

1- Um pesquisador químico está interessado em verificar como o grau de contaminação de um material (Var. gconta) influencia a o percentual de pureza de uma outra substancia (Var. impu%). Teste se o modelo linear é recomendado para o caso, se sim, qual é o modelo ? Diga se há e qual a correlação entre as vars.

The regression equation is  
 impu% = - 0,29 + 0,457 gconta

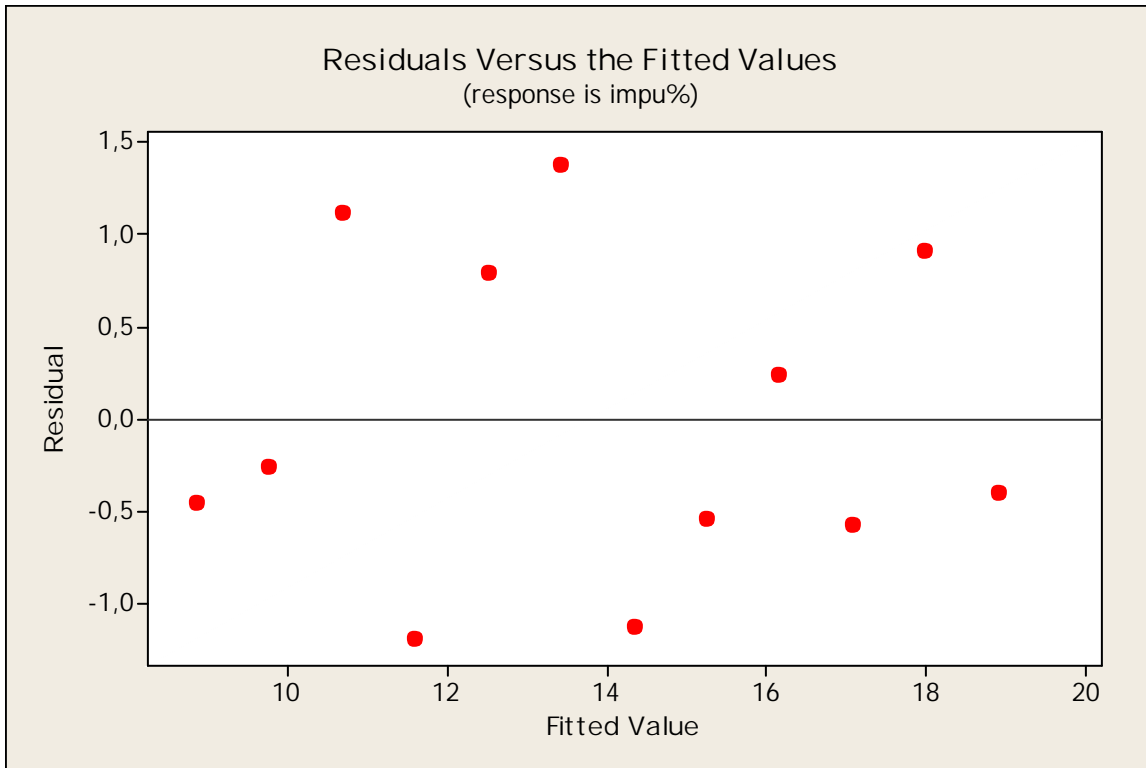
Predictor	Coef	SE Coef	T	P
Constant	-0,289	1,221	-0,24	0,817
gconta	0,45664	0,03844	11,88	0,000

S = 0,919316    R-Sq = 93,4%    R-Sq(adj) = 92,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	119,28	119,28	141,13	0,000
Residual Error	10	8,45	0,85		
Total	11	127,73			

O modelo linear é altamente significativo ( $p < 0,001$ ), com ótimo  $R^2 = 92,7\%$  e a correlação = 0,97 é excelente e significativa ( $p < 0,001$ ). Abaixo temos o gráfico de resíduos com os pontos distribuídos aleatoriamente.



2 – Verifique qual o melhor modelo para explicar a relação entre a var. preditora X e var. resposta Y.

The regression equation is  
 $Y = 12,61 - 3,238 X$

S = 2,15141    R-Sq = 24,5%    R-Sq(adj) = 16,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	14,9908	14,9908	3,24	0,102
Error	10	46,2858	4,6286		
Total	11	61,2767			

**O ajuste linear não foi bom.**

The regression equation is  
 $Y = 41,81 - 45,93 X + 14,72 X^{**2}$

S = 1,38906    R-Sq = 71,7%    R-Sq(adj) = 65,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	43,9113	21,9556	11,38	0,003

Error	9	17,3654	1,9295
Total	11	61,2767	

Sequential Analysis of Variance

Source	DF	SS	F	P
Linear	1	14,9908	3,24	0,102
Quadratic	1	28,9204	14,99	0,004

O ajuste quadrático foi significativo.

The regression equation is

$$Y = 21,72 - 1,25 X - 17,16 X^{**2} + 7,33 X^{**3}$$

S = 1,44668    R-Sq = 72,7%    R-Sq(adj) = 62,4%

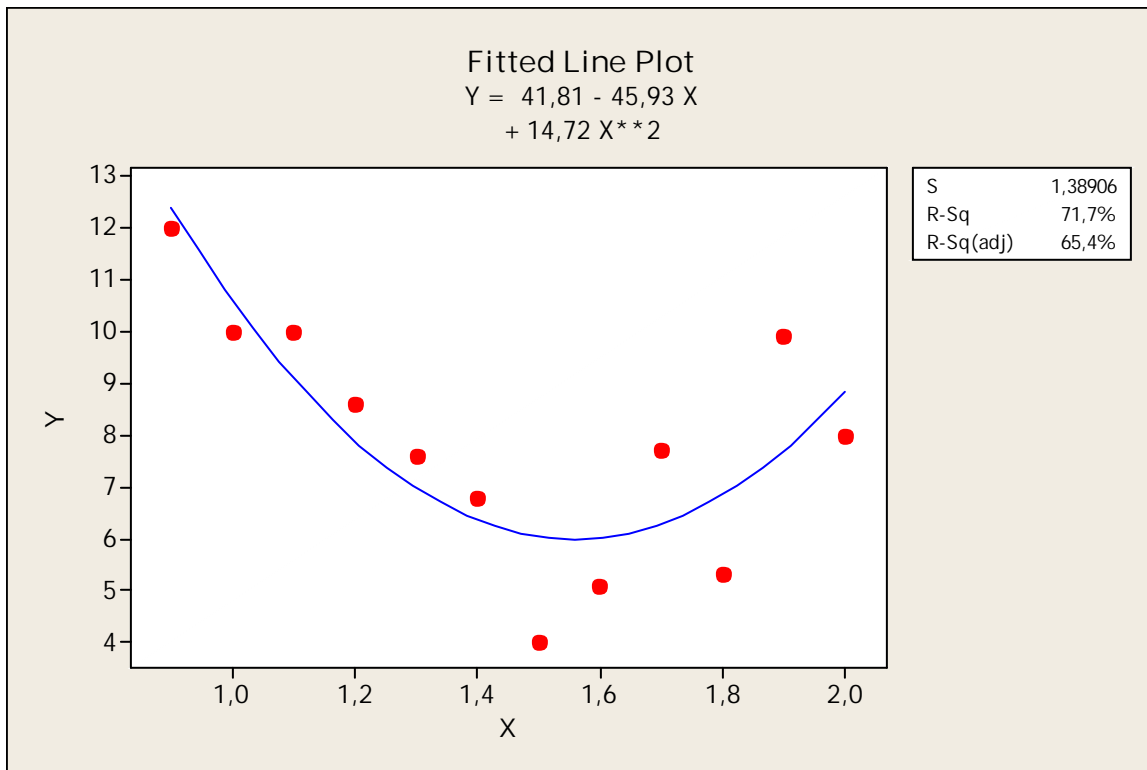
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	44,5336	14,8445	7,09	0,012
Error	8	16,7431	2,0929		
Total	11	61,2767			

Sequential Analysis of Variance

Source	DF	SS	F	P
Linear	1	14,9908	3,24	0,102
Quadratic	1	28,9204	14,99	0,004
Cubic	1	0,6223	0,30	0,600

O ajuste quadrático é realmente o melhor, abaixo a grafico.



3 – Um cientista está estudando como a presença da substancia T (var. subs) no organismo de uma pessoa está relacionada com o peso da mesma (var. peso), verifique isto para ele da maneira mais completa possível (correlação, regressão, modelo mais apropriado, gráfico de resíduos). Elabore a resposta que vc daria a ele.

O modelo linear foi o que apresentou o melhor ajuste, com correlação = -0,86 e  $p < 0,001$ , demonstrando que o aumento da substancia acarreta diminuição de peso.

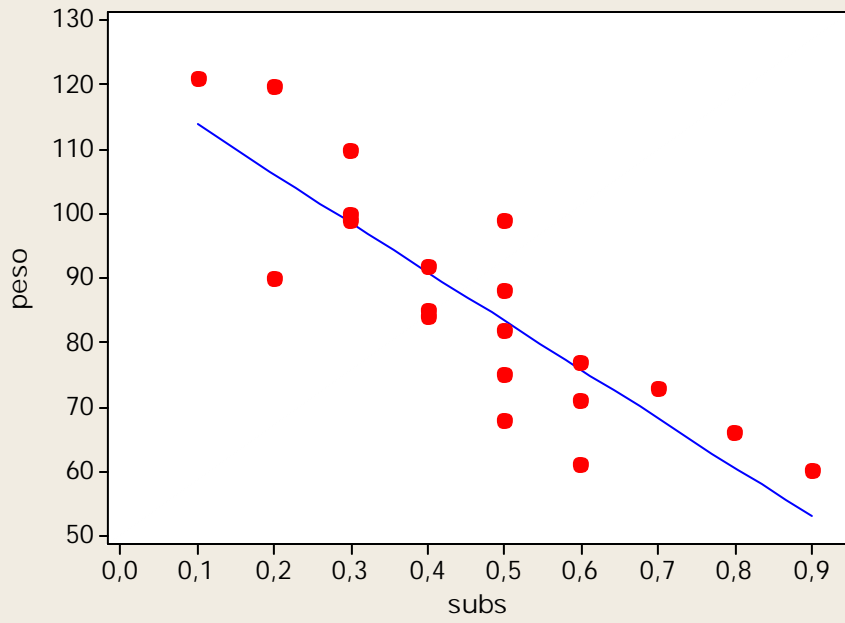
The regression equation is  
 peso = 121,4 - 76,09 subs

S = 9,55020    R-Sq = 73,5%    R-Sq(adj) = 72,0%

Analysis of Variance

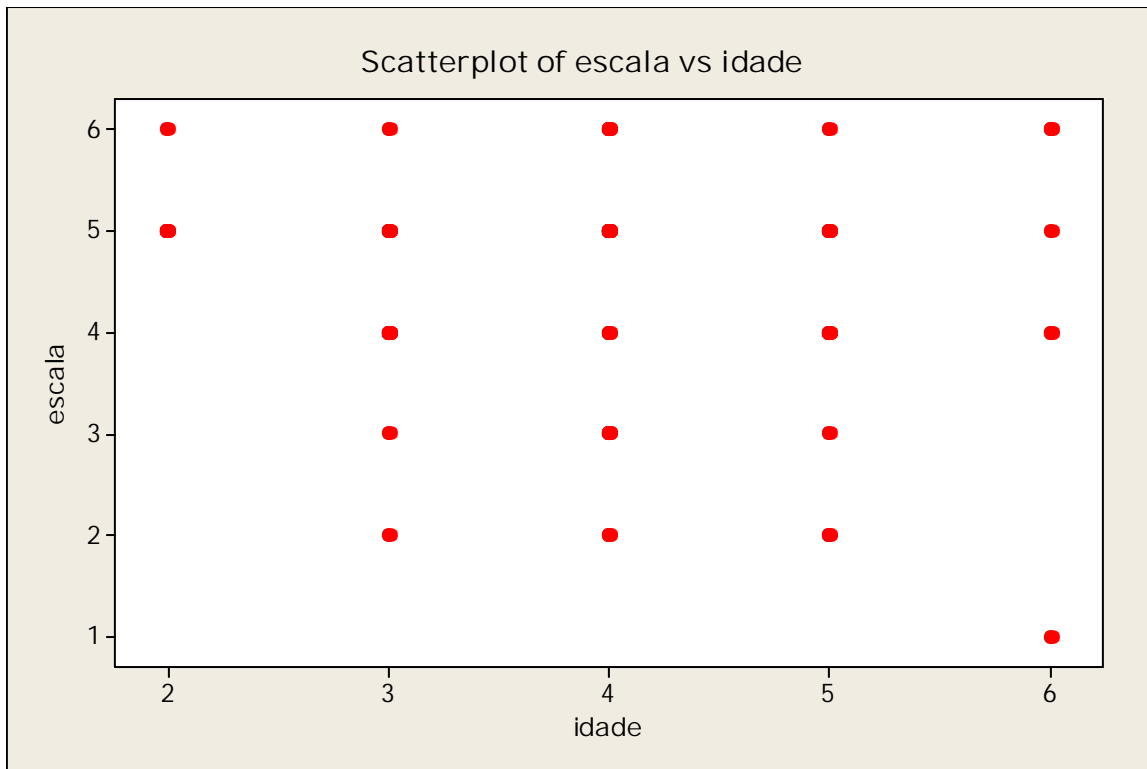
Source	DF	SS	MS	F	P
Regression	1	4547,24	4547,24	49,86	0,000
Error	18	1641,71	91,21		
Total	19	6188,95			

Fitted Line Plot  
peso = 121,4 - 76,09 subs



S	9,55020
R-Sq	73,5%
R-Sq(adj)	72,0%

4 – O que vc afirma sobre a influencia da var. Idade na var. escala ?



O gráfico de dispersão indica não haver relação entre as vars., confirmaremos pela regressão:

The regression equation is  
 $escala = 5,02 - 0,159 idade$

Predictor	Coef	SE Coef	T	P
Constant	5,0174	0,5834	8,60	0,000
idade	-0,1592	0,1411	-1,13	0,264

S = 1,20584 R-Sq = 2,1% R-Sq(adj) = 0,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1,849	1,849	1,27	0,264
Residual Error	58	84,334	1,454		
Total	59	86,183			

5 – Um sociólogo recolheu dados das variáveis deflator, gnp, unempl, armfor, popula e time; ele deseja verificar qual a influência simultânea destas variáveis sobre a var. total e estimar um modelo final somente com as vars. que realmente contribuirão para a compreensão da var. total, e contratou vc para fazer isto para ele. Boa sorte

The regression equation is  
total = - 3449892 - 0,0320 gnp - 1,97 unempl - 1,02 armfor - 0,078 popula  
+ 1814 time

Predictor	Coef	SE Coef	T	P
Constant	-3449892	828242	-4,17	0,002
gnp	-0,03196	0,02420	-1,32	0,216
unempl	-1,9721	0,3861	-5,11	0,000
armfor	-1,0200	0,1908	-5,34	0,000
popula	-0,0775	0,1616	-0,48	0,642
time	1814,1	425,3	4,27	0,002

S = 289,715 R-Sq = 99,5% R-Sq(adj) = 99,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	184169478	36833896	438,84	0,000
Residual Error	10	839348	83935		
Total	15	185008826			

Realizarei agora a análise excluindo a var. Popula.

The regression equation is  
total = - 3598729 - 0,0402 gnp - 2,09 unempl - 1,01 armfor + 1887 time

Predictor	Coef	SE Coef	T	P
Constant	-3598729	740633	-4,86	0,001
gnp	-0,04019	0,01647	-2,44	0,033
unempl	-2,0884	0,2900	-7,20	0,000
armfor	-1,0146	0,1837	-5,52	0,000
time	1887,4	382,8	4,93	0,000

S = 279,396 R-Sq = 99,5% R-Sq(adj) = 99,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	184150146	46037536	589,76	0,000
Residual Error	11	858680	78062		
Total	15	185008826			

O modelo final é o acima exposto, com todas as vars. preditoras  
senso significativas.