Econometria Aplicada

Formas funcionais

João Ricardo Costa Filho

“The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

“In God we trust. All others must bring data."

William Edwards Deming

A regressão linear

Motivação (tudo começa com uma pergunta)

Qual é a relação entre capital humano e o desenvolvimento econômico?

Referência

- Como responder a questão que motivou a nossa análise?

Referência

- Como responder a questão que motivou a nossa análise?
    - Utilizemos o capítulo 3 de Wooldridge (2006).

Modelos restritos (aninhados; *nested*)

Podemos remover variáveis dos modelos?

Muitas vezes, não temos evidências estatísticas para rejeitar a hipótese de que, individualmente, os regressores não possuem efeitos nos regressandos.

Podemos remover variáveis dos modelos?

Muitas vezes, não temos evidências estatísticas para rejeitar a hipótese de que, individualmente, os regressores não possuem efeitos nos regressandos. Mas isso significa que podemos excluí-las?

Podemos remover variáveis dos modelos?

Muitas vezes, não temos evidências estatísticas para rejeitar a hipótese de que, individualmente, os regressores não possuem efeitos nos regressandos. Mas isso significa que podemos excluí-las? Não sem antes fazer um teste!

Podemos remover variáveis dos modelos?

Muitas vezes, não temos evidências estatísticas para rejeitar a hipótese de que, individualmente, os regressores não possuem efeitos nos regressandos. Mas isso significa que podemos excluí-las? Não sem antes fazer um teste!

$\mathcal{H}_0 : \beta_{r1} = \beta_{r2} = \cdots = \beta_{rq} = 0$
$\mathcal{H}_a$: pelo menos um $\beta_{ri} \neq 0, i \in [1, q]$

$$F = \frac{SQR_{\text{restrito}} - SQR_{\text{irrestrito}}}{SQR_{\text{irrestrito}} / (n - k - 1)}.$$

Formas funcionais e não-linearidades

Modelo linear

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[Y|X = 0] = \beta_0$

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[Y|X = 0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[Y|X = 0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:
- $\frac{\Delta Y}{\Delta X} = \frac{dY}{dX} = \frac{d(\beta_0 + \beta X)}{dX} = \beta$.

Modelo linear

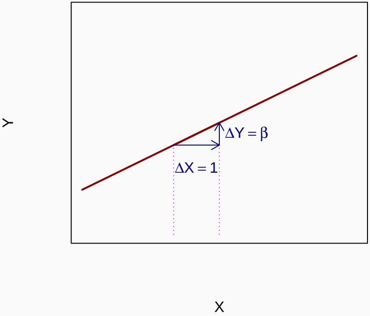
Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[Y|X=0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:
- $\frac{\Delta Y}{\Delta X} = \frac{dY}{dX} = \frac{d(\beta_0 + \beta X)}{dX} = \beta$.
- Note que a variação marginal é constante.

Modelo linear

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[Y|X=0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:
- $\frac{\Delta Y}{\Delta X} = \frac{dY}{dX} = \frac{d(\beta_0 + \beta X)}{dX} = \beta$.
- Note que a variação marginal é constante.



Se a reação entre as variáveis for não-linear, ainda podemos utilizar uma regressão linear?

Termo quadrático

Termo quadrático

Consideremos a seguinte relação entre as variáveis:

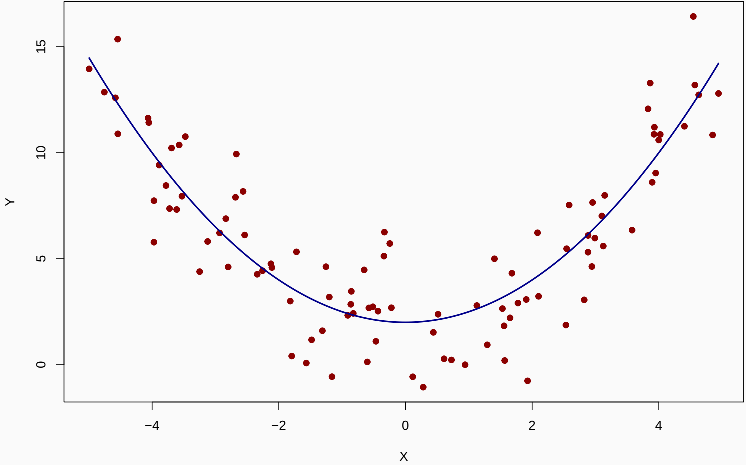
$Y_i = \beta_0 + \beta_1 X_i^2 + \epsilon_i$.

Termo quadrático

Consideremos a seguinte relação entre as variáveis:
$Y_i = \beta_0 + \beta_1 X_i^2 + \epsilon_i$. Como estimá-la?

Termo quadrático

Consideremos a seguinte relação entre as variáveis:
$Y_i = \beta_0 + \beta_1 X_i^2 + \epsilon_i$. Como estimá-la?

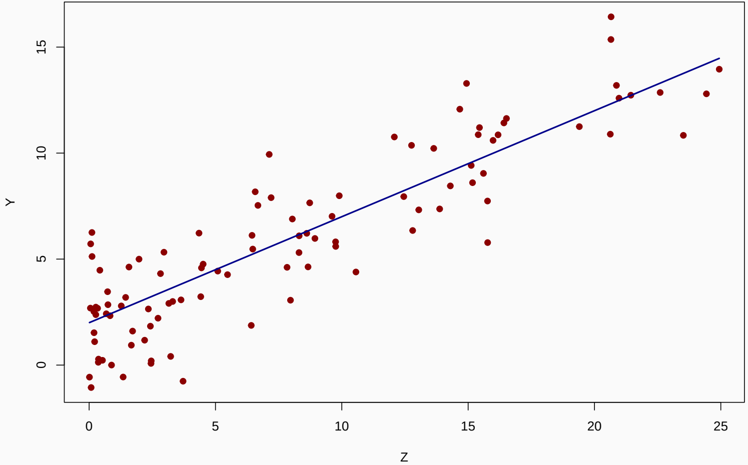


Termo quadrático

Defina $Z = X^2$ de tal forma que $Y_i = \beta_0 + \beta_1 Z_i + \epsilon_i$.

Termo quadrático

Defina $Z = X^2$ de tal forma que $Y_i = \beta_0 + \beta_1 Z_i + \epsilon_i$. Temos, portanto:



(Semi-)Elasticidade: modelo log-linear

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y)|X = 0] = \beta_0 \implies$
  $E[Y|X = 0] = e^{\beta_0}$

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y)|X=0] = \beta_0 \implies$
  $E[Y|X=0] = e^{\beta_0}$
- β_1 representa o efeito marginal
  de X em Y:

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y)|X=0] = \beta_0 \implies$
  $E[Y|X=0] = e^{\beta_0}$
- β_1 representa o efeito marginal de X em Y:

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 X_i + \varepsilon_i}$$

(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[\ln(Y)|X = 0] = \beta_0 \implies E[Y|X = 0] = e^{\beta_0}$
- β_1 representa o efeito marginal de X em Y:

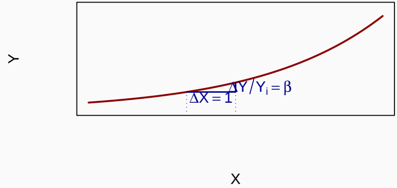
$\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 X_i + \varepsilon_i}$

$\frac{dY_i}{dX_i} = \beta_1 e^{ln(Y_i)} = \beta_1 Y_i \iff \frac{dY_i}{Y_i} = \beta_1 dX_i \approx \Delta\% Y_i$.

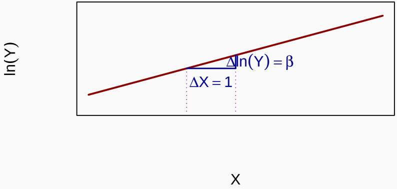
(Semi-)Elasticidade: modelo log-linear

Gericamente, temos: $\ln\left(Y_i\right) = \beta_0 + \beta_1 X_i + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[\ln(Y)|X = 0] = \beta_0 \implies E[Y|X = 0] = e^{\beta_0}$
- β_1 representa o efeito marginal de X em Y:



$\ln(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 X_i + \varepsilon_i}$



$\frac{dY_i}{dX_i} = \beta_1 e^{ln(Y_i)} = \beta_1 Y_i \iff \frac{dY_i}{Y_i} = \beta_1 dX_i \approx \Delta\% Y_i.$

Elasticidade: modelo log-log

Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y) | \ln(X) = 0] = \beta_0$

Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y) | \ln(X) = 0] = \beta_0$
- β_1 representa o efeito marginal
  de X em Y:

Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

- β_0 é "o ponto de partida":
  $E[\ln(Y) | \ln(X) = 0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 \ln(X_i) + \varepsilon_i}$$

Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[\ln(Y) | \ln(X) = 0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:

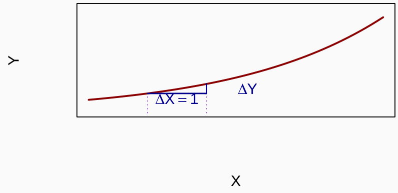
$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 \ln(X_i) + \varepsilon_i}$$

$$\frac{dY_i}{dX_i} = \beta_1 e^{ln(Y_i)} = \beta_1 Y_i \iff \frac{dY_i}{Y_i} = \beta_1 \frac{dX_i}{X_i} \approx \Delta\% Y_i = \beta_1 \Delta\% X_i.$$

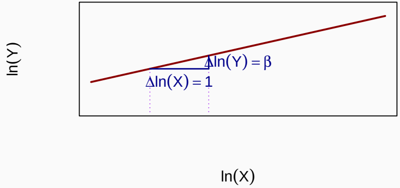
Elasticidade: modelo log-log

Gericamente, temos: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$.

- β_0 é "o ponto de partida": $E[\ln(Y) | \ln(X) = 0] = \beta_0$
- β_1 representa o efeito marginal de X em Y:



$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i \implies Y_i = e^{\beta_0 + \beta_1 \ln(X_i) + \varepsilon_i}$



$\frac{dY_i}{dX_i} = \beta_1 e^{ln(Y_i)} = \beta_1 Y_i \iff \frac{dY_i}{Y_i} = \beta_1 \frac{dX_i}{X_i} \approx \Delta\% Y_i = \beta_1 \Delta\% X_i$.

Será que podemos considerar diferentes pontos de partida?

Será que podemos considerar diferentes pontos de partida?

Gericamente, temos: $Y_i = \beta_0 + \beta_1 X_i + \delta_1 D_i + \varepsilon_i$.

Voltemos ao nosso modelo gravitacional

	Variável dependente:
	Log do Comércio
Ln PIB (Origem)	0.901***
	(0.006)
Ln PIB (Destino)	0.612***
	(0.006)
Ln Distância	−2.373***
	(0.015)
Membro da OMC (Origem)	0.971***
	(0.049)
Membro da OMC (Destino)	0.420***
	(0.050)
Observations	28,159
R^2	0.909
Adjusted R^2	0.909
Nota:	*p<0.1; **p<0.05; ***p<0.01

Interpretação dos coeficientes

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de β_1 unidades em Y.

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de β_1 unidades em Y.
- **Log-linear**: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de β_1 unidades em Y.
- **Log-linear**: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- **Log-Log**: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de β_1 unidades em Y.
- **Log-linear**: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- **Log-Log**: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.
- **Linear com dummy**: $Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, a dummy sair de 0 para 1 está associado à Y aumentar em δ unidades.

Interpretação dos coeficientes

- **Linear (nível)**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de β_1 unidades em Y.
- **Log-linear**: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\beta_1 * 100\%$ em Y.
- **Log-Log**: $\ln[Y_i] = \beta_0 + \beta_1 \ln[X_i] + \varepsilon_i$
  - Em média, o aumento de 1% em X está associado à um aumento de $\beta_1\%$ em Y.
- **Linear com dummy**: $Y_i = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, a dummy sair de 0 para 1 está associado à Y aumentar em δ unidades.
- **Log-linear com dummy**: $\ln[Y_i] = \beta_0 + \beta_1 X_i + \delta D_i + \varepsilon_i$
  - Em média, o aumento de uma unidade em X está associado à um aumento de $\exp[\delta - 1]\%$ em Y.

Referências i

Wooldridge, Jeffrey M. 2006. *Introdução à Econometria: Uma Abordagem Moderna*. Cengage Learning.