

Estadística Aplicada II

▶ Regressão Linear

Aula de hoje

- ▶ **Tópicos**

- ▶ Regressão Linear

- ▶ **Referência**

- ▶ Barrow, M. Estatística para economia, contabilidade e administração. São Paulo: Ática, 2007, Cap. 7 e 8

Aula de hoje

Objetivos:

- ▶ Entender a causalidade entre variáveis através da análise de regressão



Revisão

Covariância

- ▶ Dados n pares de valores $(x_1, y_1), \dots, (x_n, y_n)$, chamaremos de covariância entre as variáveis X e Y , na população:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

- ▶ Para calcular a covariância na amostra, devemos dividir por $n-1$ e não por n
- ▶ É a média dos produtos dos valores centrados das variáveis
- ▶ Tendo esta definição, podemos escrever o coeficiente de correlação como:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{dp(X).dp(Y)}$$

Os resultados são significantes?

➤ $H_0: r = 0$

$H_1: r \neq 0$

➤ A estatística do teste é:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

➤ A qual tem distribuição t com n-2 graus de liberdade

Teste de Hipótese

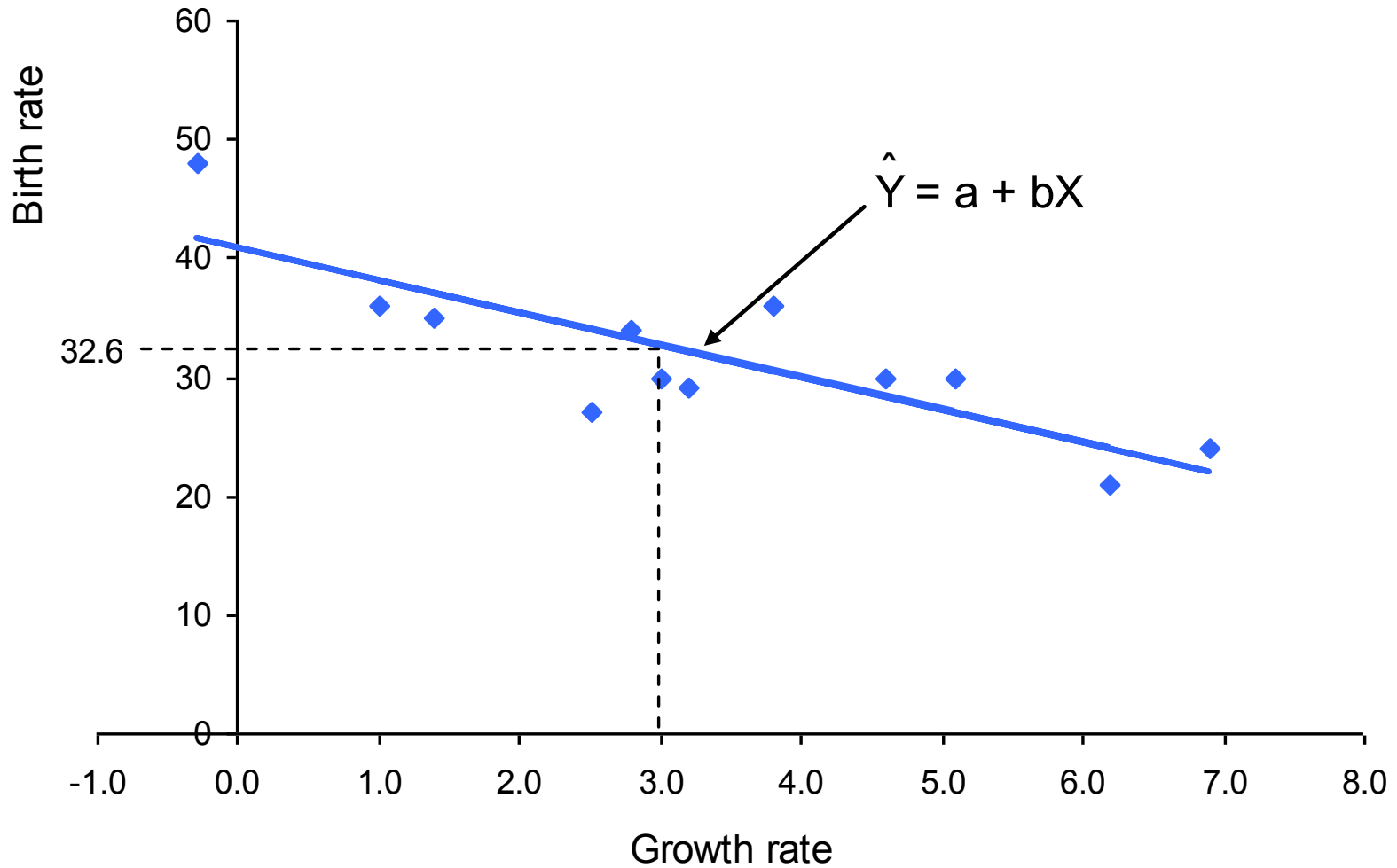
▶ As etapas do teste são:

1. Escrever as hipóteses alternativas e nulas
2. Escolher o nível de significância do teste α
3. Calcular a estatística t , conhecida como a **estatística do teste**
4. Calcular o **valor crítico** do teste t^* ,
5. Decidir: Se o valor absoluto de t for maior do que o de t^* , rejeitar H_0 com um nível de confiança de $1-\alpha$

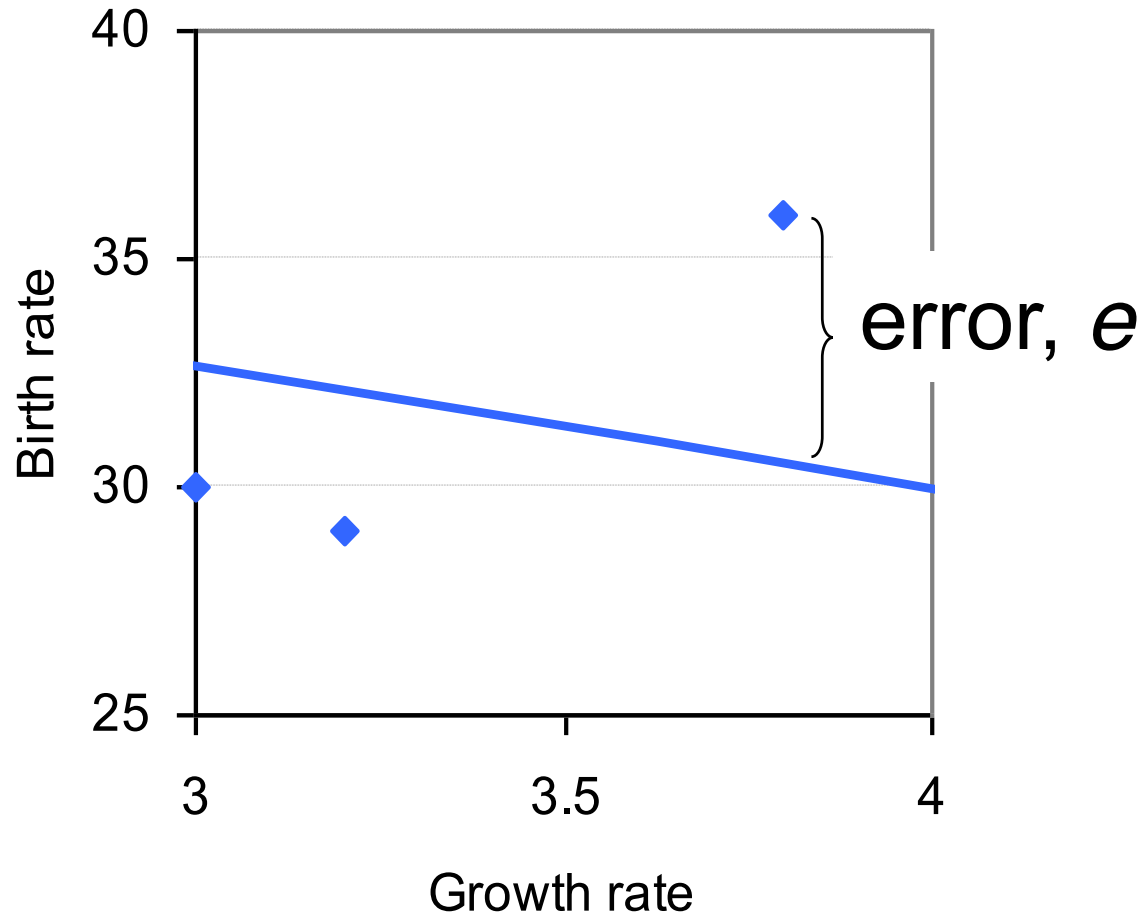
Regressão Linear Simples

- Modelo linear para explicar a variável Y , denominada variável dependente, explicada ou endógena como função da variável X , denominada variável independente, explicativa ou exógena

Regressão Linear Simples



Regressão Linear Simples



Regressão Linear Simples

- A relação entre valor observado de Y e valor previsto de Y pelo modelo é dada por:

$$Y = \hat{Y} + e$$

$$Y = a + bX + e$$

Regressão Linear Simples

➤ Os valores de a e b são dados pela minimização da soma do quadrado dos erros. Tem-se:

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

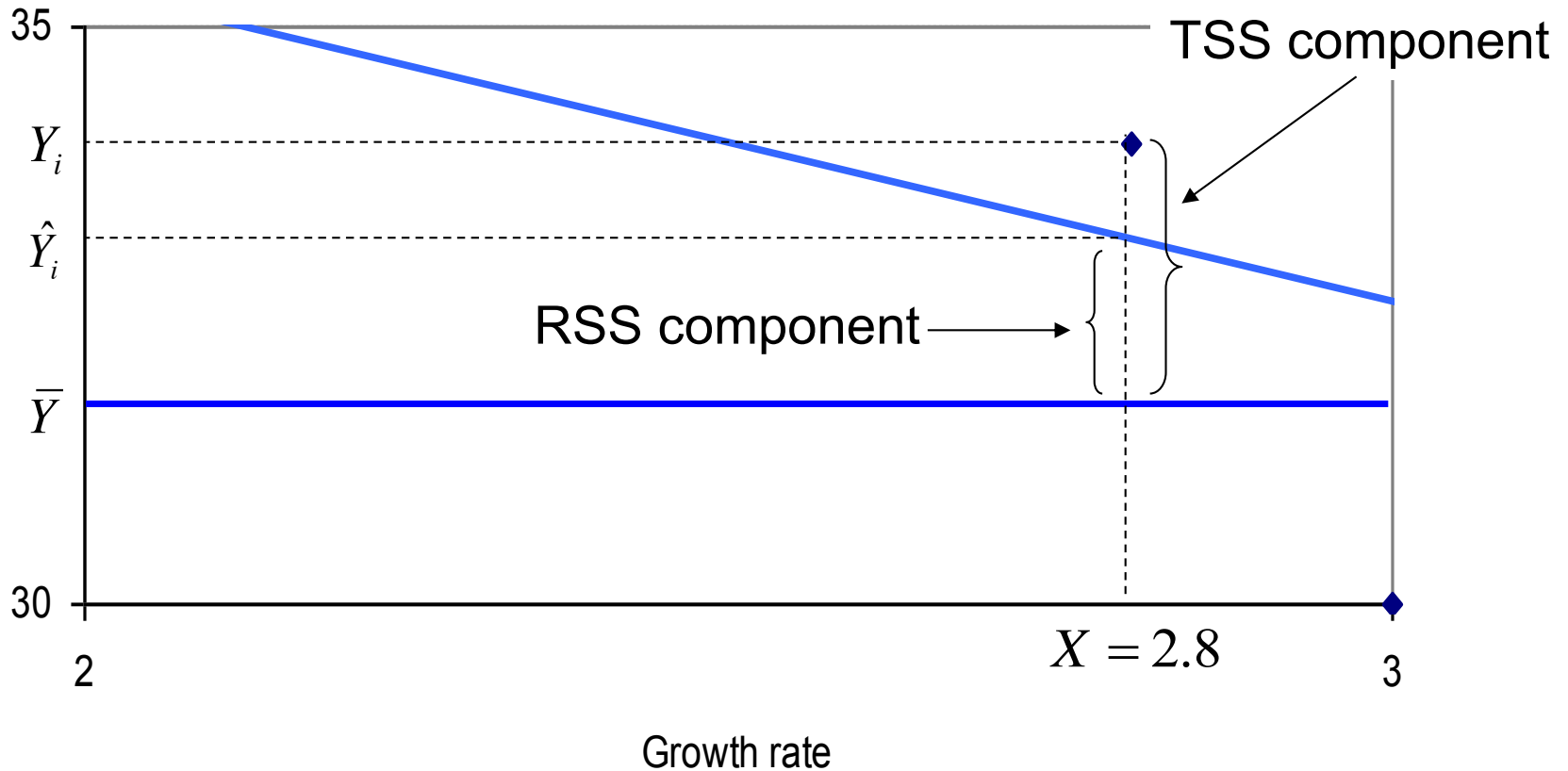
e

$$a = \bar{Y} - b\bar{X}$$

Regressão Linear Simples

Country	Birth rate Y	GNP growth X	Y^2	X^2	XY
Brazil	30	5.1	900	26.01	153.0
Colombia	29	3.2	841	10.24	92.8
Costa Rica	30	3.0	900	9.00	90.0
India	35	1.4	1,225	1.96	49.0
Mexico	36	3.8	1,296	14.44	136.8
Peru	36	1.0	1,296	1.00	36.0
Philippines	34	2.8	1,156	7.84	95.2
Senegal	48	-0.3	2,304	0.09	-14.4
South Korea	24	6.9	576	47.61	165.6
Sri Lanka	27	2.5	729	6.25	67.5
Taiwan	21	6.2	441	38.44	130.2
Thailand	30	4.6	900	21.16	138.0
Total	380	40.2	12,564	184.04	1,139.7

Regressão Linear Simples



Regressão Linear Simples

➤ Mensuração da qualidade do ajuste:

$$R^2 = \frac{RSS}{TSS}$$

Em que:

$$TSS = \sum (Y - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$$

$$ESS = \sum (Y - \hat{Y})^2 = \sum Y^2 - a\sum Y - b\sum XY$$



Inferência

Regressão Linear Simples - Inferência

➤ Considere:

$$Y_i = a + bX_i + e_i$$

- Utilizamos dados amostrais para calcular "a" e "b".
- Os valores "a" e "b" são estimativas dos verdadeiros parâmetros populacionais "alpha" e "beta".

Regressão Linear Simples - Inferência

- Podemos testar hipóteses com relação aos parâmetros populacionais
- Para isso, precisamos calcular o desvio-padrão associado a cada uma delas.
- A variância de "b" é dada por:

$$s_b^2 = \frac{s_e^2}{\sum (X_i - \bar{X})^2}$$

Regressão Linear Simples - Inferência

➤ Em que a variância estimada do erro é dada por:

$$s_e^2 = \frac{\sum e_i^2}{n-2} = \frac{ESS}{n-2}$$

➤ A variância de "a" é dada por:

$$s_a^2 = s_e^2 \times \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)$$

Regressão Linear Simples - Inferência

➤ Estamos prontos para testar hipóteses com relação aos valores dos parâmetros populacionais

➤ Exemplo:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

➤ A estatística do teste é:

$$t = \frac{b - \beta}{s_b} \sim t_{n-2}$$

Regressão Linear Simples - Inferência

- Podemos também criar intervalos de confiança para os parâmetros populacionais.
- Como exercício, crie o intervalo de confiança de 95% para "beta" em nosso exemplo.

Regressão Linear Simples - Inferência

- Podemos também testar se o valor do verdadeiro *R-squared* é igual a zero:

$$H_0: R^2 = 0$$

$$H_1: R^2 > 0$$

- Nesse caso, a estatística do teste é:

$$F = \frac{R^2/1}{(1-R^2)/(n-2)} \sim F_{1,n-2}$$

Regressão Linear Múltipla

Regressão Linear Múltipla

- Se a variável dependente for uma função de k variáveis explicativas, tem-se:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k + e$$

- Como anteriormente, os valores dos parâmetros são estimados através da minimização da soma dos quadrados dos resíduos.

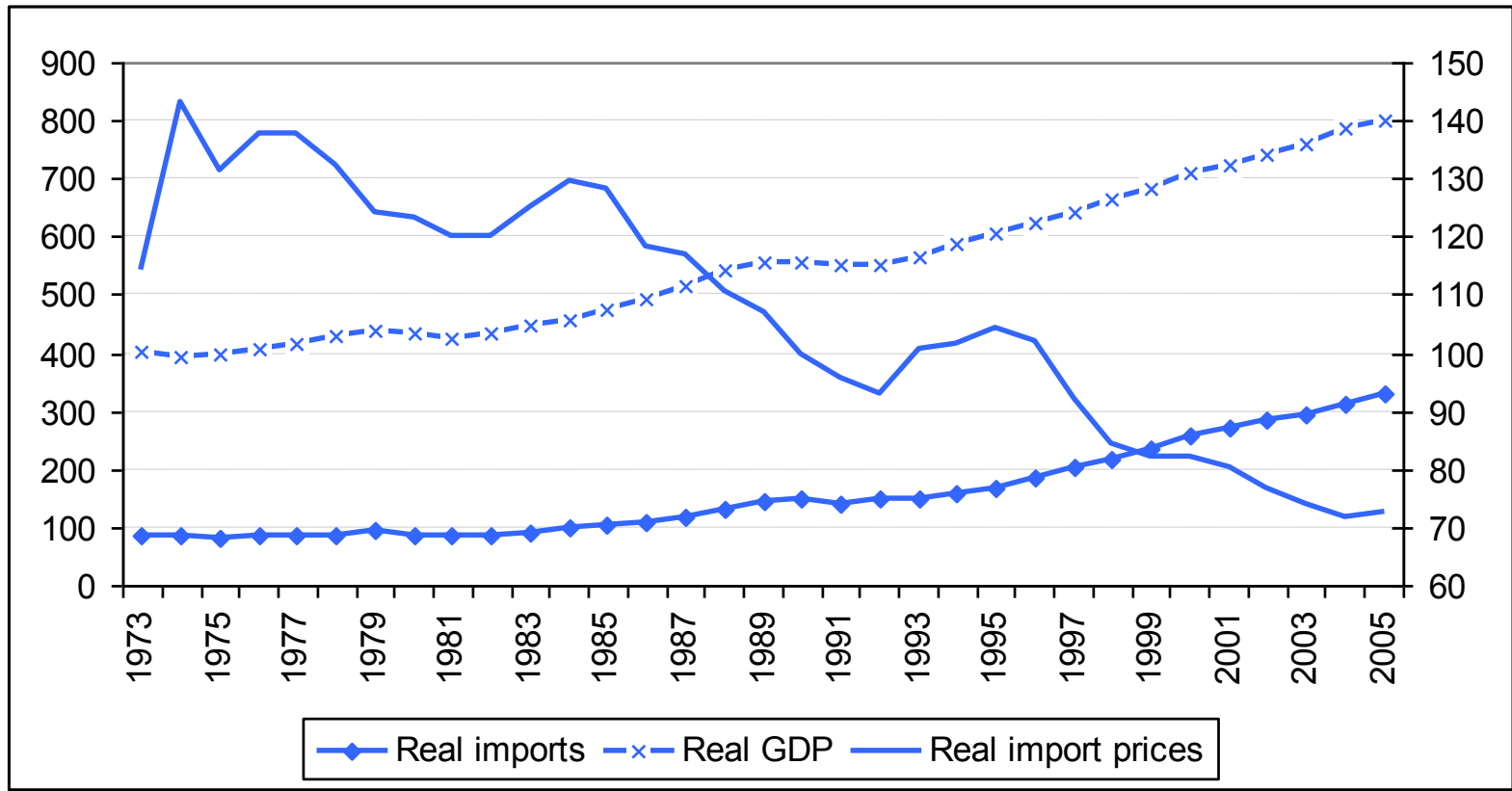
Regressão Linear Múltipla

➤ *Exemplo: Demanda por Importações*

Year	Imports	GDP	GDP deflator	Price of imports	RPI all items
1973	18.8	74.0	24.6	21.5	25.1
1974	27.0	83.8	28.7	31.3	29.1
1975	28.7	105.9	35.7	35.6	36.1
:	:	:	:	:	:
2003	314.8	1110.3	195.6	106.7	191.7
2004	333.7	1176.5	201.0	106.2	197.4
2005	366.5	1224.7	205.4	110.7	202.9

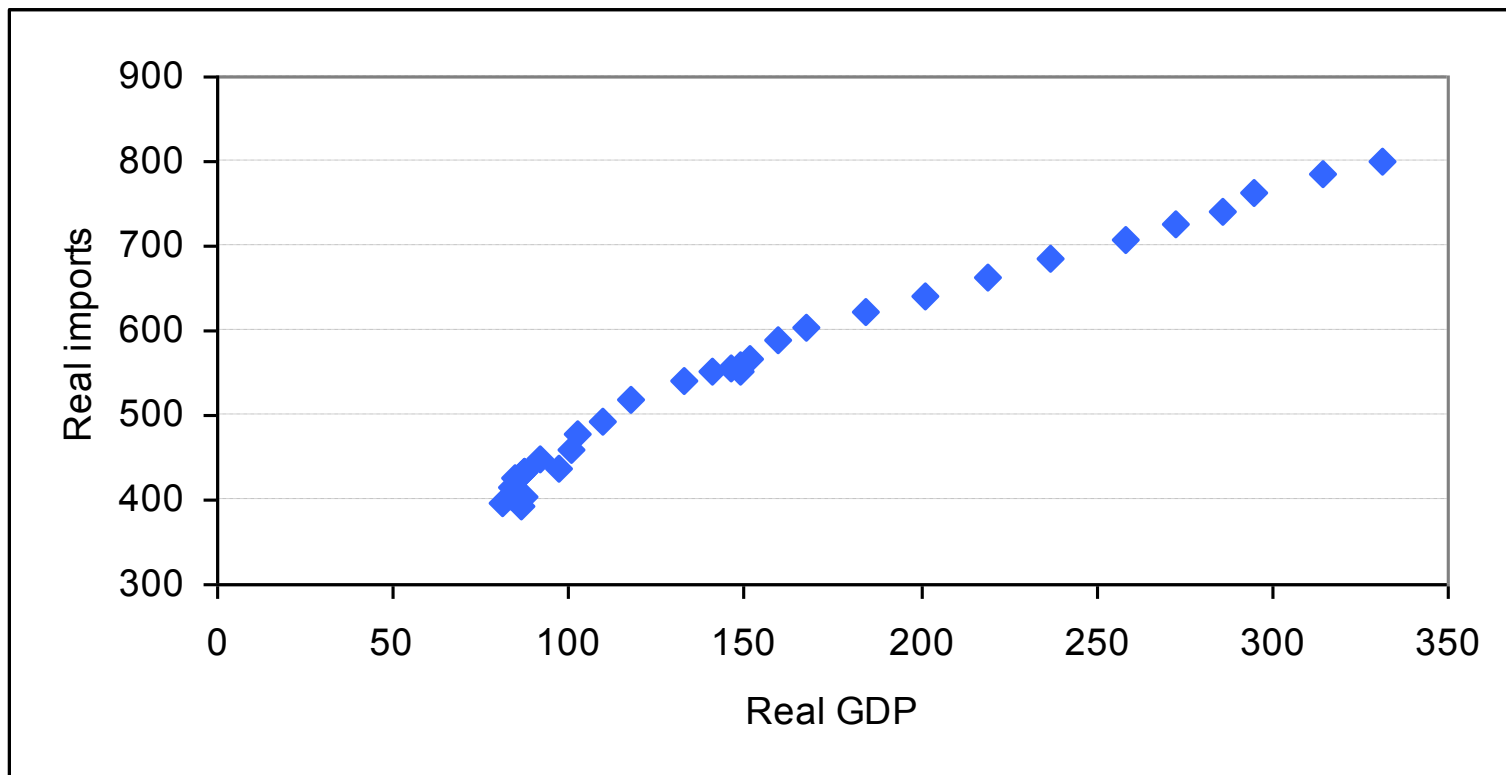
Regressão Linear Múltipla

➤ *Gráfico das séries em termos reais*



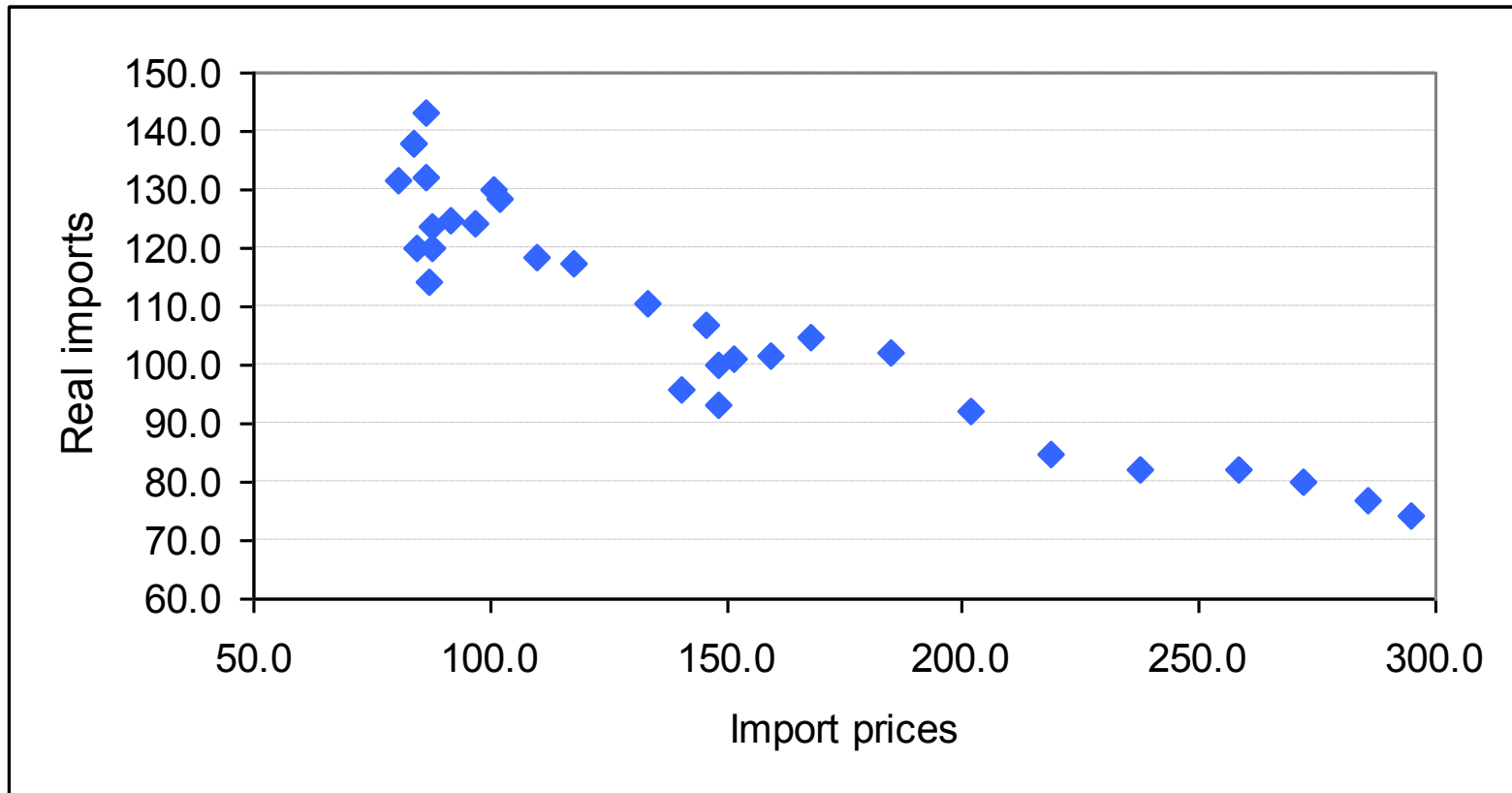
Regressão Linear Múltipla

- *Gráfico das importações e PIB, ambas em termos reais*



Regressão Linear Múltipla

- *Gráfico das importações e preços de importados, ambos em termos reais*



Regressão Linear Múltipla

- Os valores dos parâmetros estimados, bem como as informações para testes de hipóteses calculados no lab estão apresentados a seguir:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-172.61	73.33	-2.35	0.03
Real GDP	0.59	0.06	9.12	0.00
Real import prices	0.05	0.37	0.13	0.90

- Adicionalmente, o *R-squared* reportado foi de 0.96

Regressão Linear Múltipla

- O teste de significância do *R-squared* está resumido na tabela a seguir:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	129031.05	64515.52	368.23	7.82025E-21
Residual	28	4905.70	175.20		
Total	30	133936.75			

Regressão Linear Múltipla

- Estimando-se o modelo em logs, obtem-se:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-3.60	1.65	-2.17	0.04
ln GDP	1.66	0.15	11.31	0.00
ln import prices	-0.41	0.16	-2.56	0.02

- Com coeficiente de determinação de 0.98, significativamente diferente de zero.