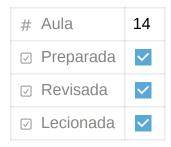


Valores Extremos a partir das Medidas de Dispersão



▼ No exemplo anterior e no nosso dia-adia...

Vimos um exemplo com poucas observações e um valor extremo (outlier) em cada grupo...

Normalmente teremos um número maior de observações, podendo ter também mais valores extremos...

Com mais observações, faz sentido utilizar decis e percentis

Aí fica a pergunta: Será que decis e percentis também são robustos (não sensíveis a valores extremos)?

- **▼ É preciso adotar então um padrão...**
 - ...além do desvio padrão!
 - ...ou melhor, um padrão acerca do que é um outlier
 - **▼** Por isso adotamos o seguinte padrão:

São considerados valores extremos (ou outliers) observações que sejam menores do que o Primeiro Quartil menos 1,5 vezes a distância interquartílica ou que sejam

maiores que o Terceiro Quartil mais 1,5 vezes a distância interquartílica.

▼ Qual a vantagem de fazer isso?

 Assumindo um critério podemos começar a comparar coisas diferentes usando este critério como referência.

▼ Dá para simplificar? Sim!

Observações menores que Q1 - 1,5 x IQR ou maiores que $Q3 + 1,5 \times IQR$

▼ E em relação ao Desvio Padrão, temos um padrão?

Importante: O Desvio Padrão diz respeito a todos os dados, não a um ou outro valor específico que pode ser ou não extremo...

Temos uma sugestão acadêmica, arbitrária:

- Se o (valor absoluto do) Coeficiente de Variação for menor ou igual a 30%, consideramos que os dados têm uma dispersão baixa...
- Se o (valor absoluto do) Coeficiente de Variação estiver entre 30% e 100%, consideramos que os dados têm uma dispersão média...
- Se o (valor absoluto do) Coeficiente de Variação for superior a 100%, consideramos que os dados têm uma dispersão alta...

▼ Para pensar: O que acontece com o Coeficiente de Variação se a média for igual a zero?

• Matematicamente a conta não se define com uma divisão por zero... Mas fora isso...