

# **AULAS 26 E 27**

# **Análise de regressão múltipla: problemas adicionais**

**Ernesto F. L. Amaral**

**12 e 14 de junho de 2012**

**Avaliação de Políticas Públicas (DCP 046)**

**Fonte:**

**Wooldridge, Jeffrey M. “Introdução à econometria: uma abordagem moderna”. São Paulo:  
Cengage Learning, 2008. pp.174-206 (capítulo 6).**

# EFEITOS DA DIMENSÃO DOS DADOS NAS ESTATÍSTICAS

- Mudanças das unidades de medida das variáveis não afeta o  $R^2$ .
- A intenção agora é de examinar o efeito do redimensionamento das variáveis dependente ou independente sobre:
  - Erros-padrão.
  - Estatísticas  $t$ .
  - Estatísticas  $F$ .
  - Intervalos de confiança.
- Escolhendo as unidades de medida, a aparência da equação estimada pode melhorar, sem alterar a essência do modelo.
- É geralmente realizada com valores monetários, especialmente quando os montantes são muito grandes.

## EXEMPLO

- *pesônas*: peso dos recém-nascidos, em onças.
- *cigs*: número médio de cigarros que a mãe fumou por dia durante a gravidez.
- *rendfam*: renda anual familiar, em milhares de dólares.
- Equação 1:

$$\widehat{pesonas} = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 rendfam$$

# EFEITOS DA DIMENSÃO DOS DADOS

Variável Dependente	(1) pesonas	(2) pesonaslb = pesonas/16	(3) pesonas
Variáveis Independentes			
<i>cigs</i>	-0,4634 (0,0916)	-0,0289 (0,0057)	-----
<i>maços = cigs/20</i>	-----	-----	-9,268 (1,832)
<i>rendfam</i>	0,0927 (0,0292)	0,0058 (0,0018)	0,0927 (0,0292)
<i>intercepto</i>	116,974 (1,049)	7,3109 (0,0656)	116,974 (1,049)
Observações	1.388	1.388	1.388
R-quadrado	0,0298	0,0298	0,0298
SQR	557.485,51	2.177,6778	557.485,51
EPR	20,063	1,2539	20,063

## MUDANÇA NA DEPENDENTE

– Não importa como a variável dependente seja medida, os efeitos da constante e coeficientes são transformados nas mesmas unidades.

– Equação 2:

$$\widehat{pessoas}/16 = \hat{\beta}_0/16 + (\hat{\beta}_1/16)cigs + (\hat{\beta}_2/16)rendfam$$

## E A SIGNIFICÂNCIA ESTATÍSTICA?

- A alteração da variável dependente de onças para libras não tem efeito sobre o quanto são estatisticamente importantes as variáveis independentes.
- Os erros-padrão na coluna (2) são 16 vezes menores que os da coluna (1).
- As estatísticas  $t$  na coluna (2) são idênticas às da coluna (1).
- Os pontos extremos dos intervalos de confiança na coluna (2) são exatamente os pontos extremos na coluna (1) divididos por 16, já que ICs mudam pelos mesmos fatores dos erros-padrão.
- IC de 95% é beta estimado +/- 1,96 erro padrão estimado.

## E O GRAU DE AJUSTE? E O SQR? E O EPR?

- Os  $R^2$  das duas regressões são idênticos, como esperado.
- A soma dos resíduos quadrados (SQR) e o erro-padrão da regressão (EPR) possuem diferentes equações.
- Quando *pesonaslb* é a variável dependente, o resíduo da observação *i* na equação (2) é:  $\hat{u}_i/16$
- O resíduo quadrado em (2) é:  $(\hat{u}_i/16)^2 = \hat{u}_i^2/256$
- Por isso, **SQR(2) = SQR(1) / 256.**
- Como:  $EPR = \hat{\sigma} = \sqrt{SQR/(n - k - 1)} = \sqrt{SQR/1.385}$
- Por isso, **EPR(2) = EPR(1) / 16.**

## REDUZIMOS O ERRO?

- O erro na equação com  $pesonas/b$  como a variável dependente tem um desvio-padrão 16 vezes menor do que o desvio-padrão do erro original.
- Isso não significa reduzir o erro por mudar a medida da variável dependente.
- O EPR menor simplesmente reflete uma diferença nas unidades de medida.



## MUDANÇA NA INDEPENDENTE

- *maços*: quantidade de maços de cigarros fumados por dia:

$$maços = cigs / 20$$

$$\begin{aligned} \widehat{pessoas} &= \hat{\beta}_0 + (20\hat{\beta}_1) \left( \frac{cigs}{20} \right) + \hat{\beta}_2 rendfam \\ &= \hat{\beta}_0 + (20\hat{\beta}_1) maços + \hat{\beta}_2 rendfam \end{aligned}$$

- O intercepto e o coeficiente de inclinação de *rendfam* não se alteraram.
- O coeficiente de *maços* é 20 vezes o de *cigs*.
- O erro-padrão de *maços* é 20 vezes o de *cigs*, o que significa que a estatística *t* é a mesma.
- Se *maços* e *cigs* fossem inseridos conjuntamente, teríamos multicolinearidade perfeita.

## COEFICIENTES BETA

- Algumas vezes, uma variável-chave é medida em uma dimensão de difícil interpretação.
- Primeiro exemplo: ao invés de perguntar o efeito sobre o salário, proveniente do aumento em dez pontos em um teste, talvez faça mais sentido perguntar sobre efeito proveniente do aumento de um desvio-padrão.
- Segundo exemplo: é o caso de variáveis criadas com análise fatorial, já que não sabemos exatamente o que a unidade de medida significa.
- Como o desvio-padrão da variável “fatorial” é geralmente próximo de uma unidade, verificamos o efeito na unidade da variável dependente (beta), após a alteração de um desvio-padrão na variável independente.

## COEFICIENTES PADRONIZADOS

- Algumas vezes é útil obter resultados de regressão quando todas as variáveis tenham sido padronizadas.
- Uma variável é padronizada pela subtração de sua média e dividindo o resultado por seu desvio-padrão.
- Ou seja, computamos a transformação z de cada variável e depois fazemos a regressão usando esses valores z.

- Portanto, partimos de:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik} + \hat{u}$$

- Novo beta = beta original \* (dp de x / dp de y)
- Intercepto (beta zero) não existe mais:

$$z_y = \hat{b}_1 z_1 + \hat{b}_2 z_2 + \dots + \hat{b}_k z_k + erro$$

- Para  $j = 1, \dots, k$ , os coeficientes são:  $\hat{b}_j = (\hat{\sigma}_j / \hat{\sigma}_y) \hat{\beta}_j$

# INTERPRETANDO COEFICIENTES PADRONIZADOS

- Os coeficientes padronizados são também chamados de coeficientes beta.
- Se  $x_1$  aumentar em um desvio-padrão, então o  $y$  predito será alterado em  $b_1$  desvios-padrão.
- Os efeitos não estão sendo medidos em termos das unidades originais de  $y$  ou de  $x_j$ , mas em unidades de desvios-padrão.
- A dimensão das variáveis independentes passa a ser irrelevante, colocando-as em igualdade.
- Quando cada  $x_j$  é padronizado, a comparação das magnitudes dos coeficientes (significância econômica) é mais convincente. Ou seja, a variável com maior coeficiente é a “mais importante”.
- O Stata apresenta os beta padronizados com opção “, beta”.

# USO DE FORMAS FUNCIONAIS LOGARÍTMICAS

- O uso de logaritmos das variáveis dependentes ou independentes é o artifício mais comum em econometria para permitir relações não lineares entre a variável explicada e as variáveis explicativas.

$$\widehat{\log}(y) = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2$$

- $\beta_1$  é a elasticidade de  $y$ , em relação a  $x_1$ :
  - Quando  $x_1$  aumenta em 1%,  $y$  aumenta em  $\beta_1\%$ , mantendo  $x_2$  fixo.
- $100*\beta_2$  é a semi-elasticidade de  $y$ , em relação a  $x_2$ :
  - Quando  $x_2$  aumenta em 1,  $y$  aumenta em  $100*[\exp(\beta_2)-1]$ , mantendo  $x_1$  fixo.
  - No entanto, podemos utilizar  $100*\beta_2$ , quando temos pequenas mudanças percentuais.

## PECULIARIDADES DO USO DE LOGARITMOS

- Com log, ignoramos unidades de medida das variáveis, pois coeficientes de inclinação não variam pelas unidades.
- Quando  $y > 0$ , os modelos que usam  $\log(y)$  satisfazem MQO mais do que os modelos que usam o nível original de  $y$ .
- Log é útil para variáveis estritamente positivas com grandes valores e distribuição concentrada, tais como: renda, vendas de empresas, população, matrículas, empregados, votação.
- Log estreita amplitude dos valores, tornando estimativas menos sensíveis a observações extremas (*outliers*).
- Variáveis medidas em anos aparecem em forma original.
- Taxas geralmente aparecem em forma original.
- Log não é usado se variável tem valor zero ou negativo.
- Não é válido comparar  $R^2$  entre modelos com  $y$  e  $\log(y)$ .

## EXEMPLO DE NÃO-LINEARIDADE

- Para cada ano adicional de educação, há um aumento fixo no salário. Esse é o aumento tanto para o primeiro ano de educação quanto para anos mais avançados:

$$\textit{salário} = \beta_0 + \beta_1 \textit{educ} + u$$

- Suponha que o aumento percentual no salário é o mesmo, dado um ano a mais de educação formal. Um modelo que gera um efeito percentual constante é dado por:

$$\log(\textit{salário}) = \beta_0 + \beta_1 \textit{educ} + u$$

- Se  $\Delta u = 0$ , então:

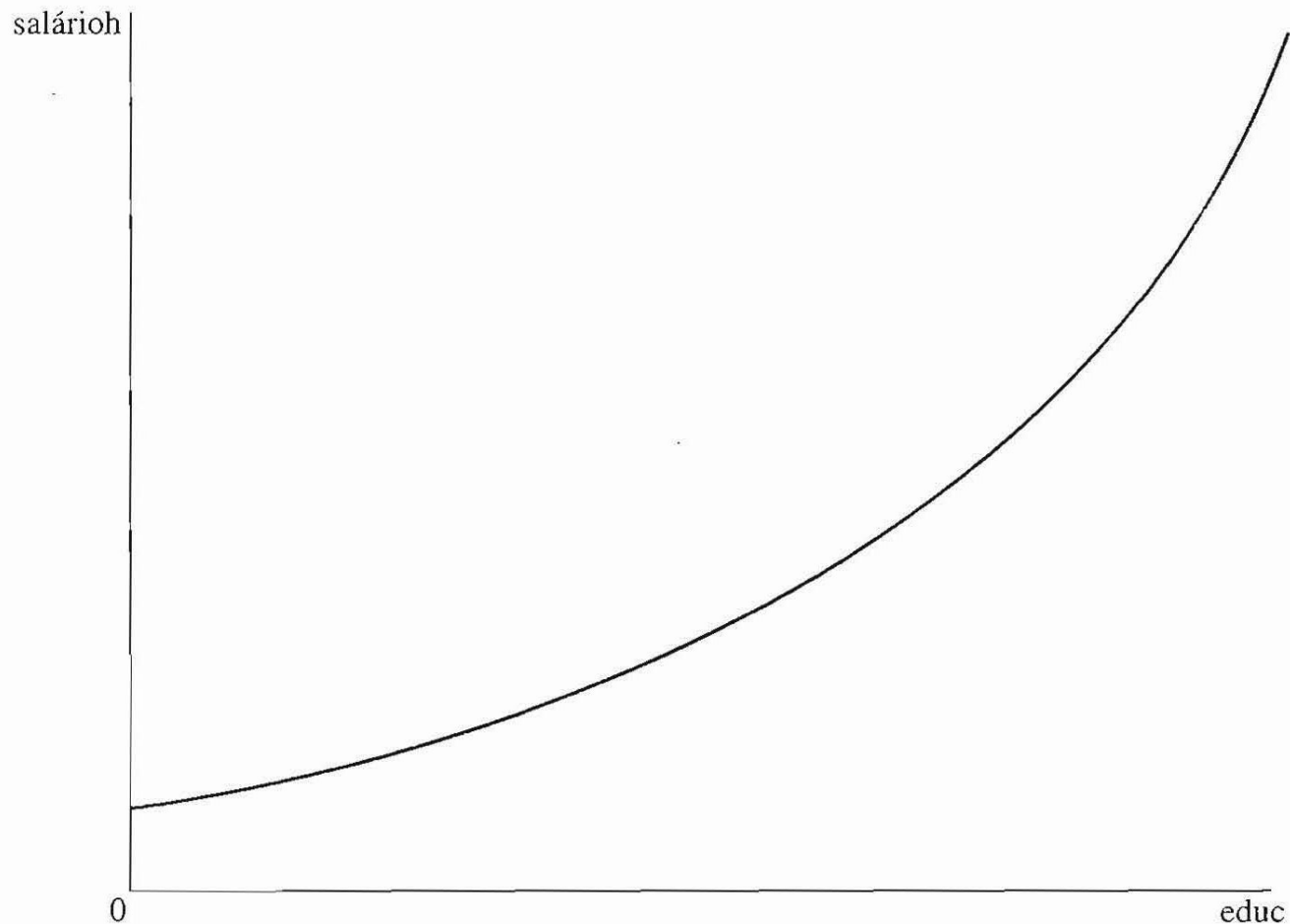
$$\% \Delta \textit{salário} = (100 * \beta_1) \Delta \textit{educ}$$

- Para cada ano adicional de educação, há um aumento de ?% sobre o salário.

- Como a variação percentual no salário é a mesma para cada ano adicional de educação, a variação no salário aumenta quando a educação formal aumenta.

**Figura 2.6**

$$\text{saláριο} = \exp(\beta_0 + \beta_1 \text{educ}), \text{ com } \beta_1 > 0.$$





## INTERPRETAÇÃO DOS COEFICIENTES

- Aumento de uma unidade em  $x$  aumenta  $y$  em  $\beta_1$  unidades:

$$y = \beta_0 + \beta_1 x + u$$

- Aumento de 1% em  $x$  aumenta  $y$  em  $(\beta_1/100)$  unidades:

$$y = \beta_0 + \beta_1 \log(x) + u$$

- Aumento de uma unidade em  $x$  aumenta  $y$  em  $(100*\beta_1)\%$ :

$$\log(y) = \beta_0 + \beta_1 x + u$$

- Aumento de 1% em  $x$  aumenta  $y$  em  $\beta_1\%$ :

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

- Este último é o modelo de elasticidade constante.
- Elasticidade é a razão entre o percentual de mudança em uma variável e o percentual de mudança em outra variável.

# FORMAS FUNCIONAIS ENVOLVENDO LOGARITMOS

Modelo	Variável Dependente	Variável Independente	Interpretação de $\beta_1$
nível-nível	y	x	$\Delta y = \beta_1 \Delta x$
nível-log	y	$\log(x)$	$\Delta y = (\beta_1 / 100) \% \Delta x$
log-nível	$\log(y)$	x	$\% \Delta y = (100 \beta_1) \Delta x$
log-log	$\log(y)$	$\log(x)$	$\% \Delta y = \beta_1 \% \Delta x$

# MODELOS COM FUNÇÕES QUADRÁTICAS

- Funções quadráticas são usadas para capturar efeitos marginais crescentes ou decrescentes.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

- Sempre existe um valor positivo de  $x$ , no qual o efeito de  $x$  sobre  $y$  é zero, chamado de ponto crítico:  $x^* = |\beta_1 / (2\beta_2)|$ .
- Interpretações: (1) após ponto crítico, a relação se inverte; (2) após/antes ponto crítico, há poucos casos; (3) falta incluir variáveis; ou (4) falta transformar variáveis.
- Quando o coeficiente de  $x$  é positivo e o coeficiente de  $x^2$  é negativo, a função quadrática tem um formato parabólico ( $\cap$ ):
  - Antes desse ponto,  $x$  tem um efeito positivo sobre  $y$ .
  - Após esse ponto,  $x$  tem um efeito negativo sobre  $y$ .
- Se  $\beta_1$  é negativo e  $\beta_2$  é positivo, função tem formato U.

## MODELOS COM TERMOS DE INTERAÇÃO

- O efeito de uma variável independente, sobre a variável dependente, pode depender de outra variável explicativa:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

- O efeito parcial de  $x_2$  sobre  $y$  é:  $\Delta y / \Delta x_2 = \beta_2 + \beta_3 x_1$ .
- $\beta_2$  é o efeito parcial de  $x_2$  sobre  $y$ , quando  $x_1 = 0$ , o que pode não ser de interesse prático.
- Podemos então reparametrizar o modelo, tal como:

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u, \text{ sendo:}$$

$\mu_1$  e  $\mu_2$  médias populacionais de  $x_1$  e  $x_2$ .

- $\delta_2$  é o efeito parcial de  $x_2$  sobre  $y$ , quando  $x_1 = \mu_1$ :

$$\delta_2 = \beta_2 + \beta_3 \mu_1$$

- É complicado interpretar modelos com termos de interação.

# GRAU DE AJUSTE E SELEÇÃO DE REGRESSORES

- Seleção de variáveis explicativas com base no tamanho do  $R^2$  pode levar a modelos absurdos.
- Nada nas hipóteses do modelo linear clássico exige que o  $R^2$  esteja acima de qualquer valor em particular.
- O  $R^2$  é simplesmente uma estimativa do quanto da variação em  $y$  é explicado por  $x_1, x_2, \dots, x_k$  na população.
- Modelos com  $R^2$  pequenos significam que não incluímos fatores importantes, mas não necessariamente significam que fatores em  $u$  estão correlacionados com os  $x$ 's.
- O tamanho de  $R^2$  não tem influência sobre a média dos resíduos ser igual a zero.
- $R^2$  pequeno sugere que variância do erro é grande em relação à variância de  $y$ , mas isso pode ser compensado por amostra grande.

## R<sup>2</sup> AJUSTADO

- Sendo  $\sigma_y^2$  a variância populacional de  $y$  e  $\sigma_u^2$  a variância populacional do erro,  $R^2$  da população é a proporção da variação em  $y$  na população, explicada pelas independentes:

$$R^2 = 1 - \sigma_u^2 / \sigma_y^2$$

- $R^2$  usual =  $SQE/SQT = 1 - SQR/SQT = 1 - (SQR/n) / (SQT/n)$
- Podemos substituir o  $SQR/n$  e  $SQT/n$ , por termos não-viesados de  $\sigma_u^2$  e  $\sigma_y^2$ , e chegamos ao  $R^2$  ajustado:

$$\begin{aligned} \bar{R}^2 &= 1 - [SQR/(n-k-1)] / [SQT/(n-1)] = 1 - \hat{\sigma}^2 / [SQT/(n-1)] \\ &= 1 - (1 - R^2)(n - 1)/(n - k - 1) \end{aligned}$$

- $R^2$  ajustado não corrige viés de  $R^2$  na estimativa do  $R^2$  da população, mas penaliza inclusão de independentes.
- $R^2$  ajustado negativo indica adaptação ruim do modelo, relativo ao número de graus de liberdade.

## $\bar{R}^2$ NA ESCOLHA DE MODELOS NÃO-ANINHADOS

- O  $R^2$  ajustado auxilia na escolha de modelo sem variáveis independentes redundantes (entre modelos não-aninhados).
- A estatística F (*test*) permite testar somente modelos aninhados.
- No exemplo do World Values Survey, podemos testar se modelo com informação se religião é muito importante (religiao) é melhor do que modelo com crença no céu (ceu):  
$$\text{tradrat5} = \beta_0 + \beta_1 \text{homem} + \beta_2 \text{religiao} + u$$
$$\text{tradrat5} = \beta_0 + \beta_1 \text{homem} + \beta_2 \text{ceu} + u$$
- Neste caso, não queremos incluir as duas variáveis em conjunto, pois teoricamente medem a mesma dimensão.
- Estes são modelos não-aninhados, exigindo comparação do  $R^2$  ajustado.

# $\bar{R}^2$ E MODELOS COM DIFERENTES FORMAS FUNCIONAIS

- Comparação dos  $R^2$  ajustados pode ser feita para optar entre modelos com formas funcionais diferentes das variáveis independentes:

$$y = \beta_0 + \beta_1 \log(x) + u$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

- Não podemos usar nem o  $R^2$  nem o  $R^2$  ajustado para escolher entre modelos não-aninhados com diferentes formas funcionais da variável dependente.
- Os  $R^2$  medem a proporção explicada do total da variação de qualquer variável dependente:
  - Portanto, diferentes funções da variável dependente terão diferentes montantes de variação a serem explicados.



# CONTROLE DE MUITOS FATORES NA REGRESSÃO

- Estamos preocupados com omissão de fatores importantes que possam estar correlacionados com as variáveis independentes.
- Se enfatizarmos o  $R^2$ , tenderemos a controlar fatores em um modelo que não deveriam ser controlados.
- Ao estudar o efeito da qualidade do ensino sobre a renda, talvez não faça sentido controlar os anos de escolaridade, pois subestimar o retorno da qualidade. Podemos estimar a equação com e sem anos de estudo.
- A questão de decidir se devemos ou não controlar certos fatores nem sempre é bem definida.
- Se nos concentrarmos na interpretação *ceteris paribus* da regressão, não incluiremos fatores no modelo, mesmo que estejam correlacionadas com a dependente.

## ADIÇÃO DE FATORES: REDUZIR VARIÂNCIA DO ERRO

- A adição de uma nova variável independente pode aumentar o problema da multicolinearidade.
- Porém, ao adicionar uma variável, estamos reduzindo a variância do erro.
- Devemos incluir variáveis independentes que afetem  $y$  e que sejam não-correlacionadas com todas variáveis independentes, pois:
  - Não induzirá multicolinearidade.
  - Reduzirá variância do erro.
  - Diminuirá erros-padrão dos coeficientes beta, gerando estimativas mais precisas (estimador com menor variância do erro amostral).

## VALORES ESTIMADOS E RESÍDUOS

- Encontrados o intercepto e a inclinação, teremos um valor estimado para  $y$  para cada observação ( $x$ ) na amostra:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- O resíduo é a diferença entre o valor verdadeiro de  $y_i$  e seu valor estimado:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

# ANÁLISE DE RESÍDUOS

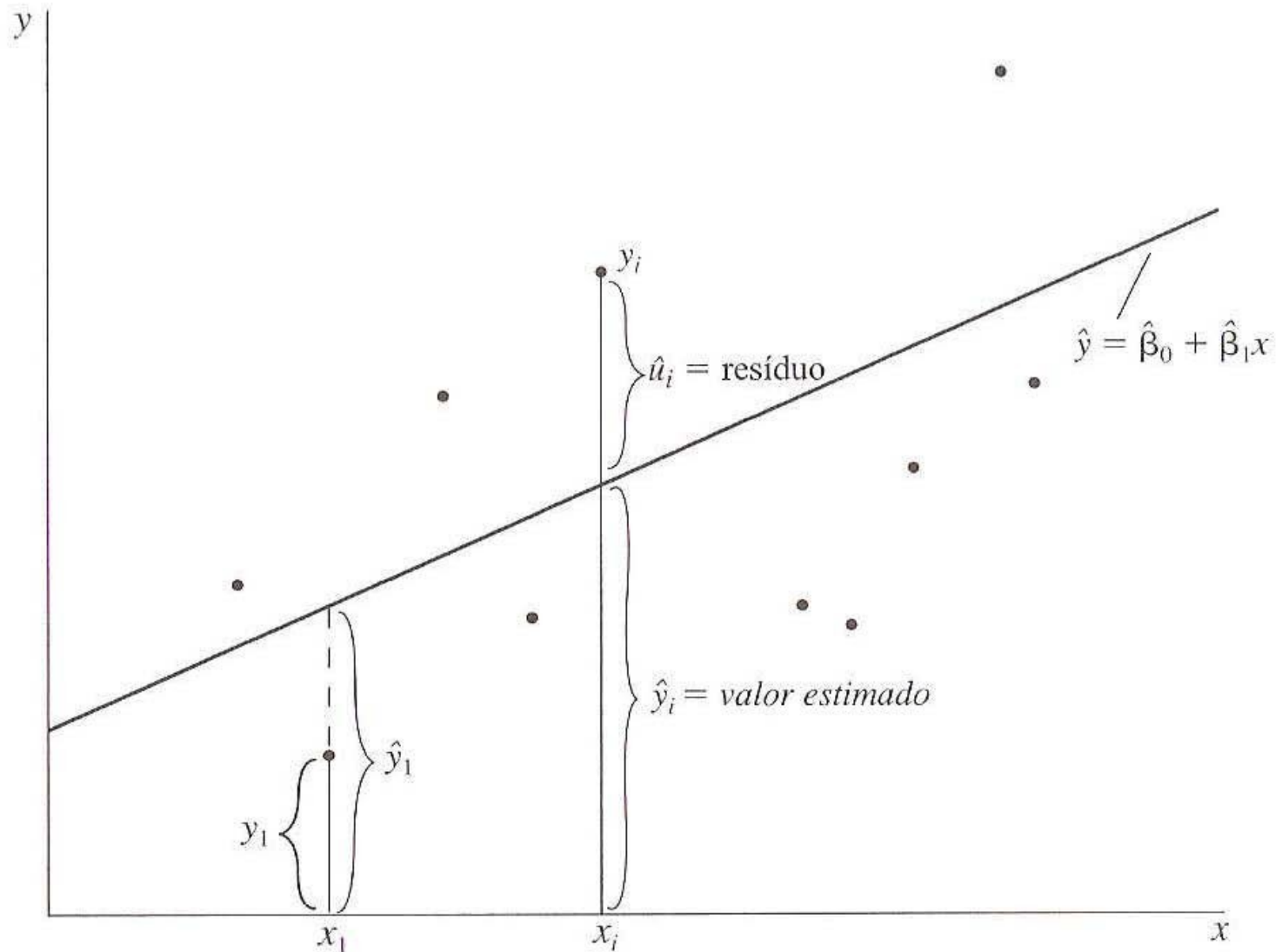
- É importante analisar os resíduos das observações individuais e examinar se valor efetivo da variável dependente está acima ou abaixo do valor previsto:

$$\hat{u}_i = y_i - \hat{y}_i$$

- Resíduo mais negativo indica valor observado mais baixo do que o previsto na regressão e vice-versa.

**Figura 2.4**

Valores estimados e resíduos.



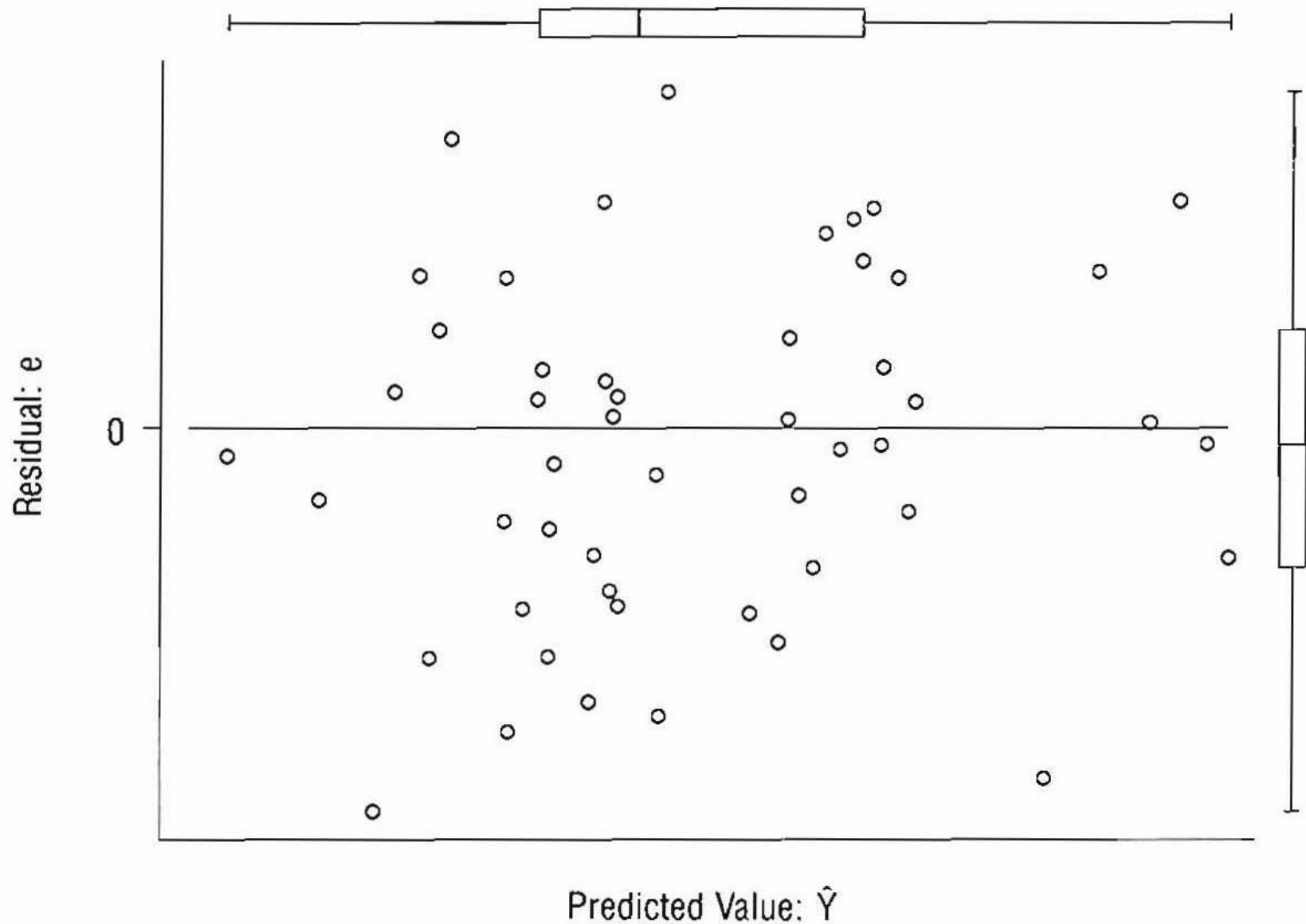
## MINIMIZANDO A SOMA DOS RESÍDUOS QUADRADOS

- Suponha que escolhemos o intercepto e a inclinação estimados com o propósito de tornar a soma dos resíduos quadrados:

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

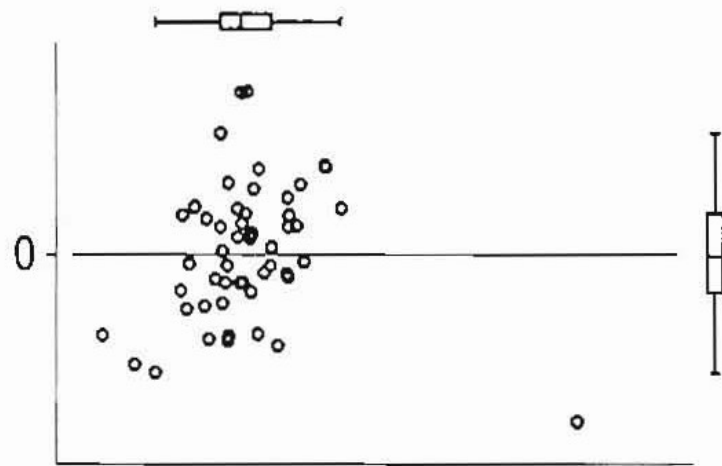
- O nome “mínimos quadrados ordinários” é utilizado porque as estimativas do intercepto e da inclinação minimizam a soma dos resíduos quadrados.
- Não é utilizada a minimização dos valores absolutos dos resíduos, porque a teoria estatística para isto seria muito complicada.

# HOMOSCEDASTICIDADE DOS RESÍDUOS

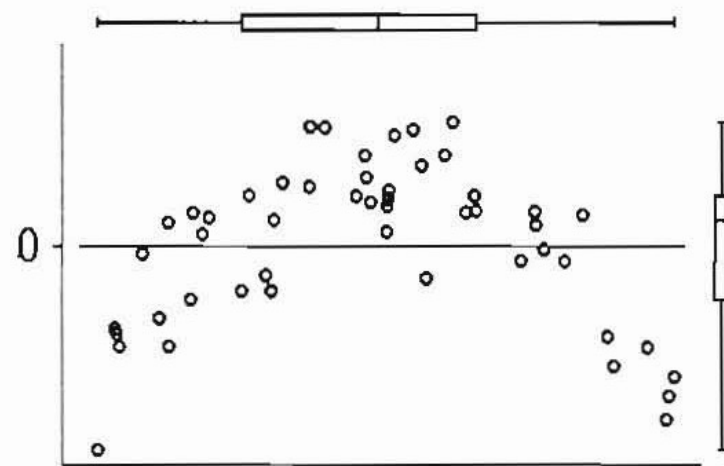


**Figure 2.10** “All clear”  $e$ -versus- $\hat{Y}$  plot (artificial data).

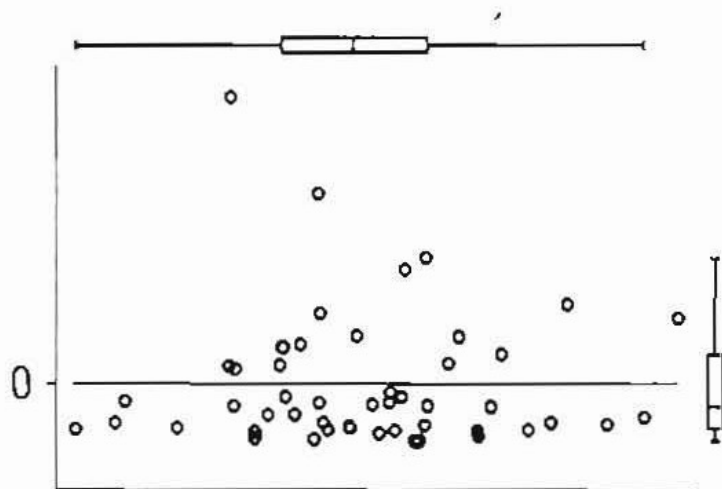
# HETOROCEDASTICIDADE DOS RESÍDUOS



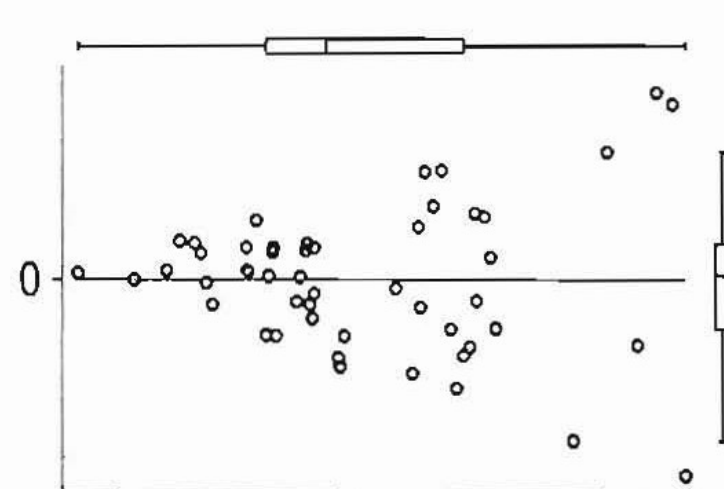
Influential Case



Curvilinear Relation



Nonnormal Residual Distribution



Heteroscedasticity

**Figure 2.11** Examples of trouble seen in  $e$ -versus- $\hat{Y}$  plots (artificial data).



## LEMBRE-SE

- “É muito mais importante tornar-se proficiente em interpretar coeficientes do que eficiente no cálculo de fórmulas.”  
(Wooldridge, 2008: 45)