Método dos Mínimos Quadrados

Uma situação que nos é apresentada com frequência em observações quantitativas de uma variável dependente y, cujos valores são função de um conjunto de variáveis independentes $(x_1, \ldots, x_\ell)$, consiste na consideração de dois ingredientes para análise:

1. **Modelo** - Uma formulação explicita da dependência de y em função das variáveis $\mathbf{x} = (x_1, \ldots, x_\ell)$, expressa em termos de um conjunto de parâmetros $\{a_0, a_1, \ldots, a_m\}$; $\quad a_k \in \mathbb{R}$ e uma família de funções $\{g_0(\mathbf{x}), g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})\}$; $g_k() : \mathbb{R}^\ell \to \mathbb{R}$. No contexto da nossa apresentação vamos considerar apenas casos em que os modelos são lineares nos parâmetros $\{a_0, a_1, \ldots, a_m\}$, ou seja:

$$y = f(\mathbf{x}) = a_0 g_0(\mathbf{x}) + a_1 g_1(\mathbf{x}) + \ldots + a_m g_m(\mathbf{x})$$

Em diversas situações a definição das funções que compreendem a família $\{g_0(x), g_1(x), \ldots, g_m(x)\}$ é resultado de uma análise teórica que resulta em um modelo geral que pode ser considerado para descrever um fenômeno em diferentes contextos. O conjunto de parâmetros $\{a_0, a_1, \ldots, a_m\}\,; \quad a_k \in \mathbb{R}$ devem ser escolhidos de forma a *ajustar* o modelo

$$y = f(x) = a_0 g_0(x) + a_1 g_1(x) + \ldots + a_m g_m(x)$$

a um contexto específico. Assim devemos considerar a família $\{g_0(x), g_1(x), \ldots, g_m(x)\}$ como dada e os valores dos parâmetros $\{a_0, a_1, \ldots, a_m\}\,; \quad a_k \in \mathbb{R}$ devem ser obtidos levando em consideração uma situação específica.

Exemplo

Um movimento de *queda livre* consiste em um movimento com aceleração constante (aceleração da gravidade) e portanto a dependência da velocidade $v(t)$ em função do tempo t é dada por uma expressão da forma:

$$v(t) = v_0 + gt$$

onde os valores de v_0 e g caracterizam um particular contexto, por exemplo um movimento de queda livre de um objeto com velocidade vertical inicial v_0 lançado em uma localidade específica do planeta Terra onde a aceleração da gravidade g tem um valor específico. Neste exemplo:

- A variável dependente é a velocidade $v(t)$
- a variável independente é o tempo t
- O conjunto de parâmetros $\{a_0, a_1\}$ são os valores de v_0 e g. $(v_0 = a_0;\ g = a_1)$
- A familia de funções $\{g_0(t), g_1(t)\}$ está definida por: $g_0(t) = 1$ e $g_1(t) = t$.
- O modelo é $v(t) = a_0 g_0(t) + a_1 g_1(t)$.

2. Resultados Empíricos (*caso discreto*)

Consiste em um conjunto de dados obtidos pela observação de valores da variável dependente y para diferentes valores das variáveis independentes x,

i	x^i	y_i
1	x^1	y_1
$\vdots$	$\vdots$	$\vdots$
n	x^n	y_n

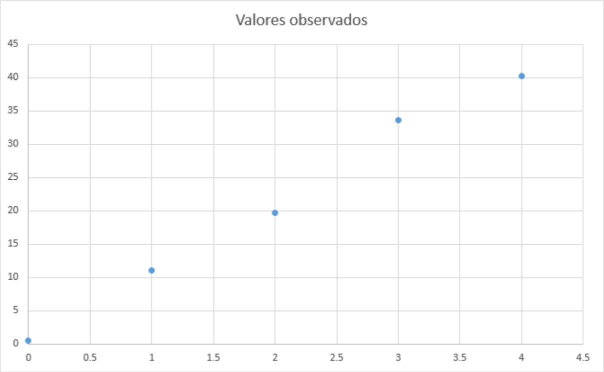
$\left\{(\mathsf{x}^i, y_i)\right\}_{i=1,\ldots,n}$; $\mathsf{x}^i \in \mathbb{R}^\ell$; $y_i \in \mathbb{R}$.

Exemplo

Na observação de um experimento de queda livre foram observados os seguintes valores para $v(t_i)$ em função de valores fixados $\{t_i\}_{i=1,\dots,5}$ da variável independente t.

i	t^i	v_i
1	0	0.51562
2	1	10.97913
3	2	19.70396
4	3	33.66647
5	4	40.2618

t_i	v_i
0	0.51562
1	10.97913
2	19.70396
3	33.66547
4	40.2618

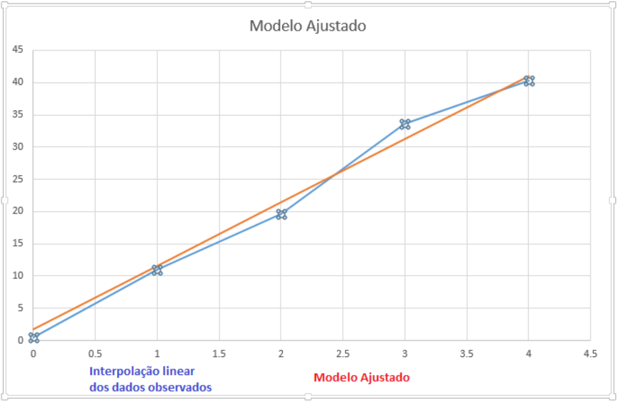


O modelo considerado,

$$v(t) = v_0 + gt$$

explicita uma dependência linear de $v(t)$ como função de t e portanto um gráfico com pontos $\left\{ \left(t^i, v(t^i) \right) \right\}_{i=1,\ldots,5}$ deve consitir de uma reta. Porém é claro que não existe uma reta que contenha o conjunto dos valores observados e portanto dos dados não definem diretamente os valores de v_0 e g.

Desta forma, a definição dos valores dos parâmetros v_0 e g requer a introdução de um critério para a *escolha* da reta que resulta no *melhor ajuste* do modelo aos dados observados (escolher uma reta equivale a definir os valores de v_0 e g).


Modelo Ajustado
45
40
35
30
25
20
15
10
5
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Interpolação linear
dos dados observados
Modelo Ajustado

Para definir um possível critério de melhor ajuste, para cada valor x^i; $i = 1, ..., n$ da variável independente, consideramos a magnitude da diferença entre os valores obtidos y_i para cada um dos valores x^i e os valores previstos pelos modelo

$$y(x^i) = a_0 g_0(x^i) + \ldots + a_m g_m(x^i)$$

(*dependentes da escolha dos parâmetros* $a_0, ..., a_m$)

$$\mid y_i - y(x^i) \mid =$$

$$\mid y_i - [a_0 g_0(x^i) + \ldots + a_m g_m(x^i)] \mid ; \quad i = 1, ..., n$$

Um critério para a escolha dos valores dos parâmetros $a_0, ..., a_m$ poderia ser, considerar a função:

$$E_{ABS}(a_0, ..., a_m) = \sum_{i=1}^{n} | \, y_i - [a_0 g_0(x^i) + \ldots + a_m g_m(x^i)] \, |$$

e escolher os valores de $a_0, ..., a_m$ correspondentes ao *melhor ajuste*, como sendo os valores para os quais $E_{ABS}(a_0, ..., a_m)$ assume seu valor mínimo.

Esta estratégia incorre em uma dificuldade técnica uma vez que a função $E_{ABS}(a_0, ..., a_m)$ é não diferenciável e por isso, o problema de determinar os valores de $a_0, ..., a_m$ para os quais ela assume seu mínimo não tem uma solução imediata.

Para contornar esta dificuldade consideramos em lugar de $E_{ABS}(a_0, ..., a_m)$ a função:

$$EQ(a_0, ..., a_m) = \sum_{i=1}^{n} \left(y_i - [a_0 g_0(x^i) + \ldots + a_m g_m(x^i)]\right)^2$$

e os valores de $a_0, ..., a_m$ para os quais $EQ(a_0, ..., a_m)$ é mínimo são obtidos resolvendo:

$$\begin{cases} \frac{\partial EQ}{\partial a_0}(a_0, ..., a_m) = 0 \\ \vdots \\ \frac{\partial EQ}{\partial a_m}(a_0, ..., a_m) = 0 \end{cases}$$

Como a função $EQ(a_0, ..., a_m)$ é uma função quadradica nas variáveis $a_0, ..., a_m$, as derivadas parciais $\frac{\partial EQ}{\partial a_j}(a_0, ..., a_m)$; $j = 0, ...m$ são funções lineares e portanto o sistema de equações acima consiste em um sistema linear com $m + 1$ equações e $m + 1$ incógnitas $a_0, ..., a_m$.