Modelos Lineares I

Regressão Linear Múltipla (RLM):

Transformações de linearização

(37ª e 38ª Aulas)

Professor: Dr. José Rodrigo de Moraes

Universidade Federal Fluminense (UFF)

Departamento de Estatística (GET)

1

Modelo de Regressão Linear:

Transformações de dados para resolver o problema de não linearidade:

Quando a análise gráfica indicar que o modelo linear não é apropriado para um conjunto de dados, existem duas opções:

1) *Descartar o modelo de regressão linear e pesquisar um mais apropriado;*

2) *Usar algum tipo de transformação nos dados de forma que o modelo linear seja válido.*

OBS: Se a não linearidade da relação entre X e Y foi identificada, e se a distribuição dos resíduos é aproximadamente normal com variância constante, então uma opção seria transformar a variável X e/ou Y.

2

Transformações de dados para resolver o problema de não linearidade (*continuação*):

- Se um modelo linear não é apropriado para um conjunto de dados, pode-se optar por fazer alguma transformação para resolver o problema de não linearidade.
- Existem funções (modelos não lineares) que através de transformações de variáveis reduzem-se a modelo lineares da forma desejada: $Y=\beta_0+\beta_1X+\varepsilon$.
- O conhecimento da forma de diversas famílias de funções não lineares (mas linearizáveis), possibilita:
  - ✓ *a escolha de um dado modelo mais apropriado*;
  - ✓ *o tipo de transformação a ser adotada em X e/ou Y*;
  - ✓ *a forma linear resultante*.

3

Transformações de dados para resolver o problema de não linearidade – *Modelos Linearizáveis*:

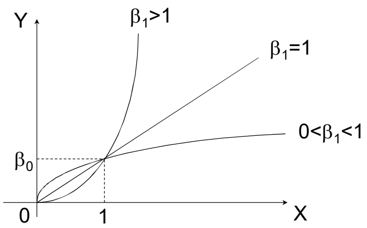
- **1º caso A:**

Função Potência

$Y=\beta_0X^{\beta_1}$

$\beta_0>0$

$\beta_1>0$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis transformadas X*=logX e Y*=logY, obtendo-se assim as estimativas dos parâmetros β_1 e β_0*=logβ_0 e, em seguida, a estimativa de $\beta_0=10^{\beta_0^*}$.

4

Transformações de dados para resolver o problema de não linearidade – *Modelos Linearizáveis*:

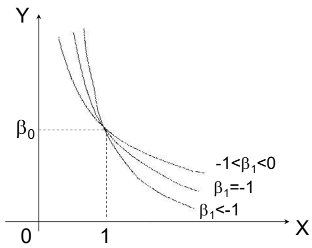
- **1º caso B:**

Função Potência

$Y=\beta_0X^{\beta_1}$

$\beta_0>0$

$\beta_1<0$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis transformadas X*=logX e Y*=logY, obtendo-se assim as estimativas dos parâmetros β_1 e β_0*=logβ_0 e, em seguida, a estimativa de $\beta_0=10^{\beta_0^*}$.

5

Modelo de Regressão Linear:

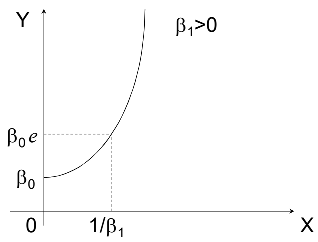
Transformações de dados para resolver o problema de não linearidade (*continuação*) – *Modelos Linearizáveis*:

- **2º caso - A:**

F. exponencial

$Y=\beta_0e^{\beta_1X}$

$\beta_1>0$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis X e Y*=lnY, obtendo-se assim as estimativas dos parâmetros β_1 e β_0^* e, em seguida, a estimativa de $\beta_0=e^{\beta_0^*}$.

6

Prof.: Dr. José Rodrigo de Moraes: Estatístico (ENCE), Mestre em Estatística Social (ENCE) e Doutor em Saúde Coletiva (IESC/UFRJ)

Modelo de Regressão Linear:

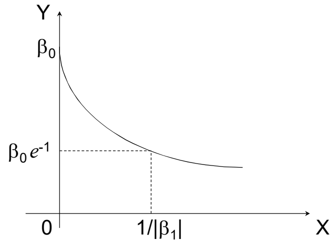
Transformações de dados para resolver o problema de não linearidade (*continuação*) – *Modelos Linearizáveis*:

- **2º caso - B:**

F. exponencial

$$Y = \beta_0 e^{\beta_1 X}$$

$$\beta_1 < 0$$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis X e Y*=lnY, obtendo-se assim as estimativas dos parâmetros β_1 e β_0^* e, em seguida, a estimativa de $\beta_0 = e^{\beta_0^*}$.

7

Modelo de Regressão Linear:

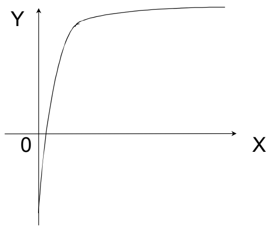
Transformações de dados para resolver o problema de não linearidade (*continuação*) – *Modelos Linearizáveis*:

- **3º caso - A:**

F. logarítmica

$$Y = \beta_0 + \beta_1 \log X$$

$$\beta_1 > 0$$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis X*=logX e Y, obtendo-se diretamente as estimativas dos parâmetros β_1 e β_0.

8

Modelo de Regressão Linear:

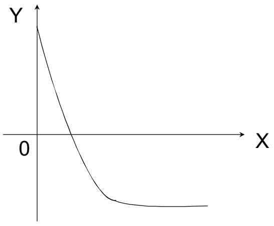
Transformações de dados para resolver o problema de não linearidade (*continuação*) – *Modelos Linearizáveis*:

- **3º caso - B:**

F. logarítmica

$$Y = \beta_0 + \beta_1 \log X$$

$$\beta_1 < 0$$



OBS: Neste caso, a teoria de regressão linear é aplicada as variáveis X*=logX e Y, obtendo-se diretamente as estimativas dos parâmetros β_1 e β_0.

9

Transformações de dados para resolver o problema de não linearidade (*Resumo*) – *Modelos linearizáveis*:

Caso	Modelo original	Transformação	Modelo transformado
1AeB	$Y = \beta_0 X^{\beta_1} \varepsilon$	Y*=logY, X*=logX	$Y^* = \log\beta_0 + \beta_1 X^* + \log\varepsilon$
2AeB	$Y = \beta_0 e^{\beta_1 X} \varepsilon$	Y*=lnY	$Y^* = \ln\beta_0 + \beta_1 X + \ln\varepsilon$
3AeB	$Y = \beta_0 + \beta_1 \log X + \varepsilon$	X*=logX	$Y = \beta_0 + \beta_1 X^* + \varepsilon$
4	$Y = \dfrac{1}{\beta_0 + \beta_1 X + \varepsilon}$	Y*=1/Y	$Y^* = \beta_0 + \beta_1 X + \varepsilon$
5	$Y = \beta_0 + \dfrac{\beta_1}{X} + \varepsilon$	X*=1/X	$Y = \beta_0 + \beta_1 X^* + \varepsilon$

10

Exemplo 1: ***Transformação - Problema de não linearidade***

A tabela 1 a seguir fornece informações sobre o peso do corpo (em kg) e do cérebro (em gramas) de 11 animais. Faça o gráfico de dispersão entre essas variáveis, e escolha e ajuste um modelo apropriado para explicar o peso do cérebro dos animais, a partir do peso de seus corpos.

11

Exemplo 1: ***Transformação - Problema de não linearidade***

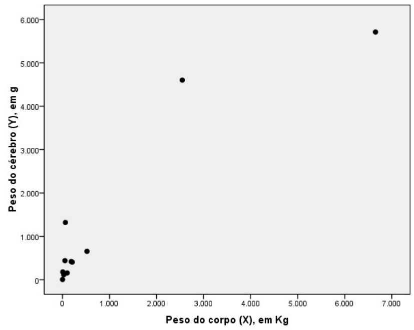
Tabela 1: Peso do corpo e do cérebro de alguns animais.

Animal	Peso do corpo (X)	Peso do cérebro (Y)
1	6.654,00	5.712,00
2	2.547,00	4603,00
3	521,00	655,00
4	207,00	406,00
5	187,00	419,00
6	100,00	157,00
7	62,00	1.320
8	52,16	440,00
9	27,66	115,00
10	6,80	179,00
11	1,04	5,50

12

Exemplo 1: ***Transformação - Problema de não linearidade***

Figura 1: Gráfico de dispersão entre o peso do corpo (X) e o peso do cérebro (Y).



13

Exemplo 1: ***Transformação - Problema de não linearidade***

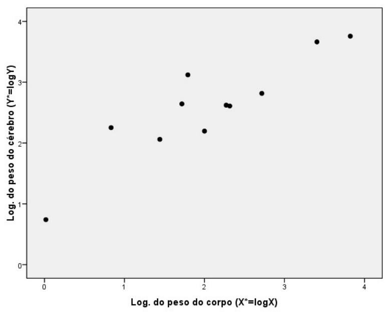
Tabela 2: Log. do peso do corpo e do cérebro de alguns animais

Animal	Log. peso do corpo (X*=logX)	Log. peso do cérebro (Y*=logY)
1	3,823	3,757
2	3,406	3,663
3	2,717	2,816
4	2,316	2,609
5	2,272	2,622
6	2,000	2,196
7	1,792	3,121
8	1,717	2,643
9	1,442	2,061
10	0,833	2,253
11	0,017	0,740

14

Exemplo 1: ***Transformação - Problema de não linearidade***

Figura 2: Gráfico de dispersão entre o log. do peso do corpo (X*=logX) e o log. do peso do cérebro (Y*=logY).



15

Exemplo 1: ***Transformação – Problema de não linearidade***

Resultados do ajuste do modelo transformado: ***log. do peso do corpo*** **(*X*=logX*)** ***e log. do peso do cérebro*** **(*Y*=logY*).**

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,900[a]	,810	,789	,378967

a. Predictors: (Constant), logX

b. Dependent Variable: logY

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5,527	1	5,527	38,488	,000[a]
	Residual	1,293	9	,144		
	Total	6,820	10			

a. Predictors: (Constant), logX

b. Dependent Variable: logY

16

Exemplo 1: ***Transformação – Problema de não linearidade***

Resultados do ajuste do modelo transformado: ***log. do peso do corpo*** **(*X*=logX*)** ***e log. do peso do cérebro*** **(*Y*=logY*).**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1,194	,252		4,733	,001	,623	1,765
	logX	,687	,111	,900	6,204	,000	,437	,938

a. Dependent Variable: logY

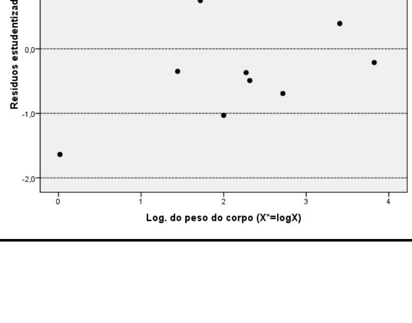
Modelo transformado :

Modelo original :

17

Exemplo 1: ***Transformação – Problema de não linearidade***

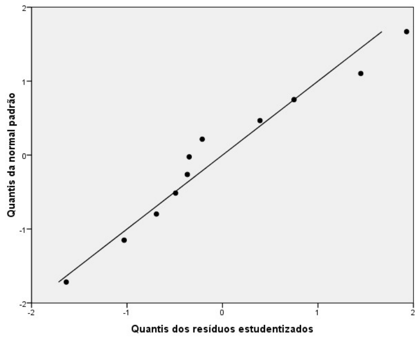
Figura 3: Gráfico de dispersão entre o log. do peso do corpo (X*=logX) ***e*** **os resíduos padronizados do modelo.**



18

Exemplo 1: ***Transformação – Problema de não linearidade***

Figura 4: QQ-Plot (*normalidade*) para os resíduos padronizados do modelo.



19

Exemplo 2: ***Transformação - Problema de não linearidade***

Uma população de bactérias é exposta a raios ultra-violeta durante 15 períodos de 6 minutos cada. Para cada período o número de bactérias sobreviventes é registrado. A teoria da sobrevivência, em geral, modela tal fenômeno pela função dada por:

$Y_t = \beta_0 e^{\beta_1 \cdot t}$, onde:

β_0 e β_1 são parâmetros desconhecidos. O parâmetro β_0 denota a população no início do experimento e β_1 é a taxa de mortalidade no instante t.

A tabela a seguir fornece os resultados do experimento com uma população inicial de 355 bactérias.

20

Exemplo 2: ***Transformação - Problema de não linearidade***

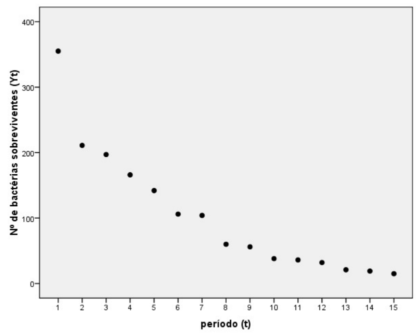
Tabela 1: Nº de bactérias sobreviventes segundo o período.

t	Y_t
1	355
2	211
3	197
4	166
5	142
6	106
7	104
8	60
9	56
10	38
11	36
12	32
13	21
14	19
15	15

21

Exemplo 2: ***Transformação - Problema de não linearidade***

Figura 1: Gráfico de dispersão entre o período e o número de bactérias sobreviventes.



22

Exemplo 2: ***Transformação - Problema de não linearidade***

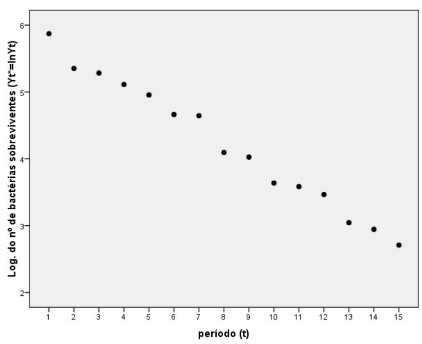
Tab.2: Log. do nº de bactérias sobreviventes segundo o período

t	Y_t	$Y_t^* = \ln Y_t$
1	355	5,872118
2	211	5,351858
3	197	5,283204
4	166	5,111988
5	142	4,955827
6	106	4,663439
7	104	4,644391
8	60	4,094345
9	56	4,025352
10	38	3,637586
11	36	3,583519
12	32	3,465736
13	21	3,044522
14	19	2,944439
15	15	2,708050

23

Exemplo 2: ***Transformação - Problema de não linearidade***

Figura 2: Gráfico de dispersão entre o período e o log. do nº bactérias sobreviventes (Y_t^*=lnY_t).



24

Exemplo 2: ***Transformação – Problema de não linearidade***

Resultados do ajuste do modelo transformado:

Log. do nº bactérias sobreviventes **($Y_t^*= lnY_t$)** ***e o período*** **(t).**

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,994[a]	,988	,987	,110016010

a. Predictors: (Constant), t

b. Dependent Variable: lnYt

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13,359	1	13,359	1103,702	,000[a]
	Residual	,157	13	,012		
	Total	13,516	14			

a. Predictors: (Constant), t

b. Dependent Variable: lnYt

Exemplo 2: ***Transformação – Problema de não linearidade***

Resultados do ajuste do modelo transformado:

Log. do nº bactérias sobreviventes **($Y_t^*=lnY_t$)** ***e o período*** **(t).**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	5,973	,060		99,922	,000	5,844	6,102
	t	-,218	,007	-,994	-33,222	,000	-,233	-,204

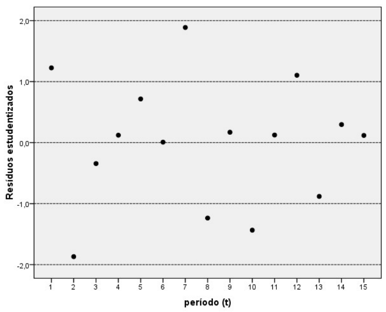
a. Dependent Variable: lnYt

Modelo transformado :

Modelo orginal :

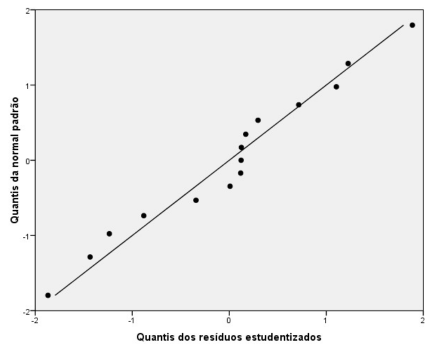
Exemplo 2: ***Transformação – Problema de não linearidade***

Figura 3: Gráfico de dispersão entre o período (t) e os resíduos padronizados do modelo.



Exemplo 2: ***Transformação – Problema de não linearidade***

Figura 4: QQ-Plot (*normalidade*) para os resíduos padronizados do modelo.



Exemplo 3: ***Outra transformação - Problema de não linearidade***

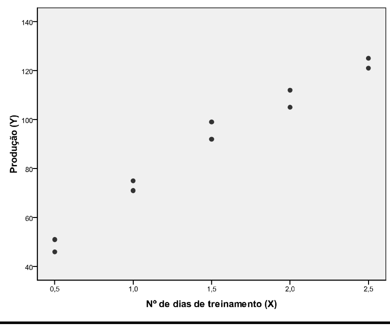
A tabela 1 abaixo fornece o número de dias de treinamento e a produção (em vendas) dos novos vendedores.

Tabela 1: Dados originais

Vendedor	Nº de dias (X)	Produção (Y)
1	0,5	46
2	0,5	51
3	1,0	71
4	1,0	75
5	1,5	92
6	1,5	99
7	2,0	105
8	2,0	112
9	2,5	121
10	2,5	125

Exemplo 3: ***Outra transformação – Problema de não linearidade***

Figura 1: Gráfico de dispersão entre o nº de dias de treinamento (X) e a produção (Y).



Prof.: Dr. José Rodrigo de Moraes: Estatístico (ENCE), Mestre em Estatística Social (ENCE) e Doutor em Saúde Coletiva (IESC/UFRJ)

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 1:

Nº de dias de treinamento (X) e produção (Y).

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,985[a]	,970	,966	5,173

a. Predictors: (Constant), nºdias

b. Dependent Variable: prod

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6808,050	1	6808,050	254,447	,000[a]
	Residual	214,050	8	26,756		
	Total	7022,100	9			

a. Predictors: (Constant), nºdias

b. Dependent Variable: prod

31

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 1:

Nº de dias de treinamento (X) e produção (Y).

Coefficients[a]

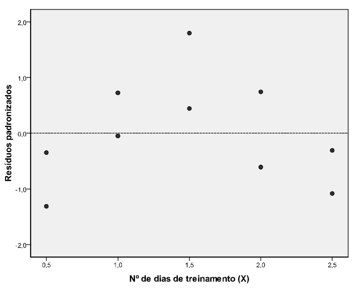
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	34,350	3,836		8,954	,000	25,504	43,196
	nºdias	36,900	2,313	,985	15,951	,000	31,566	42,234

a. Dependent Variable: prod

32

Exemplo 3: Outra transformação - *Problema de não linearidade*

Figura 2: Gráfico de dispersão entre o nº de dias de treinamento (X) e os resíduos padronizados do modelo.



33

Exemplo 3: Outra transformação – *Problema de não linearidade*

Transformação: $X^*=\sqrt{X}$

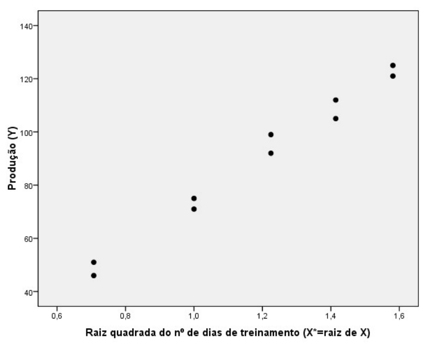
Tabela 1: Dados transformados

Vendedor	Nº de dias (X)	Raiz quadrada do nº de dias (X*)	Produção (Y)
1	0,5	0,7071	46
2	0,5	0,7071	51
3	1,0	1,0000	71
4	1,0	1,0000	75
5	1,5	1,2247	92
6	1,5	1,2247	99
7	2,0	1,4142	105
8	2,0	1,4142	112
9	2,5	1,5811	121
10	2,5	1,5811	125

34

Exemplo 3: Outra transformação – *Problema de não linearidade*

Figura 3: Gráfico de dispersão entre a raiz quadrada do nº de dias de treinamento ($X^*=\sqrt{X}$) e a produção (Y).



35

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 2: *Raiz quadrada do nº de dias de treinamento ($X^*=\sqrt{X}$) e a produção (Y).*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,993[a]	,987	,985	3,405

a. Predictors: (Constant), raiz_nºdias

b. Dependent Variable: prod

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6929,366	1	6929,366	597,785	,000[a]
	Residual	92,734	8	11,592		
	Total	7022,100	9			

a. Predictors: (Constant), raiz_nºdias

b. Dependent Variable: prod

36

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 2: ***Raiz quadrada do nº de dias de treinamento (X*=√X) e a produção (Y).***

Coefficients[a]

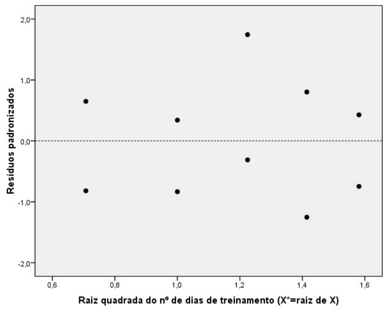
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-11,687	4,284		-2,728	,026	-21,567	-1,808
	raiz_nºdias	85,527	3,498	,993	24,450	,000	77,460	93,594

a. Dependent Variable: prod

37

Exemplo 3: Outra transformação – *Problema de não linearidade*

Figura 4: Gráfico de dispersão entre a raiz quadrada do nº de dias de treinamento (X*=√X) e os resíduos padronizados do modelo.



38

Exemplo 3: Outra transformação - *Problema de não linearidade*

Conclusões gerais:

Os gráficos construídos demonstram a existência de uma relação linear entre Y e X*=√X e que os resíduos (padronizados) se encontram aleatoriamente distribuídos em torno de zero. Além disso, a medida de qualidade do ajuste (R^2=98,7%) e os testes de significância dos parâmetros do modelo (testes t e F) indicam a adequação do modelo ajustado para os dados transformados, dado por:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \sqrt{X_i} \rightarrow \quad Y_i = -11{,}687 + 85{,}527\sqrt{X_i}$$

39

Exemplo 3: Outra transformação – *Problema de não linearidade*

Alternativa: Regressão polinomial (2º grau)

Tabela 1: Calculando a variável "nº de dias ao quadrado (X^2)".

Vendedor	Nº de dias (X)	Nº de dias ao quadrado (X^2)	Produção (Y)
1	0,5	0,25	46
2	0,5	0,25	51
3	1,0	1,00	71
4	1,0	1,00	75
5	1,5	2,25	92
6	1,5	2,25	99
7	2,0	4,00	105
8	2,0	4,00	112
9	2,5	6,25	121
10	2,5	6,25	125

40

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 3:
Regressão polinomial

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,994[a]	,987	,984	3,580

a. Predictors: (Constant), nºdias2, nºdias
b. Dependent Variable: prod

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6932,371	2	3466,186	270,408	,000[a]
	Residual	89,729	7	12,818		
	Total	7022,100	9			

a. Predictors: (Constant), nºdias2, nºdias
b. Dependent Variable: prod

41

Exemplo 3: Outra transformação - *Problema de não linearidade*

Ajuste do modelo 3:
Regressão polinomial

Coefficients[a]

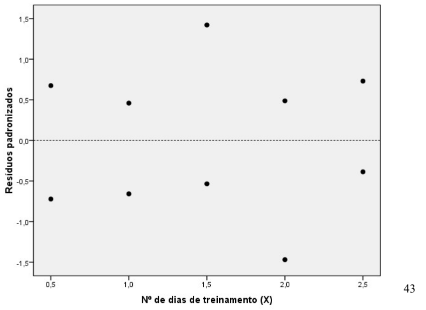
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	19,600	5,430		3,610	,009	6,761	32,439
	nºdias	62,186	8,276	1,659	7,514	,000	42,617	81,755
	nºdias2	-8,429	2,706	-,688	-3,114	,017	-14,828	-2,029

a. Dependent Variable: prod

42

Exemplo 3: Outra transformação – *Problema de não linearidade*

Figura 4: Gráfico de dispersão entre o nº de dias de treinamento (X) e os resíduos padronizados do modelo polinomial.



43

Aula prática – Exercício 1 com "*Saídas*":

Os dados apresentados na tabela a seguir se referem ao desempenho dos alunos numa prova da disciplina de mestrado profissional.

O objetivo do estudo é estudar a relação entre as horas de estudo e o percentual (%) de questões erradas (*proxy* do desempenho) para 12 alunos matriculados na disciplina.

44

Aula prática – Exercício 1 (*continuação*):

Tabela 1: *Dados sobre n=12 alunos*

Aluno	% Questões erradas	Horas de estudo
1	20	1,81
2	25	1,70
3	30	1,65
4	35	1,55
5	40	1,48
6	50	1,40
7	55	1,38
8	60	1,30
9	65	1,26
10	70	1,24
11	75	1,21
12	80	1,20

45

Aula prática - Exercício 1 (*continuação*):

a) Construa o gráfico de dispersão entre as horas de estudo e o indicador de desempenho na disciplina.

b) Ajuste um modelo estatístico, da forma $Y=\beta_0+\beta_1X+\varepsilon$, para avaliar o efeito das horas de estudo no desempenho dos alunos na disciplina. Analise o gráfico de dispersão entre os valores ajustados e os resíduos estudentizados. Avalie, usando o QQ -Plot, a hipótese de normalidade dos resíduos.

c) Proponha um novo modelo estatístico para avaliar o efeito das horas de estudo no desempenho dos alunos.

46

- **Aula prática - Exercício 1 (*continuação*):**

d) Avalie a significância individual dos parâmetros do novo modelo proposto por você considerando α=5%.

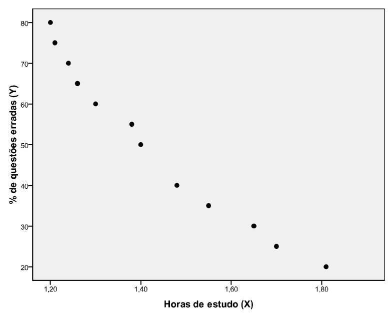
e) Avalie a necessidade de se considerar o termo quadrático no modelo utilizando agora o *Teste F de comparabilidade de modelos*. Qual a sua conclusão?

f) Refaça os gráficos da letra b). Na sua opinião o novo modelo é adequado?

47

Aula prática – Exercício 1 – letra a):

Figura 1: Gráfico de dispersão entre as horas de estudo (X) e o percentual de questões erradas (Y)



48

Aula prática – Exercício 1 – Letra b): *"Algumas saídas"*

Ajuste do modelo 1:

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4346,590	1	4346,590	246,508	,000[a]
	Residual	176,327	10	17,633		
	Total	4522,917	11			

a. Predictors: (Constant), Horas_estudo_X

b. Dependent Variable: Quest_erradas_Y

Coefficients[a]

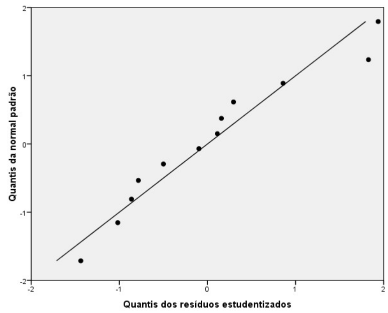
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	188,218	8,860		21,243	,000
	Horas_estudo_X	-96,252	6,130	-,980	-15,701	,000

a. Dependent Variable: Quest_erradas_Y

49

Aula prática – Exercício 1 – letra b):

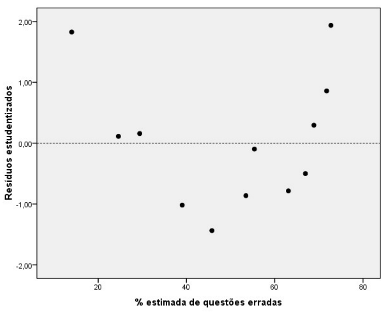
Figura 2: QQ Plot (*normalidade*) dos resíduos estudentizados do modelo



50

Aula prática – Exercício 1 – letra b):

Figura 3: Gráfico de dispersão entre % estimado de questões erradas e os resíduos estudentizados do modelo.



51

Aula prática – Exercício 1 – Letra d): *"Algumas saídas"*

Ajuste do modelo 2:

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4483,950	2	2241,975	517,816	,000[a]
	Residual	38,967	9	4,330		
	Total	4522,917	11			

a. Predictors: (Constant), X2, Horas_estudo_X

b. Dependent Variable: Quest_erradas_Y

Coefficients[a]

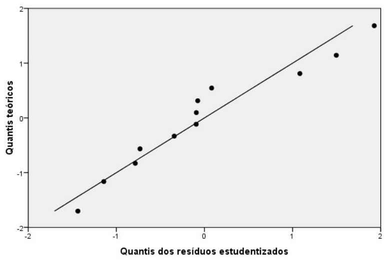
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	411,949	39,963		10,308	,000
	Horas_estudo_X	-403,901	54,705	-4,114	-7,383	,000
	X2	103,755	18,421	3,138	5,633	,000

a. Dependent Variable: Quest_erradas_Y

52

Aula prática – Exercício 1 – letra f):

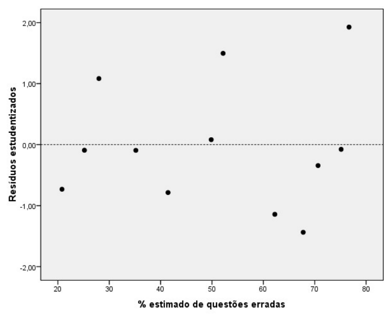
Figura 4: QQ Plot (*normalidade*) dos resíduos estudentizados do modelo



53

Aula prática – Exercício 1 – letra f):

Figura 5: Gráfico de dispersão entre % estimado de questões erradas e os resíduos estudentizados do modelo.



54

Aula prática - Exercício 2:

A associação industrial de um determinada cidade tem como objetivo verificar se existe relação entre o "número de trabalhadores (X)" e o número de supervisores dos estabelecimentos associados (Y)".

A Tabela 1 fornece essas informações para uma amostra de 27 estabelecimentos.

55

Aula prática - Exercício 2 (*continuação*):

Tabela 2: Dados sobre n=20 estabelecimentos

Estab.	Nº de trab. (X)	Nº de superv. (Y)
1	294	30
2	247	32
3	267	37
4	358	44
5	423	47
6	311	49
7	450	56
8	534	62
9	438	68
10	697	78
11	688	80
12	630	84
13	709	88
14	627	97

continuação

Estab.	Nº de trab. (X)	Nº de superv. (Y)
15	615	100
16	999	109
17	1022	114
18	1015	117
19	700	106
20	850	128
21	980	130
22	1025	160
23	1021	97
24	1200	180
25	1250	112
26	1500	210
27	1650	135

56

- **Aula prática – Exercício 2 (*continuação*):**

a) Ajuste o modelo (usando Y e lnY) e avalie a existência ou não de alguma violação das hipóteses básicas do modelo. Para tanto use análises gráficas (*utilize os resíduos estudentizados*). Qual a conclusão obtida?

b) Caso necessário, corrija a violação identificada na letra a) e analise as estimativas dos parâmetros do novo modelo. Calcule alguma medida de qualidade do ajuste.

c) Avalie as hipótese básicas do modelo.

57