Objetivos da aula

- **Essa aula objetiva fornecer algumas ferramentas descritivas úteis para escolha de uma forma funcional adequada.**
- **Por exemplo, qual seria a forma funcional adequada para estudar a relação entre salario-hora e tempo de escolaridade?**
- **Além disso, após a adoção da forma funcional, esse grupo de *slides* objetiva auxiliar o leitor na interpretação correta das estimativas dos parâmetros do modelo proposto.**
- **Aqui, cabe ressaltar que não existem regras rígidas e rápidas que funcionem para todas as situações.**

Análise de Regressão Linear Simples II

Aula 02

Gujarati e Porter – Seções 6.4 a 6.8

Wooldridge, 2011 – Seção 2.4

Introdução

Como dito na primeira aula da disciplina, as relações lineares não são, em geral, suficientes para todas as aplicações em Administração e Economia. Assim sendo, nessa aula estudaremos maneiras de incorporar algumas não-linearidades na análise de regressão simples. Para tanto, será muito importante sabermos definir apropriadamente as variáveis dependente e independente. Ainda, após tais transformações buscaremos fazer as interpretações dos parâmetros em termos do problema de interesse.

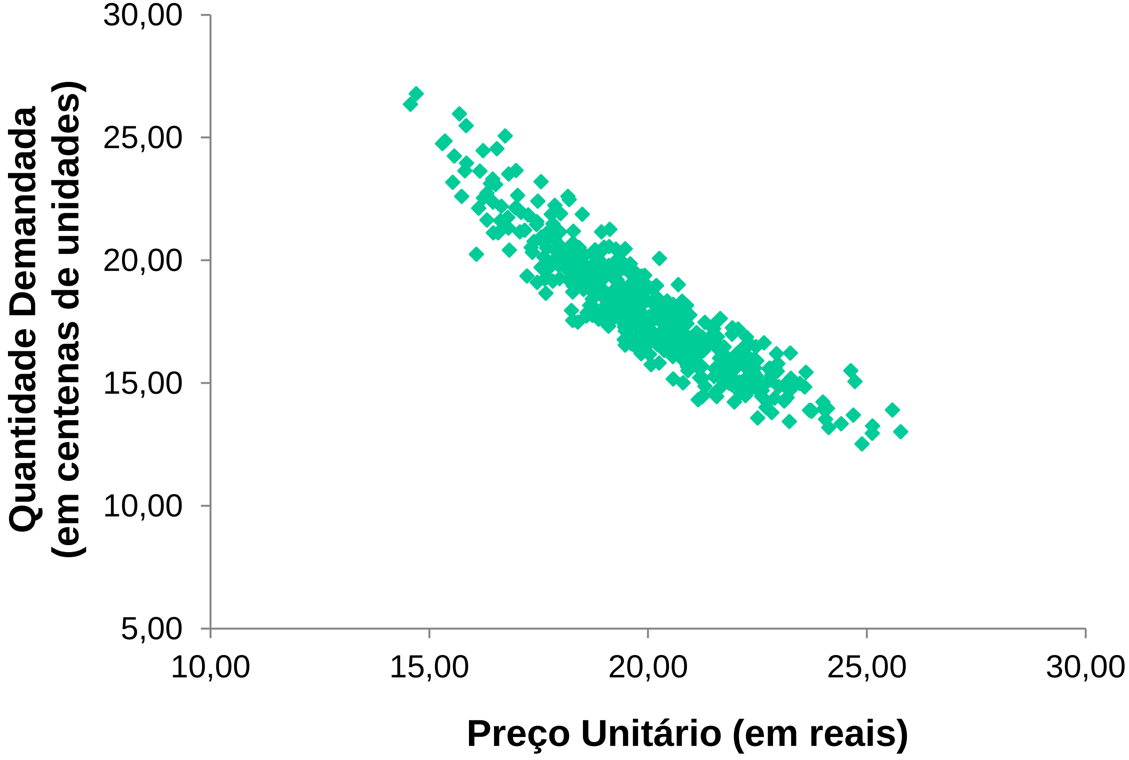
Exemplo 1

O senhor Nusdeo tem interesse em modelar a relação entre a quantidade demandada de um bem (variável resposta, y) e o seu preço unitário (variável explicativa, x).

Ainda, em posse de tal modelo, e com base numa amostra coletada ao acaso, o senhor Nusdeo gostaria de encontrar uma estimativa para a elasticidade preço da demanda por tal bem.

Exemplo 1

Via construção de um diagrama de dispersão, o senhor Nusdeo observa o seguinte comportamento entre as variáveis de interesse:



Modelo Log-Log

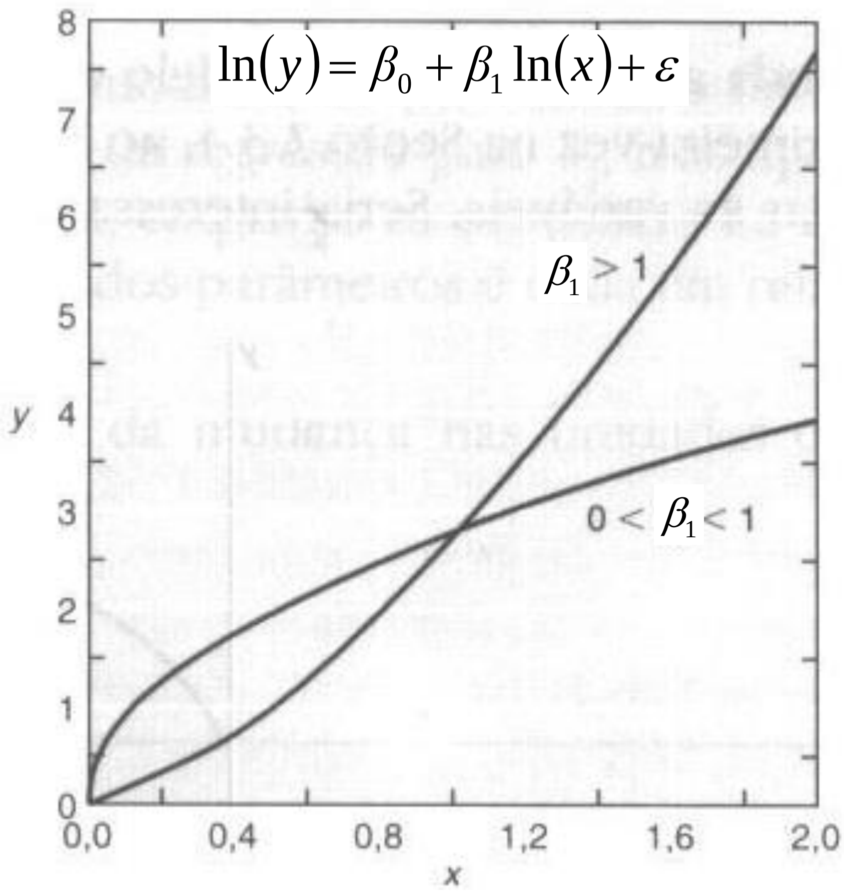
O modelo de regressão

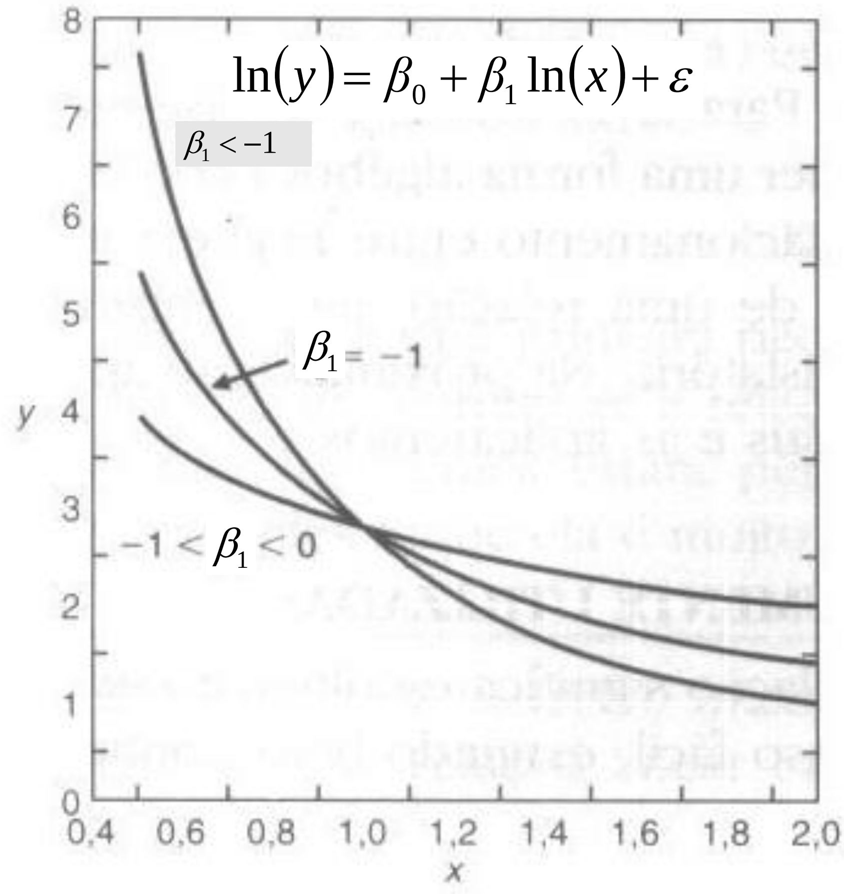
$$\ln(y) = \beta_0 + \beta_1 \ln(x) + \varepsilon$$

é conhecido na prática como modelo log-log.

Observação

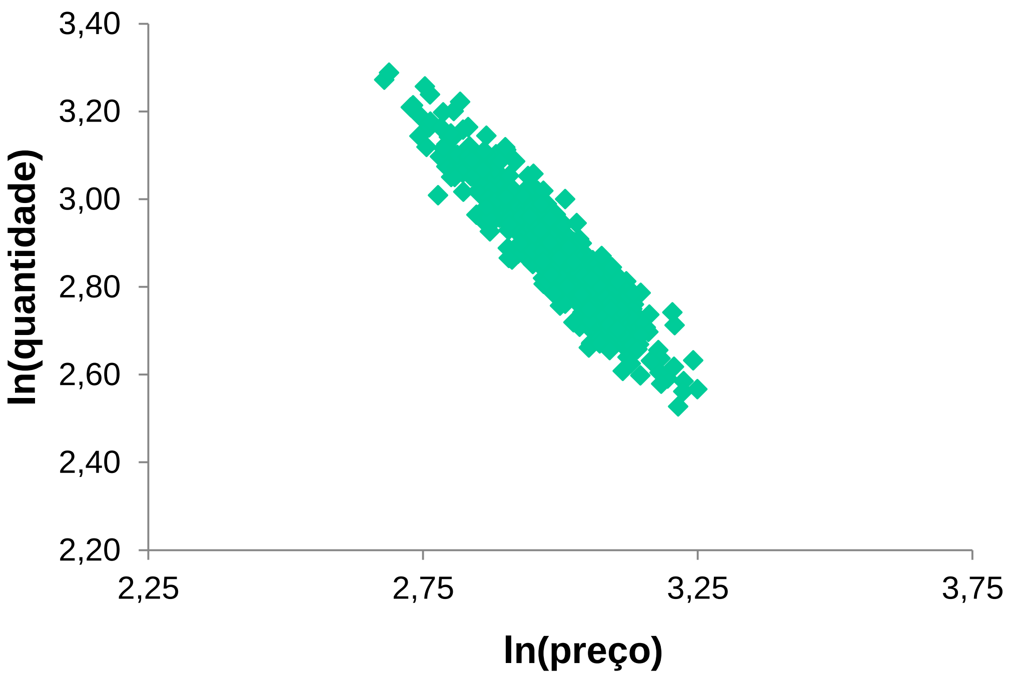
Nas figuras a seguir são apresentados alguns possíveis comportamentos entre as variáveis y e x, dado por um modelo conhecido como log-log





Voltando ao Exemplo 1

Após aplicar a transformação ln(·) em ambas as variáveis de interesse, o senhor Nusdeo observou, via diagrama de dispersão, a seguinte relação entre as variáveis de interesse:



Voltando ao Exemplo 1

Com base na amostra coletada, o senhor Nusdeo obteve os seguintes resultados:

RESUMO DOS RESULTADOS

Estatística de regressão	
R-Quadrado	0.8489
Erro padrão	0.0524
Observações	500

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	7.674	7.6741	2797.40	1.74E-206
Resíduo	498	1.366	0.0027		
Total	499	9.040			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	6.61	0.07	93.72	2.00E-300	6.47	6.75
ln(preço)	-1.25	0.02	-52.89	1.74E-206	-1.29	-1.20

Voltando ao Exemplo 1

Com base no quadro anterior, a equação estimada fica dada por

$$\widehat{\ln\left(quantidade\right)} = 6{,}61 - 1{,}25\ln(preço)$$

Finalmente, qual interpretação deve ser dada para a estimativa do parâmetro associado ao regressor, em termos do problema?

Alguns Resultados Úteis

a) Variação absoluta: $\Delta x = x_1 - x_0$

b) Variação relativa: $\dfrac{\Delta x}{x_0}$

c) Variação relativa percentual: $\%\Delta x = 100\left(\dfrac{\Delta x}{x_0}\right)$

d) Para pequenas variações em x, $100\Delta \ln(x) \approx \%\Delta x$

e) Para $x \approx 0$, $\ln(1+x) \approx x$

Modelo Log-Log

Usando os resultados do *slide* anterior, não é difícil ver que, para o modelo

$$\ln(y) = \beta_0 + \beta_1 \ln(x) + \varepsilon$$

a interpretação associada ao parâmetro β_1 pode ser expressa como

$$\frac{\%\Delta E(y \mid x)}{\%\Delta x} = \beta_1 \quad \textbf{(elasticidade)}$$

Voltando ao Exemplo 1

Voltando à equação estimada

$$\widehat{\ln(quantidade)} = 6{,}61 - 1{,}25\ln(preço)$$

não é difícil observar, após utilizar os resultados do slide anterior, que a estimativa associada ao parâmetro β_1 reflete a elasticidade preço da demanda.

Elasticidade

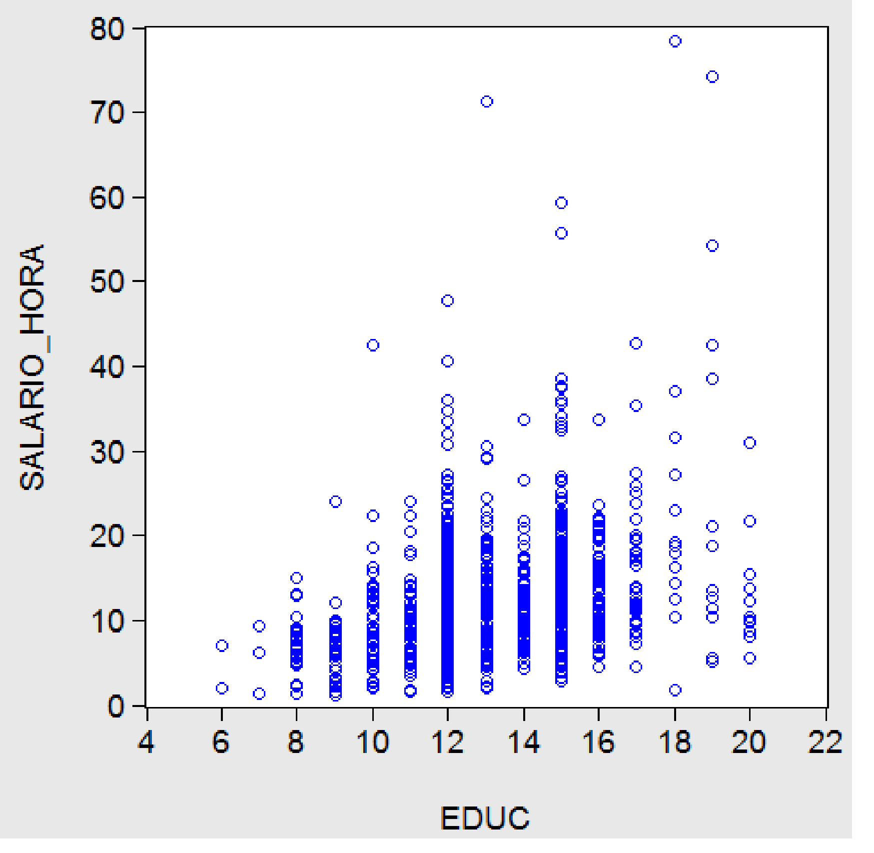
O coeficiente de elasticidade, η, da variável *y* em relação à variável *x*, que mede a variação percentual de *y* correspondente a uma dada variação percentual em *x*, pode ser calculado pela seguinte expressão:

$$\eta = \frac{dy}{dx} \cdot \frac{x}{y}$$

<u>Exercício</u>: Prove que o parâmetro β_1, no modelo log-log, é a elasticidade da variável *y* em relação à variável *x*.

Exemplo 2

Na primeira aula da disciplina tratamos de um exemplo ligado à modelagem do efeito do tempo de escolaridade no salário-hora dos norte-americanos. Para tanto, uma amostra aleatória de 1217 indivíduos foi investigada e o diagrama de dispersão encontra-se ao lado:



Exemplo 2

Após uma análise descritiva bastante detalhada, chegamos à seguinte equação estimada:

$$\widehat{\ln(salario)} = 1{,}178 + 0{,}091 educ$$

Pergunta: é estimado que cada ano a mais de escolaridade aumente quanto, em média, o salário-hora dos norte-americanos?

Modelo Log-Nível

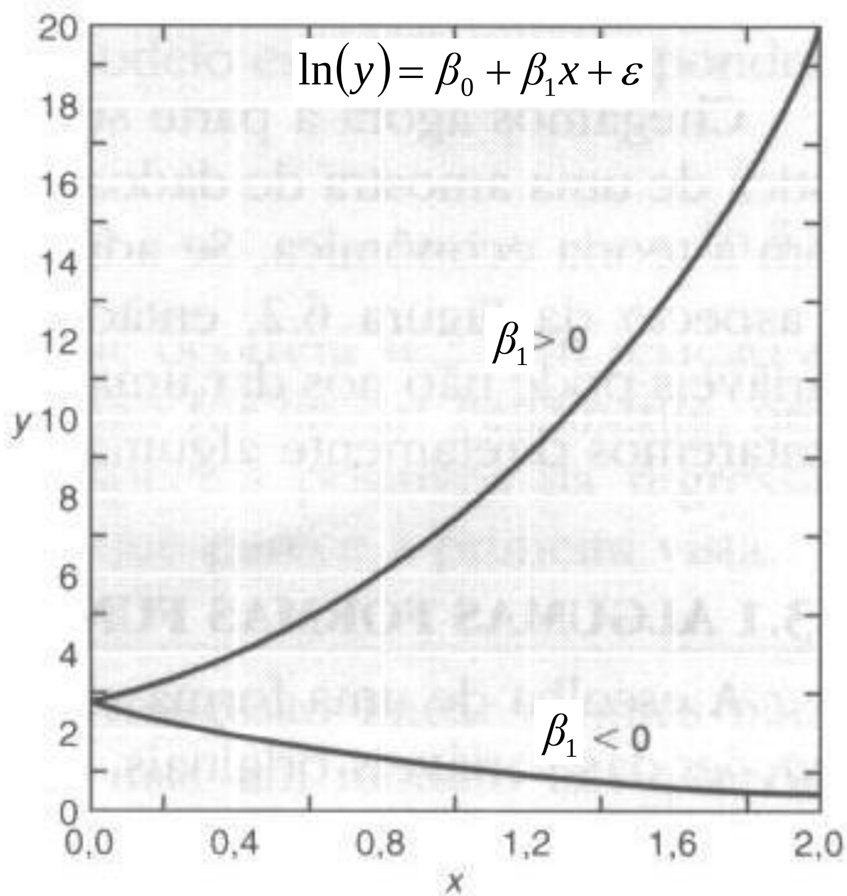
O modelo de regressão

$$\ln(y) = \beta_0 + \beta_1 x + \varepsilon$$

é conhecido na prática como modelo log-nível.

Observação

Na figura a seguir são apresentados dois possíveis comportamentos entre as variáveis y e x, dada pelo modelo log-nível, dependendo do sinal do parâmetro β_1.



Voltando aos Resultados Úteis

a) Variação absoluta: $\Delta x = x_1 - x_0$

b) Variação relativa: $\dfrac{\Delta x}{x_0}$

c) Variação relativa percentual: $\%\Delta x = 100\left(\dfrac{\Delta x}{x_0}\right)$

d) Para pequenas variações em x, $100\Delta \ln(x) \approx \%\Delta x$

e) Para $x \approx 0$, $\ln(1+x) \approx x$

Modelo Log-Nível

Usando os resultados do *slide* anterior, não é difícil ver que, para o modelo

$$\ln(y) = \beta_0 + \beta_1 x + \varepsilon$$

a interpretação associada ao parâmetro β_1 pode ser expressa como

$$\frac{\%\Delta E(y \mid x)}{\Delta x} = 100 \cdot \beta_1 \quad \textbf{(semi-elasticidade)}$$

Observação

Caso β seja um parâmetro associado a uma variável explicativa numérica no nível, num modelo em que a variável resposta se encontra no $\ln(\cdot)$, é interessante se fazer uma correção para se ter uma interpretação mais adequada dos resultados, caso as mudanças em ln(y) sejam grandes demais, já que a aproximação

$$\%\Delta y \approx 100 \Delta \ln(y)$$

pode se mostrar muito imprecisa.

Observação (cont.)

O uso de algumas propriedades algébricas das funções exponenciais e logarítmicas produzem resultados mais adequados para a porcentagem de mudança no valor de y, que é dado por

$$100\left[\exp\left(\beta\Delta x\right)-1\right].$$

(lembre-se que o mais usual é adotar Δx = 1)

Nota: Tal ajuste não é tão crucial para pequenas mudanças percentuais.

Voltando ao Exemplo 2

Após uma análise descritiva bastante detalhada, chegamos ao seguinte modelo estimado:

$$\widehat{\ln(salario)} = 1{,}178 + 0{,}091\,educ$$

Pergunta: é estimado que cada ano a mais de escolaridade aumente quanto, em média, o salário-hora dos norte-americanos?

Exemplo 3

Rose Jolie, médica de um grande hospital universitário tem interesse em estimar o valor médio da variação absoluta no peso dos recém-nascidos dada uma variação percentual na renda mensal familiar. Para responder a esse objetivo, a médica construiu uma base de dados (*bwght.wf1*) contendo informações como, por exemplo,

***pesonasc* – peso do recém-nascido, em gramas;**

***cigs* – número de cigarros fumados, ao dia, pela mãe, durante a gravidez;**

***rendafam* – renda familiar mensal, em milhares de reais;**

para uma amostra aleatória de 1388 famílias.

Exemplo 3 (cont.)

Dentre os modelos,

$$pesonasc = \beta_0 + \beta_1 rendafam + \varepsilon$$

$$\ln(pesonasc) = \beta_0 + \beta_1 \ln(rendafam) + \varepsilon$$

$$\ln(pesonasc) = \beta_0 + \beta_1 rendafam + \varepsilon$$

$$pesonasc = \beta_0 + \beta_1 \ln(rendafam) + \varepsilon$$

qual seria o mais indicado para fornecer diretamente a estimativa de interesse para a médica? Ainda, baseando-se no modelo escolhido, quanto vale a estimativa de interesse?

Modelo Nível-Log

O modelo nível-log, assim como o modelo log-nível, também é conhecido na literatura como modelo semilogarítmico.

Ainda, usando os resultados apresentados no *slide* 19, não é difícil ver que, para o modelo

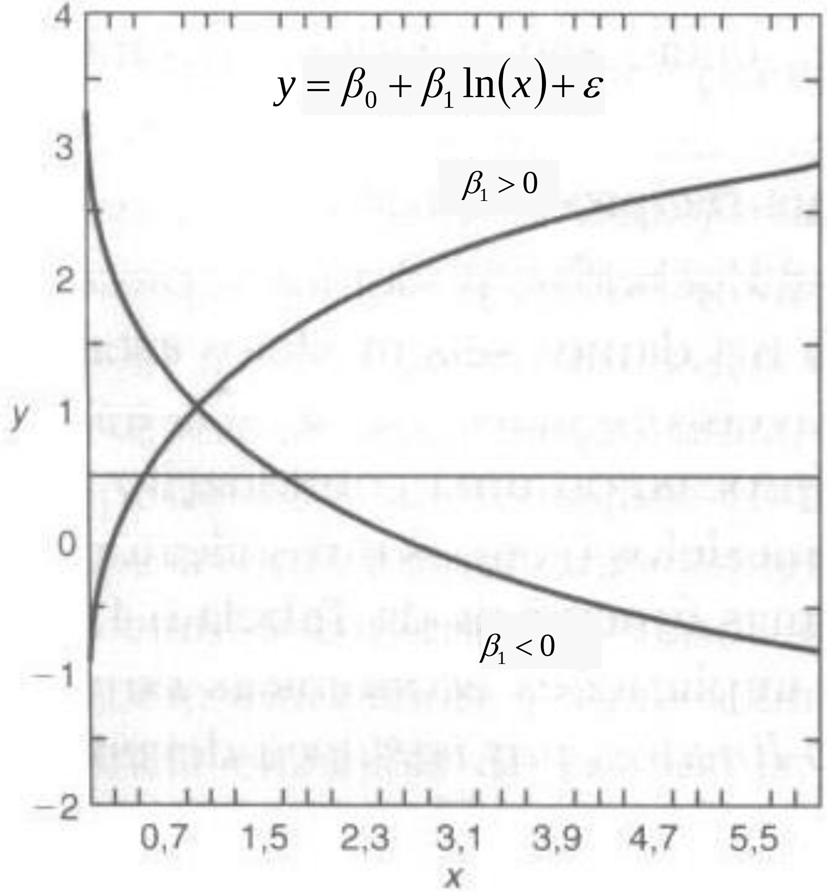
$$y = \beta_0 + \beta_1 \ln(x) + \varepsilon$$

a interpretação associada ao parâmetro β_1 pode ser expressa como

$$\frac{\Delta E(y \mid x)}{\% \Delta x} = \frac{\beta_1}{100}$$

Observação

Na figura a seguir são apresentados dois possíveis comportamentos entre as variáveis y e x, dada pelo modelo nível-log, dependendo do sinal do parâmetro β_1.



Voltando ao Exemplo 3

Adotando o modelo

$$pesonasc = \beta_0 + \beta_1 \ln(rendafam) + \varepsilon$$

quanto vale a estimativa de interesse da médica?

Dependent Variable: PESONASC
Method: Least Squares
Date: 08/17/13 Time: 13:04
Sample: 1 1388
Included observations: 1388
PESONASC=C(1)+C(2)*LOG(RENDAFAM)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	3268.045	28.91675	113.0157	0.0000
C(2)	66.56618	16.78767	3.965182	0.0001

R-squared	0.011217	Mean dependent var	3365.076
Adjusted R-squared	0.010503	S.D. dependent var	577.0252
S.E. of regression	573.9869	Akaike info criterion	15.54453
Sum squared resid	4.57E+08	Schwarz criterion	15.55207
Log likelihood	-10785.90	Hannan-Quinn criter.	15.54735
F-statistic	15.72267	Durbin-Watson stat	1.913720
Prob(F-statistic)	0.000077		

Formas Funcionais Mais Usuais

Tipo de Relação	Modelo Estatístico	Interpretação de β_1
Nivel-Nível	$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$	$\frac{\Delta E(y \mid x)}{\Delta x} = \beta_1$
Log-Log (potência)	$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$	$\frac{\%\Delta E(y \mid x)}{\%\Delta x} = \beta_1$
Log-Nível (exponencial)	$\ln(y_i) = \beta_0 + \beta_1 \beta x_i + \varepsilon_i$	$\frac{\%\Delta E(y \mid x)}{\Delta x} = 100\beta_1$
Nível-Log (semi-Log)	$y_i = \beta_0 + \beta_1 \ln(x_i) + \varepsilon_i$	$\frac{\Delta E(y \mid x)}{\%\Delta x} = \frac{\beta_1}{100}$

Leituras Complementares

Leitura Complementar I

(O Modelo Recíproco)

Exemplo

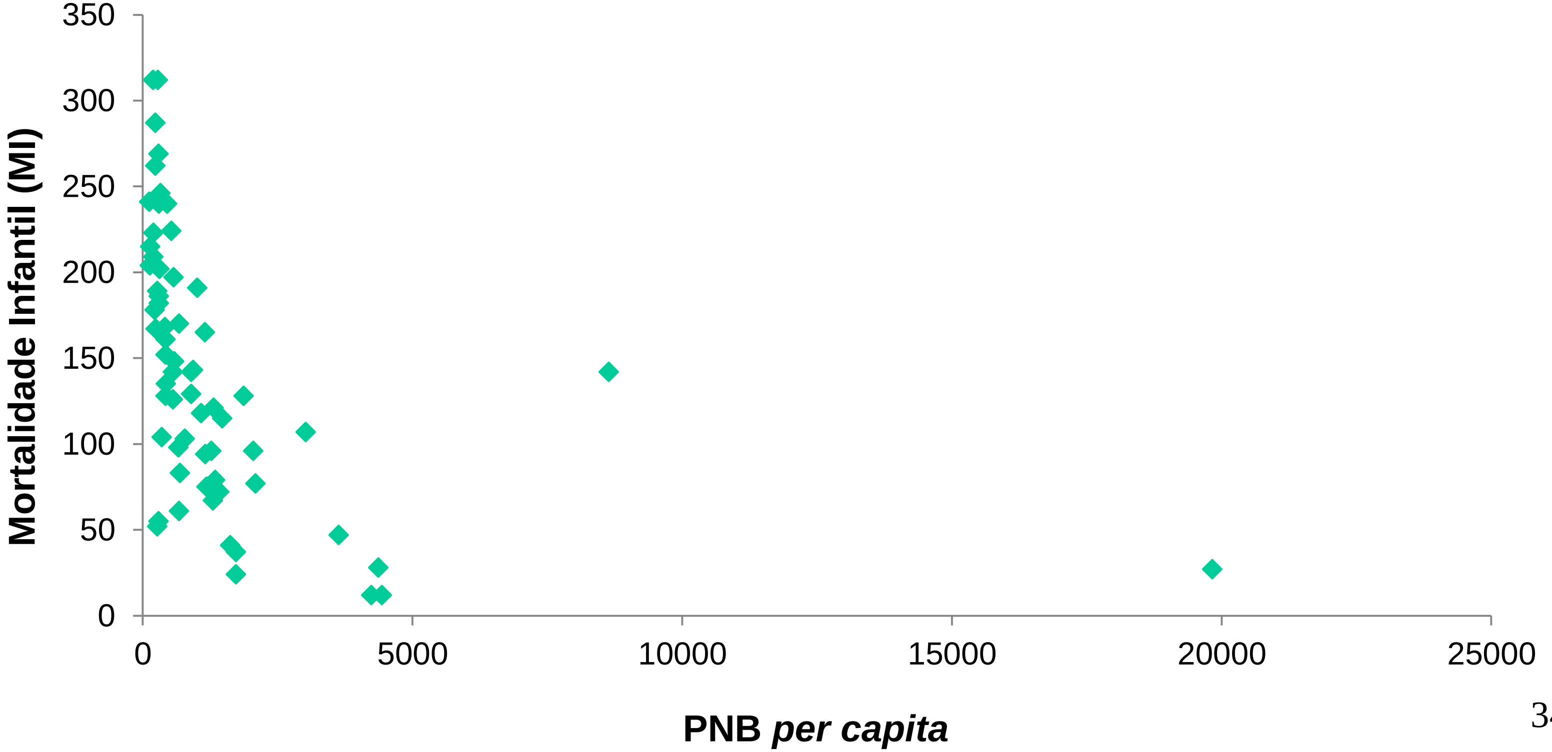
MUKHERJEE, WHITE e WHYTE (1998), coletaram informações sobre mortalidade infantil (MI – número anual de óbitos de crianças menores de 5 anos por 1000 nascidos vivos), PNB *per capita* (PNB *per capita* em 1980) e algumas outras variáveis de 64 países.

NOTA*

"O Produto Interno Bruto (PIB) difere do Produto Nacional Bruto (PNB) basicamente pela Renda Líquida Enviada ao Exterior (RLEE). A RLEE é desconsiderada no cálculo do PIB e considerada no cálculo do PNB, inclusive porque o PNB é gerado a partir da soma do PIB mais entradas e saídas de capital. A RLEE representa a diferença entre recursos enviados ao exterior (pagamento de fatores de produção internacionais alocados no país) e os recursos recebidos do exterior a partir de fatores de produção que, sendo do país considerado, encontram-se em atividade em outros países. Assim (e simplificadamente), caso um país possua empresas atuando em outros países, mas proíba a instalação de transnacionais no seu território, terá uma renda líquida enviada ao exterior negativa. Ainda, o país exemplificado terá um PNB maior que o PIB. No caso brasileiro, o PNB é menor que o PIB, uma vez que a RLEE é positiva (ou seja, envia-se mais recursos ao exterior do que se recebe)."

Voltando ao Exemplo

Como parte do estudo, os autores buscaram encontrar alguma relação entre MI e PNB, cujo diagrama de dispersão encontra-se disponível a seguir:



Voltando ao Exemplo

Do gráfico anterior, podemos observar que à medida que o PNB *per capita* aumenta, a mortalidade infantil reduz.

Tal comportamento já era esperado, uma vez que, mantendo tudo o mais constante, as pessoas podem elevar seus gastos com saúde à medida que o PNB per capita aumenta.

Todavia, a relação entre as variáveis parece ser não linear.

Assim, os autores anteriormente citados resolveram propor um modelo recíproco para analisar o comportamento entre essas duas variáveis.

O Modelo Recíproco

O modelo

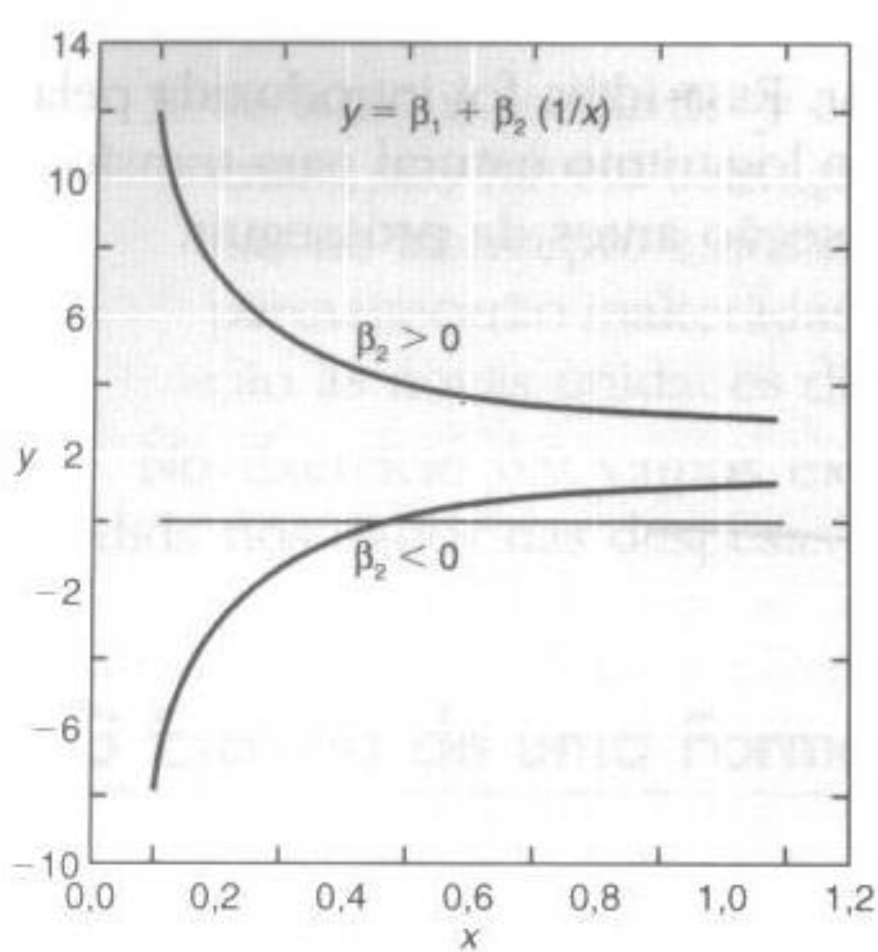
$$y = \beta_1 + \beta_2 \frac{1}{x} + \varepsilon$$

é conhecido na literatura como modelo recíproco.

Este modelo apresenta os seguintes aspectos: quando o valor do regressor aumenta indefinidamente, o termo $\beta_2(1/x)$ tende a zero e a variável resposta, portanto, aproxima-se de um valor limite ou assintótico β_1.

O Modelo Recíproco

Na figura a seguir são apresentados dois possíveis comportamentos entre as variáveis y e x, dada pelo modelo recíproco, dependendo do sinal do parâmetro β_2.



O Modelo Recíproco

Para o modelo recíproco, dado por:

$$y = \beta_1 + \beta_2 \frac{1}{x} + \varepsilon$$

não é difícil mostrar que:

$$\frac{dy}{dx} = -\beta_2 \frac{1}{x^2}$$

$$\eta = \frac{dy}{dx} \cdot \frac{x}{y} = -\beta_2 \frac{1}{x^2} \cdot \frac{x}{y} = -\beta_2 \cdot \frac{1}{xy}$$ **(elasticidade)**

Voltando ao Exemplo

Resultados obtidos após a estimação, por MQO:

$$MI = \beta_1 + \beta_2 \frac{1}{PNB} + \varepsilon$$

RESUMO DOS RESULTADOS

Estatística de regressão	
R-Quadrado	0.459
Erro padrão	56.330
Observações	64

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	166946.60	166946.60	52.61	7.82E-10
Resíduo	62	196731.40	3173.09		
Total	63	363678.00			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	81.79	10.83	7.55	2.38E-10	60.14	103.45
1/PNB	27273.17	3760.00	7.25	7.82E-10	19757.03	34789.30

Voltando ao Exemplo

Na medida em que o PNB *per capita* aumenta indefinidamente, a MI aproxima-se de seu valor assintótico de 82 óbitos anuais de crianças menores de 5 anos por 1000 nascidos vivos, em média.

Ainda, como a estimativa para o parâmetro β_2 deu positiva, isso implica que a taxa de variação da MI em relação ao PNB seja negativa.

Leitura Complementar II

(O Modelo Log-Inverso)

O Modelo Log-Inverso

A forma funcional para o modelo log-inverso é dada por:

$$\ln(y) = \beta_1 - \beta_2 \frac{1}{x} + \varepsilon$$

Ainda, não é difícil verificar que:

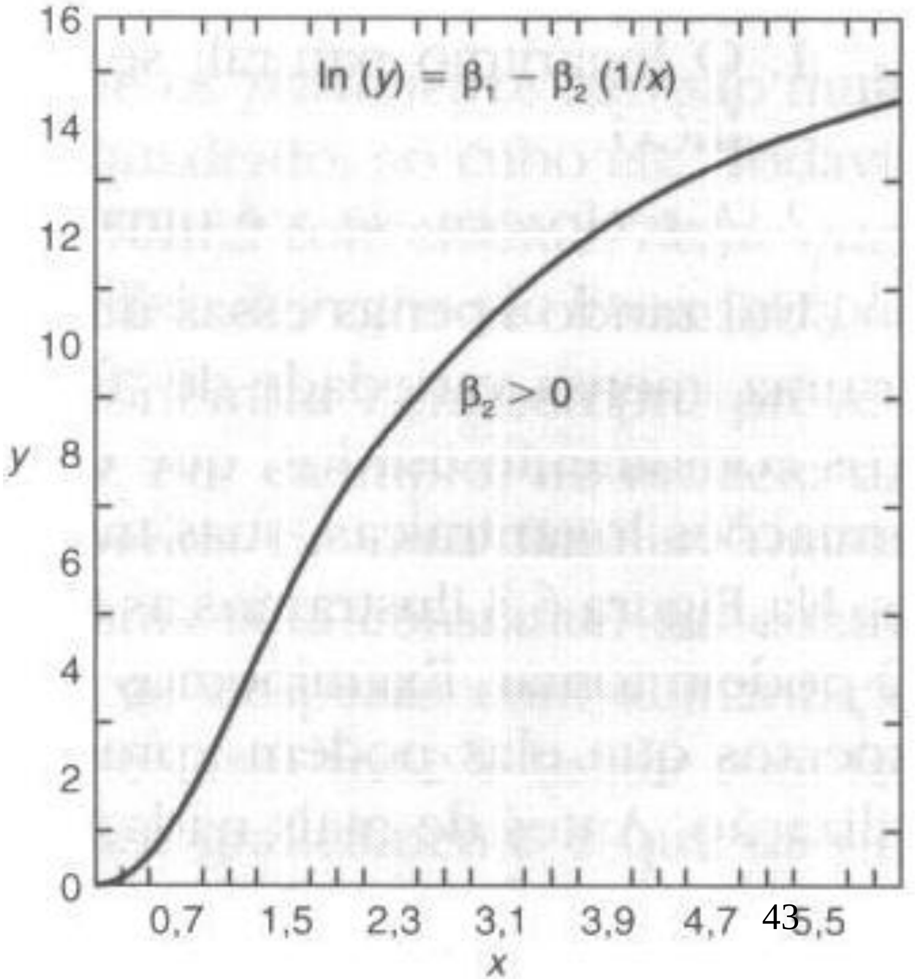
$$\frac{dy}{dx} = \beta_2 \frac{y}{x^2}$$

$$\eta = \frac{dy}{dx} \cdot \frac{x}{y} = \beta_2 \cdot \frac{1}{x}$$ **(elasticidade)**

O Modelo Log-Inverso

Observação

Apenas para fins ilustrativos, a figura ao lado apresenta o comportamento entre as variáveis y e x, dada pelo modelo log-inverso, para $\beta_2 > 0$.



Resumo dos Resultados

Quadro Resumo (Alguns Resultados Úteis)

Tipo	Modelo Estatístico	$\frac{dy}{dx}$	Elasticidade
Linear	$y_i = a+\beta x_i+e_i$	β	$\beta \frac{x_i}{y_i}$
Recíproco	$y_i = \alpha + \beta \frac{1}{x_i} + e_i$	$-\beta \frac{1}{x_i^2}$	$-\beta \frac{1}{x_i y_i}$
Log-Log	$\ln(y_i) = \alpha + \beta \ln(x_i) + e_i$	$\beta \frac{y_i}{x_i}$	β
Log-Nível	$\ln(y_i) = \alpha + \beta x_i + e_i$	βy_i	βx_i
Nível-Log	$y_i = \alpha + \beta \ln(x_i) + e_i$	$\beta \frac{1}{x_i}$	$\beta \frac{1}{y_i}$
Log-Inverso	$\ln(y_i) = \alpha - \beta \frac{1}{x_i} + e_i$	$\beta \frac{y_i}{x_i^2}$	$\beta \frac{1}{x_i}$