

Medidas de dispersão

As medidas de dispersão (ou de escala, ou de variabilidade), como o próprio nome sugere, são aquelas que nos auxiliarão a analisar o quanto a distribuição possui os dados dispersos (espalhados) em relação a sua média, ou seja, o quão distantes os dados da distribuição estão em relação à média, determinando assim, o grau de variabilidade da distribuição.

São medidas que, junto das demais, nos conferem ainda mais confiabilidade nas observações acerca da distribuição e, conseqüentemente, decisões mais seguras.

Variância

A variância é uma dessas medidas para se analisar a dispersão de uma distribuição de dados e junto com as medidas que vimos em exemplos anteriores vai potencializar ainda mais nossas observações e, conseqüentemente, dando confiança para as análises.

Agora, o grande ponto da variância é que ela vai observar a dispersão dos dados em relação a média da distribuição, ou seja, o quanto eles estão próximos ou distantes da média e, vai te resumir essa resposta em um único número, para que você possa tomar sua decisão.

O cálculo da variância vai realizar os seguintes passos:

1. Calcular a média da distribuição dos dados.
2. Subtrair a média de cada dado da distribuição, para obter a diferença.
3. Elevar ao quadrado cada uma dessas diferenças.
4. Somar os quadrados das diferenças.
5. Dividir a soma dos quadrados pela quantidade de dados.

Então, aqui temos um ponto de atenção importante! O resultado da variância é um número elevado ao quadrado! Sim! Quando você for analisar a variância precisará se lembrar disso, visto que essa análise é realizada junto a média (que não está elevada ao quadrado).

Mais a frente veremos o desvio padrão e o coeficiente de variação, que são os mais indicados para se analisar junto a média.

Agora que entendemos a variância, vamos conhecer a outra medida de dispersão.

Desvio padrão

Vou começar falando matematicamente sobre o desvio padrão, visto que é muito simples, pois o seu cálculo é a raiz quadrada da variância. Somente isso!

Pega-se o resultado da variância e obtêm-se a sua raiz quadrada. E, se o desvio padrão é baseado na variância, a observação dele é também acerca da média. **Em outras palavras é a variação média da dispersão dos dados em relação à média.**

E uma vez que temos o valor do desvio padrão, é possível analisá-lo junto a média, a partir de uma medida chamada coeficiente de variação, cujo cálculo é:

$$\text{Coeficiente de variação (CV)} = \frac{\text{desvio padrão}}{\text{média}}$$

E o guideline de observação da magnitude da dispersão, a partir do resultado do coeficiente de variação é:

- Diferença pequena (< 15%): Pode-se considerar que a distribuição possui uma baixa dispersão.
- Diferença moderada (15% - 30%): Considera-se que a distribuição dos dados possui uma dispersão moderada.
- Diferença significativa (> 30%): Nesse ponto, é possível dizer que a distribuição possui uma alta dispersão de seus dados.

Em questões de ordem prática, muitas vezes o desvio padrão é utilizado como informação junto a média.

Por exemplo:

Imagine que você está analisando a média das vendas da sua loja e o resultado foi de R\$ 1.500,00. Como você conhece de análise estatística de dados, resolveu gerar outros subsídios para sustentar sua análise e um desses foi o cálculo do desvio padrão, que gerou como resultado um de número de 142,86.

E aí, diante dessas duas medidas, é comum observar o seguinte discurso:

“A média das vendas foi de R\$ 1.500,00, podendo variar R\$ 142,86 para mais ou para menos”

Medidas de posição: Quartis, limites e outliers

E ainda, em casos mais profundos, quando a pessoa também já observou que a média é “puxada” para cima, visto as outras observações de medidas que fez, ela pode fazer a seguinte análise:

“A média das vendas foi de R\$ 1.500,00, podendo variar R\$ 142,86 para mais”

Ou seja, exclui-se o “para menos”, visto que as análises apontam para uma tendência de influência superior.

De certa forma, esses discursos não estão equivocados, dado que o desvio padrão é justamente a média da distância dos dados em relação à média distribuição. Porém, é importante destacar que, fundamentalmente falando, o desvio padrão não tem essa intenção por natureza. Atente-se a isso!

Agora, o que não pode acontecer (ou não deveria acontecer) de maneira alguma, é utilizar o desvio padrão junto a mediana, visto que ele é calculado a partir da média. Ok?

Conclusão

É importante mencionar que, quando utilizamos alguma tecnologia da informação, não há a necessidade de matemática pura, dado que existem funções específicas para isso. Inclusive, para calcular o desvio padrão, não se faz necessário calcular a variância antes, pois existe uma função específica para ele que, no Google Planilhas, chama-se STDEV (uma abreviação de standard derivation, em português, desvio padrão).

De forma geral, a análise da dispersão de uma distribuição de dados, quando estamos trabalhando com a análise estatística de dados, parte do coeficiente de variação, que faz uso do desvio padrão e da média.

Medidas de posição: Quartis, limites e outliers

Referências:

ARTES, Rinaldo. **Coeficiente de assimetria**. Insper, 11p, 2014

Parmezan. **Ciência de dados: fundamentos e aplicações**. 1. ed. 2. reimp. Rio de Janeiro: LTC, 2024.

MORETTIN, Pedro Alberto; SINGER, Júlio da Motta. **Estatística e ciência de dados**. Rio de Janeiro: LTC, 2022.