**CC0288 - Inferência Estatística I**

**Método dos Mínimos Quadrados - 14/04/2023.**

**Prof. Maurício**

# Método dos Mínimos Quadrados

Vamos analisar a seção 11.4 do livro do Bussab&Morettin.

Um dos procedimentos mais usados para obter estimadores é aquele que se baseia no princípio dos mínimos quadrados, introduzido por Gauss em 1794, mas que primeiro apareceu com esse nome no apêndice do tratado de Legendre, Nouvelles Méthodes pour la Determination des Orbites des Comètes, publicado em Paris em 1806. Gauss somente viria a publicar seus resultados em 1809, em Hamburgo. Ambos utilizaram o princípio em conexão com problemas de Astronomia e Física.

Vejamos o procedimento por meio de um exemplo simples.

**Exemplo 11.9** Um engenheiro está estudando a resistência $Y$ de uma fibra em função de seu diâmetro $X$ e notou que as variáveis são aproximadamente proporcionais, isto é, elas obedecem à relação

$$Y \approx \theta X, \qquad (11.25).$$

em que $\theta$ é o coeficiente de proporcionalidade. Agora ele deseja estimar o parâmetro $\theta$ , baseado numa amostra de cinco unidades, que, submetidas a mensuração e testes, produziram os resultados:

| | | | | | | |
|---|---|---|---|---|---|---|
| X: | 1,2 | 1,5 | 1,7 | 2 | 2,6 | $\bar{X} = 1,8$ |
| Y | 3,9 | 4,7 | 5,6 | 5,8 | 7,0 | $\bar{Y} = 5,4$ |

Vamos fazer um diagrama de dispersão :

Foi feito um pequeno programa no **R**:

```
 X=c(12,15,17,20,26)/10;X
[1] 1.2 1.5 1.7 2.0 2.6
> Y=c(39,47,56,58,70)/10;Y
[1] 3.9 4.7 5.6 5.8 7.0
> n=length(Y);n
[1] 5
>
> Xb=mean(X);Xb
```

```
[1] 1.8
>
> Yb=mean(Y);Yb
[1] 5.4
>
> plot(X,Y,xlim=c(0,3),ylim=c(0,8))
> abline(a=0,b=3,col="red")
>
```

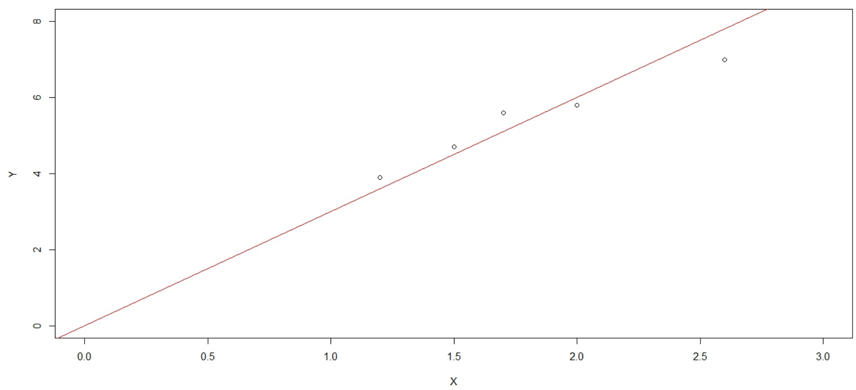


Figura 1:

O valor $a$ é o coeficiente linear e o valor $b$ o coeficiente angular da reta que passa pela origem. Inspecionando os resultados, conclui-se que $\hat{\theta} = 3$ parece ser um valor razoável.

Como verificar a qualidade dessa estimativa ? Podemos utilizar o modelo

$$\hat{Y} = 3X$$

e ver como esse prevê os valores de $Y$, para os dados valores de $X$, e como são as discrepâncias entre os valores observados e os estimados pelo modelo.

Essa análise está resumida na Tabela 11.1

| X | Y | $\hat{Y} = 3X$ | $e = Y - \hat{Y}$ | $e^2 = (Y - \hat{Y})^2$ |
|---|---|---|---|---|
| 1,2 | 3,9 | 3,6 | 0,3 | 0,09 |
| 1,5 | 4,7 | 4,5 | 0,2 | 0,04 |
| 1,7 | 5,6 | 5,1 | 0,5 | 0,25 |
| 2,0 | 5,8 | 6 | -0,2 | 0,04 |
| 2,6 | 7,0 | 7,8 | -0,8 | 0,64 |
| | | Total | 0 | 1,06 |

Os valores da coluna $(Y - 3X) = (Y - \bar{Y})$ medem a inadequação do modelo para cada observação da amostra, enquanto o valor é uma tentativa de medir o erro quadrático total da amostra?. Como em situações anteriores, elevou-se ao quadrado para evitar o problema do sinal.

Quanto menor for o erro quadrático total, melhor será a estimativa.

Isso nos sugere procurar a estimativa que torne mínima essa soma de quadrados. Matematicamente, o problema passa a ser o de encontrar o valor de $\theta$ que minimize a função

$$S(\theta) = \sum_{i=1}^{5} (Y_i - \theta X_i)^2, \qquad (11.26)$$

O mínimo da função é obtido derivando-a em relação a $\theta$, e igualando o resultado a zero (ver Morettin et al., 2005), o que resulta

$$S'(\theta) = 2 \sum_{i=1}^{5} (Y_i - \theta X_i)(-X_i),$$

$$S'(\theta) = -2 \sum_{i=1}^{5} (X_i\, Y_i - \theta X_i^2) = 0$$

$$S'(\theta) = -2 \sum_{i=1}^{5} X_i\, Y_i + 2\theta \sum_{i=1}^{5} X_i^2.$$

A derivada segunda de $S$ é dada por:

$$S^{''}(\theta) = 2\sum_{i=1}^{5} X_i^2 > 0$$

Igualando a derivada primeira a zero temos:

$$\sum_{i=1}^{5} X_i\, Y_i - \theta \sum_{i=1}^{5} X_i^2 = 0$$

Assim

$$\theta = \frac{\sum_{i=1}^{5} X_i\, Y_i}{\sum_{i=1}^{5} X_i^2},$$

que é ponto de mínimo relativo pois a derivada segunda é sempre positiva.

O estimador de mínimos quadrados de $\theta$ é dado por:

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{5} X_i\, Y_i}{\sum_{i=1}^{5} X_i^2}.$$

Usando os dados acima encontramos $\hat{\theta}_{MQ} = 2,94$ , que conduz a um valor mínimo para $S(\theta)$ de $0,94$ . Observe que esse valor é realmente menor do que o observado para $\theta = 3$, ou seja, $1,06$.

Vamos utilizar o **R** para calcular a estimativa de mínimos quadrados de $\theta$ bem como os valores previstos pelo modelo.

```
> X=c(12,15,17,20,26)/10;X
[1] 1.2 1.5 1.7 2.0 2.6
> Y=c(39,47,56,58,70)/10;Y
[1] 3.9 4.7 5.6 5.8 7.0
> n=length(Y);n
[1] 5
>
> SX=sum(X);SX
[1] 9
> X2=X^2
>
> SX2=sum(X2);SX2
[1] 17.34
> XY=X*Y
>
> SXY=sum(XY);SXY
[1] 51.05
>
> teta_est=SXY/SX2;teta_est
[1] 2.94406
>
> round(teta_est,2)  ##Estimativa de mínimos quadrados.
[1] 2.94
>
> Y_est=teta_est*X;Y_est
[1] 3.532872 4.416090 5.004902 5.888120 7.654556
>
> e=Y-Y_est
> e2=e^2
>
> sum(e2)
[1] 1.005738
>
> tabela=cbind(X,Y,X2,XY,Y_est,e,e2);tabela
X   Y   X2    XY    Y_est           e          e2
[1,] 1.2 3.9 1.44  4.68 3.532872  0.36712803 0.134782989
[2,] 1.5 4.7 2.25  7.05 4.416090  0.28391003 0.080604908
[3,] 1.7 5.6 2.89  9.52 5.004902  0.59509804 0.354141676
[4,] 2.0 5.8 4.00 11.60 5.888120 -0.08811995 0.007765126
[5,] 2.6 7.0 6.76 18.20 7.654556 -0.65455594 0.428443479
>
```

Vamos fazer agora com uma função direta do **R**:

```
> ####Fazer direto no R:
>
>
```

```
> ###O (-1) no comando indica que o coeficiente linear é zero!!!!!!
>
>
>
>
> mod1=lm(Y~X -1);mod1

Call:
lm(formula = Y ~ X - 1)

Coefficients:
X
2.944

>
> predict(mod1)
1        2        3        4        5
3.532872 4.416090 5.004902 5.888120 7.654556
>
>
> summary(mod1)

Call:
lm(formula = Y ~ X - 1)

Residuals:
1        2        3        4        5
0.36713  0.28391  0.59510 -0.08812 -0.65456

Coefficients:
Estimate Std. Error t value Pr(>|t|)
X   2.9441     0.1204   24.45 1.66e-05 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 0.5014 on 4 degrees of freedom
Multiple R-squared:  0.9934,    Adjusted R-squared:  0.9917
F-statistic: 597.7 on 1 and 4 DF,  p-value: 1.661e-05

> anova(mod1)
Analysis of Variance Table

Response: Y
Df  Sum Sq Mean Sq F value    Pr(>F)
X          1 150.294 150.294  597.75 1.661e-05 ***
Residuals  4   1.006   0.251
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
>
> ###### ##############################################################
>
> SY2=sum(Y^2);SY2
```

```
[1] 151.3
>
> ###Observe que:
>
>
>
> 150.294 + 1.006
[1] 151.3
>
```

Como foi dito, não esperávamos uma relação perfeita entre as duas variáveis, já que o diâmetro da fibra não é o único responsável pela resistência; outros fatores não controlados afetam o resultado.

Desse modo, duas amostras obtidas do mesmo diâmetro $X$ não teriam obrigatoriamente que apresentar o mesmo resultado $Y$, mas valores em torno de um valor esperado $\theta X$.

Em outras palavras, estamos supondo que, para um dado valor da variável explicativa $X$, os valores da variável resposta $Y$ seguem uma distribuição de probabilidade $f_Y(y)$, centrada em $\theta X$.

Isso equivale a afirmar que, para cada $X$, o desvio

$$\epsilon = Y - \theta X$$

segue uma distribuição centrada no zero.

Para melhor entendimento dessa proposição, veja o Capítulo 16. Podemos, então, escrever

$$E(Y|x) = \theta x, \text{ para todo valor } x.$$

É comum supor que $\epsilon$ tem a mesma distribuição, para todo valor $x$ da variável explicativa $X$. Desse modo, é comum escrever

$$Y = \theta\, x + \epsilon$$

, com $\epsilon$ seguindo a distribuição $f_\epsilon(.)$, com média zero. Como ilustração, poderíamos supor que

$$\epsilon \sim N(0, \sigma^2), \quad \text{para todo } x.$$

Quanto menor for a variância $\sigma^2$ , melhor será a ?previsão? de $Y$ como função de $x$ . Assim, parece razoável escolher $\theta$ que torna mínima a soma dos quadrados do erros: O modelo acima pode ser generalizado, de modo a envolver outras funções do parâmetro, resultando no modelo

$$Y = g(X; \theta) + \epsilon, \qquad (11.27)$$

Note que

$$E(Y) = g(X; \theta),$$

e devemos procurar o valor de $\theta$ que minimize a função

$$S(\theta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (Y_i - E(Y_i))^2 = \sum_{i=1}^{n} (Y_i - g(X_i;\theta))^2, \qquad (11.28)$$

Para uma amostra $(X_1, Y_1), \ldots, (X_n, Y_n)$ das variáveis $X$ e $Y$. A solução MQ é chamada de estimador de mínimos quadrados (EMQ) de $\theta$ .

Nos Capítulos 15 e 16, voltaremos a esse tópico e trataremos com mais detalhes os chamados modelos lineares.

Suponha que

$$Y_i|X_i = \theta X_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

com

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \ldots, n$$

Na prática usamos:

$$Y_i = \theta X_i + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

Aplicando o operador esperança temos:

$$E\left(Y_i\right) = \theta X_i + E(\epsilon_i) = \theta X_i + 0 = \theta X_i.$$

Calculando a variância temos:

$$Var\left(Y_i\right) = Var\left(\theta X_i + \epsilon_i\right) = Var(\epsilon_i) = \sigma^2.$$

Considere o estimador de MQ de $\theta$ dado por:

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{n} X_i\, Y_i}{\sum_{i=1}^{n} X_i^2} = \sum_{i=1}^{n} w_i\, Y_i,$$

com

$$w_i = \frac{X_i}{\sum_{i=1}^{n} X_i^2}.$$

Inicialmente mostre que:

$$\sum_{i=1}^{n} w_i\, X_i = 1.$$

**Prova**

Temos que :

$$\sum_{i=1}^{n} w_i \, X_i = \sum_{i=1}^{n} \frac{X_i}{\sum_{i=1}^{n} X_i^2} \times X_i.$$

$$\sum_{i=1}^{n} w_i \, X_i = \frac{1}{\sum_{i=1}^{n} X_i^2} \sum_{i=1}^{n} X_i^2 = 1.$$

Agora mostre que $\hat{\theta}_{MQ}$ é um estimador não viciado de $\theta$.

$$E\left(\hat{\theta}_{MQ}\right) = E\left(\sum_{i=1}^{n} w_i Y_i\right),$$

$$E\left(\hat{\theta}_{MQ}\right) = \sum_{i=1}^{n} E(w_i \, Y_i) = \sum_{i=1}^{n} w_i \, E(\, Y_i)$$

$$E\left(\hat{\theta}_{MQ}\right) = \sum_{i=1}^{n} w_i \, \theta \, X_i = \theta \sum_{i=1}^{n} X_i w_i$$

$$E\left(\hat{\theta}_{MQ}\right) = \theta \times 1 = \theta.$$

Agora mostre que:

$$\sum_{i=1}^{n} w_i^2 = \frac{1}{\sum\limits_{i=1}^{n} X_i^2}.$$

**Prova:**
Seja

$$A = \sum_{i=1}^{n} X_i^2$$

.

Assim

$$w_i = \frac{X_i}{A} \quad , w_i^2 = \frac{X_i^2}{A^2}$$

$$\sum_{i=1}^{n} w_i^2 = \sum_{i=1}^{n} \frac{X_i^2}{A^2} = \frac{1}{A^2} \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} w_i^2 = \frac{1}{A^2} A = \frac{1}{A} = \frac{1}{\sum\limits_{i=1}^{n} X_i^2}$$

Agora mostre que a variância do estimador de mínimos quadrados de $\theta$ é dada por:

$$Var\left(\hat{\theta}_{MQ}\right) = \frac{\sigma^2}{\sum\limits_{i=1}^{n} X_i^2}.$$

**Prova:**

$$Var\left(\hat{\theta}_{MQ}\right) = Var\left(\sum_{i=1}^{n} w_i Y_i\right) = \sum_{i=1}^{n} Var(w_i Y_i),$$

$$Var\left(\hat{\theta}_{MQ}\right) = \sum_{i=1}^{n} w_i^2 \; V(Y_i) = \sum_{i=1}^{n} w_i^2 \; \sigma^2,$$

$$Var\left(\hat{\theta}_{MQ}\right) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}.$$

Qual a distribuição Amostral do estimador de mínimos quadrados $V = \hat{\theta}_{MQ}$?

**Prova:** Sabemos que

$$Y_i \sim N(\theta X_i, \sigma^2)$$

e sua função geradora de momentos é dada por:

$$M_{Y_i}(t) = \exp(\theta X_i \; t + \frac{1}{2}\sigma^2 \; t^2)$$

A função de momentos de $V$ é dada por:

$$M_V(t) = E\left(e^{tV}\right) = E\left(e^{t \sum_{i=1}^{n} w_i Y_i}\right)$$

$$M_V(t) = E\left(e^{\sum_{i=1}^{n} w_i \; t \; Y_i}\right)$$

$$M_V(t) = \prod_{i=1}^{n} E\left(e^{wi \; t \; Y_i}\right),$$

$$M_V(t) = \prod_{i=1}^{n} M_{Y_i}(w_i t)$$

$$M_V(t) = \prod_{i=1}^{n} \exp(\theta X_i \; w_i \; t + \frac{1}{2}\sigma^2 \; w_i^2 \; t^2)$$

$$M_V(t) = \exp\left(\theta \; t \sum_{i=1}^{n} X_i \; w_i \; + \frac{1}{2}\sigma^2 \sum_{i=1}^{n} w_i^2 \; t^2\right)$$

$$M_V(t) = \exp\left(\theta \; t \; \times 1 + \frac{1}{2} \times \sigma^2 \times \frac{1}{\sum_{i=1}^{n} X_i^2} \; t^2\right)$$

$$M_V(t) = \exp\left(\theta\, t \; + \frac{1}{2} \times \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}\; t^2\right)$$

Assim

$$V = \hat{\theta}_{MQ} \sim N\left(\theta, \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}\right)$$

Qual o estimador de mínimos quadrados baseado em uma amostra aleatória de tamanho $n$, $X_1, X_2, \ldots, X_n)$, de uma variável $X$ com distribuição:

a. $X \sim Poisson(\theta)$

b. $X \sim N(\mu, 1)$

c. $X \sim Ber(p)$.

**Solução:** Devemos minimizar
Sabemos que

$$E(X_i) = \theta, \;\; i = 1, 2, \ldots, n.$$

$$S(\theta) = \sum_{i=1}^{n} (X_i - E(X_i))^2 = \sum_{i=1}^{n} (X_i - \theta)^2$$

$$S(\theta) = \sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\, \theta + n\theta^2 = a\theta^2 + b\theta + c$$

Como $a > 0$ temos

$$\theta_{min} = \frac{-b}{2a} = \frac{2\sum_{i=1}^{n} X_i}{2n} = \bar{X}$$

que é o nosso estimador de mínimos quadrados.
No caso da normal temos:
Sabemos que

$$E(X_i) = \mu, \;\; i = 1, 2, \ldots, n.$$

$$S(\mu) = \sum_{i=1}^{n} (X_i - E(X_i))^2 = \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\mu_{min} = \bar{X},$$

que é o nosso estimador de mínimos quadrados.
No caso da Bernouli temos

$$E(X_i) = p, \quad i = 1, 2, \ldots, n.$$

$$S(p) = \sum_{i=1}^{n} (X_i - E(X_i))^2 = \sum_{i=1}^{n} (X_i - p)^2$$

$$S(p) = np^2 - 2\sum_{i=1}^{n} X_i \, p + \sum_{i=1}^{n} X_i^2$$

$$S'(p) = 2pn - 2\sum_{i=1}^{n} X_i$$

$$S''(p) = 2n > 0$$

De $S'(p) = 0$ temos

$$p_{min} = \bar{X},$$

que é o nosso estimador de mínimos quadrados.
**Problemas**

6. Estamos estudando o modelo $y_t = \mu + \epsilon_t$, para o qual uma amostra de cinco elementos produziu os seguintes valores para $y_t : 3, 5, 6, 8, 16$.

(a) Calcule os valores de

$$S(\mu) = \sum_{t} (y_t - \mu)^2$$

para $\mu = 6, 7, 8, 9, 10$, e faça o gráfico de $S(\mu)$ em relação a $\mu$.

Qual o valor que parece tornar mínimo $S(\mu)$?

(b) Derivando $S(\mu)$ em relação a $\mu$ , e igualando o resultado a zero, você encontrará o EMQ de $\mu$. Usando os dados acima, encontre a estimativa para $\mu$ e compare com o resultado do item anterior.

7. Os dados abaixo referem-se ao índice de inflação ($y_t$) de 1967 a 1979.

| Ano(t) | 1967 | 1969 | 1971 | 1973 | 1975 | 1977 | 1979 |
|---|---|---|---|---|---|---|---|
| Inflação($y_t$) | 128 | 192 | 277 | 373 | 613 | 1236 | 2639 |

(a) Faça o gráfico de $y_t$ contra $t$.

(b) Considere ajustar o modelo

$$y_t = \alpha + \beta t + \epsilon_t$$

aos dados. Encontre as estimativas de mínimos quadrados de $\alpha$ e $\beta$.

(c) Qual seria a inflação em 1981?

(d) Você teria alguma restrição em adotar o modelo linear nesse caso ?

8. No Problema 7, determinamos os estimadores de mínimos quadrados para o modelo

$$y_t = f(t) + \epsilon_t$$

, no qual

$$f(t) = \alpha + \beta\, t$$

.

Suponha agora que

$$f(t) = \alpha + \beta\, x_t,\ t = 1, 2, \ldots, n,$$

ou seja, temos $n$ valores fixos $x_1, x_2, \ldots, x_n$ uma variável fixa (não aleatória) $x$x. Obtenha os EMQ de $\alpha$ e $\beta$ para esse modelo.

9. Aplique os resultados do Problema 8 para os dados a seguir:

| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_t$ | 1,5 | 1,8 | 1,6 | 2,5 | 4,0 | 3,8 | 4,5 | 5,1 | 6,5 | 6,0 |
| $y_t$ | 66,8 | 67,0 | 66,9 | 67,6 | 68,9 | 68,7 | 69,3 | 69,8 | 71,0 | 70,6 |