Notas Finais Capítulo 1

Dada a amostra bivariada de variáveis quantitativas
$\{(x_i, y_i)\}_{i=1}^n$ ***n*** dimensão da amostra

Se

- a nuvem de pontos sugere a existência de uma tendência linear entre os valores observados;
- $|r| \simeq 1$

É admissivel supor que uma recta modela bem a relação entre as variáveis

A recta é denotada por

$\boldsymbol{y = b_0 + b_1 x}$ $\longrightarrow$ e designa-se **recta de regressão** entre

- y (variável resposta e
  x (variável explicativa, regressora)

A regressão linear simples

Coeficientes determinados pelo **método dos mínimos quadrados**

$$\boldsymbol{b_1 = \frac{cov(x, y)}{s_x^2} = r\,\frac{s_y}{s_x}}$$

$$\boldsymbol{b_0 = \overline{y} - b_1\overline{x}}$$

A $\boldsymbol{b_1}$ chama-se **coeficiente de regressão de y sobre x**

e interpreta-se como **a variação esperada em y quando x aumenta uma unidade** – atenção às unidades de b_1.

A recta de regressão passa no ponto $(\overline{x}, \overline{y})$.

Precisão da recta de regressão – O coeficiente de determinação

Um dos objectivos da recta de regressão é o de **predizer** o valor de uma variável conhecendo o valor assumido pela outra **mas** é necessário avaliar o **grau de precisão** atingido pelas estimativas/previsões. Essa precisão é avaliada pelo quociente

$$R^2 = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2} \qquad 0 \leq R^2 \leq 1$$

que toma valores entre 0 e 1

Verifica-se que $R^2 = r^2$.

Interpretação — $100 \times R^2$ **mede a percentagem de variabilidade "explicada" pela regressão**

Finalmente ...

Falámos aqui de dois termos muito importantes: **Correlação** e **regressão**

Correlação – refere-se à existência de associação entre duas variáveis — se ela for linear é medida por $\boldsymbol{r}$;

Regressão — refere-se a um modelo que exprime uma variável em função da outra — considerámos o modelo linear — chama-se **recta de regressão**