

ESTATÍSTICA DESCRITIVA

A Estatística descritiva é um ramo da Estatística que se concentra na organização, resumo e interpretação de conjuntos de dados.

MÉDIA ARITIMÉTICA:

A média é calculada somando todos os valores de um conjunto de dados e dividindo o resultado pelo número total de observações.

Exemplo:

conjunto de dados {2, 4, 6, 8, 10}

MEDIA: $(2 + 4 + 6 + 8 + 10) / 5 = 6$.

MÉDIA PONDERADA:

A média ponderada é calculada atribuindo pesos diferentes a cada valor do conjunto de dados, com base na sua importância relativa. Cada valor é multiplicado pelo seu peso correspondente, e a soma desses produtos é dividida pela soma total dos pesos.

Exemplo:

Uma faculdade exerce pesos para cada unidade durante o semestre para aprovar os alunos. Sendo eles:

- U1: 0.4

- U2: 0.6

A média para passar no semestre é 6. Durante o semestre foram realizadas 2 provas, uma em cada unidade.

Fulano tirou, 5 na primeira unidade e 9 na segunda unidade.

Cálculo:

$((5 \times 0.4) = 2) + ((9 \times 0.6) = 5,4) = 7,4$. Aprovado no semestre.

MEDIANA:

Representa o valor central de um conjunto de dados ordenado.

1. Ordenamos os dados em ordem crescente ou decrescente
2. Encontramos o valor do meio.
 - a. Se o número total de observações for ímpar, a mediana será o valor central.

- b. Se o número total de observações for par, a mediana será a média dos dois valores centrais.

Exemplo:

(70, 84, 10, 12) → Par

$(84+10) / 2 = 47 \rightarrow$ Mediana

(70, 84, 10, 12, 23) → Ímpar

10 → Mediana

MODA:

É o valor que ocorre com mais frequência em um conjunto de dados.

Para encontrar a moda, devemos identificar o valor que se repete com maior frequência. É possível ter um conjunto de dados unimodal (com apenas uma moda), bimodal (com duas modas) ou multimodal (com mais de duas modas).

Exemplo:

(70, 84, 08, 09, 10, 12, 23, 23, 23, 23)

Frequência: 23

OBSERVAÇÃO: Bimodal ou Multimodal só é considerado com o mesmo numero de frequência das outras modas, fora isso, a moda permanece com a maior frequência.

MEDIDAS DE DISPERSÃO:

→ VARIÂNCIA:

- A variância mede a dispersão dos valores em relação à média. Quanto maior a variância, mais os valores se afastam da média.
- A variância é representada pelo símbolo σ^2 (sigma ao quadrado) quando se trata de uma população e por s^2 (s ao quadrado) quando se trata de uma amostra.
 - Variância para a população: $\sigma^2 = \sum (x_i - \mu)^2 / n$
 - Lendo: sigma ao quadrado = Somatório de $(x_i$ (que é o valor individual) – μ (média)) ao quadrado dividido pelo número total de observações (tamanho da população).
 - Variância para a amostra: $s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$

- Lendo: s^2 ao quadrado = Somatório de $(x_i - \bar{x})^2$ (que é o valor individual - \bar{x} (média da amostra)) ao quadrado dividido pelo tamanho da amostra - 1.

➔ **DESVIO PADRÃO:**

- Raiz quadrada da variância
- $\sqrt{s^2}$ ou s

➔ **DESVIO MÉDIO:**

- É uma medida da dispersão dos dados em relação à média. Ele é calculado somando as diferenças entre cada valor e a média, e depois dividindo pelo número total de observações. O desvio médio é menos sensível a valores extremos em comparação com a variância e o desvio padrão.

❖ **Erro padrão da média:** O erro padrão da média é uma medida de quão precisamente a média amostral representa a média populacional. Ele indica a variabilidade esperada das médias amostrais em torno da média populacional. O erro padrão da média é calculado dividindo o desvio padrão da população pela raiz quadrada do tamanho da amostra.

❖ **Coefficiente de variação:** O coeficiente de variação é uma medida de variabilidade relativa, que compara o desvio padrão com a média. É calculado como o desvio padrão dividido pela média e é expresso como uma porcentagem. O coeficiente de variação permite comparar a dispersão relativa entre diferentes conjuntos de dados, especialmente quando as médias são diferentes.

QUARTIS E INTERVALO INTERQUARTIL (IQR):

- Os quartis dividem o conjunto de dados ordenado em quatro partes iguais, representando os 25%, 50% e 75% dos dados.
- O IQR é a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1), ou seja, $IQR = Q3 - Q1$. Ele fornece uma medida da dispersão dos valores intermediários, desconsiderando os valores extremos.

PROBLEMAS:

1. **Desenho amostral inadequado:** Um desenho amostral inadequado pode levar a uma amostra não representativa da população em estudo. Por exemplo, uma amostra que não é aleatória, mas sim voluntária, pode resultar em viés de seleção, pois certos grupos podem ser sub ou super-representados.
2. **Erros de obtenção de dados:** Erros de obtenção de dados ocorrem quando há falhas na coleta, registro ou processamento dos dados. Isso pode incluir erros de medição, problemas de confiabilidade dos instrumentos de coleta de dados, erros de digitação, entre outros. Esses erros podem levar a dados imprecisos e prejudicar a validade dos resultados.

3. **Amostras viciadas**: Uma amostra viciada é aquela que não representa adequadamente a população. Isso pode ocorrer quando há sub ou super-representação de certos grupos na amostra, seja por problemas no desenho amostral, falta de resposta ou outros fatores. Amostras viciadas podem levar a conclusões errôneas e não generalizáveis.
4. **Populações mal definidas ou dimensionadas**: Populações mal definidas ocorrem quando os critérios de inclusão e exclusão não estão claramente estabelecidos, resultando em incerteza sobre a adequação dos dados coletados. Populações mal dimensionadas ocorrem quando a quantidade de dados disponíveis é insuficiente para realizar inferências confiáveis sobre a população em estudo.
5. **Metodologia de obtenção de dados incongruentes com os objetivos da pesquisa**: É essencial que a metodologia de obtenção de dados esteja alinhada aos objetivos da pesquisa. Se os procedimentos de coleta não são adequados para responder às perguntas de pesquisa, os resultados podem ser inadequados ou irrelevantes.