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 $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\delta_i - \bar{\delta}\right)^2}$ = 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), qt(1-((0.1/2)), 80))
```

```
## [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 dentro da região crítica, rejeita-se H_0 e admite-se que θ 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.

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

- ☐ *(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 β_1:

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

$y_i - \bar{y} = \beta_0 + \beta_1 x_i + u_i - (\beta_0 + \beta_1 \bar{x} + \bar{u})$

$= \beta_1 (x_1 - \bar{x}) + u_i - \bar{u}$

$= \beta_1 (x_i - \bar{x}) + \tilde{u}_i$ onde $\tilde{u}_i = u_i - \bar{u}$ (2)

(1) em (2)

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

$$= \frac{\sum_{i=1}^{n} \beta_1 (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} + \frac{\sum_{i=1}^{n} \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \beta_1 \frac{\cancel{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{\cancel{\sum_{i=1}^{n} (x_i - \bar{x})^2}} + \frac{\sum_{i=1}^{n} \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n} \tilde{u}_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Viés de β_1:

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

$$= E(\cancel{\beta_1|X}) + E\left(\frac{\sum_{i=1}^{n} \tilde{u}_i (x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}\middle|X\right) - \cancel{\beta_1} = E\left(\frac{\sum_{i=1}^{n} \tilde{u}_i (x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}\middle|X\right)$$

$$= \frac{\sum_{i=1}^{n} E[\tilde{u}_i (x_i-\bar{x})|X]}{\sum_{i=1}^{n} (x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n} E(u_i-\bar{u}|X)(x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$$

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

$$E(u_i-\bar{u}|X) = E(u_i|X) - E(\bar{u}|X) =$$

$$= E(\cancel{u_i|X}) - n^{-1} E(\cancel{u_i}|X)$$

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.

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

- ☐ *(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 $\hat{\beta}_1$:

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

$$Var(\hat{\beta}_1|X) = E\left\{\left[\frac{\sum_{i=1}^{n}(U_i-\bar{U})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}+\cancel{\beta_1}-\cancel{\beta_1}\right]^2 \middle| X\right\}$$

$$= E\left\{\left[\frac{\sum_{i=1}^{n}(U_i-\cancel{\bar{U}})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]^2 \middle| X\right\} \qquad (MCRL4)\ \bar{U} = E(U_i|X) = 0$$

$$= E\left\{\left[\frac{\sum_{i=1}^{n}U_i(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]^2 \middle| X\right\}$$

2) c.

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

$$= E\left\{\left[\sum_{i=1}^{N}u_i w_i(X)\right]^2 \middle| X\right\} = (?) = E\left[\sum_{i=1}^{N}u_i^2 w_i(X)^2 + \sum_{j\neq 1}^{N}\sum_{i=1}^{N}u_i w_i(X) u_j w_j(X) \middle| X\right]$$

☒ *(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 $\dfrac{(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$ como $W_i(X)$:

$$Var(\hat{\beta}_1|X) = E\left\{\left[\sum_{i=1}^{n} u_i w_i(X)\right]^2 \Big| X\right\} =$$

$$= (w_1 u_1 + \dots + w_n u_n)(u_1 w_1 + \dots + u_n w_n) =$$

$$= u_1 w_1 (u_1 w_1 + \dots + u_n w_n) + \dots + u_n w_n (u_1 w_1 + \dots + u_n w_n) =$$

$$= u_1 w_1^2 + u_1 w_1 (u_1 w_1 + \dots + u_n w_n) + \dots + u_n w_n^2 + u_n w_n (u_1 w_1 + \dots + u_n w_n) =$$

$$= \sum_{i=1}^{n} u_i^2 w_i^2 + \sum_{j \neq 1}^{n} \sum_{i=1}^{n} u_i w_i \, u_j w_j$$

Logo:

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

2) d.

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

☐ *(A) Hipótese de distribuição normal para o termo de perturbação*

☐ *(B) Hipótese de homocedasticidade, apenas*

☒ *(C) Hipótese de homocedasticidade e ausência de autocorrelação*

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

$$= \underbrace{\sum_{i=1}^{n} E(U_i^2|X)\, w_i(X)^2}_{\substack{\text{Hipótese da homo-}\\ \text{cedasticidade:}\\ E(U_i^2|X) = \sigma^2\\ \text{para todas as}\\ \text{obs.}}} + \underbrace{\sum_{j\neq 1}^{n} \sum_{i=1}^{n} \cancel{E(U_i U_j|X)\, w_i(X) w_j(X)}}_{\substack{\text{Hipótese da ausência}\\ \text{de autocorrelação:}\\ Cov(U_i, U_j) = 0}}$$

$$\text{Logo, } Var(\hat{\beta}_1|X) = \sum_{i=1}^{n} \sigma^2 w_i(X)^2$$

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).

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

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

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.

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

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

$$= \frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - 2\bar{y}\sum_{i=1}^{N} y_i + n\bar{y}^2\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - \cancel{2n}\bar{y}^2 \cancel{+n}\bar{y}^2\right) = \frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - n\bar{y}^2\right)$$

$$= \hat{\sigma}_y^2$$

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

$$= \frac{1}{N}\sum_{i=1}^{n}\left[y_i^2 - 2y_i\hat{\beta}_0 - 2y_i\hat{\beta}_1 x_i + \hat{\beta}_0^2 + 2\hat{\beta}_0\hat{\beta}_1 x_i + (\hat{\beta}_1 x_i)^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{n}\left[y_i^2 - 2y_i(\bar{y} - \hat{\beta}_1\bar{x}) - 2y_i\hat{\beta}_1 x_i + (\bar{y} - \hat{\beta}_1\bar{x})^2 + 2(\bar{y} - \hat{\beta}_1\bar{x})\hat{\beta}_1 x_i + (\hat{\beta}_1 x_i)^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{n}\left[y_i^2 - 2y_i\bar{y} + 2y_i\hat{\beta}_1\bar{x} - 2y_i\hat{\beta}_1 x_i + \bar{y}^2 - 2\bar{y}\bar{x}\hat{\beta}_1 + (\hat{\beta}_1\bar{x})^2 + 2\bar{y}\hat{\beta}_1 x_i - 2\hat{\beta}_1\bar{x}\hat{\beta}_1 x_i + (\hat{\beta}_1 x_i)^2\right]$$

$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + 2\hat{\beta}_1\bar{x}\sum_{i=1}^{n} y_i - 2\hat{\beta}_1\sum_{i=1}^{n} x_i y_i + n\bar{y}^2 - n2\bar{y}\bar{x}\hat{\beta}_1 + n\hat{\beta}_1^2\bar{x}^2 + 2\bar{y}\hat{\beta}_1\sum_{i=1}^{n} x_i - 2\hat{\beta}_1^2\bar{x}\sum_{i=1}^{n} x_i + \hat{\beta}_1^2\sum_{i=1}^{N} x_i^2\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i^2 - n\cancel{2}\bar{y}^2 \cancel{+ n2\hat{\beta}_1\bar{x}\bar{y}} - 2\hat{\beta}_1\sum_{i=1}^{n} x_i y_i \cancel{+ n\bar{y}^2} \cancel{- n2\bar{x}\bar{y}\hat{\beta}_1} \cancel{+ n\hat{\beta}_1^2\bar{x}^2} + n2\bar{y}\bar{x}\hat{\beta}_1 - n\cancel{2}\hat{\beta}_1^2\bar{x}^2 + \hat{\beta}_1^2\sum_{i=1}^{N} x_i^2\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - 2\hat{\beta}_1\sum_{i=1}^{n} x_i y_i + n2\hat{\beta}_1\bar{x}\bar{y} + \hat{\beta}_1^2\sum_{i=1}^{n} x_i^2 - n\hat{\beta}_1^2\bar{x}^2\right)$$

$$= \hat{\sigma}_y^2 - 2\hat{\beta}_1\hat{\sigma}_{xy} + \hat{\beta}_1^2\hat{\sigma}_x^2$$

Caso 1: EQM$_{\text{FRP}}$ = EQM$_{\text{média}}$ $\leftrightarrow x_i - \bar{x} = 0$, pois,

$$\hat{\beta}_1^2\hat{\sigma}_x^2 = \hat{\beta}^2\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N} = \hat{\beta}^2\frac{0}{N} = 0$$

$$2\hat{\beta}_1\hat{\sigma}_{xy} = 2\hat{\beta}_1\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{N} = 2\hat{\beta}_1\frac{\sum_{i=1}^{N}0(y_i-\bar{y})}{N} = 0$$

Caso 2: EQM$_{\text{FRP}}$ < EQM$_{\text{média}}$ $\leftrightarrow 2\hat{\beta}_1\hat{\sigma}_{xy} - \hat{\beta}_1^2\hat{\sigma}_x^2 > 0$, ou seja,

$$\begin{aligned}
2\hat{\beta}_1\hat{\sigma}_{xy} &> \hat{\beta}_1^2\hat{\sigma}_x^2 \\
2\frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}\hat{\sigma}_{xy} &> \left(\frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}\right)^2\hat{\sigma}_x^2 \\
2\frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x^2} &> \frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x^4}\hat{\sigma}_x^2 \\
2\frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x^2} &> \frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x^2}
\end{aligned}$$

Assim, EQM$_{\text{FRP}}$ < EQM$_{\text{média}}$ $\forall(x_i, y_i)$.

Caso 3: EQM$_{\text{FRP}}$ > EQM$_{\text{média}}$ $\leftrightarrow \hat{\beta}_1^2\hat{\sigma}_x^2 - 2\hat{\beta}_1\hat{\sigma}_{xy} > 0$. No entanto, como provado no caso 2, $\hat{\beta}_1^2\hat{\sigma}_x^2 > 2\hat{\beta}_1\hat{\sigma}_{xy}$ não é um resultado possível.

Logo, EQM$_{\text{FRP}}$ $\leq$ EQM$_{\text{média}}$ $\forall(x_i, y_i)$.

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.

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

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \,, \quad \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - 2\hat{\beta}_1 \sum_{i=1}^{n} x_i y_i + n2\hat{\beta}_1 \bar{x}\bar{y} + \hat{\beta}_1^2 \sum_{i=1}^{n} x_i^2 - n\hat{\beta}_1^2 \bar{x}^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y}^2) + \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - \underbrace{2\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \right)}_{(1)}$$

Sabendo que $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$: $(1)\ 2\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \right) = \underbrace{2\hat{\beta}_1^2 \left(\sum x_i^2 - n\bar{x}^2 \right)}_{(2)}$ pois

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

Substituindo (1) por (2):

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\hat{\beta}_1^2 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right) = \sum_{i=1}^{n} (y_i - \bar{y})^2 + \cancel{\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2} - \cancel{2}\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 =$$

$$= \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}_{(3)}$$

Substituindo (3) em R^2: $R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \cancel{1} - \left[\cancel{1} - \frac{\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \right] =$

$$\frac{\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \left(\frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \right)^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \right]^2 \cancel{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{\sum_{i=1}^{n} (x_i - \bar{x})^{\cancel{4}} \sum_{i=1}^{n} (y_i - \bar{y})^2} =$$

$$= \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$R = \sqrt{\frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$R = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \ \hat{\sigma}_y^2}}$$

Assim, caso $cov(x_i, y_i) > 0$, a relação entre o coeficiente de correlação e o de determinação será positiva.

3) c.

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

$$R^2 = 1 - \frac{EQM_{FRP}}{EQM_{\text{média}}} = 1 - \frac{\sum_{i=1}^{N}\left[y_i - \hat{E}\left(y \mid x = x_i\right)\right]^2}{\sum_{i=1}^{N}\left(y_i - y\right)^2}$$

Em que $\hat{E}\left(y \mid x = x_i\right)$ = $y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ *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.*

A regressão simples apresenta melhor desempenho preditivo pois não assume, havendo correlação entre variável dependente e variável independente, que os valores se distribuirão ao redor da média amostral. Utilizando a função de regressão amostral, ao se calcular a média condicional da variável dependente para cada valor da variável independente, leva-se em consideração a influência da variância da variável independente que, caso seja significativa, aproximará melhor os dados em uma reta de regressão, aumentando seu poder explicativo e preditivo.

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.)

Tabela 1: Coeficientes da regressão

Parâmetro	Estimativa pontual
Intercepto	3.4998133
Renda per capita	-0.002594
Anos de estudo do responsável pelo domicilio	-0.054417
Sexo feminino	0.8879468
Idade (em meses)	0.0157118

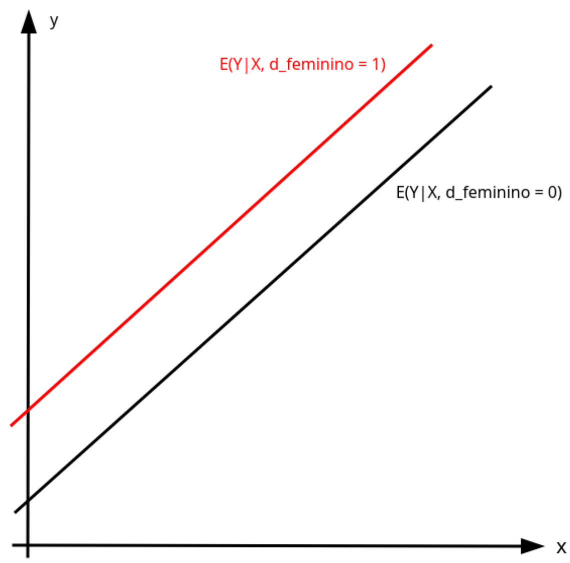
Parâmetro	Estimativa pontual
Acesso à abastecimento hídrico	-0.285436
Área rural	0.2625615
Região Norte	1.3930684
Região Nordeste	0.0366229
Região Centro-Oeste	-0.241068
Região Sudeste	-0.449075

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.*
   - $E(Y|X, \text{d\_feminino} = 0) = \beta_0 + \beta_1 X$ representando observações do sexo masculino
   - $E(Y|X, \text{d\_feminino} = 1) = \beta_0 + \beta_1 X + \beta_2 \delta$ representando observações do sexo feminino (δ)



2. *Com base no passo anterior, escreva, textualmente, a interpretação da magnitude numérica da estimativa pontual.*

O coeficiente de aproximadamente 0,89 para o sexo indica que, em média, indivíduos do sexo feminino possui 0,89 centímetros a mais de déficit de altura para a idade de crianças de cinco anos, residentes em domicílios brasileiros.

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.*

Tabela 2: Matriz de dummies

	δ_1	δ_2	δ_3	δ_4	$E(Y\|X)$	Equação
S	0	0	0	0	$E(Y\|X, \delta_1 = \cdots = \delta_4 = 0)$	$= \beta_0 + \beta_1 X$
N	1	0	0	0	$E(Y\|X, \delta_1 = 1, \delta_2 = \delta_3 = \delta_4 = 0)$	$= \beta_0 + \beta_1 X + \beta_2 \delta_1$
NE	0	1	0	0	$E(Y\|X, \delta_2 = 1, \delta_1 = \delta_3 = \delta_4 = 0)$	$= \beta_0 + \beta_1 X + \beta_3 \delta_2$
CO	0	0	1	0	$E(Y\|X, \delta_3 = 1, \delta_1 = \delta_2 = \delta_4 = 0)$	$= \beta_0 + \beta_1 X + \beta_4 \delta_3$
SE	0	0	0	1	$E(Y\|X, \delta_4 = 1, \delta_1 = \delta_2 = \delta_3 = 0)$	$= \beta_0 + \beta_1 X + \beta_5 \delta_4$

Onde X representa as demais variáveis da regressão.

2. *Com base no passo anterior, escreva, textualmente, a interpretação da magnitude numérica da estimativa pontual.*

O coeficiente de aproximadamente 1,4 para a região Norte indica que, em média, residentes da região norte possuem 1,4 centímetros a mais de déficit de altura para a idade de crianças de cinco anos, residentes em domicílios brasileiros, em comparação com a região de referência (Sul).

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.

O coeficiente para variáveis contínuas representa o efeito de uma variação na variável independente na variável dependente, ou seja, $\Delta y_i = \beta_i \Delta x_i$.

O coeficiente da variável "renda domiciliar *per capita*" indica que, *ceteris paribus*, um aumento na renda leva a uma diminuição do déficit de altura para a idade de crianças de cinco anos em -0,0026.