Medidas de Dispersão a partir da Média

# Aula	10
☑ Preparada	☑
☑ Revisada	☑
☑ Lecionada	☑

▼ Quanto erramos ao usar a Média?

No caso das idades calculamos a média (além da mediana e dos quartis), mas quanto estamos errando ao estimar a idade de todos os alunos com a média?

▼ Podemos calcular o desvio da média… e tirar a média dele!

Recordando… tínhamos que:

$$\overline{x} = \frac{1}{n}\sum_{i=i}^{n} x_i$$

Então se definirmos o desvio como sendo:

$$d_i = (x_i - \overline{x})$$

Então podemos calcular o…

▼ Desvio médio

$$DM = \frac{1}{n}\sum_{i=i}^{n}(x_i - \overline{x})$$

Mas se você observar bem a fórmula, verá um problema. Se a média é o centro de massa, quanto vai valer o desvio médio?

Z-E-R-O! A soma dos desvios vai ser igual a zero por que os desvios dos valores acima da média são compensados pelos desvios dos valores menores que a média.

Do Desvio Médio ao Desvio Padrão

Precisamos evitar a compensação de termos positivos e negativos.

Temos duas maneiras de fazer isso:

Com o desvio absoluto (usamos o módulo do desvio):

$$DA = \frac{1}{n}\sum_{i=i}^{n} |x_i - \overline{x}|$$

Com o desvio quadrático (usamos o quadrado do desvio):

$$DQ = Var(x) = \sigma^2 = \frac{1}{n}\sum_{i=i}^{n} (x_i - \overline{x})^2$$

Quais são os problemas com o desvio absoluto e com o desvio quadrático?

Em relação ao desvio absoluto:

- Embora conceitualmente simples…
- …ele é matematicamente bastante complicado de se operar…

Em relação ao desvio quadrático:

- Ele calcula o desvio numa dimensão ao quadrado…
- Como que “tira dessa realidade” a dimensão do desvio…
- Erros maiores são então exagerados…

Para resolver isso…

- **Surge o Desvio Padrão:**

$$DP = \sigma = \sqrt{\frac{1}{n}\sum_{i=i}^{n}(x_i - \overline{x})^2}$$

Quais as vantagens do Desvio Padrão?

Matematicamente é mais fácil trabalhar com o quadrado e com a raiz quadrada.

O Desvio Padrão é expresso na mesma dimensão do dado original.

Como comparar dois Desvios Padrão de amostras diferentes?

Um desvio padrão diz respeito à sua amostra, uma vez que ele está totalmente ligado à média daquela amostra.

No entanto, como o desvio padrão e a média estão na mesma dimensão, é possível padronizar, “referenciar” o desvio padrão pela média!

- **Surge então o Coeficiente de Variação:**

$$CV = \frac{\sigma}{\mu}$$

Qual a Desvantagem do Desvio Padrão?

A formulação matemática da variância e do desvio padrão em amostras nos leva a uma fórmula ligeiramente diferente dessas apresentadas…

Assim temos que:

- **Para a população:**

$$Média_{Populacional} = \mu = \frac{1}{N}\sum_{i=i}^{N} x_i$$

$$DQ_{Populacional} = Var_{Populacional}(x) = \sigma^2 = \frac{1}{N}\sum_{i=i}^{N}(x_i - \mu)^2$$

$$DP_{Populacional} = \sigma = \sqrt{\frac{1}{N}\sum_{i=i}^{N}(x_i - \mu)^2}$$

$$CV_{Populacional} = \frac{\sigma}{\mu}$$

- **Para a amostra:**

$$Média_{Amostral} = \overline{x} = \frac{1}{n}\sum_{i=i}^{n} x_i$$

$$DQ_{Amostral} = Var_{Amostral}(x) = s^2 = \frac{1}{(n-1)}\sum_{i=i}^{n}(x_i - \overline{x})^2$$

$$DP_{Amostral} = s = \sqrt{\frac{1}{(n-1)}\sum_{i=i}^{n}(x_i - \overline{x})^2}$$

$$CV_{Amostral} = \frac{s}{\overline{x}}$$

Para (tentar) deixar mais claro...

Vimos que a soma dos desvios é igual a zero:

$$\sum_{i=i}^{n} d_i = \sum_{i=i}^{n}(x_i - \overline{x}) = 0$$

▼ Então para calcular os desvios, nós não precisamos de todos os n termos, mas apenas de n - 1 …

- Exemplo: Se tivermos 2 observações e uma delas tiver o desvio igual a 3, sabemos que a outra terá o desvio igual a -3, uma vez que sabemos que a soma dos desvios é igual a zero.

▼ No caso de uma população inteira, a divisão por n pode ser aproximada e simplificada pela divisão por n - 1.

- Pense na população brasileira: Você acha que faz muita diferença calcular o desvio padrão de qualquer coisa dividindo a soma dos desvios quadráticos por 203.062.512 ou por 203.062.511?
- Mesmo que a sua população seja bem menor que a população brasileira, a diferença entre n e n - 1 será muito menos sensível na população de pesquisa do que em uma amostra de pesquisa…