Ajuste de Curvas - python

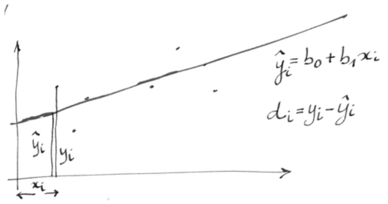
Ao estudar um fenômeno, pode-se propor um modelo matemático. Nele, definem-se

- variável resposta (ou variável dependente) – é aquela que é respondida pelo modelo, ou seja, é a variável de saída do modelo.
- variável explicativa (ou independente) – que é aquela exigida pelo modelo ou seja é a variável de entrada.

Para estudar melhor o fenômeno, deve-se construir um gráfico $x0y$ conhecido como gráfico de dispersão, para verificar visualmente como os dados se distribuem. É bom lembrar que a decisão entre quem é entrada e quem é saída é uma decisão puramente arbitrária. A matemática "lava as mãos" nessa hora. Para ela tanto faz se *o cachorro abana o rabo ou ao contrário é o rabo que abana o cachorro.*

Este procedimento é parecido com a interpolação, mas lá havia a preocupação de analisar os dados desconhecidos apenas entre os limites obtidos na experimentaão (senão seria extrapolação). Aqui não: no ajuste de curvas busca-se a expressão de uma curva que se extende ao infinito no sentido matemático.

Veja-se um exemplo de um gráfico com alguns pontos:



Cada ponto no gráfico tem as suas coordenadas x e y e buscou-se uma reta ($y = b_0 + b_1.x$) que busque reproduzir o comportamento da distribuição dos pontos examinados. Cada ponto tem sua ordenada y real (y_i) e a ordenada correspondente $\hat{y}_i = b_0 + b_1 x_i$.

Para cada ponto i, pode-se calcular a distância entre y_i e $\hat{y}_i$ através da expressão $d+i = y_i - \hat{y}_i$. A soma das distâncias de todos os pontos levantados é

$$D = \sum_{j=1}^{n} d_i$$

D é a somatória das diferenças entre o valor real de y e aquele dado pela reta $y = b_0 + b_i.x$. Como D pode ser para mais ou para menos, o somatório simples pode mascarar os erros e não serve. Uma possível solução seria tomar o módulo ($|d_i|$) que resolveria o problema do sinal, mas criaria outro: devido a suas decontinuidades o módulo não tem uma boa derivada.

Resta tomar o quadrado das diferenças (d_i^2), que resolve os 2 problemas. Daqui, vem o nome do método que o dos "mínimos quadrados".

Então,

$$D = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

. Substituindo d_i^2 por $(y_i - \hat{y}_i)^2$, chamado D de $D(b_0, b_1)$ que é a reta determinada por b_o e b_1 e substituindo $\hat{y}_i$ por $b_0 - b_1 x_i$ fica

$$D(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Derivando esta equação em relação a b_0 fica:

$$\frac{\partial D}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1.x_i)$$

Derivando a mesma equação em relação a b_1 fica:

$$\frac{\partial D}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1.x_i).x_i$$

Os valores de b_0 e b_1 em que a função $D(b_0, b_1)$ apresenta um ponto de mínimo são obtidos – as usual – igualando-se as derivadas a zero. Fica: $-2\sum_{i=1}^{n}(y_i - b_0 - b_1.x_i) = 0$ e $-2\sum_{i=1}^{n}(y_i - b_0 - b_1.x_i).x_i = 0$. Daí, $\sum y_i - \sum b_0 - \sum b_1.x_i = 0$ e $\sum x_i y_i - \sum b_0 x_i - \sum b_1.x_i^2 = 0$. Chamando $\sum$ de S fica: $Sy_i - Sb_0 - Sb_1 x_i = 0$ e
$Sy_i x_i - Sb_0 x_i - Sb_1 x_i^2 = 0$. Reescrevendo
$Sy_i = nb_0 + Sx_i b_1$ e
$Sy_i x_i = Sx_i.b_0 + Sx_i^2 b_1$ que é um sistema linear. Na sua forma matricial

$$\begin{bmatrix} n & Sx_i \\ Sx_i & Sx_i^2 \end{bmatrix} . \begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{Bmatrix} Sy_i \\ Sy_i x_i \end{Bmatrix}$$

O objetivo final é obter b_0 e b_1 que, lembrando são os coeficientes da reta procurada ($\hat{y}_i = b_o + b_1 x_i$). A solução deste sistema pelo método de escalonamento de Gauss é

$$b_1 = \frac{n.\sum x_i y_i - \sum x_i \sum y_i}{n.\sum x_i^2 - (\sum x_i)^2}$$

e

$$b_0 = \frac{\sum y_i - (\sum x_i).b_1}{n}$$

Obtidos estes parâmetros pode-se escrever a "menor" reta como

$$Y = b_0 + b_1 X$$

Seja um exemplo:

i	1	2	3	4	5
x_i	1.3	3.4	5.1	6.8	8.0
y_i	2.0	5.2	3.8	6.1	5.8

Calculando:
$n = 5$
$\sum x_i = ... = 24.6$
$\sum x_i^2 = (1.3)^2 + (3.4)^2 + ... = 149.50$
$\sum y_i = 22.9$
$\sum x_i y_i = (1.3 \times 2.0) + (3.4 \times 5.2) + ... = 127.54$

$$b_1 = \frac{5 \times 127.54 - 24.6 \times 22.9}{5 \times 149.50 - 24.6^2} = 0.522$$

e

$$b_0 = \frac{22.9 - (24.6 \times 0.522)}{5} = 2.01$$

e daqui

$$Y = 2.01 + 0.522X$$

Python

```
# minquad
def minquad(x,y):
    import numpy as np
    n=len(x)
    sy=0
    sx=0
    sx2=0
    sxy=0
    sy2=0
    i=0
    while i<n:
       sx=sx+x[i]
       sy=sy+y[i]
       sx2=sx2+(x[i]**2)
       sxy=sxy+(x[i]*y[i])
       sy2=sy2+y[i]**2
       i=i+1
    print('sy ',sy)
    print('sx ',sx)
    print('sx2 ',sx2)
    print('sxy ',sxy)
    print('n ',n)
    print('sy2 ',sy2)
    b1=((sy*sx)-(n*sxy))/((sx**2)-(sx2*n))
    b0=(sy-(sx*b1))/n
    num=(sxy-((sx*sy)/n))**2
    den=((sx2-((sx**2)/n))*(sy2-((sy**2)/n)))
    print("coef = ",num/den)

    while (1==1):
       xx=float(input("Informe x (-1 para fim)"))
       if xx==-1:
          return
       yy=b0+(b1*xx)
       print("y = ",yy)
#para chamar faca: x=[1.3,3.4,5.1,6.8,8.0] e
# y=[2.0, 5.2, 3.8, 6.1, 5.8]
#depois: minquad(x,y) e informe quantos x_i
# quiser, seguido de -1
```

Coeficiente de determinação

mede a qualidade do ajuste linear, isto é quão próxima é a reta determinada em relação ao conjunto de pontos. É um valor (R^2) que varia entre zero e um ($0 \leq R^2 \leq 1$). Quanto mais perto de 1, melhor o ajuste. A fórmula é

$$R^2 = \frac{[\sum xy - ((\sum x \sum y)/n)]^2}{[\sum x^2 - ((\sum(x))^2/n)].[\sum y^2 - ((\sum(y))^2/n)]}$$

no exemplo acima, o número que faltou calcular é $\sum y^2 = 116.33$ e fica:

$$R^2 = \frac{[127.54 - 24.6 \times 22.9/5]^2}{149.5 - 24.6^2/5] \times [116.33 - 22.9^2/5]}$$

ou $R^2 = 0.679$.

☞ Para você fazer

Para testar o algoritmo implementado, faça:

```
x = 26 31 33 36 38 41 42 44
y = 40 41 45 49 51 55 56 58
a = 9.249642005 e b = 1.103102625
para x = 45, y = 58.88  e para x = 50, y = 64.40
```

☞ Para você fazer

1. No exercício a seguir, considere valores de x=

45 46 48 49 55 61 62 65

associados aos seguintes valores de y=

279 283 299 301 346 380 384 401

Calcule os parâmetros b_0 e b_1, e depois infira os valores de y para dois valores de x dados (x_1 e x_2)

66 70

Para calcular a resposta, arredonde para 2 casas decimais

y_1	y_2

2. Em um país distante chamado LISARB, há muitos problemas econômicos. Em particular, há uma inflação persistente que APARENTEMENTE depende entre outras coisas, da emissão de moeda pelo Banco Central daquele país. Economistas que apoiam esta visão levantaram a inflação e a emissão de moeda nos últimos 10 anos, obtendo as duas séries históricas, a seguir:

```
EMISSAO ANUAL DE MOEDA NOS ULTIMOS
10 ANOS (em bilhoes de dólares)
221  331  293  398  454  505  579  745  643  895

INFLAÇAO ANUAL NOS ULTIMOS 10 ANOS
5.6  5.2  5.6  5.6  6.5  6.6  7.6 10.2  8.8  8.3
```

Para o ano próximo, o governo pretende efetuar uma emissão de

662

bilhões de dólares. Segundo este modelo econométrico, baseado no método dos mínimos quadrados, qual é a inflação esperada ?

inflação