

Exercício 1

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

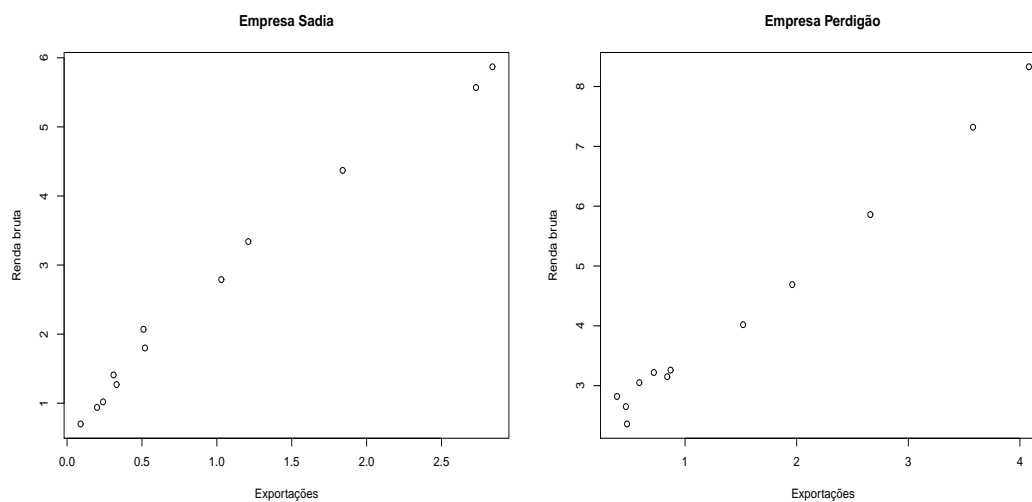


Figura 1: Diagrama de dispersão entre receita bruta (y) e receita com exportações (x) para as empresas Sadia e Perdigão.

- 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,

$$* \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{0,09+0,20+\dots+2,73+2,84}{12} = \frac{11,85}{12} = 0,9875$$

$$* \sum_{i=1}^{12} X_i^2 = 22.2703$$

$$* S_X^2 = \frac{\sum_{i=1}^{12} X_i^2 - n\bar{X}^2}{n-1} = \frac{22.2703 - 12 \times 0,9875^2}{11} = 0,9608$$

$$* S_X = \sqrt{S_X^2} = 0,9802$$

$$* \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{0,70+0,94+\dots+5,57+5,87}{12} = \frac{31,15}{12} = 2,5958$$

$$* \sum_{i=1}^{12} Y_i^2 = 117,0583$$

$$* S_Y^2 = \frac{\sum_{i=1}^{12} Y_i^2 - n\bar{Y}^2}{n-1} = \frac{117,0583 - 12 \times 2,5958^2}{11} = 3,2909$$

$$* S_Y = \sqrt{S_Y^2} = 1,8141$$

$$* \sum_{i=1}^{12} (X_i Y_i) = 50,1765$$

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

$$r = \frac{\sum_{i=1}^{12} (X_i Y_i) - n\bar{X}\bar{Y}}{(n-1)S_X S_Y} = \frac{50,1765 - 12 \times 2,5634}{(n-1)S_X S_Y} = \frac{19,4157}{11 \times 1,7781} = 0,9927$$

Denotemos por,

- W : a renda de exportação da empresa Perdigão,
- Z : a renda bruta da empresa Perdigão.

De modo que,

$$* \bar{W} = \frac{\sum_{i=1}^n W_i}{n} = \frac{0,48+0,39+\dots+3,59+4,08}{12} = \frac{18,16}{12} = 1,5133$$

$$* \sum_{i=1}^{12} W_i^2 = 45,6228$$

$$* S_W^2 = \frac{\sum_{i=1}^{12} W_i^2 - n\bar{W}^2}{n-1} = \frac{45,6228 - 12 \times 1,5133^2}{11} = 1,6493$$

$$* S_W = \sqrt{S_W^2} = 1,2842$$

$$* \bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} = \frac{50,73}{12} = 4,2275$$

$$* \sum_{i=1}^{12} Z_i^2 = 256,2329$$

$$* S_Z^2 = \frac{\sum_{i=1}^{12} Z_i^2 - n\bar{Z}^2}{n-1} = \frac{256,2329 - 12 \times 4,2275^2}{11} = 3,7974$$

$$* S_Z = \sqrt{S_Z^2} = 1,9487$$

$$* \sum_{i=1}^{12} (W_i Z_i) = 104,1606$$

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

$$r = \frac{\sum_{i=1}^{12} (W_i Z_i) - n \overline{WZ}}{(n-1)S_W S_Z} = \frac{104.1606 - 12 \times 6.3975}{(n-1)S_Z S_W} = \frac{27.3906}{11 \times 2.5025} = 0.9950$$

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,

$$\begin{aligned} b &= \frac{\sum_{i=1}^n X_i Y_i - n \overline{XY}}{(n-1)S_X^2} \\ &= \frac{50.1765 - 12 \times 2.5634}{(n-1)S_X^2} = \frac{19.4157}{11 \times 0.9608} \\ &= 1.8371 \end{aligned}$$

* a é dado por,

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

Portanto,

$$\hat{Y} = a + bX = 0.7817 + 1.8371X.$$

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,

$$\begin{aligned} d &= \frac{\sum_{i=1}^n W_i Y_i - n \overline{ZW}}{(n-1)S_W^2} \\ &= \frac{104.1606 - 12 \times 6.3975}{(n-1)1.6493} = \frac{27.3906}{11 \times 1.6493} = 1.5098 \end{aligned}$$

* e c é dado por,

$$c = \overline{Z} - d\overline{W} = 4.2275 - 1.5098 \times 1.5133 = 1.9427$$

Portanto,

$$\hat{Z} = 1.9427 + 1.5098W$$

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,

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

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,

$$\begin{aligned}\hat{Z} &= 1.9427 + 1.5098W \\ &= 1.9427 + 1.5098 \times 3.8 = 7.68\end{aligned}$$

Exercício 2

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

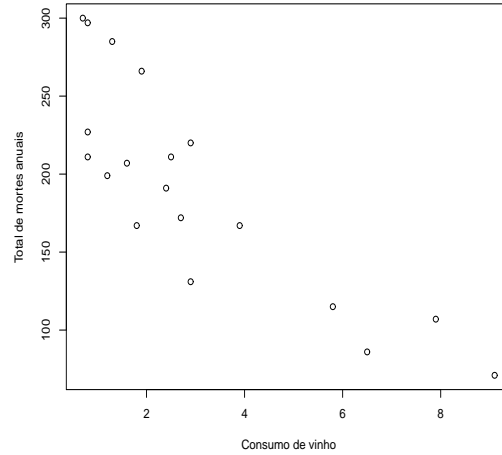


Figura 2: Diagrama de dispersão consumo de vinho (x) e taxa de morte (y).

– 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 álcool de vinho consumidos por pessoa),
- Y : a taxa de mortes anuais (por 100000 habitantes) por doenças cardíacas.

De modo que,

$$* \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{57.5}{19} = 3.026$$

$$* \sum_{i=1}^{19} X_i^2 = 287.39$$

$$* S_X^2 = \frac{\sum_{i=1}^{12} X_i^2 - n\bar{X}^2}{n-1} = \frac{287.39 - 19 \times 3.026^2}{18} = 6.3007$$

$$* S_X = \sqrt{S_X^2} = 2.51$$

$$* \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{3630}{19} = 191,0526$$

$$* \sum_{i=1}^{19} Y_i^2 = 777726$$

$$* S_Y^2 = \frac{\sum_{i=1}^{12} Y_i^2 - n\bar{Y}^2}{n-1} = \frac{777726 - 19 \times 191,0526^2}{18} = 4678.065$$

$$* S_Y = \sqrt{S_Y^2} = 68,396$$

$$* \sum_{i=1}^{19} (X_i Y_i) = 8381.4$$

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,

$$r = \frac{\sum_{i=1}^{12} (X_i Y_i) - n \bar{X} \bar{Y}}{(n-1) S_x S_y} = \frac{8381.4 - 19 \times 578,125}{(n-1) S_x S_y} = \frac{-2602.975}{18 \times 171.674} = -0.8424$$

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,

$$\begin{aligned} b &= \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{(n-1) S_x^2} \\ &= \frac{8381.4 - 19 \times 578,125}{(n-1) S_x^2} = \frac{-2602.975}{18 \times 6.3007} \\ &= -22.9514 \end{aligned}$$

* a é dado por,

$$a = \bar{Y} - b \bar{X} = 191,0526 - 3.026 \times (-22.9514) = 260,503$$

Portanto,

$$\hat{Y} = a + bX = 260,5037 - 22.9514X$$

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,

$$\begin{aligned} \hat{Y} &= 260,5037 - 22.9514X \\ &= 260,5037 - 22.9514 \times 4.1 = 166.403 \end{aligned}$$

Exercício 3

- a) Calculamos a proporção de segurados para cada faixa etária. Na Tabela abaixo podemos ver facilmente que a medida que aumenta a faixa etária a proporção de segurados cresce, sendo que na faixa etária de maiores de 35 anos encontram-se 72,2% dos segurados.

Tabela 2: Proporção de segurados para cada faixa etária

	< 25	25-29	30-35	> 35	Total
n_i	1138	2336	3007	16878	23359
Proporção	0.049	0.1	0.129	0.722	1

- b) Construímos uma tabela de contingência entre faixa-etária do segurado e distrito onde reside o segurado. Em cada cidade a distribuição dos segurados por faixa etária parecem bem

Tabela 3: Tabela de contingência entre faixa-etária do segurado e distrito onde reside o segurado.

	<25	25-29	30-35	>35	Total
1	638	1157	1396	7354	10545
	0.061	0.11	0.132	0.697	1
2	309	675	863	4806	6653
	0.047	0.101	0.13	0.722	1
3	119	335	493	3220	4167
	0.029	0.08	0.118	0.773	1
4	72	169	255	1498	1994
	0.036	0.085	0.128	0.751	1

próximas (próximas daquelas obtidas no item a considerando todas as cidades) de modo que não há indícios da correlação entre as variáveis distrito e faixa etária.

- c) Notamos que a maioria dos segurados (quase metade deles) residem no distrito 1, seguidos em ordem pelos distritos 2, 3 e por fim o 4 que possui menos de dez por cento dos segurados.

Tabela 4: Proporção de segurados para cada Distrito

	1	2	3	4	Total
n_i	10545	6653	4167	1994	23359
Proporção	0.451	0.286	0.178	0.085	

- d) Observe que não há indícios de correlação entre classe de consumo dos automóveis e o distrito. Note também que a maior propoção de segurados em todos os distritos pertecem a faixa de consumo de 1l a 1,5 litros. Parece haver uma proporção relativamente próxima de segurados com consumo inferior a 1 litro e com consumo entre 1,5 l e 2l e são poucos os segurados (menos de 10 por cento) que pertencem a classe de consumo superior a 2l.

Tabela 5: Tabela de contingência entre distrito onde reside o segurado e classe de consumo do automóvel

	<1l	1-1.5l	1.5-2l	>2l	total
1	2387 0.227	5098 0.483	2414 0.229	646 0.061	10545
2	1306 0.196	3324 0.5	1572 0.236	451 0.068	6653
3	845 0.203	2083 0.5	915 0.219	324 0.078	4167
4	409 0.205	958 0.481	469 0.235	158 0.079	1994