

## **1. O que são Outliers?**

Outliers são pontos em um conjunto de dados que se desviam drasticamente dos demais, parecendo "não pertencer" ao grupo. Eles podem distorcer médias e prejudicar análises. A ideia de usar uma regra para identificá-los foi popularizada pelo estatístico John Tukey (1977), que nos deu a ferramenta mais famosa para visualizá-los.

## **2. Situações que Podem Gerar Outliers:**

Antes de qualquer ação, é fundamental entender a origem do outlier. As três situações mais comuns são:

- Erro de Medição ou Entrada: A causa mais frequente, como um erro de digitação (ex: idade 200).
- Evento Raro e Legítimo: Um valor real, mas extremo (ex: o salário de um bilionário em uma pesquisa de bairro).
- Novidade ou Descoberta: O outlier pode sinalizar algo novo e importante, como uma fraude ou uma falha em um sistema.

## **3. O que Fazer e Possibilidades de Substituição:**

A ação correta depende da causa identificada. A regra principal, como destaca o autor Aggarwal (2017), é investigar antes de agir.

Remoção: Faça isso apenas se tiver certeza de que é um erro.

Substituição: Se não puder remover, você pode substituir o valor. As opções mais comuns são:

Pela Mediana: É mais seguro do que usar a média, pois a mediana não é afetada por valores extremos.

Por um Limite (Winsorização): Substituir os outliers pelo valor mais próximo que ainda é considerado "normal" (ex: o valor do 99º percentil).

- Transformação: Aplicar uma função (como a de log) em todos os dados para reduzir o impacto dos outliers.
- Investigação: Se o outlier for uma novidade, ele deve ser o foco principal da sua análise.

## **4. Plotando um Gráfico com Outliers:**

No arquivo.

## **Referências:**

- Aggarwal, C. C. (2017). *Outlier Analysis*. Springer.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.