabalhos utilizando funções segmentadas com o intuito de relacionar peso-idade de animais possibilitando o ajuste de modelos distintos para diferentes intervalos de idade. Geralmente são utilizados modelos não lineares usuais para tal finalidade.

2.3 Modelos segmentados

Uma alternativa para descrever a relação entre peso-idade, utilizando curvas de crescimento, é definir um modelo por partes, contemplando diferentes intervalos de idades, ou seja, estimar os parâmetros de um modelo com ajustes locais para cada segmento. Este modelo é chamado função segmentada ou regressão segmentada.

A grande vantagem das funções segmentadas é poder ajustar funções que descrevam o crescimento do animal de forma diferente para cada intervalo de idade considerado. Os nós das funções segmentadas, definidos como os pontos de junção destes intervalos, apresentam facilidade de interpretação e aplicação prática, uma vez que representam a idade em que o crescimento do animal apresenta uma mudança no comportamento.

Segundo Hernandez (2010), as funções segmentadas começaram a receber atenção na década de 50, no controle de qualidade, em que o objetivo é verificar se há alguma mudança de comportamento dos dados e em que ponto da variável independente ocorre tal alteração. Ao longo dos anos, o campo tornou-se muito mais diversificado acarretando expansão na sua utilização.

A origem do nome “spline” vem de uma régua elástica, usada em desenhos de engenharia, que pode ser curvada de forma a passar por um dado conjunto de pontos. Os nós eram fixados nas áreas de interesse causando a deformação da estrutura.

Segundo Keele (2008), “splines” são funções de regressão segmentadas em que se impõem restrições para que os pontos de cada segmento se unam nos pontos chamados nós e podem ser descritas da seguinte forma:

$$y = \sum_{j=0}^g \beta_{0j} x^j + \sum_{i=1}^c \beta_i (x - t_i)^g + \varepsilon,$$

Em que, g é o grau da função segmentada, c é o número de nós, t_i são os nós entre os segmentos e,

$$(x - t_i) = \begin{cases} x - t_i, & \text{se } x > t_i \\ 0, & \text{se } x \leq t_i \end{cases} \quad i=1, 2, \dots, c$$

Para o caso linear-linear, por exemplo, tem-se:

$$\begin{aligned} y &= \beta_{00} + \beta_{01}x + \varepsilon_i, & \text{se } x \leq t_1; \\ y &= \beta_{00} + \beta_{01}x + \beta_1(x - t_1)^1 + \varepsilon_i, & \text{se } x > t_1 \end{aligned}$$

Este modelo pertence à classe dos modelos não lineares, pois se derivando a função y em relação a t_i , considerando o nó como um parâmetro, tem-se como resultado uma função do parâmetro β_1 , caracterizando um modelo não linear.

Malhado et al. (2008a, 2008b) utilizaram a regressão não paramétrica por meio de polinômios articulados, “spline”, para descreveras tendências genéticas e fenotípicas das características de desenvolvimento ponderal em bovinos da raça Nelore.

Segundo Tholon et al.(2012), o uso de funções *splines* é uma alternativa viável ao ajuste de curvas de crescimento o que pode contribuir para uma correta determinação do desempenho do animal em idades iniciais e tardias.

2.4 Avaliadores da qualidade do ajuste de modelos

Segundo Regazzi (2003), ao determinar a relação entre as variáveis, por meio de estudos de regressão, há o problema da definição da forma matemática que será ajustada para representar o fenômeno em estudo. Frequentemente, ajusta-se mais de um modelo e, com base nos resultados dos avaliadores da qualidade do ajuste, escolhe-se qual o modelo que apresentou melhor ajuste.

Entretanto, a escolha dos melhores modelos não tem sido tarefa fácil, uma vez que diferentes avaliadores de qualidade de ajuste são propostos na literatura, e cada um preconiza determinada característica, como por exemplo, a minimização dos erros de ajustes, a habilidade de predição e a simplicidade do modelo (SILVEIRA et al., 2012).

Ainda segundo Silveira et al. (2012), tal escolha é ainda mais difícil quando se tem diferentes populações ou cruzamentos, pois muitas vezes o modelo que melhor representou o crescimento dos animais de uma população pode não ser o melhor para as demais.

Na literatura, a maioria dos trabalhos envolvendo comparação de modelos para curva de crescimento utiliza apenas o coeficiente de determinação ajustado (R^2_{aj}) e o desvio padrão residual como avaliadores da qualidade de ajuste. Tal procedimento pode ser explicado pelo fato destes avaliadores serem apresentados diretamente nos arquivos de saídas de *softwares* estatísticos como SAS® e R (SILVEIRA et al., 2009).

De acordo com Pachepsky, Haskett e Acock (1996), o melhor modelo é aquele que possuir pequeno número de parâmetros para certa precisão das estimativas, oferecer uma forma mais simplificada, além de estar em acordo com leis físicas e biológicas, apresentar os menores desvios entre valores observados e preditos, e fornecer variância mínima dos estimadores.

Tanto as funções segmentadas quanto os modelos não lineares usuais apresentam vantagens e desvantagens sob pontos de vista estatísticos. É preciso determinar qual é o mais apropriado para descrever o crescimento corporal de uma população de ovinos e neste caso os avaliadores de qualidade de ajuste são úteis para fazer tal escolha (OLIVEIRA, 2011).

Silva et al. (2011) utilizaram análise de agrupamento para classificar e selecionar modelos não lineares de crescimento de bovinos Nelore, tendo em vista os resultados de diferentes avaliadores de qualidade de ajuste.

2.5 Simulação pelo método de Monte Carlo

O método de Monte Carlo é uma ferramenta matemática comumente utilizada em diversos segmentos da ciência e da engenharia para simular problemas que podem ser representados por processos estocásticos. O crescimento deste tema se deve diretamente ao rápido avanço tecnológico dos computadores, que traz benefícios tanto à velocidade de processamento quanto à capacidade de armazenamento de informações (YORIYAZ, 2009).

Métodos de Monte Carlo são cálculos numéricos baseados em amostragem de variável aleatória. A técnica de amostragem aleatória para resolver problemas matemáticos é conhecida desde 1770, chamado "método de Monte Carlo", porque o cassino Monte Carlo era um centro famoso por jogos envolvendo sorteio. No início de 1960, o método de Monte Carlo foi utilizado por H. O. Anger para simular a resposta física de sua nova câmara de cintilação (BUVAT; CASTIGLIONI, 2002).

Segundo Garcia, Lustosa e Barros (2010), com o crescimento da complexidade dos problemas reais e a evolução dos sistemas computacionais, a simulação aparece como um instrumento cada vez mais utilizado nas mais variadas áreas de conhecimento, em estudo de problemas, geralmente

complexos, para os quais não se dispõe de solução analítica. Consiste no desenvolvimento de um modelo ou representação de uma situação real (ou ainda por existir) e, por intermédio do uso do computador, possibilita a realização de experimentos com vários cenários.

Garcia, Lustosa e Barros (2010) utilizaram o método de Simulação de Monte Carlo para prever variações nos custos de produção em um período pós-privatização da Companhia Vale do Rio Doce – CVRD, considerando a sua privatização ocorrida em 1997.

Uma breve revisão das principais aplicações do Método de Monte Carlo foi realizada por Yoriyaz (2009), abrangendo as aplicações nas diversas modalidades, acompanhado dos principais códigos computacionais. Com o intuito de introduzir o tema àqueles que desejam conhecer o método, é apresentada uma breve descrição dos seus conceitos básicos e de suas potencialidades.

Entre os muitos algoritmos desenvolvidos ao longo dos últimos anos para distribuições de dose e no planejamento da radioterapia, os algoritmos baseados em métodos de Monte Carlo têm provado ser muito promissores em termos de precisão, proporcionando resultados mais realistas. Os métodos permitem simular o transporte de radiação ionizante em configurações complexas (OLIVEIRA et al., 2013).

Pereira Filho e Ferreira (2012) propuseram um teste de normalidade univariado que pode ser empregado para qualquer tamanho de amostra, incluindo os casos de amostras superiores a 5.000. A validação do procedimento proposto foi por meio da avaliação das taxas de erro tipo I e poder do teste, utilizando para isso os métodos Monte Carlo, considerando diferentes tamanhos de amostras e níveis nominais de significância, além de comparar os resultados do desempenho do teste proposto com os do teste de Shapiro-Wilk, considerado um teste de normalidade com propriedades ótimas.

Em muitas situações, parâmetros de interesse dos modelos de regressão não linear ou funções dos mesmos são quantidades sujeitas à investigação. Surge então a preocupação de como fazer inferência sobre os parâmetros. Zeviani (2013) utilizou procedimentos de simulação Monte Carlo além do método delta e *bootstrap* para fazer inferências sobre os parâmetros.

Biase e Ferreira (2011) propuseram uma abordagem bayesiana para realizar um teste global de igualdade de várias proporções binomiais e um teste de comparações múltiplas de proporções binomiais. O desempenho desses testes foi avaliado utilizando simulação Monte Carlo.

Foi avaliado, via simulação Monte Carlo, o comportamento da variância genética aditiva de populações selecionadas e não selecionadas por seis gerações sucessivas. Por meio de um modelo genético aditivo, foram simuladas populações de 40 animais (20 machos e 20 fêmeas) sob seleção e acasalamento aleatório (REIS et al., 2009).

Quanto às bovinoculturas de leite e de corte, devido ao grande intervalo de geração, há necessidade de muito tempo para se obter dados que possam ser utilizados em pesquisas que visem comparações de metodologias, utilizando dados de sucessivas gerações em populações sob seleção. Nesse caso, a simulação de dados via método Monte Carlo figura-se como uma alternativa viável para o problema, pois além de fornecer resultados confiáveis, possibilita também uma redução de tempo e custo inerentes às pesquisas desenvolvidas nessa área (SILVA; AQUINO; OLIVEIRA, 2001).

2.6 Propriedades dos Estimadores dos Parâmetros

Uma importante aplicação de um modelo de regressão é fazer previsão ou estimar novos valores da variável dependente, a partir de valores observados ou atribuídos à variável independente. Para que essas previsões sejam

adequadas, é essencial que os parâmetros envolvidos no modelo de regressão tenham estimadores que satisfaçam algumas propriedades tais como não tendenciosidade e consistência (DOMINGOS FILHO, 2012).

A consistência do estimador é, geralmente, considerada uma propriedade essencial de um estimador razoável, pois significa que o estimador é não viciado (não contém erro) e que sua variabilidade converge para zero quando o tamanho da amostra tende ao infinito (LARSEN; MARX, 1986).

Com o grande desenvolvimento dos recursos computacionais, o uso de simulações consiste em uma alternativa viável para o estudo de propriedades estatísticas de interesse. Esses estudos são baseados em informações reais e utilizados como repetições de um experimento, sendo igualmente aplicáveis para variáveis respostas contínuas ou discretas (VIOLA et al., 2000).

Em análise de regressão, linear ou não linear, uma suposição geralmente considerada é a de homocedasticidade dos erros, ou seja, assume-se que os erros do modelo adotado possuem variâncias estatisticamente idênticas (MAZUCHELI; SOUZA; PHILIPPSEN, 2011).

Se à medida que o valor da variável dependente aumenta, os erros de predição também aumentam, tem-se heterogeneidade na variância. Fundamentalmente, a violação desse pressuposto é preocupante na medida em que afeta a confiabilidade dos testes de significância e intervalos de confiança (FIGUEIREDO FILHO et al., 2011).

Maciel, Vieira e Marinho (2013) utilizaram simulações de Monte Carlo para avaliar os desempenhos dos estimadores dos Métodos dos Momentos Modificados e suas versões corrigidas para os parâmetros α e β da distribuição Birnbaum-Saunders, através da verificação de algumas propriedades desses estimadores. Segundo os mesmos autores, um dos principais problemas da inferência estatística é o da estimação, os estimadores deverão ser consistentes e de preferência com o menor viés possível.

Segundo Cribari-Neto e Soares (2003), no que se refere à análise dos vieses dos estimadores de matrizes de covariância, o estimador de mínimos quadrados ordinários é não viesado sob homocedasticidade, mas fortemente viesado quando esta propriedade não se verifica, o que, segundo os mesmos autores, era esperado.

Domingos Filho (2012) mostrou um estudo detalhado das propriedades dos estimadores de mínimos quadrados sobre os parâmetros do modelo de calibração estrutural, quando a suposição de normalidade for eliminada.

Caetano e Machado (2009) procuraram descrever as propriedades dos estimadores de GoM, (método Grade of Membership), amplamente utilizado no Brasil na área de Demografia e na Saúde Pública. Estes mostraram que o modelo dispõe de propriedades estatísticas importantes, enfatizando a consistência e a identificabilidade, na busca de um modelo que convirja para uma solução única, contudo, o critério de suficiência estatística não foi abordado no trabalho.

3 MATERIAL E MÉTODOS

3.1 Origem dos dados

Os dados utilizados neste trabalho são provenientes de um experimento realizado na Estação Experimental de Jaguaquara, pertencente à Empresa Baiana de Desenvolvimento Agrícola S.A., localizada no Município de Jaguaquara, BA, microrregião administrativa de Jequié, BA.

Foram utilizados dados de pesagens de ovinos, de três grupos genéticos, obtidos pelos cruzamentos das raças Dorper x Morada Nova (DMN); Dorper x Rabo Largo (DRL) e Dorper x Santa Inês (DSI), no período de 2003 a 2005.

Foram utilizados 36 animais DMN (26 machos e 10 fêmeas), 61 DRL (36 machos e 25 fêmeas) e 59 DSI (31 machos e 28 fêmeas). Após o desmame, foram mantidos em sistema de produção semi-intensivo, durante o dia com pastejo e sal mineral, sendo recolhidos ao final da tarde (SOUZA JÚNIOR et al., 2009).

Foram realizadas pesagens do nascimento até os 337 dias de idade em intervalos de tempo irregulares entre as pesagens. Para ajuste dos modelos de curvas de crescimento, considerou-se a média dos pesos por idade em razão da grande irregularidade nos intervalos de pesagens.

3.2 Modelos segmentados

Para os modelos segmentados não lineares, os nós, assim como os demais parâmetros, foram estimados por meio do método dos mínimos quadrados cujas soluções foram obtidas por meio do processo iterativo de Gauss-Newton.

No método iterativo de Gauss-Newton, as soluções são obtidas por meio de algoritmos computacionais que se baseiam numa aproximação por série de Taylor de primeira ordem para produzir uma linearização da função não linear com termos lineares e, então, aplica-se o método dos mínimos quadrados ordinários para estimar os parâmetros.

No caso em que os nós são pré-definidos pelo pesquisador, tem-se uma regressão segmentada linear e a obtenção dos parâmetros é por meio de uma regressão múltipla, utilizando o método de mínimos quadrados.

Os nós foram pré-determinados de acordo com Souza Júnior et al. (2009) que avaliaram o crescimento diferencial dos cortes da carcaça de cordeiros cruzados Dorper com as raças Rabo Largo e Santa Inês. No estudo foram utilizados 20 cordeiros de cada grupo genético, em cinco idades de abate: 84 (desmame), 126, 168, 220 e 272 dias de idade. Como são necessárias duas idades para definir os nós, foram utilizadas 84 (desmame) e 168 dias.

Neste trabalho, foram estimados os parâmetros para as funções segmentadas linear-linear (LL) e linear-linear-linear (LLL), além da regressão linear simples, (Tabela 1), devido ao comportamento dos dados analisados pelo diagrama de dispersão.

Tabela 1 Modelos de regressão segmentada para descrever curvas de crescimento de ovinos cruzados, y representa o peso e x a idade do animal

Modelos	Equações
Linear	$y = \beta_{00} + \beta_{01}x + \varepsilon,$
Linear- Linear	$y = \beta_{00} + \beta_{01}x + \varepsilon, \text{ se } x < t_1$
	$y = \beta_{00} + \beta_{01}x + \beta_1(x - t_1) + \varepsilon, \text{ se } x \geq t_1$
Linear- Linear- Linear	$y = \beta_{00} + \beta_{01}x + \varepsilon, \text{ se } x < t_1$
	$y = \beta_{00} + \beta_{01}x + \beta_1(x - t_1) + \varepsilon, \text{ se } t_1 \leq x < t_2$
	$y = \beta_{00} + \beta_{01}x + \beta_1(x - t_1) + \beta_2(x - t_2) + \varepsilon, \text{ se } x \geq t_2$

Para os modelos ilustrados na Tabela 1, tem-se que y representa a observação da variável dependente (peso), x a observação da variável independente (idade).

Pode-se interpretar biologicamente os parâmetros das regressões segmentadas como intercepto com o eixo y (β_{00}), ou peso ao nascer no caso em que a idade zero for observada, ganho de peso diário na primeira fase (β_{01}), variação no ganho de peso da primeira fase para a segunda (β_1), variação no ganho de peso da segunda fase para a terceira (β_2), e os nós, idades de mudança no comportamento do crescimento, (t_1 e t_2), ganho de peso na segunda fase ($\beta_{01} + \beta_1$), ganho de peso na terceira fase ($\beta_{01} + \beta_1 + \beta_2$).

3.3 Estimação dos parâmetros das regressões segmentadas não lineares

Nos modelos não lineares, não é possível obter a forma analítica para os estimadores de mínimos quadrados ou de máxima verossimilhança. Métodos numéricos devem ser usados juntamente com os métodos de estimação tradicionais através de cálculos computacionais intensivos.

Para obter as equações normais para o modelo não linear:

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

Em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, f(\mathbf{x}, \boldsymbol{\theta}) = \begin{bmatrix} f(x_1, \boldsymbol{\theta}) \\ f(x_2, \boldsymbol{\theta}) \\ \vdots \\ f(x_n, \boldsymbol{\theta}) \end{bmatrix} e \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Em que, \mathbf{y} é vetor de variável resposta, $f(\mathbf{x}, \boldsymbol{\theta})$ é uma função não linear, \mathbf{x} é o vetor da variável preditora e $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_p]$ é o vetor de parâmetros p dimensional, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ representa o efeito do erro aleatório não observável.

Minimiza-se, em relação aos parâmetros, a equação:

$$SQE(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2$$

Na notação vetorial:

$$SQE(\boldsymbol{\theta}) = \|\mathbf{y} - f(\mathbf{X}, \boldsymbol{\theta})\|^2$$

$$SQE(\boldsymbol{\theta}) = [\mathbf{y} - f(\mathbf{X}, \boldsymbol{\theta})]'[\mathbf{y} - f(\mathbf{X}, \boldsymbol{\theta})]$$

Em que: SQE é a soma dos quadrados dos erros que deverá ser minimizada por $\hat{\theta}$, $\|y-f(\theta)\|$ é a norma euclidiana de $y-f(\theta)$.

Desta forma o estimador $\hat{\theta}$ de mínimos quadrados de θ é obtido por meio da minimização da soma de quadrados residuais, descrita acima.

A solução das equações normais usualmente requer métodos numéricos iterativos uma vez que a solução analítica geralmente não pode ser encontrada.

Para obtenção dos estimadores de mínimos quadrados, utilizando-se o processo de derivação, têm-se as seguintes notações de diferenciação matricial:

- a) $f(\theta)$ é uma função vetor coluna $n \times 1$ de um argumento p dimensional

$$f(\theta) = \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \\ \vdots \\ f_n(\theta) \end{bmatrix}_n$$

- b) $F(\theta)$ é matriz jacobiana de $f(\theta)$

$$F(\theta) = \frac{\partial f(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial f_1(x, \theta)}{\partial \theta_1} & \frac{\partial f_1(x, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_1(x, \theta)}{\partial \theta_p} \\ \frac{\partial f_2(x, \theta)}{\partial \theta_1} & \frac{\partial f_2(x, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_2(x, \theta)}{\partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n(x, \theta)}{\partial \theta_1} & \frac{\partial f_n(x, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_n(x, \theta)}{\partial \theta_p} \end{bmatrix}_{n \times p}$$

De forma que, o estimador de mínimos quadrados, $\hat{\boldsymbol{\theta}}$, que satisfaz

$$\frac{\partial SQE}{\partial \boldsymbol{\theta}'} = \mathbf{0}$$

Sendo:

$$\begin{aligned} \frac{\partial SQE}{\partial \boldsymbol{\theta}'} &= \frac{\partial}{\partial \boldsymbol{\theta}'} [\mathbf{Y} - f(\boldsymbol{\theta})]' [\mathbf{Y} - f(\boldsymbol{\theta})] \\ \frac{\partial SQE}{\partial \boldsymbol{\theta}'} &= 2[\mathbf{Y} - f(\boldsymbol{\theta})]' \left(-\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) = -2[\mathbf{Y} - f(\boldsymbol{\theta})]' F(\boldsymbol{\theta}) \end{aligned}$$

Tem-se, então:

$$\begin{aligned} \Rightarrow -2[\mathbf{Y} - f(\hat{\boldsymbol{\theta}})]' F(\hat{\boldsymbol{\theta}}) &= 0 \Rightarrow -2F'(\hat{\boldsymbol{\theta}})[\mathbf{Y} - f(\hat{\boldsymbol{\theta}})] = 0 \\ \Rightarrow F'(\hat{\boldsymbol{\theta}})[\mathbf{Y} - f(\hat{\boldsymbol{\theta}})] &= \mathbf{0} \end{aligned}$$

Portanto, o sistema de equações normais (SEN), para o modelo não linear é dado por:

$$\left[\begin{array}{ccc} \frac{\partial f_1(x, \boldsymbol{\theta})}{\partial \theta_1} & \dots & \frac{\partial f_n(x, \boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial f_1(x, \boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_n(x, \boldsymbol{\theta})}{\partial \theta_2} \\ \vdots & & \vdots \\ \frac{\partial f_1(x, \boldsymbol{\theta})}{\partial \theta_p} & \dots & \frac{\partial f_n(x, \boldsymbol{\theta})}{\partial \theta_p} \end{array} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} \cdot \left(\begin{array}{c} \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] - \left[\begin{array}{c} f_1(\boldsymbol{\theta}) \\ f_2(\boldsymbol{\theta}) \\ \vdots \\ f_n(\boldsymbol{\theta}) \end{array} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0} \end{array} \right) = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \end{array} \right]$$

As equações, anteriores, do SEN são não lineares nos parâmetros (são função de algum parâmetro) e, em geral, não existe uma solução explícita, sendo necessário o uso de processos iterativos.

3.3.1 Processo Iterativo

O procedimento numérico por meio de métodos iterativos é uma maneira de resolver problemas de mínimos quadrados, quando temos um modelo não linear nos parâmetros, na qual é determinada a solução do sistema de equações normais.

No método de Gauss-Newton, é feita a linearização da função não linear, descrita abaixo.

Seja $y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i$ o modelo não linear e $\hat{\boldsymbol{\theta}}_0$ o vetor de valores iniciais tal que:

$$F'(\hat{\boldsymbol{\theta}}_0)[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0)] \approx \mathbf{0} .$$

Por expansão de série de Taylor de primeira ordem pelo vetor $\hat{\boldsymbol{\theta}}_0$:

$$f(\hat{\boldsymbol{\theta}}) \approx f(\hat{\boldsymbol{\theta}}_0) + F(\hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0) \quad (1)$$

$$F'(\hat{\boldsymbol{\theta}})[\mathbf{Y} - f(\hat{\boldsymbol{\theta}})] \approx \mathbf{0} \quad (2)$$

Aplicando (1) em (2):

$$F'(\hat{\boldsymbol{\theta}})[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0) - F(\hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0)] \approx \mathbf{0}$$

A teoria de inferência estatística que envolve modelo de regressão não linear baseia-se na aproximação linear por série de Taylor, por essa razão, como no caso linear, temos:

$F(\hat{\boldsymbol{\theta}}) = \mathbf{X}$, a matriz de derivadas parciais. Desta forma:

$$\begin{aligned} F'(\hat{\boldsymbol{\theta}})[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0) - F(\hat{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0)] &\approx \mathbf{0} \\ \mathbf{X}'[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0)] - \mathbf{X}'[\mathbf{X}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0)] &\approx \mathbf{0} \\ \mathbf{X}'(\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0)) &= \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0) \end{aligned}$$

Multiplicando, à esquerda, ambos os lados da igualdade por $(\mathbf{X}'\mathbf{X})^{-1}$, temos:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_0)] = \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0$$

Seja $\hat{\boldsymbol{\theta}}_0$ o vetor dos valores iniciais dos parâmetros. Observe que as derivadas, assim como f , são avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0$

Dessa forma, tomando $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_t$ e $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{t+1}$ obtém-se:

$$\hat{\boldsymbol{\theta}}_{t+1} \approx \hat{\boldsymbol{\theta}}_t + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{Y} - f(\hat{\boldsymbol{\theta}}_t)]$$

que é a expressão para a t-ésima iteração do processo iterativo de Gauss-Newton.

Isso gera um processo iterativo que prossegue até que algum critério adotado para convergência seja atingido.

Um problema dos métodos iterativos é determinar os valores iniciais $\hat{\boldsymbol{\theta}}_0$ das estimativas. Utilizar as interpretações biológicas dos parâmetros para

encontrar os valores iniciais dos estimadores iniciais dos parâmetros pode ser uma alternativa. A escolha de valores de $\hat{\theta}_0$ próximos aos verdadeiros valores dos parâmetros facilita a convergência do processo iterativo. Uma alternativa para isso é buscar na literatura valores que representem a realidade, como peso à maturidade, para ser utilizados como valores iniciais, uma vez que, os parâmetros apresentam interpretações biológicas.

3.3.2 Derivadas parciais para funções segmentadas não lineares

Após a convergência do procedimento iterativo de Gauss Newton para um vetor $\hat{\theta}$ de estimativas, podemos obter o quadrado médio residual que é um estimador da variância dos erros aleatórios dado por:

$$QME = \frac{\sum_{i=1}^n [y_i - f(x_i, \hat{\theta})]^2}{n - p}$$

em que, n é o número de observações (número de pesagem), e p é o número de parâmetros.

As derivadas parciais, em relação aos parâmetros, da regressão segmentada linear-linear abaixo, para o peso (y) em função da idade (x):

$$y = \beta_{00} + \beta_{01}x + \varepsilon, \text{ se } x < t_1$$

$$y = \beta_{00} + \beta_{01}x + \beta_1(x - t_1) + \varepsilon, \text{ se } x \geq t_1$$

São dadas por:

$$\text{Se } x < t_1 \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_{00}} = 1; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_{01}} = x; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_1} = 0; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial t_1} = 0$$

$$\text{Se } x \geq t_1 \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_{00}} = 1; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_{01}} = x; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \beta_1} = x - t_1; \quad \frac{\partial f(\boldsymbol{\theta})}{\partial t_1} = -\beta_1$$

Dessa forma, tem-se que a matriz de derivadas parciais, $F(\theta) = {}_n X_p$, sendo n o número de pesagens, ou seja, os n tempos em que os animais foram pesados e p o número de parâmetros do modelo e $x_m < t_1 < x_{m+1}$ para algum índice m , será dada por:

$$F(\boldsymbol{\theta}) = {}_n X_p = \left[\frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \begin{pmatrix} \frac{\partial f(\mathbf{X}_1, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_1, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_1, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_1, \boldsymbol{\theta})}{\partial t_1} \\ \frac{\partial f(\mathbf{X}_2, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_2, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_2, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_2, \boldsymbol{\theta})}{\partial t_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f(\mathbf{X}_m, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_m, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_m, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_m, \boldsymbol{\theta})}{\partial t_1} \\ \frac{\partial f(\mathbf{X}_{m+1}, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_{m+1}, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_{m+1}, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_{m+1}, \boldsymbol{\theta})}{\partial t_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f(\mathbf{X}_{n-1}, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_{n-1}, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_{n-1}, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_{n-1}, \boldsymbol{\theta})}{\partial t_1} \\ \frac{\partial f(\mathbf{X}_n, \boldsymbol{\theta})}{\partial \beta_{00}} & \frac{\partial f(\mathbf{X}_n, \boldsymbol{\theta})}{\partial \beta_{01}} & \frac{\partial f(\mathbf{X}_n, \boldsymbol{\theta})}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_n, \boldsymbol{\theta})}{\partial t_1} \end{pmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$F(\boldsymbol{\theta}) = {}_n X_4 = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & 0 & 0 \\ 1 & x_{m+1} & (x_{m+1} - t_1) & -\beta_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & (x_{n-1} - t_1) & -\beta_1 \\ 1 & x_n & (x_n - t_1) & -\beta_1 \end{pmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

A matriz $X'X$ é da seguinte forma:

$$\begin{aligned}
 a_{11} &= n \\
 a_{12} &= a_{21} = \sum_{i=1}^n x_i \\
 a_{13} &= a_{31} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n (x_i - t_1) \\
 a_{14} &= a_{41} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n -\beta_1 \\
 a_{22} &= \sum_{i=1}^n x_i^2 \\
 a_{23} &= a_{32} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n x_i(x_i - t_1) \\
 a_{24} &= a_{42} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n x_i(-\beta_1) \\
 a_{33} &= \sum_{i=1}^m 0 + \sum_{i=m+1}^n (x_i - t_1)^2 \\
 a_{34} &= a_{43} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n -\beta_1(x_i - t_1) \\
 a_{34} &= a_{43} = \sum_{i=1}^m 0 + \sum_{i=m+1}^n -\beta_1(x_i - t_1) \\
 a_{44} &= \sum_{i=1}^m 0 + \sum_{i=m+1}^n (-\beta_1)^2
 \end{aligned}$$

Segundo Draper e Smith (1998 citados por SAVIAN; MUNIZ, 2007) a matriz assintótica de variância dos estimadores é estimada da seguinte forma:

$$\text{vâr}(\hat{\boldsymbol{\theta}}) = (\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}^2 = (\mathbf{X}'\mathbf{X})^{-1} QME$$

$$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \widehat{\text{var}}(\hat{\theta}_1) & \widehat{\text{cov}}(\hat{\theta}_1, \hat{\theta}_2) & \cdots & \widehat{\text{cov}}(\hat{\theta}_1, \hat{\theta}_p) \\ \widehat{\text{cov}}(\hat{\theta}_2, \hat{\theta}_1) & \widehat{\text{var}}(\hat{\theta}_2) & \cdots & \widehat{\text{cov}}(\hat{\theta}_2, \hat{\theta}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\text{cov}}(\hat{\theta}_p, \hat{\theta}_1) & \vdots & \ddots & \widehat{\text{var}}(\hat{\theta}_p, \hat{\theta}_p) \end{bmatrix}$$

em que, p é o número de parâmetros, \mathbf{X} é a matriz de derivadas parciais do modelo em relação aos parâmetros e QME é o quadrado médio do resíduo.

Dessa forma, define-se o intervalo de confiança assintótico para θ_i como:

$$ic(\theta_i) = \hat{\theta}_i \pm t_{\left(gl, \frac{\alpha}{2}\right)} \sqrt{\widehat{\text{var}}(\hat{\theta}_i)}$$

em que, t = valor crítico de t de student, gl = é o número de graus de liberdade do erro, α = é o nível de significância.

Para os demais, os tipos de regressões segmentadas, a matriz de derivadas parciais, a matriz $X'X$, a matriz de variâncias e covariâncias, assim como o intervalo de confiança são obtidos de modo análogo ao da regressão segmentada linear-linear.

3.4 Estimação dos parâmetros da regressão segmentada linear nos parâmetros

No caso em que os nós das regressões segmentadas são conhecidos ou pré-determinados pelo pesquisador, tem-se o caso de regressão segmentada linear nos parâmetros. Um dos casos de regressão segmentada é quando os

dados podem ser descritos por duas funções lineares simples que são unidas por um ponto de junção, o nó.

O modelo para dois segmentos de primeiro grau, linear-linear, com um nó t_1 , é dado por:

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i, & \text{para } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \varepsilon_i, & \text{para } x_{1i} \geq t_1 \end{cases}$$

Esses dois modelos podem ser escritos como um modelo de regressão múltipla se outra variável independente x_2 for definida (KAPS; LAMBERSON, 2004) por:

$$\begin{cases} x_{2i} = 0 & \text{para } x_{1i} < t_1 \\ x_{2i} = (x_{1i} - t_1) & \text{para } x_{1i} \geq t_1 \end{cases}$$

Dessa forma o novo modelo de regressão múltipla é dado por:

$$y = \beta_{00} + \beta_{01}x_{1i} + \beta_1x_{2i} + \varepsilon_i \quad (3)$$

Em que a matriz \mathbf{X} , para $x_m < t_1 < x_{m+1}$, é definida por:

$$\begin{pmatrix} 1 & x_{11} & 0 \\ 1 & x_{12} & 0 \\ 1 & x_{13} & 0 \\ & \vdots & \vdots \\ 1 & x_{1m} & 0 \\ 1 & x_{1(m+1)} & (x_{1(m+1)} - t_1) \\ & \vdots & \vdots \\ 1 & x_{1n} & (x_{1n} - t_1) \end{pmatrix}$$

A terceira coluna recebe zero para valores de x_{1i} menores que t_1 e recebe $[x_{1i} - t_1]$ para valores de x_{1i} maiores que t_1 .

Logo, o modelo de regressão (3) pode ser expresso, utilizando o valor do nó determinado anteriormente, por:

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \beta_1(0) + \varepsilon_i & \text{para } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1x_{2i} + \varepsilon_i & \text{para } x_{1i} \geq t_1 \end{cases}$$

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i & \text{para } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \varepsilon_i & \text{para } x_{1i} \geq t_1 \end{cases}$$

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i & \text{para } x_{1i} < t_1 \\ (\beta_{00} - \beta_1t_1) + (\beta_{01} + \beta_1)x_{1i} + \varepsilon_i & \text{para } x_{1i} \geq t_1 \end{cases}$$

Com essa parametrização os dois segmentos lineares se interceptam no ponto de abscissa igual a t_1 .

O modelo para três segmentos, linear-linear-linear, com dois nós t_1 e t_2 , é dado por:

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i, & \text{para } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \varepsilon_i, & \text{para } t_1 \leq x_{1i} < t_2 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \beta_2(x_{1i} - t_2) + \varepsilon_i, & \text{para } x_{1i} \geq t_2 \end{cases}$$

Se outras variáveis independentes x_2 e x_3 forem definidas, esses três modelos podem ser escritos como um modelo de regressão múltipla dado por:

$$y = \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{2i}) + \beta_2(x_{3i}) + \varepsilon_i$$

Em que:

$$\begin{cases} x_{2i} = 0 \text{ para } x_{1i} < t_1 \\ x_{2i} = (x_{1i} - t_1) \text{ para } x_{1i} \geq t_1 \end{cases}$$

$$\begin{cases} x_{3i} = 0 \text{ para } x_{1i} < t_2 \\ x_{3i} = (x_{1i} - t_2) \text{ para } x_{1i} \geq t_2 \end{cases}$$

A matriz \mathbf{X} , para $x_m < t_1 < x_{m+1}$ e $x_r < t_2 < x_{r+1}$, $r > m$ pode, então, ser dada por:

$$\begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{12} & 0 & 0 \\ 1 & x_{13} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1m} & 0 & 0 \\ 1 & x_{1(m+1)}[x_{1(m+1)} - t_1] & 0 & \\ & \vdots & \vdots & \ddots \\ & 1 & x_{1r}[x_{1r} - t_1] & 0 \\ 1 & x_{1(r+1)}[x_{1(r+1)} - t_1][x_{1(r+1)} - t_2] & & \\ & \vdots & \vdots & \ddots \\ 1 & x_n[x_{1n} - t_1][x_{1n} - t_2] & & \end{pmatrix}$$

A segunda coluna recebe zero para valores de x_{1i} menores que t_1 e assume $[x_{1i} - t_1]$ para valores maiores que t_1 , terceira coluna recebe zero para valores de x_{1i} menores que t_2 e recebe $[x_{1i} - t_2]$ para valores de x_{1i} maiores que t_2 , sendo $t_1 < t_2$ os nós.

Logo, modelo de regressão pode ser expresso utilizando-se os valores dos nós determinados anteriormente, por:

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \beta_1 \cdot 0 + \beta_2 \cdot 0 + \varepsilon_i & \text{para } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \beta_2 \cdot 0 + \varepsilon_i, & \text{para } t_1 \leq x_{1i} < t_2 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \beta_2(x_{1i} - t_2) + \varepsilon_i, & \text{para } t_2 \leq x_{1i} \end{cases}$$

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i & \text{para } x_{1i} < t_1 \\ (\beta_{00} - \beta_1 t_1) + (\beta_{01} + \beta_1)x_{1i} + \varepsilon_i & \text{para } t_1 \leq x_{1i} < t_2 \\ (\beta_{00} - \beta_1 t_1 - \beta_2 t_2) + (\beta_{01} + \beta_1 + \beta_2)x_{1i} + \varepsilon_i, & \text{para } t_2 \leq x_{1i} \end{cases}$$

Com essa parametrização, as regressões lineares se interceptam nos pontos de abscissa igual a t_1 e t_2 .

O modelo para dois segmentos, linear-platô, uma de primeiro grau e outra platô, com o nó t_1 é dado por:

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + \varepsilon_i, & \text{para } x_{1i} \leq t_1 \\ \beta_{00} + \beta_{01}t_1 + \varepsilon_i, & \text{para } x_{1i} > t_1 \end{cases}$$

A primeira equação pode ser estimada por meio de um modelo de regressão linear simples em que a variável independente é dada pelos valores de x_{1i} menores que t_1 .

A segunda equação é uma constante dada pela primeira equação avaliada no ponto de abscissa t_1 , predeterminado pelo pesquisador.

3.5 Soma de quadrados extra

A soma de quadrados extra será utilizada para verificar o número de segmentos significativos das regressões segmentadas lineares.

No método de soma de quadrados extra, é verificada a redução na soma de quadrados do erro e o acréscimo na soma de quadrados da regressão quando

uma ou mais variáveis preditoras são acrescentadas no modelo de regressão múltipla, dado que outras variáveis preditoras já estão contidas no modelo.

A soma de quadrados do erro para a regressão múltipla com as variáveis X_1 e X_2 incluídas no modelo $SQE(X_1, X_2)$, é menor do que a soma de quadrados do erro quando apenas a variável X_1 está no modelo $SQE(X_1)$. A diferença entre $SQE(X_1)$ e $SQE(X_1, X_2)$ é denominada de soma de quadrados extra.

Esta redução na soma de quadrados do erro é o resultado da adição de X_2 dado que X_1 já está contida no modelo. Esta soma de quadrados extra representada por $SQR(X_2|X_1)$ mede o efeito marginal da adição de X_2 no modelo de regressão quando X_1 já está incluída no modelo.

Equivalentemente, pode-se calcular a soma de quadrados extra como:

$$SQR(X_2 | X_1) = SQR(X_1, X_2) - SQR(X_1) ;$$

em que $SQR(X_1, X_2)$ é a soma de quadrados da regressão quando X_1 e X_2 estão incluídas no modelo e $SQR(X_1)$ é a soma de quadrados da regressão quando apenas X_1 está no modelo.

Da mesma forma, pode-se considerar a soma de quadrados extra de X_3 dado que X_1 e X_2 já estão incluídas no modelo, representada por:

$$SQE(X_3 | X_1, X_2) = SQE(X_1, X_2) - SQE(X_1, X_2, X_3)$$

em que, $SQE(X_1, X_2, X_3)$ é a soma de quadrados do erro quando X_1 , X_2 e X_3 estão incluídas no modelo e $SQE(X_1, X_2)$ é a soma de quadrados do erro quando apenas X_1 e X_2 estão no modelo.

De forma equivalente, a soma de quadrados extra de X_3 dado que X_1 e X_2 já estão incluídas no modelo também pode ser representada por:

$$SQR(X_3 | X_1, X_2) = SQR(X_1, X_2, X_3) - SQR(X_1, X_2)$$

em que, $SQR(X_1, X_2, X_3)$ é a soma de quadrados da regressão quando X_1 , X_2 e X_3 estão incluídas no modelo e $SQR(X_1, X_2)$ é a soma de quadrados da regressão quando apenas X_1 e X_2 estão no modelo.

Outra soma de quadrados extra que mede o efeito da adição de X_2 e X_3 ao modelo quando X_1 já está no modelo é:

$$SQE(X_2, X_3 | X_1) = SQE(X_1) - SQE(X_1, X_2, X_3)$$

De forma análoga:

$$SQR(X_2, X_3 | X_1) = SQR(X_1, X_2, X_3) - SQR(X_1)$$

A decomposição da soma de quadrado de regressão em soma de quadrados extra é dada por:

$$SQR(X_1, X_2, X_3) = SQR(X_1) + SQR(X_2 | X_1) + SQR(X_3 | X_1, X_2)$$

A Tabela 2 da análise da variância a seguir contém a decomposição da soma de quadrado de regressão, SQR , para o caso de três variáveis explanatórias.

Tabela 2 Esquema para análise da variância com a decomposição da soma de quadrados da regressão para três variáveis explanatórias

Causas de Variação	Soma de Quadrados	Graus de Liberdade	Quadrados Médios
Regressão	SQR (X_1, X_2, X_3)	3	QMR (X_1, X_2, X_3)
X_1	SQR (X_1)	1	QMR (X_1)
$X_2 X_1$	SQR ($X_2 X_1$)	1	QMR ($X_2 X_1$)
$X_3 X_1, X_2$	SQR ($X_3 X_1, X_2$)	1	QMR ($X_3 X_1, X_2$)
Erro	SQE (X_1, X_2, X_3)	n-4	QME (X_1, X_2, X_3)
Total	SQ Total	n-1	

A importância do cálculo das somas de quadrados extra é a possibilidade de se fazer testes de hipóteses sobre os coeficientes de regressão em que o interesse é saber se certas variáveis explanatórias podem ser retiradas ou incluídas no modelo de regressão.

No estudo em questão, o teste F será utilizado para verificar se certas variáveis podem ser retiradas do modelo de regressão múltipla, ou seja, se os dados podem ser descritos por um ou dois ou três segmentos, uma vez que o modelo segmentado linear pode ser descrito por uma regressão linear múltipla, com as variáveis criadas pela diferença entre a variável independente e os nós das regressões segmentadas.

3.6 Testes de hipóteses sobre os coeficientes de regressão usando as somas de quadrados extras

Deseja-se saber se o termo $\beta_k X_k$ pode ser retirado do modelo, ou seja, deseja-se testar as seguintes hipóteses (KAPS; LAMBERSON, 2004):

$$H_0 : \beta_k = 0$$

$$H_a : \beta_k \neq 0$$

Para o modelo completo de primeira ordem, com três variáveis preditoras (linear-linear-linear):

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Ajusta-se o modelo completo e obtém-se a soma de quadrados do erro, $SQE(completo) = SQE(X_1, X_2, X_3)$, com $n-4$ graus de liberdade, uma vez que há 4 parâmetros no modelo.

Para avaliar se X_3 deve ser descartada, tem-se as hipóteses:

$$H_0 : \beta_3 = 0$$

$$H_a : \beta_3 \neq 0$$

Sob a hipótese nula, o modelo com dois segmentos é dado por:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Ajusta-se o modelo reduzido anterior e obtém-se $SQE(reduzido) = SQE(X_1, X_2)$, com $n-3$ graus de liberdade.

Lembrando que:

$$\hat{\beta} = (X'X)^{-1}X'Y;$$

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

O teste estatístico é dado por:

$$F = \frac{SQE(\text{Reduzido}) - SQE(\text{Completo})}{gl_{\text{Reduzido}} - gl_{\text{Completo}}} \div \frac{SQE(\text{Completo})}{gl_{\text{Completo}}}$$

$$F = \frac{SQE(X_1, X_2) - SQE(X_1, X_2, X_3)}{(n-3) - (n-4)} \div \frac{SQE(X_1, X_2, X_3)}{(n-4)}$$

Percebe-se que no numerador tem-se a soma de quadrados extra devido a X_3 :

$$SQE(X_1, X_2) - SQE(X_1, X_2, X_3) = SQE(X_3 | X_1, X_2)$$

Dessa forma, o teste estatístico pode ser escrito por:

$$F = \frac{\frac{SQE(X_3 | X_1, X_2)}{(1)}}{\frac{SQE(X_1, X_2, X_3)}{n-4}} = \frac{QMR(X_3 | X_1, X_2)}{QME(X_1, X_2, X_3)}$$

Se a hipótese nula não for rejeitada, conclui-se que a variável X_3 pode ser retirada do modelo que já contém X_1 , X_2 , ou seja, a regressão segmentada não apresenta o terceiro segmento.

Em resumo, ajusta-se os modelos de regressão linear acrescentando sequencialmente as variáveis. A cada entrada de variável, é realizado o teste estatístico F, avaliando a Soma de Quadrados Extra devida à variável que está sendo adicionada no modelo e é verificado se esta variável pode ou não ser

retirada do modelo, ou seja, se mais um segmento pode ou não ser incluído na regressão segmentada.

3.7 Avaliadores da qualidade do ajuste

Para comparar a qualidade de ajuste, foram utilizadas medidas de comparação da qualidade do modelo, o critério de informação de Akaike (AIC) (AKAIKE, 1974), critério de informação bayesiano (BIC), baseados no máximo da função de verossimilhança sendo dependentes do número de observações e parâmetros do modelo em estudo, e coeficiente de determinação ajustado (R_{aj}^2), cujas equações são dadas, respectivamente, por:

$$AIC = -2 \log[L(X; \hat{\theta})] + 2p$$

$$BIC = -2 \log[L(X; \hat{\theta})] + p \log(n)$$

$$R_{aj}^2 = R^2 - \left(\frac{p-1}{n-p} \right) (1 - R^2)$$

em que, $\log[L(X; \theta)]$ é o valor do logaritmo natural da função de verossimilhança conjunta maximizada, n é o número de observações utilizadas e p é o número de parâmetros do modelo.

No modelo de regressão segmentada linear-linear, tem-se para o nó t_1 , em que $x_m < t_1 < x_{m+1}$, para algum índice m , que o logaritmo natural da função de verossimilhança conjunta (HERNANDEZ, 2010) é dada por:

$$\begin{aligned} \ln_{LL}(y_i) = & -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m [y_i - (\beta_{00} + \beta_{01}x_i)]^2 \\ & - \frac{1}{2\sigma^2} \sum_{i=m+1}^n \{y_i - [\beta_{00} + \beta_{01}x_i + \beta_1(x_i - t_1)]\}^2 \end{aligned}$$

De modo análogo, no modelo de regressão segmentado linear-linear-linear, tem-se, para os nós t_1 e t_2 em que $x_m < t_1 < x_{m+1}$, para algum índice m e $x_r < t_2 < x_{r+1}$, para algum índice $r > m$, que o logaritmo natural da função de verossimilhança conjunta é dada por:

$$\begin{aligned} \ln_{LLL}(y_i) = & -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m [y_i - (\beta_{00} + \beta_{01}x_{1i})]^2 \\ & - \frac{1}{2\sigma^2} \sum_{i=m+1}^r \{y_i - [\beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1)]\}^2 \\ & - \frac{1}{2\sigma^2} \sum_{i=r+1}^n \{y_i - [\beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \beta_2(x_{1i} - t_2)]\}^2 \end{aligned}$$

Para o modelo linear-platô com o nó t_1 em que $x_m < t_1 < x_{m+1}$, para algum índice m , temos que o logaritmo natural da função de verossimilhança conjunta é representado por:

$$\begin{aligned} \ln_{LP}(y_i) = & -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m [y_i - (\beta_{00} + \beta_{01}x_i)]^2 \\ & - \frac{1}{2\sigma^2} \sum_{i=m+1}^n \{y_i - [\beta_{00} + \beta_{01}t_1]\}^2 \end{aligned}$$

3.8 Simulação de dados

Para a realização da metodologia proposta, implementou-se um programa utilizando *software* livre R (R DEVELOPMENT CORE TEAM, 2013). Utilizou-se 1.000 iterações e as idades utilizadas para gerar os pesos foi do grupo genético DRL, com tamanho da amostra de 195 observações.

Realizou-se a simulação Monte Carlo de dados de pesagem de ovinos para avaliar as propriedades dos estimadores dos parâmetros, como consistência e tendência, da regressão segmentada linear nos parâmetros. Os resultados

apresentados foram gerados considerando os ajustes linear, linear-linear e linear-linear-linear e correspondem aos modelos apresentados abaixo, respectivamente:

$$y = \beta_{00} + \beta_{01}x_{1i} + e$$

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + e, & \text{se } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + e, & \text{se } x_{1i} \geq t_1 \end{cases}$$

$$y = \begin{cases} \beta_{00} + \beta_{01}x_{1i} + e, & \text{se } x_{1i} < t_1 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + e, & \text{se } t_1 \leq x_{1i} < t_2 \\ \beta_{00} + \beta_{01}x_{1i} + \beta_1(x_{1i} - t_1) + \beta_2(x_{1i} - t_2) + e, & \text{se } x_{1i} > t_2 \end{cases}$$

em que, os valores paramétricos assumidos na simulação, $\beta_{00} = 4,48$; $\beta_{01} = 0,11$; $\beta_1 = 0,13$; $\beta_2 = -0,10$, foram obtidos por meio de ajustes dos dados experimentais do grupo genético DRL. Os pontos de junção, t_1 e t_2 , dados respectivamente por $t_1 = 84$ e $t_2 = 168$ foram considerados de acordo com Souza Júnior et al. (2009).

Com o intuito de retratar as características heterocedásticas encontradas em dados de crescimento animal (MAZUCHELI; SOUZA; PHILIPPSEN, 2011), foram avaliados cenários, cujos pesos dos animais foram simulados considerando a variância do erro crescente no decorrer do tempo, de acordo com

$$e \sim N(0; 0,005(idade^{1,67}))$$

No primeiro cenário foram gerados pesos considerando o modelo com um segmento, no segundo cenário, o modelo com dois segmentos e no terceiro, com três segmentos. Em todos os casos, foram realizados os ajustes, com os

dados gerados, considerando um, dois e três segmentos com o intuito de verificar se a soma de quadrados extras é uma ferramenta estatística apropriada para determinar o número de segmentos significativos nas regressões segmentadas lineares nos parâmetros.

Posteriormente, foi avaliado, em cada ajuste, o AIC, BIC e R_{aj}^2 para avaliar a qualidade do ajuste das regressões segmentadas, além do valor-p do teste estatístico F para avaliar a soma de quadrado extra.

Foram observados também, os valores da esperança matemática dos estimadores a fim de verificar se os mesmos são não viesados.

Com o intuito de avaliar a consistência dos estimadores dos parâmetros da regressão segmentada linear nos parâmetros, aumentou-se gradativamente o tamanho das amostras (número de pesagens) a fim de se observar se as distribuições amostrais dos estimadores se tornaram cada vez mais concentrada em torno do verdadeiro parâmetro populacional.

4 RESULTADOS E DISCUSSÃO

4.1 Regressão segmentada ajustada a dados de ovinos

Nas Tabelas 3, 4 e 5 são mostradas as estimativas dos parâmetros das funções segmentadas não linear, para os três grupos genéticos analisados, assim como os respectivos intervalos de confiança, com nível de confiança de 95%.

Tabela 3 Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DMN, com os respectivos intervalos de confiança

Grupo				
Genético	Parâmetros	L-L	L-P	L-L-L
DMN	β_{00}	2,96 (1,55;4,37)	4,61 (3,57;5,65)	3,11 (1,86;4,36)
	β_{01}	0,12 (0,10; 0,14)	0,10 (0,08;0,12)	0,12 (0,10;0,14)
	β_1	-0,05 (-0,07;-0,03)	-	-0,08 (-0,14;0,02)
	β_2	-	-	0,07 (-0,008;0,15)
	t_1	118 (83,66;152;34)	246 (220,52;271,48)	144 (114,80;173,20)
	t_2	-	-	211 (158,96;263,04)

Tabela 4 Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DSI, com os respectivos intervalos de confiança

	β_{00}	4,45 (3,12;5,78)	4,49 (3,14;5,84)	4,55 (3,24;5,86)
	β_{01}	0,16 (0,14;0,18)	0,155 (0,14;0,17)	0,153 (0,133;0,17)
DSI	β_1	-0,17 (-0,19;-0,15)	-	-0,135 (-0,17;-0,09)
	β_2	-	-	-0,17 (-0,27;-0,07)
	t_1	162 (148,73;175,27)	157,1 (145,79;168,41)	156 (137,71;174,29)
	t_2	-	-	267,5 (243,12;291,88)

Tabela 5 Estimativas dos parâmetros das funções segmentadas linear-linear (L-L), linear-platô (L-P) e linear-linear-linear (L-L-L) não linear nos parâmetros (nó desconhecido) para a descrição dos pesos (kg) de ovinos do grupo genético DRL, com os respectivos intervalos de confiança

DRL	β_{00}	4,37 (3,33;5,41)	4,57 (3,55;5,59)	3,92 (1,88;5,96)
	β_{01}	0,12 (0,10;0,14)	0,11 (0,09;0,13)	0,132 (0,07;0,19)
	β_1	-0,111 (-0,13;-0,09)	-	-0,02 (-0,08;0,04)
	β_2	-	-	-0,12 (-0,14;-0,10)
	t_1	188 (167,07;208,93)	197,9 (183,24;212,56)	49 (-51,43;149,43)
	t_2	-	-	202,9 (180,97;224,83)

t_1 é o nó entre a primeira e a segunda fase; t_2 é o nó entre a segunda e terceira fase.

O intervalo de confiança para os parâmetros das regressões segmentadas linear-linear e linear-platô, do grupo genético DMN, não contém o zero, entretanto para a regressão segmentada linear-linear-linear os intervalos para os parâmetros β_1 e β_2 contém o zero, indicando que, para esse grupo genético, estes parâmetros não são significativos.

Para o grupo genético DSI, nenhum dos intervalos de confiança para os parâmetros das regressões segmentadas linear-linear, linear-platô e linear-linear-linear contém o zero, dessa forma, todos estes ajustes podem ser utilizados para descrever o crescimento dos animais deste grupo.

Os parâmetros das regressões segmentadas linear-linear e linear-platô, do grupo genético DRL, apresentam intervalo de confiança que não contém o zero, no entanto, para a regressão segmentada linear-linear-linear, os intervalos para os parâmetros β_1 e t_1 contém o zero, indicando que, para esse grupo genético, o modelo com três segmentos não se ajusta bem.

Para as funções segmentadas linear-linear, considerando os três grupos genéticos, observou-se que o ganho de peso diário para o primeiro segmento variou de 120g/dia a 160g/dia (parâmetro β_{01}), sendo o grupo genético DSI o que apresentou maior estimativa de ganho de peso nessa fase.

Nas regressões segmentadas linear-platô, o ganho de peso diário estimado, até o peso adulto, ficou entre 100g/dia e 155g/dia, sendo também o grupo DSI (Dorper x Santa Inês) o que apresentou maior estimativa.

Esses resultados corroboram com Sousa e Moraes (2000) quando afirmaram que o alto tamanho corporal na raça Santa Inês tem sido objeto de propaganda.

Nas regressões segmentadas linear-linear-linear, o ganho de peso estimado na primeira fase foi de 120g/dia a 153g/dia (parâmetro β_{00}).

Resultados semelhantes foram encontrados por Piola Júnior et al. (2009) que avaliaram a influência de diferentes níveis de energia na ração sobre o desempenho e características da carcaça de cordeiros mestiços Texel. Os autores obtiveram as estimativas do ganho de peso médio diário dos animais entre 120 e 208g/dia.

Por outro lado, o ganho de peso médio diário estimado foi inferior ao observado por Brochier e Carvalho (2008) que avaliaram o consumo de nutrientes, o ganho de peso e a conversão alimentar de cordeiros machos, não castrados da raça Texel, alimentados em confinamento com dietas contendo diferentes níveis de energia e obtiveram ganhos de peso médio diário entre 212 e 257g/dia.

As estimativas superiores, para os pesos dos animais, encontrados por estes autores, pode ser decorrência de que os animais em estudo são machos e segundo Costa Júnior et al. (2006), a diferença entre machos e fêmeas, para peso corporal como para medidas morfométricas, acentuam-se à medida que as idades dos animais aumentam.

A mudança no comportamento do crescimento dos animais, estimado na regressão segmentada linear-linear, aconteceu quando os animais estavam com idade de 118, 162 e 188 dias, respectivamente para os grupos DMN, DSI e DRL (parâmetro t_1). Após essas idades os animais apresentaram um menor desenvolvimento que nas idades iniciais ($\hat{\beta}_1 < 0$).

Esses valores são superiores aos obtidos por Souza et al. (2011) que analisaram os modelos não lineares usuais Brody, Von Bertalanffy, Richards, Logístico e Gompertz para descrever o crescimento em ovinos da raça Morada Nova criados no estado da Bahia e observaram que o ganho de peso foi crescente até atingir o máximo, aproximadamente 116 g/dia aos 38 dias, após esse período os animais tiveram uma redução no ganho de peso. Posteriormente aos 120 dias, o ganho de peso foi muito pequeno, resultado semelhante foi encontrado no grupo genético DMN em que os animais praticamente não ganharam peso após os 118 dias.

O peso adulto dos animais (peso no platô) foi de 29,21; 28,84 e 26,34kg, respectivamente para os grupos DMN, DSI, DRL. Estes resultados concordam com os encontrados por outros autores, Carneiro et al. (2007) estimaram curvas de crescimento individuais, por meio do modelo Logístico, e as estimativas para o peso adulto dos grupos DSI, DMN, DRL foram 28,680; 27,144 e 26,085kg respectivamente.

Zapata et al. (2001) estudaram as características de rendimento e composição da carcaça de ovinos do nordeste brasileiro, utilizados para produção de carne e afirmaram que os ovinos adultos do nordeste brasileiro

podem apresentar peso vivo de até 30 kg, com média geralmente inferior à de animais de raças de clima temperado criados na região sul.

A idade em que os animais dos grupos genéticos DSI, DRL e DMN atingiram o peso adulto foram 157, 198 e 246 dias (parâmetro t_1 das regressões segmentadas linear-platô), respectivamente. Assim, verifica-se que o grupo DSI apresenta maior precocidade.

Esses resultados são próximos aos observados por Sarmiento et al. (2006) que, através da taxa de crescimento absoluta, comprovou que o ganho de peso dos animais atingiu o máximo em torno de 0,170 kg, aos 52 dias, nas idades posteriores o ganho de peso decresceu, em torno dos seis meses os animais praticamente não ganharam peso.

Para a função segmentada linear-linear-linear, tem-se que o ganho de peso diário ficou entre 120 e 150g/dia para o primeiro segmento (parâmetro β_{01}) e para o segundo, esses valores ficaram entre 18,1 a 109,8 g/dia ($\beta_{01} + \beta_1$), com idades 144, 156 e 49 dias respectivamente para os grupos DMN, DSI e DRL. No terceiro segmento, o ganho de peso diário ($\beta_{01} + \beta_1 + \beta_2$) apresentou valores negativos -0,1543 kg/dia e -0,0082 kg/dia para os grupos DSI e DRL, respectivamente, enquanto que para o grupo genético DMN foi de 0,104 kg/dia.

Na Tabela 6, estão apresentados os avaliadores da qualidade de ajuste AIC, BIC e R_{aj}^2 para os grupos genéticos DMN, DRL e DSI, das regressões segmentadas não lineares ajustadas aos dados de ovinos cruzados.

Tabela 6 Resultados dos avaliadores da qualidade de ajuste, critério de informação de Akaike (AIC), critério de informação bayesiano (BIC) e coeficiente de determinação ajustado (R^2_{aj}) para as funções segmentadas não lineares, considerando os três grupos genéticos (DMN, DSI e DRL)

Avaliadores	Modelos	Grupos genéticos		
		DMN	DSI	DRL
AIC	L-L	773,69	1185,22	1028,59
	L-P	649,81	1184,99	1026,60
	L-L-L	774,46	1177,42	1031,56
BIC	L-L	785,76	1198,57	1041,68
	L-P	658,86	1195,01	1036,42
	L-L-L	792,57	1197,45	1051,19
R^2_{aj} %	L-L	82,80	79,13	81,23
	L-P	81,24	79,05	81,33
	L-L-L	91,08	90,09	89,50

De acordo com os resultados apresentados na Tabela 6, para o grupo DMN, a função segmentada linear-platô apresentou melhor qualidade de ajuste, uma vez que as estimativas do AIC e BIC apresentaram os menores valores para esse modelo, além de apresentar R^2_{aj} superior a 80%.

Através das estimativas do R^2_{aj} e AIC, utilizadas na indicação do melhor modelo, a função segmentada linear-linear-linear se ajustou melhor aos dados de peso-idade para o grupo genético DSI. Embora o AIC e BIC, nesse grupo genético, apresentaram estimativas bem próximas para as três regressões segmentadas.

O mesmo é percebido nos intervalos de confiança para os parâmetros desse grupo, uma vez que, em nenhum intervalo está contido o zero e, portanto, todos os modelos podem ser utilizados (Tabela 4).

Entretanto, a função segmentada linear-platô apresentou menor valor para o BIC, no grupo DSI, e os resultados dos demais avaliadores da qualidade de ajuste foram próximos dos encontrados na função segmentada linear-linear-linear, com isso a função segmentada linear-platô também apresentou um bom ajuste para esse grupo.

Ao comparar as estimativas dos avaliadores da qualidade do ajuste pode-se verificar que para o grupo genético DRL, a função segmentada linear-platô foi a que melhor se ajustou aos dados, uma vez que para esse modelo o AIC e BIC apresentaram menores valores e R_{aj}^2 de 81,34%.

Os avaliadores da qualidade de ajuste AIC e BIC (Tabela 6) indicaram a regressão segmentada linear-linear-linear como o pior ajuste para o grupo genético DMN, o que está de acordo com os resultados encontrados no intervalo de confiança para os parâmetros desse modelo (Tabela 3) uma vez que o zero estava presente no intervalo do parâmetro β_2 indicando que o terceiro segmento não se adéqua bem aos dados, entretanto o coeficiente de determinação ajustado indicou os três segmentos como o de melhor ajuste. O mesmo ocorreu com o grupo DRL.

Resultados semelhantes foram encontrados por Oliveira (2010) que utilizou o mesmo conjunto de dados e empregou o AIC, BIC e R_{aj}^2 como avaliadores da qualidade de ajuste para avaliar o desempenho dos modelos não lineares usuais utilizados para descrever o crescimento de ovinos dos mesmos grupos genéticos. Para o grupo DMN, os valores do AIC ficaram entre 772,5 e 794,45, para o BIC, 781,62 a 806,52 e 82,36 a 82,82 para o R_{aj}^2 . Para o grupo DSI, o AIC ficou entre 1.192,82 e 1.206,96, para o BIC, 1.206,17 a 1.216,97 e 76,72 a 78,35 para o R_a^2 . Já para o grupo DRL os valores ficaram entre 1.034,59

e 1.043,30 para AIC, 1.047,5 e 1.053,12 para o BIC e 79,66 a 80,65 para o R^2_{aj} . Estes resultados demonstram o melhor desempenho das regressões segmentadas na descrição do crescimento de ovinos em relação aos modelos não lineares usuais.

Nas Tabelas 7, 8 e 9, encontram-se as estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os três grupos genéticos.

Tabela 7 Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DMN

Grupo Genético	Parâmetros	L	L-L	L-L-L
DMN	β_{00}	4,86 (3,88;5,84)	2,43 (0,87;3,99)	2,68 (1,09;4,27)
	β_{01}	0,09 (0,09;0,10)	0,14 (0,12;0,16)	0,13 (0,11;0,16)
	β_1	-	-0,06 (-0,09;-0,03)	-0,04 (-0,08;0,005)
	β_2	-	-	-0,03 (-0,07;0,01)
	t_1	-	84	84
	t_2	-	-	168

Tabela 8 Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DSI

DSI	β_{00}	9,07 (7,73;10,40)	2,65 (0,75;4,56)	4,11 (2,39;5,83)
	β_{01}	0,09 (0,08;0,10)	0,21 (0,18;0,24)	0,16 (0,13;0,19)
	β_1	-	-0,16 (-0,19;-0,12)	-0,02 (-0,07;0,02)
	β_2	-	-	-0,16 (-0,20;-0,12)
	t_1	-	84	84
	t_2	-	-	168

Tabela 9 Estimativas dos parâmetros das funções segmentadas linear (nó conhecido) para os dados experimentais do grupo genético DRL

	β_{00}	6,71 (5,72;7,71)	3,52 (1,97;5,06)	4,49 (3,03;5,94)
	β_{01}	0,09 (0,08;0,09)	0,15 (0,12;0,17)	0,11 (0,08;0,14)
DRL	β_1	-	-0,08 (-0,11;-0,05)	0,013 (-0,03;0,05)
	β_2	-	-	-0,10 (-0,14;-0,07)
	t_1	-	84	84
	t_2	-	-	168

t_1 é o nó entre a primeira e segunda fase; t_2 é o nó entre a segunda e terceira fase.

Os nós, 84 e 168 (t_1 e t_2), apresentados acima, foram previamente definidos, portanto, não se caracterizam como parâmetro e consequentemente não possuem os intervalos de confiança atribuídos a eles.

Os parâmetros das funções segmentadas linear e linear-linear, do grupo genético DMN, apresentam intervalos de confiança que não contém o zero, entretanto, para a função segmentada linear-linear-linear, os intervalos para os parâmetros β_1 e β_2 contém o zero, indicando que, para esse grupo genético, o ajuste com três segmentos é não significativo.

O mesmo resultado foi encontrado quando o ajuste foi realizado por meio da regressão segmentada não linear (Tabela 3).

Para o grupo genético DSI, nenhum dos intervalos de confiança para os parâmetros das funções segmentadas linear-linear, linear-platô contém o zero, contudo, para a função segmentada linear-linear-linear o intervalo para o parâmetro β_1 contém o zero indicando que o ajuste com três segmentos não deve ser utilizado para descrever o crescimento dos animais deste grupo.

Utilizando o ajuste não linear, nesse grupo genético, nenhum intervalo de confiança para os parâmetros dos três modelos ajustados contém o zero (Tabela 4) diferente do que ocorreu no caso linear.

As funções segmentadas linear-linear e linear-platô, do grupo genético DRL, apresentam intervalo de confiança que não contém o zero, no entanto, para a função segmentada linear-linear-linear os intervalos para os parâmetros β_1 contém o zero, indicando que, para esse grupo genético, o modelo com três segmentos não se ajusta bem nesse caso, o mesmo ocorreu no ajuste não linear.

Nas Tabelas 10, 11 e 12, encontram-se as somas de quadrado de regressão das funções segmentadas linear (nó conhecido) para os três grupos genéticos, a fim de definir o número de segmentos significativos para cada regressão segmentada, por meio da soma de quadrado extra.

Tabela 10 Análise da variância para o peso de ovinos do grupo genético DMN com a decomposição da soma de quadrados da regressão, considerando 1.000 repetições

Causas de Variação	Soma de Quadrados	G. L.	Quadrados Médios	F	p-valor
Regressão	(6954,75)	3	QMR (X_1, X_2, X_3)		
X_1	6794,88	1	6794,878	711,059	<0,0001
$X_2 X_1$	139,57	1	139,572	14,606	0,0001
$X_3 X_1, X_2$	20,30	1	20,302	2,125	0,1471
Erro	1404,78	147	9,556		
Total	8359,53	150	-		

Através dos resultados apresentados na Tabela 10, pode-se perceber que os testes estatísticos para os parâmetros β_{01} e β_1 são significativos, portanto, os dois segmentos devem ser mantidos no modelo, entretanto, o mesmo não ocorre com o parâmetro β_2 sendo esse não significativo e, portanto, o terceiro

segmento deve ser retirado da função segmentada, dessa forma, o modelo que descreve melhor o crescimento dos animais do grupo genético DMN apresenta dois segmentos.

Resultados semelhantes podem ser visualizados por meio do intervalo de confiança. Para os parâmetros da regressão com três segmentos (Tabela 7); pode-se perceber que, para alguns parâmetros os intervalos contêm o zero, sugerindo que esse modelo foi não significativo para este grupo genético.

Tabela 11 Análise da variância para os pesos de ovinos do grupo genético DRL com a decomposição da soma de quadrados da regressão considerando, 1.000 repetições

Causas de Variação	Soma de Quadrados	G. L.	Quadrados Médios	F	p-valor
Regressão	(9403,749)	3	QMR (X_1, X_2, X_3)		
X1	8630,67	1	8630,671	753,61	<0,0001
X ₂ X ₁	352,67	1	352,666	30,79	< 0,0001
X ₃ X ₁ , X ₂	420,41	1	420,412	36,71	< 0,0001
Erro	2187,41	191	11,452		
Total	11591,162	194	-		

Os resultados da Tabela 11 mostram que os testes estatísticos, para os três parâmetros, foram significativos indicando que os mesmos não puderam ser retirados do modelo. Sendo assim, a relação entre peso e idade dos animais do grupo genético DRL é mais bem descrita pela regressão com três segmentos.

O zero está contido no intervalo de confiança do parâmetro β_1 da regressão com três segmentos do grupo genético DRL, (Tabela 9), indicando que, para esse grupo genético, o ajuste com três segmentos não se ajusta bem, o mesmo não ocorreu com os resultados encontrados pela soma de quadrados extras.

Tabela 12 Análise da variância para os pesos de ovinos do grupo genético DSI com a decomposição da soma de quadrados da regressão, considerando 1.000 repetições

Causas de Variação	Soma de Quadrados	G.L.	Quadrados Médios	F	p-valor
Regressão	(13491,79)	3	QMR (X_1, X_2, X_3)		
X_1	10945,91	1	10945,91	636,4715254	< 0,0001
$X_2 X_1$	1529,33	1	1529,33	88,92590913	< 0,0001
$X_3 X_1, X_2$	1016,55	1	1016,55	59,10930468	< 0,0001
Erro	3508,346	204	17,1978		
Total	17000,136	207	-		

Os testes estatísticos para os três parâmetros foram significativos indicando que nenhum dos mesmos deve ser retirado do ajuste, sendo assim, o modelo que melhor estima o crescimento dos animais do grupo genético DSI é representado por três segmentos, como mostrado na Tabela 12.

Este resultado não está de acordo com o encontrado pelos intervalos de confiança (Tabela 8) que indicaram a regressão com três segmentos não significativos.

Encontram-se, na Tabela 13, alguns avaliadores da qualidade do ajuste para estimativas dos parâmetros das funções segmentadas linear nos parâmetros considerando um, dois e três segmentos, para os três grupos genéticos.

Tabela 13 Avaliadores da qualidade do ajuste para estimativas dos parâmetros das funções segmentadas linear nos parâmetros para os três grupos genéticos

Grupo Genético	Avaliador da Qualidade de Ajuste	Linear	Linear-Linear	Linear-Linear-Linear
DMN	AIC	787,58	775,47	775,30
	BIC	796,63	787,54	790,39
	$R^2_{aj}(\%)$	81,16%	82,72	82,85%
DSI	AIC	1297,44	1238,88	1187,95
	BIC	1307,45	1252,23	1204,64
	$R^2_{aj}(\%)$	64,21	73,12	79,06%
DRL	AIC	1089,81	1067,07	1034,79
	BIC	1099,63	1080,16	1051,159
	$R^2_{aj}(\%)$	74,33	77,27	80,83%

Os resultados dos avaliadores da qualidade do ajuste, apresentados na Tabela 13 acima, estão de acordo com os encontrados pelos testes estatísticos utilizando as somas de quadrados extras, uma vez que, para o grupo genético DMN foi indicado, como o melhor ajuste, a função segmentada que apresenta dois segmentos e para os grupos DSI e DRL, os modelos que melhor se adequaram possuem três segmentos.

Essa concordância, entre os resultados apresentados pelos avaliadores da qualidade de ajuste e pelos testes estatísticos, nos faz acreditar que a soma de quadrados extra é uma boa ferramenta estatística para indicar o número de segmentos significativos na regressão segmentada. As Figuras 1, 2 e 3 ilustram os gráficos de dispersão do peso em função da idade de ovinos respectivamente para os grupos genéticos DMN, DSI e DRL, além das funções segmentadas

linear, linear-linear e linear-linear-linear estimadas para os mesmos grupos genéticos.

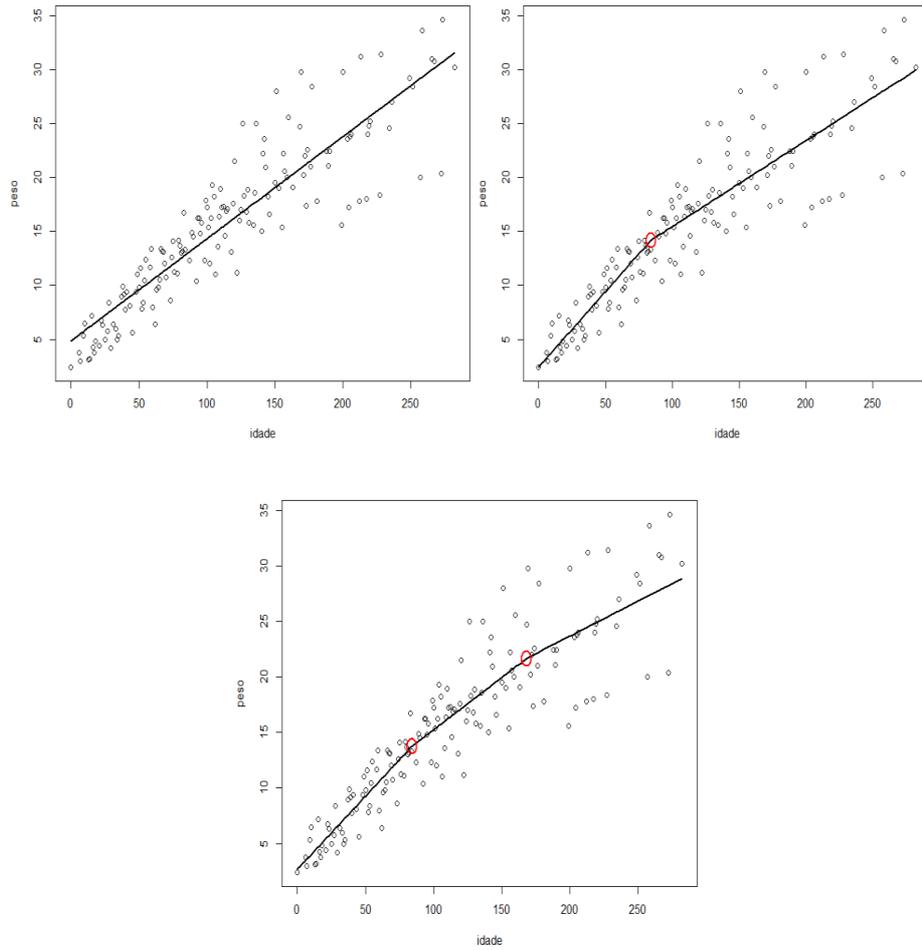


Figura 1 Modelo linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DMN

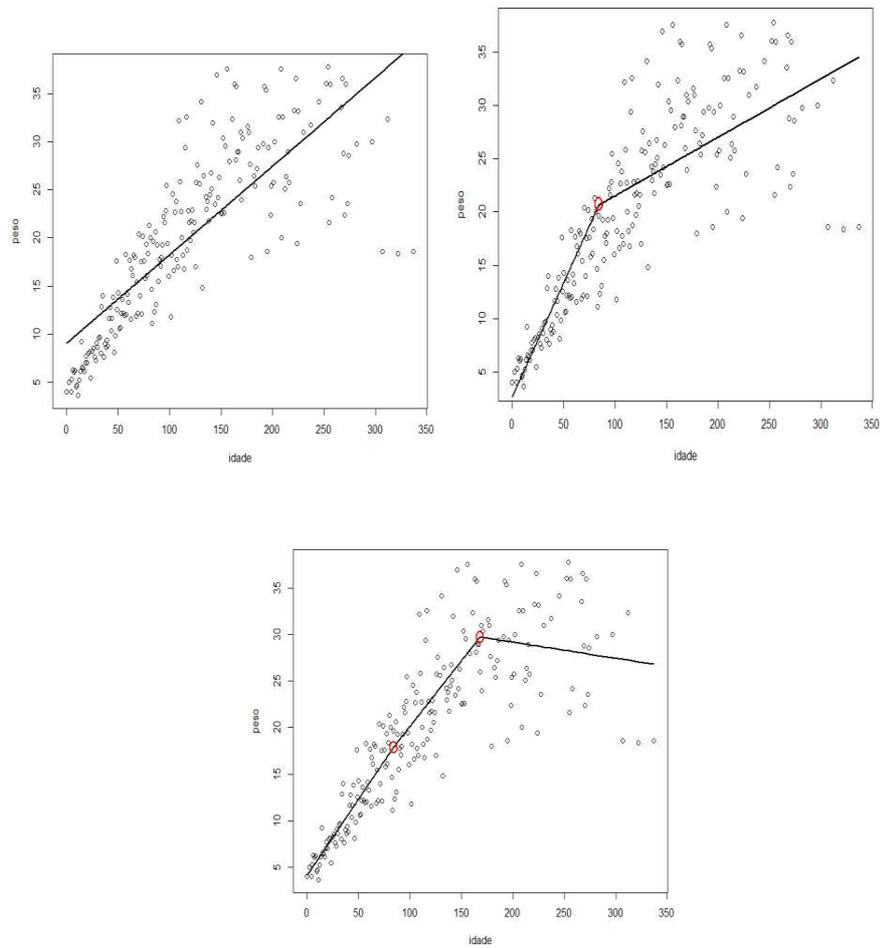


Figura 2 Modelo linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DSI

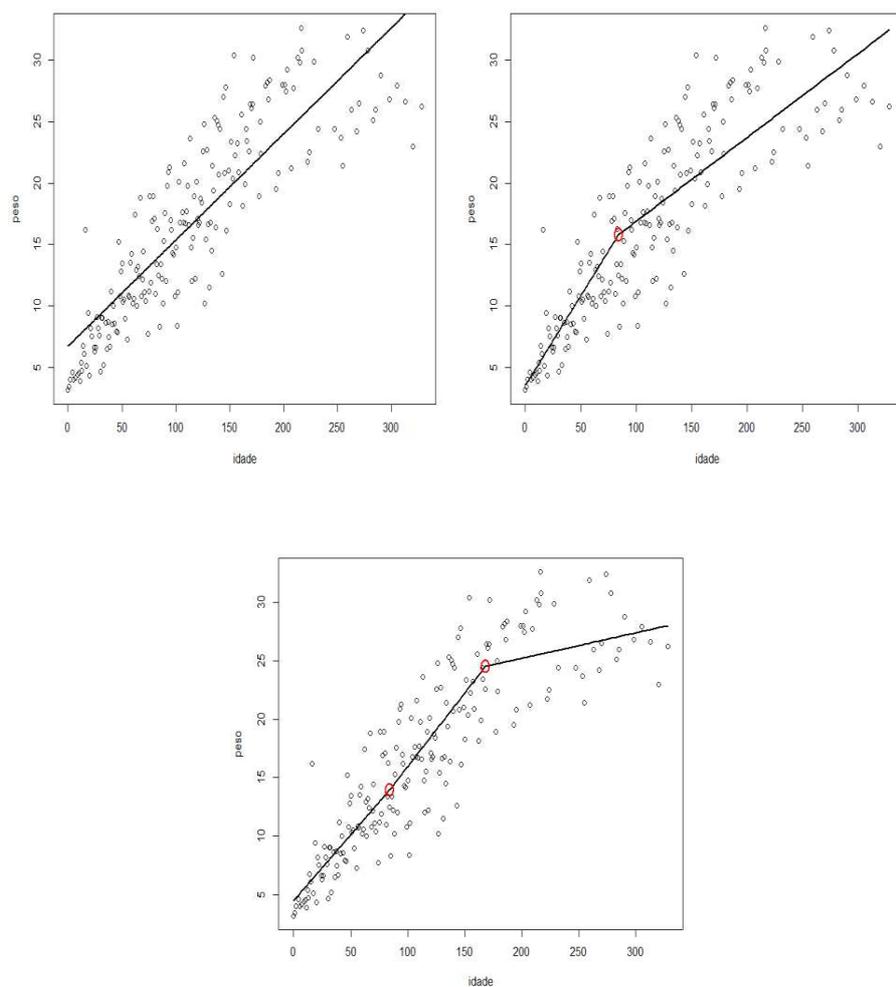


Figura 3 Modelos linear, linear-linear e linear-linear-linear considerando o peso em função da idade dos animais do grupo genético DRL

4.2 Regressão segmentada ajustada a dados simulados

Na Tabela 14 abaixo, estão apresentados os resultados dos avaliadores da qualidade do ajuste da regressão segmentada linear nos parâmetros ajustados a partir dos dados simulados e considerando 1.000 repetições.

Tabela 14 Valores médios do AIC, BIC e R_{aj}^2 para os 1.000 modelos simulados

Crescimento	Modelo	AIC	BIC	R_{aj}^2
	Linear	1047,0	1056,8	85,27
Linear	Linear- Linear	1048,2	1061,3	85,24
	Linear- Linear- Linear	1048,5	1064,9	85,29
	Linear	1114	1124	93,75
Linear-Linear	Linear- Linear	1047,6	1060,7	95,55
	Linear- Linear- Linear	1047,8	1064,2	95,57
	Linear	1082,0	1092	92,56
Linear-Linear- Linear	Linear- Linear	1080,0	1094	92,64
	Linear- Linear- Linear	1048,0	1064,4	93,8

Os resultados dispostos na Tabela 14 demonstram um bom desempenho dos avaliadores da qualidade de ajuste utilizados, uma vez que, indicaram como melhor comportamento os modelos que foram simulados de forma adequada aos dados.

O coeficiente de determinação ajustado foi o avaliador da qualidade de ajuste com pior desempenho, no caso em que foi simulado o crescimento com um segmento o mesmo indicou um modelo com três segmentos e, no caso em que foi simulado um modelo com dois segmentos, o indicado pelo coeficiente de determinação ajustado foi uma função segmentada com três segmentos.

No entanto, os valores encontrados para os coeficientes de determinação ajustado nos modelos simulados e nos modelos indicados como melhores foram muito próximos, o que indica que o erro cometido por esse avaliador da qualidade de ajuste não foi grande.

Nas Tabelas 15, 16, 17 e 18 estão apresentados os valores-p da decomposição da soma de quadrados extra a fim de verificar se o teste F indica de forma correta o número de segmentos significativos, para tanto foram considerados os tamanhos de amostra 50, 100, 150 e 195.

Tabela 15 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 50 observações

Causas de Variação	Simulado	valor-p
X_1		1
$X_2 X_1$	Linear	1
$X_3 X_1, X_2$		1
X_1		< 0,0001
$X_2 X_1$	Linear-Linear	0,0240
$X_3 X_1, X_2$		0,7919
X_1		< 0,0001
$X_2 X_1$	Linear-Linear-	0,6257
$X_3 X_1, X_2$	Linear	0,1376

Com amostra de tamanho 50, percebe-se que o teste F não indicou, de forma correta, nenhum dos três casos testados.

Tabela 16 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 100 observações

Causas de Variação	Simulado	valor-p
X_1		0,5730
$X_2 X_1$	Linear	0,6771
$X_3 X_1, X_2$		0,6332
X_1		< 0,001
$X_2 X_1$	Linear-Linear	< 0,001
$X_3 X_1, X_2$		0,8091
X_1		< 0,001
$X_2 X_1$	Linear-Linear-	0,3416
$X_3 X_1, X_2$	Linear	0,0063

No caso em que a simulação foi realizada com amostra de tamanho 100, o teste F indicou, de forma correta, apenas o número de segmentos da regressão segmentada linear-linear.

Tabela 17 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 150 observações

Causas de Variação	Simulado	valor-p
X_1		< 0,0001
$X_2 X_1$	Linear	0,4157
$X_3 X_1, X_2$		0,3199
X_1		< 0,0001
$X_2 X_1$	Linear-Linear	< 0,0001
$X_3 X_1, X_2$		0,4513
X_1		< 0,0001
$X_2 X_1$	Linear-Linear-	0,1789
$X_3 X_1, X_2$	Linear	< 0,0001

No caso em que a simulação foi realizada com amostra de tamanho 150, o teste F indicou, de forma correta, o número de segmentos da regressão segmentada linear e linear-linear, errando apenas para o caso linear-linear-linear.

Tabela 18 ANOVA com a decomposição da soma de quadrados da regressão linear considerando o ajuste a partir dos dados simulados com 195 observações

Causas de Variação	Simulado	valor-p
X_1		< 0,0001
$X_2 X_1$	Linear	0,4842
$X_3 X_1, X_2$		0,3551
X_1		< 0,0001
$X_2 X_1$	Linear-Linear	< 0,0001
$X_3 X_1, X_2$		0,3436
X_1		< 0,0001
	Linear-Linear-	< 0,005
$X_2 X_1$	Linear	
$X_3 X_1, X_2$		< 0,0001

No primeiro cenário da simulação, considerando o tamanho de amostra 195, foi gerado o crescimento com apenas um segmento e de acordo com os resultados dos testes estatísticos (Tabela 18), baseados nas somas de quadrados extras, pode-se concluir que o primeiro segmento foi significativo (p -valor < 1%) e os dois últimos segmentos foram não significativos (p -valor > 1%) podendo, dessa forma, ser retirados do modelo, nesse caso, temos que a soma de quadrado indicou de forma correta o número de segmentos significativos para o modelo.

Foi simulado, no segundo cenário, o crescimento considerando dois segmentos e de acordo com os resultados, baseados nas somas de quadrados extras, (Tabela 18) pode-se perceber que o primeiro e segundo segmentos foram significativos (p -valor menor que 1%) e apenas o último segmento foi não significativo (p -valor maior que 1%) podendo, dessa forma, ser retirado do modelo, o que indica novamente o bom desempenho do teste estatístico, baseado

nas somas de quadrados extras, como um indicador do número de segmentos significativos nas regressões segmentadas lineares.

No terceiro cenário, as somas de quadrados extras também indicaram, de forma correta, o número de segmentos significativos, uma vez que, foi simulado o crescimento com três segmentos e o teste estatístico foi significativo (p-valor < 1%) para os três parâmetros, concluindo que os três segmentos devem ser mantidos no ajuste.

Tabela 19 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 50 observações

Simulado	Parâmetros	Valores paramétricos	Esperança	Viés
Linear	β_{00}	4,4800	4,484	0,0037
	β_{01}	0,1100	0,1099	-0,0001
Linear-Linear	β_{00}	4,4800	4,4810	0,001
	β_{01}	0,1100	0,1098	-0,0002
	β_1	0,1300	0,1301	0,0001
Linear- Linear-Linear	β_{00}	4,4800	4,470	-0,010
	β_{01}	0,1100	0,1104	0,0004
	β_1	0,1300	0,1293	-0,0007
	β_2	-0,1000	-0,0992	0,0008

Tabela 20 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 100 observações

Simulado	Parâmetros	Valores paramétricos	Esperança	Viés
Linear	β_{00}	4,4800	4,482	0,002
	β_{01}	0,1100	0,1099	-0,0001
Linear-Linear	β_{00}	4,4800	4,4830	0,0030
	β_{01}	0,1100	0,1099	-0,0001
	β_1	0,1300	0,1304	0,0004
Linear-Linear-Linear	β_{00}	4,4800	4,4820	0,0020
	β_{01}	0,1100	0,1098	-0,0002
	β_1	0,1300	0,1304	0,0004
	β_2	-0,1000	-0,1003	-0,0003

Tabela 21 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1000 repetições e 150 observações

Simulado	Parâmetros	Valores paramétricos	Esperança	Viés
Linear	β_{00}	4,4800	4,465	-0,015
	β_{01}	0,1100	0,1099	0,0001
Linear-Linear	β_{00}	4,4800	4,4770	-0,0030
	β_{01}	0,1100	0,1102	0,0002
	β_1	0,1300	0,1296	0,0004
Linear- Linear-Linear	β_{00}	4,4800	4,4720	0,0080
	β_{01}	0,1100	0,1102	0,0002
	β_1	0,1300	0,1299	-0,0001
	β_2	-0,1000	-0,0999	0,0001

Tabela 22 Valores dos parâmetros simulados, esperanças matemática e vieses dos estimadores considerando 1.000 repetições e 195 observações

Simulado	Parâmetros	Valores paramétricos	Esperança	Viés
Linear	β_{00}	4,4800	4,4820	0,0020
	β_{01}	0,1100	0,1099	-0,0001
Linear-Linear	β_{00}	4,4800	4,4700	-0,0100
	β_{01}	0,1100	0,1101	0,0001
	β_1	0,1300	0,1296	-0,0004
Linear- Linear-Linear	β_{00}	4,4800	4,4690	-0,0110
	β_{01}	0,1100	0,1103	0,0003
	β_1	0,1300	0,1295	-0,0005
	β_2	- 0,1000	-0,0998	0,0002

Por meio dos resultados apresentados nas Tabelas 19, 20, 21 e 22, pode-se perceber que a criação de novas variáveis transformando uma regressão segmentada em uma regressão múltipla, e utilizando as estimativas dos parâmetros da regressão múltipla com as devidas reparametrização para determinar o ajuste das funções segmentadas linear (nó conhecido), apresenta-se como uma metodologia bastante adequada. Os vieses encontrados em todas as estimativas foram muito pequenos, mesmo com os dados gerados com erros dependentes e variâncias heterocedásticas.

Os aspectos observados nos estimadores \hat{a} , \hat{b} , \hat{c} e \hat{d} , das Figuras 4, 5 e 6, retratam uma situação de inconsistência dos mesmos. Foram considerados, respectivamente, os ajustes linear, linear-linear e linear-linear-linear utilizando dados de crescimento gerados com erros correlacionados e heterocedásticos, o que retrata melhor a realidade do crescimento animal.

Segundo Mazucheli, Souza e Philippsen (2011), em muitas situações práticas, em que a variável resposta é o peso ao longo do tempo, a suposição de homocedasticidade pode não ser razoável.

O ajuste realizado por meio dos dados simulados com erros correlacionados e heterocedásticos gerou vieses pequenos, mas, no entanto, apresentou inconsistência. Este resultado pode ser devido ao fato de que a violação de cada pressuposto está associada a um determinado problema (FIGUEIREDO FILHO et al., 2011).

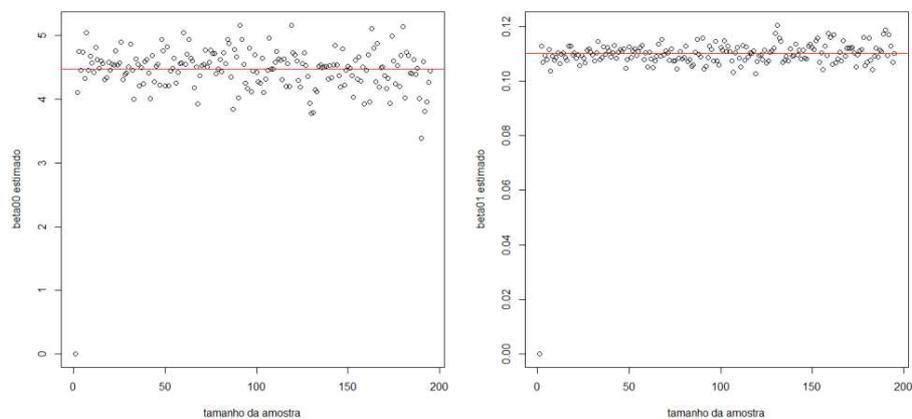


Figura 4 Estimativas dos parâmetros β_{00} e β_{01} em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear com erros correlacionados e variâncias heterocedásticas

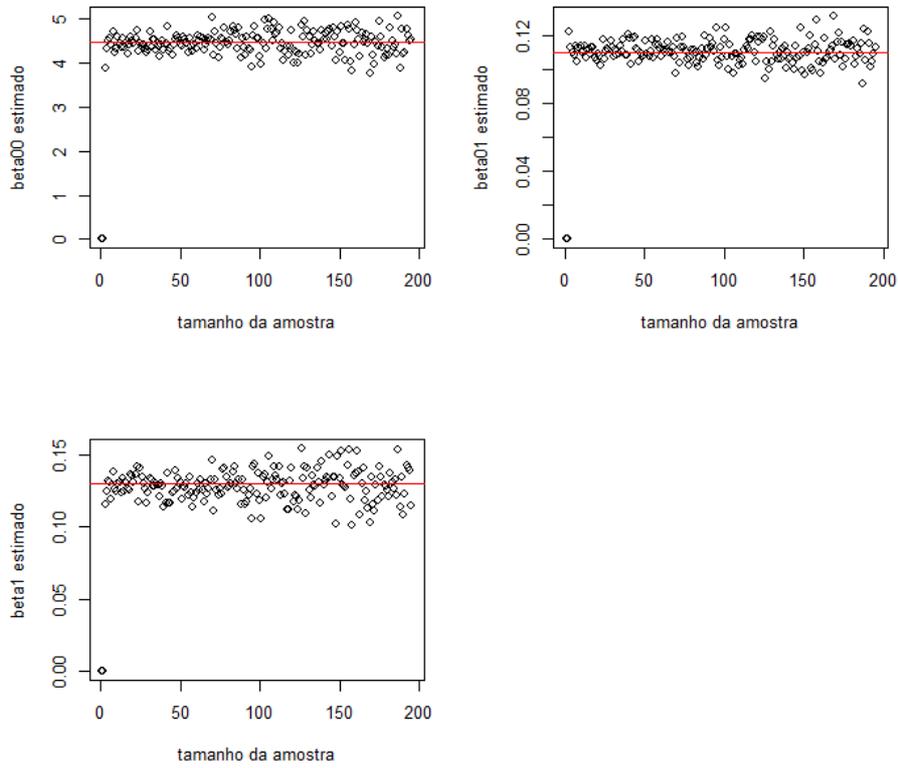


Figura 5 Estimativas dos parâmetros β_{00} , β_{01} e β_1 , em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear-linear com erros correlacionados e variâncias heterocedásticas

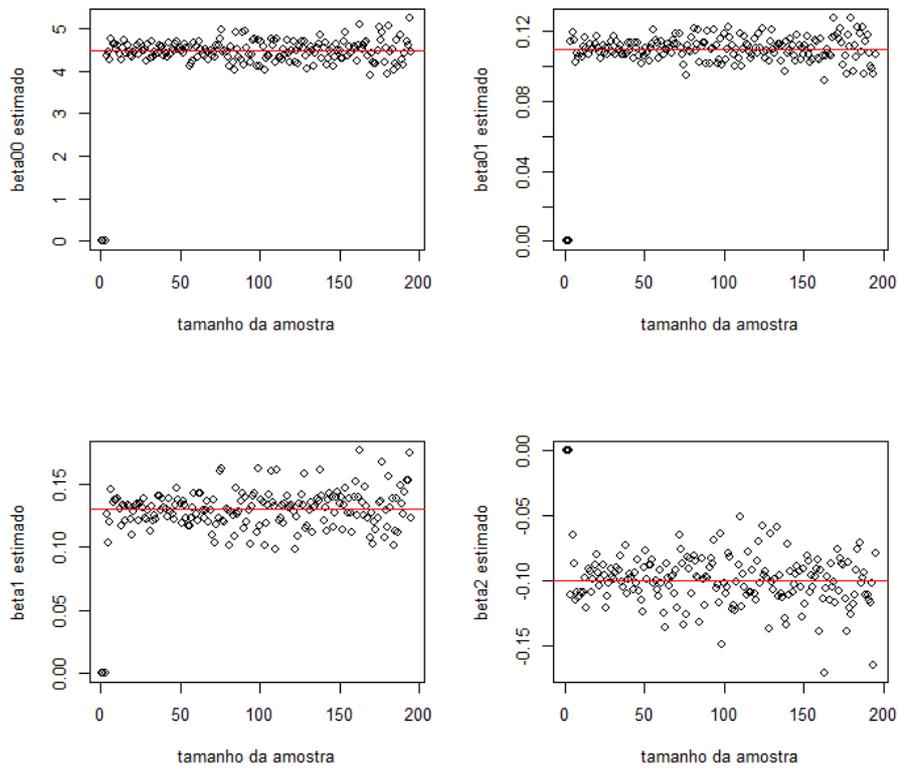


Figura 6 Estimativas dos parâmetros β_{00} , β_{01} , β_1 e β_2 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo segmentado linear-linear-linear com erros correlacionados e variâncias heterocedásticas

Os aspectos observados nos estimadores \hat{a} , \hat{b} , \hat{c} e \hat{d} , nas Figuras 7, 8 e 9 retratam uma situação de consistência dos mesmos. Neste caso, foram considerados, respectivamente, os ajustes linear, linear-linear e linear-linear-linear utilizando dados de crescimento gerados com erros normais e homocedásticos.

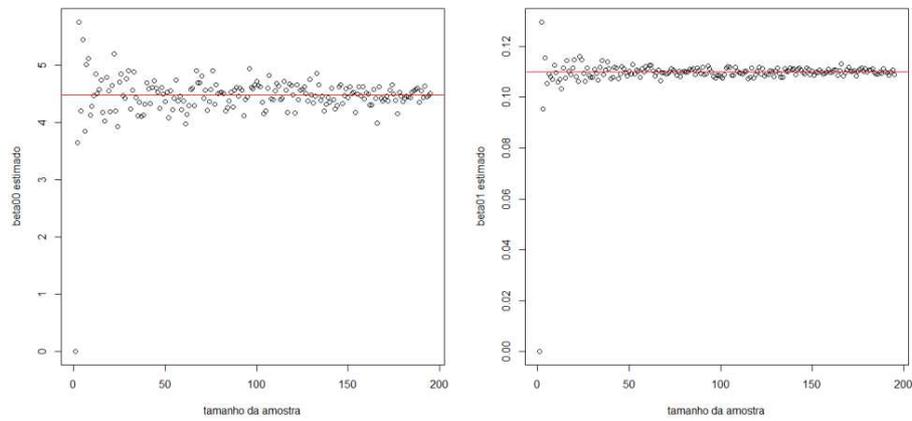


Figura 7 Estimativas dos parâmetros β_{00} e β_{01} em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear com erros normais e homocedásticos

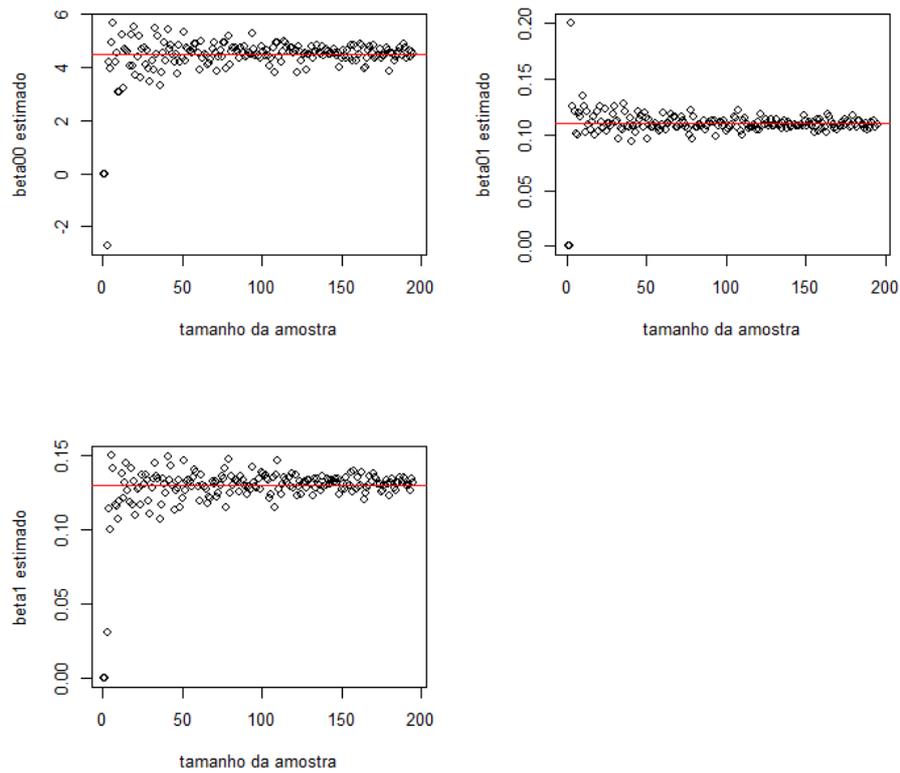


Figura 8 Estimativas dos parâmetros β_{00} , β_{01} e β_1 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear-linear com erros normais e homocedásticos

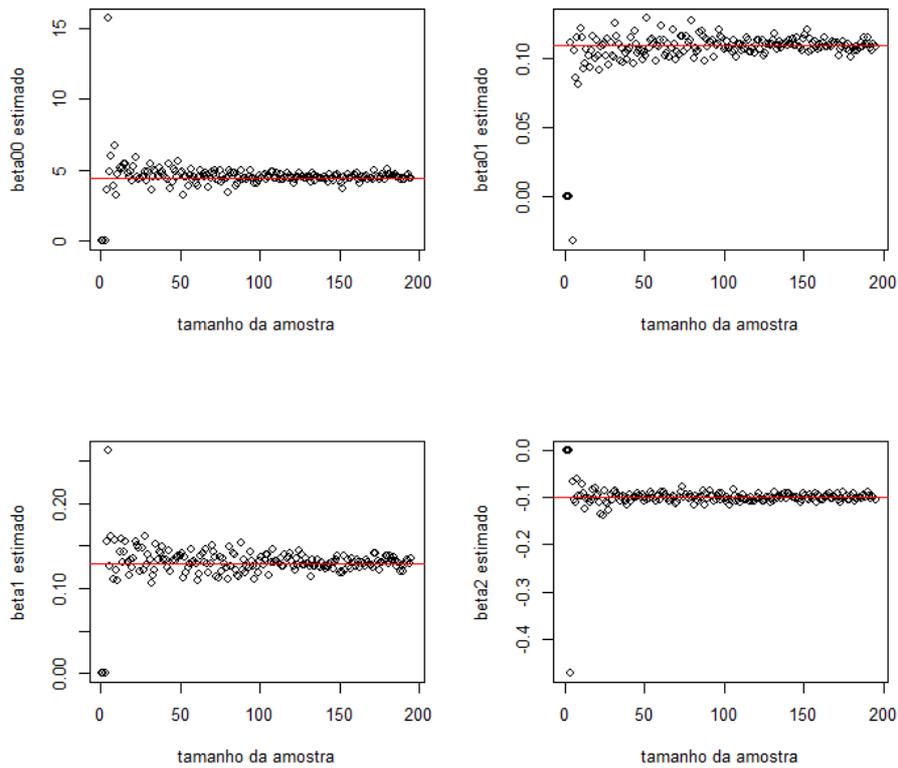


Figura 9 Estimativas dos parâmetros β_{00} , β_{01} , β_1 e β_2 em função de diferentes tamanhos de amostra, levando em consideração o ajuste do modelo linear-linear-linear com erros normais e homocedásticos

Com respeito à consistência, é perceptível a melhora no ajuste do modelo, uma vez que as suposições de normalidade e homocedasticidade da variância foram atendidas.

Resultados semelhantes foram encontrados por Mazucheli, Souza e Philippsen (2011), estes autores afirmaram que a qualidade das estimativas, depende da não violação de várias suposições e em análise de regressão, seja

linear ou não linear, uma violação muito comum é a falta de homocedasticidade de variâncias presente principalmente, na análise de dados de peso-idade de animais.

5 CONCLUSÕES

As funções segmentadas mostraram-se adequadas para descrever curvas de crescimento dos três grupos genéticos de ovinos cruzados.

A função segmentada linear-platô apresentou melhor ajuste nas regressões segmentadas não lineares, a linear-linear-linear proporcionou um bom desempenho podendo também ser utilizada para descrever as curvas de crescimento dos ovinos em estudo.

O teste estatístico, F, baseado nas somas de quadrado extra, mostrou-se como uma boa ferramenta para determinar o número de segmentos significativos nas regressões segmentadas lineares mas desde que o tamanho da amostra seja bastante grande, o mesmo não ocorre com amostras pequenas.

A redução da regressão segmentada linear nos parâmetros em uma regressão múltipla, por meio da criação de novas variáveis, mostrou-se uma metodologia estatística apropriada para determinar os estimadores das regressões segmentadas linear, uma vez que apresentou pequeno viés e são consistentes quando os erros apresentam variâncias homocedásticas e são independentes.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transaction on Automatic Control**, New York, v. 19, p. 716-723, 1974.
- BERTALANFFY, L. von. Quantitative laws for metabolism and growth. **Quarterly Review of Biology**, Chicago, v. 32, p. 217-231, 1957.
- BIASE, N. G.; FERREIRA, F. D. F. Testes de igualdade e de comparações múltiplas para várias proporções binomiais independentes. **Revista Brasileira de Biometria**, São Paulo, v. 29, n. 4, p. 549-570, 2011.
- BROCHIER, M. A.; CARVALHO, S. Consumo, ganho de peso e análise econômica da terminação de cordeiros em confinamento com dietas contendo diferentes proporções de resíduo úmido de cervejaria. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 60, n. 5, p. 1205-1212, out. 2008.
- BRODY, S. **Bioenergetics and growth**. New York: Rheinhold, 1945. 1023 p.
- BUVAT, I.; CASTIGLIONI, I. Monte Carlo simulation in SPET and PET. **Quarterly Journal of Nuclear Medicine**, Torino, v. 46, n. 1, p. 48-61, Mar. 2002.
- CAETANO, A.; MACHADO, C. J. Consistência e identificabilidade no Modelo *Grade Of Membership*: uma nota metodológica. **Revista Brasileira de Estudos de População**, Rio de Janeiro, v. 26, n. 1, p. 145-149, 2009.
- CARNEIRO, A. P. S. et al. Identidade de modelos não lineares para comparar curvas de crescimento de bovinos da raça Tabapuã. **Pesquisa Agropecuária Brasileira**, Brasília, v. 49, n. 1, p. 57-62, jan. 2014.
- CARNEIRO, P. L. S. et al. Desenvolvimento ponderal e diversidade fenotípica entre cruzamentos de ovinos Dorper com raças locais. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 7, p. 991-998, jul. 2007.
- COSTA JÚNIOR, G. S. et al. Caracterização morfométrica de ovinos da raça Santa Inês criados nas microrregiões de Teresina e Campo Maior, Piauí. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 35, n. 6, p. 2260-2267, 2006.

CRIBARI-NETO, F.; SOARES, A. C. N. Inferência em modelos heterocedásticos. **Revista Brasileira de Economia**, Rio de Janeiro, v. 57, n. 2, p. 319-335, 2003.

DOMINGOS FILHO, M. Estimadores de mínimos quadrados no modelo de calibração estrutural sem suposição de normalidade. **Revista Brasileira de Biometria**, São Paulo, v. 30, n. 2, p. 239-257, 2012.

DUARTE, F. A. M. **Estudo da curva de crescimento de animais da raça Nelore (*Bos taurus indicus*) através de cinco modelos estocásticos**. 1975. 284 f. Tese (Livre-Docência em Genética) - Universidade de São Paulo, Ribeirão Preto, 1975.

FERNANDES, F. M. N. Situação da ovinocultura de São Paulo. In: SIMPÓSIO PAULISTA DE OVINOCULTURA, 1., 1998, Botucatu. **Anais...** Botucatu: Fundação Cargill, 1998. p. 3-10.

FIGUEIREDO FILHO, D. et al. O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO). **Revista Política Hoje**, Recife, v. 20, n. 1, p. 44-99, 2011.

FITZHUGH JUNIOR, H. A. Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, Champaign, v. 42, n. 4, p. 1036-1051, Apr. 1976.

GARCIA, S.; LUSTOSA, P. R. B.; BARROS, N. R. Aplicabilidade do método de simulação de Monte Carlo na previsão dos custos de produção de companhias industriais: o caso da companhia Vale do Rio Doce. **RCO - Revista de Contabilidade e Organizações**, São Paulo, v. 4, n. 10, p. 172-173, set./dez. 2010.

GOTTSCHALL, C. S. Impacto nutricional na produção de carne-curva de crescimento. In: LOBATO, J. F. P.; BARCELLOS, J. O. J.; KESSLER, A. M. (Ed.). **Produção de bovinos de corte**. Porto Alegre: EDIPUCRS, 1999. p. 169-192.

GUEDES, M. H. P. et al. Análise Bayesiana da curva de crescimento de cordeiros da raça Santa Inês. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 57, n. 3, p. 415-417, 2005.

GUEDES, M. H. P. et al. Estudo das curvas de crescimento de cordeiros das raças Santa Inês e Bergamácia considerando heterogeneidade de variâncias. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 2, p. 381-388, mar./abr. 2004.

HERNANDEZ, E. L. **Parameter estimation in linear-linear segmented regression**. 2010. 93 f. Dissertation (Master in Statistic) - Faculty of Brigham Young University, Brigham, 2010.

HOLANDA JUNIOR, E. V.; AS, J. L.; ARAUJO, G. G. L. Articulação dos segmentos da cadeia produtiva de caprinos e ovinos: os fluxos alternativos de comercialização. In: SIMPÓSIO INTERNACIONAL SOBRE CAPRINOS E OVINOS DE CORTE, 2., 2003, João Pessoa. **Anais...** João Pessoa: UFPB, 2003. p. 83-93.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção da pecuária municipal rebanho bovino brasileiro**: efetivo de rebanhos 2011. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 10 dez. 2013.

KAPS, M.; LAMBERSON, W. **Biostatistics for animal science**. Oxford: CABI, 2004. 444 p.

KEELE, L. J. **Semiparametric regression Estatística for the science**. New York: J. Wiley, 2008. 230 p.

LARSEN, R. J.; MARX, M. L. **An introduction to mathematical statistics and its applications**. 2nd ed. New Jersey: Prentice Hall, 1986. 433 p.

LOBO, R. N. B. et al. Parâmetros genéticos de características estimadas da curva de crescimento de ovinos da raça Santa Inês. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 35, n. 3, p. 1012-1019, 2006. Suplemento.

MACIEL, D. B. M.; VIEIRA, G. I. A.; MARINHO, P. R. D. Avaliação dos desempenhos de estimadores para os parâmetros da distribuição Birnbaum-Saunders. **TEMA**, São Carlos, v. 14, n. 3, p. 373-382, 2013.

MALHADO, C. H. M. et al. Growth curves in Dorper sheep crossed with the local Brazilian breeds, Morada Nova, Rabo Largo, and Santa Inês. **Small Ruminant Research**, v. 84, p. 16-21, June 2009.

MALHADO, C. H. M. et al. Melhoramento e estrutura populacional em bubalinos da raça Mediterrâneo no Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 2, p. 215-220, fev. 2008a.

MALHADO, C. H. M. et al. Progresso genético e estrutura populacional do rebanho Nelore no estado da Bahia. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 9, p. 1163-1169, set. 2008b.

MAZUCHELI, J.; SOUZA, R. M.; PHILIPPSEN, A. S. Modelo de crescimento de Gompertz na presença de erros normais heterocedásticos: um estudo de caso. **Revista Brasileira de Biometria**, São Paulo, v. 29, n. 1, p. 91-101, 2011.

MAZZINI, A. R. A. et al. Análise da curva de crescimento de machos Hereford. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 5, p. 1105-1112, set./out. 2003.

MAZZINI, A. R. A. et al. Curvas de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.

MEDEIROS, J. X. **Situação atual das cadeias produtivas**. Brasília: MCT/CNPq, 2001. 55 p. Relatório final.

MENDES, P. N. et al. Modelo logístico difásico no estudo do crescimento de fêmeas da raça Hereford. **Ciência Rural**, Santa Maria, v. 38, n. 7, p. 1984-1990, 2008.

NOCCHI, E. D. G. **Os efeitos da crise da lã no mercado internacional e os impactos sócio-econômicos no município de Santana do Livramento, RS, Brasil**. 2001. 71 p. Dissertação (Mestrado em Integração e Cooperação Internacional) - Universidade Nacional de Rosário, Rosário, 2001.

OLIVEIRA, A. C. H. et al. Monte Carlo simulation of a medical linear accelerator for generation of phase spaces. In: INTERNATIONAL NUCLEAR ATLANTIC CONFERENCE, 1., 2013, Recife. **Anais...** Recife: INAC, 2013. 1 CD-ROM.

OLIVEIRA, D. C. **Funções splines para estudo de curvas de crescimento em ovinos cruzados**. 2011. 69 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2011.

OLIVEIRA, T. S. Ovinocultura de corte no Brasil: abordagem mercadológica. **Revista da Terra**, Pindorama, v. 1, n. 126, p. 76-78, jun. 2010.

PACHEPSKY, L. B.; HASKETT, J. D.; ACOCK, B. An adequate model of photosynthesis. I., parameterization, validation and comparison of models. **Agricultural Systems**, Essex, v. 50, n. 2, p. 209-225, Aug. 1996.

PEREIRA FILHO, N. de A.; FERREIRA, D. F. Teste Monte Carlo de normalidade univariado baseado em distâncias. **Revista Brasileira de Biometria**, São Paulo, v. 30, n. 3, p. 401-416, 2012.

PIOLA JÚNIOR, W. et al. Ganho de peso e características da carcaça de cordeiros recebendo diferentes níveis de energia na ração. **Ciências Agrárias**, Londrina, v. 30, n. 4, p. 935-944, out./dez. 2009.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 out. 2014.

RAMOS, A. de A. et al. Caracterização fenotípica e genética da produção de leite e do intervalo entre partos em bubalinos da raça Murrah. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, n. 8, p. 1261-1267, ago. 2006.

REGAZZI, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear. **Revista Ceres**, Viçosa, MG, v. 50, n. 287, p. 9-26, 2003.

REGAZZI, A. J.; SILVA, C. H. O. Testes para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear em dados de experimento com delineamento em blocos casualizados. **Revista Ceres**, Viçosa, MG, v. 57, n. 3, p. 315-320, 2010.

REIS, R. L. et al. Estimativas de variância genética aditiva em populações selecionadas e não-selecionadas via simulação Monte Carlo utilizando o software R. **Ciência e Agrotecnologia**, Lavras, v. 33, n. 1, p. 285-291, jan./fev. 2009.

RICHARDS, F. J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Oxford, v. 10, p. 290-300, 1959.

SARMENTO, J. L. R. et al. Estudo da curva de crescimento de ovinos Santa Inês. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 35, n. 2, p. 435-442, 2006.

SAVIAN, T. V.; MUNIZ, J. A. A study of *in situ* degradability: heterogeneity of variances and correlated errors. **Scientia Agricola**, Piracicaba, v. 64, n. 5, p. 548-554, 2007.

SILVA, F. F. e; AQUINO, L. H. de; OLIVEIRA, A. I. G. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado Nelore. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 5, p. 1195-1205, set./out. 2001.

SILVA, N. A. M. et al. Seleção e classificação multivariada de modelos de crescimento não lineares para bovinos Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 63, n. 2, p. 364-371, 2011.

SILVEIRA, F. G. et al. Análise de agrupamento na seleção de modelos de regressão não-lineares para curvas de crescimento de ovinos cruzados. **Ciência Rural**, Santa Maria, v. 41, n. 4, p. 692-698, abr. 2011.

SILVEIRA, F. G. et al. Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte. In: REUNIÃO ANNUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 54.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 13., 2009, São Carlos. **Anais...** São Carlos: UFSCar, 2009. 1 CD-ROM.

SILVEIRA, F. G. et al. Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 13, n. 1, p. 62-73, jan./mar. 2012.

SIQUEIRA, E. R.; SIMÕES, C. D.; FERNANDES, S. Efeito do sexo e do peso ao abate sobre a produção de carne de cordeiro. Morfometria da carcaça, pesos dos cortes, composição tecidual e componentes não constituintes da carcaça. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 30, n. 4, p. 1299-1307, 2001.

SORIO, A. **Carne ovina e caprina**: produção e consumo no Brasil e nas Américas. Disponível em: <<http://www.neapl.sedotec.se.gov.br/module/news/article.php>>. Acesso em: 25 jul. 2010.

SOUSA, W. H.; LEITE, P. R. M. **Ovinos de corte**: a raça Dorper. João Pessoa: EMEPA-PB, 2000. 76 p.

SOUSA, W. H.; MORAIS, O. R. Programa de melhoramento genético para ovinos deslanados do Brasil: ovinos da raça Santa Inês. In: SIMPÓSIO INTERNACIONAL SOBRE CAPRINOS E OVINOS DE CORTE, 1., 2000, João Pessoa. **Anais...** João Pessoa: Empresa Estadual de Pesquisa Agropecuária da Paraíba, 2000. p. 223-229.

SOUZA, J. C.; BIANCHINI SOBRINHO, E. Estimativa do peso de bovinos de corte, aos 24 meses, da raça Nelore, usando curvas de crescimento. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, MG, v. 23, n. 1, p. 85-91, jan./fev. 1994.

SOUZA, L. A. et al. Curvas de crescimento em ovinos da raça morada nova criados no Estado da Bahia. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 40, n. 8, p. 1700-1705, 2011.

SOUZA JUNIOR, A. A. O. et al. Estudo alométrico dos cortes da carcaça de cordeiros cruzados Dorper com as raças Rabo Largo e Santa Inês. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 10, n. 2, p. 423-433, jul. 2009.

THOLON, P. et al. Utilização de funções lineares e não lineares para ajuste do crescimento de bovinos Santa Gertrudes, criados a pasto. **ARS Veterinária**, Jaboticabal, v. 28, n. 4, p. 234-239, 2012.

VIANA, J. G. A. **Governança da cadeia produtiva da ovinocultura no Rio Grande do Sul**: estudo de caso à luz dos custos de transação e produção. 2008. 137 p. Dissertação (Mestrado em Extensão Rural) - Universidade Federal de Santa Maria, Santa Maria, 2008.

VIDAL, M. F. et al. Análise econômica da produção de ovinos em lotação rotativa em pastagem de capim tanzânia (*Panicum maximum* (Jacq)). **Revista de Economia e Sociologia Rural**, Brasília, v. 44, n. 4, p. 801-818, dez. 2006.

VIOLA, D. N. et al. Uma avaliação do estimador de pseudo verossimilhança para modelos autoligísticos espaciais. **Revista de Matemática e Estatística**, São Paulo, v. 20, n. 1, p. 1-10, 2000.

YORIYAZ, H. Método de Monte Carlo: princípios e aplicações em física médica. **Revista Brasileira de Física Médica**, Florianópolis, v. 3, n. 1, p. 141-149, 2009.

ZAPATA, J. F. F. et al. Características de carcaça de pequenos ruminantes no Nordeste do Brasil. **Ciência Animal**, Goiânia, v. 11, n. 2, p. 79-86, 2001.

ZEVIANI, W. M. **Parametrizações interpretáveis em modelos não lineares**. 2013. 146 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2013.

APÊNDICE

APÊNDICE A – Dados experimentais do grupo genético DRL.

Seguem abaixo os dados assim como os scripts utilizados para as análises e simulação apresentadas na tese.

Os dados são do grupo genético DRL e foram utilizados na simulação, com as idades foram gerados os y (peso) em cada cenário.

Dados Experimentais:

idade	peso
0	3.185000000
1	3.400000000
2	4.000000000
4	4.600000000
5	4.000000000
7	4.120000000
9	4.400000000
10	4.500000000
11	3.900000000
12	5.400000000
13	4.733333333
14	6.750000000
15	6.080000000
16	16.200000000
17	5.125000000
19	9.400000000
20	4.300000000
21	8.200000000
22	7.500000000
24	6.625000000
25	6.266666667
26	6.600000000
27	9.100000000
28	8.200000000
29	7.600000000
30	4.650000000
31	9.000000000
32	9.000000000
33	5.200000000

35	8.600000000
36	6.466666667
37	8.666666667
38	7.450000000
39	6.680000000
40	11.200000000
41	8.500000000
42	10.000000000
43	8.577777778
45	7.920000000
46	7.850000000
47	15.200000000
48	10.800000000
49	12.800000000
50	13.466666670
51	10.325000000
52	10.550000000
53	8.960000000
55	7.233333333
56	10.866666670
57	10.733333330
58	13.500000000
59	14.240000000
60	10.200000000
61	10.571428570
62	17.400000000
63	12.920000000
64	10.000000000
65	13.200000000
66	12.428571430
67	18.800000000
68	10.800000000
69	12.150000000
70	14.450000000
71	11.100000000
72	10.400000000
74	7.733333333
75	11.200000000
76	18.933333330
77	11.880000000
78	16.933333330
79	18.933333330

80	17.133333330
81	11.000000000
82	13.400000000
83	16.250000000
84	12.500000000
85	8.300000000
86	13.400000000
87	12.200000000
88	10.200000000
89	15.250000000
90	17.550000000
91	12.040000000
92	19.800000000
93	20.900000000
94	21.300000000
95	17.000000000
96	16.200000000
97	14.333333330
98	14.200000000
99	10.800000000
100	14.750000000
101	8.400000000
102	11.100000000
103	20.100000000
104	16.800000000
106	17.640000000
107	16.750000000
108	21.600000000
109	16.733333330
110	17.700000000
111	19.800000000
112	16.600000000
113	23.600000000
114	14.760000000
115	12.000000000
116	15.566666670
117	18.900000000
118	12.200000000
119	20.100000000
120	17.133333330
121	16.600000000
122	16.771428570

123	18.733333330
124	18.400000000
125	22.600000000
126	24.800000000
127	10.200000000
128	15.400000000
129	22.733333330
130	16.640000000
131	11.500000000
132	16.800000000
133	14.500000000
134	21.400000000
135	19.400000000
136	25.333333330
137	16.400000000
138	25.075000000
139	24.700000000
140	20.666666670
141	24.400000000
143	12.600000000
144	27.000000000
145	20.800000000
146	27.800000000
147	16.120000000
149	21.000000000
150	18.266666670
151	23.350000000
153	20.350000000
154	30.400000000
155	22.266666670
157	23.200000000
158	20.900000000
161	25.600000000
162	18.133333330
164	19.920000000
165	24.400000000
166	23.400000000
168	22.550000000
169	26.400000000
170	26.066666670
171	26.400000000
172	30.200000000

177	18.900000000
178	25.000000000
179	22.400000000
183	27.900000000
185	28.200000000
186	26.800000000
187	28.400000000
193	19.500000000
195	20.800000000
199	28.000000000
201	28.000000000
202	27.500000000
203	29.200000000
207	21.200000000
209	27.700000000
213	30.200000000
215	29.800000000
216	32.600000000
217	30.800000000
222	21.700000000
224	22.500000000
228	29.900000000
232	24.400000000
247	24.400000000
253	23.700000000
255	21.400000000
259	31.900000000
263	26.000000000
268	24.200000000
270	26.500000000
274	32.400000000
278	30.800000000
283	25.100000000
285	26.000000000
290	28.800000000
298	26.800000000
305	27.900000000
313	26.600000000
320	23.000000000
328	26.200000000

ANEXOS

ANEXO A – Script da simulação

```

Script:
# a: parâmetro para estrutura de covariância. Coloque zero para medidas
repetidas independentes.
# mu: média real
# n: número tempos distintos onde foram realizadas medidas
# m: número de indivíduos
# pho: expoente do tempo indicando o acréscimo de variância no tempo
library(MASS)
library(mvtnorm)

gera_heteroced_2<-
function(a=0,sigma=1,n=10,mu=10,m=200,pho=5/3)
{

  tempo<- 1:n
  Sigma<- matrix(0,n,n)
  diag(Sigma)<- sigma*(tempo^pho)
  expoentes<- 1:((n*n-n)/2)

  for(i in 2:n){Sigma[i,1:(i-1)]<- a^(2/expoentes)[1:(i-1)]; Sigma[1:(i-
1),i]<- a^(2/expoentes)[1:(i-1)]} # covariância.

  Y<- matrix(replicate(m,rmvnorm(1,rep(mu,n),Sigma)),ncol=1) #
gerando dados de uma Normal e empilhando em um único vetor
  Y<- cbind(Y,rep(1:n,m))

  # plot(Y[,1]~Y[,2],xlab="Tempo")

  return(Y)
}

dados = read.table("drl.txt",header=TRUE)
dados

idade = dados$idade

#parâmetros escolhidos para gerar os pesos
a= 4.48
b= 0.11
c= 0.13

```

```

d= -0.10
x0=84
x00= 186
# número de parâmetros estimados em cada regressão segmentada
p=1
q=2
r=3

n=length(idade)
n

# número de simulações
nsim = 1000

# preparando as matrizes para receber os avaliadores da qualidade de
ajuste, matriz formada de zeros com nsim (1000) linhas e três colunas em que
cada coluna apresenta o avaliador da qualidade do ajuste para o caso linear,
linear-linear e linear-linear-linear respectivamente.
resAIC = matrix(0,nsim,3)
resBIC = matrix(0,nsim,3)
resR2 = matrix(0,nsim,3)
resPV = matrix(0,nsim,3)
resPar = matrix(0,nsim,2)

for(i in 1:nsim){
#gerando erros heterocedásticos e dependentes.
erro =
(gera_heteroced_2(a=0,n=length(idade),mu=0,m=1,pho=5/3,sigma=0.005))[,1]

# gerando o y (peso) seguindo ajuste linear
y = (a+b*idade)+erro

## linear (utilizando o y gerado e fazendo o ajuste linear)

reg = lm(y ~ idade)

#y (peso) estimado
yest = predict(reg)

```

```

#soma de quadrado do erro
errl = (y-vest)^2
SQerrl = sum(errl)

#soma de quadrado da regressão
SQRegl = var(yest)*(n-1)

#soma de quadrado total
SQTotall = var(y)*(n-1)

# cálculo dos avaliadores da qualidade do ajuste
aicl = AIC(reg)

bicl = BIC(reg)

r2l=1-((n-1)/(n-(p+1)))*(1-(SQRegl/SQTotall))

## linear-linear
# criando a variável x2= 0, para idade menor que x0 e x2= idade-x0,
para idade maior que x0
x2=as.numeric(idade>x0)

x2star = (idade-x0)*x2

# utilizando o y gerado (peso), no caso com um segmento, e fazendo o
ajuste com 2 segmentos
regll = lm(y ~ idade + x2star)

yestll = predict(regll)

#soma de quadrado do erro
errll = (y-yestll)^2
SQerrll = sum(errll)

#soma de quadrado da regressão
SQRegll = var(yestll)*(n-1)

#soma de quadrado total
SQTotalll = var(y)*(n-1)

#avaliadores da qualidade do ajuste
aicll = AIC(regll)

```

```

bicll = BIC(regll)

r2ll=1-((n-1)/(n-(q+1)))*(1-(SQRegll/SQTotalll))

## linear-linear-linear
# criando as variáveis x2 e x3, em que x2 segue da forma que foi
definida anteriormente e x3=0 se idade menor do que o segundo nó e x3=idade-
x00, se idade maior que segundo nó (x0=primeiro nó e x00 segundo nó)

x2lll=as.numeric(idade>x0)

x2lllstar = (idade-x0)*x2lll

x3=as.numeric(idade>x00)

x3star = (idade-x00)*x3

#Utilizando o y gerado (como um segmento) e fazendo o ajuste com três
segmentos
reglll = lm(y ~ idade + x2lllstar + x3star)

yestlll = predict(reglll)

#soma de quadrado do erro
errlll = (y-yestlll)^2
SQerrlll = sum(errlll)

#soma de quadrado da regressão
SQReglll = var(yestlll)*(n-1)

#soma de quadrado total
SQTotallll = var(y)*(n-1)

#avaliadores da qualidade do ajuste
aiclll = AIC(reglll)
biclll = BIC(reglll)
r2lll=1-((n-1)/(n-(r+1)))*(1-(SQReglll/SQTotalll))

## soma de quadrados extras
# SSE(X1), valor F e p-valor para SSE(X1)

```

```

sq1=SQRegI
f1=(sq1/1)/(SQerrlll/n-4)
pvalor1= 1-pf(f1,1,n-4)

# SSE(X2| X1) valor F e p-valor para SSE(X2| X1)
sq2= SQerrl-SQerrll
sq22=SQRegll-SQRegI
f2=(sq22/1)/(SQerrlll/n-4)
pvalor2= 1-pf(f2,1,n-4)

# SSE(X3|X2, X1) valor F e p-valor para SSE(X3|X2, X1)
sq3= SQerrll-SQerrlll
sq33=SQReglll-SQRegll
f3=(sq33/1)/(SQerrlll/n-4)
pvalor3= 1-pf(f3,1,n-4)

## Estimando os parâmetros da spline com um segmento, utilizando a
idade do grupo genético DRL e o y simulado (com um segmento), por meio do
método de mínimo quadrados

X0 = rep(1,length(idade))
X = matrix(c(X0,idade),ncol=2)
res = solve(t(X)%*%X)%*%(t(X)%*%y)

## guardando os resultados, primeira coluna o linear, segunda coluna o
linear- linear e terceira coluna linear- linear- linear
#avaliadores da qualidade de ajuste
resAIC[i,1] = aicl ; resAIC[i,2] = aicll ; resAIC[i,3] = aiclll

resBIC[i,1] = bicl ; resBIC[i,2] = bicll ; resBIC[i,3] = biclll

resR2[i,1] = r2l ; resR2[i,2] = r2ll ; resR2[i,3] = r2lll
# p-valor para as somas de quadrados extras
resPV[i,1] = pvalor1 ; resPV[i,2] = pvalor2 ; resPV[i,3] = pvalor3

#parâmetros estimados para a regressão com dois segmentos
resPar[i,1] = res[1] ; resPar[i,2] = res[2]
}

```

```
resAIC  
resBIC  
resR2  
resPV  
resPar
```

```
# estatística descritiva dos resultados  
summary(resAIC)  
summary(resBIC)  
summary(resR2)  
summary(resPV)  
summary(resPar)
```

Os demais cenários seguiram de forma análoga: no segundo, foi gerado o y (peso) com dois segmentos e foram ajustados com um, dois e três segmentos e as mesmas análises foram realizadas, no terceiro o y gerado possuía três segmentos.