



UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Iago Renan Valentim da Silva

**Validação de modelos de crescimento aplicados
ao perfil de mudanças fisiológicas do sexo
feminino**

Campina Grande - PB

Dezembro 2019

Iago Renan Valentim da Silva

**Validação de modelos de crescimento aplicados ao perfil
de mudanças fisiológicas do sexo feminino**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Profa. Dra. Ana Patrícia Bastos Peixoto

Coorientador: Profa. Dra. Glaucia Amorim Faria

Campina Grande - PB

Dezembro 2019

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586v Silva, Iago Renan Valentim da.
Validação de modelos de crescimento aplicados ao perfil de mudanças fisiológicas do sexo feminino [manuscrito] / Iago Renan Valentim da Silva. - 2019.
36 p. : il. colorido.
Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2019.
"Orientação : Profa. Dra. Ana Patrícia Bastos Peixoto, Coordenação do Curso de Estatística - CCT."
"Coorientação: Profa. Dra. Gláucia Amorim Faria, UNESP - Universidade Estadual Paulista"
1. Modelos não lineares. 2. Curvas de crescimento. 3. Métodos de estimação. I. Título
21. ed. CDD 519.5

Iago Renan Valentim da Silva

Validação de modelos de crescimento aplicados ao perfil de mudanças fisiológicas do sexo feminino

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 04/12/2019.

BANCA EXAMINADORA



Profa. Dra. Ana Patrícia Bastos Peixoto
Universidade Estadual da Paraíba



Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba



Prof. Dr. Silvio Fernando Alves Xavier
Júnior
Universidade Estadual da Paraíba

Resumo

O presente trabalho expõe a aplicação de modelos de crescimento no que se refere ao desenvolvimento fisiológico em pessoas do sexo feminino, caracterizado pela puberdade, sendo esta, o processo de determinação biológica para todos os seres humanos, na qual se restringe ao período de transição da criança para fase adulta, popularmente chamada de adolescência. Foram ajustadas curvas de crescimento associadas a altura e peso de meninas dos oito aos dezoito anos de idade e os dados em estudo foram submetidos ao ajuste dos modelos Logístico, Gompertz e Von Bertalanffy. O método utilizado para o ajuste foi o processo iterativo de Gauss Newton, conseqüentemente foi realizada a interpretação biológica dos parâmetros, o coeficiente de determinação, o intervalo de confiança a 95% das estimativas, seguida da análise dos resíduos projetados. Também foi empregado o critério de Informação de Akaike e o critério de Informação de Bayes e análise das distâncias de Cook para selecionar o modelo que mais se adequou aos dados. O modelo Logístico se mostrou o mais adequado, sendo o mais indicado para representar curvas de crescimento médio da altura e do peso de meninas.

Palavras-chaves: Modelos não lineares. Curvas de crescimento. Critério de seleção.

Abstract

This paper exposes the application of growth models with regard to physiological development in females, characterized by puberty, which is the biological determination process for all human beings, which is restricted to the transition period of child to adulthood, popularly called adolescence. Growth curves associated with height and weight of girls from eight to eighteen years old were adjusted and the data under study were adjusted to the Logistic, Gompertz and Von Bertalanffy models. The method used for the adjustment was the iterative Gauss Newton, consequently the biological interpretation of the parameters, the coefficient of determination, the confidence interval at 95 % of the estimates, followed by the analysis of the projected residues. The Akaike Information criterion (AIC) and Bayes Information criterion (BIC) and Cook distances analysis were also used to select the model that best fit the data. The Logistic model was the most appropriate, being the most suitable to represent average growth curves of height and weight of girls.

Key-words: Nonlinear models. Growth curves. Selection criteria.

Lista de ilustrações

Figura 1 – Representação da curva $(2^\beta 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5	13
Figura 2 – Representação da curva $(2^{\log \phi} 3^{\log \phi})^T$ com ϕ variando em espaçamentos iguais a 1,0.	14
Figura 3 – Gráfico de dispersão para a altura (a) e peso (b) das meninas durante seu desenvolvimento de acordo com a Idade (anos).	20
Figura 4 – Boxplot para a altura(a) e peso(b) das meninas durante seu desenvolvimento de acordo com a Idade(anos).	21
Figura 5 – Ajuste dos modelos Logístico, Gompertz e Von Bertalanffy para a altura das meninas durante a puberdade.	24
Figura 6 – Histogramas e o gráfico de envelope dos resíduos normalizados da media da altura das meninas para cada modelo.	25
Figura 7 – Histogramas e o gráfico de envelope dos resíduos normalizados da media do peso das meninas para cada modelo.	26
Figura 8 – Ajuste dos modelos Logístico, Gompertz e Von Bertalanffy para as médias semestrais da altura e peso das meninas durante a puberdade.	26
Figura 9 – Resíduos Projetados dos ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para a altura das meninas.	27
Figura 10 – Resíduos Projetados dos ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para o peso das meninas.	27
Figura 11 – Gráficos das distâncias de Cook para os ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para a altura das meninas.	28
Figura 12 – Gráficos das distâncias de Cook para os ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para o peso das meninas.	28
Figura 13 – Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo logístico, (a) θ_1 (b) θ_2 (c) θ_3 relacionado a altura e (d) θ_1 (e) θ_2 (f) θ_3 ao peso.	32
Figura 14 – Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo Logístico aos dados da altura (a) e do peso (b), e envelope simulado.	33

Lista de tabelas

Tabela 1 – Modelos de regressão não linear utilizados para descrever curvas de crescimento	16
Tabela 2 – Médias da altura e do peso das meninas durante seu desenvolvimento no intervalo de dez anos (dos 8 aos 18 anos).	22
Tabela 3 – Estimativas dos parâmetros para o modelo Logístico, Gompertz e Von Bertalanffy, erro padrão da estimativa (E.P.E.), valor- p para o teste t e intervalos de confiança (IC) de 95% para altura das meninas durante a puberdade	23
Tabela 4 – Estimativas dos parâmetros para o modelo Logístico, Gompertz e Von Bertalanffy, erro padrão da estimativa (E.P.E.), valor- p para o teste t e intervalos de confiança (IC) de 95% para o peso das meninas durante a puberdade	24
Tabela 5 – Valores dos Coeficientes de determinação ajustado (R_a^2), Critério de Informação de Akaike (AIC), critério de informação Bayesiano (BIC) e quadrado médio dos resíduos (QMRes) para os modelos Logístico, Gompertz e Von Bertalanffy para as variáveis altura e peso.	28
Tabela 6 – Medidas de curvatura quadrática média (RMS) para os modelos Logístico, Gompertz e Von Bertalanffy referente a altura e peso das meninas.	29
Tabela 7 – Cinco primeiras amostras da altura das meninas simuladas para o modelo logístico dos 8 aos 18 anos	30
Tabela 8 – Cinco primeiras Amostras para o peso das meninas simuladas para o modelo logístico dos 8 aos 18 anos	31
Tabela 9 – Estimativas dos parâmetros do modelo logístico para as cinco primeiras amostras simuladas do ajuste da altura e do peso.	31
Tabela 10 – Média, variância e erro quadrático médio das estimativa dos parâmetros do ajuste dos modelos da altura e do peso para as 1000 amostras simuladas	32
Tabela 11 – Estatísticas para as curvaturas dos modelos logístico para altura e peso, considerando as 1000 amostras simuladas.	33

Sumário

1	INTRODUÇÃO	8
2	FUNDAMENTAÇÃO TEÓRICA	9
2.1	Modelos de Regressão não Linear	9
2.2	Métodos de Estimação	9
2.2.1	Método dos Mínimos Quadrados	10
2.2.2	Método da Máxima Verossimilhança	10
2.2.3	Métodos Iterativos	11
2.2.3.1	Método Newton-Raphson	11
2.2.4	Método Gauss-Newton	12
2.3	Medidas de não linearidade	12
2.3.1	Medidas de curvatura de Bates e Watts	14
2.4	Modelos de Crescimento	15
2.5	Técnicas de Diagnóstico	16
2.5.1	Distância de Cook	17
2.5.2	Modelo Adequado	17
2.5.2.1	Coeficiente de determinação	17
2.5.2.2	Critério de Informação de Akaike (AIC)	18
2.5.2.3	Critério de Informação (BIC)	18
2.5.3	Resíduos Projetados	19
3	MATERIAL E MÉTODOS	20
4	RESULTADOS E DISCUSSÃO	23
4.1	Ajuste dos modelos para as médias da altura e do peso	23
4.2	Análise através de simulação	30
5	CONCLUSÃO	34
	REFERÊNCIAS	35

1 Introdução

A puberdade é um processo de determinação biológica para todos os seres humanos, na qual se restringe ao período de transição da criança para fase adulta, popularmente chamamos esse período de adolescência. A adolescência compreende a faixa etária situada entre os dez e dezoito anos de idade, já o processo de puberdade ocorre como uma parte dela, variando de pessoa para pessoa, pois deve-se levar em consideração fatores genéticos e socioeconômicos. As mudanças ocorridas no desenvolvimento do sexo feminino e masculino ocorrem de maneira diferente e em períodos diferentes. Para as meninas, a puberdade começa entre oito anos e meio e 10 anos e dura cerca de quatro anos.

De acordo com a ORGANIZAÇÃO MUNDIAL DA SAÚDE (2019) a adolescência constitui como a fase mais importante do crescimento e desenvolvimento de um indivíduo, marcada pelas transformações tanto no aspecto físico quanto psíquico e sociais. É durante a puberdade que as meninas são marcadas por todo o seu “amadurecimento”, algumas características são: o crescimento esquelético linear, ocasionando o aumento da estatura; a alteração da fisiologia e composição corporal no qual o formato do corpo muda, o percentual de gordura corporal aumenta e se acumula nos quadris e coxas.

Através do estudo de Regressão não linear é possível analisar fases de desenvolvimento utilizando curvas de crescimento envolvendo a adolescência feminina, possuindo variações de altura e peso, tornando-a uma possível fonte de estudo para fins estatísticos. Com isso, avaliar a possível relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns nesses estudos.

Segundo Bucur e Gazzola (2011) um modelo não linear é considerado o melhor, em comparação a outros, se ele apresentar o comportamento mais linear. Para avaliar a não linearidade de um modelo, as ferramentas utilizadas são as medidas de curvatura de Bates e Watts e o viés de Box. Na prática, a medida de vício de Box e as medidas de curvatura de Bates e Watts são as ferramentas mais utilizadas na avaliação da não linearidade de um modelo de regressão não linear, podendo, também, ser usadas em problemas de discriminação, uma vez que o melhor modelo possível, dentre todos os propostos, pode ser considerado como aquele que apresenta o comportamento mais próximo do comportamento linear (RATKOWSKY et al., 1983).

O presente trabalho tem o intuito de realizar um estudo da não linearidade de modelos de crescimento por meio de medidas de não linearidade, assim como caracterizar as curvas com suas devidas interpretações de acordo com os modelos ajustados, dentre eles os Logístico, Gompertz e o Von Bertalanffy propiciando inferências necessárias para o estudo em questão.

2 Fundamentação Teórica

2.1 Modelos de Regressão não Linear

Com o objetivo de estudar consequências físicas e biológicas, diferentes tipos de modelos surgiram em meados do século XX para solucionar equações diferenciais em regressão, buscando principalmente descrever fenômenos advindos de processos naturais, uma área na qual outrora nos deparamos com comportamentos lineares e também não lineares para parâmetros de um determinado modelo (PAINÉ et al., 2012). Comumente identifica-se que um modelo segue comportamento não linear, quando os estimadores de mínimos quadrados dos seus parâmetros não possuem propriedades desejáveis, ou seja, os coeficientes estimados são muito tendenciosos, têm um excesso considerável, variância acima do limite mínimo de variação e tem distribuição marcadamente distorcida. Assim, propriamente o denominamos “não linear”, pelo fato desses modelos de regressão diferirem muito da linearidade a medida em que os estimadores demonstram manifestações de comportamento não-linear.

Por definição os modelos de regressão não linear padrão possuem a seguinte forma:

$$y_i = f(\mathbf{X}_i, \theta) + \epsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

em que temos a variável resposta y_i , sendo $f(\mathbf{X}_i, \theta)$ a função das variáveis preditoras, \mathbf{X}_i o vetor de variáveis explicativas e θ os parâmetros a serem estimados. Os erros ϵ_i assim como nos modelos lineares também são considerados *iid* seguindo uma distribuição $\mathbf{N} \sim (0, \sigma^2 \mathbf{I})$.

Nesse estudo é importante saber que os modelos não lineares são caracterizados como exposto da equação (2.1) em que a função $f(\mathbf{X}_i, \theta)$ não apresenta linearidade em pelo menos um dos parâmetros θ_i , ou seja, para termos certeza de que um modelo é não linear basta que ao menos uma das derivadas da função $f(\mathbf{X}_i, \theta)$ com relação a θ dependa de ao menos um dos parâmetros θ_i .

2.2 Métodos de Estimação

Nas subseções a seguir, apresentamos um conjunto de métodos de estimação que podem ser utilizados para obter as estimativas dos parâmetros em regressão não linear, entre os vários métodos, o *método dos mínimos quadrados* é o mais popular. Na aplicação de modelos não lineares, o sistema de equações normais não é resolvido de maneira fácil ou ágil, mas graças a evolução da computação, métodos iterativos se tornaram cada vez mais utilizados, dentre os quais temos os métodos de *Gauss-Newton* e *Newton-Raphson* sendo os mais trabalhados atualmente, em um deles é possível estimar parâmetros desconhecidos

sem calcular a matriz hessiana e no outro seu uso é necessário (COELHO-BARROS et al., 2008).

2.2.1 Método dos Mínimos Quadrados

Admitindo o modelo apresentado em (2.1), o estimador de mínimos quadrados para θ , denotado por $\hat{\theta}$ é um ponto no espaço de parâmetros tais que $f(\hat{\theta})$ é mais próximo de y no espaço amostral entre todos os possíveis $f(\theta)$ na solução local (HUANG; HUANG, 2010). O estimador de mínimos quadrados é denominado assim, pois minimiza a soma de quadrados dos resíduos.

$$SQRes(\theta) = \sum_{i=1}^n \{y_i - f(x_i; \theta)\}^2, \theta \in \Theta \subset \mathbb{R}^p. \quad (2.2)$$

Quando f exposta na equação (2.2) é diferenciável em relação a cada um dos parâmetros θ , podemos encontrar o estimador de mínimos quadrados $\hat{\theta}$ a partir do seguinte sistema de equações

$$\left. \frac{\partial SQRes(\theta)}{\partial \theta_r} \right|_{\theta=\hat{\theta}} = 0, r = 1, \dots, p.$$

Assim, o sistema de *equações normais* para um modelo não linear é dado por,

$$\sum_{i=1}^n \left. \frac{\partial f(x_i; \theta)}{\partial \theta_r} \{y_i - f(x_i; \theta)\} \right|_{\theta=\hat{\theta}} = 0, r = 1, \dots, p. \quad (2.3)$$

Muitas vezes esse sistema de *equações normais* determinado em (2.3) não possui uma solução analítica, sendo necessário procedimentos iterativos.

2.2.2 Método da Máxima Verossimilhança

Considere uma distribuição normal independente e identicamente distribuída, logo os erros ϵ_i se distribuem da seguinte forma:

$$y_i - f(x; \theta) = \epsilon_i \sim N(0, \sigma^2 \mathbf{I}),$$

em que a função de log-verossimilhança é dada por

$$l(\theta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{SQRes(\theta)}{2\sigma^2}. \quad (2.4)$$

Quando em (2.4) σ^2 é conhecido, é maximizado em respeito ao parâmetro θ que é o mesmo que minimizar o $SQRes(\theta)$ (SEBER; WILD, 2003). Assim, o estimador de máxima verossimilhança θ_{mle} é o mesmo estimador dos mínimos quadrados $\hat{\theta}$. Por outro lado, quando σ^2 é desconhecido, $\partial l / \partial \sigma^2 = 0$ que tem solução

$$\sigma^2(\theta) = \frac{SQRes(\theta)}{n}.$$

Substituindo $\sigma^2(\boldsymbol{\theta})$ na log-verossimilhança obtêm-se o seguinte resultado

$$l(\boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta})) = \kappa - \frac{n}{2} \log(SQRes(\boldsymbol{\theta})),$$

no qual κ é alguma constante. Maximizando $l(\boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta}))$ com relação a $\boldsymbol{\theta}$, temos

$$\boldsymbol{\theta}_{mle} \quad e \quad \sigma_{mle}^2 = \frac{SQRes(\boldsymbol{\theta})}{n}.$$

Em resumo, quando a estrutura do ruído segue uma distribuição normal, a estimativa de máxima verossimilhança é igual às estimativas dos mínimos quadrados.

2.2.3 Métodos Iterativos

Afim de minimizar o $SQRes(\boldsymbol{\theta})$ em (2.2), tanto quanto maximizar $L(\boldsymbol{\theta}, \sigma^2)$, é necessário o uso de algum método numérico iterativo, pois geralmente essas equações não possuem solução analítica. Deste modo, será apresentado o método de *Newton-Raphson* e *Gauss-Newton*, que foi utilizado computacionalmente nesse estudo para aproximação da estimativa dos mínimos quadrados em modelos não-lineares (MATTOS, 2013).

2.2.3.1 Método Newton-Raphson

Tal método é usado para encontrar raízes de uma função. Portanto, não é um método de maximização ou minimização de funções, mas usa o fato equivalente de que a derivada no ponto crítico de uma função é zero (HARTMANN, 2005).

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

Seja $SQRes(\boldsymbol{\theta})$, efetuamos chutes geralmente com base na análise gráfica residual, e denotamos $\boldsymbol{\theta}^{(a)}$, com o objetivo de minimizar a soma de quadrados dos resíduos, iremos então aproximar da seguinte forma

$$SQRes(\boldsymbol{\theta}) \approx SQRes(\boldsymbol{\theta}^{(a)}) + \mathbf{U}'(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})' \frac{\mathbf{H}(\boldsymbol{\theta}^{(a)})}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}),$$

Sendo $\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial SQRes(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ o vetor gradiente e $\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 SQRes(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ a matriz Hessiana de $SQRes(\boldsymbol{\theta})$. A aproximação da função de soma de quadrado terá um ponto estacionário quando seu gradiente é zero, isto é, quando

$$\mathbf{U}(\boldsymbol{\theta}^{(a)}) + \mathbf{H}(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) = 0,$$

e este ponto estacionário será um mínimo se $\mathbf{H}(\boldsymbol{\theta}^{(a)})$ é positiva definida. Admitindo que $\mathbf{H}(\boldsymbol{\theta}^{(a)})$ é positiva definida, o próximo passo do algoritmo é

$$\boldsymbol{\theta}^{(\theta+1)} = \boldsymbol{\theta}^{(a)} = [\mathbf{H}(\boldsymbol{\theta}^{(a)})]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(a)}). \quad (2.5)$$

Dentre os muito existentes, um critério para verificar a convergência poderia ser

$$|SQRes(\boldsymbol{\theta}^{(a+1)}) - SQRes(\boldsymbol{\theta}^{(a)})| < 0,000001.$$

O método de *Newton-Raphson* funciona bem em muitas situações, mas pode apresentar desempenho ruim em alguns casos. Em particular, se a função não for aproximadamente quadrática ou se a estimativa corrente se encontrar muito distante do ponto ótimo, pode ocorrer problemas de convergências

2.2.4 Método Gauss-Newton

Suponha que $\boldsymbol{\theta}^{(a)}$ é uma aproximação da estimativa de mínimos quadrados $\hat{\boldsymbol{\theta}}$ de um modelo não linear (FRERY; CRIBARI-NETO, 2005). Para $\boldsymbol{\theta}$ próximo de $\boldsymbol{\theta}^a$, considera-se uma expansão em série de Taylor de primeira ordem como

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \approx \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}^{(a)}) + \mathbf{F}(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$$

onde $\mathbf{F}(\boldsymbol{\theta}) = \left[\frac{\partial(x_i, \theta)}{\partial \theta_j} \right]$. Definindo $\mathbf{r}(\boldsymbol{\theta})$ como sendo um vetor de resíduos,

$$\mathbf{r}\boldsymbol{\theta} \approx \mathbf{r}(\boldsymbol{\theta}^{(a)}) - \mathbf{F}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$$

substituindo $\mathbf{r}(\boldsymbol{\theta})'\mathbf{r}(\boldsymbol{\theta})$ em $SQres(\boldsymbol{\theta})$ pelo método de mínimos quadrados, tem-se que esta aproximação resulta em

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + [\mathbf{F}'^{(a)}\mathbf{F}^{(a)}]^{-1}\mathbf{F}'^{(a)}\mathbf{r}(\boldsymbol{\theta}^{(a)}), \quad (2.6)$$

O processo iterativo conhecido como Método de Gauss-Newton é repetido até obter a convergência.

2.3 Medidas de não linearidade

A medida que o tamanho da amostra cresce, o comportamento dos estimadores de mínimos quadrados para os modelos não lineares se torna cada vez mais próximo do comportamento dos estimadores lineares, mas, até recentemente, não havia prontamente métodos aplicáveis para quantificar o comportamento não linear. As expressões que indicam a adequação de uma aproximação linear e seus efeitos nas inferências são chamadas medidas de não linearidade (BATES; WATTS, 1980).

Ratkowsky et al. (1983) compararam algumas formas paramétricas para diversos modelos normais não lineares por meio de simulações e utilizou as medidas de Box e de Bates e Watts. Para se ter uma ideia mais clara dos conceitos de não linearidade intrínseca e de não linearidade paramétrica, serão comparados, a seguir, um modelo linear e um modelo não linear para o caso de $n = 2$ e $p = 1$, p é a quantidade de parâmetros. Considere,

inicialmente, o modelo linear simples $Y_i = \beta x_i + \epsilon_i, i = 1, 2$, em que x é uma covariável qualquer e β um parâmetro desconhecido. Nesse caso, o espaço de estimação tem dimensão igual a um, sendo formado pelos pontos

$$x\beta = (x_1, x_2)\beta, \beta \in \mathbb{R},$$

ou seja, é uma reta em R^2 . Além disso, para qualquer conjunto de soluções β^1, β^2, \dots , tais que $\beta^{(i+1)} - \beta^{(i)} = \Delta$, onde Δ é uma constante arbitrária, as soluções possíveis para $X\beta$ serão tais que

$$x\beta^{(i+1)} - x\beta^{(i)} = (x_1, x_2)\Delta, i = 1, 2, \dots,$$

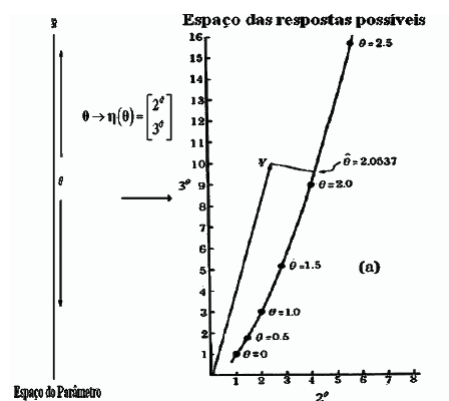
portanto, se as soluções para β forem igualmente espaçadas, então os valores ajustados correspondentes serão, também, igualmente espaçados.

Considere agora, o modelo normal não linear $y_i = x_i^\beta + \epsilon_i, i = 1, 2$ e os dados apresentados em (ZEVIANI, 2009)

$$\mathbf{y} = (2, 5, 10)^T \text{ e } \mathbf{x} = (2, 3)^T$$

Nesse caso, o espaço de estimação não é mais uma reta, e sim uma curva ao redor da estimativa de máxima verossimilhança $\hat{\beta} = 2,05$. A curva correspondente aos pontos $(2^\beta, 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5. Note que os pontos do espaço de estimação não são igualmente espaçados como ocorre no modelo linear (Figura 1). Assim, quanto mais essa curva se afasta da reta tangente em $\hat{\beta}$ maior será o que Bates e Watts (1980) chamam de não linearidade intrínseca do modelo, e quanto mais desiguais forem os espaçamentos entre os pontos do espaço de estimação, maior será o que ambos autores chamam de não linearidade aparente causada pela parametrização do modelo (PRUDENTE, 2009).

Figura 1 – Representação da curva $(2^\beta, 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5



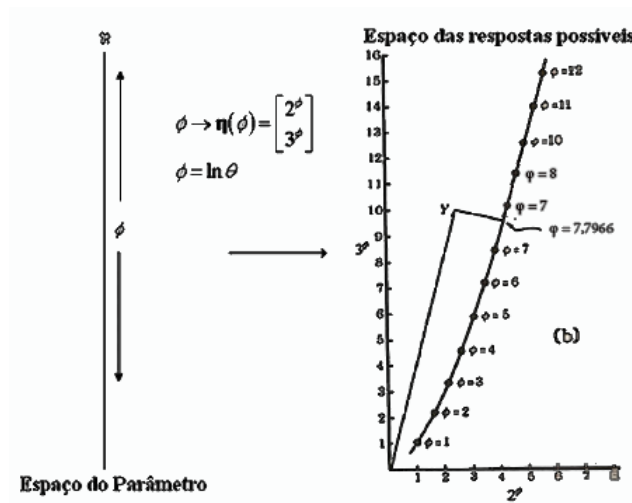
Fonte: (RATKOWSKY et al., 1983)

Portanto, a não linearidade de um modelo pode ser diversas causas. A primeira é a curvatura real do modelo ou intrínseca, como definem Bates e Watts (1980), que é invariante com qualquer tipo de reparametrização. A segunda é a curvatura devida à forma como os parâmetros aparecem no modelo. Essa última pode ser eliminada ou pelo menos reduzida por meio da reparametrização. Para ilustrar esse fato, considere o modelo normal não-linear descrito anteriormente com a seguinte reparametrização:

$$y_i = x_i^{\log \phi} + \epsilon_i, \quad i = 1, 2 \tag{2.7}$$

em que, $\phi = \exp(\beta)$. A Figura 2 mostra os pontos da curva $(2^{\log \phi}, 3^{\log \phi})^T$ com espaçamentos iguais a 1,0 para ϕ . Nota-se que os espaçamentos entre os pontos correspondentes são praticamente iguais, indicando que o grau de não linearidade aparente foi, substancialmente, reduzido com essa reparametrização. Entretanto, a curvatura do espaço de estimação continua com a mesma forma anterior, com era de se esperar.

Figura 2 – Representação da curva $(2^{\log \phi}, 3^{\log \phi})^T$ com ϕ variando em espaçamentos iguais a 1,0.



Fonte: (RATKOWSKY et al., 1983)

2.3.1 Medidas de curvatura de Bates e Watts

Um tratamento explicitamente mais geométrico para a não linearidade foi apresentado por Bates e Watts (1980), que propuseram a quantificação da não linearidade em duas componentes.

Considere o modelo de regressão normal não linear $Y = g(\beta; X) + \epsilon = \mu(\beta) + \epsilon$. Uma reta no espaço paramétrico passando por $\hat{\beta}$, pode ser expressa usando um parâmetro escalar b , dado por

$$\beta b = \hat{\beta} + b\mathbf{h},$$

em que $\mathbf{h} = (h_1, \dots, h_p)^T$ é um vetor de valores não-nulos. Essa reta gera uma curva, sobre o espaço, definida por

$$\mu_h(b) = \mu(\hat{\beta} + b\mathbf{h}).$$

Aceleração da curva μ_h é estimada por

$$\hat{\mu}_h = \mathbf{h}^T \hat{\mathbf{W}} \mathbf{h}$$

em que \mathbf{W} é uma matriz de dimensão $n \times (p \times p)$ com i -ésima face expressa como $W_i = \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}$, $i = 1, \dots, n$ e $r, s = 1, \dots, p$. Portanto, cada elemento indexado por i é da forma $\mathbf{h}^T \hat{\mathbf{W}}_i \mathbf{h}$, para $i = 1, \dots, n$. Bates e Watts (1980), descreveram as seguintes curvaturas:

Curvatura intrínseca definida por:

$$K_h^{IN} = \frac{\|\hat{\mu}^{IN}\|}{\|\hat{\mu}\|^2}$$

Curvatura devida à parametrização definida por

$$K_h^{PE} = \frac{\|\hat{\mu}^{PE}\|}{\|\hat{\mu}\|^2}$$

A reparametrização não mudará o formato da superfície esperada. Entretanto, pode tornar as propriedades dos modelos não lineares mais similares aquelas dos modelos lineares com relação a questão das variações iguais em β não produzem variações iguais em $\mu(\beta)$.

Essas curvaturas podem ser padronizadas de modo que fiquem invariantes a mudanças de escala. Após padronizadas as medidas relativas podem ser usadas não somente para comparar diferentes parametrizações de um determinado modelo, mas também diferentes conjuntos de dados para o mesmo modelo ou para modelos diferentes. Para efetuar os cálculos não há, em geral, fórmulas explícitas, sendo necessário recorrer a algum processo iterativo.

Segundo Bates e Watts (1980), as reparametrizações não mudam a forma da superfície esperada, nem o efeito da curvatura intrínseca, sendo invariante com qualquer tipo de reparametrização, que é uma propriedade do local da solução. Assim, uma boa reparametrização pode mudar consideravelmente a curvatura de efeito parâmetro, sendo eliminada ou pelo menos reduzida, tornando mais válidos os resultados das inferências que foram obtidos pela aproximação linear e pela teoria assintótica.

2.4 Modelos de Crescimento

O estudo de modelos de crescimento pode ser utilizado em diferentes áreas de pesquisa, contudo se torna bastante interessante quando partimos geralmente de dados

biológicos (PAINÉ et al., 2012). Na biologia podemos estar interessados em analisar o crescimento de animais e plantas com objetivo de compreender seus comportamentos e utilizarmos os resultados esperados, por exemplo, para melhorar a economia de uma empresa através de experimentos que resultem na diminuição de tempo e custeio até seu faturamento, para plantas uma das variáveis interessantes é o tamanho em que pode assumir uma assíntota por causa da limitação dos recursos do subsolo ou alterações ontogenéticas, como o início da floração.

As curvas de crescimento esboçam modelos assintóticos que geralmente representam o peso e altura durante toda vida útil t (Aggrey (2002), Kimura (1980)). Para modelar esse tipo de comportamento existem diversos modelos dentre os quais se destacam:

Tabela 1 – Modelos de regressão não linear utilizados para descrever curvas de crescimento

Modelo	Função
Logístico	$y_t = \theta_1(1 + \theta_2 e^{-\theta_3 t})^{-m}$
Gompertz	$y_t = \theta_1 e^{-\theta_2 e^{-\theta_3 t}}$
Von Bertalanffy	$y_t = \theta_1(1 - \theta_2 e^{-\theta_3 t})^3$

De acordo como foram apresentados na Tabela 1, a função y_t é a resposta média, θ_1 é o parâmetro que representa o valor assintótico da curva, isto é, o valor médio livre das variações sazonais, θ_2 é uma constante que esta relacionada aos valores iniciais, indicando a proporção do crescimento assintótico a ser ganho, θ_3 e a taxa de crescimento da variável de interesse, determina a eficiência do crescimento, t e o tempo em dias, m representa o parâmetro que define a forma da curva nos modelos e conseqüentemente determina o ponto na qual passa de uma fase de crescimento inibitório, indicando o ponto que ele passa a crescer com menor e ciência.

2.5 Técnicas de Diagnóstico

Diagnosticar modelos de regressão não-linear é bastante importante para a análise de dados. Os principais fatores a serem observados são a não-linearidade do modelo e as características dos dados. Depois de obter as estimativas de mínimos quadrados, algumas verificações de diagnóstico da magnitude da não linearidade intrínseca por efeito de parâmetros são necessários para avaliar o desempenho dos estimadores e o modelo ajustado. Alguns métodos práticos podem ser discutidos afim de validar a qualidade e performasse dos ajustes dos modelos em questão.

- (i) *O número de iterações* necessárias para convergência no algoritmo de *Newton-Raphson*. Uma grande quantidade de iterações até a convergência ou divergência do algoritmo serve como uma indicação de que o modelo e a combinação dos dados não

são próximos do linear e as propriedades de normalidade assintótica na convergência no ponto não é confiável.

(ii) *Medidas de curvatura na convergência e reparametrização.*

Conforme exposto por Bates e Watts (1980) definir limites nos valores máximos intrínsecos e nas medidas de não-linearidade por efeitos de parâmetro é importante para determinar se o estimador atinge um mínimo global na soma de quadrados dos resíduos. Além disso, eles ofereceram uma outra maneira para validar os efeitos de parâmetros, propondo uma re-parametrização do modelo estudado (BATES; WATTS, 1980).

2.5.1 Distância de Cook

Uma das mais conhecidas medidas de influência é a distância de Cook, que foi proposta por (COOK, 1977), inicialmente aplicada em modelos normais lineares e foi rapidamente estendida para classes de modelos não lineares, definida por

$$D_i = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})'(\hat{\mathbf{F}}'\hat{\mathbf{F}})(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})/p\hat{\sigma}^2. \quad (2.8)$$

Desenvolvendo a expressão (2.8), obtém-se a forma aproximada

$$D_i^1 = \frac{t_i^2 \hat{p}_{ii}}{p(1 - \hat{p}_{ii})}, i = 1, \dots, n.$$

2.5.2 Modelo Adequado

Para escolher o modelo adequado para seu experimento é essencial obter bons resultados, ou seja, o modelo selecionado deve representar muito bem o comportamento da variável resposta. Existem métodos que são capazes de medir a proporção da variação no modelo ajustado, uma delas é o coeficiente de determinação ajustado R_a^2 muito utilizado em regressão linear, embora seja uma ferramenta comum na aplicação de regressão linear, pode não ser interessante aplicá-lo em modelos não lineares, pois a quantidade de parâmetros nem sempre está relacionado com o número de variáveis explicativas. Alternativamente, é essencial outros testes de ajuste como o critério de informação Akaike (AIC) e o critério de informação bayesiano (BIC).

2.5.2.1 Coeficiente de determinação

Para regressão linear e não-linear, o coeficiente de determinação é possivelmente a estatística mais utilizada para avaliar a adequação dos modelos empíricos ajustados aos dados. Isto é porque o valor de R^2 é fornecido por todos os *softwares* atuais para análise de regressão (CORNELL; BERGER, 1987). Porém para o modelo não linear não é tão simples a aplicação desta definição, no entanto, um dos critérios para sua definição requer

a presença do intercepto no modelo, na qual, nem sempre está presente no modelo de regressão. Deste modo, consideraremos o R_a^2 como sendo uma medida próxima ao R^2 . Para o cálculo do R_a^2 temos

$$R_a^2 = 1 - \frac{SQE(\hat{\theta})}{\|y - \hat{y}\|^2},$$

em que, $SQE(\hat{\theta})$ é a soma dos quadrados dos resíduos avaliados em $\hat{\theta}$, y é o valor observado e \hat{y} indica o valor predito. Entretanto, apenas o valor do R_a^2 não serve como um critério adequado para verificação de ajuste dos modelos, em muitos casos o ajuste de modelos não lineares, é comum a obtenção de R_a^2 assintóticos altos e similares (RESENDE, 2007).

2.5.2.2 Critério de Informação de Akaike (AIC)

O critério de informação de Akaike é um método popular para comparar a adequação de vários modelos possivelmente ideais (AKAIKE, 1973). A seleção de modelos pode ser vista como uma forma de identificar o modelo que entre um conjunto de candidatos, está mais próximo da realidade. Procurando por um equilíbrio entre precisão e simplicidade.

O cálculo é feito com uma relação simples entre a log-verossimilhança da função avaliada em $(\hat{\theta})$ e o número de parâmetros (K):

$$AIC = -2 \ln(\hat{\theta}) + 2K.$$

Sendo assim, a partir dos modelos concorrentes opta-se por aquele que apresentar o menor valor de (AIC). Apesar da maioria entre os pesquisadores utilizarem o (AIC), alguns acreditam que é muito liberal e tendencioso selecionando modelos mais complexos. Foi apontado que (AIC) negligencia a variabilidade amostral dos parâmetros estimados. Uma outra alternativa popular que serve como critério de seleção de modelos é a informação bayesiana critério ou (BIC) (BURNHAM; ANDERSON, 2004).

2.5.2.3 Critério de Informação (BIC)

Uma abordagem diferente, mas simples, é a critério de informação Bayesiano (BIC) Schwarz et al. (1978), formulado como

$$BIC = -2 \ln(\hat{\theta}) + K \log n$$

no qual, n é o número de observações que entram no cálculo de probabilidade. O (BIC) é uma aproximação assintótica para seleção de modelo Bayesiano (BMS), em qual integra no espaço de parâmetros. Especificamente, o (BMS) requer o cálculo da probabilidade de dados fornecidos pelo modelo (POSADA; BUCKLEY, 2004). O valor mínimo do (BIC) estimado para um determinado modelo de um subsistema é utilizado para selecionar qual o modelo mais corresponde aos dados.

2.5.3 Resíduos Projetados

De acordo com Cook e Tsai (1985) pode-se empregar uma expansão em série de Taylor para aproximar e investigar o comportamento dos resíduos ordinários na regressão não linear. Entretanto, em alguns casos, esses resíduos podem produzir resultados enganosos quando usados em métodos de diagnósticos análogos aos da regressão linear, não refletindo de forma correta a distribuição dos erros. Sugere-se então a utilização dos resíduos projetados, sendo ele superior em qualidade de diagnóstico (MARTINS, 2013).

Seja $\tilde{\mathbf{X}} = \frac{\partial \mu(\beta)}{\partial \beta_r}$, $i = 1, \dots, n$ e $r = 1, \dots, p$ a matriz de derivadas de primeira ordem avaliada em θ (Jacobiana) também denominada por matriz modelo local. A matriz \mathbf{W} é uma matriz de dimensão $(n \times p) \times p$ de derivadas de segunda ordem avaliada em θ (Hessiana). As estimativas dos resíduos projetados pode ser calculados por

$$(\mathbf{I} - \mathbf{H}_2)\mathbf{r} = (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon} - (\mathbf{I} - \mathbf{H}_1)\boldsymbol{\epsilon}$$

em que, $\mathbf{H}_1 = \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}$ é operador de projeção ortogonal do espaço gerado sobre a projeção $\mathbf{S} = (\mathbf{I} - \mathbf{H})\mathbf{T}$, em que, \mathbf{T} é uma matriz $n \times q$ obtida pelos vetores não nulos, de derivadas de segunda ordem e $\mathbf{H} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T$. Sendo \mathbf{V} uma matriz $n \times (p + q)$ definida como $\mathbf{V} = (\tilde{\mathbf{X}}, \mathbf{S})$ e $\mathbf{H}_2 = \mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T$ o operador de projeção ortogonal em $C(\mathbf{V})$

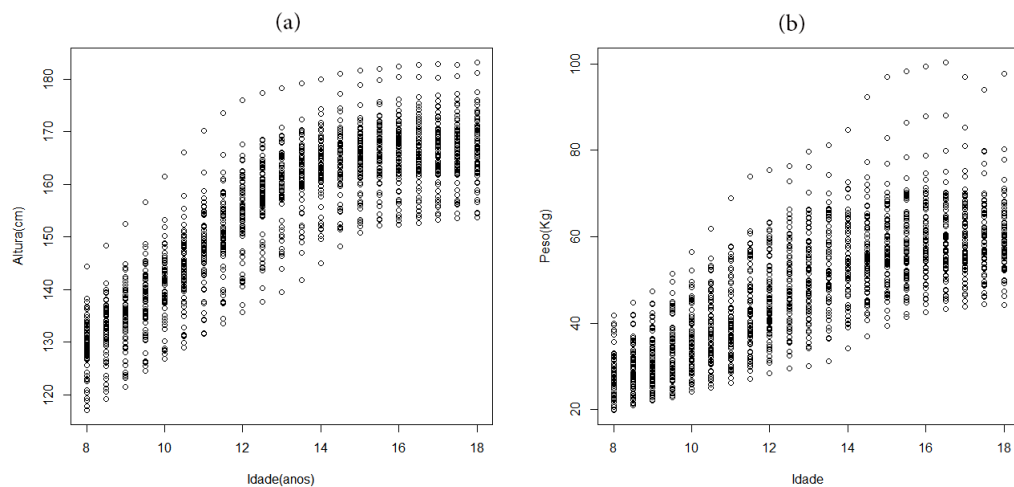
3 Material e Métodos

Os dados utilizados para esse trabalho foram originalmente apresentados por (TUD-DENHAM, 1954), na qual o Berkeley Guidance Study matriculou crianças nascidas em Berkeley, Califórnia, entre janeiro de 1928 e junho de 1929, e as avaliaram periodicamente até os dezoito anos de idade. As informações contidas no banco de dados possuem medidas antropométricas longitudinais de 136 crianças desde o nascimento até os 18 anos. As crianças foram medidas quanto à altura, peso (despido), comprimento do tronco, diâmetro biacromial, diâmetro bi-ilíaco, circunferência da perna e força dinamométrica.

A aplicação foi feita para a classe de modelos de crescimento discutida na metodologia, neste estudo objetivou-se analisar especificamente a altura(cm) e o peso(kg) de 70 meninas durante o período que antecede e sucede a puberdade que é justamente entre os 8 e 18 anos. Toda parte computacional foi resolvida utilizando o (R Core Team, 2019).

É importante saber que determinadas mudanças fisiológicas ocorridas no sexo feminino se iniciam geralmente pós menarca. Através do estudo de Tanner (1981) é evidenciado que a menarca acontece geralmente entre 10 e 12 anos, em alguns casos podendo ocorrer aos 15 anos. Existem poucos estudos que avaliam o envolvimento do crescimento após a ocorrência da primeira menstruação, período em que são extremamente visíveis mudanças na estrutura corporal feminina. Na Figura 3 é mostrado o gráfico de dispersão da altura (a) e peso (b) respectivamente para o desenvolvimento das meninas ao longo do intervalo pré estabelecido.

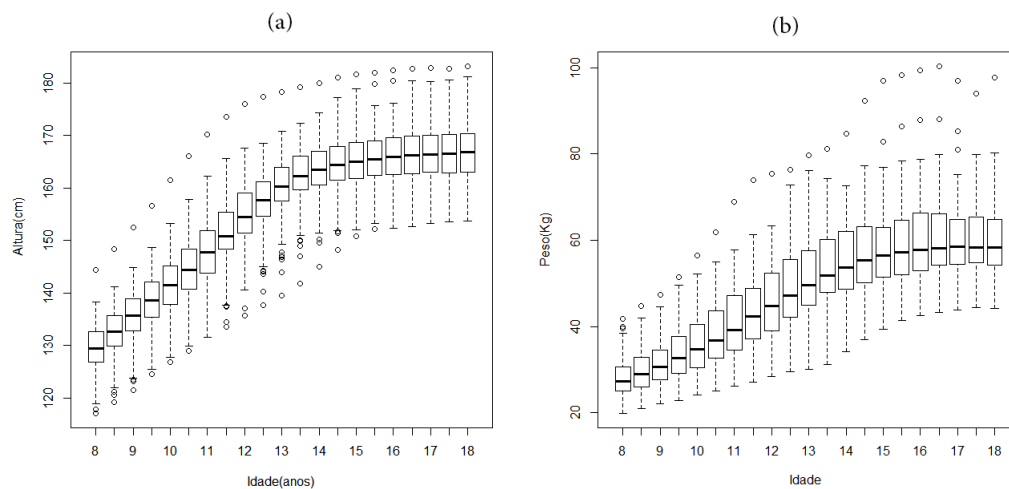
Figura 3 – Gráfico de dispersão para a altura (a) e peso (b) das meninas durante seu desenvolvimento de acordo com a Idade (anos).



Analisando descritivamente é possível observar que após os 16 anos de idade a altura das crianças (agora adolescentes), estabiliza, a partir do qual decorre menores incrementos em sua estatura, ou seja, as meninas apresentam taxa de crescimento até os 16,5 anos, algo que é biologicamente esperado. Para a dispersão do peso é interessante saber que existe uma forte relação entre a altura e o peso, por isso a taxa de crescimento do peso se estabiliza na mesma faixa de idade, podendo ser comprovado estatisticamente utilizando o coeficiente de correlação de *Spearman* que obteve $\rho = 0,9929$, portanto as variáveis são fortemente relacionadas e ambas crescem no mesmo ritmo.

No gráfico de *boxplot*, apresentado na Figura 4, observa-se que após os 16 anos a variabilidade da altura e do peso pouco se altera até a chegada dos 18 anos.

Figura 4 – Boxplot para a altura(a) e peso(b) das meninas durante seu desenvolvimento de acordo com a Idade(anos).



Posteriormente foi utilizado o valor médio entre todas as meninas com idade entre 8 e 8,5 e assim sucessivamente (até os 18 anos) para altura e o peso das meninas, tornando possível o ajuste dos modelos de crescimento, conforme exposto na Tabela 2.

Tabela 2 – Médias da altura e do peso das meninas durante seu desenvolvimento no intervalo de dez anos (dos 8 aos 18 anos).

Idade	Altura	Peso
8	129,16	28
8,5	132,15	29,76
9	135,12	31,62
9,5	138,02	33,76
10	140,94	35,88
10,5	144,20	38,17
11	147,62	40,66
11,5	151,12	43,46
12	154,46	46,15
12,5	157,37	48,80
13	159,75	51,29
13,5	161,61	53,31
14	163,07	54,92
14,5	164,10	56,46
15	164,85	57,82
15,5	165,35	58,94
16	164,74	59,76
16,5	166,02	60,07
17	166,22	60,02
17,5	166,37	59,89
18	166,54	59,77

No estudo utilizou-se os modelos de regressão não linear Logístico, Gompertz e Von Bertalanffy apresentados na Tabela 1 afim de inferir a altura e o peso das meninas ao longo dos anos e comparar qual modelo teve melhor ajuste, além de verificar o grau de linearidade de cada modelo.

Para a estimação dos parâmetros do modelo não linear foi utilizada a técnica dos mínimos quadrados ordinários, com o uso do método iterativo de *Gauss-Newton* (BATES; WATTS, 1980). Dentre as estatísticas fornecidas pelo procedimento de estimação, foram obtidos intervalos de confiança para os parâmetros, bem como o coeficiente de determinação assintótico (R_a^2). Para a comparação dos modelos foram utilizado os critérios de informação AIC e BIC. Após obtenção das estatísticas citadas acima, procedeu-se a análise gráfica do modelo selecionado para verificar a qualidade do ajuste e a obtenção das medidas de não linearidade. Os ajustes inferenciais foram realizados com o uso do *software R* (R Core Team, 2019).

4 Resultados e Discussão

A seguir, são apresentados os principais resultados inferenciais obtidos, assim como os encontrados por meio de simulação.

4.1 Ajuste dos modelos para as médias da altura e do peso

Para realização do método iterativo de *Gaus-Newton*, foram obtidos os valores iniciais dos parâmetros por meio da linearização do modelo, sendo assim possível ajustar os modelos aos dados. A seguir são apresentadas as estimativas encontradas para os parâmetros dos modelos não lineares Logístico, Gompertz e o Von Bertalanffy, conforme a metodologia estabelecida, bem como os valores do erro padrão da estimativa, o teste t e os intervalos de confiança a 95%.

Observando a Tabela 3, sabe-se que o parâmetro θ_1 é a estatura máxima para o ajuste, θ_2 é uma constante que está relacionada aos valores iniciais, indicando a proporção de crescimento a ser ganho e θ_3 é o quanto as meninas crescem (cm) por mês. Também percebe-se que as estimativas dos parâmetros de todos os modelos foram significativas através do teste t , considerando um nível nominal de 5% de significância, encontrando-se dentro do intervalo de confiança assintótico de 95%.

Tabela 3 – Estimativas dos parâmetros para o modelo Logístico, Gompertz e Von Bertalanffy, erro padrão da estimativa (E.P.E.), valor- p para o teste t e intervalos de confiança (IC) de 95% para altura das meninas durante a puberdade

Modelos	Parâmetros	Estimativas	E.P.E.	Valor- p	IC (95%)
Logístico	θ_1	171,8286	1,4684	<0,0001	[169,2505; 175,1712]
	θ_2	1,2243	0,1768	0,0001	[0,8920; 1,5863]
	θ_3	0,2825	0,0228	0,0001	[0,2387; 0,3283]
Gompertz	θ_1	172,8328	1,7687	<0,0001	[169,7955; 176,9543]
	θ_2	0,2490	0,1787	0,0001	[0,4896; 1,1875]
	θ_3	0,0406	0,0235	0,0001	[0,2040; 0,2960]
Von Bertalanffy	θ_1	173,2195	1,8890	<0,0001	[170,0020; 177,6622]
	θ_2	0,6643	0,1191	0,0001	[0,4765; 0,9584]
	θ_3	0,2379	0,0237	0,0001	[0,1925; 0,2853]

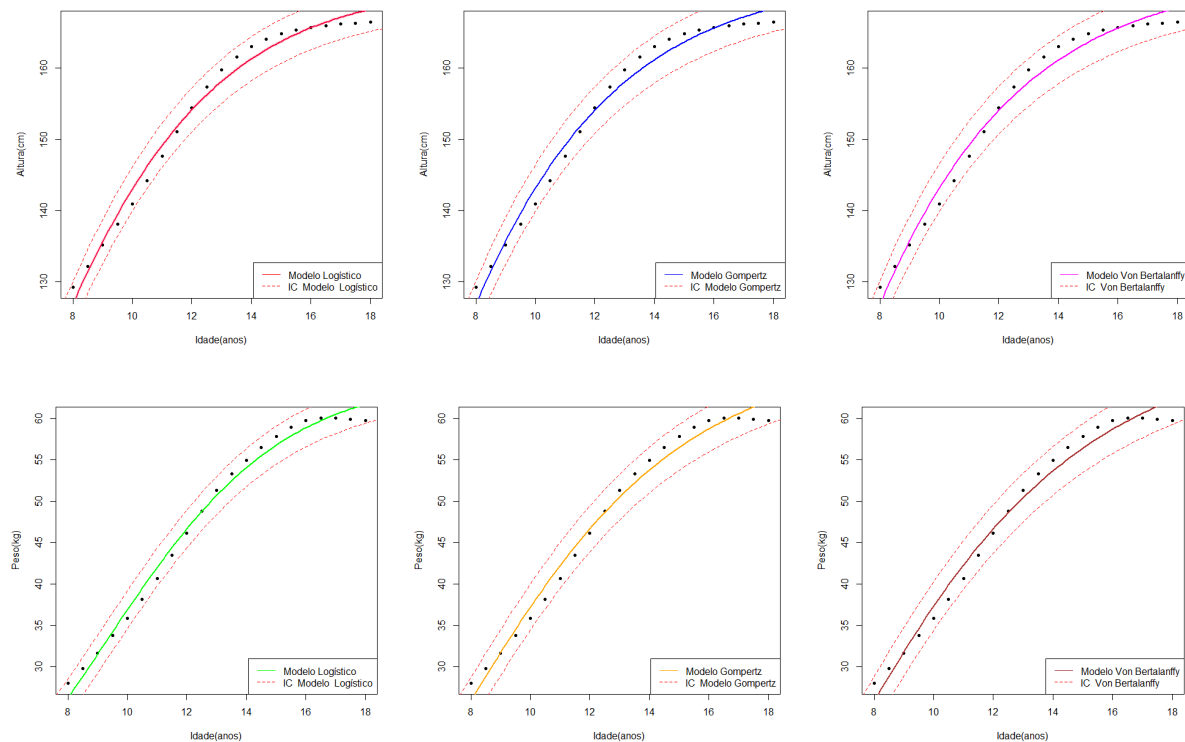
O mesmo raciocínio segue para a Tabela 4, o parâmetro θ_1 é o peso máximo para o ajuste, θ_2 é uma constante que está relacionada aos valores iniciais, indicando a proporção de crescimento a ser ganho e θ_3 é o quanto as meninas ganham peso (kg) por mês. As estimativas dos parâmetros de todos os modelos também foram significativas através do teste t , considerando um nível nominal de 5% de significância, e encontram-se dentro do intervalo de confiança assintótico de 95%.

Tabela 4 – Estimativas dos parâmetros para o modelo Logístico, Gompertz e Von Bertalanffy, erro padrão da estimativa (E.P.E.), valor- p para o teste t e intervalos de confiança (IC) de 95% para o peso das meninas durante a puberdade

Modelos	Parâmetros	Estimativas	E.P.E.	Valor- p	IC (95%)
Logístico	θ_1	65,0106	1,1704	<0,0001	[62,9235; 67,6417]
	θ_2	3,04182	0,1707	0,0001	[2,7121; 3,3959]
	θ_3	0,3312	0,0203	0,0001	[0,2916; 0,3724]
Gompertz	θ_1	68,0585	2,0576	<0,0001	[64,6202; 73,0100]
	θ_2	1,8456	0,1747	0,0001	[1,5179; 2,2054]
	θ_3	0,2350	0,0222	0,0001	[0,1921; 0,2794]
Von Bertalanffy	θ_1	69,6372	2,5780	<0,0001	[65,4534; 76,0719]
	θ_2	1,4321	0,2513	0,0001	[1,0343; 2,0539]
	θ_3	0,2032	0,0228	0,0001	[0,1593; 0,2486]

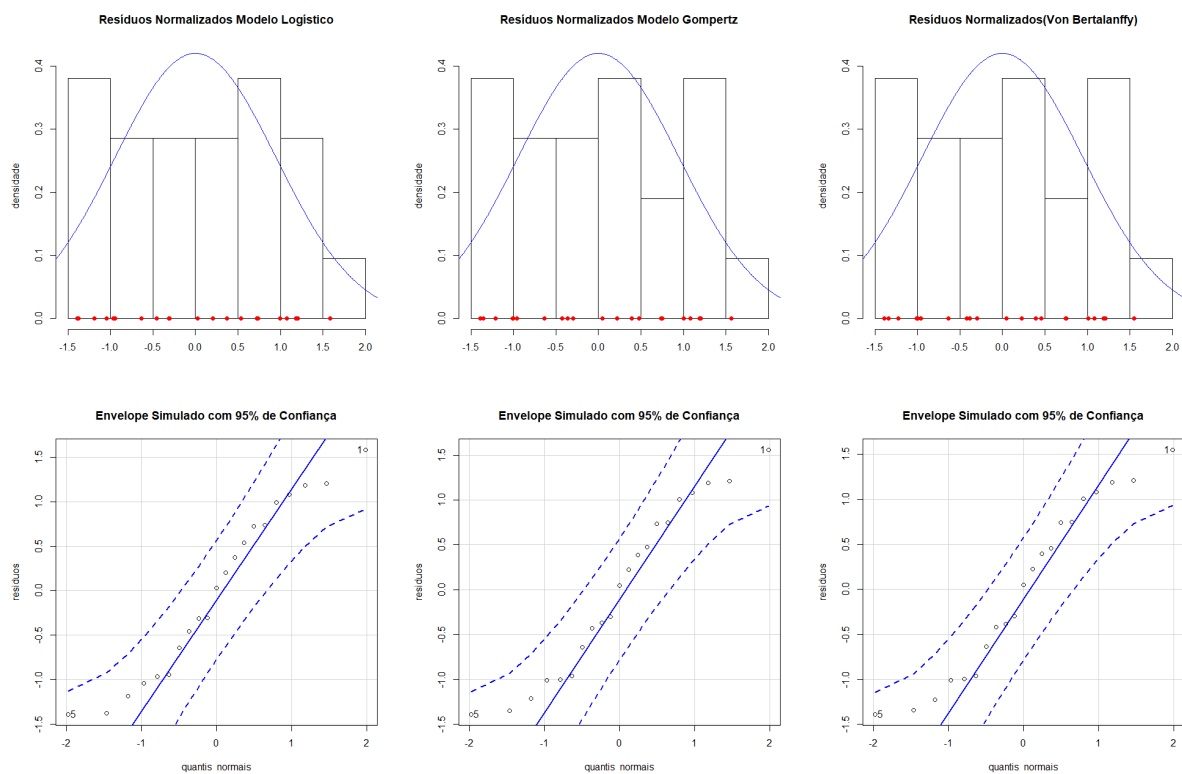
A Figura 5 expõe as curvas ajustadas para os três modelos e os seus respectivos intervalos com 95% de confiança, nota-se que todos os pontos estão dentro do intervalo, evidenciado que os ajustes se mostram satisfatórios. Contudo, o fato de os parâmetros serem significativos não é suficiente para validação do ajuste dos modelos, é necessário realizar a verificação da qualidade do ajuste por meio de uma análise de diagnóstico.

Figura 5 – Ajuste dos modelos Logístico, Gompertz e Von Bertalanffy para a altura das meninas durante a puberdade.



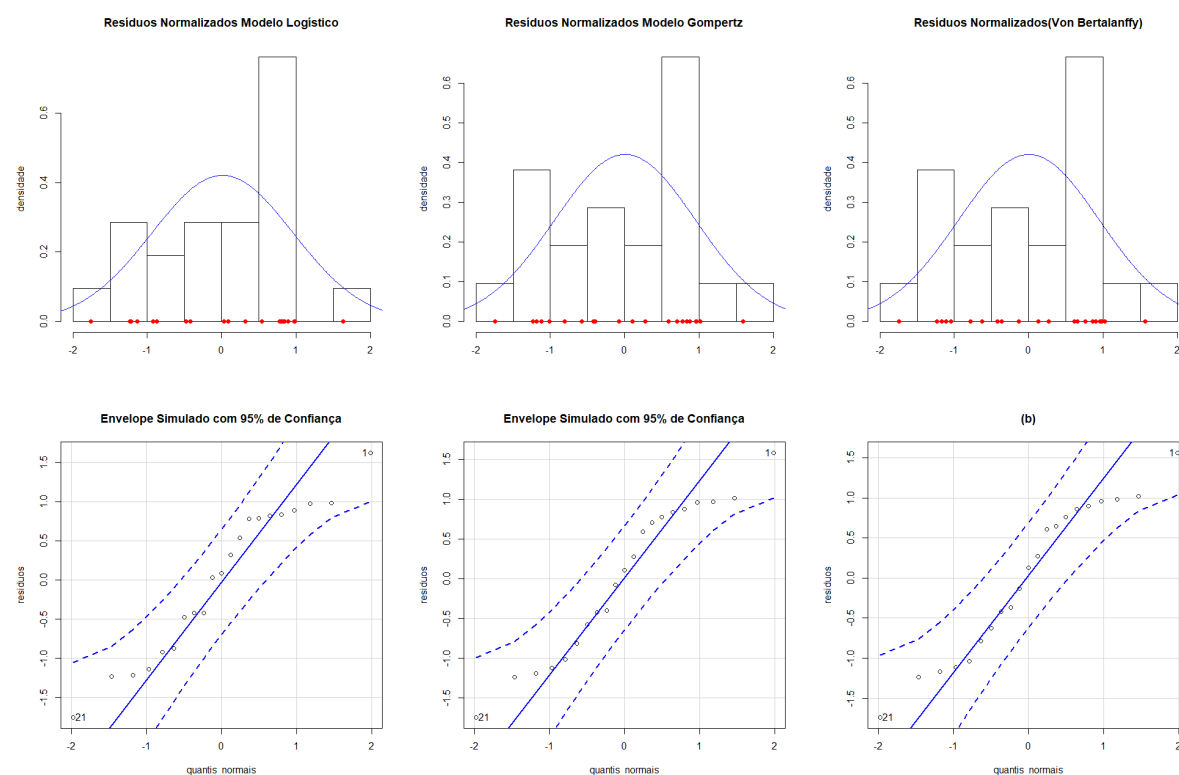
Na Figura 6 são apresentados os histogramas dos resíduos normalizados e a densidade de uma distribuição normal com mesma média e variância da altura (cm), e logo abaixo os envelopes simulados com 95% de confiança dos resíduos normalizados, para cada ajuste. Por meio da inspeção dos gráficos não é possível avaliar se os resíduos de todos os modelos seguem normalidade. Para confirmar estatisticamente se os resíduos de cada ajuste seguem normalidade foi realizado o teste de Shapiro-Wilk no qual os p -valores obtidos foram 0,21, 0,18 e 0,17 respectivamente, ou seja, não existem evidências para rejeitar a hipótese nula de que os resíduos seguem normalidade.

Figura 6 – Histogramas e o gráfico de envelope dos resíduos normalizados da media da altura das meninas para cada modelo.



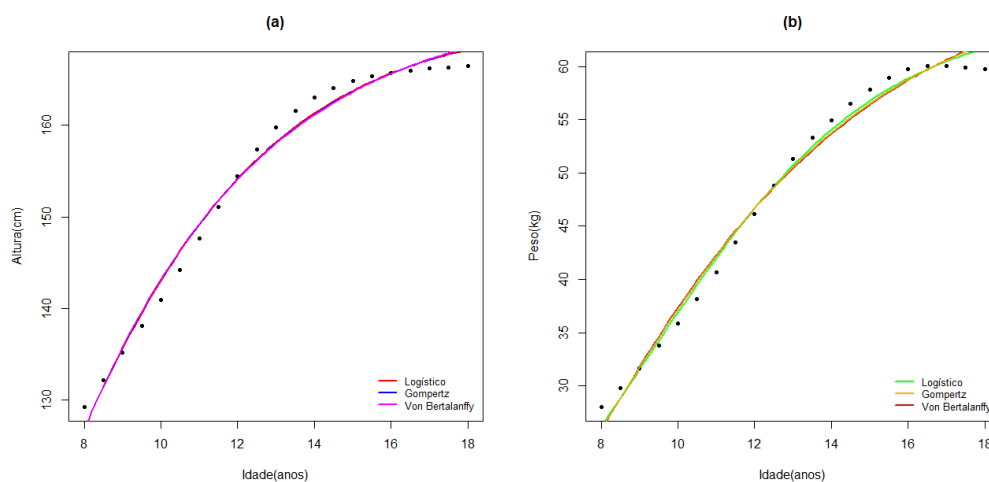
Da mesma forma são apresentados na Figura 7 os histogramas dos resíduos normalizados e a densidade de uma distribuição normal com mesma média e variância do peso (kg), e logo abaixo os envelopes simulados com 95% de confiança dos resíduos normalizados, para cada ajuste. Realizando o teste de *Shapiro-Wilk* no qual os p -valores obtidos foram 0,20, 0,22 e 0,21 respectivamente, ou seja, não existem evidências para rejeitar a hipótese nula de que os resíduos seguem normalidade.

Figura 7 – Histogramas e o gráfico de envelope dos resíduos normalizados da media do peso das meninas para cada modelo.



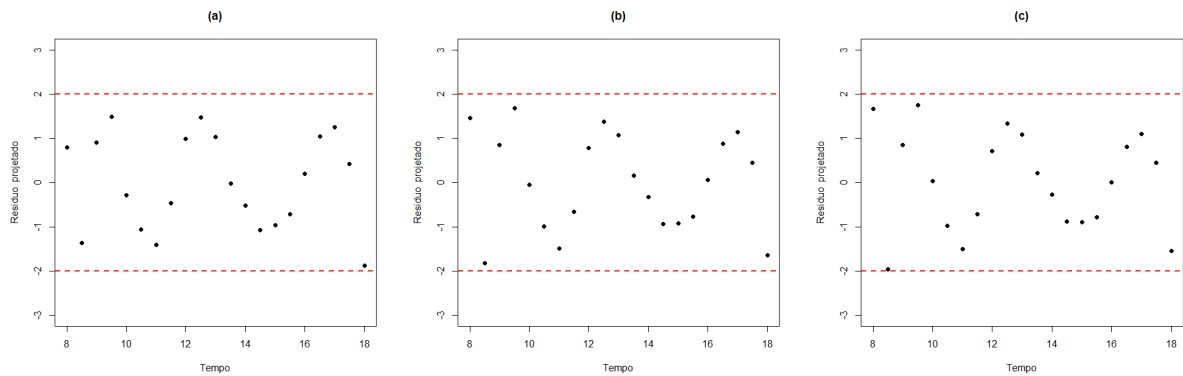
A Figura 8 apresenta as curvas ajustadas para os pontos médios da altura (a) e do peso (b), juntamente com os três modelos ajustados. Nota-se que as curvas ajustadas para altura ficaram muito semelhantes, dificultando julgar o modelo mais adequado, o mesmo raciocínio segue para as curvas ajustadas do peso.

Figura 8 – Ajuste dos modelos Logístico, Gompertz e Von Bertalanffy para as médias semestrais da altura e peso das meninas durante a puberdade.



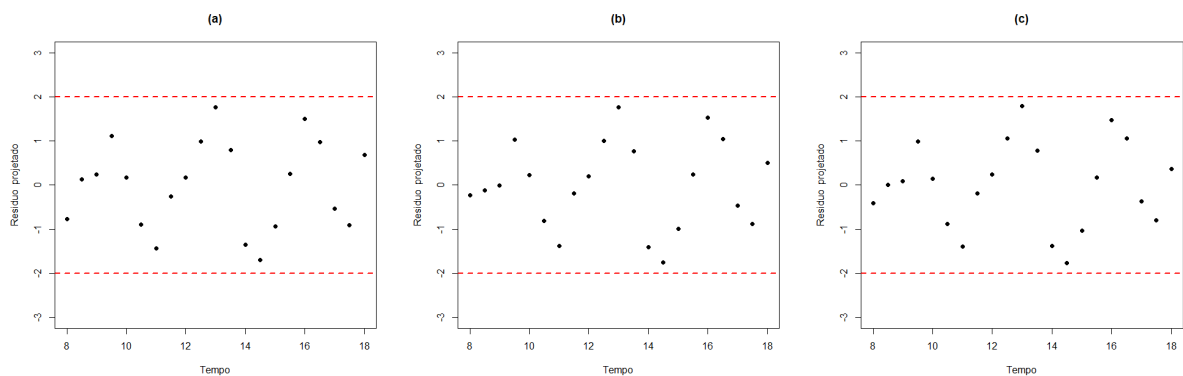
A seguir é apresentado o gráfico para os resíduos projetados dos modelos ajustados para altura na Figura 9. Pode ser observado que os resíduos projetados se apresentam de forma aleatória em todos os ajustes, descartando a suspeita de pontos discrepantes nos dados observados, pois nenhum dos valores está fora do intervalo $(-2, 2)$.

Figura 9 – Resíduos Projetados dos ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para a altura das meninas.



De maneira semelhante na Figura 10, os resíduos projetados dos modelos ajustados para o peso se apresentam de forma aleatória em todos os ajustes, também descartando a suspeita de pontos discrepantes nos dados observados, pois nenhum dos valores está fora do intervalo $(-2, 2)$.

Figura 10 – Resíduos Projetados dos ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para o peso das meninas.



Ainda como ferramenta de diagnóstico de ajuste dos modelos propostos no trabalho, são apresentados, nas Figuras 11 e 12, os gráficos da distância de Cook. Para altura, verifica-se a presença de um possível ponto discrepante na 1ª observação. Por outro lado, observando as distâncias de Cook para os ajustes do peso, a 1ª observação também expõe o fato de ser um ponto discrepante, mas com um adicional, no modelo Gompertz a 21ª observação é tida como um ponto influente. Contudo, a retirada desse ponto teria que ser profundamente analisada, pois certamente causaria variação nas estimativas dos parâmetros.

Figura 11 – Gráficos das distâncias de Cook para os ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para a altura das meninas.

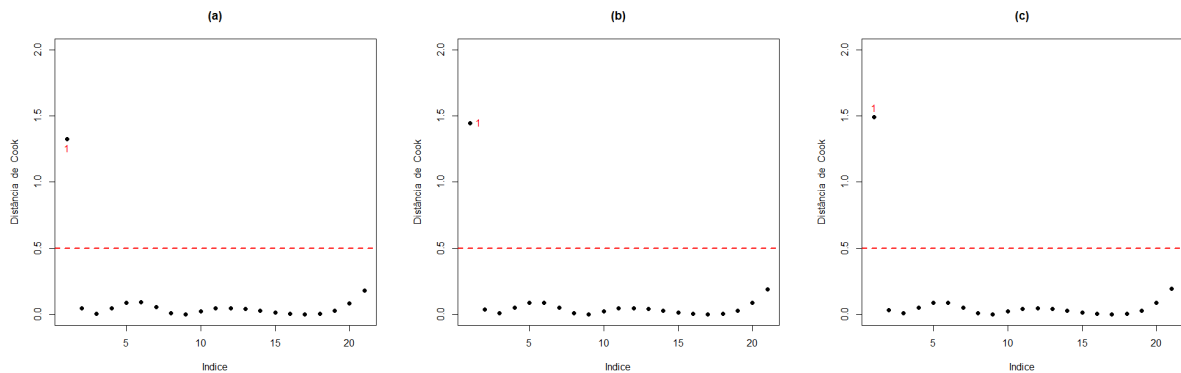
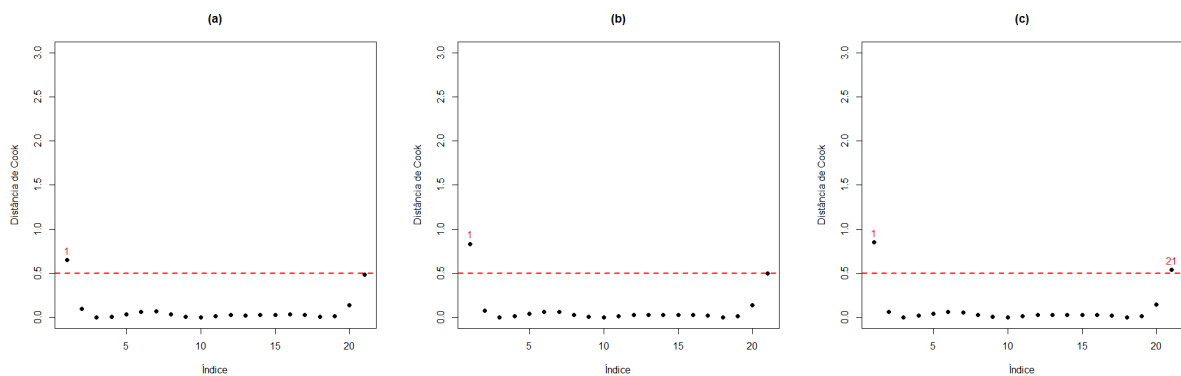


Figura 12 – Gráficos das distâncias de Cook para os ajustes dos modelos (a) Logístico, (b) Gompertz e (c) Von Bertalanffy para o peso das meninas.



Uma vez realizadas as análises, foram empregados critérios para verificar a qualidade do ajuste, além de comparar os modelos. A validação da qualidade de ajuste pode ser vista na Tabela 5, indicando que o modelo Logístico apresentou maior R^2 e menor erro quadrático médio. Os valores de AIC, BIC e QMRes apontam que o modelo é mais adequado, pois o mesmo apresentou valores menores em relação aos outros modelos ajustados aos dados.

Tabela 5 – Valores dos Coeficientes de determinação ajustado (R_a^2), Critério de Informação de Akaike (AIC), critério de informação Bayesiano (BIC) e quadrado médio dos resíduos (QMRes) para os modelos Logístico, Gompertz e Von Bertalanffy para as variáveis altura e peso.

Variável	Modelo	R_a^2	AIC	BIC	QMRes
Altura	Logístico	0,9882	80,6477	84,8257	2,1720
	Gompertz	0,9866	83,2251	87,4032	2,4557
	Von Bertalanffy	0,9860	84,0611	88,2392	2,5554
Peso	Logístico	0,9919	67,9237	72,1018	1,1850
	Gompertz	0,9881	75,9653	80,1433	1,7379
	Von Bertalanffy	0,9866	78,4445	82,6225	1,9557

Na Tabela 6 estão apresentados os valores das curvaturas quadrática média (RMS), propostas por Bates e Watts (1980) para analisar o quão próximo está um modelo não linear de um modelo linear, por meio de dois tipos de curvatura: a curvatura de efeito parâmetro (c^θ) e a curvatura intrínseca (c^ι). A curvatura intrínseca foi menor do que 0,5, para todos os modelos, indicando-se que o local da solução na vizinhança das estimativas é relativamente plano, tornando-se, assim, verdadeira a aproximação linear por um plano.

Tabela 6 – Medidas de curvatura quadrática média (RMS) para os modelos Logístico, Gompertz e Von Bertalanffy referente a altura e peso das meninas.

Variável	Modelo	c^θ	c^ι
Altura	Logístico	0,7984	0,0556
	Gompertz	0,9567	0,0547
	Von Bertalanffy	2,3140	0,0554
Peso	Logístico	0,6150	0,0452
	Gompertz	1,2319	0,0501
	Von Bertalanffy	2,1676	0,0521

Por outro lado, a curvatura devido ao efeito de parâmetros apresentou valores que excederam a cota 0,5, para os modelos Gompertz e Von Bertalanffy. Tendo em vista o processo inferencial, tais resultados indicam que nem sempre seria aceitável assumir uma boa aproximação da superfície esperada por um plano de coordenadas paralelas e uniformes no local da solução (RATKOWSKY, 1983).

Há também o caso em que mesmo com uma curvatura intrínseca considerada baixa, ocorre uma curvatura de efeito do parâmetros muito alta, acima de 1, indicando que ou a forma com que o parâmetro aparece no modelo são inadequadas ou o modelo pode ser inadequado para esse conjunto de dados, tendo em vista que as inferências obtidas por meio de aproximação linear e teoria assintótica podem ser invalidadas, pois as propriedades dos estimadores referentes a este modelo estão muito distantes em relação as propriedades dos estimadores lineares. O resultado não favorável da curvatura de efeito de parâmetro pode também ser justificado pelo comportamento do modelo em relação aos dados.

4.2 Análise através de simulação

Os estimadores dos parâmetros, em modelos de regressão linear, não possuem algumas das propriedades apresentadas pelos estimadores dos parâmetros em modelos de regressão não linear. Assim sendo, existe a necessidade de estudar o comportamento do modelo, segundo (RATKOWSKY et al., 1983), isso pode ser feito por meio de estudos de simulação.

Para obtenção das amostras simuladas foram consideradas como valores paramétricos as estimativas dos parâmetros apenas para o modelo logístico, por ter sido o que melhor se ajustou tanto para altura quanto para o peso das meninas. Foram gerados 1000 amostras a partir de cada um dos modelos, cada uma com 21 observações, e considerando $\epsilon_i \sim N(0, \sigma^2)$ em que σ^2 é a estimativa da variância residual obtida, também, dos ajustes dos modelos aos dados experimentais ($\sigma_{t1}^2 = 1,4745$ e $\sigma_{t2}^2 = 1,0893$)

Nas Tabelas 7 e 8 estão apresentadas as cinco primeiras amostras simuladas por meio do modelo logístico, para altura e peso, respectivamente.

Tabela 7 – Cinco primeiras amostras da altura das meninas simuladas para o modelo logístico dos 8 aos 18 anos

Idade (anos)	Amostras				
	1	2	3	4	5
8	127,6478	129,5379	127,6657	125,5196	128,8778
8,5	130,9496	131,8899	133,7436	132,8567	131,2413
9	138,187	134,0279	134,8784	134,212	136,1482
9,5	139,7228	140,3611	138,3758	138,6003	137,8499
10	144,6821	143,2175	141,4883	144,6936	145,3598
10,5	146,843	144,0784	143,3728	147,9169	147,7108
11	150,9712	146,1692	149,5669	148,8052	149,7184
11,5	152,1463	150,2387	153,874	150,1947	150,9634
12	153,6267	153,0918	154,5683	154,0749	152,3757
12,5	156,5558	156,2704	157,3966	155,3553	154,2708
13	157,9121	159,2776	157,6537	155,5206	156,0802
13,5	159,3193	159,4185	159,0376	159,59	159,4662
14	159,8995	160,5498	161,6717	160,0011	159,5192
14,5	160,5241	162,2335	162,1972	163,2034	162,9517
15	163,1427	162,8981	162,552	165,2101	163,8905
15,5	164,5215	164,3167	165,3085	165,3488	164,7671
16	166,6315	165,2108	166,0263	165,4599	165,1558
16,5	167,0001	165,344	167,5148	166,0077	166,0707
17	167,7593	165,9199	167,888	166,3641	167,3847
17,5	168,6805	167,9168	168,0754	166,3709	167,903
18	168,8043	170,638	168,2087	167,6075	167,9788

Tabela 8 – Cinco primeiras Amostras para o peso das meninas simuladas para o modelo logístico dos 8 aos 18 anos

Idade (anos)	Amostras				
	1	2	3	4	5
8	26,8274	28,2238	26,8407	25,2552	27,7361
8,5	28,5458	29,2405	30,6099	29,9547	28,7613
9	33,4618	30,3892	31,0175	30,5252	31,9556
9,5	34,4223	34,8939	33,4273	33,5931	33,0387
10	38,1298	37,0479	35,7704	38,1384	38,6305
10,5	39,946	37,9038	37,3825	40,7395	40,5873
11	43,3502	39,8026	42,3127	41,7499	42,4246
11,5	44,6672	43,2579	45,9435	43,2254	43,7933
12	46,2690	45,8738	46,9646	46,6001	45,3448
12,5	49,5107	48,7589	49,5909	48,2050	47,2817
13	49,9718	51,6256	50,3218	48,6238	49,1593
13,5	52,5754	52,0474	51,8701	52,5819	52,1867
14	52,9683	53,4845	54,3133	52,7754	52,7231
14,5	53,4636	55,1879	55,1611	56,3211	55,7184
15	56,2763	56,0956	55,8399	57,3869	56,8287
15,5	57,6667	57,5153	58,7783	58,3599	58,4623
16	59,8367	58,6013	59,1038	59,3550	58,7038
16,5	60,3534	58,7872	60,4807	59,5170	58,8111
17	60,3857	59,5554	60,4893	59,6442	60,3932
17,5	61,8944	61,2387	61,4543	59,8283	61,2845
18	61,9784	63,4245	61,5313	60,8022	61,4040

Na Tabela 9 encontram-se as estimativas dos parâmetros para cada uma das amostras bem como o valor paramétrico utilizado no estudo de simulação. São apresentados, média, variância e erro quadrático médio (EQM) dos parâmetros, para cada um dos modelos ajustados Tabela 10.

Tabela 9 – Estimativas dos parâmetros do modelo logístico para as cinco primeiras amostras simuladas do ajuste da altura e do peso.

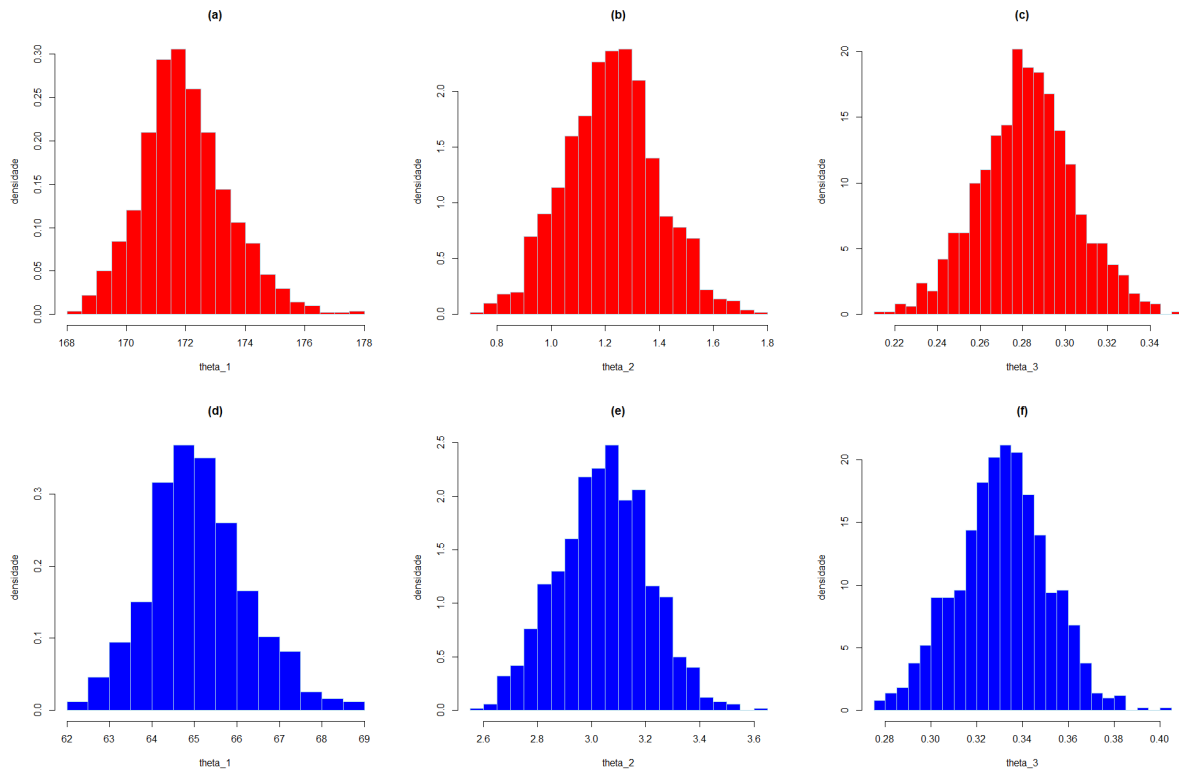
	Amostras	θ_1	θ_2	θ_3
Altura	1	172,1565	1,0986	0,2712
	2	174,8302	0,85	0,233
	3	172,9196	1,1035	0,2664
	4	170,5226	1,3108	0,2956
	5	172,7153	0,979	0,2549
Peso	1	66,0101	2,7994	0,3065
	2	67,1593	2,7441	0,2925
	3	65,5487	2,9880	0,3238
	4	64,0830	3,0988	0,3398
	5	66,0891	2,7711	0,3011
	6	64,6191	3,0539	0,3361

Tabela 10 – Média, variância e erro quadrático médio das estimativa dos parâmetros do ajuste dos modelos da altura e do peso para as 1000 amostras simuladas

	Parâmetros	Média	EQM	Variância
Altura	θ_1	171,9799	2,1910	2,1718
	θ_2	1,2228	0,0298	0,0299
	θ_3	0,2820	0,0005	0,0005
	Parâmetros	Média	EQM	Variância
Peso	θ_1	65,1049	1,3458	1,3393
	θ_2	3,0435	0,0280	0,0281
	θ_3	0,3312	0,0004	0,0004

Inspecionando as distribuições dos estimadores dos parâmetros, apenas o parâmetro θ_1 não converge para uma distribuição normal. Este fato pode ser confirmado pelos histogramas apresentados na Figura 13 e pelo teste de normalidade de Shapiro Wilk, em que, os p-valores obtidos em θ_1 para altura e peso foi $3,43e - 07$ e $1,304e - 06$ respectivamente, os outros parâmetros (θ_2 e θ_3) tiveram o nível descritivo maior que 0,05, não rejeitando a hipótese nula de que os estimadores seguem distribuição normal.

Figura 13 – Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo logístico, (a) θ_1 (b) θ_2 (c) θ_3 relacionado a altura e (d) θ_1 (e) θ_2 (f) θ_3 ao peso.



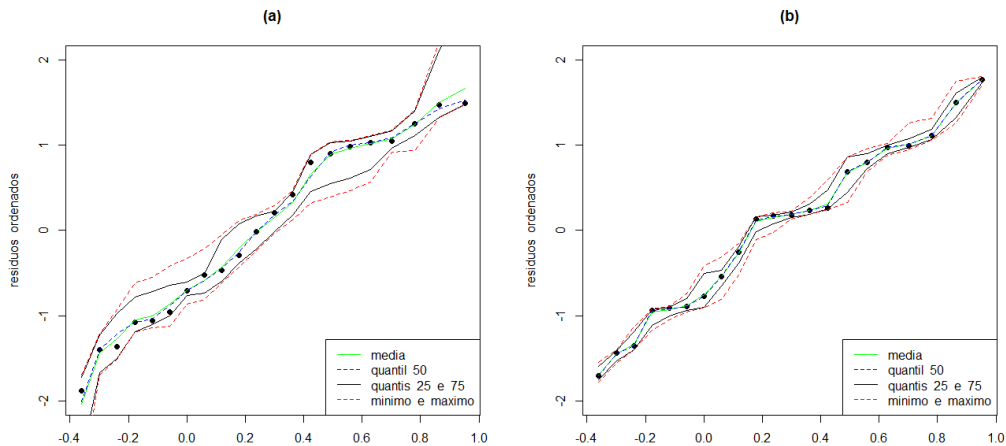
É importante lembrar que o valor do quadrado médio residual obtido para os dados da altura e do peso através do modelo logístico foram menores que o valor obtido pelos ajustes Gompertz e Von Bertalanffy, conforme foi apresentado na Tabela 5. Como a simulação foi realizada considerando esses valores para os dados originais, por consequência a amplitude dos resíduos encontrados permanecerá sendo melhor.

No gráfico de envelope simulado Figura 14 é possível perceber que o modelo logístico se ajustou melhor para variável peso, pois a amplitude dos resíduos encontrados para todas as curvas simuladas foi menor, sendo confirmado na Tabela 11, no qual expõe as medidas de curvatura mínima e máxima de todas as simulações.

Tabela 11 – Estatísticas para as curvaturas dos modelos logístico para altura e peso, considerando as 1000 amostras simuladas.

	Curvaturas	Mín	Máx
Altura	c^θ	0,1828	1,035
	c^t	0,0129	0,0465
	Curvaturas	Mín	Máx
Peso	c^θ	0,1496	0,7645
	c^t	0,1111	0,0393

Figura 14 – Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo Logístico aos dados da altura (a) e do peso (b), e envelope simulado.



5 Conclusão

A utilização de modelos não lineares foi adequada sendo capaz de apresentar parâmetros com interpretação biológica, comprovando sua importância no estudo do crescimento antropométrico feminino. Os modelos Logístico, Gompertz e o Von Bertalanffy apresentaram-se apropriados para descrever curvas de crescimento para os dados apresentados.

A análise dos resíduos projetados mostrou-se viável na aplicação em modelos de regressão não linear, por apresentar maior qualidade de diagnóstico do que outras técnicas. Também foi possível observar a existência de pontos discrepantes nos dados observados.

Terminadas as análises dos dados, através dos avaliadores da qualidade de ajuste, o modelo Logístico se mostrou o melhor para estes dados, sendo o mais indicado para representar curvas de crescimento médio da altura e do peso de meninas. Além disso foi possível observar que quanto maior a medida de curvatura melhor será a utilização do resíduo projetado, o que pôde ser comprovado por meio do estudo de simulação.

Referências

- AGGREY, S. Comparison of three nonlinear and spline regression models for describing chicken growth curves. *Poultry Science*, Oxford University Press Oxford, UK, v. 81, n. 12, p. 1782–1788, 2002. Citado na página 16.
- AKAIKE, H. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, Oxford University Press, v. 60, n. 2, p. 255–265, 1973. Citado na página 18.
- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–25, 1980. Citado 6 vezes nas páginas 12, 13, 14, 15, 17 e 22.
- BUCUR, D.; GAZZOLA, F. The first biharmonic steklov eigenvalue: positivity preserving and shape optimization. *Milan Journal of Mathematics*, Springer, v. 79, n. 1, p. 247, 2011. Citado na página 8.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 33, n. 2, p. 261–304, 2004. Citado na página 18.
- COELHO-BARROS, E. A.; SIMÕES, P. A.; ACHCAR, J. A.; MARTINEZ, E. Z.; SHIMANO, A. C. Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos. *Revista Colombiana de Estadística*, Universidad Nacional de Colombia, v. 31, n. 1, p. 111–129, 2008. Citado na página 10.
- COOK, R.; TSAI, C.-L. Residuals in nonlinear regression. *Biometrika*, Oxford University Press, v. 72, n. 1, p. 23–29, 1985. Citado na página 19.
- COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, Taylor & Francis Group, v. 19, n. 1, p. 15–18, 1977. Citado na página 17.
- CORNELL, J.; BERGER, R. Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. *Phytopathology*, American Phytopathological Society, v. 77, n. 1, p. 63–70, 1987. Citado na página 17.
- FRERY, A. C.; CRIBARI-NETO, F. *Elementos de estatística computacional usando plataformas de software livre/gratuito*. [S.l.]: IMPA, 2005. Citado na página 12.
- HARTMANN, S. A remark on the application of the newton-raphson method in non-linear finite element analysis. *Computational Mechanics*, Springer, v. 36, n. 2, p. 100–116, 2005. Citado na página 11.
- HUANG, H.-H.; HUANG, S.-Y. Nonlinear regression analysis. *International Encyclopedia of Education*, p. 339–346, 2010. Citado na página 10.
- KIMURA, D. K. Likelihood methods for the von bertalanffy growth curve. *Fishery bulletin*, v. 77, n. 4, p. 765–776, 1980. Citado na página 16.

- MARTINS, M. C. *Análise de resíduos projetados em modelos não-lineares de regressão: uma aplicação em crescimento de frutos*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 19.
- MATTOS, T. Modelos não lineares e suas aplicações. *Trabalho de conclusão de curso de bacharelado em Estatística. Universidade Federal de Juiz de Fora*, 2013. Citado na página 11.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. *Constituição da Organização Mundial da Saúde (OMS/WHO)*. 2019. Disponível em: <<https://www.who.int/eportuguese/countries/bra/pt/>>. Citado na página 8.
- PAINE, C. T.; MARTHEWS, T. R.; VOGT, D. R.; PURVES, D.; REES, M.; HECTOR, A.; TURNBULL, L. A. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods in Ecology and Evolution*, Wiley Online Library, v. 3, n. 2, p. 245–256, 2012. Citado 2 vezes nas páginas 9 e 16.
- POSADA, D.; BUCKLEY, T. R. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology*, Society of Systematic Zoology, v. 53, n. 5, p. 793–808, 2004. Citado na página 18.
- PRUDENTE, A. A. Modelos não-lineares de regressão: Alguns aspectos de teoria assintótica. *Universidade Federal Rural de Pernambuco. Recife*, 2009. Citado na página 13.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 20 e 22.
- RATKOWSKY, D.; LOWRY, R.; MCMEEKIN, T.; STOKES, A.; CHANDLER, R. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *Journal of bacteriology*, Am Soc Microbiol, v. 154, p. 1222–1226, 1983. Citado 5 vezes nas páginas 8, 12, 13, 14 e 30.
- RESENDE, M. D. V. de. Matemática e estatística na análise de experimentos e no melhoramento genético. *Embrapa Florestas, Colombo*, v. 362, 2007. Citado na página 18.
- SCHWARZ, G. et al. Estimating the dimension of a model. *The annals of statistics*, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citado na página 18.
- SEBER, G.; WILD, C. Measures of curvatures and nonlinearity. *Nonlinear Regression. Hoboken, NJ: John Wiley & Sons*, p. 127–190, 2003. Citado na página 10.
- TANNER, J. M. Growth and maturation during adolescence. *Nutrition reviews*, Wiley Online Library, v. 39, n. 2, p. 43–55, 1981. Citado na página 20.
- TUDDENHAM, R. D. Physical growth of california boys and girls from birth to eighteen years. *University of California publications in child development*, v. 1, p. 183–364, 1954. Citado na página 20.
- ZEVIANI, W. *Avaliação de modelos de regressão não linear na cinética de liberação de potássio de resíduos orgânicos. Lavras, Universidade Federal de Lavras, 2009. 85p*. Tese (Doutorado) — Dissertação de Mestrado, 2009. Citado na página 13.