

Analizando média e mediana

Já sabemos que para calcular uma média usamos o comando `mean` e para calcular a mediana usamos `median`. O cálculo da mediana era mais trabalhoso antes da invenção dos pacotes estatísticos, pois para conhecermos a mediana precisamos ordenar os dados do menor para o maior. Por não precisar dessa ordenação, a média era uma medida mais rápida de ser obtida. Como não temos mais essa dificuldade de ordenação, a mediana é uma medida tão rápida de se conseguir quanto a média e com algumas vantagens. Qual das alternativas abaixo faz a melhor comparação entre média e mediana?

A

A média pode ser "distorcida", ou enviesada, por valores discrepantes. Por exemplo, o conjunto de dados '(5,5,5,5,5,5,5,5,100)' tem média 14,5. A mediana é menos afetada por valores discrepantes. No exemplo dado, a mediana é 5.



Correto. É muito comum termos uma média distorcida por conta de valores discrepantes. Na nossa aula, vimos que a média de tempo que os alunos levam para completar um curso é de 48 dias, enquanto a mediana fica em 8 dias. Essa diferença na média ocorre por conta dos valores discrepantes, casos em que o aluno levou 200, 300, 400 dias para concluir um curso.

B

A média é tão robusta quanto a mediana, além disso, se tivermos valores discrepantes para os dois lados, eles se cancelam. Por exemplo, o conjunto de dados '(0,5,5,5,5,5,5,5,10)' tem média e mediana iguais a 5.

**C**

A mediana é uma medida mais robusta e não muito mais trabalhosa de se conseguir, pois a ordenação dos dados para depois calculá-la é feita automaticamente pelos pacotes estatísticos tais como o R.



Correto. A mediana já foi muito mais trabalhosa de se conseguir por causa da necessidade de ordenação dos dados. Porém, com os pacotes estatísticos, nós nem percebemos essa diferença, pois eles fazem essa ordenação automaticamente.