Escola Superior de Agricultura “Luiz de Queiroz”
Universidade de São Paulo
Departamento de Ciências Exatas
Pós-Graduação em Estatística e Experimentação Agropecuária

Regressão Não Linear

Gustavo Thomas

Relatório sobre regressão não linear para a disciplina de Regressão e Covariância.

PIRACICABA-SP
2016

Sumário

1 Introdução **1**

2 Regressão Não Linear **2**
2.1 Modelos não lineares 2
2.2 Linearização 3

3 Estimação e diagnóstico de um modelo não linear **4**
3.1 Estimação por Mínimos Quadrados Ordinários 4
3.2 Estimação por Mínimos Quadrados Ordinários Generalizados 4
3.3 Método iterativo de Gauss-Newton 5
3.4 Propriedades dos estimadores 6
3.5 Técnicas para obtenção de estimativas iniciais 6
3.6 Análises de diagnóstico e comparação de modelos 7

4 Exemplos de aplicação **8**
4.1 Exemplo 1 8
4.2 Exemplo 2 10
4.3 Exemplo 3 12

5 Considerações finais **16**

1 Introdução

Os modelos não lineares surgiram com os modelos de regressão geral na década de 20 com Ronald Fisher, Ronald Aylmer e Winifred Mackenzie. Entretanto, somente a partir dos anos 70 ganharam mais popularidade e passaram a ser melhor estudados com os avanços computacionais.

Apareceram novos algoritmos para ajuste dos modelos e estes se tornaram mais complexos. Com isso, a popularidade dos modelos não lineares e a valorização dos processos inferenciais por eles gerados também aumentou. Com exceção de alguns trabalhos mais antigos, a maior parte da literatura sobre modelos não lineares surgiu depois dos anos 80 e desde então tem se mantido uma área bastante ativa.

Atualmente, modelos não lineares são usados por pesquisadores para modelar variáveis nas mais diversas áreas do conhecimento, como agricultura, biologia, econometria, engenharia, química, etc.

2 Regressão Não Linear

Os modelos de regressão assumem, de modo geral, a seguinte forma:

$$\mathbf{Y} = f(\mathbf{X}, \theta) + \varepsilon = E(\mathbf{Y}|\mathbf{X}) + \varepsilon \tag{1}$$

em que $f(\mathbf{X}, \theta)$ é a função das variáveis preditoras $\mathbf{X}$ e parâmetros θ a serem estimados (que representa também a esperança da variável resposta dadas as preditoras) e ε são os erros aleatórios, sobre os quais são assumidos os pressupostos usuais.

Enquanto que os modelos lineares são compostos por uma relação linear da variável dependente com as preditoras, os modelos não lineares são caracterizados por funções $f(\mathbf{X}, \theta)$ não lineares em pelo menos um dos parâmetros θ_i. Em outras palavras, *para que um modelo seja não linear é necessário que ao menos uma das derivadas da* $f(\mathbf{x}, \theta)$ *com relação a* θ_i *dependa de ao menos um dos parâmetros* θ_i.

2.1 Modelos não lineares

Entre os exemplos de modelos não-lineares mais conhecidos, pode-se citar os seguintes:

- Modelo de Mitscherlich
$$\mathbf{Y} = \alpha(1 - 10^{-\gamma(\mathbf{X}+\theta)}) + \varepsilon$$
usado especialmente no estudo da variação do crescimento de vegetais ou produção de determinada cultura (Y) em função da quantidade de nutriente fornecida (X);

- Modelo Logístico
$$\mathbf{Y} = \frac{\theta}{1 + e^{-(\alpha+\beta\mathbf{X})}} + \varepsilon$$
usado para modelar crescimento de populações humanas, de animais e vegetais. Usualmente $\theta > 0$ e $\beta > 0$, em que a função logística tem a aparência de uma curva sigmóide com assíntotais em 0 e α;

- Modelo de Gompertz
$$\mathbf{Y} = e^{\alpha+\beta\rho\mathbf{X}} + \varepsilon$$
em que $\beta < 0$ e $0 < \rho < 1$ e a função de Gompertz $exp(\alpha + \beta\rho\mathbf{X})$ é monotonicamente crescente;

- Modelo de Michaelis-Mentel
$$\mathbf{Y} = \frac{\theta_1\mathbf{X}}{\mathbf{X} + \theta_2} + \varepsilon$$
usado em estudos da taxa de reações químicas catalizadas por enzimas.

2.2 Linearização

Existem modelos não lineares aos quais pode ser aplicada alguma reparametrização que torne a função $f(\mathbf{X},\theta)$ linear nos parâmetros θ. A esses modelos dá-se o nome de modelos transformavelmente lineares.

Exemplo 1: O modelo de Michaelis-Mentel pode ser linearizado aplicando a transformação recíproca, de modo a obter:

$$\frac{1}{f} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1 x} + \frac{1}{\varepsilon} <=> f^* = \beta_1 + \beta_2 u + \varepsilon^*$$

em que $f^* = \frac{1}{f}$, $\beta_1 = \frac{1}{\theta_1}$, $\beta_2 = \frac{\theta_2}{\theta_1}$ e $u = \frac{1}{x}$.

Exemplo 2: Seja o modelo não-linear com erro multiplicativo

$$\mathbf{Y} = e^{\alpha+\beta\mathbf{X}}\varepsilon.$$

Este pode ser linearizado com a transformação logarítmica

$$log(\mathbf{Y}) = \alpha + \beta\mathbf{X} + log(\varepsilon).$$

Caso o erro transformado $\varepsilon^* = log(\varepsilon)$ atenda aos pressupostos usuais, é possível fazer uma regressão linear em $log(\mathbf{Y})$. Entretanto, sabe-se que se ε segue uma distribuição normal, $\varepsilon^* = log(\varepsilon)$ já não tem mais essa propriedade.

Dessa forma, é importante destacar que ao linearizar um modelo não-linear o vetor de erros também é transformado, logo os pressupostos usuais assumidos sobre os mesmos deixam de ser válidos na maioria dos casos (ver Bates(1988)).

Todavia, a abordagem de linearização é interessante para obter valores iniciais para o método iterativo de ajuste utilizado na regressão não-linear, como será descrito a seguir.

3 Estimação e diagnóstico de um modelo não linear

3.1 Estimação por Mínimos Quadrados Ordinários

Sabe-se da teoria da estimação por mínimos quadrados ordinários (MQO) que as estimativas obtidas por esse método para θ para um modelo não linear na forma (1) são aquelas que minimizam a soma de quadrados dos erros

$$S(\theta) = \Sigma_{i=1}^{n}[\mathbf{Y}_i - f(\mathbf{X}_i - \theta)]^2 \tag{2}$$

Para encontrar os estimadores que geram tais estimativas, é preciso derivar (2) com relação a cada elemento do vetor θ e igualar a 0, obtendo-se o seguinte vetor de equações normais:

$$\Sigma_{i=1}^{n}[\mathbf{Y}_i - f(\mathbf{X}_i - \theta)][\frac{\delta f(\mathbf{X}_i;\theta)}{\delta \theta_j}]_{\theta=\widehat{\theta}} = 0 \tag{3}$$

Para modelos lineares esse sistema normalmente possui um vetor único de soluções para os parâmetros θ, em que cada parâmetro depende unicamente dos dados observados. Já para modelos não lineares, os estimadores $\widehat{\theta}$ resultantes de (3) em geral não apresentam forma analítica e as soluções precisam ser obtidas por algum método de aproximações iterativo.

Dentre os principais, pode-se citar o método de Gauss-Newton, método de Davidon-Fletcher-Powell, método da descida mais íngreme (*steepest descent*), método da procura pelo gradiente conjugado, método da procura de Nelder-Mead, algoritmo de Golub-Pereyra, entre outros. A seguir, será descrito o método iterativo de Gauss-Newton.

3.2 Estimação por Mínimos Quadrados Ordinários Generalizados

Para analisar modelos não-lineares em que há heteroscedasticidade ou autocorrelação residual, pode-se utilizar a estimação dos parâmetros pelo método de mínimos quadrados generalizados (MQG). Esse método é muito similar ao método de MQO tradicional, sendo que nesse caso a soma a ser minimizada é dada por

$$S(\theta) = \Sigma_{i=1}^{n}[\mathbf{Y}_i - f(\mathbf{X}_i;\theta)]'\mathbf{V}^{-1}[\mathbf{Y}_i - f(\mathbf{X}_i - \theta)]$$

em que a matriz $\mathbf{V}$ é a matriz de variâncias e covariâncias dos resíduos, que pode ser qualquer matriz positiva definida conhecida. Quando $\mathbf{V} = \mathbf{I}_n$, tem-se novamente o método de estimação por MQO.

Para obter as estimativas de mínimos quadrados generalizados, multiplica-se o modelo não linear original em (1) por $\mathbf{R} = (\mathbf{U}')^{-1}$, em que $\mathbf{V} = \mathbf{U}'\mathbf{U}$ é uma decomposição de Cholesky de $\mathbf{V}$. Assim, transforma-se (1) em

$$\mathbf{z} = \mathbf{k}(\theta) + \eta$$

em que $\mathbf{z} = \mathbf{R}\mathbf{y}$, $\mathbf{k}(\theta) = \mathbf{R}f(\mathbf{X},\theta)$ e $\eta = \mathbf{R}\varepsilon$. Com essa transformação, a soma de quadrados a ser minimizada passa a ser dada por

$$S(\theta) = [\mathbf{z} - \mathbf{k}(\theta)]'[\mathbf{z} = \mathbf{k}(\theta)]$$

que é análoga à soma em (2). Dessa forma, esta pode ser minimizada por um procedimento padrão de mínimos quadrados ordinários tratando os dados como um caso de variâncias homogêneas e observações independentes. No software **R**, uma regressão não linear por MQG pode ser ajustada através da função `gnls()` do pacote `nlme`. Maiores informações sobre esse método podem ser encontradas em Seber(2003).

3.3 Método iterativo de Gauss-Newton

O método iterativo de Gauss-Newton se baseia em aproximações lineares para a função $f(\mathbf{X};\theta)$ a cada passo, da seguinte forma.

Tomando $\widehat{\theta} \approx \theta^{(a)}$, tem-se que a expansão em série de Taylor de primeira ordem de $f(\mathbf{X};\theta)$ para $\widehat{\theta}$ em torno de $\theta^{(a)}$ é dada por

$$f(\mathbf{X};\theta) \approx f(\mathbf{X};\theta^{(a)}) + \frac{\delta f(\mathbf{X}_i;\theta)}{\delta\theta_j}(\theta - \theta^{(a)}). \tag{4}$$

Seja $F(\theta^{(a)}) = \frac{\delta f(\mathbf{X}_i;\theta)}{\delta\theta_j}$ a matriz com as derivadas da função não linear em relação a cada um dos parâmetros e sejam $\mathbf{r}(\theta) = \mathbf{Y} - f(\mathbf{X};\theta)$ os resíduos ordinários de θ. Assim, de forma análoga, esses resíduos podem ser aproximados por

$$\mathbf{r}(\theta) \approx \mathbf{r}(\theta^{(a)}) - F(\theta^{(a)})(\theta - \theta^{(a)}). \tag{5}$$

Como o objetivo é minimizar a SQRes=$\mathbf{r}(\theta)'\mathbf{r}(\theta)$ pelo método de mínimos quadrados, tem-se que esta, por sua vez, é aproximada por

$$SQRes = \mathbf{r}(\theta)'\mathbf{r}(\theta) \approx [\mathbf{r}(\theta^{(a)}) - F(\theta^{(a)})(\theta - \theta^{(a)})]'[\mathbf{r}(\theta^{(a)}) - F(\theta^{(a)})(\theta - \theta^{(a)})] = \tag{6}$$

$$\mathbf{r}(\theta^{(a)})'\mathbf{r}(\theta^{(a)}) - 2\mathbf{r}(\theta^{(a)})F(\theta^{(a)})(\theta - \theta^{(a)}) + (\theta - \theta^{(a)})'F(\theta^{(a)})'F(\theta^{(a)})(\theta - \theta^{(a)}) \tag{7}$$

em que a expressão (7) pode ser vista como a expressão matricial para soma de quadrados de uma regressão linear simples com $(\theta - \theta^{(a)}) = \beta$, $F(\theta^{(a)}) = \mathbf{X}$ e $F(\theta^{(a)}) = \mathbf{y}$. Assim, pode-se deduzir que esta será minimizada quando

$$(\theta - \theta^{(a)}) = [F(\theta^{(a)})'F(\theta^{(a)})]^{-1}F(\theta^{(a)})'\mathbf{r}(\theta^{(a)}). \tag{8}$$

A cada iteração é gerada uma nova aproximação $\theta^{(a+1)}$ para o vetor de parâmetros $\theta^{(a)}$, através de

$$\theta^{(a+1)} = \theta^{(a)} + [F(\theta^{(a)})'F(\theta^{(a)})]^{-1}F(\theta^{(a)})'F(\theta^{(a)}). \tag{9}$$

O processo (9) é repetido até que se obtenha convergência definida por algum critério, como $(\theta^{(a+1)} - \theta^{(a)}) < 0,000001$, por exemplo.

O método de Gauss-Newton equivale a uma simplificação do método de Newton-Raphson, necessitando de um número maior de iterações que este, entretanto os cálculos realizados em cada iteração são mais simples. Para maiores detalhes, ver Hoffmann(2013).

Apesar de ser numericamente estável, esse método pode apresentar convergência bastante lenta quando uma precisão grande for exigida, convergir para um máximo local ao invés do máximo global ou mesmo não convergir, especialmente para um tamanho amostral pequeno e/ou estimativas iniciais ruins Bates(1988).

3.4 Propriedades dos estimadores

Para o modelo de regressão linear, sabe-se através do Teorema de Gauss que os estimadores de MQO são normalmente distribuídos, não viesados e de variância uniformemente mínima (dentre todos os estimadores possíveis) para os parâmetros do modelo.

Já para modelos não lineares, essas propriedades são válidas somente assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Em geral, para pequenas amostras, essas propriedades são desconhecidas. Porém, afirma-se que quanto maior o tamanho amostral, mais aplicáveis se tornam os resultados assintóticos (Mazucheli(2002)).

Assumidas as suposições usuais sobre os erros aleatórios, pode-se demonstrar, assintoticamente e sob certas condições de regularidade, as seguintes propriedades dos estimadores de MQO na regressão não linear:

(i) $\widehat{\theta} \sim N(\theta, \sigma^2\mathbf{C})$, em que $\mathbf{C} = [F(\theta^{(a)})'F(\theta^{(a)})]^{-1}$;

(ii) $\frac{(n-p)s^2}{\sigma^2} \approx \chi^2_{n-p}$;

(iii) $\widehat{\theta}$ é independente de σ^2 e

(iv) $\frac{[SQRes(\theta)-SQRes(\widehat{\theta})]\backslash p}{SQRes(\widehat{\theta})\backslash(n-p)} \approx F_{p,n-p}$.

Uma forma de avaliar a precisão desses resultados assintóticos é através de medidas de não-linearidade, como as propostas por Box em 1971 e por Bates e Watts em 1980. Quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos e, consequentemente, mais confiáveis serão as inferências. Mais detalhes sobre estas podem ser encontrados nas respectivas referências.

3.5 Técnicas para obtenção de estimativas iniciais

Todos os processos iterativos dependem da escolha de boas estimativas iniciais e os métodos de ajuste de uma regressão não linear não são exceção. Como já comentado, a escolha de uma estimativa inicial ruim pode levar o algoritmo a convergir para um máximo local ao invés do máximo global ou mesmo divergir, além de se tornar mais demorado e custoso computacionalmente. Entre as técnicas mais conhecidas para escolha de estimativas iniciais de um modelo não linear estão as seguintes:

Linearização: como já comentado anteriormente, essa técnica consiste em realizar uma transformação da função esperança afim de torná-la linear nos parâmetros. Quando isso for possível, pode-se realizar uma regressão linear da variável transformada sobre os novos parâmetros e usar as estimativas (após aplicar a transformação inversa) dessa regressão como estimativas iniciais para o modelo não linear, ignorando o erro aleatório.

Através das propriedades do modelo: outra abordagem é a de estudar o comportamento limite da função esperança quando x_i para pra 0 ou infinito e substituir os parâmetros por valores iniciais que os representem nas condições na escala e contexto do problema.

Análise gráfica: ainda, outra técnica muito usada é a simples análise visual dos dados para "chutar" valores iniciais para os parâmetros do modelo. Nesse caso, o conhecimento do pesquisador e a familiaridade com o modelo utilizado são de grande valia.

3.6 Análises de diagnóstico e comparação de modelos

Definido(s) o(s) modelo(s), obtidas as estimativas iniciais e realizados os ajustes com convergência, é necessário avaliar os modelos obtidos.

Inicialmente, pode-se testar os erros-padrão dos parâmetros e calcular uma estatística t para cada caso (usualmente a função `nls()` do software `R` já retorna esses testes), tal que

$$t = \frac{\widehat{\theta}}{\widehat{s_\theta} \backslash \sqrt{n}} \sim t_\nu. \tag{10}$$

Se o valor da estatística não for significativo é recomendado que se tire o parâmetro em questão e que o ajuste seja refeito sem o mesmo.

Quando um modelo final é obtido, deve-se realizar a análise de resíduos para verificar a qualidade do ajuste. De forma análoga aos modelos lineares, esta é feita através do gráfico de valores observados versus ajustados, resíduos versus valores ajustados e verificação da normalidade. Para a análise residual, Cook e Tsai criaram, em 1985, uma modificação dos resíduos padronizados tradicionais para modelos não lineares, que chamaram de *resíduos projetados*, porém os originais ainda são mais frequentemente utilizados por serem algebricamente menos complexos.

Para o caso de se ter dois modelos igualmente bem ajustados, é preciso fazer uma comparação dos mesmos para decidir qual é o preferido. Se estes forem aninhados, pode-se utilizar um teste de razão de verossimilhanças, análogo ao usado para modelos lineares. Se os modelos não são aninhados, é necessário que se encontre um equilíbrio entre o melhor modelo e o preferido/mais utilizado pelo pesquisador. Para comparar dois modelos não aninhados quanto ao ajuste, pode-se usar critérios como a magnitude do quadrado médio do resíduo e/ou das estatísticas AIC, BIC, entre outras.

4 Exemplos de aplicação

Para os exemplos de aplicação foi utilizado o *software* `R` e a função `nls()` do pacote `nlstools`. Essa função tem como estrutura e principais argumentos os seguintes:

$$nls(formula, data, start, control(list), algorithm, trace)$$

em que

`formula` é a fórmula do modelo que será utilizada. Entretanto, nesse caso a fórmula é definida como uma expressão realmente matemática ao invés da notação Wilkinson-Rogers utilizada para definir modelos lineares como utilizada, por exemplo, com a função `lm()`. Por exemplo, para o modelo logístico na parametrização tradicional esse argumento seria dado por `theta1/(1 + exp(-(theta2 + theta3*x)))`.

`data` é o conjunto de dados a ser utilizado.

`start` é uma lista com valores iniciais para cada um dos parâmetros.

`control` é uma função que controla características do algoritmo, como o número máximo de iterações e o critério de tolerância. Por exemplo, poderia-se definir `control=nls.control(maxiter=40, tol=0.00001)`.

`algorithm` define o tipo de algoritmo utilizado. O *default* é o algoritmo de Gauss-Newton. Outras opções são o algoritmo para modelos parcialmente lineares Golub-Pereyra (opção `plinear`) e a opção `port` para um algoritmo que considera restrições nos parâmetros do modelo.

`trace` define se são apresentados a SQRes e as estimativas dos parâmetros a cada passo ou não (`trace=TRUE` ou `trace=FALSE`).

Os códigos utilizados nos exemplos de aplicação serão apresentados juntamente com a descrição e discussão dos mesmos. Para maiores detalhes sobre a função `nls()` ver Fox e Weisberg (2010).

4.1 Exemplo 1

No primeiro exemplo será demonstrado de forma bastante simples o ajuste de um modelo de regressão não linear através do método de Gauss-Newton. Para tanto, será usado um modelo com apenas uma variável regressora (x) e a variável resposta (y), com 3 observações de cada, como apresentado no código abaixo.

```
y = c(0.8,0.45,0.04)
x = c(1,4,16)
```

O modelo não linear que será ajustado para este conjunto de dados é dado por $y_i = e^{-\theta x_i} + \varepsilon_i$. Para começar, é necessário obter uma estimativa inicial para o parâmetro θ, que nesse caso pode ser obtida linearizando o modelo através da transformação logarítmica. Para tanto, tem-se que o modelo linearizado é dado por

$$ln(y_i) = y_i^* = -\theta x_i + \varepsilon_i^* \tag{11}$$

e a respectiva soma de quadrados a ser minimizada para esse modelo é dada por

$$SQRes = S(\theta) = \Sigma_{i=1}^{3}(y_i^* + \theta x_i). \tag{12}$$

Após derivar (12) em relação a θ e isolá-lo, obtém-se o seguinte estimador

$$\theta_0 = \frac{\Sigma_{i=1}^{3} y_i^* x_i}{\Sigma_{i=1}^{3} x_i^2}. \tag{13}$$

Realizando os cálculos, obtém-se por (13) a estimativa inicial `theta0 = 0,2012`. Fazendo uso desta e procedendo com o método iterativo de Gauss-Newton, é necessário obter os valores estimados da função $f(\mathbf{X};\theta)$, os respectivos resíduos e os valores estimados com a derivada da função $f(\mathbf{X};\theta)$. Esses valores podem ser calculados no software `R`, respectivamente, como segue.

```
theta0 = 0.2012
f = function(x,theta){exp(-theta*x)}
(f_theta0 = f(x=x,theta=theta0))
0.81774887 0.44717735 0.03998704
(r_theta0 = y - f_theta0)
-0.0177488654  0.0028226469  0.0000129649
F. = function(x,theta){-x*exp(-theta*x)}
(F_theta0=F.(x=x,theta=theta0))
-0.8177489 -1.7887094 -0.6397926
```

De posse dessas estimativas, pode-se aplicar a expressão (9) e obter o valor estimado após a primeira iteração de Gauss-Newton.

```
(theta_1 = theta0+solve(t(F_theta0)%*%F_theta0)%*%F_theta0%*%r_theta0)
0.2034108
```

Assim, usando $\theta_0 = 0,2012$ como estimativa inicial, obtém-se $\theta_1 = 0,2034$ como resultado da primeira iteração. Observa-se que a diferença entre os valores já está na terceira casa decimal, o que demonstra que o método da linearização resultou em uma estimativa inicial bastante próxima da final para o parâmetro θ.

Realizando as iterações seguintes com a função `nls()` obtemos os seguintes resultados.

```
modelo = nls(y~exp(-theta*x),start=c(theta=0.2012),trace=T)
0.0003229897 :  0.2012
0.0003018397 :  0.2034108
0.0003018337 :  0.2034484
0.0003018337 :  0.2034489
```

em que a primeira coluna reporta a soma de quadrados residual e a segunda a estimativa do parâmetro por iteração realizada (linha). Dessa forma, temos que a estimativa final para o parâmetro é dada por $\widehat{\theta} = 0,2034489$ e a soma de quadrados residual final por $SQRes = 0.0003018337$. Como esse é um valor quase nulo, tem-se um ajuste quase perfeito, como pode ser observado pelo gráfico de dispersão com a reta ajustada apresentado na Figura 1 abaixo.

```
plot(y~x)
curve(exp(-0.2034489*x),16,1,add=T,col="blue")
```

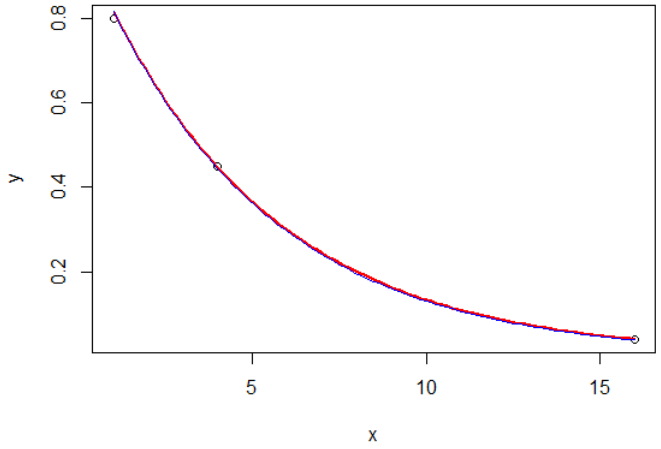

Figura 1: Curva ajustada considerando $\widehat{\theta} = 0,2034489$.

4.2 Exemplo 2

Para o segundo exemplo será utilizado o banco de dados `USPop` do pacote `car` do `R`. Esse banco de dados contém as populações dos Estados Unidos da década de 1790 até a década de 2000, em milhões de habitantes. Portanto, trata-se de uma variável dependente (*population*) e uma variável independente (*years*), que na verdade contém as 22 décadas compreendidas entre 1790 e 2000. Na figura abaixo é apresentado o gráfico de dispersão e a linha de um ajuste linear da população pelas décadas.

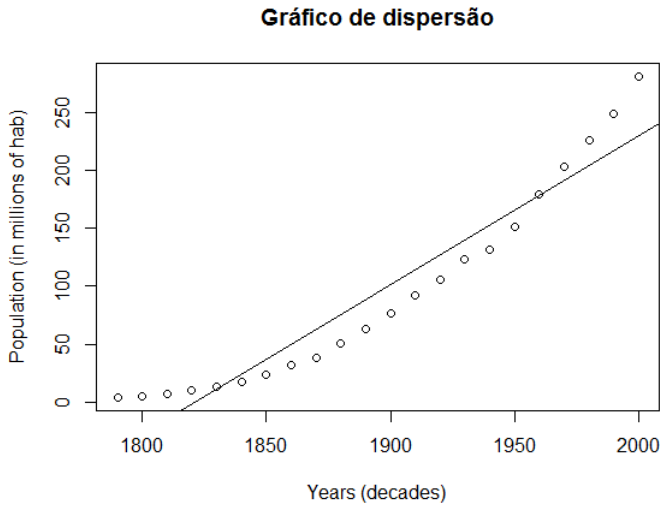

Figura 2: Gráfico de dispersão da população dos EUA por década

A partir da Figura 2 acima vemos que um modelo linear (linha ajustada) não se ajustaria aos dados, visto que o crescimento da população não ocorre a taxas constantes como pressupõe um comportamento linear.

Já o modelo logístico de três variáveis apresentado na subseção 2.1 pode ser uma boa opção para modelar esse comportamento. Para encontrar estimativas iniciais para esse modelo, pode-se utilizar novamente a técnica da linearização, visto que o modelo logístico é linearizável. Para tanto, realiza-se os seguintes passos:

$$\mathbf{y} = \frac{\theta_1}{1+e^{-(\theta_2+\theta_3\mathbf{x})}}$$
$$\frac{\mathbf{y}}{\theta_1} = \frac{1}{1+e^{-(\theta_2+\theta_3\mathbf{x})}}$$
$$log[\frac{\mathbf{y}\backslash\theta_1}{1-\mathbf{y}\backslash\theta_1}] = \theta_2+\theta_3\mathbf{x}.$$

A primeira das 3 equações acima é o modelo logístico tracidional de 3 parâmetros. Na segunda equação foi feita a divisão de todos os termos por θ_1. Como θ_1 é a assíntota do modelo logístico, tem-se que o lado esquerdo da segunda equação é um número entre 0 e 1. Já na terceira equação foi feita a transformação logito (muito usada como função de ligação em modelos lineares generalizados com resposta binomial), tornando o modelo linear em θ_2 e θ_3.

Dessa forma, de posse de uma estimativa inicial para o parâmetro θ_1 pode-se obter estimativas iniciais para os demais através de uma regressão linear, como realizado abaixo. A estimativa inicial utilizada para θ_1 foi de 400, visto que como assíntota superior uma escolha razoável é um valor maior do que o máximo amostral, que pela Figura 2 é em torno de 300.

```
library('car'); library('nlstools'); attach(USPop)
```

```
Coefficients:
```

```
(Intercept)         year
```

```
-49.24991      0.02507
```

De onde obtém-se que as estimativas iniciais a serem utilizadas para θ_2 e θ_3 são, respectivamente, -49 e 0.025. Assim, pode-se ajustar o modelo não linear através da função `nls()` com o seguinte código.

```
pop.mod = nls(population ~ theta1/(1 + exp(-(theta2 + theta3*year))),
 start=list(theta1=400,theta2=-49,theta3=0.025))

summary(pop.mod)

       Estimate Std. Error t value Pr(>t)
theta1 440.8333 35.0001    12.6    1.1e-10
theta2 -42.7069 1.83914   -23.2    2.1e-15
theta3  0.02161 0.00101    21.4    8.9e-15
```

De onde pode-se observar que a estimativa máxima para a população dos EUA é de $\widehat{\theta_1} = 440.8$, ou seja, de aproximadamente 441 milhões de habitantes. Na segunda coluna acima são apresentados os erros-padrão das estimativas, cujo valor para o primeiro parâmetro é relativamente grande, refletindo a incerteza na estimativa da assíntota visto que todos os valores observados são menores que esta. Na Figura 3 abaixo é apresentado novamente o gráfico de dispersão entre as variáveis com a linha ajustada agora pelo modelo não linear em questão.

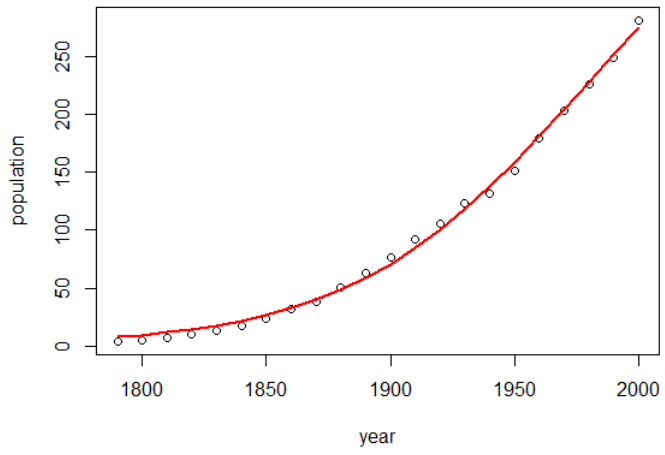

Figura 3: Gráfico de dispersão com linha do ajuste não linear

4.3 Exemplo 3

Nesse exemplo será feito uso do banco de dados `O2K` do pacote `nlstools` do `R`. Tratam-se de observações da quantidade de oxigênio ingerida (`VO2`, em mL/min) por um paciente com problemas pulmonares em vários momentos durante um experimento que consiste em 5,83 minutos de descanso inicial seguidos de 6 minutos de exercícios. Há interesse em se estudar a

quantidade de oxigênio ingerida durante a fase de descanso (`VO2rest`) e relacionar com a quantidade máxima consumida durante o exercício (`VO2peak`). Um modelo não-linear conhecido para descrever tal situação é dado por

$$VO_2(t) = VO_2rest + (VO_2peak - VO_2rest)(1 - e^{-(t-5.83)/\mu})$$

em que t é o tempo decorrido durante o experimento, $VO_2(t)$ a respectiva quantidade de oxigênio ingerida pelo paciente e $\mu > 0$ a inclinação da taxa de aumento do consumo de oxigênio. Como essa equação modela somente o consumo de oxigênio durante a fase de exercício do experimento, ela é válida somente quando $t > 5,83$. Antes disso, durante a fase de preparação (0 a 5,83 minutos), sabe-se que $VO_2(t) = VO_2rest$.

Antes de ajustar o modelo é necessário obter novamente estimativas iniciais para os parâmetros `VO2rest`, `VO2peak` e μ. Para tal, pode-se fazer uso da função `preview` do pacote `nlstools`, que gera o gráfico de valores preditos para um modelo especificado e a respectiva soma de quadrados de resíduos dados "chutes"para os valores iniciais deste. Para o chute `(VO2rest=400,VO2peak=1600,mu=1)`, tem-se os resultados que seguem.

```
formula = as.formula(VO2 ~ (t <= 5.883) * VO2rest + (t > 5.883) *
 (VO2rest + (VO2peak - VO2rest) * (1 - exp(-(t - 5.883) / mu))))

preview(formulaExp,data=O2K,start=list(VO2rest=400,VO2peak=1600,mu=1))
```

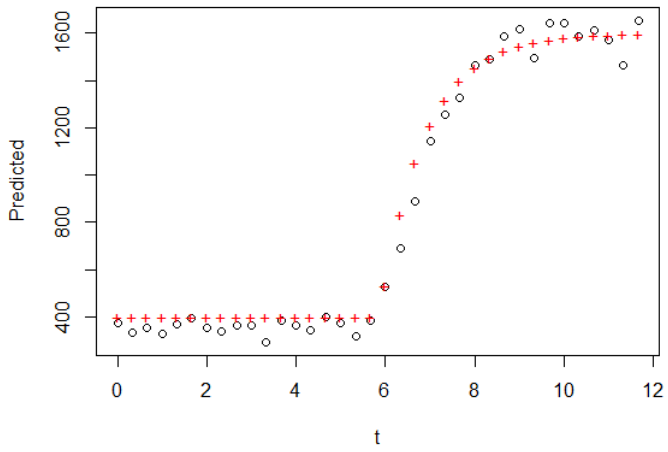

Figura 4: Gráfico de valores observados e ajustados com a função *preview*

Pelo código e pela Figura 3 acima pode-se observar que os valores iniciais são, respectivamente, o valor aproximado da ingestão de oxigênio para a fase de descanso, um valor máximo aproximado de ingestão durante o exercício e uma taxa μ unitária. Através da função `preview()` foi possível observar que o chute de valores iniciais parece ser satisfatório. Dessa forma, pode-se ajustar um modelo não linear com essas estimativas iniciais para os dados como segue.

```
modelo=nls(formula, start=list(VO2rest=400, VO2peak = 1600,mu = 1),data = O2K)
summary(modelo)

        Estimate   Std. Error  t value Pr(>t)
VO2rest 3.568e+02  1.141e+01   31.26   <2e-16 ***
```

```
VO2peak 1.631e+03  2.149e+01   75.88   <2e-16 ***
mu      1.186e+00  7.661e-02   15.48   <2e-16 ***
```

Assim, pode-se concluir que as ingestões médias de oxigênio estimadas são dadas por 356,8 mL/min para a fase de descanso e 1631 mL/min para o consumo máximo durante o exercício. Todos os parâmetros apresentaram contribuição significativa para o modelo.

Para avaliar a qualidade do ajuste pela análise de resíduos, pode-se utilizar as funções `nlsResiduals`, `test.nlsResiduals` e `qqPlot` para extrair os resíduos ordinários, realizar testes de normalidade e de ausência de autocorrelação residual (*Runs Test*) e gerar um gráfico quantil-quantil com envelope simulado dos resíduos, respectivamente. Os resultados são apresentados como segue.

```
res_mod = nlsResiduals(modelo)
```

```
test.nlsResiduals(res_mod)
```

```
Shapiro-Wilk normality test
```

```
W = 0.95205, p-value = 0.1214
```

```
Runs Test
```

```
Standard Normal = 0.76123, p-value = 0.4465
```

```
qqPlot(residuals(modelo))
```

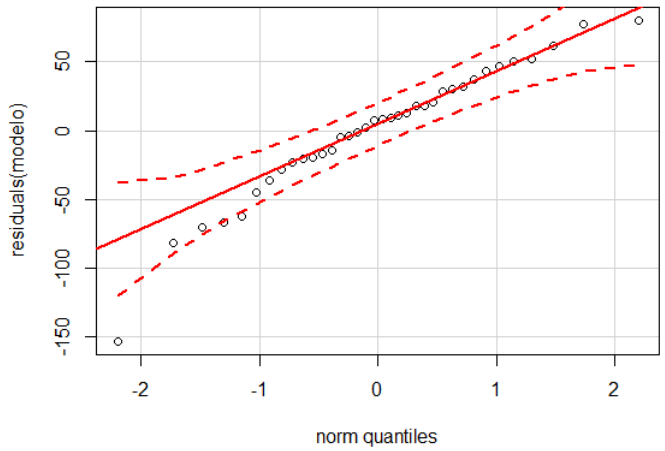

Figura 5: Gráfico quantil-quantil com envelope simulado

Pelos resultados apresentados observa-se que não há evidências para rejeitar a hipótese de normalidade dos resíduos nem de ausência de autocorrelação. Ainda, pela Figura 5 fica mais evidente que a suposição de normalidade está satisfeita.

Para avaliar a correlação entre parâmetros do modelo e detectar pontos de mínimo locais ao invés de globais pode-se criar regiões de confiança bidimensionais. No programa `R`, a função

`nlsContourRSS` gera justamente esse tipo de gráficos bidimensionais, mandando as estimativas para os demais parâmetros do modelo (que não são os 2 usados no gráfico) fixados em suas estimativas de mínimos quadrados. Para o modelo ajustado, tais regiões podem ser obtidas da seguinte forma:

```
reg_conf = nlsContourRSS(modelo)
plot(reg_conf, nlev = 19).
```

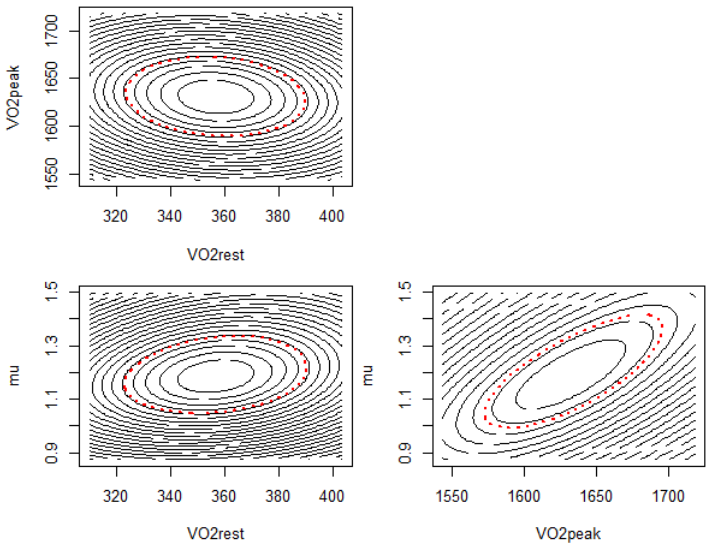

Figura 6: Gráfico com as regiões de confiança para cada 2 parâmetros do modelo

De forma análoga aos intervalos de confiança unidimensionais, as linhas pontilhadas nos gráficos da Figura 6 representam as regiões de confiança para os parâmetros em questão de forma bidimensional. As regiões de confiança nos três gráficos apresentados são elípticas e não apresentam outras regiões de mínimo próximas, sendo assim indicativas de bom ajuste. Ainda, pode-se observar no gráfico entre μ e $VO2peak$ que estas apresentam uma correlação um pouco maior que os outros dois pares de parâmetros, o que também é compreensível, visto que enquanto a taxa de crescimento da ingestão de oxigênio é positiva o consumo máximo ingerido também cresce. Finalmente, o gráfico com os valores observados e a linha dos preditos pelo modelo é apresentado abaixo.

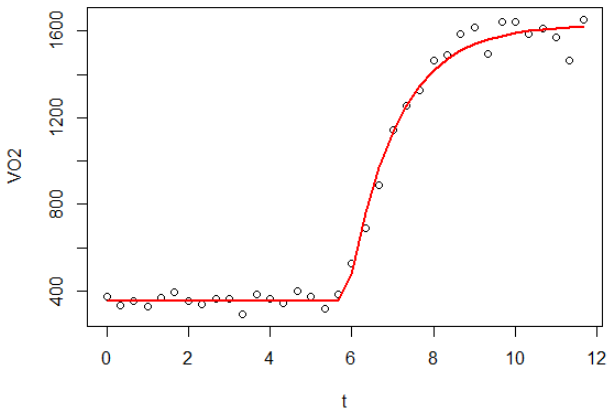

Figura 7: Gráfico de dispersão com linha do ajuste não linear

5 Considerações finais

Esse relatório teve por finalidade introduzir os conceitos básicos de regressão não linear e mostrar sua utilidade prática.

Inicialmente essa classe de modelos foi apresentada e comparada com os modelos lineares clássicos. Foram descritos os principais métodos de estimação e o método iterativo de Gauss-Newton para ajuste dos modelos. Também foram citadas as principais técnicas para avaliar os modelos e classificá-los.

Finalmente, foram apresentados 3 exemplos de aplicação de modelos não lineares. O primeiro deles fez uso de dados fictícios com o intuito de demonstrar o processo de iteração de uma forma mais prática e detalhada. Os dois seguintes usaram bancos de dados do programa `R` e com o auxílio da função `nls()` foram realizados os ajustes e discutidas as principais características desse tipo de modelos.

Referências

Bates, D. & Watts, D. Sons, J. W. &. **Nonlinear Regression Analysis and Its Applications**. John Wiley & Sons, 1988.

Baty, F.; Ritz, C.; Charles, S.; Brutsche, M.; Flandrois, J.-P. & Delignette-Muller, M.-L. **A Toolbox for Nonlinear Regression in R: The Package nlstools**. Journal of Statistical Software, 2015.

Fox, J. & Weisberg, S. **Nonlinear Regression and Nonlinear Least Squares in R**. An R Companion to Applied Regression, 2010.

Gallant, R. A. Series, W. **Nonlinear Statistical Models**. Wiley Series, 1987.

Hoffmann, R. Pronto, L. **Análise Estatística de Relações Lineares e Não Lineares**. Livro Pronto Editora, 2002.

Mazucheli, J. e Jorge Alberto Achcar, J. M. **Algumas considerações em regressão não linear**. Acta Scientiarum, 2002.

Thalita do Bem Mattos, T. **Modelos Não Lineares e suas Aplicações**. Trabalho de conclusão de curso de bacharelado em Estatística. Universidade Federal de Juiz de Fora, 2013.

Seber, G. A. F. & Wild, C. J. Sons, J. W. &. **Nonlinear Regression**. John Wiley & Sons, 2003.