

## O modelo normal não linear

### The nonlinear normal model

### El modelo normal no lineal

Recebido: 23/01/2022 | Revisado: 02/09/2022 | Aceito: 04/09/2022 | Publicado: 11/09/2022

**Joeziley Ambrózio da Fonseca**

ORCID: <https://orcid.org/0000-0002-5658-5912>  
Universidade Federal de Campina Grande, Brasil  
E-mail: [joeziley@gmail.com](mailto:joeziley@gmail.com)

#### Resumo

Quando nos referimos à análise de dados, existe uma gama de modelos que se ajustam aos dados, como modelos de regressão linear e também modelos de regressão não linear. Na estatística a regressão não linear é uma forma da análise de regressão, em que inicialmente uma certa função matemática é ajustada aos dados, é importante ressaltar que esse ajustamento pode ser feito por meio qualitativo usando um diagrama de dispersão. O objetivo principal deste artigo será abordar a técnica de regressão normal não linear, onde será abordado alguns métodos para a estimação dos parâmetros do modelo normal não linear, a precisão dos estimadores, intervalos de confiança, testes de hipóteses, alguns resultados assintóticos e a estimação da variância. Por fim, será analisado um conjunto de dados referente a Coelhos Europeus para demonstrar a aplicabilidade do modelo não linear, em que inicialmente tentamos ajustar o conjunto de dados ao modelo normal linear onde não é obtido êxito no ajustamento e em seguida é considerado o modelo normal não linear, que foi o modelo que melhor se ajustou aos dados.

**Palavras-chave:** Regressão normal não linear; Estimação de parâmetros; Intervalos de confiança; Testes de hipóteses; Modelos não lineares.

#### Abstract

When referring to data analysis, there is a range of models that fit the data, such as linear regression models and also non-linear regression models. In statistics, nonlinear regression is a form of regression analysis, in which initially a certain mathematical function is adjusted to the data, it is important to emphasize that this adjustment can be made qualitatively using a scatter diagram. The main objective of this article will be to approach the nonlinear normal regression technique, where some methods for estimating the parameters of the nonlinear normal model, the precision the estimators, confidence intervals, hypothesis tests, some asymptotic results and the estimation of variance. Finally, a set of data regarding European Rabbits will be analyzed to demonstrate the applicability of the non-linear model, in which initially we try to adjust the data set to the linear normal model where the adjustment is not successful and then the non-linear normal model is considered, which was the model that best fit the data.

**Keywords:** Normal non-linear regression; Estimation of parameters; Confidence intervals; Hypothesis tests; Nonlinear models.

#### Resumen

Cuando se hace referencia al análisis de datos, existe una gama de modelos que se ajustan a los datos, como los modelos de regresión lineal y también los modelos de regresión no lineal. En estadística, la regresión no lineal es una forma de análisis de regresión, en la que inicialmente se ajusta una determinada función matemática a los datos, es importante enfatizar que este ajuste se puede realizar cualitativamente mediante un diagrama de dispersión. El objetivo principal de este artículo será abordar la técnica de regresión normal no lineal, donde se encuentran algunos métodos para estimar los parámetros del modelo normal no lineal, la precisión de los estimadores, los intervalos de confianza, las pruebas de hipótesis, algunos resultados asintóticos y la estimación de la varianza. Por último, se analizará un conjunto de datos sobre conejos europeos para demostrar la aplicabilidad del modelo no lineal, en el cual inicialmente tratamos de ajustar el conjunto de datos al modelo lineal normal donde el ajuste no es exitoso y luego se considera el modelo normal no lineal, que fue el modelo que mejor se ajustó a los datos.

**Palabras clave:** Regresión no lineal normal; Estimación de parámetros; Intervalos de confianza; Pruebas de hipótesis; Modelos no lineales.

## 1. Introdução

Avaliar a possível relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns em análise estatística. Pode-se atingir este objetivo usando os modelos de regressão lineares ou os modelos de regressão não lineares.

Em estatística, a regressão não linear é uma forma de análise observacional em que os dados são modelados por uma função que é uma combinação não linear de parâmetros do modelo e depende de uma ou mais variáveis independentes. Os dados são ajustados geralmente pelo Método dos mínimos quadrados ou por algum método de aproximações sucessivas (Mazucheli & Achcar, 2002).

Os métodos mais utilizados para a estimação dos parâmetros do modelo em regressão não linear são o método de mínimos quadrados e o método de máxima verossimilhança. Segundo a afirmação de Gallant (1987), o método de estimação por mínimos quadrados utilizado para encontrar os parâmetros de um modelo de regressão não linear é o mesmo utilizado em modelos lineares. Draper e Smith (1998) afirmam que a estimativa de mínimos quadrados de  $\theta$  é a mesma estimativa encontrada por meio do método de máxima verossimilhança.

No referido trabalho, mostra-se que o modelo de regressão linear simples não é o mais apropriado para a análise dos dados, porém, pela análise gráfica, foi possível ajustar uma curva de crescimento que melhor se adéque aos dados. O termo curva de crescimento é geralmente utilizado para descrever curvas que apresentam o comportamento de medidas de dimensão ao longo do tempo (Silva et al., 2021). Curvas de crescimento de animais, plantas e frutos são comumente estudadas por regressão não lineares, como em Fernandes et al. (2019), Jane et al. (2019) e Silva et al. (2020), respectivamente. E no artigo em questão trata-se do crescimento do rebrote em pastagem contra o tempo desde o último pastejo.

Com base no exposto, o presente estudo teve como objetivo geral estudar os modelos não lineares, com foco no modelo normal não linear. Métodos de estimação, resultados assintóticos, intervalos de confiança e testes de hipóteses serão discutidos para um maior aprofundamento do modelo.

## 2. Metodologia

Para uma maior motivação, analisamos um conjunto de dados onde, apesar de o modelo de regressão linear se ajustar ao banco de dados, podemos ver que esse ajuste pode ser melhorado. Surge então a necessidade de utilizar modelos mais precisos e eficientes. O modelo de regressão não linear é uma proposta alternativa para algumas situações onde o modelo linear não é satisfatório. Para uma melhor prática dos modelos não lineares, uma sugestão são os diversos exemplos dados por Ratkowsky (1983) e Bates e Watts (1988), nesse artigo foi apresentado o exemplo estudado por Ratkowsky (1983).

### 2.1 Exemplo do crescimento de Pastagem

Neste exemplo, estudado por Ratkowsky (1983), observamos o crescimento do rebrote em pastagem contra o tempo desde o último pastejo, ou seja, a variável explicativa é o tempo e a variável resposta é a produção. Os dados da Tabela 1 foram retirados do livro *Statistical Tools for Nonlinear Regression: A practical Guide With S-PLUS an R examples* de Huet et al. (2003) e apresenta a produção de rebrota de pastagem em cada tempo.

**Tabela 1.** Dados de produção de rebrota de pastagem versus tempo (as unidades não são mencionadas).

Tempo após pastagem	9	14	21	28	42	57	63	70	79
Produção	8.93	10.8	18.59	22.33	39.35	56.11	61.73	64.62	67.08

Fonte: Huet et al. (2003).

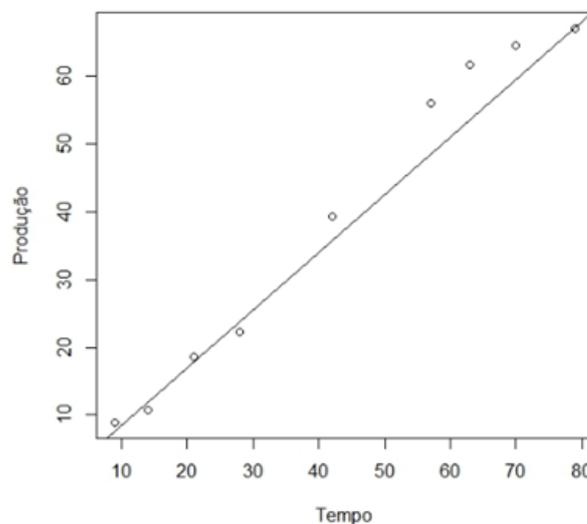
O modelo de regressão linear não normal será dado pela equação 2.1:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (2.1)$$

Onde:  $y_i$  é a variável resposta e representa a produção e  $x_i$  é a variável explicativa que representa o tempo de pastagem, e temos que  $\varepsilon_i \sim N(0, \sigma^2)$ , em que os  $\varepsilon_i$  são independentes e identicamente distribuídos.

O Gráfico de dispersão da variável resposta Produção contra a variável Tempo de Pastagem é exibido na Figura 1:

**Figura 1.** Gráfico dos valores da variável resposta contra a variável tempo.



Fonte: Autor.

A partir da Figura 1 podemos ver que o modelo linear normal se adequou ao banco de dados, pois a reta ajustada do modelo linear normal acomodou bem os dados, mas podemos sugerir um modelo que se adequa melhor, usando a regressão não linear.

Surge então a necessidade de utilizar técnicas mais avançadas para ajustar modelos a bancos de dados como este, regressão linear se adequa, porém, o ajuste não é dos melhores. Como já mencionado uma alternativa é o uso da Regressão não linear. Primeiramente, iremos identificar as características do modelo em estudo, que são as seguintes:

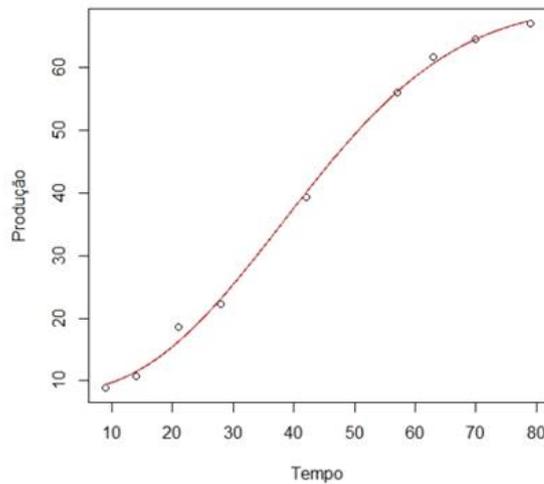
$$\left. \begin{array}{l} Y_i = f(x_i, \vartheta) + \varepsilon_i \\ \text{Var}(\varepsilon_i) = \sigma^2, E(\varepsilon_i) = 0 \end{array} \right\} \quad (2.2)$$

Neste exemplo, uma possível escolha para a função  $f(x_i, \vartheta)$  é o modelo de crescimento dado por:

$$f(x, \vartheta) = \vartheta_1 - \vartheta_2 \exp(-\exp(\vartheta_3 + \vartheta_4 \log x)) \quad (2.3)$$

Após ajustar este modelo ao banco de dados, obteve-se o resultado apresentado na Figura 2. Neste gráfico, podemos ver claramente que o modelo não linear se ajusta bem ao conjunto de dados.

**Figura 2.** Gráfico de dispersão da variável resposta contra a variável tempo, ajustando uma curva.



Fonte: Autor.

## 2.2 O modelo de regressão não linear paramétrico

O modelo de regressão não linear é frequentemente utilizado por pesquisadores em diversas áreas do conhecimento, tais como, agricultura, biologia, econometria, engenharia, química etc. O modelo de regressão não linear pode ser escrito como:

$$Y_i = f(x_i, \theta) + \varepsilon_i, i = 1, \dots, n, \quad (2.4)$$

Onde  $Y_i$  é a variável resposta,  $f(.)$  é uma função não linear contínua, com forma conhecida, do vetor de variáveis explicativas  $x_i$ , e dos parâmetros desconhecidos  $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ . Iremos considerar os erros aleatórios,  $\varepsilon_i$ , independentes e identicamente distribuídos, normais com media zero e variância constante  $\sigma^2$ .

## 2.3 Estimação

Como sabemos, o vetor de parâmetros  $\theta$  geralmente é desconhecido e precisamos estimá-lo. Uma solução natural é escolher o valor de  $\theta$  que minimiza as distâncias entre os valores de  $f(x, \vartheta)$  e as observações  $y_1, y_2, \dots, y_n$  de Y. Por exemplo, pode-se escolher o valor de  $\theta$  que minimiza a soma dos quadrados  $C(\theta)$  definida por:

$$C(\theta) = \sum_{i=1}^n (Y_i - f(x_i, \theta))^2 \quad (2.5)$$

Considerando  $\hat{\theta}$  o estimador de mínimos quadrados de  $\theta$ . Se assumirmos que  $Var(\varepsilon_i) = \sigma^2$ , uma estimativa de  $\sigma^2$  é obtida como  $\hat{\sigma}^2 = \frac{C(\hat{\theta})}{n}$ .

Sob os pressupostos enunciados no parágrafo anterior,  $\hat{\theta}$  é também a solução do conjunto de  $p$  equações, Para  $a = 1, \dots, p$ , onde  $\partial f / \partial \theta_a$  é a derivada parcial de  $f$  com relação à  $\theta_a$ .

$$-2 \sum_{i=1}^n \frac{\partial f}{\partial \theta_a} (x_i, \theta) (Y_i - f(x_i, \theta)) = 0 \quad (2.6)$$

Como  $f$  é não linear em  $\theta$ , nenhuma solução explícita pode ser calculada. Um procedimento iterativo é necessário. Alguns dos procedimentos que podem ser usados são: método de estimação de Newton-Raphson, método BFGS, L-BFGS-B, o método nls(). Aderimos o método nls() para realizar as estimativas dos parâmetros não lineares do modelo, esta função encontra-se disponível no *software* R.

Os chutes iniciais são de extrema importância, no método nls() é preciso utilizar o chute inicial, tivemos alguns problemas em obter uma solução rápida das estimativas dos parâmetros, pois para que haja uma convergência rápida devemos chutar bons valores iniciais, o que não é nada simples na análise de regressão não linear.

O estimador de mínimos quadrados de  $\theta$  também é o estimador de máxima verossimilhança, no caso de observações Gaussianas. Seja  $Y$  uma variável aleatória distribuída como uma variável Gaussiana com esperança  $f(x, \theta)$  e variância  $\sigma^2$ . A função densidade de probabilidade de  $Y$  no ponto  $y$  é definida como segue:

$$\ell(y, x, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-f(x,\theta))^2}{2\sigma^2}\right) = * \quad (2.7)$$

Onde observamos  $n$  variáveis Gaussianas independentes:  $Y_i, i = 1, \dots, n$ , com esperança:  $E(Y_i) = f(x_i, \theta)$  e variância:  $Var(Y_i) = \sigma^2$ . A densidade de probabilidade das observações calculadas em:  $y_i, i = 1, \dots, n$ , é dada pela seguinte fórmula:

$$\prod_{i=1}^n \ell(y_i, x_i, \theta, \sigma^2). \quad (2.8)$$

Sejam  $\varepsilon_i$  variáveis aleatórias independentes e identicamente distribuídas, com distribuição  $N(\theta, \sigma^2)$ . Então a função de verossimilhança  $L$  é dada por:

$$L(\theta, \sigma^2) = \prod_{i=1}^n \ell(y_i, x_i, \theta, \sigma^2), \quad (2.9)$$

Onde  $\ell(y_i, x_i, \theta, \sigma^2) = *$ .

Os estimadores de máxima verossimilhança para  $\theta$  e  $\sigma^2$  são os valores que maximizam a função de verossimilhança, ou equivalentemente, seu logaritmo. Considerando o caso em que  $Y_i$  tem distribuição Gaussiana com esperança  $f(x_i, \theta)$  e variância  $\sigma^2$ , temos a seguinte função log verossimilhança:

$$\ell(\theta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - f(x_i, \theta))^2 \quad (2.10)$$

É fácil ver que a maximização de  $\ell(\theta, \sigma^2)$  em  $\theta$  é equivalente a minimizar  $C(\theta)$ . O estimador de máxima verossimilhança de  $\sigma^2$  irá satisfazer a condição de que a derivada de  $\ell(\theta, \sigma^2)$  em relação à  $\sigma^2$  é igual a 0. Ou seja, temos o seguinte:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} C(\theta) = 0. \quad (2.11)$$

A partir, o estimador de máxima verossimilhança de  $\sigma^2$  é

$$\widehat{\sigma^2} = \frac{c(\hat{\theta})}{n}, \quad (2.12)$$

Onde  $\hat{\theta}$  é o estimador de máxima verossimilhança de  $\theta$ .

### 3. Precisão dos Estimadores, Intervalo de Confiança e Testes de Hipóteses

Para determinar a precisão dos estimadores que obtivemos anteriormente, demonstraremos como calcular intervalos de confiança e realizar testes. Então introduzimos a metodologia, descrevendo inicialmente alguns procedimentos assintóticos clássicos e dois testes assintóticos, o teste de Wald e o teste da razão de verossimilhanças, e então apresentamos procedimentos e testes baseados em um método de reamostragem, o bootstrap.

#### 3.1 Formulação do problema

O modelo de regressão não linear foi definido na seção 2.1, seja  $\hat{\theta}$  o estimador de mínimos quadrados de  $\theta$  quando temos variâncias homogêneas.

Seja  $\lambda$  uma função de parâmetros e  $\hat{\lambda}$  um estimador tal que  $\hat{\lambda} = \lambda(\hat{\theta})$ . Deve-se entender como calcular um intervalo de confiança para  $\lambda$ , e como fazer um teste de hipóteses para este parâmetro. A função  $\lambda$  deve satisfazer algumas condições de regularidade. Deve ser uma função contínua de  $\theta$  com derivadas parciais contínuas em relação à  $\theta$ .

#### 3.2 Soluções

##### 3.2.1 Resultados assintóticos clássicos

$\hat{\theta}$  é uma função da variável resposta  $Y_i$ , e quando o número de observações tende ao infinito, sua distribuição é conhecida: Segundo Huet et al. (2004),  $\hat{\theta} - \theta$  tende a zero, e a distribuição limite de  $V_{\hat{\theta}}^{-1/2}(\hat{\theta} - \theta)$  é uma distribuição normal padrão (Gaussiana)  $p$  dimensional  $N_p(0, I_p)$ , onde  $I_p$  é uma matriz identidade  $p \times p$  e  $V_{\hat{\theta}}$  é a matriz de covariância assintótica estimada de  $\hat{\theta}$ . Assim, para  $n$  suficientemente grande, a distribuição de  $\hat{\theta}$  pode ser aproximada pela distribuição normal  $N_p(\theta, V_{\hat{\theta}})$ . Precisamos de um resultado para  $\hat{\lambda} = \lambda(\hat{\theta})$ . A distribuição limite (quando  $n$  tende ao infinito) será uma distribuição normal padrão,  $N(0, 1)$ , onde  $\hat{S}$  é uma estimativa do erro padrão de  $\hat{\lambda}$ :  $\hat{T} = \frac{\hat{\lambda} - \lambda}{\hat{S}}$

Notações e fórmulas:  $f(x_i, \theta)$  será denotada por  $f_i$ . O vetor de  $p$  derivadas de  $f$  com relação à  $\theta$  calculado em  $x_i$  é denotado por  $\partial f_i / \partial \theta$ . Os componentes de  $\partial f_i / \partial \theta$  são  $[(\partial f_i / \partial \theta_a)(x_i, \theta)]$   $a = 1, \dots, p$ .

Seja  $\Gamma_{\theta}$  a matriz  $p \times p$  definida como segue:

$$\Gamma_{\theta} = \frac{1}{\sigma^2} \frac{1}{n} \sum_{i=1}^k n_i \frac{\partial f_i}{\partial \theta} \left( \frac{\partial f_i}{\partial \theta} \right)^T, \quad (3.1)$$

Esta é a fórmula geral, veja que  $\sigma^2$  não varia, é constante, e o expoente  $T$  significa que o vetor é transposto. Os elementos  $(a, b)$  de  $\Gamma_{\theta}$  são:

$$\Gamma_{\theta, ab} = \frac{1}{\sigma^2} \frac{1}{n} \sum_{i=1}^k n_i \frac{\partial f_i}{\partial \theta_a} \left( \frac{\partial f_i}{\partial \theta_b} \right)^T, \quad (3.2)$$

Onde  $a = 1, \dots, p$  e  $b = 1, \dots, p$ .

Seja  $\Delta_\theta$  a matriz  $p \times p$  definida como segue:

$$\frac{1}{n} \sum_{i=1}^k \frac{n_i}{\sigma_i^2} \frac{\partial f_i}{\partial \theta} \left( \frac{\partial f_i}{\partial \theta} \right)^T, \quad (3.3)$$

Neste caso  $\sigma_i^2$  varia, este é um modelo mais complexo.

Seja  $\hat{f}_i = f(x_i, \hat{\theta})$ ,  $\partial \hat{f}_i / \partial \theta$  o vetor com componentes  $\partial f / \partial \theta_\alpha(x_i, \hat{\theta})$  e  $\Gamma_{\hat{\theta}}$  e  $\Delta_{\hat{\theta}}$  sejam as matrizes  $\Gamma_\theta$  e  $\Delta_\theta$ , onde os parâmetros desconhecidos são substituídos pelos seus estimadores:

$$\Gamma_{\hat{\theta}} = \frac{1}{\hat{\sigma}^2} \frac{1}{n} \sum_{i=1}^k n_i \frac{\partial \hat{f}_i}{\partial \theta} \left( \frac{\partial \hat{f}_i}{\partial \theta} \right)^T,$$

$$\Delta_{\hat{\theta}} = \frac{1}{n} \sum_{i=1}^k \frac{n_i}{\hat{\sigma}^2} \frac{\partial \hat{f}_i}{\partial \theta} \left( \frac{\partial \hat{f}_i}{\partial \theta} \right)^T,$$

E  $\hat{\sigma}^2 = C(\hat{\theta})/n$ ,  $\hat{\sigma}_i^2 = S_i^2$ . Onde  $V_{\hat{\theta}}$  é o estimador de  $V_\theta$ , a matriz de covariância assintótica de  $\hat{\theta}$ : No caso  $Var(\epsilon_{ij}) = \sigma^2$ ,  $V_{\hat{\theta}} = \frac{1}{n} \Gamma_{\hat{\theta}}^{-1}$ , no caso  $Var(\epsilon_{ij}) = \sigma_i^2$ ,  $V_{\hat{\theta}} = \frac{1}{n} \Delta_{\hat{\theta}}^{-1}$ .

Como  $\hat{\theta} - \theta$  é pequeno, temos a distribuição limite de  $\hat{\lambda}$  pela aproximação  $\lambda(\hat{\theta}) - \lambda(\theta)$  por uma função linear de  $\hat{\theta} - \theta$ :

$$\left( \frac{\partial \lambda}{\partial \theta} \right)^T (\hat{\theta} - \theta) = \sum_{\alpha=1}^p \frac{\partial \lambda}{\partial \theta_\alpha} (\hat{\theta}_\alpha - \theta_\alpha). \quad (3.4)$$

Assim,  $(\hat{\lambda} - \lambda) / S_{\hat{\theta}}$  é distribuído, quando  $n$  tende ao infinito, como uma  $N(0; 1)$ , onde:

$$S_{\hat{\theta}}^2 = \left( \frac{\partial \lambda}{\partial \theta} \right)^T V_\theta \frac{\partial \lambda}{\partial \theta} = \sum_{\alpha=1}^p \sum_{b=1}^p \frac{\partial \lambda}{\partial \theta_\alpha} \frac{\partial \lambda}{\partial \theta_b} V_{\theta, \alpha b} \quad (3.5)$$

Dessa forma,  $\hat{S} = S_{\hat{\theta}}$  é a estimativa assintótica do erro padrão.

### 3.2.2 Intervalos de confiança assintótico para $\lambda$

Se a distribuição de  $\hat{T}$  for conhecida e não depender dos parâmetros, ou seja,  $\hat{T}$  tem que ser uma quantidade pivotal. Digamos  $F(u) = P(\hat{T} \leq u)$ , nós calcularíamos os percentis  $\alpha/2$  e  $1 - \alpha/2$  de  $\hat{T}$  digamos  $u_\alpha$  e  $u_{1-\alpha/2}$ . Seja o intervalo:

$$\hat{I} = [\hat{\lambda} - u_{1-\alpha/2} \hat{S}; \hat{\lambda} - u_{\alpha/2} \hat{S}]$$

Este um intervalo de confiança para  $\lambda$ , com nível de confiança  $1 - \alpha$ . Neste caso, a probabilidade de  $\hat{I}$  conter  $\lambda$ , é igual a  $1 - \alpha$ . Entretanto, como vimos no parágrafo anterior, podemos apenas aproximar, quando  $n$  é suficientemente grande, a distribuição de  $\hat{T}$ . Portanto, usamos essa aproximação para calcular intervalos de confiança com probabilidade de cobertura próximo a  $1 - \alpha$ .

Seja  $N$  uma variável distribuída como uma  $N(0; 1)$ . Seja  $v_\alpha$  o percentil  $\alpha$  de  $N$ . Do resultado da seção 3.2.1, podemos deduzir um intervalo de confiança para  $\lambda$ , que segue:

$$\hat{I}_N = [\hat{\lambda} - v_{1-\alpha/2} \hat{S}; \hat{\lambda} + v_{1-\alpha/2} \hat{S}]$$

Este intervalo é simétrico em torno de  $\hat{\lambda}$  para  $v_\alpha = -v_{1-\alpha}$ . A probabilidade de que  $\hat{T}$  seja menor que  $v_\alpha$  tende a  $\alpha$  quando  $n$  tende ao infinito; a probabilidade de  $\lambda$  estar no intervalo  $\hat{I}_N$  tende a  $1 - \alpha$  quando  $n$  tende ao infinito. Dizemos que  $\hat{I}_N$  tem nível assintótico  $1 - \alpha$ .

*Observações:*

1. Por analogia com o caso de regressão linear normal, no modelo de regressão não linear com variância homogênea, definimos um intervalo de confiança alternativo para  $\lambda$ . Na equação (3.1), nós substituímos  $v_\alpha$  por  $\sqrt{n/(n-p)}t_\alpha$ , onde  $t_\alpha$  é o percentil  $\alpha$  de uma distribuição t-student com  $n-p$  graus de liberdade:

$$\hat{I}_T = \left[ \hat{\lambda} - \sqrt{\frac{n}{n-p}} t_{1-\alpha/2} \hat{S}; \quad \hat{\lambda} + \sqrt{\frac{n}{n-p}} t_{1-\alpha/2} \hat{S} \right]. \quad (3.6)$$

Onde  $\hat{I}_T$  tem o mesmo nível assintótico de  $\hat{I}_N$ , mas  $\hat{I}_T$  é mais largo que  $\hat{I}_N$  e sua probabilidade de cobertura será maior. Segundo Huet et al. (1989), alguns estudos têm mostrado que  $\hat{I}_T$  tem probabilidade de cobertura mais próxima de  $1 - \alpha$  do que  $\hat{I}_N$ .

2. Os intervalos “ $\hat{I}_N$ , e  $\hat{I}_T$ ” são simétricos em torno de  $\hat{\lambda}$ . Em algumas aplicações, uma parte do intervalo de confiança simétrico pode não coincidir com o conjunto de variações do parâmetro  $\lambda$ . Por exemplo, considere  $\lambda = \exp \theta_3$  no exemplo do crescimento do pasto. Segundo (HUET et AL., 1989), se a estimativa do erro padrão de  $\hat{\lambda}$ , digamos  $\hat{S}$ , é maior que  $\lambda/v_{1-\alpha/2}$ , então o limite inferior de  $\hat{I}_N$  é negativo apesar de  $\lambda$  ser estritamente positivo. Neste caso, é fácil ver que é mais apropriado calcular um intervalo de confiança para  $\theta_3$  e então transformar esse intervalo tomando a exponencial de seus limites para encontrar um intervalo de confiança para  $\lambda$ . Mais geral, seja  $\hat{S}_3$  a estimativa do intervalo de erro padrão para  $\theta_3$ , e seja  $g$  uma função estritamente crescente de  $\theta_3$ . Se  $\theta_3$  está em

$$[\hat{\theta}_3 - v_{1-\alpha/2} \hat{S}_3; \hat{\theta}_3 + v_{1-\alpha/2} \hat{S}_3],$$

Então,  $\lambda = g(\theta_3)$  está em:

$$[g(\hat{\theta}_3 - v_{1-\alpha/2} \hat{S}_3); g(\hat{\theta}_3 + v_{1-\alpha/2} \hat{S}_3)].$$

### 3.2.3 Teste assintótico para $\lambda = \lambda_0$ contra $\lambda \neq \lambda_0$

Seja  $\lambda_0$  um intervalo fixo de  $\lambda$  e considere a hipótese de interesse  $H_0: \lambda = \lambda_0$ , contra a hipótese alternativa  $H_1: \lambda \neq \lambda_0$ .

#### Teste de Wald

Quando  $H_0$  é verdadeira, a distribuição limite de  $(\hat{\lambda} - \lambda_0)/\hat{S}$  segue uma  $N(0; 1)$ . Assim, a distribuição limite da estatística de teste é:  $S_W = \frac{(\hat{\lambda} - \lambda_0)^2}{\hat{S}^2}$

Onde  $\chi^2$  possui 1 grau de liberdade. A hipótese nula  $H_0$  será rejeitada para valores grandes de  $S_W$ , digamos  $S_W > c$ , onde  $c$  é escolhido de modo que  $P(Z_1 \leq c) = 1 - \alpha$ , onde  $Z_1$  tem distribuição  $\chi^2$  com 1 grau de liberdade.

Dessa forma, quando  $H_0$  é verdadeira, a probabilidade de  $S_W$  ser maior que  $C$  (em outras palavras, a probabilidade de a hipótese  $H_0$  ser rejeitada quando deveria ser aceita) tende a  $\alpha$  quando  $n$  tende ao infinito. Dizemos que este teste tem um erro assintótico do tipo I, igual a  $\alpha$ . Suponha agora que  $H_0$  é falsa. Então o poder do teste definido como a probabilidade de  $S_W$  ser maior que  $C$  (ou seja, a probabilidade de rejeitar a hipótese  $H_0$  quando  $H_0$  é falsa) tende a 1 quando  $n$  tende ao infinito. Nós dizemos que este teste é consistente.

*Observação:* Como na seção 3:2:2, variâncias homogêneas podem ser consideradas separadamente; a hipótese  $H_0$  será rejeitada se:  $\frac{n-p}{n} S_w > C$ ,

Onde  $C$  é escolhido tal que  $Pr(F_{1,n-p} \leq C) = 1 - \alpha$ , onde  $F_{1,n-p}$  é distribuída como uma variável de Fisher com 1 e  $n - p$  graus de liberdade.

#### Teste da razão de verossimilhanças

Outra ideia é estimar os parâmetros sob a restrição de que  $\lambda = \lambda_0$ , digamos  $\hat{\theta}_{H_0}$ ; então estimá-los sem a restrição, digamos  $\hat{\theta}_{H_1}$ ; e depois comparar os critérios de estimação (2)  $C(\hat{\theta}_{H_0})$  e  $C(\hat{\theta}_{H_1})$  no caso de variâncias homogêneas. Se  $H_0$  é verdadeira, a diferença entre  $C(\hat{\theta}_{H_0})$  e  $C(\hat{\theta}_{H_1})$  será pequena. Seja a estatística de teste:

$$S_L = n \log C(\hat{\theta}_{H_0}) - n \log C(\hat{\theta}_{H_1})$$

Quando  $n$  tende ao infinito, pode-se demonstrar que a distribuição de  $S_L$  é uma  $\chi^2$  com 1 grau de liberdade. A hipótese  $H_0$  será rejeitada quando  $S_L > C$ , onde  $C$  é escolhido de modo que  $P(Z_1 \leq C) = 1 - \alpha$ , e  $Z_1 \sim \chi_1^2$ .

Conforme Huet et al. (2004), este teste baseado em  $S_L$  é chamado de teste de razão de verossimilhanças. Tem as mesmas propriedades assintóticas do teste de Wald. Embora o teste de Wald seja mais fácil de calcular, alguns argumentos teóricos favorecem o teste da razão de verossimilhanças. Para verificar a precisão desses resultados assintóticos é através das medidas de não linearidade (Box, 1971).

### 3.2.4 Estimativas Bootstrap

Métodos de reamostragem como o Bootstrap é especialmente útil para estimar a precisão de um estimador. Sejam  $Y_1, Y_2, \dots, Y_n$  observações; escolhemos um modelo de regressão não linear paramétrico com vetor de parâmetros  $\theta$ , e encontramos uma estimativa para uma função de  $\theta$ , digamos  $\lambda(\theta)$ . Temos  $\hat{\lambda} = (\hat{\theta})$ , mas estamos interessados em calcular a precisão de  $\hat{\lambda}$  ou, mais precisamente, em conhecer a sua distribuição (ou algumas das suas características). Se pudéssemos repetir o experimento sob exatamente as mesmas condições, observaríamos  $Y_1^1, Y_2^1, \dots, Y_n^1$ , e da mesma forma que para  $\hat{\lambda}$  calcularíamos  $\hat{\lambda}^1$ . Poderíamos repeti-lo novamente e calcular  $\hat{\lambda}^2$  com  $Y_1^2, Y_2^2, \dots, Y_n^2$ . Dessa forma,  $\hat{\lambda}^1, \hat{\lambda}^2, \dots, \hat{\lambda}^n$  seria uma amostra de variáveis aleatórias distribuídas como  $\hat{\lambda}$ . Esta amostra aproximaria a distribuição de  $\hat{\lambda}$ . Em resumo, os métodos de reamostragem são uma maneira de imitar a repetição do experimento.

As estimativas de Bootstrap baseiam-se em estimativas  $\hat{\lambda}^* = \lambda(\theta^*)$  calculadas a partir de amostras artificiais de bootstrap  $(x_i Y_i^*), i = 1, \dots, n$ , onde

$$Y_i^* = f(x_i, \hat{\theta}) + \epsilon_i^*.$$

Os erros  $\epsilon_i^*$  são simulados da seguinte maneira: Sejam  $\hat{\epsilon}_i = Y_i - f(x_i, \hat{\theta})$  os resíduos, e  $\tilde{\epsilon}_i = \hat{\epsilon}_i - \hat{\epsilon}_\bullet$ , os resíduos centrados, onde  $\hat{\epsilon}_\bullet$  é a média da amostra, e  $\hat{\epsilon}_\bullet = \sum_i \hat{\epsilon}_i / n$ . O conjunto de  $\epsilon_{ij}^*$ , para  $i = 1, \dots, n$  é uma amostra aleatória da função de distribuição empírica baseada em  $\tilde{\epsilon}_i$  ( $n \tilde{\epsilon}_i$  são retirados com reposição, cada uma com probabilidade  $1/n$ ). Há  $n^n$  amostras diferentes, que são selecionados pelos resíduos.

$\hat{\theta}$  será o valor de  $\vartheta$  que minimiza

$$C^*(\theta) = \sum_{i=1}^n (Y_i^* - f(x_i, \theta))^2.$$

A estimativa de bootstrap de  $\lambda$  e  $\hat{\lambda}^* = (\hat{\theta}^*)$ .

Seja  $B$  o número de simulações de bootstrap  $\hat{\lambda}^{*,b} = \lambda(\hat{\theta}^{*,b})$ ,  $b = 1, \dots, B$ .  $B$  é uma amostra de estimadores de bootstrap de  $\lambda$ . A escolha de  $B$  será discutida no final desta seção. O resultado importante é que a distribuição de  $\hat{\lambda}^*$ , estimada pela função de distribuição empírica de  $(\hat{\lambda}^{*,b}, b = 1, \dots, B)$ , aproxima-se da distribuição de  $\hat{\lambda}$ . Seja:

$$\hat{T}^* = \frac{\hat{\lambda}^* - \hat{\lambda}}{S_{\hat{\theta}^*}}.$$

Grosso modo, a diferença entre a função de distribuição de  $\hat{T}$  e  $\hat{T}^*$  tende à zero quando o número de observações  $n$  é grande; assim podemos usar o quantil de  $\hat{T}^*$  em vez de  $\hat{T}$  para obter intervalos de confiança ou testes.

Enfatizamos que o uso da distribuição de bootstrap para aproximar a distribuição de  $\hat{T}$  é teoricamente justificado quando  $n$  é grande e é uma alternativa à distribuição normal padrão apresentada na seção 3:2:1. Em conjunto de dados reais, o número de observações é fixo e pode ser pequeno. Nenhum resultado teórico é conhecido sobre a superioridade de uma dessas aproximações sobre as outras.

#### Intervalo de confiança Bootstrap para $\lambda$

Seja  $(\hat{T}^{*,b}, b = 1, \dots, B)$  uma amostra de tamanho  $B$  de  $\hat{T}^*$ ;  $\hat{T}^*$  é calculado da mesma forma que  $\hat{T}$ , substituindo  $Y_i$  por  $Y_i^*$ . Pode-se mostrar que  $Pr(\hat{T} \leq b_\alpha)$  tende à  $\alpha$  quando  $n$  tende ao infinito. Isto nos dá o seguinte intervalo de confiança bootstrap para  $\lambda$ :

$$\hat{I}_B = [\hat{\lambda} - b_{1-\alpha/2}\hat{S}; \hat{\lambda} - b_{\alpha/2}\hat{S}].$$

Para  $n$  e  $B$  grandes, a probabilidade de cobertura de  $\hat{I}_B$  é próximo à  $1 - \alpha$ .

#### *Estimativa bootstrap da precisão de $\hat{\lambda}$*

A variância, e até mesmo o viés, de  $\hat{\lambda}$  pode ser infinita ou indefinida. No entanto, suas estimativas (usando resultados assintóticos da seção 3:2:2 ou o bootstrap) medem a localização e dispersão da distribuição de  $\hat{\lambda}$ .

*Variância:* A estimativa de bootstrap da variância é calculada usando a variância empírica da amostra  $(\hat{\lambda}^{*,b}, b = 1, \dots, B)$ :

$$\hat{S}^{*2} = \sum_{b=1}^B \frac{1}{B-1} (\hat{\lambda}^{*,b} - \hat{\lambda}^{*,\bullet})^2, \quad (3.7)$$

Onde  $\hat{\lambda}^{*,\bullet}$  é a média da amostra, dada por  $\hat{\lambda}^{*,\bullet} = \sum_{b=1}^B \hat{\lambda}^{*,b} / B$ .

*Viés:* Como observamos na seção 3:2:2, a esperança de  $\hat{\lambda}$ ,  $E(\hat{\lambda})$ , está próxima de  $\lambda$  quando temos grandes valores de  $n$ . Em outras palavras, o viés de  $\hat{\lambda}$ ,  $BIAS = E(\hat{\lambda}) - \lambda$ , está próximo de zero. Podemos usar a amostra bootstrap para estimar esse viés:

$$\widehat{BIAS}^* = \hat{\lambda}^{*,\bullet} - \hat{\lambda}.$$

*Erro quadrático médio:* Podemos estimar o erro quadrático médio (MSE) de uma forma semelhante:  $MSE = E(\hat{\lambda} - \lambda)^2 = S^2 + BIAS^2$  e a variância de  $\hat{\lambda}$ ; e é estimado por:  $\widehat{MSE}^* = \hat{S}^{*2} + \widehat{BIAS}^{*2}$ .

*Mediana:* como é sempre definido o erro mediano, a mediana de  $\hat{\lambda} - \lambda$ , é em especial de grande interesse. Sua estimativa de bootstrap,  $\widehat{MED}^*$ , é a mediana dos  $B$  valores de  $|\hat{\lambda}^{*,b} - \hat{\lambda}|$ .

*Observações:*

1. Vimos que o número de diferentes amostras de bootstrap é igual a  $n$ . Obviamente, nunca escolhemos para  $B$  um valor que rapidamente se torna não usual ( $8^8 = 16777216!$ ). Segundo Huet et al. (1989), na prática, no entanto, um número moderado geralmente é suficiente: Se  $B$  é cerca de 50, podemos estimar a característica precisamente, e se  $B$  é cerca de 200, podemos calcular um intervalo de confiança, por exemplo.

2. Outros métodos de reamostragem, como o jackknife, estão disponíveis na literatura, especialmente para estimar a precisão da característica, ver Wu (1986) e Bunke (1990) para mais detalhes.

3. Existe também o método de bootstrap para o caso de variâncias heterogêneas.

## 4. Estimação da Variância

### 4.1 Modelagem paramétrica da variância

Antes de estimar as variâncias, devemos escolher uma função de variância. Em alguns casos, esta escolha pode ser feita numa base mais ou menos teórica. Frequentemente, temos apenas indicações qualitativas: ou a variância das observações cresce com sua esperança, ou a relação entre a variância das observações e sua esperança, pode ser representada por um par abola, por exemplo. Em ambos os casos, a escolha do modelo é baseada nos resíduos dentro de uma coleção restrita de modelos possíveis.

Para uma variância crescente, segundo Huet et al. (2004), existem essencialmente dois modelos. Ou a variação varia como uma potência da resposta, como segue:

$$\sigma^2 = \sigma^2 g(x_i, \theta, \tau) = \sigma^2 f(x_i, \theta)^\tau,$$

Ou a variância varia como uma função linear da resposta, como segue:

$$\sigma^2 = \sigma^2 g(x_i, \theta, \tau) = \sigma^2 (1 + \tau f(x_i, \theta))$$

Note que temos que ter cuidado ao usar alguns desses modelos. A variância é positiva por definição, mas as funções mencionadas anteriormente podem ser negativas para alguns valores da variável independente  $x$  e dos parâmetros. Consideremos o modelo de variância dado pela equação (4.1), por exemplo. Mesmo que a função  $f(x, \theta)$  não seja negativa para qualquer valor de  $x$  e  $\theta$ , a função  $\sigma^2(1 + \tau f(x, \theta))$  é negativa quando  $\tau < -1/f(x, \theta)$ .

### 4.2 Estimação

#### 4.2.1 Estimação por máxima verossimilhança

O método de máxima verossimilhança é amplamente conhecido e é utilizado no campo da inferência paramétrica. Considere o modelo de regressão não linear:

$$\left. \begin{aligned} Y_i &= f(x_i, \theta) + \varepsilon_i \\ \text{Var}(\varepsilon_i) &= \sigma^2 g(x_i, \theta, \tau), E(\varepsilon_i) = 0 \end{aligned} \right\} \quad (4.1)$$

Onde:  $\varepsilon_i$  são assumidos como variáveis aleatórias gaussianas independentes para  $i$  variando de 1 a  $n$ . Note que os parâmetros  $\theta$  podem modelar tanto as funções de regressão como de variância, enquanto que os parâmetros  $\tau$  modelam apenas a função de variância.

Para facilidade de manuseio, geralmente usamos o logaritmo da verossimilhança em vez da própria verossimilhança. Para o modelo de regressão heterocedástico precedente, o logaritmo da função de verossimilhança é definido como segue:

$$\ell(\theta, \sigma^2, \tau) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^n \left[ \log \sigma^2 g(x_i, \theta, \tau) + \frac{(Y_i - f(x_i, \theta))^2}{\sigma^2 g(x_i, \theta, \tau)} \right]. \quad (4.2)$$

Para um dado conjunto de observações  $Y_1, \dots, Y_n$ , a log-verossimilhança é uma função dos parâmetros. Se  $p$  é a dimensão de  $\theta$  e  $q$  é a dimensão de  $\tau$ , o modelo depende de  $p + q + 1$  parâmetros. O estimador de máxima verossimilhança

$\theta, \sigma^2, \hat{\tau}$  maximiza a log-verossimilhança. Os valores de  $\hat{\theta}, \hat{\sigma}^2, \hat{\tau}$  não podem ser dados explicitamente; eles são obtidos por computação numérica.

#### 4.2.2 Estimação por quase-verossimilhança

O estimador de máxima verossimilhança definido anteriormente baseia-se na suposição de que os erros  $\varepsilon_i$  para  $i$  variando de 1 a  $n$  são distribuídos como variáveis normais. Em alguns casos esta suposição não é apropriada. Por exemplo, quando observamos contagens ou proporções, é mais apropriado modelar a distribuição dos erros  $\varepsilon_i$  com uma distribuição Poisson ou binomial. Em outras situações, poderíamos nos satisfazer considerando a regressão não linear (expressão 4.2) sem qualquer suposição adicional sobre a distribuição de erros. Para estimar os parâmetros, podemos usar o método de quase-verossimilhança baseado no conhecimento das funções de regressão e de variância apenas. Distinguiremos dois casos de acordo com a modelagem paramétrica da função de variância. Mais precisamente, consideramos primeiro o caso em que o parâmetro  $\tau$  é conhecido e então o caso em que o parâmetro  $\tau$  é desconhecido e tem que ser estimado. Quando o parâmetro  $\tau$  é conhecido, assumimos que a variância do  $\varepsilon_i$  é proporcional a uma função conhecida  $g$  que depende da variável independente  $x_i$  e dos parâmetros desconhecidos  $\theta$ . Consideremos o modelo de regressão não linear.

##### Caso 1:

$$\left. \begin{aligned} Y_i &= f(x_i, \theta) - \varepsilon_i \\ \text{Var}(\varepsilon_i) &= \sigma^2 g(x_i, \theta), E(\varepsilon_i) = 0 \end{aligned} \right\}$$

Onde os  $\varepsilon_i$  são assumidos como variáveis aleatórias com  $i$  variando de 1 a  $n$ .

As equações de quase-verossimilhança são dadas pelas fórmulas

$$U_a(\theta) = \sum_{i=1}^n \frac{\partial f}{\partial \theta_a}(x_i, \theta) \frac{Y_i - f(x_i, \theta)}{g(x_i, \theta)}$$

Onde para cada  $a = 1, \dots, p$ ,

$$\frac{\partial f}{\partial \theta_a}(x_i, \theta)$$

Denota a derivada de  $f$  em relação à  $\theta_a$  calculada em  $x_i$  e  $\theta$ .

Os estimadores de quase-verossimilhança de  $\theta$  e  $\sigma^2$  denotados por  $\hat{\theta}_{QL}$  e  $\hat{\sigma}_{QL}^2$ , são definidos como segue:

$$\begin{aligned} U_a(\hat{\theta}_{QL}) &= 0 \\ \hat{\sigma}_{QL}^2 &= \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - f(x_i, \hat{\theta}_{QL}))^2}{g(x_i, \hat{\theta}_{QL})}. \end{aligned}$$

Note que, se a variância das observações é constante ( $g(x, \theta, \tau) = 1$ ), a estimativa de quase-verossimilhança e o estimativa de mínimos quadrados são as mesmas.

##### Caso 2:

Quando o parâmetro  $\tau$  é desconhecido: Vamos voltar ao modelo de regressão não linear, temos que estimar  $\theta$  com  $p$  componentes,  $\tau$  com  $q$  componentes, e  $\sigma^2$ . As equações de quase-verossimilhança são dadas pelas fórmulas:

$$U_a(\theta, \tau) = \sum_{i=1}^n \frac{\partial f}{\partial \theta_a}(x_i, \theta) \frac{Y_i - f(x_i, \theta, \tau)}{g(x_i, \theta, \tau)} \quad (4.3)$$

Para  $a = 1, \dots, p$ .

$$U_{p+b}(\theta, \tau) = \sum_{i=1}^n \frac{\partial g}{\partial \tau_b}(x_i, \theta, \tau) \frac{(Y_i - f(x_i, \theta))^2 - \sigma^2 g(x_i, \theta, \tau)}{g(x_i, \theta, \tau)}, \quad (4.4)$$

Para  $b = 1, \dots, q$ . Os estimadores de quase-verossimilhanças  $\hat{\theta}_{QL}, \hat{\tau}_{QL}, \hat{\sigma}_{QL}^2$  satisfazem:

$$\left. \begin{aligned} U_a(\hat{\theta}_{QL}, \hat{\tau}_{QL}) &= 0 \\ U_{p+b}(\hat{\theta}_{QL}, \hat{\tau}_{QL}) &= 0 \end{aligned} \right\}$$

Para  $a = 1, \dots, p$ , e  $b = 1, \dots, q$ , e

$$\hat{\sigma}_{QL}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - f(x_i, \hat{\theta}_{QL}))^2}{g(x_i, \hat{\theta}_{QL}, \hat{\tau}_{QL})}. \quad (4.5)$$

### 4.2.3 Estimação por três etapas

Para alguns casos, como no caso 1 apresentado na seção 4.2.2, o número de parâmetros é grande, tornando difícil a maximização numérica direta da log verossimilhança em relação ao conjunto completo de parâmetros ou a resolução do conjunto de equações (4.5). Para superar esta dificuldade, podemos usar o método de três etapas. Este método divide o processo de estimativa em sucessivas etapas de fácil manipulação, conforme descrevemos quando estimamos os parâmetros pelo método de quase-verossimilhança.

*Etapa 1:* Estimativa dos mínimos quadrados ordinários de  $\theta$ . Um primeiro estimador de  $\theta$  é  $\hat{\theta}_{OLS(1)}$ , que minimiza:

$$C(\theta) = \sum_{i=1}^n (Y_i - f(x_i, \theta))^2 \quad (4.6)$$

*Etapa 2:* A estimativa de  $\tau$  pelo método da quase-verossimilhança assume que o valor do parâmetro  $\theta$  é conhecido e é dado pela sua estimativa a partir do passo anterior, digamos  $\hat{\theta}_{OLS(1)}$ . Então estimamos  $\tau$  usando o método de quase verossimilhança:  $\hat{\tau}_{QL(2)}$  satisfaz  $U_{p=b}(\hat{\theta}_{OLS(1)}, \hat{\tau}_{QL(2)}) = 0$  para  $b = 1, \dots, q$ .

*Etapa 3:* Estimativa de  $\theta$  pelo método de quase-verossimilhança. Finalmente, estimamos  $\theta$  e  $\sigma^2$  usando o método de quase verossimilhança, assumindo que o valor de  $\tau$  é conhecido e é igual a  $\hat{\tau}_{QL(2)}$ . Mais exatamente, o estimador  $\hat{\theta}_{QL(3)}$  é definido por  $U_a(\hat{\theta}_{QL(3)}, \hat{\tau}_{QL(2)}) = 0$  para  $a = 1, \dots, p$  e:

$$\hat{\sigma}_{QL(3)}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - f(x_i, \hat{\theta}_{QL(3)}))^2}{g(x_i, \hat{\theta}_{QL(3)}, \hat{\tau}_{QL(2)})}. \quad (4.7)$$

Este método de estimativa de três etapas tem boas propriedades: podemos obter resultados assintóticos clássicos análogos aos descritos na seção 3.2.1. Na segunda e terceira etapas do método alternativo de três etapas, o método de máxima verossimilhança pode ser usado em vez do método da quase-verossimilhança.

## 4.3 Testes

### 4.3.1 O teste de Wald

Cada um dos métodos descritos na seção anterior tem boas propriedades: Podemos obter resultados assintóticos clássicos análogos aos descritos na seção 3.2.1. Como consequência, para uma função real contínua dos parâmetros  $\lambda =$

$(\theta, \sigma^2, \tau)$ , temos os seguintes resultados. Sejam  $\hat{\theta}, \hat{\sigma}^2, \hat{\tau}$  as estimativas de  $\theta, \sigma^2, \tau$ , obtidas por quaisquer dos métodos que consideramos na seção anterior. Denote  $\hat{\lambda} = \lambda(\hat{\theta}, \hat{\sigma}^2, \hat{\tau})$  e  $\lambda = (\theta, \sigma^2, \tau)$ . Então  $\hat{\lambda} - \lambda$  tende a 0 quando  $n$  tende ao infinito. Além disso, existe uma estimativa assintótica  $\hat{S}$  do erro padrão de  $\hat{\lambda}$  tal que a distribuição de  $\hat{S}^{-1}(\hat{\lambda} - \lambda)$  pode ser aproximada pela distribuição Normal padrão  $N(0, 1)$ . Como na seção 3.2.3, podemos testar a hipótese nula  $H_0: \{\lambda = \lambda_0\}$  contra a hipótese alternativa  $H_1: \{\lambda \neq \lambda_0\}$ , usando a estatística de Wald.

$$S_W = \frac{(\hat{\lambda} - \lambda_0)^2}{\hat{S}^2}.$$

Para esse fim, usamos a regra de decisão a seguir. Escolhemos um nível de risco  $\alpha, \alpha < 1$ ; então calculamos o valor crítico  $C$  tal que  $P_r\{Z_1 > C\} = \alpha$ , onde  $Z_1$  é uma variável aleatória distribuída como uma  $\chi^2$  com um grau de liberdade; então comparamos  $S_W$  e  $C$ ; se  $S_W > C$ , rejeitamos a hipótese nula  $H_0$ ; se  $S_W \leq C$ , não a rejeitamos. Este teste tem um nível  $\alpha$  assintótico.

#### 4.3.2 O teste da razão de verossimilhança

Na seção 3.2.3, a estatística do teste de razão de verossimilhança  $S_L$  foi introduzida para um modelo de regressão com variâncias homogêneas, onde o método de estimação é o de mínimos quadrados ordinários. Nesse caso, o cálculo da estatística “ $S_L$ ” foi um caso particular da aplicação da metodologia de inferência geral baseada no princípio da máxima verossimilhança. Nesta seção, descrevemos como esse princípio se aplica quando nos deparamos com o modelo de regressão heterocedástico. Vamos introduzir alguma notação. Chamamos  $\phi$  a todo o conjunto de parâmetros  $\phi^T = (\theta^T, \sigma^T, \tau^T)$ . Chamamos  $q$  a dimensão de  $\tau$ , então  $\phi$  é de dimensão  $p + q + 1$ . Por exemplo, no modelo descrito pelas equações  $f(x, \theta) = 100/1 + \exp(\theta_2(x - \theta_1))$  e  $\sigma_1^2 = \sigma^2 + \sigma^2 \tau f(x_i, \theta)[100 - f(x_i, \theta)] = \sigma^2 g(x_i, \theta, \tau)$ ,  $p = 2$ ,  $q = 1$  e  $\phi = (\theta_1, \theta_2, \sigma^2, \tau)^T$  é um vetor de dimensão quatro. Em muitas situações,  $q = 0$ , por exemplo, se descrevemos a variância por  $\sigma_1^2 = \sigma^2 f(x_i, \theta)$  ou por  $\sigma^2 f(x_i, \theta)^2$ . Consideramos agora uma situação bastante geral em que a hipótese nula  $H_0$  é descrita por uma restrição linear em um subconjunto de todo o conjunto de parâmetros  $\phi$ . A restrição é expressa como  $\Lambda_\phi = L_0$ , onde  $\Lambda$  é uma matriz  $r \times (p + q + 1)$ , ( $r < p + q + 1$ ) e  $L_0$  é um vetor constante com dimensão  $r$ .  $r$  é também assumido como sendo o posto da matriz  $\Lambda$ . Seja  $\hat{\phi}_H^T = (\hat{\phi}_H^T, \hat{\sigma}_H^2, \hat{\tau}_H^T)$  os estimadores dos parâmetros para a hipótese nula  $H_0$ , o que significa que a restrição  $\Lambda \hat{\phi}_H = L_0$  é preenchida, enquanto que  $\hat{\theta}, \hat{\sigma}^2, \hat{\tau}$ , os estimadores de máxima verossimilhança clássica dos parâmetros do modelo (4.2) são os estimadores de parâmetros para a hipótese alternativa não restritiva  $H_1$ . A estatística de razão de verossimilhança é:

$$S_L = -2 \log \frac{L_n(Y_n; \hat{\theta}, \hat{\sigma}^2, \hat{\tau})}{L_n(Y_1, \dots, Y_n; \hat{\theta}_H, \hat{\sigma}_H^2, \hat{\tau}_H)} \quad (4.8)$$

Com nossas anotações, isso acaba por ser:

$$S_L = \sum_{i=1}^n \log \frac{\hat{\sigma}_H^2 g(x_i; \hat{\theta}_H, \hat{\tau}_H)}{\hat{\sigma}^2 g(x_i; \hat{\theta}, \hat{\tau})}. \quad (4.9)$$

A hipótese nula  $H_0$  é rejeitada se  $S_L > C$ , onde  $C$  é definido por

$$P_r\{Z_r \leq C\} = 1 - \alpha.$$

$Z_r$  é distribuída como uma  $\chi^2$  com  $r$  graus de liberdade, e  $\alpha$  é o nível (assintótico) do teste.

*Observações:*

1. A justificativa intuitiva desta regra de decisão é clara. Por um lado, se  $H_0$  for verdadeira, então  $\hat{\theta}_{H_0}$ ,  $\hat{\sigma}_{H_0}^2$ ,  $\hat{\tau}_{H_0}$  e  $\hat{\theta}$ ,  $\hat{\sigma}^2$ ,  $\hat{\tau}$  são estimadores adequados dos parâmetros do modelo: Quando o número de observações é grande, eles estão próximos do verdadeiro valor dos parâmetros. Portanto, eles tomam valores semelhantes e a razão de verossimilhança é próxima de 1, ou, de forma equivalente, seu logaritmo é próximo de 0. Por outro lado, se  $H_0$  for falsa, então  $\hat{\theta}_{H_0}$ ,  $\hat{\sigma}_{H_0}^2$ ,  $\hat{\tau}_{H_0}$  não estimam o verdadeiro valor dos parâmetros e espera-se que tenham valores muito diferentes de  $\hat{\theta}$ ,  $\hat{\sigma}^2$ ,  $\hat{\tau}$ . Portanto,  $S_L$  tem um valor diferente de 0.

2. A estatística do teste,  $S_L$ , para o caso em que as variâncias são homogêneas é dada por:

$$S_L = n \log C(\hat{\theta}_{H_0}) - n \log C(\hat{\theta}_{H_1}).$$

Quando “ $n$ ” tende ao infinito, é possível mostrar que a distribuição limite de  $S_L$  é uma  $\chi^2$  com um grau de liberdade.

## 5. Aplicação

Nesta seção apresentaremos uma aplicação, da metodologia apresentada, a um conjunto de dados reais. A parte computacional foi feita no *software* R.

Os dados dessa aplicação foram originalmente apresentados em Dudzinski e Mykytowycz (1961) e estudados posteriormente por Ratkowsky (1983) baseados em um modelo de regressão normal não linear. Os dados consistem em um conjunto de 71 observações em que a variável resposta apresenta o peso das lentes em mg ( $y$ ) dos olhos de coelhos europeus na Austrália e a variável explicativa corresponde a idade em dias ( $x$ ) dos coelhos. Inicialmente ajustamos o modelo normal linear, que pode ser escrito da seguinte forma:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (5.1)$$

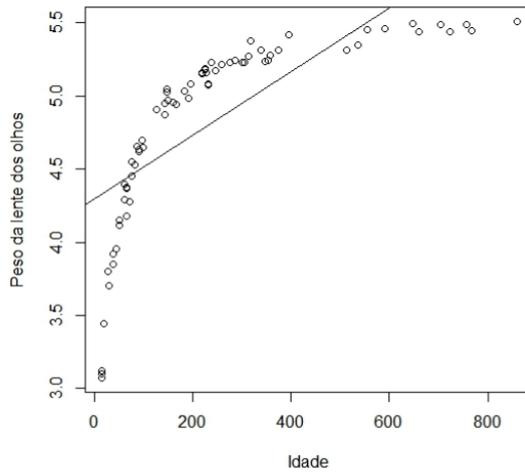
Em que  $y_i$  é o peso das lentes em mg dos olhos dos coelhos e  $x_i$  é a idade em dias dos coelhos e consideramos os erros independentes e identicamente distribuídos, após estimar os parâmetros do modelo, a expressão (5.1) ficará escrita da seguinte forma, que será o modelo normal linear estimado:

$$\hat{y}_i = 4,3003 + 0,0022x_i,$$

Após usar o *software* R para a estimação dos parâmetros, obtivemos  $R^2 = 0,55$ , como não é próximo de 1, podemos concluir que o modelo não é bom. Obtivemos  $R_{ajustado}^2 = 54\%$  o que nos diz que a variável regressora explica apenas 54% da variável resposta.

Prosseguindo a análise com o modelo normal linear, vejamos o gráfico de dispersão, da variável “peso” da lente dos olhos em função da idade apresentado na Figura 3.

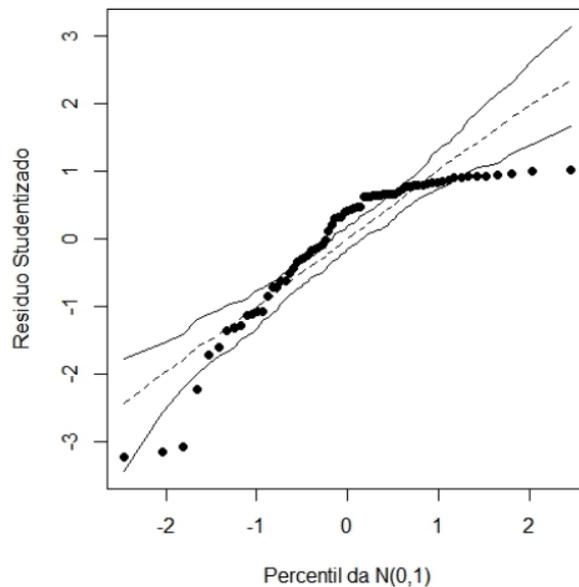
**Figura 3.** Gráfico de dispersão do peso das lentes dos olhos contra a idade.



Fonte: Autor.

Analisando graficamente, podemos ver um comportamento não linear através da Figura 3. Veja que o modelo Normal linear não se ajusta aos dados, daí vem a necessidade do ajuste de um modelo de regressão não linear. Apresentamos o gráfico de envelope a seguir na Figura 4.

**Figura 4.** Gráfico do envelope do modelo normal linear.



Fonte: Autor.

A partir da Figura 4, chegamos à conclusão de que o modelo linear normal não se ajusta aos dados, pois muitas observações ficaram fora do intervalo de confiança, logo, temos a necessidade de usar a regressão não linear para fazer as devidas análises. Partindo para o modelo não linear normal, o modelo utilizado é dado por:

$$\log(Y_i) = \theta_1 - \frac{\theta_2}{x_i + \theta_3} + \varepsilon_i, i = 1, \dots, 71 \quad (5.2)$$

Onde  $Cov(\epsilon_i, \epsilon_j) = 0$  para  $i \neq j$ . Segundo Cysneiros e Ferrari (2006), assumiram que  $\epsilon_i \sim N(0, \sigma^2)$  e estimaram os parâmetros do modelo por máxima verossimilhança, obtendo as seguintes estimativas:  $\hat{\theta}_1 = 5,640$ ,  $\hat{\theta}_2 = 130,58$ ,  $\hat{\theta}_3 = 37,603$  e  $\sigma^2 = 0,004$ , os resultados estão presentes na Tabela 2.

**Tabela 2.** Estimativas para o modelo (5.2) com os respectivos erros padrões (assintóticos).

Parâmetro	nls()	Erro Padrão
$\theta_1$	5,64	-0,02
$\theta_2$	130,584	-5,725
$\theta_3$	37,603	-2,323
$\sigma^2$	0,004	
$SQRes$	0,269	

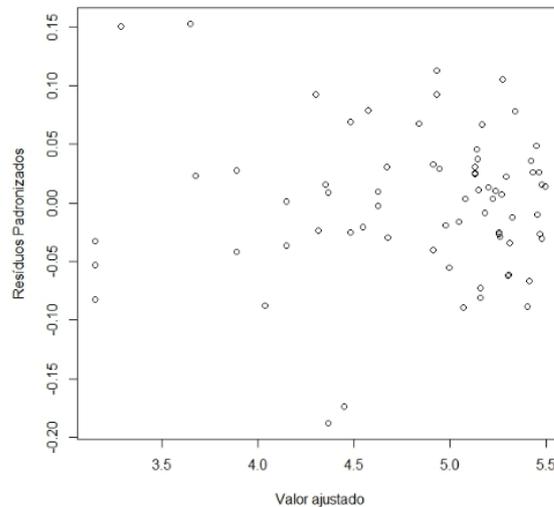
Fonte: Autor.

Após ter os valores estimados dos parâmetros do modelo, podemos escrever  $f(x, \theta)$  da seguinte forma:

$$f(x, \hat{\theta}) = 5,640 - \frac{130,58}{x + 37,603}.$$

Prosseguindo, temos o gráfico da análise de resíduos do modelo normal não linear na Figura 5:

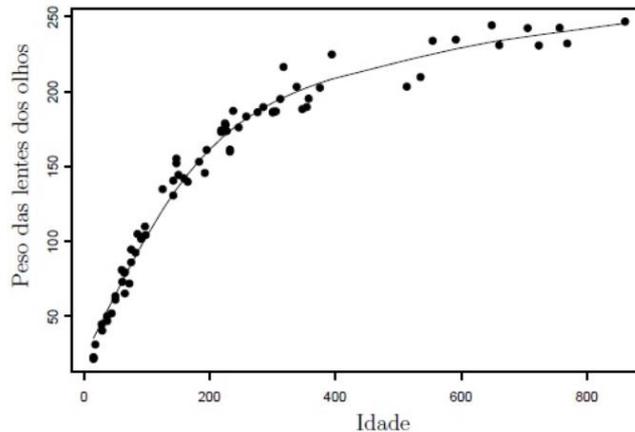
**Figura 5.** Gráfico dos valores ajustados contra resíduos padronizados.



Fonte: Autor.

Na Figura 5, podemos ver que as observações seguem um padrão homocedástico, pois a diferença na variância,  $\sigma_i - \sigma_{i+1}$ , é pequena. Portanto, há uma alta confiança na homogeneidade aproximada das variâncias, e podemos assumir que  $Var(\epsilon_{ij}) = \sigma^2$ . A partir da Figura 6 podemos ver que o modelo de regressão não linear dado pela expressão (5.2) se ajusta bem aos dados. Veja que os dados estão comportados sob a curva ajustada, portanto segue que o modelo (5.2) está adequado. À medida que a idade dos coelhos aumenta podemos ver que o peso das lentes dos olhos dos coelhos vai aumentando de forma não linear.

**Figura 6.** Gráfico da dispersão do peso das lentes dos olhos contra a idade.



Fonte: Autor.

### 5.1 Calculando os intervalos de confiança para $\lambda$

Após estimar os parâmetros do modelo não linear, escolhemos o parâmetro  $\lambda(\theta) = \theta_2$  para calcular alguns intervalos de confiança, em que os mesmos serão calculados com nível assintótico de 95%.

Seja  $N$  uma variável distribuída como uma  $N(0, 1)$ . Seja  $v_\alpha$  o percentil  $\alpha$  de  $N$ . Um intervalo de confiança para  $\lambda$  é dado por:

$$\hat{I}_N = [\hat{\lambda} - v_{1-\alpha/2}\hat{S}; \hat{\lambda} + v_{1-\alpha/2}\hat{S}].$$

O resultado será o seguinte:

$$\hat{I}_N = [130,58 - 1,96 \times 5,72; 130,58 + 1,96 \times 5,72]$$

$$\hat{I}_N = [119,3688; 141,7912].$$

Podemos afirmar com confiabilidade de 95% de que o verdadeiro valor do parâmetro  $\theta_2$  pertence ao intervalo  $\hat{I}_N$ .

Calculamos também um intervalo de confiança alternativo para  $\lambda$ , dado por:

$$\hat{I}_\tau = \left[ \hat{\lambda} - \sqrt{\frac{n}{n-p}} t_{1-\alpha/2} \hat{S}; \hat{\lambda} + \sqrt{\frac{n}{n-p}} t_{1-\alpha/2} \hat{S} \right]$$

Daí, teremos o seguinte resultado para este intervalo de confiança:

$$\hat{I}_\tau = \left[ 130,58 - \sqrt{\frac{71}{71-3}} \times 1,995469 \times 5,72; 130,58 + \sqrt{\frac{71}{71-3}} \times 1,995469 \times 5,72 \right]$$

$$\hat{I}_\tau = [118,9169; 142,2431].$$

Veja que a partir do intervalo de confiança alternativo tivemos um intervalo maior, temos 95% de confiança de que o verdadeiro valor do parâmetro está em  $\hat{I}_\tau$ . Uma observação é que os dois intervalos ficaram muito próximos. No entanto, o intervalo  $\hat{I}_N$  é mais preciso do que o intervalo  $\hat{I}_\tau$ , por ter uma amplitude menor, com o mesmo nível de confiança de 95%.

## 6. Considerações Finais

Após o estudo da análise de regressão não linear, podemos perceber a grande importância dos modelos não lineares, pois nos deparamos com bancos de dados que o modelo normal linear não se ajusta, com isso a técnica dos modelos não lineares e dos modelos lineares generalizados são possíveis escolhas.

No trabalho utilizamos o método da regressão não linear e obtivemos um bom ajuste. Tivemos dificuldades computacionais, mas estas estão cada vez menores com o avanço computacional que é uma ferramenta de extrema importância para ajustar esses modelos, e um dos livros que ajudou na parte de programação foi “Nonlinear Regression with R” de Ritz (2008). Trabalhamos com o banco de dados dos *coelhos europeus*, em que o ajuste não linear ficou adequado e calculamos alguns intervalos de confiança e obtemos resultados bastante precisos.

## Referências

- Bard, Y. (1974). *Nonlinear Parameter Estimation*. New York: Academic Press.
- Bates, D.M., & Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
- Bates, D. M., & Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, 1–25.
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression: iterative estimation and linear approximations*. [S.l.], Wiley Online Library.
- Box, M.J. (1971). Bias in nonlinear estimation. *Journal of Royal Statistical Society Serie B. Methodological*, London, 33(2), 171-201.
- Bunke, O. (1990). Estimating the accuracy of estimators under regression models. *Technical report*, Humboldt University, Berlin.
- Cysneiros, A.H., & Ferrari, S.L. (2006). An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. *Statistics and probability letters*, Elsevier, 76(3), 255-265.
- Cysneiros, F. J. A., & Vanegas, L. H. (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*, Elsevier, 78(18), 3269–3273.
- Dudzinski, M., & Mykytowycz, R. (1961). The eye lens as an indicator of age in the wild rabbit in australian. *Wildlife Research*, CSIRO, 6(2), 156-159.
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York: J. Wiley, 1–706.
- Fernandes, F. A., Fernandes, T. J., Pereira, A. A., Meirelles, S. L. C., & Costa, A. C. (2019). Growth curves of meat-producing mammals by von Bertalanffy's model. *Pesquisa Agropecuária Brasileira*, S4, 1-S. doi: 10.1590/S1678-3921.pab2019.v54.01162.
- Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: J. Wiley and Sons, 1–610.
- Huet, S., Jolivet, E., & Mess, A. E. (1989). Some simulation results about confidence intervals and bootstrap methods in nonlinear regression. *Statistics*.
- Huet, S., Bouvier, A., Poursat, M.A., & Jolivet, E. (2004). *Tools for Nonlinear Regression: A practical Guide With S-PLUS and R examples*. New York: Springer.
- Jane, S. A., Fernandes, F. A., Silva, E. M., Muniz, J. A., & Fernandes, T. J. (2019). Comparação dos modelos polinomial e não lineares na descrição do crescimento de pimenta. *Revista Brasileira de Ciências Agrárias*, 14(4), 1-7, doi:10.5039/agraria.v14i4a7180.
- Mazucheli, & Achcar. (2002). Algumas considerações em regressão não linear. *Acta Scientiarum, Maringá*, 24(6), 1761-1770.
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, JSTOR, 370–384.
- Ratnkowsky, D. A. (1983). *Nonlinear regression modeling*. Dekker, New York.
- Ratnkowsky, D. A. (1990). *Handbook of nonlinear regression models*. Marcel Dekker, New York.
- Ritz, C., & Streibig, J. C. (2008). *Nonlinear Regression with R*. Springer, 1.
- Silva, E. M., Fruhauf, A. C., Silva, E. M., Muniz, J. A., Fernandes, J. F., & Silva, V. F. (2021). Evaluation of the critical points of the most adequate nonlinear model in adjusting growth data of green dwarf coconut fruits. *Revista Brasileira de Fruticultura*, 43(1). doi: <http://dx.doi.org/10.1590/0100-29452021726>.
- Silva, E. M. da Tadeu, M. H.; Silva, V. P. da, Pio, R., Fernandes, T. J., & Muniz, J. A. (2020). Description of blackberry fruit growth by nonlinear regression models. *Revista Brasileira de Fruticultura*, 42(2).
- Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*.