Regressão Múltipla – Tópicos Adicionais

HO 231 – Econometria

Profa. Rosangela Ballini

Instituto de Economia - UNICAMP

Ementa

- Propriedades dos Estimadores
- Linearidade por Anamorfose
- Normalidade dos Erros
- Previsão de *Y* com Regressor em Logaritmo
- Comparação de Modelos (Teste de Wald e R^2 ajustado)

Bibliografia

Wooldridge, J. M. 2003. *Introductory Econometrics*. Cap. 5 e 6.

Propriedades dos Estimadores

A escolha do melhor estimador para um parâmetro baseia-se nas seguintes propriedades:

1) Exatidão: capacidade do estimador de, em média, igualar-se ao real parâmetro da população;

$$E(\hat{\theta}) = \theta$$

2) Eficência: medida de precisão de um estimador exato. Pode ser relativa (quando é menor que a de outro estimador) ou absoluta (quando é maior que qualquer outro estimador);

$$Var(\hat{\theta}) < Var(\hat{\theta}^*)$$

3) Consistência: capacidade do estimador de convergir para o real parâmetro da população em grandes amostras;

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta \qquad \lim_{n \to \infty} Var(\hat{\theta}) = 0$$

Teorema de Gauss-Markov

- Para que os estimadores de MQO sejam os Melhores Estimadores Lineares Não Viesados (MELNV, ou *BLUE*):

Modelo Clássico de Regressão Linear

1) Relação Linear entre *X* e *Y;*

2) Os valores de *X* são fixos em repetidas amostras, e não aleatórios;

3) Colineariedade não perfeita entre os regressores

4) Os erros possuem média condicional zero, ou seja, $E(u_i)=0$

5) A variabilidade dos erros é constante, qualquer que seja *X*:

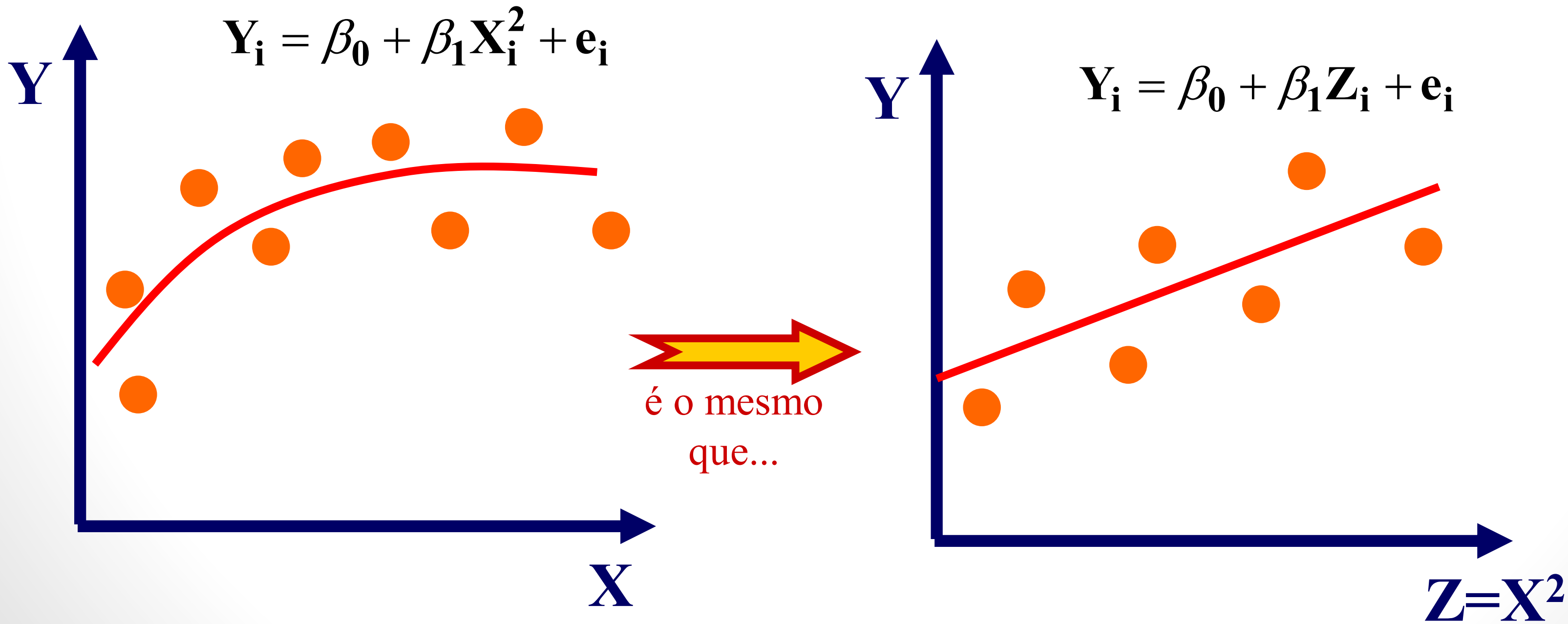
6) Os erros são não autocorrelacionados, ou seja, $E(e_i e_j)=0$:

7) Os erros apresentam distribuição normal:

Não é um pressuposto necessário para que os estimadores de MQO sejam MELNV, mas necessário para que estes tenham distribuição normal.

Linearidade por Anamorfose

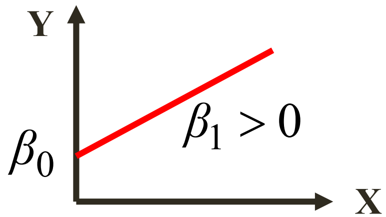
- Relações não lineares nas variáveis podem ser transformadas em relações lineares por anamorfose, ou seja, por transformações de suas variáveis originais;
- A escolha da forma funcional (tipo de transformação das variáveis) dependerá da análise da distribuição dos valores e, principalmente, do conhecimento prévio das relações por parte do pesquisador;



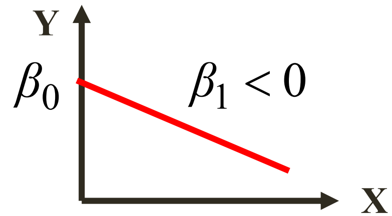
Formas Funcionais - Exemplos

1) Modelo linear:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

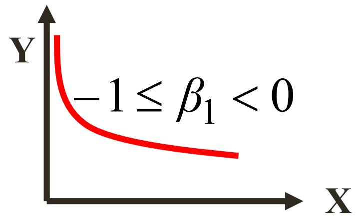


ou

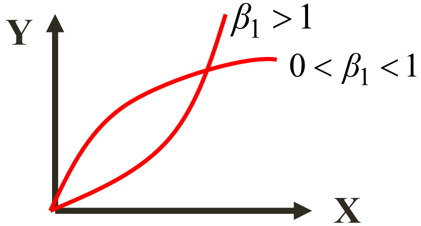


2) Modelo log-log:

$$ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + e_i$$

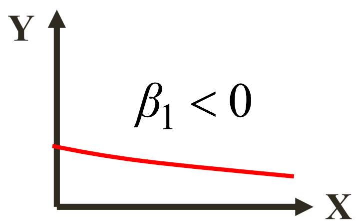


ou

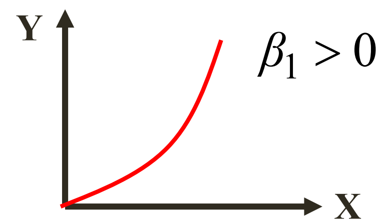


3) Modelo log-lin:

$$ln(Y_i) = \beta_0 + \beta_1 X_i + e_i$$

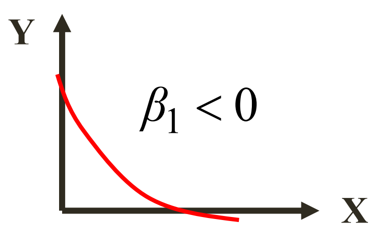


ou

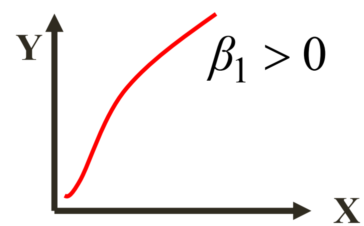


4) Modelo lin-log:

$$Y_i = \beta_0 + \beta_1 ln(X_i) + e_i$$

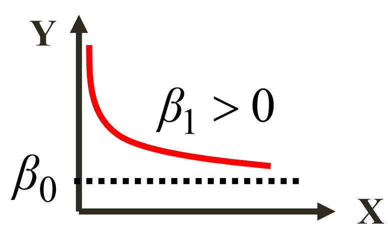


ou

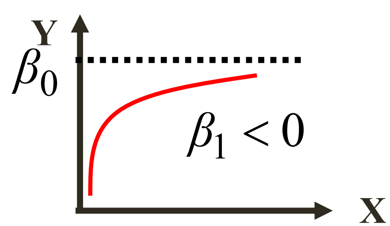


5) Modelo inverso:

$$Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + e_i$$

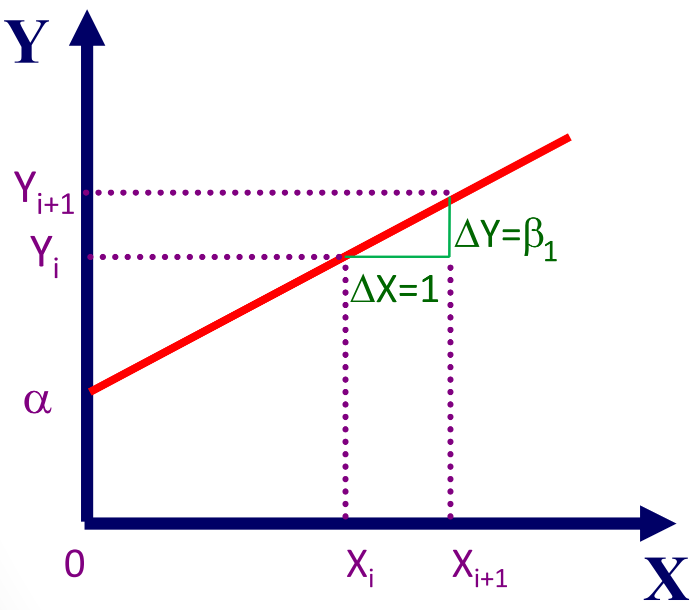


ou



Modelo Linear - Interpretação

- Um modelo linear pressupõe que *Y* apresente variações absolutas constantes para variações absolutas constantes em *X*;
- A variação marginal no valor esperado de *Y* é a mesma para qualquer valor de *X*;



$$E[Y|X=0]=\beta_0$$

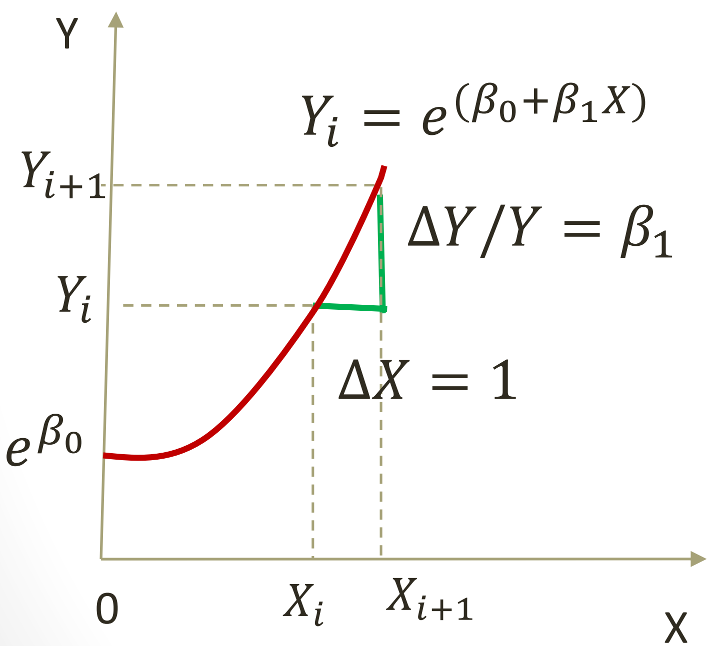
β_0 é o valor esperado de *Y* para valores nulos de *X*

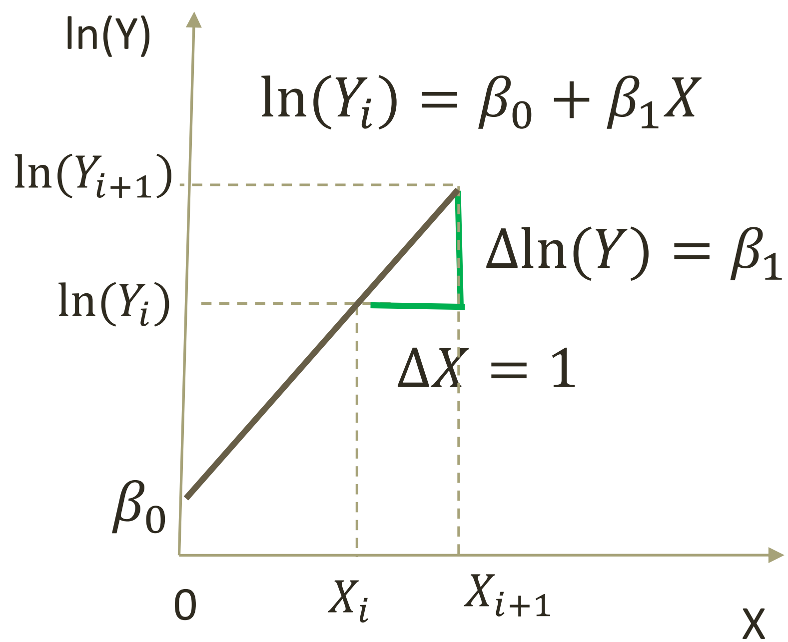
$$\frac{\Delta Y}{\Delta X}=\frac{dY}{dX}=\frac{d(\beta_0+\beta_1 X)}{dX}=\beta_1$$

β_1 é a variação marginal absoluta no valor esperado de *Y* (ΔY) dada uma variação unitária em *X* ($\Delta X=1$).

Modelo Log-Lin - Interpretação

- Pressupõe que *Y* apresente variações relativas constantes em função de variações absolutas constantes de X;
- Exemplo: uma variação de 0,01 em ln(*Y*) significa uma variação de 1% no valor de *Y*;

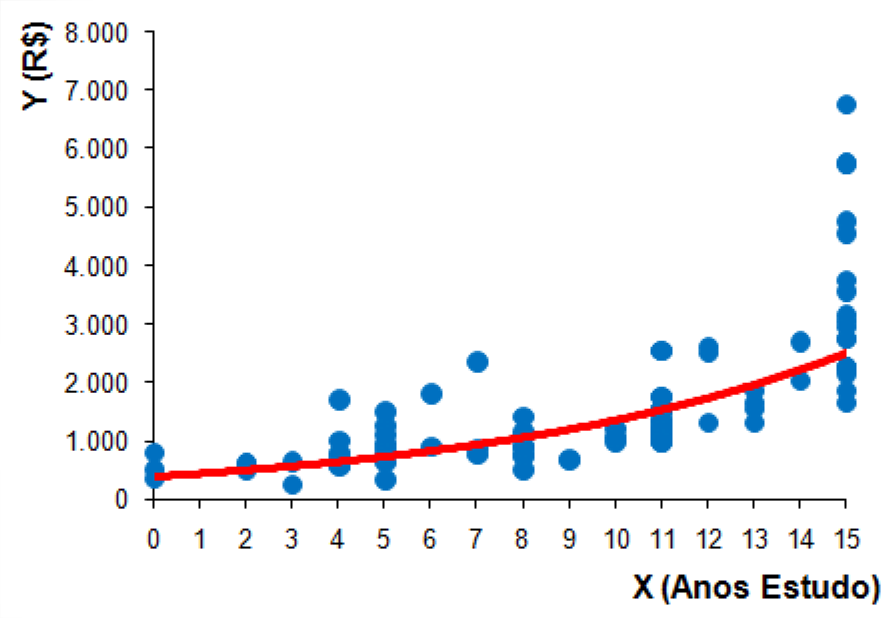




$$\frac{\Delta \ln(Y)}{\Delta Y} = \frac{1}{Y} \Rightarrow \Delta \ln(Y) = \frac{\Delta Y}{Y} \quad \Longrightarrow \quad \frac{\Delta \ln(Y)}{\Delta X} = \frac{\Delta Y/Y}{\Delta X} = \beta_1$$

Modelo Log-Lin - Exemplo

Qual o impacto da escolaridade sobre a renda do trabalho?



Pressupondo que a renda do trabalho (Y, em R$ mensais) cresça exponencialmente com os anos de escolaridade (X):

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + e_i$$

Modelo estimado:

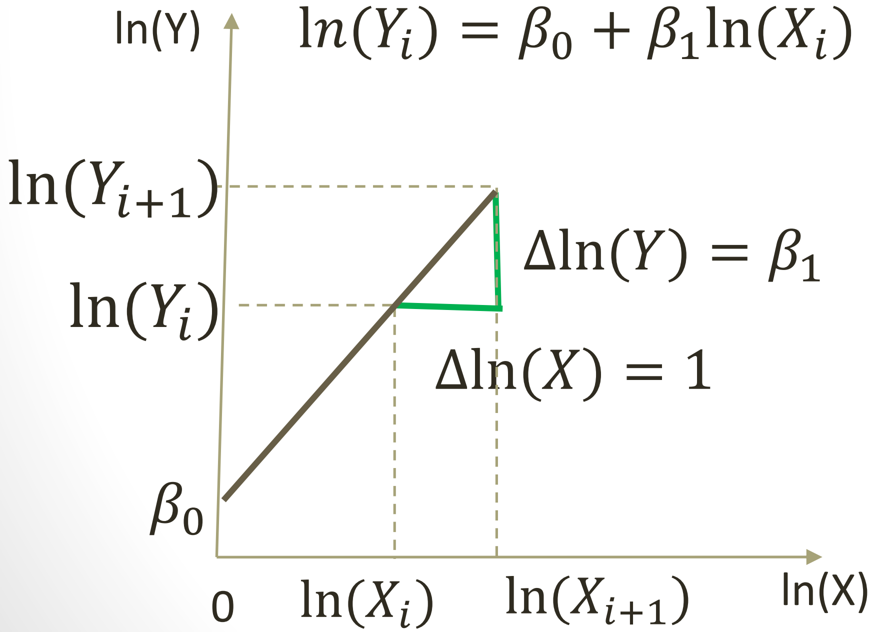
$$\ln(Y_i) = 6{,}006 + 0{,}121 X_i + \hat{e}_i$$

Espera-se, para cada ano adicional de escolaridade (ΔX=1), um incremento relativo de 12,1% no rendimento do trabalho (ΔY=0,121 Y_i).
O rendimento esperado para aqueles sem escolaridade (X=0) seria de aproximadamente 406 reais ($e^{6{,}006}$).

Modelo Log-Log - Interpretação

- Modelo log-log pressupõe que Y apresente variações relativas constantes em função de variações relativas constantes de X;
- Coeficiente angular será uma medida constante da elasticidade de *Y* em relação a *X*, ou seja, será a variação percentual em *Y* para uma variação de 1% em *X*.
- Exemplo: Função de Cobb-Douglas: $Y_i = \alpha X_i^{\beta_1}$

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i)$$

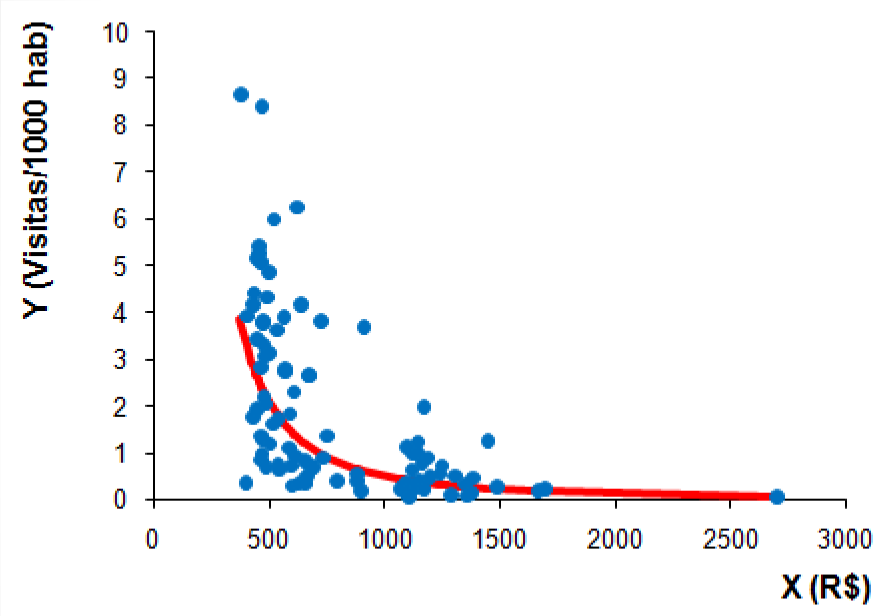


$$\frac{\Delta \ln(Y)}{\Delta \ln(X)} = \frac{d[\beta_0 + \beta_1 \ln(X)]}{d ln(X)} = \beta_1$$

$$\frac{\Delta \ln(Y)}{\Delta \ln(X)} = \beta_1 \Rightarrow \frac{\Delta Y / Y_i}{\Delta X / X_i} = \beta_1$$

Modelo Log-Log - Exemplo

Como o custo de viagem afeta a taxa de visitação a um parque nacional?



Pressupõe-se uma elasticidade constante entre taxa de visitação (Y, em visitas/100.000 habitantes) e custo de viagem do município de origem ao parque (X, em R$), ou seja, incrementos percentuais no custo de viagem gerariam reduções percentuais na taxa de visitação:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i)$$

Modelo estimado:

$$\ln(Y_i) = 13{,}492 - 2{,}049 \ln(X_i) + \hat{e}_i$$

O coeficiente angular sugere uma demanda relativamente elástica às variações no custo de viagem. Para cada aumento de percentual no custo de viagem, espera-se uma redução 2,05% na taxa de visitação.

Modelos Logarítmicos - Exemplo

- A partir do arquivo Dados_CO2.xlsx estime:

1. $CO2 = \beta_0 + \beta_1 PIB + \beta_2 Setor2 + u$
2. $\ln(CO2) = \beta_0 + \beta_1 PIB + \beta_2 Setor2 + u$
3. $\ln(CO2) = \beta_0 + \beta_1 \ln(PIB) + \beta_2 Setor2 + u$

Modelos Logarítmicos - Exemplo

- Modelos com mesmo regressor podem ser comparados com o R^2;

Variable	linear	loglin	loglog
pib	.00032095	.00009541	
	12.27	10.30	
	0.0000	0.0000	
setor2	.10508293	.04189997	.02460615
	5.29	5.96	5.26
	0.0000	0.0000	0.0000
lnpib			.82701542
			21.54
			0.0000
_cons	-1.1262519	-1.3743163	-6.5469295
	-1.68	-5.79	-21.14
	0.0948	0.0000	0.0000
F	87.854872	69.502992	270.07999
r2	.51420759	.45574838	.76492579

legend: b/t/p

Variações absolutas constantes: para cada variação de 1 US$ no PIB per capita, espera-se um acréscimo de 0,3 kg nas emissões de CO^2;

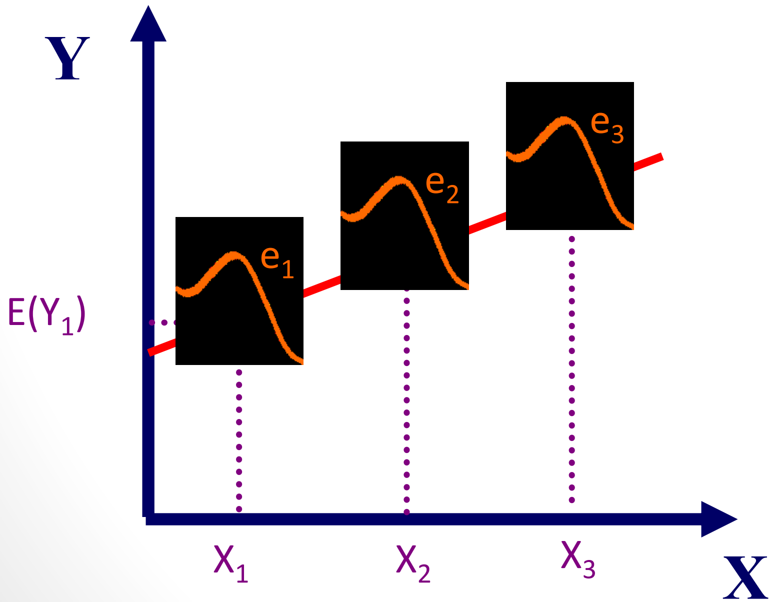
Variações absolutas em X e relativas em Y: Para cada variação de 1 US$ no PIB per capita, espera-se um acréscimo de 0,009% nas emissões de CO^2;

Variações relativas em *X* e *Y*: Para cada variação de 1% no PIB, espera-se uma variação de 0,83% nas emissões de CO^2.

O R2 do modelo linear com os dos demais modelos não são diretamente comparáveis. Entretanto, a expressiva diferença entre o R2 do modelo log-log em relação aos demais, sugere a maior qualidade deste ajuste. Outras estatísticas podem, entretanto, auxiliar na decisão...

Normalidade dos Erros

- Seja o modelo de regressão dado por : $Y = \alpha + \beta X + e$
- Pressupõe-se que o erro seja a soma de inúmeros fatores com efeito mínimo sobre a variável dependente, e que esteja aleatoria e normalmente distribuídos em torno da reta de regressão;
- É o mesmo que pressupor que os valores de Y estejam normalmente distribuídos em torno de sua esperança condicional;



O teste *F* para o modelo e os testes *t* para os coeficientes angulares baseiam-se no pressuposto da normalidade dos erros. Caso esses não estejam normalmente distribuídos os testes serão inválidos para amostras pequenas. **Para amostras grandes, o pressuposto de normalidade não é necessário para que as estatísticas de teste sejam válidas**.

Normalidade dos Erros - Histograma

- O objeto lm contém uma lista de componentes, entre eles **residuals**;

```
> res=loglog$residuals
```

res é uma variável que recebe os valores dos resíduos do modelo loglog ajustado.

- O histograma permite visualizar a forma de distribuição observada dos dados e compará-la com uma distribuição teórica;
- O comando **hist.()** elabora um histograma da variável desejada.
- A opção **density** adiciona a função densidade à frequência observada;

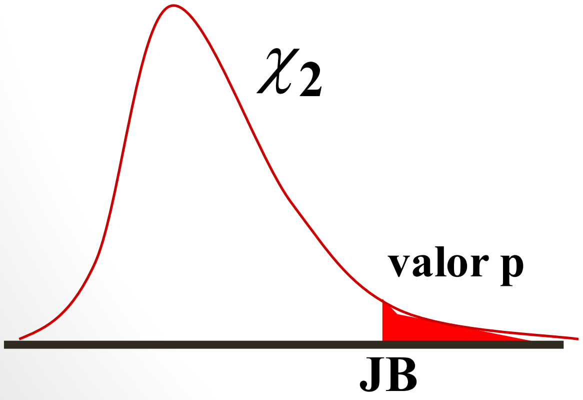
```
> hist(res, freq=F)
> par(new=TRUE)
> plot(density(res),main='',xlab='',ylab='‘)
```

Teste de Jarque-Bera

O teste Jarque-Bera de normalidade é um **teste assintótico** (para amostras grandes) baseado nas medidas de assimetria (AS) e curtose (K) dos resíduos. O valor da estatística é calculado por :

$$JB = n\left[\frac{AS^2}{6} + \frac{(K-3)^2}{24}\right]$$

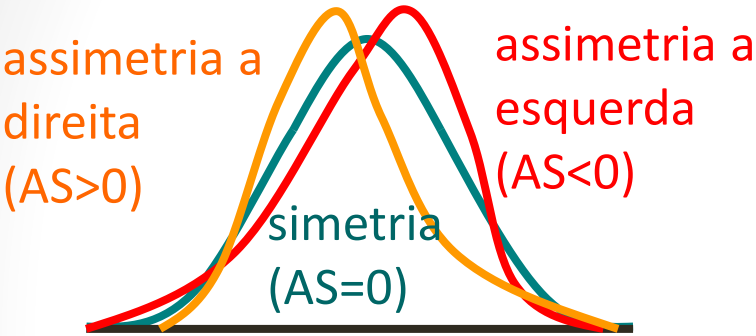
Sob a hipótese nula de normalidade dos erros, a estatística JB segue uma distribuição qui-quadrado com 2 graus de liberdade. Em outras palavras, dadas as hipóteses:



$$\begin{cases} H_0: \text{Erros são normalmente distribuídos} \\ H_1: \text{Erros não são normalmente distribuídos} \end{cases}$$

Medidas de Assimetria e Curtose

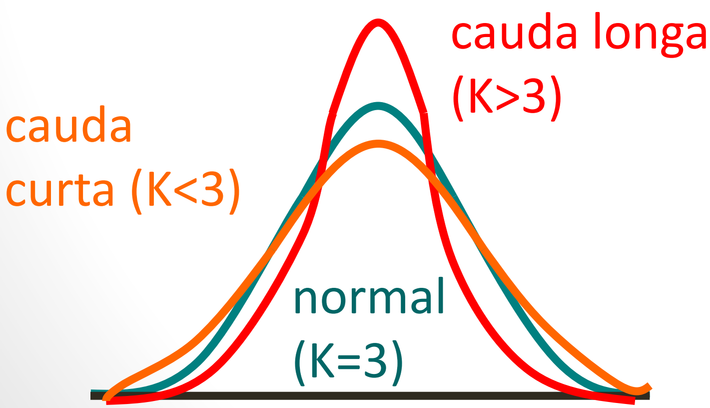
1) Assimetria:



Simetria refere-se a quão simetricamente os resíduos estão distribuídos em torno da média zero, e será dado por:

$$AS = \frac{1}{n}\sum\left(\frac{e_i - \bar{e}}{\hat{\sigma}}\right)^3$$

2) Curtose



Curtose refere-se ao achatamento da distribuição em relação a uma curva normal, e será dado por:

$$K = \frac{1}{n}\sum\left(\frac{e_i - \bar{e}}{\hat{\sigma}}\right)^4$$

Teste de Shapiro-Wilk

Teste de Shapiro-Wilk (1965): utilizado para amostras de qualquer tamanho.

O teste e baseado no cálculo da estatística W que verifica se uma amostra aleatória de tamanho n provem de uma distribuição normal:

$$W = \frac{b}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Sendo X_i valores da amostra e b uma constante determinada a partir das médias, variâncias de uma amostra de tamanho n. As hipóteses nula e alternativa são:

H0: a amostra provem de uma População Normal

H1: a amostra não provem de uma População Normal

Teste de Normalidade dos Erros

$$\begin{cases} H_0: & \text{Erros normais} \\ H_1: & \text{Erros não normais} \end{cases}$$

Rejeitar H_0 significa encontrar evidências que os erros não estão normalmente distribuídos;

- Entre os diversos testes de normalidade, o teste de Shapiro-Wilk apresenta um bom poder (capacidade de rejeitar H_0 em populações não normais);

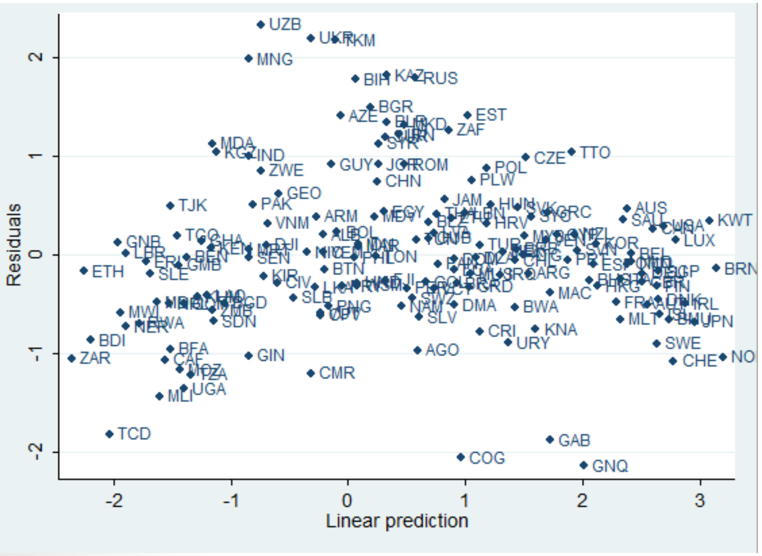
```
> shapiro.test(res)

Shapiro-Wilk normality test data: res
       W = 0.97117, p-value = 0.001372
```

Há evidências para rejeitar H_0, ou seja, para afirmar que os erros do modelo não estejam normalmente distribuídos.
Assim, a validade das estatísticas *t* e *F* estará condicionada às propriedades assintóticas dos estimadores (para grandes amostras). Em outras palavras, as estatísticas *t* e *F* não seriam apropriadas para amostras pequenas.

Análise Exploratória – *scatter*

- A dispersão dos resíduos em torno dos valores previstos permite, por exemplo, identificar eventuais valores extremos ou falhas de especificação;

O comando **fitted.values**, fornece os valores previstos na variável *ln(co2)*.



A dispersão dos resíduos em torno dos valores previstos sugere um possível comportamente quadrático do regressando em função dos regressores (provavelmente PIB per capita).

Há ainda valores extremos, como Congo (COG) que possui um valor de lnco2 muito inferior ao esperado, e Uzbesquitão (UZB), com um valor muito superior ao esperado.

Previsão de *Y* em Equações (Semi-)Logarítmicas

- Seja a relação semi-logarítmica: $\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$

- Os valores previstos de ln(*Y*) serão:

$$\widehat{\ln(Y)} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

- A simples previsão de *Y* pelo antilogaritmo de ln(*Y*) será viesada:

$$\hat{Y} \neq e^{\widehat{\ln(Y)}}$$

- Se o modelo satisfaz as hipóteses RLM1 a RLM6, tem-se:

$$E(Y|X_1, X_2) = \exp\left(\frac{\sigma^2}{2}\right).\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

- Sendo a previsão dada por: $$\hat{Y} = \exp\left(\frac{\hat{\sigma}^2}{2}\right)\exp(\widehat{\ln(Y)})$$

A qual é uma estimativa consistente. Porém, depende da normalidade do termo erro.

- Se o termo erro é independente das variáveis explicativas, tem-se:

$$E(Y|X_1, X_2) = \alpha_0 . \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

em que α_0 é o valor esperado de *exp(u)*.

- Dessa forma, o valor previsto de y é dada por:

$$\hat{Y} = \hat{\alpha}_0 \exp(\widehat{\ln(Y)})$$

- Duas formas para estimar α_0:

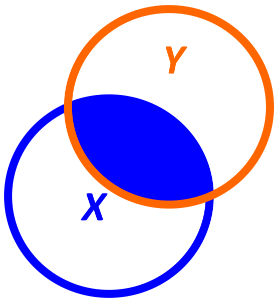
1. Estimador do método dos momentos:
$$\hat{\alpha}_0 = \frac{\sum \exp(\hat{u})}{n}$$
2. Ou usando as estimativas de MQO:
$$\check{\alpha}_0 = \left(\sum \widehat{m}_i^2\right)^{-1} \left(\sum \widehat{m}_i y_i\right)$$

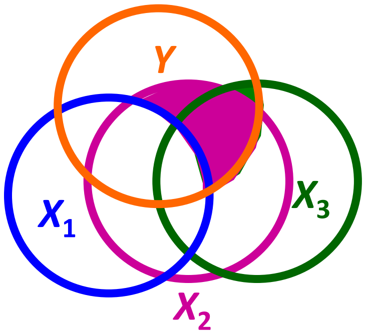
em que $\widehat{m}_i = \exp(\widehat{\ln(y)})$

Comparação de Modelos *Nested*

- Sejam dois modelos, um deles contendo um sub-conjunto de regressores do outro (*nested models*):

$$Y = \alpha + \beta X_1 + e \qquad Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$





Pergunta: Podemos de fato afirmar que as variáveis adicionais X_2 e X_3 acrescentem informação relevante para explicar *Y*?

Teste de Wald:

- Teste usado para verificar se um conjunto de q regressores contribui para explicar a variável dependente:

$$\begin{cases} H_0 : \beta_{j+1} = .. = \beta_{j+q} = 0 \\ H_1 : \beta_{j+s} \neq 0 \end{cases}$$

Teste F: incorpora a restrição no procedimento de estimativa dos parâmetros.

Esta questão é verificada por meio do teste F :

$$F = \frac{\dfrac{(SQRes_R - SQRes_U)}{(g_R - g_U)}}{SQRes_U / g_U}$$

Sendo, g_R graus de liberdade do modelo restrito e g_U graus de liberdade irrestrito.

- Testes para grandes amostras: estatística multipliador de Lagrange (LM)
- LM: exige apenas a estimação do modelo restrito
- Etapas para obtermos LM:

1. Regrida *Y* sobre o conjunto restrito de *q* variáveis e salve os resíduos $\tilde{u}$
2. Regrida $\tilde{u}$ sobre todas as variáveis independents e obtenha o R_u^2
3. Calcule $LM = n * R_u^2$
4. Para *q* graus de liberdade, a estatística LM tem distribuição χ_q^2, obtenha o p-valor. Se p-valor menor que um determinado nível de significância rejeita-se H_0

Modelos *Nested* - Exemplo

- Sejam dois nested-models para o logaritmo das emissões:

$$\ln(co2) = \alpha + \beta setor2 + e \qquad \ln(co2) = \alpha + \beta_1 setor2 + \beta_2 pib + \beta_3 pib^2 + e$$

- Para sabermos se as variáveis *pib* e *pib2* acrescentam informações relevantes à previsão das emissões, devemos testar:

$$\begin{cases} H_0 : \beta_2 = \beta_3 = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

- Teste de Wald no R: função **linearHypothesis**:

```
> linearHypothesis(unrestr,c("pib=0","pib2=0"),test="F")

> linearHypothesis(unrestr,c("pib=0","pib2=0"),test="Chisq")
```

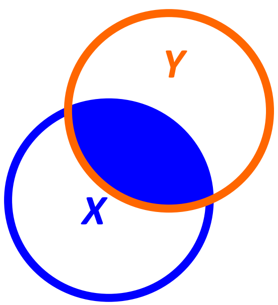
Testa a hipótese que os coeficientes associados às variáveis *pib* e *pib2* são, simultaneamente, iguais a zero.

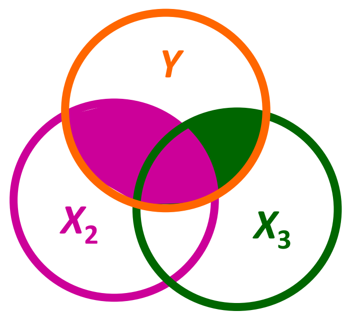
A probabilidade de erro ao afirmarmos que as variáveis acrescentam informação relevante é praticamente nula.

Comparação de Modelos *Non-Nested*

- Sejam dois modelos, contendo conjuntos distintos de regressores (modelos *non-nested*):

$$Y = \alpha + \beta X_1 + e \qquad\qquad Y = \alpha + \beta_2 X_2 + \beta_3 X_3 + e$$





Pergunta: Qual é o melhor modelo para explicar o comportamento de *Y*?

Coeficiente de Determinação Ajustado ($\bar{R}^2$):

$$\bar{R}^2 = 1 - \frac{SQRes/[n-(k+1)]}{STQ/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-(k+1)}$$

O R^2 ajustado pondera o coeficiente de determinação (R^2) pelo número de variáveis explicativas e pelo número de observações da amostra.

Particularmente útil quando compara-se modelos de regressão múltipla que prevêem a mesma variável dependente, pois penaliza aquele modelo com maior número de variáveis independentes.

Modelos *Non-Nested* - Exemplo

- Sejam dois non-nested models para o logaritmo das emissões:

$$\ln(co2) = \alpha + \beta \ln(pib) + e \qquad\qquad \ln(co2) = \alpha + \beta_2 pib + \beta_3 pib^2 + e$$

- O R^2 ajustado é apresentado juntamente às estimativas da tabela ANOVA:

```
----------------------------------------------------------------
Observations              169                    169
R2                        0.72571                0.47437
Adjusted R2               0.72407                0.46804
F Statistic               441.85050***           74.90614***
                          (df = 1; 167)          (df = 2; 166)
================================================================
```

O primeiro modelo tem um maior poder de explicação da variabilidade do *lnco*2. Após poderarmos pelo número de variáveis de cada modelo, o poder de explicação do primeiro modelo é de 72,4% e o do segundo modelo é de 46,8%.

Exercícios

1) A partir do arquivo *wage2.dta*:
   a) Compare a qualidade do ajuste de um modelo linear e outro log-lin para a renda (*wage*) como função linear dos anos de educação (*educ*), quoeficiente de inteligência (*IQ*) e experiência profissional (*exper*);
   b) Compare a normalidade dos resíduos nos dois ajustes;
   c) Realize previsões para a renda (*wage*) a partir do modelo log-lin;
   d) Analise a contribuição de um termo quadrático para a experiência profissional;

Exercícios

2) A partir do arquivo Dados_PIB.XLS, que contém informações do WDB sobre sobre gastos com P&D, PIB per capita, razão de dependência para jovens e gastos com educação dos países em 2010, pede-se:
   a) Compare a qualidade dos ajustes linear, log-lin e log-log;
   b) Compare a normalidade dos resíduos de cada ajuste;
   c) Realize previsões para o PIB per capita nas funções logarítmicas;
   d) Avalie a necessidade de um termo quadrático para a razão de dependência;