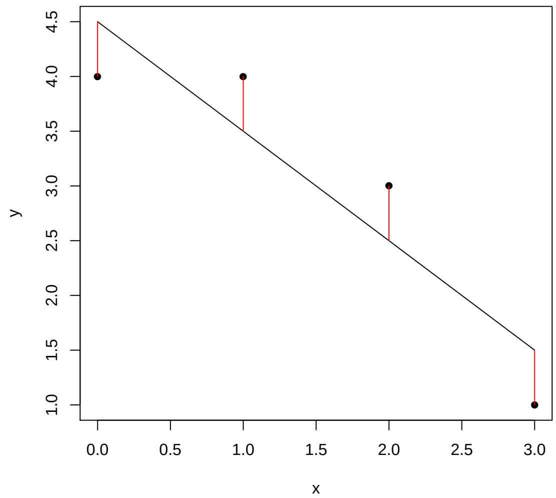
10.9 Soluções

10.9.1 Solução dos exercícios sobre Formas de dependência, etc.

10.9.2 Soluções dos exercícios sobre o método de mínimos quadrados

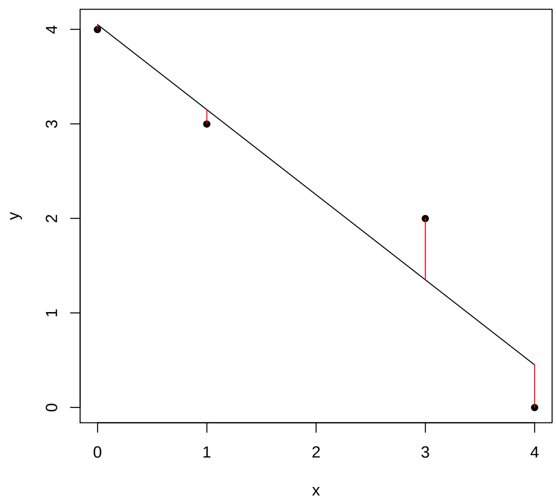
Solução do Exc. 192
(a)



Valores dos resíduos:

1	2	3	4
-0.5	0.5	0.5	-0.5

Soma dos quadrados dos resíduos igual a 1.
(b)

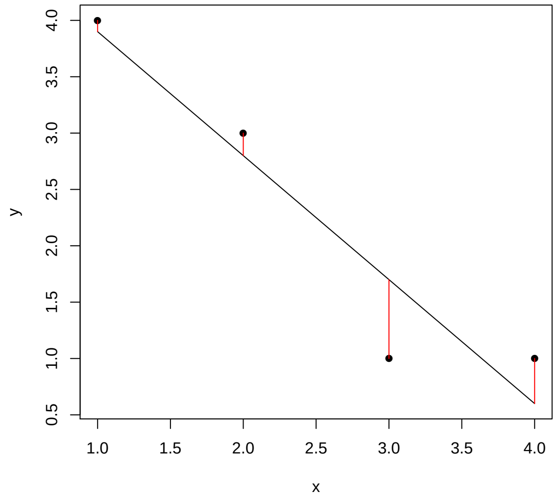


Valores dos resíduos:

1	2	3	4
-0.05	-0.15	0.65	-0.45

Soma dos quadrados dos resíduos igual a $0,65$.

(c)

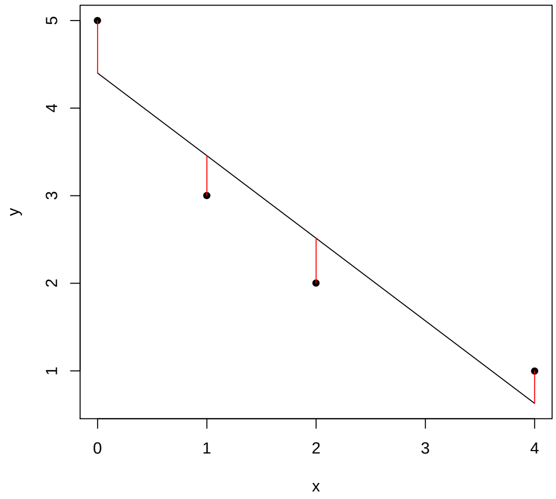


Valores dos resíduos:

1	2	3	4
0.1	0.2	-0.7	0.4

Soma dos quadrados dos resíduos igual a $0,7$.

(d)



Valores dos resíduos:

1	2	3	4
0.6	-0.4571	-0.5143	0.3714

Soma dos quadrados dos resíduos igual a $0,971$.

Solução do Exc. 193(a). Como ambos os valores de Y dos dois pontos dados são 0, então a reta de regressão feita só para estas dois pontos é horizontal. Precisamos acrescentar um ponto que incline esta reta da maneira que seu intercepto (b) seja negativo. Espero que o Exercício 192 serviu para criar a intuição que leve a solução. Uma das possíveis soluções é $(4, -1)$.

Também, podemos abordar este problema da seguinte maneira. Como nosso objetivo é garantir o sinal positivo para b e como o denumerador da expressão para b é sempre positivo, então podemos reformular o problema asim: achar (x_3, y_3) da maneira que

$$\sum_{i=1}^{3} x_i y_i - 3\overline{x}\overline{y}$$

seja negativo, onde $(x_1, y_1) = (0; 0)$ e $(x_2, y_2) = (3; 0)$. Agora observamos que devido ao fato que $y_1 = 0$ e $y_2 = 0$, temos que $x_1y_1 = 0$ e $x_2y_2 = 0$. Vamos então escolher $x_3 = -3$ para garantir que $\overline{x} = 0$. Com isto, a expressão acima reduz-se a simplesmente $(-3)y_3$. Isto garante que ao escolher para y_3 qualquer valor positivo garantimos que toda a expressão tenha valor negativo. Uma das possibilidades é $(x_3, y_3) = (-3; 2)$.

O método aplicado no parágrafo acima não é genérico. Ele funcionou bem no caso particular do expercício, pois houve só dois pontos e a coordenada y dos ambos era 0.

Solução do Exc. 194. Dica: Ao compor o exercício, achei que sua consepção intuitiva desenvolvida na solução do Ex. 192 deve ser suficiente para lhe sugerir a posição do ponto solicitado. Como então a solução espera-se partir de considerações qualitativas e intuitivas, pedi que essa seja confirmada por uma caminho rigoroso. Há também a seguinte abordagem ao problema: Denotar por (u, v) as coordenadas do terceiro ponto, escrever β como função de u e v, e achar valores numéricos destes – e, portanto, a posição do terceio ponto – que fazem a valer a desigualdade $\beta < 0$. Confesso que mandar aluno por este caminho não é minha intenção pois as contas são monstruosas.

Solução do Exc. 195 O sinal do valor de b da reta ajustada aos três pontos do enunciado é dado pelo numerador da expressão para b. O numerados é igual a

$$(0 \times 0 + 0,5 \times 0,7 + 1 \times 1) - 3 \times \frac{0 + 0,5 + 1}{3} \times \frac{0 + 0,7 + 1}{3} = 1,35 - 0,85 > 0$$

e isto garante que $b > 0$ (se continuarmos as contas, chegaremos a descobrir que $b = 1$).

Quanto à tarefa de acrescimo do quarto ponto, espero que o Exercício 192 serviu para criar a intuição sobre a posição da reta de mínimos quadrados em função dasposições dos pontos para os quais esta reta foi traçada. Esta intuição sugere que (x_4, y_4), o ponto a adicionar, seja $(2; -2)$. Vamos verificar se esta sugestão resolve a tarefa. O numerador da expressão para b é

$$(0 \times 0 + 0,5 \times 0,7 + 1 \times 1 + 2 \times (-2)) - 4 \times \frac{0 + 0,5 + 1 + 2}{4} \times \frac{0 + 0,7 + 1 + (-2)}{4}$$

Este é igual a $-2,3875$, que é negativo, o que implica $b < 0$.

Solução do Exc. 196 Observe que a coordenada y do ponto da reta $y = a+bx$ que corresponde ao valor $x = x_i$ é $a + bx_i$. Então, para qualquer ponto (x_i, y_i) do conjunto, seu resíduo é a distância entre y_i e $a + bx_i$. O quadrado do resíduo é então $[y_i - (a + bx_i)]^2$. Esta fórmula é a base para a solução do exercício.

Substituindo $(0; 0)$ no lugar de (x_i, y_i) da fórmula e usando a informação que o quadrado do resíduo até este ponto é 4, conlcuimos que

$$4 = a^2, \text{ quer dizer, } a \text{ é } 2 \text{ ou } -2$$

Substituindo $(4;1)$ e usando o fato que o quadrado do resíduo até este ponto é 9, concluimos que

$$9 = [1 - (a + 4b)]^2$$

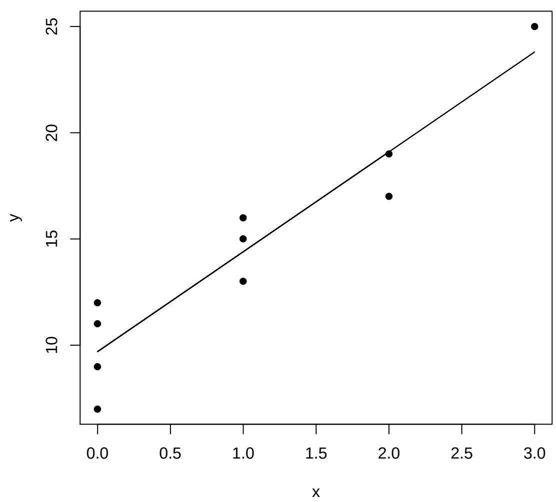
As duas equações enolvendo a e b têm 4 soluções:

$$a = 2, b = 1/2, \qquad a = 2, b = -1, \qquad a = -2, b = 0, \qquad a = -2, b = 3/2$$

Como o enunciado afirmou que β é negativo, então dentre os quatro soluções acima listadas, só $a = 2, b = -1$ atende às condições do exercício. Então, a resposta final é que a reta ajustada tem a seguinte forma: $y = 2 - x$.

10.9.3 Soluções dos exercícios sobre regressão

Solução do Exc. 200 (a)



O diagrama de dispersão é a figura acima porém sem a reta (esta reta ilustra o resultado da solução do item (c)). As posições dos pontos sugerem que X e Y possuem relação da seguinte natureza: o aumento do valor de X causa o aumento do valor de Y. Até podemos cogitar que a relação esteja próxima da relação linear. A “quantificação” da confiança neste tipo de relação é dada pelo valor do coeficiente de correlação linear de Pearson considerado no item seguinte.

(b) Utilizando a fórmula:

$$r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2) \cdot (\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2)}},$$

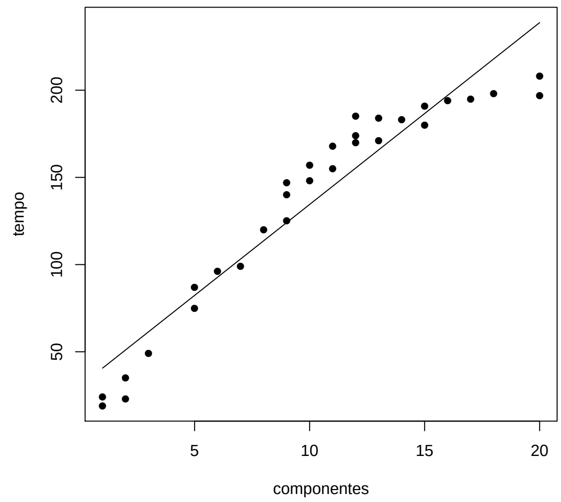
obtemos $r_{x,y} = 0,947$. Sua proximidade de 1 sugere que X e Y têm relação linear “estragada” por um pequeno ruido aleatório (isto é: $Y = a + bX + ruido$).

(c) Calculando

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot (\overline{x})^2}; \qquad a = \overline{y} - b\overline{x}.$$

temos $b = 4,7$ e $a = 9,7$. Esta reta está desenhada no diagrama de dispersão.

(d) Assumindo que o modelo de regressão linear é válido para a relação entre X e Y, o valor de b adquire a seguinte interpretação: para cada vez a mais que a caixa é transferida, há um aumento médio de 4,7 ampolas quebradas.



Solução do Exc. 201(e) O diagrama de dispersão é a figura acima sem a reta. As posições dos pontos sugerem que há uma relação entre as variáveis X e Y seguindo a qual Y aumenta com o aumento de X.

Os pontos "desenham" uma minhoca, mas esta não é muito entortada o que deixa a esperança que talvéz, haja a relação linear entre X e Y. A esperança terá seu respaldo no resultado de cálculo do próximo item.

(f) Utilizando a fórmula:

$$r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2) \cdot (\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2)}},$$

obtemos $r_{x,y} = 0,953$. A proximidade deste valor a 1 é uma forte indicação à existência de relação linear (com ruido pequeno) entre X e Y.

(g) Calculando

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot (\overline{x})^2}; \qquad a = \overline{y} - b\overline{x}.$$

temos $b = 10,429$ e $a = 30,194$. Uma das interpretações de b (que é a única ensinada nesta disciplina) é assim: para cada componente a mais reparado, temos um aumento médio de 10,429 minutos no tempo de reparo.

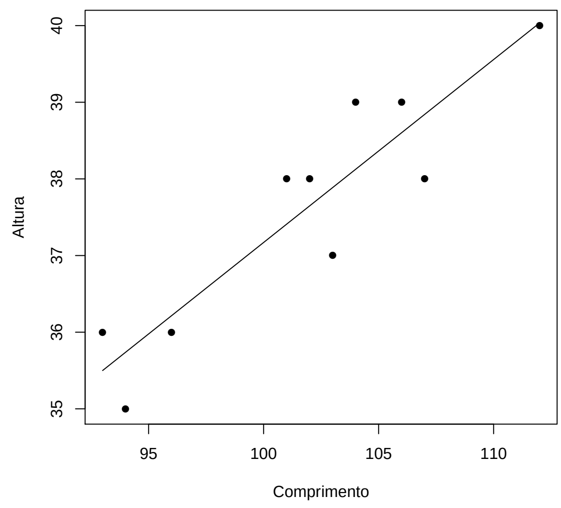
(h) Para $X = 8$, o valor de $a + bX$, com a e b calculados acima, é

$$30,194 + 8 \times 10,429 = 113,626.$$

Este valor chama-se *o tempo esperado (*ou *médio) para reparo de computador com 8 componentes a serem reparados ou substituídos*. Na teoria estatística, a notação tradicional para este valor é $\hat{Y}(8)$.

O valor esperado calculado acima não precisa ser necessariamente igual a um dos valores observados cuja coordenada X é 8. Esta desigualdade vem do fato que os valores observados são "corrumpidos" pelo erro: $Y = a + bX + erro$, enquanto que o valor esperado desconsidera o erro: $\hat{Y} = a + bX$. Esta eliminação justifica-se pelo fato que $\hat{Y}(x)$ é a média de todas as observações com $X = x$, e que a média do erro – seguindo a própria definição do modelo de regressão linear – é zero.

Solução do Exc. 203 A figura abaixo apresenta o diagrama de dispersão junto com a reta de regressão linear (que será calculada no item (c)).



O diagrama indica fortemente que a variável Y é uma função crescente da variável X. A grosso modo, esta relação é linera, conforme indica o valor do coeficiente de correlação linear de Pearson:

$$r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2) \cdot (\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2)}} = 0,913.$$

Os valores numéricos dos parâmetros a e b da reta ajustada aos pontos pelo método de mínimos quadrados são

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot (\overline{x})^2} = 0,24; \qquad a = \overline{y} - b\overline{x} = 13,3.$$

O fato que $b > 0$ confirma nossa observação feita acima de que Y é função crescente de X. Já que $b = 0,24$, podemos dizer que para o aumento de 1cm de Comprimento, há um aumento médio de 0,24cm na Altura das tartarugas.

Também com o uso da reta podemos fazer previsões médias: para os tartarugas de Comprimento 95cm, sua Altura média (ou, em outras palavras, esperada) é

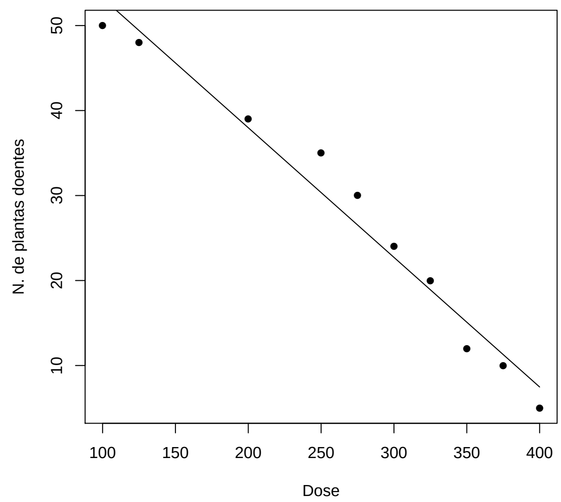
$$13,3 + 95 \times 0,24 = 36,1cm,$$

e para os tartarugas com Comprimento de 105cm, temos a seguinte Altura esperada:

$$13,3 + 105 \times 0,24 = 38,5cm.$$

Solução do Exc. 204 A figura abaixo apresenta o diagrama de dispersão junto com a reta de regressão linear (a equação da reta será calculada mais adiante As posições dos pontos do diagrama indicam que Y é uma função decrescente de X. Ainda mais, a proximidade das posições a uma reta imaginária sugere que Y é uma função linear de X com pequeno ruido, isto é: $Y = a + bX + ruido$. Isto está confirmado pela proximidade a -1 do valor do coeficiente de correlação linear de Pearson entre X e Y:

$$\begin{aligned} r_{x,y} &= \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2) \cdot (\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2)}} \\ &= \frac{59450 - 2700 \cdot 273 \cdot 1/10}{\sqrt{(822500 - 2700^2/10) \cdot (9695 - 273^2/10)}} = -0,985 \end{aligned}$$



A reta de ajustas aos pontos pelo método de minimos quadrados tem os seguintes parâmetros:

$$b = \frac{\sum_{i=1}^n x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^n x_i^2 - n \cdot (\overline{x})^2} = \frac{59450 - 2700 \cdot 273 \cdot 1/10}{822500 - 2700^2/10} = -0,1525;$$

$$a = \overline{y} - b\overline{x} = 273/10 - (-0,1525) \cdot 2700/10 = 68,478.$$

Vale notar que o valor de b é negativo, o que está de acordo com o fato já observado que Y é decrescente em função de X. O valor numérico de B significa o seguinte: para cada aumento de uma dose, há uma queda média de 0,1525 no número de plantas doentes.

Usando a reta ajustada, temos que o número esperado de plantas doentes para cada 100 plantas tratadas na dose de 260 gr/ha dado por:

$$68,478 + (-0,1525) \times 260 = 28,828.$$

Solução do Exc. 205. (a) Diagrama de dispersão para as variáveis X e Y com a reta de regressão:
(b) Calculemos as médias e desvios padrão para cada uma das variáveis:

$$\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} \approx 5,64 \text{ meses}$$

$$\sigma_x^2 = \frac{\sum_{i=1}^{11} x_i^2 - 11(\bar{x})^2}{10} \to \sigma_x^2 \approx 11,41 \text{ (meses)}^2$$

$$\sigma_x = \sqrt{\sigma_x^2} \approx 3,38 \text{ meses}$$

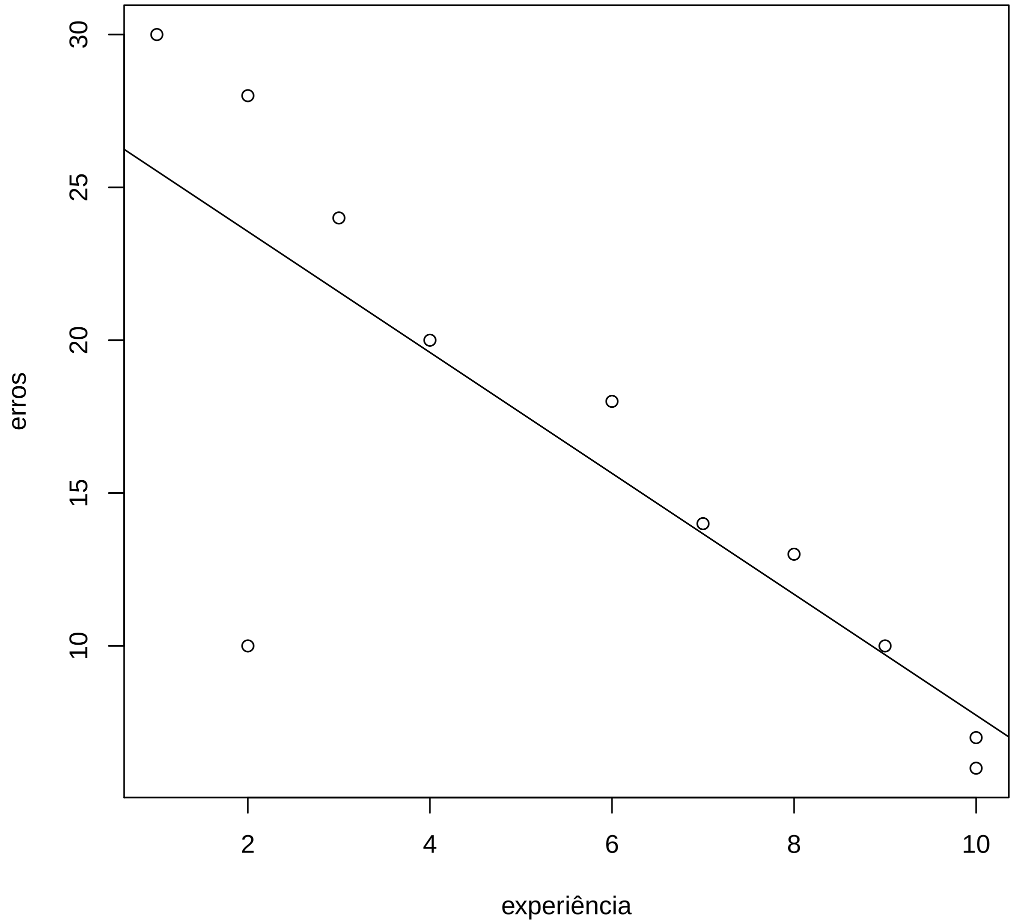
$$\bar{y} = \frac{\sum_{i=1}^{11} y_i}{11} \approx 16,36 \text{ erros}$$

$$\sigma_y^2 = \frac{\sum_{i=1}^{11} y_i^2 - 11(\bar{y})^2}{10} \to \sigma_y^2 \approx 68,99 \text{ erros}^2$$

$$\sigma_y = \sqrt{\sigma_x^2} \approx 8,31 \text{ erros}$$

Calculemos, então, o coeficiente de correlação linear de Pearson entre X e Y:

$$r = \sum_{i=1}^{11} \frac{(x_i - \bar{x})(y_i - \bar{y})}{10.\sigma_x.\sigma_y} = \frac{\sum_{i=1}^{11} x_i.y_i - 11.\bar{x}.\bar{y}}{10.\sigma_x.\sigma_y} \approx -0,81$$



O coeficiente de correlação entre as variáveis X e Y é $r \approx -0,81$ e, assim, pode-se dizer que X e Y têm correlação linear negativa forte, uma vez que se aproxima de $-1,0$.
(c) Para construir a reta de regressão, obtém-se, primeiramente,a equação de regressão de Y em função de X dada pela forma: $\hat{y} = a + b.x$. Obtém-se os valores de a e b:

$$\begin{aligned} b &= \frac{\sum_{i=1}^{11} x_i.y_i - 11.\bar{x}.\bar{y}}{10.\sigma_x^2} \approx -1,99 \\ a &= \bar{y} - b.\bar{x} \approx 27,51 \end{aligned}$$

Assim, tem-se: $\hat{y} = 27,51 - 1,99.x$.
(d) O coeficiente $b \approx -1,99$ indica que a cada aumento de uma unidade em x, ocorre em média um aumento de $-1,99$ unidades em y.

(e) O datilógrafo com comportamento diferente dos demais é aquele em que temos: $(x, y) = (2, 10)$. Assim, o novo coeficiente de correlação e a nova reta ajustada sem essa observação serão:

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^{10} x_i}{10} \approx 6,0 \text{ meses} \\ \sigma_x^2 &= \frac{\sum_{i=1}^{10} x_i^2 - 10(\bar{x})^2}{9} \rightarrow \sigma_x^2 \approx 11,11 \text{ (meses)}^2 \\ \sigma_x &= \sqrt{\sigma_x^2} \approx 3,33 \text{ meses} \\ \bar{y} &= \frac{\sum_{i=1}^{10} y_i}{10} \approx 17,0 \text{ erros} \\ \sigma_y^2 &= \frac{\sum_{i=1}^{10} y_i^2 - 10(\bar{y})^2}{9} \rightarrow \sigma_y^2 \approx 71,55 \text{ (erros)}^2 \end{aligned}$$

$$\sigma_y \;=\; \sqrt{\sigma_x^2} \approx 8,46 \text{ erros}$$

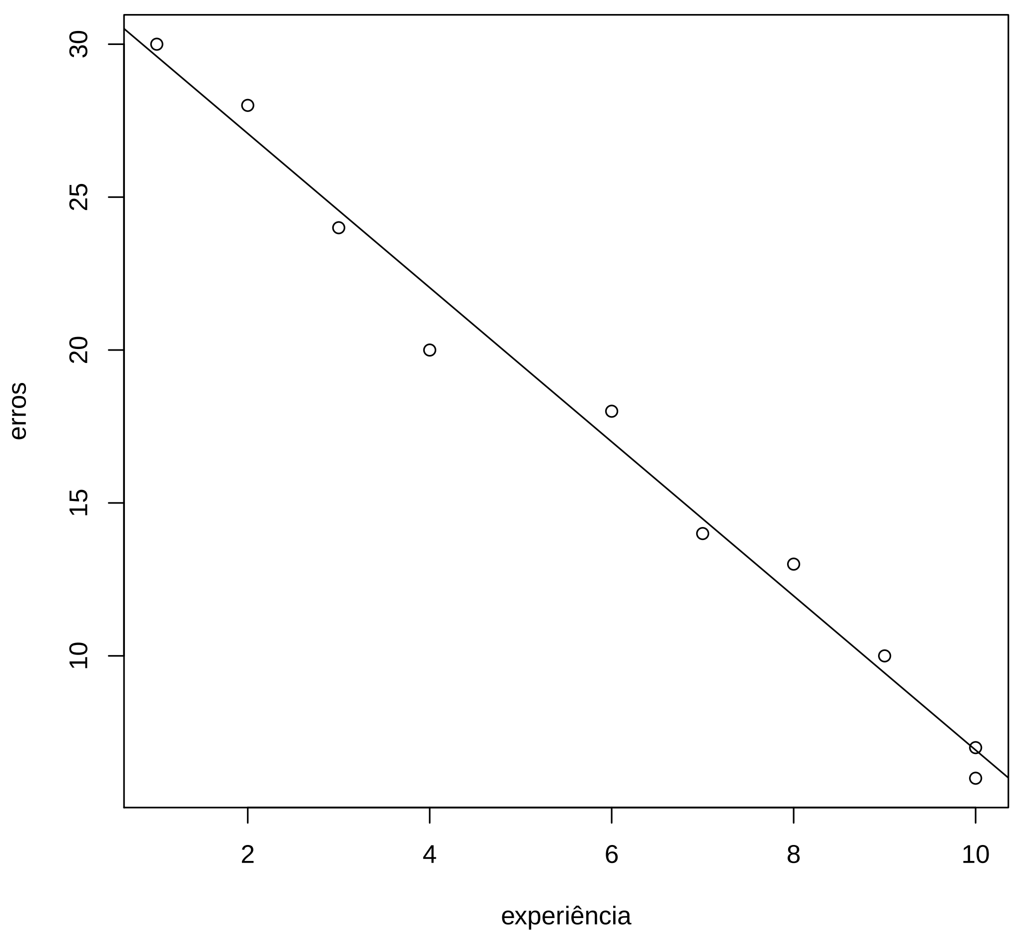
Calculemos, então, o novo coeficiente de correlação linear de Pearson entre X e Y:

$$r = \sum_{i=1}^{10} \frac{(x_i - \bar{x})(y_i - \bar{y})}{9.\sigma_x.\sigma_y} = \frac{\sum_{i=1}^{10} x_i.y_i - 10.\bar{x}.\bar{y}}{9.\sigma_x.\sigma_y} \approx -0,99$$

Para a nova reta de regressão, calculemos a nova equação de regressão $\hat{y} = a + b.x$:

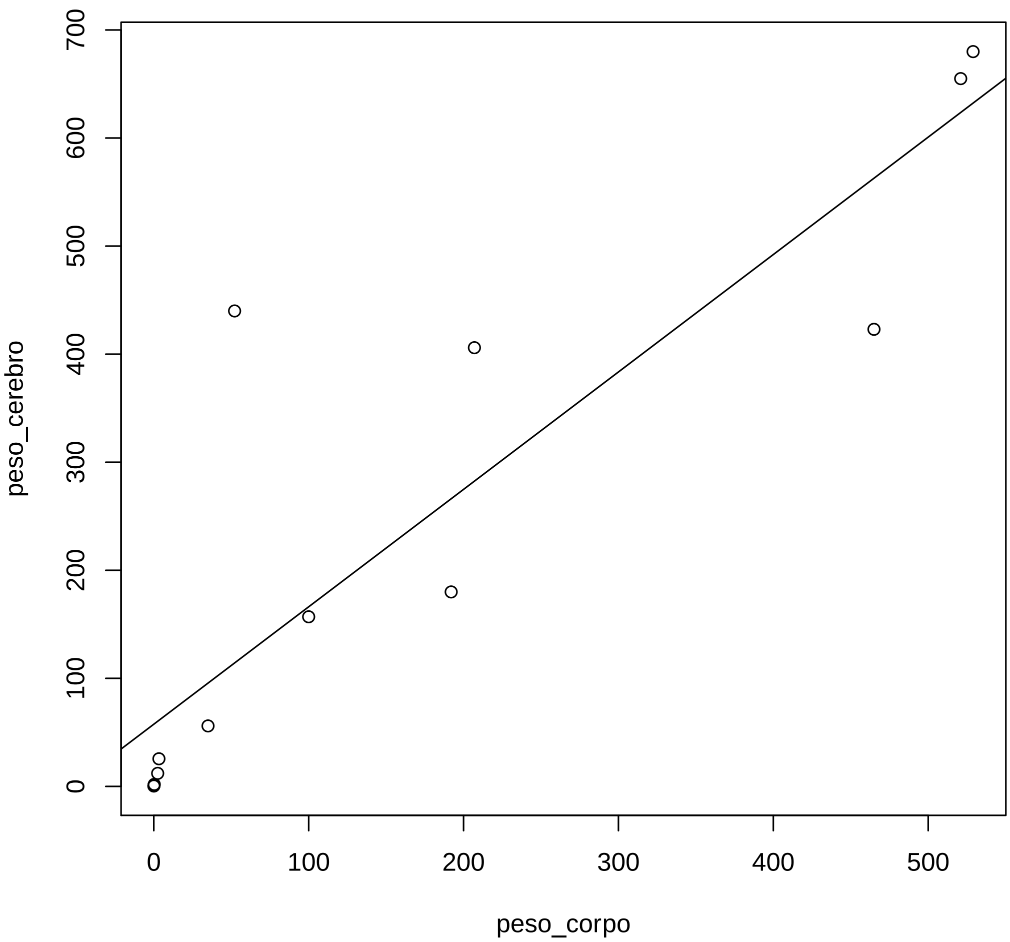
$$\begin{aligned} b &= \frac{\sum_{i=1}^{10} x_i.y_i - 10.\bar{x}.\bar{y}}{9.\sigma_x^2} \approx -2,52 \\ a &= \bar{y} - b.\bar{x} \approx 32,12 \end{aligned}$$

Assim: $\hat{y} = 32,12 - 2,52.x$ que está na figura abaixo.



F

A observação eliminada tinha um comportamento diferente dos demais, isto é, tratava-se de um datilógrafo com apenas 2 meses de experiência tendo uma quantidade de erros muito próxima à quantidade de erros de um profissional mais experiente (10 erros). Assim, ao refazermos os cálculos eliminando essa observação, o módulo tanto do coeficiente de correlação linear de Pearson quanto da inclinação da reta aumentam, pois as variáveis meses de experiência(X) e erros(Y) ficam com uma relação linear mais forte.



Solução do Exc. 206. **(a)** O diagrama de dispersão para as variáveis X e Y está na figura abaixo. Calculemos as médias e desvios padrão para cada uma das variáveis:

$$\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{13} x_i}{13} \approx 162,11 \text{ kg} \\
\sigma_x^2 &= \frac{\sum_{i=1}^{13} x_i^2 - 13(\bar{x})^2}{12} \to \sigma_x^2 \approx 43295,62 \text{ (kg)}^2 \\
\sigma_x &= \sqrt{\sigma_x^2} \approx 208,08 \text{ kg} \\
\bar{y} &= \frac{\sum_{i=1}^{13} y_i}{13} \approx 233,69 \text{ g} \\
\sigma_y^2 &= \frac{\sum_{i=1}^{13} y_i^2 - 13(\bar{y})^2}{12} \to \sigma_y^2 \approx 64986,53 \text{ g} \\
\sigma_y &= \sqrt{\sigma_x^2} \approx 254,92 \text{ g}
\end{aligned}$$

Calculemos, então, o coeficiente de correlação linear de Pearson entre X e Y:

$$r = \sum_{i=1}^{13} \frac{(x_i - \bar{x})(y_i - \bar{y})}{12.\sigma_x.\sigma_y} = \frac{\sum_{i=1}^{13} x_i.y_i - 13.\bar{x}.\bar{y}}{12.\sigma_x.\sigma_y} \approx 0,89$$

O coeficiente de correlação entre as variáveis peso médio do corpo e peso médio do cérebro para as 13 espécies de animais é $r \approx 0,89$. Então, tem-se uma correlação linear positiva forte, aproximando-se de $1,0$.

(b) A equação de regressão do peso médio do cérebro (Y) em função do peso médio do corpo (X) é $\hat{y} = a + b.x$ onde

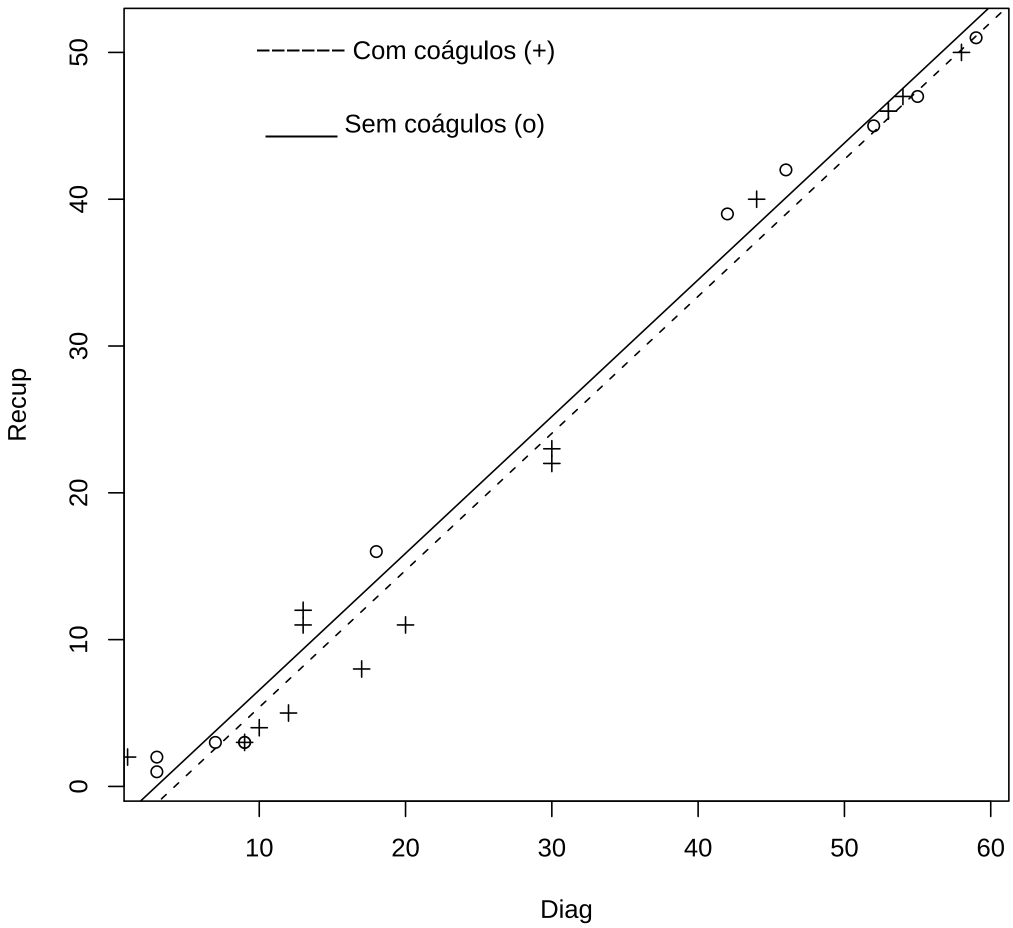
$$\begin{aligned} b &= \frac{\sum_{i=1}^{13} x_i.y_i - 13.\bar{x}.\bar{y}}{12.\sigma_x^2} \approx 1,09 \\ a &= \bar{y} - b.\bar{x} \approx 57,56 \end{aligned}$$

Assim, tem-se: $\hat{y} = 57,56 + 1,09.x$.

(c) Uma previsão do peso médio do cérebro para animais com peso médio corporal igual a 62 kg será dada pela equação de regressão: $\hat{y} = 57,56 + 1,09.x \rightarrow \hat{y} = 57,56 + 1,09.62 \rightarrow \hat{y} = 125,0\ g$ Assim, a previsão é um peso médio de $125,0\ g$ para o cérebro de animais com peso médio corporal de 62 kg.

Pelo item (b), pode-se concluir que a relação entre o peso do cérebro e o peso do corpo para o homem não segue a relação linear existente entre os pesos cerebral e corpóreo para os animais.

Solução do Exc. 207 (a) O diagrama de dispersão para as variáveis Diag(X) e Recup(Y) em pacientes com coágulos presentes e ausentes está na figura abaixo:



Para os pacientes com coágulos, (os de números 3, 4, 8, 9, 11, 15, 16, 17, 18, 19, 20, 21, 23, 25, 26), temos:

Calculemos as médias e desvios padrão para cada uma das variáveis Diag(X) e Recup(Y):

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = 27,8 \text{ horas}$$

$$
\begin{aligned}
\sigma_x^2 &= \frac{\sum_{i=1}^{15} x_i^2 - 15(\bar{x})^2}{14} \rightarrow \sigma_x^2 \approx 385,31 \text{ (horas)}^2 \\
\sigma_x &= \sqrt{\sigma_x^2} \approx 19,63 \text{ horas} \\
\bar{y} &= \frac{\sum_{i=1}^{15} y_i}{15} = 22 \text{ horas} \\
\sigma_y^2 &= \frac{\sum_{i=1}^{15} y_i^2 - 15(\bar{y})^2}{14} \rightarrow \sigma_y^2 \approx 342,71 \text{ (horas)}^2 \\
\sigma_y &= \sqrt{\sigma_x^2} \approx 18,51 \text{ horas}
\end{aligned}
$$

Calculemos, então, o coeficiente de correlação linear de Pearson entre X e Y: $r = \sum_{i=1}^{15} \frac{(x_i-\bar{x})(y_i-\bar{y})}{14.\sigma_x.\sigma_y} = \frac{\sum_{i=1}^{15} x_i.y_i - 15.\bar{x}.\bar{y}}{14.\sigma_x.\sigma_y} \approx 1,0$

O coeficiente de correlação entre as variáveis Diag(X) e Recup(Y), medidas em horas, para os pacientes que apresentaram coágulos no momento de admissão é $r \approx 1,0$ e, assim, pode-se dizer que X e Y têm correlação linear positiva e perfeita para tal caso.

Para construir a reta de regressão, obtém-se, primeiramente, os valores de a e b:

$$
\begin{aligned}
b &= \frac{\sum_{i=1}^{15} x_i.y_i - 15.\bar{x}.\bar{y}}{14.\sigma_x^2} \approx 0,93 \\
a &= \bar{y} - b.\bar{x} \approx -3,94
\end{aligned}
$$

Assim, tem-se a reta: $\hat{y} = -3,94 + 0,93.x$

Para os pacientes sem coágulos (os com os números 1, 2, 5, 6, 7, 10, 12, 13, 14, 22, 24), temos as as médias e desvios padrão calculados para cada uma das variáveis Diag(X) e Recup(Y):

$$
\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{11} x_i}{11} = 27,55 \text{ horas} \\
\sigma_x^2 &= \frac{\sum_{i=1}^{11} x_i^2 - 11(\bar{x})^2}{10} \rightarrow \sigma_x^2 \approx 529,4 \text{ (horas)}^2 \\
\sigma_x &= \sqrt{\sigma_x^2} \approx 23,0 \text{ horas} \\
\bar{y} &= \frac{\sum_{i=1}^{11} y_i}{11} = 22,91 \text{ horas} \\
\sigma_y^2 &= \frac{\sum_{i=1}^{11} y_i^2 - 11(\bar{y})^2}{10} \rightarrow \sigma_y^2 \approx 463,45 \text{ horas}^2 \\
\sigma_y &= \sqrt{\sigma_x^2} \approx 21,53 \text{ horas}
\end{aligned}
$$

Calculemos, então, o coeficiente de correlação linear de Pearson entre X e Y: $r = \sum_{i=1}^{11} \frac{(x_i-\bar{x})(y_i-\bar{y})}{10.\sigma_x.\sigma_y} = \frac{\sum_{i=1}^{11} x_i.y_i - 11.\bar{x}.\bar{y}}{10.\sigma_x.\sigma_y} \approx 1,0$ O coeficiente de correlação entre as variáveis Diag(X) e Recup(Y), medidas em horas, para os pacientes que apresentaram coágulos no momento de admissão é $r \approx 1,0$ e, assim, pode-se dizer que X e Y têm correlação linear positiva e prefeita para tal caso.

Para construir a reta de regressão, obtém-se, primeiramente, os valores de a e b:

$$
\begin{aligned}
b &= \frac{\sum_{i=1}^{11} x_i.y_i - 11.\bar{x}.\bar{y}}{10.\sigma_x^2} \approx 0,93 \\
a &= \bar{y} - b.\bar{x} \approx -2,75
\end{aligned}
$$

Assim, tem-se a equação: $\hat{y} = -2,75 + 0,93.x$

Pode-se concluir, então, que o relacionamento entre as variáveis Diag(X) e Recup(Y) não depende da ocorrência de coágulos, uma vez que em ambos o coeficiente de correlação linear de Pearson vale aproximadamente $1,0$ e a inclinação da reta de regressão é $b \approx 0,93$.