

Gabarito da Lista 2 – Tópico 2.3 e 2.4

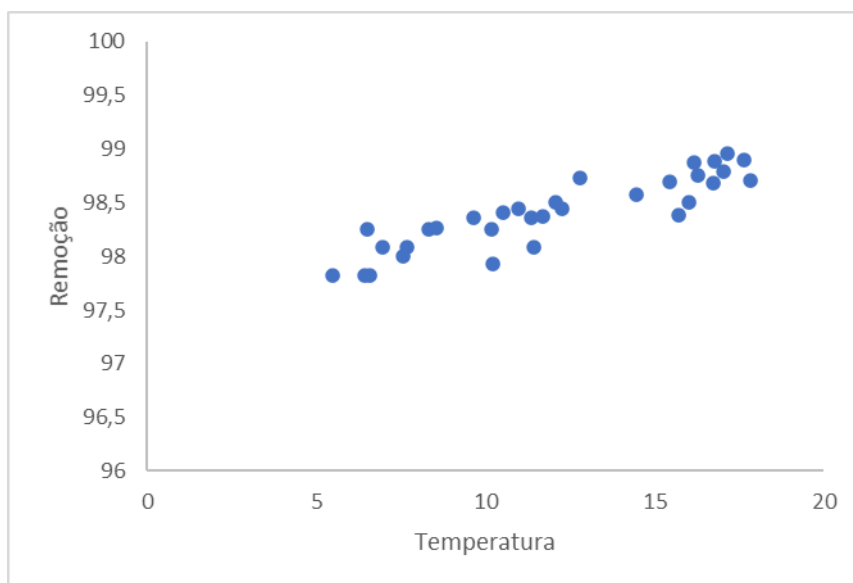
1.

a) Classifique as variáveis e identifique as escalas de medida.

Temperatura inicial (°C): variável numérica contínua, escala intervalar

Eficiência da remoção (%): variável numérica contínua, escala de razão

b) Utilizando a planilha Excel, construa o diagrama de dispersão bivariada de y e x. Este diagrama sugere uma associação entre as variáveis? Se sim, de que tipo?



O diagrama sugere uma associação linear positiva, pois os valores das variáveis crescem conjuntamente, formando uma elipse ascendente.

c) Utilizando a planilha Excel, construa a tabela auxiliar.

Tabela Auxiliar

i	x	y	x^2	y^2	xy
1	7,68	98,09	58,98	9621,65	753,33
2	6,51	98,25	42,38	9653,06	639,61
3	6,43	97,82	41,34	9568,75	628,98
4	5,48	97,82	30,03	9568,75	536,05
5	6,57	97,82	43,16	9568,75	642,68
6	10,22	97,93	104,45	9590,28	1000,84
7	15,69	98,38	246,18	9678,62	1543,58
8	16,77	98,89	281,23	9779,23	1658,39
9	17,13	98,96	293,44	9793,08	1695,18
10	17,63	98,90	310,82	9781,21	1743,61
11	16,72	98,68	279,56	9737,74	1649,93
12	15,45	98,69	238,70	9739,72	1524,76
13	12,06	98,51	145,44	9704,22	1188,03
14	11,44	98,09	130,87	9621,65	1122,15

15	10,17	98,25	103,43	9653,06	999,20
16	9,64	98,36	92,93	9674,69	948,19
17	8,55	98,27	73,10	9656,99	840,21
18	7,57	98,00	57,30	9604,00	741,86
19	6,94	98,09	48,16	9621,65	680,74
20	8,32	98,25	69,22	9653,06	817,44
21	10,50	98,41	110,25	9684,53	1033,31
22	16,02	98,51	256,64	9704,22	1578,13
23	17,83	98,71	317,91	9743,66	1760,00
24	17,03	98,79	290,02	9759,46	1682,39
25	16,18	98,87	261,79	9775,28	1599,72
26	16,26	98,76	264,39	9753,54	1605,84
27	14,44	98,58	208,51	9718,02	1423,50
28	12,78	98,73	163,33	9747,61	1261,77
29	12,25	98,45	150,06	9692,40	1206,01
30	11,69	98,37	136,66	9676,66	1149,95
31	11,34	98,36	128,60	9674,69	1115,40
32	10,97	98,45	120,34	9692,40	1080,00
Soma	384,26	3149,04	5099,2412	309892,65	37850,776
Média	12,00812	98,4076			

Cálculos intermediários:

$$SQX = 5099,2412 - 32 \times 12,00812^2 = 485$$

$$SQY = 309892,65 - 32 \times 98,4076^2 = 3,501$$

$$SPXY = 37850,776 - 32 \times 12,00812 \times 98,4076 = 36,71$$

d) Calcule a covariância para y e x (indicando as unidades de medida) e interprete-a.

$$s_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n-1} = 36,71 / 31 = 1,184 \text{ } ^\circ\text{C } \%$$

Interpretação: O sinal positivo indica que as variáveis temperatura e eficiência de remoção crescem conjuntamente.

e) Calcule o coeficiente de correlação linear para y e x e interprete-o.

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}} = 36,71 / \text{Raiz } (485 \times 3,501) = 0,8909$$

Interpretação: Correlação **forte e positiva** entre as variáveis temperatura e eficiência de remoção, ou seja, existe uma forte tendência de valores altos de temperatura estarem associados a valores altos de eficiência de remoção.

f) Estabeleça a hipótese de interesse a respeito do coeficiente de correlação linear de y e x, teste-a e redija a conclusão. Use $\alpha = 0,05$.

1. Pressuposição

- Distribuição da variável (X,Y) é normal bivariada;

2. Hipóteses estatísticas

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

3. Taxa de erro

$$\alpha = 0,05$$

4. Estatística do teste

$$t = \frac{\hat{\beta}_1}{\sqrt{\frac{\sum \hat{e}_i^2}{n-2}}} = 10,74$$

5. Decisão e conclusão

Como $t_{\alpha/2(30)} = 2,042 < t = 10,74$, rejeitamos H_0 .

Concluimos, ao nível de 5% de significância, que o coeficiente de correlação populacional difere de zero. Portanto, existe uma correlação linear positiva entre a temperatura e a eficiência de remoção. Valores altos de temperatura estão associados a valores altos de eficiência de remoção.

- g) Construa o intervalo de confiança, ao nível de 95%, para o coeficiente de correlação linear de y e x.

$$IC(\rho; 0,95): \left[\frac{\exp \left\{ 2 \left[\frac{1}{2} \ln \left(\frac{1 + 0,8909}{1 - 0,8909} \right) - \frac{1,96}{\sqrt{32 - 3}} \right] \right\} - 1}{\exp \left\{ 2 \left[\frac{1}{2} \ln \left(\frac{1 + 0,8909}{1 - 0,8909} \right) - \frac{1,96}{\sqrt{32 - 3}} \right] \right\} + 1}, \frac{\exp \left\{ 2 \left[\frac{1}{2} \ln \left(\frac{1 + 0,8909}{1 - 0,8909} \right) + \frac{1,96}{\sqrt{32 - 3}} \right] \right\} - 1}{\exp \left\{ 2 \left[\frac{1}{2} \ln \left(\frac{1 + 0,8909}{1 - 0,8909} \right) + \frac{1,96}{\sqrt{32 - 3}} \right] \right\} + 1} \right]$$

$$IC(\rho; 0,95): [0,7865; 0,9458]$$

Concluimos com 95% de confiança que os limites 0,7865 e 0,9458 compreendem o coeficiente de correlação populacional.

2.

- a) $r = -0,684$

Existe correlação negativa moderada entre a taxa de mortalidade e a taxa de alfabetização nos municípios amostrados. Isso significa que valores altos de mortalidade estão moderadamente associados a valores baixos de alfabetização.

- b) Alfabetização x Mortalidade

Mortalidade (x_i)	Alfabetização (y_i)	Posto de x_i	Posto de y_i	d_i	d_i^2
15,69	89,28	1	11	-10	100
16,62	77,54	2	6	-4	16
23,19	86,23	3	10	-7	49
31,71	83,38	4	9	-5	25
32,81	90,43	5	12	-7	49
37,04	81,82	6	8	-2	4
44,18	69,95	7	5	2	4
47,08	65,81	8	4	4	16
51,57	59,72	9	1	8	64
56,56	63	10	2	8	64
63,32	63,64	11	3	8	64
66,05	79,33	12	7	5	25
				Soma	480

$$r_s = -0,6783$$

Existe correlação negativa moderada entre os postos da a mortalidade até um ano de idade e da taxa de alfabetização nos municípios amostrados.

EspVida x Renda

EspVida (x_i)	Renda (y_i)	Posto de x_i	Posto de y_i	d_i	d_i^2
58,96	65,34	1	3	-2	4
59,58	66,96	2	4	-2	4
61,19	74,79	3	5	-2	4
62,45	58,68	4	1	3	9
63,65	60	5	2	3	9
64,46	80,69	6	6	0	0
67,42	125,75	7	7	0	0
67,69	188,29	8	10	-2	4
68,1	173,38	9	9	0	0
68,68	196,51	10	11	-1	1
71,01	150,67	11	8	3	9
71,36	264,55	12	12	0	0
				Soma	44

$$r_s = 0,8465$$

Existe correlação positiva e forte entre os postos da renda *per capita* e da esperança de vida ao nascer nos municípios amostrados.

c)

$$\begin{cases} H_0 : \rho_s = 0 \\ H_1 : \rho_s \neq 0 \end{cases}$$

$$n=12 \text{ e } \alpha=0,05$$

$$r_s \text{ crítico} = 0,587$$

Alfab x MorInf

$$|-0,6783| > 0,587 - \text{Rejeita-se } H_0$$

Concluimos, ao nível de 5% de significância, que o coeficiente de correlação de postos populacional difere de zero. Portanto, existe uma correlação negativa significativa entre os postos da mortalidade até um ano de idade e da taxa de alfabetização nos municípios amostrados. Valores altos de mortalidade estão moderadamente associados a valores baixos de alfabetização nos municípios.

EspVida x Renda

$$|0,8465| > 0,587 - \text{Rejeita-se } H_0$$

Concluimos, ao nível de 5% de significância, que o coeficiente de correlação de postos populacional difere de zero. Portanto, existe uma correlação positiva significativa entre os postos da renda *per capita* e da esperança de vida ao nascer nos municípios. Valores altos de renda estão fortemente associados a valores altos de esperança de vida ao nascer nos municípios.