Exercícios

1. Com os dados ao lado
   (a) Calcule a covariância de X e Y
   (b) Calcule o coeficiente de correlação de Pearson
   (c) Ao nível de 5% de significância, é possível afirmar que existe correlação entre X e Y?

X	Y
4	9
6	10
8	12
9	9
10	11
13	10
15	16
18	21
19	15

X	Y	(x - 11,33)	(y - 12,56)	(x - 11,33)^2	(y - 12,56)^2	(x - 11,33)(y - 12,56)
4	9	-7,33	-3,56	53,73	12,67	26,09
6	10	-5,33	-2,56	28,41	6,55	13,64
8	12	-3,33	-0,56	11,09	0,31	1,86
9	9	-2,33	-3,56	5,43	12,67	8,29
10	11	-1,33	-1,56	1,77	2,43	2,07
13	10	1,67	-2,56	2,79	6,55	-4,28
15	16	3,67	3,44	13,47	11,83	12,62
18	21	6,67	8,44	44,49	71,23	56,29
19	15	7,67	2,44	58,83	5,95	18,71
			Σ	**220,01**	**130,19**	**135,29**
			Σ /(N-1)	**23,45**	**13,47**	**14,03**

$\bar{x} = 11{,}33$
$\bar{y} = 12{,}56$
$Covar(x;y) = 14{,}03$

$$R = \frac{Covar(x;y)}{s_x s_y} = \frac{Covar(x;y)}{\sqrt{s_x^2 s_y^2}} = \frac{14{,}03}{\sqrt{23{,}45 \times 13{,}47}} = 0{,}789$$

$$t_{Calculado} = R\sqrt{\frac{(N-2)}{(1-R^2)}} = 0{,}789\sqrt{\frac{(9-2)}{(1-0{,}789^2)}} = 3{,}40$$

$t_{Crítico} = t_{N-2;\propto} = t_{7;5\%} = 2{,}365$ → Sim, é possível afirmar ao nível de 5% que existe correlação linear entre X e Y

2. Com os dados ao lado
   (a) Calcule a reta de regressão entre X e Y
   (b) A regressão é significante ao nível de 5%
   (c) O intercepto é significante ao nível de 5%
   (d) O coeficiente angular é significante ao nível de 5%
   (e) Construa um intervalo com 99% de confiança para o intercepto
   (f) Construa um intervalo com 99% de confiança para o coeficiente angular
   (g) Qual a estimativa do valor esperado de Y, quando X=20?
   (h) Construa um intervalo com 99% de confiança para valor esperado de Y quando X=24.
   (i) Construa um intervalo com 99% de confiança para valor de Y quando X=24.

X	Y
9	90
10	65
12	80
14	82
15	70
16	55
19	74
20	50
22	70
23	40
25	30
26	35
28	20
30	35
32	45

$$\sum x = 301$$
$$\sum x^2 = 6.785$$
$$\sum xy = 15.074$$
$$\sum y = 841$$

$$N = 15$$
$$\bar{x} = 20{,}07$$
$$s_x^2 = 53{,}21$$
$$\bar{y} = 56{,}07$$

(a) Calcule a reta de regressão entre X e Y: $f(x) = 104{,}63 - 2{,}42x$

X	Y	f(x)	(Ymedio - Y)^2	(Ymedio - f(x))^2	(Y- f(x))^2
9	90	82,85	1.151,24	717,17	51,12
10	65	80,43	79,74	593,41	238,08
12	80	75,59	572,64	381,03	19,45
14	82	70,75	672,36	215,50	126,56
15	70	68,33	194,04	150,31	2,79
16	55	65,91	1,14	96,83	119,03
19	74	58,65	321,48	6,66	235,62
20	50	56,23	36,84	0,03	38,81
22	70	51,39	194,04	21,90	346,33
23	40	48,97	258,24	50,41	80,46
25	30	44,13	679,64	142,56	199,66
26	35	41,71	443,94	206,21	45,02
28	20	36,87	1.301,04	368,64	284,60
30	35	32,03	443,94	577,92	8,82
32	45	27,19	122,54	834,05	317,20
			6.472,86	**4.362,63**	**2.113,55**

(b) A regressão é significante ao nível de 5% → Sim. Ver a ANOVA abaixo

Fonte de Variação	GL	Soma de Quadrados	Quadrado Médio	$F_{Calculado}$	$F_{Crítico}$
Regressão	1	4.362,63	4.362,63	28,90	4,600
Erro	14	2.113,55	150,97		
Total	13	6.472,86			

$$s_e = \sqrt{150{,}97} = 12{,}29$$

$$s_a = s_e \sqrt{\frac{1}{N} + \frac{\left(\frac{\sum x}{N}\right)^2}{\sum x^2 - \frac{(\sum x)^2}{N}}} = 12{,}29 \sqrt{\frac{1}{15} + \frac{\left(\frac{301}{15}\right)^2}{6.785 - \frac{(301)^2}{15}}} = 9{,}58$$

$$s_b = s_e \sqrt{\frac{N}{N\sum x^2 - (\sum x)^2}} = 12{,}29 \sqrt{\frac{15}{15 \times 6.785 - (301)^2}} = 0{,}45$$

(c) O intercepto é significante ao nível de 5%

$$t_{Calculado} = \frac{a - \alpha_0}{s_a} = \frac{104{,}63 - 0}{9{,}58} = 10{,}92$$

$$t_{Crítico} = t_{N-2;5\%} = 2{,}160$$

Sim. O intercepto é estatisticamente significante

(d) O coeficiente angular é significante ao nível de 5%

$$t_{Calculado} = \frac{b - \beta_0}{s_b} = \frac{-2{,}42 - 0}{0{,}45} = -5{,}38$$

$$t_{Crítico} = t_{N-2;5\%} = 2{,}160$$

Sim. O coeficiente angular é estatisticamente significante

(e) Construa um intervalo com 99% de confiança para o intercepto

$$\alpha = a + t_{N-2;\alpha} s_a = 104{,}63 \pm 3{,}102 \times 9{,}58$$

$$\alpha = 104{,}63 \pm 28{,}85$$

$$75{,}78 \leq \alpha \leq 133{,}48$$

(f) Construa um intervalo com 99% de confiança para o coeficiente angular

$$\beta = b + t_{N-2;\alpha} s_b = -2{,}42 \pm 3{,}102 \times 0{,}45$$

$$\beta = -2{,}42 \pm 1{,}36$$

$$-3{,}78 \leq \beta \leq -1{,}06$$

(g) Qual a estimativa do valor esperado de Y, quando X=20?

$$f(20) = 104{,}63 - 2{,}42 \times 20 = 56{,}23$$

(h) Construa um intervalo com 99% de confiança para valor esperado de Y quando X=24.

$$\mu(y|x) = f(x) \pm t_{N-2;\alpha/2} s_e \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{N\sum x^2 - (\sum x)^2}}$$

$$\mu(y|x = 24) = 46{,}55 \pm 3{,}012 \times 12{,}29 \sqrt{\frac{1}{15} + \frac{(24 - 20{,}07)^2}{15 \times 6.785 - (301)^2}}$$

$$\mu(y|x = 24) = 46{,}55 \pm 0{,}26$$

(i) Construa um intervalo com 99% de confiança para valor de Y quando X=24.

$$y = f(x) \pm t_{N-2;\alpha/2} s_e \sqrt{1 + \frac{1}{N} + \frac{(x-\bar{x})^2}{N\sum x^2 - (\sum x)^2}}$$

$$y = 46{,}55 \pm 3{,}012 \times 12{,}29 \sqrt{1 + \frac{1}{15} + \frac{(24-20{,}07)^2}{15 \times 6.785 - (301)^2}}$$

$$\mu(y|x = 24) = 46{,}55 \pm 1{,}03$$

Com uma amostra de 10 elementos foi feita uma tentativa de regressão linear. Os dados e resultados estão nas tabelas abaixo.

Id	X	Y	f(x)	(y - f(x))^2	(y - ymed)^2	(f(x) - ymed)^2
1	12	56	68,04	145,067	4515,840	3042,144
2	15	107	83,51	551,848	262,440	1575,410
3	18	80	98,97	359,966	1866,240	586,959
4	20	135	109,28	661,404	139,240	193,704
5	23	137	124,75	150,150	190,440	2,391
6	25	65	135,06	4907,827	3387,240	140,562
7	26	178	140,21	1428,038	3003,040	289,361
8	28	125	150,52	651,274	3,240	746,387
9	29	169	155,67	177,561	2097,640	1054,613
10	31	180	165,98	196,441	3226,240	1830,494
	Σ	**1.232**		**9.229,574**	**18.691,600**	**9.462,026**

Estatística de regressão	
R múltiplo	0,71149
R-Quadrado	0,506218
R-quadrado ajustado	0,444495
Erro padrão	33,96611
Observações	10

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	9.462,026	9.462,026	8,201484	0,021028
Erro	8	9.229,574	1.153,697		
Total	9	18.691,6			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*	*Inferior 95,0%*	*Superior 95,0%*
Interseção	6,187588	42,24704	0,146462	0,88718	-91,2343	103,6094	-91,2343	103,6094
X	5,154732	1,799948	2,863823	0,021028	1,004045	9,305418	1,004045	9,305418

3. Ao afirmar que existe regressão linear entre X e Y, a probabilidade de cometer um erro é de:

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	9.462,026	9.462,026	8,201484	***0,021028***
Erro	8	9.229,574	1.153,697		
Total	9	18.691,6			

4. Ao afirmar que o coeficiente linear da reta de regressão linear é diferente de zero, a probabilidade de cometer um erro é de:

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*	*Inferior 95,0%*	*Superior 95,0%*
Interseção	6,187588	42,24704	0,146462	***0,88718***	-91,2343	103,6094	-91,2343	103,6094
X	5,154732	1,799948	2,863823	0,021028	1,004045	9,305418	1,004045	9,305418

5. Ao afirmar que o coeficiente angular da reta de regressão linear é maior que zero, a probabilidade de cometer um erro é de:

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*	*Inferior 95,0%*	*Superior 95,0%*
Interseção	6,187588	42,24704	0,146462	0,88718	-91,2343	103,6094	-91,2343	103,6094
X	5,154732	1,799948	2,863823	***0,021028***	1,004045	9,305418	1,004045	9,305418

6. O modelo de regressão explica quanto do comportamento de Y?

Estatística de regressão	
R múltiplo	0,71149
R-Quadrado	***0,506218***
R-quadrado ajustado	0,444495
Erro padrão	33,96611
Observações	10

7. Qual o valor do coeficiente de correlação linear de Pearson entre X e Y?

Estatística de regressão	
R múltiplo	***0,71149***
R-Quadrado	0,506218
R-quadrado ajustado	0,444495
Erro padrão	33,96611
Observações	10

8. Considere a amostra abaixo. Faça uma regressão pela função $f(x) = a*e^{(bx)}$

X	Y
5	12
6	32
7	89
8	43
9	80
10	229
11	298
12	629

Regressão Linear

X	Y
5	2,485
6	3,466
7	4,489
8	3,761
9	4,382
10	5,434
11	5,697
12	6,444

N = 8
$\Sigma x = 68$
$\Sigma x^2 = 620$
$\Sigma y = 36,158$
$\Sigma xy = 328,505$
$f(x) = 0,236 + 0,504\ x$

$f(x) = 1,266e^{0,504\ x}$