01. Os dados a seguir correspondem às variáveis renda familiar (X) e gasto com alimentação(Y) numa amostra de dez famílias representadas em unidades monetárias

X	3	5	10	20	30	50	70	100	150	200
Y	1,5	2	6	10	15	20	25	40	60	80

a. Faça um diagrama de dispersão.

b. Ajuste um modelo de regressão da forma:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

Com $E(\epsilon_i) = 0 \quad e \quad Var(\epsilon_i) = \sigma^2$.
Note que:

$$E(Y_i) = \beta_0 + \beta_1 \, X_i,$$

e a reta de regressão estimada

$$\hat{Y}_i = b_0 + b_1 \, X_i.$$

Vamos calcular e responder as perguntas:

c. Qual a previsão do gasto com alimentação para uma família com renda de 170 unidades monetárias?

d. Um acréscimo, em média, de 20 unidades monetárias na renda de uma família provoca que efeito no gasto médio com alimentação?

e. Qual a previsão do gasto com alimentação para uma família com renda excepcional de 1000 unidades monetárias? Você acha este valor razoável? Por que?

f. Se você respondeu que o valor obtido em (e) não é razoável, encontre uma explicação para o ocorrido.
Sugestão; Interprete a natureza das variáveis X e Y e os valores de Y para grandes valores de X.

Aproveite a saída do R:

```
> X=c(3,5,10,20,30,50,70,100,150,200)
> Y=c(1.5,2,6,10,15,20,25,40,60,80)
> n=length(X);n
[1] 10
> SX=sum(X);SY=sum(Y);SX2=sum(X^2);SY2=sum(Y^2);SXY=sum(X*Y)
> SX;SY;SX2;SY2;SXY
[1] 638
[1] 259.5
[1] 81334
[1] 12992.25
[1] 32474.5
> Xb=mean(X);Yb=mean(Y);Xb;Yb
[1] 63.8
[1] 25.95
>
```

```
> mod1=lm(Y~X);mod1

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)            X
0.9536       0.3918

> summary(mod1)

Call:
lm(formula = Y ~ X)

Residuals:
Min      1Q  Median      3Q     Max
-3.3791 -0.6075  0.0723  1.0183  2.2926

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.953595   0.733558    1.30     0.23
X           0.391793   0.008134   48.17 3.82e-11 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 1.64 on 8 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9961
F-statistic:  2320 on 1 and 8 DF,  p-value: 3.817e-11

> anova(mod1)
Analysis of Variance Table

Response: Y
Df Sum Sq Mean Sq F value    Pr(>F)
X          1 6236.7  6236.7  2320.2 3.817e-11 ***
Residuals  8   21.5     2.7
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
>
> b_0=0.9536;b_1=0.3918
>
>
> ##Xh=170
>
> Xh=170
> Y_prev=b_0+b_1*Xh;Y_prev
[1] 67.5596
>
>
>
>
> ##Xh=1000
```

```
>
> Xh=1000
> Y_prev=b_0+b_1*Xh;Y_prev
[1] 392.7536
>
>
```

Os dados a seguir referem-se a meses de experiência de dez digitadores e o número de erros cometidos na digitação de determinado texto.

Meses(X)	1	2	3	4	5	6	7	8	9	10
Erros(Y)	30	28	24	20	18	14	13	10	7	6

Dados:

$$\sum_{i=1}^{10} X_i = 60; \ ;\sum_{i=1}^{10} X_i^2 = 460 \ ;\sum_{i=1}^{10} Y_i = 170; \ ;\sum_{i=1}^{10} X_iY_i = 768.$$

a. Represente graficamente esse conjunto de dados.
b. Assumindo que um modelo de regressão linear é adequado, determine os coeficientes da equação pelo método dos mínimos quadrados.
   Explicite o modelo com as suposições.
c. Represente a reta de regressão no gráfico feiro anteriormente.
d. Qual a posição do ponto $(\bar{X}, \bar{Y})$ em relação à reta de regressão?
e. Qual o número esperado de erros para um digitador com 5 meses de experiência?
f. Faça tudo no **R**. Não esqueça d calcular as somas de quadrados.