Gabarito - Lista 7 – Tópicos 2.4, 3.4 e 4.4

1.

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) Alfab x MorInf

MorInf (x_i)	Alfab (y_i)	Posto de x_i	Posto de y_i	di	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	di	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)

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

$n=12$ e $\alpha=0,05$

r_s crítico = 0,587

Alfab x MorInf

$|-0,6783| > 0,587$ – Rejeita-se H_0

Concluímos, 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

$|0,8465| > 0,587$ – Rejeita-se H_0

Concluímos, 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.

2.1.

a) Obter os coeficientes de correlação linear entre y e x.

Matriz de correlações de Pearson

Corr	**ST**	**DQO**
ST	1	0,9554794
DQO	0,9554794	1

Existe uma correlação forte e positiva entre as variáveis redução de sólidos totais (x) e redução da demanda bioquímica de oxigênio (y), ou seja, existe uma forte tendência de valores altos de redução de ST estarem associados a valores altos de redução de DQO.

Teste de significância para o coeficiente de correlação

	Var_I	**Var_J**	**Correl**	**Coef_Det**	**T**	**Valor_p**	**Sig**	**Extr_Inf**	**Extr_Sup**
1	DQO	ST	0,95548	0,91294	18,02996	0	1%	0,911	0,97799

b) Obter os parâmetros do modelo.

Parametro	Estimativa	ErrPadrao	T	p	Inf95	Sup95	EstPadr
Intercep β_0	3,829633	1,768447	2,1655	0,038158	0,2228611	7,436405	-
ST β_1	0,9036432	0,05011897	18,03	0	0,8014249	1,005862	0,9554794

c) Testar a hipótese de interesse sobre o coeficiente de regressão utilizando o teste t e concluir com base no valor p.

Hipóteses estatísticas

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

Taxa de erro: $\alpha = 0,05$

Parametro	Estimativa	ErrPadrao	T	p	Inf95	Sup95	EstPadr
Intercep β_0	3,829633	1,768447	2,1655	0,038158	0,2228611	7,436405	-
ST β_1	0,9036432	0,05011897	18,03	0	0,8014249	1,005862	0,9554794

Como valor $p < \alpha$, rejeitamos H_0

Concluímos ao nível de 5% de significância que o coeficiente de regressão populacional difere de zero. Portanto, existe relação linear significativa entre a redução de ST e a redução da DQO.

d) Testar a hipóteses de interesse sobre o coeficiente de regressão utilizando a análise da variância e concluir com base no valor p.

Hipóteses estatísticas

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

Taxa de erro: $\alpha = 0,05$

Tabela da análise da variância

Fontes	GL	SQ	QM	F	p
Regressão	1	3390,5515	3390,5515	325,0795	0
Resíduo	31	323,32731	10,429913	-	-
Total	32	3713,8788	-	-	-

Como valor $p < \alpha$, rejeitamos H_0

Concluímos ao nível de 5% de significância que o coeficiente de regressão populacional difere de zero. Portanto, existe relação linear significativa entre a redução de ST e a redução da DQO.

e) Obter o coeficiente de determinação.

Estatísticas Auxiliares			
Estat	**DesvPadr**	**CoefDet**	**CDetAjust**
Valor	3,2295	0,91294	0,91013

Concluímos que 91% da variação da resposta (redução da DQO) é explicada pela variação da preditora (redução de ST).

f) Obter os limites de confiança, ao nível de 95%, para o intercepto e o coeficiente de regressão.

Parametro	Estimativa	ErrPadrao	T	p	Inf95	Sup95	EstPadr
Intercep β_0	3,829633	1,768447	2,1655	0,038158	0,2228611	7,436405	-
ST β_1	0,9036432	0,05011897	18,03	0	0,8014249	1,005862	0,9554794

Concluímos, com o nível de 95% de confiança, que o intercepto populacional (β_0) é coberto pelos limites 0,2229 a 7,436 e o coeficiente de regressão populacional (β_1) é coberto pelos limites 0,8014 e 1,006.

g) Obtenha a estimativa da média de y para x=30.

	DQO	ST	Estim	DiagH	Residuo	RSExt	DFitS	TipoObs
1	5	3	6,540563	0,2537	-1,540563	-0,5459	-0,3183	DI
2	11	7	10,15514	0,1989	0,8448643	0,2879	0,1434	DI
3	21	11	13,76971	0,1517	7,230291	2,658	1,124	DI
4	16	15	17,38428	0,1123	-1,384281	-0,449	-0,1597	Ok
5	16	18	20,09521	0,08783	-4,095211	-1,345	-0,4173	Ok
6	28	27	28,228	0,04034	-0,2279999	-0,0709	-0,01454	Ok
7	27	29	30,03529	0,03508	-3,035286	-0,9554	-0,1822	Ok
8	25	30	30,93893	0,03318	-5,93893	-1,953	-0,3618	Ok
9	35	30	30,93893	0,03318	4,06107	1,293	0,2395	Ok
10	30	31	31,84257	0,03175	-1,842573	-0,5735	-0,1039	Ok

O valor esperado para DQO quando a ST for igual a 30 é 30,94.

2.2.

a) Obter os coeficientes de correlação linear entre y e x1 e y e x2

Variaveis: Correlações entre as variáveis do modelo			
Variaveis	**ALUMINATO**	**SILICATO**	**calor**
ALUMINATO	1	0,22858	0,73072
SILICATO	0,22858	1	0,81625
calor	0,73072	0,81625	1

b) Obter os parâmetros do modelo.

Parametro	Estimativa	ErrPadrao	T	p	Inf95	Sup95
Intercep β_0	52,57735	2,286174	22,998	5,4566E-010	47,48344	57,67126
ALUMINATO β_1	1,468306	0,1213009	12,105	2,6922E-007	1,19803	1,738581
SILICATO β_2	0,6622505	0,04585472	14,442	5,029E-008	0,5600798	0,7644212

c) Efetuar a análise da variância para testar a hipótese geral de linearidade da relação entre y, x1 e x2; e concluir com base no valor p.

Hipótese estatística

$$\begin{cases} H_0 : \beta_i = 0, \text{ sendo } i = 1, 2 \\ H_1 : \beta_i \neq 0, \text{ para pelo menos um } i \ (i = 1 \text{ e/ou } 2) \end{cases}$$

Taxa de erro: $\alpha = 0,05$

Tabela da análise da variância

Quadro da Análise da Variação para calor					
Fontes	**GL**	**SQ**	**QM**	**F**	**p**
Regressão	2	2657,8586	1328,9293	229,5037	4,407E-009
Resíduo	10	57,904483	5,7904483	-	-
Total	12	2715,7631	-	-	-

Como valor p < α, rejeitamos H_0

Concluímos ao nível de 5% de significância que pelo menos um dos coeficientes de regressão parciais difere de zero. Portanto, existe relação linear significativa entre a quantidade de calor e pelo menos uma das variáveis quantidade de aluminato tricálcico e quantidade de silicato tricálcico.

d) Obter o coeficiente de determinação corrigido.

Estat: Estatísticas Auxiliares			
Estat	**DesvPadr**	**CoefDet**	**CDetAjust**
Valor	2,4063	0,97868	0,97441

Significa que 97% da variação da quantidade de calor é explicado pelo modelo de regressão múltipla com as variáveis preditoras quantidade de Aluminato e quantidade de Silicato.

e) Testar as hipóteses parciais sobre os coeficientes de regressão; e concluir com base no valor p.

Hipóteses estatísticas

$$\begin{cases} H_0^1 : \beta_1 = 0 \\ H_1^1 : \beta_1 \neq 0 \end{cases}$$

$$\begin{cases} H_0^2 : \beta_2 = 0 \\ H_1^2 : \beta_2 \neq 0 \end{cases}$$

Taxa de erro: α = 0,05

Parametro	**Estimativa**	**ErrPadrao**	**T**	**p**	**Inf95**	**Sup95**
Intercep β_0	52,57735	2,286174	22,998	5,4566E-010	47,48344	57,67126
ALUMINATO β_1	1,468306	0,1213009	12,105	2,6922E-007	1,19803	1,738581
SILICATO β_2	0,6622505	0,04585472	14,442	5,029E-008	0,5600798	0,7644212

Como ambos os valores p são menores que α, rejeitamos as duas hipóteses de nulidade testadas.

Concluímos, ao nível de 5% de significância, que os coeficientes de regressão parciais β_1 e β_2 diferem de zero. Portanto, existe efeito significativo da quantidade de silicato tricálcico sobre a quantidade de calor, adicional ao efeito da quantidade de aluminato tricálcico; também existe efeito significativo da quantidade de aluminato tricálcico sobre a quantidade de calor, adicional ao efeito da quantidade de silicato de tricálcico.

f) Obter os limites de confiança, ao nível de 95%, para os coeficientes de regressão parciais.

Parametro	**Estimativa**	**ErrPadrao**	**T**	**p**	**Inf95**	**Sup95**
Intercep β_0	52,57735	2,286174	22,998	5,4566E-010	47,48344	57,67126
ALUMINATO β_1	1,468306	0,1213009	12,105	2,6922E-007	1,19803	1,738581
SILICATO β_2	0,6622505	0,04585472	14,442	5,029E-008	0,5600798	0,7644212

Concluímos, com o nível de 95% de confiança, que o coeficiente de regressão parcial β_1 é coberto pelos limites 1,198 a 1,739 e o coeficiente de regressão parcial β_2 é coberto pelos limites 0,5601 e 0,7644.