REGRESSÃO LINEAR SIMPLES ($\hat{Y} = \hat{a} + \hat{b}\, X$ **é a recta de regressão estimada)**

MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS

Um resíduo ou erro é a diferença entre o valor observado e o valor estimado pelo modelo de regressão, ou seja, $e_i = y_i - \hat{y}_i = y_i - \hat{a} - \hat{b}\, x_i$.

De acordo com o método dos mínimos quadrados os estimadores dos parâmetros a e b, ou seja $\hat{a}$ e $\hat{b}$, consiste na minimização da soma quadrática dos erros.
A soma quadrática dos erros é designada por $SQ_{Erro} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}\, x_i)^2$.
Esta minimização que se pretende resulta da resolução do seguinte sistema:

$$\begin{cases} \dfrac{\partial}{\partial a} SQ_{Erro} = 0 \\ \dfrac{\partial}{\partial b} SQ_{Erro} = 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{\partial}{\partial a} \sum_{i=1}^{n} \left(y_i - a - b\, x_i \right)^2 = 0 \\ \dfrac{\partial}{\partial b} \sum_{i=1}^{n} \left(y_i - a - b\, x_i \right)^2 = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \sum_{i=1}^{n} \left(y_i - a - b\, x_i \right) = 0 \\ -2 \sum_{i=1}^{n} x_i \left(y_i - a - b\, x_i \right) = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \sum_{i=1}^{n} y_i - n\, a - b \sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n} x_i\, y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 = 0 \end{cases} \Leftrightarrow \begin{cases} n\, \overline{y} - n\, a - b\, n\, \overline{x} = 0 \\ \sum_{i=1}^{n} x_i\, y_i - a\, n\, \overline{x} - b \sum_{i=1}^{n} x_i^2 = 0 \end{cases} \Leftrightarrow \begin{cases} a = \overline{y} - b\, \overline{x} \\ \sum_{i=1}^{n} x_i\, y_i - (\overline{y} - b\, \overline{x})\, n\, \overline{x} - b \sum_{i=1}^{n} x_i^2 = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} - \\ \sum_{i=1}^{n} x_i\, y_i - n\, \overline{x}\, \overline{y} + n\, b\, \overline{x}^2 - b \sum_{i=1}^{n} x_i^2 = 0 \end{cases} \Leftrightarrow \begin{cases} - \\ \sum_{i=1}^{n} x_i\, y_i - n\, \overline{x}\, \overline{y} + b \left(n\, \overline{x}^2 - \sum_{i=1}^{n} x_i^2 \right) = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} - \\ b = \dfrac{- \sum_{i=1}^{n} x_i\, y_i + n\, \overline{x}\, \overline{y}}{n\, \overline{x}^2 - \sum_{i=1}^{n} x_i^2} \end{cases} \Leftrightarrow \begin{cases} - \\ b = \dfrac{\sum_{i=1}^{n} x_i\, y_i - n\, \overline{x}\, \overline{y}}{\sum_{i=1}^{n} x_i^2 - n\, \overline{x}^2} \end{cases} \Leftrightarrow \begin{cases} - \\ b = \dfrac{\frac{1}{n} \sum_{i=1}^{n} x_i\, y_i - \overline{x}\, \overline{y}}{\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2} \end{cases} \Leftrightarrow \begin{cases} \hat{a} = \overline{y} - \hat{b}\, \overline{x} \\ \hat{b} = \dfrac{Cov(X,Y)}{Var(X)} \end{cases}$$

A variável Y é chamada de variável resposta, explicada ou dependente e a variável X é chamada de variável controlada, explicativa ou independente.

Covariância entre as variáveis X e Y:

$$Cov(X,Y) = S_{XY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \overset{*}{=} \frac{1}{n} \sum_{i=1}^{n} x_i\, y_i - \overline{x}\, \overline{y}$$

Variância da variável X :

$$Var(X) = S_{XX} = S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \overset{*}{=} \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

Coeficiente angular da recta (declive da recta):

$$b = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \overset{*}{=} \frac{Cov(X,Y)}{VAR(X)} = \frac{S_{xy}}{S_x^2}$$

Ordenada na origem da recta: $a = \overline{y} - b\overline{x}$

Coeficiente de correlação linear amostral ou índice de ajustamento:

$$r_{XY} = \frac{n\sum_{i=1}^{n} x_i\, y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\ \sqrt{n\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2}} \overset{*}{=} \frac{\mathrm{Cov(X,Y)}}{\sqrt{\mathrm{Var(X)}\,\mathrm{Var(Y)}}} \overset{*}{=} b\,\frac{S_x}{S_y}$$

Coeficiente de determinação:

$$R^2 = r_{XY}^2 = \frac{(Cov(X,Y))^2}{\mathrm{Var(X)}\,\mathrm{Var(Y)}} \overset{*}{=} 1 - \frac{SQ_{Erro}}{Var(Y)}$$

(*) requer uma demonstração