Exercício 1

Na tabela abaixo são apresentadas a receita bruta e a receita com exportações (em milhões de reais) das empresas Sadia e Perdigão de 1994 a 2005. Resolva os itens a seguir.

Tabela 1: Receita Bruta e Receita com exportações (em milhões de reais) das empresas Sadia e Perdigão de 1994 a 2005.

Ano	Sadia		Perdigão	
	R Bruta	Exportações	R. Bruta	Exportações
1994	0.70	0.09	2.36	0.48
1995	0.94	0.20	2.82	0.39
1996	1.02	0.24	3.05	0.59
1997	1.27	0.33	3.22	0.72
1998	1.41	0.31	2.65	0.47
1999	1.80	0.52	3.15	0.84
2000	2.07	0.51	3.26	0.87
2001	2.79	1.03	4.02	1.52
2002	3.34	1.21	4.69	1.96
2003	4.37	1.84	5.86	2.66
2004	5.57	2.73	7.32	3.58
2005	5.87	2.84	8.33	4.08

(a) Construa o diagrama de dispersão entre receita bruta (y) e receita com exportações (x) para as duas empresas. Comente.

(b) Calcule o coeficiente de correlação linear de Pearson entre as duas variáveis para as duas empresas. Comente.

(c) Ajuste as retas de regressão linear entre as duas variáveis para as duas empresas. Interprete o coeficiente angular b para cada empresa.

(d) Se as exportações para Sadia e Perdigão num ano seguinte estão estimadas em 2,7 e 3,8 milhões de reais, respectivamente, estime pelas retas de regressão as respectivas receitas brutas.

Solução

a) Diagrama de dispersão entre receita bruta e receita com exportações das empresas Sadia e perdigão.

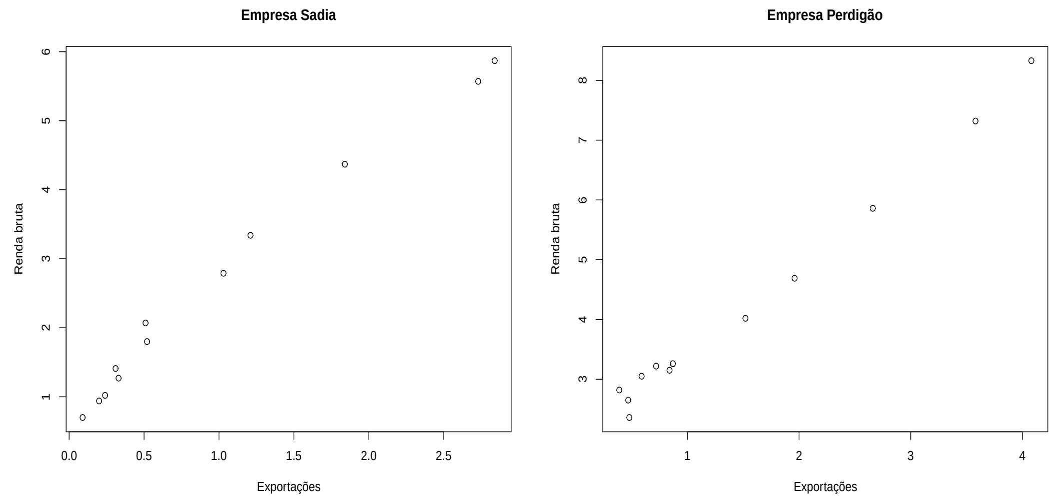

Figura 1: Diagrama de dispersão entre receita bruta (y) e receita com exportações (x) para as empresas Sadia e Perdigão.

- Observamos que tanto na Sadia quanto na perdigão a medida que a receita com exportações aumenta a receita bruta tende a aumentar. Nota-se também fortes indícios de uma tendência linear em ambos os casos.

b) Denotemos por,

- X : a renda de exportação da empresa Sadia,
- Y : a renda bruta da empresa Sadia.
  De modo que,
  * $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{0{,}09+0{,}20+\cdots+2{,}73+2{,}84}{12} = \frac{11.85}{12} = 0.9875$
  * $\sum_{i=1}^{12} X_i^2 = 22.2703$
  * $S_X^2 = \frac{\sum_{i=1}^{12} X_i^2 - n\overline{X}^2}{n} = \frac{22.2703-12\times 0.9875^2}{12} = 0.8807$
  * $S_X = \sqrt{S_X^2} = 0.9385$
  * $\overline{Y} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{0{,}70+0{,}94+\cdots+5{,}57+5{,}87}{12} = \frac{31.15}{12} = 2.5958$
  * $\sum_{i=1}^{12} Y_i^2 = 117.0583$
  * $S_Y^2 = \frac{\sum_{i=1}^{12} Y_i^2 - n\overline{Y}^2}{n} = \frac{117.0583-12\times 2.5958^2}{12} = 3.0166$
  * $S_Y = \sqrt{S_Y^2} = 1.7368$

* $\sum_{i=1}^{12}(X_iY_i) = 50.1765$

Portanto o coeficiente de correlação linear entre a receita bruta e a receita com exportações da empresa Sadia é dada por,

$$r = \frac{\sum_{i=1}^{12}(X_iY_i) - n\overline{XY}}{\text{n } S_x S_y} = \frac{50.1765 - 12 \times 2.5634}{\text{n} S_x S_y} = \frac{19.4157}{12 \times 1.7781} = 0.9099$$

Denotemos por,

- W : a renda de exportação da empresa Perdigão,
- Z : a renda bruta da empresa Perdigão.
  De modo que,
  * $\overline{W} = \frac{\sum_{i=1}^{n} W_i}{n} = \frac{0{,}48+0{,}39+\cdots 3{,}59+4.08}{12} = \frac{18.16}{12} = 1.5133$
  * $\sum_{i=1}^{12} W_i^2 = 45.6228$
  * $S_W^2 = \frac{\sum_{i=1}^{12} W_i^2 - n\overline{W}^2}{n} = \frac{45.6228 - 12 \times 1.5133^2}{12} = 1.5118$
  * $S_W = \sqrt{S_W^2} = 1.2295$
  * $\overline{Z} = \frac{\sum_{i=1}^{n} Z_i}{n} = \frac{50.73}{12} = 4.2275$
  * $\sum_{i=1}^{12} Z_i^2 = 256.2329$
  * $S_Z^2 = \frac{\sum_{i=1}^{12} Z_i^2 - n\overline{Z}^2}{n} = \frac{256.2329 - 12 \times 4.2275^2}{12} = 3.4809$
  * $S_Z = \sqrt{S_Z^2} = 1.8657$
  * $\sum_{i=1}^{12}(W_iZ_i) = 104.1606$

Portanto o coeficiente de correlação linear entre a receita bruta e a receita com exportações da empresa perdigão é dada por,

$$r = \frac{\sum_{i=1}^{12}(W_iZ_i) - n\overline{WZ}}{\text{n}S_W S_Z} = \frac{104.1606 - 12 \times 6.3975}{\text{n } S_Z S_W} = \frac{27.3906}{12 \times 2.5025} = 0.9121$$

Note que ambos os coeficientes de correlação linear estão realmente próximos de 1 o que indica uma relação linear positiva praticamente perfeita. Estes coeficientes colaboram com as conclusões obtidas a partir dos diagramas de dispersão.

c) – A reta de regressão linear entre as duas variáveis para a empresa sadia é, $\hat{Y} = a + bX$ onde,
  * b é dado por,

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} X_iY_i - n\overline{XY}}{\text{nS}} \\ &= \frac{50.1765 - 12 \times 2.5634}{\text{n S}} = \frac{19.4157}{12 \times 0.8807} \\ &= 1.8371 \end{aligned}$$

* a é dado por,

$$a = \overline{Y} - b\overline{X} = 2.5958 - 1.8371 \times 0.9875 = 0.7817$$

Portanto,

$$Y = a + bX = 0.7817 + 1.8371X.$$

Interpretação de b: Para cada aumento de uma unidade (em milhões de reais) da receita com exportações temos um aumento médio de 1.8371 unidades (em milhões de reais) da receita bruta da Sadia.

– A reta de regressão linear entre as duas variáveis para a empresa Perdigão é, $\hat{Z} = c + dW$ onde,

* d é dado por,

$$\begin{aligned} d &= \frac{\sum_{i=1}^{n} W_i Y_i - n\overline{ZW}}{\text{n } S_W} \\ &= \frac{104.1606 - 12 \times 6.3975}{\text{n x } 1.2295} = \frac{27.3906}{12 \times 1.2295} = 1.8565 \end{aligned}$$

* e c é dado por,

$$c = \overline{Z} - d\overline{W} = 4.2275 - 1.8565 \times 1.5133 = 1.4180$$

Portanto,

$$Z = 1.4180 + 1.8565W$$

Interpretação de d: Para cada aumento de uma unidade (em milhões de reais) da receita com exportações temos um aumento médio de 1.5098 unidades (em milhões de reais) da receita bruta da Perdigão.

d) A receita bruta estimada pela reta de regressão quando a receita de exportação da Sadia é estimada em 2.7 milhões de reais é dada por,

$$\begin{aligned} Y &= 0.7817 + 1.8371X \\ &= 0.7817 + 1.8371 \times 2.7 = 5.742 \end{aligned}$$

A receita bruta estimada pela reta de regressão quando a receita de exportação da Sadia é estimada em 3.8 milhões de reais é dada por,

$$\begin{aligned} Z &= 1.4180 + 1.8565W \\ &= 1.9427 + 1.5098 \times 3.8 = 8.47 \end{aligned}$$

Exercício 2

Os dados descritos a seguir referem-se ao consumo anual de vinho (litros de alcool de vinho consumidos por pessoa) e o total de mortes anuais (por 100.000 habitantes) por doenças cardíacas em 19 países.

C. Vinho	T. Mortes	C.Vinho	T. Mortes
2,5	211	1,8	167
3,9	167	1,9	266
2,9	131	0,8	227
2,4	191	6,5	86
2,9	220	1,6	207
0,8	297	5,8	115
9,1	71	1,3	285
0,8	211	1,2	199
0,7	300	2,7	172
7,9	107		

Resolva os itens a seguir.

(a) Construa o diagrama de dispersão entre o consumo de vinho (x) e o total de mortes anuais (y). Comente.

(b) Calcule o coeficiente de correlação linear de Pearson entre as duas variáveis. Comente.

(c) Ajuste a reta de regressão linear entre as duas variáveis e interprete o coeficiente angular b.

(d) Supor que um país que não está na amostra tenha tido um consumo de álcool de vinho de 4,1 litros por habitante num determinado ano. Estime pela reta de regressão o total previsto de mortes (por 100.000 habitantes) no ano.

Solução

a) Diagrama de dispersão entre consumo de vinho e total de mortes anuais.

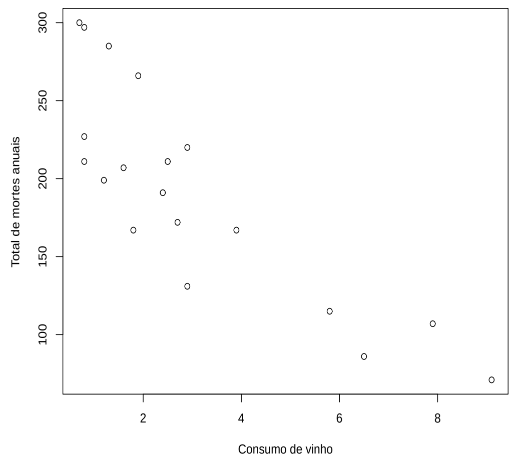

Figura 2: Diagrama de dispersão consumo de vinho (x) e taxa de morte (y).

– Observamos que a medida que aumenta o consumo de vinho (em litros por pessoa) diminui o total de mortes anuais (por 100000 habitantes)por doenças cardíacas. Nota-se também indícios de uma tendência linear.

b) Denotemos por,

– X : Consumo de vinho (litros de alcool de vinho consumidos por pessoa),

– Y : a taxa de mortes anuais (por 100000 habitantes) por doenças cardíacas.
De modo que,

* $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{57.5}{19} = 3.026$
* $\sum_{i=1}^{19} X_i^2 = 287.39$
* $S_X^2 = \frac{\sum_{i=1}^{12} X_i^2 - n\overline{X}^2}{n} = \frac{287.39 - 19 \times 3.026^2}{19} = 5.9691$
* $S_X = \sqrt{S_X^2} = 2.44$
* $\overline{Y} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3630}{19} = 191,0526$
* $\sum_{i=1}^{19} Y_i^2 = 777726$
* $S_Y^2 = \frac{\sum_{i=1}^{12} Y_i^2 - n\overline{Y}^2}{n} = \frac{777726 - 19 \times 191,0526^2}{19} = 4431.8514$
* $S_Y = \sqrt{S_Y^2} = 66,572$
* $\sum_{i=1}^{19}(X_i Y_i) = 8381.4$

Portanto o coeficiente de correlação linear entre o consumo de vinho (litros por pessoa) e a taxa de mortes anuais (por 100.00 habitantes) por doenças cardíacas em 19 países é dada por,

$$r = \frac{\sum_{i=1}^{12}(X_i Y_i) - n\overline{XY}}{\text{n}\, S_x S_y} = \frac{8381.4 - 19 \times 578,125}{\text{n} S_x S_y} = \frac{-2602.975}{19 \times 171.674} = -0.7980$$

O coeficiente de correlação linear obtido indica uma forte correlação linear (negativa) entre as variáveis. O que colabora com as conclusões tomadas a partir do diagrama de dispersão.

c) – A reta de regressão linear entre o consumo de vinho e o total de mortes anuais é, $\hat{Y} = a + bX$ onde,

  * b é dado por,

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\text{n S}} \\ &= \frac{8381.4 - 19 \times 578,125}{\text{nS}} = \frac{-2602.975}{19 \times 6.3007} \\ &= -21.7434 \end{aligned}$$

  * a é dado por,

$$a = \overline{Y} - b\overline{X} = 191,0526 - 3.026 \times (-21.7434) = 256,848$$

Portanto,

$$Y = a + bX = 256{,}848 - 21.7434X$$

Interpretação de b: Para cada aumento de uma unidade do consumo de vinho (1 litro de vinho consumido por pessoa) temos uma diminuição média de 22.9514 unidades de mortes anuais (por 100000 habitantes) por doenças cardíacas.

d) A taxa de mortes anuais (por 100000 habitantes) por doenças cardíacas estimada pela reta de regressão quando o consumo de vinho estimado é 4,1 litros por habitante num determinado ano é dada por,

$$\begin{aligned} Y &= 256{,}848 - 21.7434X \\ &= 256{,}848 - 21.7434 \times 4.1 = 167.7000 \end{aligned}$$

Exercício 3

Uma amostra de 10 casais e seus respectivos salários anuais (em u.m.) foi colhida num certo bairro conforme vemos na tabela abaixo

	Casal nº	1	2	3	4	5	6	7	8	9	10
Salário	Homem (X)	10	10	10	15	15	15	15	20	20	20
	Mulher (Y)	5	10	10	5	10	10	15	10	10	15

Sabe-se: $\sum x = 150$, $\sum x^2 = 2400$
$\sum y = 100$, $\sum y^2 = 1100$
$\sum xy = 1550$.

(a) Encontre o salário anual médio dos homens e desvio padrão do salário anual dos homens.

(b) Encontre o salário anual médio das mulheres e o desvio padrão do salário anual das mulheres.

(c) Construa o diagrama de dispersão.

(d) Encontre a correlação linear entre o salário anual dos homens e o das mulheres. Comente.

(e) Qual o salário médio familiar? E o desvio padrão?

(f) Se o homem é descontado em 8% e a mulher em 6% qual o salário líquido anual médio familiar? E o desvio padrão?

Solução

a) O salário médio anual dos homens é dado por,

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{150}{10} = 15$$

A variância é,

$$S_x^2 = \frac{1}{10}\left(\sum_{i=1}^{10} x_i{}^2 - 10\overline{x}^2\right) = \frac{1}{10}\left(2400 - 10 \times 15^2\right) = \frac{1}{10}(2400 - 2250) = \frac{150}{10} = \frac{15}{1}.$$

Donde,

$$S_x = \sqrt{S_x^2} = \sqrt{\frac{15}{1}} = 3.8730$$

b) O salário médio anual das mulheres é dado por,

$$\overline{y} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{100}{10} = 10$$

A variância é,

$$S_y^2 = \frac{1}{10}\left(\sum_{i=1}^{10} y_i{}^2 - 10\overline{y}^2\right) = \frac{1}{10}\left(1100 - 10 \times 10^2\right) = \frac{1}{10}(1100 - 1000) = \frac{10}{1}.$$

Donde,

$$S_y = \sqrt{S_y^2} = 3.1622$$

c) O diagrama de dispersão entre salário anual dos homens (X) e o salário das mulheres (Y) é dado por,

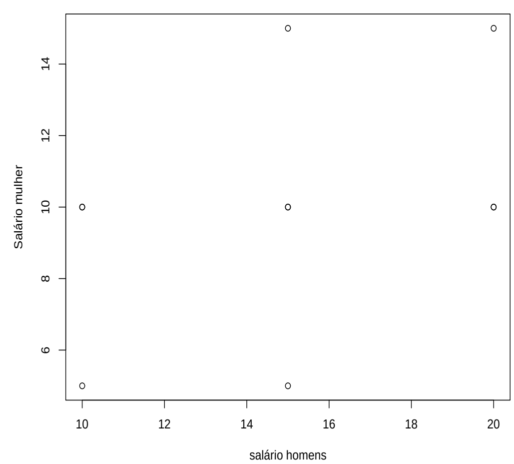

Figura 3: Diagrama de dispersão entre salário anual dos homens (x) e salário das mulheres (y)

d) Temos que,

$$\sum_{i=1}^{10} x_i y_i = 50 + 100 + 100 + 75 + 150 + 150 + 225 + 200 + 200 + 300 = 1550$$

Portanto o coeficiente de correlação linear entre o salário anual dos homens e mulheres,

$$r = \frac{\sum_{i=1}^{12}(X_i Y_i) - n\overline{XY}}{\text{n}S_x S_y} = \frac{1550 - 10 \times 15 \times 10}{10 \times 3.8730 \times 3.1622} = 0.4082$$

Não há indícios de correlação linear entre o salário anual dos homens e mulheres como já era esperado a partir da Figura 3.

e) Defina, $z_i = x_i + y_i$, $i = 1, \cdots, 10$. O salário médio anual familiar é dado por,

$$\overline{z} = \frac{\sum_{i=1}^{10} z_i}{10} = \overline{z} = \frac{\sum_{i=1}^{10} x_i + y_i}{10} = \frac{\sum_{i=1}^{10} x_i}{10} + \frac{\sum_{i=1}^{10} y_i}{10} = \frac{150 + 100}{10} = 25$$

A variância é,

$$\begin{aligned}S_z^2 &= \frac{1}{10}\left(\sum_{i=1}^{10} {z_i}^2 - 10\overline{z}^2\right) = \frac{1}{10}\left(\sum_{i=1}^{10} (x_i + y_i)^2 - 10 \times 25^2\right) = \frac{1}{10}\left(\sum_{i=1}^{10}(x_i^2 + 2x_iy_i + y_i^2) - 10 \times 25^2\right) \\ &= \frac{1}{10}\left(\sum_{i=1}^{10} x_i^2 + 2\sum_{i=1}^{10} x_iy_i + \sum_{i=1}^{10} y_i^2 - 10 \times 25^2\right) = \frac{1}{10}(2400 + 2 \times 1550 + 1100 - 10 \times 625) = \frac{350}{10}\end{aligned}$$

Donde,

$$S_z = \sqrt{S_z^2} = \sqrt{\frac{350}{10}} = 5.9160$$

e) Defina o salário líquido anual do homem como, $h_i = x_i - \frac{8}{100} \times x_i$, $i = 1, \cdots, 10$.
$m_i = y_i - \frac{6}{100} \times y_i$, $i = 1, \cdots, 10$.
De modo que o salário líquido familiar é definido como, $f_i = h_i + m_i$. Logo,

Tabela 2: Salário líquido anual

	Casal nº	1	2	3	4	5	6	7	8	9	10
Salário	Homem (H)	9,2	9,2	9,2	13,8	13.8	13.8	13.8	18.4	18.4	18.4
	Mulher (M)	4.7	9.4	9.4	4.7	9.4	9.4	14.1	9.4	9.4	14.1
	familiar(F)	13.9	18.6	18.6	18.5	23.2	23.2	27.9	27.8	27.8	32.5

Então,

$$\overline{f} = \frac{\sum_{i=1}^{10} f_i}{10} = \frac{232}{10} = 23,2$$

A variância é,

$$S_f^2 = \frac{1}{10}\left(\sum_{i=1}^{10} {f_i}^2 - 10\overline{f}^2\right) = \frac{1}{10}\left(5684.2 - 10 \times 23.2^2\right) = \frac{301.8}{10}$$

Donde,

$$S_f = \sqrt{S_f^2} = \sqrt{\frac{301.8}{10}} = 5.4936.$$