Exemplo dado em sala do livro do Professor Rodolfo Hoffmann.

X	Y
0	3
1	2
1	3
2	5
3	4
3	4
4	7
5	6
5	7
6	9

a. Faça um diagrama de dispersão. Podemos usar um modelo de regressão linear? Explicite o modelo bem como suas suposições.
b. Obtenha as estimativas de mínimos quadrados dos coeficientes linear e angular detalhando um quadro com seus cálculos.

Solução: Vamos completar nosso quadro:

i	X	Y	XY	X^2	Y^2
1	0	3	0	0	9
2	1	2	2	1	4
3	1	3	3 7	1	9
4	2	5	10	4	25
5	3	4	12	9	16
6	3	4	12	9	16
7	4	7	28	16	49
8	5	6	30	25	36
9	5	7	35	25	49
10	6	9	54	36	81
Soma	30	50	186	126	294

Temos uma amostra de $n = 10$ pares e as principais somas de quadrados são dadas por:

$$\sum_{i=1}^{10} X_i = 30\ ; \sum_{i=1}^{10} X_i^2 = 126, \quad \sum_{i=1}^{10} X_i\, Y_i = 186\ ;$$

$$\sum_{i=1}^{10} Y_i = 50\ ; \sum_{i=1}^{10} Y_i^2 = 294$$

As médias amostrais de X e Y são dadas por:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{30}{10} = 3.$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{50}{10} = 5.$$

Vamos centrar as variáveis:

$$x = X - \bar{X} = X - 5, y = Y - \bar{Y}$$

Vamos analisar a saída do R a seguir:

```
> X=c(0,1,1,2,3,3,4,5,5,6)
>
>
> n=length(X);n
[1] 10
>
> SX=sum(X);SX;SX2=sum(X^2);SX2
[1] 30
[1] 126
>
> Y=c(3,2,3,5,4,4,7,6,7,9)
>
> SY=sum(Y);SY;SY2=sum(Y^2);SY2
[1] 50
[1] 294
>
>
> XY=X*Y;X2=X^2;Y2=Y^2
> mean(X);mean(Y)
[1] 3
[1] 5
>
> x=X-mean(X);y=Y-mean(Y)
>
> xy=x*y;x2=x^2;y2=y^2
>
> i=1:n
>
> tab=cbind(i,X,Y,XY,X2,Y2,x,y,xy,x2,y2)
> tab
i X Y XY X2 Y2  x  y xy x2 y2
[1,]  1 0 3  0  0  9 -3 -2  6  9  4
[2,]  2 1 2  2  1  4 -2 -3  6  4  9
[3,]  3 1 3  3  1  9 -2 -2  4  4  4
[4,]  4 2 5 10  4 25 -1  0  0  1  0
[5,]  5 3 4 12  9 16  0 -1  0  0  1
[
[7,]  7 4 7 28 16 49  1  2  2  1  4
[8,]  8 5 6 30 25 36  2  1  2  4  1
[9,]  9 5 7 35 25 49  2  2  4  4  4
```

```
[10,] 10 6 9 54 36 81  3  4 12  9 16
>
> Sx=sum(x);Sy=sum(y);Sxy=sum(xy);Sx2=sum(x2);Sy2=sum(y^2)
> Sx;Sy;Sxy;Sx2;Sy2
[1] 0
[1] 0
[1] 36
[1] 36
[1] 44
>
```

Vamos estimar o coeficiente angular β_1

$$b_1 = \frac{n \times SXY - SX \times SY}{n \times SX2 - SX^2} = \frac{10 \times 186 - 30 \times 50}{10 \times 126 - 30^2} = \frac{360}{360} = 1$$

Você poderia ter calculado direto:

$$b_1 = \frac{Sxy}{Sx2} = \frac{36}{36} = 1.$$

em que

$$Sxy = SXY - \frac{SXSY}{n} = SXY - n\bar{X}\bar{Y} = 186 - 10 \times 3 \times 5 = 186 - 150 = 36.$$

$$Sx2 = SX2 - \frac{SX^2}{n} = SXY - n\bar{X}^2 = 126 - 10 \times 9 = 126 - 90 = 36.$$

A estimativa do coeficiente linear é dada por:

$$b_0 = \bar{Y} - b_1 \times \bar{X} = 5 - 1 \times 3 = 2.$$

A reta de regressão estimada é:

$$\hat{Y}_i = 2 + X_i.$$

Resolva matricialmente: Vamos usar a notação da disciplina de Regressâo:

```
> A=matrix(c(n,SX,SX,SX2),ncol=2);A
[,1] [,2]
[1,]   10   30
[2,]   30  126
> det(A)
[1] 360
>
> IA=solve(A);IA
[,1]        [,2]
[1,]  0.35000000 -0.08333333
[2,] -0.08333333  0.02777778
>
```

```
> require(MASS)
>
> fractions(IA)
[,1]  [,2]
[1,]  7/20 -1/12
[2,] -1/12  1/36
>
> B=matrix(c(SY,SXY),ncol=1);B
[,1]
[1,]   50
[2,]  186
>
>
> b=IA%*%B;b
[,1]
[1,]    2
[2,]    1
>
> b_0=b[1];b_0
[1] 2
>
> b_1=b[2];b_1
[1] 1
>
```

Calcule as somas de quadrados total ,de regressão e Residual:

$$SQT = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = 294 - 250 = 44.$$

$$SQreg = b_1 \times Sxy = 1 \times 36 = 36.$$

A soma de quadrados residual é dada por:

$$SRes = SQT - SQReg = 44 - 36 = 8.$$

Veja na tabela

```
> Y_est=b0+b1*X
>
>
> e=Y-Y_est;e
[1]  1 -1  0  1 -1 -1  1 -1  0  1
>
> Se=sum(e);Se
[1] 0
>
> Se2=sum(e^2);Se2
[1] 8
```

```
>
> SQT=SY2-SY^2/n;SQT
[1] 44
>
> SQReg=b1*Sxy;SQReg
[1] 36
>
> SQRes=SQT-SQReg;SQRes
[1] 8
> S2=SQRes/(n-2);S2
[1] 1
>
> tab1=cbind(i,Y,Y_est,e);tab1
      i Y    Y_est  e
[1,]  1 3      2   1
[2,]  2 2      3  -1
[3,]  3 3      3   0
[4,]  4 5      4   1
[5,]  5 4      5  -1
[6,]  6 4      5  -1
[7,]  7 7      6   1
[8,]  8 6      7  -1
[9,]  9 7      7   0
[10,] 10 9     8   1
```

Agora fazer tudo no R:

```
> mod=lm(Y~X);mod

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)            X
2            1

> summary(mod)

Call:
lm(formula = Y ~ X)

Residuals:
Min     1Q Median     3Q    Max
-1     -1      0      1      1

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.0000     0.5916   3.381 0.009633 **
X             1.0000     0.1667   6.000 0.000323 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

```
Residual standard error: 1 on 8 degrees of freedom
Multiple R-squared:  0.8182,    Adjusted R-squared:  0.7955
F-statistic:    36 on 1 and 8 DF,  p-value: 0.0003234

> anova(mod)
Analysis of Variance Table

Response: Y
Df Sum Sq Mean Sq F value    Pr(>F)
X          1     36      36      36 0.0003234 ***
Residuals  8      8       1
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
>
```