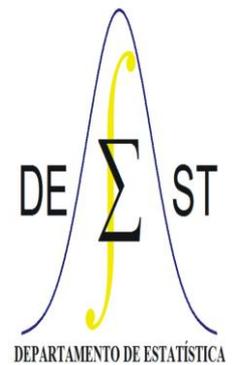




**UNIVERSIDADE FEDERAL DE OURO PRETO**  
**INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS**  
**DEPARTAMENTO DE ESTATÍSTICA**  
**BACHARELADO EM ESTATÍSTICA**



# **Modelo de Regressão Não Linear em Dados da COVID-19**

**Amanda Helena Lopes Valente**

Ouro Preto – MG

Fevereiro - 2024

**Amanda Helena Lopes Valente**

## **Modelo de Regressão Não Linear em Dados da COVID-19**

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador(a)

Prof. Dr. Diana Campos de Oliveira

UNIVERSIDADE FEDERAL DE OURO PRETO – UFOP

DEPARTAMENTO DE ESTATÍSTICA – DEEST

Ouro Preto - MG

Fevereiro – 2024



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DE OURO PRETO  
REITORIA  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
COLEGIADO DO CURSO DE ESTATÍSTICA



## FOLHA DE APROVAÇÃO

**Amanda Helena Lopes Valente**

### **Modelo de Regressão Não Linear em Dados da COVID-19.**

Monografia apresentada ao Curso de Estatística da Universidade Federal  
de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 22 de fevereiro de 2024

Membros da banca

Dra. Diana Campos de Oliveira - Orientadora (Universidade Federal de Ouro Preto) Dra.  
Graziela Dutra Rocha Gouvêa (Universidade Federal de Ouro Preto)  
Dr. Tiago Martins Pereira (Universidade Federal de Ouro Preto)

Professora Dra. Diana campos de Oliveira, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 22/02/2024



Documento assinado eletronicamente por **Diana Campos de Oliveira, PROFESSOR DE MAGISTERIO SUPERIOR**, em 01/04/2024, às 10:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0692378** e o código CRC **65576CB5**.

**Referência:** Caso responda este documento, indicar expressamente o Processo nº 23109.004245/2024-89 SEI nº 0692378

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163

Telefone: (31)3559-1312 - [www.ufop.br](http://www.ufop.br)

## AGRADECIMENTOS

Primeiramente, agradeço a Deus, por essa oportunidade de estudar em uma das maiores Universidades de Minas Gerais e por ter me dado foco, força e fé para continuar e não desistir desta importante jornada no meu caminho. Também agradeço aos meus pais e irmãos por toda força durante o curso e por me apoiarem em um dos momentos mais difíceis da minha jornada da Universidade, que foi mudar para o curso de Estatística, foi aqui que eu realmente me encontrei e é o que eu realmente amo fazer. Também agradeço à minha amiga Kethelen que sempre esteve do meu lado e sempre me apoiou em tudo.

Agradeço também à minha orientadora Diana, por toda dedicação neste projeto e por todo conhecimento passado durante este tempo. Sem você essa monografia não teria saído. Também agradeço a todos os professores que passaram nesta minha jornada, por todo conhecimento passado, por toda paciência e compreensão e principalmente por ter me ajudado a me tornar a profissional que sou hoje.

Não posso esquecer também dos meus colegas e professores do curso de matemática, onde fiquei por sete períodos e grande parte da bagagem que tenho hoje veio dessa experiência maravilhosa que tive.

Por fim, mas não menos importante agradeço a oportunidade de estagiar na Secretaria de Desenvolvimento Econômico na Prefeitura de Ouro Preto, pois a experiência na prática foi através desta incrível oportunidade, onde participei de diversas pesquisas de campo, análise de dados, Dashboards, modelos estatísticos, entre outros. Só tenho a agradecer por todas as oportunidades e aprendizados que passaram no meu caminho durante essa jornada.

## RESUMO

Os modelos de regressão desempenham um papel fundamental na análise de dados, mas em situações em que a aplicação de modelos de regressão linear é inviável, os modelos de regressão não linear surgem como alternativa. Esses modelos oferecem a capacidade de representar relações mais complexas, como curvas, oscilações, interações não lineares e comportamentos caóticos. Este estudo propõe a utilização de dois modelos de regressão não linear, o modelo logístico e também do modelo de Richards para descrever o padrão de crescimento de casos de COVID-19. Os dados utilizados foram obtidos a partir do site da Organização Mundial da Saúde (OMS). Em uma tentativa de solucionar os problemas da variância não constante e não normalidade dos erros, foi feita uma transformação na variável resposta ( $y^3$ ). A aplicação dos modelos de regressão não linear incluiu uma análise de variância (ANOVA), revelando que o modelo logístico se ajustou melhor aos dados e a curva resultante do modelo logístico ajustou-se de maneira satisfatória aos dados. Também foi utilizado os critérios AIC e BIC para verificar qual dos modelos se ajustou melhor aos dados. Embora o coeficiente de determinação ( $R^2$ ) tenha sido significativamente elevado, a análise de resíduos não indicou a aceitação completa das suposições em nenhum dos dois modelos.

**Palavras-chave:** regressão não linear, modelo logístico, modelo de richards, coronavírus.

## ABSTRACT

Regression models play a fundamental role in data analysis, but in situations where the application of linear regression models is unfeasible, nonlinear regression models emerge as an alternative. These models offer the capability to represent more complex relationships, such as curves, oscillations, nonlinear interactions, and chaotic behaviors. This study proposes the use of two nonlinear regression models, the logistic model and the Richards model, to describe the growth pattern of COVID-19 cases. The data used were obtained from the World Health Organization (WHO) website. In an attempt to address the issues of non-constant variance and non-normality of errors, a transformation was applied to the response variable ( $y^3$ ). The application of nonlinear regression models included an analysis of variance (ANOVA), revealing that the logistic model fit the data better, and the resulting curve from the logistic model fit the data satisfactorily. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were also used to determine which of the models best fit the data. Although the coefficient of determination ( $R^2$ ) was significantly high, residual analysis did not indicate complete acceptance of assumptions in either of the two models.

**Keywords:** nonlinear regression, logistic model, Richards model, coronavirus.

# SUMÁRIO

.....	1
1. INTRODUÇÃO .....	1
2. REVISÃO DE LITERATURA .....	2
2.1 Coronavírus .....	2
2.1.1 Epidemiologia .....	2
2.1.2 Vias de Transmissão .....	3
2.1.3 Período de incubação .....	3
2.1.4 Apresentação clínica .....	4
2.1.5 Vacina .....	4
2.2 Modelo de regressão não linear .....	5
2.3 Definição .....	7
2.4 Métodos para Estimações dos parâmetros do modelo de regressão não linear .....	9
2.4.1 Método dos Mínimos Quadrados .....	10
2.4.2 Métodos Iterativos .....	12
2.5 Qualidade do ajuste de regressão .....	14
2.5.1 Coeficiente de determinação ( $R^2$ ) .....	15
2.5.2 Critério de informação de Akaike (AIC) .....	16
2.5.3 Critério de informação bayesiano .....	17
2.6 Análise de Resíduos .....	18
2.6.1 Teste de homocedasticidade .....	18
2.6.2 Teste de normalidade .....	19
2.6.3 Independência dos Resíduos .....	20
2.7 Modelo Logístico .....	20
2.7.1 Características do Gráfico do Modelo Logístico .....	21
2.8 Modelo de Richards .....	22
3. MATERIAL E MÉTODOS .....	25
3.1 Origem dos Dados .....	25
4. Resultados e discussões .....	27
4.1 Modelo Logístico .....	27
4.2 Modelo de Richards .....	33
4.3 Comparação dos modelos .....	36
5. RESULTADOS E CONCLUSÕES .....	37

6. Referências.....	38
APÊNDICES .....	40

## LISTA DE TABELAS

Tabela 1: Exemplos de Modelos de crescimento .....	7
Tabela 2: Modelos de Crescimento Logístico e Richards .....	26
Tabela 3: Tabela de Coeficientes de Regressão do modelo Logístico .....	27
Tabela 4: Análise da Anova em relação aos dados mencionados .....	28
Tabela 5 : Tabela de Coeficientes de Regressão do modelo Logístico após a transformação .	29
Tabela 6: Análise da Anova em relação aos dados transformados.....	31
Tabela 7: Tabela de Coeficientes de Regressão para o modelo de Richards .....	33
Tabela 8: Análise da Anova do modelo de Richards.....	34
Tabela 9: Comparação dos valores do AIC e BIC nos modelos .....	36

## LISTA DE FIGURAS

Figura 1: Gráfico do Modelo Logístico .....	21
Figura 2: Gráfico do Modelo de Richards .....	23
Figura 3: Diagrama de dispersão dos dados .....	25
Figura 4: Análise de Resíduos do modelo logístico .....	28
Figura 5: Gráfico de dispersão dos dados após a transformação.....	29
Figura 6: Curva do modelo logístico ajustada aos dados transformados .....	30
Figura 7: Análise de Resíduos do modelo logístico .....	32
Figura 8: Curva do modelo de Richards ajustada aos dados .....	34
Figura 9: Análise de Resíduos do modelo de Richards .....	35

## 1. INTRODUÇÃO

Em estatística, regressão é uma das técnicas mais utilizadas para análises de dados, ela permite modelar e inferir a relação de uma variável dependente (variável de resposta) com variáveis independentes (variáveis explicativas).

Há vários métodos de estimação de coeficientes de regressão tais como método dos mínimos quadrados, método dos momentos generalizado, método da máxima verossimilhança, logit. A escolha do modelo depende do comportamento das variáveis e dos dados.

Existem situações em que não é desejável, ou até mesmo possível, descrever um fenômeno através de um modelo de regressão linear. Este fato impulsionou o estudo de novas técnicas estatísticas, surgindo uma nova classe de modelos, os denominados modelos não lineares.

O objetivo desse trabalho foi utilizar modelos não lineares para descrever os números de casos de COVID-19 no Brasil, ao longo dos 750 dias, com data inicial em 3 de janeiro de 2020, com o auxílio de modelos de regressão não lineares. Esses modelos são muito utilizados para descrever curvas que não seguem uma relação linear como a curva de crescimento da COVID-19.

Os estudos envolvendo a COVID-19 são de extrema importância por várias razões e uma das mais importantes envolvendo os modelos de regressão não lineares é a ajuda que esses modelos podem dar para capturar padrões mais complexos e não lineares na disseminação do vírus.

A compreensão dos padrões de propagação é crucial para a implementação de medidas de controle, contenção, para a prever futuros picos, entender fatores que contribuem para a propagação da doença, entre outros.

## 2. REVISÃO DE LITERATURA

### 2.1 Coronavírus

Todas as informações sobre a COVID-19 neste trabalho foram retiradas da (OMS, 2023). Foi declarado em 30 de janeiro de 2020, que o surto do novo coronavírus constituía uma Emergência de Saúde Pública de Importância Internacional (ESPII), considerado o nível mais alto de alerta da Organização, sendo “um evento extraordinário que pode constituir um risco de saúde pública para outros países devido a disseminação internacional de doenças; e potencialmente requer uma resposta internacional coordenada e imediata”.

Já em 11 de março de 2020, declarou a COVID-19 como uma pandemia, que se entende como a distribuição geográfica da doença.

É muito variável o número de casos e mortes causadas pelo vírus SARS-COV-2 nos diversos países pelo mundo. Essa variabilidade é causada por diversos fatores como, distanciamento social, medidas governamentais, realização de testes de diagnósticos, entre outros.

A velocidade que aumenta o número de casos é assustadora, devido ao fato de o vírus ser extremamente transmissível, por exemplo, se cada indivíduo infectado transmite para outras duas pessoas e essas duas pessoas para mais duas pessoas cada uma, essa propagação fará o número de casos aumentar de uma forma muito rápida.

O Brasil teve o seu primeiro caso confirmado em fevereiro de 2020, crescendo de uma maneira um pouco controlada, pois as medidas de prevenção estavam sendo cumpridas. Entretanto, esse crescimento desacelerado não perdurou por muito tempo, visto que em setembro de 2020 já haviam cerca de 5.000.000 de casos confirmados e mais de 142.000 mortes.

#### 2.1.1 Epidemiologia

A epidemiologia da COVID-19 varia entre países devido às diferentes medidas de prevenção que afetam diretamente o número de casos e óbitos. Vários fatores, incluindo aspectos biológicos, sociodemográficos, econômicos, organizacionais e recursos do sistema de saúde, gravidez, estão associados à propagação da doença.

Homens são mais frequentemente e gravemente afetados do que mulheres, com uma idade média de 47 anos. A maioria das mortes ocorre em indivíduos acima de 70 anos e com doenças crônicas subjacentes, embora casos leves sejam mais comuns em crianças e

adolescentes. No entanto, há relatos de casos de Síndrome Inflamatória Multissistêmica na Infância associada à COVID-19, alguns dos quais fatais.

Foram utilizadas algumas estratégias destinadas a controlar a propagação da doença, como isolamento social, higienização, entre outros, até que surgisse uma vacina ou tratamento para a doença. Uma outra estratégia também seria o controle da epidemia pela imunidade adquirida de forma gradual. Visto que, o sistema de saúde não suportaria um grande fluxo de pessoas doentes.

Então, alguns países conseguiram aliar medidas de isolamento social em conjunto com a testagem universal. Os países que realizaram a testagem universal em conjunto com o distanciamento social, higienização e uso de máscaras controlaram a epidemia e reduziram as taxas de letalidade do vírus, como foi o caso da Coreia do Sul.

No Brasil, eram testados somente casos graves, ou seja, a estimativa de detecção dos casos confirmados era um em cada cinco casos e um em cada dez casos confirmados, sendo que foi estimado que 79% das infecções eram transmitidas por pessoas sem sintomas, ou seja, assintomáticos.

### 2.1.2 Vias de Transmissão

A principal transmissão do SARS-COV-2 é através de gotículas de uma pessoa contaminada. O contato com superfícies ou fontes contaminadas por gotículas respiratórias liberadas via tosse ou espirro, entre outros e leva as mãos ao nariz, boca ou olhos também seria uma possível fonte de transmissão. Porém, não há relatos específicos que comprovem esse tipo de transmissão.

### 2.1.3 Período de incubação

Em média, o período de incubação do SARS-COV-2, é de aproximadamente 5 dias, tendo uma variação de 0 a 14 dias. Em 97,5% dos pacientes os sintomas se desenvolvem em 11,5 dias de infecção. Um período de incubação longo pode favorecer o aumento do risco de transmissão.

#### 2.1.4 Apresentação clínica

Os pacientes infectados por SARS-COV-2 apresentam um quadro clínico muito variável podendo ser pacientes assintomáticos, sintomas leves ou até mesmo um quadro mais grave. Porém, 80% dos casos são leves a moderados com cura espontânea.

Os sintomas mais comuns são febres, tosse e fadiga, com febre sendo o sintoma mais frequente nos pacientes. Alguns outros menos comuns são dor de cabeça, congestão/descarga nasal, conjuntivite, diarreia, vômito, dor abdominal, entre outros. Um estudo mostrou que manifestações cardíacas, como arritmias, e disfunção hepática aconteciam em até 50% dos casos.

#### 2.1.5 Vacina

Todas as informações sobre as vacinas contra a covid foram retiradas (Instituto Butantan. (2022, 17 de janeiro)). As primeiras vacinas contra a covid disponibilizadas para a população no Brasil foram a Corona Vac, produzida pela Sinovac na China e pelo Instituto Butantan no Brasil, e a Oxford, produzida pela AstraZeneca.

No Brasil, dia 17 de janeiro de 2021 foi aplicada a primeira vacina contra a Covid-19. A escolhida para receber a dose foi a enfermeira Mônica Calazans, que, desde o começo da pandemia, trabalha na linha de frente, no Instituto Emílio Ribas, em São Paulo.

Logo em seguida, diversos profissionais da área de saúde, idosos, indígenas, quilombolas e adultos com comorbidades receberam a primeira dose da Corona Vac.

Além da AstraZeneca e da CoronaVac, a Anvisa autorizou o uso de outras duas vacinas no Brasil: a Janssen e a Pfizer.

Com o início do Projeto S, em fevereiro de 2021, que se iniciou com vacinação de voluntários para o estudo de efetividade da vacina Corona Vac, ocorreu na cidade Serrana, em São Paulo. Os resultados da pesquisa foram positivos, pois mostraram uma queda drástica em hospitalizações e mortes, até mesmo em idosos, as principais vítimas da covid-19.

Em maio de 2021, os resultados prévios do Projeto S indicaram que os casos sintomáticos caíram 80%, as internações, 86%, e as mortes, 95%, para os voluntários que se vacinaram com a vacina Corona Vac.

O Brasil, em julho de 2021, alcançou a marca de 96 milhões de brasileiros vacinados pelo menos com uma dose contra a Covid-19. Com o progresso da vacinação no Brasil, houve

uma queda de 42% das mortes causadas pelo Covid-19, segundo dados do LocalizaSUS, do Ministério da Saúde.

Em 2022, após dois anos da pandemia de Covid-19, o mundo estava o mais próximo da normalidade desde o início da pandemia. Apesar disso, o vírus ainda circulava, mas a taxa de morte caiu drasticamente no Brasil, a vacinação evoluiu, e o uso de máscara deixou de ser obrigatório. Mas foi em 5 de maio de 2023, que a Organização Mundial da Saúde (OMS) declarou o fim da Emergência de Saúde Pública de Importância Internacional (ESPII) relacionado à COVID-19.

A queda nas hospitalizações, internações e mortes relacionadas à COVID-19, e um grande número de pessoas imunizadas ao SARS-COV-2, levaram os membros da 15ª sessão deliberativa do Comitê, a recomendar o fim da pandemia para o diretor-geral da OMS, que em seguida tomou a decisão.

Com tudo isso, o fim da pandemia causado pelo vírus SARS-COV-2, o COVID-19, não quer dizer que o vírus tenha deixado de existir e continua sendo uma ameaça à saúde. E como disse, Jarbas Barbosa, diretor da Organização Pan-Americana da Saúde (OPAS), “Não devemos baixar a guarda, precisamos continuar vacinando os grupos vulneráveis e fortalecendo a vigilância. Também é hora de nos concentrarmos em nos preparar melhor para futuras emergências e reconstruir melhor para um futuro mais saudável e sustentável”.

## 2.2 Modelo de regressão não linear

Os Modelos não lineares não seguem uma relação linear simples entre as variáveis independentes, explicativas ou regressora, que podem ser fixas (controladas) ou aleatórias, e a variável dependente ou resposta. Em vez disso, eles podem apresentar relações complexas, como curvas, oscilações, interações não lineares ou comportamentos caóticos (REGAZZI, 2003).

Esses modelos são comumente utilizados em diversas áreas, como física, química, biologia, engenharia, finanças e ciências para descrever fenômenos complexos e realizar previsões mais precisas.

O modelo é dito não linear quando ele não é linear em relação aos seus parâmetros, ou seja, pelo menos uma das derivadas parciais em relação a um certo parâmetro é função de algum parâmetro desconhecido.

Em geral, não há expressão de forma fechada para os parâmetros como há na regressão linear. Geralmente, algoritmos de otimização numérica são aplicados para determinar os parâmetros ajustados.

Ao contrário dos modelos lineares, os modelos não lineares geralmente exigem técnicas mais avançadas de estimação e inferência estatística. Por exemplo, a regressão linear pode ser resolvida com um simples ajuste de mínimos quadrados, enquanto a regressão não linear pode exigir métodos mais complexos.

Um das vantagens dos modelos não lineares é que em algumas situações eles fornecem uma relação mais específica entre as variáveis envolvidas no estudo, apresentam parâmetros que são interpretáveis biologicamente e, em geral, são mais parcimoniosos, em relação aos modelos lineares.

Entretanto os estimadores de mínimo quadrados para os modelos não lineares não apresentam propriedades de não tendenciosidade e variância mínima, estas propriedades desejáveis são satisfeitas assintoticamente, ou seja, quando o tamanho da amostra é suficientemente grande.

Nos modelos não lineares como os parâmetros resultantes são interpretáveis biologicamente é importante que se tenha um conhecimento prévio sobre a relação a ser modelada.

Algumas das desvantagens dos modelos não lineares é que os métodos de inferência são aproximados e requerem métodos iterativos de estimação, mas com o desenvolvimento computacional ficou contornado este problema.

Uma classe de modelos que tem proporcionado uma grande motivação para o desenvolvimento estatístico em modelos de regressão não lineares são os modelos de crescimento, utilizados em diversas áreas do conhecimento.

Na agricultura permite conhecer como a produção cresce, assim como a velocidade do crescimento, a depender dos tratamentos ambientais em que elas foram submetidas. Na medicina permite um maior conhecimento do crescimento infantil. Em demografia pode se conhecer melhor o crescimento populacional. Na química há interesse na formulação de produtos de uma reação química ao longo do tempo.

Para modelos de crescimento, a literatura apresenta várias equações, ver Seber e Wild (2003). Na

Tabela 1, são apresentados alguns modelos para descrever crescimento, como de animais e plantas na tentativa de compreender seus mecanismos essenciais.

Tabela 1: Exemplos de Modelos de crescimento

Modelo	Função
Weibull	$f(x) = \alpha - \beta e^{-\gamma x^\delta}$
Von Bertalanffy	$f(x) = \alpha[1 - e^{-\gamma(x-\delta)}]$
Gompertz	$f(x) = \alpha e^{-e^{\beta-\gamma x}}$
Morgan-Mercer-Flodin (MMF)	$f(x) = \frac{\beta\gamma + \alpha x^\delta}{(\gamma + x^\delta)}$

Para os modelos de crescimento temos o parâmetro  $\alpha$  como o limite máximo esperado para a resposta, ou seja, a assíntota. Já o parâmetro  $\beta$  está relacionado ao intercepto, ou seja, ao valor de  $\mu = E(y)$  quando  $x = 0$ . O parâmetro  $\gamma$ , por sua vez, está relacionado à taxa média de crescimento da curva e, finalmente, alguns modelos também utilizam o parâmetro  $\delta$  para aumentar a flexibilidade do ajuste dos dados.

### 2.3 Definição

A relação variável resposta/variável independente, nos modelos não lineares pode ser representada pela equação:

$$y_i = f(x_i, \theta) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

em que:

$y_i$  representa a observação da variável dependente;

$f(x_i, \theta)$  é a função de regressão ou função resposta;

$x_i$  representa a observação da variável independente;

$\theta$  é o vetor de parâmetros desconhecidos;

$\varepsilon_i$  representa o efeito do erro aleatório não observável considerados independentes e identicamente distribuídos, normais com média zero e variância constante  $\sigma^2$ .

Ao analisar os modelos

( 2.2) e

(2.3), a seguir:

$$f(x, \theta_0, \theta_1, \theta_2) = (\theta_0 + \theta_1 x + \theta_2 x^2) \quad (2.2)$$

$$f(x, \theta_a, \theta_e, \theta_c) = \theta_a(1 - \exp - \theta_e(x - \theta_c)) \quad (2.3)$$

Percebemos que as derivadas parciais de

( 2.2) são:

$$\frac{\partial f}{\partial \theta_0} = 1, \quad \frac{\partial f}{\partial \theta_1} = x, \quad \frac{\partial f}{\partial \theta_2} = x^2 \quad (2.4)$$

Verifica-se que as derivadas parciais não envolvem parâmetros, mas apenas as quantidades conhecidas em

( 2.2),  $x$  e  $x^2$ . Por isso, a equação

( 2.2) é um modelo linear nos parâmetros. Por outro lado, ao avaliar-se as derivadas parciais da equação

(2.3), temos:

$$\frac{\partial f}{\partial \theta_a} = 1 - \exp\{\theta_e(x - \theta_c)\}$$

$$\frac{\partial f}{\partial \theta_e} = -\theta_a(\theta_c - x) \exp\{-\theta_e(x - \theta_c)\}$$

$$\frac{\partial f}{\partial \theta_c} = -\theta_a \theta_b \exp\{-\theta_b(x - \theta_c)\},$$
(2.5)

Podemos notar que temos pelo menos dois parâmetros presentes nas derivadas parciais e então,

(2.3) é um modelo não linear.

Um modelo que se torna linear nos parâmetros através de uma transformação é chamado de modelo intrinsecamente linear.

Um exemplo de um modelo intrinsecamente linear é:

$$f(x, \theta) = \exp\{\theta_0 + \theta_1 x\}$$
(2.6)

Após realizar a transformação log, o modelo se torna linear nos parâmetros, dado por:

$$\log f(x, \theta) = \theta_0 + \theta_1 x$$
(2.7)

Já um modelo em que não se torna linear nos parâmetros após transformações é chamado de intrinsecamente não linear, que é o que ocorre em

(2.3).

## 2.4 Métodos para Estimções dos parâmetros do modelo de regressão não linear

Os métodos de estimação para os parâmetros de uma regressão são vários, os mais utilizados são o método dos mínimos quadrados e o método da máxima verossimilhança.

Em modelos lineares, para obter as estimativas dos parâmetros, um método amplamente utilizado é o dos mínimos quadrados, esse procedimento também pode ser utilizado em modelos não lineares. Entretanto, a minimização das somas de quadrados do erro produz equações normais que não são lineares nos parâmetros o que impossibilita uma solução analítica. Uma alternativa é a utilização de métodos iterativos que permitem encontrar soluções analíticas aproximadas.

### 2.4.1 Método dos Mínimos Quadrados

O Método dos Mínimos Quadrados (Damodar N. Gujarati, Dawn C. Porter (2011)) é uma técnica de otimização matemática que busca encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos) de forma a maximizar o grau de ajuste do modelo aos dados observados.

Acredita-se que criador das bases fundamentais do método dos mínimos quadrados tenha sido Carl Friedrich Gauss, em 1795, quando Gauss tinha somente dezoito anos. Entretanto, em 1805 Adrien-Marie Legendre foi o primeiro a publicar o método, em seu *Nouvelles méthodes pour la détermination des orbites des comètes*. Gauss publicou suas conclusões apenas em 1809.

Segue o modelo de regressão não linear:

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.8)$$

com

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, f(x, \boldsymbol{\theta}) = \begin{bmatrix} f(x, \theta_1) \\ f(x, \theta_2) \\ \vdots \\ f(x, \theta_n) \end{bmatrix} e \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.9)$$

Este método é uma extensão dos mínimos quadrados ordinários (MQO) que é usado em modelos lineares. No método dos mínimos quadrados em modelos não lineares, também busca-se minimizar a soma dos resíduos quadrados entre as previsões do modelo e os valores observados. Os MQO minimizarão a soma dos quadrados dos resíduos (SQR).

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2 \quad (2.10)$$

Para notação vetorial, temos:

$$SQRes(\theta) = \|y - f(\hat{\theta})\|^2$$

$$SQRes(\theta) = [y - f(\hat{\theta})]^T [y - f(\hat{\theta})] \quad (2.11)$$

Com

- $SQRes$  é a soma dos quadrados dos erros aleatórios que deverá ser minimizada por  $\theta$ ;
- $\|y - f(\hat{\theta})\|$  é a norma euclidiana de  $y - f(\hat{\theta})$

Ao fazer a diferenciação de (2.10) com relação a cada um dos parâmetros e igualar as equações a zero, temos,

$$\left( \frac{\partial SQRes(\theta)}{\partial \theta_l} \right) \Big|_{\theta} = 0 \quad (l = 1, 2, \dots, r)$$

$$\sum_{i=1}^n [y_i - f(x_i, \theta)] \frac{\partial f(x_i, \theta)}{\partial \theta_l} \Big|_{\theta} = 0 \quad (i = 1, 2, \dots, r)$$
(2.12)

Para  $f(\theta)$ , tem-se uma função vetor coluna  $n \times 1$ , que é dada a seguir:

$$f(\theta) = \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \\ \vdots \\ f_n(\theta) \end{bmatrix}_{n \times 1}$$
(2.13)

A matriz  $F(\theta)$  é a matriz jacobiana de  $f(\theta)$ , dada por:

$$F(\theta) = \frac{\partial f(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial f_1(\theta)}{\partial \theta_1} & \frac{\partial f_1(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_1(\theta)}{\partial \theta_p} \\ \frac{\partial f_2(\theta)}{\partial \theta_1} & \frac{\partial f_2(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\theta)}{\partial \theta_1} & \frac{\partial f_n(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_n(\theta)}{\partial \theta_p} \end{bmatrix}_{n \times p}$$
(2.14)

Sendo assim,  $\hat{\theta}$ , o estimador de mínimos quadrados, exerce a equação  $\frac{SQRes}{\partial \theta^T} = \phi$  caracteriza a minimização em estudo, assim:

$$\frac{QMR}{\partial \theta^r} = \frac{\partial}{\partial \theta^r} [y - f(\theta)]^T [y - f(\theta)] = 2[y - f(\theta)]^T \left( \frac{-\partial f(\theta)}{\partial \theta^r} \right) = -2[y - f(\theta)]^T F(\theta) \quad (2.15)$$

Logo, tem-se:

$$\begin{aligned} -2[y - f(\hat{\theta})]^T F(\hat{\theta}) &= \phi \\ -2F^T(\hat{\theta})[y - f(\hat{\theta})] &= \phi \\ F^T(\hat{\theta})[y - f(\hat{\theta})] &= \phi \end{aligned} \quad (2.16)$$

Com isso, o modelo não linear é dado pelo sistema de equações normais a seguir:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_p} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_p} \end{bmatrix}_{n \times p} \left( \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\theta}) \\ f_2(\hat{\theta}) \\ \vdots \\ f_n(\hat{\theta}) \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.17)$$

Ao seguir este método, é possível obter um sistema de  $p$  equações normais (2.12) para o modelo não linear.

No entanto, mesmo com a aplicação do MQO, não é possível obter soluções explícitas para as incógnitas, são feitas as iterações até que alcance a convergência seguindo algum critério escolhido. É importante mencionar que os MQO aplicados aos modelos de regressão não linear são chamados de mínimos quadrados não lineares (MQNL).

## 2.4.2 Métodos Iterativos

São diversos métodos iterativos, tais como Gauss Newton, Newton-Raphson, Método de Jacobi, Método de Gauss-Seidel, Método do gradiente conjugado, dentre outros.

Um método iterativo é um algoritmo que gera uma sequência de soluções aproximadas que vão melhorando conforme iterações são executadas, cada uma das aproximações é obtida das anteriores pela repetição do mesmo processo que resolvem uma classe de problemas estabelecida. Para tal procedimento se faz necessário uma aproximação inicial escolhida pelo pesquisador.

Uma implementação específica de um método iterativo, incluindo o critério para a parada é um algoritmo iterativo. Um método iterativo é considerado convergente se a sequência correspondente converge, dado uma tolerância inicial de aproximação.

Os métodos iterativos pelo qual se encontra a solução do Sistema de equações normais é uma forma de resolver os problemas de mínimos quadrados não lineares.

Um método iterativo amplamente utilizado para regressão não linear é o Método de Gauss Newton, no qual é feita a linearização da função não linear.

Seja  $y_i = f(x_i, \theta) + \varepsilon_i$  o modelo não linear e  $\hat{\theta}_0$  um valor tal que:

$$F^T(\hat{\theta}_0)[y - f(\hat{\theta}_0)] = \emptyset.$$

Por expansão de série de Taylor de primeira ordem pelo ponto  $\hat{\theta}_0$  tem-se:

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

Sendo assim, aplicando a aproximação anterior em  $F^T(\hat{\theta})[y - f(\hat{\theta})] = \emptyset$  obtém-se:

$$F^T(\hat{\theta})[y - f(\hat{\theta}_0) - F(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0)] = \emptyset.$$

Ao multiplicar  $F^T(\hat{\theta})^{-1}$  à esquerda de ambos os lados se tem:

$$y - f(\hat{\theta}_0) - F(\hat{\theta}_0)\hat{\theta} + F(\hat{\theta}_0)\hat{\theta}_0 \approx \emptyset, \text{ dessa forma:}$$

$F(\hat{\theta}_0)\hat{\theta} \approx F(\hat{\theta}_0)\hat{\theta}_0 + [y - f(\hat{\theta}_0)]$  agora multiplicando por  $[F(\hat{\theta}_0)]^{-1}$  a esquerda ambos os lados se têm:

$$\hat{\theta} = \hat{\theta}_0 + [F(\hat{\theta}_0)]^{-1}[y - f(\hat{\theta}_0)], \text{ agora tomando } \hat{\theta}_0 = \hat{\theta}_t \text{ e } \hat{\theta} = \hat{\theta}_{t+1} \text{ obtém-se:}$$

$\hat{\theta}_{t+1} \approx \hat{\theta}_t + [F(\hat{\theta}_t)]^{-1}[y - f(\hat{\theta}_t)]$  que é a expressão para a t-ésima iteração do processo iterativo de Gauss-Newton.

Com isto gera um processo iterativo para determinar  $\hat{\theta}$  onde se pretende que  $SQE(\hat{\theta}_1)$  seja menor que  $SQE(\hat{\theta}_2)$ , com isso  $\hat{\theta}_1$  pode ser um estimador melhor que  $\hat{\theta}_2$ .

O processo iterativo de Gauss-Newton continua até a convergência, isto é, até que:

$$\frac{|(\hat{\theta}_{j,t+1} - \hat{\theta}_{j,t})|}{\hat{\theta}_{j,t}} < \delta, u = 1, 2, \dots, p \quad (2.18)$$

Onde  $\delta$  é algum número pequeno em que a cada iteração o  $SQE(\hat{\theta}_t)$ , pode ser avaliado para garantir que foi obtida uma redução neste valor, com relação ao anterior.

Após a convergência do procedimento para um vetor  $\hat{\theta}$  de estimativas, podemos obter o quadrado médio residual dado por:  $QMR = \frac{\sum_{i=1}^n [y_i - f(x_i, \hat{\theta})]^2}{n-p}$ , que é um estimador de  $\sigma^2$ .

Não é fácil encontrar os valores iniciais dos parâmetros, para o ajuste de modelos de regressão não linear. Uma vez que escolhido mal os valores iniciais, pode necessitar de um número grande de iterações para a convergência. Assim como, pode ocorrer a convergência para um mínimo local ou não ocorrer a convergência.

Não se tem uma forma padrão para se obter os valores iniciais para as estimativas dos parâmetros, uma maneira prática é utilizar valores concordantes com as interpretações biológicas, necessitando de conhecimento prévio do fenômeno que se estuda.

Nos modelos não lineares utilizados para descrever o crescimento animal existe a assíntota do modelo que representa o parâmetro relativo ao peso que o animal atinge na maturidade, se os dados são, por exemplo, de pesos e idade de bovinos um bom valor inicial para este parâmetro é baseado no peso que o animal chega na idade adulta.

## 2.5 Qualidade do ajuste de regressão

Ao determinar a relação entre as variáveis, por meio de estudo de regressão, há o problema da definição da forma matemática que será ajustada para melhor representar o fenômeno em estudo.

Em geral, segundo (Silveira, 2010) alguns modelos são ajustados e a partir dos resultados dos avaliadores da qualidade dos modelos é escolhido a equação que melhor se adequa para descrever a relação entre as variáveis envolvidas.

Dentre as medidas estatísticas que avaliam a qualidade do ajuste de regressão, uma das mais utilizadas é o coeficiente de determinação ( $R^2$ ). Essa medida indica o quão bem um

modelo de regressão linear se ajusta aos dados observados, ou seja, ele mede proporção da variação total dos dados que é explicada pelo modelo de regressão.

### 2.5.1 Coeficiente de determinação ( $R^2$ )

O coeficiente de determinação ( $R^2$ ), é uma medida estatística que avalia a qualidade do ajuste de regressão. Essa medida indica o quão bem um modelo de regressão linear se ajusta aos dados observados, ou seja, ele mede proporção da variação total dos dados que é explicada pelo modelo de regressão.

O coeficiente de determinação ajustado ( $R_{aj}^2$ ) leva em consideração o número de parâmetros no modelo, por isso o  $R_{aj}^2$  é o mais indicado para seleção de modelos mais parcimoniosos.

É dada a seguir a expressão do  $R_{aj}^2$ :

$$R_{aj}^2 = R^2 - \left( \frac{p-1}{n-p} \right) (1 - R^2) \quad (2.19)$$

em que

$p$  é o número de parâmetros;

$n$  é o número de observações;

$R^2$  é o coeficiente de determinação dado por:

$$R^2 = \frac{\sum_{i=1}^n y_i y_i^{\wedge} - \frac{(\sum_{i=1}^n y_i) (\sum_{i=1}^n y_i^{\wedge})}{n}}{\left[ \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right] \left[ \sum_{i=1}^n y_i^{\wedge 2} - \frac{(\sum_{i=1}^n y_i^{\wedge})^2}{n} \right]} \quad (2.20)$$

$y_i$  é o peso observado;

$y_i^{\wedge}$  é o peso estimado.

O  $R^2$  varia de 0 a 1, sendo que quanto mais próximo de 1, melhor é o ajuste do modelo. No entanto, é importante ressaltar que o  $R^2$  não indica a qualidade do modelo em si, mas apenas a qualidade do ajuste aos dados.

Além disso, é importante lembrar que o coeficiente de determinação  $R^2$  não é indicado para seleção de modelos, pois quanto maior o número de parâmetros, maior tende a ser o  $R^2$ , levando a escolha de um modelo mais complexo.

Para modelos de regressão mais complexos, existem outras medidas de avaliação de ajuste como o coeficiente de determinação ajustado ( $R^2_{aj}$ ) que é mais adequado.

O coeficiente de determinação ajustado ( $R^2_{aj}$ ) leva em consideração o número de parâmetros no modelo, por isso o  $R^2_{aj}$  é o mais indicado para seleção de modelos mais parcimoniosos.

### 2.5.2 Critério de informação de Akaike (AIC)

O valor do AIC, de acordo com (Akaike, 1974), para um modelo é calculado a partir do estimador de máxima verossimilhança e do número de parâmetros utilizados pelo modelo, e quanto menor sua estimativa melhor é o modelo de regressão. É importante lembrar que o AIC é utilizado para comparar diferentes modelos e não como uma medida absoluta da qualidade do modelo. É dada a seguir a expressão do AIC:

$$AIC = -2\loglike + 2p \quad (2.21)$$

em que  $\loglike$  é o valor do logaritmo natural da função de verossimilhança dada por:

$$\loglike = \frac{-n}{2} \ln(2\pi) - \frac{-n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \theta)) \quad (2.22)$$

$p$  é o número de parâmetros do modelo.

O critério de informação de Akaike, AIC, é um avaliador utilizado para avaliar a qualidade relativa de diferentes modelos estatísticos em relação a um conjunto de dados. O AIC fornece uma estimativa da qualidade de um modelo em termos da sua capacidade de prever corretamente novos dados e quanto menor sua estimativa melhor é o modelo de regressão.

### 2.5.3 Critério de informação bayesiano

O critério de informação Bayesiano, BIC, (SCHWARZ, 1978) é uma extensão do critério de informação de Akaike (AIC) que leva em consideração o tamanho da amostra em relação ao número de parâmetros do modelo.

O BIC é calculado a partir da função de verossimilhança do modelo, do número de parâmetros e do tamanho da amostra. O BIC penaliza modelos com muitos parâmetros em relação ao tamanho da amostra, e com isso evita a escolha de modelos mais complexos que se ajustam bem aos dados.

Assim como o AIC, o modelo com menor valor de BIC é considerado o modelo com melhor relação entre ajuste e complexidade. É dada a seguir a expressão do BIC:

$$BIC = -2\loglike + pln(n) \quad (2.23)$$

em que *loglike* é o valor do logaritmo natural da função de verossimilhança dada por:

$$\loglike = \frac{-n}{2} \ln(2\pi) - \frac{-n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \theta)) \quad (2.24)$$

*p* é o número de parâmetros;

*n* é o número de observações.

Além dos avaliadores da qualidade de ajuste citados acima existem vários outros como o quadrado médio do erro, índice assintótico.

Porém, na maioria das vezes, a escolha dos melhores modelos não é tão fácil. Por causa do grande número de avaliadores de qualidade de ajustes existentes na literatura, muitas vezes há discordância entre eles para a definição do melhor modelo a ser utilizado.

Para Silveira et al., 2012, considera-se simultaneamente todos os avaliadores de qualidade utilizando análise de agrupamento a fim de agrupar modelos cujos resultados dos avaliadores de qualidade de ajuste sejam similares dentro de cada população e assim estabelecer qual modelo melhor se ajusta aos dados de crescimento.

## 2.6 Análise de Resíduos

Os resíduos representam as diferenças entre os valores observados e os valores previstos pelo modelo. Para analisar os resíduos no modelo não linear existem algumas técnicas que podem ser aplicadas.

A investigação dos resíduos é uma etapa obrigatória de qualquer análise de regressão. Se o modelo é adequado, os resíduos devem se apresentar de forma aleatória, isto é, eles não devem conter nenhum padrão evidente. Desta forma, a verificação do modelo pode ser realizada pela análise de gráficos e testes de hipóteses a respeito dos resíduos.

Os resíduos contêm informação sobre o motivo do modelo não ter se ajustado bem aos dados, além de identificar se uma ou mais suposições do modelo foram violadas

A análise de resíduos em modelos não lineares é uma etapa muito importante na validação e diagnóstico desses modelos. Ao analisar os resíduos, é possível verificar se o modelo está capturando adequadamente os padrões e estruturas presentes nos dados.

Para verificar possíveis suposições do modelo, como homogeneidade da variância (teste de homoscedasticidade), normalidade dos resíduos (teste de normalidade) e independência dos resíduos, existem vários testes estatísticos. Esses testes são usados para complementar as análises gráficas.

### 2.6.1 Teste de homocedasticidade

A suposição de homocedasticidade, em relação as variáveis explicativas, resulta na variância dos erros constante ao longo das observações. Para detectar a heterocedasticidade pode se utilizar análises gráficas e os testes estatísticos específicos como os testes de Breusch-Pagan, Teste de Bartlett, Teste da Razão de Verossimilhança.

Uma análise gráfica importante é plotar os resíduos em relação aos valores ajustados (previstos) pelo modelo. Um padrão aleatório de dispersão dos pontos em torno de zero indica que o modelo é apropriado para os dados. No entanto, se houver uma tendência (por exemplo, pontos formando uma linha reta crescente ou decrescente) ou uma forma não linear, isso pode indicar problemas de falta de ajuste.

Como consequência da heterocedasticidade, tem-se o viés do estimador da variância de  $\hat{\beta}$ . Com isso, as estatísticas de teste t e F deixam de ser válidas, pois dependem da variância do estimador, mesmo em amostras grandes.

Se a variância não for constante o estimador de Mínimos Quadrados não terá variância mínima, e os erros-padrão dos coeficientes serão maiores que no caso que se tem a variância constante.

Portanto, a falha na suposição da homocedasticidade é mais grave do que a falha na suposição da normalidade.

Soluções para a heterocedasticidade pode ser utilizar transformação de variáveis, procedimento de mínimos quadrados ponderados.

## 2.6.2 Teste de normalidade

A verificação da normalidade dos resíduos é uma suposição importante que pode ser feita por meio de várias técnicas, como os testes de hipóteses de Shapiro-Wilk, Kolmogorov-Smirnov (K-S), Anderson-Darling, dentre outros.

Pode também utilizar análise visual de gráficos de resíduos. Para a suposição que os resíduos seguem uma distribuição normal temos que os resíduos devem se distribuir em torno do zero de maneira simétrica.

Quando os resíduos são normalmente distribuídos, quer dizer que as estimativas dos parâmetros do modelo não são tendenciosas e os testes de hipóteses e os intervalos de confiança são válidos, além do mais, indica que o modelo está capturando adequadamente a estrutura dos dados. Mas, caso os resíduos não sigam uma distribuição normal é possível realizar transformações nos dados.

Para analisar a normalidade dos dados pela análise visual, é importante ressaltar o tamanho da amostra, pois em amostras grandes, desvios sutis da normalidade podem ser detectados, já em amostras pequenas, a análise visual dos dados pode não ser tão precisa.

O histograma permite verificar se os dados possuem uma forma aproximadamente simétrica (distribuição normal), e caso possua, recomenda-se que os dados são normalmente distribuídos. Mas, se houver desvios significativos da forma característica da distribuição normal, pode ser um indício da violação da normalidade.

Também pode ser empregado testes estatísticos para a verificação da normalidade dos dados, como o teste de Shapiro-Wilk, que é bastante utilizado. A hipótese nula do teste é que os dados seguem uma distribuição normal. No caso do modelo não linear o teste pode ser aplicado aos resíduos do modelo.

### 2.6.3 Independência dos Resíduos

A suposição da independência dos resíduos implica que os resíduos não estão correlacionados entre si, ou seja, cada observação fornece informações únicas e não influencia as outras observações.

Para verificar a presença de autocorrelação nos resíduos, existem alguns métodos e deve ser escolhido de acordo com o contexto específico do modelo e dos dados envolvidos.

A violação de independência e variância constante pode tornar o modelo instável e diferentes amostras podem resultar em modelos diferentes significativamente, levando a conclusões errôneas.

## 2.7 Modelo Logístico

O modelo de regressão logístico (Verhulst, 1838, 1961) não linear é muito utilizado para modelar fenômenos de crescimento como população, PIB, dentre outros. Sua característica principal é a variável dependente apresentar um crescimento inicial rápido, seguido de uma desaceleração, e então formando uma curva em formato de “S”.

Na medicina, o modelo logístico é utilizado para modelar a propagação de doenças ou a resposta dos pacientes a tratamentos.

Em econometria, permite explicar uma variável discreta, como por exemplo as intenções de voto em atos eleitorais.

Em seguradora, permite encontrar fração dos clientes que sejam sensíveis a determinada política securitária em relação a um dado risco particular;

Em instituições financeiras, pode detectar os grupos de risco para a subscrição de um crédito;

A forma geral de um modelo de regressão logístico não linear é dada por:

$$f(x) = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 x)}}$$

(2.25)

tem-se:

- $f(x)$  a variável dependente;
- $x$  é a variável independente;
- $\beta_1$  é o valor máximo que a variável  $y$  pode atingir;
- $\beta_2$  é valor da variável independente no ponto de inflexão da curva
- $\beta_3$  é a taxa de crescimento;

A estimação dos parâmetros no modelo de regressão não linear pode ser feita utilizando o método de Máxima Verossimilhança ou o método dos mínimos quadrados não lineares.

O seguinte trabalho foi considerado o modelo logístico, por se tratar de dados da Covid-19, onde se tinha um crescimento inicial rápido e logo em seguida uma desaceleração. A estimação dos parâmetros foi feita utilizando a técnica dos mínimos quadrados não lineares e o método iterativo de Gauss Newton.

### 2.7.1 Características do Gráfico do Modelo Logístico

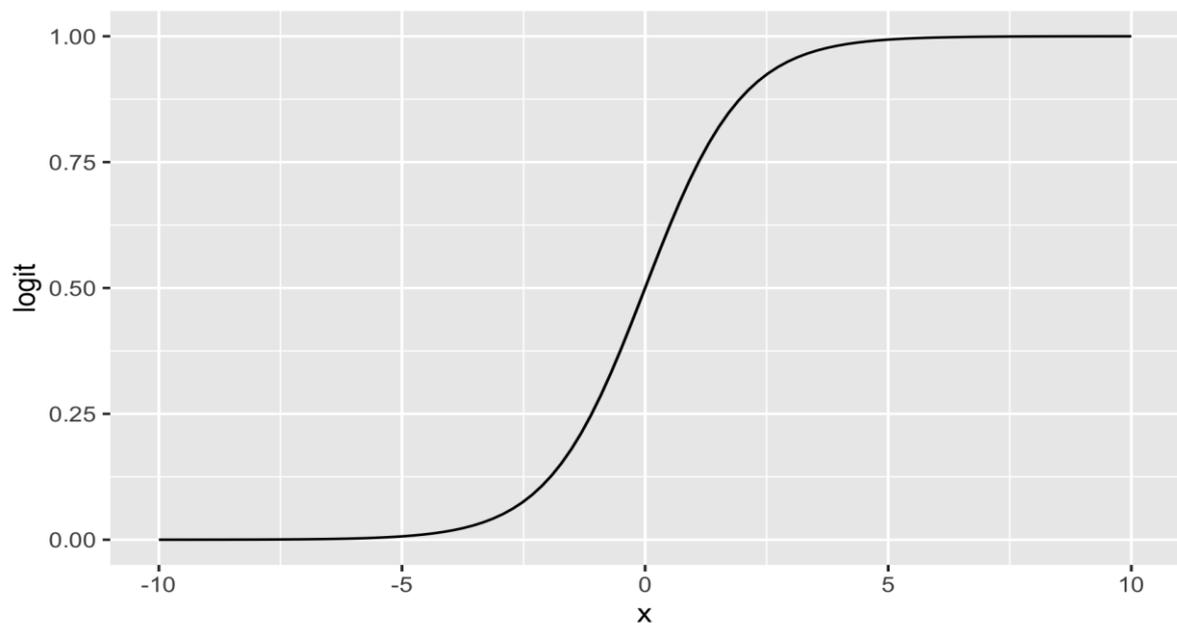


Figura 1: Gráfico do Modelo Logístico

O modelo logístico tem uma forma característica em formato de “S”. No início, há um rápido crescimento. Após esse aumento inicial, o crescimento diminui e ocasionalmente se estabiliza, e então se tem uma curva sigmoide.

Na Figura 1, existe o ponto de inflexão no gráfico, no qual a curva muda sua concavidade para cima para a concavidade para baixo. É exatamente neste ponto que a taxa de crescimento da variável dependente é máxima, ou seja, é o ponto onde a segunda derivada da função muda de sinal. Portanto, trata-se de um ponto de grande relevância para interpretação do modelo de regressão Logístico.

Na Figura 1, o modelo tem duas assíntotas horizontais. A primeira assíntota é próxima ao valor mínimo da variável dependente, e a segunda é próxima ao valor máximo, que é conhecido como a capacidade da carga do sistema, na equação logística é representado por  $\beta_1$ .

A taxa de crescimento, representada por  $\beta_3$ , atinge a inclinação da curva logo no início. Já o ponto de inflexão, representada por  $\beta_2$ , é influenciado pela localização da variável independente. E é a capacidade de carga que define a assíntota superior da curva.

## 2.8 Modelo de Richards

O modelo de Richards (Richards, 1959), também chamado de “Modelo Logístico Generalizado”, esse modelo é uma extensão do modelo logístico. Ele é bastante utilizado para descrever o crescimento de populações biológicas ou o desenvolvimento de fenômenos em que as curvas de crescimento exibem padrões sigmóides.

O modelo de Richards é utilizado quando os dados de crescimento exibem uma fase inicial de crescimento rápido, seguida por uma fase de desaceleração até atingir um limite superior. Esse modelo tem uma flexibilidade que permite capturar uma variedade de padrões de crescimento, tornando-o um modelo bastante valioso.

Esse modelo é utilizado em biologia, ecologia, agronomia, entre outras disciplinas, para descrever padrões de crescimento complexos.

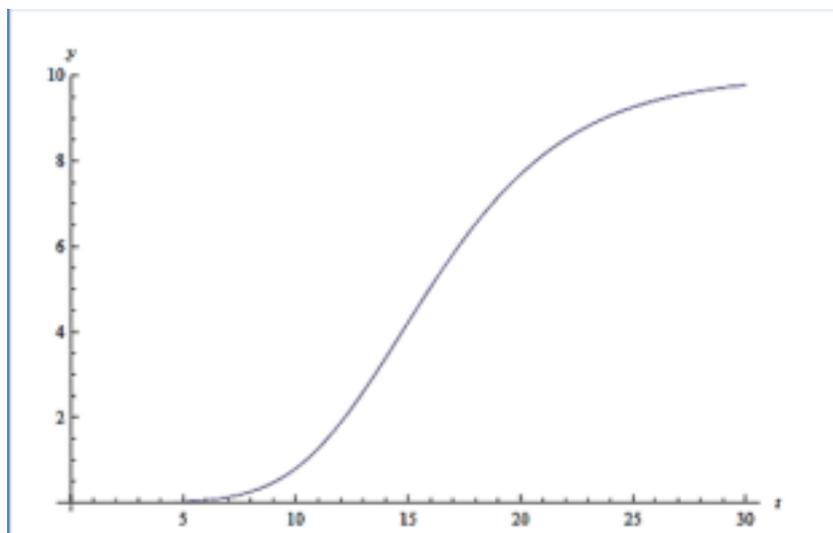
A equação do modelo de Richards é dada por:

$$f(x) = \beta_1 + \frac{(\beta_2 - \beta_1)}{(1 + \beta_3 \cdot \exp(-\beta_4 \cdot (x - \beta_5)))^{\frac{1}{\beta_6}}}$$

(2.26)

A seguir estão os parâmetros e seus significados:

- $f(x)$  a variável dependente;
- $x$ : A variável independente,
- $\beta_1$  : Limite inferior assintótico da curva, que representa o valor final que a função atinge,
- $\beta_2$  : Limite superior assintótico da curva, que representa o valor máximo que a função pode atingir,
- $\beta_3$  : Influência da inclinação inicial da curva,
- $\beta_4$  : Taxa de crescimento ou declínio inicial,
- $\beta_5$  : Ponto de inflexão, onde a taxa de crescimento muda,
- $\beta_6$  : Parâmetro de curvatura que controla a forma da curva.



*Figura 2: Gráfico do Modelo de Richards*

O comportamento específico do gráfico pode ser modificado ajustando os parâmetros da equação de Richards. Parâmetros como taxa de crescimento, ponto de inflexão e limites superior e inferior influenciam a forma da curva.

Pequenas mudanças nos parâmetros do modelo podem ter impactos significativos na forma da curva, destacando a sensibilidade do modelo às condições iniciais e aos valores dos parâmetros.

### 3. MATERIAL E MÉTODOS

#### 3.1 Origem dos Dados

Os dados utilizados se referem ao número de casos da COVID-19 no Brasil (variável dependente) ao longo dos primeiros 750 dias (variável independente). Foi considerado como primeiro dia, 3 de janeiro de 2020.

Os dados foram retirados do site da Organização Mundial da Saúde (OMS); disponível em: <https://data.who.int/dashboards/covid19/data?n=c>; acessado em janeiro de 2024. A OMS é uma organização internacional que fornece informações e dados sobre a Covid-19 do mundo todo.

A partir dessa base foi filtrado somente os dados do Brasil. Somente as variáveis data e casos acumulados foram mantidos no banco de dados e a variável data foi transformada em dias. Todos os aspectos relacionados à computação foram abordados por meio do programa R, e as instruções correspondentes estão disponíveis no Apêndice A.

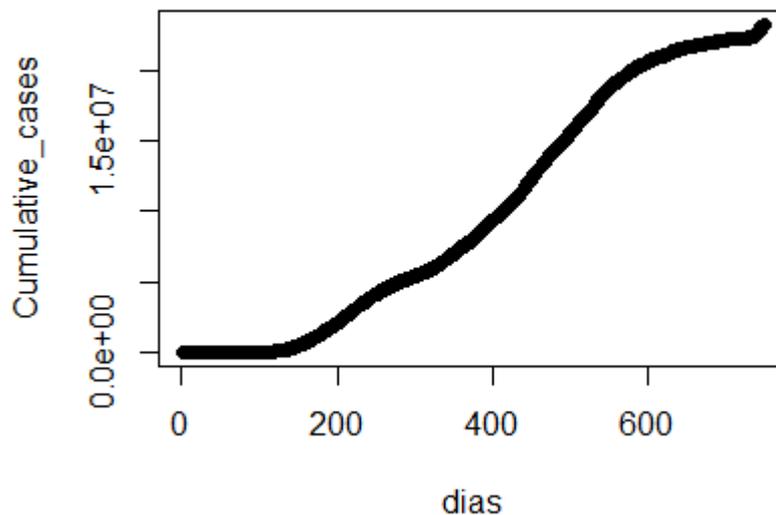


Figura 3: Diagrama de dispersão dos dados

Os modelos de regressão não linear utilizados foram o modelo logístico e o de Richards para descrever a progressão da doença COVID-19, doença causada por um novo coronavírus denominado SARS-CoV-2, ao longo dos dias (Tabela 2).

Tabela 2: Modelos de Crescimento Logístico e Richards

Modelo	Função
Richards	$f(x) = \beta_1 + \frac{(\beta_2 - \beta_1)}{(1 + \beta_3 \cdot \exp(-\beta_4 \cdot (x - \beta_5)))^{\frac{1}{\beta_6}}}$
Logístico	$f(x) = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 x)}}$

Ambos os modelos tinham como objetivo descrever os números de casos da COVID-19 no Brasil, com o auxílio de modelos de regressão não lineares. Neste estudo foram feitas tentativas de capturar padrões mais complexos, quais foram os picos da doença, a taxa de crescimento da curva, entre outros.

Os parâmetros desses modelos foram estimados utilizando o método dos quadrados mínimos ordinários por meio do processo iterativo de Gauss-Newton, que em geral esse método utilizado pelo software R incorporam rotinas adicionais de otimização numérica, diferentemente dos passos descritos neste trabalho sobre o processo iterativo de Gauss-Newton, maiores informações nas publicações de AGUIAR (2012) e de SILVA et al. (2019). Nesse método, são fornecidas as estimativas iniciais dos parâmetros e a forma da função  $f(X, \theta)$ , utilizando o software R (R Team, 1995).

## 4. RESULTADOS E DISCUSSÕES

### 4.1 Modelo Logístico

Neste trabalho um dos modelos realizados foi o modelo logístico ajustado aos dados da Covid-19, onde casos acumulados é a variável dependente e dias a variável explicativa (ou independente).

O modelo logístico, foi definido pela equação (4.1), e foi ajustado aos dados.

$$f(x) = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 x)}} \quad (4.1)$$

A estimativa dos parâmetros segue na Tabela 3.

Tabela 3: Tabela de Coeficientes de Regressão do modelo Logístico

	Estimativa	Erro Padrão	Valor t	Pr(> t )
$\beta_1$	$2,391 \times 10^7$	$8,178 \times 10^4$	292,4	<2e-16 ***
$\beta_2$	4,450	$2,741 \times 10^{-2}$	162,3	<2e-16 ***
$\beta_3$	$1,019 \times 10^{-2}$	$7,700 \times 10^{-5}$	132,4	<2e-16 ***

A estimativa do parâmetro  $\beta_1$  igual a 23.910.000 como limite máximo esperado para os casos acumulados da COVID-19.

Para  $\beta_2$  está relacionado ao intercepto, ou seja, espera-se que os casos iniciais sejam por volta de 4,450.

Já para  $\beta_3$ , tem se o parâmetro 0,01019, que está relacionado à taxa média de crescimento da curva, ou seja, a taxa de crescimento é positiva (0,01).

Analisou se cada um dos  $\beta$  individualmente. Para isso, utilizou o teste de hipóteses com a hipótese nula sendo  $\beta$  estatisticamente igual a zero e hipótese alternativa se  $\beta$  for estatisticamente diferente de zero. Todos os parâmetros do modelo foram significativos (Tabela 4).

Tabela 4: Análise da Anova em relação aos dados mencionados

	GL	SQ	QM	F	Pr (>F)
Regressão	2	$5,216085 \times 10^{16}$	$2,608042 \times 10^{16}$	98202,06	0
Resíduo	746	$1,981221 \times 10^{14}$	$2,655792 \times 10^{11}$		

O valor do coeficiente de determinação ( $R^2$ ) foi bem alto (0,9962161), indicando que o modelo se ajustou bem aos dados (Tabela 4).

Posteriormente foi realizado a análise de resíduos do modelo afim de verificar se o modelo foi adequado, ou seja, as suposições foram satisfeitas.

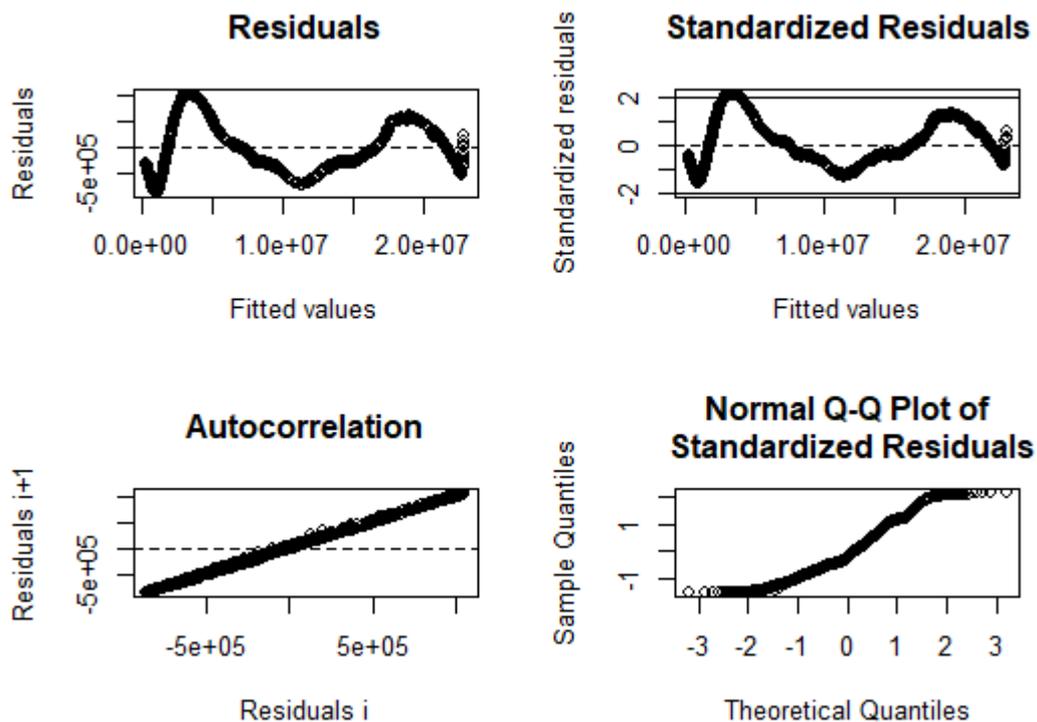


Figura 4: Análise de Resíduos do modelo logístico

Porém, os critérios da análise de resíduos, como normalidade, variância e independência não foram satisfeitas como pode se observar na Figura 4.

Na tentativa de solucionar o problema de variância não constante, foram feitas algumas transformações na variável resposta, mas a transformação que melhor se adequou aos dados foi a transformação cúbica e então seguimos com a análise do modelo Logístico após ter feito a transformação cúbica na variável resposta.

Na Figura 5, segue o gráfico de dispersão dos dados após a transformação. Podemos notar que a curva ficou mais parecida com o formato de “S” que é característico do modelo Logístico.

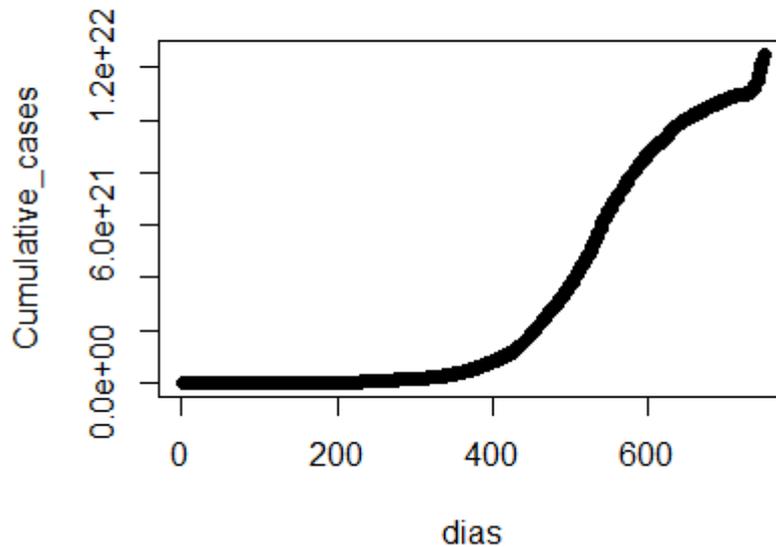


Figura 5: Gráfico de dispersão dos dados após a transformação

Na Tabela 5 seguem as estimativas dos coeficientes do modelo Logístico aos dados transformados.

Tabela 5 : Tabela de Coeficientes de Regressão do modelo Logístico após a transformação

	Estimativa	Erro Padrão	Valor t	Pr(> t )
$\beta_1$	$1,140 \times 10^{22}$	$1,760 \times 10^{19}$	647,8	<2e-16 ***
$\beta_2$	9,997	$3,950 \times 10^{-2}$	253,1	<2e-16 ***
$\beta_3$	$1,861 \times 10^{-2}$	$7,890 \times 10^{-5}$	235,8	<2e-16 ***

Para o modelo logístico ajustado acima, temos  $\beta_1$  igual a  $1,140 \times 10^{22}$  como limite máximo esperado para casos acumulados, porém como foi feita uma transformação então  $\beta_1$  é igual a  $\sqrt[3]{1,140 \times 10^{22}} = 22.508.573$ , ou seja, o número máximo de casos acumulados da COVID-19 no Brasil é esperado como 22.508.573 durante o período avaliado.

Para  $\beta_2$  está relacionado ao intercepto, ou seja, o número de casos iniciais é por volta de 9,997, para voltar para a escala original basta fazer  $\sqrt[3]{9,997} \approx 2$ .

Já para  $\beta_3$ , tem se o parâmetro igual a 0.01861, e para voltar para a escala original basta fazer  $\sqrt[3]{0,01861} \approx 0,2650$ , que está relacionado à taxa média de crescimento da curva, ou seja, a taxa de crescimento é positiva (0,2650).

Com estes parâmetros, foi possível ajustar a curva aos dados observados e para analisar a significância do modelo foi feita uma Análise de Variância (ANOVA), cuja técnica é utilizada para testar a significância do modelo bem como a significância individual de cada parâmetro do modelo, analisando a variação explicada pelos diferentes termos do modelo em relação à variação não explicada.

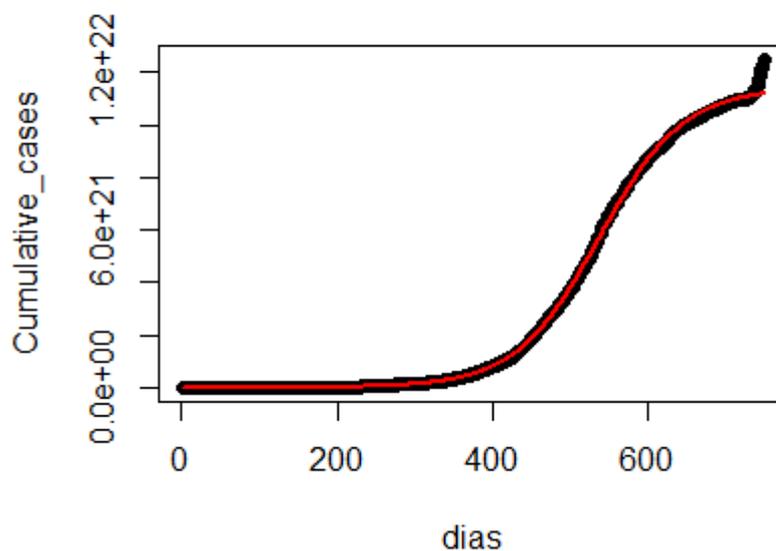


Figura 6: Curva do modelo logístico ajustada aos dados transformados

Na

Tabela 6 estão os resultados para a Análise de Variância do modelo. A estatística F é usada para testar a significância global do modelo. Analisando as estatísticas F desta ANOVA, um valor relativamente baixo, observa-se que pelo menos um dos coeficientes na regressão é significativamente diferente de zero, sugerindo que o modelo global é significativo.

Tabela 6: Análise da Anova em relação aos dados transformados

	GL	SQ	QM	F	Pr(>F)
Regressão	2	$1,298898 \times 10^{46}$	$6,494492 \times 10^{45}$	485618,9	0
Resíduo	746	$9,976735 \times 10^{42}$	$1,337364 \times 10^{40}$		

Na Tabela 6, pode se observar que o valor do coeficiente de determinação ( $R^2$ ) foi bem alto, aproximadamente 0,999 indicando um bom ajuste do modelo.

$$R^2 = 0,9992325$$

Temos os indicadores de qualidade do modelo logístico, como o AIC igual a 71333,75 e o BIC igual a 71352,2 que é uma extensão do critério de informação de Akaike (AIC). Após a estimação dos parâmetros do modelo de Richards será feito uma comparação dos valores do AIC e do BIC para identificar o modelo que melhor se ajustou aos dados.

A Análise de Resíduos é uma etapa muito importante na avaliação da qualidade do ajuste de um modelo estatístico.

a)

b)

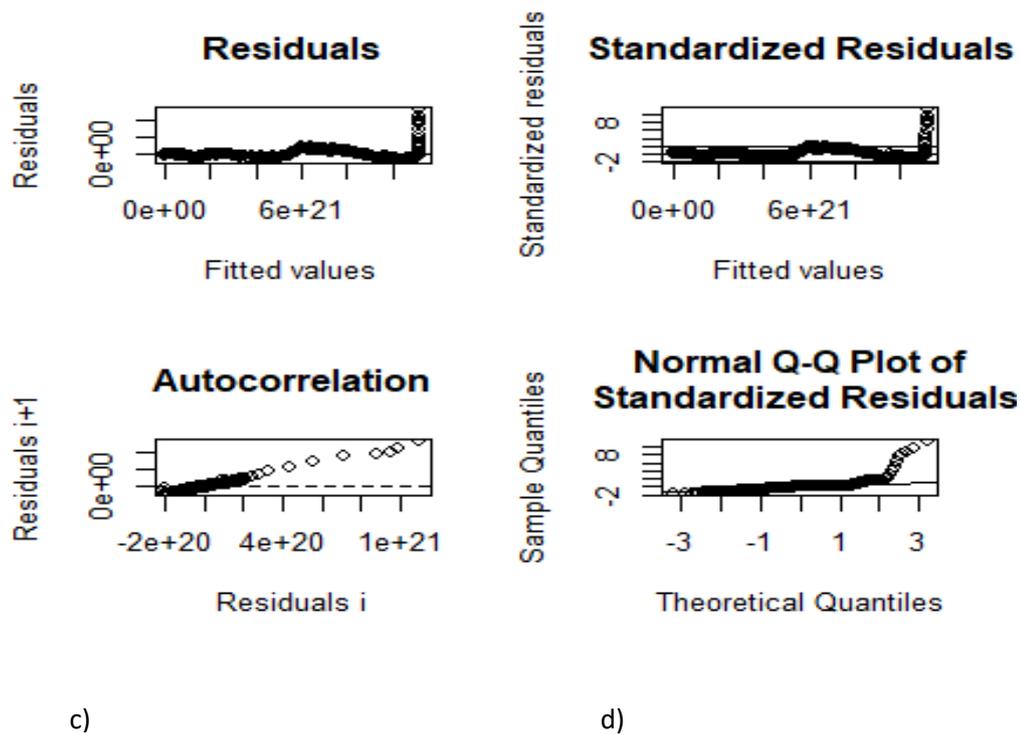


Figura 7: Análise de Resíduos do modelo logístico

A Figura 7 a) e b) mostram os resíduos versus os valores ajustados do modelo. Este gráfico é utilizado para verificar se a suposição de homocedasticidade dos erros é válida. Nota-se que esse gráfico segue uma tendência, logo não é uma nuvem aleatória, nesta sequência o critério de homocedasticidade não foi atendido.

A Figura 7 c) mostra resíduos por ordem. Esse gráfico é utilizado para verificar se a suposição de independência dos erros é válida. Nota-se que esse gráfico não parece ser uma nuvem aleatória em torno de zero, ou seja, é possível identificar uma tendência. Com isso podemos considerar que a suposição de independência não foi atendida.

Por fim, na Figura 7 o último gráfico d) compara os quantis teóricos da distribuição normal com os quantis empíricos obtidos a partir dos resíduos e nota-se que os pontos não estão aderentes à reta que faz com que a suposição de normalidade dos erros não seja aceita.

Também foi utilizado o teste de Shapiro-Wilk para verificar se os dados seguem uma distribuição normal. Como no caso o modelo é não linear, o teste foi aplicado aos resíduos do modelo. A hipótese nula desse teste é que a população está normalmente distribuída. Assim, como o valor de p foi menor que  $2,2 \times 10^{-16}$ , um valor bem próximo de zero, a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos como apontou a análise gráfica.

## 4.2 Modelo de Richards

Outro modelo utilizado neste trabalho foi o modelo de Richards também ajustado aos dados da Covid-19, onde casos acumulados é a variável dependente e dias a variável explicativa (ou independente).

O modelo de Richards, foi definido pela equação (4.2).

$$f(x) = \beta_1 + \frac{(\beta_2 - \beta_1)}{(1 + \beta_3 \cdot \exp(-\beta_4 \cdot (x - \beta_5)))^{\frac{1}{\beta_6}}}$$
(4.2)

Também não foram satisfeitas as suposições na análise de resíduos no modelo de Richards. Com isso, também foi feita uma transformação cúbica aos dados na tentativa de encontrar os parâmetros ideais. Foi através dessa transformação que foram encontrados a estimativa dos parâmetros.

Tabela 7: Tabela de Coeficientes de Regressão para o modelo de Richards

	Estimativa	Erro Padrão	Valor t	Pr(> t )
$\beta_1$	$2,957 \times 10^{20}$	$4,941 \times 10^{19}$	5,985	3,37e-09 ***
$\beta_2$	$9,227 \times 10^{21}$	$7,257 \times 10^{19}$	127,149	< 2e-16 ***
$\beta_3$	$9,410 \times 10^{21}$	$2,099 \times 10^{27}$	0,000	1,00000
$\beta_4$	$9,599 \times 10^{-2}$	$2,125 \times 10^{-2}$	4,516	7,31e-06 ***
$\beta_5$	-5,776	$2,324 \times 10^6$	0,000	1,00000
$\beta_6$	2,324	$8,015 \times 10^{-1}$	2,900	0,00385 **

Para o modelo de Richards ajustado aos dados mencionados, temos  $\beta_1$  igual a  $2,957 \times 10^{20}$ , para voltar para a escala original basta fazer  $\sqrt[3]{2,957 \times 10^{20}} \approx 6.662.191$  casos acumulados que representa o limite inferior assintótico da curva, onde é o valor final que a função atinge.

Para  $\beta_2$  tem-se  $9,227 \times 10^{21}$ , e voltando a escala original temos  $\sqrt[3]{9,227 \times 10^{21}} \approx 20.974.269$  casos acumulados que representa o limite superior assintótico da curva, onde é o valor máximo que a função pode atingir.

Para  $\beta_3$ , tem se a estimativa do parâmetro igual a  $9,410 \times 10^{21}$ , e voltando a escala original temos  $\sqrt[3]{9,410 \times 10^{21}} \approx 21.112.024$  que está relacionado à inclinação inicial da curva.

A estimativa do parâmetro  $\beta_4$  é igual 0,09599, e voltando para a escala original temos  $\sqrt[3]{0,09599} \approx 0,4578698$  que está relacionado a taxa de crescimento do número de casos.

O valor de  $\beta_5$  estimado é igual a -5,776, voltando para a escala original temos  $\sqrt[3]{-5,776} \approx -1,79$  e este é o ponto de inflexão do modelo.

Por último temos o  $\beta_6$  igual a 2,324, voltando para a escala original temos  $\sqrt[3]{2,324} \approx 1,324582$  que é o parâmetro de curvatura que controla a forma da curva.

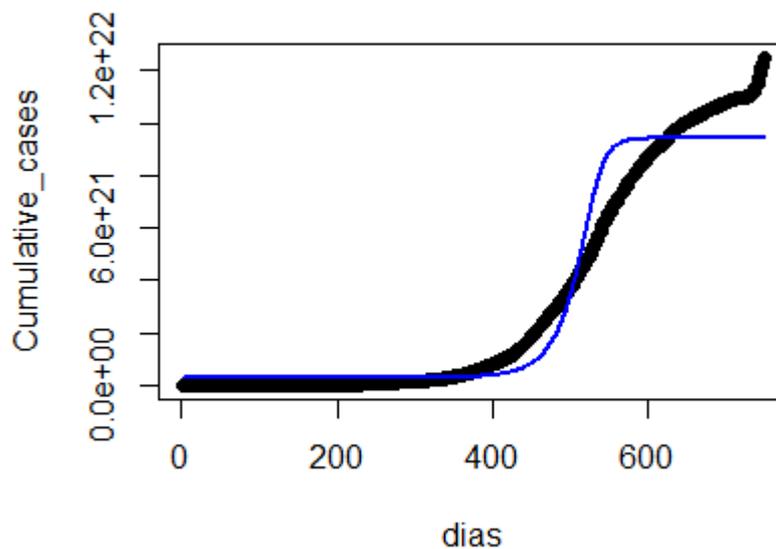


Figura 8: Curva do modelo de Richards ajustada aos dados

Com a estimação destes parâmetros foi possível ajustar a curva aos dados observados e para analisar a significância dos coeficientes do modelo foi feita uma Análise de Variância (ANOVA), cuja técnica é utilizada para testar a significância do modelo bem como a significância individual de cada parâmetro do modelo, analisando a variação explicada pelos diferentes termos do modelo em relação à variação não explicada.

Tabela 8: Análise da Anova do modelo de Richards

	GL	SQ	QM	F	Pr(>F)
Regressão	5	$1.2262466 \times 10^{46}$	$2.452491 \times 10^{45}$	2474.119	0

Resíduo	743	$7.365051 \times 10^{44}$	$9.912585 \times 10^{41}$
---------	-----	---------------------------	---------------------------

$$R^2 = 0,9433412$$

Pela Tabela 8 podemos ver que o p-valor é significativo, ou seja, há pelo menos um coeficiente do modelo significativo e, portanto, o modelo como um todo é considerado significativo. O coeficiente de determinação ( $R^2$ ) foi, aproximadamente 0,94 que é um valor bom, porém o  $R^2$  do modelo logístico foi maior (0,999).

Temos os indicadores de qualidade do modelo de Richards, como o AIC igual a 74561,69, lembrando que o valor do AIC é utilizado para comparação de modelos e não como um avaliador de qualidade do modelo, e o BIC igual a 74594,02.

*Como no modelo logístico, as suposições de homoscedasticidade, independência e normalidade também não foram atendidas no modelo de Richards como pode se observar na*

Figura 9.

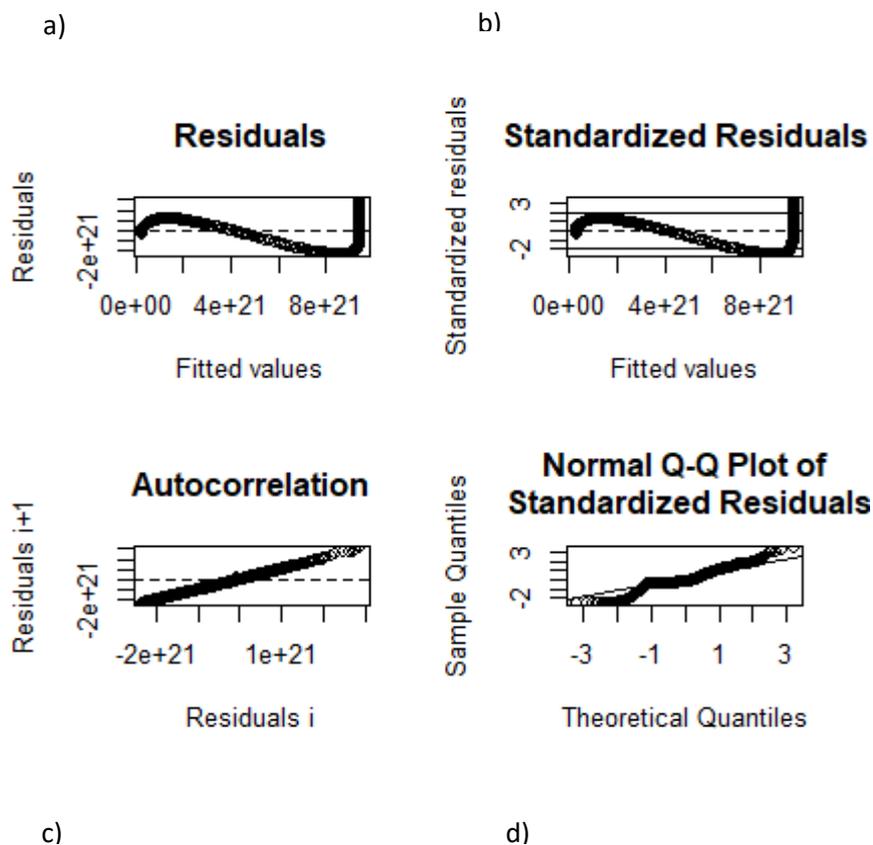


Figura 9: Análise de Resíduos do modelo de Richards

Então, como o valor de  $p$  foi bem próximo de zero (**Erro! Fonte de referência não encontrada.**), a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos como apontou a análise gráfica.

Como no modelo logístico, também foi utilizado o teste de Shapiro-Wilk para verificar se os dados seguem uma distribuição normal. Assim, como o valor de  $p$  foi menor do que  $2,2 \times 10^{-16}$ , um valor bem próximo de zero, a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos como apontou a análise gráfica.

### 4.3 Comparação dos modelos

Ao realizar a análise dos modelos logístico e de Richards podemos observar através dos estimadores de qualidade como  $R^2$ , AIC e BIC que o modelo logístico se ajustou melhor aos dados.

Comparando os valores do AIC e BIC do modelo de Richards com o modelo logístico, como pode observar na Tabela 9, o modelo que melhor se ajustou aos dados foi o modelo logístico. É importante lembrar que o AIC e o BIC são utilizados para comparar diferentes modelos, quanto menor a sua estimativa, melhor o ajuste.

Tabela 9: Comparação dos valores do AIC e BIC nos modelos

Modelo	AIC	BIC
Logístico	71333,75	71352,22
Richards	74561,69	74594,02

Então podemos concluir analisando também os valores do coeficiente de determinação ( $R^2$ ) em ambos os modelos que o modelo logístico ( $R^2 = 0,9992325$ ) se ajustou melhor aos dados, pois o valor do  $R^2$  foi maior do que o modelo de Richards ( $R^2 = 0,9433412$ ).

## 5. RESULTADOS E CONCLUSÕES

Este trabalho teve como propósito apresentar uma análise de regressão não linear sobre os dados da COVID-19, empregando dois modelos de crescimento para estimar os casos acumulados ao longo do tempo, o modelo Logístico e o modelo de Richards. Após ter a seleção do chute inicial utilizou-se o Método dos Mínimos Quadrados, uma técnica de otimização matemática, utilizando o algoritmo Gauss-Newton. A avaliação da qualidade dos ajustes revelou que o modelo Logístico superou o modelo de Richards.

Em uma tentativa de estabilizar a variância, foi realizada uma transformação na variável dependente ( $y^3$ ), porém esta abordagem não foi bem sucedida, sugerindo a presença de possível heterocedasticidade em ambos os modelos.

O modelo logístico se ajustou melhor aos dados em relação ao modelo de Richards. A superioridade do modelo Logístico evidenciou-se não apenas pela excelente adaptação da curva, mas também pelos indicadores. O valor do coeficiente de determinação  $R^2$  foi mais elevado, enquanto os critérios AIC e o BIC apresentaram valores menores, todos apontando para o modelo Logístico como melhor escolha.

Embora o modelo Logístico tenha se ajustado melhor aos dados, a análise de resíduos indicou que as suposições de variância, normalidade e independência não foram plenamente satisfeitas.

Para trabalhos futuros, recomenda-se a incorporação de outra transformação de variável a fim de se obter as suposições para os modelos de regressão.

## 6. REFERÊNCIAS

AGUIAR, A.A. Análise de convergência local do método de Gauss-Newton sob condição Lipschitz. Goiânia, 2012. 47p. (Monografia de Especialização). Instituto de Matemática e Estatística, Universidade Federal de Goiás.

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

Instituto Butantan. (2022, 17 de janeiro). Vacinação contra COVID-19 no Brasil completa 1 ano com grande impacto da CoronaVac na redução de hospitalizações e mortes. Recuperado de <https://butantan.gov.br/noticias/vacinacao-contracovid-19-no-brasil-completa-1-ano-com-grande-impacto-da-coronavac-na-reducao-de-hospitalizacoes-e-mortes#:~:text=O%20avan%C3%A7o%20da%20vacina%C3%A7%C3%A3o%20no,et%C3%A1ria%20que%20recebeu%20a%20CoronaVac>.

Gujarati, D.N., & Porter, D.C. (2011). *Econometria Básica*. LTDA.

MDPI. (2022). Título do Artigo. *International Journal of Environmental Research and Public Health*, 19(21), 13902. <https://www.mdpi.com/1660-4601/19/21/13902>.

Organização Mundial da Saúde [OMS]. (2023). OMS declara fim da emergência de saúde pública de importância internacional. Recuperado de <https://www.paho.org/pt/noticias/5-5-2023-oms-declara-fim-da-emergencia-saude-publica-importancia-internacional-referente>.

Organização Pan-Americana da Saúde (OPAS). Histórico da pandemia COVID-19. Recuperado de <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>.

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

R Core Team. (1995). R: A Linguagem de Programação e Ambiente de Desenvolvimento para Análise Estatística. Disponível em: <https://www.r-project.org/>.

Richards, FJ (1959). Um crescimento flexível funções para uso empírico. *J. Exp. Robô.*, 10: 290 -300.

SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v.6, p.461-464, 1978.

SEBER, G.; WILD, C. *Nonlinear Regression*. [S.l.]: Wiley, 2003. (Wiley Series in Probability and Statistics). ISBN 9780471471356.

SILVA, E.M.; FRUHAUF, A.C.; FERNANDES, F.A.; PAULA, G.S.; MUNIZ, J.A.; FERNANDES, T.J. Método de Newton e Gauss-Newton na estimação dos parâmetros de modelo de regressão não linear. *Sigmae*, v.8, p.728-734, 2019.

SILVEIRA, F.G. Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte, 2010, 59f. Dissertação (Mestrado) Universidade Federal de Viçosa– Departamento de Estatística, Viçosa – Minas Gerais.

SILVEIRA, F.G.; SILVA, F.F.; CARNEIRO, P.L.S.; MALHADO, C.H.M. Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte. *Revista Brasileira de Saúde e Produção Animal* [online], v.13, n1, p.62-73, 2012.

SOUZA, G.S. Introdução aos modelos de regressão linear e não-linear. Brasília: Embrapa-SPI, 1998. 505p.

Verhulst, P. (1838). Aviso sobre a lei que o a população continua a aumentar. *Corresp. Matemática. Física*. 10, 113–121.

## APÊNDICES

APÊNDICE A – Rotinas utilizadas no software RStudio considerando os modelos Logístico e de Richards.

```
library(lattice)
library(latticeExtra)
library(car)
library(alr4)
library(nlstools)
library(nls2)
library(rootSolve)
library(readxl)
library(devtools)
install_github(repo = "walmes/wzRfun", ref = "master")
library(wzRfun)
library(tidyverse)
library(manipulate)
library(growthmodels)
library(minpack.lm)
generalisedRichard
richard

dados1 = read_excel('dados.xlsx')
dados1 = dados1 %>% filter(dias < 750 )
str(dados1)
summary(dados1)

par(mfrow = c(1,1))
plot(Cumulative_cases~dias, dados1, pch = 19)

# Modelo logístico
logi <- function(x, b1, b2, b3){
```

```

    b1 / (1 + exp(b2 - b3 * x))
  }
start=list()
manipulate({
  plot(Cumulative_cases~dias,data = dados1)
  curve(logi(x, b1=b1,b2=b2,b3=b3),add=TRUE, col='red')
  start<-list(b1=b1,b2=b2,b3=b3)},
  b1=slider(1,max(dados1$Cumulative_cases),initial=max(dados1$Cumulative_cases)),
  b2=slider(5, 100,initial=19, step = 0.01),
  b3=slider(0,1.5,initial=0.2, step = 0.001)
)
Start

n0 <- nls(Cumulative_cases ~ (b1 / (1 + exp(b2 - b3 * dias))),
  data = dados1,
  start = start)
summary(n0)
as.list(coef(n0))

plot(Cumulative_cases~dias, dados1, pch = 19)
lines(logi(dados1$dias, b1=as.list(coef(n0))$b1,b2=as.list(coef(n0))$b2,b3=as.list(coef(n0))$b3), col =
'red', lwd = 2)

## residuos
R2nls(n0)
res_n0 = residuals(n0)
shapiro.test(res_n0)
res_n01 = nlsResiduals(n0)
AIC(n0)
BIC(n0)
plot(res_n01)

```

```

### Transformação cúbica da variável casos cumulativos

dados1 = read_excel('dados.xlsx')

dados1 = dados1 %>% filter(dias < 750 )

dados1$Cumulative_cases = (dados1$Cumulative_cases)^3

str(dados1)

summary(dados1)

par(mfrow = c(1,1))

plot(Cumulative_cases~dias, dados1, pch = 19)

# Modelo logístico

logi <- function(x, b1, b2, b3){
  b1 / (1 + exp(b2 - b3 * x))
}

start=list()

manipulate({
  plot(Cumulative_cases~dias,data = dados1)
  curve(logi(x, b1=b1,b2=b2,b3=b3),add=TRUE, col='red')
  start<<-list(b1=b1,b2=b2,b3=b3)},
  b1=slider(1,max(dados1$Cumulative_cases),initial=max(dados1$Cumulative_cases)),
  b2=slider(5, 100,initial=19, step = 0.01),
  b3=slider(0,1.5,initial=0.2, step = 0.001)
)

start

### outro modelo

modRic = function(x,b1,b2,b3,b4,b5,b6){
  b1 + (b2 - b1) / (1 + b3 * exp(-b4 * (x - b5)))^(1/b6)
}

startModRic=list()

```

```

manipulate({
  plot(Cumulative_cases~dias,data = dados1)
  curve(modRic(x, b1=b1,b2=b2,b3=b3,b4=b4,b5=b5,b6=b6),add=TRUE, col='red')
  startModRic<-list(b1=b1,b2=b2,b3=b3,b4=b4,b5=b5,b6=b6)},
  b1=slider(0,max(dados1$Cumulative_cases),initial=0),
  b2=slider(1,max(dados1$Cumulative_cases),initial=max(dados1$Cumulative_cases)),
  b3=slider(0, max(dados1$Cumulative_cases),initial=1, step = 0.01),
  b4=slider(-1,10,initial=0.2, step = 0.001),
  b5=slider(0,60,initial = 0,step = 1),
  b6=slider(-1,10,initial = 0.1,step = 0.01)
)
startModRic

n0 <- nls(Cumulative_cases ~ (b1 / (1 + exp(b2 - b3 * dias))),
  data = dados1,
  start = start)
summary(n0)
as.list(coef(n0))

n2 = nlsLM(Cumulative_cases ~ (b1 + (b2 - b1) / (1 + b3 * exp(-b4 * (dias - b5)))^(1/b6)),
  data = dados1,
  start = startModRic,
  control = list(maxiter=100, maxfev = 100))
summary(n2)
as.list(coef(n2))

plot(Cumulative_cases~dias, dados1, pch = 19)
lines(modRic(dados1$dias, b1=as.list(coef(n2))$b1,b2=as.list(coef(n2))$b2,b3=as.list(coef(n2))$b3,
  b4=as.list(coef(n2))$b4, b5=as.list(coef(n2))$b5, b6=as.list(coef(n2))$b6), col = 'blue', lwd = 2)
lines(logi(dados1$dias, b1=as.list(coef(n0))$b1,b2=as.list(coef(n0))$b2,b3=as.list(coef(n0))$b3), col =
'red', lwd = 2)

```

```
## residuos  
R2nls(n0)  
R2nls(n2)  
res_n0 = residuals(n0)  
shapiro.test(res_n0)  
res_n01 = nlsResiduals(n0)  
plot(res_n01)  
res_n2 = residuals(n2)  
shapiro.test(res_n2)  
res_n02 = nlsResiduals(n2)  
plot(res_n02)
```