AJUSTE DE CURVAS POR MÍNIMOS QUADRADOS: COMPARAÇÃO ENTRE METODOLOGIAS E DESAFIOS

GUSTAVO WEIRICH CORRÊA[1]; ALEXANDRE MOLTER[2].

[1]*Universidade Federal de Pelotas – correa.gw@hotmail.com*
[2]*Universidade Federal de Pelotas – alexandre.molter@yahoo.com.br*

1. INTRODUÇÃO

O ajuste de curvas por mínimos quadrados é muito utilizado em diversas áreas do conhecimento, como na biologia, economia e astronomia. Existem duas formas de realizar esse ajuste de curvas, uma via regressão linear e outra sem regressão linear. Nesse contexto, serão mostradas algumas diferenças entre esses dois métodos, os quais foram aplicados ao modelo de Verhulst, tomando como base os dados obtidos na pesquisa realizada por PITOL, et al (2021).

Dependendo dos dados utilizados, e a função escolhida no ajuste de curvas, uma das metodologias, ou ainda as duas, poderão se tornar impraticáveis. Um exemplo disto quando aparece uma divisão por zero no cálculo dos parâmetros. Outro exemplo quando não se consegue obter as soluções do sistema algébrico não linear advindo do processo de minimização. Além disto, existem conjuntos de dados e formas de funções que podem gerar um grande conjunto de soluções, tornando sua resolução sem o auxílio computacional quase impraticável.

Para obter uma função que aproxime uma tabela de pontos experimentais, dados em coordenadas (x_i, y_i) , é comum utilizar o ajuste de curvas por mínimos quadrados e regressão linear (BASSANEZI, 2002; SOUZA, acesso em 2022), onde y_i é obtido por meio de experimentos, e desejamos obter a expressão analítica de uma curva $y=f(x)$ que melhor se ajusta a esses pontos. O objetivo da regressão linear é conseguir traçar uma reta por meio de dados de um diagrama de dispersão. Já sem regressão linear, o ajuste é feito diretamente para a função $f(x)$.

2. METODOLOGIA

O método dos quadrados mínimos consiste em determinar parâmetros de modo que “minimize” o valor de

$$S=\sum_{i=1}^{n}(y_i-\overline{y}_i)^2=\sum_{i=1}^{n}[(f\,\overline{x}_i;a_1,a_{2,}\ldots,a_k)-\overline{y}_i]^2, \qquad (1)$$

isto é, devemos minimizar a soma dos quadrados dos desvios entre os valores $\overline{y}_i$ observados e os valores $y_i=f(x_i,a_{1,}a_{2,}\ldots,a_k)$ ajustados.

Um ajuste é linear se for da forma $y(x)=f(x;a,b)=ax+b$. Nesse caso, devemos encontrar os valores dos parâmetros a e b que tornam mínimo o valor da soma dos quadrados dos desvios:

$$S=S(b,a)=\sum_{i=1}^{n}(b+a\,\overline{x}_i-\overline{y}_i)^2. \qquad (2)$$

Tais valores devem satisfazer, necessariamente, às condições:

$$\begin{aligned}\frac{\partial S}{\partial b}=0\Leftrightarrow\sum_{i=1}^{n}2(b+a\,\overline{x}_i-\overline{y}_i)=0,\\ \frac{\partial S}{\partial a}=0\Leftrightarrow\sum_{i=1}^{n}2\,\overline{x}_i(b+a\,\overline{x}_i-\overline{y}_i)=0.\end{aligned} \qquad (3)$$

Desenvolvendo as derivadas parciais e calculando o sistema de equações, chegamos nos seguintes parâmetros:

$$a=\frac{\sum \overline{x_i}\,\overline{y_i}-n\,\overline{xy}}{\sum \overline{x_i}^{\,2}-n\,\overline{x}^{\,2}},$$
$$b=\frac{\sum \overline{x_i}^{\,2}\sum \overline{y_i}-\sum \overline{x_i}\sum \overline{x_i}\,\overline{y_i}}{n\sum \overline{x_i}^{\,2}-\left(\sum \overline{x}\right)^2}. \quad (4)$$

Onde $\overline{x}$ é a média dos valores x_i e $\overline{y}$ a média dos valores de y_i .

O modelo de Verhulst supõe que uma população tende a estabilizar-se, ou seja, a espécie cresce até um limite máximo sustentável, aqui denominado k . Este modelo pode ser considerado como um caso particular do modelo de Lotka-Volterra:

$$\frac{dy}{dt}=ay-by^2, \quad (5)$$

que também pode ser escrita da seguinte forma:

$$\frac{dy}{dt}=ay\left(1-\frac{b}{a}y\right) \text{ ou } \frac{dy}{dt}=ay\left(1-\frac{y}{k}\right), \quad (6)$$

onde o limite máximo sustentável é calculado por $k=\frac{a}{b}$.

Ajuste via regressão linear

A equação de Verhulst é chamada de equação logística e tem como solução geral:

$$y(t)=\frac{k}{\frac{1}{c}e^{-at}+1}. \quad (7)$$

Sendo que para $t=0$ tem-se que $c=\frac{y_0}{k\left(1-\frac{y_0}{k}\right)}$, então $y_0=\frac{ck}{c+1}$. Substituindo c na equação (7), obtém-se:

$$y(t)=\frac{k}{\left(\frac{k-y_0}{y_0}\right)e^{-at}+1}, \text{ ou } y(t)=\frac{k}{\left(\frac{k}{y_0}-1\right)e^{-at}+1}. \quad (8)$$

A partir da equação (8), temos que:

$$y(t)=\frac{k}{\frac{1}{c}e^{-at}+1}. \quad (9)$$

Em seguida são feitas algumas adequações, e é possível chegar na seguinte forma:

$$\frac{y}{k\left(1-\frac{y}{k}\right)}=ce^{at}. \quad (10)$$

Após a equação (10), podemos efetuar a linearização da equação. Para isso, são aplicadas propriedades dos logaritmos, conforme mostrado a seguir:

$$\ln\left[\frac{y}{k\left(1-\frac{y}{k}\right)}\right]=\ln\left(ce^{at}\right), \text{ ou } \ln(y)-\ln\left[k\left(1-\frac{y}{k}\right)\right]=at+\ln c. \quad (11)$$

Chega-se então a forma linearizada da equação de Verhulst:

$$\overline{y}=\overline{a}+\overline{b}\,t, \quad (12)$$

em que $\bar{y}=\ln(y)-\ln\left[k\left(1-\frac{y}{k}\right)\right]$, $\bar{a}=\ln c$ e $\bar{b}=a$.

O estudo foi realizado a partir de dados experimentais coletados da pesquisa de PITOL, et al (2021). Os dados tomados como referência foram os da substância $NaNO_2$, encontrados na análise, e estão dispostos na seguinte tabela:

Tabela 1: dados experimentais $NaNO_2$.

t_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13
y_i	0,64	0,64	0,64	1,17	2,31	3,33	5,08	6,06	7,05	7,5	7,78	7,83	7,97	8

Fonte: (PITOL, et. al., 2021).

A partir dessa tabela, podemos realizar as operações para encontrar a curva de Verhulst. Para isso, admitimos $k=8,5$ como limite máximo sustentável. Os valores encontrados para os parâmetros *a* e *b* são $a=0,0797$ e $b=-2,9966$.

Ajuste sem regressão linear

Para ajustar os pontos da tabela 1 ao modelo de Verhulst sem o uso da regressão linear, é necessário utilizar o método dos mínimos quadrados, conforme a equação (1). Portanto, ao incluirmos (5) em (1) temos a equação de mínimos quadrados para o modelo estudado.

Com o intuito de encontrar os parâmetros a e b do ajuste sem regressão linear, utilizamos o procedimento descrito em (2). A partir disso, realizamos os cálculos das derivadas geradas e por fim calcula-se os parâmetros a e b do sistema algébrico não linear resultante.

Ao desenvolver os cálculos, o programa utilizado não suportou a inserção de 10 ou mais dados, devido a limitação de soluções. Por essa razão, os coeficientes foram calculados a partir dos 9 primeiros dados dispostos na tabela (1). O resultado encontrado para a e b utilizando mínimos quadrados, sem regressão linear, para os 9 primeiros dados, foi $a=0,4592$ e $b=0,0553$.

3. RESULTADOS E DISCUSSÃO

Com a finalidade de identificar o ajuste mais adequado, foi calculado o coeficiente de correlação de Pearson para as funções. O valor deste coeficiente varia no intervalo [–1, 1], sendo que a correlação será melhor quanto mais próxima de 1 ou –1, e pior quanto mais perto de 0. Se o coeficiente r for igual a 1 ou –1, então a correlação é perfeita, e caso $r=0$, não há correlação (BASSANEZI, 2002).

Para calcular o coeficiente de correlação foi utilizada a seguinte fórmula:

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\left(\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2\right)^{1/2}}. \qquad (13)$$

Nesse contexto, a partir das equações geradas para o modelo de Verhulst, foi feita uma comparação entre os dois métodos. Com a finalidade de investigar a equação ideal para esse caso, foi calculado o coeficiente de correlação de Pearson e também o desvio padrão dos dois casos.

Tabela 2: medidas de comparação.

Caso	Desvio Padrão	Pearson
Regressão linear	0.4168279	0.9802182
Sem regressão linear	0.4589146	0.9619822

Também podemos perceber a diferença entre as duas curvas quando observamos o gráfico comparativo que segue abaixo.

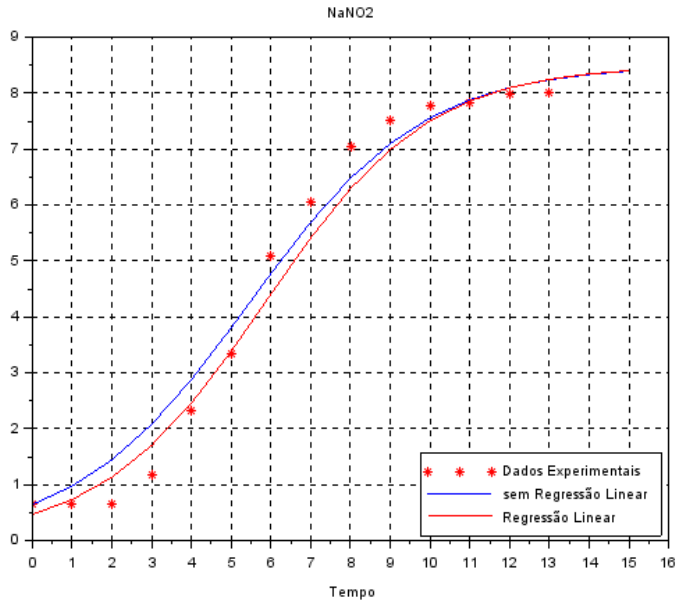

Figura 1: gráfico comparativo de ajustes.

4. CONCLUSÕES

Um dos fatores dificultadores do trabalho foi a impossibilidade do programa utilizado realizar a solução completa do ajuste de todos os pontos. Entretanto, como podemos observar na figura 1, há uma boa proximidade entre as duas curvas, mesmo tendo utilizado somente 9 dos 13 dados da tabela (1) no cálculo dos coeficientes.

O desafio em realizar esse tipo de ajuste é que em certos casos essa metodologia torna-se inviável, dependendo do modelo estudado e dos dados disponibilizados. Além disso, algumas resoluções do sistema algébrico não linear advindo da otimização possuem um número muito grande de soluções, dificultando a utilização de todos os dados experimentais.

A perspectiva de continuação do estudo é fazer a análise do método de mínimos quadrados para aproximação de superfícies.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BASSANEZI, R. C. **Ensino-aprendizagem com Modelagem matemática**. 4ª edição. Campinas, Editora Contexto, 2002, 392 p.

PITOL, L., WEIMANN, T., STROSCHEIN, S. D., ALMANÇA, M. A. K., DE OLIVEIRA, C. S., STIEHL R. Dinâmica populacional de fungos causadores de doenças de tronco em videiras sob a ação de possíveis agentes inibidores. **REMAT Revista Eletrônica da Matemática**. Bento Gonçalves, RS, v. 7, n. 1, p.30, 2021, 12 jan. 2021. DOI: https://doi.org/10.35819/remat2021v7i1id4432.

SOUZA, M. J. F. **Ajuste de Curvas pelo Método dos Quadrados Mínimos.** Notas de aula de Métodos Numéricos, Departamento de Computação/ICEB/UFOP, Disponível em: https://www.decom.ufop.br/prof/marcone/ Acessado em: 07 de mar. 2022.