Econometria Aplicada

Regressão linear múltipla e 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

O que explica o comércio bilateral entre os países?

A equação de 33 trilhões de dólares

O modelo gravitacional é muito conhecido nos cursos de Economia Internacional e já figura no Capítulo 2 de Krugman, Obstfeld, and Melitz (2023). Na sua forma mais simples, a equação que sintetiza as trocas comerciais bilaterais entre países pode ser escrita como:

$$T_{o,d} = \frac{Y_o \times Y_d}{D_{o,d}}$$

onde $T_{i,j}$ representa o fluxo comercial bilateral entre o país de origem (o) e o país de destino (d), Y_o é o PIB do país de origem, Y_d é o PIB do país de destino e $D_{o,d}$ a distância entre eles.

A equação de 33 trilhões de dólares

O modelo gravitacional é muito conhecido nos cursos de Economia Internacional e já figura no Capítulo 2 de Krugman, Obstfeld, and Melitz (2023). Na sua forma mais simples, a equação que sintetiza as trocas comerciais bilaterais entre países pode ser escrita como:

$$T_{o,d} = \frac{Y_o \times Y_d}{D_{o,d}}$$

onde $T_{i,j}$ representa o fluxo comercial bilateral entre o país de origem (o) e o país de destino (d), Y_o é o PIB do país de origem, Y_d é o PIB do país de destino e $D_{o,d}$ a distância entre eles. Como podemos estimar essa relação não-linear com uma regressão linear?

A equação de 33 trilhões de dólares

O primeiro passo, será escrevê-la de uma maneira um pouco mais genérica:

A equação de 33 trilhões de dólares

O primeiro passo, será escrevê-la de uma maneira um pouco mais genérica:

$$T_{o,d} = \frac{Y_o^{\beta_1} \times Y_d^{\beta_2}}{D_{o,d}^{\beta_3}}$$

Regressão múltipla

Agora, podemos estimar a equação do slide anterior:

Regressão múltipla

Agora, podemos estimar a equação do slide anterior:

$$\ln T_{o,d} = \beta_1 Y_o + \beta_2 Y_d + \beta_3 \ln Dist_{i,j} + \varepsilon_{o,d}$$

Regressão múltipla

	Variável dependente:
	Log do Comércio
Ln PIB (Origem)	0.944^{***}
	(0.006)
Ln PIB (Destino)	0.622^{***}
	(0.006)
Ln Distância	-2.344^{***}
	(0.015)
Observations	28,159
R^2	0.907
Adjusted R^2	0.907
Nota:	*p<0.1; **p<0.05; ***p<0.01

Inferência

A relação entre o PIB do país de origem e o fluxo de comércio bilateral é estatísticamente significativa?

Inferência

A relação entre o PIB do país de origem e o fluxo de comércio bilateral é estatísticamente significativa?

Com a estatística $\hat{\beta}_1$, podemos realizar o seguinte teste:

Inferência

A relação entre o PIB do país de origem e o fluxo de comércio bilateral é estatísticamente significativa?

Com a estatística $\hat{\beta}_1$, podemos realizar o seguinte teste:

$$\mathcal{H}_0 : \beta_1 = \mu$$
$$\mathcal{H}_a : \beta_1 \neq \mu$$

Inferência

E a significância conjunta?

E a significância conjunta? Com a estatística F, podemos realizar o seguinte teste:

Inferência

E a significância conjunta? Com a estatística F, podemos realizar o seguinte teste:

$$\mathcal{H}_0 : \beta_0 = \beta_1 = 0$$
$$\mathcal{H}_a : \beta_j \neq 0, \text{ para } j = [0, 1]$$

Vamos olhar para os resíduos.

Formas funcionais e não-linearidades

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

Termo quadrático

Termo quadrático

Consideremos uma regressão para verificar o retorno da educação na qual há um controle pela experiência.

Termo quadrático

Consideremos uma regressão para verificar o retorno da educação na qual há um controle pela experiência. Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes?

Termo quadrático

Consideremos uma regressão para verificar o retorno da educação na qual há um controle pela experiência. Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes? Como estimar?

Termo quadrático

Consideremos uma regressão para verificar o retorno da educação na qual há um controle pela experiência. Será que a experiência possui retornos marginais constantes, crescentes ou decrescentes? Como estimar?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 Exper_i + \beta_3 Exper_i^2 + \varepsilon_i$$

(Semi-)Elasticidade

Imagine que estamos interessados na semi-elasticidade salário-educação e na elasticidade salário-experiência. Como estimar essas relações?

(Semi-)Elasticidade

Imagine que estamos interessados na semi-elasticidade salário-educação e na elasticidade salário-experiência. Como estimar essas relações?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \varepsilon_i$$

Interpretação dos coeficientes

Quais são os significados de β_1 e β_2 agora?

Interpretação dos coeficientes

Quais são os significados de β_1 e β_2 agora?

- β_1 (educação): Em média, se aumentarmos a **educação** de um(a) trabalhador(a) em 1 ano, o salário aumentará em $\beta_1 * 100\%$ dólares por hora.

Interpretação dos coeficientes

Quais são os significados de β_1 e β_2 agora?

- β_1 (educação): Em média, se aumentarmos a **educação** de um(a) trabalhador(a) em 1 ano, o salário aumentará em $\beta_1 * 100\%$ dólares por hora.
- β_2 (experiência): Em média, se aumentarmos a **experiência** de um(a) trabalhador(a) em 1%, o salário aumentará em $\beta_1\%$.

Dummy

Será que há alguma diferença entre trabalhadores casos e solteiros?

Dummy

Será que há alguma diferença entre trabalhadores casos e solteiros?

$$\ln[salario_i] = \beta_0 + \beta_1 Educ_i + \beta_2 \ln[Exper_i] + \delta D_i + \varepsilon_i$$

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): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - 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.

Vamos para a atividade em grupo!

Referências i

Krugman, Paul, Maurice Obstfeld, and Marc J Melitz. 2023. *Economia Internacional: Teoria e Política*. 12ª edição. Pearson.