



Passo-a-passo

ETAPA 4. Medidas de Tendência Central

Prof. Pedro Feliú

INTRODUÇÃO

Nesta etapa executaremos comandos referentes à descrição estatística dos dados, utilizando medidas de tendência central. Essa etapa está conectada com o vídeo conceitual da aula 3, parte II.

PASSO 1: Medidas de tendência central

Vamos abrir agora o banco de dados “mundo” no Rstudio, com os mesmos passos já aprendidos na etapa anterior. Use a função **attach()**. Na sequência digite os comandos:

```
attach(mundo)  
mean (populacao, na.rm=TRUE)  
median (populacao, na.rm=TRUE)  
mean (floresta, na.rm = TRUE)  
median (floresta, na.rm = TRUE)  
mean (pobreza, na.rm=TRUE)  
median (pobreza, na.rm=TRUE)  
subset(table(macro_regiao),table(macro_regiao)==max(table(macro_regiao)))
```

```
> mundo <- read_excel("Documents/Estatística I/mundo.xls")  
> View(mundo)  
> attach(mundo)  
> mean (populacao, na.rm=TRUE)  
[1] 37686820  
> median (populacao, na.rm=TRUE)  
[1] 8331248  
> mean (floresta, na.rm = TRUE)  
[1] 207735.3  
> median (floresta, na.rm = TRUE)  
[1] 26935  
> mean (pobreza, na.rm=TRUE)  
[1] 25.36078  
> median (pobreza, na.rm=TRUE)  
[1] 23.3  
> subset(table(macro_regiao),table(macro_regiao)==max(table(macro_regiao)))  
Africa  
56  
~
```

A função **mean** retorna a média da variável selecionada nos parênteses. Como é muito comum em bancos de dados grades como o “mundo”, há muitos dados ausentes (missing), ou seja, determinada variável para determinado país não existe sistematizada. Na visualização do banco de dados no Rstudio, os dados missing aparecem como NA. Como NA é texto e estamos lidando com variáveis numéricas, é preciso inserir no comando da média e mediana o complemento `na.rm=TRUE`, que diz ao R que esse dado é ausente. No exemplo da figura acima, a média populacional dos países é 37686820. A função **median** retorna a mediana. A mediana da população dos países é 8331248, muito menor do que a média. Isso indica que valores muito elevados, como China e Índia, estão puxando a média para cima, como descrito no vídeo conceitual. O caso da cobertura florestal em km (floresta) é o mesmo, alguns países têm boa parte das florestas do mundo, com uma diferença muito grande entre a média (207735,3) e mediana (26935). Já no caso da porcentagem da população em estado de pobreza (menos de 2 dólares por dia), a média é muito semelhante a mediana. Não há valores muito extremos deturpando a média (25,3% de população pobre, ¼ do mundo). A função **subset** retorna a moda, o valor mais frequente de uma distribuição. Nesse caso, aplicamos aos continentes do mundo, uma variável qualitativa nominal, apenas para

Estatística I

efeito didático. O continente que mais possui países é a África com 56 países, sendo o continente que mais se repete na base de dados.