

Temos três maneiras de medir o centro de uma distribuição:

1. A média de nossos dados
2. A mediana
3. A moda

Vamos examinar esses três um pouco mais de perto.

A Média:

A média, ou a média de nosso conjunto de dados, é melhor usada como medida de centro se nossos dados forem aproximadamente simétricos e não contiverem valores discrepantes.

A fórmula para a média é $\frac{\sum_{i=1}^n x_i}{n}$. Essa fórmula nos diz para adicionar todos os valores em nosso conjunto de dados e dividi-los pelo número de valores que adicionamos.

A média é muito sensível a outliers. A média sempre será puxada na direção do maior outlier

A Mediana:

A mediana é melhor usada como uma medida de centro quando os valores discrepantes estão presentes nos dados. Isso ocorre porque a mediana não será afetada por observações extremamente pequenas ou grandes.

A mediana são os pontos de dados em que 50% das observações estão acima e abaixo desse ponto de dados. Para encontrar a posição da mediana de um conjunto de dados com n observações, temos de considerar dois casos separados:

1. O conjunto de dados tem um número ímpar de observações:
 - a. Se este for o caso, então, a posição da mediana tomando $\frac{n+1}{2}$
2. O conjunto de dados tem um número par de observações:
 - a. Neste caso, encontramos as duas observações intermediárias e as medimos.

Exemplo da mediana:

Use este conjunto de dados como exemplo:

12	6	7	14	8	18	2	1	2	22	15
----	---	---	----	---	----	---	---	---	----	----

O primeiro passo que precisamos dar para encontrar a mediana é pedir nosso conjunto de dados:

1	2	2	6	7	8	12	14	15	18	22
---	---	---	---	---	---	----	----	----	----	----

Agora, como $n = 11$, neste caso, podemos encontrar a mediana tomando $n + 1/2$, que então se torna 6. Assim, contamos com a 6ª observação em nosso conjunto de dados, que é o número 8.

1	2	2	6	7	8	12	14	15	18	22
---	---	---	---	---	---	----	----	----	----	----

Se em vez disso tivéssemos 10 observações, calcularíamos a média dos dois valores do meio:

1	2	2	6	7	8	12	14	15	18
---	---	---	---	---	---	----	----	----	----

Isso faria com que nossa mediana fosse 7.5.

A Moda:

Esta medida de centro é melhor usada ao analisar conjuntos de dados categóricos. O número é o número, intervalo de números ou categoria que ocorre com mais frequência.

A moda também é muito resistente a outliers, pois depende de qual observação ocorre mais e não do valor real da observação.

