

UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE DESENVOLVIMENTO TECNOLÓGICO

Exercícios

Métodos Estatísticos Aplicados a Recursos Hídricos

Clause Fátima de Brum Piana

Pelotas, 2016.

UNIDADE 2. CORRELAÇÃO LINEAR SIMPLES

1. O proprietário de uma grande cadeia de sorveterias gostaria de estudar os efeitos da temperatura atmosférica (x) sobre as vendas (y) durante a temporada do verão. Uma amostra de 21 dias consecutivos foi selecionada, com os seguintes resultados:

i	x	y	i	x	y
1	17,22	2933,6	12	23,89	3705,6
2	21,11	3242,4	13	36,67	6562,0
3	22,78	3474,0	14	37,78	6330,4
4	23,89	3956,5	15	33,33	6118,1
5	26,67	4554,8	16	30,56	5461,9
6	27,78	4342,5	17	28,89	4979,4
7	29,44	5172,4	18	31,11	5519,8
8	31,11	5597,0	19	26,67	4361,8
9	32,22	6060,2	20	27,78	4130,2
10	32,78	5905,8	21	24,44	3821,4
11	33,33	6253,2			

A partir dos dados acima:

- Construa o diagrama de dispersão (x,y);
- Trace uma reta para indicar a média de X e uma reta para indicar a média de Y no diagrama de dispersão e comente sobre a dispersão dos pontos;
- Calcule a covariância amostral (s_{xy}). O que ela indica?
- Calcule o coeficiente de correlação linear da amostra (r_{xy}) e interprete seu significado.
- Verifique se a correlação entre as variáveis x e y é significativa, ou seja, teste a hipótese de que $\rho_{xy}=0$.
- Construa o intervalo de confiança, ao nível de 95%, para o coeficiente de correlação linear populacional.
- Use o programa WinStat para proceder à análise de correlação para esses dados.

2. Um engenheiro químico está investigando se a temperatura (em °C) de operação do processo está relacionada com o rendimento do produto. O estudo resultou nos dados da tabela seguinte:

Temperatura (x)	Rendimento (y)
100	45
110	51
120	46
130	55
140	66
150	75
160	74
170	78
180	85
190	89

- Construa o diagrama de dispersão (x,y) e interprete-o.
- Calcule o coeficiente de correlação linear da amostra (r) e interprete-o.
- Teste a hipótese de que $\rho_{xy} = 0$.
- Construa o intervalo de confiança, ao nível de 95%, para o coeficiente de correlação linear populacional.
- Use o programa WinStat para proceder à análise de correlação para esses dados.

3. Uma pesquisa foi realizada com o objetivo de verificar se existe associação entre a falta de sono e a capacidade de as pessoas resolverem problemas simples. Foram testadas 10 pessoas, mantendo-se sem dormir por um determinado número de horas. Após cada um destes períodos, cada pessoa teve de resolver um teste com adições simples, anotando-se então os erros cometidos. Os dados resultantes são os seguintes:

Número de erros	6	8	6	10	8	14	12	14	12	16
Número de horas sem dormir	8	8	12	12	16	16	20	20	24	24

- Calcule o coeficiente de correlação e interprete-o.
- Teste a hipótese de que $\rho_{xy} = 0$.
- Construa o intervalo de confiança, ao nível de 95%, para o coeficiente de correlação linear populacional.
- Use o programa WinStat para proceder à análise de correlação para esses dados.

UNIDADE 3. REGRESSÃO LINEAR SIMPLES

1. Um engenheiro está estudando a bacia do rio São Francisco por meio de algumas variáveis que são observadas em 22 estações fluviométricas do alto rio. Um dos objetivos da pesquisa é identificar se existe relação entre a vazão média (em m^3/s) de longo termo e a área de drenagem. Os valores observados para essas variáveis foram os seguintes:

Estação	Área (x)	Vazão (y)	Estação	Área (x)	Vazão (y)
1	83,9	1,32	12	3727,4	65,30
2	188,3	2,29	13	4142,9	75,00
3	279,4	4,24	14	4874,2	77,20
4	481,3	7,34	15	5235,0	77,50
5	675,7	8,17	16	5414,2	86,80
6	769,7	8,49	17	5680,4	85,70
7	875,8	18,90	18	8734,0	128,00
8	964,2	18,30	19	10191,5	152,00
9	1206,9	19,30	20	13881,8	224,00
10	1743,5	34,20	21	14180,1	241,00
11	2242,4	40,90	22	29366,2	455,00

- Construa o diagrama de dispersão dos valores observados de X e Y.
- Supondo que a relação entre X e Y é linear, estabeleça o modelo estatístico que exprime essa relação, diga o significado de cada termo e liste as pressuposições associadas.
- Estime os parâmetros do modelo (β_0 e β_1).
- Ajuste a equação da reta e trace a reta ajustada no diagrama de dispersão.
- Teste a hipótese de interesse a respeito de β_1 utilizando a análise da variância.
- Verifique a qualidade do ajustamento do modelo aos dados observados.
- Construa os intervalos de confiança para β_1 e β_0 e interprete-os.
- Use o programa WinStat para proceder à análise de regressão linear para esses dados.

2. Os dados abaixo referem-se aos pesos iniciais e ganhos de peso (em gramas) de 15 ratos fêmeas entre 24 a 84 dias de idade, submetidas a uma dieta com altos teores de proteína. O objetivo deste estudo é verificar se o ganho de peso está relacionado com o peso inicial.

Rato	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Peso inicial (x)	50	64	76	64	74	60	69	68	56	48	57	59	46	45	65
Ganho de peso (y)	128	159	158	119	133	112	96	126	132	118	107	106	82	103	104

- Supondo que a relação entre X e Y é linear, estabeleça o modelo estatístico que exprime essa relação, diga o significado de cada termo.
- Estime os parâmetros do modelo (β_0 e β_1).
- Ajuste a equação da reta.
- Teste a hipótese de interesse a respeito de β_1 utilizando o teste t.
- Calcule o coeficiente de determinação e verifique a qualidade do ajustamento do modelo.
- Construa o intervalo de confiança para β_1 e interprete-o.
- Use o programa Winstat para proceder à análise de regressão linear para esses dados.

3. Um experimento mediu o rendimento de trigo (em toneladas por hectare) em porções de terra (parcelas) submetidas a sete níveis de Nitrogênio (em kg por hectare). Os resultados foram os seguintes:

Parcela	1	2	3	4	5	6	7
Quantidade de N (x)	40	60	80	100	120	140	160
Rendimento (y)	15,9	18,8	21,6	25,2	28,7	30,4	30,7

- Construa o diagrama de dispersão dos valores observados de X e Y.
- Supondo que a relação entre X e Y é linear, estabeleça o modelo estatístico que exprime essa relação, diga o significado de cada termo.
- Estime os parâmetros do modelo (β_0 e β_1).
- Ajuste a equação da reta e trace a reta ajustada no diagrama de dispersão.
- Teste a hipótese de interesse a respeito de β_1 utilizando a análise da variância.
- Verifique a qualidade do ajustamento do modelo aos dados observados.
- Faça a predição de Y para o x_4 e interprete esse valor.
- Faça a previsão de Y para o x_4 e interprete esse valor.
- Use o programa WinStat para proceder à análise de regressão linear para esses dados.

4. Em geral, experimentos afirmam que a percentagem de frutos atacados por larvas de mariposa é maior em macieiras anãs. Aparentemente, a densidade de mariposas não tem relação com o tamanho do pomar ou porte da planta, de modo que os riscos de ataque em um fruto em particular seriam aumentados quando o número de frutos ou plantas for pequeno. Os dados abaixo se referem aos resultados de um experimento contendo evidências sobre esse fenômeno.

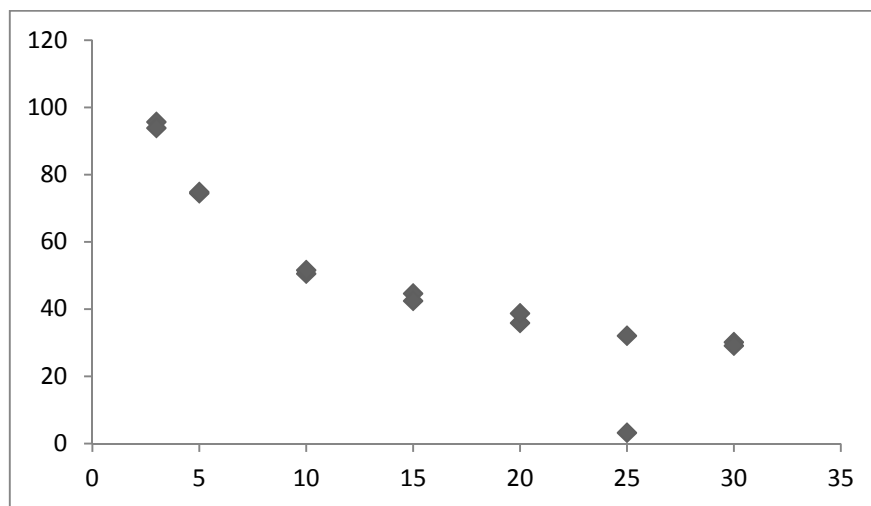
Planta	1	2	3	4	5	6	7	8	9	10	11	12
Número de frutos (em centenas)	8	6	11	22	14	17	18	24	19	23	26	40
Percentual de frutos atacados	59	58	56	53	50	45	43	42	39	38	30	27

- Supondo que a relação entre X e Y é linear, estabeleça o modelo estatístico que exprime essa relação, diga o significado de cada termo.
- Estime os parâmetros do modelo (β_0 e β_1).
- Ajuste a equação da reta.
- Teste a hipótese de interesse a respeito de β_1 utilizando o teste t.
- Use o programa Winstat para proceder à análise de regressão linear para esses dados.

5. Um sofisticado simulador estocástico de tráfego fornece a velocidade média (Y) em avenidas de uma metrópole em função do volume de automóveis (X). O resultado de 14 simulações revelou o seguinte:

i	Y	X
1	95,6	3
2	93,8	3
3	74,4	5
4	74,8	5
5	50,5	10
6	51,5	10
7	44,6	15
8	42,4	15
9	35,8	20
10	38,7	20
11	32,0	25
12	3,2	25
13	30,1	30
14	29,1	30
Soma	696,5	216
Média	49,75	15,43

- Calcule o valor do coeficiente de correlação e interprete-o.
- Teste a hipótese de que $\rho_{xy}=0$.
- Considerando que a variável volume de tráfego (X) tem efeito linear sobre a variável velocidade média (Y), estabeleça o modelo que expressa essa relação e explique o significado de cada termo.
- Estime os parâmetros do modelo (β_0 e β_1) e ajuste a equação da reta.
- Trace a reta da regressão no gráfico de dispersão dos valores observados.



- Estabeleça a hipótese de interesse sobre β_1 e conclua sobre esta hipótese a partir da tabela de análise de variância abaixo. Conclua com base no valor p.

Fontes	GL	SQ	S ²	F	Valor p
Regressão	1	7095,79	7095,79	45,36	0,00002089
Resíduo	12	1877,35	156,45	-	-
Total	13	8973,14	-	-	-

- g) Faça a **predição** de y para x=10 (intervalo com 95% de confiança).
- h) Faça a **previsão** de y para x=10 (intervalo com 95% de confiança).
- i) Use o programa WinStat para proceder à análise de regressão linear para esses dados.

UNIDADE 4. REGRESSÃO LINEAR MÚLTIPLA (2 VARIÁVEIS)

1. A tabela que segue apresenta os dados de peso (Y), em kg, altura (X₁), em cm, e idade (X₂), em meses, de cordeiros de um rebanho:

Peso (Y):	64	71	53	67	55	58	77	57	56	51	76	68
Altura (X ₁):	57	59	49	62	51	50	55	48	52	42	61	57
Idade (X ₂):	8	10	6	11	8	7	10	9	10	6	12	9

- a) Ajuste a equação do plano de regressão $E(Y) = \mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.
- b) Efetue a análise da variância e teste a hipótese de linearidade da relação entre Y e (X₁, X₂), ou seja:

$$\begin{cases} H_0 : \beta_i = 0, & i = 1, 2 \\ H_1 : \beta_i \neq 0, & i = 1 \text{ e/ou } 2 \end{cases}$$
- c) Calcule as estimativas das variâncias de $\hat{\beta}_1$ e $\hat{\beta}_2$ e teste as duas hipóteses parciais:

$$\begin{cases} H_0^1 : \beta_1 = 0 \\ H_1^1 : \beta_1 \neq 0 \end{cases} \quad \text{e} \quad \begin{cases} H_0^2 : \beta_2 = 0 \\ H_1^2 : \beta_2 \neq 0 \end{cases}$$
- d) Com base em resultados obtidos nos itens anteriores, indique: i) o acréscimo médio de peso por cm de aumento da altura, para cordeiros de uma mesma idade; ii) o acréscimo médio de peso por mês, para cordeiros de uma mesma altura.
- e) Calcule o coeficiente de determinação corrigido e interprete-o.
- f) Use o programa WinStat para proceder à análise de regressão linear múltipla para esses dados.

2. Os resultados que seguem foram obtidos de medidas tomadas em cada uma de 10 parcelas de 100 m² com uma cultivar de arroz, correspondentes a produção de grãos em kg (y), altura da planta em cm (x₁) e número de perfilhos por touceira (x₂):

$$\bar{y} = 66; \bar{x}_1 = 96; \bar{x}_2 = 17;$$

$$\sum (y_j - \bar{y})^2 = 321; \sum (x_{1j} - \bar{x}_1)^2 = 1.754; \sum (x_{2j} - \bar{x}_2)^2 = 32;$$

$$\sum (x_{1j} - \bar{x}_1)(y_j - \bar{y}) = -652; \sum (x_{2j} - \bar{x}_2)(y_j - \bar{y}) = 72; \sum (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) = -157.$$

- a) Estime uma equação de regressão linear múltipla que relacione a produção de grãos, Y, com a altura da planta, X₁, e o número de perfilhos, X₂.
- b) Na relação linear entre a variável Y e as variáveis X₁ e X₂, obtida no item a, explique o significado da estimativa do coeficiente de regressão parcial correspondente a X₁.
- c) Efetue a análise da variância e o teste F para a hipótese geral de linearidade da relação entre Y e (X₁, X₂).
- d) Calcule o coeficiente de determinação corrigido e explique o seu significado.
- e) Para a equação de regressão considerada no item a, teste as hipóteses parciais.

- f) Considerando os resultados do item anterior, qual das variáveis explanatórias, X_1 e X_2 , você escolheria, se tivesse que expressar o peso da produção de grãos através de uma equação de regressão linear simples. Seria, nesse caso, razoável a utilização de uma equação com uma única variável preditora?
- g) Use o programa WinStat para proceder à análise de regressão linear múltipla para esses dados.

3. Um estudo foi realizado para identificar o modelo que melhor representa a relação entre a variável resposta vazão mínima média (m^3/s) e as possíveis preditoras declividade de drenagem (m/km) e densidade de drenagem ($junções/km^2$). O primeiro modelo a ser testado é o modelo de regressão linear múltipla. Os dados observados em 15 estações fluviométricas da bacia do rio Paraopeba são apresentados na tabela abaixo.

Estação (j)	Vazão mínima média (Y)	Declividade de drenagem (X_1)	Densidade de drenagem (X_2)
1	2,60	2,69	0,098
2	1,49	3,94	0,079
3	1,43	7,20	0,119
4	3,44	3,18	0,102
5	1,37	2,44	0,123
6	2,53	1,25	0,136
7	15,12	1,81	0,121
8	16,21	1,59	0,137
9	21,16	1,21	0,134
10	30,26	1,08	0,018
11	28,53	1,00	0,141
12	1,33	4,52	0,064
13	0,43	10,27	0,131
14	39,12	0,66	0,143
15	45,00	0,60	0,133
Soma	210,02	43,44	1,679
Média	14,00	2,896	0,1119

- a) Estime os parâmetros da equação de regressão linear múltipla que relacione a vazão mínima média, Y, com a declividade, X_1 , e a densidade de drenagem, X_2 .
- b) Na relação linear entre a variável Y e as variáveis X_1 e X_2 , obtida no item anterior, explique o significado da estimativa do coeficiente de regressão parcial correspondente a X_1 .
- c) Efetue a análise da variância e o teste F para a hipótese geral de linearidade da relação entre Y e (X_1, X_2).
- d) Calcule o coeficiente de determinação corrigido e explique o seu significado.
- e) Para a equação de regressão considerada no item a, teste as hipóteses parciais.
- f) Considerando os resultados do item e, qual das variáveis explanatórias, X_1 e X_2 , você escolheria, se tivesse que expressar vazão mínima média através de uma equação de regressão linear simples. Seria, nesse caso, razoável a utilização de uma equação com apenas uma das variáveis preditoras?
- g) Use o programa WinStat para proceder à análise de regressão linear múltipla para esses dados.

UNIDADE 5. MODELOS DE CLASSIFICAÇÃO SIMPLES E DUPLA

1. Um biólogo deseja estudar o efeito do etanol sobre o tempo de sono. Uma amostra de 16 ratos, homogêneos em relação à idade e outras características importantes, foi selecionada e a cada rato foi dada uma injeção oral de etanol. Quatro concentrações, 0,0 (controle), 0,5, 1,0 e 1,5 mg/kg de peso vivo, de etanol foram utilizadas (em quatro grupos de tamanho quatro). A variável observada foi o tempo de sono com movimentos rápidos de olho (REM), registrado num período de 24 horas. Os resultados foram os seguintes:

Dose	Repetição			
	1	2	3	4
0,0 (controle)	88,6	73,2	91,4	68,0
0,5 mg/kg	63,0	73,9	69,2	80,1
1,0 mg/kg	64,2	69,0	59,1	72,4
1,5 mg/kg	44,9	59,5	70,2	76,3

- Identifique o fator de tratamento e seus níveis e a variável resposta.
- Estabeleça o modelo, diga o significado de cada termo e indique as pressuposições associadas.
- Defina os desvios importantes no processo de decomposição da variação total.
- Estabeleça e teste essa hipótese que corresponde aos objetivos do experimento. (Hipótese estatística, teste e conclusão).
- Faça a análise da variância desses dados utilizando o programa WinStat.

2. Na tabela abaixo é apresentado o conteúdo de Nitrogênio (mg) em plantas de trevo vermelho inoculadas com culturas de *Rhizobium trifolli* (do trevo vermelho) mais uma combinação de cinco raças de *R. meliloti*. (da alfafa). Foram utilizados seis tratamentos: cada uma das cinco culturas de *R. trifolli* testada individualmente com a combinação de raças de *R. meliloti*, (1, 2, 3, 4, 5) e mais uma combinação de todas as raças *R. trifolli* com todas as de *R. meliloti* (6). Foram utilizados cinco vasos (repetições) por tratamento.

Tratamentos	Repetição					Média
	1	2	3	4	5	
1 - 3Dok1	19,4	32,6	27,0	32,1	33,0	28,8
2 - 3Dok5	17,7	24,8	27,9	25,2	24,3	24,0
3 - 3Dok4	17,0	19,4	9,1	11,9	15,8	14,6
4 - 3Dok7	20,7	21,0	20,5	18,8	18,6	19,9
5 - 3Dok13	14,3	14,4	11,8	11,6	14,2	13,3
6 - Combinação	17,3	19,4	19,1	16,9	20,8	18,7

- Identifique o fator de tratamento e seus níveis e a variável resposta.
- Estabeleça o modelo, diga o significado de cada termo.
- Defina os desvios importantes no processo de decomposição da variação total.
- Estabeleça e teste essa hipótese que corresponde aos objetivos do experimento. (Hipótese estatística, teste e conclusão).
- Use o teste DMS de Fisher para discriminar a variação de tratamento ($\alpha=0,05$).
- Efetue a análise da variância e a discriminação da variação de tratamento usando o programa Winstat.

3. Um experimento foi conduzido para testar o efeito da dieta em leitões da raça Duroc. O pesquisador utilizou um conjunto de 30 leitões da raça Duroc, os quais foram casualizados aos cinco tratamentos. Os ganhos de peso dos animais (em kg) durante o período experimental foram os seguintes:

Dieta	1	2	3	4	5	6	Médias
Ração A	4,1	3,3	3,1	4,2	3,6	4,4	3,783
Ração B	5,2	4,8	4,5	6,8	5,5	6,2	5,500
Ração C	6,3	6,5	7,2	7,4	7,8	6,7	6,983
Ração D	6,5	6,8	7,3	7,4	6,9	7,0	7,000
Ração E	9,5	9,6	9,2	9,1	9,8	9,1	9,387
Média geral	-	-	-	-	-	-	6,530

- Identifique o fator de tratamento e seus níveis e a variável resposta.
- Estabeleça o modelo, diga o significado de cada termo e indique as pressuposições associadas.
- Estabeleça a hipótese de interesse.
- Complete a tabela da análise da variância para o experimento e conclua sobre a hipótese testada.

Tabela da análise da variância:

Fonte de variação	v	SQ	s ²	F	F tabelado
Dieta					
Resíduo				-	
Total		110,74	-	-	

- Compare as médias dos tratamentos pelo teste de Tukey, ao nível $\alpha=0,05$, apresentando todos os passos do processo, e redija as conclusões.
- Efetue a análise da variância e a discriminação da variação de tratamento usando o programa WinStat.

4. Num experimento destinado a comparar quatro tipos de teclado de computador, mediu-se a velocidade (em termos de palavras por minuto) com que cada digitador concluía o texto. Para aumentar a amplitude das conclusões, foi inserido um novo fator que correspondia a três tipos de parágrafos: um texto legal, um texto simples e um texto técnico. Cada combinação foi repetida duas vezes e o sorteio das combinações foi feito inteiramente ao acaso. Realizado o experimento, foram observados os seguintes resultados:

Tabela de médias

Parágrafo	Tipo de teclado				Médias
	A	B	C	D	
Legal	64,5	76,0	70,0	75,5	71,50
Comum	56,0	62,0	54,5	58,5	57,75
Técnico	54,5	66,5	57,0	72,0	62,50
Médias	58,33	68,17	60,50	68,67	63,92

- Identifique o fator de tratamento e seus níveis e a variável resposta.
- Estabeleça o modelo, diga o significado de cada termo e indique as pressuposições associadas.

- c) Defina os desvios importantes no processo de decomposição da variação total.
- d) Estabeleça as hipóteses que correspondem aos objetivos do experimento.
- e) Complete a tabela de análise da variância e conclua sobre as hipóteses.

Tabela da análise da variância

Fonte de variação	v	SQ	s ²	F
Teclado		500,83		
Parágrafo				
Teclado x Parágrafo				
Resíduo		57,00		
Total		1477,83		

- f) Teste os efeitos simples dos fatores usando o teste de Tukey.
- g) Faça a análise da variância e a discriminação da variação de tratamento usando o programa WinStat.

5 Um engenheiro suspeita que o acabamento de uma superfície de peças metálicas seja influenciado pelo tipo de tinta usada e pelo tempo de secagem. Ele seleciona dois tempos de secagem e usa três tipos de tinta. Três peças são testadas com cada combinação de tipo de tinta e tempo de secagem. Os dados observados são apresentados a seguir.

Acabamento de uma superfície, segundo o tratamento.

Tinta	Tempo de secagem (min)					
	20			30		
A	74	64	50	78	85	92
B	92	86	68	66	45	85
C	54	73	60	77	80	69

Tabela de médias.

Material	Tempo de secagem (min)		Médias marginais
	20	30	
A	62,67	85,00	73,83
B	82,00	65,33	73,67
C	62,33	75,33	68,83
Médias marginais	69,00	75,22	72,11

- a) Liste os fatores e correspondentes níveis e os tratamentos do experimento.
- b) Estabeleça o modelo estatístico e diga o significado de cada termo.
- c) Estabeleça as hipóteses de interesse.
- d) Construa o gráfico da interação.
- e) A partir da tabela da análise da variância conclua sobre as hipóteses testadas ($\alpha=0,05$).

Fonte de variação	v	SQ	s ²	F	Valor p
Material		96,78	48,39	0,3309	0,7246
Tempo		174,22	174,22	1,1915	0,2965
Material x Tempo		1244,11	622,06	4,2542	0,0401
Resíduo		1754,67	146,22	-	-
Total		3269,78	-	-	-

- f) Teste o efeito simples do tempo de secagem dentro de cada nível do fator Material, usando o teste DMS de Fisher, ao nível $\alpha=0,05$.
- g) Faça a análise da variância e a discriminação da variação de tratamento usando o programa Winstat.

Respostas dos exercícios

2. b) $r = 0,9736$. Correlação muito forte e positiva entre as variáveis temperatura do processo rendimento do produto, ou seja, existe uma forte tendência de valores altos de temperatura a valores altos de rendimento.

c) $t = 12,06$ e $t_{\alpha/2} = 2,306$. Rejeita-se H_0 .

d) IC (ρ ; 0,95): [0,88883 0,9939]

3. a) $r = 0,8015$. Correlação forte e positiva entre as variáveis número de erros e número de horas sem dormir.

b) $t = 3,79$ e $t_{\alpha/2} = 2,306$. Rejeita-se H_0 .

c) IC (ρ ; 0,95): [0,34688; 0,95113]