

UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE DESENVOLVIMENTO TECNOLÓGICO
MÉTODOS ESTATÍSTICOS APLICADOS A RECURSOS HÍDRICOS - 1ª PROVA

Nome: _____

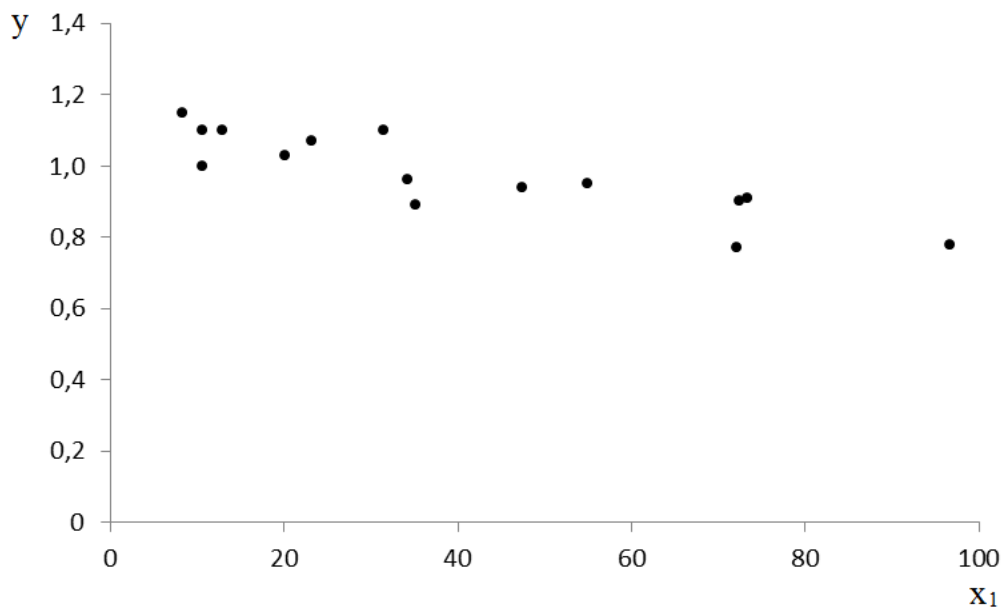
Data: 11/05/2015

Um estudo foi realizado em um caminhão de cargas leves movido a diesel para verificar se a umidade (em %) e a temperatura (em °C) do ar influenciam a emissão de óxido nitroso (em ppm). As medições das emissões foram tomadas em momentos diferentes, com condições experimentais variadas. Os dados observados estão na tabela abaixo.

j	Dados observados			Tabela auxiliar					
	Óxido nitroso (y)	Umidade (x ₁)	Temperatura (x ₂)	y ²	x ₁ ²	x ₂ ²	x ₁ y	x ₂ y	x ₁ x ₂
1	0,90	72,4	24,6	0,81	5241,76	605,71	65,16	22,15	1781,84
2	0,96	34,3	25,1	0,92	1176,49	627,78	32,93	24,05	859,41
3	0,89	35,1	20,0	0,79	1232,01	400,00	31,24	17,80	702,00
4	1,00	10,7	26,1	1,00	114,49	681,79	10,70	26,11	279,39
5	1,10	12,9	19,7	1,21	166,41	386,78	14,19	21,63	253,70
6	1,15	8,3	19,3	1,32	68,89	373,78	9,55	22,23	160,47
7	1,03	20,1	24,9	1,06	404,01	622,23	20,70	25,69	501,38
8	0,77	72,2	25,4	0,59	5212,84	644,60	55,59	19,55	1833,08
9	1,07	23,2	24,9	1,14	538,24	619,46	24,82	26,63	577,42
10	0,94	47,4	30,3	0,88	2246,76	920,11	44,56	28,51	1437,80
11	1,10	31,5	24,9	1,21	992,25	622,23	34,65	27,44	785,75
12	1,10	10,6	30,2	1,21	112,36	910,03	11,66	33,18	319,77
13	0,91	73,3	24,6	0,83	5372,89	605,71	66,70	22,40	1803,99
14	0,78	96,6	25,9	0,61	9331,56	673,11	75,35	20,24	2506,23
15	0,95	54,9	21,6	0,90	3014,01	467,04	52,16	20,53	1186,45
Soma	14,65	603,5	367,6	14,50	35224,97	9160,34	549,96	358,15	14988,68
Média	0,9767	40,23	24,51						

Questão 1 (5,0). Considere apenas as variáveis y e x₁.

a) (0,2) Interprete o diagrama de dispersão desses dados.



- b) (0,5) Calcule o coeficiente de correlação linear simples e interprete-o.
- c) (0,6) Estabeleça a hipótese de interesse a respeito do coeficiente de correlação linear simples, teste-a e redija a conclusão. Use $\alpha = 0,05$.
- d) (0,2) Interprete o seguinte intervalo de confiança: IC (ρ ; 0,95): [-0,9551; -0,6380].
- e) (0,8) Assumindo que o modelo de regressão linear simples é adequado para descrever a relação entre x_1 e y , estabeleça o modelo, diga o significado de cada termo (contextualizando) e indique as pressuposições associadas ao modelo.
- f) (0,5) Estime os parâmetros do modelo de regressão linear simples.
- g) (0,6) Ajuste a equação da reta, calcule $\hat{\mu}_5$ e $\hat{\epsilon}_5$ e trace a reta ajustada no diagrama de dispersão.
- h) (0,8) Estabeleça a hipótese de interesse a respeito do coeficiente de regressão e teste-a utilizando a análise da variância. Use $\alpha = 0,05$. (Construa a tabela da análise da variância e redija a conclusão.)
- i) (0,4) Faça a **predição** de y para $x_1=60$ (intervalo com 99% de confiança) e interprete o resultado.
- j) (0,4) Faça a **previsão** de y para $x_1=60$ (intervalo com 99% de confiança) e interprete o resultado.

Questão 2 (4,0). Considere as variáveis y , x_1 e x_2 .

- a) (0,8) Assumindo que o modelo de regressão linear múltipla é adequado para descrever a relação entre y , x_1 e x_2 , estime os parâmetros do modelo e ajuste a equação do plano.
- b) (0,2) Estime a média de óxido nitroso emitido para $x_1 = 50\%$ e $x_2 = 28^\circ\text{C}$.
- c) (0,2) Explique o significado da estimativa do coeficiente de regressão parcial β_1 neste modelo.
- d) (0,8) Efetue a análise da variância para testar a hipótese geral de linearidade da relação entre y , x_1 e x_2 . (Estabeleça a hipótese de interesse, construa a tabela da análise da variância e redija a conclusão. Use $\alpha = 0,05$)
- e) (0,4) Calcule o coeficiente de determinação corrigido e explique o seu significado.
- f) (0,8) Teste as hipóteses parciais, usando $\alpha = 0,05$. (Relacione todos os passos do teste de hipóteses e redija as conclusões.)
- g) (0,6) Considerando os resultados do item anterior, qual é o modelo mais adequado para explicar a variação do óxido nitroso emitido por um caminhão de cargas leves movido a diesel? Por quê? Escreva a equação ajustada deste modelo e o seu coeficiente de determinação.
- h) (0,2) Estime a média de óxido nitroso emitido para $x_1 = 50\%$.

Questão 3 (1,0). Complete as afirmações com V (verdadeiro) ou F (falso) e **corrija** as falsas.

- a) () O método científico compreende uma sequência de operações, em que a primeira etapa consiste na formulação de hipóteses científicas.
- b) () Os métodos de pesquisa mais utilizados na engenharia são: o estudo observacional e o experimento. No experimento não há intervenção do pesquisador na amostra, mas no estudo observacional há.
- c) () Processos hidrológicos determinísticos ocorrem com alta frequência em rios naturais.
- d) () O modelo de regressão linear é adequado quando a parte aleatória do modelo é preponderante sobre a parte fixa.

- e) () O coeficiente de correlação linear não permite estabelecer relação de causa e efeito entre as variáveis estudadas.
- f) () Se em um teste de hipóteses o valor p for igual a 0,28 devemos rejeitar H_0 .
- g) () O método dos mínimos quadrados tem como objetivo obter estimativas de β_1 e β_0 de tal forma que a soma dos erros seja o menor valor possível.
- h) () A amplitude do intervalo de confiança para μ diminui quando x_i se afasta da média \bar{x} .
- i) () A temperatura é um exemplo de variável discreta medida em escala intervalar.
- j) () Não é recomendada a extrapolação da equação de regressão para além dos limites dos dados amostrais utilizados na estimativa dos parâmetros do modelo de regressão. Uma das razões para isso é que a relação entre as variáveis X e Y pode não ser linear para valores que extrapolam os dados utilizados na regressão.

Tabela 1. Limites da distribuição t de Student.

Graus de Liberdade (ν)	Limites bilaterais: $P(t > t_{\alpha/2})$							
	Nível de Significância (α)							
	0,50	0,20	0,10	0,05	0,025	0,02	0,01	0,005
12	0,695	1,356	1,782	2,179	2,560	2,681	3,055	3,428
13	0,694	1,350	1,771	2,160	2,533	2,650	3,012	3,373
14	0,692	1,345	1,761	2,145	2,510	2,624	2,977	3,326
15	0,691	1,341	1,753	2,132	2,490	2,602	2,947	3,286
16	0,690	1,337	1,746	2,120	2,473	2,583	2,921	3,252
17	0,689	1,333	1,740	2,110	2,458	2,567	2,898	3,223
18	0,688	1,330	1,734	2,101	2,445	2,552	2,878	3,197

Tabela 2. Limites unilaterais superiores da distribuição F: $P[F > f_\alpha]$

ν_2	α	ν_1																			
		1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120	Inf.
10	0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
	0,025	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
	0,01	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,78	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
	0,001	21,04	14,91	12,55	11,28	10,48	9,92	9,52	9,20	8,96	8,75	8,59	8,45	8,13	7,80	7,64	7,47	7,30	7,12	6,94	6,76
11	0,05	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
	0,025	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
	0,01	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
	0,001	19,69	13,81	11,56	10,35	9,58	9,05	8,66	8,35	8,12	7,92	7,76	7,63	7,32	7,01	6,85	6,68	6,52	6,35	6,17	6,00
12	0,05	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
	0,025	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
	0,01	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
	0,001	18,64	12,97	10,80	9,63	8,89	8,38	8,00	7,71	7,48	7,29	7,14	7,00	6,71	6,40	6,25	6,09	5,93	5,76	5,59	5,42
13	0,05	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
	0,025	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
	0,01	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
	0,001	17,81	12,31	10,21	9,07	8,35	7,86	7,49	7,21	6,98	6,80	6,65	6,52	6,23	5,93	5,78	5,63	5,47	5,30	5,14	4,97

Formulário

Cálculos preliminares

$$SQX_1 = \sum x_{1j}^2 - n\bar{x}_1^2$$

$$SQX_2 = \sum x_{2j}^2 - n\bar{x}_2^2$$

$$SQY = \sum y_j^2 - n\bar{y}^2$$

$$SPX_1X_2 = \sum x_{1j}x_{2j} - n\bar{x}_1\bar{x}_2$$

$$SPX_1Y = \sum x_{1j}y_j - n\bar{x}_1\bar{y}$$

$$SPX_2Y = \sum x_{2j}y_j - n\bar{x}_2\bar{y}$$

Regressão linear simples

$$\hat{\beta}_1 = \frac{SPXY}{SQX}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$SQ_{Total} = SQY$$

$$SQ_{Reg} = \hat{\beta}_1^2 SQX$$

$$SQ_{Res} = \sum \hat{e}_i^2 \text{ (por diferença)}$$

$$r^2 = \frac{SQ_{Reg}}{SQ_{Total}}$$

$$S^2(\hat{\beta}_1) = \frac{S^2 Res}{SQX}$$

$$S^2(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{SQX} \right) (S^2 Res)$$

$$S^2(\hat{\mu}_i) = \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SQX} \right) (S^2 Res)$$

$$S^2(y_i) = \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SQX} \right) (S^2 Res)$$

$$IC(\beta_1; 1-\alpha) : \hat{\beta}_1 \pm t_{\alpha/2} \sqrt{S^2(\hat{\beta}_1)}$$

$$IC(\beta_0; 1-\alpha) : \hat{\beta}_0 \pm t_{\alpha/2} \sqrt{S^2(\hat{\beta}_0)}$$

$$IC(\mu_i; 1-\alpha) : \hat{\mu}_i \pm t_{\alpha/2} \sqrt{S^2(\hat{\mu}_i)} \text{ (predição)}$$

$$IC(y_i; 1-\alpha) : \hat{y}_i \pm t_{\alpha/2} \sqrt{S^2(y_i)} \text{ (previsão)}$$

Teste t

$$T = \frac{\hat{\theta}}{S(\hat{\theta})} \sim t(v), \text{ onde } v = n - p$$

Correlação linear

$$r_{xy} = \frac{SPXY}{\sqrt{SQX}\sqrt{SQY}}$$

$$S^2(r_{xy}) = \frac{1 - r_{xy}^2}{n - 2}$$

Regressão linear múltipla

$$\begin{cases} \hat{\beta}_1 SQX_1 + \hat{\beta}_2 SPX_1X_2 = SPX_1Y \\ \hat{\beta}_1 SPX_1X_2 + \hat{\beta}_2 SQX_2 = SPX_2Y \end{cases}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2$$

$$S^2(\hat{\beta}_1) = \frac{SQX_2}{SQX_1 \times SQX_2 - (SPX_1X_2)^2} (S^2 Res)$$

$$S^2(\hat{\beta}_2) = \frac{SQX_1}{SQX_1 \times SQX_2 - (SPX_1X_2)^2} (S^2 Res)$$

$$SQ_{Total} = SQY$$

$$SQ_{Reg} = \hat{\beta}_1 SPX_1Y + \hat{\beta}_2 SPX_2Y$$

$$SQ_{Res} = \sum \hat{e}_j^2 \text{ (por diferença)}$$

$$r_C^2 = r^2 - \frac{2}{n-3}(1-r^2)$$