

# Desvendando Quartis e Boxplots: Uma Análise Visual de Dados com Google Colab

Na análise de dados, compreender a distribuição e a dispersão de um conjunto de informações é fundamental. Duas ferramentas estatísticas poderosas para essa tarefa são os **quartis** e o **boxplot**. Este guia completo irá desmistificar esses conceitos, mostrando como calculá-los e visualizá-los de forma prática utilizando o Google Colab.

## O que são Quartis?

Imagine que você ordenou um conjunto de dados do menor para o maior valor. Os quartis são três pontos que dividem esses dados em quatro partes iguais, cada uma contendo 25% das observações.

- **Primeiro Quartil (Q1):** Também conhecido como o 25º percentil, este é o valor que deixa 25% dos dados abaixo dele e 75% acima.
- **Segundo Quartil (Q2):** Este é o valor central dos dados, mais conhecido como a **mediana** ou o 50º percentil. Ele divide o conjunto de dados exatamente ao meio.
- **Terceiro Quartil (Q3):** Correspondendo ao 75º percentil, este valor separa os 75% menores dados dos 25% maiores.

## Método Passo a Passo para Encontrar a Posição

1. **Conte o número total de dados (n):** Primeiro, veja quantos pontos de dados você tem no seu conjunto.
2. **Ordene os dados:** É **essencial** que os dados estejam em ordem crescente, do menor para o maior.
3. **Calcule a Posição:** Use a fórmula abaixo para encontrar a posição do percentil que você deseja. Para o primeiro quartil (Q1), estamos procurando o 25º percentil ( $p=25$ ).

A fórmula para a posição é:

$$\text{Posição} = p/100 \times (n+1)$$

Onde:

- $p$  é o percentil desejado
- $n$  é o número total de dados.

## O Que Significa uma Posição "Quebrada" (4.25)?

Aqui está o ponto crucial que explica por que diferentes programas podem dar resultados ligeiramente diferentes. Uma posição de **4.25** significa que o valor do primeiro quartil (Q1) não é exatamente o 4º número nem o 5º, mas está **entre** eles.

A diferença entre o terceiro e o primeiro quartil ( $Q3 - Q1$ ) é chamada de **Amplitude Interquartil (AIQ)** ou *Interquartile Range (IQR)*. Essa medida é crucial, pois representa a dispersão dos 50% centrais dos dados, sendo menos sensível a valores extremos (outliers).

Para encontrar o valor exato, nós fazemos uma **interpolação linear**:

**Parte Inteira (4): Pegue o valor na 4ª posição.**

**Parte Fracionária (0.25): Pegue a diferença entre o próximo valor (o 5º) e o valor atual (o 4º) e multiplique pela parte fracionária.**

**Some os dois resultados**

## O que é um Boxplot?

O boxplot, ou diagrama de caixa, é uma representação gráfica que resume a distribuição de um conjunto de dados com base nos seus quartis. Ele oferece uma visualização rápida da tendência central, dispersão e assimetria dos dados.

Os principais componentes de um boxplot são:

- **A Caixa:** Representa a amplitude interquartil (AIQ). A linha inferior da caixa é o primeiro quartil ( $Q1$ ) e a linha superior é o terceiro quartil ( $Q3$ ).
- **A Linha Dentro da Caixa:** Marca a mediana ( $Q2$ ).
- **Os "Bigodes" (Whiskers):** As linhas que se estendem a partir da caixa indicam a variabilidade fora dos quartis superior e inferior. Geralmente, eles se estendem até 1.5 vezes a AIQ a partir de  $Q1$  e  $Q3$ .
- **Outliers:** Pontos de dados que caem fora dos "bigodes" são considerados valores atípicos ou outliers e são plotados individualmente.