

# Análise Exploratória de Dados - Avaliação Presencial - 2023/01

Prof. Hugo Carvalho

25/05/2023

**Questão 1:** O objetivo dessa questão é explorar a quantidade denominada *desvio mediano absoluto*. Primeiramente, seja  $x = \{x_1, \dots, x_n\}$  um conjunto de dados. Lembre que pensamos na mediana de  $x$ , denotada por  $\text{med}(x)$ , como a observação central após colocar  $x$  em ordem crescente. Matematicamente, temos que:

$$\text{med}(x) = x_{(\frac{n+1}{2})}, \text{ se } n \text{ é ímpar},$$

$$\text{med}(x) = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, \text{ se } n \text{ é par}.$$

Uma vez que temos a mediana de  $x$ , defina as quantidades abaixo:

$$y_i = |x_i - \text{med}(x)|, \text{ para } i = 1, \dots, n.$$

Finalmente, denotando por  $y$  o conjunto  $\{y_1, \dots, y_n\}$ , defina o *desvio mediano absoluto* de  $x$ , denotado por  $\text{dma}(x)$  como

$$\begin{aligned} \text{dma}(x) &= \text{med}(y) \\ &= \text{med}(\{|x_1 - \text{med}(x)|, \dots, |x_n - \text{med}(x)|\}). \end{aligned}$$

- a) Com base na explicação acima, interprete o que tal quantidade visa medir, em relação ao conjunto de dados  $x$ .  
*Dica: Comece olhando a quantidade “de dentro para fora”: o que significa  $\text{med}(x)$ , depois fale sobre os  $y_i$  e finalmente fale sobre sua mediana.*
- b) Essa é uma medida de centralidade, dispersão, ou nenhum dos dois, em relação ao conjunto  $x$ ? Justifique adequadamente a sua resposta.
- c) Essa métrica parece ser muito ou pouco sensível em relação à valores discrepantes? Justifique adequadamente a sua resposta.
- d) Escreva uma comparação do desvio mediano absoluto com o *desvio médio absoluto em relação à mediana*, definido por

$$\text{dmarm}(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \text{med}(x)|.$$

- e) Se  $\text{dma}(x) = 0$ , o que você pode dizer sobre o conjunto de dados  $x$ ?  
*Dica: Aqui recomendo fazer o inverso do item a), ou seja, comece olhando a igualdade “de fora para dentro”.*

— QUESTÃO DOIS NO VERSO DA PÁGINA —

**Questão 2:** A Figura 1 contém *box-plots* referentes às idades de lutadores/as do torneio de artes marciais mistas UFC (*Ultimate Fighting Championship*), separados/as por categoria e sexo, em determinado ano (que não foi informado). Caso fique ruim na impressão, os números dentro dos retângulos dos *box-plots* representam os valores das respectivas medianas, e os nomes indicam os/as lutadores/as de menor e maior idade dentro de sua respectiva categoria. Com base nisso escreva um texto de aproximadamente uma página discorrendo sobre as informações contidas nesse gráfico. Seu texto deve tratar dos pontos abaixo (mas não necessariamente se limitando a eles):

- Estudar alguma relação entre a idade e a categoria
- Tentar encontrar algum padrão nas idades dos campeões (marcados com uma estrela em cada *box-plot*)
- Pontos positivos dessa visualização, que facilitam a transmissão da informação desejada
- Pontos negativos dessa visualização, que podem confundir o interlocutor, e possíveis pontos de melhoria

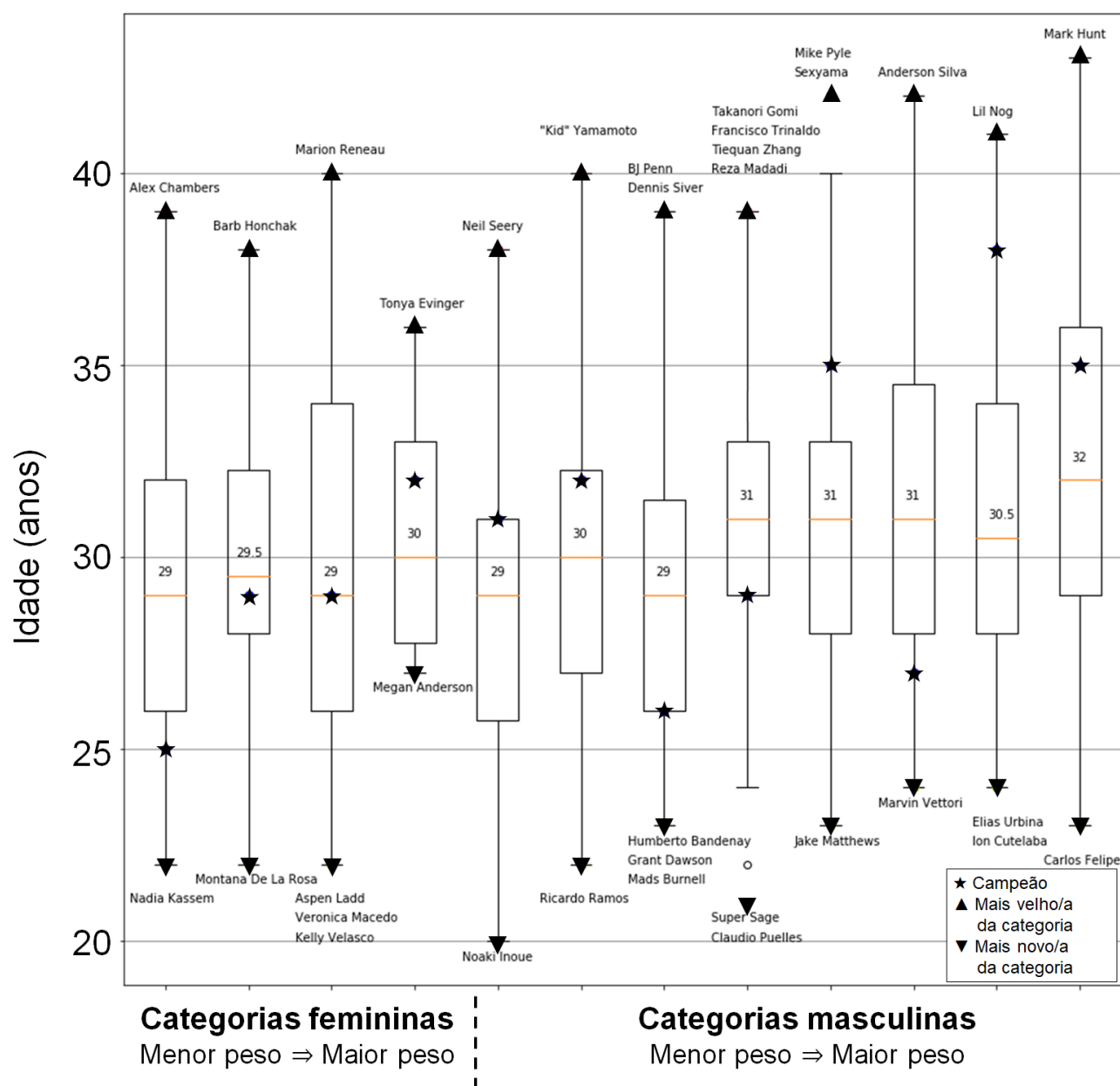


Figure 1: *Box-plot* das idades de lutadores/as UFC, separados/as por categoria e sexo.