

# Identificação de *outliers* em conjuntos de dados

Prof. Rafael H. P. Lima

Blog Aprendendo Gestão  
[www.aprendendogestao.com.br](http://www.aprendendogestao.com.br)

# Introdução

## O que são outliers?

Outliers são pontos que se distinguem dos demais;

Devemos procurar explicações para os outliers

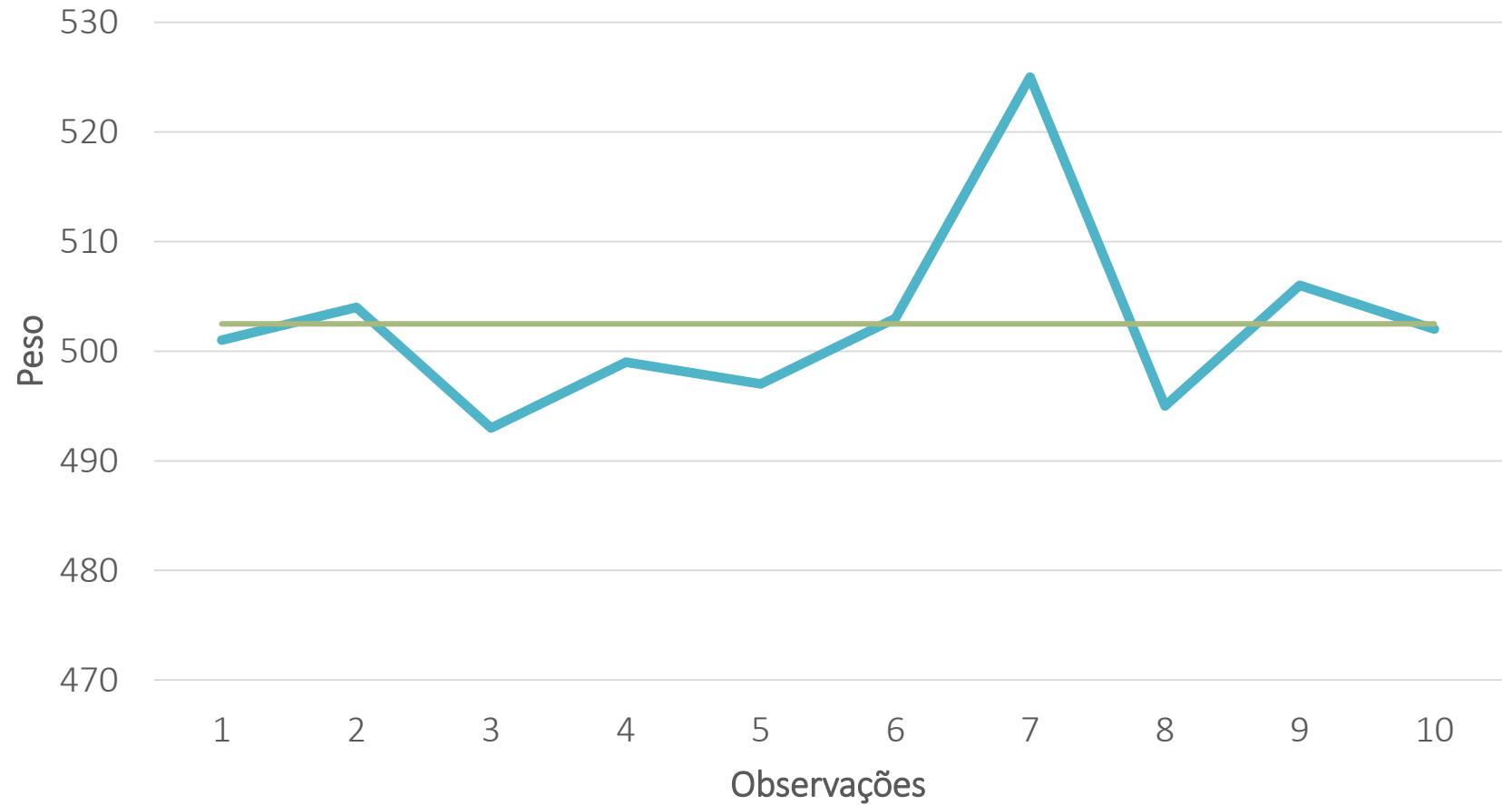
Em estimativas estatísticas, costuma-se eliminar os outliers para evitar o viés sobre as estimativas

**Exemplo:** estimativa de peso dos produtos em uma fábrica de alimentos

Observação	Valor
1	501
2	504
3	493
4	499
5	497
6	503
7	525
8	495
9	506
10	502

Vamos analisar esses valores individuais com relação à média da amostra

Observação	Valor
1	501
2	504
3	493
4	499
5	497
6	503
7	525
8	495
9	506
10	502
<b>Média</b>	<b>502,5</b>



# Identificação de Outliers

## Método para Identificação de Outliers

Vamos utilizar o método baseado na amplitude interquartil (IQR)

$$IQR = Q_3 - Q_1$$

Podemos usar essa estimativa de variabilidade para calcular os limites inferior e superior para identificação de *outliers*

$$L_{Sup} = \bar{x} + 1,5 \times IQR$$

$$L_{Inf} = \bar{x} - 1,5 \times IQR$$