

Análise Exploratória de Dados - Avaliação Presencial 02 - 2024/01

Prof. Hugo Carvalho

02/07/2024

Questão 1: O objetivo dessa questão é explorar a quantidade denominada *coeficiente de variação*. Primeiramente, seja $x = \{x_1, \dots, x_n\}$ um conjunto de dados. Lembre que pensamos na média de x como uma medida de centralidade do conjunto de dados, e no desvio padrão de x como uma medida que determina a dispersão dos dados em torno da média. Tais quantidades são calculadas, respectivamente, por:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad \text{dp}(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

A partir disso, definimos o *coeficiente de variação* de x como

$$\text{cv}(x) = \frac{\text{dp}(x)}{\bar{x}}.$$

Com base nisso, faça o que se pede abaixo.

- Interprete o que mede o coeficiente de variação. Mais especificamente, explique se $\text{cv}(x)$ é uma medida de centralidade, dispersão, ambos ou nenhum, e explique **intuitivamente** o que ela significa (sem recorrer a fórmulas matemáticas, como se estivesse explicando para alguém leigo em Estatística).
- Mostre/argumente que o coeficiente de variação é uma quantidade *adimensional*, ou seja, sem unidades de medida.
- Aponte vantagens e desvantagens do coeficiente de variação.
- Alguém te faz a seguinte afirmativa: “Estou tentando estimar a distância entre os pontos A e B, através de diversas medições. Meus dados me dão uma distância média de 100 km, com desvio padrão de 500 m, de modo que posso considerar que minha estimativa está relativamente acurada”. Analise essa afirmação à luz do coeficiente de variação.
- Você tem um conjunto de dados x com desvio padrão igual a 2. Interprete o coeficiente de variação nos seguintes cenários para a média: $\bar{x} = 5$, $\bar{x} = 1$, $\bar{x} = 0$ e $\bar{x} = -2$.

Questão 2: A Figura 1 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, apontando também pontos positivos e negativos dela.

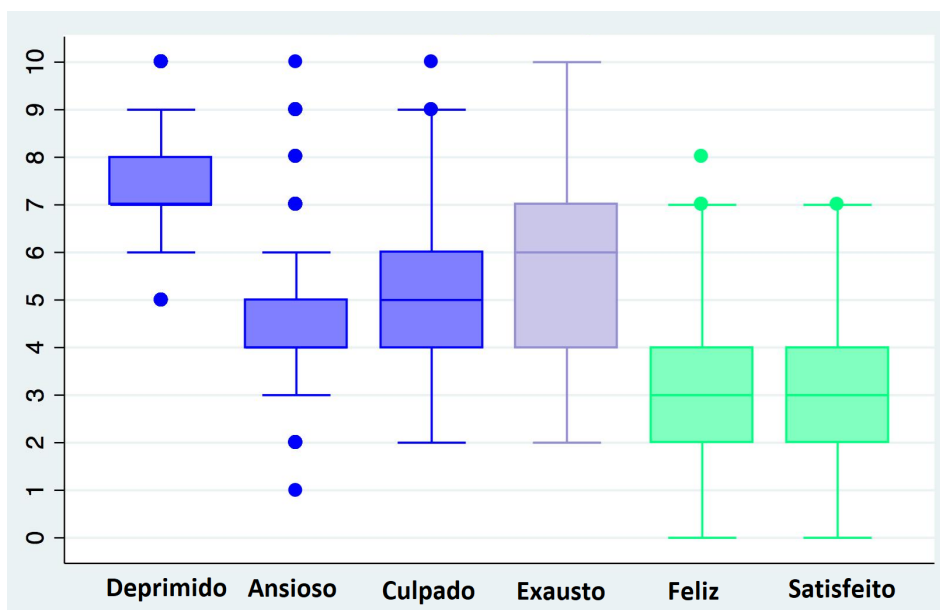


Figure 1: *Box-plots* referentes ao humor de um usuário do Reddit que sofre de depressão crônica, ao longo de um ano.