

# Análise Exploratória de Dados

## Segunda chamada da avaliação presencial - 2023/01

Prof. Hugo Carvalho

04/07/2023

**Questão 1:** O objetivo desta questão é explorar a quantidade denominada *média interquartil*. Primeiramente, seja  $x = \{x_1, \dots, x_n\}$  um conjunto de dados, e lembre que denotamos por  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  o conjunto de dados  $x$  após ele ser ordenado de forma crescente. Definimos então a *média interquartil* de  $x$ , denotada por  $x_{\text{MIQ}}$  como

$$x_{\text{MIQ}} = \frac{2}{n} \sum_{i=n/4}^{3n/4} x_{(i)}.$$

Repare que os limites inferior e superior do somatório podem ser números fracionários. Nesse caso, simplesmente comece o somatório no próximo inteiro e termine no inteiro anterior ao limite superior. Repare também que o fator multiplicativo na frente é  $2/n$ , e não  $1/n$ , como seria esperado na média aritmética usual.

- Explique o que a fórmula para  $x_{\text{MIQ}}$  visa calcular. Em particular, compare-a com a média de  $x$ , denotada por  $\bar{x}$ .
- Aponte vantagens e desvantagens da quantidade  $x_{\text{MIQ}}$ , quando comparada com a quantidade  $\bar{x}$ .
- Crie conjuntos de dados (pequenos!) que ilustrem a comparação que você fez no item b) acima.
- Essa métrica parece ser muito ou pouco sensível em relação à valores discrepantes? Justifique adequadamente a sua resposta.
- Se  $x_{\text{MIQ}} = 0$ , o que você pode dizer sobre o conjunto de dados  $x$ ?

**Questão 2:** Analise a Figura 1 abaixo. Ela contém *box-plots* referentes aos dados de sentimentos de um usuário do Reddit que sofre de depressão crônica coletou ao longo de um ano. A forma de coleta e tratamento dos dados foi: ao final de cada dia o usuário atribuiu uma nota de 0 até 10 para cada um dos seis sentimentos apresentados na figura, em um aplicativo de controle de humor; ao final de um ano ele utilizou algum *software* para gerar tais gráficos. Com base nessas informações, escreva um texto de aproximadamente uma página falando sobre a imagem na Figura 1.

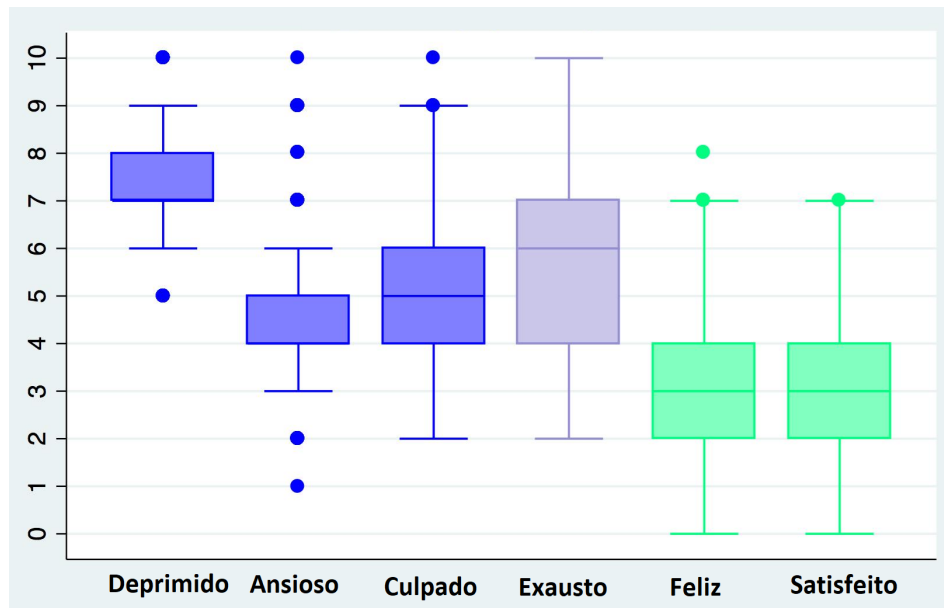


Figure 1: *Box-plots* referentes ao humor de um usuário do Reddit que sofre de depressão crônica, ao longo de um ano.