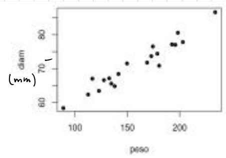
1.45. (*Teste 7.11.2018*) As maçãs Fugi são pagas aos agricultores em função do seu calibre (diâmetro, em mm). Nas centrais fruteiras, os calibradores electrónicos recorrem ao peso dos frutos para estimar os seus diâmetros. Para estabelecer a relação entre pesos (em gramas) e diâmetros (em mm), pesaram-se e mediram-se 20 maçãs obtidas no mercado. Introduzidos os dados no R e obtiveram-se os seguintes resultados:

```
> cor(peso,diam)^2
[1] 0.9298733

Coefficients:
(Intercept)        peso
  42.4601        0.1815
```



a) Escreva a equação da recta de regressão obtida no R. Qual a precisão dessa recta? Comente.

b) Parece-lhe adequada a utilização do modelo linear para descrever a relação entre o diâmetro e o peso dos frutos analisados? Justifique.

c) Sabendo que a média e o desvio padrão do peso dos frutos são 158.25 g e 37.09 g, respectivamente, determine a média e o desvio padrão do diâmetro dos frutos observados.

d) Qual a variação esperada no diâmetro dos frutos associada ao aumento de 1 grama no peso, estimada pela regressão?

e) Determine a recta de regressão dos mínimos quadrados do diâmetro sobre o peso, sendo agora o peso dado em kg. Qual a precisão desta recta?

FORMULÁRIO →

Regressão linear simples

$\widehat{y}_i = b_0 + b_1 x_i \qquad \widehat{y}_i - \overline{y} = b_1(x_i - \overline{x})$

$y_i = b_0 + b_1 x_i + e_i$

$$\begin{cases} b_1 = \dfrac{cov(x,y)}{s_x^2} = r\dfrac{s_y}{s_x}, & s_x \neq 0 \\ b_0 = \overline{y} - b_1\overline{x} \end{cases}$$

$$\sum_{i=1}^n (y_i-\overline{y})^2 = \sum_{i=1}^n (y_i-\hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i-\overline{y})^2$$
$$\Leftrightarrow SQ_T = SQ_E + SQ_R$$

$$R^2 = \frac{SQ_R}{SQ_T} = \frac{cov^2(x,y)}{s_x^2\, s_y^2} = r^2$$

Indicadores para dados bivariados

$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$

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

$r = r_{x,y} = \dfrac{cov(x,y)}{s_x s_y} \qquad$ se $s_x \neq 0, s_y \neq 0$

se $x_i' = a + bx_i$ e $y_i' = c + dy_i \qquad b,d \neq 0$

$cov(x',y') = bd\, cov(x,y)$

$$r_{x',y'} = \begin{cases} r_{x,y} & \text{se } bd > 0 \\ -r_{x,y} & \text{se } bd < 0 \end{cases}$$

a) Recta de regressão

$y = 42.4601 + 0.1815x$

$b_1 = 0.1815\ mm/g$ É O COEF. DE REGRESSÃO (DECLIVE DA RECTA)

$b_0 = 42.4601\ mm$ É O VALOR DA RECTA QUANDO $x=0$

PRECISÃO DA RECTA É DADA PELO COEF. DE DETERMINAÇÃO (R^2) QUE PODE SER OBTIDO DE 2 FORMAS:

$R^2 = \dfrac{SQR}{SQT}$ É A PROPORÇÃO DA VARIABILIDADE DA VARIÁVEL RESPOSTA (y) QUE É EXPLICADA PELA RECTA DE REGRESSÃO

$R^2 = (r_{xy})^2 = 0.9298 = \underline{92.28\%}$

ISTO É, 92.98% DA VARIABILIDADE DOS DIÂMETROS DAS MAÇÃS É EXPLICADA PELA RECTA DE REGRESSÃO

b) SIM, É ADEQUADA POIS:

1. OS PONTOS DISPÕEM-SE PRÓXIMO DE UMA RECTA (NÃO HÁ CURVATURA)
2. O R^2 É ELEVADO, O QUE INDICA UMA FORTE CORRELAÇÃO LINEAR.

c) $\overline{x} = 158.25\ g$

$s_x = 37.09\ g$

$\overline{y} = ?$

COMO $\overline{y} = b_0 + b_1\overline{x}$

ENTÃO

$\overline{y} = 42.4601 + 0.1815 \times 158.25$

$= \underline{71.18}\ mm$

NOTA: $\underset{mm}{\overline{y}} = \underset{mm}{b_0} + \underset{mm/g}{b_1} \times \underset{g}{\overline{x}}$

$s_y = ?$

COMO $R^2 = 0.9298$, ENTÃO $r = +\sqrt{0.9298} = 0.9642$

↑ POIS O DECLIVE DA RECTA É POSITIVO

FORMULÁRIO: $b_1 = r\dfrac{s_y}{s_x}$

OU SEJA $s_y = \dfrac{b_1 s_x}{r} = \dfrac{0.1815 \times 37.09}{0.9642}$

$s_y = \underline{6.98}\ mm$

d) $b_1 = 0.1815\ mm/g$ É O DECLIVE DA RECTA E É INTERPRETADO COMO O AUMENTO MÉDIO DO DIÂMETRO (mm) QUANDO O PESO AUMENTA UMA UNIDADE (1g)

e) EM PRIMEIRO, DETERMINA-SE A RELAÇÃO ENTRE A VARIÁVEL ORIGINAL (x) E A NOVA VARIÁVEL (x')

x' É O PESO EM kg ; x É O PESO EM GRAMAS

ENTÃO, $\underbrace{x\ g}_{\text{obs. original}} = \boxed{x \times 0.001}\ kg = \boxed{x'}\ kg$. PORTANTO, $\boxed{x' = 0.001\,x}$

$1\ g = 0.001\ kg$

SABENDO QUE $x' = 0.001x$, DETERMINA-SE A NOVA RECTA DE REGRESSÃO:

$\underset{mm}{y} = \underset{mm}{b_0'} + \underset{mm/kg}{b_1'}\ \underset{kg}{x'}$ NOVA RECTA

$$b_1' = \frac{cov(x',y)}{s_{x'}^2} = \frac{0.001 \times cov(x,y)}{0.001^2\ s_x^2} = \frac{1}{0.001}\underbrace{\left(\frac{cov(x,y)}{s_x^2}\right)}_{b_1} = 1000\ b_1 = 181.5\ mm/kg$$

$x' = 0 + 0.001x$

(a) (b)

$$b_0' = \overline{y} - b_1'\overline{x}' = \overline{y} - 1000\, b_1 \times 0.001\,\overline{x} = \overline{y} - b_1\overline{x} = b_0 = 42.4601\ mm$$

Indicadores para dados bivariados

$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$

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

$r = r_{x,y} = \dfrac{cov(x,y)}{s_x s_y} \qquad$ se $s_x \neq 0, s_y \neq 0$

se $x_i' = a + bx_i$ e $y_i' = c + dy_i \qquad b,d \neq 0$

$cov(x',y') = bd\, cov(x,y)$

$$r_{x',y'} = \begin{cases} r_{x,y} & \text{se } bd > 0 \\ -r_{x,y} & \text{se } bd < 0 \end{cases}$$

$$cov(x', y') = bd\, cov(x, y)$$

$$r_{x',y'} = \begin{cases} r_{x,y} & \text{se } bd > 0 \\ -r_{x,y} & \text{se } bd < 0 \end{cases}$$

$$b_0' = \bar{y} - b_1' \overline{x'} \underset{(a)}{=} \bar{y} - \cancel{1000}\, b_1 \times 0.\cancel{00}1\, \bar{x} \underset{(b)}{=} \bar{y} - b_1 \bar{x} = b_0 = 42.4601 \text{ mm}$$

C.A. $\overline{x'} = \dfrac{\sum_{i=1}^{n} x_i'}{n} = \dfrac{\sum_{i=1}^{n} 0.001\, x}{n} = 0.001 \dfrac{\sum x}{n} = 0.001\, \bar{x}$

PROP.: $\overline{a + bx} = a + b\bar{x}$

FINALMENTE, CALCULA-SE O NOVO COEF. DE DETERMINAÇÃO:

COEF. DE DETERMINAÇÃO $R^{2'} = (r_{x'y})^2 = (r_{xy})^2 = R^2 = 0.9298$

COMO SE ESPERAVA, É IGUAL AO COEF. DE DETERMINAÇÃO ORIGINAL.

$$\boxed{\begin{aligned} &\text{C.A. } r_{x'y} = \frac{cov(x', y)}{s_{x'}\, s_y} = \frac{\cancel{0.001}\, cov(x,y)}{\cancel{0.001}\, s_x\, s_y} = \frac{cov(x,y)}{s_x\, s_y} = r_{xy} \\ &\text{PROP: } S_{a+bx} = |b|\, S_x \quad \text{(DESVIO-PADRÃO)} \end{aligned}}$$