



**UEPB**

**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I - CAMPINA GRANDE  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE GRADUAÇÃO EM BACHARELADO EM ESTATÍSTICA**

**ANDERSON SEVERINO DE SALES**

**AVALIAÇÃO DAS PROPRIEDADES DOS ESTIMADORES DOS MODELOS NÃO  
LINEARES POR MEIO DE SIMULAÇÃO**

**CAMPINA GRANDE - PB  
2023**

ANDERSON SEVERINO DE SALES

**AVALIAÇÃO DAS PROPRIEDADES DOS ESTIMADORES DOS MODELOS NÃO LINEARES POR MEIO DE SIMULAÇÃO**

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, como requisito parcial à obtenção do título de bacharel em Estatística.

**Orientador:** Profa. Dra. Ana Patricia Bastos Peixoto  
**Coorientador:** Profa. Débora de Sousa Cordeiro

**CAMPINA GRANDE - PB  
2023**

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.

- 519.5 Sales, Anderson Severino de.  
Avaliação das propriedades dos estimadores dos modelos não lineares por meio de simulação [manuscrito] / Anderson Severino de Sales. - 2023.  
29 p. : il. colorido.
- Digitado.  
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.  
"Orientação : Profa. Dra. Ana Patricia Bastos Peixoto, Coordenação do Curso de Estatística - CCT. "  
"Coorientação: Profa. Esp. Débora de Sousa Cordeiro , Departamento de Estatística - CCT."
1. Modelo não linear. 2. Curva de crescimento. 3. Bootstrap. 4. Confiabilidade. I. Título
21. ed. CDD S163a

**ANDERSON SEVERINO SALES**

**AVALIAÇÃO DAS PROPRIEDADES DOS ESTIMADORES DOS MODELOS NÃO LINEARES POR MEIO DE SIMULAÇÃO.**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciência e Tecnologia da Universidade Estadual da Paraíba como requisito parcial à obtenção do título de Bacharel em Estatística.

**Área de concentração:** Estatística.

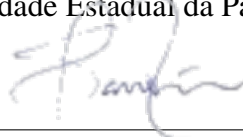
Aprovada em: 07/12/2023.

**BANCA EXAMINADORA**



---

Profa. Dra. Ana Patrícia Bastos Peixoto (Orientador)  
Universidade Estadual da Paraíba (UEPB)



---

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



---

Prof. Pedro Augusto Ferreira da Silva  
Universidade Estadual da Paraíba (UEPB)

À memória do meu querido pai, Severino, que me ensinou valores. A minha mãe, Severina, que sempre me ensinou a ser forte e corajoso.  
DEDICO

## **AGRADECIMENTOS**

Agradeço primeiramente à Deus, força minha, que nos momentos mais difíceis sempre esteve ao meu lado.

Agradeço aos meus pais, por todo amor, parceria e compreensão. Por ter me ensinado valores que dinheiro nenhum compra.

Agradeço à minha noiva, por todo suporte, respeito e "paciência".

Agradeço aos meus irmãos Carlos, Sonia, Eliane e Giane e aos meus amados sobrinhos e cunhados, por todo respeito, admiração e cada palavra de incentivo.

Agradeço à minha orientadora Ana Patrícia Bastos Peixoto e a minha coorientadora Débora Cordeiro por toda a atenção, paciência e apoio na realização desse trabalho.

Agradeço aos meus professores Vitória, Ricardo, Cleanderson e Tiago por todos seus ensinamentos e exemplo de profissionalismo.

Por fim, agradeço aos meus queridos colegas e amigos da graduação, em especial Welinton, Gilber, João, Rayza, Rosângela, Andreza e Fernanda por todo companherismo e amizade.  
"VAI DAR CERTO!"

## RESUMO

A estimação de parâmetros é um processo fundamental em estatística, pois permite a construção de modelos matemáticos que podem ser usados para fazer previsões ou inferências sobre os dados. Os modelos lineares são os mais comuns, mas podem não ser adequados quando os dados não seguem seus pressupostos. Nesses casos, os modelos não lineares podem ser mais adequados, pois são mais flexíveis e podem acomodar melhor a variação dos dados. Neste trabalho, para avaliação dos estimadores, foi avaliado o peso do bulbo da cebola ao longo dos meses. Os dados foram usados para estimar os parâmetros de três modelos de curva de crescimento: Gompertz, Logístico e Von Bertalanffy. A simulação *bootstrap* foi usada para avaliar a confiabilidade dos estimadores dos parâmetros. Os resultados mostraram que todos os três modelos foram capazes de ajustar os dados de forma satisfatória. No entanto, houve algumas diferenças entre os modelos. O modelo Logístico foi o que apresentou o melhor ajuste, seguido do modelo Von Bertalanffy e do modelo de Gompertz. Os resultados deste trabalho indicam que a simulação *bootstrap* é uma ferramenta eficaz para avaliar as propriedades dos estimadores de modelos não lineares. Este método pode ser usado para verificar a confiabilidade dos estimadores e dos modelos, bem como comparar diferentes modelos.

**Palavras-chaves:** modelo não linear; curva de crescimento; *Bootstrap*; confiabilidade.

## ABSTRACT

Parameter estimation is a fundamental process in statistics, as it allows the construction of mathematical models that can be used to make predictions or inferences about data. Linear models are the most common, but they may not be appropriate when the data do not follow their assumptions. In these cases, nonlinear models may be more appropriate, as they are more flexible and can better accommodate the variation of the data. In this study, to evaluate the estimators, the weight of the onion bulb was evaluated over the months. The data were used to estimate the parameters of three growth curve models: Gompertz, Logistic, and Von Bertalanffy. Bootstrap simulation was used to evaluate the reliability of the parameter estimators. The results showed that all three models were able to fit the data satisfactorily. However, there were some differences between the models. The Logistic model showed the best fit, followed by the Von Bertalanffy model and the Gompertz model. The results of this study indicate that bootstrap simulation is an effective tool for evaluating the properties of estimators for nonlinear models. This method can be used to verify the reliability of the estimators and the models, as well as to compare different models

**Keywords:** nonlinear model; growth curve; *Bootstrap*; reliability.



## SUMÁRIO

1	INTRODUÇÃO . . . . .	8
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	9
2.1	Modelos de regressão não lineares . . . . .	9
2.2	Propriedades dos estimadores . . . . .	10
2.3	Métodos dos Mínimos Quadrados . . . . .	11
2.4	Método dos Mínimos Quadrados Ordinários . . . . .	11
2.5	Estimadores de Máxima Verossimilhança . . . . .	12
2.5.1	<i>Função de Verossimilhança</i> . . . . .	12
2.5.2	<i>Estimador de Máxima Verossimilhança</i> . . . . .	13
2.6	Métodos iterativos . . . . .	13
2.6.1	<i>Método de Newton-Raphson</i> . . . . .	13
2.6.2	<i>Método de Gauss-Newton</i> . . . . .	14
2.7	Análise de Tendências no Vício . . . . .	15
2.7.1	<i>Vício de box</i> . . . . .	15
2.8	Modelos de crescimento . . . . .	15
2.8.1	<i>Modelo de regressão logístico</i> . . . . .	16
2.8.2	<i>Modelo Von Bertalanffy</i> . . . . .	16
2.8.3	<i>Modelo Gompertz</i> . . . . .	17
2.9	Método de <i>bootstrap</i> . . . . .	17
2.10	<i>Bootstrap</i> paramétrico . . . . .	18
2.11	Intervalos de confiança <i>bootstrap</i> . . . . .	18
2.12	Teste de normalidade <i>bootstrap</i> . . . . .	18
3	MATERIAL E MÉTODOS . . . . .	20
4	RESULTADOS E DISCUSSÃO . . . . .	22
4.1	Simulação <i>bootstrap</i> . . . . .	24
5	CONCLUSÃO . . . . .	27
	REFERÊNCIAS . . . . .	28

## 1 INTRODUÇÃO

A Estatística é fundamental para sociedade, sendo uma ferramenta fundamental para analisar e compreender os dados que permeiam diversas áreas da vida cotidiana. Em um mundo cada vez mais orientado por dados, a Estatística fornece os meios para coletar, organizar e interpretar informações complexas. Ela oferece recursos e métodos que permitem extrair significado dos números, transformando-os em conhecimento. Além disso, a Estatística é fundamental na pesquisa científica, contribuindo para a validade e confiabilidade dos resultados. Em resumo, esta ciência é um componente essencial na compreensão do mundo ao nosso redor. Seja na formulação de políticas, na tomada de decisões empresariais, na pesquisa científica ou na interpretação de eventos cotidianos, a estatística oferece as ferramentas necessárias para transformar dados aparentemente complexos em conhecimento significativo e aplicável Morettin e Bussab (2017).

Os modelos de regressão desempenham são essenciais na estatística e são amplamente utilizados para entender e modelar as relações entre características de interesse que chamamos variáveis. A regressão é uma técnica estatística que busca quantificar a relação entre uma variável dependente e uma ou mais variáveis independentes. Essa relação é expressa por meio de uma equação matemática que descreve como as variáveis independentes influenciam a variável dependente. Existem diferentes tipos de modelos de regressão, cada um adequado para diferentes situações e tipos de dados, dentre eles, tem-se aqueles que são chamados não lineares.

Os modelos de regressão não lineares modelam relações mais complexas que não podem ser representadas por uma equação linear. Pode envolver funções exponenciais, logarítmicas, sigmóides, entre outras Zeviani P. J. W.; Bonat (2013). A escolha desse modelo é geralmente feita com base em algum conhecimento prévio sobre a relação entre as variáveis, como afirma Zeviani P. J. W.; Bonat (2013). Esse conhecimento pode ser obtido a partir de uma análise de um diagrama de dispersão, de restrições de forma da função, da solução de uma equação diferencial ou da interpretação dos parâmetros do modelo. Se os pontos do diagrama se distribuem de forma linear, um modelo linear pode ser suficiente. No entanto, se os pontos se distribuem de forma curva, por exemplo, um modelo não linear pode ser necessário. As restrições de forma da função, por sua vez, podem ser impostas ao modelo para garantir que a relação entre as variáveis siga um determinado padrão.

A determinação precisa dos parâmetros é um elemento crucial no processo de encontrar os valores que melhor se ajustam aos dados observados. Entre as abordagens estatísticas, destacam-se as categorias dos modelos lineares e não lineares. Nos modelos lineares, a estimação dos parâmetros envolve a resolução de um sistema de equações lineares em relação aos coeficientes de regressão desconhecidos. Nesse contexto, uma única solução é obtida, proporcionando uma forma analítica para a estimação dos parâmetros. No entanto, em muitas situações, descrever um fenômeno por meio de uma regressão linear não é desejável, nem mesmo possível (ZEVIANI P. J. W.; BONAT, 2013).

Geralmente, quando os estimadores de mínimos quadrados dos parâmetros de um modelo não apresentam as características desejadas, como viés excessivo e alta variância, costuma-se concluir que o modelo exibe um comportamento não linear, acima do limite mínimo de variação e tem distribuição marcadamente distorcida (RENAM, 2019). Os estimadores para o modelo não linear são obtidos e avaliados de diversas formas, dentre elas tem-se, máxima verossimilhança, mínimos quadrados, mínimos quadrados ordinários, entre outros. Para se obter com precisão os estimadores, usa-se os métodos iterativos para otimização, dos quais alguns serão abordados neste trabalho, como Gaws Newton (MAZUCHELI; ACHCAR, 2002) e Newton Raphson (AMARAL; SOUZA; CATALAN, 2015). Nesse sentido, esse trabalho teve por objetivo avaliar as propriedades dos estimadores dos modelos não lineares por meio de simulação *bootstrap* para avaliar o comportamento da curva de crescimento e a confiabilidade destes estimadores.

## 2 FUNDAMENTAÇÃO TEÓRICA

Aqui serão expostos os principais resultados importantes para a construção desse estudo. Destacando os principais métodos de grande impacto e relevância inerentes aos resultados.

### 2.1 Modelos de regressão não lineares

Os modelos de regressão constituem uma ferramenta estatística essencial para compreender e quantificar as relações entre variáveis. Esses modelos buscam representar matematicamente a relação entre uma variável dependente e uma ou mais variáveis independentes. De forma geral, segundo Mazucheli e Achcar (2002), os modelos não lineares são importantes porque permitem que os pesquisadores modelem fenômenos do mundo real que não podem ser adequadamente descritos por modelos lineares. O modelo de regressão não linear pode ser descrito como:

$$y_i = f(x_i, \theta) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

em que,  $y_i$  é a variável resposta,  $f(\cdot)$  é uma função não linear contínua, com forma conhecida, do vetor de variáveis explicativas  $X_i$  e dos parâmetros desconhecidos  $\theta$ . Os erros aleatórios neste trabalho,  $\varepsilon_i$ , são considerados independentes e identicamente distribuídos, normais com média zero e variância constante  $\sigma^2$ , isto é,  $\varepsilon_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma_i^2)$ .

A função esperança ou função de regressão é dada por

$$E(Y_i|X_i) = f(X_i, \theta) \quad (2)$$

Para Bates e Watts (1988), nos modelos de regressão não linear, pelo menos uma das derivadas parciais da função  $f(x)$  em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo. Segundo Zeviani P. J. W.; Bonat (2013), para entender o significado de um parâmetro, é preciso saber sua dimensão e como a função se comporta em termos dele. A dimensionalidade pode ser determinada a partir das unidades de medida dos termos conhecidos do modelo. O estudo da função em termos do parâmetro inclui o estudo de seus limites, derivadas, integrais, inversas e avaliações em certos pontos.

Ainda de acordo com Zeviani P. J. W.; Bonat (2013), uma vantagem fundamental do modelo não linear em comparação com o modelo linear, reside no fato de que sua seleção está relacionada ao entendimento prévio da relação que está sendo modelada e, geralmente, permite interpretações práticas dos parâmetros. Isso implica que é possível ajustar modelos que não poderiam ser adequadamente ajustados por meio de modelos lineares tradicionais. Essa vantagem se traduz em uma maior flexibilidade na modelagem de relacionamentos complexos entre variáveis, permitindo que os parâmetros do modelo tenham significados mais intuitivos, o que pode ser valioso em diversas aplicações.

Em seu estudo, Mattos (2013) destaca que os modelos não lineares (MNL) oferecem vantagens substanciais em relação aos modelos lineares (ML). Essas vantagens incluem a capacidade de embasar sua escolha em teoria, princípios mecânicos ou informações prévias, como dados físicos, químicos ou biológicos. Além disso, certos parâmetros dos MNL possuem interpretações práticas e são de grande interesse para os pesquisadores. Os MNL também permitem a realização de previsões fora do domínio observado das variáveis independentes, o que os torna uma ferramenta valiosa em diversas aplicações. Esses modelos tendem a ser mais parcimoniosos, frequentemente exigindo a estimativa de menos parâmetros em comparação com os modelos lineares.

Por outro lado, Zeviani P. J. W.; Bonat (2013) ressaltam que os MNL têm desvantagens, como a necessidade de procedimentos iterativos para a estimação, que dependem da inicialização

dos parâmetros e o uso de métodos de inferência aproximados. Portanto, é fundamental ponderar as vantagens e desvantagens dos MNL ao aplicá-los em pesquisa e modelagem. A escolha do modelo não linear adequado é crucial para obter resultados precisos. Isso ocorre porque diferentes modelos não lineares podem representar a mesma relação de maneira diferente. A escolha do modelo errado pode levar a estimativas enviesadas e a previsões incorretas.

A obtenção das estimativas dos parâmetros em modelos de regressão não lineares pode ser abordada de diversas maneiras, sendo o método dos mínimos quadrados e o método da máxima verossimilhança os mais comuns. No entanto, nos casos de modelos não lineares, a resolução direta das equações normais se torna uma tarefa desafiadora, tornando necessária a aplicação de métodos iterativos para obter essas estimativas (MAZUCHELI; ACHCAR, 2002). A seguir, serão apresentados alguns dos métodos amplamente empregados para realizar essas estimativas.

## 2.2 Propriedades dos estimadores

Em estatística, uma propriedade de um estimador é uma característica que o estimador possui ou não. Essas propriedades são importantes porque ajudam a avaliar a qualidade do mesmo. Dentre essas propriedades, podemos citar como mais importantes:

- i) Não viesado: Um estimador é não tendencioso quando seu valor esperado é igual ao valor real do parâmetro que está sendo estimado;
- ii) Eficiência: Ocorre quando um estimador tem a menor variância possível entre todos os estimadores não tendenciosos do mesmo parâmetro;
- iii) Consistência: Se dar quando o valor da estimação converge para o valor real do parâmetro à medida que o tamanho da amostra aumenta.
- iv) Linearidade: Um estimador é linear quando ele pode ser expresso como uma combinação linear das observações da amostra;
- v) Independência: Há independência quando seus valores não são correlacionados entre si;
- vi) Normalidade: Um estimador é normal quando sua distribuição é próxima da distribuição normal.

A escolha do estimador apropriado para um determinado problema depende de uma série de fatores, incluindo a natureza do parâmetro que está sendo estimado, a quantidade de dados disponíveis e as propriedades desejadas do estimador (BATES; WATTS, 1988). Como exemplo, podem ser citados, de acordo com algumas das propriedades, os seguintes estimadores

1. A média amostral que representa um estimador não tendencioso da média populacional;
2. A variância amostral, por sua vez, é um estimador não tendencioso da variância populacional;
3. O intervalo de confiança de um estimador é um estimador consistente do parâmetro que está sendo estimado.

Nesse sentido, para ajustar um modelo não linear aos dados observados, é necessário estimar os parâmetros desconhecidos que definem a forma da função não linear. A estimação de parâmetros é geralmente realizada por meio de métodos iterativos, como o Método dos Mínimos

Quadrados Não Lineares (NLS), ou algoritmos de otimização. A obtenção das estimativas dos parâmetros pode ser realizada por meio de diversas abordagens, entre as quais se destacam o método dos mínimos quadrados e o método da máxima verossimilhança (MATTOS, 2013). No contexto dos modelos de regressão não lineares, a resolução do sistema de equações normais não se revela uma tarefa trivial, tornando necessária a aplicação de métodos iterativos para a obtenção dessas estimativas. Neste ponto, são apresentadas algumas técnicas de estimação que se mostram relevantes e frequentemente utilizadas.

### 2.3 Métodos dos Mínimos Quadrados

No método de mínimos quadrados (MMQ), não é necessário conhecer a forma da distribuição dos erros. O objetivo é estimar os parâmetros  $\beta_0$  e  $\beta_1$ , de modo que os desvios ( $\varepsilon_i$ ) entre os valores observados e estimados sejam mínimos. A estimativa de mínimos quadrados de  $\theta$ , denotado por  $b(\theta)$ , minimiza a soma de quadrados dos erros

$$SQRes(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (3)$$

em que,  $\theta \in \Theta$ . Dessa forma, para encontrar o estimador de mínimos quadrados, o processo envolve a diferenciação das equações em relação a cada um dos parâmetros e, em seguida, igualar as equações a zero, de forma que

$$\frac{\partial SQRes(\theta)}{\partial \theta_r} = 0 \quad (r = 1, 2, \dots, p)$$

$$\sum_{i=1}^n [y_i - f(x_i, \theta)] \frac{\partial f(x_i, \theta)}{\partial \theta_r} = 0 \quad (r = 1, 2, \dots, p) \quad (4)$$

Ao seguir esse procedimento, obtém-se um sistema de  $p$  equações normais para o modelo não linear. No entanto, para a maioria dos modelos não lineares, esse sistema não pode ser resolvido analiticamente, ou seja, não há uma fórmula geral que possa ser usada para encontrar as soluções. Portanto, é necessário usar métodos iterativos para encontrar soluções aproximadas para o sistema (MATTOS, 2013).

### 2.4 Método dos Mínimos Quadrados Ordinários

De acordo com a teoria de estimação usando o método de mínimos quadrados ordinários (MQO), as estimativas para o parâmetro  $\theta$  em um modelo não linear na forma (18) são aquelas que buscam minimizar a soma dos quadrados dos erros (THOMAS, 2016). Essa soma é dada por:

$$S(\theta) = \sum_{i=1}^n [Y_i - f(X_i - \theta)]^2. \quad (5)$$

Assim, para determinar os estimadores que conduzem a essas estimativas, é necessário aplicar o cálculo das derivadas parciais da equação em relação a cada componente do vetor  $\theta$  e igualar essas derivadas a zero. Isso resultará no conjunto de equações normais que seguem:

$$\sum_{i=1}^n [Y_i - f(X_i - \theta)] \frac{\delta f(X_i; \theta)}{\delta \theta_j} = 0. \quad (6)$$

Para modelos lineares, é comum que o sistema de equações tenha uma solução única para os parâmetros  $\theta$ , no qual cada parâmetro pode ser expresso de forma direta em termos

dos dados observados. Por outro lado, em modelos não lineares, os estimadores  $\theta$  geralmente não podem ser obtidos de maneira analítica, o que significa que não existe uma fórmula direta para essas estimativas. Nesse contexto, recorreremos a métodos iterativos de aproximação para encontrar as soluções. Esses métodos envolvem um processo de iteração, no qual os valores dos parâmetros são gradualmente ajustados até que uma solução que satisfaça os critérios desejados seja alcançada.

Dentre as metodologias de otimização utilizadas na estimação de parâmetros, destacam-se diversas abordagens importantes, tais como o método de Gauss-Newton (AMARAL; SOUZA; CATALAN, 2015), o método de Davidon-Fletcher-Powell (HOFFMANN, 2016), a técnica da descida mais íngreme (steepest descent) (SANTOS, 2011), o método da busca com gradiente conjugado (ZEVIANI P. J. W.; BONAT, 2013), a abordagem de busca de Nelder-Mead (HOFFMANN, 2016), o algoritmo de Golub-Pereyra (ZEVIANI P. J. W.; BONAT, 2013), entre outras. Um estimador bastante utilizado nesse cenário dos modelos não lineares é o estimador de Máxima Verossimilhança.

## 2.5 Estimadores de Máxima Verossimilhança

O Método da Máxima Verossimilhança é uma abordagem essencial para a obtenção da estimativa mais provável do parâmetro populacional desconhecido com base em uma amostra. Essencialmente, esse método busca encontrar o valor do estimador que torna a ocorrência da amostra em questão mais provável. Para fazer isso, a função de probabilidade das variáveis aleatórias que representam a amostra é tratada como uma função do parâmetro populacional correspondente, permitindo assim a estimativa desse parâmetro desconhecido. A seção subsequente enfatizará a importância dessa função, chamada de função de verossimilhança, no processo de estimativa.

### 2.5.1 Função de Verossimilhança

Quando temos uma amostra aleatória simples de uma variável  $X$  com uma função de densidade  $f(x; \theta)$ , em que  $\theta$  é um parâmetro desconhecido, introduzimos a função  $L(\theta; x)$ . Essa função avalia a probabilidade da amostra com base em diferentes valores possíveis de  $\theta$ , ajudando-nos a encontrar o valor de  $\theta$  que melhor se encaixa nos dados. Basicamente, buscamos o  $\theta$  que maximiza a verossimilhança de nossos dados amostrais. A função é expressa por:

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta, \quad (7)$$

denotada como Função de Verossimilhança. Essa função depende do parâmetro populacional e os valores da amostra estão fixos.  $\phi$  é o espaço paramétrico de  $\theta$ , o conjunto de valores que o parâmetro pode assumir. O logaritmo natural da função de verossimilhança  $l(\theta; x)$ , chamado de Função de Log-Verossimilhança, é frequentemente preferido por sua conveniência em cálculos, ele é definido por:

$$l(\theta; x) = \log L(\theta; x) = \log \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \log f(x_i; \theta). \quad (8)$$

O logaritmo é uma função que preserva a ordem dos valores, tornando mais fácil lidar com otimização. Portanto, neste contexto, optamos por trabalhar com a função de log-verossimilhança em vez da verossimilhança diretamente. Isso ocorre porque a função de log-verossimilhança simplifica o processo de otimização e facilita os cálculos, mantendo a mesma relação de ordem entre os valores.

### 2.5.2 Estimador de Máxima Verossimilhança

O estimador de máxima verossimilhança de um parâmetro  $\theta$  é o valor que torna  $\theta$  mais provável com base na amostra observada. Portanto, é possível definir o estimador de máxima verossimilhança de  $\theta$  (E.M.V.) como qualquer valor  $\hat{\theta}$ , tal que  $L(\theta; x)$  é máxima. Com isso,

$$\hat{L}(\theta; x) \geq L(\theta; x) \quad \forall x \in \mathbb{R}, \forall \theta \in \Theta.$$

O valor do parâmetro que torna a função de verossimilhança máxima é o mesmo que faz a função de log-verossimilhança atingir o seu ponto máximo. Isso ocorre porque o logaritmo é uma função crescente, então os extremos da função de log-verossimilhança correspondem aos extremos da função de verossimilhança. Para encontrar esse valor, procuramos o ponto onde a primeira derivada se anula e verificamos se a matriz Hessiana é negativa nesse ponto. Normalmente, a função de log-verossimilhança é uma função côncava. Portanto, o ponto em que a primeira derivada se iguala a zero é o ponto onde ocorre o máximo da função.

Para melhor precisão do estimador, se usa os métodos iterativos. Esses métodos envolvem a iteração sequencial por meio de cálculos repetidos para se aproximar da solução desejada. A cada passo, uma estimativa melhor da solução é obtida com base na iteração anterior e esse processo continua até que uma condição de convergência seja satisfeita.

## 2.6 Métodos iterativos

Para otimizar a minimização da  $SQRes(\theta)$  nas equações ou maximizar  $L(\theta, x)$ , é essencial empregar métodos numéricos iterativos. Na literatura, existem várias opções de métodos iterativos propostos, como o método Newton-Raphson, método Gauss-Newton, método *Steepest-Descent*, método de Marquardt, entre outros. O método de Gauss-Newton é o algoritmo mais amplamente utilizado para estimar regressões não lineares e sua abordagem fundamental envolve a criação de aproximações lineares para a função de expectativa  $f(x_i, \theta)$  em cada iteração (MATTOS, 2013).

### 2.6.1 Método de Newton-Raphson

Isaac Newton e Joseph Raphson, dois matemáticos do século XVII, desenvolveram independentemente o método de Newton-Raphson, um método poderoso e eficiente para encontrar raízes de equações não lineares. O método é baseado na ideia de aproximar a raiz da equação por meio de uma reta tangente ao gráfico da função em um ponto próximo à raiz. O método de Newton-Raphson é um método iterativo para encontrar raízes de equações não lineares. O método funciona escolhendo um valor inicial  $e$ , a partir dele, calculando uma nova aproximação usando uma fórmula que leva em consideração a derivada da função. O processo é repetido até que a aproximação seja suficientemente precisa (AMARAL; SOUZA; CATALAN, 2015).

Próximo a  $\theta^{(a)}$ , aproximamos

$$SQRes(\theta) \approx SQRes(\theta^{(a)}) + U_0(\theta^{(a)})(\theta - \theta^{(a)}) + (\theta - \theta^{(a)})' \frac{H(\theta^{(a)})}{2} (\theta - \theta^{(a)}). \quad (9)$$

A aproximação da função de soma de quadrados terá um ponto estacionário quando seu gradiente for igual a zero. Isso significa que:

$$U(\theta) = \frac{\partial SQRes(\theta)}{\partial \theta} = 0.$$

Além disso, a matriz Hessiana da  $SQRes(\theta)$ , que é denotada como  $H(\theta)$ , deve satisfazer a seguinte condição:

$$H(\theta) = \frac{\partial^2 SQRes(\theta)}{\partial \theta \partial \theta^T}, \text{ é positiva (semi)definida.}$$

Essas condições indicam que, para encontrar um ponto estacionário da função da soma de quadrados, é necessidade de procurar um valor de  $\theta$ , em que o gradiente seja zero ( $U(\theta) = 0$ ) e a matriz Hessiana seja positiva (semi) definida.

Se  $H(\theta^{(a)})$  é positiva definida, o passo de Newton-Raphson é

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

O Processo é obtido até obter a convergência. Dentre os muitos existentes, um critério para verificar a convergência poderia ser:

$$\left| SQRes(\theta^{(a+1)}) - SQRes(\theta^{(a)}) \right| < 0.000001.$$

Outro método iterativo a ser abordado, é o método iterativo de Gauss-Newton. O método de Gauss-Newton é uma versão simplificada do método de Newton-Raphson. Embora exija um número maior de iterações do que o Newton-Raphson, as operações realizadas em cada iteração são mais simples (THOMAS, 2016).

### 2.6.2 Método de Gauss-Newton

Embora seja numericamente estável, este método pode experimentar uma taxa de convergência bastante lenta quando uma alta precisão é necessária, ou pode convergir para um máximo local em vez do máximo global, ou até mesmo não convergir. Isso é especialmente observado quando se lida com um tamanho amostral pequeno e/ou estimativas iniciais de má qualidade (BATES; WATTS, 1988).

Suponha que  $\theta^{(a)}$  seja uma estimativa de mínimos quadrados  $b\theta$  de um modelo não linear. Quando  $\theta$  está próxima de  $\theta^{(a)}$ , é comum considerar uma expansão em série de Taylor de primeira ordem expresso, a seguir:

$$f(x, \theta) \approx f(x, \theta^{(a)}) + F(\theta^{(a)})(\theta - \theta^{(a)}), \quad (10)$$

em que,  $F(\theta) = \frac{\partial f(x, \theta)}{\partial \theta_j} = \left\langle \frac{\partial f(x_i, \theta)}{\partial \theta_j} \right\rangle_i$ . Definindo  $r(\theta)$  como um vetor de resíduos, segue que:

$$r(\theta) = y - f(x, \theta) \approx r(\theta^{(a)}) - F^{(a)}(\theta - \theta^{(a)}),$$

em que  $F^{(a)} = F(\theta^{(a)})$ . Substituindo  $r'(\theta)r(\theta)$  em (9) obtém-se,

$$SQRes(\theta) \approx r'(\theta^{(a)})r(\theta^{(a)}) - 2r'(\theta^{(a)})F^{(a)}(\theta - \theta^{(a)}) + (\theta - \theta^{(a)})^T F^{(a)}(\theta - \theta^{(a)}). \quad (11)$$

Assim, a minimização de  $SQRes(\theta)$  ocorrerá quando

$$\theta - \theta^{(a)} = [F^{(a)T} F^{(a)}]^{-1} F^{(a)T} r(\theta^{(a)}). \quad (12)$$

Portanto, devido à estimativa inicial  $\theta^{(a)}$ , a próxima iteração é obtida da seguinte maneira:

$$\theta^{(a+1)} = \theta^{(a)} + [F^{(a)T} F^{(a)}]^{-1} F^{(a)T} r(\theta^{(a)}), \quad (13)$$

resultando no processo iterativo. O processo é repetido até obter a convergência (MATTOS, 2013).



É importante observar que, mesmo quando critérios de término são definidos para o processo iterativo, a convergência não é assegurada. Se os valores iniciais forem escolhidos de maneira inadequada, pode ser necessário um grande número de iterações para que a convergência ocorra. Além disso, como já citado, há a possibilidade de convergir para um mínimo local em vez do mínimo global desejado, ou ainda, pode não ocorrer convergência (MATTOS, 2013).

No entanto, quando os valores iniciais dos parâmetros são escolhidos de forma apropriada e estão suficientemente próximos dos valores verdadeiros dos parâmetros, o método iterativo tende a convergir rapidamente, mesmo em situações em que existam múltiplos mínimos locais no espaço de parâmetros. Assim, a seleção cuidadosa dos valores iniciais desempenha um papel fundamental na estimativa de parâmetros em contextos não lineares. Ela pode influenciar significativamente a eficiência do processo iterativo e a probabilidade de alcançar um mínimo global desejado.

## 2.7 Análise de Tendências no Vício

A análise de tendências no vício é um procedimento estatístico que visa compreender como variáveis independentes influenciam desvios em relação a padrões desejados, especialmente no contexto de vícios em modelos estatísticos. Essa análise identifica padrões não lineares nos dados, avalia relações complexas entre variáveis e detecta desvios sistemáticos, sendo fundamental na validação de modelos (MAZUCHELI; ACHCAR, 2002). Essa abordagem proporciona uma compreensão mais aprofundada das nuances dos dados, contribuindo para uma modelagem mais precisa e informada.

### 2.7.1 Vício de box

O termo "vício de Box" refere-se a uma técnica estatística chamada "diagnóstico de Box", que é utilizada para identificar quais parâmetros têm uma contribuição mais significativa no desvio da linearidade em um modelo estatístico. Em outras palavras, a análise de Box destaca quais variáveis ou fatores têm um impacto mais pronunciado na não conformidade do modelo com uma relação linear. Para Mazucheli e Achcar (2002) a estatística utilizada para examinar os desvios nos estimadores de mínimos quadrados dos parâmetros será caracterizada pela expressão:

$$\text{vício } \hat{\theta} = -\frac{\sigma^2}{2} \left[ \sum_{i=1}^n F(\theta) F_t(\theta)^{-1} \right] - \sum_{i=1}^n F(\theta). \quad (14)$$

$$\text{traço} = \left( \sum_{i=1}^n F(\theta) F_t(\theta)^{-1} H(\theta) \right),$$

em que  $F(\theta)$  representa o vetor  $(p \times 1)$  de primeiras derivadas da  $f(x_i; \theta)$ , denominado de vetor velocidade e  $H(\theta)$  é uma matriz  $(p \times p)$  das segundas derivadas em relação a cada elemento de  $\theta$ . Para resolver a equação (14), inicialmente assume-se que  $\hat{\theta}$  e  $\hat{\sigma}^2$  são os valores verdadeiros de  $\theta$  e  $\sigma^2$ , respectivamente. Dessa forma, o vetor de dimensões  $p \times 1$  representa a discrepância entre as estimativas e os valores reais dos parâmetros.

## 2.8 Modelos de crescimento

Existem diversos modelos não lineares e, entre eles, os de curvas de crescimento são notáveis. Destacam-se os modelos Gompertz, Von Bertalanffy e Logístico, que compartilham

diversos parâmetros em comum. Esses parâmetros permitem atribuir significado biológico específico a cada um, ampliando a compreensão desses modelos (MATTOS, 2013).

Conforme mencionado por Santos (2011), as curvas de crescimento são frequentemente associadas a padrões sigmoidais, representando o tempo de vida de medições dimensionais, como altura e peso. A modelagem desses dados de crescimento busca extrair informações físicas dos parâmetros, com o objetivo de estabelecer um modelo padrão que represente as observações em estudo.

### 2.8.1 *Modelo de regressão logístico*

O modelo logístico original foi desenvolvido em 1838 por Pierre Verhulst, um matemático Belga. Modelo esse que depois seria mais aprofundado pelo matemático francês, Cauchy, que estudou as propriedades da função logística. Assim, com suas propriedades já definidas, o matemático e biólogo americano Alfred Lotka usou o modelo logístico para estudar o crescimento populacional. Segundo Souza et al. (2010), o modelo logístico é empregado por meio de uma função que possibilita a modelagem do crescimento de uma população. Ele é comumente utilizado para representar crescimentos sigmoidais, caracterizados pelo ponto de inflexão que ocorre aproximadamente na metade do último valor medido.

De acordo com Zeviani P. J. W.; Bonat (2013), a resposta média do modelo linear pode ser ajustada para uma escala mais adequada para o problema em questão, por meio de uma função de ligação que é dada por:

$$y_t = \alpha (1 + \beta e^{-\gamma t})^{-\tau} + \varepsilon \quad (15)$$

Os pontos de inflexão da curva são os valores de  $y$ , em que a curva muda de concavidade. No caso da curva exponencial, os pontos de inflexão são  $y = \ln \alpha k$  e  $E(y) = 2\alpha$ . Esses valores são independentes dos valores de  $y$ , ou seja, eles não dependem do valor específico de  $y$  em que a curva está sendo avaliada. Os parâmetros da curva exponencial podem ser estimados pelo método da máxima verossimilhança. Esse método busca encontrar os valores dos parâmetros que tornam a curva mais provável, dada a distribuição dos dados observados.

Os parâmetros da curva exponencial são  $\alpha$ ,  $\beta$  e  $\gamma$ . O  $\alpha$  representa o valor assintótico da variável resposta.  $\beta$  refere-se a uma constante que está relacionada ao valor observado inicial da variável resposta.  $\gamma$  é a taxa de crescimento da variável resposta, ou seja, a velocidade com que a variável resposta se aproxima do seu valor assintótico. Por sua vez,  $\tau$  é o parâmetro que molda a curvatura nos modelos, determinando o momento em que a transição ocorre de uma fase de crescimento inibitório para uma de crescimento menos restrito, indicando o ponto em que o crescimento se torna menos previsível e  $t$  o tempo em meses.

### 2.8.2 *Modelo Von Bertalanffy*

Segundo Oliveira et al. (2007), o modelo não linear Von Bertalanffy, é bastante utilizado para o estudo de crescimento em peso de peixes, isso porque pode explicar a variação de comprimento de uma determinada variável ao longo do tempo. No modelo de Von Bertalanffy,  $y$  cresce para uma assíntota horizontal superior, sendo a distância que falta percorrer dada por uma exponencial decrescente, logo a especificação do modelo na família Von Bertalanffy composta por 3 parâmetros será definida por:

$$y_t = \alpha (1 - \beta e^{-\gamma t})^3 + \varepsilon, \quad (16)$$

De acordo com Santos (2011), o modelo acima ajusta processos de crescimento sigmoidais, no qual o ponto de inflexão está localizado aproximadamente em 30% do último valor

medido. Os parâmetros  $\alpha$  representam o valor assintótico da variável resposta, enquanto  $\beta$  refere-se a uma constante que esta relacionada ao valor observado inicial e  $\gamma$  é a taxa de crescimento da variável resposta, que indica a velocidade com que o valor observado se aproxima do valor máximo, determinando a eficiência do crescimento.

### 2.8.3 Modelo Gompertz

O modelo de Gompertz é uma fórmula matemática que descreve como uma população ou sistema cresce ao longo do tempo. O crescimento começa rapidamente, mas depois diminui à medida que a população ou sistema se aproxima de um limite máximo (MÁXIMO, 2023). O modelo de Gompertz pode ser usado para estudar uma variedade de fenômenos, incluindo o crescimento populacional, o desenvolvimento de doenças e o aprendizado de novos conceitos. Tal modelo é dado por:

$$y_t = \alpha e^{-\beta e^{-\gamma t}} \quad (17)$$

Na biologia tem-se usado o modelo de Gompertz para estudar o crescimento populacional de diversas espécies e para prever o crescimento em diferentes cenários (DOMINGUES, 2011).

## 2.9 Método de *bootstrap*

A técnica de *bootstrap* é um método estatístico que usa a amostragem aleatória para estimar parâmetros. Ela foi desenvolvida por Bradley Efron em 1979 e é uma ferramenta poderosa que pode ser usada em uma variedade de aplicações. Ela funciona da seguinte forma, a partir de uma amostra original, são geradas várias amostras *bootstrap*, cada uma do mesmo tamanho da amostra original. Essas amostras *bootstrap* são geradas usando reposição, o que significa que cada observação da amostra original pode ser selecionada mais de uma vez. Assim, após gerar as amostras *bootstrap*, os parâmetros de interesse são estimados para cada amostra. A distribuição de amostragem dos parâmetros é então estimada usando os valores estimados das amostras *bootstrap*.

Segundo Araujo et al. (2013), a técnica de *bootstrap* pode ser usada para estimar uma variedade de parâmetros, incluindo a média, a variância, a correlação e os coeficientes de regressão. Ela também pode ser usada para construir intervalos de confiança e realizar testes de hipóteses. Essa abordagem tem algumas vantagens importantes sobre outros métodos estatísticos. Ela é não paramétrica, o que significa que não requer que a distribuição da população seja conhecida. Ela também é robusta, o que significa que é relativamente insensível a violações das suposições dos métodos paramétricos.

A técnica de *bootstrap* é uma ferramenta poderosa que pode ser usada em uma variedade de aplicações. Ela é uma alternativa viável aos métodos estatísticos tradicionais em muitos casos (EFRON; TIBSHIRANI, 1986). Vejamos alguns exemplos do seu uso:

- i) Estimar a média de uma população quando a distribuição da população é desconhecida.
- ii) Construir um intervalo de confiança para a correlação entre duas variáveis.
- iii) Testar a hipótese de que a média de uma população é igual a um valor especificado.
- iv) Avaliar o desempenho de um modelo de regressão.

O método de *bootstrap* é uma técnica de reamostragem que pode ser usada para fazer inferências sobre um determinado parâmetro. O processo consiste em repetir várias vezes a amostra original, gerando novas amostras com reposição Araujo et al. (2013). A média das novas amostras é então calculada e o erro padrão da média é estimado pelo desvio padrão.

## 2.10 *Bootstrap* paramétrico

O método *bootstrap* paramétrico consiste em gerar várias reamostras da amostra original, utilizando a distribuição de probabilidade dos dados originais. Para cada reamostra, são estimados os parâmetros do modelo de regressão não linear. Os valores estimados dos parâmetros são então usados para construir uma distribuição de probabilidade para os estimadores de mínimos quadrados. Essa distribuição de probabilidade pode ser utilizada para estimar intervalos de confiança para os estimadores de mínimos quadrados, bem como para testar hipóteses sobre os parâmetros do modelo.

## 2.11 Intervalos de confiança *bootstrap*

Segundo Souza et al. (2010), a metodologia *bootstrap* pode ser utilizada para obter intervalos de confiança confiáveis para os parâmetros de um modelo de regressão não linear. Isso é importante porque permite aos pesquisadores estimar com confiança a região em que os parâmetros do modelo provavelmente estão localizados. Efron e Tibshirani (1986) descrevem dois métodos para a obtenção de intervalos de confiança *bootstrap*: o método percentual e o método percentual t. Os intervalos obtidos via método percentil t tem a forma:

$$\gamma^* - t_{1-\frac{\alpha}{2}} \cdot \frac{s(\gamma^*)}{\sqrt{n}} \leq \gamma \leq \gamma^* + t_{1-\frac{\alpha}{2}} \cdot \frac{s(\gamma^*)}{\sqrt{n}}. \quad (18)$$

Em que  $s(\gamma^*)$  é o desvio padrão (estimado) de  $\gamma^*$ . Os  $t_i$  são determinados com base na distribuição *bootstrap* de  $\gamma^*$ .

## 2.12 Teste de normalidade *bootstrap*

Segundo Razali, Wah et al. (2011) o teste de Shapiro-Wilk é um teste não paramétrico, o que significa que ele não faz nenhuma suposição sobre a distribuição dos dados. Isso o torna um teste robusto, o que significa que ele é menos sensível a desvios da normalidade do que outros testes de normalidade. A hipótese do teste é definida como :

$$\begin{cases} H_0 : \text{Os parâmetros provém de uma distribuição Normal} \\ H_1 : \text{Os parâmetros não provém de uma distribuição Normal} \end{cases} \quad (19)$$

Segundo Araujo et al. (2013), a assimetria é uma medida da falta de simetria de uma distribuição. Ela mede a diferença entre a mediana e a média da distribuição. No caso da distribuição normal, a média e a mediana coincidem e, portanto, a distribuição é simétrica. Isso significa que a metade das observações fica de um lado da média e a outra metade fica do outro lado. Quando a distribuição não é simétrica, a média e a mediana não coincidem. Nesse caso, a assimetria pode ser positiva ou negativa.

Uma assimetria positiva indica que a distribuição tem mais observações com valores maiores do que a média do que observações com valores menores do que a média. Uma assimetria negativa, por sua vez, indica que a distribuição tem mais observações com valores menores do que a média do que observações com valores maiores do que a média. O teste de assimetria é um teste estatístico que avalia se uma distribuição é simétrica. Ele é baseado na estatística  $p$ , que é uma medida da assimetria da distribuição. O teste de assimetria é baseado na seguinte hipótese nula:

$$\begin{cases} H_0 : \text{Coeficientes da assimetria} = 0, \\ H_1 : \text{Coeficientes da assimetria} \neq 0. \end{cases} \quad (20)$$

No mesmo sentido o coeficiente de curtose deve ser igual a 3 para que a distribuição dos dados seja normal. A hipótese do teste é definida como:

$$\begin{cases} H_0 : \text{Coeficientes da curtose} = 3, \\ H_1 : \text{Coeficientes da curtose} \neq 3. \end{cases} \quad (21)$$

### 3 MATERIAL E MÉTODOS

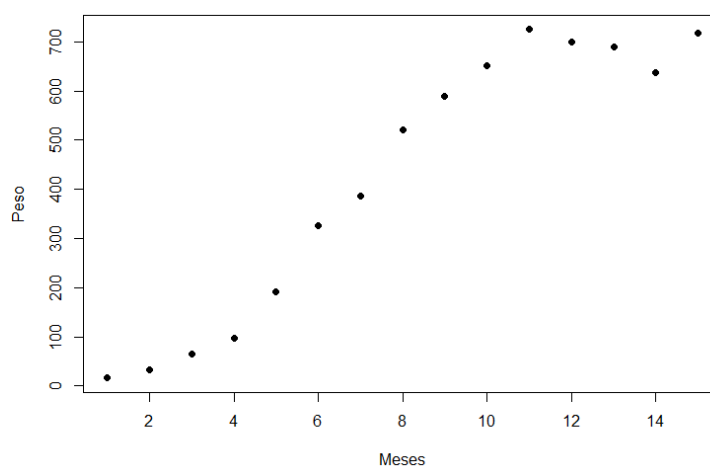
Com a intenção de aplicar a teoria abordada, foi utilizada uma amostra de 15 observações do peso seco dos bulbos de cebola e topos e o seu tempo de crescimento, obtidos em Ratkowsky (1983). Considerando que a variável resposta  $y$  refere-se ao peso seco dos bulbos de cebola, e a variável preditora ( $X$ ) é o tempo de crescimento. A Tabela 1 apresenta a relação dos 15 pesos secos dos bulbos de cebola com seus respectivos tempos de crescimento.

Tabela 1 – Relação do peso(g) e tempo de crescimento(meses) dos bulbos de cebola.

Tempo	Peso
1	16,08
2	33,83
3	65,80
4	97,20
5	191,55
6	362,20
7	386,87
8	520,53
9	590,03
10	651,92
11	724,93
12	699,56
13	689,96
14	637,56
15	717,41

Fonte: Ratkowsky (1983).

A Figura 1 apresenta a curva observada do peso em função do tempo. Através do gráfico percebe-se que o crescimento é não linear. Sendo assim, recomenda-se a necessidade de usar as técnicas do modelo não linear para melhor descrever os dados sob estudo.



Fonte: Elaborado pelo autor via software R.

Figura 1 – Relação entre o peso dos bulbos ao longo dos meses.

Nesse sentido, foram ajustados os modelos Logístico, Gompertz e o Von Bertalanffy, com

o propósito de estimar o peso dos bulbos de cebola em relação ao tempo. Para os modelos, foram calculadas as estimativas dos parâmetros, o erro padrão, a estatística do teste t para verificar se os parâmetros são significativos e os intervalos de confiança dos respectivos modelos. Na sequência, foi aplicada a técnica *bootstrap* com a finalidade de verificar, nos estimadores, o seu erro padrão e o viés, além de verificar a qualidade do ajuste entre os modelos, validando-os com confiança nos resultados. Por meio da função “nls” do Software R versão 4.3.1 interface do RStudio, foi possível obter as estimativas dos parâmetros.

O ajuste do modelo foi realizado por meio do método iterativo de Gauss-Newton. Este método começa por atribuir valores iniciais aos parâmetros do modelo, visando aproximar o modelo não linear a termos lineares. A partir desses valores iniciais, são realizadas várias iterações até que o processo convirja. Em seguida, é aplicado o método de mínimos quadrados para encontrar as estimativas aproximadas dos estimadores dos parâmetros, obtendo-se, assim, o problema de regressão não linear. O viés de Box permite verificar se existe algum parâmetro que contribua para o desvio da linearidade. Para avaliar o modelo, foi utilizados os critérios de determinação ( $R^2$ ) e informação de Akaike (AIC). O ( $R^2$ ) é uma medida da proporção da variabilidade da resposta que é explicada pelo modelo, enquanto o AIC é uma medida da complexidade do modelo em relação à sua capacidade de ajuste aos dados.

O método *bootstrap* paramétrico foi iniciado a partir dos valores estimados assintoticamente dos parâmetros da amostra original. Para as 1000 estimativas *bootstrap* obtidas na simulação, foram calculados o viés relativo *bootstrap*, os intervalos de confiança e realizados os testes de normalidade dos parâmetros, que foram: Shapiro-Wilk, assimetria e curtose. Esses testes foram realizados com o objetivo de medir a dispersão dos dados e verificar a normalidade da distribuição para cada modelo utilizado. As medidas desses testes foram obtidas a partir dos resíduos das distribuições, pois, assumindo que os modelos a serem testados são os mais adequados, os resíduos refletiriam as propriedades admitidas pela variável de erro.

#### 4 RESULTADOS E DISCUSSÃO

Levando em conta os dados observados, foi possível ajustar o modelo não linear em função do comportamentos dos dados e obter os estimadores para os parâmetros do modelo. Desse modo, segue os resultados inerentes aos modelos.

Tabela 2 – Estimativas (E) dos parâmetros (P), o erro padrão (E.P.), teste t e seus respectivos intervalos de confiança IC a 95% dos modelos.

Modelo	P.	E.	E.P.	valor $t$	p-valor	IC
Von Bertalanffy	$\alpha$	752,28381	30,38797	24.756	< 0,001	[698,3057; 825,1986]
	$\beta$	1,77290	0,30651	5.784	< 0,001	[1,2139; 2,1541]
	$\gamma$	0,33357	0,03651	9.136	< 0,001	[0,2574; 0,3877]
Logístico	$\alpha$	702,87140	13,93972	50,42	< 0,001	[673,9843; 733,9568]
	$\beta$	4,4425	0,3508	12,66	< 0,001	[3,7801; 5,2739]
	$\gamma$	0,68856	0,05737	12,00	< 0,001	[0,5807; 0,8228]
Gompertz	$\alpha$	723,10872	22,06050	32.778	< 0,001	[680,1908; 775,9797]
	$\beta$	2,50018	0,28469	8.782	< 0,001	[1,9374; 3,2739]
	$\gamma$	0,45010	0,05178	8.693	< 0,001	[0,3487; 0,5832]

Fonte: Elaborado pelo autor via software R.

Por meio da Tabela 2, é possível perceber que todos os parâmetros dos modelos são significativo a um nível de 0,05 de significância pelo teste t, com p valor menor que o nível de significância estabelecido. E encontram-se dentro do intervalo de confiança assintótico de 95%. As estimativas para os parâmetros foram diferentes, portanto o resultado foi diferente. Porém, percebe-se que o modelo logístico sobressaiu com relação aos modelos. Tendo um erro padrão menor e a amplitude do intervalo menor. Além do teste, os parâmetros encontram-se dentro do intervalo de confiança assintótico de 95%.

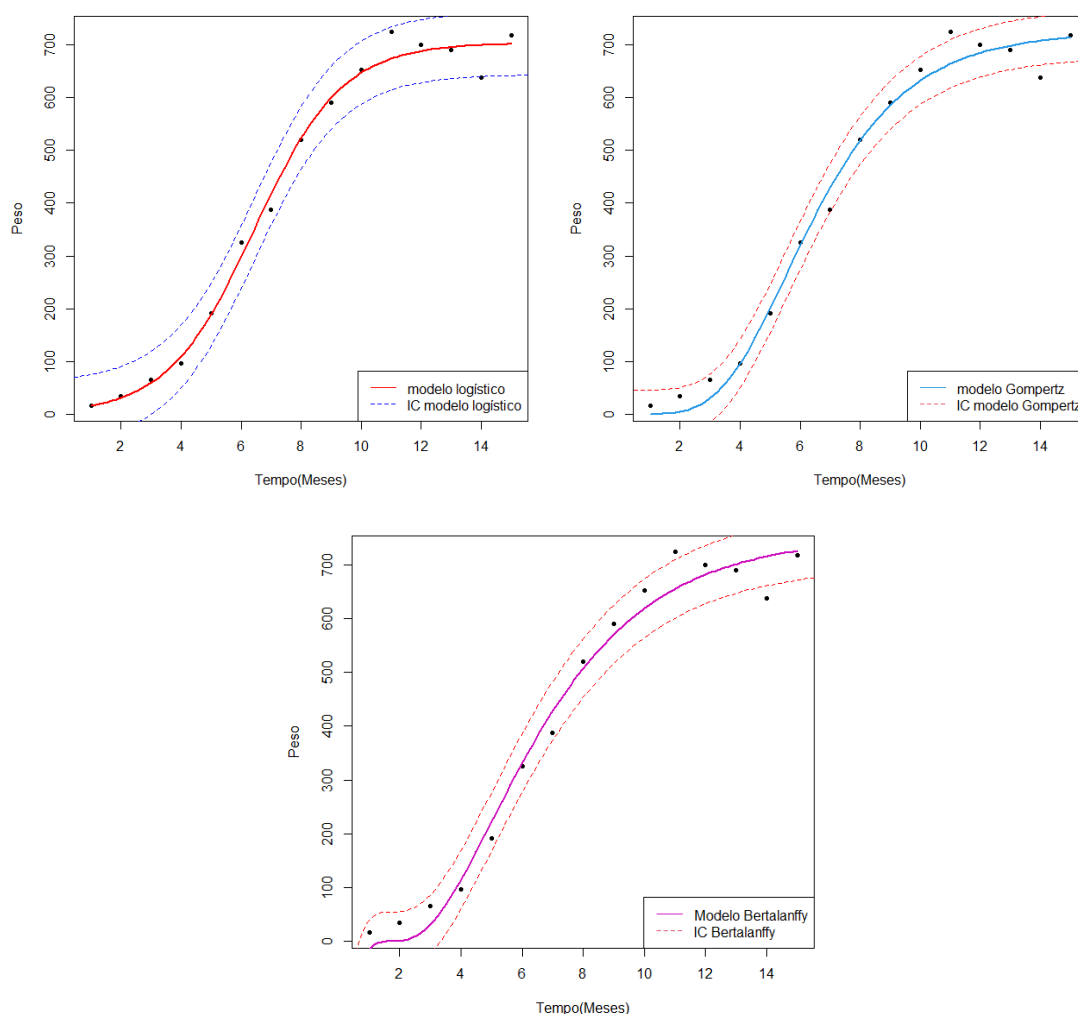
A qualidade de ajuste dos modelos aos dados, foi analisada por dois critérios: Informação de Akaike (AIC) e coeficiente de determinação  $R^2$ . Na Tabela 3, encontram-se os valores das estatística  $R^2$  e AIC. Assim, o melhor modelo ajustado é aquele que apresentar menor valor para AIC e maior valor para  $R^2$ .

Tabela 3 – Estatísticas da qualidade de informação de ajuste.

	Logístico	Gompertz	Von Bertalanffy
$R^2$	0,9918	0,9886	0,9837
AIC	146,4048	152,7216	157,9225

Na Tabela 3, é notável que os coeficientes de determinação são elevados, em torno de 99%. No entanto, mesmo com coeficientes de determinação muito próximos para os três modelos, destaca-se que o modelo logístico apresentou o maior coeficiente e o menor valor para o Critério de Informação de Akaike (AIC), sendo o modelo ligeiramente mais adequado aos dados sob estudo. Nesse sentido, a Figura 2 exibe a curva de crescimento do peso ajustada pelos modelos Von Bertalanffy, Logístico e Gompertz. Observa-se que o modelo logístico possui um intervalo de confiança mais apropriado, evidenciado pela faixa de confiança que engloba todos os dados observados.





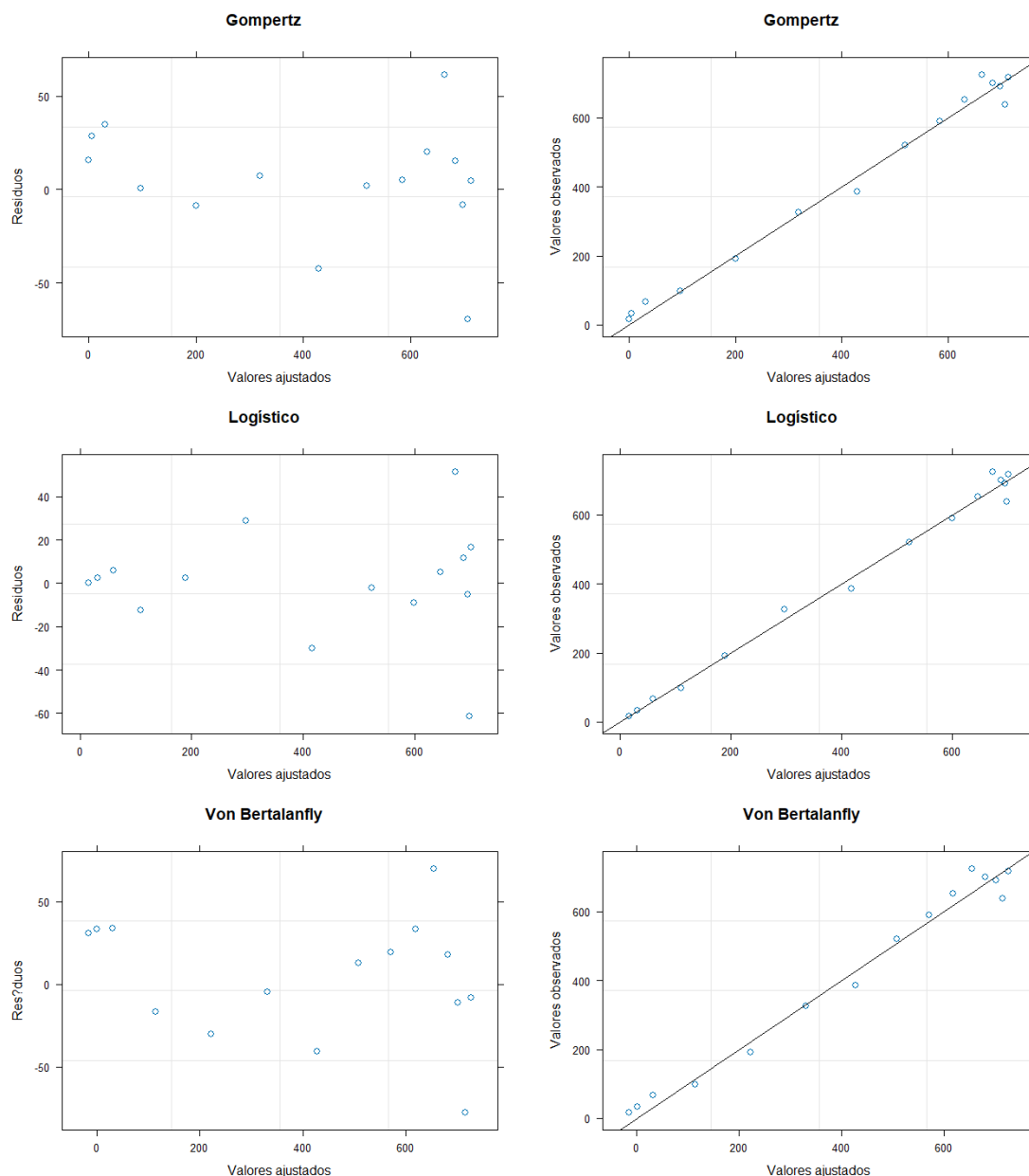
Fonte: Elaborado pelo autor via software R.

Figura 2 – Ajuste do peso do bulbo em relação ao tempo de crescimento para os modelos Logístico, Von Bertalanffy e Gompertz.

Após a seleção do modelo mais apropriado aos dados, a análise prosseguiu com a avaliação gráfica da homocedasticidade e normalidade dos resíduos, conforme ilustrado na Figura 3, para os modelos Gompertz, Von Bertalanffy e Logístico. Nessa figura, são apresentados os respectivos gráficos dos resíduos em relação ao crescimento (meses) do bulbo da cebola e dos valores observados em comparação com os valores preditos.

Na Figura 3, os gráficos de resíduos em relação aos valores ajustados, alocados no lado esquerdo, abordam a questão da homocedasticidade dos resíduos. Observa-se uma variância que tem o comportamento mais constante, com os resíduos distribuídos aleatoriamente ao redor da média zero. Essa homogeneidade de variâncias ao longo do tempo sugere que o modelo de regressão não linear é apropriado.

Ao analisar o comportamento gráfico no lado direito da Figura 3, que exhibe valores preditos em relação a valores observados, referente à normalidade, destaca-se a proximidade dos pontos em relação à reta, indicando evidências de que os resíduos seguem uma distribuição normal. Nota-se que o ajuste e a análise de resíduos dos três modelos são notavelmente semelhantes.



Fonte: Elaborado pelo autor via software R.

Figura 3 – Gráfico da Análise dos Resíduos.

Os estimadores dos parâmetros, em modelos de regressão não linear, não possuem algumas das propriedades apresentadas pelos estimadores dos parâmetros em modelos de regressão linear. Assim sendo, existe a necessidade de estudar o comportamento do modelo e isso pode ser feito por meio de estudos de simulação. Para obtenção das amostras simuladas foram consideradas, como valores paramétricos, as estimativas dos parâmetros do modelo logístico, no qual obteve o melhor ajuste para o peso do bulbo. Foram gerados 1000 amostras a partir de cada um dos modelos, cada uma com 15 observações.

#### 4.1 Simulação *bootstrap*

Na Tabela 4 são apresentadas as estimativas dos parâmetros pelo *bootstrap* nos três modelos, seus respectivos valores do vício relativo a estimativa *bootstrap* (V.R.B.), vício de Box

e o Intervalo de confiança.

Tabela 4 – Estimativas *bootstrap* (E.B.), vício relativo a estimativa *bootstrap* (V.R.B.), vício de Box, intervalo de confiança *bootstrap* a 95% para os parâmetros ( $p$ ).

Modelo	$p$	E.B.	V.R.B	Vício de Box	Intervalo de Confiança
Von Bert.	$\alpha$	754,8833	0,6816	5,1278	[707,8004 ;824,3946]
	$\beta$	1,7743	-0,9952	-0,0176	[ 1,2601 ; 2,1855]
	$\gamma$	0,3334	-1,0875	-0,0036	[0,2616 ; 0,3821]
Logístico	$\alpha$	703,2009	0,0635	0,4468	[676,8470; 727,4211]
	$\beta$	4,4576	1,0325	0,04587	[ 3,9127 ; 5,1729]
	$\gamma$	0,6930	0,9453	0,0065	[ 0,6026 ; 0,8082]
Gompertz	$\alpha$	724,4153	0,1807	1,3066	[688,0420 ;764,3047]
	$\beta$	12,7989	5,0410	0,6142	[ 7,7340 ; 21,9733]
	$\gamma$	0,4525	0,5381	0,0024	[0,3644 ; 0,5499]

Fonte: Elaborado pelo autor via software R.

Na Tabela 4 observa-se que os valores das estimativas dos parâmetros e seus respectivos intervalos de confiança *bootstrap* são muito próximos dos valores assintóticos. Os resultados da simulação *bootstrap* indicam que o modelo logístico é o melhor modelo para descrever o crescimento do peso do bulbo de cebola. O modelo logístico apresenta um bom ajuste aos dados, além de ser o modelo com o menor intervalo de confiança para os seus parâmetros.

O modelo Gompertz também apresenta um bom ajuste aos dados, mas o seu intervalo de confiança para os parâmetros é maior do que o do modelo logístico. O modelo de von Bertalanffy apresenta um ajuste inferior aos dois modelos anteriores.

Tabela 5 – Testes de normalidade *bootstrap*.

Modelo	$p$	Shapiro-Wilk	$p$ -valor	Assimetria	$p$ -valor	Curtose	$p$ -valor
Von Bert.	$\alpha$	0,9813 <sub>ns</sub>	< 0,0001	0,1255 <sub>ns</sub>	< 0,0001	3,4210*	0,0182
	$\beta$	0,9829 <sub>ns</sub>	< 0,0001	-0,3588 <sub>ns</sub>	< 0,0001	2,5658*	0,0005
	$\gamma$	0,9838 <sub>ns</sub>	< 0,0001	-0,4842 <sub>ns</sub>	< 0,0001	3,0488 <sub>ns</sub>	0,6735
Logístico	$\alpha$	0,9934 <sub>ns</sub>	< 0,0001	0,5666 <sub>ns</sub>	0,0029	3,7086*	0,0003
	$\beta$	0,9899 <sub>ns</sub>	< 0,0001	0,4046 <sub>ns</sub>	< 0,0001	3,2404 <sub>ns</sub>	0,1284
	$\gamma$	0,9915 <sub>ns</sub>	< 0,0001	0,3693 <sub>ns</sub>	< 0,0001	3,1690 <sub>ns</sub>	0,2570
Gompertz	$\alpha$	0,9967 <sub>ns</sub>	0,0405	0,1037*	0,1037	3,4503*	0,0112
	$\beta$	0,8750 <sub>ns</sub>	< 0,0001	2,0250 <sub>ns</sub>	< 0,0001	12,7553*	< 0,0001
	$\gamma$	0,9945 <sub>ns</sub>	0,0010	0,2595 <sub>ns</sub>	< 0,0001	3,3663*	0,0316

\* significativo ao nível de 5% de probabilidade, *ns* não significativo

Na Tabela 5 são apresentados os resultados dos testes de normalidade para os parâmetros via método *bootstrap*, no qual encontram-se os valores do teste de Shapiro-Wilk, do teste de assimetria e do teste de curtose, com seus respectivos  $p$ -valores. Nota-se para o teste Shapiro-Wilk, que os parâmetros não significativos aceitam a hipótese nula do teste, já os parâmetros significativos rejeitam a hipótese nula, em que esta hipótese nula evidencia que os parâmetros seguem uma distribuição normal. Neste caso, os três modelos apresentam seus respectivos parâmetros com índice de normalidade.

No teste de assimetria, a rejeição da hipótese nula ocorre quando os parâmetros são estatisticamente significativos, indicando uma falta de simetria na distribuição. Em contraste, a aceitação da hipótese nula acontece quando os parâmetros não são significativos, sugerindo

simetria e, portanto, uma distribuição aproximadamente normal. Para o teste de assimetria apenas o parâmetro  $\alpha$  dos modelos Von Bert. e Logístico apresentaram simetria entre os dados. No teste de curtose, a hipótese de igualdade é aceita quando os parâmetros não são significativos, indicando que a curtose é próxima ou igual a 3, refletindo um achatamento semelhante ao da distribuição normal. Portanto, dos modelos sob estudo, o modelo logístico teve um achatamento mais próximo da normal.

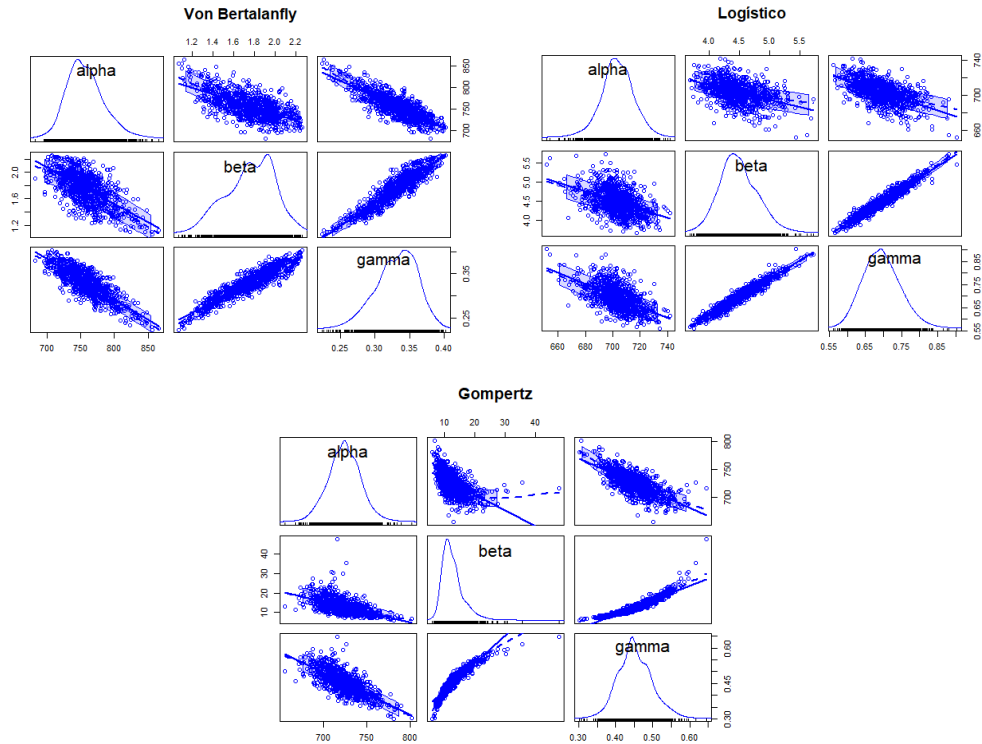


Figura 4 – Matrizes de dispersão conjunta de estimativas *bootstrap* dos parâmetros e os histogramas de frequência para os modelos analisados.

Conforme os testes relatados na tabela 5, é possível perceber que na Figura 4, os gráficos evidenciam os resultados inerentes as estimativas e testes expostos nas tabelas, acima, enfatizando o melhor ajuste do modelo Logístico aos dados do peso do bulbo da cebola.

## 5 CONCLUSÃO

Neste trabalho, foram avaliados três modelos de curva de crescimento para o peso do bulbo de cebola: Gompertz, Logístico e Von Bertalanffy. Os dados realizado em campo por Ratkowsky, com 15 observações ao longo de 15 meses.

Os resultados da simulação *bootstrap* mostraram que todos os três modelos foram capazes de ajustar os dados de forma satisfatória e tiveram resultados confiáveis. No entanto, o modelo Logístico apresentou o melhor ajuste, seguido do modelo Gompertz e do modelo de Von Bertalanffy.

O modelo Logístico apresentou um bom ajuste aos dados de acordo com o AIC e o  $R^2$ , além de ser o modelo com o menor intervalo de confiança para os seus parâmetros. O modelo de Gompertz também apresentou um bom ajuste aos dados, mas o seu intervalo de confiança para os parâmetros é mais amplo que o do modelo Logístico. O modelo de Von Bertalanffy apresentou um ajuste inferior aos dois modelos anteriores. Portanto, os resultados do estudo indicam que o modelo Logístico é o melhor modelo para descrever o crescimento do peso do bulbo de cebola. E, nesse sentido, uma vez que tal modelo se ajustou aos dados do peso de bulbo de cebola, podemos afirmar que o seu peso possui relação com o tempo de crescimento das mesmas.

## REFERÊNCIAS

- AMARAL, C.; SOUZA, M.; CATALAN, T. Um estudo do método de newton-raphson. *Revista eletrônica matemática e estatística em foco*, p. 42, 2015. Citado 3 vezes nas páginas 8, 12 e 13.
- ARAÚJO, F. M. d. et al. Uso da técnica bootstrap em modelos de regressão não linear. *Programa e Resumos*, 2013. Citado 2 vezes nas páginas 17 e 18.
- BATES, D. M.; WATTS, D. G. Nonlinear regression: iterative estimation and linear approximations. *Nonlinear regression analysis and its applications*, Wiley Hoboken, NJ, p. 32–66, 1988. Citado 3 vezes nas páginas 9, 10 e 14.
- DOMINGUES, J. S. Análise do modelo de gompertz no crescimento de tumores sólidos e inserção de um fator de tratamento. *Biomatemática*, v. 21, p. 103–12, 2011. Citado na página 17.
- EFRON, B.; TIBSHIRANI, R. Bootstrap methods for standard errors. *Confidence Intervals, and Other Measures of Statistical Accuracy*, v. 1, 1986. Citado 2 vezes nas páginas 17 e 18.
- HOFFMANN, R. *Análise estatística de relações lineares e não lineares*. [S.l.]: USP, 2016. Citado na página 12.
- MATTOS, T. do B. Modelos não lineares e suas aplicações. Juíz de Fora, 2013. Citado 6 vezes nas páginas 9, 11, 13, 14, 15 e 16.
- MÁXIMO, K. A. M. Compreendendo o modelo de gompertz e suas aplicações. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, v. 9, n. 3, p. 2134–2141, 2023. Citado na página 17.
- MAZUCHELI, J.; ACHCAR, J. A. Algumas considerações em regressão não linear. *Acta Scientiarum*, v. 24, n. 6, p. 1761–1770, 2002. Citado 4 vezes nas páginas 8, 9, 10 e 15.
- MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. [S.l.]: Saraiva Educação SA, 2017. Citado na página 8.
- OLIVEIRA, L. de et al. Modelo de von bertalanffy generalizado aplicado ao crescimento de suínos de corte. *IMECC–UNICAMP*, p. 101, 2007. Citado na página 16.
- RATKOWSKY, D. Nonlinear regression modelling. New York, Marcel Dekker, p. 62–88, 1983. Disponível em: <<https://www.itl.nist.gov/div898/strd/nls/data/ratkowsky3.shtml>>. Acessado em: 15 de novembro de 2023. Citado na página 20.
- RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, v. 2, n. 1, p. 21–33, 2011. Citado na página 18.
- RENAM, I. *Validação de modelos de crescimento aplicados ao perfil de mudanças fisiológicas do sexo feminino*. Tcc (trabalho de conclusão de curso). Campina Grande-PB, Brasil, 2019. Citado na página 8.
- SANTOS, A. L. P. Estudo de modelo de crescimento via dados simulados. *Monografia (Graduação) — Universidade Estadual da Paraíba, Campina*, 2011. Citado 2 vezes nas páginas 12 e 16.

SOUZA, E. M. de et al. Modelagem não linear da extração de zinco de um solo tratado com lodo de esgoto. *Acta Scientiarum. Technology*, Universidade Estadual de Maringa, v. 32, n. 2, p. 193–199, 2010. Citado 2 vezes nas páginas 16 e 18.

THOMAS, G. Regressão não linear. *Trabalho de conclusao do curso de pós-graduação em estatística e experimentação agropecuária. Escola superior de Agricultura Luiz de Queiroz-USP*, 2016. Citado 2 vezes nas páginas 11 e 14.

ZEVIANI P. J. W.; BONAT, W. *Modelos de regressão não linear*. Curitiba: W.M.Z., P.J.R.Jr W.H.B., 2013. 101 p. Citado 4 vezes nas páginas 8, 9, 12 e 16.