

Praticando: Média, Mediana e Moda

Na última aula aprendemos sobre ***Tendências Centrais***: *Média*, *Mediana* e *Moda*. Agora veremos este assunto aplicado ao **R**.

Escrevemos uma sequência de números quaisquer no **R**:

```
> numeros <- c(1, 2, 3, 4, 5)
```

Revendo: ao digitar "numeros", conseguimos visualizar essa lista:

```
> numeros  
[1] 1 2 3 4 5
```

Para que o programa calcule a *Média Aritmética* desses números, digitamos "mean()":

```
> mean(numeros)  
[1] 3
```

A palavra *Mean* significa *Média* em inglês.

Para o cálculo da *Mediana*, digitamos "median()":

```
> median(numeros)  
[1] 3
```

Nesse caso em particular a *Média* é a *Mediana* são iguais.

Vamos agora gravar uma nova lista de números:

```
> lista <- c(2, 3, 7, 8, 1, 3, 4, 8, 22, 67, 19)
```

Calculando a Média e a Mediana:

```
> mean(lista)  
[1] 13.09091  
> median(lista)  
[1] 7
```

Perceba que o **R** calcula bem rápido essas medidas. Neste momento somos nós que estamos ditando esses dados de entrada. Em aulas futuras saberemos como importá-los de arquivos CSV.

Infelizmente o **R** não possui uma função própria para a *Moda*, porém podemos escrevê-la manualmente. Agora não é o momento, pois irá requerer maior conhecimento de **R** de sua parte e esse não é o foco da aula.

Vimos nas aulas passadas que é importante saber se a distribuição é *Normal* ou não. No *R* rodamos um *Teste de Hipótese* (veremos em outro momento o que é esse teste) para saber se os dados obedecem uma Normal. Para tal digitamos "shapiro.test()":

```
> shapiro.test(numeros)
      Shapiro-Wilk normality test
data:  numeros
W = 0.9868, p-value = 0.9672
```

Devemos observar o valor de "*p-value*". Se esse número for muito pequeno, algo em torno de 0,05, não é Normal. Nesse caso ele é próximo de 1, logo devemos acreditar que é uma Normal.

Façamos o teste com "lista":

```
> shapiro.test(lista)
      Shapiro-Wilk normality test
data:  lista
W = 0.634, p-value = 6.77e-05
```

Perceba que para "lista" o número é bem pequeno: 6.77e-05 (o "e-05" representa o número 10 *elevado a -5*, que está multiplicando 6,77). Isso significa que "lista" não é uma distribuição normal.

Outro método que é de costume ser usado no *R* é o "summary()", que nos passa várias informações importantes como a *Média* e a *Mediana*:

```
> summary(lista)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   3.00    7.00   13.09   13.50   67.00
```

Esse método resume o que já sabíamos anteriormente: *Média: 13,09 Mediana: 7 Valor mínimo: 1 Valor máximo: 67*

Em outras aulas descobriremos o que são o 1º e 3º quartis.

Revendo:

Nessa aula vimos na prática como se utiliza o *R* para: *Cálculo de Média e Mediana utilizando, respectivamente "mean()" e "median()"*; Descobrir se a distribuição é Normal ou não utilizando shapiro.test(): *Se p-value for pequeno, não é Normal Se p-value for próximo de 1, é Normal*; * Resumir as informações de uma lista de dados utilizando "summary ()".

