MEDIDAS SEPARATRIZES

As separatrizes são medidas que dividem (ou separam) uma série ordenada em duas ou mais partes, cada uma contendo a mesma quantidade de elementos. Nesse caso, o nome da medida separatriz é definido de acordo com a quantidade de partes em que a série é dividida:

- **mediana**: divide uma série ordenada em **duas partes** iguais, cada uma contendo **50%** dos valores da sequência;
- **quartis**: dividem uma série ordenada em **quatro partes** iguais, cada uma contendo **25%** dos valores da sequência;
- **quintis**: dividem uma série ordenada em **cinco partes** iguais, cada uma contendo **20%** dos valores da sequência;
- **decis**: dividem uma série ordenada em **dez partes** iguais, cada uma contendo **10%** dos valores da sequência; e
- **percentis**: dividem uma série ordenada em **cem partes** iguais, cada uma contendo **1%** dos valores da sequência.

Ao longo dessa aula, vamos estudar a mediana, os quartis, os decis e os percentis. Os quintis, por não serem tão explorados em provas de concurso, não serão abordados.

MEDIANA

A mediana é, simultaneamente, uma MEDIDA DE POSIÇÃO, de TENDÊNCIA CENTRAL e SEPARATRIZ. Ela caracteriza a **posição central** de uma série de valores. Além disso, também **separa uma série de valores em duas partes de tamanhos iguais**, **cada uma contendo o mesmo número de elementos**. Muitas vezes, a mediana é designada como **valor mediano**, sendo representada pelos símbolos $\boldsymbol{M_d}$ ou, em menor frequência, $\tilde{\boldsymbol{x}}$.

Mediana para dados não-agrupados.

O método para determinação da mediana envolve a realização de uma etapa anterior, que consiste na ordenação do conjunto de dados. Feito isso, **a mediana é o elemento que ocupa a POSIÇÃO CENTRAL de uma série de observações ordenada segundo suas grandezas (isto é, dados brutos organizados em rol crescente ou decrescente)**.

Por exemplo, vamos determinar a mediana da seguinte série de valores:

$$\{3, 17, 13, 19, 2, 5, 7, 1, 8, 21, 9\}.$$

Em conformidade com a definição da mediana, a primeira etapa consiste na ordenação (crescente ou decrescente) da série de valores. Desse modo, obtemos:

$$\{\mathbf{1, 2, 3, 5, 7, 8, 9, 13, 17, 19, 21}\}.$$

Agora, determinaremos o elemento que ocupa a posição central desse conjunto de dados. Para isso, devemos encontrar o termo que possui o mesmo número de elementos tanto à sua esquerda quanto à sua direita. Em nosso exemplo, esse valor é o 8, pois existem cinco elementos antes dele e cinco após ele.

$$\underbrace{\mathbf{1, 2, 3, 5, 7,}}_{\textit{5 elementos antes}} \quad \underbrace{\mathbf{8,}}_{\textit{elemento central}} \quad \underbrace{\mathbf{9, 13, 17, 19, 21}}_{\textit{5 elementos depois}}.$$

É importante notarmos que **essa série possui um número ímpar de elementos**. Quando isso acontece, isto é, **quando uma série possui um número ímpar de elementos, a mediana SEMPRE coincide com o elemento central do conjunto de dados**. Portanto, temos:

$$M_d = 8$$

Contudo, se porventura **a série tivesse um número par de elementos, POR CONVENÇÃO, a mediana seria a média aritmética dos dois termos centrais**. Assim, caso adicionássemos o número 23 ao conjunto de dados apresentado anteriormente, teríamos a seguinte situação:

$$\underbrace{\mathbf{1, 2, 3, 5, 7,}}_{\textit{5 elementos antes}} \quad \underbrace{\mathbf{8, 9,}}_{\textit{elementos centrais}} \quad \underbrace{\mathbf{13, 17, 19, 21, 23}}_{\textit{5 elementos depois}}.$$

Nesse caso, em que temos um número par de elementos, a mediana é definida como a média aritmética dos termos centrais, que são os números 8 e 9. Assim, temos:

$$M_d = \frac{8+9}{2} = \frac{17}{2} = 8{,}5$$

Note que, quando o número é ímpar, o termo central sempre ocupa a posição $\boldsymbol{\frac{n+1}{2}}$. Por outro lado, quando o número de termos é par, existem dois termos centrais, sendo que o primeiro ocupa a posição $\boldsymbol{\frac{n}{2}}$; e o segundo ocupa a posição imediatamente seguinte, ou seja, $\boldsymbol{\frac{n}{2} + 1}$.

Essas relações são importantes porque nem sempre conseguiremos identificar a posição central tão facilmente. Por exemplo, se tivermos uma série composta por 501 elementos, podemos afirmar que o termo central será o elemento ocupando a posição $\frac{n+1}{2} = \frac{501+1}{2} = 251$, sem precisar recorrer a qualquer outro método. Logo, a mediana terá o mesmo valor do termo central:

$$M_d = x_{251}.$$

Vejamos a disposição do termo central em relação aos demais elementos da série:

$$\underbrace{\boldsymbol{x_1, x_2, \cdots, x_{250},}}_{\textit{250 elementos antes}} \quad \underbrace{\boldsymbol{x_{251},}}_{\textit{termo central}} \quad \underbrace{\boldsymbol{x_{252}, x_{253}, \cdots, x_{501}}}_{\textit{250 elementos depois}}.$$

Agora, se tivermos uma série composta por 500 elementos, os termos centrais serão os elementos ocupando as posições:

$$\frac{n}{2} = \frac{500}{2} = 250 \quad \text{e} \quad \frac{n}{2} + 1 = \frac{500}{2} + 1 = 251.$$

Vejamos a disposição dos termos centrais em relação aos demais elementos da série:

$$\underbrace{x_1, x_2, \cdots, x_{249},}_{250\ elementos\ antes} \quad \underbrace{x_{250}, x_{251},}_{termos\ centrais} \quad \underbrace{x_{252}, x_{253}, \cdots, x_{500}}_{250\ elementos\ depois}.$$

Nessa situação, por convenção, a mediana será a média aritmética entre os termos centrais,

$$M_d = \frac{x_{250} + x_{251}}{2}.$$

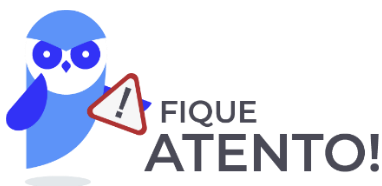
Portanto, podemos estabelecer que a mediana de um conjunto composto por n elementos ordenados de forma crescente ou decrescente será:

a) se n for ímpar, o termo de ordem $\frac{n+1}{2}$, isto é,

$$\boxed{\boldsymbol{M_d = x_{\frac{n+1}{2}}}}$$

b) se n for par, a média aritmética dos termos de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$, isto é,

$$\boxed{\boldsymbol{M_d = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}}}$$



A mediana nem sempre coincidirá com um elemento da série de dados. Isso somente acontecerá quando o número de elementos da série de dados for ímpar, pois haverá coincidência entre os valores da mediana e do termo que ocupa a posição central. Contudo, quando o número de elementos for par, não existirá essa coincidência.

PRESTE MAIS ATENÇÃO!

Quando o número de elementos do conjunto é ÍMPAR, o valor da mediana é único e igual ao termo central. Porém, quando o número de elementos é **PAR**, **a mediana pode ser QUALQUER VALOR ENTRE OS TERMOS CENTRAIS, havendo infinitos valores possíveis para a mediana.** Para exemplificar, imaginemos os números 1 e 2 como termos centrais. Entre esses dois números temos infinitas possibilidades de escolha, a exemplo de 1,01; 1,2; 1,673; etc**. A mediana poderia ser qualquer desses valores, contudo, POR CONVENÇÃO, adotamos a média aritmética dos valores centrais**.

INDO MAIS FUNDO!

Seja $\{x_n\}$ uma série de dados estatísticos composta por n elementos ordenados de forma crescente ou decrescente, isto é, $\{x_n\} = \{x_1, x_2, x_3, \cdots, x_n\}$. A mediana desse conjunto de dados será:

a) se n for ímpar, o termo de ordem $\frac{n+1}{2}$, isto é, $M_d = x_{\frac{n+1}{2}}$

b) se n for par, a média aritmética dos termos de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$, isto é, $M_d = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

EXEMPLIFICANDO

Calcular a mediana dos seguintes conjuntos:

a) seja $\{x_n\}$ uma série de dados composta pelos seguintes valores:

$$\{x_n\} = \{1, 3, 5, 6, 7, 9, 11, 13, 17, 18, 20\}$$

Como o número de elementos é ímpar, $n = 11$, temos:

$$M_d = x_{\frac{11+1}{2}} = x_{\frac{12}{2}} = x_6$$

Logo, a mediana é o sexto elemento da série, isto é:

$$M_d = x_6 \Rightarrow \ M_d = 9$$

b) seja $\{y_n\}$ uma série de dados composta pelos seguintes elementos:

$$\{y_n\} = \{11, 14, 15, 16, 18, 19, 21, 23\}$$

Como o número de elementos é par, $n = 8$, temos:

$$M_d = \frac{y_{\frac{n}{2}} \ + \ y_{\frac{n}{2}+1}}{2} = \frac{y_{\frac{8}{2}} \ + \ y_{\frac{8}{2}+1}}{2} = \frac{y_4 \ + \ y_{4+1}}{2} = \frac{y_4 \ + \ y_5}{2}$$

Logo, a mediana será a média aritmética entre o quarto e o quinto elementos da série, isto é:

$$M_d = \frac{y_4 + y_5}{2} = \frac{16 + 18}{2} = \frac{34}{2} = 17 \Rightarrow M_d = 17$$

Como vimos, a mediana depende apenas do termo que ocupa a posição central em um conjunto de dados, e não dos valores de todos os elementos que compõem a série. Por isso, dizemos que **a mediana não sofre tanta influência pela presença de valores extremos/discrepantes quanto à média**. Essa é, inclusive, uma das principais diferenças entre essas duas medidas.

Podemos constatar essa propriedade da mediana por meio de um exemplo. Considere que tenhamos inicialmente a seguinte série:

$$\{1, 2, 4, 6, 7, 9, 10, 11, 13\}$$

A média aritmética desses valores é:

$$\bar{x} = \frac{1 + 2 + 4 + 6 + 7 + 9 + 10 + 11 + 13}{9} = \frac{63}{9} = 7$$

Como o número de elementos é ímpar, $n = 9$, a mediana será o elemento que ocupa a posição:

$$\frac{n + 1}{2} = \frac{9 + 1}{2} = \frac{10}{2} = 5.$$

O **quinto termo** é 7, portanto:

$$M_d = x_5 \Rightarrow \ M_d = 7.$$

Agora, considere que o elemento de valor 13 tenha sido alterado para 130.000. Veja o que acontece com a média aritmética desse conjunto:

$$\overline{x} = \frac{1 + 2 + 4 + 6 + 7 + 9 + 10 + 11 + 130.000}{9} = 14.450 \Rightarrow \ \overline{x} = 14.450$$

Como o número de elementos permanece inalterado, $n = 9$, a mediana continua sendo o elemento que ocupa a posição:

$$\frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5.$$

Logo, a mediana ainda é representada pelo **quinto** termo da série:

$$M_d = x_5 \Rightarrow \ M_d = 7.$$

Portanto, a alteração no valor de um único elemento do conjunto de dados causou um impacto significativo na média, ao passo que a mediana permaneceu inalterada. Por isso, **dizemos que a média é mais influenciada pela presença de valores extremos que a mediana**.

PRESTE MAIS
ATENÇÃO!

A mediana depende da apenas posição e não dos valores dos elementos de uma série ordenada. Essa é uma das principais diferenças entre a média e a mediana, pois a primeira é muito impactada pela presença de valores extremos enquanto a última não.

FIQUE ATENTO!

Em geral, os valores da média aritmética e da mediana são diferentes. Por exemplo, a média da série $\{x_n\} = \{1, 2, 4, 6, 7, 9, 10, 11, 13\}$ é:

$$\overline{x} = \frac{1 + 2 + 4 + 6 + 7 + 9 + 10 + 11 + 130.000}{9} = 14.450,$$

enquanto sua mediana é:

$$M_d = 7.$$

HORA DE PRATICAR!

(CESPE/IPHAN/2018) Define-se estatística descritiva como a etapa inicial da análise utilizada para descrever e resumir dados. Em relação às medidas descritivas, julgue o item a seguir.

A mediana é o valor que ocupa a posição central da série de observações de uma variável, dividindo-se o conjunto de valores ordenados em partes assimétricas desiguais.

Comentários:

A mediana é o valor que ocupa a posição central da série de observações de uma variável, dividindo o conjunto em duas partes com a mesma quantidade de valores. As partes não serão necessariamente assimétricas desiguais. Se tivéssemos um conjunto formado apenas por elementos repetidos, por exemplo, a mediana dividiria os valores em partes simétricas.

Gabarito: Errado.

(CESPE/IPHAN/2018) Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}. Com referência a esses dados, julgue o item seguinte.

A mediana do conjunto é igual a 3.

Comentários:

A mediana divide um conjunto ao meio e ocupa a posição central. Temos 11 elementos na amostra. Então, a mediana ocupará a posição:

$$\frac{n+1}{2} = \frac{11+1}{2} = \frac{12}{2} = 6$$

Vejamos:

$$\{1, 2, 2, 3, 3, \underbrace{3}_{6^{\underline{o}}\ termo\ =\ termo\ central}, 4, 4, 4, 4, 4\}$$

Portanto, a mediana corresponde ao sexto termo da série de observações. Logo:

$$M_d = 3.$$

Gabarito: Certo.

(CESPE/FUB/2018) A tabela seguinte mostra as quantidades de livros de uma biblioteca que foram emprestados em cada um dos seis primeiros meses de 2017.

	Mês					
	1	2	3	4	5	6
Quantidade	50	150	250	250	300	200

A partir dessa tabela, julgue o próximo item.

A mediana dos números correspondentes às quantidades de livros emprestados no primeiro semestre de 2017 é igual a 200.

Comentários:

A mediana é o termo central de uma amostra ou população. Se temos 6 meses, então a mediana poderá ser encontrada pela média dos termos que ocupam as posições 3 e 4, pois, nesse caso, não há apenas um termo central.

Assim, o primeiro passo a se dar é colocar os dados da tabela em ordem crescente (isto é, em rol crescente):

$$50 \quad 150 \quad \underbrace{200 \quad 250}_{termos\ centrais} \quad 250 \quad 300$$

Agora, basta encontrarmos a média aritmética dos termos nas posições 3 e 4:

$$M_d = \frac{200+250}{2} = 225$$

Gabarito: Errado.

Mediana para dados agrupados sem intervalos de classe

O raciocínio adotado no cálculo da mediana para **dados agrupados por valor (sem intervalos de classe)** é similar ao empregado no caso dos dados não-agrupados. Basicamente, teremos que encontrar um valor que dividirá a distribuição de frequências em duas partes contendo o mesmo número de elementos.

Considere a seguinte situação hipotética: uma empresa realizou uma pesquisa para medir o nível de satisfação dos clientes com relação ao seu atendimento. Os clientes puderam atribuir notas de 0 a 5 no que diz respeito ao nível de satisfação, resultando na seguinte distribuição de frequências:

Nível de Satisfação (X_i)	Frequência (f_i)
0	3
1	5
2	8
3	10
4	13
5	10

O total de clientes entrevistados foi de:

$$3 + 5 + 8 + 10 + 13 + 10 = 49.$$

Como o número de entrevistados é ímpar, $n = 49$, a mediana será o termo que ocupa a posição de ordem:

$$\frac{n+1}{2} = \frac{49+1}{2} = \frac{50}{2} = 25$$

Em outros termos, a mediana será o elemento que ocupa a **vigésima quinta** posição. Para chegarmos a esse elemento, precisamos percorrer cada um dos níveis de satisfação. Reparem que três clientes atribuíram a nota 0 (zero); cinco atribuíram a nota 1 (um); e oito atribuíram a nota 2 (dois). Portanto, até esse ponto, temos um total de 16 avaliações:

$$3 + 5 + 8 = 16$$

Vejam que ainda não chegamos na posição desejada, isto é, na **vigésima quinta**. Contudo, sabemos que o próximo nível de satisfação, referente à nota 3 (três), teve frequência absoluta igual a 10. Se somarmos essas dez novas avaliações com o total obtido anteriormente, chegaremos a um valor que ultrapassa a posição procurada ($16 + 10 = 26$). Assim, descobrimos que a mediana está localizada nessa faixa de avaliação. Portanto,

$$M_d = x_{25} = 3$$

Esse procedimento pode ser simplificado com a introdução de uma coluna adicional para armazenar as frequências acumuladas. Já vimos que, para calcularmos a frequência acumulada, devemos repetir a primeira frequência e somar as frequências subsequentes, exibindo os resultados a cada linha. Observem:

Nível de Satisfação (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})	Memória de cálculo
0	3	3	= 3
1	5	8	= 3+5 = 8
2	8	16	= 8 + 8 = 16
3	10	26	= 16 + 10 = 26
4	13	39	= 26 + 13 = 39
5	10	49	= 39 + 10 = 49

Vamos remover a memória de cálculo para simplificar a tabela:

Nível de Satisfação (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	3	3
1	5	8
2	8	16
3	10	26
4	13	39
5	10	49

Reparem que o número 16, na terceira linha da coluna de frequências acumuladas, representa a soma das frequências absolutas simples das três primeiras linhas, isto é, $3 + 5 + 8$. Assim, concluímos que 16 clientes avaliaram o atendimento da empresa com nota igual ou inferior a 2. De forma análoga, como 49 clientes participaram da pesquisa, podemos afirmar que 33 avaliaram o atendimento com nota igual ou superior a 3.

Observem que a introdução da coluna de frequências acumuladas torna possível calcularmos a mediana de forma praticamente imediata. Nesse sentido, **se n for ímpar, basta identificarmos o valor da variável correspondente à primeira frequência acumulada imediatamente igual ou superior à posição de ordem**

$\frac{n+1}{2}$; e, se n for par, basta identificarmos os dois valores correspondentes às frequências acumuladas imediatamente iguais ou superiores às posições de ordens $\frac{n}{2}$ e $\frac{n}{2}+1$, respectivamente.

Em nosso exemplo, como a frequência total é ímpar, teremos que buscar pela posição $\frac{n+1}{2} = \frac{49+1}{2} = 25$. A mediana será o valor da variável correspondente à primeira frequência acumulada maior ou igual a essa posição, portanto, $M_d = 3$. Vejamos:

Nível de Satisfação (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	3	3
1	5	8
2	8	16
3	**10**	**26 (> 25)**
4	13	39
5	10	49

TOME NOTA!

Assim, podemos estabelecer que a mediana de uma tabela de frequências composta por um total de n elementos será:

a) se n for ímpar, o valor identificado na tabela correspondente à frequência acumulada imediatamente igual ou superior à posição de ordem $\frac{n+1}{2}$, isto é,

$$\boxed{M_d = X_{\frac{n+1}{2}}}$$

b) se n for par, a média aritmética dos valores identificados na tabela correspondentes às frequências acumuladas imediatamente iguais ou superiores às posições de ordens $\frac{n}{2}$ e $\frac{n}{2}+1$, isto é,

$$\boxed{M_d = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}}$$

EXEMPLIFICANDO

Calcular a mediana da seguinte tabela de frequências:

Nota (X_i)	Frequência (f_i)
6	5
7	15
8	10
9	7
10	3

Vamos construir a coluna da frequência acumulada para calcular a mediana.

Nota (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})
6	5	5
7	15	20
8	10	30
9	7	37
10	3	40
TOTAL	**40**	

Como o número de elementos é par, $n = 40$, temos dois termos ocupando as posições centrais. O primeiro termo ocupa a posição $\frac{n}{2} = \frac{40}{2} = 20$; o segundo, a posição $\frac{n}{2} + 1 = 21$. Assim, a mediana será a média aritmética dos termos que ocupam essas duas posições.

A frequência acumulada indica que 20 elementos foram contados até a segunda linha. Portanto,

$$x_{20} = 7$$

Logo, o termo de posição 21 está na linha seguinte:

$$x_{21} = 8$$

Assim, a mediana é:

$$M_d = \frac{x_{20} + x_{21}}{2} = \frac{7 + 8}{2} = 7{,}5$$

(CESPE/Pref. SL/2017)

Texto 11A2CCC

A tabela a seguir apresenta uma comparação entre a evolução populacional ocorrida na cidade de São Luís, no estado do Maranhão e no Brasil, a cada cinco anos, de 1985 a 2010.

Ano	São Luís (em milhares)	Maranhão (em milhões)	Brasil (em milhões)
1985	640	4,3	137
1990	700	4,9	146
1995	780	5,2	156
2000	870	5,6	171
2005	960	6,1	183
2010	1.000	6,6	192

IBGE (com adaptações).

Com base na tabela do texto 11A2CCC, considerando-se a sequência dos seis valores correspondentes à população de São Luís, infere-se que a mediana desses valores é igual a

a) 725.000.

b) 775.000.

c) 825.000.

d) 875.000.

e) 700.000.

Comentários:

A mediana é o termo central de uma amostra ou população. Se temos 6 observações representadas na tabela, então a mediana poderá ser encontrada pela média dos termos que ocupam as posições 3 e 4 pois nesse caso não há apenas um termo central. Os dados já estão ordenados em ordem crescente. Então:

$$M_d = \frac{780 + 870}{2} = \frac{1650}{2}$$

$$M_d = 825$$

Gabarito: C.

(CESPE/TCE-PA/2016)

Número diário de denúncias registradas (X)	Frequência Relativa
0	0,3
1	0,1
2	0,2
3	0,1
4	0,3
Total	**1,0**

A tabela precedente apresenta a distribuição de frequências relativas da variável X, que representa o número diário de denúncias registradas na ouvidoria de determinada instituição pública. A partir das informações dessa tabela, julgue o item seguinte.

A mediana do número diário de denúncias registradas é igual a 2.

Comentários:

A mediana é o valor associado a uma frequência relativa acumulada de 50%, ou seja, que separa os 50% menores dos 50% maiores. Vamos calcular a frequência relativa acumulada.

Número de denúncias (X_i)	Frequência Relativa (f_r)	Frequência Relativa Acumulada (f_{rac})
0	0,3 = 30%	30%
1	0,1 = 10%	40%
2	0,2 = 20%	60% (> 50%)
3	0,1 = 10%	70%
4	0,3 = 30%	100%

A primeira linha apresenta uma frequência acumulada de 30%, indicando que 30% dos valores são iguais a zero.

A segunda linha apresenta uma frequência acumulada de 40%, indicando que 40% dos valores são menores ou iguais a 1.

A terceira linha apresenta uma frequência acumulada de 60%, indicando que 60% dos valores são menores ou iguais a 2.

Observe que o patamar de 50% foi ultrapassado na terceira linha. Portanto, a mediana é igual a 2.

Gabarito: Certo.

Mediana para dados agrupados em classes

O raciocínio adotado no cálculo da mediana para dados agrupados em classes é muito similar ao empregado no tópico anterior. Agora, contudo, **não nos importaremos com o número de elementos da série. Adotaremos um único procedimento de cálculo, independentemente de termos um número par ou ímpar de elementos**.

Considere a distribuição de frequências descrita a seguir, que resume as idades de um grupo de 50 alunos do Estratégia Concursos:

Idades	Frequência (f_i)
23 ⊢ 26	3
26 ⊢ 29	4
29 ⊢ 32	10
32 ⊢ 35	13
35 ⊢ 38	10
38 ⊢ 41	6
41 ⊢ 44	4
TOTAL	**50**

A etapa inicial do cálculo da mediana consiste na construção da coluna de frequências acumuladas:

Idades	Frequência (f_i)	Frequência Acumulada (f_{ac})
23 ⊢ 26	3	3
26 ⊢ 29	4	7
29 ⊢ 32	10	17
32 ⊢ 35	13	30
35 ⊢ 38	10	40
38 ⊢ 41	6	46
41 ⊢ 44	4	50
TOTAL	**50**	

Para calcular a mediana de dados que estão agrupados por intervalo de classes, precisamos identificar a classe em que se encontra a mediana, a chamada **classe mediana**, **que corresponde à frequência acumulada imediatamente igual ou superior à metade da frequência total, ou seja, metade da soma das frequências simples, $\sum f_i/2$**. Em nosso exemplo, temos:

$$\frac{\sum f_i}{2} = \frac{50}{2} = 25$$

Agora, devemos comparar o valor encontrado com os valores presentes na coluna de frequências acumuladas, percorrendo-os de cima para baixo. A classe mediana será a primeira classe em que a frequência acumulada for igual ou superior a 25. Assim, teremos que analisar o seguinte:

- a primeira frequência acumulada (3) é maior ou igual a 25? Não;
- a segunda frequência acumulada (7) é maior ou igual a 25? Não;
- a terceira frequência acumulada (17) é maior ou igual a 25? Não;
- a quarta frequência acumulada (30) é maior ou igual a 25? Sim.

Pronto, encontramos a **classe mediana**. Nesse ponto, paramos a comparação e verificamos que a mediana se encontra na quarta classe, isto é, no intervalo entre 32 e 35.

Conhecendo a classe mediana, podemos aplicar a fórmula da mediana, a seguir:

$$\boxed{\boldsymbol{M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h}}$$

em que:

l_{inf} é o limite inferior da classe mediana;

$f_{ac_{ant}}$ é a frequência acumulada da classe anterior à classe mediana;

f_i é a frequência simples da classe mediana; e

h é a amplitude do intervalo da classe mediana.

Já sabemos que a amplitude é a diferença entre os limites da classe. Assim, temos:

$$h = 35 - 32 = 3.$$

Os demais elementos da fórmula são ilustrados a seguir:

Idades	Frequência (f_i)	Frequência Acumulada (f_{ac})
23 ⊢ 26	3	3
26 ⊢ 29	4	7
29 ⊢ 32	10	17
32 ⊢ 35	13	30
35 ⊢ 38	10	40
38 ⊢ 41	6	46
41 ⊢ 44	4	50
TOTAL	**50**	

$f_{ac_{ant}}$

classe mediana

l_{inf}

$\sum f_i$

f_i

Após identificarmos os elementos, precisamos aplicá-los na fórmula mostrada anteriormente:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

$$M_d = 32 + \left[\frac{\left(\frac{50}{2}\right) - 17}{13}\right] \times 3$$

$$M_d = 32 + \left(\frac{25 - 17}{13}\right) \times 3$$

$$M_d = 32 + \left(\frac{8}{13}\right) \times 3 \cong 33{,}85$$

Sendo assim, podemos afirmar que:

a) 50% dos valores estão entre 23 e 33,85;

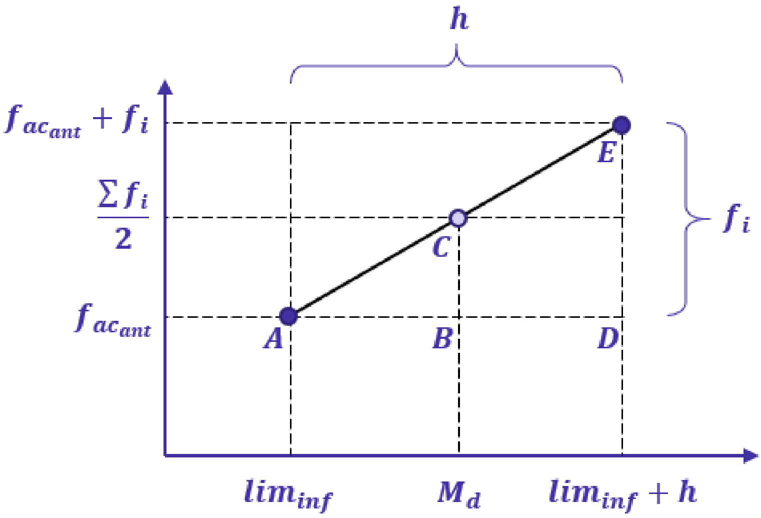
b) 50% dos valores estão entre 33,85 e 44.

TOME NOTA!

Com a memorização da fórmula da mediana para dados agrupados em classes, você conseguirá compreender, com mais facilidade, a aplicação dos **quartis**, **decis** e **percentis**. As fórmulas dessas medidas sofrem poucas alterações em relação à fórmula da mediana. Por isso, recomendo fortemente que internalizem a expressão:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

A fórmula apresentada anteriormente é obtida pelo **método de interpolação linear**. Esse método consiste, basicamente, em utilizar valores conhecidos para estimar valores desconhecidos de forma linear, isto é, por meio de uma reta. No caso da mediana, a reta se inicia no ponto $\left(lim_{inf}, f_{ac_{ant}}\right)$; passa pelo ponto $\left(M_d, \sum f_i/2\right)$ e termina no ponto $\left(lim_{inf}+h, f_{ac_{ant}}+f_i\right)$. Vejamos:



Em virtude da semelhança entre os triângulos ABC e ADE, podemos estabelecer a seguinte relação de proporcionalidade:

$$\frac{M_d - l_{inf}}{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}} = \frac{\left(l_{inf}+h\right) - l_{inf}}{\left(f_{ac_{ant}} + f_i\right) - f_{ac_{ant}}}$$

$$\frac{M_d - l_{inf}}{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}} = \frac{h}{f_i}$$

$$M_d - l_{inf} = \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

Assim, chegamos à fórmula mostrada anteriormente:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

EXEMPLIFICANDO

Calcular a mediana da distribuição de frequências apresentada a seguir, referente às estaturas de um grupo de 40 alunos:

Estaturas	Frequência (f_i)
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
TOTAL	**40**

Como sabemos, o primeiro passo é construir a coluna de frequências acumuladas:

Estaturas	Frequência (f_i)	Frequência Acumulada (f_{ac})
150 ⊢ 154	4	4
154 ⊢ 158	9	13
158 ⊢ 162	11	24
162 ⊢ 166	8	32
166 ⊢ 170	5	37
170 ⊢ 174	3	40
TOTAL	**40**	

Agora, precisamos identificar a classe mediana. Para tanto, vamos calcular sua posição por meio da expressão:

$$\frac{\sum f_i}{2} = \frac{40}{2} = 20$$

Devemos comparar esse valor com os existentes na coluna de frequências acumuladas, da mesma maneira que fizemos anteriormente. A classe mediana será a primeira classe em que a frequência acumulada for igual ou superior a 20.

Assim, teremos:

- a primeira frequência acumulada (4) é maior ou igual a 20? Não;
- a segunda frequência acumulada (13) é maior ou igual a 20? Não;
- a terceira frequência acumulada (24) é maior ou igual a 20? Sim.

Nesse ponto, paramos a comparação e verificamos que a mediana se encontra na terceira classe, isto é, no intervalo entre 158 e 162.

Conhecendo a classe mediana, podemos identificar os termos empregados na fórmula da mediana: $\sum f_i = 40;\ l_{inf} = 158;\ f_{ac_{ant}} = 13; f_i = 11;$ e $h = 162 - 158 = 4$.

Finalmente, vamos aplicar a fórmula da mediana:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

$$M_d = 158 + \left[\frac{\left(\frac{40}{2}\right) - 13}{11}\right] \times 4$$

$$M_d = 158 + \left(\frac{20 - 13}{11}\right) \times 4$$

$$M_d = 158 + \left(\frac{7}{11}\right) \times 4 \cong 160{,}54$$

Portanto, metade dos valores estão entre 150 e 160,54; e metade estão entre 160,54 e 174.

Reparem que o método de interpolação linear nada mais é que uma regra de três simples: a frequência $\boldsymbol{f_i = 11}$ está para a amplitude $\boldsymbol{h = 4}$, assim como a diferença $\left[\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}\right] = 7$ está para $\boldsymbol{x}$.

$$\begin{array}{ccc} \boldsymbol{f_i} & \longrightarrow & \boldsymbol{h} \\ \left[\left(\frac{\sum \boldsymbol{f_i}}{\boldsymbol{2}}\right) - \boldsymbol{f_{ac_{ant}}}\right] & \longrightarrow & \boldsymbol{x} \end{array} \quad \Longrightarrow \quad \begin{array}{ccc} \boldsymbol{11} & \longrightarrow & \boldsymbol{4} \\ \boldsymbol{7} & \longrightarrow & \boldsymbol{x} \end{array}$$

$$11x = 4 \times 7 \qquad \Longrightarrow \qquad x = \left(\frac{7}{11}\right) \times 4 = 2{,}54$$

Agora, para encontrar a mediana $\boldsymbol{M_d}$, basta adicionar o valor de $\boldsymbol{x}$ ao limite inferior, $\boldsymbol{l_{inf} = 158}$:

$$M_d = l_{inf} + x = 158 + 2{,}54 = 160{,}54$$

HORA DE PRATICAR!

(CESPE/DEPEN/2015)

Idade (x)	Percentual
$18 \leq x < 25$	30%
$25 \leq x < 30$	25%
$30 \leq x < 35$	20%
$35 \leq x < 45$	15%
$45 \leq x < 60$	10%
Total	100%

Felipe M. Monteiro, Gabriela R. Cardoso e Rafael da Silva. A seletividade do sistema prisional brasileiro e as políticas de segurança pública. *In*: XV Congresso Brasileiro de Sociologia, 26 a 29 de julho de 2011. Curitiba (PR). Grupo de Trabalhos - Violência e Sociedade (com adaptações).

Com base nos dados dessa tabela, julgue o item a seguir.

A mediana da distribuição mostrada é igual ou superior a 30 anos, pois as idades mínima e máxima na população prisional brasileira em 2010 foram, respectivamente, 18 e 60 anos.

Comentários:

Nessa questão, podemos afirmar que o item está errado, pois os valores mínimo e máximo não são suficientes para determinarmos o valor da mediana. Assim, ainda que o valor da mediana estivesse correto, o item estaria errado.

De todo modo, vamos calcular o valor da mediana para treinar. Primeiro, construiremos a coluna da frequência acumulada e descobriremos a classe mediana.

Idade (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})
18 - 25	30%	30%
25 - 30	**25%**	**55%**
30 - 35	20%	75%
35 - 45	15%	90%
45 - 60	10%	100%
Total	**100%**	

A classe mediana é a primeira classe com frequência acumulada maior ou igual a 50%. Dessa forma, a classe mediana é a segunda, pois $55\% \geq 50\%$. Assim, a mediana é um número entre 25 e 30.

Sabendo disso, vamos calcular o valor da mediana pelo método da interpolação:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

em que:

- o somatório das frequências é $\sum f_i = 100\%$.
- o limite inferior da classe é $l_{inf} = 25$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 30\%$.
- a frequência da própria classe é $f_i = 25\%$.
- a amplitude da classe é $h = 30 - 25 = 5$.

Agora podemos aplicar a fórmula:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i}\right] \times h$$

$$M_d = 25 + \left[\frac{50\% - 30\%}{25\%}\right] \times 5$$

$$M_d = 29$$

Gabarito: Errado.

(FCC/CNMP/2015) A tabela de frequências absolutas abaixo corresponde à distribuição dos valores dos salários dos funcionários de nível médio lotados em um órgão público no mês de dezembro de 2014.

Classe de Salários (R$)	Frequências Absolutas
$1.500 \vdash 2.500$	f_1
$2.500 \vdash 3.500$	f_2
$3.500 \vdash 4.500$	f_3
$4.500 \vdash 5.500$	f_4
$5.500 \vdash 6.500$	f_5
$6.500 \vdash 7.500$	f_6

Observação: $f_i = -i^2 + 10i + 1, 1 \leq i \leq 6$.

O valor da mediana destes salários, obtido pelo método da interpolação linear, é, em R$, igual a

a) 5.320,00.

b) 5.040,00.

c) 5.260,00.

d) 4.900,00.

e) 5.400,00.

Comentários:

Nosso primeiro passo será calcular as frequências absolutas por meio da equação apresentada ($f_i = -i^2 + 10i + 1$) no problema. Em seguida, completamos a tabela com as novas informações e calculamos as frequências acumuladas. Assim:

Classe de Salários (R$)	Frequências Absolutas	Frequências Acumuladas
$1.500 \vdash 2.500$	$f_1 = -1^2 + 10 + 1 = 10$	$\mathbf{0} + 10 = \mathbf{10}$
$2.500 \vdash 3.500$	$f_2 = -2^2 + 10 \times 2 + 2 = 17$	$\mathbf{10} + 17 = \mathbf{27}$
$3.500 \vdash 4.500$	$f_3 = -3^2 + 10 \times 3 + 3 = 22$	$\mathbf{27} + 22 = \mathbf{49}$
$4.500 \vdash 5.500$	$f_4 = -4^2 + 10 \times 4 + 4 = 25$	$\mathbf{49} + 25 = \mathbf{74}$
$5.500 \vdash 6.500$	$f_5 = -5^2 + 10 \times 5 + 5 = 26$	$\mathbf{74} + 26 = \mathbf{100}$
$6.500 \vdash 7.500$	$f_6 = -6^2 + 10 \times 6 + 6 = 25$	$\mathbf{100} + 25 = \mathbf{125}$
Total	**125**	**125**

Somando as frequências acumuladas chegamos a um total de 125 observações. A mediana é o termo central do conjunto. Portanto, temos:

$$\frac{125}{2} = 62{,}5.$$

A quarta classe é a primeira a superar esse valor em termos de frequências acumuladas, portanto, ela será nossa classe mediana. Vejamos:

Classe de Salários (R$)	Frequências Absolutas	Frequências Acumuladas
$1.500 \vdash 2.500$	10	10
$2.500 \vdash 3.500$	17	27
$3.500 \vdash 4.500$	22	49
$\mathbf{4.500 \vdash 5.500}$	$\mathbf{25}$	$\mathbf{74}\ (> \mathbf{62,5})$
$5.500 \vdash 6.500$	26	100
$6.500 \vdash 7.500$	25	125
Total	**125**	**125**

Sabendo disso, podemos calcular a mediana por meio do método de interpolação linear. Temos:

$$M_d = l_{inf} + \left[\frac{\left(\frac{\sum f_i}{2}\right) - f_{ac_{ant}}}{f_i} \right] \times h$$

em que:

o limite inferior da classe é $l_{inf} = 4.500$.

a frequência acumulada da classe anterior é $f_{ac_{ant}} = 49$.

a frequência da própria classe é $f_i = 25$.

a amplitude da classe é $h = 5.500 - 4.500 = 1.000$.

Aplicando a fórmula, temos:

$$M_d = 4.500 + \left[\frac{\left(\frac{125}{2}\right) - 49}{25}\right] \times 1.000$$

$$M_d = 4.500 + \left(\frac{13,5}{25}\right) \times 1.000$$

$$M_d = 4.500 + 13,5 \times 40$$

$$M_d = 4.500 + 540$$

$$M_d = 5040$$

Gabarito: B.

Propriedades da Mediana

A seguir, estudaremos algumas propriedades importantes sobre a mediana.

TOME NOTA!

1º Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, a mediana do conjunto fica aumentada (ou diminuída) dessa constante.**

EXEMPLIFICANDO

Para explicar essa propriedade, vamos tomar como exemplo a sequência $\{\boldsymbol{x_n}\} = \{3, 6, 8, 9, 10\}$, cuja mediana é:

$$M_{d_x} = x_{\frac{n+1}{2}} = x_{\frac{5+1}{2}} = x_{\frac{6}{2}} = x_3 = 8$$

$$M_{d_x} = 8$$

Se adicionarmos o número 5 a cada um dos termos da sequência, iremos obter uma nova lista $\{y_n\} = \{x_n + 5\} = \{8, 11, 13, 14, 15\}$, cuja mediana é:

$$M_{d_y} = y_{\frac{n+1}{2}} = y_{\frac{5+1}{2}} = y_{\frac{6}{2}} = y_3 = 13$$

$$M_{d_y} = 13$$

Veja que a adição do número 5 a cada um dos termos da sequência fez com que a mediana também aumentasse em 5 unidades, indo de 8 para 13.

TOME NOTA!

2º Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a mediana do conjunto fica multiplicada (ou dividida) por esta constante.**

EXEMPLIFICANDO

Para explicar essa propriedade, vamos tomar como exemplo a sequência $\{\boldsymbol{x_n}\} = \{3, 6, 8, 9, 10\}$, cuja mediana é:

$$M_{d_x} = x_{\frac{n+1}{2}} = x_{\frac{5+1}{2}} = x_{\frac{6}{2}} = x_3 = 8$$

$$M_{d_x} = 8$$

Se multiplicarmos cada um dos termos da sequência por 5, iremos obter uma nova lista $\{y_n\} = \{x_n \times 5\} = \{15, 30, 40, 45, 50\}$, cuja mediana é:

$$M_{d_y} = y_{\frac{n+1}{2}} = y_{\frac{5+1}{2}} = y_{\frac{6}{2}} = y_3 = 40$$

$$M_{d_y} = 40$$

Veja que a multiplicação de cada um dos termos da sequência por 5 fez com que a mediana também fosse multiplicada por 5, aumentando de 8 para 40.

TOME NOTA!

3º Propriedade

- **A soma dos desvios absolutos de uma sequência de números, em relação a um número a, é mínima quando a é a mediana dos números.**

EXEMPLIFICANDO

Para explicar essa propriedade, vamos tomar como exemplo a série $\{x_n\} = \{1, 2, 3, 4, 6, 7, 8, 9\}$. Como o número de termos é par, a mediana será, por convenção, a média aritmética dos dois termos centrais:

$$M_d = \frac{4+6}{2} = 5.$$

O desvio em relação à mediana corresponde à diferença entre cada elemento da sequência e a mediana. Como são 8 números, teremos a mesma quantidade de desvios para calcular. Logo, basta encontrarmos a diferença entre cada número e a mediana:

$$d_1 = x_1 - M_d = 1 - 5 = -4$$

$$d_2 = x_2 - M_d = 2 - 5 = -3$$

$$d_3 = x_3 - M_d = 3 - 5 = -2$$

$$d_4 = x_4 - M_d = 4 - 5 = -1$$

$$d_5 = x_5 - M_d = 6 - 5 = 1$$

$$d_6 = x_6 - M_d = 7 - 5 = 2$$

$$d_7 = x_7 - M_d = 8 - 5 = 3$$

$$d_8 = x_8 - M_d = 9 - 5 = 4$$

Agora, precisamos somar os valores absolutos (valores positivos) desses desvios:

$$\sum_{i=1}^{8} |d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8}|d_i| = |-4| + |-3| + |-2| + |-1| + |1| + |2| + |3| + |4|$$

$$\sum_{i=1}^{8}|d_i| = 4 + 3 + 2 + 1 + 1 + 2 + 3 + 4 = 20$$

A propriedade garante que, ao calcularmos a soma dos desvios absolutos em relação à mediana, o menor valor que encontraremos para essa sequência será 20.

Há um detalhe importante que precisamos esclarecer. Como vimos anteriormente, **quando o número de elementos do conjunto é ímpar, o valor da mediana é único e igual ao termo central**. Porém, **quando o número de elementos é par, a mediana pode ser qualquer valor entre os termos centrais, havendo infinitos valores possíveis para a mediana**. **Por convenção, contudo, adotamos a média aritmética dos valores centrais**.

Certo, o que isso tem a ver com a propriedade que estamos estudando? Significa dizer que, se calcularmos a soma dos desvios absolutos para qualquer valor entre 4 e 6, que são os termos centrais, o valor dos desvios absolutos em relação a mediana também será mínimo. A título exemplificativo, vamos calcular os desvios em relação ao valor 4,5:

$$d_1 = x_1 - 4{,}5 = 1 - 4{,}5 = -3{,}5$$

$$d_2 = x_2 - 4{,}5 = 2 - 4{,}5 = -2{,}5$$

$$d_3 = x_3 - 4{,}5 = 3 - 4{,}5 = -1{,}5$$

$$d_4 = x_4 - 4{,}5 = 4 - 4{,}5 = -0{,}5$$

$$d_5 = x_5 - 4{,}5 = 6 - 4{,}5 = 1{,}5$$

$$d_6 = x_6 - 4{,}5 = 7 - 4{,}5 = 2{,}5$$

$$d_7 = x_7 - 4{,}5 = 8 - 4{,}5 = 3{,}5$$

$$d_8 = x_8 - 4{,}5 = 9 - 4{,}5 = 4{,}5$$

Somando os valores absolutos desses desvios:

$$\sum_{i=1}^{8}|d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8}|d_i| = |-3{,}5| + |-2{,}5| + |-1{,}5| + |-0{,}5| + |1{,}5| + |2{,}5| + |3{,}5| + |4{,}5|$$

$$\sum_{i=1}^{8}|d_i| = 3{,}5 + 2{,}5 + 1{,}5 + 0{,}5 + 1{,}5 + 2{,}5 + 3{,}5 + 4{,}5 = 20$$

Como havíamos previsto, o valor também foi igual ao valor mínimo, 20.

Por último, a propriedade também garante que, para qualquer valor fora do intervalo entre 4 e 6, encontraremos um valor maior que o mínimo. Por exemplo, vamos calcular os desvios em relação ao valor 7:

$$d_1 = x_1 - 7 = 1 - 7 = -6$$

$$d_2 = x_2 - 7 = 2 - 7 = -5$$

$$d_3 = x_3 - 7 = 3 - 7 = -4$$

$$d_4 = x_4 - 7 = 4 - 7 = -3$$

$$d_5 = x_5 - 7 = 6 - 7 = -1$$

$$d_6 = x_6 - 7 = 7 - 7 = 0$$

$$d_7 = x_7 - 7 = 8 - 7 = 1$$

$$d_8 = x_8 - 7 = 9 - 7 = 2$$

Somando os valores absolutos desses desvios:

$$\sum_{i=1}^{8} |d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8} |d_i| = |-6| + |-5| + |-4| + |-3| + |-1| + |0| + |1| + |2|$$

$$\sum_{i=1}^{8} |d_i| = 6 + 5 + 4 + 3 + 1 + 0 + 1 + 2 = 22$$

Portanto, como havíamos previsto anteriormente, o valor foi maior que o mínimo.