

# Econometria 1

Pedro Henrique Rocha Mendes \*

## Prova 1

1)

Ao final da década de 90, houve uma expressiva valorização do Real, e há um debate acerca das consequências para a indústria brasileira. O efeito da valorização sobre o investimento industrial é ambíguo. Por um lado, a valorização exerce influência positiva sobre o investimento, barateando importações de bens de capital. Por outro lado, a valorização exerce influência negativa, reduzindo a rentabilidade das exportações de manufaturados. A teoria, pois, não permite concluir quanto ao efeito líquido destas duas influências, ele pode ser positivo ou negativo. Seja  $I_{i0}$  o valor real do investimento realizado pela  $i$ -ésima empresa industrial antes da valorização e  $I_{i1}$  o valor deste investimento após a valorização. Com base em uma amostra de 81 ( $9^2$ ) empresas industriais, coletada no período a que este enunciado se refere, é possível testar, a um nível de significância de 10%, a hipótese de que a média populacional para a variação do investimento foi nula no período (ou seja, para a empresa média, efeitos positivo e negativo cancelaram-se). O valor observado, na amostra, da média para o indicador de impacto,  $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \delta_i = \frac{1}{N} \sum_{i=1}^N \left( \frac{I_{i1} - I_{i0}}{I_{i0}} \right)$ , é de -0,3, com desvio padrão  $s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\delta_i - \bar{\delta})^2} = 1,5$ . Preencha as lacunas abaixo:

1) a.

Calcule o valor observado da estatística do teste:

```
t <- (-0.3-0) / (1.5/9)
t
```

```
## [1] -1.8
```

---

\*RA: 11201811516, Turno: Manhã

**1) b.**

*Apresente a região crítica do teste e o p-valor do teste:*

Para  $\alpha = 0,1$ :

```
c(qt((0.1/2), 80, lower.tail = TRUE), qt(1-((0.1/2)), 80, lower.tail = TRUE))
```

```
## [1] -1.664125 1.664125
```

```
2*pt(t, 80, lower.tail = TRUE)
```

```
## [1] 0.07563041
```

**1) c.**

*Decida quanto à hipótese nula:*

- $H_0: \theta = 0$
- $H_1: \theta \neq 0$

Como a estatística de teste caiu fora da região crítica, rejeita-se  $H_0$  e admite-se que  $\theta$  é diferente de 0 a um nível de significância de 10%.

**1) d.**

*Explique a implicação da decisão anterior para o debate acerca das consequências para a indústria brasileira:*

Rejeita-se a hipótese de que a média populacional para a variação do investimento foi nula no período, havendo a possibilidade de variações positivas ou negativas na taxa de investimento pós-valorização da moeda.

**1) e.**

*Informe os limites inferior e superior do intervalo com nível de confiança de 90% para o indicador de impacto.*

```
t <- qt(1-((0.1/2)), 80, lower.tail = F)
```

```
c((-0.3-t*(1.5/9)), (-0.3+t*(1.5/9)))
```

```
## [1] -0.0226459 -0.5773541
```

2)

Em cada um dos quatro itens a seguir há pelo menos um dos passos lógicos que compõem o procedimento de simplificação da variância do estimador de mínimos quadrados ordinários para o coeficiente de uma função de regressão populacional. Selecione a alternativa que descreve corretamente a definição, propriedade ou hipótese que justifica o passo lógico indicado pelo símbolo (?). Justifique sua resposta.

2) a.

$$\text{Var}(\hat{\beta}_1|X) = E\left\{\left[\hat{\beta}_1 - E(\hat{\beta}_1|X)\right]^2 \mid X\right\} = (?) = E\left[\left(\hat{\beta}_1 - \beta_1\right)^2 \mid X\right]$$

- ☐ (A) Propriedade de ausência de viés do estimador para o intercepto
- ☒ (B) Propriedade de ausência de viés do estimador para o coeficiente
- ☐ (C) Definição do estimador

Estimador de  $\beta_1$ :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

$$\begin{aligned} y_i - \bar{y} &= \beta_0 + \beta_1 x_i + u_i - (\beta_0 + \beta_1 \bar{x} + \bar{u}) \\ &= \beta_1 (x_i - \bar{x}) + u_i - \bar{u} \\ &= \beta_1 (x_i - \bar{x}) + \tilde{u}_i \quad \text{onde} \\ \tilde{u}_i &= u_i - \bar{u} \quad (2) \end{aligned}$$

(1) em (2)

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n [\beta_1 (x_i - \bar{x}) + \tilde{u}_i] (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n [\beta_1 (x_i - \bar{x})^2 + \tilde{u}_i (x_i - \bar{x})]}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n \beta_1 (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Viés de  $\beta_1$ :

$$\begin{aligned} B(\hat{\beta}_1|X) &= E\left(\beta_1 + \frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \middle| X\right) - \beta_1 \\ &= \cancel{E(\beta_1|X)} + E\left(\frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \middle| X\right) - \cancel{\beta_1} = E\left(\frac{\sum_{i=1}^n \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \middle| X\right) \\ &= \frac{\sum_{i=1}^n E[\tilde{u}_i (x_i - \bar{x})|X]}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n E(u_i - \bar{u}|X)(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Assumindo  $E(u_i|X) = 0$  (MCRL 4):

$$\begin{aligned} E(u_i - \bar{u}|X) &= E(u_i|X) - E(\bar{u}|X) = \\ &= \cancel{E(u_i|X)} - n^{-1} \cancel{E(u_i|X)} \end{aligned}$$

$$\text{Logo: } B(\hat{\beta}_1|X) = \frac{\sum_{i=1}^n 0 (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0$$

Portanto,  $E(\hat{\beta}_1|X) = \beta_1$ , o que justifica a passagem.

2) b.

$$\text{Var}(\hat{\beta}_1|X) = E[(\hat{\beta}_1 - \beta_1)^2|X] = (?) = E\left\{\left[\frac{\sum_{i=1}^N u_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}\right]^2 \middle| X\right\}$$

- ☐ (A) O estimador do intercepto é aproximadamente equivalente à covariância amostral, exceto pela multiplicação pelo termo de perturbação
- ☒ (B) Fórmula do estimador do coeficiente
- ☐ (C) Ausência de viés para o estimador do coeficiente

Estimador de  $\hat{\beta}_1$ :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta_1$$

$$\begin{aligned}
\text{Var}(\hat{\beta}_1|X) &= E \left\{ \left[ \frac{\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \cancel{\beta_1} - \cancel{\beta_1} \right]^2 \middle| X \right\} \\
&= E \left\{ \left[ \frac{\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^2 \middle| X \right\} \quad \text{MCRL 4} \\
&\quad \bar{u} = E(u_i|X) = 0 \\
&= E \left\{ \left[ \frac{\sum_{i=1}^n u_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^2 \middle| X \right\}
\end{aligned}$$

2) c.

$$\begin{aligned}
\text{Var}(\hat{\beta}_1|X) &= E \left\{ \left[ \frac{\sum_{i=1}^N u_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \right]^2 \middle| X \right\} \\
&= E \left\{ \left[ \sum_{i=1}^N u_i w_i(X) \right]^2 \middle| X \right\} = (?) = E \left[ \sum_{i=1}^N u_i^2 w_i(X)^2 + \sum_{j \neq 1}^N \sum_{i=1}^N u_i w_i(X) u_j w_j(X) \middle| X \right]
\end{aligned}$$

- ☒ (A) Aplicação do quadrado perfeito a um somatório
- ☐ (B) Variância da soma é igual à soma das variâncias
- ☐ (C) Propriedade de linearidade da expectativa

Definindo  $\frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$  como  $w_i(X)$ :

$$\begin{aligned}
\text{Var}(\hat{\beta}_1|X) &= E \left\{ \left[ \sum_{i=1}^n u_i w_i(X) \right]^2 \middle| X \right\} = \\
&= (w_1 u_1 + \dots + w_n u_n) (u_1 w_1 + \dots + u_n w_n) = \\
&= u_1 w_1 (u_1 w_1 + \dots + u_n w_n) + \dots + u_n w_n (u_1 w_1 + \dots + u_n w_n) = \\
&= u_1^2 w_1^2 + u_1 w_1 (u_2 w_2 + \dots + u_n w_n) + \dots + \\
&+ u_n w_n (u_1 w_1 + \dots + u_n w_n) = \\
&= \sum_{i=1}^n u_i^2 w_i^2 + \sum_{j \neq i} \sum_{i=1}^n u_i w_i u_j w_j
\end{aligned}$$

Logo:

$$\text{Var}(\hat{\beta}_1|X) = E \left[ \sum_{i=1}^n u_i^2 w_i^2(X) + \sum_{j \neq i} \sum_{i=1}^n u_i w_i(X) u_j w_j(X) \middle| X \right]$$

2) d.

$$\begin{aligned}
\text{Var}(\hat{\beta}_1|X) &= E \left[ \sum_{i=1}^N u_i^2 w_i(X)^2 + \sum_{j \neq i} \sum_{i=1}^N u_i w_i(X) u_j w_j(X) \middle| X \right] \\
&= (?) = \sum_{i=1}^N \sigma^2 w_i(X)^2
\end{aligned}$$

- ☐ (A) Hipótese de distribuição normal para o termo de perturbação
- ☐ (B) Hipótese de homocedasticidade, apenas
- ☐ (C) Hipótese de homocedasticidade e ausência de autocorrelação

3)

Um pesquisador considera dois procedimentos para prever o padrão de variação da variável remuneração ao longo de uma amostra de trabalhadores:

1. A previsão pela média, ou seja, toma-se por base a média amostral da remuneração, assumindo-se, pois, que o nível salarial individual se mantém satisfatoriamente próximo de tal tendência central;

2. A previsão pela regressão simples, ou seja, toma-se por base uma regressão simples em que a variável explicativa é o nível educacional, assumindo-se, pois, que este exibe padrão de variação satisfatoriamente próximo do observado para o nível de remuneração.

A qualidade da previsão gerada por cada maneira, ou seja, o desempenho de cada maneira, é medida a partir do erro quadrático médio, calculado como segue (Obs.: notar que as medidas são inversamente proporcionais ao grau de desempenho).

$$EQM_{\text{média}} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$EQM_{FRP} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

**3) a.**

Demonstre que o erro quadrático médio cometido pela regressão nunca é maior do que aquele cometido pela média. Ou seja, a primeira pode ter desempenho no mínimo equivalente à segunda, mas nunca inferior. Dica: a regressão atingiria o menor grau de desempenho possível no caso em que a correlação entre educação e remuneração na população e na amostra fosse nula. Tomando por base as fórmulas dos estimadores de MQO para os parâmetros, reconsidere a fórmula no interior do erro quadrático médio para a regressão como base para responder a este item.

**3) b.**

Considere a fórmula do coeficiente de determinação da regressão simples abaixo. Continuando o raciocínio utilizado na resposta ao item anterior, demonstre que a relação entre coeficiente de determinação e coeficiente de correlação entre educação (x) e remuneração (y) é positiva.

$$r^2 = 1 - \frac{EQM_{FRP}}{EQM_{\text{média}}} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

**3) c.**

Outra forma de entender a fórmula do coeficiente de determinação é escrevendo-a como segue:

$$r^2 = 1 - \frac{EQM_{FRP}}{EQM_{\text{média}}} = 1 - \frac{\sum_{i=1}^N [y_i - \hat{E}(y | x = x_i)]^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Em que  $\hat{E}(y | x = x_i) = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$  é a função de regressão amostral, tratando-se, pois, da contrapartida amostral da expectativa condicional. Com base nisso explique por que, havendo correlação considerável entre educação ( $x$ ) e remuneração ( $y$ ), a regressão simples em questão tende a apresentar um melhor desempenho preditivo do que a média incondicional. Sua resposta deve estar embasada na diferença conceitual entre média condicional, tal como se encontra no numerador da expressão, e média incondicional, esta contida no denominador. Leve em conta, adicionalmente, que a média condicional é calculada a partir de dois passos, o primeiro consistindo na seleção de um grupo amostral em função de um determinado nível educacional. Já o segundo compreende o cálculo da média exclusivamente para o grupo selecionado – tal cálculo é repetido para todos os grupos amostrais.

4)

Foi estimada, como parte de uma monografia do Bacharelado em Ciências Econômicas da UFABC, a regressão abaixo em que o déficit de altura para a idade de crianças de cinco anos, residentes em domicílios brasileiros, é explicado em função de características do domicílio e da criança. A variável dependente está medida em centímetros. OBS: nenhuma das variáveis está em forma logarítmica.

Parâmetro	Estimativa pontual
Intercepto	3.4998133
Renda per capita	-0.002594
Anos de estudo do responsável pelo domicílio	-0.054417
Sexo feminino	0.8879468
Idade (em meses)	0.0157118
Acesso à abastecimento hídrico	-0.285436
Área rural	0.2625615
Região Norte	1.3930684
Região Nordeste	0.0366229
Região Centro-Oeste	-0.241068
Região Sudeste	-0.449075

Obs.: foram omitidas as estimativas pontuais para as seguintes variáveis: status do domicílio em relação à conexão à rede de suprimento de eletricidade, à rede de esgoto e acesso a serviços de saúde.



**4) a.**

*Qual é o significado do valor numérico da estimativa pontual para o sexo feminino? Para responder:*

- 1. Apresente a definição formal do coeficiente em questão tomando por base a diferença entre as expectativas condicionais da variável dependente referentes a cada um dos dois grupos sociais em questão, ceteris paribus nas demais variáveis independentes.*
- 2. Com base no passo anterior, escreva, textualmente, a interpretação da magnitude numérica da estimativa pontual.*

**4) b.**

*Qual é o significado do valor numérico da estimativa pontual para a região Norte? Para responder:*

- 1. Apresente a definição formal do coeficiente em questão tomando por base a diferença entre as expectativas condicionais da variável dependente referentes aos grupos regionais em questão, ceteris paribus nas demais variáveis independentes.*
- 2. Com base no passo anterior, escreva, textualmente, a interpretação da magnitude numérica da estimativa pontual.*

**4) c.**

*Agora, tomando o coeficiente da renda domiciliar per capita, uma variável medida em Reais (R\$), explique o significado da estimativa pontual correspondente. Continue a tomar por base uma comparação entre dois grupos sociais equivalentes em todas as demais variáveis.*