PROFISSÃO POLICIAL

Estatística

Professor Rodolfo Schmit

ESTATÍSTICA
Professor Rodolfo Schmit

Sumário

ESTATÍSTICA DESCRITIVA: MEDIDAS DESCRITIVAS

1 MEDIDAS DESCRITIVAS

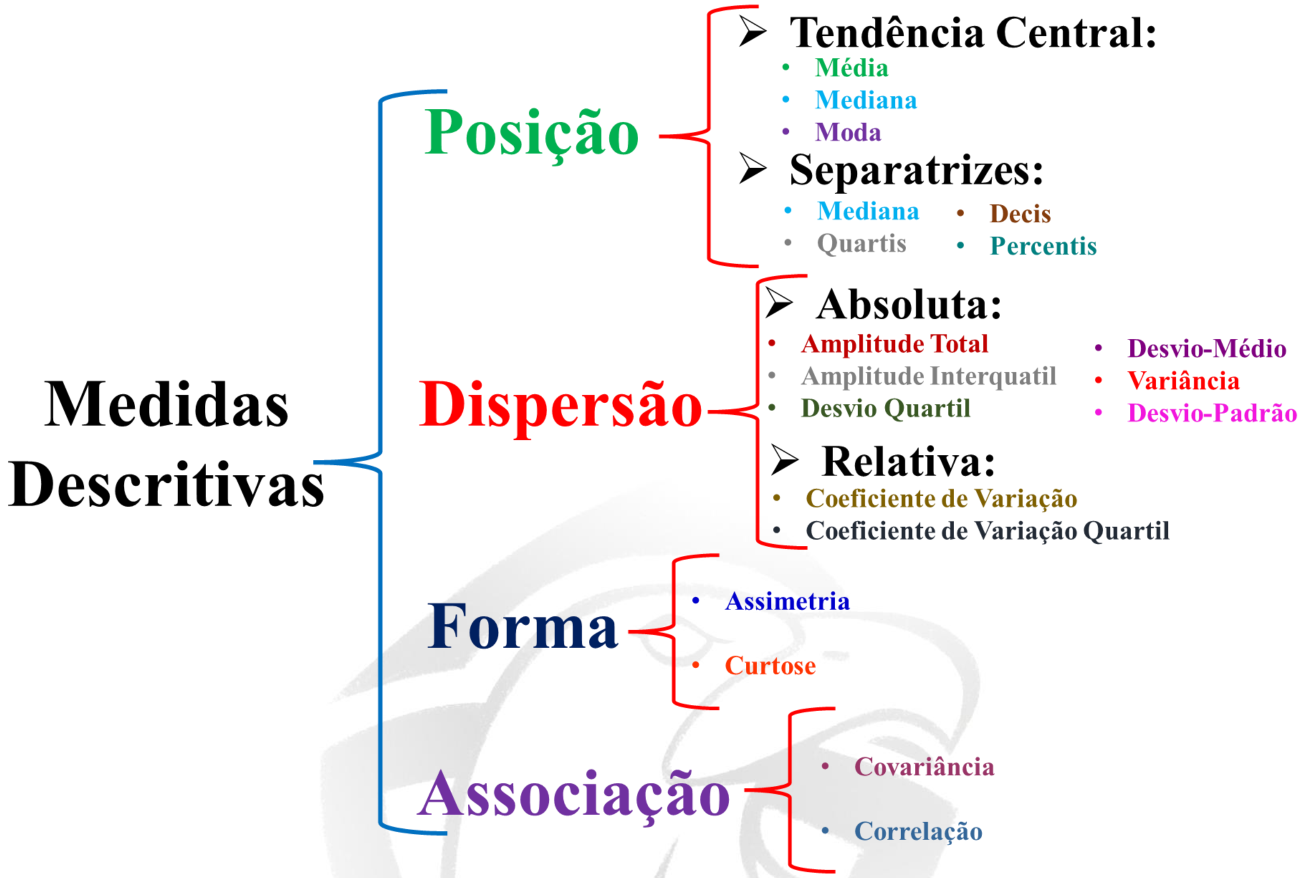
As medidas descritivas são resumos numéricos que tentam expressar o comportamento observado no conjunto de dados. São informações obtidas por cálculos matemáticos que resumem, descrevem e interpretam o fenômeno em estudo. Perante um conjunto de dados grande, elas são altamente eficientes para tornar a informação manejável e, com isso, relacionar os dados e levantar hipóteses.

No entanto, como acontece sempre que se resume algo, esse processo implica a perda de alguma informação mais detalhada. Por isso, conhecer as informações que podem ser obtidas por cada medida descritiva, bem como as informações perdidas é essencial para uma análise descritiva e exploratória. Para alcançar sua completude, as medidas descritivas devem ser analisadas em conjunto, pois cada uma extrai uma informação distinta em relação ao conjunto de dados e quando juntas permitem uma interpretação mais satisfatória. Por exemplo, o valor da Média (medida de tendência central) é frequentemente apresentado em associação com o valor do Desvio Padrão (medida de dispersão).

Sobretudo, as medidas descritivas são classificadas de acordo com o tipo de informação gerada. Desse modo, os tipos de medidas descritivas são: de posição (tendência central e separatrizes); de dispersão (absolutas e relativas); de forma; e de associação. Desse modo, existem quatro grandes grupos de medidas descritivas que direcionam a informações distintas.

- **Posição:** indicam a posição do conjunto de dados na escala de valores que a variável pode assumir. Podem indicar o centro dos dados ou outra posição específica;
- **Dispersão:** indicam se os dados observados são concentrados (próximos uns dos outros) ou dispersos, quantificando a magnitude dessa dispersão;
- **Forma:** indicam a distribuição do conjunto de dados, e associam o formato da representação gráfica que esse conjunto de dados irá assumir;
- **Associação:** indicam a variação associada entre duas variáveis distintas, indicando se há alguma relação entre duas caraterísticas.

Desse modo, as principais medidas descritivas estudadas dentro de cada grupo serão:



2 MEDIDAS DE POSIÇÃO: TENDÊNCIA CENTRAL

Os números permitem que valores quantitativos sejam representados por um dado central e encontrados através de conjuntos finitos e infinitos. Nesse sentido, as medidas de posição referem-se à “localização” do conjunto de dados em relação aos valores que a variável pode assumir, isto é, indicam um valor que está posicionado em algum ponto específico da escala numérica.

As medidas de posição de tendência central informam valores que tendem a estar posicionados no centro do conjunto de dados, ou próximo a isso. As principais medidas de tendência são a média, a mediana e a moda. De forma prática, a utilização dessas três medidas varia consoante o tipo de informação que pretendemos resumir ou descrever.

Em outras palavras, **média**, **moda e mediana** são medidas descritivas usadas para simplificar um conjunto de dados a **um único elemento**, e tentam direcionar para uma informação sobre a tendência central do fenômeno estudado.

Medidas de Tendência Central

Média $(\bar{X})$: **soma de todas observações da variável X** $(\sum X_i)$ **dividido pela quantidade de observações** (n);

$$\bar{X} = \frac{\sum X_i}{n}$$

Mediana (Me): **observação central do conjunto de dados que separa em 50% de elementos abaixo e acima desse valor; Observação de X com Frequência acumulada de 50% (n/2).**

$$Me \rightarrow F(X_i) = 50\% = \frac{n}{2}$$

Moda (Mo): **observação que mais se repete, com maior frequência absoluta ou relativa;**

$$Mo \rightarrow X_i \; com \; maior \; f(X_i)$$

2.1 MÉDIA

A média tenta resumir o conjunto de dados, a partir de operações matemáticas que levam em consideração **todos os valores observados,** para alcançar um único elemento numérico que os representem. Com esse propósito, é a medida que tem o intuito de quantificar o desempenho central (médio) da variável estudada. Logo, a média é considerada como um número que tem a faculdade de representar uma série de valores.

Existem diferentes metodologias de cálculo matemático para obter a média, contudo, todas sempre vão levar em seu cálculo cada valor observado. A metodologia mais simples e comum é a **média aritmética.** Na Estatística, essa média é de longe a mais aplicada e cobrada em concursos. Nesse sentido, o enfoque dessa aula será na metodologia de cálculo da média aritmética. Existem também outras metodologias de cálculo, como média geométrica e harmônica, que também serão abordadas para conhecimento.

Agora, vamos estudar como obter e calcular a média a partir de um conjunto de dados em análise. Os procedimentos serão diferentes de acordo com a apresentação dos dados (dados brutos, ponderados e agrupados).

2.1.1 Cálculo da Média em Dados Brutos

Para o estudo da média em dados brutos, vamos trabalhar com o seguinte exemplo:

Objeto de estudo

Quantidade X de substâncias apreendidas (em mil kg) por operação policial na região do centro oeste brasileiro.

$$X = \{10;\ 18;\ 14;\ 12;\ 17;\ 22;\ 12\}$$
$$n = 7$$

Para obter a média, deve-se **somar todas as observações de X e dividir pela quantidade total de elementos (n)**. O cálculo da média pode ser simbolicamente expresso por:

$$\bar{X} = \frac{\sum X_i}{n}$$

Em que X_i corresponde ao valor de cada observação (na i-ésima observação):

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots X_n}{n}$$

Desse modo, cada valor observado da variável X deve ser somado e depois dividido pelo número de elementos que foram somados. Nesse exemplo, deve ser somado cada quantidade de substâncias apreendidas que foi observada (10, 18, 14, etc.) e dividir por 7, o número total de elementos observados. Veja:

$$\bar{X} = \frac{10 + 18 + 14 + 12 + 17 + 22 + 12}{7}$$

$$\bar{X} = \frac{105}{7} = 15\, mil\; kg/operação$$

Com isso, a média será obtida captando toda a quantidade de substâncias apreendidas e distribuindo igualmente em cada operação policial. Assim, podemos inferir que, em média, foram apreendidas 15 mil kg de substâncias por operação.

A quantidade de substâncias apreendidas foi variável conforme cada operação, contudo, se for necessário representar em um único valor, a média seria uma opção viável.

Outro detalhe importante é que a simbologia para representar a média muitas vezes consiste em realizar um traço em cima do símbolo que representa a variável. Nesse caso, como a variável é representada por X, e média de X será $\bar{X}$.

2.1.2 Cálculo da Média em Dados Ponderados

Quando os dados estão organizados em uma tabela de frequência por valor (dados ponderados), o procedimento matemático para obter a média será um pouco diferente. Segue um exemplo:

Objeto de estudo

Estudo sobre número de indiciados por inquérito policial (variável X) no estado de Roraima no mês de março de 2021. Foram analisados 50 inquéritos policiais

Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
0	3	6%	3	6%
1	15	30%	18	36%
2	20	40%	38	76%
3	10	20%	48	96%
4	2	4%	50	100%
Soma (Σ)	50	100%	-	-

Nesse exemplo, foram analisados 50 inquéritos policiais ao qual foram distribuídos entre os valores de 0 a 4 indiciados por inquérito. Entenda que o conjunto de dados é composto por **50 observações.**

A média deve ser obtida analisando a **frequência absoluta ou relativa**. Para obter a média, é preciso entender que cada observação de X se repete a quantidade de vezes da sua frequência. Então, é preciso somar cada valor de X vezes a sua respectiva frequência e, em seguida, dividir por n.

$$\bar{X} = \frac{\sum X_i \times f_i}{n}$$

$$\bar{X} = \frac{X_1 \times f_1 + X_2 \times f_2 + \cdots + X_n \times f_n}{n}$$

Entenda que cada observação de X se repete tantas vezes conforme indicado pela sua respectiva frequência. Então, ao invés de somar o mesmo valor várias vezes, você multiplica pela quantidade de vezes que ele repete (sua frequência).

Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)
0	3
1	15
2	20
3	10
4	2
Soma (Σ)	50

$$\bar{X} = \frac{0 \times 3 + 1 \times 15 + 2 \times 20 + 3 \times 10 + 4 \times 2}{50}$$

$$\bar{X} = \frac{93}{50} = 1,86\ indiciados/inquérito$$

Em média, os inquéritos policiais indiciam 1,86 suspeitos por inquérito. Ponto interessante a destacar aqui é que apesar da variável ser de natureza discreta, a média pode assumir um valor particionado.

Na essência, o cálculo da média continua o mesmo. A frequência foi utilizada para simplificar o cálculo, pois seria muito mais trabalhoso somar todas as 50 observações uma a uma. É importante que o aluno entenda que a tabela de frequência simplificou e organizou o conjunto de dados, mas é preciso também como os dados seriam apresentados na forma bruta.

Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)
0	3
1	15
2	20
3	10
4	2
Soma (Σ)	50

Dados Brutos

$$0; 0; 0$$

$$1; 1; 1; 1; 1; 1; 1; 1; 1; 1;$$
$$1; 1; 1; 1; 1$$

$$2; 2; 2; 2; 2; 2; 2; 2; 2; 2;$$
$$2; 2; 2; 2; 2; 2; 2; 2; 2; 2$$

$$3; 3; 3; 3; 3; 3; 3; 3; 3; 3$$

$$4; 4$$

$$n = 50$$

O cálculo também pode ser feito usando a frequência relativa, nesse caso, é possível utilizar o número fracionado, ou então, a percentagem (no segundo caso, deve-se dividir por 100 no final). A segunda opção é mais simples e comum, veja:

Número de indiciados por inquérito (X_i)	Frequência Relativa (fr_i)
0	6%
1	30%
2	40%
3	20%
4	4%
Soma (Σ)	100%

$$\bar{X} = \frac{\sum X_i \times fr_i}{100}$$

(fórmula para quando usar frequência relativa em porcentagem)

$$\bar{X} = \frac{0 \times 6 + 1 \times 30 + 2 \times 40 + 3 \times 20 + 4 \times 4}{100}$$

$$\bar{X} = \frac{186}{100} = 1,86\ indiciados/inquérito$$

2.1.3 Cálculo da Média em Dados Agrupados

Para essa demonstração será utilizada o seguinte exemplo:

Objeto de estudo

Dados referentes a quantidade de drogas, em quilogramas, apreendidas por semana em uma delegacia de polícia. O estudo foi analisado no decorrer de 9 semanas.

Valor Observado (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
0⊢10	2	22%	2	22%
10⊢20	4	45%	6	67%
20⊢⊣30	3	33%	9	100%
Soma (Σ_i)	9	100%	-	-

Quando os dados estão agrupados ocorre uma **perda na precisão da informação**, pois não é possível inferir quais são as observações presentes dentro de cada intervalo. Desse modo, para o cálculo da média, assume-se que as observações coincidem com **o ponto médio de cada classe**.

O ponto médio de cada classe é calculado de forma simples, basta somar os limites de cada classe e dividir por dois, da seguinte forma:

$$Pm_i = \frac{LS_i + LI_i}{2}$$

Em que Pm_i corresponde ao ponto médio de uma determinada classe (i-ésima classe); LS_i o limite superior da respectiva classe; LI_i o limite inferior da respectiva classe. Assim, os pontos médios para cada classe são:

$$Pm_{1^{\underline{a}}} = \frac{10 + 0}{2} = 5$$

$$Pm_{2^{\underline{a}}} = \frac{20 + 10}{2} = 15$$

$$Pm_{3^{a}} = \frac{30 + 20}{2} = 25$$

Com isso, a média será calculada da mesma forma que os dados ponderados. No entanto, ao invés de utilizar o valor de cada observação, considera-se que os valores coincidiram com o ponto médio da classe. Assim, associamos cada ponto médio com a respectiva frequência da classe.

Valor Observado (X_i)	Ponto Médio (Pm_i)	Frequência Absoluta (f_i)
0⊢10	5	2
10⊢20	15	4
20⊢⊣30	25	3
Soma (Σ_i)	-	9

Desse modo, trabalha-se com a ideia de que foram observados duas vezes 5kg/semana de droga apreendida; quatro vezes 10 kg/semana; e 3 vezes 15 kg/semana. Por fim, basta somente calcular a média:

$$\bar{X} = \frac{\sum Pm_i \times f_i}{n}$$

$$\bar{X} = \frac{5 \times 2 + 15 \times 4 + 25 \times 3}{9}$$

$$\bar{X} = \frac{145}{9}$$

$$\bar{X} = 16,11\ kg/semana$$

O cálculo também pode ser aplicado com uso da frequência relativa, dessa vez vamos utilizar o valor fracionado e não em porcentagem.

Valor Observado (X_i)	Ponto Médio (Pm_i)	Frequência Relativa (fr_i)
0⊢10	5	22%
10⊢20	10	45%
20⊢30	15	33%
Soma (Σ_i)	-	100%

$$\bar{X} = \sum_{i=1}^{n} Pm_i \times fr_i$$

(fórmula para quando usar frequência relativa em decimal ou fração)

$$\overline{X} = 5 \times 0,22 + 15 \times 0,45 + 25 \times 0,33$$

$$\overline{X} \cong 16,11\ kg/semana$$

O cálculo da média com perda na precisão dos dados gerará uma média diferente caso fosse calculada com os dados completos (brutos ou ponderados). Quando se trabalha com grande número de observações, principalmente em variáveis contínuas, simplificar o conjunto de dados torna-se uma alternativa vantajosa para organização dos dados, mesmo com imprecisão no valor exato da média.

2.1.4 Outros Tipos de Média

A média de um conjunto de dados pode ser obtida por outros cálculos matemáticos, como já apontado. Apesar de cálculos distintos, todas as metodologias tentam resumir uma ideia de centralidade. Vamos estudos outros tipos de média para conhecimento. Além da média aritmética que trabalhamos até o momento existem também a média geométrica e média harmônica.

TIPOS DE MÉDIA
- **Média Aritmética ($\overline{X}$)**
- **Média Geométrica ($\overline{G}$)**
- **Média Harmônica ($\overline{H}$)**

2.1.4.1 Média Aritmética ($\overline{X}$):

Como já abordado, a média aritmética é obtida pela soma de todas as observações do conjunto de dados dividido pelo número de observações. Assim, pode ser representada matematicamente:

$$\overline{X} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + X_3 + \cdots X_n}{n}$$

Essa média é amplamente utilizada para os fenômenos em estudo estatístico.

2.1.4.2 Média Geométrica ($\overline{G}$):

Esse método de cálculo considera o princípio da multiplicação. O cálculo é efetuado multiplicando cada observação e extraindo a raiz quadrada na potência equivalente ao número de observações. Assim:

$$\overline{G} = \sqrt[n]{\prod X_i} = \sqrt[n]{X_1 \times X_2 \times X_3 \times \ldots X_n}$$

Essa média é aplicada no estudo de fenômenos quem apresentam crescimento e variações na escala de progressões geométricas. Foi criada com essa finalidade e é aplicada em estudos mais avançados, em que considerar a média geométrica irá representar melhor o desempenho do fenômeno estudado.

2.1.4.3 Média Harmônica ($\overline{H}$):

O cálculo da média harmônica é efetuado invertendo a fração de cada observação e invertendo a fração principal do cálculo de uma média aritmética simples. Assim, temos que o total de observações (n) é dividido pelo somatório do inverso das observações. Portanto, pode ser calculada da seguinte forma:

$$\bar{H} = \frac{n}{\sum \frac{1}{X_i}} = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \cdots \frac{1}{X_n}}$$

Basicamente, esse tipo de cálculo é recomendado quando envolve fenômenos que possuem a relação de ser inversamente proporcionais (exemplo, velocidade e tempo) uns aos outros. É também comum de ser aplicada em estudos de economia que envolvem taxas, geralmente a média de taxas é feita por uma média harmônica.

2.1.4.4 Relação entre os três tipos de média

Dos três tipos de médias apresentadas (aritmética, geométrica e harmônica), a mais importante e utilizada é a **média aritmética**. Para a maioria das provas de Estatística, os cálculos da média geométrica e harmônica não são cobrados. O mais importante é entender que todas as metodologias de cálculos estão preocupadas em quantificar uma tendência central do conjunto de dados. Em últimos casos, existe uma relação entre os três tipos de médias que podem ser essenciais para solucionar uma questão que envolvam esse conteúdo.

Para exemplificar essa relação, vamos calcular os três tipos de médias utilizado o seguinte conjunto de dados:

$$X = \{1, 3, 9\}$$

- Média Aritmética Simples:

$$\bar{X} = \frac{1+3+9}{3} = \frac{13}{3} = 4,33$$

- Média Geométrica:

$$\bar{G} = \sqrt[3]{1 \times 3 \times 9} = \sqrt[3]{27} = 3$$

- Média Harmônica:

$$\bar{H} = \frac{3}{\frac{1}{1} + \frac{1}{3} + \frac{1}{9}} = \frac{3}{\frac{13}{9}} = \frac{3 \times 9}{13} = \frac{27}{13} = 2,08$$

Com base no exemplo abordado, entenda que para o mesmo conjunto de dados a relação entre os tipos de média será:

$$\bar{X} \geq \bar{G} \geq \bar{H}$$

A média aritmética será sempre maior que a média geométrica que, por sua vez, será maior que a média harmônica. Somente serão iguais, quando os valores do conjunto de dados forem idênticos entre si, por exemplo, X = {2, 2, 2, 2, 2}. Essa relação pode não ser verdadeira quando a variável assumir ao menos um valor negativo.

2.1.5 Média Simples e Ponderada

A média também pode ser classificada em simples ou ponderada de acordo com sua metodologia de cálculo. A diferença está no fato de atribuir **pesos diferentes** para cada um dos valores observados, ou não. Nesse sentido, quando os valores possuem todo o mesmo peso (todos os valores com peso igual a 1), teremos o cálculo da **média simples**. Em contrapartida, quando atribuímos um peso de importância para cada valor observado, utilizaremos a metodologia de cálculo de **média ponderada**.

A média ponderada é frequentemente utilizada para obter notas ou conceitos atribuídos em avaliações. Assim, cada tipo de avaliação possui um peso de acordo com a sua relevância. Quanto maior o peso, maior será a relevância daquele valor para o cálculo da média. A partir de um exemplo, vamos aplicar o cálculo da média aritmética por meio de metodologia simples e ponderada.

Exemplo:

Para obter a média semestral de um ensino superior, um aluno realizou cinco metodologias de avaliação diferentes: prova objetiva, apresentação oral, redação argumentativa, trabalho em grupo e atividades de fixação. Nesse sentido, o professor atribuiu pesos iguais a 3, 2, 2, 1, 1, respectivamente a cada metodologia de avaliação diferentes. Em cada avaliação o aluno obtinha uma nota de 0 a 10. O desempenho do aluno é apresentado na tabela a seguir.

METODOLOGIA DE AVALIAÇÃO	PESO (P)	NOTA (X)
Prova Objetiva	3	8
Apresentação	2	4
Redação	2	9
Trabalho em Grupo	1	8
Atividades	1	6

Nessa situação, para se obter a média ponderada ($\bar{X}_p$) do desempenho desse aluno, é necessário multiplicar a nota por seu respectivo peso e dividir pelo somatório dos pesos de cada nota. Dessa forma o cálculo da média ponderada seria o seguinte:

$$\bar{X}_p = \frac{\sum X_i \times P_i}{\sum P_i}$$

$$\bar{X}_p = \frac{X_1 \times P_1 + X_2 \times P_2 + \cdots + X_n \times P_n}{P_1 + P_2 + \cdots + P_n}$$

Dessa forma, de acordo com o exemplo, temos o seguinte cálculo de média ponderada:

$$\bar{X}_p = \frac{8 \times 3 + 4 \times 2 + 9 \times 2 + 8 \times 1 + 6 \times 1}{3 + 2 + 2 + 1 + 1}$$

$$\bar{X}_p = \frac{24 + 8 + 18 + 8 + 6}{9}$$

$$\bar{X}_p = \frac{64}{9} = 7,11$$

Desse modo, a média ponderada do desempenho semestral do aluno é de 7,11. Observe que o raciocínio matemático da média ponderada é semelhante do aplicado a média em dados ponderados.

No mesmo exemplo, se fosse aplicado o cálculo da média simples ($\bar{X}$), não seria considerado os diferentes pesos atribuídos a cada nota. Basicamente, na média simples, soma-se cada nota e divide-se pelo total de observações (o equivalente a cada nota ter peso igual a 1). Portanto, o cálculo da média aritmética simples seria igual ao já aplicado nos exemplos anteriores:

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots X_n}{n}$$

$$\bar{X} = \frac{8 + 4 + 9 + 8 + 6}{5}$$

$$\bar{X} = \frac{35}{5} = 7$$

Destarte, o valor da média semestral, nesse exemplo, para uma metodologia simples seria igual a 7.

2.1.6 Características da Média

Primeiramente, a média é um valor numérico que sempre ficará localizada entre os valores extremos da variável. Assim:

$$\textit{Valor Mínimo} \leq \textit{Média} \leq \textit{Valor Máximo}$$

A média é uma medida que tenta quantificar a centralidade do conjunto de dados, e leva em consideração **todos os valores observados em seu cálculo.**

Desse modo, é também a medida de posição **mais sensível** com a presença de valores discrepantes (extremos) em comparação aos outros valores observados. Nessa situação, ela não vai representar a centralidade do conjunto dos dados e torna-se uma medida descritiva não interessante para representar o centro do conjunto de dados.

Veja o exemplo de cálculo de média **sem dados discrepantes**:

$$X = \{10;\ 18;\ 14;\ 12;\ 17;\ 22;\ 12\}$$
$$n = 7$$

$$\bar{X} = \frac{10 + 18 + 14 + 12 + 17 + 22 + 12}{7}$$

$$\bar{X} = \frac{105}{7} = 15$$

Agora, veja o que acontece com o valor da média com **dados discrepantes**, isto é, caso seja observado um valor muito atípico dos demais:

$$X = \{10;\ 18;\ 14;\ 12;\ 94;\ 22;\ 12\}$$
$$n = 7$$

$$\bar{X} = \frac{10 + 18 + 14 + 12 + 94 + 22 + 12}{7}$$

$$\bar{X} = \frac{182}{7} = 26$$

No primeiro caso, com valores de X variando de 10 a 22, observa-se que a média foi igual a 15, e tem uma melhor representatividade do centro dos dados, pois possui quatro observações abaixo da média 15 {10, 12, 12, 14} e três observações acima da média 15 {17, 18, 22}.

No segundo caso, a média teve valor igual a 26. Veja que abaixo da média há seis observações {10, 12, 12, 18, 22}, enquanto acima de média há apenas um observações de X {94}. Isso ocorreu porque o valor 94 é muito discrepante dos valores observados. Assim, a média se tornou um valor muito superior a maioria das observações. Já não representa a centralidade do conjunto de dados. Portanto, pode ser afirmado que **a média é influenciada por cada um e por todos os valores da variável.**

Outra característica da média é que o valor da média **não necessariamente coincide com um dos valores que a compõem**. Como foi visto nos exemplos anteriores, a média 15 ou 26, não é observada no conjunto de dados X.

2.1.7 Média Global

Outro conceito aplicado a média chama-se **média global**, que consiste em calcular a média das médias de 2 ou mais grupos.

É muito comum, por exemplo, aplicar a média global quando o aluno precisa obter a média do seu ano letivo, nesse caso, o aluno calcula a média das médias de todas as disciplinas. Assim, ele irá obter a média global do seu ano letivo.

Imagine um exemplo em que, no ano letivo, um aluno tem notas de duas disciplinas X_1 e X_2. Para cada disciplina, o aluno possui uma média $\bar{X}_1$ e $\bar{X}_2$:

$$X_1 = \{10; 8; 8; 7; 5; 4\} \quad n = 6 \qquad \bar{X}_1 = \frac{42}{6} = 7$$

$$X_2 = \{5; 7; 0; 4; 9\} \quad n = 5 \qquad \bar{X}_2 = \frac{25}{5} = 5$$

Para a disciplina X_1, o aluno teve seis avaliações e obteve média igual a 7. Já na disciplina X_2, o aluno teve cinco avaliações e obteve média igual a cinco. Nesse sentido, a média global ($\bar{X}_G$) do ano letivo foi de:

$$\bar{X}_G = \frac{7 + 5}{2} = 6$$

Logo, a média das médias de cada disciplina foi igual a 6, que corresponde ao desempenho do ano letivo do aluno.

2.2 MEDIANA

A mediana é uma medida que separa o conjunto de dados em exatamente **50% para cada lado**, por isso ela é um valor de referência para indicar a observação que está exatamente no centro. Isto é, a mediana é um valor que separa a metade superior e inferior de um conjunto de dados ordenados (nos diz que a quantidade de valores é a mesma tanto antes quanto depois da mediana). Nesse sentido, a mediana é uma medida ideal para ser utilizada quando o objetivo for classificar os elementos avaliados e distingui-los quanto ao desempenho na metade.

Para calcular e identificar a mediana, é necessário que o conjunto de dados estejam ordenado de forma crescente, **em rol**. Além disso, se a quantidade de elementos for ímpar, o valor da mediana corresponde ao valor de central do conjunto de dados. Todavia, se a quantidade de elementos for par, é preciso obter a média dos valores centrais para obter a mediana.

2.2.1 Cálculo da Mediana em Dados Brutos

Objeto de estudo:

Quantidade de substâncias apreendidas (em mil kg) apreendidas por operação policial na região do centro oeste brasileiro.

$$X = \{10;\ 18;\ 14;\ 12;\ 17;\ 22;\ 12\}$$

$$n = 7$$

Inicialmente, os dados devem ser colocados em ordem crescente, após isso deve ser identificada a posição central. A identificação pode ser de forma visual ou calculando a posição do centro. A mediana só pode ser obtida nessa condição, os **dados brutos devem estar ordenados**.

$$\xrightarrow{\textit{Rol Crescente}}$$

$$X = \{10; 12; 12; 14;\ 17; 18; 22\}$$

Ao analisar um pequeno conjunto de dados, como esse exemplo de apenas 7 elementos, visualmente conseguimos detectar a posição central do conjunto de dados e indicar qual valor corresponde a mediana. Veja:

$$Me = 14$$

$$X = \{\underbrace{\underset{1^{a}}{10}; \underset{2^{a}}{12}; \underset{3^{a}}{12}}_{\substack{\text{50\% Inferior} \\ \text{(3 elementos)}}}; \underset{4^{a}}{14}; \underbrace{\underset{5^{a}}{17}; \underset{6^{a}}{18}; \underset{7^{a}}{22}}_{\substack{\text{50\% Superior} \\ \text{(3 elementos)}}}\}$$

$$Me = 14\ mil\ kg/\ operação$$

Observe que a mediana corresponde ao valor 14 que está na quarta posição de um total de sete elementos ordenados. Nesse sentido, considerando o valor 14 a mediana (o centro dos dados), observe que, antes da mediana, há 3 elementos e, após a mediana, há também 3 elementos. Isto é, o conjunto de dados foi particionado de modo que tenha 50% dos dados mais inferiores e 50% de dados superiores.

Quando for obter a mediana em um conjunto de dados muito extenso, identificar visualmente o centro pode ser um pouco difícil, para torna-se mais interessante calcular **a posição central** dos dados e identificar a mediana. Para qualquer conjunto de dados brutos, você pode achar a posição central (da mediana P_{Me}) pela seguinte expressão:

$$P_{Me} = \frac{n+1}{2}$$

$$P_{Me} = \frac{7+1}{2} = 4^{a}\ posição$$

Logo, o dado observado que estiver na 4ª posição **(quando ordenado),** corresponde a mediana. A mediana não é o valor 4, e sim a quarta observação ordenada, não cometa esse equívoco. Veja novamente:

$$Me = 14$$

$$X = \{\underset{1^a}{10}; \underset{2^a}{12}; \underset{3^a}{12}; \underset{4^a}{14}; \underset{5^a}{17}; \underset{6^a}{18}; \underset{7^a}{22}\}$$

$$Me = 14\ mil\ kg/\ operação$$

Ao interpretar essa medida descritiva, sabe-se que a mediana é igual a 14 mil kg de substâncias apreendidas por operação policial. Nesse sentido, podemos inferir que 50% das operações policiais registradas tiveram valor igual ou menor que 14 mil kg, bem como, 50% das operações tiveram valor igual ou maior que 14 mil kg.

Nesse o exemplo abordado, temos um total de 7 observações (n=7) que é um número ímpar. Nessa situação, o valor central será um dado observado do conjunto de dados. Contudo, quando o total de observações for um número par, teremos dois elementos centrais, e a mediana será a **média desse duas observações**. Suponha que o exemplo anterior tenha 6 observações:

$$X = \{10; 12; 12; 14; 17; 22\}$$
$$n = 6$$

Desse modo, a posição central do conjunto de dados será a seguinte:

$$P_{Me} = \frac{6+1}{2} = 3,5$$

O valor da posição foi de 3,5. Assim, entende-se que o centro do conjunto dos dados ordenados está **entre o terceiro e quarto elemento**. Logo, obtemos a média entre esses dois valores centrais para calcular a média.

$$Me = \frac{12+14}{2} = 13$$

$$X = \{10; 12; 12; 14; 17; 22\}$$

1ª 2ª 3ª 4ª 5ª 6ª

$$Me = 13\ mil\ kg/\ operação$$

2.2.2 Cálculo da Mediana em Dados Ponderados

Para obter o valor da mediana em dados ponderados, a melhor informação que indica a posição do conjunto de dados está na **frequência acumulada**. Isso porque essa frequência acumula os valores das observações anteriores e, de certa forma, contabiliza o número de elementos e indica sua posição.

Desse modo, basta identificar, na frequência acumulada, onde está a posição central e ver qual observação corresponde a essa posição. Se a mediana separa o conjunto de dados ordenados no meio, logo, ela acumula 50% da frequência acumulada (n/2):

$$Me \rightarrow F(X_i) = 50\% = \frac{n}{2}$$

Na tabela de frequência não procuramos a posição central, e sim o valor de X que acumula 50% dos dados. Logo, nessa situação, não precisamos somar **n+1,** como nos dados brutos. Apenas é preciso, achar a metade dos dados, ou seja, n/2.

Vamos aplicar em um exemplo:

Objeto de estudo

Estudo sobre número de indiciados por inquérito policial (variável X) no estado de Roraima no mês de março de 2021. Foram analisados 50 inquéritos policiais

Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
0	3	6%	3	6%
1	15	30%	18	36%
2	20	40%	38	76%
3	10	20%	48	96%
4	2	4%	50	100%
Soma (Σ)	50	100%	-	-

Para 50 observações, a mediana será o valor de X que acumula **50%** dos dados ou que acumula **25 observações** (50/2). Ao analisar a tabela de frequência, dando enfoque apenas para frequências acumuladas, podemos alcançar o valor da mediana, veja:

	Número de indiciados por inquérito (X_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
	0	3	6%
	1	18	36%
Me	$\boxed{2}$	38	76%
	3	48	96%
	4	50	100%

$$Me = 2\ indiciados/\ inquérito$$

O valor de X=1 acumula até 18 observações, logo, ele está abaixo da metade do conjunto de dados. Nesse raciocínio, quando observamos o valor X=2, é acumulado 38 observações. Logo, o valor de X=2 contempla a 25ª que seria a metade dos dados. Para visualizar melhor isso, imagine que as cinquentas observações estão listadas de forma bruta em ordem crescente, com isso, seria possível constatar que o elemento presente na 25ª posição seria o valor 2 indiciados/ inquérito.

$$\mathrm{Me} = 2$$

$$X = \{\underset{1^a}{0}; \underset{2^a}{0}; \underset{3^a}{0}; \underset{4^a}{1}; \underset{5^a}{1}; \underset{6^a}{1}; \ldots; \underset{19^a}{2}; \underset{20^a}{2}; \underset{21^a}{2}; \underset{22^a}{2}; \underset{23^a}{2}; \underset{24^a}{2}; \underset{25^a}{2} \mid \underset{26^a}{2}; \underset{27^a}{2}; \ldots\}$$

$$n = 50$$

Portanto, 50% dos inquéritos policiais dessa região indiciam 2 ou menos suspeitos, bem como, 50% dos inquéritos indiciam 2 ou mais suspeitos.

2.2.3 Cálculo da Mediana em Dados Agrupados

Para demonstrar esse cálculo, vamos aplicar o seguinte exemplo:

Objeto de estudo

Dados referentes a quantidade de drogas, em quilogramas, apreendidas por semana em uma delegacia de polícia. O estudo foi analisado no decorrer de 9 semanas.

Valor Observado (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
$0 \vdash 10$	2	22%	2	22%
$10 \vdash 20$	4	45%	6	67%
$20 \vdash\dashv 30$	3	33%	9	100%
Soma (Σ_i)	9	100%	-	-

Para calcular a mediana em conjunto de dados agrupados, é preciso inicialmente identificar a classe mediana, ou seja, a classe com o intervalo de valores que engloba a mediana.

	Valor Observado (X_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
	$0 \vdash 10$	2	22%
Classe Mediana	$\boxed{10 \vdash 20}$	6	67%
	$20 \vdash\dashv 30$	9	100%

A identificação da classe mediana funciona da mesma forma que a identificação da mediana nos dados ponderados. Como os dados são agrupados não é possível identificar o valor exato da mediana, e sim a classe em que ela se encontra. Para isso, basta localizar a classe imediatamente superior que acumula metade do total dos elementos ($\frac{\sum f_i}{2}$ ou $\frac{n}{2}$) na frequência acumulada, ou então, que acumula um pouco mais de 50% na frequência acumulada relativa.

Com isso, é possível inferir que a mediana está localizada entre o valor 10 kg/semana até 20 kg/semana. Para calcular o valor exato da mediana, é necessário utilizar a **interpolação linear** (uma técnica para estimar a mediana de um conjunto de dados agrupados em classes). Para calcular a mediana por interpolação linear, você precisa saber o número total de observações, a frequência acumulada até a classe mediana e a amplitude da classe. Quando se trabalha com dados agrupados, utiliza-se esse método de cálculo para estimar o valor dentro do intervalo que corresponde proporcionalmente o valor que acumula a metade dos dados.

O cálculo da interpolação linear trabalha com a ideia de que existe uma razão constante entre **a diferença dos valores observados com a diferença de sua respectiva frequência acumulada (ou acumulada relativa).** Veja a relação matemática:

$$\frac{20 - 10}{6 - 2}$$

$20 - 10$ → Diferença entre os valores observados

$6 - 2$ → Diferença entre frequências acumuladas dos respectivas valores de X

Entenda que, pela tabela de frequência dos dados agrupados, podemos associar que com 10 kg/semana acumula-se 2 observações e com 20 kg/semana acumula-se 6 observações. Essa conclusão é obtida conferindo o limite superior da classe e sua respectiva frequência acumulada. Nesse mesmo raciocínio, sabemos que o valor da mediana (ainda não conhecido) certamente possui uma frequência acumulada de 50% ou nesse exemplo em que n=9, de n/2, isto é, 4,5.

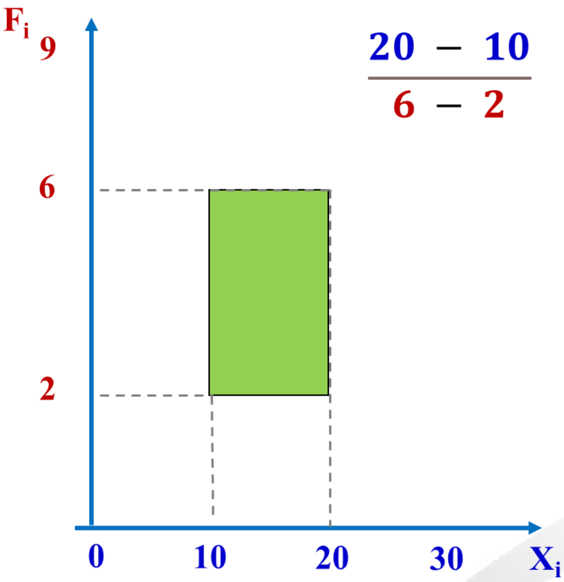
$$F(Me) = \frac{n}{2} = \frac{9}{2} = 4,5 \; ou$$

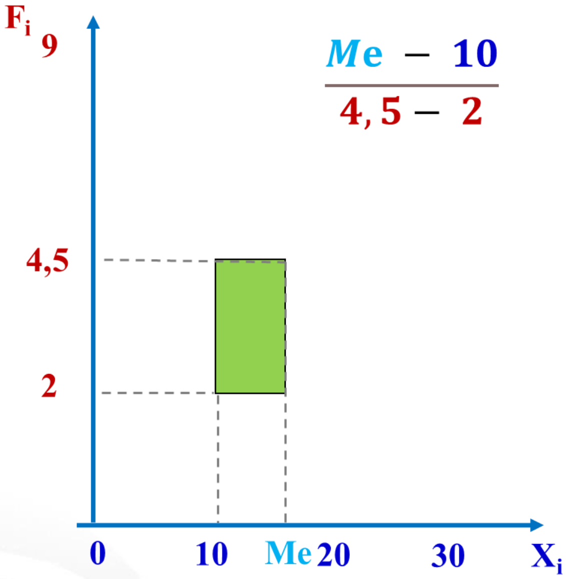
$$F(Me) = 50\%$$

Desse modo, ao associar os valores de X e sua respectiva frequência acumulada, tem-se:

Valor Observado (X_i)	Frequência Acumulada (F_i)
10	2
Me	4,5
20	6

A divisão dessas diferenças estabelece uma relação de proporção com qualquer outra relação nesse conjunto de dados. Assim, com o conhecimento da técnica de interpolação linear, pode ser estabelecido uma igualdade, que envolva a mediana, entre razões da diferença dos valores X sob a diferença das suas respectivas frequências acumuladas. Veja uma possibilidade:





Pela técnica de interpolação linear, essas duas possíveis razões vão gerar sempre a mesma constante. Portanto, é possível igualar as duas razões e descobrir o valor da mediana.

$$\frac{20 - 10}{6 - 2} = \frac{Me - 10}{4,5 - 2}$$

$$\frac{10}{4} = \frac{Me - 10}{2,5}$$

$$2,5 = \frac{Me - 10}{2,5}$$

$$2,5 \times 2,5 = Me - 10$$

$$6,25 = Me - 10$$

$$Me = 10 + 6,25 = 16,25$$

$$Me = 16,25\ \text{kg/semana}$$

Enquanto esse cálculo é efetuado, o aluno deve entender **que nunca obterá um valor que extrapole o limite da classe mediana**, assim se porventura ocorrer algum erro no cálculo que passe desse valor, é interessante revisar os cálculos, pois certamente houve algum erro.

2.2.4 Mediana em Variáveis Qualitativas

A mediana somente pode ser aplicada para variáveis ao qual realizar ordenação. Nesse sentido, variáveis **qualitativas ordinais** obviamente são possíveis de se obter a mediana. Da mesma forma, podemos concluir que a mediana não pode ser aplicada as variáveis qualitativas nominais, uma vez que não é possível ordenar os dados.

Sobretudo, a metodologia para se obter a mediana em variáveis qualitativas ordinais não se diferencia da forma aplicada para variáveis quantitativas. Só é preciso entender que cada valor não numérico, uma vez ordenado, terá uma posição central que divide o conjunto de dados no meio. Vamos aplicar em um exemplo:

Objeto de estudo

Um estudo aplicado a 50 pessoas apresenta a distribuição do nível de experiência de trabalho em tecnologia da informação.

Nível de Experiência (X_i)	Frequência (f_i)
Amador	10
Intermediário	12
Avançado	17
Profissional	11
Total	50

A variável qualitativa nível de experiência apresenta os valores {amador, intermediário, avançado, profissional}. Desses valores, observa-se a distribuição da frequência ao analisar 50 pessoas. Para obter a mediana, basta encontrar o valor da variável que acumula 50% dos dados observados (que nesse exemplo, corresponde a n/2=25). Com isso, basta a frequência acumulada e encontrar a mediana:

	Nível de Experiência (X_i)	Frequência (f_i)	Frequência Acumulada (F_i)
	Amador	10	10
	Intermediário	12	22
Mediana	Avançado	17	39
	Profissional	11	50
	Total	50	-

Por fim, observa-se que a mediana corresponde ao nível de experiência avançado. Logo, podemos afirmar que 50% das pessoas analisadas são amadores, intermediários ou avançados, e que as outras 50% das pessoas possuem nível de experiência avançado ou profissional.

2.2.5 Características da Mediana

A mediana **não necessariamente será um valor presente no conjunto de dados**. Quando apresentar um tamanho de conjunto de dados de valor ímpar certamente será um valor observado, porém, quando for par, será a média dos valores centrais e pode não ser um valor observado.

A mediana, ao contrário da média, não depende de todos os valores observados, além disso, **sofre baixa influência de valores extremos**. Apenas algumas observações discrepantes podem alterar levemente a posição e o valor da mediana, mas não alterará em grande escala como ocorre com a média. A mediana não fornecerá resultados arbitrariamente grandes ou pequenos desde que **mais da metade dos dados não esteja contaminada**.

Portanto, a mediana é adequada quando os dados observados apresentam **grande variabilidade**, distribuição assimétrica, além de valores extremos indefinidos. A vantagem da mediana em relação à média é que a mediana pode dar uma ideia melhor de **um valor típico** porque não é tão distorcida por valores extremamente altos ou baixos.

2.3 MODA

A moda é o valor de uma variável que mais se repete no conjunto de dados. Em outras palavras, é o valor com a **maior frequência (maior incidência)**, ou então, valor com maior **probabilidade** de ocorrer.

2.3.1 Cálculo da Moda em Dados Brutos

Na situação de um conjunto de dados na forma bruta, para obter a moda, basta identificar visualmente o valor da variável que mais se repete no conjunto de dados. Vamos verificar em um exemplo:

Objeto de estudo:

Quantidade de substâncias apreendidas (em mil kg) apreendidas por operação policial na região do centro oeste brasileiro.

$$X = \{10;\ 18;\ 14;\ 12;\ 17;\ 22;\ 12\}$$
$$n = 7$$

Nesse exemplo, com uma pequena quantidade de dados, pode ser constatado que a moda é igual a 12 kg/semana. Veja:

$$X = \{10;\ 18;\ 14;\ 12;\ 17;\ 22;\ 12\}$$

$$Mo = 12\ mil\ kg/\ operação$$

Logo, a quantidade de substância apreendida mais comum de ser apreendida no decorrer das 7 operações policiais foi de 12 mil kg de drogas.

2.3.2 Cálculo da Moda em Dados Ponderados

Em situação de dados ponderados (em tabela de frequência), para identificar a observação que corresponde à moda, devem-se utilizar as informações presente na tabela de **frequência absoluta ou relativa**. Praticamente, a observação que possuir maior valor de frequência absoluta ou relativa será a moda. Vamos aplicar em um exemplo:

Objeto de estudo

Estudo sobre número de indiciados por inquérito policial (variável X) no estado de Roraima no mês de março de 2021. Foram analisados 50 inquéritos policiais

Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
0	3	6%	3	6%
1	15	30%	18	36%
2	20	40%	38	76%
3	10	20%	48	96%
4	2	4%	50	100%
Soma (Σ)	50	100%	-	-

Para obter a moda na tabela de frequência, basta procurar o valor de X com maior frequência (maior quantidade de repetições). Assim, para a moda, devemos analisar a **frequência absoluta ou relativa**.

Nesse exemplo, temos:

	Número de indiciados por inquérito (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)
	0	3	6%
	1	15	30%
Mo	$\boxed{2}$	20	40%
	3	10	20%
	4	2	4%

$$Mo = 2\ indiciados/\ inquérito$$

Logo, o mais comum nos inquéritos policiais é indiciar 2 suspeitos.

2.3.3 Cálculo da Moda em Dados Agrupados

Para demonstrar esse cálculo, vamos aplicar o seguinte exemplo:

Objeto de estudo

Dados referentes a quantidade de drogas, em quilogramas, apreendidas por semana em uma delegacia de polícia. O estudo foi analisado no decorrer de 9 semanas.

Valor Observado (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)	Frequência Acumulada (F_i)	Frequência Acumulada Relativa (Fr_i)
0⊢10	2	22%	2	22%
10⊢20	4	45%	6	67%
20⊢⊣30	3	33%	9	100%
Soma (Σ_i)	9	100%	-	-

Como as observações estão agrupadas em classes, é necessário, primeiramente, identificar a classe que engloba a moda, denominada de **classe modal**. Para isso, basta identificar a classe com maior frequência absoluta ou relativa (igualmente como para dados ponderados). Portanto:

	Valor Observado (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)
	$0 \vdash 10$	2	22%
Classe Modal	$\boxed{10 \vdash 20}$	4	45%
	$20 \vdash\!\dashv 30$	3	33%
	Soma (Σ_i)	9	100%

Após essa etapa, é preciso calcular o valor pontual da moda, que estará dentro dos limites da classe modal. Para isso, existem quatro principais metodologias matemáticas diferentes que podem ser utilizadas.

Em se tratando de dados agrupados, a moda é fortemente afetada pela maneira como as classes são constituídas. Isto faz com que distribuições de frequência do mesmo conjunto de dados elaboradas de formas diferentes (com número de classes diferentes) podem representar valores modais diferentes.

2.3.3.1 Método da Moda Bruta:

É o método mais simples; consiste em tomar como Moda o ponto médio da classe modal. Assim:

$$Mo = \frac{10 + 20}{2} = 15\ kg/semana$$

2.3.3.2 Método da Moda Pearson

É calculada por meio da média e da mediana a partir da seguinte expressão:

$$Mo = 3Me - 2\overline{X}$$

É a diferença entre três vezes o valor da mediana e duas vezes o valor da média. Logo, para o exemplo em questão, os valores da média e da mediana são respectivamente $\overline{X}$ = 16,11 e Me = 16,25. Assim, a moda será:

$$Mo = 3 \times 16,25 - 2 \times 16,11$$

$$Mo = 48,75 - 32,22 = 16,53\ kg/semana$$

2.3.3.3 Método da Moda de Czuber

Essa metodologia estima a moda baseado nos valores de frequência das classes modal, anterior a modal e posterior a modal. O cálculo é feito pela seguinte fórmula:

$$Mo = Li + h\frac{f_{Modal} - f_{Ant.}}{2f_{Modal} - (f_{Ant.} + f_{Post.})}$$

Li: corresponde ao limite inferior da classe modal; Li = 10
h: corresponde à amplitude da classe modal; h =10
f_{Modal} : frequência absoluta da classe modal; $f_{Modal} = 4$
$f_{Ant.}$: frequência da classe anterior à classe modal; $f_{Ant.} = 2$
$f_{Post.}$: frequência da classe posterior à classe modal; $f_{Post.} = 3$

Desse modo, os valores correspondentes a cada frequência podem ser encontrados:

h=10

Valor Observado (X_i)	Frequência Absoluta (f_i)	Frequência Relativa (fr_i)
0⊢10	$f_{Ant.}$ ← 2	22%
Li ← 10⊢20	$\boxed{4}$ → f_{Modal}	45%
20⊢30	$f_{Post.}$ ← 3	33%
Soma (Σ_i)	9	100%

Com isso, o cálculo é efetuado da seguinte forma:

$$Mo = 10 + 10 \times \frac{4 - 2}{2 \times 4 - (2 + 3)}$$

$$Mo = 10 + 10 \times \frac{2}{8 - 5}$$

$$Mo = 10 + \frac{20}{3}$$

$$Mo = 10 + 6,666 = 16,66\ kg/semana$$

Essa fórmula de Czuber também pode ser apresentada da seguinte forma:

$$Mo = Li + h\,\frac{d_A}{d_A + d_P}$$

Em que os termos d_A e d_P representam as distâncias da frequência da classe modal com a frequência da classe anterior a modal e com a frequência da classe posterior a modal, respectivamente. Isto é:

$$d_A = f_{Modal} - f_{Ant.}$$

$$d_P = f_{Modal} - f_{Post.}$$

Em relação a fórmula anterior apresentada, se for colocado as distâncias em evidência, matematicamente, serão as mesmas fórmulas. O aluno pode encontrar qualquer uma das duas fórmulas em seus estudos. Particularmente, essa segunda fórmula é mais fácil de memorizar. Assim, ao resolver a fórmula, será obtido o mesmo resultado.

$$d_A = 4 - 2 = 2$$

$$d_P = 4 - 3 = 1$$

$$Mo = 10 + 10\,\frac{2}{2 + 1}$$

$$Mo = 10 + \frac{20}{3} = 16,66\ kg/semana$$

É interessante observar que a fórmula do Czuber consiste em multiplicar a amplitude da classe (h) por uma fração que terá valor positivo menor do que 1 $\left(\frac{d_A}{d_A+d_P}\right)$. Após isso a fórmula soma com a amplitude do limite inferior da classe modal.

Com isso, podemos concluir que o valor da moda, obtida por essa metodologia, depende dos valores das frequências das classes anterior e posterior a classe modal. Se a frequência da classe anterior a modal for maior que a frequência posterior a modal, a moda estará mais próxima do limite inferior. O raciocínio inverso também será verdadeiro, isto é, se a frequência posterior a modal for maior que a anterior a modal, o valor da moda estará mais próximo do limite superior da classe modal.

No exemplo em questão, sabemos que moda está contida no intervalo entre 10 a 20. Ao observar que a frequência da classe posterior é 3 e a frequência da classe anterior é 2, sabemos que a moda será mais próxima do limite superior 20, portanto, a moda será maior que a média dessa classe, isto é, maior que 15. Nesse sentido, o resultado do cálculo da moda foi de 16,66. Quanto maior for a diferença entre as frequências anterior e posterior, mais na extremidade da classe o valor da moda estará.

2.3.3.4 Método da Moda de King

Estima a moda baseado nos valores de frequência das classes anterior à modal e posterior à modal. O cálculo é feito pela seguinte fórmula:

$$Mo = Li + h \frac{f_{Post.}}{(f_{Ant.} + f_{Post.})}$$

Assim, o cálculo é procedido da seguinte maneira:

$$Mo = 10 + 10 \times \frac{3}{(2+3)}$$

$$Mo = 10 + \frac{30}{5} = 16\ kg/semana$$

A metodologia de King, apesar de ser uma fórmula diferente, tem o mesmo raciocínio da metodologia de Czuber. Sendo assim, os valores das frequências anterior e posterior a modal são determinantes para quantificar o valor da moda e indicar se está mais próximo do limite inferior ou superior da classe modal.

2.3.4 Moda em Variáveis Qualitativas

A Moda é especialmente útil quando se trata de variáveis qualitativas, principalmente, **variáveis qualitativas nominais**, pois não é possível obter a média e a mediana nesses casos. Nessa situação, a moda é a única medida descritiva pode ser definida e torna-se uma informação muito relevante para descrever o conjunto de dados. Veja um exemplo:

Objeto de estudo

Um estudo analisa a nacionalidade de um grupo de 8 pessoas que estão envolvidas em uma investigação policial.

Indivíduo	Nacionalidade
1	Português
2	Francês
3	Mexicano
4	Equatoriano
5	Francês
6	Brasileiro
7	Francês
8	Equatoriano

No exemplo, a variável em estudo é a nacionalidade das pessoas investigadas. Essa variável assume valores como “brasileiro”, “mexicano” etc. Não são valores numéricos, logo, não é possível calcular um média. Como também são valores não ordenáveis, não possível definir a mediana. Contudo, é possível definir o valor que corresponde a moda. Ao observar a frequência dos valores, observa-se que a nacionalidade “francês” é observada três vezes, com maior frequência. Logo, a moda é o valor “francês”.

$$Mo = Francês$$

2.3.5 Características da Moda

A moda, em um conjunto de dados, pode ser um valor único, como também, pode haver duas ou mais modas. Ainda, pode ocorrer da moda não existir em um conjunto de dados.

Um conjunto de dados será **Unimodal**, quando somente um valor tem maior frequência que os demais, exemplo:

$$X = \{2, 3, 4, 4, 4, 5, 8\} \quad Mo = 4$$

Um conjunto de dados será **Bimodal** quando dois valores possuem maior frequência do que os demais, exemplo:

$$X = \{2, 3, 4, 4, 4, 6, 7, 7, 7\} \quad Mo = 4\,, 7$$

No mesmo raciocínio, um conjunto de dados será **Trimodal**, quando houver três modas. Normalmente, a partir de três modas ou mais, o conjunto de dados é classificado como **Multimodal**.

$$X = \{3, 4, 4, 4, 6, 7, 7, 7, 9, 9, 9\} \quad Mo = 4, 7\,, 9$$

Quando o conjunto de dados não tem um valor que se repete, não existe moda e classifica-se como **amodal**, exemplo:

$$X = \{2, 4, 7, 8, 9, 10, 15\} \qquad Mo = \varnothing$$

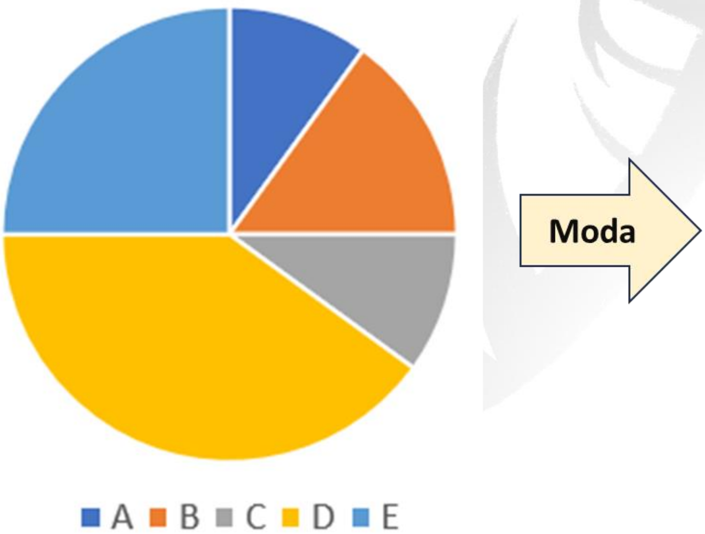
Ao contrário da Média e da Mediana, a Moda tem de ser **obrigatoriamente um valor existente no conjunto de dados**.

Por fim, a Moda **não é afetada (sensível) pelos valores extremos** da distribuição, desde que esses valores não constituam o valor modal (algo que pode ocorrer em pequenas amostras, mas não é comum).

2.3.6 Identificação da Moda em Gráficos

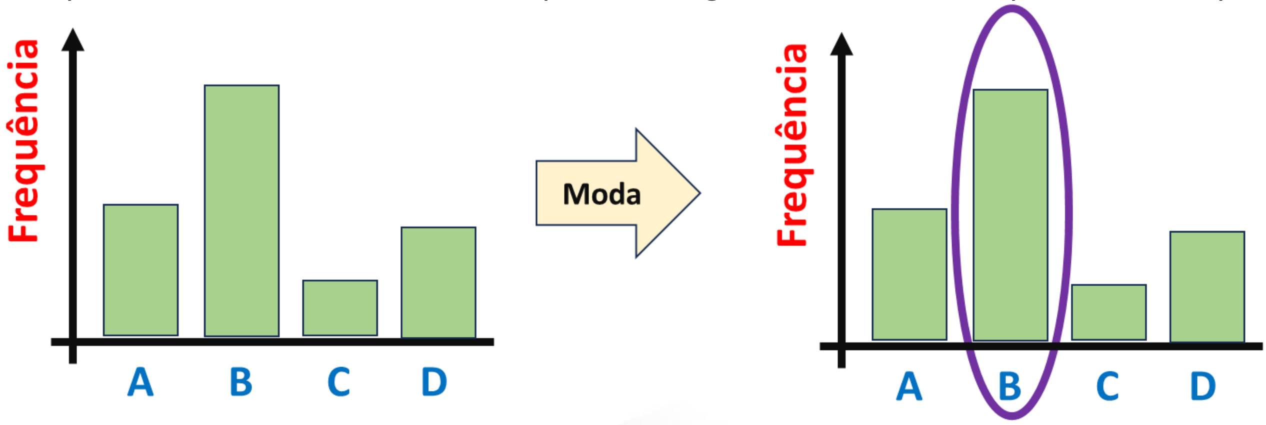
A moda é muitas vezes uma medida identificada visualmente. Dessa forma, é uma medida descritiva que pode ser facilmente identificada em um gráfico, principalmente, um gráfico que apresenta a distribuição de frequência. Para encontrar a moda, deve-se sempre procurar pelo valor com maior expressão de frequência no gráfico.

No gráfico de setores (ou de pizza), normalmente utilizado por variáveis qualitativas, deve-se encontrar pelo valor com maior área do setor (em termos simples, que possui a maior fatia da pizza).

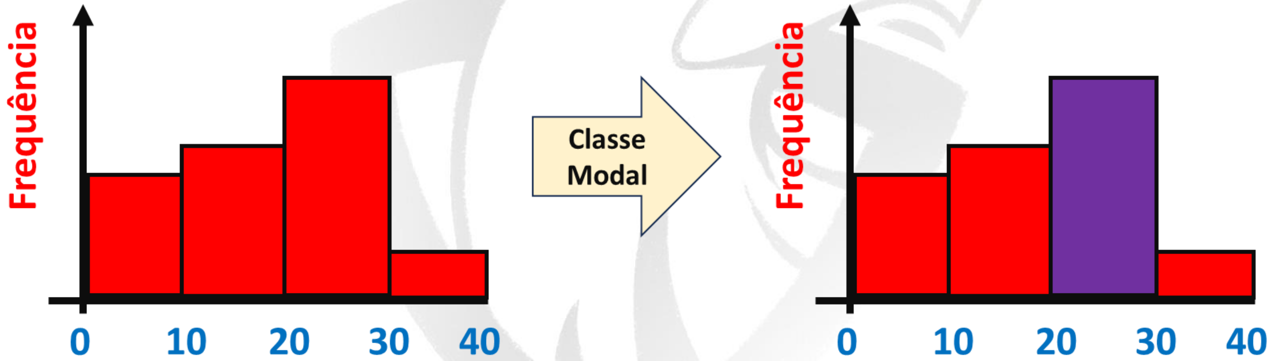


$$Mo = D$$

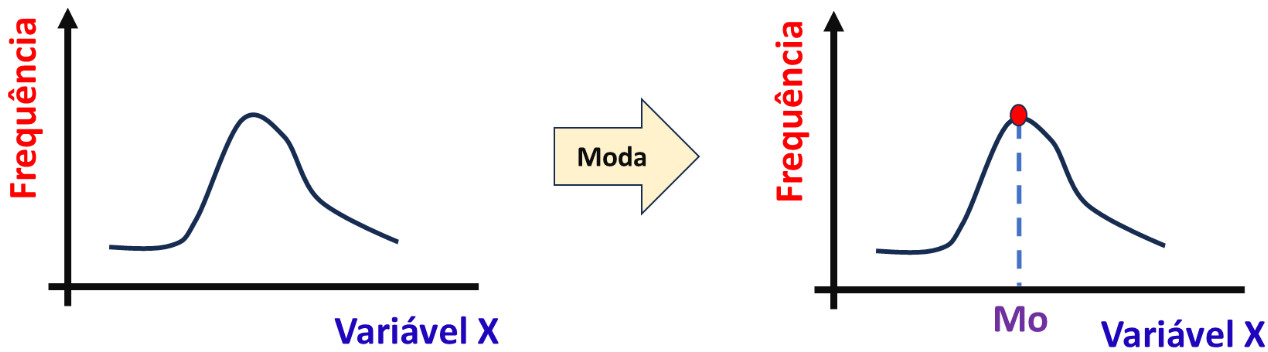
No gráfico de colunas, para identificar a moda, basta encontrar o valor da variável que apresenta a maior coluna de frequência. Logo, ele será o valor que mais se repete.



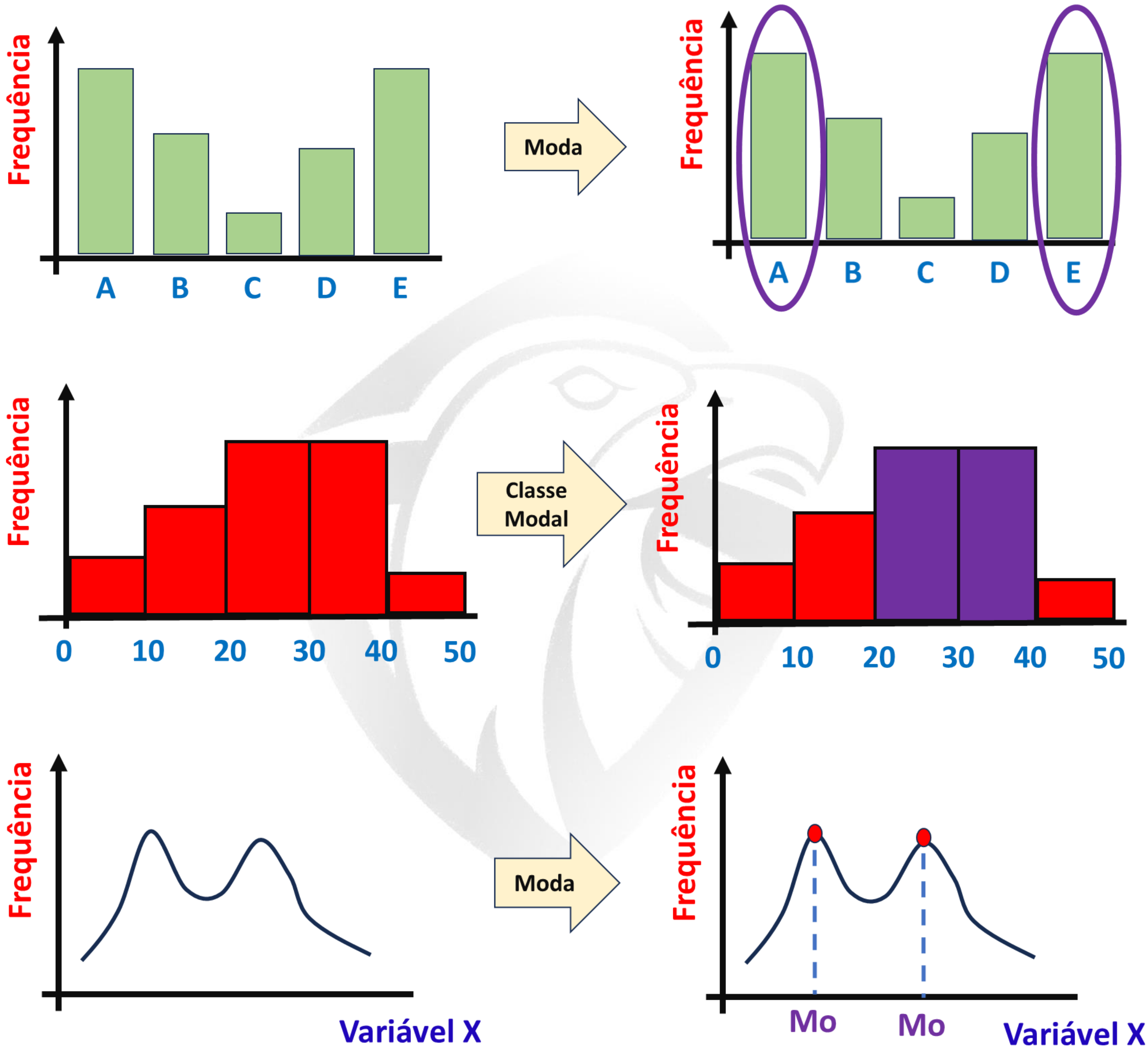
O mesmo raciocínio pode ser aplicado no histograma, porém, como é um gráfico que trabalha com dados agrupados, a maior coluna indicará a classe modal. Desse modo, para encontrar o valor pontual da moda será necessário aplicar alguma das metodologias estudas, como, Czuber, King etc.



Para um gráfico de polígonos, como a curva de frequência, para detectar a moda, é necessário encontrar o valor da variável que corresponde ao pico da curva (o ponto mais alto).



Por fim, é interessante que o aluno visualize algumas representações gráficas que evidenciam distribuições de frequências não unimodais. Nesses casos, serão visualizadas duas ou mais expressões gráficas com maior frequência que as demais.



QUESTÕES DE RENDIMENTO

01 (CESPE|2018|POLÍCIA FEDERAL|PERITO CRIMINAL)

Considerando que a análise de uma amostra de minério de chumbo tenha apresentado os seguintes resultados percentuais (%): 8,10; 8,32; 8,12; 8,22; 7,99; 8,31, julgue o item a seguir, relativo a esses dados.

O valor médio do teor de chumbo presente na amostra foi superior a 8%.

Certo ( ) Errado ( )

Resolução

Para resolver essa questão, nem é necessário efetuar o cálculo da média. **Basta observar o conjunto de dados analisado!** Veja que das 6 observações que compõem o conjunto, 5 delas são superiores a 8,00%, enquanto apenas uma observação (7,99%) é inferior a 8,00%.

$$\boldsymbol{Amostra\ de\ Minério\ de\ Chumbo = \{7,99;\ 8,10;\ 8,12;\ 8,22;\ 8,31;\ 8,32\}}$$

Logo, se a média é uma medida descritiva que resume um conjunto de dados, obtendo um único valor que capta a **tendência central,** obviamente, para esse conjunto de dados, a média será superior a 8,00%. Desse modo, o valor médio do teor de chumbo com certeza será superior a 8%.

Para conferir o procedimento matemático para obter essa informação, confira:

$$\bar{X} = \frac{7,99 + 8,10 + 8,12 + 8,22 + 8,31 + 8,32}{6}$$

$$\bar{X} = \frac{49,06}{6} = 8,17\%$$

Como já afirmamos, a média é um valor superior a 8%, alternativa correta.

É sempre interessante observar o conjunto de dados analisados, associando com a informação da medida descritiva. Nesse exemplo, só a essência do conhecimento estatístico é necessária para resolver a questão!

CERTA.

---

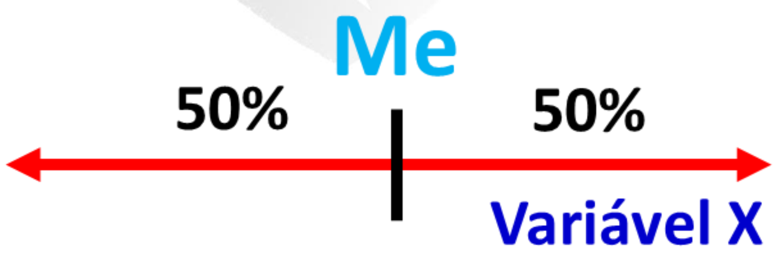
02 (INSTITUTO AOCP|2021|PC-PA|INVESTIGADOR)

Uma testemunha de um roubo afirma que o ladrão tem uma estatura mediana de 1,70 m. Então, pode-se esperar que, em termos de probabilidade, as alturas X de possíveis suspeitos se situem em:

a) $P(X \leq 1{,}70) = P(X \geq 1{,}70) = 50\%$.
b) $P(X = 1{,}70) = 50\%$.
c) $P(X > 1{,}70) = 75\%$.
d) $P(X < 1{,}70) = 25\%$.
e) $P(X < 1{,}70) = P(X > 1{,}70) = 25\%$.

Resolução

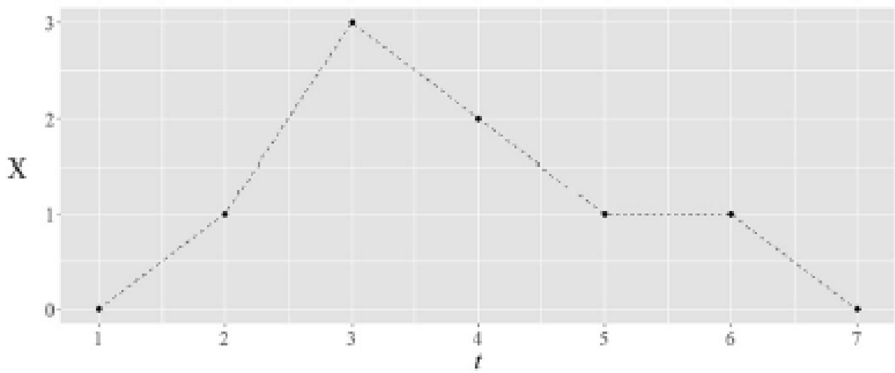
A mediana defini o valor de X que separa o conjunto de dados exatamente ao meio. Logo, ao saber que a mediana da altura X é igual a 1,70 m, em termos de probabilidade, podemos afirmar certamente que as chances de encontrar um valor abaixo da mediana é de 50% [P(X<1,70)=50%], da mesma forma que a chances de encontrar um valor acima da mediana também será 50% [P(X>1,70)=50%].



ALTERNATIVA CERTA: LETRA A.

---

03 (CESPE|2022|PC-RO|INVESTIGADOR)



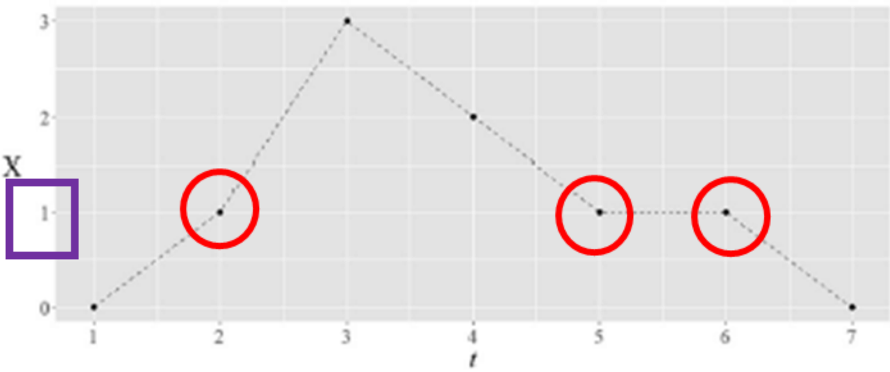
Conforme a figura precedente, que mostra os valores da variável *X* obtidos em uma amostra de tamanho n = 7, tem-se que a moda de *X* é igual a

a) 0.
b) 1.
c) 1,5.
d) 3.
e) 2.

Resolução

A moda consiste na medida descritiva que defini o valor da variável X que mais se repete no conjunto de dados. Para isso, basta identificar no gráfico qual valor de X é o mais incidente. **Cuidado! O gráfico apresentado não é um gráfico de frequência.** A questão tenta induzir ao erro. Veja que o eixo das ordenadas (vertical) representa os valores da variável X, enquanto o eixo das abscissas (horizontal) indica cada um dos valores obtidos na amostra, que no gráfico foi representado pela letra t. Por essa razão, o eixo horizontal do gráfico é numerado de 1 a 7, já que o tamanho da amostra é igual a 7 (n=7). Praticamente, o gráfico representa os dados na **forma bruta,** cada ponto indica uma observação de X.

Portanto, é necessário identificar que os valores de X estão no eixo vertical e variam entre 0 a 3. Com isso, perceba que o valor que mais frequente é a observação X=1 que ocorre três vezes no decorrer das 7 observações. Veja a indicação no gráfico:



$$Mo = 1$$

A questão tem objetivo de confundir a leitura do gráfico, e interpretar que o valor 3 que tem o maior pico é aquele com maior frequência. Por isso, a interpretação das informações no gráfico sempre é um procedimento inicial imprescindível. Ao transformar esse gráfico de dados brutos em tabela de frequência, o correspondente seira:

Variável X	Frequência
0	2
1	3
2	1
3	1
Total	**7**

Portanto, valor com maior frequência é X=1, a moda.

ALTERNATIVA CERTA: LETRA B.

04 (CESPE|2022|PETROBRAS|PROFISSIONAL SUPERIOR)

O item a seguir é apresentada uma situação hipotética seguida de uma assertiva a ser julgada a respeito de probabilidade e estatística.

Os preços de um determinado produto em 10 diferentes lojas são dados na tabela a seguir.

N.º de lojas	2	3	1	2	2
Preço (R$)	195	210	220	235	240

A média aritmética dos preços encontrados foi de R$ 219,00.

Certo ( ) Errado ( )

Resolução

A questão apresenta uma tabela de frequência para a variável preço de um determinado produto, analisado em um total de 10 diferentes lojas. Tem-se nessa questão, apresentação de **dados ponderados,** tabela de frequência para cada valor de preço. Assim, para obter a média dos preços, é necessário somar todos os valores de preços (ponderando com suas respectivas frequências) e dividir pelo total de observações (n=10). Logo:

$$\bar{X} = \frac{195 \times 2 + 210 \times 3 + 220 \times 1 + 235 \times 2 + 240 \times 2}{10}$$

$$\bar{X} = \frac{390 + 630 + 220 + 470 + 480}{10}$$

$$\bar{X} = \frac{2190}{10} = R\$\,219,00$$

CERTA.

05 (IBFC|2022|EBSERH|ADMINISTRATIVO)

A tabela apresenta as idades dos 80 funcionários de uma loja de departamentos.

Idade	18 ├20	20 ├22	22 ├24	24 ├26	26 ├28
Frequência absoluta	8	22	25	15	10

De acordo com a tabela, a idade mediana dos funcionários dessa loja é igual a:

a) 22,5
b) 23
c) 23,2
d) 22,8
e) 22,4

Resolução

Nessa questão, tem-se o estudo da variável idades dos funcionários, sendo analisado ao todo 80 funcionários (n=80). O conjunto de dados é apresentado na forma de uma tabela de frequência com **dados agrupados**. Assim, para obter a mediana de conjunto de 80 dados, é necessário incialmente identificar a classe mediana e, em seguida, realizar a técnica da **interpolação linear**.

Desse modo, a localização da mediana será o valor da idade que acumula 50% dados, ou então, que acumula 80/2 = 40 observações. Logo:

$$F(Me) = \frac{n}{2} = \frac{80}{2} = 40 \text{ ou } F(Me) = 50\%$$

Para isso, precisamos obter a frequência acumulada das idades, isto é, para cada classe, acumular os valores da frequência absoluta. Logo:

Idade	18 ├20	20 ├22	22 ├24	24 ├26	26 ├28
Frequência absoluta	8	22	25	15	10
Frequência Acumulada	**8**	**30**	**55**	**70**	**80**

Com isso, basta identificar a classe mediana que será aquela que acumula 40 observações. Veja:

			Classe Mediana		
Idade	18 ├20	20 ├22	22 ├24	24 ├26	26 ├28
Frequência absoluta	8	22	25	15	10
Frequência Acumulada	**8**	**30**	**55**	**70**	**80**

Dessa forma, sabemos que a Mediana está contida no intervalo de 22 a 24. Para definir o valor pontual, é preciso aplicar a interpolação linear. Nesse sentido, deve-se associar o valor de idade com sua respectiva frequência acumulada. Até a idade 22 tem-se acumulado 30 observações; até a idade 24, acumula-se 55 observações; e a mediana, certamente acumula 40 observações (50%). Assim:

Valor Observado (X_i)	**Frequência Acumulada (F_i)**
22	**30**
Me	**40**
24	**55**

$$\frac{24 - 22}{55 - 30} = \frac{Me - 22}{40 - 30}$$

$$\frac{2}{25} = \frac{Me - 22}{10}$$

$$\frac{20}{25} = Me - 22$$

$$Me = 22 + 0,8 = 22,8$$

Portanto, o valor da mediana será de 22,8 anos.

ALTERNATIVA CERTA: LETRA D.

06 (CESPE|2015|DEPEN|AGENTE PENITENCIÁRIO)

Considerando os dados da tabela abaixo, que apresenta a distribuição populacional da quantidade diária de incidentes (N) em determinada penitenciária, julgue os itens que se seguem.

quantidade diária de incidentes (N)	frequência relativa
0	0,1
1	0,2
2	0,5
3	0,0
4	0,2
total	1

A moda da distribuição de N é igual a 4, pois esse valor representa a maior quantidade diária de incidentes que pode ser registrada nessa penitenciária.

Certo ( ) Errado ( )

Resolução

A questão tenta confundir conceitos das medidas descritivas da Estatística. A moda é a observação que mais se repete, ou então aquela com maior frequência. Para esse conjunto de dados, a moda é igual a 2 incidentes/dia, isso porque existe uma frequência de 50% (ou 0,5) dos dados observados com esse valor.

	quantidade diária de incidentes (*N*)	frequência relativa
	0	0,1
	1	0,2
Moda	$\boxed{2}$	0,5
	3	0,0
	4	0,2
	total	1

A maior quantidade diária de incidentes $(X = 4)$ corresponde a observação máxima (**X**$_{\text{máx}}$) e não necessariamente coincidirá com a moda. Nesse caso, não corresponde a moda, pois a observação 4 possui uma frequência relativa de 20%, isto é, repete-se com menor frequência em comparação a observação $X = 2$.

ERRADA.

07 (CESPE|2004|POLÍCIA FEDERAL|ESTATÍSTICO)

Em determinada semana, certa região foi dividida em 200 setores disjuntos para o estudo da distribuição espacial da incidência de um certo tipo de crime. Cada setor possui a forma de um quadrado de 4 km^2 de área. Acredita-se que a ocorrência do crime seja aleatória. A tabela abaixo apresenta o percentual de setores em que foi registrada a incidência X (número de ocorrências observadas no setor) do crime investigado.

X	percentual de setores em que se registrou a incidência X (em %)
0	10
1	25
2	35
3	25
4	5
total	100

A mediana de X é menor que 1.

Certo ( ) Errado ( )

Resolução

A questão apresenta **dados ponderados,** isto é, uma tabela de frequência sem intervalo, em que cada observação possui sua respectiva frequência. A questão nos fornece a **frequência relativa** do número de ocorrências de um certo crime. Contudo, para encontrar a mediana, a forma mais rápida é obtendo a **frequência acumulada relativa** desse conjunto de dados. Com isso, sabe-se que a observação que acumular **50%** dos dados corresponde ao valor da mediana.

Vamos construir os valores da frequência acumulada relativa e localizar a mediana:

Mediana

X	percentual de setores em que se registrou a incidência X (em %)	*Fac.* (%)
0	10	10
1	25	35
2	35	70
3	25	95
4	5	100
total	100	

O registro de 2 ocorrências de um certo crime acumula 70% das observações, logo, o valor 2, em suas repetições observadas, contempla 50% dos dados. Portanto, a mediana desse conjunto de dados é igual a 2. Por fim, a questão está errada uma vez que afirma que o valor da mediana é menor do que 1.

ERRADA.

---

09 (CESPE|2010|ABIN|TÉCNICO DE INTELIGÊNCIA)

Considerando que o diagrama de ramos-e-folhas abaixo mostra a distribuição das idades (em anos) dos servidores de determinada repartição pública, julgue os próximos itens.

2	1 3 2 6
3	4 3 5 8 7
4	6 2 1 9 6
5	4 2 0 5

A mediana das idades dos servidores é igual a 39,5 anos.

Certo ( ) Errado ( )

Resolução

A distribuição da idade dos servidores está representada em um diagrama de ramos e folhas. Esse diagrama organiza e ilustra o conjunto de dados de forma mais sintética. Os dados observados correspondem a concatenação dos ramos com as respectivas folhas. Porém, tenha cuidado! **Não necessariamente as folhas precisam**

estar ordenadas em rol crescente. Essa é a grande pegadinha da questão, pois primeiramente precisamos ordenar todo o conjunto de dados (ordenar as folhas). Ao se falar de mediana, os dados **precisam sempre estar em rol crescente.**

Vamos ordenar essa representação gráfica em rol crescente:

Ramos	Folhas
2	1 2 3 6
3	3 4 5 7 8
4	1 2 6 6 9
5	0 2 4 5

Isso corresponde a seguinte forma de dados brutos:

$$\{21, 22, 23, 26, 33, 34, 35, 37, 38, 41, 42, 46, 46, 49, 50, 52, 54, 55\}$$

Portanto, temos um total de 18 observações ($\boldsymbol{n = 18}$). A posição central desse conjunto de dados é igual a:

$$P_{Me} = \frac{n+1}{2}$$

$$P_{Me} = \frac{18+1}{2} = 9,5$$

Com isso, o valor da mediana é igual a **média** da observação na 9ª posição com a observação localizada na 10ª posição. Entenda:

Ramos	Folhas
2	1 2 3 6
3	3 4 5 7 8
4	1 2 6 6 9
5	0 2 4 5

{21, 22, 23, 26, 33, 34, 35, 37, 38, 41, 42, 46, 46, 49, 50, 52, 54, 55}

$$Me = \frac{38 + 41}{2} = \frac{79}{2} = 39,5$$

Por fim, a questão está correta já que afirma que a mediana é exatamente o valor de 39,5 anos.

CERTA.

---

10 (CESPE|2022|PM-MG|SOLDADO)

A tabela abaixo registra, em horas, o tempo de permanência, em determinado dia, de um grupo de pessoas no aplicativo *WhatsApp*.

Nº DE HORAS (x_i)	Nº DE USUÁRIOS (f_i)
8	10
9	7
11	12
14	14
15	20
20	12
SOMA	75

Considerando as informações da tabela, é **CORRETO** afirmar que a mediana, a média e a moda, em horas, são respectivamente:

a) 12,5; 12,5; 20.
b) 15; 12,48; 13.
c) 12,5; 13,48; 12.
d) 14; 13,48; 15.

Resolução

A mediana é o valor da variável X que acumula n/2 (50%) observações, que para um conjunto de 75 observações corresponde a uma frequência acumulada de 37,5. Para isso é interessante identificar a frequência acumulada dessa tabela de frequência:

Nº DE HORAS (x_i)	Nº DE USUÁRIOS (f_i)	$Fac.(\%)$
8	10	10
9	7	17
11	12	29
Mediana $\boxed{14}$	14	43
15	20	63
20	12	75
SOMA	75	

A mediana corresponde a observação 14 horas, pois acumula 43 observações que inclui a observação de 37,5 (50% dos dados).

A média para dados ponderados é obtida pelo cálculo de uma média ponderada, isto é, considera cada observação com sua respectiva frequência. Veja:

$$\bar{X} = \frac{\sum X_i \times f_i}{n}$$

$$\bar{X} = \frac{8 \times 10 + 9 \times 7 + 11 \times 12 + 14 \times 14 + 15 \times 20 + 20 \times 12}{75}$$

$$\bar{X} = \frac{1011}{75} = 13,48\ horas$$

Por fim, a moda consiste na observação que apresenta a maior frequência absoluta, logo, corresponde ao valor $Mo = 15$. Entenda:

Moda

Nº DE HORAS (x_i)	Nº DE USUÁRIOS (f_i)
8	10
9	7
11	12
14	14
15	20
20	12
SOMA	75

Assim, a questão correta é alternativa D. Veja que para essa questão o cálculo da média seria muito trabalhoso e longo de se fazer. Contudo, bastaria identificar a moda e a mediana para encontrar a alternativa correta.

ALTERNATIVA CERTA: LETRA D.

---

11 (CESPE|2022|PC-RO|DATILOSCOPISTA)

x	frequência acumulada relativa
0	0,0
1	0,1
2	0,3
3	0,6
4	1,0

Sabendo-se que, nessa tabela, é apresentada a distribuição de frequência acumulada relativa de uma variável quantitativa discreta X, é correto afirmar que, se A, B e C representam, respectivamente, a média, a mediana e a moda da distribuição de X, então a soma A + B + C é igual a:

a) 6.
b) 7.
c) 8.
d) 9.
e) 10.

Resolução

A questão apresenta a distribuição da frequência acumulada de uma variável discreta X. Para responder essa questão, precisa-se obter o valor das três medidas descritivas de tendência central, isto é, a média, a mediana, e a moda. Inicialmente vamos obter a mediana, pois é preciso apenas identificar o valor X que acumula 50% das observações, assim:

	x	frequência acumulada relativa
	0	0,0
	1	0,1
	2	0,3
Me	$\boxed{3}$	0,6
	4	1,0

Em seguida, para obter a média e a moda, é necessário obter a frequência relativa não acumulada de cada valor de X. Dessa forma, é preciso subtrair cada valor de frequência acumulada em relação a frequência acumula do valor de X anterior. Com isso, já conseguimos detectar a moda, pois será o valor de X com maior frequência relativa. Veja:

	x	frequência acumulada relativa	Frequência Relativa
	0	0,0	$0,0$
	1	0,1	$0,1 - 0,0 = 0,1$
	2	0,3	$0,3 - 0,1 = 0,2$
	3	0,6	$0,6 - 0,3 = 0,3$
Mo	4	1,0	$1,0 - 0,6 = 0,4$

Agora, é possível obter a média somando cada valor x multiplicado por sua respectiva frequência relativa. Assim:

$$\bar{X} = 1 \times 0,1 + 2 \times 0,2 + 3 \times 0,3 + 4 \times 0,4$$

$$\bar{X} = 0,1 + 0,4 + 0,9 + 1,6 = 3$$

Por fim, a questão pede o somatório das medidas de tendência central, logo, a alternativa correta é a letra E.

$$3 + 4 + 3 = 10$$

ALTERNATIVA CERTA: LETRA E.

---

12 (FUNDATEC|2022|SEPOG-RS|ANALISTA)

Considerando a tabela de frequência apresentada abaixo, referente à distribuição de uma determinada variável X, pode-se dizer que o valor aproximado da média de X é de:

X	Frequência
0 - 10	25
10 - 20	10
20 - 30	49
30 - 40	88
40 - 50	28
TOTAL	200

a) 39,2.
b) 34,2.
c) 29,2.
d) 24,2.
e) 22,2.

Resolução

Nessa questão, temos a distribuição de frequência de uma variável X com **dados agrupados**. A questão solicita o valor aproximado da média, para isso, devemos trabalhar com a ideia que os valores observados de X vão coincidir com o ponto médio de cada classe. Portanto, devemos iniciar calculando o ponto médio das classes:

$$Pm_{1^{\underline{a}}} = \frac{10 + 0}{2} = 5$$

$$Pm_{2^{\underline{a}}} = \frac{20 + 10}{2} = 15$$

$$Pm_{3^{\underline{a}}} = \frac{30 + 20}{2} = 25$$

$$Pm_{4^{\underline{a}}} = \frac{40 + 30}{2} = 35$$

$$Pm_{5^{\underline{a}}} = \frac{50 + 40}{2} = 45$$

Desse modo, vamos associar cada ponto médio das classes com suas respectivas frequências. Assim, realizaremos o cálculo da média somando todos os valores de X (ponderado a sua respectiva frequência) e depois dividindo pelo total de observações n=200.

Valor Observado (X_i)	Ponto Médio (Pm_i)	Frequência Absoluta (f_i)	$Pm_i \times f_i$
0⊢10	5	25	125
10⊢20	15	10	150
20⊢30	25	49	1225
30⊢40	35	88	3080
40⊢⊣50	45	28	1260
Soma (Σ_i)	-	200	5840

$$\bar{X} = \frac{5840}{200} = 29,2$$

ALTERNATIVA CERTA: LETRA C.

13 (IBFC|2022|EBSERH|ADMINISTRATIVO)

Numa pesquisa sobre o "peso", em quilogramas, de crianças atendidas num hospital, se chegou ao seguinte resultado indicado na tabela a seguir:

Peso (Kg)	Frequência
20 ├ 22,5	2
22,5 ├ 25	4
25 ├ 27,5	6
27,5 ├ 30	10
30 ├ 32,5	9

De acordo com a tabela, a moda da distribuição, pelo método de Czuber, é igual a:

a) 28,5
b) 29,5
c) 28
d) 27,75
e) 28,75

Resolução

A variável Peso é apresentada em uma tabela de frequência com dados agrupados. Para obter a moda nessa situação, é necessário identificar a classe modal e, em seguida, aplicar a metodologia de Czuber. Dessa forma, a classe modal corresponde a classe de peso que possui o maior valor de frequência. Isto é:

Peso (Kg)	Frequência
20 ├ 22,5	2
22,5 ├ 25	4
25 ├ 27,5	6
$\boxed{27,5 ├ 30}$	10
30 ├ 32,5	9

Classe Modal

Nesse raciocínio, sabemos que a moda está compreendida no intervalo de 27,5 a 30 kg. Agora, para saber o valor pontual da moda devemos aplicar a fórmula de Czuber:

$$Mo = Li + h\,\frac{d_A}{d_A + d_P}$$

$$d_A = f_{Modal} - f_{Ant.}$$

$$d_P = f_{Modal} - f_{Post.}$$

Assim, é preciso identificar o limite inferior da classe modal (Li), a amplitude da classe modal (h), e as frequências da classe modal e das classes anterior e posterior a modal. Ao analisar a tabela de frequência, tem-se:

h=2,5

	Peso (Kg)	Frequência	
	20 ├ 22,5	2	
	22,5 ├ 25	4	
	25 ├ 27,5	6	$f_{Ant.}$
Li	27,5 ├ 30	$\boxed{10}$	f_{Modal}
	30 ├ 32,5	9	$f_{Post.}$

Com isso, ao aplicar a metodologia, o valor da moda será:

$$d_A = 10 - 6 = 4$$

$$d_P = 10 - 9 = 1$$

$$Mo = 27,5 + 2,5\,\frac{4}{4+1}$$

$$Mo = 27,5 + \frac{10}{5}$$

$$Mo = 27,5 + 2 = 29,5$$

ALTERNATIVA CERTA: LETRA B.

---

14 (CESPE|2022|DPE RO|ANALISTA)

x	frequência relativa
0	0,23
1	0,22
2	0,50
3	0,05

Considerando que a tabela acima mostra a distribuição de frequências de uma variável X obtida com base em uma amostra aleatória simples de tamanho igual a n, julgue o item que se segue.

A média amostral da variável X é inferior a 1,5.

Certo ( ) Errado ( )

Resolução

A questão apresenta a distribuição de frequência da variável X em **dados ponderados.** Com isso, devemos entender que cada valor de X é observado sobre uma respectiva frequência relativa conforme indicado na tabela. Para obter a média, é necessário realizar o somatório de todas as observações de X multiplica por sua respectiva frequência relativa. Logo:

$$\bar{X} = 0 \times 0,23 + 1 \times 0,22 + 2 \times 0,5 + 3 \times 0,05$$

$$\bar{X} = 0,22 + 1,0 + 0,15 = 1,37$$

Portanto, a questão está correta, pois a média é inferior a 1,5.

CERTA.

PROFISSÃO POLICIAL

CONCURSEIRO QUE PRETENDE SER POLICIAL NÃO FAZ RATEIO