

1 - FOUNDATIONS

Aprendizagem 2024/2025

BOXPLOT

- **Interquartile range:** $IQR = q_n(75) - q_n(25)$
- **Limites do boxplot:** $[q_n(25) - 1.5IQR; q_n(75) + 1.5IQR]$

CORRELAÇÕES

- **Pearson** (mede força da relação linear; quando dados seguem dist. normal):
$$\text{pearson}(y1, y2) = \frac{\sum y1y2 - \frac{\sum y1 \sum y2}{n}}{\sqrt{\left(\sum y1^2 - \frac{(\sum y1)^2}{n}\right)\left(\sum y2^2 - \frac{(\sum y2)^2}{n}\right)}}$$
- **Spearman** (ideal para variáveis ordinais e para lidar com presença de outliers; quando dados não seguem dist. normal):
$$\text{spearman}(y1, y2) = \text{pearson}(r1, r2)$$

OUTRAS FÓRMULAS

- **Média:** $\mu = \frac{\sum x_i}{n}$
- **Variância:** $\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$ (população); $\sigma^2 = \frac{\sum (x_i - \mu)^2}{n-1}$ (amostra)
- **Desvio padrão:** $\sigma = \sqrt{\sigma^2}$

ENTROPIA E GANHO DE INFORMAÇÃO

- **Entropia** (medida de pureza dos dados, 0 significa determinismo): $H(y)$
- **Ganho de informação** (quantidade de informação que a feature y dá acerca do target z): $IG(y)$

SUMÁRIO

- Ficha 1

FICHA 1, EXERCÍCIO I.5

