

02

Média, Mediana e Moda: Como encontrar a tendência central?

Pronto para a segunda aula de Estatística? Vamos lá!

Média Aritmética

A primeira coisa que iremos mostrar é a *Média Aritmética*. Muito provavelmente você já deve ter estudado esse assunto na escola. Se pedirmos para você calcular a Média, por exemplo, de 1, 2, 3 e 4, você fará:

$$(1 + 2 + 3 + 4)/4 = 10/4 = 2,5$$

De modo geral:

$$(x_1 + x_2 + x_3 + \dots + x_n)/n$$

Agora façamos uma pergunta: será mesmo que para todo conjunto de números devemos utilizar a *Média Aritmética* para encontrarmos uma *Tendência Central*? Talvez. Devemos saber que tipo de dado possuímos. Lembra-se da primeira aula de Estatística? A *Média* serve para dados *Ordinais*, ou *Intervalares*? Vejamos um exemplo:

Imagine que você é um professor e no final da aula você entregue um formulário para os alunos preencherem com a nota que eles te dariam, de 1 a 10:

Nota: ()1 ()2 ()3 ()4 ()5 ()6 ()7 ()8 ()9 ()10 [dado do tipo Ordinal]

Segue que: 1º aluno: 10 2º aluno: 1 3º aluno: 1 4º aluno: 1

Fazendo a Média:

$$(10 + 1 + 1 + 1)/4 = 3,25$$

Não é estranho? O 10 dado pelo primeiro aluno não parece exceção? Você é, claramente, um professor nota 1! Perceba que escolher a *Média Aritmética*, sem pensar no tipo de dado, pode te levar para uma informação estranha.

Veja outro exemplo que pode nos enganar:

Em uma empresa os salários dos funcionários são como se segue:

- 1º funcionário: 2000 reais
- 2º funcionário: 2000 reais
- 3º funcionário: 2000 reais
- 4º funcionário: 100.000 reais

Calculando a Média:

$$100.000 + 2000 + 2000 + 2000 / 4 = 26500$$

Você diria que a média dos salários da empresa é de 26500 reais? Não, porque o 100.000 reais está atrapalhando a média real.

Chegamos à conclusão que temos que pensar em outras soluções para o cálculo da *Tendência Central* de um conjunto de dados. A *Média Aritmética* é totalmente sensível aos valores que chamamos de *outliers*, ou pontos fora da curva.

Veremos agora outras maneiras de calcular a tendência central de um conjunto de dados.

A Mediana

Para o cálculo da *Mediana*, colocamos os valores dados em ordem crescente e escolhemos aquele valor que é central. Isto é fácil para quantidades ímpares de dados, pois haverá um número que estará bem no meio da amostra. Veja um exemplo:

1 1 6 1 5 10 1 1 1

Em ordem crescente:

1 1 1 1 1 5 6 10

O valor central é o 1, que é a nossa *Mediana*.

Para quantidades pares de dados, pegamos os dois valores centrais e calculamos sua *Média Aritmética*:

1 1 5 1 2 10 1 6

Em ordem crescente:

1 1 1 1 2 5 6 10

Os valores centrais são, respectivamente, o 1 e o 2. A média entre ele é 1,5, a *Mediana* que queríamos encontrar.

Perceba que a *Mediana* é menos suscetível aos *outliers*. Lógico que isso não significa que a *Média Aritmética* não é uma boa Medida de Tendência Central. Se sua distribuição é estável, a Média pode ser boa.

Então como saber qual das duas usar? Verifique se os *outliers* são muito grandes e distantes do resto da amostra. Outra regra é que se o tipo de dado for *Ordinal*, a Média não é um bom método.

A Moda

A *Moda* é aquele elemento que mais se repete na distribuição:

1 1 2 2 2 2 3 3 5 5 5

Nessa amostra, o número 2 é o que mais se repete, logo ele é a nossa *Moda*.

Então, verifique qual é o número que mais se repete na distribuição, pois pode ser uma maneira honesta de se calcular a Medida de Tendência Central.

Você irá perceber que numa amostra bem distribuída, a **Média**, a **Mediana** e a **Moda** são iguais ou possuem valores muito próximos.

Dicas:

- Com dados *Ordinais* usamos **Mediana** ou **Moda**;
- Utilizamos a **Média** em dados *Intervalares*, salvo aqueles que possuam *outliers*;
- Desenhe um **Histograma**, como vimos na aula passada. Verifique se consegue desenhar uma linha parecida com a *Normal*. Se você tiver uma **distribuição normal**, o cálculo da **Média** é um método razoável.