Resolução da primeira parte da terceira lista de exercícios do curso MAE0328 disponível no Texto de Regressão (Paula, 2023).

16. Considere o seguinte modelo de regressão linear $y_i = \beta_1+\beta_2(x_i-\bar{x})+\epsilon_i$, em que $\epsilon_i \sim \mathrm{N}(0,\sigma^2)$ independentes, para $i=1,\ldots,n$. Note que tal modelo pode ser expresso na forma matricial como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, em que $\mathbf{y} = (y_1,\ldots,y_n)^\top$, $\mathbf{X} = [\mathbf{X}_1^\top,\cdots,\mathbf{X}_n^\top]^\top$, com $\mathbf{X}_i = [1, x_i-\bar{x}]$, $\boldsymbol{\beta} = (\beta_1,\beta_2)^\top$ e $\boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_n)^\top$. Sabemos, para $k>0$, que

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= (\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_2)^{-1}\mathbf{X}^\top\mathbf{y}\\ &= \begin{bmatrix} n+k & 0\\ 0 & S_{xx}+k\end{bmatrix}^{-1}\begin{bmatrix} n\bar{y}\\ S_{xy}\end{bmatrix} = \begin{bmatrix} n\bar{y}/(n+k)\\ S_{xy}/(S_{xx}+k)\end{bmatrix},\end{aligned}$$

e também que,

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}_R) &= \sigma^2(\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_2)^{-1}(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_2)^{-1}\\ &= \sigma^2\begin{bmatrix} n+k & 0\\ 0 & S_{xx}+k\end{bmatrix}^{-1}\begin{bmatrix} n & 0\\ 0 & S_{xx}\end{bmatrix}\begin{bmatrix} n+k & 0\\ 0 & S_{xx}+k\end{bmatrix}^{-1}\\ &= \sigma^2\begin{bmatrix} n/(n+k)^2 & 0\\ 0 & S_{xx}/(S_{xx}+k)^2\end{bmatrix}.\end{aligned}$$

Note, para $Z_k = (\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_p)^{-1}(\mathbf{X}^\top\mathbf{X})$, que

$$\begin{aligned}\mathrm{E}\{\hat{\boldsymbol{\beta}}_R\} &= \mathrm{E}\{(\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_p)^{-1}\mathbf{X}^\top\mathbf{y}\}\\ &= \mathrm{E}\{(\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_p)^{-1}(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\}\\ &= \mathrm{E}\{(\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_p)^{-1}(\mathbf{X}^\top\mathbf{X})\hat{\boldsymbol{\beta}}\}\\ &= (\mathbf{X}^\top\mathbf{X}+k\mathbf{I}_p)^{-1}(\mathbf{X}^\top\mathbf{X})\boldsymbol{\beta} = Z_k\boldsymbol{\beta},\end{aligned}$$

portanto $\hat{\boldsymbol{\beta}}_R$ é viesado para $\boldsymbol{\beta}$ com $k>0$.

18. Considere o seguinte modelo de regressão linear múltipla $y_i = \mathbf{x}_i^\top\boldsymbol{\beta}+\epsilon_i$, em que $\epsilon_i \sim \mathrm{N}(0,\sigma^2)$ independentes, para $i=1,\ldots,n$. Note que tal modelo pode ser expresso na forma matricial como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$, em que $\mathbf{y} = (y_1,\ldots,y_n)^\top$, $\mathbf{X} = [\mathbf{X}_1^\top,\cdots,\mathbf{X}_n^\top]^\top$, com $\mathbf{X}_i = \mathbf{x}_i^\top$, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top$ e $\boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_n)^\top$. Assumindo que $r(X) = p$, temos

$$\begin{aligned}SQRes(k) - SQRes &= (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_R)^\top(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_R) - (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\\ &= -2\mathbf{y}^\top\mathbf{X}\hat{\boldsymbol{\beta}}_R + \hat{\boldsymbol{\beta}}_R^\top\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}}_R + 2\mathbf{y}^\top\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}}\\ &= -2(\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}})^\top\mathbf{X}^\top\mathbf{y} + (\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}})^\top(\mathbf{X}^\top\mathbf{X})(\hat{\boldsymbol{\beta}}_R+\hat{\boldsymbol{\beta}})\\ &= (\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}})^\top\{-2\mathbf{X}^\top\mathbf{y} + (\mathbf{X}^\top\mathbf{X})(\hat{\boldsymbol{\beta}}_R+\hat{\boldsymbol{\beta}})\}\\ &= (\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}})^\top\{-(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} + (\mathbf{X}^\top\mathbf{X})\hat{\boldsymbol{\beta}}_R\}\\ &= (\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}})^\top(\mathbf{X}^\top\mathbf{X})(\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}}) \geq 0,\end{aligned}$$

pois $(\mathbf{X}^\top\mathbf{X})$ é positiva semi-definida. Dessa forma $SQRes(k) \geq SQRes$.

19. Considere o seguinte modelo de regressão linear $y_i = \alpha + \beta x_i + \epsilon_i$, em que $\epsilon_i \sim \mathrm{N}(0, d_i\sigma^2)$ independentes, com $d_i > 0$, para $i = 1, \ldots, n$. Note que tal modelo pode ser expresso na forma matricial como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\mathbf{X} = [\mathbf{X}_1^\top, \cdots, \mathbf{X}_n^\top]^\top$, com $\mathbf{X}_i = [1, x_i]$, $\boldsymbol{\beta} = (\alpha, \beta)^\top$ e $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ com $\boldsymbol{\epsilon} \sim \mathrm{N}_n(0_n, \sigma^2\mathbf{V}^{-1})$, com $\mathbf{V} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$. Sabemos que

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}\mathbf{y} \\
&= \begin{bmatrix} \sum_{i=1}^n d_i^{-1} & \sum_{i=1}^n x_i d_i^{-1} \\ \sum_{i=1}^n x_i d_i^{-1} & \sum_{i=1}^n x_i^2 d_i^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i d_i^{-1} \\ \sum_{i=1}^n x_i y_i d_i^{-1} \end{bmatrix} \\
&= \frac{1}{c}\begin{bmatrix} (\sum_{i=1}^n x_i^2 d_i^{-1})(\sum_{i=1}^n y_i d_i^{-1}) - (\sum_{i=1}^n x_i y_i d_i^{-1})(\sum_{i=1}^n y_i d_i^{-1}) \\ (\sum_{i=1}^n d_i^{-1})(\sum_{i=1}^n y_i x_i d_i^{-1}) - (\sum_{i=1}^n x_i d_i^{-1})(\sum_{i=1}^n y_i d_i^{-1}) \end{bmatrix},
\end{aligned}
$$

com $c = (\sum_{i=1}^n d_i^{-1})(\sum_{i=1}^n x_i^2 d_i^{-1}) - (\sum_{i=1}^n x_i d_i^{-1})^2$. Também, para $C = \begin{bmatrix} 0 & 1 \end{bmatrix}$ temos que

$$
\begin{aligned}
ASQ(C\boldsymbol{\beta} = 0) &= \hat{\boldsymbol{\beta}}^\top C^\top \{C(\mathbf{X}^\top\mathbf{V}\mathbf{X})^{-1}C^\top\}^{-1} C\hat{\boldsymbol{\beta}} \\
&= \hat{\beta}_1^2 c \left\{ C \begin{pmatrix} \sum_{i=1}^n x_i^2 d_i^{-1} & -\sum_{i=1}^n x_i d_i^{-1} \\ -\sum_{i=1}^n x_i d_i^{-1} & \sum_{i=1}^n d_i^{-1} \end{pmatrix} C^\top \right\}^{-1} \\
&= \frac{c\hat{\beta}_1^2}{\sum_{i=1}^n d_i^{-1}}.
\end{aligned}
$$

Referências

Paula GA (2023) Regressão linear múltipla (versão parcial preliminar). Notas de aula atualizadas em 03-23 do curso MAE0328