INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO EXPERIMENTAL– 2024-25

Resoluções de exercícios teóricos de Regressão Linear - Abordagem Descritiva

1. Tem-se:

   (a) $\sum_{i=1}^{n}(x_i - \overline{x}) \;=\; \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} \;=\; n\,\overline{x} - n\,\overline{x} \;=\; 0.$

   (b) Por definição, $(n-1)cov_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$. Distribuindo o primeiro factor de cada parcela pelas parcelas do segundo factor e utilizando o resultado da alínea anterior, temos:

   $$(n-1)cov_{xy} \;=\; \sum_{i=1}^{n}(x_i-\overline{x})y_i - \sum_{i=1}^{n}(x_i-\overline{x})\overline{y} \;=\; \sum_{i=1}^{n}(x_i-\overline{x})y_i - \overline{y}\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})}_{=\,0} \;=\; \sum_{i=1}^{n}(x_i-\overline{x})y_i$$

   Trocando o papel das variáveis x e y, mostra-se que $(n-1)cov_{xy} = \sum_{i=1}^{n} x_i(y_i - \overline{y})$.

   (c) Por definição, $(n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})$. Procedendo como na alínea anterior, tem-se $(n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \overline{x})x_i - \overline{x}\underbrace{\sum_{i=1}^{n}(x_i - \overline{x})}_{=\,0} \;=\; \sum_{i=1}^{n}(x_i - \overline{x})x_i$

2. (a) Tendo em conta que os valores ajustados de y são dados por $\hat{y}_i = b_0 + b_1\,x_i$, tem-se que a média dos valores ajustados é dada por:

   $$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \;=\; \frac{1}{n}\sum_{i=1}^{n}(b_0 + b_1\,x_i) \;=\; \frac{1}{n}\sum_{i=1}^{n} b_0 + \frac{1}{n}\sum_{i=1}^{n} b_1 x_i \;=\; b_0 + b_1\,\overline{x}\;.$$

   Mas a ordenada de origem duma recta de regressão é dada por $b_0 = \overline{y} - b_1\,\overline{x}$, pelo que a última expressão equivale à média $\overline{y}$ dos valores observados de y.

   (b) Tem-se, por definição, que $e_i = y_i - \hat{y}_i$. Logo (e tendo em conta a alínea anterior),

   $$\overline{e} \;=\; \frac{1}{n}\sum_{i=1}^{n} e_i \;=\; \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \;=\; \overline{y} - \overline{y} \;=\; 0.$$

   (c) Por definição, a a variância amostral de n observações z_i quaisquer, é dada por $s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})^2$. Assim, a variância amostral das n observações y_i da variável resposta é dada por: $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$. O somatório nesta expressão é, por definição, a Soma de Quadrados Total, SQT. Logo, $SQT = (n-1) \times s_y^2$, como se pedia para justificar. De forma análoga, a variância dos n valores ajustados da variável resposta, os $\hat{y}_i$,é dada por: $s_{\hat{y}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$. Repare-se que, nesta última expressão, a média dos valores $\hat{y}_i$ é igual à média dos n valores observados y_i, como se viu na alínea 2a. Mas o somatório nesta expressão é, por definição, SQR. Logo, $SQR = (n-1) \times s^2_{\hat{y}}$. Finalmente, a variância amostral dos n resíduos e_i é, por definição, $s^2_e = \frac{1}{n-1} \sum\limits_{i=1}^{n} (e_i - \overline{e})^2$. Mas a média dos n resíduos é nula, como se viu na alínea 2b, e $\overline{e}=0$ implica que o somatório nesta última expressão é apenas a Soma de Quadrados Residual, $SQRE = \sum\limits_{i=1}^{n} e^2_i$. Logo, $SQRE = (n-1) \times s^2_e$.

(d) Ora, recordando a definição dos valores ajustados de y e a expressão da ordenada na origem da recta de regressão, b_0, temos que $\hat{y}_i = b_0 + b_1\, x_i = \overline{y} + b_1(x_i - \overline{x})$. Logo,

$$SQR \;=\; \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \;=\; \sum_{i=1}^{n}[b_1\,(x_i - \overline{x})]^2 \;=\; b_1^2 \sum_{i=1}^{n}(x_i - \overline{x})^2 \;=\; b_1^2\,(n-1)s^2_x\;.$$

(e) Na expressão que define SQT vamos introduzir um par de parcelas de soma zero, que nos ajudarão nas contas subsequentes $(-\hat{y}_i + \hat{y}_i = 0)$:

$$\begin{aligned} SQT &= \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}\left[(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})\right]^2 \\ &= \sum_{i=1}^{n}\Big[(y_i - \hat{y}_i)^2 + (\hat{y}_i - \overline{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \overline{y})\Big] \\ &= \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{=SQRE} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}_{=SQR} + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) \qquad (1) \end{aligned}$$

Para que a igualdade pedida se verifique, é preciso que a última parcela na expressão (1) seja nula. Viu-se acima que $\hat{y}_i = b_0 + b_1\, x_i = \overline{y} + b_1(x_i - \overline{x})$. Logo, o somatório na última parcela da equação (1) pode ser re-escrito como:

$$\begin{aligned} \sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) &= \sum_{i=1}^{n}\left[(y_i - \overline{y}) - b_1\,(x_i - \overline{x})\right]\, b_1\,(x_i - \overline{x}) \\ &= b_1[\underbrace{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}_{=(n-1)\,cov_{xy}} - b_1\,\underbrace{\sum_{i=1}^{n}(x_i - \overline{x})^2}_{=(n-1)\,s^2_x}] \end{aligned}$$

Tendo em conta que $b_1 = \frac{cov_{xy}}{s^2_x}$, tem-se $b_1\, s^2_x = cov_{xy}$. Logo, a diferença acima anula-se.

(f) Pela definição de coeficiente de correlação entre x e y, tem-se:

$$r_{xy} \;=\; \frac{cov_{xy}}{s_x \cdot s_y} \;=\; \frac{cov_{xy}}{s^2_x} \cdot \frac{s_x}{s_y} \;=\; b_1 \cdot \frac{s_x}{s_y}$$

(g) Por definição, $R^2 = \frac{SQR}{SQT}$. Tendo em conta que $SQT = (n-1)s^2_y$, que $SQR = b^2_1(n-1)s^2_x$ (Exercício 2d) e o resultado da alínea anterior, tem-se $R^2 = \frac{b^2_1\, s^2_x}{s^2_y} = (r_{xy})^2$.

(h) Os valores ajustados $\hat{y}_i$ são dados por uma mesma transformação linear (afim) dos valores do preditor: $\hat{y}_i = b_0 + b_1\, x_i$. São conhecidas as propriedades destas transformações sobre a covariância e a variância. Assim,

$$r^2_{y\hat{y}} = \left(\frac{cov_{y\hat{y}}}{s_y s_{\hat{y}}}\right)^2 = \frac{cov^2_{y,b_0+b_1x}}{s^2_y\, s^2_{b_0+b_1x}} = \frac{(b_1 cov_{y,x})^2}{s^2_y\, b^2_1 s^2_x} = \frac{\cancel{b^2_1} cov^2_{xy}}{\cancel{b^2_1} s^2_x\, s^2_y} = r^2_{xy} = R^2.$$

Assim, o coeficiente de determinação duma regressão linear simples é também o quadrado do coeficiente de correlação linear entre os valores observados e os valores ajustados de y. Esta propriedade estende-se às regressões lineares múltiplas, embora seja necessário adaptar a justificação.

3. Comecemos por recordar alguns resultados já previamente discutidos:

- Viu-se no Exercício 1b) da Regressão Linear Simples que, para qualquer conjunto de n pares de observações se tem: $(n-1)\,\text{cov}_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i$. Distribuindo y_i e o somatório pela diferença, tem-se:

$$(n\text{–}1)\,\text{cov}_{xy} = \sum_{i=1}^{n} x_i y_i - \overline{x} \underbrace{\sum_{i=1}^{n} y_i}_{=n\,\overline{y}} = \sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y} \quad \Leftrightarrow \quad \sum_{i=1}^{n} x_i y_i = (n\text{–}1)\,\text{cov}_{xy} + n\,\overline{x}\,\overline{y}. \quad (2)$$

- Tomando $y_i = x_i$, para todo o i, na fórmula anterior, obtém-se:

$$(n-1)\,s^2_x = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x^2_i - n\,\overline{x}^2 \quad \Leftrightarrow \quad \sum_{i=1}^{n} x^2_i = (n-1)\,s^2_x + n\,\overline{x}^2 \qquad (3)$$

- O *produto de matrizes* AB só é possível quando o número de colunas da matriz A fôr igual ao número de linhas da matriz B (matrizes *compatíveis* para a multiplicação). Se A é de dimensão $p \times q$ e B de dimensão $q \times r$, o produto AB é de dimensão $p \times r$.
- O elemento na linha i, coluna j, dum produto matricial AB, é dado pelo *produto interno* da linha i de A com a coluna j de B: $(AB)_{ij} = (a_{i1}\ a_{i2}\ ....\ a_{iq}) \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{qj} \end{pmatrix} = \sum_{k=1}^{q} a_{ik}b_{kj}$.
- O produto interno de dois vectores n-dimensionais $\vec{\mathbf{x}}$ e $\vec{\mathbf{y}}$ é dado por $\vec{\mathbf{x}}^t\vec{\mathbf{y}} = \sum_{i=1}^{n} x_i\, y_i$. No caso de um dos vectores ser o vector de n uns, $\vec{\mathbf{1}}_n$, o produto interno resulta na soma dos elementos do outro vector, ou seja, em n vezes a média dos elementos do outro vector: $\vec{\mathbf{1}}^t_n\vec{\mathbf{x}} = \sum_{i=1}^{n} x_i = n\,\overline{x}$.
- A *matriz inversa* duma matriz $n \times n$ A é definida (caso exista) como a matriz (única) A^{-1}, também de dimensão $n \times n$, tal que $AA^{-1} = \mathbf{I}_n$, onde $\mathbf{I}_n$ é a matriz identidade de dimensão $n \times n$ (recorde-se que uma matriz identidade é uma matriz quadrada com todos os elementos diagonais iguais a 1 e todos os elementos não diagonais iguais a zero).

- No caso de A ser uma matriz 2×2, de elementos $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, a matriz inversa é dada (verifique!) por:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{4}$$

esta matriz inversa existe *se e só se* o *determinante* $ad - bc \neq 0$.

Com estes resultados prévios, as contas do exercício resultam de forma simples:

(a) A matriz do modelo $\mathbf{X}$ é de dimensão $n \times (p{+}1)$, que no caso duma regressão linear simples ($p{=}1$), significa $n{\times}2$. Tem uma primeira coluna de uns (o vector $\vec{\mathbf{1}}_n$) e uma segunda coluna com os n valores observados da variável preditora x, coluna essa que designamos pelo vector $\vec{\mathbf{x}}$. Logo, a sua transposta $\mathbf{X}^t$ é de dimensão $2 \times n$. Como o vector $\vec{\mathbf{y}}$ é de dimensão $n \times 1$, o produto $\mathbf{X}\vec{\mathbf{y}}$ é possível e o resultado é um vector de dimensão 2×1. O primeiro elemento (na posição (1,1)) desse produto é dada pelo produto interno da primeira linha de $\mathbf{X}^t$ com a primeira e única coluna de $\vec{\mathbf{y}}$, ou seja, por $\vec{\mathbf{1}}_n^t \vec{\mathbf{y}} = \sum\limits_{i=1}^{n} y_i = n\,\overline{y}$. O segundo elemento (posição (2,1)) desse vector é dado pelo produto interno da segunda linha de $\mathbf{X}^t$ e a única coluna de $\vec{\mathbf{y}}$, ou seja, por $\vec{\mathbf{x}}^t\vec{\mathbf{y}} = \sum\limits_{i=1}^{n} x_i\, y_i = (n{-}1)\,\mathrm{cov}_{xy} + n\,\overline{x}\,\overline{y}$, tendo em conta a equação (2).

(b) Tendo em conta que $\mathbf{X}^t$ é de dimensão $2 \times n$ e $\mathbf{X}$ é de dimensão $n \times 2$, o produto $\mathbf{X}^t\mathbf{X}$ é possível e de dimensão 2×2. O elemento na posição $(1,1)$ é o produto interno da primeira linha de $\mathbf{X}^t$ ($\vec{\mathbf{1}}_n$) com a primeira coluna de $\mathbf{X}$ (igualmente $\vec{\mathbf{1}}_n$), logo é: $\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n = n$. O elemento na posição (1,2) é o produto interno da primeira linha de $\mathbf{X}^t$ ($\vec{\mathbf{1}}_n$) e segunda coluna de $\mathbf{X}$ ($\vec{\mathbf{x}}$), logo é $\vec{\mathbf{1}}_n^t\vec{\mathbf{x}} = \sum\limits_{i=1}^{n} x_i = n\,\overline{x}$. O elemento na posição (2,1) é o produto interno da segunda linha de $\mathbf{X}^t$ ($\vec{\mathbf{x}}$) com a primeira coluna de $\mathbf{X}$ ($\vec{\mathbf{1}}_n$), logo é também $n\,\overline{x}$. Finalmente, o elemento na posição (2,2) é o produto interno da segunda linha de $\mathbf{X}^t$ ($\vec{\mathbf{x}}$) com a segunda coluna de $\mathbf{X}$ ($\vec{\mathbf{x}}$), ou seja, $\vec{\mathbf{x}}^t\vec{\mathbf{x}} = \sum\limits_{i=1}^{n} x_i^2 {=} (n{-}1)\, s_x^2 + n\,\overline{x}^2$. Fica assim provado o resultado do enunciado.

(c) A expressão da inversa dada no enunciado vem directamente de aplicar a fórmula (4) à matriz ($\mathbf{X}^t\mathbf{X}$) obtida na alínea anterior. Apenas há que confirmar a expressão do determinante $ad{-}bc = n \sum\limits_{i=1}^{n} x_i^2 {-} \left(\sum\limits_{i=1}^{n} x_i\right)^2 {=} n \sum\limits_{i=1}^{n} x_i^2 {-} (n\,\overline{x})^2 {=} n \left(\sum\limits_{i=1}^{n} x_i^2 {-} n\overline{x}^2\right) {=} n(n{-}1)\, s_x^2$, tendo em conta a fórmula (3).

(d) Usando as expressões finais obtidas nas alíneas (c) e (a), obtém-se

$$\begin{aligned}
(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{y}} &= \frac{1}{n(n{-}1)s_x^2} \begin{bmatrix} (n{-}1)s_x^2 + n\overline{x}^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix} \begin{bmatrix} n\overline{y} \\ (n{-}1)cov_{xy} + n\,\overline{x}\,\overline{y} \end{bmatrix} \\
&= \frac{1}{n(n-1)\, s_x^2} \begin{bmatrix} (n{-}1)\, s_x^2\, n\,\overline{y} + \cancel{n^2\overline{x}^2\,\overline{y}} - n\,\overline{x}\,(n{-}1)\, cov_{xy} - \cancel{n^2\,\overline{x}^2\,\overline{y}} \\ -\cancel{n^2\,\overline{x}\,\overline{y}} + n(n{-}1)\, cov_{xy} + \cancel{n^2\overline{x}\,\overline{y}} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\cancel{n(n-1)}\, \cancel{s_x^2}\ \overline{y}}{\cancel{n(n-1)}\, \cancel{s_x^2}} - \frac{\cancel{n(n-1)} cov_{xy}\, \overline{x}}{\cancel{n(n-1)}\, s_x^2} \\ \frac{\cancel{n(n-1)}\ cov_{xy}}{\cancel{n(n-1)}\ s_x^2} \end{bmatrix} = \begin{bmatrix} \overline{y} - b_1\overline{x} \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.
\end{aligned}$$

4. Sabemos que a matriz de projecção ortogonal referida é dada por $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, onde $\mathbf{X}$ é a matriz do modelo, ou seja, a matriz de dimensões $n \times (p+1)$ que tem na primeira coluna, n uns, e em cada uma das p restantes colunas, as n observações de cada variável preditora. Ora,

(a) A idempotência é fácil de verificar, tendo em conta que $(\mathbf{X}^t\mathbf{X})^{-1}$ é a matriz inversa de $\mathbf{X}^t\mathbf{X}$:

$$\mathbf{H}\,\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\,\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\cancel{\mathbf{X}^t\mathbf{X}}\cancel{(\mathbf{X}^t\mathbf{X})^{-1}}\mathbf{X}^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{H}\ .$$

A simetria resulta de três propriedade conhecidas de matrizes: a transposta duma matriz transposta é a matriz original ($(\mathbf{A}^t)^t = \mathbf{A}$); a transposta dum produto de matrizes é o produto das correspondentes transpostas, pela ordem inversa ($(\mathbf{AB})^t = \mathbf{B}^t\mathbf{A}^t$); e a transposta duma matriz inversa é a inversa da transposta ($(\mathbf{A}^{-1})^t = (\mathbf{A}^t)^{-1}$). De facto, tem-se:

$$\mathbf{H}^t = [\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]^t = \mathbf{X}[(\mathbf{X}^t\mathbf{X})^{-1}]^t\mathbf{X}^t = \mathbf{X}[(\mathbf{X}^t\mathbf{X})^t]^{-1}\mathbf{X}^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{H}\ .$$

(b) Como foi visto nas aulas teóricas, qualquer vector do subespaço das colunas da matriz $\mathbf{X}$, ou seja, do subespaço $\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^n$, se pode escrever como $\mathbf{X}\vec{\mathbf{a}}$, onde $\vec{\mathbf{a}} \in \mathbb{R}^{p+1}$ é o vector dos $p+1$ coeficientes na combinação linear das colunas de $\mathbf{X}$. Ora, a projecção ortogonal deste vector sobre o subespaço $\mathcal{C}(\mathbf{X})$ (que já o contém) é dada por

$$\mathbf{H}\mathbf{X}\vec{\mathbf{a}} \ = \ \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{X}\vec{\mathbf{a}}) \ = \ \mathbf{X}\cancel{(\mathbf{X}^t\mathbf{X})^{-1}}\cancel{(\mathbf{X}^t\mathbf{X})}\vec{\mathbf{a}} \ = \ \mathbf{X}\vec{\mathbf{a}}\ .$$

Assim, o vector $\mathbf{X}\vec{\mathbf{a}}$ fica igual após a projecção.

(c) Por definição, o vector dos valores ajustados é dado por $\hat{\vec{\mathbf{y}}} = \mathbf{H}\vec{\mathbf{y}}$. Ora, a média desses valores ajustados, que podemos representar por $\overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i$, pode ser calculado tomando o produto interno do vector $\vec{\mathbf{1}}_n$ de n uns com o vector $\hat{\vec{\mathbf{y}}}$, uma vez que esse produto interno devolve a soma dos elementos de $\hat{\vec{\mathbf{y}}}$. Assim, a média dos valores ajustados é $\overline{\hat{y}} = \frac{1}{n}\vec{\mathbf{1}}_n^t\hat{\vec{\mathbf{y}}} = \frac{1}{n}\vec{\mathbf{1}}_n^t\mathbf{H}\vec{\mathbf{y}} = \frac{1}{n}(\mathbf{H}\vec{\mathbf{1}}_n)^t\vec{\mathbf{y}} = \frac{1}{n}\vec{\mathbf{1}}_n^t\vec{\mathbf{y}}$, uma vez que $\mathbf{H}\vec{\mathbf{1}}_n = \vec{\mathbf{1}}_n$. Esta última afirmação decorre directamente da alínea anterior, uma vez que o vector $\vec{\mathbf{1}}_n$ pertence ao subespaço das colunas de $\mathbf{X}$, sendo a primeira das colunas dessa matriz (usando a notação da alínea anterior, tem-se $\vec{\mathbf{1}}_n = \mathbf{X}\vec{\mathbf{a}}$ com $\vec{\mathbf{a}} = (1, 0, 0, ..., 0)^t$). Mas a expressão final obtida, $\frac{1}{n}\vec{\mathbf{1}}_n^t\vec{\mathbf{y}}$, é a média $\overline{y}$ dos valores observados de Y (já que $\vec{\mathbf{1}}_n^t\vec{\mathbf{y}}$ devolve a soma dos elementos do vector dessas observações, $\vec{\mathbf{y}}$). Assim, também na regressão linear múltipla, valores observados de Y e correspondentes valores ajustados partilham o mesmo valor médio.

5. **[Opcional]**

A matriz de projecção ortogonal $\mathbf{P} = \vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t$ é de dimensão $n \times n$ (confirme!), uma vez que o vector $\vec{\mathbf{1}}_n$ é $n \times 1$. Mas o seu cálculo é facilitado pelo facto de que $\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n$ é, neste caso, um escalar. Concretamente, $\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n = n$, pelo que $(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1} = \frac{1}{n}$. Logo $\mathbf{P} = \frac{1}{n}\vec{\mathbf{1}}_n\vec{\mathbf{1}}_n^t$. O produto $\vec{\mathbf{1}}_n\vec{\mathbf{1}}_n^t$ resulta numa matriz $n \times n$ com todos os elementos iguais a 1 (não confundir com o produto pela ordem inversa, $\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n$: recorde-se que o produto de matrizes **não** é comutativo). Assim,

$$\mathbf{P} \ = \ \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

Tem-se $\mathbf{P}^t = \left[\frac{1}{n}\vec{\mathbf{1}}_n\vec{\mathbf{1}}_n^t\right]^t = \frac{1}{n}[\vec{\mathbf{1}}_n^t]^t[\vec{\mathbf{1}}_n]^t = \frac{1}{n}\vec{\mathbf{1}}_n\vec{\mathbf{1}}_n^t = \mathbf{P}$, logo $\mathbf{P}$ é simétrica. Por outro lado, $\mathbf{PP} = \vec{\mathbf{1}}_n\cancel{(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t \cdot \vec{\mathbf{1}}_n}(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t = \vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t = \mathbf{P}$, logo $\mathbf{P}$ é idempotente.

(b) O vector $\mathbf{P}\vec{\mathbf{Y}}$ repete n vezes o valor médio $\overline{Y}$ (confirme!). Logo, $\vec{\mathbf{Y}}-\mathbf{P}\vec{\mathbf{Y}}$ é o vector centrado das observações de $\vec{\mathbf{Y}}$:

$$\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}} \;=\; \begin{bmatrix} Y_1 - \overline{Y} \\ Y_2 - \overline{Y} \\ Y_3 - \overline{Y} \\ \vdots \\ Y_n - \overline{Y} \end{bmatrix} \;=\; \vec{\mathbf{Y}}^c$$

A norma ao quadrado dum qualquer vector $\vec{\mathbf{z}} = (z_1, z_2, ..., z_n)^t$ pode ser escrita de duas formas equivalentes: $\|\vec{\mathbf{z}}\|^2 = \sum\limits_{i=1}^{n} z_i^2$, ou $\|\vec{\mathbf{z}}\|^2 = \vec{\mathbf{z}}^t\vec{\mathbf{z}}$. Assim, tem-se $\|\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}}\|^2 = \sum\limits_{i=1}^{n}(Y_i - \overline{Y})^2 = SQT$. Tendo em conta as propriedades relativas a matrizes, incluindo a simetria e idempotência das matrizes $\mathbf{I}$ (de verificação trivial) e $\mathbf{P}$ (alínea anterior), tem-se também:

$$\begin{aligned} SQT \;=\; \|\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}}\|^2 \;&=\; \|(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}}\|^2 \;=\; [(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}}]^t(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{I}-\mathbf{P})^t(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{I}^t-\mathbf{P}^t)(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{I}^2-\mathbf{IP}-\mathbf{PI}+\mathbf{P}^2)\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{I}-\mathbf{P}-\mathbf{P}+\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{I}-\mathbf{P})\vec{\mathbf{Y}} \end{aligned}$$

De forma análoga, e como o vector $\hat{\vec{\mathbf{Y}}}$ dos valores ajustados é dado por $\hat{\vec{\mathbf{Y}}} = \mathbf{X}\hat{\vec{\boldsymbol{\beta}}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{Y}} = \mathbf{H}\vec{\mathbf{Y}}$, temos que o vector $\mathbf{H}\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}}$ tem como elementos $\hat{Y}_i - \overline{Y}$:

$$\mathbf{H}\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}} \;=\; \begin{bmatrix} \hat{Y}_1 - \overline{Y} \\ \hat{Y}_2 - \overline{Y} \\ \hat{Y}_3 - \overline{Y} \\ \vdots \\ \hat{Y}_n - \overline{Y} \end{bmatrix}$$

pelo que o quadrado da sua norma é $SQR = \sum\limits_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$.

Uma expressão alternativa para SQR resulta de considerar (como no caso de SQT) a definição duma norma ao quadrado, e usar as propriedades de matrizes, incluindo a simetria e idempotência de $\mathbf{P}$ e da matriz $\mathbf{H}$ (ver Exercício 4), bem como a propriedade $\mathbf{HP} = \mathbf{PH} = \mathbf{P}$. Esta última passagem resulta do facto de $\mathbf{P} = \vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t$, pelo que $\mathbf{HP} = \mathbf{H}\vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t$. Mas, como se viu no Exercício 4, a projecção dum vector sobre um subespaço ao qual esse vector pertence, deixa o vector invariante. Ora $\mathbf{H}$ projecta ortogonalmente sobre o espaço das colunas da matriz $\mathbf{X}$, $\mathcal{C}(\mathbf{X})$, e o vector $\vec{\mathbf{1}}_n$ pertence a esse subespaço, uma vez que é a primeira coluna da matriz $\mathbf{X}$. Logo, $\mathbf{H}\vec{\mathbf{1}}_n = \vec{\mathbf{1}}_n$, pelo que $\mathbf{HP} = \mathbf{H}\vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t = \vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t = \mathbf{P}$. Por outro lado, a simetria de $\mathbf{P}$ (e de $\mathbf{H}$) implica que $\mathbf{P} = \mathbf{P}^t = (\mathbf{HP})^t = \mathbf{P}^t\mathbf{H}^t = \mathbf{PH}$. Logo,

$$\begin{aligned} SQR \;=\; \|\mathbf{H}\vec{\mathbf{Y}} - \mathbf{P}\vec{\mathbf{Y}}\|^2 \;&=\; \|(\mathbf{H}-\mathbf{P})\vec{\mathbf{Y}}\|^2 \;=\; \vec{\mathbf{Y}}^t(\mathbf{H}-\mathbf{P})^t(\mathbf{H}-\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{H}^t-\mathbf{P}^t)(\mathbf{H}-\mathbf{P})\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{H}-\mathbf{P})(\mathbf{H}-\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{H}^2-\mathbf{HP}-\mathbf{PH}+\mathbf{P}^2)\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{H}-\mathbf{P}-\mathbf{P}+\mathbf{P})\vec{\mathbf{Y}} \\ &=\; \vec{\mathbf{Y}}^t(\mathbf{H}-\mathbf{P})\vec{\mathbf{Y}}\ . \end{aligned}$$

Finalmente, o vector $\vec{\mathbf{Y}} - \mathbf{H}\vec{\mathbf{Y}} = \vec{\mathbf{Y}} - \hat{\vec{\mathbf{Y}}}$ é o vector dos resíduos, e a sua norma ao quadrado é $SQRE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$. Mas, ao mesmo tempo, tem-se:

$$\begin{aligned}
SQRE \;=\; \|\vec{\mathbf{Y}} - \mathbf{H}\vec{\mathbf{Y}}\|^2 &= \|(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}}\|^2 \;=\; \vec{\mathbf{Y}}^t(\mathbf{I} - \mathbf{H})^t(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}} \\
&= \vec{\mathbf{Y}}^t(\mathbf{I}^t - \mathbf{H}^t)(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}} \\
&= \vec{\mathbf{Y}}^t(\mathbf{I}^2 - \mathbf{IH} - \mathbf{HI} + \mathbf{H}^2)\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t(\mathbf{I} - \mathbf{H} - \mathbf{H} + \mathbf{H})\vec{\mathbf{Y}} \\
&= \vec{\mathbf{Y}}^t(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}} \;.
\end{aligned}$$

(c) Pelas expressões da alínea anterior, pondo em evidência $\vec{\mathbf{Y}}^t$ à esquerda e $\vec{\mathbf{Y}}$ à direita tem-se:

$$\begin{aligned}
SQRE + SQR &= \vec{\mathbf{Y}}^t(\mathbf{I} - \mathbf{H})\vec{\mathbf{Y}} + \vec{\mathbf{Y}}^t(\mathbf{H} - \mathbf{P})\vec{\mathbf{Y}} \;=\; \vec{\mathbf{Y}}^t[(\mathbf{I} - \mathbf{H}) + (\mathbf{H} - \mathbf{P})]\vec{\mathbf{Y}} \\
&= \vec{\mathbf{Y}}^t[\mathbf{I} - \mathbf{P}]\vec{\mathbf{Y}} \;=\; SQT \;.
\end{aligned}$$