By @kakashi_copiador

Índice

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 da 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.

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

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:

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

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:

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

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:

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

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,

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

FIQUE ATENTO!

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 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:

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

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. Organizando os dados da tabela em ordem crescente (isto é, em rol crescente), temos:

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

Encontrando a média 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, e tirarmos a média aritmética desses dois valores.

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,

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

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,

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

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:

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

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 $\rightarrow f_{ac_{ant}}$
32 ⊢ 35	13	30 $\rightarrow$ *classe mediana*
35 ⊢ 38	10	40
38 ⊢ 41	6	46
41 ⊢ 44	4	50
TOTAL	**50**	

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;

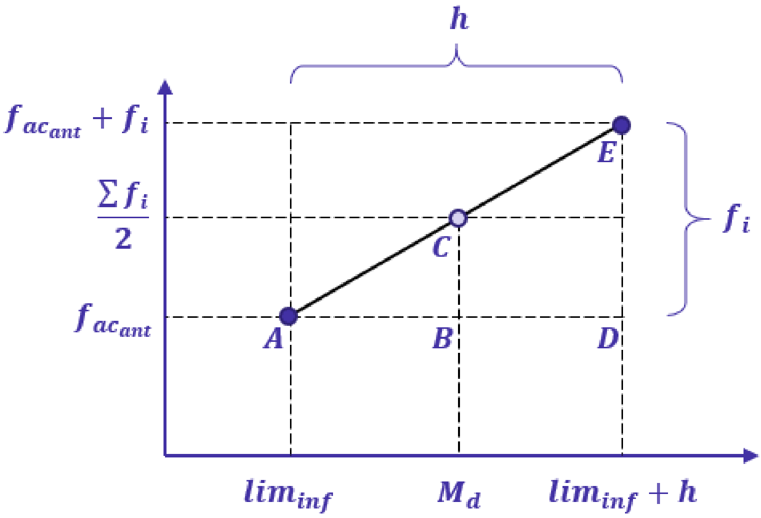
b) 50% dos valores estão entre 33,85 e 44.

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, como veremos nos próximos tópicos. 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:

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

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

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

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, podemos concluir que:

a) metade dos valores estão entre 150 e 160,54;

b) metade dos valores estão entre 160,54 e 174.

(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: $\boldsymbol{f_i = -i^2 + 10i + 1, 1 \leq i \leq 6}$ **.**

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
$\mathbf{4.500 \vdash 5.500}$	$\mathbf{25}$	$\mathbf{74\ (> 62,5)}$
$5.500 \vdash 6.500$	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.

QUARTIL, DECIL E PERCENTIL

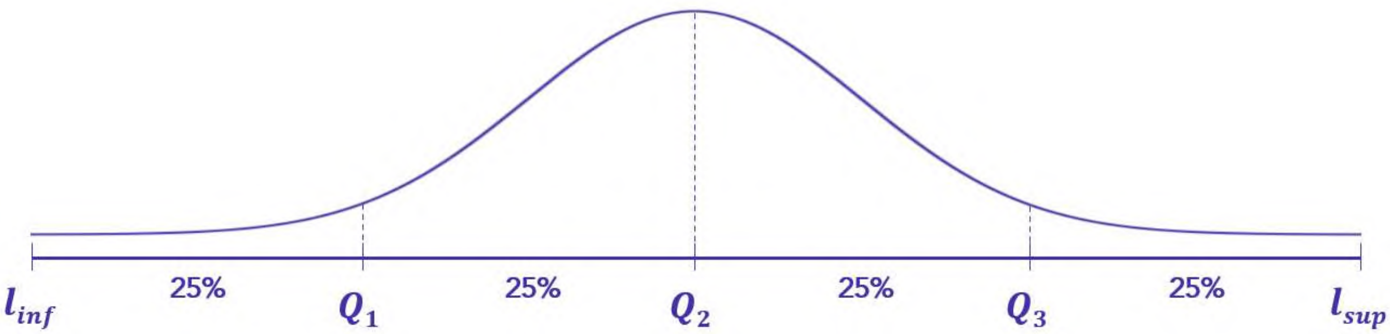
Já vimos que a mediana separa uma série em duas partes iguais, cada uma contendo o mesmo número de elementos. Contudo, uma série também pode ser dividida em um número maior de partes, todas compostas por quantidades iguais de elementos. Nesse caso, **o nome da medida separatriz é atribuído de acordo com a quantidade de partes em que a série é dividida**:

- **quartil**: divide uma série em **quatro partes** iguais (Q_1 , Q_2 , Q_3);
- **quintil**: divide uma série em **cinco partes** iguais (Qt_1 , Qt_2 , $\cdots$, Qt_5);
- **decil**: divide uma série em **dez partes** iguais (D_1 , D_2 , $\cdots$, D_9);
- **percentil**: divide uma série em **cem partes** iguais (P_1 , P_2 , $\cdots$, P_{99}).

Nessa seção, vamos detalhar algumas medidas separatrizes que também são muito exploradas em provas de concursos: os quartis, os decis e os percentis.

Quartis

Denominamos de **quartis** os valores de uma série que a dividem em **quatro partes iguais**, isto é, **quatro partes contendo o mesmo número de elementos (25%).** A imagem a seguir mostra os quartis de uma distribuição hipotética:



Temos, então, **3 quartis** (Q_1 , Q_2 e Q_3) para dividir uma série em **quatro partes** iguais:

- Q_1: o **primeiro quartil** corresponde à separação dos primeiros 25% de elementos da série;
- Q_2: o **segundo quartil** corresponde à separação de metade dos elementos da série, **coincidindo com a mediana ($Q_2 = M_d$)**;
- Q_3: o **terceiro quartil** corresponde à separação dos primeiros 75% de elementos da série, ou dos últimos 25% de elementos da série.

Para o cálculo dos **quartis**, empregaremos a mesma fórmula adotada no cálculo da mediana, apenas substituindo a expressão $\frac{\sum f_i}{2}$ por $\frac{k \times \sum f_i}{4}$, em que k indica a ordem do quartil e assume valores inteiros no intervalo de 1 a 3.

Quartil para dados não-agrupados

O cálculo do **quartil para dados não-agrupados** é realizado, de forma aproximada, por meio das etapas descritas a seguir:

- 1.ª etapa: determinamos a posição do quartil, por meio da expressão:

$$\boldsymbol{P_{Q_k} = \frac{k \times n}{4} \quad (k = 1, 2, 3);}$$

- 2.ª etapa: identificamos a posição mais próxima do rol;
- 3.ª etapa: verificamos o valor que está ocupando essa posição.

Sempre que houver necessidade, teremos que organizar o conjunto de valores por ordem de magnitude.

FIQUE ATENTO!

Embora fórmula anterior possa ser utilizada para o cálculo da posição de Q_2, por depender de uma aproximação, nem sempre o valor do segundo quartil resultará no valor convencionado para a mediana. Por isso, para o cálculo de Q_2, vamos adotar o procedimento utilizado para encontrar a mediana. Isto é, se o número de elementos for ímpar, Q_2 será representado pelo elemento que ocupar a posição central, $\frac{n+1}{2}$. Se o número de elementos do conjunto for par, Q_2 será representado pela média aritmética entre os elementos que ocuparem as posições centrais, $\frac{n}{2}$ e $\frac{n}{2} + 1$.

EXEMPLIFICANDO

Calcular os quartis Q_1 , Q_2 e Q_3 para o seguinte conjunto de valores:

$$\{11, 15, 18, 12, 19, 25, 26, 16, 15, 21, 22, 27, 14, 18, 20, 24\}$$

Reparem que os valores não estão organizados, portanto, nossa primeira tarefa será colocá-los em ordem de magnitude (rol). Assim, teremos:

$$\{11, 12, 14, 15, 15, 16, 18, 18, 19, 20, 21, 22, 24, 25, 26, 27\}$$

a) Cálculo de Q_1:

Começamos determinando a posição de Q_1:

$$P_{Q_1} = \frac{1 \times 16}{4} = 4$$

Depois, identificamos a posição mais próxima no rol. Como o resultado foi um número inteiro, a posição mais próxima coincidirá com o valor encontrado, não havendo necessidade de aproximação.

Em seguida, verificamos o valor que está ocupando a posição identificada:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
11	12	14	**15**	15	16	18	18	19	20	21	22	24	25	26	27

Portanto, o valor 15 corresponde a 25% do rol.

b) Cálculo de Q_2:

Para o cálculo de Q_2, precisamos verificar o número de elementos do conjunto. Se o número de elementos for ímpar, Q_2 será representado pelo elemento que ocupar a posição $\frac{n+1}{2}$. Se o número de elementos do conjunto for par, Q_2 será representado pela média aritmética entre os elementos que ocuparem as posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

No nosso exemplo, como o conjunto é formado por um número par de observações, Q_2 será calculado como a média aritmética dos valores que ocupam as posições 8 e 9.

$$Q_2 = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_{\frac{16}{2}} + X_{\frac{16}{2}+1}}{2} = \frac{X_8 + X_9}{2} = \frac{18 + 19}{2} = 18{,}5$$

Portanto, o valor 18,5 corresponde a 50% do rol.

c) Cálculo de Q_3:

Começamos determinando a posição de Q_3:

$$P_{Q_3} = \frac{3 \times 16}{4} = 12$$

Depois, identificamos a posição mais próxima no rol. Como o resultado foi um número inteiro, a posição mais próxima coincidirá com o valor encontrado, não havendo necessidade de aproximação.

Em seguida, verificamos o valor que está ocupando a posição identificada:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
11	12	14	15	15	16	18	18	19	20	21	**22**	24	25	26	27

Portanto, o valor 22 corresponde a 75% do rol.

Quartil para dados agrupados sem intervalos de classe

O cálculo do **quartil para dados agrupados sem intervalos de classe** é realizado, de forma aproximada, por meio das etapas descritas a seguir:

- 1.ª etapa: determinamos a **posição do quartil**, por meio da expressão:

$$\boldsymbol{P_{Q_k} = \frac{k \times \sum fi}{4} \quad (k = 1, 2, 3)}$$

em que $\sum fi$ é a soma das frequências simples;

- 2.ª etapa: identificamos a posição do **quartil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do quartil**;
- 3.ª etapa: verificamos o valor da variável que corresponde a essa posição.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

FIQUE ATENTO!

Embora fórmula anterior possa ser utilizada para o cálculo da posição de Q_2, por depender de uma aproximação, nem sempre o valor do segundo quartil resultará no valor convencionado para a mediana. Por isso, para o cálculo de Q_2, vamos adotar o procedimento utilizado para encontrar a mediana. Isto é, se o número de elementos for ímpar, Q_2 será representado pelo elemento que ocupar a posição central, $\frac{n+1}{2}$. Se o número de elementos do conjunto for par, Q_2 será representado pela média aritmética entre os elementos que ocuparem as posições centrais, $\frac{n}{2}$ e $\frac{n}{2}+1$.

EXEMPLIFICANDO

Calcular Q_1, Q_2 e Q_3 da tabela de frequências a seguir, que representa a quantidade de filhos dos professores do Estratégia Concursos:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	4	4
1	6	10
2	9	19
3	5	24
4	4	28
TOTAL	**28**	

a) Cálculo de Q_1:

Começamos determinando a posição de Q_1:

$$P_{Q_1} = \frac{1 \times 28}{4} = 7$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	**Frequência (f_i)**	**Frequência Acumulada (f_{ac})**
0	4	4
1	**6**	**10 ($\geq$ 7)**
2	9	19
3	5	24
4	4	28
TOTAL	**28**	

Portanto, a quantidade de 1 filho corresponde a 25% do rol.

b) Cálculo de Q_2:

Para o cálculo de Q_2, precisamos verificar o número de elementos do conjunto. Se o número de elementos for ímpar, Q_2 será representado pelo elemento que ocupar a posição $\frac{n+1}{2}$. Se o número de elementos do conjunto for par, Q_2 será representado pela média aritmética entre os elementos que ocuparem as posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

No nosso exemplo, como o conjunto é formado por um número par de observações, Q_2 será calculado como a média aritmética dos valores que ocupam as posições 14 e 15.

$$Q_2 = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_{\frac{28}{2}} + X_{\frac{28}{2}+1}}{2} = \frac{X_{14} + X_{15}}{2} = \frac{2+2}{2} = 2$$

Portanto, a quantidade de 2 filhos corresponde a 50% do rol.

c) Cálculo de Q_3:

Começamos determinando a posição de Q_3:

$$P_{Q_3} = \frac{3 \times 28}{4} = 21$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	**Frequência (f_i)**	**Frequência Acumulada (f_{ac})**
0	4	4
1	6	10
2	9	19
3	**5**	**24 (≥ 21)**
4	4	28
TOTAL	**28**	

Portanto, a quantidade de 3 filhos corresponde a 75% do rol.

Quartil para dados agrupados em classes

O cálculo do **quartil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinamos a **posição do quartil**, por meio da expressão:

$$P_{Q_k} = \frac{k \times \sum fi}{4} \quad (k = 1, 2, 3);$$

em que:

$k =$ índice do quartil;

$\sum fi =$ somatório das frequências simples.

- 2.ª etapa: identificamos a posição do **quartil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do quartil**;

- 3.ª etapa: verificamos as informações referentes à classe correspondente a essa posição; e

- 4.ª etapa: calculamos o valor do **quartil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes**, mudando-se apenas o $\frac{k \times \sum fi}{4}$:

$$\boldsymbol{Q_k = l_{inf_{Q_k}} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac_{ant}}}{f_{Q_k}}\right] \times h_{Q_k}}$$

em que:

$l_{inf_{Q_k}} =$ limite inferior da classe do quartil considerado;

$f_{ac_{ant}} =$ frequência acumulada da classe anterior à classe do quartil considerado;

$h_{Q_k} =$ amplitude do intervalo de classe do quartil considerado;

$f_{Q_k} =$ frequência simples da classe do quartil considerado.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

EXEMPLIFICANDO

Calcular Q_1, Q_2 e Q_3 da distribuição de frequências dos pesos de um grupo de 180 alunos do Estratégia Concursos:

i	Pesos (kg)	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	10 ⊢ 30	10	10
2	30 ⊢ 50	24	34
3	50 ⊢ 70	57	91
4	70 ⊢ 90	44	135
5	90 ⊢ 110	29	164
6	110 ⊢ 130	16	180
	TOTAL	180	

a) Cálculo de Q_1:

Começamos determinando a posição de Q_1:

$$P_{Q_1} = \frac{1 \times \sum fi}{4} = \frac{180}{4} = 45$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

	i	Pesos (kg)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
	1	10 ⊢ 30	10	10	
	2	30 ⊢ 50	24	34	$f_{ac_{ant}}$
	3	50 ⊢ 70	57	91 (> 45)	*classe considerada*
l_{inf}	4	70 ⊢ 90	44	135	f_i
	5	90 ⊢ 110	29	164	
	6	110 ⊢ 130	16	180	
		TOTAL	180		$\sum f_i$

Em seguida, encontramos o valor numérico de Q_1 utilizando a expressão:

$$Q_1 = l_{inf_{Q_1}} + \left[\frac{\frac{1 \times \sum fi}{4} - f_{ac_{ant}}}{f_{Q_1}}\right] \times h_{Q_1}$$

$$Q_1 = 50 + \left[\frac{45 - 34}{57}\right] \times 20 \cong 53{,}9\ Kg$$

b) Cálculo de Q_2:

Começamos determinando a posição de Q_2:

$$P_{Q_2} = \frac{2 \times \sum f_i}{4} = \frac{2 \times 180}{4} = 90$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

	i	Pesos (kg)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
	1	10 ⊢ 30	10	10	
	2	30 ⊢ 50	24	34	$f_{ac_{ant}}$
	3	50 ⊢ 70	57	91 (> 90)	*classe considerada*
	4	70 ⊢ 90	44	135	f_i
l_{inf}	5	90 ⊢ 110	29	164	
	6	110 ⊢ 130	16	180	
		TOTAL	180		$\sum f_i$

Em seguida, encontramos o valor numérico de Q_2 utilizando a expressão:

$$Q_2 = l_{inf_{Q_2}} + \left[\frac{\frac{2 \times \sum fi}{4} - f_{ac_{ant}}}{f_{Q_2}}\right] \times h_{Q_2}$$

$$Q_2 = 50 + \left[\frac{90 - 34}{57}\right] \times 20 \cong 69{,}7\ Kg$$

c) Cálculo de Q_3:

Começamos determinando a posição de Q_3:

$$P_{Q_3} = \frac{3 \times \sum f_i}{4} = \frac{3 \times 180}{4} = 135$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

i	Pesos (kg)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
1	10 ⊢ 30	10	10	
2	30 ⊢ 50	24	34	
3	50 ⊢ 70	57	91	$f_{ac_{ant}}$
4	70 ⊢ 90	44	135 (= 135)	*classe considerada*
5	90 ⊢ 110	29	164	f_i
6	110 ⊢ 130	16	180	
	TOTAL	180		$\sum f_i$

l_{inf}

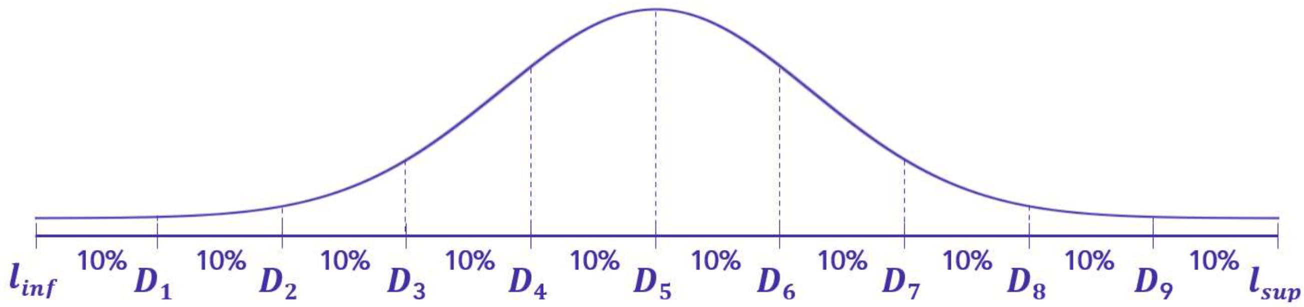
Em seguida, encontramos o valor numérico de Q_3 utilizando a expressão:

$$Q_3 = l_{inf_{Q_3}} + \left[\frac{\frac{3 \times \sum fi}{4} - f_{ac_{ant}}}{f_{Q_3}}\right] \times h_{Q_3}$$

$$Q_3 = 70 + \left[\frac{135 - 91}{44}\right] \times 20 = 90\ Kg$$

Decis

Denominamos de **decis** os valores de uma série que a dividem em **dez partes iguais**, isto é, **dez partes contendo o mesmo número de elementos (10%)**. A imagem a seguir mostra os decis de uma distribuição hipotética:



Temos, então, **9 decis** (D_1 , D_2, $\cdots$, D_9) para dividir uma série em **dez partes** iguais:

- D_1: o **primeiro decil** corresponde à separação dos primeiros 10% de elementos da série;
- D_5: o **quinto decil** corresponde à separação de metade dos elementos da série, **coincidindo com a mediana ($D_5 = M_d$)**;
- D_9: o **nono decil** corresponde à separação dos primeiros 90% de elementos da série, ou dos últimos 10% de elementos da série.

Para o cálculo dos **decis**, empregaremos a mesma fórmula adotada no cálculo da mediana, apenas substituindo a expressão $\frac{\sum f_i}{2}$ por $\frac{k \times \sum f_i}{10}$, em que k indica a ordem do decil e assume valores inteiros no intervalo de 1 a 9.

Decil para dados não-agrupados

O cálculo do **decil** segue o mesmo raciocínio empregado no cálculo do quartil para dados não-agrupados. A primeira tarefa que devemos realizar, se houver necessidade, é organizar o conjunto de valores por ordem de magnitude. Depois disso, procedemos conforme as seguintes etapas:

- 1.ª etapa: determinamos a posição do decil, por meio da expressão:

$$\boldsymbol{P_{D_k} = \frac{k \times n}{10} \quad (k = 1, 2, \cdots, 9);}$$

- 2.ª etapa: identificamos a posição mais próxima do rol;
- 3.ª etapa: verificamos o valor que está ocupando essa posição.

EXEMPLIFICANDO

Calcularemos os quartis D_1 e D_8 para o seguinte conjunto de valores:

$$\{5, 12, 15, 20, 2, 3, 4, 18, 10, 22\}$$

Reparem que os valores não estão organizados. Portanto, nossa primeira tarefa será colocá-los em ordem de magnitude (rol):

$$\{2, 3, 4, 5, 10, 12, 15, 18, 20, 22\}$$

a) Cálculo de D_1:

Começamos determinando a posição de D_1:

$$P_{D_1} = \frac{1 \times 10}{10} = 1$$

Depois, identificamos a posição mais próxima no rol. Como o resultado foi um número inteiro, a posição mais próxima coincidirá com o valor encontrado, não havendo necessidade de aproximação.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
2	3	4	5	10	12	15	18	20	22

Portanto, o valor 2 corresponde a 10% do rol.

b) Cálculo de D_8:

Começamos determinando a posição de D_8:

$$P_{D_8} = \frac{8 \times 10}{10} = 8$$

Depois, identificamos a posição mais próxima no rol. Como o resultado foi um número inteiro, a posição mais próxima coincidirá com o valor encontrado, não havendo necessidade de aproximação.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
2	3	4	5	10	12	15	**18**	20	22

Portanto, o valor 18 corresponde a 80% do rol.

Decil para dados agrupados sem intervalos de classe.

O cálculo do **decil para dados agrupados sem intervalos de classe** será realizado por meio das etapas descritas a seguir:

- 1.ª etapa: determinamos a **posição do decil**, por meio da expressão:

$$P_{D_k} = \frac{k \times \sum fi}{10} \quad (k = 1, 2, \cdots, 9)$$

em que $\sum fi$ é a soma das frequências simples;

- 2.ª etapa: identificamos a posição do **decil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do decil**;
- 3.ª etapa: verificamos o valor da variável que corresponde a essa posição.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

EXEMPLIFICANDO

Vamos calcular D_3 e D_8 da tabela de frequências a seguir, que representa a quantidade de filhos dos professores do Estratégia Concursos:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	18	18
1	35	53
2	46	99
3	28	127
4	25	152
5	10	162
6	5	167
7	3	170
TOTAL	**170**	

a) Cálculo de D_3:

Começamos determinando a posição de D_3:

$$P_{D_3} = \frac{3 \times 170}{10} = 51$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	18	18
1	**35**	**53 (≥ 51)**
2	46	99
3	28	127
4	25	152
5	10	162
6	5	167
7	3	170
TOTAL	**170**	

Portanto, a quantidade de 1 filho corresponde a 30% do rol.

b) Cálculo de D_8:

Começamos determinando a posição de D_8:

$$P_{D_8} = \frac{8 \times 170}{10} = 136$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	18	18
1	35	53
2	46	99
3	28	127
4	**25**	**152 (≥ 136)**
5	10	162
6	5	167
7	3	170
TOTAL	**170**	

Portanto, a quantidade de 4 filhos corresponde a 80% do rol.

Decil para dados agrupados em classes

O cálculo do **decil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinamos a **posição do decil**, por meio da expressão:

$$P_{D_k} = \frac{k \times \sum fi}{10} \quad (k = 1, 2, 3, \cdots, 9);$$

em que:

k = índice do decil;

$\sum fi$ = somatório das frequências simples.

- 2.ª etapa: identificamos a posição do **decil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do decil**;

- 3.ª etapa: verificamos as informações referentes à classe correspondente a essa posição; e

- 4.ª etapa: calculamos o valor do **decil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes**, mudando-se apenas o $\frac{k \times \sum fi}{10}$:

$$\boldsymbol{D_k = l_{inf\,D_k} + \left[\frac{\frac{k \times \sum f_i}{10} - f_{ac\,ant}}{f_{D_k}}\right] \times h_{D_k}}$$

em que:

$l_{inf\,D_k}$ = limite inferior da classe do decil considerado;

$f_{ac\,ant}$ = frequência acumulada da classe anterior à classe do decil considerado;

h_{D_k} = amplitude do intervalo de classe do decil considerado;

f_{D_k} = frequência simples da classe do decil considerado.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

EXEMPLIFICANDO

Calcular D_1, D_2 e D_7 da distribuição de frequências das estaturas de um grupo de 54 alunos do Estratégia Concursos:

i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	120 ⊢ 128	6	6
2	128 ⊢ 136	12	18
3	136 ⊢ 144	16	34
4	144 ⊢ 152	13	47
5	152 ⊢ 160	7	54
	TOTAL	**54**	

a) Cálculo de D_1:

Começamos determinando a posição de D_1:

$$P_{D_1} = \frac{1 \times \sum f_i}{10} = \frac{54}{10} = 5{,}4$$

Depois disso, identificamos a posição na coluna de frequências acumuladas.

	i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
	1	120 ⊢ 128	6	6 (> 5,4)	*classe considerada*
l_{inf}	2	128 ⊢ 136	12	18	f_i
	3	136 ⊢ 144	16	34	
	4	144 ⊢ 152	13	47	
	5	152 ⊢ 160	7	54	
		TOTAL	54		$\sum f_i$

Em seguida, encontramos o valor numérico de D_1 utilizando a expressão:

$$D_1 = l_{inf_{D_1}} + \left[\frac{\frac{1 \times \sum f_i}{10} - f_{ac_{ant}}}{f_{D_1}}\right] \times h_{D_1}$$

$$D_1 = 120 + \left[\frac{5{,}4 - 0}{6}\right] \times 8 = 127{,}5\ cm$$

b) Cálculo de D_2:

Começamos determinando a posição de D_2:

$$P_{D_2} = \frac{2 \times \sum f_i}{10} = \frac{2 \times 54}{10} = 10{,}8$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

	i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
	1	120 ⊢ 128	6	6	$f_{ac_{ant}}$
	2	128 ⊢ 136	12	18 (> 10,8)	*classe considerada*
l_{inf}	3	136 ⊢ 144	16	34	f_i
	4	144 ⊢ 152	13	47	
	5	152 ⊢ 160	7	54	
		TOTAL	54		$\sum f_i$

Em seguida, encontramos o valor numérico de D_2 utilizando a expressão:

$$D_2 = l_{inf_{D_2}} + \left[\frac{\frac{2 \times \sum f_i}{10} - f_{ac_{ant}}}{f_{D_2}}\right] \times h_{D_2}$$

$$D_2 = 128 + \left[\frac{10{,}8 - 6}{12}\right] \times 8 = 131{,}2\ cm$$

c) Cálculo de D_7:

Começamos determinando a posição de D_7:

$$P_{D_7} = \frac{7 \times \sum f_i}{10} = \frac{7 \times 54}{10} = 37{,}8$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

	i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
	1	120 ⊢ 128	6	6	
	2	128 ⊢ 136	12	18	
	3	136 ⊢ 144	16	34	$f_{ac_{ant}}$
	4	144 ⊢ 152	13	47 (> 37,8)	*classe considerada*
l_{inf}	5	152 ⊢ 160	7	54	f_i
		TOTAL	54		$\sum f_i$

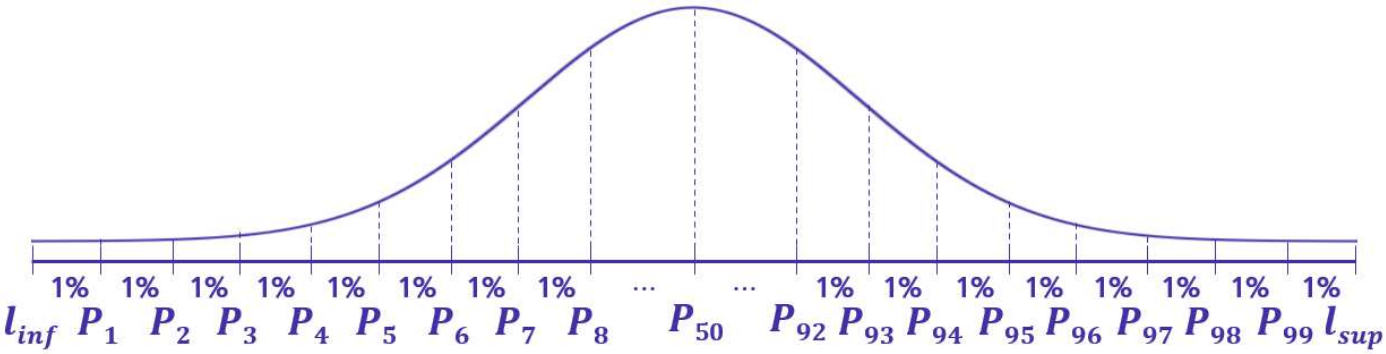
Em seguida, encontramos o valor numérico de D_7 utilizando a expressão:

$$D_7 = l_{inf\,D_7} + \left[\frac{\frac{7 \times \sum f_i}{10} - f_{ac_{ant}}}{f_{D_7}}\right] \times h_{D_7}$$

$$D_7 = 144 + \left[\frac{37{,}8 - 34}{13}\right] \times 8 = 146{,}3\ cm$$

Percentis

Denominamos de percentis os valores de uma série que a dividem em **cem partes iguais**, isto é, **cem partes contendo o mesmo número de elementos (1%)**. A imagem a seguir mostra os percentis de uma distribuição hipotética:



Temos, então, **99 percentis** (P_1 , P_2, ⋯, P_{99}) para dividir uma série em **cem partes** iguais:

- P_1: o **primeiro percentil** corresponde à separação do primeiro 1% de elementos da série;
- P_{50}: o **quinquagésimo percentil** corresponde à separação de metade dos elementos da série, **coincidindo com a mediana ($P_{50} = M_d$)**;
- P_{99}: o **nonagésimo nono percentil** corresponde à separação dos primeiros 99% de elementos da série, ou do último 1% de elementos da série.

Para o cálculo dos **percentis**, empregaremos a mesma fórmula adotada no cálculo da mediana, apenas substituindo a expressão $\frac{\sum f_i}{2}$ por $\frac{k \times \sum f_i}{100}$, em que k indica a ordem do percentil e assume valores inteiros no intervalo de 1 a 99.

Percentil para dados não-agrupados

O cálculo do **percentil** segue o mesmo raciocínio empregado nos cálculos do quartil e do decil para dados não-agrupados. A primeira tarefa que devemos realizar, se houver necessidade, é organizar o conjunto de valores por ordem de magnitude. Depois disso, colocamos em prática as seguintes etapas:

- 1.ª etapa: determinamos a posição do percentil, por meio da expressão:

$$P_{P_k} = \frac{k \times n}{100} \quad (k = 1, 2, \cdots, 99);$$

- 2.ª etapa: identificamos a posição mais próxima do rol;
- 3.ª etapa: verificamos o valor que está ocupando essa posição.

EXEMPLIFICANDO

Calcularemos os percentis P_{27} e P_{83} para o seguinte conjunto de valores:

$$\{15, 2, 4, 6, 10, 12, 13, 7, 21, 18, 20\}$$

Reparem que os valores não estão organizados. Portanto, nossa primeira tarefa será colocá-los em ordem de magnitude (rol):

$$\{2, 4, 6, 7, 10, 12, 13, 15, 18, 20, 21\}$$

a) Cálculo de P_{27}:

Começamos determinando a posição de P_{27}:

$$P_{P_{27}} = \frac{27 \times 11}{100} = 2{,}97$$

Depois, identificamos a posição mais próxima no rol:

$$P_{P_{27}} = 3$$

Em seguida, verificamos o valor que está ocupando essa posição.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
2	4	6	7	10	12	13	15	18	20	21

Portanto, o valor 6 corresponde a 27% do rol.

b) Cálculo de P_{83}:

Começamos determinando a posição de P_{83}:

$$P_{P_{83}} = \frac{83 \times 11}{100} = 9{,}13$$

Depois, identificamos a posição mais próxima no rol:

$$P_{P_{27}} = 9$$

Em seguida, verificamos o valor que está ocupando essa posição.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
2	4	6	7	10	12	13	15	**18**	20	21

Portanto, o valor 18 corresponde a 83% do rol.

Percentil para dados agrupados sem intervalos de classe

O cálculo do **percentil para dados agrupados sem intervalos de classe** será realizado por meio das etapas descritas a seguir:

- 1.ª etapa: determinamos a **posição do percentil**, por meio da expressão:

$$P_{P_k} = \frac{k \times \sum fi}{100} \qquad (k = 1, 2, \cdots, 99)$$

em que $\sum fi$ é a soma das frequências simples;

- 2.ª etapa: identificamos a posição do **percentil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do percentil**;
- 3.ª etapa: verificamos o valor da variável que corresponde a essa posição.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

EXEMPLIFICANDO

Vamos calcular P_{45} e P_{93} da tabela de frequências a seguir, que representa a quantidade de filhos dos professores do Estratégia Concursos:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	15	15
2	30	45
3	20	65
4	12	77
5	10	87
6	8	95
TOTAL	**95**	

a) Cálculo de P_{45}:

Começamos determinando a posição de P_{45}:

$$P_{P_{45}} = \frac{45 \times 95}{100} = 42{,}75$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	15	15
2	**30**	**45 (≥ 43)**
3	20	65
4	12	77
5	10	87
6	8	95
TOTAL	**95**	

Portanto, a quantidade de 2 filhos corresponde a 45% do rol.

b) Cálculo de P_{93}:

Começamos determinando a posição de P_{93}:

$$P_{P_{93}} = \frac{93 \times 95}{100} = 88{,}35$$

Depois disso, identificamos a posição na coluna de frequências acumuladas. Em seguida, verificamos a variável que corresponde a essa posição:

Filhos	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	15	15
2	30	45
3	20	65
4	12	77
5	10	87
6	**8**	**95 ($\geq$ 88)**
TOTAL	**95**	

Portanto, a quantidade de 6 filhos corresponde a 93% do rol.

Percentil para dados agrupados em classes

O cálculo do **percentil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinamos a **posição do percentil**, por meio da expressão:

$$P_{P_k} = \frac{k \times \sum fi}{100} \quad (k = 1, 2, 3, \cdots, 99);$$

em que:

$k =$ índice do percentil;

$\sum fi =$ somatório das frequências simples.

- 2.ª etapa: identificamos a posição do **percentil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do percentil**;

- 3.ª etapa: verificamos as informações referentes à classe correspondente a essa posição; e

- 4.ª etapa: calculamos o valor do **percentil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes**, mudando-se apenas o $\frac{k \times \sum fi}{100}$:

$$\boxed{\boldsymbol{P_k = l_{inf_{P_k}} + \left[\frac{\frac{k \times \sum f_i}{100} - f_{ac_{ant}}}{f_{P_k}}\right] \times h_{P_k}}}$$

em que:

$l_{inf_{P_k}} =$ limite inferior da classe do percentil considerado;

$f_{ac_{ant}} =$ frequência acumulada da classe anterior à classe do percentil considerado;

$h_{P_k} =$ amplitude do intervalo de classe do percentil considerado;

$f_{P_k} =$ frequência simples da classe do percentil considerado.

Sempre que houver necessidade, teremos que incluir uma coluna de **frequências acumuladas**.

EXEMPLIFICANDO

Calcular P_{45} e P_{77} da distribuição de frequências das estaturas de um grupo de 54 alunos do Estratégia Concursos:

i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})
1	120 ⊢ 128	5	5
2	128 ⊢ 136	13	18
3	136 ⊢ 144	16	34
4	144 ⊢ 152	13	47
5	152 ⊢ 160	7	54
	TOTAL	**54**	

a) Cálculo de P_{45}:

Começamos determinando a posição de P_{45}:

$$P_{P_{45}} = \frac{45 \times \sum f_i}{100} = \frac{45 \times 54}{100} = 24{,}30$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
1	120 ⊢ 128	5	5	
2	128 ⊢ 136	13	18	→ $f_{ac_{ant}}$
3	136 ⊢ 144 (l_{inf})	16 (f_i)	34 (> 24,30)	→ *classe considerada*
4	144 ⊢ 152	13	47	
5	152 ⊢ 160	7	54	
	TOTAL	54		→ $\sum f_i$

Em seguida, encontramos o valor numérico de P_{45} utilizando a expressão:

$$P_{45} = l_{inf\,P_{45}} + \left[\frac{\frac{45 \times \sum fi}{100} - f_{ac_{ant}}}{f_{P_{45}}}\right] \times h_{P_{45}}$$

$$P_{45} = 136 + \left[\frac{24{,}30 - 18}{16}\right] \times 8 = 139{,}15\ cm$$

b) Cálculo de P_{77}:

Começamos determinando a posição de P_{77}:

$$P_{P_{77}} = \frac{77 \times \sum f_i}{100} = \frac{77 \times 54}{100} = 41{,}04$$

Depois, identificamos a classe que representa essa posição na coluna de frequências acumuladas:

i	Estaturas (cm)	Frequência (f_i)	Frequência Acumulada (f_{ac})	
1	120 ⊢ 128	5	5	
2	128 ⊢ 136	13	18	
3	136 ⊢ 144	16	34	$f_{ac_{ant}}$
4	144 ⊢ 152	13	47 (> 41,04)	$classe\ considerada$
5	152 ⊢ 160	7	54	f_i
	TOTAL	54		$\sum f_i$

l_{inf}

Em seguida, encontramos o valor numérico de P_{77} utilizando a expressão:

$$P_{77} = l_{inf\,P_{77}} + \left[\frac{\frac{77 \times \sum fi}{100} - f_{ac_{ant}}}{f_{P_{77}}}\right] \times h_{P_{77}}$$

$$P_{77} = 144 + \left[\frac{41{,}58 - 34}{13}\right] \times 8 = 148{,}66\ cm$$

(FCC/BANRISUL/2019) As idades dos 120 funcionários lotados em uma repartição pública estão distribuídas conforme a tabela de frequências absolutas abaixo.

Idade (x) em anos	Número de funcionários
$20 < x \leq 30$	40
$30 < x \leq 40$	50
$40 < x \leq 50$	20
$50 < x \leq 60$	10
Total	120

Utilizando o método da interpolação linear, obteve-se o primeiro quartil ($Q1$) e a mediana (Md) desta distribuição em anos. A amplitude do intervalo [$Q1$,Md] é então igual a

a) 4,0.

b) 6,5.

c) 10,0.

d) 3,5.

e) 7,5.

Comentários:

Para calcularmos o primeiro quartil, precisamos encontrar a posição:

$$\frac{1n}{4} = \frac{1 \times 120}{4} = 30$$

Agora, vamos determinar a classe em que se encontra o primeiro quartil. Devemos procurar a primeira frequência acumulada que é maior do que ou igual a 30. Vejam que a primeira frequência acumulada já é maior do que 30. Portanto, o primeiro quartil está na primeira classe.

Idade em anos	Número de funcionários	Frequência acumulada
20 - 30	40	40 (> 30)
30 - 40	50	90
40 - 50	20	110
50 - 60	10	120
Total	120	

Agora é só aplicar a fórmula do Q_1:

$$Q_1 = l_{inf} + \left[\frac{\frac{1n}{4} - fac_{ant}}{f_i}\right] \times h$$

- o limite inferior da classe é $l_{inf} = 20$.
- a frequência acumulada da classe anterior é $fac_{ant} = 0$, pois não há classe anterior.
- a frequência da própria classe é $f_i = 40$.
- a amplitude da classe é $h = 30 - 20 = 10$.

Jogando esses valores na fórmula, teremos:

$$Q_1 = 20 + \left[\frac{30 - 0}{40}\right] \times 10$$

$$Q_1 = 27{,}5$$

Pronto, agora vamos calcular a mediana. Para tanto, precisamos descobrir a posição:

$$\frac{n}{2} = \frac{120}{2} = 60$$

De posse dessa informação, podemos determinar a classe em que se encontra a mediana. Devemos procurar a primeira frequência acumulada que é maior do que ou igual a 60. A segunda frequência acumulada é maior que 60. Portanto, a mediana está na segunda classe.

Idade em anos	Número de funcionários	Frequência acumulada
20 - 30	40	40
30 - 40	50	90 (> 60)
40 - 50	20	110
50 - 60	10	120
Total	120	

Agora é só aplicar a fórmula da M_d:

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

- o limite inferior da classe é $l_{inf} = 30$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 40$.
- a frequência da própria classe é $f_i = 50$.
- a amplitude da classe é $h = 40 - 30 = 10$.

Jogando esses valores na fórmula, teremos:

$$M_d = 30 + \left[\frac{60 - 40}{50}\right] \times 10$$

$$M_d = 34$$

A amplitude do intervalo $[Q_1, M_d]$ é dada pela diferença entre os extremos.

$$M_d - Q_1 = 34 - 27,50 = 6,50$$

Gabarito: B.

(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.

O valor do primeiro quartil do conjunto de dados (Q1/4) é igual a 3.

Comentários:

Podemos usar a seguinte fórmula para encontrar a posição do primeiro quartil:

$$P_{Q_1} = \frac{1 \times n}{4}$$

Assim, temos:

$$1 \times \frac{11}{4} = 2,75$$

Logo, o primeiro quartil está entre as posições 2 e 3. De acordo com o conjunto, o valor do primeiro quartil é 2.

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.

O valor do terceiro quartil do conjunto de dados (Q3/4) é igual a 4.

Comentários:

Podemos usar a seguinte fórmula para encontrar a posição do terceiro quartil:

$$P_{Q_3} = \frac{3 \times n}{4}$$

Assim, temos:

$$3 \times \frac{11}{4} = 8{,}25$$

Logo, o terceiro quartil está entre as posições 8 e 9. De acordo com o conjunto, o valor do terceiro quartil é 4.

Gabarito: Certo.

(FGV/DPE-RJ/2014) Um levantamento sobre o local de residência das pessoas que recorrem à Defensoria Pública indicou que 30% moram a menos de 10 quilômetros distância de um dos pontos de atendimento, 70% a menos de 20 quilômetros, 90% a menos de 30 quilômetros e que nenhuma delas mora a mais de 40 quilômetros de distância. Supondo ainda que a distribuição das distâncias, dentro das faixas especificadas (10 Km), é uniforme pode-se afirmar que

a) A distância mediana é de 15 quilômetros e o percentil 75 é 22,5 quilômetros.

b) A distância mediana é de 20 quilômetros e o desvio-padrão é de 8 quilômetros.

c) A distância mediana é de 20 quilômetros a moda é de 18 quilômetros.

d) A distância mediana é de 15 quilômetros e o decil 9 é de 35 quilômetros.

e) A distância mediana é de 15 quilômetros e a média é de 15 quilômetros.

Comentários:

Para entendermos a questão vamos montar uma tabela com os dados fornecidos pelo enunciado.

Distância (x)	Freq. Relativa Acumulada
$0 \leq x < 10$	30%
$\mathbf{10 \leq x < 20}$	**70% (> 50%)**
$20 \leq x < 30$	90%
$30 \leq x < 40$	100%

Para identificar a mediana, precisamos encontrar o termo central da amostra. Como temos uma frequência acumulada total de 100%, a mediana ocupará a posição representada por 50%. Dessa forma, nossa classe mediana será o intervalo entre 10 e 20 quilômetros, pois é o primeiro a superar a marca de 50%.

Sabendo disso, podemos aplicar a fórmula da mediana:

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

- o limite inferior da classe é $l_{inf} = 10$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 30\%$.
- a frequência da própria classe é $f_i = 40\%$.
- a amplitude da classe é $h = 20 - 10 = 10$.

Jogando esses valores na fórmula, teremos:

$$M_d = 10 + \left(\frac{50\% - 30\%}{40\%}\right) \times 10$$

$$M_d = 10 + \left(\frac{20\%}{40\%}\right) \times 10$$

$$M_d = 10 + \left(\frac{1}{2}\right) \times 10$$

$$M_d = 10 + 5 = 15$$

Portanto, eliminamos as alternativas B e C, pois a mediana é 15 e não 20 quilômetros.

Agora, vamos calcular o septuagésimo quinto percentil (P_{75}). Para isso, precisamos identificar a classe que corresponde à frequência acumulada de 75%. Vejamos:

Distância (x)	Freq. Relativa Acumulada
$0 \leq x < 10$	30%
$10 \leq x < 20$	70%
$\mathbf{20 \leq x < 30}$	**90% (> 75%)**
$30 \leq x < 40$	100%

Pronto, agora podemos aplicar a fórmula do percentil:

$$P_{75} = l_{inf} + \left[\frac{\frac{75 \times \sum fi}{100} - f_{ac_{ant}}}{f_i}\right] \times h$$

- o limite inferior da classe é $l_{inf} = 20$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 70\%$.

- a frequência da própria classe é $f_i = 20\%$.
- a amplitude da classe é $h = 30 - 20 = 10$.

Jogando esses valores na fórmula, teremos:

$$P_{75} = 20 + \left[\frac{75\% - 70\%}{20\%}\right] \times 10$$

$$P_{75} = 20 + \left[\frac{5\%}{20\%}\right] \times 10$$

$$P_{75} = 20 + \left(\frac{1}{4}\right) \times 10$$

$$P_{75} = 20 + 2{,}5 = 22{,}5$$

Analisando as alternativas, observamos que a letra A está correta.

Para não restar dúvidas, vamos analisar as alternativas D e E:

- Letra D: Errada. A alternativa se refere ao nono decil, que é o valor cuja frequência relativa acumulada vale 90%. A tabela nos mostra que esse valor corresponde a 30 km, e não a 35 km.

- Letra E: Errada. Precisamos calcular as frequências relativas simples para sabermos a média. Para isso tomamos a diferença entre duas frequências acumuladas seguidas para termos a frequência simples. Montando a tabela, temos:

Distância	Ponto Médio (X)	Frequência Simples (f)	$X \times f$
$0 \leq x < 10$	5	$0{,}30$	$1{,}5$
$10 \leq x < 20$	15	$0{,}40$	6
$20 \leq x < 30$	25	$0{,}20$	5
$30 \leq x < 40$	35	$0{,}10$	$3{,}5$
Total		$\mathbf{1}$	$\mathbf{16}$

Então, a média vale 16 km.

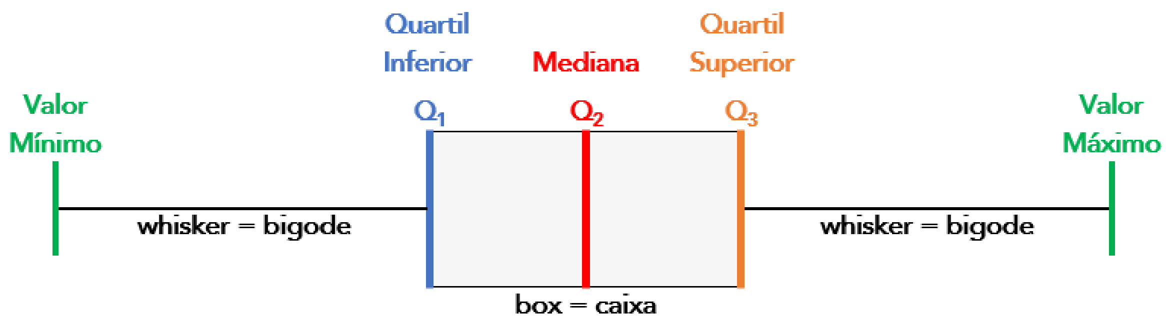
Gabarito: A.

Box Plot

Um **boxplot** (também chamado de **box-and-whisker plot**) é uma ferramenta gráfica frequentemente utilizada na análise exploratória de dados que permite visualizar a distribuição dos dados e os valores discrepantes (outliers), assim como a distância dos valores extremos em relação à maioria dos dados. Essa ferramenta **resume cinco medidas descritivas** de um conjunto de dados, incluindo: **o valor mínimo, o primeiro quartil, a mediana, o terceiro quartil e o valor máximo**.

Para construir um gráfico de *boxplot*, usamos uma haste horizontal ou vertical e uma caixa retangular (*box*). O local em que a haste começa (da esquerda para a direita) indica o valor mínimo e o ponto em que a haste termina indica o valor máximo.

A caixa retangular, localizada **no meio da haste**, em geral, **possui três linhas**. **A primeira linha, na extremidade esquerda da caixa, indica o primeiro quartil**. **A terceira linha, na extremidade direita, indica o terceiro quartil**. A **linha do meio, no interior da caixa, indica o segundo quartil ou a mediana**. O segundo quartil pode estar entre o primeiro e o terceiro quartis, ou pode coincidir com um, ou outro, ou ambos.

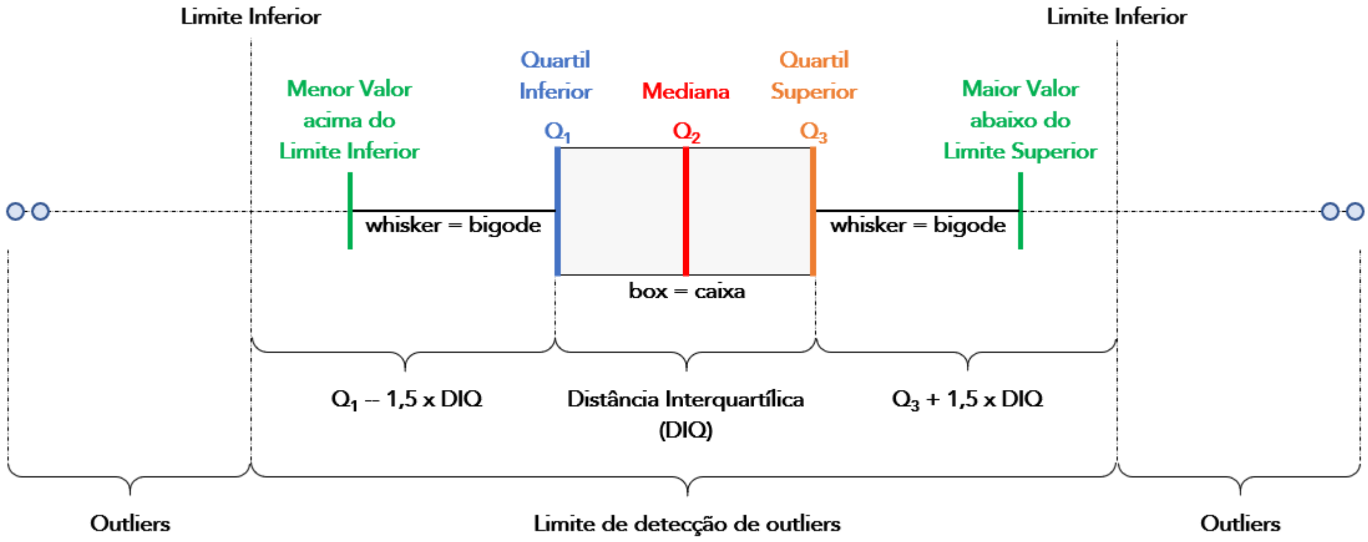


Além disso, há dois traços, chamados de ***whiskers*** (ou bigodes), ligando o valor mínimo à extremidade esquerda da caixa e o valor máximo à extremidade direita da caixa. **Cada um desses traços comporta, aproximadamente, 25% dos dados. O restante, cerca de 50%, está distribuído no interior da caixa.**

Também **podemos encontrar gráficos de *box plot* com pontos ou asteriscos marcando valores discrepantes (*outliers*)**. Nesses casos, os *whiskers* não se estendem aos valores mínimo e máximo do conjunto de dados, mas ficam limitados a um comprimento máximo de $1{,}5 \times DIQ$, em que $\boldsymbol{DIQ}$ **é a distância interquartílica**.

A **distância interquartílica** (ou amplitude interquartílica, ou intervalo interquartílico) é calculada pela fórmula:

$$\boxed{\boldsymbol{DIQ = Q_3 - Q_1}}$$



Dessa forma, **valores menores que $Q_1 - 1,5 \times DIQ$ ou maiores que $Q_3 + 1,5 \times DIQ$ são considerados VALORES DISCREPANTES (*OUTLIERS*) e representados por PONTOS ou ASTERISCOS**.

FIQUE ATENTO!

É importante salientarmos que a fórmula da **distância interquartílica** se parece muito com a do **desvio quartílico (ou amplitude semi-interquartílica)**, podendo ser facilmente confundida. O **desvio quartílico** é calculado pela expressão apresentada a seguir:

$$\boxed{D_q = \frac{Q_3 - Q_1}{2}}$$

Essa medida será abordada de forma mais detalhada na aula de **medidas de dispersão**.

EXEMPLIFICANDO

Vamos ver como representar um conjunto de dados em forma de *box plot*. Considere a seguinte série, que elenca os pesos de um grupo de **40 alunos** do Estratégia Concursos:

{58, 65, 66, 67, 68, 68, 68, 69, 69, 69, 69, 70, 70, 70, 70, 70, 70, 70, 70, 71, 71, 72, 72, 73, 73, 74, 75, 75, 75, 75, 75, 76, 76, 77, 77, 78, 79, 79, 80, 87}.

Primeiro, vamos identificar as medidas descritivas desse conjunto:

a) valor mínimo:

$$mín = 58$$

b) primeiro quartil:

Determinando a posição de Q_1:

$$P_{Q_1} = \frac{1 \times 40}{4} = 10$$

Em seguida, verificamos o valor que está ocupando essa posição

$$Q_1 = 69$$

c) mediana:

Como o conjunto é formado por um número par de elementos, o segundo quartil (mediana) será calculado como a média aritmética dos termos centrais. Os termos centrais ocupam as posições 20 e 21, então, calculando a média aritmética, descobrimos que o segundo quartil vale:

$$Q_2 = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{40}{2}} + x_{\frac{40}{2}+1}}{2} = \frac{x_{20} + x_{21}}{2} = \frac{71 + 71}{2} = 71$$

d) terceiro quartil:

Determinando a posição de Q_3:

$$P_{Q_3} = \frac{3 \times 40}{4} = 30$$

Em seguida, verificamos o valor que está ocupando essa posição

$$Q_3 = 75$$

e) valor máximo:

$$máx = 87$$

f) distância interquartílica:

$$DIQ = Q_3 - Q_1 = 75 - 69 = 6$$

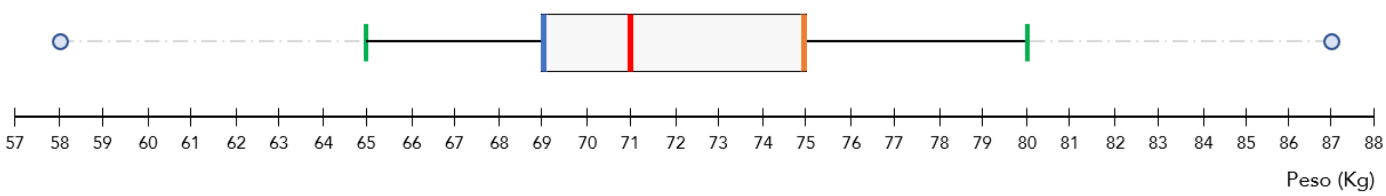
g) limite inferior para detecção de outliers:

$$Q_1 - 1{,}5 \times DIQ = 69 - 1{,}5 \times 6 = 60$$

h) limite superior para detecção de outliers:

$$Q_3 + 1{,}5 \times DIQ = 75 + 1{,}5 \times 6 = 82$$

Agora, podemos reunir essas informações em um gráfico de *box plot.* Reparem que o valor mínimo está localizado em 65, porque esse é o **MENOR VALOR do conjunto de dados IGUAL OU SUPERIOR ao limite inferior para detecção de *outliers*** (60). De igual modo, notem que o valor máximo está localizado em 80, porque esse é o **MAIOR VALOR do conjunto de dados IGUAL OU INFERIOR ao limite superior para detecção de outliers** (82). Vejam também que os valores abaixo de 60 e acima de 82 são considerados discrepantes, sendo sinalizados por pontos.



Agora, vamos analisar uma segunda série de dados, que representa os pesos de um grupo de **40 alunas** do Estratégia Concursos.

{59, 60, 61, 62, 62, 63, 64, 64, 64, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 66, 66, 66, 66, 66, 66, 66, 67, 67, 67, 67, 67, 67, 67, 68, 68, 68, 68, 68, 69, 69}

Vamos identificar as medidas descritivas desse conjunto:

a) valor mínimo:

$$mín = 59$$

b) primeiro quartil:

Determinando a posição de Q_1:

$$P_{Q_1} = \frac{1 \times 40}{4} = 10$$

Em seguida, verificamos o valor que está ocupando essa posição

$$Q_1 = 65$$

c) mediana:

Como o conjunto é formado por um número par de elementos, o segundo quartil (mediana) será calculado como a média aritmética dos termos centrais. Os termos centrais ocupam as posições 20 e 21, então, calculando a média aritmética, descobrimos que o segundo quartil vale:

$$Q_2 = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{40}{2}} + x_{\frac{40}{2}+1}}{2} = \frac{x_{20} + x_{21}}{2} = \frac{66 + 66}{2} = 66$$

d) terceiro quartil:

Determinando a posição de Q_3:

$$P_{Q_3} = \frac{3 \times 40}{4} = 30$$

Em seguida, verificamos o valor que está ocupando essa posição

$$Q_3 = 67$$

e) valor máximo:

$$máx = 69$$

f) distância interquartílica:

$$DIQ = Q_3 - Q_1 = 67 - 65 = 2$$

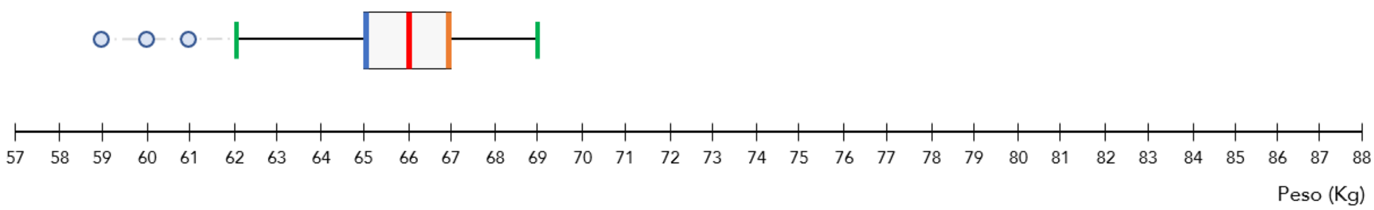
g) limite inferior para detecção de outliers:

$$Q_1 - 1{,}5 \times DIQ = 65 - 1{,}5 \times 2 = 62$$

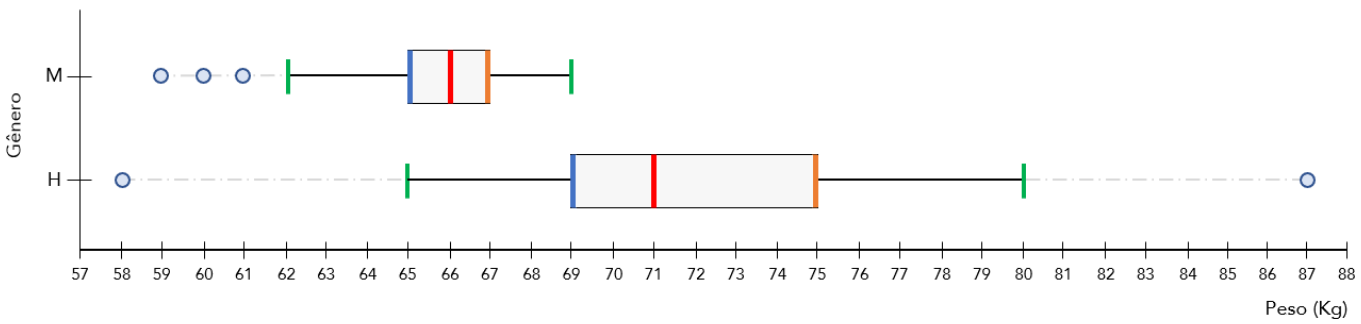
h) limite superior para detecção de outliers:

$$Q_3 + 1{,}5 \times DIQ = 67 + 1{,}5 \times 2 = 70$$

Nesse ponto, podemos reunir essas informações em um segundo gráfico de *box plot.* Reparem que o valor mínimo está localizado em 62, porque esse é o **MENOR VALOR do conjunto de dados IGUAL OU SUPERIOR ao limite inferior para detecção de *outliers*** (62). De igual modo, notem que o valor máximo está localizado em 69, porque esse é o **MAIOR VALOR do conjunto de dados IGUAL OU INFERIOR ao limite superior para detecção de outliers** (70). Vejam também que os valores abaixo de 62 são considerados discrepantes, sendo sinalizados por pontos.



Podemos, também, construir um único gráfico de *box plot* reunindo duas ou mais distribuições, o que torna o processo de comparação mais simples e visual.



Por fim, gostaria de chamar atenção para um detalhe importante. A depender do conjunto de dados, teremos valores iguais para máximos, mínimos, primeiro quartil, mediana e terceiro quartil. Nesse caso, o aspecto visual do *box plot* sofrerá alterações.

Por exemplo, considere a distribuição a seguir, em que a mediana e o terceiro quartil são iguais:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	1	1	1	6	6	6	6	6	6	8

a) valor mínimo: 1;

b) primeiro quartil:

$$Posição: 1 \times (11/4) = 2{,}75 \longrightarrow \ Valor\ mais\ próximo: x_3 = 1$$

c) segundo quartil (mediana):

Como o número de elementos do conjunto é ímpar, o segundo quartil (mediana) coincidirá com o termo central da série. Assim, calculando a posição da mediana teremos:

$$Q_2 = x_{\frac{n+1}{2}} = x_{\frac{11+1}{2}} = x_{\frac{12}{2}} = x_6$$

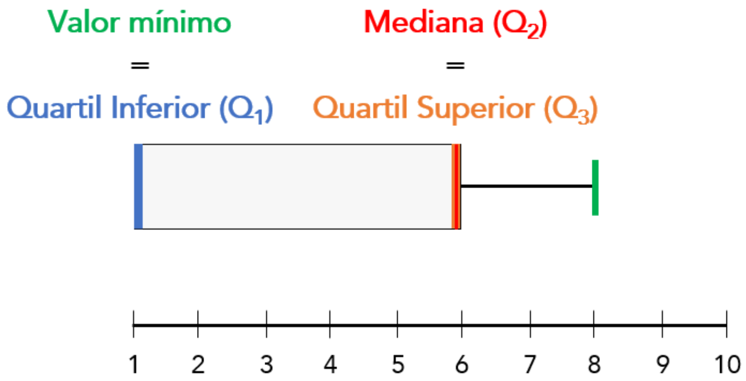
Portanto, o segundo quartil (mediana) é igual a $x_6 = 6$

d) terceiro quartil:

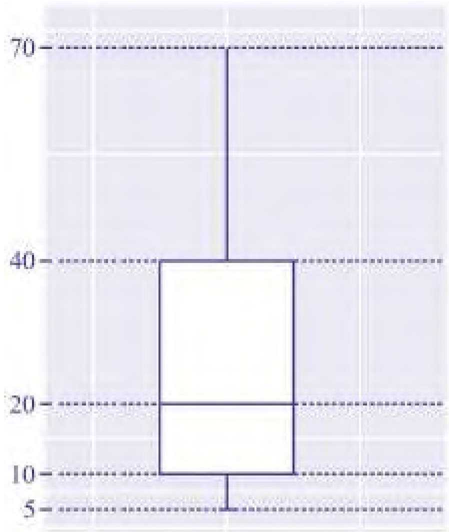
$$Posição: 3 \ \times \ (11/4) = 8{,}25 \longrightarrow Valor\ mais\ próximo: x_8 = 6$$

e) valor máximo: 8

Nessa situação, a caixa retangular não exibirá três linhas em seu interior, pois a linha da mediana irá se sobrepor a do quartil superior. Além disso, também não haverá um traço ligando o valor mínimo ao quartil inferior, vez que ambos terão a mesma magnitude. O gráfico ficará assim:


Valor mínimo
=
Quartil Inferior (Q1)
Mediana (Q2)
=
Quartil Superior (Q3)
1
2
3
4
5
6
7
8
9
10

(CESPE/TJ-PA/2020)



Considerando que o desenho esquemático (boxplot) antecedente se refere a uma variável quantitativa X, assinale a opção correta.

a) O intervalo interquartil é igual a 65.

b) Metade da distribuição da variável X se encontra entre os valores 20 e 40.

c) Os valores da variável X que se encontram no intervalo [5;10] representam 5% da distribuição de X.

d) A mediana de X é igual a 25.

e) O primeiro quartil da distribuição de X é igual a 10.

Comentários:

Analisando cada uma das alternativas:

- Letra A: **Errada**. O intervalo interquartil vale $DIQ = Q_3 - Q_1 = 40 - 10 = 30$.
- Letra B: **Errada**. O diagrama mostra que a mediana ocorre em 20, logo, a metade da distribuição está entre 20 e 70.
- Letra C: **Errada**. O diagrama é distribuído em 4 quartis, 4 partes com a mesma quantidade de valores, cada quartil representando 25% das observações, sendo o primeiro quartil representado pelo intervalo [5;10].
- Letra D: **Errada**. O diagrama mostra que a mediana vale 20.
- Letra E: **Correta**. O diagrama mostra que o primeiro quartil vale 10, ocupando o intervalo [5;10].

Gabarito: E.

(CESPE/IPHAN/2018) Julgue o item subsequente, referente à análise exploratória de dados.

O BOXPLOT representa os dados em um retângulo construído com o primeiro e o segundo quartil, fornecendo informação sobre valores médios.

Comentários:

O boxplot oferece informações sobre os quartis (1.º, 2.º e 3.º), o limite mínimo, o limite máximo e os outliers (valores discrepantes). O 2.º quartil coincide com o valor da mediana. Portanto, a questão erra ao afirmar que as informações são sobre os valores médios.

Gabarito: Errado.

(FGV/MPE-BA/2017) Em uma amostra desconfia-se de que três valores sejam, na verdade, " outliers" e que deveriam ser descartados. Para tal avaliação o estatístico dispõe apenas dos valores dos 1º e 3º quartil da distribuição. Os números são os seguintes:

$\boldsymbol{Q_1(X) = 21}$, $\boldsymbol{Q_3(X) = 33}$, $\boldsymbol{X_1 = 58}$, $\boldsymbol{X_2 = 2}$ **e** $\boldsymbol{X_3 = 43}$

Onde $\boldsymbol{Q_{is}}$ **são os quartis e os** $\boldsymbol{X_{is}}$ **os valores em análise.**

Assim, é correto afirmar que:

a) Todos os valores são "outliers";

b) Os valores X_1 e X_3 são "outliers";

c) Nenhum dos valores é "outliers";

d) Apenas o valor X_2 é "outlier";

e) Os valores X_1 e X_2 são "outliers".

Comentários:

Para resolvermos a questão, precisamos calcular os limites inferior e superior da amostra. Assim:

$$l_{inf} = Q_1 - 1{,}5 \times (Q_3 - Q_1)$$

$$l_{inf} = 21 - 1{,}5 \times (33 - 21)$$

$$l_{inf} = 21 - 18 = 3$$

$$l_{sup} = Q_3 + 1{,}5 \times (Q_3 - Q_1)$$

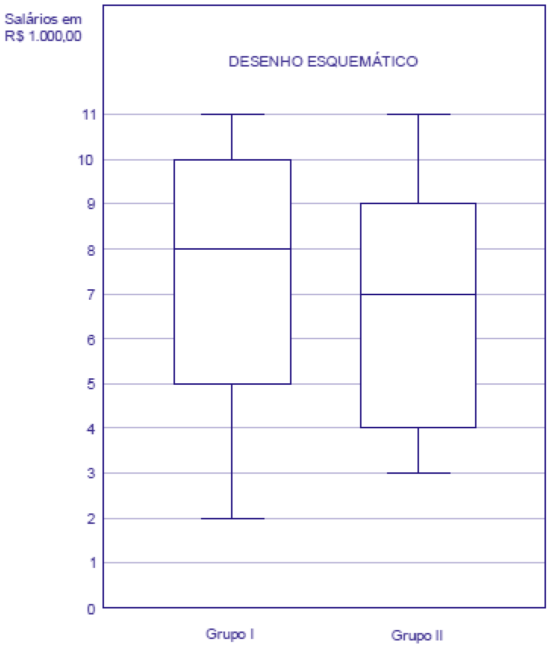
$$l_{sup} = 33 + 1{,}5 \times (33 - 21)$$

$$l_{sup} = 33 + 18 = 51$$

Como $2 < 3$ e $58 > 51$, então podemos afirmar que os valores X_1 e X_2 são *outliers*.

Gabarito: E.

(FCC/TRT 3ª Região/2015) Seja uma representação gráfica de dados de acordo com o desenho esquemático abaixo (box-plot) que foi preparado para comparar todos os salários dos funcionários do sexo masculino (Grupo I) com todos os salários dos funcionários do sexo feminino (Grupo II) lotados em um órgão público.



Neste desenho esquemático

a) O número de elementos do Grupo I é superior ao número de elementos do Grupo II.

b) O módulo da diferença entre as medianas dos 2 grupos é igual a 25% do menor salário deste órgão público.

c) Mais da metade dos elementos do Grupo I possui um salário inferior a R$ 5.000,00 ou superior a R$ 10.000,00.

d) O valor do menor salário do Grupo II corresponde a 37,5% do valor da mediana do Grupo I.

e) A diferença interquartil do Grupo I é superior à diferença interquartil do Grupo II.

Comentários:

Vamos analisar cada uma das alternativas:

- Letra A: **Errada**. O diagrama não mostra a quantidade de elementos de cada grupo, portanto, não podemos afirmar que o Grupo I tem mais elementos que o Grupo II.

- Letra B: **Errada**. O diagrama mostra que as medianas valem 8 e 7 para cada grupo, portanto, a diferença entre elas é igual a 1. Podemos observar também que o menor salário é 2. Logo, a diferença entre as medianas corresponde a 50% do menor salário.

- Letra C: **Errada**. Se separarmos o Grupo I em quartis, cada quartil corresponderá a 25% do total de elementos. Somando o último e o primeiro quartil teremos 50%, exatamente a metade, dos elementos. Portanto, não podemos afirmar que mais da metade dos elementos do Grupo I possui um salário inferior a R$ 5.000,00 ou superior a R$ 10.000,00.

- Letra D: **Correta**. O menor salário do Grupo II vale 3. A mediana do Grupo I vale 8. Então, $\frac{3}{8} = 0{,}375 = 37{,}5\%$.

- Letra E: **Errada**. Para o primeiro grupo $Q_3 = 10,\ Q_1 = 5 \rightarrow Q_3 - Q_1 = 10 - 5 = 5$ para o segundo grupo $Q_3 = 9\,, Q_1 = 4 \rightarrow Q_3 - Q_1 = 9 - 4 = 5$. Portanto, as distâncias são iguais.

Gabarito: D.

RESUMO DA AULA

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	QUARTIS	QUINTIS
Divide uma série ordenada em duas partes iguais, cada uma contendo 50% dos valores da sequência	Dividem uma série ordenada em quatro partes iguais, cada uma contendo 25% dos valores da sequência.	Dividem uma série ordenada em cinco partes iguais, cada uma contendo 20% dos valores da sequência

DECIS	PERCENTIS
Dividem uma série ordenada em dez partes iguais, cada uma contendo 10% dos valores da sequência.	Dividem uma série ordenada em cem partes iguais, cada uma contendo 1% dos valores da sequência.

MEDIANA

A mediana é, simultaneamente, uma **MEDIDA DE POSIÇÃO, de TENDÊNCIA CENTRAL e SEPARATRIZ**. Ela separa uma série de valores em **duas** partes de **tamanhos iguais**, cada uma contendo o **mesmo número de elementos**. Sendo representada pelos símbolos M_d ou, em menor frequência, $\widetilde{x}$.

Mediana para dados não-agrupados.

A **mediana** é o elemento que ocupa a **POSIÇÃO CENTRAL** de uma série de observações ordenada segundo suas grandezas. Para estabelecer que a mediana de um conjunto composto por n elementos ordenados de forma crescente ou decrescente temos:

a) se n for ímpar: $\boldsymbol{M_d = x_{\frac{n+1}{2}}}$

b) se n for par: $\boldsymbol{M_d = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}}$

Mediana para dados agrupados sem intervalos de classe

A **mediana** é um valor que dividirá a distribuição de frequências em **duas partes** contendo o **mesmo número de elementos**. Para estabelecer que a mediana de uma tabela de frequências composta por um total de n elementos:

a) se n for ímpar: $\boldsymbol{M_d = x_{\frac{n+1}{2}}}$

b) se n for par: $\boldsymbol{M_d = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}}$

Mediana para dados agrupados em classes

I - Para calcular a mediana de dados que estão agrupados por intervalo de classes, identificamos a **classe mediana**, **que corresponde à frequência acumulada imediatamente igual ou superior à metade da frequência total, ou seja, metade da soma das frequências simples, $\boldsymbol{\sum f_i/2}$.**

II - Conhecendo a classe mediana, podemos aplicar a fórmula da mediana:

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

III - **método de interpolação linear** consiste em utilizar valores conhecidos para estimar valores desconhecidos de forma linear, isto é, por meio de uma reta.

Propriedades da Mediana

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.**

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.**

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.**

QUARTIL, DECIL E PERCENTIL

O nome da medida separatriz é atribuído de acordo com a quantidade de partes em que a série é dividida:

Quartis

QUARTIS

Valores de uma série que a dividem em QUATRO PARTES IGUAIS, isto é, quatro partes contendo o mesmo número de elementos (25%).

Q_1: o primeiro quartil corresponde à separação dos primeiros 25% de elementos da série.

Q_2: o segundo quartil corresponde à separação de metade dos elementos da série, coincidindo com a mediana ($Q_2 = M_d$)

Q_3: o terceiro quartil corresponde à separação dos primeiros 75% de elementos da série, ou dos últimos 25% de elementos da série

Quartil para dados não-agrupados

O cálculo do **quartil para dados não-agrupados** será realizado por meio de três etapas:

- 1.ª etapa: determinar a posição do quartil, por meio da expressão:

$$P_{Q_k} = \frac{k \times n}{4} \quad (k = 1, 2, 3);$$

- 2.ª etapa: identificar a posição mais próxima do rol;
- 3.ª etapa: verificar o valor que está ocupando essa posição.

Quartil para dados agrupados sem intervalos de classe

O cálculo do **quartil para dados agrupados sem intervalos de classe** será realizado por meio de três etapas:

- 1.ª etapa: determinar a **posição do quartil**, por meio da expressão:

$$P_{Q_k} = \frac{k \times \sum fi}{4} \quad (k = 1, 2, 3)$$

- 2.ª etapa: identificar a posição do **quartil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do quartil**;
- 3.ª etapa: verificar o valor da variável que corresponde a essa posição.

Quartil para dados agrupados em classes

O cálculo do **quartil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinar a **posição do quartil**, por meio da expressão:

$$P_{Q_k} = \frac{k \times \sum fi}{4} \quad (k = 1, 2, 3)$$

- 2.ª etapa: identificar a posição do **quartil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do quartil**;
- 3.ª etapa: verificar as informações referentes à classe correspondente a essa posição; e
- 4.ª etapa: calcular o valor do **quartil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes**:

$$Q_k = l_{inf_{Q_k}} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac_{ant}}}{f_{Q_k}} \right] \times h_{Q_k}$$

Decis

DECIS

Valores de uma série que a dividem em DEZ PARTES IGUAIS, isto é, dez partes contendo o mesmo número de elementos (10%).

D_1: o primeiro decil corresponde à separação dos primeiros 10% de elementos da série;

D_5: o quinto decil corresponde à separação de metade dos elementos da série, coincidindo com a mediana ($D_5 = M_d$);

D_9: o nono decil corresponde à separação dos primeiros 90% de elementos da série, ou dos últimos 10% de elementos da série.

Decil para dados não-agrupados

O cálculo do **decil para dados não-agrupados** será realizado por meio de três etapas:

- 1.ª etapa: determinar a posição do decil, por meio da expressão:

$$P_{D_k} = \frac{k \times n}{10} \quad (k = 1, 2, \cdots, 9);$$

- 2.ª etapa: identificar a posição mais próxima do rol;
- 3.ª etapa: verificar o valor que está ocupando essa posição.

Decil para dados agrupados sem intervalos de classe.

O cálculo do **decil para dados agrupados sem intervalos de classe** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinamos a **posição do decil**, por meio da expressão:

$$P_{D_k} = \frac{k \times \sum fi}{10} \quad (k = 1, 2, \cdots, 9)$$

- 2.ª etapa: identificamos a posição do **decil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do decil**;
- 3.ª etapa: verificamos o valor da variável que corresponde a essa posição.

Decil para dados agrupados em classes

O cálculo do **decil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinamos a **posição do decil**, por meio da expressão:

$$P_{D_k} = \frac{k \times \sum fi}{10} \quad (k = 1, 2, \cdots, 9)$$

- 2.ª etapa: identificamos a posição do **decil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do decil**;
- 3.ª etapa: verificamos as informações referentes à classe correspondente a essa posição; e
- 4.ª etapa: calculamos o valor do **decil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes:**

$$D_k = l_{inf_{D_k}} + \left[\frac{\frac{k \times \sum f_i}{10} - f_{ac_{ant}}}{f_{D_k}}\right] \times h_{D_k}$$

Percentis

PERCENTIS

Valores de uma série que a dividem em CEM PARTES IGUAIS, isto é, cem partes contendo o mesmo número de elementos (1%)

P_1: o primeiro percentil corresponde à separação do primeiro 1% de elementos da série;

P_{50}: o quinquagésimo percentil corresponde à separação de metade dos elementos da série, coincidindo com a mediana ($P_{50} = M_d$);

P_{99}: o nonagésimo nono percentil corresponde à separação dos primeiros 99% de elementos da série, ou do último 1% de elementos da série.

Percentil para dados não-agrupados

O cálculo do **Percentil para dados não-agrupados** será realizado por meio de três etapas:

- 1.ª etapa: determinar a posição do percentil, por meio da expressão:

$$P_{P_k} = \frac{k \times n}{100} (k = 1, 2, \cdots, 99);$$

- 2.ª etapa: identificar a posição mais próxima do rol;
- 3.ª etapa: verificar o valor que está ocupando essa posição.

Percentil para dados agrupados sem intervalos de classe

O cálculo do **percentil para dados agrupados sem intervalos de classe** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinar a **posição do percentil**, por meio da expressão:

$$P_{P_k} = \frac{k \times \sum fi}{100} \quad (k = 1, 2, \cdots, 99)$$

- 2.ª etapa: identificar a posição do **percentil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do percentil**;
- 3.ª etapa: verificar o valor da variável que corresponde a essa posição.

Percentil para dados agrupados em classes

O cálculo do **percentil para dados agrupados em classes** será realizado por meio das seguintes etapas:

- 1.ª etapa: determinar a **posição do percentil**, por meio da expressão:

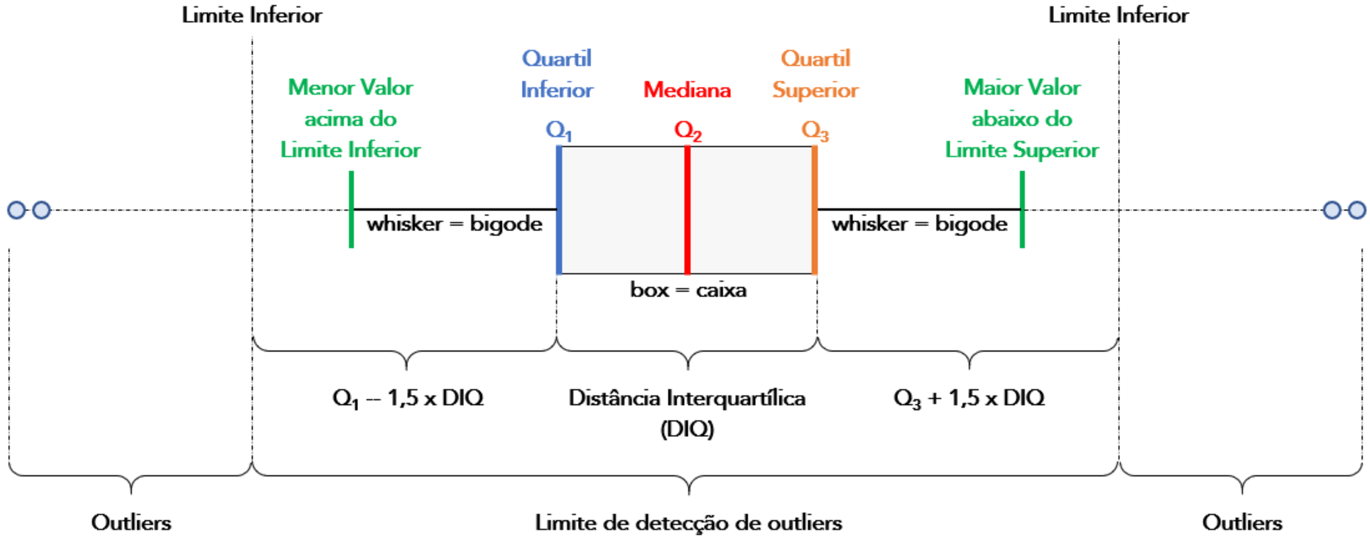
$$P_{P_k} = \frac{k \times \sum fi}{100} \quad (k = 1, 2, \cdots, 99)$$

- 2.ª etapa: identificar a posição do **percentil** na coluna de frequências acumuladas, isto é, **a frequência acumulada imediatamente igual ou superior à posição do percentil**;
- 3.ª etapa: verificar as informações referentes à classe correspondente a essa posição; e
- 4.ª etapa: calculamos o valor do **percentil** por meio da fórmula apresentada a seguir, que **consiste em uma variação da fórmula da mediana para dados agrupados em classes**:

$$P_k = l_{inf_{P_k}} + \left[\frac{\frac{k \times \sum f_i}{100} - f_{ac_{ant}}}{f_{P_k}} \right] \times h_{P_k}$$

BOX PLOT

Um **boxplot** ou **box-and-whisker plot** é uma ferramenta gráfica frequentemente utilizada na análise exploratória de dados que permite visualizar a distribuição dos dados e os valores discrepantes (outliers). Essa ferramenta **resume cinco medidas descritivas** de um conjunto de dados, incluindo: **o valor mínimo, o primeiro quartil, a mediana, o terceiro quartil e o valor máximo**.



A **distância interquartílica** (ou amplitude interquartílica, ou intervalo interquartílico) é calculada pela fórmula:

$$DIQ = Q_3 - Q_1$$

QUESTÕES COMENTADAS

Mediana

1. (CESPE/PETROBRAS/2022)

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 obtida com base em uma amostra aleatória simples de tamanho igual a n, julgue o item que se segue.

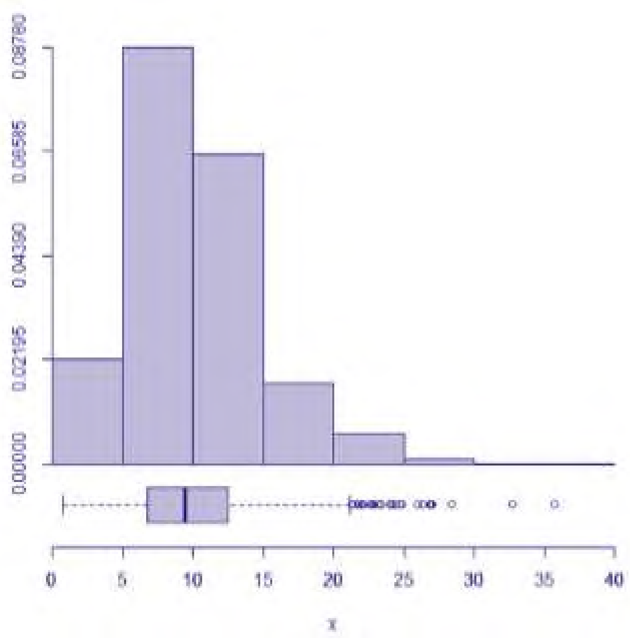
A mediana de x é igual a 1,5.

Comentários:

A mediana é o termo central da amostra, ou seja, em um conjunto de dados ordenados, a mediana denota a posição que corresponde a 50% da amostra. Observando a tabela, percebemos que o valor 2 possui uma frequência relativa de 0,50, o que corresponde a 50% dos dados. Logo, podemos concluir que a mediana dos dados apresentados na tabela é 2.

Gabarito: Errado.

2. (CESPE/PETROBRAS/2022) A figura seguinte mostra o histograma como uma estimativa da função de densidade de uma distribuição X, juntamente com o diagrama boxplot correspondente a esse conjunto de dados.



Tamanho da amostra	1.000
Média amostral	10
Desvio padrão amostral	4,7

Considerando a figura e as informações apresentadas no quadro, julgue o item que se segue.

A diferença entre a média amostral e a mediana amostral é superior a 0.

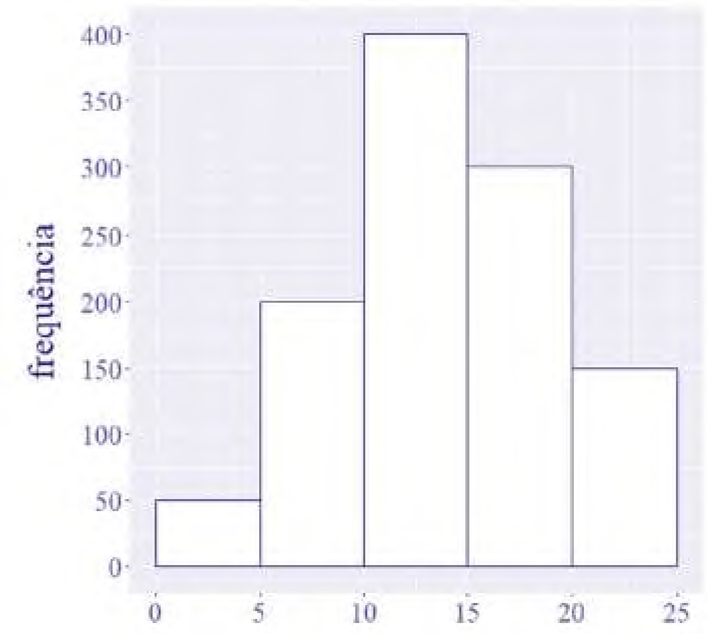
Comentários:

Observando os dados do histograma, podemos concluir que se trata de uma distribuição assimétrica à direita. Nesse caso, o valor da média é maior que o da mediana que, por sua vez, é maior que o da moda.

Conforme a tabela apresentada na questão, a média vale 10. Como a média vale 10, a mediana deverá assumir um valor positivo e inferior a 10. Logo, a diferença entre a média e a mediana será maior que 0.

Gabarito: Certo.

3. (CESPE/TELEBRAS/2022)



Considerando que o histograma apresentado descreve a distribuição de uma variável quantitativa X por meio de frequências absolutas, julgue o item que se segue.

A mediana de X se encontra na classe modal.

Comentários:

A classe modal é a que apresenta a maior frequência, isto é, a terceira classe, representada pelo intervalo que varia de 10 a 15. Agora, vamos encontrar a posição da mediana. Transcrevendo os dados para uma tabela e acrescentando as frequências acumuladas, temos:

Classes (x)	Frequências	Freq. acumulada
$0 \vdash 5$	50	50
$5 \vdash 10$	200	250
$10 \vdash 15$	400	650
$15 \vdash 20$	300	950
$20 \vdash 25$	150	1100
Total	**1100**	

A classe mediana corresponde à classe com frequência acumulada imediatamente igual ou superior à metade da frequência total:

$$\frac{\sum f_i}{2} = \frac{1100}{2} = 550$$

Assim, concluímos que a classe mediana está na posição 550, que corresponde à terceira classe (de 10 a 15).

Logo, a mediana se encontra na classe modal.

Gabarito: Certo.

4. (FGV/CBM-AM/2022) A soma de 11 números inteiros estritamente positivos, não necessariamente distintos, é 2022.

O maior valor que a mediana desses 11 números pode ter é

a) 335.

b) 336.

c) 337.

d) 338.

e) 339.

Comentários:

A mediana é o termo central da amostra. Ela divide o conjunto de dados em duas partes, separando os valores inferiores à mediana dos valores superiores à mediana. Então, se a mediana é o termo central da amostra, em um conjunto com 11 termos, ela será o 6º termo do conjunto de dados ordenados.

Então, se a soma desses 11 números é igual a 2022, para encontramos o valor da maior mediana basta considerarmos que os cinco primeiros números são iguais a 1, menor valor inteiro e positivo inferior à mediana. Vamos esquematizar para melhor compreensão:

$$\underbrace{1+1+1+1+1}_{valores\ inferiores} + \underbrace{M_d}_{mediana} + \underbrace{x_7 + x_8 + x_9 + x_{10} + x_{11}}_{valores\ superiores} = 2022$$

Como não conhecemos o valor da mediana, e como também não sabemos quais são os valores dos demais números, vamos considerar que todos possuem o mesmo valor, igual a x. Agora, basta resolvermos a equação:

$$5 + x + x + x + x + x + x = 2022$$

$$6x = 2022 - 5$$

$$x = \frac{2017}{6}$$

Vejam que essa divisão não é exata, portanto, restará uma unidade. Logo, temos que:

$$x = 336$$

Agora, podemos atribuir a sobra da divisão ao último termo da amostra, assim:

$$\underbrace{1+1+1+1+1}_{valores\ inferiores} + \underbrace{336}_{mediana} + \underbrace{336+336+336+336+337}_{valores\ superiores} = 2022$$

Logo, o maior valor da mediana é 336.

Gabarito: B.

5. (FGV/SEFAZ ES/2022) As notas de nove candidatos num certo exame foram:

54, 48, 46, 51, 38, 50, 44, 58, 32.

A mediana dessas notas é igual a

a) 44.

b) 46.

c) 48.

d) 50.

e) 51.

Comentários:

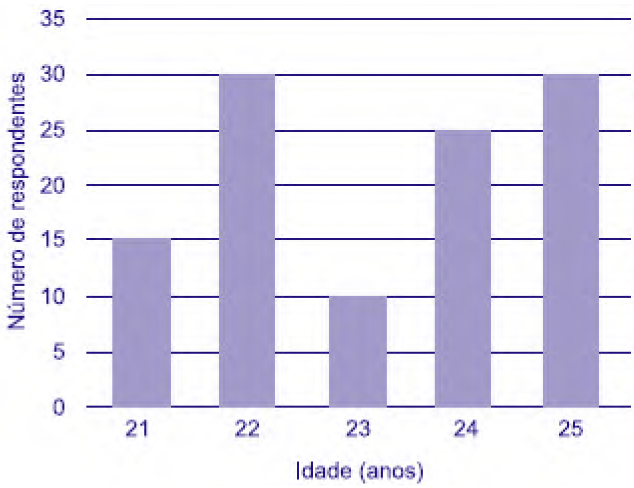
A mediana divide uma série ordenada em duas partes iguais. Como temos 9 termos na amostra, a mediana é determinada pelo termo central, isto é, o quinto termo da amostra. É importante observar que, para encontrarmos a mediana, os dados devem estar ordenados. Assim, temos:

$$32, 38, 44, 46, \underbrace{\mathbf{48}}_{\boldsymbol{termo\ central}}, 50, 51, 54, 58$$

Portanto, a mediana é 48.

Gabarito: C.

6. (VUNESP/PM-SP/2022) Uma pesquisa foi realizada com um grupo de pessoas cujas idades, em anos, pertencem ao conjunto {21, 22, 23, 24, 25, 26}. O gráfico registra as frequências absolutas dos entrevistados com menos de 26 anos.



Sabendo que a mediana das idades do conjunto completo de dados (incluindo as pessoas com 26 anos) é igual a 24 anos, o número máximo de pessoas com 26 anos que participaram da pesquisa foi

a) 19.

b) 25.

c) 35.

d) 49.

e) 55.

Comentários:

A mediana é o termo central da amostra, ela divide uma série ordenada de dados em duas partes iguais, separando os valores inferiores à mediana dos valores superiores à mediana.

O gráfico nos apresenta as quantidades de pessoas e suas respectivas idades, com exceção das pessoas com 26 anos. O enunciado também nos informa que a mediana das idades é 24. Ora, se a mediana divide a série em duas partes, basta sabermos quantas pessoas estão concentradas na parte inferior à mediana para descobrirmos quantas pessoas estarão do lado superior. Assim:

$$15 + 30 + 10 + 25 = 80$$

Logo, temos 80 pessoas com idades inferiores ou iguais à mediana. Nesse ponto, devemos lembrar de subtrair a posição ocupada pela mediana. Dessa forma, teremos 79 pessoas com idades inferiores ou iguais mediana.

A metade superior deverá conter a mesma quantidade de pessoas da metade inferior, isto é, 79 pessoas. Então, se temos 30 pessoas com 25 anos, podemos calcular quantas pessoas teremos com 26 anos:

$$30 + x = 79$$

$$x = 79 - 30$$

$$x = 49$$

Portanto, teremos 49 pessoas com 26 anos.

Gabarito: D.

7. (AOCP/PC PA/2021) 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) $P(X < 1{,}70) = P(X > 1{,}70) = 50\%$.

b) $P(X = 1{,}70) = 50\%$.

c) $P(X > 1{,}70) = 75\%$.

d) $P(X < 1{,}70) = 25\%$.

e) $P(X < 1{,}70) = P(X > 1{,}70) = 25\%$.

Comentários:

A mediana divide uma série ordenada em duas partes iguais, cada uma contendo 50% dos valores da sequência. Logo, a alternativa que melhor representa essa relação é:

$$P(X < 1{,}70) = P(X > 1{,}70) = 50\%.$$

Gabarito: A.

8. (CESGRANRIO/CEF/2021) Seis candidatos, aprovados para a penúltima etapa de um processo seletivo, foram submetidos a um teste de conhecimentos gerais com 10 itens do tipo "verdadeiro/falso". Os dois primeiros candidatos acertaram 8 itens cada, o terceiro acertou 9, o quarto acertou 7, e os dois últimos, 5 cada. Pelas regras do concurso, passariam, para a etapa final da seleção, os candidatos cujo número de acertos fosse maior ou igual à mediana do número de acertos dos seis participantes.

Quantos candidatos passaram para a etapa final?

a) 2

b) 3

c) 4

d) 5

e) 6

Comentários:

Primeiro vamos encontrar o valor da mediana das notas obtidas pelos candidatos. Para isso, temos que ordenar os valores das notas:

$$5, 5, \underbrace{7, 8}_{termos\ centrais}, 8, 9$$

Como o número de termos é par, a mediana será dada pela média dos dois termos centrais:

$$M_d = \frac{7+8}{2} = \frac{15}{2} = 7{,}5$$

Logo, vão para a etapa final os candidatos com notas maiores ou iguais a 7,5.

Dessa forma, passaram apenas 3 candidatos: os dois primeiros com nota 8, e o terceiro com nota 9.

Gabarito: B.

9. (CESPE/PG-DF/2021)

Estatística	X, em R$ milhões
Mínimo	0,5
Primeiro quartil (Q_1)	1
Segundo quartil (Q_2)	2
Terceiro quartil (Q_3)	5
Máximo	20

O quadro apresentado mostra estatísticas descritivas produzidas por um estudo acerca de despesas públicas (X, em R$ milhões) ocorridos no ano de 2019 em uma amostra aleatória simples de 100 contratos.

Com base nessas informações, julgue o item que se segue.

A mediana da variável *X* foi igual a R$ 2 milhões.

Comentários:

Analisando a tabela apresentada na questão, verificamos que o segundo quartil (Q_2) foi igual a 2 milhões. Como a mediana é equivalente ao segundo quartil, podemos afirmar que a mediana de x também foi igual a 2 milhões.

Gabarito: Certo.

10. (CESPE/CBM AL/2021) Em determinado dia, em uma região atendida por uma unidade do corpo de bombeiros, ocorreram 16 acidentes, que resultaram em 48 vítimas, socorridas pelos

bombeiros nos próprios locais de acidente. Entre essas vítimas, 4 vieram a óbito no momento do atendimento, e as demais sobreviveram.

Com base nessa situação hipotética, julgue o item a seguir.

Suponha que as idades das vítimas que vieram a óbito sejam 12, 50, 30 e 20 anos de idade. Nesse caso, a mediana das idades é maior que 26 anos.

Comentários:

A mediana divide uma série ordenada em duas partes de tamanhos iguais. Como o número de termos da amostra em questão é par, a mediana será dada pela média aritmética dos dois termos centrais:

$$12 - \underbrace{20 - 30}_{termos\ centrais} - 50$$

$$Md = \frac{30 + 20}{2} = \frac{50}{2} = 25$$

Gabarito: Errado.

11. (CESPE/BANESE/2021) A respeito do conjunto de dados {11, 6, 28, 51, 49, 32, 33}, julgue o item a seguir.

A mediana desse conjunto de dados é igual a 51.

Comentários:

A mediana é o termo central da amostra. Ela divide o conjunto de dados em duas partes, separando os valores inferiores à mediana dos valores superiores à mediana. Como o número de termos do conjunto é 7, a mediana será o termo na posição 4. É importante destacar que para encontrarmos a mediana de um conjunto, os dados devem estar ordenados. Então, temos:

$$6, 11, 28, \underbrace{32}_{mediana}, 33, 49, 51$$

Portanto, a mediana é 32.

Gabarito: Errado.

12. (FGV/IMBEL/2021) Considere a lista de cinco números reais: 2, 9, 4, 10, x.

Sabe-se que a mediana desses números é igual à média deles.

A soma dos possíveis valores de x é:

a) 22,5.

b) 21,25.

c) 20,75.

d) 19,5.

e) 17,5.

Comentários:

A mediana é o termo central da amostra. Ela divide os dados ordenados em duas partes iguais, de um lado teremos valores inferiores à mediana e de outro lado teremos valores superiores à mediana.

No caso apresentado, como são 5 termos, a mediana ocupará a terceira posição do conjunto de dados. Ocorre que não sabemos o valor de x, portanto, temos 5 possibilidades para o termo que representa a mediana. Ordenando os dados, temos:

$$2, 4, \underbrace{9}_{mediana}, 10, x$$

$$2, 4, \underbrace{9}_{mediana}, x, 10$$

$$2, 4, \underbrace{x}_{mediana}, 9, 10$$

$$2, x, \underbrace{4}_{mediana}, 9, 10$$

$$x, 2, \underbrace{4}_{mediana}, 9, 10$$

Dos possíveis arranjos, observamos apenas 3 opções para a mediana: 4, 9 ou x.

A questão também nos informa que o valor da mediana é igual ao valor da média. Assim, calculando a média para os dados apresentados, temos:

$$\bar{X} = \frac{2+4+9+10+x}{5} = \frac{25+x}{5}$$

Agora, considerando as possibilidades para a mediana, temos:

a) para uma mediana igual a 9:

$$9 = \frac{25+x}{5}$$

$$25 + x = 45$$

$$x = 45 - 25$$

$$x = 20$$

b) para uma mediana igual a x:

$$x = \frac{25+x}{5}$$

$$5x = 25 + x$$

$$4x = 25$$

$$x = 6{,}25$$

c) para uma mediana igual a 4:

$$4 = \frac{25 + x}{5}$$

$$25 + x = 20$$

$$x = 20 - 25$$

$$x = -5$$

Logo, temos os possíveis valores de x:

$$20;\ 6{,}25;\ -5$$

Somando os valores, temos:

$$20 + \ 6{,}25 + (-5) = 21{,}25$$

Gabarito: B.

13. (FGV/FunSaúde CE/2021) Sabe-se que x é maior do que 11 e que a diferença entre a média e a mediana dos cinco números 2, x, 11, 16, 5 é igual a 2.

O valor de x é

a) 12.

b) 16.

c) 21.

d) 26.

e) 31.

Comentários:

A mediana é termo central do conjunto de dados. Vamos ordenar os dados para encontrarmos a mediana:

$$2, 5, \underbrace{11}_{mediana}, x, 16$$

Calculando a média para os dados apresentados, temos:

$$\bar{X} = \frac{2 + 5 + 11 + x + 16}{5} = \frac{34 + x}{5}$$

Se a diferença entre a média e a mediana é igual a 2, então temos:

$$\frac{34 + x}{5} - 11 = 2$$

$$\frac{34+x}{5} = 2+11$$

$$34+x = 13 \times 5$$

$$34+x = 65$$

$$x = 65-34$$

$$x = 31$$

Portanto, x é igual a 31.

Gabarito: E.

14. (FGV/FunSaúde CE/2021) A mediana dos sete números 9, 2, 5, 3, 13, x, 5 é x.

A média desses números é

a) 5.

b) 5,5.

c) 6.

d) 6,5.

e) 7.

Comentários:

A mediana é o termo central da amostra. Ela divide o conjunto de dados em duas partes, separando os valores inferiores à mediana dos valores superiores à mediana. Se a mediana é o termo central da amostra, em um conjunto com 7 termos, a mediana será representada pelo 4º termo do conjunto de dados ordenados.

Ordenando os termos, temos:

$$2, 3, 5, \underbrace{x}_{mediana}, 5, 9, 13$$

A mediana está entre dois números iguais, no caso, 5. Logo, a mediana também deve assumir o valor 5 no conjunto. Então, teremos os seguintes dados:

$$2, 3, 5, \underbrace{5}_{mediana}, 5, 9, 13$$

Agora, basta calcularmos a média:

$$\bar{x} = \frac{2+3+5+5+5+9+13}{7} = \frac{42}{7} = 6$$

Logo, a média é igual a 6.

Gabarito: C.

15. (IBFC/IBGE/2021) Marcos pretende determinar a mediana referente aos dados brutos coletados e relacionados abaixo:

23 - 22 - 21 - 22 - 32 - 33

41 - 21 - 20 - 32 - 42 - 38

De acordo com os dados, o resultado encontrado por Marcos é igual a:

a) 37

b) 33

c) 41

d) 27,5

e) 28

Comentários:

A mediana divide a amostra ao meio, representando o valor central da amostra. Para encontrarmos a mediana, os dados devem estar em ordem crescente. Na amostra em questão, temos 12 termos, portanto, a mediana será dada pela média dos dois termos centrais, os quais ocupam as posições 6 e 7:

$$20 - 21 - 21 - 22 - 22 - \underbrace{23 - 32}_{termos\ centrais} - 32 - 33 - 38 - 41 - 42$$

Assim, temos:

$$Md = \frac{23 + 32}{2}$$

$$Md = \frac{55}{2} = 27{,}5$$

Gabarito: D.

16. (CESPE/ME/2020)

Mínimo	**5**
Moda	**9**
Mediana	**9**
Média	**10**

Máximo	15

Um levantamento amostral proporcionou as estatísticas precedentes, referentes a determinada variável quantitativa X.

Considerando essas informações e que a variável X é composta por 1240 observações, julgue o item subsequente.

A quantidade de observações da variável X maiores ou iguais a 9 é igual ou superior a 620.

Comentários:

Na tabela observamos que a mediana da amostra é igual a 9. A mediana é o termo central da amostra, que divide o conjunto de dados em duas partes, separando os valores inferiores à mediana dos valores superiores à mediana.

A questão nos diz que a variável x tem 1240 observações, então, dividindo em duas partes, temos:

$$\frac{1240}{2} = 620$$

Portanto, temos 620 observações maiores ou iguais a 9.

Gabarito: Certo.

17. (CESPE/SEFAZ DF/2020) A partir de uma amostra aleatória simples de tamanho n, sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue o item que se segue.

A mediana amostral da variável X foi igual a 2,5.

Comentários:

Em outra questão desta prova, mostramos que 80% das observações são iguais a +4 e que 20% são iguais a -1. Sabemos que a mediana divide um conjunto de dados ordenados em duas partes iguais, cada uma contendo 50% da distribuição. Como os únicos valores possíveis para a variável X são -1 e +4, a mediana necessariamente estará contida nos 80% correspondentes aos valores iguais a 4.

Gabarito: Errado.

18. (FUNDATEC/Pref. Porto Alegre/2020) Se a mediana de um determinado processo for igual a 7, isso quer dizer que:

a) O 7º valor da amostra representará a mediana.

b) O 7º valor da amostra ordenada representará a mediana.

c) A posição mediana é 7.

d) 50% dos valores da amostra são iguais a 7.

e) 50% dos valores da amostra são menores ou iguais a 7.

Comentários:

A mediana é o termo central da amostra, ou seja, em um conjunto de dados ordenados, a mediana denota a posição que corresponde a 50% da amostra. Portanto, se a mediana de um determinado processo é igual a 7, significa dizer que 50% dos valores da amostra são menores ou iguais a 7.

Gabarito: E.

19. (VUNESP/EsFCEx/2020) A tabela de frequências relativas acumuladas a seguir refere-se à distribuição dos salários dos empregados de uma empresa, sendo que não foi fornecida a correspondente frequência relativa acumulada do terceiro intervalo de classe (denotado por X na tabela).

CLASSE DE SALÁRIOS (R$)	FREQUÊNCIAS RELATIVAS ACUMULADAS (f_i)
$\mathbf{1.000,00 \vdash 3.000,00}$	10
$\mathbf{3.000,00 \vdash 5.000,00}$	30
$\mathbf{5.000,00 \vdash 7.000,00}$	X
$\mathbf{7.000,00 \vdash 9.000,00}$	80
$\mathbf{9.000,00 \vdash 11.000,00}$	100

Dado que o valor da mediana dessa distribuição obtida pelo método da interpolação linear apresentou um valor igual a R$ 6.600,00, obtém-se que X é igual a

a) 65.

b) 50.

c) 60.

d) 55.

e) 70.

Comentários:

O enunciado diz que a mediana da distribuição apresentada foi obtida pelo método da interpolação linear, portanto, temos que empregar a fórmula a seguir:

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

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.

Além disso, a questão informou o valor da mediana dessa distribuição é de R$ 6.600,00. Essa informação nos permite afirmar que a classe mediana corresponde à terceira classe ($5000 \vdash 7000$). Assim, temos os seguintes dados:

$$M_d = 6.600$$

$$l_{inf} = 5.000$$

$$f_{ac_{ant}} = 30$$

$$f_i = ?$$

$$h = 2.000$$

Então, aplicando a fórmula:

$$6600 = 5000 + \frac{\left(\frac{100}{2} - 30\right)}{f_i} \times 2000$$

$$6600 = 5000 + \frac{(50 - 30)}{f_i} \times 2000$$

$$6600 = 5000 + \frac{20}{f_i} \times 2000$$

$$1600 = \frac{40000}{f_i}$$

$$f_i = \frac{40000}{1600}$$

$$f_i = 25$$

Logo, a frequência reativa da classe mediana vale 25%.

Agora, devemos observar que X é uma frequência acumulada, portanto, precisamos somar f_i com a frequência acumulada anterior. Então, o valor de X será:

$$X = 25\% + 30\% = 55\%$$

Gabarito: D.

20. (VUNESP/Pref Ilhabela/2020) Uma pesquisa permitiu obter do mercado a tabela de frequências relativas abaixo, correspondente à distribuição dos preços de venda (p) referente a

um determinado equipamento. As frequências relativas da 1a e 2a classes por algum motivo não foram fornecidas e foram denotadas na tabela por x e y, respectivamente.

Classes de preços em R$	Frequências relativas (%)
50 < p ≤ 150	x
150 < p ≤ 250	y
250 < p ≤ 350	25
350 < p ≤ 450	30
450 < p ≤ 550	15
TOTAL	100

Utilizando o método da interpolação linear, encontra-se que o valor da mediana é igual a

a) R$ 290,00.

b) R$ 300,00.

c) R$ 310,00.

d) R$ 320,00.

e) R$ 330,00.

Comentários:

Primeiro, devemos saber que a aplicação do método da interpolação linear corresponde à utilização da fórmula da mediana. Segundo a fórmula, temos que a mediana é dada por:

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

Em que:

l_{inf} → limite inferior da classe mediana;

$\sum f_i$ → somatório das frequência simples;

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

h → amplitude da classe mediana;

f_i → frequência simples da classe mediana.

Para descobrirmos a classe mediana, precisamos calcular $\frac{\sum f_i}{2}$

$$\frac{\sum f_i}{2} = \frac{100\%}{2} = 50\%$$

Logo, a classe mediana será a terceira classe:

Classes de preços em R$	Frequências relativas (%)	Frequências Acumuladas (%)
50 < p ≤ 150	x	x
150 < p ≤ 250	y	x+y
250 < p ≤ 350	25	55 (>50)
350 < p ≤ 450	30	85
450 < p ≤ 550	15	100
TOTAL	100	

Substituindo os valores na fórmula, temos:

$$M_d = 250 + \left[\frac{50 - 30}{25}\right] \times 100$$

$$M_d = 250 + \left[\frac{20}{25}\right] \times 100$$

$$M_d = 250 + 80$$

$$M_d = 330$$

Gabarito: E.

21. (VUNESP/Pref Sorocaba/2020) Uma empresa de varejo decidiu usar o seguinte método com funcionários em treinamento: calcularia a média de vendas de todos os funcionários após um mês e só seriam efetivados os que estivessem acima da média, sendo demitidos aqueles que estivessem abaixo. Assim, os diretores da empresa imaginavam que ficariam com cerca de metade dos funcionários em treinamento. No entanto, ao final do treinamento, apenas um funcionário estava acima da média.

No método empregado pela empresa, o que foi determinante para que a média esperada de funcionários aprovados não fosse atingida?

a) Não se levou em conta a pequena variância nas vendas.

b) A média utilizada foi a geométrica.

c) Obviamente foi um erro de conta no cálculo da média.

d) O resultado deveria ter sido padronizado.

e) Deveria ter sido usada a mediana em vez da média.

Comentários:

A mediana é o valor que divide uma amostra, população ou distribuição de probabilidade em duas partes iguais, cada uma com 50% das observações. Assim, para garantir que metade dos funcionários em treinamento fossem aprovados, a empresa deveria ter utilizado a mediana em vez da média.

Gabarito: E.

22. (CESPE/Pref. São Cristóvão/2019) A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

Idade (em anos)	9	10	11	12	13	14
Quantidade de estudantes	6	22	0	1	0	1

A partir dessa tabela, julgue o item.

A mediana das idades é igual a 11,5 anos.

Comentários:

A mediana é o termo central de uma amostra ou população. Se temos 30 estudantes, então a mediana poderá ser encontrada pela média aritmética dos termos que ocupam as posições 15 e 16, pois, nesse caso, não há apenas um termo central.

Na tabela, observamos que há 22 estudantes com idades iguais a 10. Logo, os termos que ocupam as posições 15 e 16 são, respectivamente, 10 e 10. Vejamos:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	...
9	9	9	9	9	9	10	10	10	10	10	10	10	10	10	10	10	...

Então, a mediana é:

$$M_d = \frac{10 + 10}{2} = 10$$

Gabarito: Errado.

23. (CESPE/UNCISAL/2019) A SELIC é uma taxa referencial de juros estabelecida pelo Banco Central do Brasil como parâmetro para as taxas de juros cobradas pelos bancos comerciais no Brasil. A tabela seguinte mostra a evolução da SELIC, em porcentagem, no mês de janeiro dos anos de 2003 a 2012.

Disponível em: www.bcb.gov.br. Acesso em: nov. 2016 (adaptado).

O valor da mediana dos valores da SELIC mostrados no gráfico é igual a

a) 11,25.

b) 12,125.

c) 12,875.

d) 13,00.

e) 14,44.

Comentários:

A mediana é o termo central de uma amostra ou população. Se temos 10 anos representados no gráfico, então a mediana poderá ser encontrada pela média dos termos que ocupam as posições 5 e 6, pois nesse caso não há apenas um termo central.

Vamos ordenar os valores de cada ano em ordem crescente, em seguida acharemos os termos centrais.

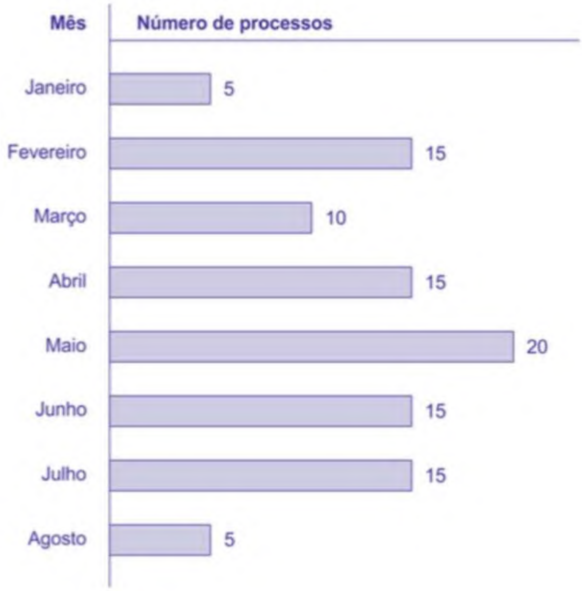
$$8{,}75 \quad 10{,}5 \quad 11{,}25 \quad 11{,}25 \quad \underbrace{12{,}75 \quad 13}_{termos\ centrais} \quad 16{,}25 \quad 17{,}25 \quad 18{,}25 \quad 25{,}15$$

Fazendo a média aritmética dos termos centrais, temos:

$$M_d = \frac{12{,}75 + 13}{2} = 12{,}875$$

Gabarito: C.

24. (FCC/BANRISUL/2019) Os números de processos com uma determinada característica autuados em um órgão público, de janeiro a agosto de 2018, podem ser visualizados pelo gráfico abaixo.



A respectiva média aritmética (número de processos por mês) está para a mediana assim como

a) 1 está para 16.

b) 2 está para 3.

c) 1 está para 8.

d) 5 está para 6.

e) 4 está para 3.

Comentários:

Para calcular a média aritmética, devemos somar os valores e dividir pelo número de meses:

$$\bar{x} = \frac{5 + 15 + 10 + 15 + 20 + 15 + 15 + 5}{8}$$

$$\bar{x} = \frac{100}{8} = 12{,}5$$

Agora, calcularemos a mediana. Para isso, organizaremos os números em ordem crescente (isto é, em rol):

$$5, 5, 10, \underbrace{15, 15}_{termos\ centrais}, 15, 15, 20$$

A mediana é a média dos termos centrais.

$$M_d = \frac{15 + 15}{2} = 15$$

Queremos calcular a razão entre a média e a mediana.

$$\frac{\bar{x}}{M_d} = \frac{12{,}5}{15}$$

Para retirar a casa decimal do numerador, vamos multiplicar numerador e denominador por 10.

$$\frac{\bar{x}}{M_d} = \frac{125}{150}$$

Agora, podemos simplificar por 25.

$$\frac{\bar{x}}{M_d} = \frac{5}{6}$$

Gabarito: D.

25. (FCC/Prefeitura do Recife/2019) A empresa Sigma apresenta pela tabela abaixo a distribuição dos salários registrados de seus 100 empregados em reais.

Salários (R$)	2.000	4.000	5.000	10.000	15.000	Total
Número de Empregados	0	10	40	X	Y	100

Não foram fornecidos os números de empregados que ganham R$ 10.000,00 e R$ 15.000,00 (denotados na tabela por x e y, respectivamente), mas sabe-se que a média aritmética dos salários é igual a R$ 8.400,00. O valor da soma da respectiva moda e da respectiva mediana desses salários é, em reais, igual a

a) 600y.

b) 625y.

c) 1.000y.

d) 750y.

e) 500y.

Comentários:

Conforme o enunciado, o número de empregados é igual a 100. Logo, temos que:

$$0 + 10 + 40 + x + y = 100$$

$$50 + x + y = 100$$

$$x + y = 50$$

Além disso, a média aritmética dos salários é igual a R$ 8.400,00. Como sabemos, para calcular essa média, devemos multiplicar cada salário pela sua respectiva frequência, somar todos os resultados, e dividir por 100, que é o total de funcionários.

$$\bar{x} = 8.400$$

$$\frac{2.000 \times 0 + 4.000 \times 10 + 5.000 \times 40 + 10.000 \times x + 15.000 \times y}{100} = 8400$$

$$\frac{0 + 40.000 + 20.000 + 10.000 \times x + 15.000 \times y}{100} = 8400$$

Nesse momento, podemos dividir todas as parcelas do numerador por 100.

$$400 + 2.000 + 100x + 150y = 8.400$$
$$2.400 + 100x + 150y = 8.400$$
$$100x + 150y = 6.000$$

Assim, teremos um sistema de equações lineares:

$$\begin{cases} x + y = 50 \\ 100x + 150y = 6.000 \end{cases}$$

Podemos resolver esse sistema multiplicando a primeira equação por (-100) para cancelar a incógnita x.

$$\begin{cases} -100x - 100y = -5.000 \\ 100x + 150y = 6.000 \end{cases}$$

Somando as duas equações, temos:

$$-100y + 150y = -5.000 + 6.000$$
$$50y = 1.000$$
$$y = 20$$

Como $x + y = 50$, temos:

$$x + 20 = 50$$
$$x = 30$$

Substituindo esses valores na tabela:

Salários	**Número de empregados**
2.000	0
4.000	10
5.000	40
10.000	30
15.000	20
Total	**100**

A moda é o termo que aparece com a maior frequência:

$$M_o = 5.000$$

Como o número de termos é par, $n = 100$, a mediana será a média aritmética dos dois termos centrais. Como são 100 termos, os termos centrais são os termos de posição $\frac{n}{2}$ e $\frac{n}{2} + 1$. Assim, os termos centrais são os termos de posição:

$$\frac{100}{2} = 50 \ e \ \frac{100}{2} + 1 = 51$$

A tabela indica que o número 4.000 apareceu 10 vezes, que o número 5.000 apareceu 40 vezes, e assim por diante

$$\left\{\underbrace{4.000, 4.000, \cdots, 4.000}_{10\ termos}, \underbrace{5.000, 5.000, \cdots, 5.000}_{40\ termos}, \underbrace{10.000, \cdots, 10.000}_{30\ termos}, \underbrace{15.000, \cdots, 15.000}_{20\ termos}\right\}$$

Assim, o termo de posição 50 é 5.000 e o termo de posição 51 é o número 10.000 (termos centrais).

$$\left\{4.000, 4.000, \cdots, 4.000, 5.000, 5.000, \cdots, \underbrace{5.000, 10.000}_{termos\ centrais}, \cdots, 10.000, 15.000, \cdots, 15.000\right\}$$

A mediana é a média dos termos centrais.

$$M_d = \frac{5.000 + 10.000}{2} = 7.500$$

Portanto, a soma da moda com a mediana é:

$$M_o + M_d = 5.000 + 7.500 = 12.500$$

Agora, para encontrarmos a resposta, dividiremos essa soma por y:

$$\frac{M_o + M_d}{y} = \frac{12.500}{20} = 625$$

$$625 \times 20 = 12.500$$

Gabarito: B.

26. (FCC/Prefeitura do Recife/2019) Com o objetivo de analisar a distribuição dos salários dos empregados de uma empresa, verificou-se que 10 empregados ganham, cada um, R$ 15.000,00; 20 ganham, cada um, R$ 2.500,00; 25 ganham, cada um, R$ 12.000,00; 60 ganham, cada um, R$ 5.000,00 e os restantes ganham, cada um, R$ 8.000,00. Sabendo-se que a mediana dos salários apresentou um valor igual a R$ 6.500,00, obtém-se que o valor da média aritmética supera o da moda em

a) R$ 2.750.00.

b) R$ 3.250,00.

c) R$ 3.000,00.

d) R$ 2.250,00.

e) R$ 2.500,00.

Comentários:

Observe a distribuição dos salários dos empregados:

Salários	Frequência
2.500	20
5.000	60

8.000	f
12.000	25
15.000	10
Total	**115 + *f***

O total de empregados é $115 + f$, que corresponde ao somatório das frequências. Quando o número de termos é ímpar, a mediana é exatamente o termo que fica no meio, ou seja, o termo de posição $\frac{n+1}{2}$. Por sua vez, quando o número de termos é par, a mediana é a média aritmética dos termos centrais, ou seja, a média aritmética entre os termos de posição $\frac{n}{2}$ e $\frac{n}{2} + 1$.

O enunciado nos disse que a mediana é igual a 6.500. Como nenhum funcionário ganha exatamente R$ 6.500,00, concluímos que a mediana será a média aritmética entre os dois termos centrais (o número de pessoas é par).

Observe que a média entre 5.000 e 8.000 é 6.500:

$$\frac{5.000 + 8.000}{2} = 6.500$$

Portanto, os dois termos centrais são 5.000 e 8.000.

Concluímos que o último 5.000 corresponde ao termo de posição $\frac{n}{2}$ e o primeiro 8.000 corresponde ao termo de ordem $\frac{n}{2} + 1$. Ora, o último 5.000 é o termo de posição 80. Basta perceber que o número 2.500 aparece 20 vezes e o número 5.000 aparece 60 vezes.

Portanto,

$$\frac{n}{2} = 80$$

$$n = 2 \times 80$$

$$n = 160$$

O total de pessoas é 160. Logo,

$$115 + f = 160$$

$$f = 45$$

Concluímos que 45 pessoas recebem 8 mil reais.

Também queremos calcular o valor da média aritmética e da moda. Para tanto, vamos reconstruir a tabela com o valor de f, que estava faltando.

Salários	Frequência
2.500	20
5.000	60

8.000	45
12.000	25
15.000	10
Total	**160**

A moda é o termo de maior frequência. Portanto,

$$M_o = 5.000$$

Vamos agora calcular a média aritmética. Para tanto, devemos multiplicar cada salário pela sua respectiva média, somar os resultados e dividir por 160, que é o total de pessoas.

Salários	Frequência
2.500	20
5.000	60
8.000	45
12.000	25
15.000	10
Total	**160**

Portanto, a média vale:

$$\bar{x} = \frac{2.500 \times 20 + 5.000 \times 60 + 8.000 \times 45 + 12.000 \times 25 + 15.000 \times 10}{160}$$

$$\bar{x} = \frac{50.000 + 300.000 + 360.000 + 300.000 + 150.000}{160}$$

$$\bar{x} = \frac{1.160.000}{160}$$

$$\bar{x} = 7.250$$

A questão pede a diferença entre a média e a moda (o quanto a média supera a moda).

$$\bar{x} - M_o = 7.250 - 5.000$$

$$= 2.250$$

Gabarito: D.

27. (FCC/Prefeitura do Recife/2019) Durante 40 dias, foi registrado o número de pessoas atendidas por dia em um guichê de uma repartição. A tabela abaixo apresentou os dados

observados sendo que não foram fornecidas as quantidades de dias em que foram atendidas uma pessoa por dia e duas pessoas por dia, indicadas na tabela por $q1$ e $q2$, respectivamente.

Número de pessoas atendidas	Quantidade de dias
0	9
1	q_1
2	q_2
3	5
4	1
Total	**40**

Sabendo-se que a mediana correspondente foi igual 1,5, tem-se que a soma da moda e da média aritmética (número de pessoas atendidas por dia) foi igual a

a) 3,00.

b) 2,80.

c) 3,45.

d) 3,20.

e) 2,95.

Comentários:

Conforme o enunciado, a soma das frequências é 40.

$$9 + q_1 + q_2 + 5 + 1 = 40$$

$$q_1 + q_2 = 25$$

A questão informou que a mediana é 1,5. Isso só é possível se os termos centrais (20º e 21º) forem 1 e 2, pois $\frac{1+2}{2} = 1{,}5$. Para que o 20º termo seja 1 e o 21º termo seja 2, devemos ter $q_1 = 11$, pois 9 + 11 = 20 (frequência da primeira classe + frequência da segunda classe).

Como $q_1 = 11$, então:

$$q_1 + q_2 = 25$$

$$11 + q_2 = 25$$

$$q_2 = 14$$

Vamos reescrever a tabela.

Número de pessoas	Quantidade

atendidas (x_i)	de dias (f_i)
0	9
1	11
2	14
3	5
4	1
Total	**40**

A moda é o termo de maior frequência. Como a maior frequência é 14, então a moda é 2.

$$M_o = 2$$

Vamos agora calcular a média. Termos que multiplicar cada valor x_i pela sua respectiva frequência e somar os resultados. Em seguida, basta dividirmos o total por 40.

Número de pessoas atendidas (x_i)	Quantidade de dias (f_i)	$x_i \times f_i$
0	9	$0 \times 9 = 0$
1	11	$1 \times 11 = 11$
2	14	$2 \times 14 = 28$
3	5	$3 \times 5 = 15$
4	1	$4 \times 1 = 4$
Total	**40**	

Portanto,

$$\bar{x} = \frac{58}{40} = 1{,}45$$

A soma da moda com a média é:

$$M_o + \bar{x} = 2 + 1{,}45 = 3{,}45$$

Gabarito: C.

28. (FUNDATEC/ESE/2019) A taxa de desemprego na grande São Paulo nos últimos 10 anos, em percentual da população, é dada pela tabela abaixo:

ANO	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
TOTAL	13,8	11,9	10,5	10,9	10,4	10,8	13,2	16,8	18,0	17,3

Nesse caso, podemos dizer que a mediana é:

a) 10,60.

b) 11,09.

c) 10,40.

d) 12,55.

e) 13,36.

Comentários:

A mediana caracteriza a posição central de uma série de valores. Ela divide uma série de valores em duas partes de tamanhos iguais, cada uma contendo o mesmo número de elementos. Quando temos um número par de observações, a mediana, por convenção, é definida como a média aritmética dos termos centrais.

Assim, ordenando os valores das taxas em ordem crescente, temos:

$$10{,}4 \quad 10{,}5 \quad 10{,}8 \quad 10{,}9 \quad \underbrace{11{,}9 \quad 13{,}2}_{termos\ centrais} \quad 13{,}8 \quad 16{,}8 \quad 17{,}3 \quad 18{,}0.$$

Como a quantidade de elementos é par, a mediana é calculada pela média dos termos centrais, ou seja:

$$M_d = \frac{11{,}9 + 13{,}2}{2} = 12{,}55.$$

Gabarito: D.

29. (VUNESP/MP-SP/2019) Considere o seguinte conjunto de dados numéricos para estatística.

8	30	5	6	4	8	12	6	13	8

Então, a soma da moda com a mediana e a média é igual a:

a) 22.

b) 24.

c) 26.

d) 28.

e) 30.

Comentários:

A moda é o termo que mais aparece. Logo:

$$M_o = 8$$

Para calcular a média, devemos somar todos os números e dividir pela quantidade de termos.

$$\bar{x} = \frac{8 + 30 + 5 + 6 + 4 + 8 + 12 + 6 + 13 + 8}{10} = \frac{100}{10} = 10$$

Para calcular a mediana, precisamos dispor os números em ordem crescente (rol).

$$4, 5, 6, 6, 8, 8, 8, 12, 13, 30$$

O número de termos é par (n = 10). Assim, temos duas posições centrais. A primeira posição central é a de posição $\frac{n}{2} = \frac{10}{2} = 5$. A outra posição central é a próxima (6º termo).

$$4, 5, 6, 6, \quad \underbrace{8, 8,}_{termos\ centrais} \quad 8, 12, 13, 30$$

Por convenção, a mediana será a média dos dois termos centrais.

$$M_d = \frac{x_5 + x_6}{2} = \frac{8 + 8}{2} = \frac{16}{2} = 8$$

A soma da moda com a mediana e a média é igual a:

$$M_o + M_d + \bar{x} = 8 + 8 + 10 = 26$$

Gabarito: C.

30. (VUNESP/PM-SP/2019) Na tabela, são apresentadas informações sobre o número de armas apreendidas pela Polícia Militar do Estado de São Paulo, no segundo semestre de 2018.

MÊS	NÚMERO DE ARMAS APREENDIDAS
Julho	688
Agosto	702
Setembro	680
Outubro	638
Novembro	695
Dezembro	629

(http://www.policiamilitar.sp.gov.br)

Se uma pesquisa utilizar a média aritmética simples do número de armas apreendidas, mensalmente, no segundo semestre, pela Polícia Militar do Estado de São Paulo, e outra pesquisa utilizar a mediana do número de armas apreendidas no segundo semestre, a diferença entre a mediana e a média será de

a) 30 armas.

b) 21 armas.

c) 12 armas.

d) 10 armas.

e) 08 armas.

Comentários:

A questão pede a diferença entre a mediana e a média de um conjunto de dados. Iniciaremos pelo cálculo da média do número de armas apreendidas. Para isso, basta somarmos todas as ocorrências no segundo semestre e dividirmos por seis:

$$\bar{x} = \frac{688 + 702 + 680 + 638 + 695 + 629}{6} = \frac{4032}{6} = 672$$

Agora, vamos calcular a mediana. Vimos que a mediana é o termo central de uma amostra ou população. Como temos um número par de meses, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 3 e 4. Organizando os dados da amostra em ordem crescente, temos:

$$629, 638, \quad \underbrace{680, 688,}_{termos\ centrais} \quad 695, 702$$

Encontrando a média dos termos nas posições 3 e 4:

$$M_d = \frac{680 + 688}{2} = 684$$

Então, a diferença entre a mediana e a média é:

$$M_d - \bar{x} = \ ?$$

$$684 - 672 = 12$$

Gabarito: C.

31. (VUNESP/Pref. Cerquilho/2019) A tabela apresenta informações sobre o número de funcionários em um escritório e os salários que são pagos a eles. Utilize as informações para responder a questão seguinte.

NÚMERO DE FUNCIONÁRIOS	SALÁRIOS
4	R$ 1.200,00
2	R$ 1.500,00
3	R$ 1.800,00
1	R$ 2.100,00

A mediana dos salários dos funcionários desse escritório é igual a

a) R$ 2.100,00.

b) R$ 1.800,00.

c) R$ 1.700,00.

d) R$ 1.500,00.

e) R$ 1.200,00.

Comentários:

A mediana é o termo central de uma amostra ou população. Como temos um número par de funcionários, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 5 e 6. Organizando os dados da amostra em ordem crescente, temos:

$$1200, 1200, 1200, 1200, \underbrace{1500, 1500,}_{termos\ centrais} 1800, 1800, 1800, 2100$$

Encontrando a média dos termos nas posições 5 e 6:

$$M_d = \frac{1500 + 1500}{2} = 1500$$

$$M_d = R\$\ 1.500,00$$

Gabarito: D.

32. (VUNESP/TRANSERP/2019) Considerando a taxa de desemprego hipotética de 11,7%(maio); 10,3% (junho); 8,9% (julho); 7,7% (agosto); 6,3% (setembro) e 4,9% (outubro) a mediana do período é

a) 8,3%

b) 7,9%

c) 7,7%

d) 7,5%

e) 7,2%

Comentários:

A mediana é o termo central de uma amostra ou população. Como temos um número par de meses, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 3 e 4. Organizando os dados da amostra em ordem crescente, temos:

$$4{,}9\%, 6{,}3\%, \underbrace{7{,}7\%, 8{,}9\%}_{termos\ centrais}, 10{,}3\%, 11{,}7\%$$

Encontrando a média dos termos nas posições 3 e 4:

$$M_d = \frac{7{,}7 + 8{,}9}{2} = 8{,}3$$

$$M_d = 8{,}3\%$$

Gabarito: A.

33. (VUNESP/TJ-SP/2019) Durante um período, decidiu-se analisar o comportamento do número de processos especiais autuados por dia em uma repartição pública. A tabela a seguir apresenta os resultados obtidos, sendo k a quantidade de dias em que não foram autuados processos.

NÚMERO DE PROCESSOS	0	1	2	3	4	5	TOTAL
QUANTIDADE DE DIAS	k	14	18	24	14	2	10k

Com relação a esta tabela, foram obtidos os respectivos valores da moda (Mo), mediana (Md) e média aritmética (Me), em número de processos por dia. Verifica-se então que (Mo + Md + Me) é igual a

a) 6,30

b) 7,85

c) 6,80

d) 6,85

e) 7,35

Comentários:

A soma de todas as frequências é igual a $10k$.

$$k + 14 + 18 + 24 + 14 + 2 = 10k$$

$$k + 72 = 10k$$

$$9k = 72 \rightarrow k = 8$$

Assim, temos a seguinte tabela.

Número de Processos (X_i)	Quantidade de Dias (f_i)
0	8
1	14
2	18
3	24
4	14
5	2
Total	**80**

A moda é o termo que mais aparece, ou seja, é o termo que possui a maior frequência. Portanto,

$$M_o = 3$$

Vamos agora calcular a média aritmética. Para tanto, vamos multiplicar cada valor pela respectiva frequência.

Número de Processos (X_i)	**Quantidade de Dias (f_i)**	$X_i \times f_i$
0	8	0 x 8 = 0
1	14	1 x 14 = 14
2	18	2 x 18 = 36
3	24	3 x 24 = 72
4	14	4 x 14 = 56
5	2	5 x 2 = 10
Total	**80**	**188**

Para calcular a média, basta dividir 188 pelo total de observações (80).

$$\bar{x} = \frac{\Sigma x_i \times f_i}{n} = \frac{188}{80} = 2{,}35$$

Agora, vamos calcular a mediana. Como $\boldsymbol{n}$ é par, $n = 80$, há duas posições centrais. A primeira posição central é a de posição $\frac{n}{2} = \frac{80}{2} = 40$. A outra posição central é a próxima (41º termo). Por convenção, a mediana será a média dos dois termos centrais.

$$M_d = \frac{x_{40} + x_{41}}{2}$$

Vamos construir a coluna da frequência acumulada.

Número de Processos (X_i)	**Frequência (f_i)**	**Frequência Acumulada (f_{ac})**
0	8	0 + 8 = **8**
1	14	8 + 14 = **22**
2	18	22 + 18 = **40**
3	24	40 + 24 = **64**
4	14	64 + 14 = **78**
5	2	78 + 2 = **80**
Total	**80**	

Observe que a frequência acumulada da terceira linha é igual a 40. Isso quer dizer que já foram escritos 40 números até a terceira linha. Portanto, $x_{40} = 2$. Na próxima linha da tabela, vamos escrever mais 24 números. Isso quer dizer que $x_{41} = 3$. Logo,

$$M_d = \frac{x_{40} + x_{41}}{2} = \frac{2+3}{2} = \frac{5}{2} = 2{,}5$$

A questão pede o valor de $M_o + M_d + M_e$.

$$M_o + M_d + M_e = 3 + 2{,}5 + 2{,}35 = 7{,}85$$

Gabarito: B.

34. (CESPE/Polícia Federal/2018)

	DIA				
	1	2	3	4	5
X (quantidade diária de drogas apreendidas, em kg)	10	22	18	22	28

Tendo em vista que, diariamente, a Polícia Federal apreende uma quantidade X, em kg, de drogas em determinado aeroporto do Brasil, e considerando os dados hipotéticos da tabela precedente, que apresenta os valores observados da variável X em uma amostra aleatória de 5 dias de apreensões no citado aeroporto, julgue o item.

A mediana das quantidades X observadas na amostra em questão foi igual a 18 kg.

Comentários:

Temos 5 elementos na amostra, a mediana divide o conjunto ao meio e ocupa a posição central. Então, a mediana ocupará a posição 3 do conjunto. Organizando os dados da tabela em ordem crescente, temos:

$$10 \quad 18 \quad \underbrace{22}_{termo\ central} \quad 22 \quad 28$$

Logo, a mediana é o número 22, termo que ocupa a posição 3.

$$M_d = 22$$

Gabarito: Errado.

35. (FCC/ALESE/2018) Em um grupo de pessoas encontramos as seguintes idades: 20, 30, 50, 39, 20, 25, 41, 47, 36, 45, 41, 52, 18, 41. A mediana e a moda são, respectivamente,

a) 36 e 45.

b) 40 e 41.

c) 41 e 20.

d) 42 e 39.

e) 39 e 42.

Comentários:

A questão pede a mediana e a moda de um conjunto de dados não agrupados. Primeiro, temos que organizar os dados em ordem crescente:

$$18, 20, 20, 25, 30, 36, 39, \mathbf{41}, \mathbf{41}, \mathbf{41}, 45, 47, 50, 52$$

A moda é o termo que possui maior frequência, ou seja, que aparece mais vezes. O número mais frequente é o 41. Portanto,

$$M_o = 41$$

Agora, para calcularmos a mediana, devemos levar em consideração que, quando o número de termos n é ímpar, a mediana é o termo que ocupa central, ou seja, é o termo de posição $\frac{n+1}{2}$. Por outro lado, quando o número de termos n é par, a mediana é calculada pela média dos dois termos centrais, que ocupam as posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

No nosso caso, temos 14 números. Como a quantidade de termos é par, então a mediana será a média aritmética entre os dois termos centrais: o sétimo e o oitavo:

$$18, 20, 20, 25, 30, 36, \underbrace{39, 41,}_{termos\ centrais} 41, 41, 45, 47, 50, 52$$

O sétimo termo é 39 e o oitavo termo é 41. Portanto,

$$M_d = \frac{39 + 41}{2} = 40$$

Gabarito: B.

36. (FCC/ISS-São Luís/2018) Um levantamento foi realizado com 40 instituições financeiras, localizadas em uma região, com relação às taxas mensais de juros aplicadas para financiamento de veículos. Verificou-se que cinco instituições aplicam a taxa de 0,80% ao mês, duas aplicam a taxa de 1,20% ao mês, oito aplicam a taxa de 1,25% ao mês, x aplicam a taxa de 1,12% ao mês e y aplicam a taxa de 0,96% ao mês. Se a média aritmética destas taxas foi igual a 1,05%, então a soma da mediana e a moda correspondentes foi de

a) 2,00%

b) 2,24%

c) 2,08%

d) 2,16%

e) 1,92%

Comentários:

Vamos organizar os dados em uma tabela:

Taxa (%)	f_i
0,80	5
0,96	y
1,12	x
1,20	2
1,25	8

A média aritmética dessas taxas foi de 1,05%. Para calcular a média, devemos multiplicar cada taxa pela sua respectiva frequência e somar todos os resultados. Em seguida, devemos dividir pelo total de observações, que é 40.

Taxa (%)	f_i	$x_i \times f_i$
0,80	5	$0{,}8 \times 5 = 4$
0,96	y	$0{,}96 \times y$
1,12	x	$1{,}12 \times x$
1,20	2	$1{,}20 \times 2 = 2{,}40$
1,25	8	$1{,}25 \times 8 = 10$
Total	**40**	$\mathbf{16,4 + 0,96 \times y + 1,12 \times x}$

A média é 1,05%. Como desprezamos o símbolo % na tabela, vamos igualar a média a 1,05:

$$\frac{16{,}4 + 0{,}96y + 1{,}12x}{40} = 1{,}05$$

$$16{,}4 + 0{,}96y + 1{,}12x = 40 \times 1{,}05$$

$$16{,}4 + 0{,}96y + 1{,}12x = 42$$

$$0{,}96y + 1{,}12x = 25{,}6 \qquad (Eq. 1)$$

A soma das frequências é 40:

$$5 + y + x + 2 + 8 = 40$$

$$y + x = 25$$

$$y = 25 - x$$

Vamos substituir y por $25 - x$ na Eq. 1.

$$0{,}96y + 1{,}12x = 25{,}6$$

$$0{,}96 \times (25 - x) + 1{,}12x = 25{,}6$$

$$24 - 0{,}96x + 1{,}12x = 25{,}6$$

$$0{,}16x = 1{,}6$$

$$x = 10$$

Como $y + x = 25$, então:

$$y + 10 = 25$$

$$y = 15$$

Vamos reescrever a tabela.

Taxa (%)	f_i
0,80	5
0,96	15
1,12	10
1,20	2
1,25	8

A moda é o termo de maior frequência. Como a maior frequência é 15, então a moda é 0,96%.

$$M_o = 0{,}96\%$$

Como são 40 termos, então a mediana será a média dos dois termos centrais (20º e 21º). Como há 5 termos na primeira linha e 15 termos na segunda linha, concluímos que o vigésimo termo é 0,96%. O 21º termo estará na próxima linha: 1,12%.

Portanto,

$$M_d = \frac{0{,}96\% + 1{,}12\%}{2} = 1{,}04\%$$

A soma da mediana e da moda é:

$$M_d \ + \ M_o = 0{,}96\% + 1{,}04\% = 2\%$$

Gabarito: A.

37. (FCC/SEDU-ES/2018) As notas dos dez alunos de uma sala foram: 1, 2, 4, 6, 6, 7, 8, 8, 8, 10. A diferença entre a moda e a mediana dessas notas é

a) 1,5.

b) 2,5.

c) 0,5.

d) 2,0.

e) 1,0.

Comentários:

A moda é o termo que aparece mais vezes em um conjunto de dados. O termo mais frequente no conjunto apresentado é o número 8, logo:

$$M_o = 8$$

Como são 10 alunos, a mediana vai ser a média entre os dois termos centrais (5º e 6º termos).

$$M_d = \frac{6+7}{2} = 6{,}5$$

A diferença entre a moda e a mediana é:

$$M_o - M_d = 8 - 6{,}5 = 1{,}5$$

Gabarito: A.

38. (FCC/TRT 2ª Região/2018) Considerando na tabela abaixo a distribuição de frequências absolutas, referente aos salários dos n empregados de uma empresa, em R$ 1.000,00, observa-se que além do total dos empregados (n) não é fornecida também a frequência correspondente ao intervalo da 4ª classe (f_4).

Classes de salários (R$ 1.000,00)	$1 \vdash 3$	$3 \vdash 5$	$5 \vdash 7$	$7 \vdash 9$	$9 \vdash 11$	Total
Frequências	4	8	10	f_4	2	n

O valor da média aritmética destes salários, obtido considerando que todos os valores incluídos num certo intervalo de classe são coincidentes com o ponto médio deste intervalo, é igual a R$ 6.200,00. O valor da mediana em R$, obtido pelo método da interpolação linear, é igual a

a) $400{,}0\ f_4$

b) $412{,}5\ f_4$

c) $387{,}5\ f_4$

d) $350{,}0\ f_4$

e) $375{,}0\ f_4$

Comentários:

Nessa questão, precisamos encontrar o valor da mediana em função de f_4. A média aritmética é 6.200. Sabemos que a média aritmética de dados agrupados em intervalos de classe é calculada pela razão entre média ponderada dos pontos médios de cada classe e a frequência de cada classe.

Vamos reorganizar a tabela:

Classes de salários (R$ 1.000,00)	Ponto Médio (PM)	Frequências (f_i)	$PM \times f_i$
$1 \vdash 3$	2	4	2×4

$3 \vdash 5$	4	8	4×8
$5 \vdash 7$	6	10	6×10
$7 \vdash 9$	8	f_4	$8 \times f_4$
$9 \vdash 11$	10	2	10×2
Total		$\mathbf{4+8+10+f_4+2}$	

Dadas as informações da nova tabela, temos:

$$\bar{x} = \frac{2 \times 4 + 4 \times 8 + 6 \times 10 + 8 \times f_4 + 10 \times 2}{4 + 8 + 10 + f_4 + 2}$$

$$6{,}2 = \frac{120 + 8f_4}{24 + f_4}$$

$$148{,}8 + 6{,}2f_4 = 120 + 8f_4$$

$$1{,}8f_4 = 28{,}8$$

$$f_4 = 16$$

Agora que já sabemos o valor de f_4, podemos calcular o valor de n na tabela. Assim:

$$4 + 8 + 10 + \mathbf{16} + 2 = 40 \rightarrow n = 40$$

Já descobrimos as informações que faltavam, mas ainda não sabemos qual é a classe mediana. Para isso, vamos construir uma coluna de frequências acumuladas:

Classes de salários (R$ 1.000,00)	**Ponto Médio (PM)**	**Frequências (f_i)**	**Frequências Acumuladas (f_{ac})**
$1 \vdash 3$	2	4	4
$3 \vdash 5$	4	8	12
$\mathbf{5 \vdash 7}$	**6**	**10**	$\mathbf{22}$ $(> \mathbf{20})$
$7 \vdash 9$	8	16	38
$9 \vdash 11$	10	2	40
Total		**40**	

A classe mediana corresponde à primeira classe cuja frequência acumulada supera metade da frequência total $\left(\frac{\sum f_i}{2}\right)$.

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 $l_{inf} = 5$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 12$.
- a frequência da própria classe é $f_i = 10$.
- a amplitude da classe é $h = 7 - 5 = 2$.

Aplicando a fórmula, temos:

$$M_d = 5 + \left[\frac{\left(\frac{40}{2}\right) - 12}{10}\right] \times 2$$

$$M_d = 5 + \left(\frac{20 - 12}{10}\right) \times 2$$

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

$$M_d = 5 + 1{,}6$$

$$M_d = 6{,}6 \ \rightarrow R\$\ 6.600{,}00$$

A questão pede o resultado em função de f_4, logo:

$$M_d = \frac{6.600}{16}$$

$$M_d = 412{,}5 \times f_4$$

Gabarito: B.

39. (FGV/ALE-RO/2018) Sejam x, y e z, respectivamente, a média, a mediana e a moda dos sete valores 9, 10, 6, 5, 20, 9 e 4. É correto concluir que

a) x < y < z .

b) x < y = z

c) x = y < z

d) y < z = x

e) x = y = z

Comentários:

A média é calculada pela soma dos valores dividida pela quantidade de valores:

$$x = \frac{9 + 10 + 6 + 5 + 20 + 9 + 4}{7} = 9$$

Agora, para calcular a mediana, precisamos organizar os números em ordem crescente:

$$4, 5, 6, \quad \underbrace{9,}_{termo\ central} \quad 9, 10, 20$$

A mediana é o termo que ocupa a posição central. Portanto,

$$y = 9$$

A moda é o termo que aparece em maior frequência. O número que aparece mais vezes é o 9, portanto:

$$z = 9$$

Assim, concluímos que:

$$x = y = z$$

Gabarito: E.

40. (FGV/ALE-RO/2018) A tabela a seguir mostra o número de gols sofridos por um time de futebol nas dez primeiras partidas de um campeonato:

Jogo	1	2	3	4	5	6	7	8	9	10
Gols Sofridos	0	1	2	0	1	2	1	0	3	2

A média e a mediana do número de gols sofridos nesses jogos são respectivamente

a) 1,2 e 1,0.

b) 1,2 e 1,5.

c) 1,1 e 1,0.

d) 1,0 e 1,0.

e) 1,0 e 1,5.

Comentários:

Calculamos a média somando os valores do conjunto e dividindo pela quantidade de números somados. No caso da tabela, somamos todos os gols sofridos e dividimos pela quantidade de jogos. Assim:

$$\bar{x} = \frac{0 + 1 + 2 + 0 + 1 + 2 + 1 + 0 + 3 + 2}{10}$$

$$\bar{x} = \frac{12}{10}$$

$$\bar{x} = 1{,}2$$

Para calcular a mediana precisamos achar o valor central do conjunto. Para isso, vamos colocar os valores da amostra em ordem crescente e encontrar seu termo central.

$$0, 0, 0, 1, \underbrace{1, 1}_{termos\ centrais}, 2, 2, 2, 3$$

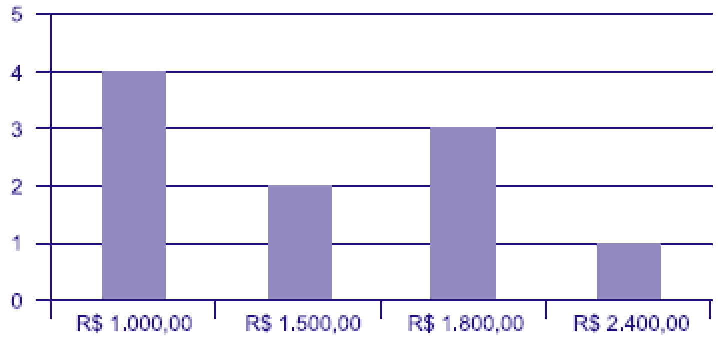
No caso da nossa amostra, temos 10 valores, então teremos 2 termos centrais. Vamos calcular a média desses dois termos e teremos a mediana:

$$M_d = \frac{1+1}{2}$$

$$M_d = 1$$

Gabarito: A.

41. (VUNESP/Pref. Serrana/2018) O gráfico apresenta informações sobre o número de funcionários em um pequeno comércio e os salários que são pagos a eles. Utilize as informações para responder à questão.



A mediana dos salários dos funcionários desse comércio é igual a

a) R$ 2.400,00.

b) R$ 2.000,00.

c) R$ 1.800,00.

d) R$ 1.500,00.

e) R$ 1.000,00.

Comentários:

A mediana é o termo central de uma amostra ou população. Como temos um número par de funcionários, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 5 e 6. Organizando os dados da amostra em ordem crescente, temos:

$$1000, 1000, 1000, 1000, \underbrace{1500, 1500,}_{termos\ centrais} 1800, 1800, 1800, 2400$$

Encontrando a média dos termos nas posições 5 e 6:

$$M_d = \frac{1500 + 1500}{2} = 1500$$

$$M_d = R\$\ 1.500,00$$

Gabarito: D.

42. (VUNESP/Pref. Sertãozinho/2018) Para responder à questão seguinte, considere os números 1,5; 1,5; 1,6; 1,6; 1,6; 1,7; 1,7; 1,8; 1,9 e 1,9 como as alturas, em metros, de 10 pessoas.

A mediana das alturas dessas pessoas

a) É 1,8 metro.

b) Está entre 1,7 e 1,8 metro.

c) É 1,7 metro.

d) Está entre 1,6 e 1,7 metro.

e) É 1,6 metro.

Comentários:

A mediana é o termo central de uma amostra ou população. Como temos um número par de pessoas, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 5 e 6. Organizando os dados da amostra em ordem crescente, temos:

$$1{,}5 \quad 1{,}5 \quad 1{,}6 \quad 1{,}6 \quad \underbrace{1{,}6 \quad 1{,}7}_{termos\ centrais} \quad 1{,}7 \quad 1{,}8 \quad 1{,}9 \quad 1{,}9$$

Encontrando a média dos termos nas posições 5 e 6:

$$M_d = \frac{1{,}6 + 1{,}7}{2} = 1{,}65$$

Analisando as alternativas, temos que a mediana está entre 1,6 e 1,7.

Gabarito: D.

43. (FCC/SEFAZ-MA/2016) Os registros da temperatura máxima diária dos primeiros 6 dias de uma semana foram: 25 °C; 26 °C, 28,5 °C; 26,8 °C; 25 °C; 25,6 °C. Incluindo também o registro da temperatura máxima diária do 7º dia dessa semana, o conjunto dos sete dados numéricos será unimodal com moda igual a 25 °C, e terá mediana igual a 26 °C. De acordo com os dados, é correto afirmar que, necessariamente, a temperatura máxima diária do 7º dia foi

a) Inferior a 25 °C.

b) Superior a 26,8 °C.

c) Igual a 26 °C.

d) Inferior a 25,6 °C.

e) Superior a 26 °C.

Comentários:

Para resolvermos a questão, vamos primeiro organizar os valores das temperaturas em ordem crescente. Assim:

$$25 \quad 25 \quad 25{,}6 \quad 26 \quad 26{,}8 \quad 28{,}5$$

O enunciado afirmou que a mediana vale 26. Como a mediana corresponde ao termo central das observações, a temperatura do 7° dia dividirá as observações em duas partes, a primeira contendo 3 valores menores que a mediana, e a segunda contendo 3 valores maiores que a mediana. Assim:

$$25 \quad 25 \quad 25{,}6 \quad \underbrace{26}_{Mediana} \quad X_7 \quad 26{,}8 \quad 28{,}5$$

ou

$$25 \quad 25 \quad 25{,}6 \quad \underbrace{26}_{Mediana} \quad 26{,}8 \quad X_7 \quad 28{,}5$$

ou

$$25 \quad 25 \quad 25{,}6 \quad \underbrace{26}_{Mediana} \quad 26{,}8 \quad 28{,}5 \quad X_7$$

Sabendo que o conjunto também é unimodal, isto é, que possui uma única moda, X_7 não poderá ser igual a nenhum outro valor existente no conjunto.

Logo, X_7 deverá assumir valores maiores que 26 necessariamente.

Gabarito: E.

44. (FGV/CODEBA/2016) Uma das características principais da mediana é

a) A invariância à unidade de medida utilizada.

b) A robustez à presença de outliers.

c) A identificação da observação mais frequente.

d) O fato de, em seu cálculo, dar mais peso às observações mais frequentes.

e) A normalização pelos desvios em relação à média.

Comentários:

Vamos analisar cada uma das alternativas:

- Letra A: **Errada**. Se mudarmos a unidade de medida utilizada a mediana também mudará.
- Letra B: **Correta**. A presença de outliers pouco impacta na mediana;
- Letra C: **Errada**. A mediana representa o termo central do conjunto. As observações frequentes estão relacionadas à moda.
- Letra D: **Errada**. A mediana representa o termo central do conjunto. As observações frequentes estão relacionadas à moda.
- Letra E: **Errada**. A mediana representa o termo central do conjunto ou da amostra. Contudo, a mediana não é uma medida de dispersão.

Gabarito: B.

45. (FGV/IBGE/2016) Após a extração de uma amostra, as observações obtidas são tabuladas, gerando a seguinte distribuição de frequências:

Valor	3	5	9	13
Frequência	5	9	10	3

Considerando que E(X) = Média de X, Mo(X) = Moda de X e Me(X) = Mediana de X, é correto afirmar que:

a) E(X) = 7 e Mo(X) = 10;

b) Me(X) = 5 e E(X) = 6,3;

c) Mo(X) = 9 e Me(X) = 9;

d) Me(X) = 9 e E(X) = 6,3;

e) Mo(X) = 9 e E(X) = 7.

Comentários:

A moda é, por definição, o valor que aparece em maior frequência. Portanto, o valor que tem frequência máxima é $M_o(X) = 9$.

O número total de termos é $5 + 9 + 10 + 3 = 27$. Como o número de termos é ímpar, a mediana é o termo de ordem:

$$\frac{(27+1)}{2} = 14.$$

Organizando os termos de forma ascendente, o 14º termo é o número 5. Veja que o número 3 aparece 5 vezes e o número 5 aparece 9 vezes. Portanto,

$$M_d(X) = 5.$$

Agora, calcularemos o valor da média. Para tanto, multiplicaremos cada termo pela sua respectiva frequência e dividiremos o resultado pela soma das frequências.

$$E(X) = \frac{3 \times 5 + 5 \times 9 + 9 \times 10 + 13 \times 3}{27} = \frac{189}{27} = 7.$$

Gabarito: E.

46. (VUNESP/CM Registro/2016) Foi construída uma amostra com 10 funcionários de uma determinada repartição para a verificação do consumo de papel A4 por mês naquela área. A verificação do consumo desses funcionários mostrou-se conforme a tabela a seguir.

FUNCIONÁRIO	CONSUMO EM FOLHAS DE PAPEL POR MÊS
A	15
B	20

C	18
D	10
E	16
F	22
G	24
H	30
I	8
J	12

Assinale a alternativa que contém, correta e respectivamente, os valores do consumo médio e do consumo mediano de papel A4 dessa repartição.

a) 15 e 16.

b) 17 e 17,5.

c) 17,5 e 17.

d) 18 e 19.

e) 19,5 e 18.

Comentários:

A questão pede o consumo médio e o consumo mediano de papel A4 em uma determinada repartição. Vamos iniciar pelo cálculo da média. Para isso, basta somarmos todas as ocorrências e dividirmos pelo total delas:

$$\bar{x} = \frac{15 + 20 + 18 + 10 + 16 + 22 + 24 + 30 + 8 + 12}{10} = \frac{175}{10} = 17{,}5$$

A mediana é o termo central de uma amostra ou população. Como temos um número par de funcionários, por convenção, a mediana deve ser encontrada pela média dos termos centrais, que ocupam as posições 5 e 6. Organizando os dados da amostra em ordem crescente, temos:

$$8 \quad 10 \quad 12 \quad 15 \qquad \underbrace{16 \quad 18}_{termos\ centrais} \qquad 20 \quad 22 \quad 24 \quad 30$$

Encontrando a média dos termos na posição 5 e 6:

$$M_d = \frac{16 + 18}{2} = 17$$

Gabarito: C.

47. (CESPE/TELEBRAS/2015) Considerando que os possíveis valores de um indicador X, elaborado para monitorar a qualidade de um serviço de cabeamento residencial para a

comunicação de dados, sejam elementos do conjunto {0, 1, 2, 3, 4, 5} e que uma amostra aleatória de 5 residências tenha apontado os seguintes indicadores: 4, 4, 5, 4 e 3, julgue o próximo item.

A mediana e a moda dos indicadores registrados na amostra foram iguais a 4.

Comentários:

A mediana é o termo central de uma amostra ou população. Os dados devem estar dispostos em ordem crescente. Então, temos:

$$3, 4, \underbrace{4}_{termo\ central}, 4, 5$$

Portanto, o valor mediano é o termo central $M_d = 4$.

A moda é o termo que possui maior frequência, ou seja, é o termo que aparece mais vezes. Portanto, a moda também é igual a 4.

Gabarito: Certo.

48. (CESPE/DEPEN/2015)

Quantidade diária de incidentes (N)	Frequência relativa
0	0,1
1	0,2
2	0,5
3	0,0
4	0,2
total	1

Considerando os dados da tabela mostrada, que apresenta a distribuição populacional da quantidade diária de incidentes (N) em determinada penitenciária, julgue o item que se segue.

A distribuição de N não é simétrica em torno da média, apesar de a média e a mediana serem iguais.

Comentários:

Apenas analisando a tabela percebemos que de fato a distribuição não é simétrica. Para que fosse simétrica as frequências acima da frequência central deveriam ser iguais às que estão abaixo desta. Não é o caso da tabela em questão.

Agora, vamos verificar se a média e a mediana são iguais. Para calcular a média basta multiplicarmos cada quantidade (N) por sua frequência relativa, a soma de todos os resultados será a média:

$$\bar{x} = \frac{(0 \times 0{,}1) + (1 \times 0{,}2) + (2 \times 0{,}5) + (3 \times 0{,}0) + (4 \times 0{,}2)}{0{,}1 + 0{,}2 + 0{,}5 + 0{,}0 + 0{,}2}$$

$$\bar{x} = \frac{0{,}2 + 1 + 0{,}8}{1{,}0}$$

$$\bar{x} = 2$$

Como sabemos, a mediana é o termo central da amostra. Na tabela em questão, a mediana é representada pelo termo que tem frequência acumulada de 50%. Podemos observar que as duas primeiras frequências acumulam um total de 30% da amostra. Somando a terceira frequência às duas anteriores, chegaremos a um total de 80% da amostra.

Sendo assim, podemos concluir que a terceira frequência, que corresponde à quantidade de incidentes $N = 2$, é a primeira a superar a faixa de 50%. Dessa forma, a frequência acumulada de 50% é representada pela quantidade de incidentes igual a 2:

$$M_d = 2$$

Portanto, a média e a mediana, de fato, são iguais.

Gabarito: Certo.

49. (FCC/DPE SP/2015) A tabela abaixo corresponde às frequências absolutas dos salários de todos os homens e de todas as mulheres que são empregados de uma empresa.

Classe de Salários (R$)	Frequências Absolutas	
	Homens	Mulheres
1.500 ⊢ 2.500	10	5
2.500 ⊢ 3.500	40	20
3.500 ⊢ 4.500	$2k$	$(K + 10)$
4.500 ⊢ 5.500	15	10
5.500 ⊢ 6.500	5	5
Total	$70 + 2K$	$50 + K$

Utilizando o método da interpolação linear para o cálculo da mediana, tem-se que o valor da mediana dos homens é igual a R$ 3.750,00 e o das mulheres é igual a

a) R$ 3.875,00.

b) R$ 4.025,00.

c) R$ 3.925,00.

d) R$ 3.825,00.

e) R$ 4.000,00.

Comentários:

Inicialmente, precisamos saber as frequências acumuladas para homens e mulheres, para isso vamos montar a tabela com as novas informações:

Classe de Salários (R$)	Frequências Absolutas		Frequências Acumuladas	
	Homens	Mulheres	Homens	Mulheres
$1.500 \vdash 2.500$	10	5	10	5
$2.500 \vdash 3.500$	40	20	50	25
$3.500 \vdash 4.500$	$2k$	$(K + 10)$	$50 + 2K$	$(K + 35)$
$4.500 \vdash 5.500$	15	10	$65 + 2K$	$(K + 45)$
$5.500 \vdash 6.500$	5	5	$70 + 2K$	$(K + 50)$
Total	$\mathbf{70 + 2K}$	$\mathbf{50 + K}$		

O enunciado da questão diz que a mediana para os homens vale 3750. Se a frequência total para homens é $70 + 2K$, logo a mediana ocupará a posição $35 + K$. Assim, temos as seguintes informações:

- classe mediana dos homens $\rightarrow$ $3.500 \vdash 4.500$;
- limite inferior à classe mediana $\rightarrow 3.500$;
- limite superior à classe mediana $\rightarrow 4.500$.

Sabendo disso, encontraremos o valor de K usando a fórmula da mediana, pelo método da interpolação linear. Assim, 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 $l_{inf} = 3.500$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 50$.
- a frequência da própria classe é $f_i = 2k$.
- a amplitude da classe é $h = 1.000$.

Aplicando a fórmula, temos:

$$3.750 = 3.500 + \left[\frac{\left(\frac{70+2k}{2}\right) - 50}{2k}\right] \times 1.000$$

$$3.750 = 3.500 + \left(\frac{k-15}{2k}\right) \times 1.000$$

$$\frac{250}{1.000} = \frac{k-15}{2k}$$

$$\frac{1}{4} = \frac{k-15}{2k}$$

$$2k = 4k - 60$$

$$2k = 60$$

$$k = 30$$

Agora, substituindo o valor de K na tabela conseguimos achar as frequências acumuladas para mulheres. Se a frequência total para mulheres é $50 + \mathbf{30} = 80$, logo, a mediana ocupará a posição 40, pois $\frac{80}{2} = 40$. Assim, temos as seguintes informações:

- classe mediana das mulheres → $3.500 \vdash 4.500$;
- limite inferior à classe mediana → 3.500;
- limite superior à classe mediana → 4.500.

Aplicando a fórmula, temos:

$$M_d = 3.500 + \left[\frac{\left(\frac{80}{2}\right) - 25}{40}\right] \times 1.000$$

$$M_d = 3.500 + \left(\frac{15}{40}\right) \times 1.000$$

$$M_d = 3.500 + \left(\frac{3}{8}\right) \times 1.000$$

$$M_d = 3.500 + 3 \times 125$$

$$M_d = 3.875$$

Gabarito: A.

50. (CESPE/ANTAQ/2014)

Nota atribuída pelo passageiro	Frequência
0	15
1	30

2	45
3	50
4	35
5	5

A tabela acima apresenta os resultados de uma pesquisa de satisfação realizada em uma amostra de usuários dos serviços de transporte fluvial prestados por uma empresa. Com base nessas informações e na tabela, julgue o próximo item.

A mediana da série de notas obtidas pela empresa é 3.

Comentários:

Vamos construir a coluna da frequência acumulada.

Nota atribuída pelo passageiro	Frequência (f_i)	Frequência Acumulada (f_{ac})
0	15	15
1	30	45
2	45	90
3	50	140
4	35	175
5	5	180
Total	**180**	

São 180 números. Como n é par, então a mediana será a média dos dois termos centrais. O primeiro termo central é o termo de posição $\frac{n}{2} = \frac{180}{2} = 90$. Assim, devemos calcular a média entre os termos de ordem 90 e 91.

A primeira frequência acumulada indica que 15 números são iguais a 0. A segunda frequência acumulada indica que 45 números são menores do que ou iguais a 1. A terceira frequência acumulada indica que 90 números são menores do que ou iguais a 2. Portanto, o 90º termo é 2.

O 91º termo estará na próxima linha da tabela. Assim, $x_{91} = 3$.

Vamos calcular a mediana:

$$M_d = \frac{x_{90} + x_{91}}{2} = \frac{2 + 3}{2} = 2{,}5$$

Gabarito: Errado.

51. (FCC/SEFAZ RJ/2014) O Departamento de Pessoal de certo órgão público fez um levantamento dos salários, em número de salários mínimos (SM), dos seus 400 funcionários, obtendo os seguintes resultados:

Salários (em número de SM)	Frequência absoluta
$4 \vdash 6$	48
$6 \vdash 8$	100
$8 \vdash 10$	x
$10 \vdash 12$	y
$12 \vdash 16$	40
Total	**400**

Sabe-se que a mediana dos salários desses funcionários calculada por meio dessa tabela pelo método da interpolação linear é igual a 8,8 SM. Nessas condições, o salário médio desses 400 funcionários, em número de salários mínimos, considerando que todos os valores incluídos em um intervalo de classe são coincidentes com o ponto médio do intervalo, é igual a

a) 8,54

b) 8,83

c) 8,62

d) 8,93

e) 8,72

Comentários:

Primeiro, precisamos encontrar os valores de X e Y. Para isso, precisamos acrescentar a coluna de frequências acumuladas à tabela anterior:

Salários (em número de SM)	Frequências Absolutas	Frequências Acumuladas
$4 \vdash 6$	48	48

$6 \vdash 8$	100	148
$8 \vdash 10$	**x**	**$148 + x$**
$10 \vdash 12$	y	$148 + x + y$
$12 \vdash 16$	40	$188 + x + y$
Total	**400**	

O enunciado informou que a mediana é **8,8**, portanto, a **classe mediana** será o intervalo que engloba o valor da mediana, isto é, a classe com limite inferior igual a **8** e limite superior igual a **10**. Além disso, como a tabela representa os salários médios de 400 funcionários, temos que a posição da mediana é indicada pela frequência de valor 200, pois $\frac{400}{2} = 200$.

Sabendo disso, encontraremos o valor de X usando a fórmula da mediana, pelo método da interpolação linear. Assim, 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 $l_{inf} = 8{,}0$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 148$.
- a frequência da própria classe é $f_i = X$.
- a amplitude da classe é $h = 10 - 8 = 2$.

Aplicando a fórmula, temos:

$$8{,}8 = 8{,}0 + \left[\frac{\left(\frac{400}{2}\right) - 148}{X}\right] \times (10 - 8)$$

$$8{,}8 = 8{,}0 + \left(\frac{200 - 148}{X}\right) \times 2$$

$$0{,}8 = \left(\frac{52}{X}\right) \times (2)$$

$$X = \frac{104}{0{,}8} = 130$$

Sabendo o valor de X, podemos encontrar o valor de Y:

$$48 + 100 + 130 + Y + 40 = 400$$

$$318 + Y = 400$$

$$Y = 400 - 318$$

$$Y = 82$$

Agora vamos calcular o salário médio dos funcionários. Para tanto, reconstruiremos a tabela anterior, incluindo os pontos médios:

Salários (em número de SM)	**Ponto Médio (PM)**	**Frequências Absolutas (f_i)**
$4 \vdash 6$	$\frac{4+6}{2} = 5$	48
$6 \vdash 8$	$\frac{6+8}{2} = 7$	100
$8 \vdash 10$	$\frac{8+10}{2} = 9$	**130**
$10 \vdash 12$	$\frac{10+12}{2} = 11$	82
$12 \vdash 16$	$\frac{12+16}{2} = 14$	40
Total		**400**

Para calcular a média de uma distribuição de frequências agrupada em intervalos de classe, devemos multiplicar cada ponto médio por sua respectiva frequência, somar tudo e dividir o resultado pela frequência total. Vejamos:

Salários	**Ponto médio (PM_i)**	**Frequência absoluta(f_i)**	$PM_i \times f_i$
$4 \vdash 6$	5	48	240
$6 \vdash 8$	7	100	700
$8 \vdash 10$	9	130	1170
$10 \vdash 12$	11	82	902
$12 \vdash 16$	14	40	560
Total		**400**	**3572**

Então, a média aritmética dos salários é:

$$\bar{x} = \frac{\sum_{i=1}^{n}(PM_i \times f_i)}{\sum_{i=1}^{n} f_i} = \frac{3.572}{400} = 8{,}93$$

Gabarito: D.

52. (FGV/AL-BA/2014) Os dados a seguir são uma amostra de 11 salários mensais (aproximados) em reais:

2.080 1.830 2.480 3.010 1.450 1.650 2.500 1.740 3.600 1.900 2.840

A mediana desses salários, em reais, é

a) 1.990.

b) 2.080.

c) 1.650.

d) 2.000.

e) 2.220.

Comentários:

A mediana é o termo central de uma amostra. Dividimos o conjunto ao meio e o termo que ocupar a posição central será a mediana. Para isso os dados da amostra devem estar em ordem crescente. Ordenando os salários, temos:

$$1450,\ 1650, 1740, 1830, 1900, \underbrace{2080}_{termo\ central}, 2480, 2500, 2840, 3010, 3600$$

Portanto, o valor mediano é 2080.

Gabarito: B.

53. (FGV/CGE-MA/2014) Sobre uma amostra com uma quantidade ímpar de valores, todos diferentes de uma variável aleatória, sabe-se que a média é maior que a mediana.

Com relação aos valores dessa amostra é necessariamente verdade que

a) Há mais valores acima da média do que abaixo da média.

b) Há mais valores abaixo da média do que acima da média.

c) Há mais valores acima da média do que abaixo da mediana.

d) Há mais valores acima da mediana do que abaixo da média.

e) A quantidade de valores acima da média é igual à quantidade de valores abaixo da média.

Comentários:

O enunciado informa que a amostra tem uma quantidade ímpar de valores. Logo, teremos um único valor central que corresponderá à mediana. Antes de adentrarmos na análise das alternativas, tomemos como exemplo o conjunto de valores $\{1, 2, 3, 5, 9\}$:

$$M_d = 3$$

$$\bar{x} = \frac{1+2+3+5+9}{5} = 4$$

Nosso exemplo está em conformidade com o que é dito no enunciado, vez que a média é maior que a mediana. Sendo assim, vamos analisar as alternativas:

- Letra A: **Errado**. Pelo nosso exemplo observamos que há mais valores abaixo da média do que acima da média.
- Letra B: **Correto**. Há mais valores abaixo da média.
- Letra C: **Errado**. A mediana ocupa a posição central da amostra, portanto, divide a amostra em duas partes iguais.
- Letra D: **Errado**. Há mais valores abaixo da média.
- Letra E: **Errado**. A alternativa se refere à mediana e não à média.

Gabarito: B.

54. (FGV/Pref. Recife/2014) A seguinte amostra de idades foi obtida:

19; 25; 39; 20; 16; 27; 40; 38; 28; 32; 30.

Assinale a opção que indica a mediana dessas idades.

a) 27

b) 28

c) 29

d) 30

e) 31

Comentários:

A mediana é o termo central de uma amostra. Para encontrar a mediana, temos que dividir o conjunto e identificar o termo que ocupa a posição central. Para isso, os dados da amostra devem estar em ordem crescente.

Ordenando as idades, temos:

$$16, 19, 20, 25, 27, \underbrace{28}_{termo\ central}, 30, 32, 38, 39, 40$$

A mediana é 28.

Gabarito: B.

55. (CESPE/PRF/2013)

Considerando os dados apresentados no gráfico, julgue o item seguinte.

A média do número de acidentes ocorridos no período de 2007 a 2010 é inferior à mediana da sequência de dados apresentada no gráfico.

Comentários:

Os valores referentes ao período de 2007 a 2010 são:

$$129, 141, 159\ e\ 183.$$

Para calcular a média desses números, devemos somar os quatro valores e dividir por 4.

$$\bar{x} = \frac{129 + 141 + 159 + 183}{4} = 153$$

Vamos, agora, calcular a mediana da sequência de dados apresentada no gráfico:

$$110, 111, 129, \underbrace{141}_{termo\ central}, 159, 183, 189$$

O valor mediano é o termo que ocupa a posição central, logo, corresponde ao número 141.

A média não é inferior à mediana. Logo, o item está errado.

Gabarito: Errado.

56. (CESPE/Polícia Federal/2012) Com relação a estatística, julgue o item seguinte.

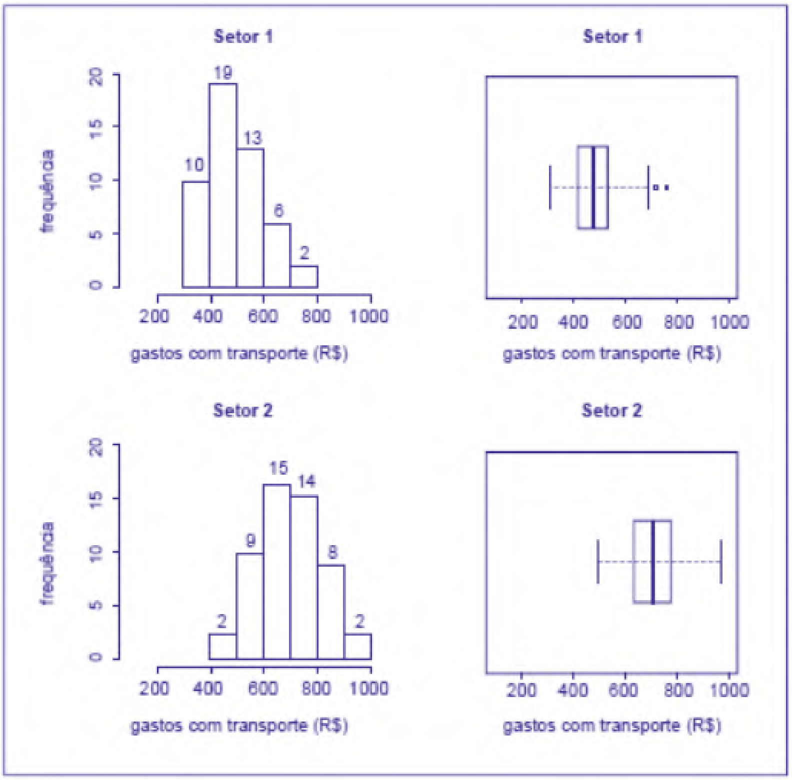
Ao contrário da mediana amostral, a média aritmética é menos sensível à presença de valores extremos (ou valores atípicos ou *outliers*).

Comentários:

A mediana não é sensível a valores extremos. Por outro lado, a média é bastante influenciada por valores extremos. Logo, o item está errado.

Gabarito: Errado

57. (CESPE/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
723,79	767,53				

Setor 2

488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56
663,81	670,12	671,90	673,78	684,69	685,98
693,35	698,58	708,78	719,80	721,16	734,84

735,94 746,34 754,83 756,10 756,96 760,80
762,29 766,24 770,11 797,73 804,06 805,97
807,29 832,83 844,00 866,77 878,27 897,09
943,10 963,25

Considerando essas informações, julgue o item.

A mediana das despesas registradas pelos servidores do setor 2 é igual a R$ 693,35.

Comentários:

Há 50 registros para cada setor. Como o número de termos é par, a mediana será a média aritmética dos dois termos centrais. A primeira posição central é $\frac{n}{2} = \frac{50}{2} = 25$. A outra posição central é $\frac{n}{2} + 1 = 26$.

Assim, devemos procurar os termos de posição 25 e 26:

488,37	493,73	547,72	552,66	567,94	571,49	572,26	582,00	583,63	594,77
598,46	619,25	624,20	631,03	634,51	637,21	655,70	657,56	663,81	670,12
671,90	673,78	684,69	685,98	**693,35**	**698,58**	708,78	719,80	721,16	734,84
735,94	746,34	754,83	756,10	756,96	760,80	762,29	766,24	770,11	797,73
804,06	805,97	807,29	832,83	844,00	866,77	878,27	897,09	943,10	963,25

O enunciado diz que a mediana é exatamente 693,35, mas certamente será maior que esse valor, vez que resultará da média aritmética entre 693,35 e 698,58. Vejamos:

$$M_d = \frac{693{,}35 + 698{,}58}{2} = 695{,}96$$

Gabarito: Errado.

58. (FUNDATEC/SEFAZ-RS/2009) A tabela a seguir representa a distribuição de frequências da idade de uma amostra de moradores de um asilo. Utilize para resolver a questão.

X_i	f_i
70 ⊢ 74	7
74 ⊢ 78	19
78 ⊢ 82	13
82 ⊢ 86	11
86 ⊢ 90	6

90 ⊢ 94	4
Total	**60**

A idade aproximada da mediana é:

a) 78,22.

b) 80,00.

c) 79,38.

d) 78,55.

e) 79,23.

Comentários:

O enunciado pede a mediana de uma distribuição de frequências de idades de uma amostra de moradores de um asilo. Para resolver a questão, o primeiro passo é montar a coluna de frequências acumuladas:

Classes	Frequência simples	Frequência acumulada
70 a 74	7	7
74 a 78	19	26
78 a 82	**13**	**39 (> 30)**
82 a 86	11	50
86 a 90	6	56
90 a 94	4	60

Assim, temos os seguintes dados:

$$M_d = ?$$

$$l_{inf} = 78$$

$$f_{ac_{ant}} = 26$$

$$f_i = 13$$

$$h = 4$$

Sabendo disso, podemos empregar a fórmula a seguir:

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

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.

Então, aplicando a fórmula:

$$M_d = 78 + \left[\frac{\left(\frac{60}{2}\right) - 26}{13} \right] \times 4$$

$$M_d = 78 + \left[\frac{30 - 26}{13} \right] \times 4$$

$$M_d = 78 + \left[\frac{4}{13} \right] \times 4$$

$$M_d = 78 + \frac{16}{13} \cong 79{,}23$$

Outra forma de responder essa questão é usando o método da interpolação linear. Da tabela anterior, podemos tirar as seguintes informações:

- o valor 78 corresponde à frequência acumulada 26.
- a mediana (M_d) é o valor que corresponde à frequência acumulada 30 (pois a mediana tem frequência acumulada igual à metade da quantidade de dados).
- o valor 82 tem a frequência acumulada 39.

Podemos montar o seguinte quadro:

	Valor	Frequência acumulada
1ª Linha	78	26
2ª Linha	M_d	30
3ª Linha	82	39

A interpolação linear nos diz que podemos fazer o seguinte: fazemos a segunda linha menos a primeira; fazemos a terceira linha menos a primeira. As diferenças são proporcionais:

$$\frac{M_d - 78}{82 - 78} = \frac{30 - 26}{39 - 26}$$

$$\frac{M_d - 78}{4} = \frac{4}{13}$$

$$M_d = 78 + \frac{16}{13}$$

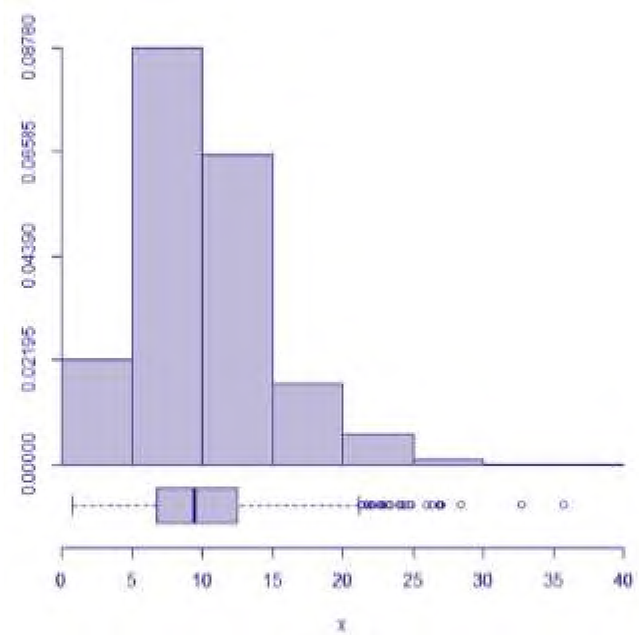
$$M_d \cong 79{,}23$$

Gabarito: E.

QUESTÕES COMENTADAS

Quartil, Decil e Percentil

1. (CESPE/PETROBRAS/2022) A figura seguinte mostra o histograma como uma estimativa da função de densidade de uma distribuição X, juntamente com o diagrama boxplot correspondente a esse conjunto de dados.



Tamanho da amostra	**1.000**
Média amostral	**10**
Desvio padrão amostral	**4,7**

Considerando a figura e as informações apresentadas no quadro, julgue o item que se segue.

O primeiro decil da distribuição do conjunto de dados em tela é igual ou inferior a 5.

Comentários:

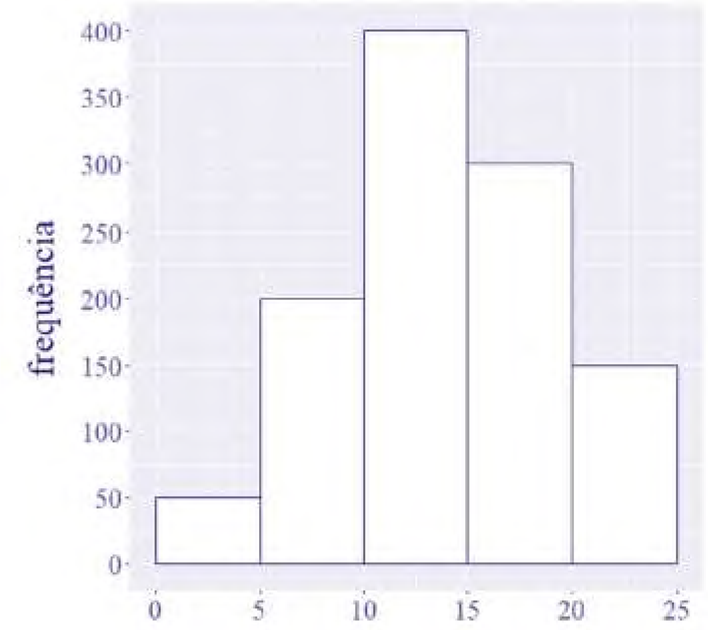
Os decis dividem uma série ordenada em dez partes iguais, cada uma contendo 10% dos valores da sequência. O gráfico apresentado traz muitas informações, mas vamos nos concentrar na primeira barra do histograma, pois a questão pede apenas o primeiro decil. Desta forma, basta multiplicarmos os valores referentes à primeira barra:

$$5 \times 0{,}02195 = 0{,}10975 \implies 10{,}975\%$$

Assim, podemos afirmar que 10,975% dos dados são menores ou iguais a 5. Logo, o primeiro decil da distribuição do conjunto de dados é igual ou inferior a 5.

Gabarito: Certo.

2. (CESPE/TELEBRAS/2022)



Considerando que o histograma apresentado descreve a distribuição de uma variável quantitativa X por meio de frequências absolutas, julgue o item que se segue.

O segundo decil da distribuição da variável X é igual a 20.

Comentários:

Para facilitar a resolução, podemos reorganizar os dados do histograma em uma tabela e acrescentar as frequências acumuladas, da seguinte forma:

Classes (x)	Frequências	Freq. acumulada
$0 \vdash 5$	50	50
$5 \vdash 10$	200	250
$10 \vdash 15$	400	650
$15 \vdash 20$	300	950
$20 \vdash 25$	150	1100

Total	1100	

Agora, para calcularmos o decil de dados agrupados em classe, precisamos encontrar a posição ocupada pelo decil. Para isso, utilizamos a seguinte fórmula:

$$P_{D_k} = \frac{k \times \sum fi}{10}$$

Queremos encontrar a posição do 2º decil:

$$P_{D_2} = \frac{2 \times 1100}{10} = 220$$

Logo, o segundo decil está na posição 220.

Observando as frequências acumuladas na tabela, descobrimos que ele pertence à segunda classe de 5 a 10. Assim, podemos aplicar a fórmula para o cálculo do decil:

$$D_k = l_{inf\,D_k} + \left[\frac{\frac{k \times \sum f_i}{10} - f_{ac\,ant}}{f_{D_k}}\right] \times h_{D_k}$$

em que:

$l_{inf\,D_k}$ → limite inferior à classe do decil;

fac_{ant} → frequência acumulada da classe anterior ao decil;

h_{D_k} → amplitude da classe;

f_{D_k} → frequência da classe.

Observem que já calculamos a posição do decil, então basta substituir em $\frac{k \times \sum f_i}{10}$:

$$D_2 = 5 + \left[\frac{220 - 50}{200}\right] \times 5$$

$$D_2 = 5 + \left[\frac{170}{200}\right] \times 5$$

$$D_2 = 5 + 0{,}85 \times 5$$

$$D_2 = 5 + 4{,}25$$

$$D_2 = 9{,}25$$

Portanto, o segundo decil da distribuição é 9,25.

Gabarito: Errado.

3. (CESPE/BANESE/2021) A respeito do conjunto de dados {11, 6, 28, 51, 49, 32, 33}, julgue o item a seguir.

O primeiro quartil do conjunto de dados em tela é igual ou superior a 33.

Comentários:

Os quartis dividem uma série ordenada em quatro partes iguais, cada uma contendo 25% dos valores da sequência. Para encontrarmos a posição do primeiro quartil, podemos usar a fórmula:

$$P_{Q_1} = \frac{n}{4}$$

em que n é igual ao número de termos.

Assim, temos:

$$P_{Q_1} = \frac{7}{4} = 1{,}75$$

Portanto, o primeiro quartil está mais próximo do segundo elemento do conjunto, cujo valor é 11.

Gabarito: Errado.

4. (IBFC/IBGE/2021) Uma pesquisa foi realizada com 40 pessoas e o terceiro quartil da distribuição de frequência é igual a 21,25. Sabendo que a amplitude da classe é 2, a frequência acumulada da classe anterior é 20, então a frequência da classe referente ao terceiro quartil é igual a:

a) 12

b) 16

c) 20

d) 8

e) 10

Comentários:

A questão nos forneceu algumas informações importantes para o cálculo do quartil. Então, podemos aplicar a fórmula:

$$Q_k = l_{inf_{Q_k}} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac_{ant}}}{f_{Q_k}}\right] \times h_{Q_k}$$

em que:

$Q_k = 21{,}25 \rightarrow$ valor do quartil;

$l_{inf_{Q_k}} = 40 - 20 = 20 \rightarrow$ limite inferior à classe quartílica;

$fac_{ant} = 20 \rightarrow$ frequência acumulada da classe anterior à classe quartílica;

$h_{Q_k} = 2 \rightarrow$ amplitude da classe;

$f_{Q_k} = ? \rightarrow$ frequência da classe.

Atribuindo os valores fornecidos no enunciado, temos:

$$21{,}25 = 20 + \left[\frac{\frac{3 \times 40}{4} - 20}{f_{Q_3}}\right] \times 2$$

$$21{,}25 = 20 + \left[\frac{30 - 20}{f_{Q_3}}\right] \times 2$$

$$21{,}25 - 20 = \left[\frac{10}{f_{Q_3}}\right] \times 2$$

$$1{,}25 = \left[\frac{10}{f_{Q_3}}\right] \times 2$$

$$\frac{10}{f_{Q_3}} = \frac{1{,}25}{2}$$

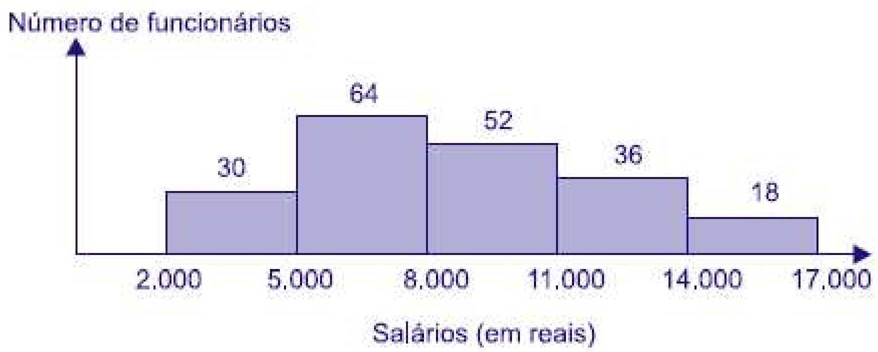
$$1{,}25 f_{Q_3} = 20$$

$$f_{Q_3} = \frac{20}{1{,}25}$$

$$f_{Q_3} = 16$$

Gabarito: B.

5. (VUNESP/EsFCEx/2021) O gráfico apresenta informações sobre a distribuição dos funcionários em relação aos salários pagos em uma empresa.



Com base nas informações apresentadas, é correto afirmar que um funcionário do grupo dos 25% dos maiores salários nessa empresa não ganha menos do que

a) R$ 11.333,33.

b) R$ 11.666,66.

c) R$ 11.527,00.

d) R$ 11.478,25.

e) R$ 11.745,75.

Comentários:

Basicamente, a questão nos pede calcularmos o terceiro quartil da distribuição. Sabemos que o terceiro quartil representa 75% dos menores salários, logo, ao encontrarmos o salário limite do terceiro quartil, também vamos descobrir o menor salário dos 25% que têm os maiores salários.

Para facilitar a resolução da questão, podemos reorganizar os dados do histograma em uma tabela e acrescentar as frequências acumuladas:

Salários (x)	Frequências	Freq. acumulada
2 ⊢ 5	30	30
5 ⊢ 8	64	94
8 ⊢ 11	52	146
11 ⊢ 14	36	182
14 ⊢ 17	18	200
Total	**200**	

Calculando a posição do quartil:

$$P_{Q_k} = \frac{k \times \sum fi}{4}$$

$$P_{Q_k} = \frac{3 \times 200}{4} = 150$$

Logo, o terceiro quartil está na posição 150, o que corresponde à 4ª classe 11 ⊢ 14.

Agora, podemos aplicar a fórmula:

$$Q_k = l_{inf\,Q_k} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac\,ant}}{f_{Q_k}}\right] \times h_{Q_k}$$

em que:

$l_{inf\,Q_k}$ → Limite inferior à classe quartílica;

fac_{ant} → frequência acumulada da classe anterior à classe quartílica;

h_{Q_k} → amplitude da classe;

$f_{Q_k} \rightarrow$ frequência da classe.

$$Q_3 = 11 + \left[\frac{150 - 146}{36}\right] \times 3$$

$$Q_3 = 11 + \left[\frac{4}{36}\right] \times 3$$

$$Q_3 = 11 + 0{,}11111 \times 3$$

$$Q_3 = 11 + 0{,}33333$$

$$Q_k = 11{,}33333$$

Convertendo novamente para a casa dos milhares, temos:

$$11{,}33333 \times 1000 = 11.333{,}33$$

Gabarito: A.

6. (CESPE/ME/2020)

Mínimo	**5**
Moda	**9**
Mediana	**9**
Média	**10**
Máximo	**15**

Um levantamento amostral proporcionou as estatísticas precedentes, referentes a determinada variável quantitativa X.

Considerando essas informações e que a variável X é composta por 1240 observações, julgue o item subsequente.

O terceiro quartil da variável X foi inferior a 9.

Comentários:

Com as informações do enunciado, podemos analisar possíveis distribuições de valores para a variável x, como a apresentada a seguir. Reparem que a mediana do conjunto é 9; a moda também é 9; a média é 10; o valor mínimo é 5; e o máximo é 15:

1	2	3	⋯	620	621	622	623	624	⋯	1239	1240
5	9	9	⋯	**9**	**9**	11	11	11	⋯	13	15

Como a mediana é 9, sabemos que as posições 620 e 621 obrigatoriamente assumem o valor 9. Assim, todos os números anteriores a essas posições devem ser iguais ou inferiores a 9.

Por outro lado, para que a média da distribuição seja 10, é necessário que os demais elementos sejam maiores que 9. Então, nessa situação, o terceiro quartil da variável x obrigatoriamente será um valor maior que o valor da mediana, que vale 9.

Gabarito: Errado.

7. (MARINHA DO BRASIL/2020) Sabe-se que o valor da amplitude semi-interquartílica de um conjunto de dados é 180 e que o valor do 3° quartil é 4 vezes o valor do 1° quartil. Assinale a opção que apresenta o valor do 1 ° e do 3° quartil, respectivamente.

a) 120 e 480

b) 90 e 360

c) 180 e 480

d) 120 e 360

e) 180 e 720

Comentários:

Temos que a amplitude semi-interquartílica é dada por:

$$\frac{Q_3 - Q_1}{2}$$

Assim,

$$\frac{Q_3 - Q_1}{2} = 180$$

Agora, sabemos que o 3º quartil é 4 vezes o valor do 1º quartil:

$$Q_3 = 4Q_1$$

$$\frac{4Q_1 - Q_1}{2} = 180$$

$$4Q_1 - Q_1 = 180 \times 2$$

$$3Q_1 = 360$$

$$Q_1 = \frac{360}{3}$$

$$Q_1 = 120$$

Substituindo Q_3, temos:

$$Q_3 = 4Q_1$$

$$Q_3 = 4 \times 120$$

$$Q_3 = 480$$

Gabarito: A.

8. (MARINHA DO BRASIL/2020) Dada a distribuição amostral abaixo, determine o 4° Decil e 66º Percentil e assinale a opção correta.

Classes	Frequência
$2 \vdash 6$	6
$6 \vdash 10$	12
$10 \vdash 14$	8
$14 \vdash 18$	14
$18 \vdash 22$	10

a) 8 e 18

b) 8 e 20

c) 9 e 20

d) 11 e 16

e) 11 e 14

Comentários:

Para calcularmos os decis, precisamos dividir a amostra em dez partes iguais. Inicialmente, precisamos determinar a posição do decil que queremos encontrar. Assim, temos que:

$$P_{D_k} = \frac{K \times \sum f_i}{10} = P_{D_4} = \frac{4 \times 50}{10} = 20$$

Depois, temos que identificar a posição do decil na coluna de frequências acumuladas, isto é, a frequência acumulada imediatamente igual ou superior à posição do decil. A posição do 4º decil está na frequência acumulada de valor igual a 20. Logo, vamos considerar a 3ª classe como a classe do 4º decil.

O decil é dado por:

$$D_4 = l_{inf\,D_4} + \left[\frac{\frac{4 \times \sum f_i}{10} - f_{ac\,ant}}{f_{D_4}}\right] \times h_{D_4}$$

$$D_4 = 10 + \left[\frac{\frac{4 \times 50}{10} - 18}{8}\right] \times 4$$

$$D_4 = 10 + \left[\frac{20 - 18}{8}\right] \times 4$$

$$D_4 = 10 + 0{,}25 \times 4$$

$$D_4 = 11$$

Já o percentil divide a amostra em cem partes Iguais. Da mesma forma que fizemos para o decil, precisamos determinar a posição do percentil:

$$P_{P_k} = \frac{K \times \sum f_i}{100} = P_{P_{66}} = \frac{66 \times 50}{100} = 33$$

Depois, identificamos a posição do percentil na coluna de frequências acumuladas, isto é, a frequência acumulada imediatamente igual ou superior à posição do percentil. Assim, a posição do 66º percentil está na frequência acumulada 33. Logo, vamos considerar a 4ª classe como a classe do 66º percentil.

Portanto, o percentil é dado por:

$$P_{66} = l_{inf\,P_{66}} + \left[\frac{\frac{66 \times \sum f_i}{100} - f_{ac\,ant}}{f_{P_{66}}}\right] \times h_{P_{66}}$$

$$P_{66} = 14 + \left[\frac{\frac{66 \times 50}{100} - 26}{14}\right] \times 4$$

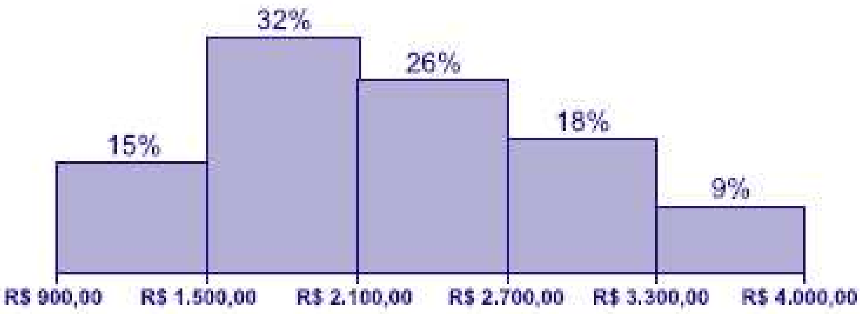
$$P_{66} = 14 + \left[\frac{33 - 26}{14}\right] \times 4$$

$$P_{66} = 14 + 0{,}5 \times 4$$

$$P_{66} = 16$$

Gabarito: D.

9. (VUNESP/UNICAMP/2019) O gráfico apresenta a distribuição dos salários dos funcionários de um escritório.



Sabendo-se que, em cada classe a distribuição de salários é uniforme, 30% dos salários mais baixos desse grupo variam de R$ 900,00 a

a) R$ 1.768,25.

b) R$ 1.779,25.

c) R$ 1.781,25.

d) R$ 1.795,25.

e) R$ 1.801,25.

Comentários:

A questão fornece uma distribuição de salários e pede o limite superior do intervalo em que estão contidos os 30% salários mais baixos desse grupo. Basicamente, o enunciado está pedindo para encontrarmos o trigésimo percentil dessa distribuição, que calcularemos por meio da seguinte fórmula:

$$P_k = l_{inf_{P_k}} + \left[\frac{\frac{k \times \sum fi}{100} - f_{ac_{ant}}}{f_{P_k}}\right] \times h_{P_k}$$

Em que:

l_{P_k} = limite inferior da classe do percentil considerado;

$f_{ac_{ant}}$ = frequência acumulada da classe anterior do percentil considerado;

h_{P_k} = amplitude do intervalo de classe do percentil considerado;

f_{P_k} = frequência simples da classe do percentil considerado.

Logo, precisamos descobrir que classe corresponde ao trigésimo percentil. Analisando o histograma, observamos que a primeira classe contém apenas 15% dos salários. A segunda, por sua vez, contempla 32% dos salários. Somando as duas, percebemos que a frequência desejada (15% + 32% > 30%) já foi superada, indicando que o trigésimo percentil está contido na segunda classe, com frequência relativa de 32%, que varia de R$ 1.500 a R$ 2.100.

Assim, temos os seguintes dados:

$$k = 30$$

$$P_k = ?$$

$$l_{inf_{P_k}} = 1.500$$

$$f_{ac_{ant}} = 15\%$$

$$f_{P_k} = 32\%$$

$$h_{P_k} = 600$$

Aplicando a fórmula, temos:

$$P_{30\%} = 1.500 + \left[\frac{\frac{30 \times 100\%}{100} - 15\%}{32\%}\right] \times 600$$

$$P_{30\%} = 1.500 + \left(\frac{15}{32}\right) \times 600$$

$$P_{30\%} = 1.500 + 281{,}25 = 1781{,}25$$

Gabarito: C.

10. (FGV/TJ-AL/2018) Para avaliar a produtividade de um dado conjunto de varas da justiça, é extraída uma amostra do número de audiências efetivamente realizadas durante um determinado período.

Os dados foram tratados, obtendo-se as seguintes estatísticas:

Me (A.) = 22, Q_1=19 e Q_3 =27

Essas estatísticas representam os Quartis da distribuição.

Adotando a técnica de Box-Plot para fins da identificação de outliers, sobre os valores A1 = 6, A2 = 11 e A3 = 40 tem-se que:

a) Todos são outliers;

b) Os dois primeiros são outliers;

c) Apenas A3 é um outlier;

d) A1 e A3 são outliers;

e) Nenhum deles é outlier.

Comentários:

Vamos, inicialmente, calcular a diferença interquartílica. Assim:

$$d = Q_3 - Q_1 = 27 - 19 = 8$$

Agora, calculamos o limite superior:

$$l_{sup} = Q_3 + 1{,}5 \times d$$

$$l_{sup} = 27 + 1{,}5 \times 8 = 39$$

O limite inferior fica:

$$l_{inf} = Q_1 - 1{,}5 \times d$$

$$l_{inf} = 19 - 1{,}5 \times 8 = 7$$

Desta forma, temos que todos os valores que estiverem acima de 39 e abaixo de 7 serão outliers (dados discrepantes), a exemplo dos valores A1 e A3. E todos os valores que estiverem entre 7 e 39 serão considerados normais, a exemplo de A2.

Gabarito: D.

11. (CESPE/SEE-DF/2017) Um levantamento estatístico, feito em determinada região do país, mostrou que jovens com idades entre 4 e 17 anos assistem à televisão, em média, durante 6 horas por dia. A tabela a seguir apresenta outras estatísticas produzidas por esse levantamento.

	Distribuição dos tempos gastos assistindo televisão (*T, em horas*)
1° quartil	2
2° quartil	4
3° quartil	8
1° decil	1
9° decil	10

Tendo como referência essas informações, julgue o seguinte item.

O desvio quartílico dos tempos T foi igual a 3.

Comentários:

Vamos aproveitar essa questão para relembrar a diferença entre intervalo interquartílico e desvio quartílico. O intervalo interquartílico (ou amplitude interquartílica, ou distância interquartil) é calculado pela fórmula:

$$Q_3 - Q_1$$

Enquanto o desvio quartílico (ou amplitude semi-interquartílica) é dado pela expressão:

$$\frac{Q_3 - Q_1}{2}$$

Os valores dos quartis foram fornecidos na tabela. Portanto, o desvio quartílico é igual a

$$\frac{8 - 2}{2} = 3$$

Gabarito: Certo.

12. (CESPE/SEE-DF/2017) Um levantamento estatístico, feito em determinada região do país, mostrou que jovens com idades entre 4 e 17 anos assistem à televisão, em média, durante 6 horas por dia. A tabela a seguir apresenta outras estatísticas produzidas por esse levantamento.

	Distribuição dos tempos gastos assistindo televisão (*T, em horas*)
1° quartil	2
2° quartil	4
3° quartil	8
1° decil	1
9° decil	10

Segundo esse levantamento, metade dos jovens com idades entre 4 e 17 anos assistem à televisão durante 8 horas por dia.

Comentários:

O segundo quartil corresponde ao valor da mediana e ambos valem 4. Então, podemos afirmar que 50% dos jovens assistem à televisão durante, pelo menos, 4 horas por dia.

Gabarito: Errado.

13. (FGV/IBGE/2016) Adotando-se para as estatísticas de posição de uma dada distribuição de frequências as convenções, Q_k = Quartil de ordem k, D_k = Decil de ordem k, Qt_k = Quintil de ordem k e P_k = Pencentil de ordem k, é correto afirmar que:

a) $Q_3 \geq D_6 \geq Qt_4 = P_{80}$;

b) $Qt_2 \leq P_{55} \leq D_6 \leq Q_3$;

c) $D_9 \geq P_{85} \geq Q_3 = Qt_3$;

d) $Q_1 \geq Qt_2 = P_{20} \leq D_3$;

e) $D_6 \leq Q_3 = P_{75} \leq Qt_3$.

Comentários:

Para responder essa questão, devíamos saber que:

a) cada **quartil** delimita 25% das observações de uma distribuição. Logo:

- $Q_1 = 1 \times 25\% = 25\%$ $\Longrightarrow$ até o primeiro quartil temos 25% das observações;
- $Q_3 = 3 \times 25\% = 75\%$ $\Longrightarrow$ até o terceiro quartil temos 75% das observações;

b) cada **quintil** delimita 20% das observações de uma distribuição. Logo:

- $Qt_2 = 2 \times 20\% = 40\% \implies$ até o segundo quintil temos 40% das observações;
- $Qt_3 = 3 \times 20\% = 60\% \implies$ até o terceiro quintil temos 60% das observações;
- $Qt_4 = 4 \times 20\% = 80\% \implies$ até o quarto quintil temos 80% das observações;

c) cada **decil** delimita 25% das observações de uma distribuição. Logo:

- $D_3 = 3 \times 10\% = 30\% \implies$ até o terceiro decil temos 30% das observações;
- $D_6 = 6 \times 10\% = 60\% \implies$ até o sexto decil temos 60% das observações;
- $D_9 = 9 \times 10\% = 90\% \implies$ até o nono decil temos 90% das observações;

d) cada **percentil** delimita 1% das observações de uma distribuição. Logo:

- $P_{20} = 20 \times 1\% = 20\% \implies$ até o vigésimo percentil temos 20% das observações;
- $P_{55} = 55 \times 1\% = 55\% \implies$ até o quinquagésimo quinto percentil temos 55% das observações;
- $P_{75} = 75 \times 1\% = 75\% \implies$ até o septuagésimo quinto percentil temos 75% das observações;
- $P_{80} = 80 \times 1\% = 80\% \implies$ até o octogésimo percentil temos 80% das observações;
- $P_{85} = 85 \times 1\% = 85\% \implies$ até o octogésimo quinto percentil temos 85% das observações;

Agora, vamos avaliar cada uma das alternativas:

- Letra A: **Errada**. Se Qt_4 é maior que 80% das observações e D_6 é superior a 60% das observações, então obrigatoriamente $Qt_4 \geq D_6$;
- Letra B: **Correta**. Se Qt_2 delimita 40% das observações; P_{55} delimita 55%; D_6 delimita 60% e Q_3 delimita 75% das observações, então $Qt_2 \leq P_{55} \leq D_6 \leq Q_3$;
- Letra C: **Errada**. Se Q_3 é maior que 75% das observações; e Qt_3 é maior que 60% das observações, então, eles delimitam porções diferentes do conjunto de dados. Logo, não podemos garantir que sejam iguais entre si;
- Letra D: **Errada**. Se Q_1 delimita 25% das observações e Qt_2 delimita 40% das observações, então $Qt_2 \geq Q_1$;
- Letra E: **Errada**. Se P_{75} é maior que 75% das observações e Qt_3 é maior que 60%, então, $P_{75} \geq Qt_3$.

Gabarito: B.

14. (FCC/TRE RR/2015) A distribuição dos valores dos salários, em dezembro de 2014, dos 200 funcionários em um órgão público é representada por uma tabela de frequências absolutas, com todos os intervalos de classe apresentando a mesma amplitude, sendo fechados à esquerda e abertos à direita. O valor da mediana, obtido pelo método da interpolação linear, foi igual a R$ 5.600,00 e pertencente ao intervalo de classe, em reais, [5.000,00 ; 6.500,00). Se 80 funcionários possuem um salário inferior a R$ 5.000,00, então a porcentagem dos funcionários que apresentam um salário igual ou superior a R$ 6.500,00 é, em %, igual a

a) 45.

b) 30.

c) 50.

d) 25.

e) 35.

Comentários:

Organizando as informações do enunciado temos o seguinte:

- nº de funcionários $\rightarrow 200$;
- mediana $\rightarrow R\$\ 5.600{,}00$;
- frequência acumulada em $R\$\ 5.600 \rightarrow \frac{200}{2} = 100$;
- frequência acumulada em $R\$\ 5.000{,}00 \rightarrow \underbrace{80}_{freq\ acumulada}\ funcionários.$

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:

- a mediana é $M_d = 5.600$.
- o limite inferior da classe é $l_{inf} = 5.000$.
- a frequência acumulada da classe anterior é $f_{ac_{ant}} = 80$.
- a amplitude da classe é $h = 6.500 - 5.000 = 1.500$.

Aplicando a fórmula, temos:

$$5.600 = 5.000 + \left[\frac{\left(\frac{200}{2}\right) - 80}{f_i}\right] \times (6.500 - 5.000)$$

$$600 = \left(\frac{100 - 80}{f_i}\right) \times 1.500$$

$$600 = \frac{20 \times 1.500}{f_i}$$

$$f_i = \frac{20 \times 15\cancel{00}}{6\cancel{00}}$$

$$f_i = 50$$

Portanto, descobrimos que a frequência acumulada em 6.500 é $80 + 50 = 130$. Isso quer dizer que 130 funcionários têm salários inferior a R$ 6.500,00. Vamos montar uma tabela para melhor visualizarmos:

Valor dos salários em (R$)	Frequência acumulada
5000	80
5600	100

6500	130

Ora, se temos um total de 200 funcionários e 130 representa a quantidade que recebe até R$ 6.500,00, então basta subtrairmos 130 de 200 para identificarmos a quantidade de funcionários que ganha acima desse valor. Assim:

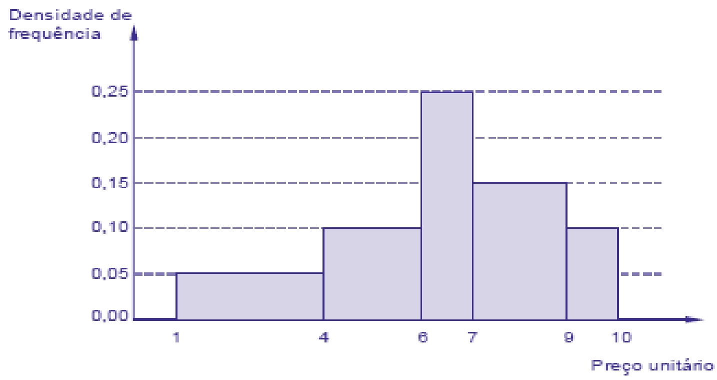
$$200 - 130 = 70$$

Logo, 70 funcionários recebem um salário superior a R$ 6.500,00. Isso representa 35% do total de funcionários:

$$\frac{70}{200} = 0{,}35 \rightarrow 35\%$$

Gabarito: E.

15. (FCC/TRE RR/2015) O histograma abaixo representa a distribuição dos preços unitários de custo, em R$, de determinado equipamento de informática no mercado. No eixo das abscissas constam os intervalos de classe, em R$, e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)^{-1}$.



Observação:
Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.

Considerando os intervalos de classe fechados à esquerda e abertos à direita, se 105 preços apresentam valores menores que R$ 6,00, então o número de preços que apresentam valores iguais ou superiores a R$ 4,00 é

a) 240.

b) 195.

c) 215.

d) 230.

e) 255.

Comentários:

Vamos analisar o histograma para entendermos a questão.

Tomando o primeiro intervalo (1 a 4), temos que essa classe representa 15% dos preços:

$$(4 - 1) \times 0{,}05 = 0{,}15 \; ou \; 15\%$$

Se pegarmos o intervalo de (1 a 6) temos que as duas primeiras classes representam 35% dos preços:

$$(6-4) \times 0{,}1 + (4-1) \times 0{,}05 = 0{,}35 \; ou \; 35\%$$

O enunciado afirma que 105 preços apresentam valores menores que R$ 6,00. Além disso, já descobrimos que a frequência acumulada até o valor de R$ 6,00 é de 35%. Portanto, 105 preços correspondem a uma frequência acumulada de 35%.

Também já sabemos que 15% dos preços são inferiores a R$ 4,00. Queremos descobrir o número de preços que apresentam valores iguais ou superiores a R$ 4,00. Sendo assim, podemos denominar de x o número de preços iguais ou superiores a esse valor, que correspondem a um percentual de 85% (= 100% - 15%).

Fazendo uma regra de três simples, temos:

$$105 \rightarrow 35\%$$
$$x \rightarrow 85\%$$
$$\frac{105}{x} = \frac{35}{85} \rightarrow 35x = 8925 \rightarrow x = 255$$

Gabarito: E.

16. (CESPE/BACEN/2013)

2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários. Com base nesses dados, julgue o próximo item.

A mediana é maior que o 50º percentil.

Comentários:

A mediana divide os dados em duas partes de mesma frequência. Assim, a mediana separa os 50% menores dos 50% maiores. O 50º percentil faz exatamente a mesma coisa. Portanto:

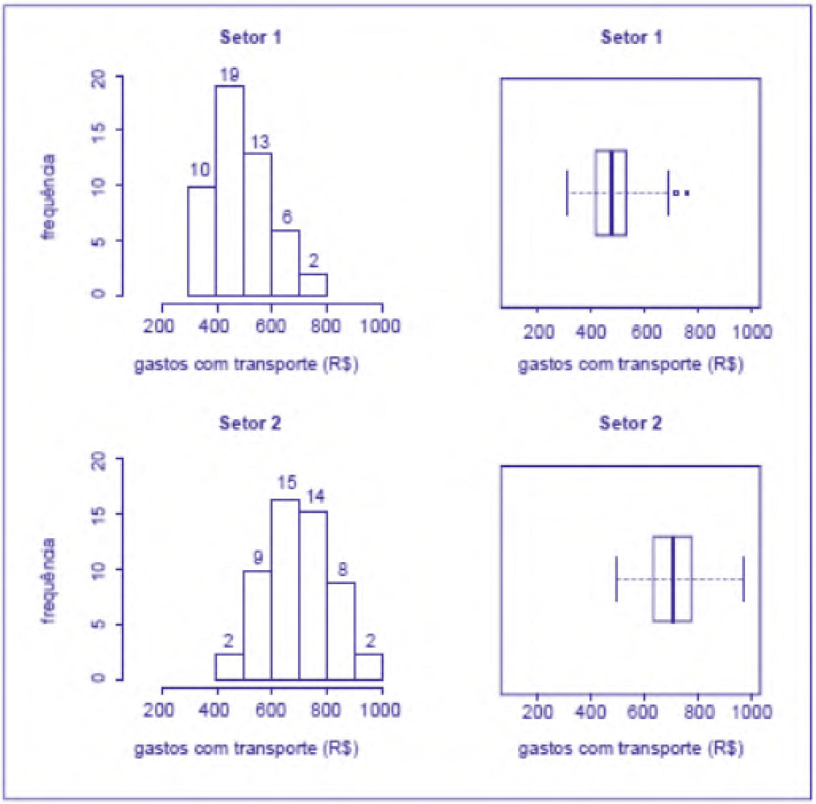
$$M_d = P_{50}.$$

Ainda, é importante lembrarmos que esses valores também coincidem com o segundo quartil e com o quinto decil. Portanto:

$$Q_2 = M_d = D_5 = P_{50}$$

Gabarito: Errado

17. (CESPE/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores 1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
723,79	767,53				

Setor 2

488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56
663,81	670,12	671,90	673,78	684,69	685,98

693,35	**698,58**	**708,78**	**719,80**	**721,16**	**734,84**
735,94	**746,34**	**754,83**	**756,10**	**756,96**	**760,80**
762,29	**766,24**	**770,11**	**797,73**	**804,06**	**805,97**
807,29	**832,83**	**844,00**	**866,77**	**878,27**	**897,09**
943,10	**963,25**				

Considerando essas informações, julgue o item.

Trinta por cento das despesas dos servidores do setor 1 correspondem a um valor superior a R$ 525,00.

Comentários:

Observamos que no setor 1 há exatamente 15 elementos superiores a 525:

526,42	528,76	531,53	547,91	572,66
591,43	596,99	609,44	632,15	639,71
677,48	683,76	688,76	723,79	767,53

De um total de 50 elementos, isso representa:

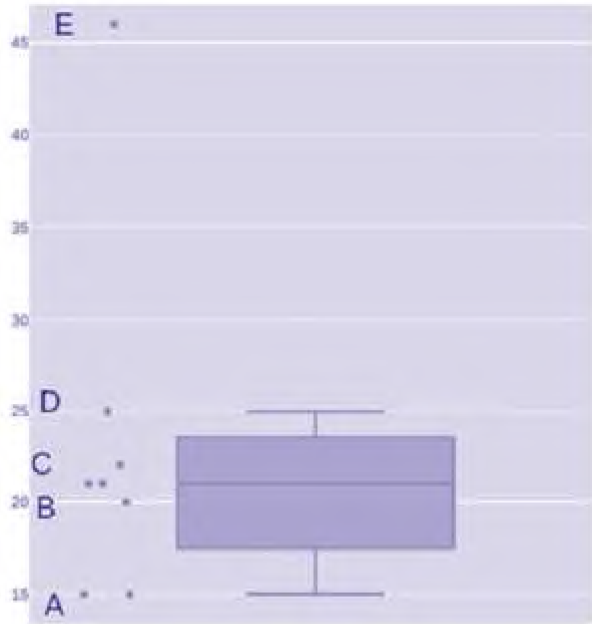
$$\frac{15}{50} = 0{,}3 \; ou \; 30\%$$

Gabarito: Certo.

Questões Comentadas

Box Plot

1. (CESPE/DPE RO/2022)



No gráfico boxplot anteriormente apresentado, o outlier do conjunto de dados é representado pelo ponto

a) A.

b) E.

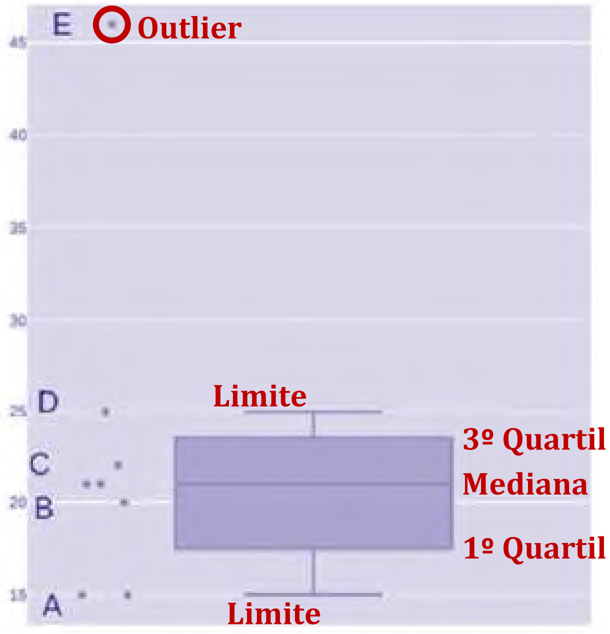
c) B.

d) C.

e) D.

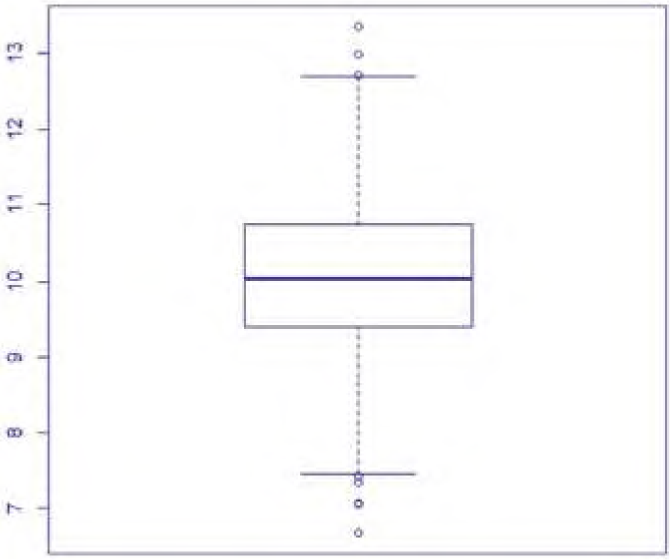
Comentários:

Os *outliers* são observações discrepantes do conjunto de dados. No diagrama de *boxplot*, os outliers aparecem abaixo do limite inferior ou acima do limite superior do *boxplot* e, geralmente, são representados por pontos, círculos ou asteriscos. Conforme mostra a figura a seguir, o ponto E representa um *outlier* do conjunto, enquanto os pontos A e D indicam os limites inferior e superior do *boxplot*.



Gabarito: B.

2. (CESPE/PETROBRAS/2022)

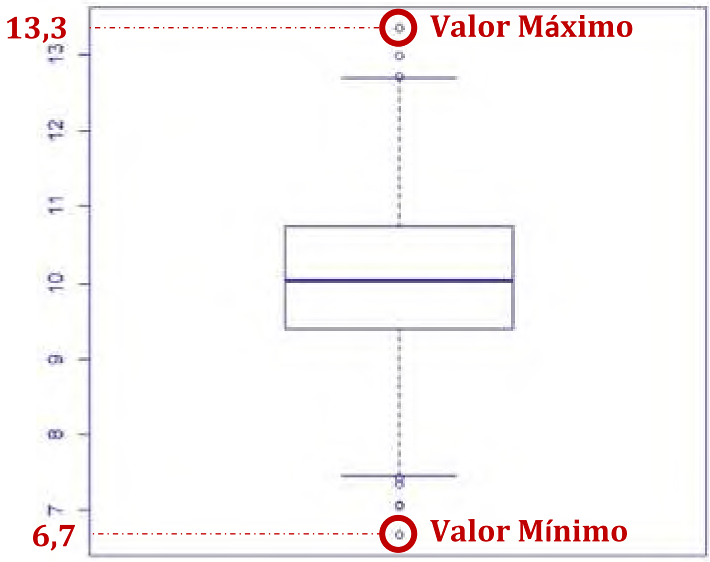


Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

A amplitude total dos dados em tela é inferior a 6.

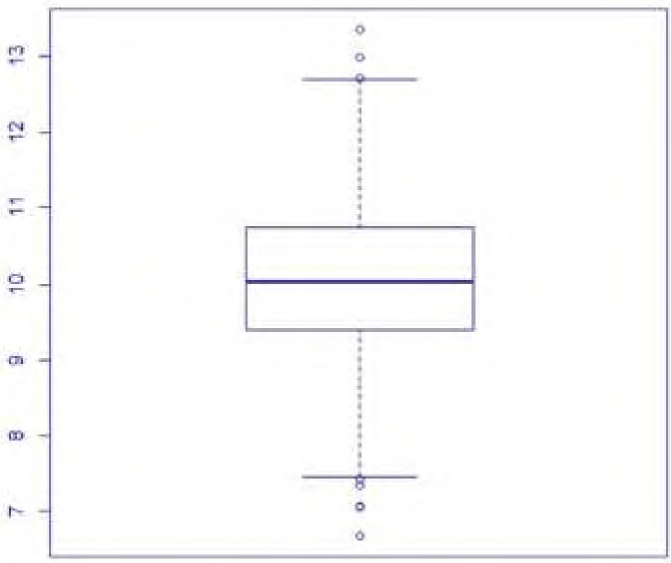
Comentários:

A amplitude total é superior a 6, pois o valor mínimo é inferior a 7 e o valor máximo é superior a 13, resultando em uma diferença superior a 6. Vejamos:



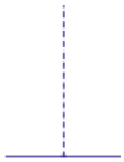
Gabarito: Errado.

3. (CESPE/PETROBRAS/2022)



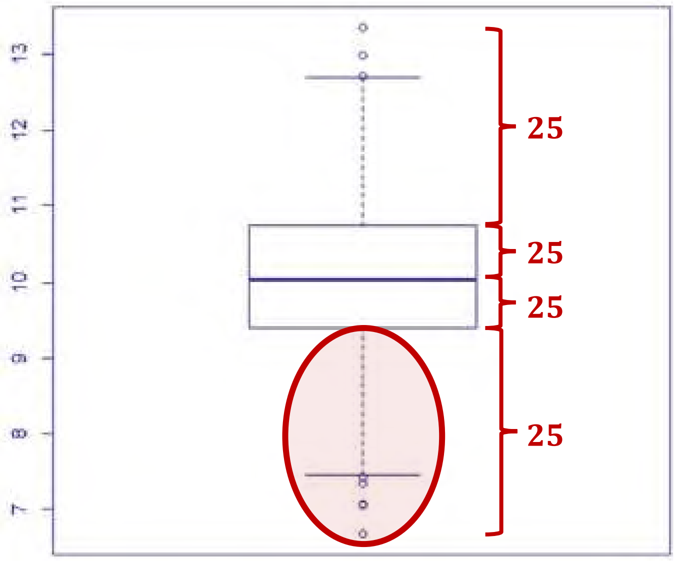
Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

Nesse diagrama, a porção da distribuição dos dados representada pela parte inferior do diagrama mostrada a seguir representa exatamente 25% dos dados em questão.



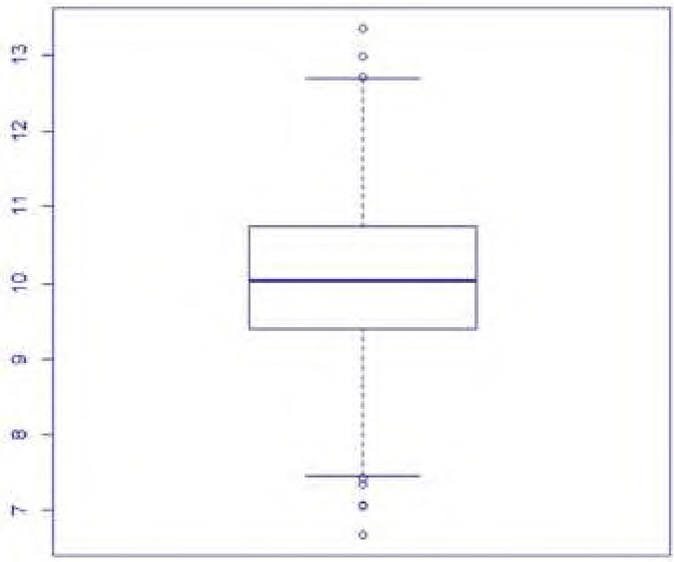
Comentários:

Em um diagrama de boxplot, as linhas no retângulo (caixa) indicam os quartis. Da menor observação até o primeiro quartil, temos 25% dos dados. Do terceiro quartil até a maior observação, também temos 25% dos dados. Essa regra continua válida na presença de valores discrepantes. Vejamos:



Gabarito: Errado.

4. (CESPE/PETROBRAS/2022)

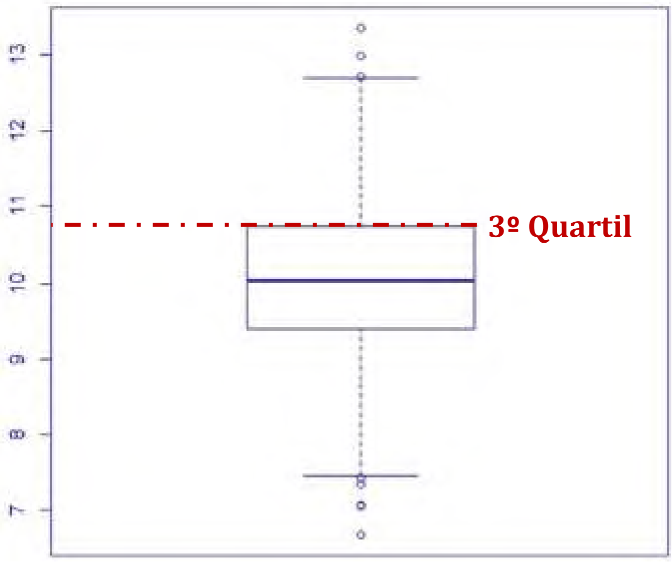


Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

O terceiro quartil é inferior a 11 e superior a 10.

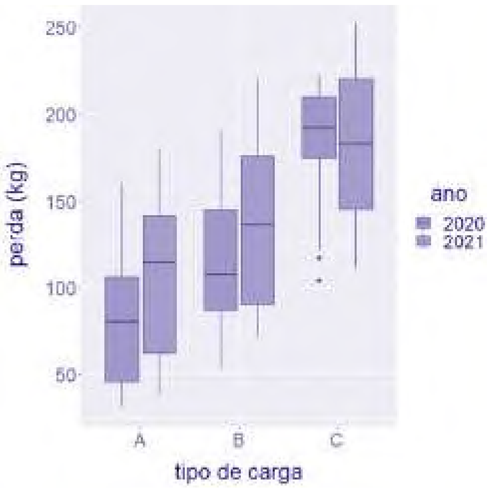
Comentários:

Como podemos observar na imagem abaixo, é exatamente isso: o terceiro quartil está localizado no intervalo entre 10 e 11.



Gabarito: Correto.

5. (CESPE/PETROBRAS/2022)



Considerando a figura precedente, que mostra desenhos esquemáticos das distribuições das quantidades de cargas perdidas nos anos de 2020 e 2021, segundo o tipo de carga transportada por uma mineradora, julgue o item que se segue.

No que se refere à distribuição da quantidade de carga do tipo B perdida em 2021, observa-se que o valor da perda mínima foi superior a Q_1 - 1,5Dq.

Comentários:

Conforme observamos na figura a seguir, o primeiro quartil é representado pela base da caixa; enquanto o topo representa o terceiro quartil. Para a carga do tipo B em 2021, os valores dos quartis são: $Q_1 = 90\ kg$ e $Q_3 = 175\ kg$.

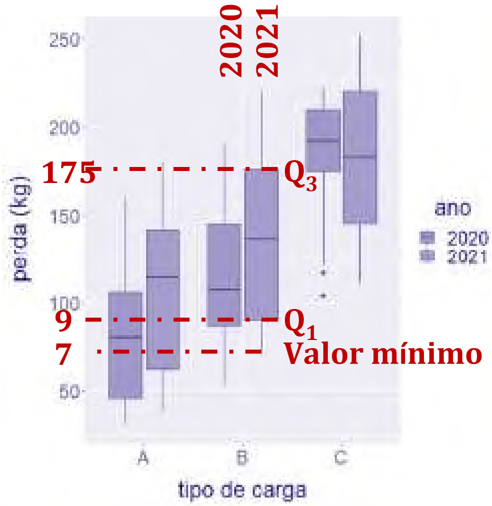
O intervalo interquartil é a diferença entre o terceiro e o primeiro quartil. Temos, portanto:

$$Dq = 175 - 90 = 85\ kg.$$

Calculando o valor de $Q_1 - 1{,}5Dq$:

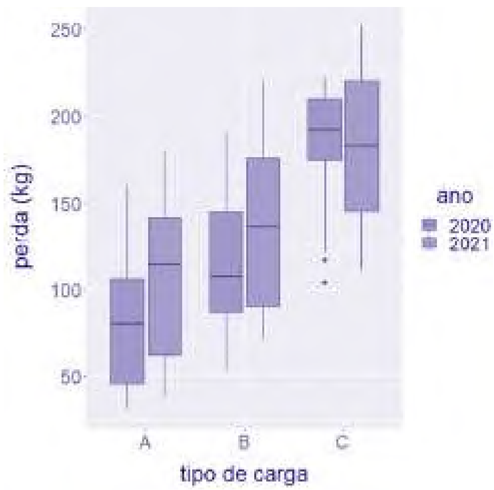
$$90 - 1{,}5 \times 85 = -37{,}5$$

A perda mínima da carga B em 2021 ficou próximo de 73 kg, portanto, um valor superior ao calculado acima.



Gabarito: Errado.

6. (CESPE/PETROBRAS/2022)

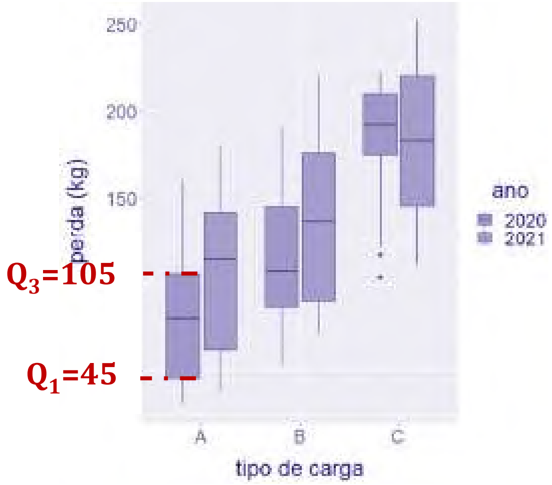


Considerando a figura precedente, que mostra desenhos esquemáticos das distribuições das quantidades de cargas perdidas nos anos de 2020 e 2021, segundo o tipo de carga transportada por uma mineradora, julgue o item que se segue.

Na distribuição da quantidade de carga do tipo A perdida em 2020, observa-se que o primeiro quartil foi superior a 100 kg, enquanto o terceiro quartil foi inferior a 50 kg.

Comentários:

Conforme podemos observar na imagem abaixo, a assertiva inverteu os valores dos quartis. Na distribuição da quantidade de carga do tipo A perdida em 2020, o primeiro quartil ficou próximo de 45 kg e o terceiro quartil foi de, aproximadamente, 105 kg. Portanto, o primeiro quartil foi inferior a 50 kg e o terceiro quartil foi superior a 100kg.



Gabarito: Errado.

7. (CESPE/TELEBRAS/2022) Com respeito ao conjunto de dados {0, 0, 1, 1, 1, 3}, julgue o item que se segue.

Se esse conjunto de dados fosse representado por um diagrama de box-plot, então os valores 0 e 3 seriam chamados valores exteriores, ou, ainda, discrepantes, atípicos ou outliers.

Comentários:

Os valores menores que $Q_1 - 1{,}5 \times DIQ$ ou maiores que $Q_3 + 1{,}5 \times DIQ$ são considerados valores discrepantes (*outliers*). Sendo assim, inicialmente, precisamos calcular os quartis:

$$P_{Q_1} = \frac{1}{4} \times 6 = 1{,}5 \Rightarrow Q_1 = 0$$

$$P_{Q_3} = \frac{3}{4} \times 6 = 4{,}5 \Rightarrow Q_3 = 1$$

Agora, calculamos a distância interquartílica:

$$DIQ = Q_3 - Q_1 = 1 - 0 = 1$$

Calculando os limites para detecção de outliers:

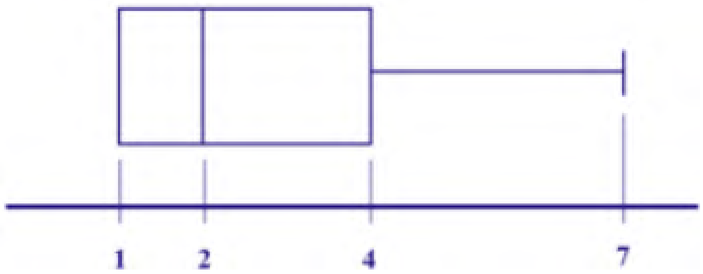
$$Q_1 - 1{,}5 \times DIQ = 1 - 1{,}5 \times 1 = -0{,}5$$

$$Q_3 + 1{,}5 \times DIQ = 1 + 1{,}5 \times 1 = 2{,}5$$

Portanto, valores menores que -0,5 ou maiores que 2,5 são considerados outliers.

Gabarito: Errado.

8. (CESPE/TCE-RJ/2021)



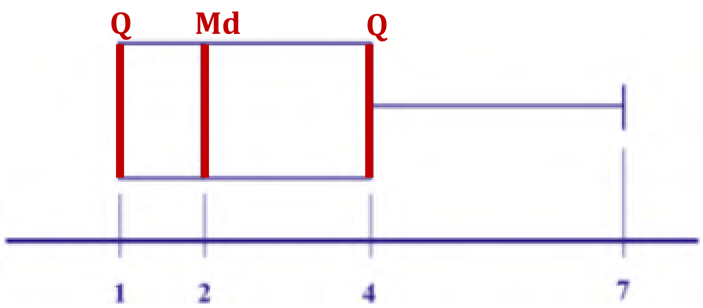
Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

A mediana da variável X é igual a 4.

Comentários:

Um diagrama de boxplot é uma ferramenta gráfica frequentemente utilizada na análise exploratória de dados que permite visualizar a distribuição dos dados e os valores discrepantes (outliers). Essa ferramenta resume cinco medidas descritivas de um conjunto de dados: valor mínimo, primeiro quartil, mediana, terceiro quartil e valor máximo.

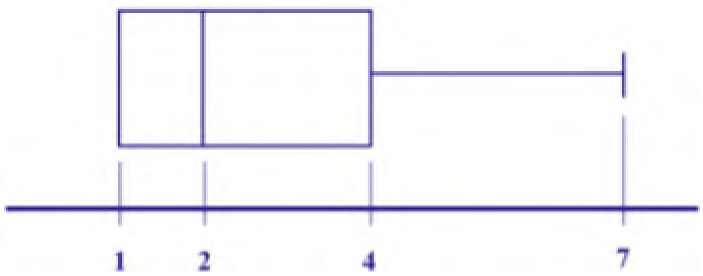
Observando o diagrama apresentado na questão, podemos identificar o primeiro quartil, a mediana, e o terceiro quartil, representados respectivamente pelos traços em 1, 2 e 4, conforme mostra a figura a seguir:



Logo, a mediana da variável x é igual a 2.

Gabarito: Errado.

9. (CESPE/TCE-RJ/2021)

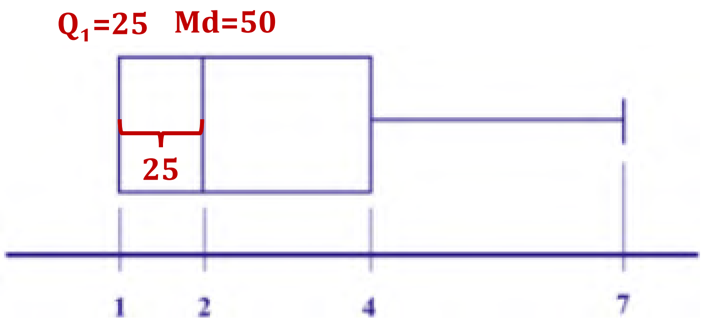


Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

Se f denota a frequência relativa do valor 1, então 0,25 ≤ f < 0,50.

Comentários:

No diagrama de boxplot, são apresentados o primeiro, segundo e terceiro quartis. Eles correspondem, respectivamente, ao primeiro, segundo e terceiro traços da caixa, sendo que a mediana corresponde ao segundo quartil:



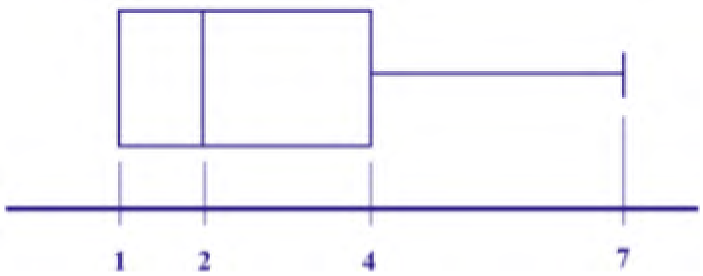
Como a mediana tem valor igual a 2, a frequência relativa f do valor 1 necessariamente precisa ser inferior a 50%, isto é, $f < 50\%$. Caso esse valor ocorresse em 50% ou mais dos casos, a própria mediana seria valor igual a 1 e o diagrama de boxplot ficaria representado de forma diferente.

Além disso, o primeiro quartil delimita 25% dos dados ordenados. No caso em questão, como o limite inferior coincide com o primeiro quartil, vez que não há o bigode esquerdo no diagrama de boxplot, significa dizer que não existem valores menores que 1 e, portanto, $f \geq 25\%$.

Logo, temos que $25\% \leq f < 50\%$.

Gabarito: Errado.

10. (CESPE/TCE-RJ/2021)

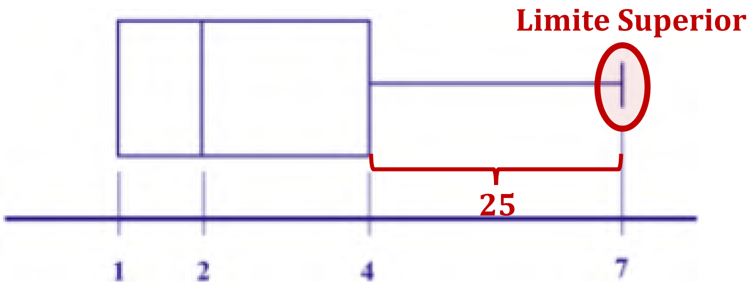


Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

$\frac{1}{3}$ das observações da variável X são iguais a 7.

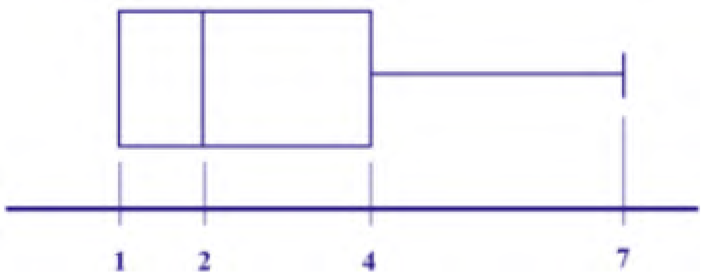
Comentários:

Conforme observamos no diagrama de boxplot, o valor 7 corresponde ao limite superior (e valor máximo) da distribuição da variável X. O valor 4, por sua vez, indica o terceiro quartil da distribuição, separando os 75% menores valores dos 25% maiores. Sabendo disso, podemos concluir que no máximo 25% das observações são iguais a 7:



Gabarito: Errado.

11. (CESPE/TCE-RJ/2021)

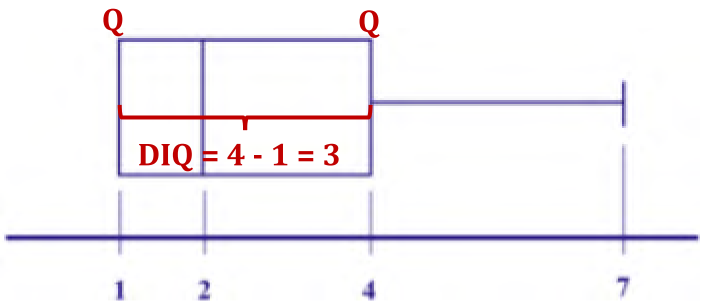


Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

O diagrama *boxplot* indica que o intervalo interquartil (ou interquartílico) da distribuição da variável X é igual a 3.

Comentários:

No diagrama de boxplot, estão representados o primeiro, segundo e terceiro quartis, sendo que a mediana corresponde ao segundo quartil:



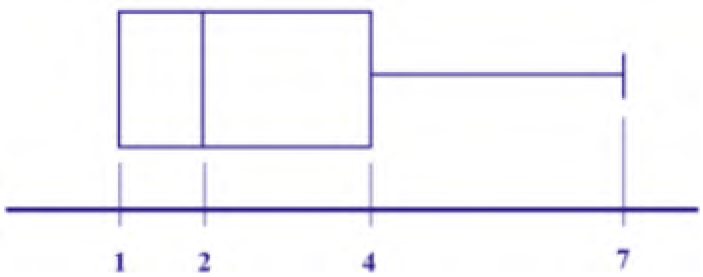
O intervalo interquartílico é definido como a diferença entre o terceiro e o primeiro quartil:

$$Q_3 - Q_1 = 4 - 1 = 3$$

Portanto, o intervalo interquartílico da distribuição de X é igual a 3.

Gabarito: Certo.

12. (CESPE/TCE-RJ/2021)

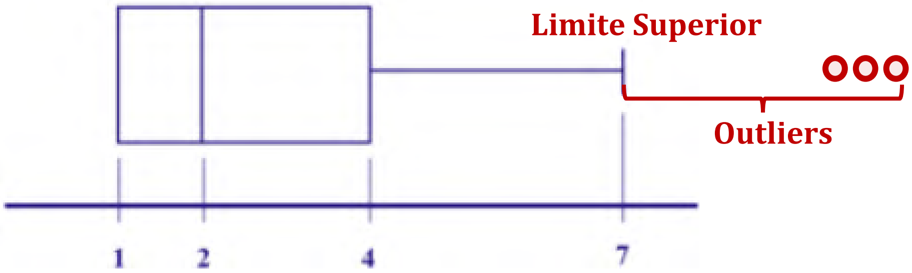


Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

As observações da variável X que assumem valores iguais a 7, com base nesse diagrama boxplot, são considerados *outliers.*

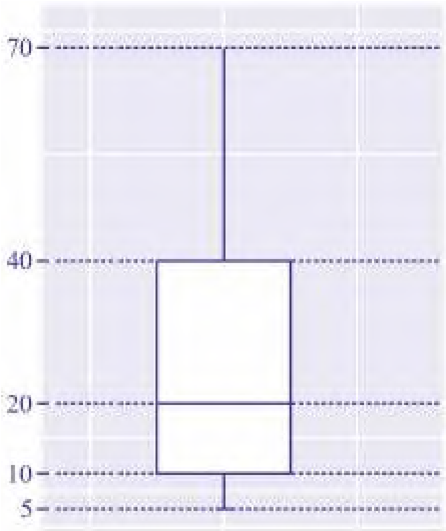
Comentários:

Conforme observamos no diagrama de boxplot, o valor 7 corresponde ao limite superior (e valor máximo) da distribuição da variável X. Os outliers são valores discrepantes (atípicos), que ocorrem acima do limite superior ou abaixo do limite superior. Portanto, para serem considerados outliers, os valores devem ser superiores a 7.



Gabarito: Errado.

13. (CESPE/TJ PA/2020)



Considerando que o desenho esquemático (boxplot) antecedente se refere a uma variável quantitativa X, assinale a opção correta.

a) O intervalo interquartil é igual a 65.

b) Metade da distribuição da variável X se encontra entre os valores 20 e 40.

c) Os valores da variável X que se encontram no intervalo [5;10] representam 5% da distribuição de X.

d) A mediana de X é igual a 25.

e) O primeiro quartil da distribuição de X é igual a 10.

Comentários:

Vamos analisar cada alternativa:

Alternativa A: **Incorreta**. O intervalo interquartil corresponde à diferença entre o terceiro e o primeiro quartil, sendo que, para a distribuição em questão, seu valor é igual a 30:

$$DIQ = Q_3 - Q_1$$

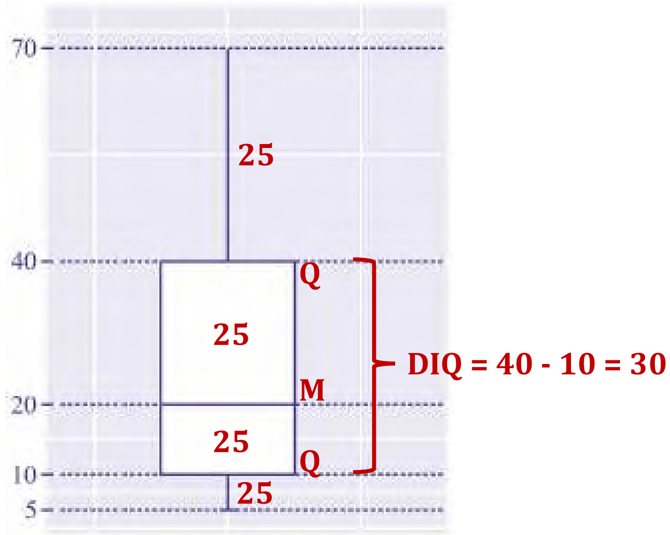
$$DIQ = 40 - 10 = 30$$

Alternativa B: **Incorreta**. Como a mediana é igual a 20, metade da distribuição se encontra entre os valores 20 e 70; e outra metade entre 5 e 20;

Alternativa C: **Incorreta**. Entre 5 e 10 temos a distância do limite inferior ao primeiro quartil da distribuição. Como os quartis dividem uma distribuição em 4 partes iguais (25% da distribuição), temos 25% da distribuição da variável X no intervalo em questão;

Alternativa D: **Incorreta**. A mediana de X é igual a 25;

Alternativa E: **Correta**. O primeiro quartil da distribuição de X é igual a 10, conforme apresenta a figura a seguir:



Gabarito: E.

14. (CESPE/ME/2020) Acerca de visualização e análise exploratória de dados, julgue o item seguinte.

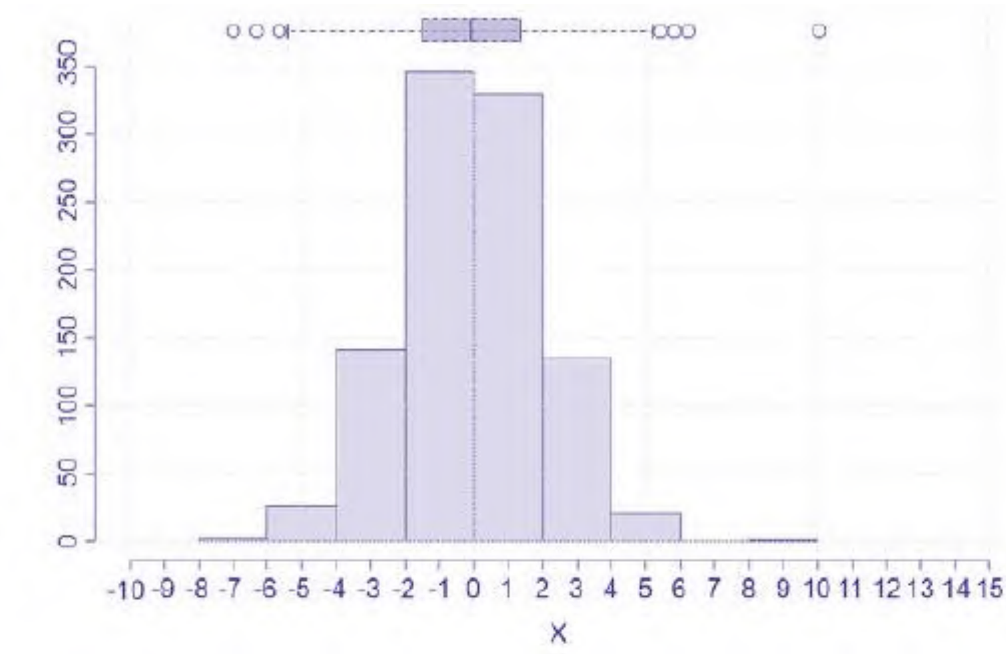
Outlier ou anomalias são padrões nos dados que não estão de acordo com uma noção bem definida de comportamento normal.

Comentários:

Os outliers são valores atípicos que escapam ao comportamento normal da distribuição e, a depender da análise, podem ser descartados para que se consiga um melhor ajuste do modelo. Normalmente, valores menores que $Q_1 - 1{,}5 \times DIQ$ ou maiores que $Q_3 + 1{,}5 \times DIQ$ são considerados discrepantes (*outliers*).

Gabarito: Certo.

15. (CESPE/ME/2020)

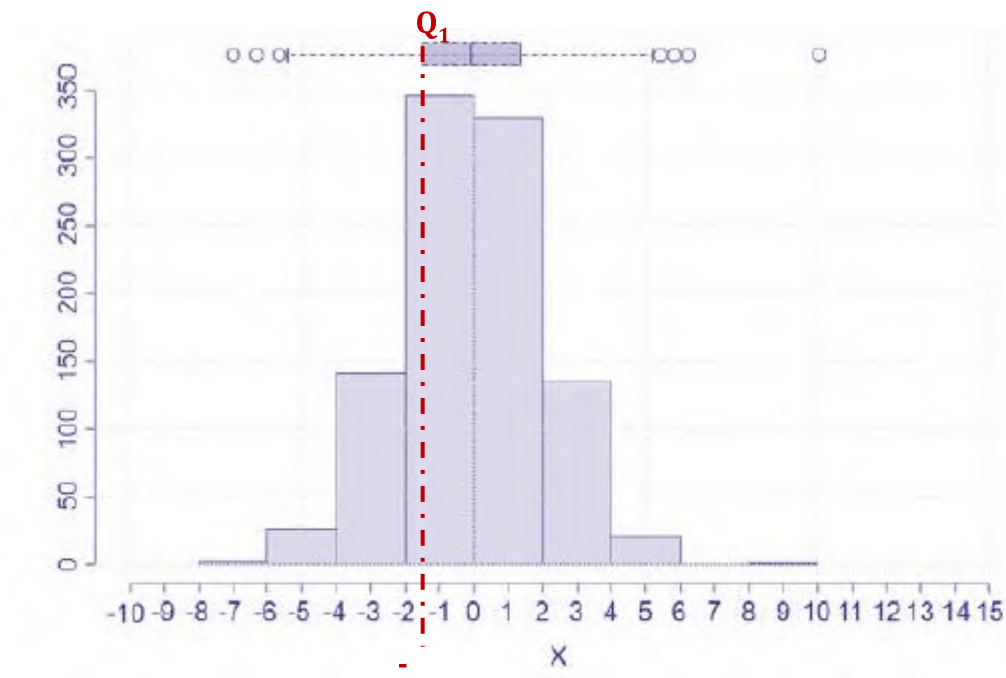


Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

O primeiro quartil da distribuição de X é inferior a −2.

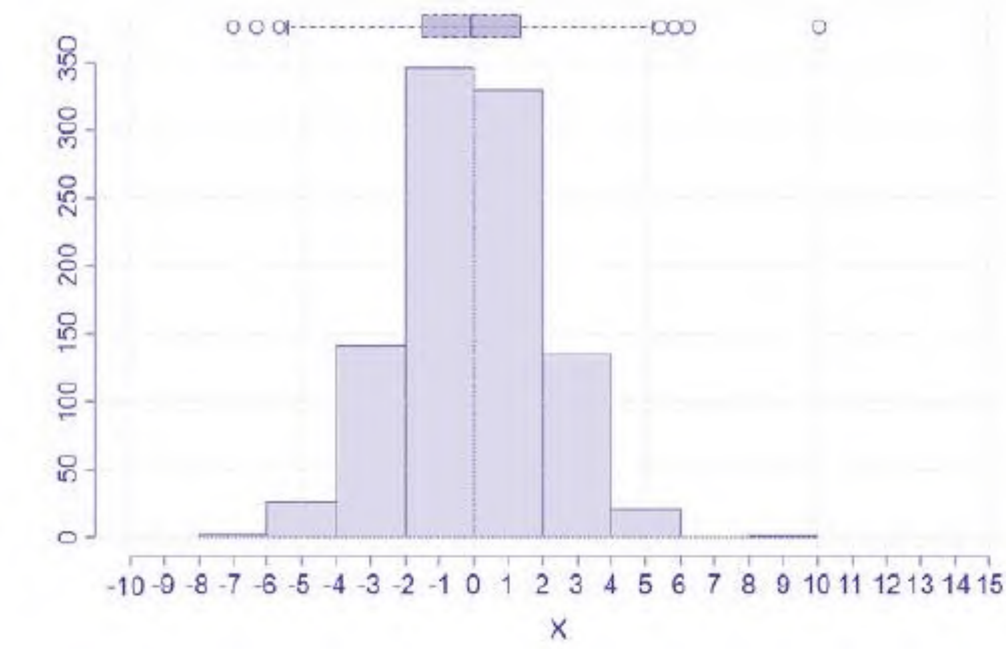
Comentários:

Com base no diagrama de boxplot existente na parte superior do histograma, podemos concluir que o primeiro quartil é a linha que inicia a caixa do diagrama. Portanto, como observamos na figura a seguir, o primeiro quartil da distribuição de X é superior a -2:



Gabarito: Errado.

16. (CESPE/ME/2020)

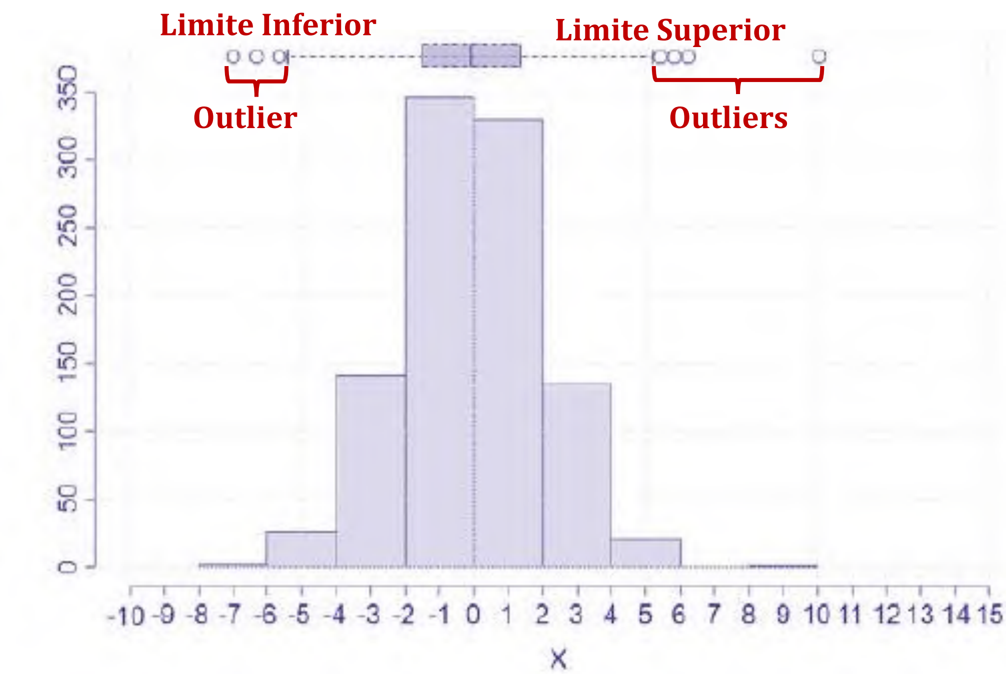


Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

No diagrama de *boxplot*, os pontos indicados pelo símbolo ○ representam *outliers.*

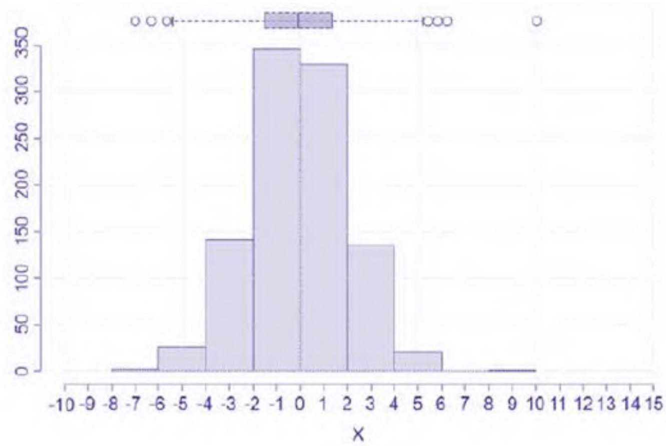
Comentários:

Os outliers são valores atípicos que escapam ao comportamento normal da distribuição, situando-se abaixo do limite inferior ou acima do limite superior do *boxplot*. No diagrama em questão, estão representados como círculos (pontos), encontrando-se à esquerda do limite inferior e à direita do limite superior:



Gabarito: Certo.

17. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

O percentual das observações da variável X que se encontram no intervalo [−2,+2] é igual ou superior a 50%.

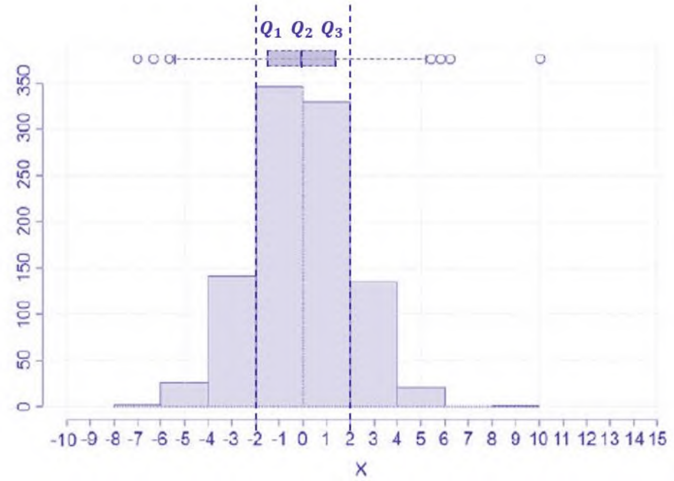
Comentários:

Em um diagrama de *boxplot*, temos:

- até o primeiro quartil: 25% dos dados;
- entre o primeiro e o segundo quartil: 25% dos dados;
- entre o segundo e o terceiro quartil: 25% dos dados; e

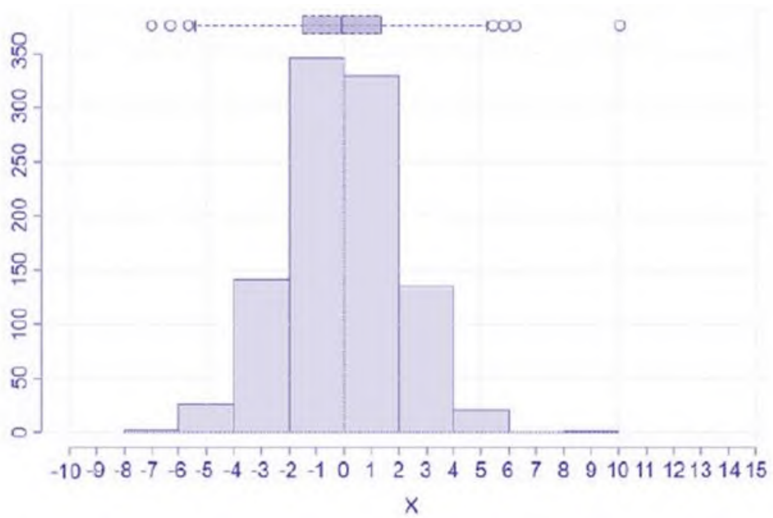
- após o terceiro quartil: 25% dos dados.

Portanto, o retângulo entre Q1 e Q3 representa 50% dos dados. No gráfico apresentado, notamos que o retângulo está contido no intervalo [−2,+2], contemplando, dessa forma, um percentual de observações superior a 50%. Vejamos a imagem a seguir:



Gabarito: Certo.

18. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

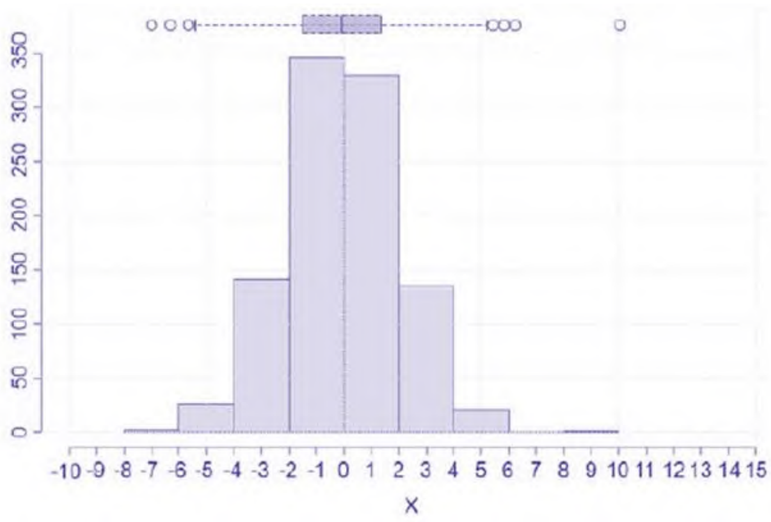
O comprimento da caixa do diagrama de *boxplot*, representado pela figura , é obtido multiplicando-se o desvio padrão de X por 2.

Comentários:

A questão trata de distância interquartílica ou intervalo interquartílico, que é equivalente à subtração entre o terceiro quartil e o primeiro quartil (Q3-Q1).

Gabarito: Errado.

19. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

A mediana de X se encontra no intervalo [−1,+1].

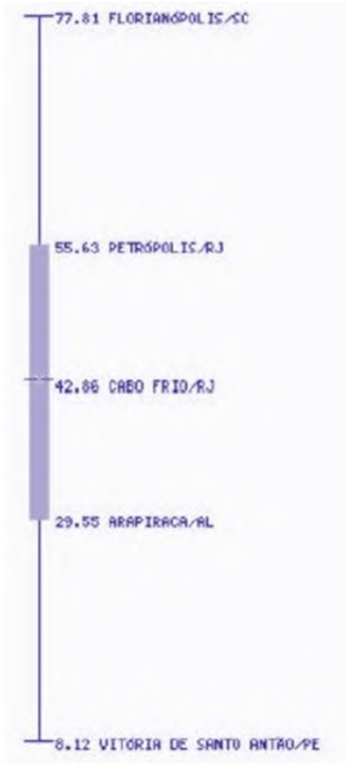
Comentários:

Em um diagrama de boxplot, a linha no interior do retângulo representa o segundo quartil, que é equivalente à mediana dos dados. Podemos verificar que a mediana (ou segundo quartil) encontra-se exatamente no intervalo [−1,+1].

Gabarito: Certo.

20. (CESPE/ME/2020) Acerca de visualização e análise exploratória de dados, julgue o item seguinte.

O gráfico apresentado a seguir é denominado caixa de barra.

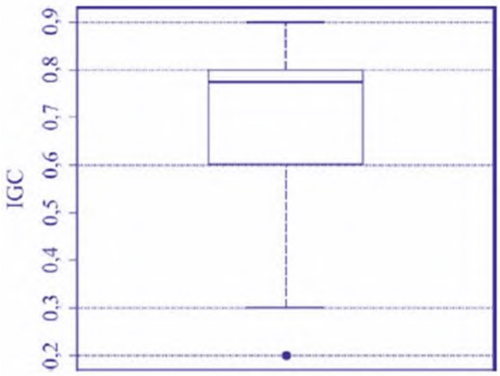


Comentários:

O nome do diagrama apresentado é **boxplot** ou **gráfico de caixa** ou **diagrama de caixa**.

Gabarito: Errado.

21. (CESPE/TCE-PR/2016)



Com base na figura antecedente, que apresenta a distribuição dos indicadores de governança corporativa (IGC) observados em uma amostra de empresas prestadoras de serviços terceirizados, assinale a opção correta.

a) O menor e o maior IGC observados na amostra foram, respectivamente, iguais a 0,3 e 0,9.

b) O diagrama mostrado na figura em questão é denominado curva de frequência.

c) O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

d) Na amostra considerada, a mediana dos indicadores observados foi inferior a 0,7.

e) A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

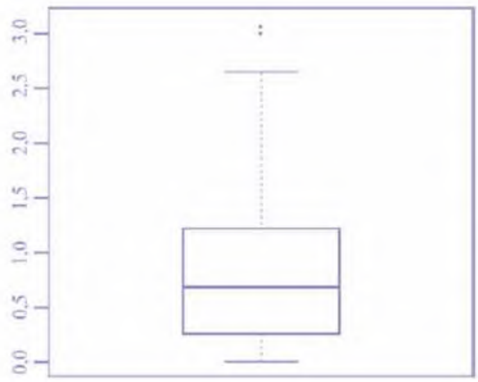
Comentários:

Vamos analisar as alternativas:

- Letra A: Errada. Observamos que o limite superior é 0,9, portanto, esse é o maior valor. O limite inferior é 0,3, porém, há um valor discrepante (outlier), simbolizando por um ponto, marcando o valor de 0,2, sendo esse o menor valor do boxplot.
- Letra B: Errada. O diagrama apresentado é um boxplot ou gráfico de caixa ou diagrama de caixa.
- Letra C: Errada. Observando o diagrama verificamos que o primeiro quartil vale 0,6, a caixa no centro do diagrama indica os quartis.
- Letra D: Errada. No segundo quartil do diagrama observamos a mediana, que, no presente caso, é superior a 0,7.
- Letra E: Correta. O ponto que vale 0,2 realmente é uma observação atípica (outlier) das demais. Item certo.

Gabarito: E.

22. (CESPE/TCE-PA/2016)



média amostral	0,80
desvio padrão amostral	0,70
primeiro quartil	0,25
mediana	0,70
terceiro quartil	1,20
mínimo	0
máximo	3,10

Um indicador de desempenho X permite avaliar a qualidade dos processos de governança de instituições públicas. A figura mostra, esquematicamente, a sua distribuição, obtida mediante estudo amostral feito por determinada agência de pesquisa. A tabela apresenta estatísticas descritivas referentes a essa distribuição.

Com base nessas informações, julgue o item a seguir.

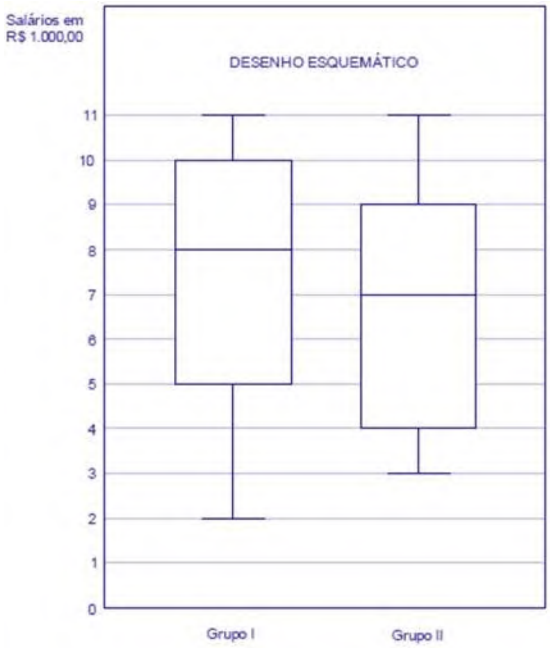
O diagrama boxplot mostrado na figura sugere a existência de pelo menos duas observações atípicas.

Comentários:

Geralmente, observações atípicas no diagrama boxplot são representados por pontos fora dos quartis. É o que observamos no diagrama em questão, em que temos dois valores discrepantes (pontos) na parte superior do diagrama.

Gabarito: Certo.

23. (FCC/TRT 3ª Região/2015) Seja uma representação gráfica de dados de acordo com o desenho esquemático abaixo (box-plot) que foi preparado para comparar todos os salários dos funcionários do sexo masculino (Grupo I) com todos os salários dos funcionários do sexo feminino (Grupo II) lotados em um órgão público.



Neste desenho esquemático

a) O número de elementos do Grupo I é superior ao número de elementos do Grupo II.

b) O módulo da diferença entre as medianas dos 2 grupos é igual a 25% do menor salário deste órgão público.

c) Mais da metade dos elementos do Grupo I possui um salário inferior a R$ 5.000,00 ou superior a R$ 10.000,00.

d) O valor do menor salário do Grupo II corresponde a 37,5% do valor da mediana do Grupo I.

e) A diferença interquartil do Grupo I é superior à diferença interquartil do Grupo II.

Comentários:

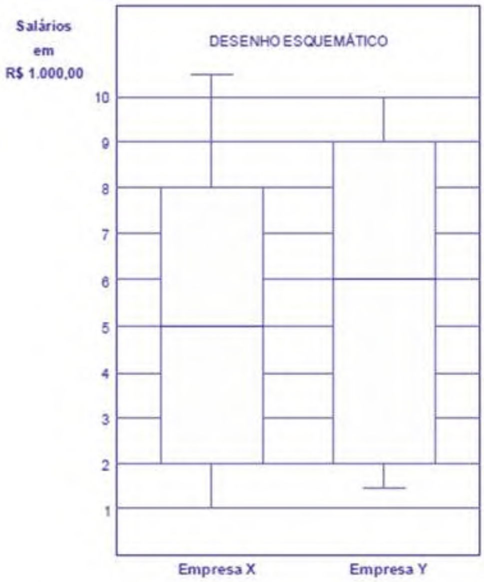
Vamos analisar cada uma das alternativas:

- Letra A: **Errada**. O diagrama não mostra a quantidade de elementos de cada grupo, portanto, não podemos afirmar que o Grupo I tem mais elementos que o Grupo II.
- Letra B: **Errada**. O diagrama mostra que as medianas valem 8 e 7 para cada grupo, portanto, a diferença entre elas é igual a 1. Podemos observar também que o menor salário é 2. Logo, a diferença entre as medianas corresponde a 50% do menor salário.
- Letra C: **Errada**. Se separarmos o Grupo I em quartis, cada quartil corresponderá a 25% do total de elementos. Somando o último e o primeiro quartil teremos 50%, exatamente a metade, dos elementos. Portanto, não podemos afirmar que mais da metade dos elementos do Grupo I possui um salário inferior a R$ 5.000,00 ou superior a R$ 10.000,00.
- Letra D: **Correta**. O menor salário do Grupo II vale 3. A mediana do Grupo I vale 8. Então, $\frac{3}{8} = 0{,}375 = 37{,}5\%$.

- Letra E: **Errada**. Para o primeiro grupo $Q_3 = 10,\ Q_1 = 5 \rightarrow Q_3 - Q_1 = 10 - 5 = 5$ para o segundo grupo $Q_3 = 9\ , Q_1 = 4 \rightarrow Q_3 - Q_1 = 9 - 4 = 5$. Portanto, as distâncias são iguais.

Gabarito: D.

24. (FCC/TRT 19ª Região/2014) Para comparar os salários dos empregados de duas empresas X e Y, considerou-se o desenho esquemático abaixo com os valores dos salários em R$ 1.000,00.



De acordo com o desenho esquemático apresentado, é correto afirmar que

a) O número de empregados da empresa X é maior que o número de empregados da empresa Y.

b) A distância interquartil da empresa Y é superior à distância interquartil da empresa X.

c) O menor salário verificado tanto na empresa X como na empresa Y é igual a R$ 2.000,00.

d) O valor da mediana da empresa Y supera o da empresa X em 10%.

e) O maior salário verificado na empresa X é inferior ao maior salário verificado na empresa Y.

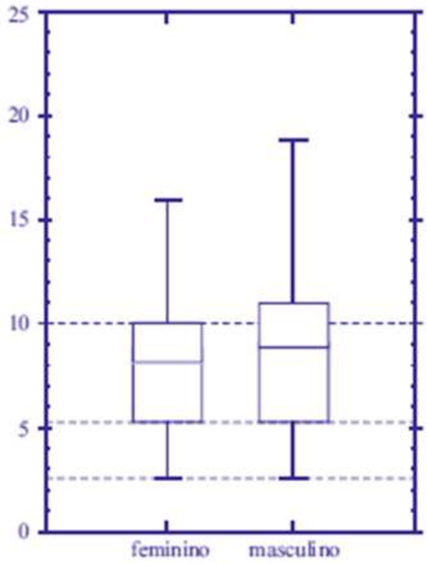
Comentários:

Vamos analisar cada uma das alternativas:

- Letra A: **Errada**. O diagrama não mostra a quantidade de empregados de cada empresa, portanto, não podemos afirmar que a empresa X tem mais empregados que a empresa Y.
- Letra B: **Correta**. A distância interquartil é dada pela diferença entre Q_3 e Q_1. Para a empresa Y, temos: $Q_3 = 9, Q_1 = 2 \rightarrow Q_3 - Q_1 = 9 - 2 = 7$; para a empresa X, temos: $Q_3 = 8\ , Q_1 = 2 \rightarrow Q_3 - Q_1 = 8 - 2 = 6$. Logo, $Y > X$.
- Letra C: **Errada**. O limite inferior das duas empresas está abaixo dos 2 mil reais, e não igual como afirma a alternativa.
- Letra D: **Errada**. A mediana da empresa X vale 5 e a da empresa Y vale 6. Então, $\frac{6}{5} = 1{,}20$. Logo, Y supera X em 20%.
- Letra E: **Errada**. O diagrama mostra que o limite superior de X é maior que 10.000, e o limite superior de Y é 10.000.

Gabarito: B.

25. (CESPE/PRF/2012)



M. M. B. Paolielo et al.. In: ***Saúde Pública****, 1997 (com adaptações).*

Em decorrência do desenvolvimento urbano e tecnológico, tem-se a preocupação de monitorar os efeitos nocivos da poluição ambiental sobre a saúde da população urbana. A figura acima mostra o diagrama de caixa (box-plot) que descreve a distribuição da concentração de chumbo no sangue, em µg.dL-1, obtida com base em uma amostra aleatória de 200 pessoas do sexo masculino e de 100 pessoas do sexo feminino que trabalham em postos de combustível localizados em determinado município brasileiro.

Com base nessas informações, julgue o item.

Com base nas linhas horizontais que cortam as caixas do diagrama apresentado, conclui-se corretamente que a média das concentrações de chumbo encontradas no sangue das pessoas do sexo feminino que trabalham em postos de combustível do referido município brasileiro é inferior à média das concentrações dessa mesma substância no sangue das pessoas do sexo masculino que trabalham nesses postos de combustível.

Comentários:

O diagrama boxplot fornece informações de medidas de separatrizes, distribuição de dados, assimetria, limites superior e inferior, mas não fornece informações sobre a média propriamente dita.

Gabarito: Errado.

LISTA DE QUESTÕES

Mediana

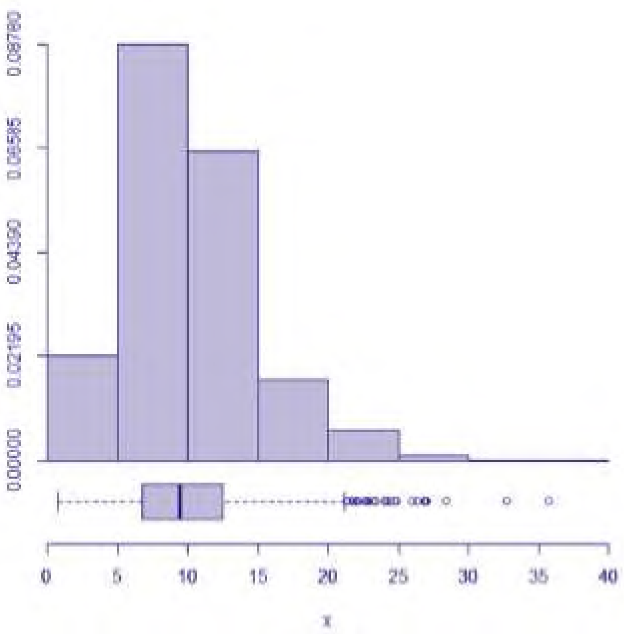
1. (CESPE/PETROBRAS/2022)

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 obtida com base em uma amostra aleatória simples de tamanho igual a n, julgue o item que se segue.

A mediana de x é igual a 1,5.

2. (CESPE/PETROBRAS/2022) A figura seguinte mostra o histograma como uma estimativa da função de densidade de uma distribuição X, juntamente com o diagrama boxplot correspondente a esse conjunto de dados.

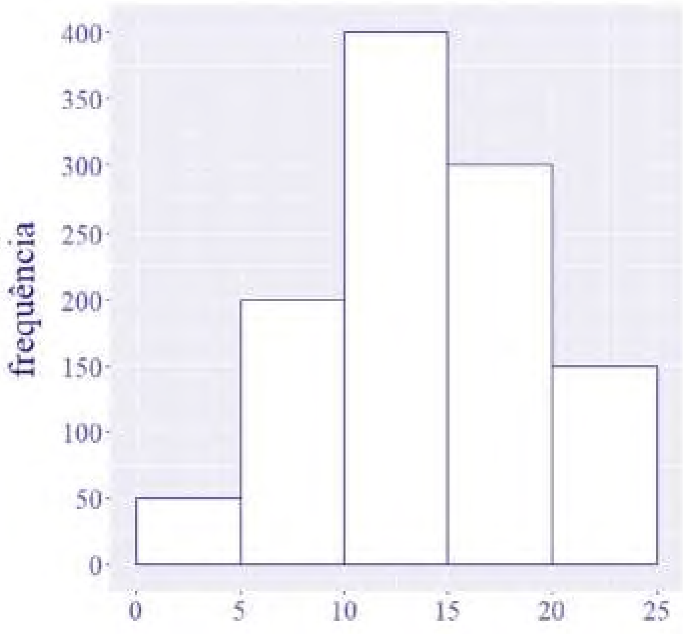


Tamanho da amostra	1.000
Média amostral	10
Desvio padrão amostral	4,7

Considerando a figura e as informações apresentadas no quadro, julgue o item que se segue.

A diferença entre a média amostral e a mediana amostral é superior a 0.

3. (CESPE/TELEBRAS/2022)



Considerando que o histograma apresentado descreve a distribuição de uma variável quantitativa X por meio de frequências absolutas, julgue o item que se segue.

A mediana de X se encontra na classe modal.

4. (FGV/CBM-AM/2022) A soma de 11 números inteiros estritamente positivos, não necessariamente distintos, é 2022.

O maior valor que a mediana desses 11 números pode ter é

a) 335.

b) 336.

c) 337.

d) 338.

e) 339.

5. (FGV/SEFAZ ES/2022) As notas de nove candidatos num certo exame foram:

54, 48, 46, 51, 38, 50, 44, 58, 32.

A mediana dessas notas é igual a

a) 44.

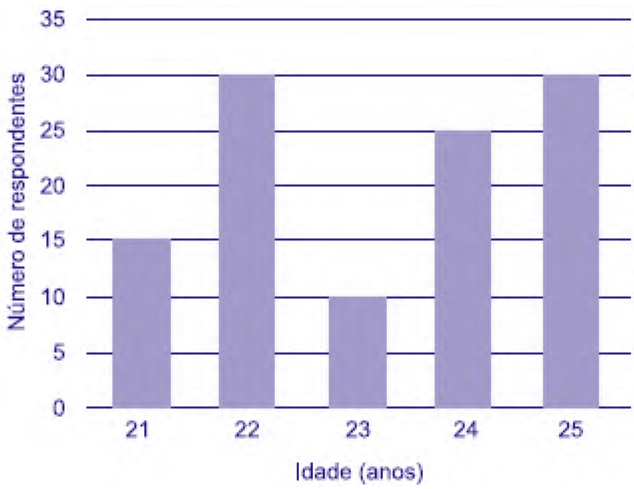
b) 46.

c) 48.

d) 50.

e) 51.

6. (VUNESP/PM-SP/2022) Uma pesquisa foi realizada com um grupo de pessoas cujas idades, em anos, pertencem ao conjunto {21, 22, 23, 24, 25, 26}. O gráfico registra as frequências absolutas dos entrevistados com menos de 26 anos.



Sabendo que a mediana das idades do conjunto completo de dados (incluindo as pessoas com 26 anos) é igual a 24 anos, o número máximo de pessoas com 26 anos que participaram da pesquisa foi

a) 19.

b) 25.

c) 35.

d) 49.

e) 55.

7. (AOCP/PC PA/2021) 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) $P(X < 1{,}70) = P(X > 1{,}70) = 50\%$.

b) $P(X = 1{,}70) = 50\%$.

c) $P(X > 1{,}70) = 75\%$.

d) $P(X < 1{,}70) = 25\%$.

e) $P(X < 1{,}70) = P(X > 1{,}70) = 25\%$.

8. (CESGRANRIO/CEF/2021) Seis candidatos, aprovados para a penúltima etapa de um processo seletivo, foram submetidos a um teste de conhecimentos gerais com 10 itens do tipo "verdadeiro/falso". Os dois primeiros candidatos acertaram 8 itens cada, o terceiro acertou 9, o quarto acertou 7, e os dois últimos, 5 cada. Pelas regras do concurso, passariam, para a etapa final da seleção, os candidatos cujo número de acertos fosse maior ou igual à mediana do número de acertos dos seis participantes.

Quantos candidatos passaram para a etapa final?

a) 2

b) 3

c) 4

d) 5

e) 6

9. (CESPE/PG-DF/2021)

Estatística	X, em R$ milhões
Mínimo	0,5
Primeiro quartil (Q_1)	1
Segundo quartil (Q_2)	2
Terceiro quartil (Q_3)	5
Máximo	20

O quadro apresentado mostra estatísticas descritivas produzidas por um estudo acerca de despesas públicas (X, em R$ milhões) ocorridos no ano de 2019 em uma amostra aleatória simples de 100 contratos.

Com base nessas informações, julgue o item que se segue.

A mediana da variável *X* foi igual a R$ 2 milhões.

10. (CESPE/CBM AL/2021) Em determinado dia, em uma região atendida por uma unidade do corpo de bombeiros, ocorreram 16 acidentes, que resultaram em 48 vítimas, socorridas pelos bombeiros nos próprios locais de acidente. Entre essas vítimas, 4 vieram a óbito no momento do atendimento, e as demais sobreviveram.

Com base nessa situação hipotética, julgue o item a seguir.

Suponha que as idades das vítimas que vieram a óbito sejam 12, 50, 30 e 20 anos de idade. Nesse caso, a mediana das idades é maior que 26 anos.

11. (CESPE/BANESE/2021) A respeito do conjunto de dados {11, 6, 28, 51, 49, 32, 33}, julgue o item a seguir.

A mediana desse conjunto de dados é igual a 51.

12. (FGV/IMBEL/2021) Considere a lista de cinco números reais: 2, 9, 4, 10, x.

Sabe-se que a mediana desses números é igual à média deles.

A soma dos possíveis valores de x é:

a) 22,5.

b) 21,25.

c) 20,75.

d) 19,5.

e) 17,5.

13. (FGV/FunSaúde CE/2021) Sabe-se que x é maior do que 11 e que a diferença entre a média e a mediana dos cinco números 2, x, 11, 16, 5 é igual a 2.

O valor de x é

a) 12.

b) 16.

c) 21.

d) 26.

e) 31.

14. (FGV/FunSaúde CE/2021) A mediana dos sete números 9, 2, 5, 3, 13, x, 5 é x.

A média desses números é

a) 5.

b) 5,5.

c) 6.

d) 6,5.

e) 7.

15. (IBFC/IBGE/2021) Marcos pretende determinar a mediana referente aos dados brutos coletados e relacionados abaixo:

23 - 22 - 21 - 22 - 32 - 33

41 - 21 - 20 - 32 - 42 - 38

De acordo com os dados, o resultado encontrado por Marcos é igual a:

a) 37

b) 33

c) 41

d) 27,5

e) 28

16. (CESPE/ME/2020)

Mínimo	5
Moda	9
Mediana	9
Média	10
Máximo	15

Um levantamento amostral proporcionou as estatísticas precedentes, referentes a determinada variável quantitativa X.

Considerando essas informações e que a variável X é composta por 1240 observações, julgue o item subsequente.

A quantidade de observações da variável X maiores ou iguais a 9 é igual ou superior a 620.

17. (CESPE/SEFAZ DF/2020) A partir de uma amostra aleatória simples de tamanho n, sabe-se que a média aritmética de uma variável X foi igual a 3. Considerando que os valores possíveis para a variável X sejam -1 e +4, julgue o item que se segue.

A mediana amostral da variável X foi igual a 2,5.

18. (FUNDATEC/Pref. Porto Alegre/2020) Se a mediana de um determinado processo for igual a 7, isso quer dizer que:

a) O 7º valor da amostra representará a mediana.

b) O 7º valor da amostra ordenada representará a mediana.

c) A posição mediana é 7.

d) 50% dos valores da amostra são iguais a 7.

e) 50% dos valores da amostra são menores ou iguais a 7.

19. (VUNESP/EsFCEx/2020) A tabela de frequências relativas acumuladas a seguir refere-se à distribuição dos salários dos empregados de uma empresa, sendo que não foi fornecida a correspondente frequência relativa acumulada do terceiro intervalo de classe (denotado por X na tabela).

CLASSE DE SALÁRIOS (R$)	FREQUÊNCIAS RELATIVAS ACUMULADAS (f_i)
$\mathbf{1.000,00 \vdash 3.000,00}$	10
$\mathbf{3.000,00 \vdash 5.000,00}$	30
$\mathbf{5.000,00 \vdash 7.000,00}$	X
$\mathbf{7.000,00 \vdash 9.000,00}$	80
$\mathbf{9.000,00 \vdash 11.000,00}$	100

Dado que o valor da mediana dessa distribuição obtida pelo método da interpolação linear apresentou um valor igual a R$ 6.600,00, obtém-se que X é igual a

a) 65.

b) 50.

c) 60.

d) 55.

e) 70.

20. (VUNESP/Pref Ilhabela/2020) Uma pesquisa permitiu obter do mercado a tabela de frequências relativas abaixo, correspondente à distribuição dos preços de venda (p) referente a um determinado equipamento. As frequências relativas da 1a e 2a classes por algum motivo não foram fornecidas e foram denotadas na tabela por x e y, respectivamente.

Classes de preços em R$	Frequências relativas (%)
$50 < p \leq 150$	x
$150 < p \leq 250$	y
$250 < p \leq 350$	25
$350 < p \leq 450$	30
$450 < p \leq 550$	15
TOTAL	100

Utilizando o método da interpolação linear, encontra-se que o valor da mediana é igual a

a) R$ 290,00.

b) R$ 300,00.

c) R$ 310,00.

d) R$ 320,00.

e) R$ 330,00.

21. (VUNESP/Pref Sorocaba/2020) Uma empresa de varejo decidiu usar o seguinte método com funcionários em treinamento: calcularia a média de vendas de todos os funcionários após um mês e só seriam efetivados os que estivessem acima da média, sendo demitidos aqueles que estivessem abaixo. Assim, os diretores da empresa imaginavam que ficariam com cerca de metade dos funcionários em treinamento. No entanto, ao final do treinamento, apenas um funcionário estava acima da média.

No método empregado pela empresa, o que foi determinante para que a média esperada de funcionários aprovados não fosse atingida?

a) Não se levou em conta a pequena variância nas vendas.

b) A média utilizada foi a geométrica.

c) Obviamente foi um erro de conta no cálculo da média.

d) O resultado deveria ter sido padronizado.

e) Deveria ter sido usada a mediana em vez da média.

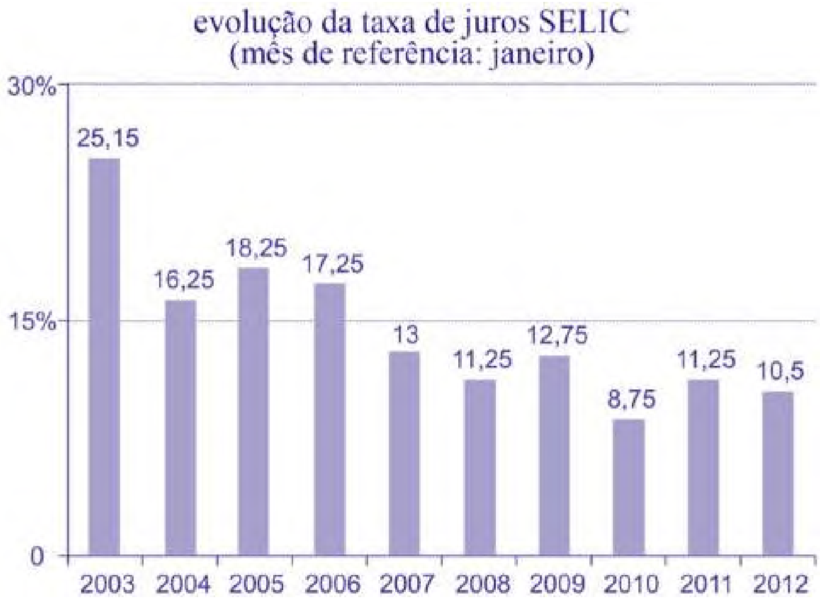
22. (CESPE/Pref. São Cristóvão/2019) A tabela seguinte mostra a distribuição das idades dos 30 alunos da turma A do quinto ano de uma escola de ensino fundamental.

Idade (em anos)	9	10	11	12	13	14
Quantidade de estudantes	6	22	0	1	0	1

A partir dessa tabela, julgue o item.

A mediana das idades é igual a 11,5 anos.

23. (CESPE/UNCISAL/2019) A SELIC é uma taxa referencial de juros estabelecida pelo Banco Central do Brasil como parâmetro para as taxas de juros cobradas pelos bancos comerciais no Brasil. A tabela seguinte mostra a evolução da SELIC, em porcentagem, no mês de janeiro dos anos de 2003 a 2012.



Disponível em: www.bcb.gov.br. Acesso em: nov. 2016 (adaptado).

O valor da mediana dos valores da SELIC mostrados no gráfico é igual a

a) 11,25.

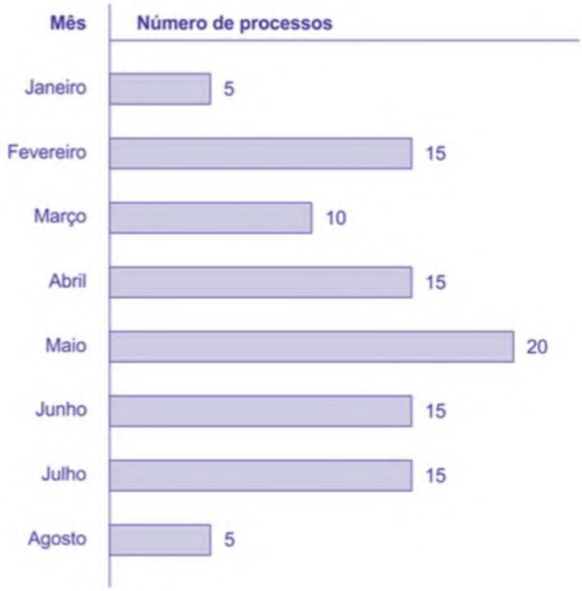
b) 12,125.

c) 12,875.

d) 13,00.

e) 14,44.

24. (FCC/BANRISUL/2019) Os números de processos com uma determinada característica autuados em um órgão público, de janeiro a agosto de 2018, podem ser visualizados pelo gráfico abaixo.



A respectiva média aritmética (número de processos por mês) está para a mediana assim como

a) 1 está para 16.

b) 2 está para 3.

c) 1 está para 8.

d) 5 está para 6.

e) 4 está para 3.

25. (FCC/Prefeitura do Recife/2019) A empresa Sigma apresenta pela tabela abaixo a distribuição dos salários registrados de seus 100 empregados em reais.

Salários (R$)	2.000	4.000	5.000	10.000	15.000	Total
Número de Empregados	0	10	40	X	Y	100

Não foram fornecidos os números de empregados que ganham R$ 10.000,00 e R$ 15.000,00 (denotados na tabela por x e y, respectivamente), mas sabe-se que a média aritmética dos salários é igual a R$ 8.400,00. O valor da soma da respectiva moda e da respectiva mediana desses salários é, em reais, igual a

a) 600y.

b) 625y.

c) 1.000y.

d) 750y.

e) 500y.

26. (FCC/Prefeitura do Recife/2019) Com o objetivo de analisar a distribuição dos salários dos empregados de uma empresa, verificou-se que 10 empregados ganham, cada um, R$ 15.000,00; 20 ganham, cada um, R$ 2.500,00; 25 ganham, cada um, R$ 12.000,00; 60 ganham, cada um, R$ 5.000,00 e os restantes ganham, cada um, R$ 8.000,00. Sabendo-se que a mediana dos salários apresentou um valor igual a R$ 6.500,00, obtém-se que o valor da média aritmética supera o da moda em

a) R$ 2.750.00.

b) R$ 3.250,00.

c) R$ 3.000,00.

d) R$ 2.250,00.

e) R$ 2.500,00.

27. (FCC/Prefeitura do Recife/2019) Durante 40 dias, foi registrado o número de pessoas atendidas por dia em um guichê de uma repartição. A tabela abaixo apresentou os dados observados sendo que não foram fornecidas as quantidades de dias em que foram atendidas uma pessoa por dia e duas pessoas por dia, indicadas na tabela por $q1$ e $q2$, respectivamente.

Número de pessoas atendidas	Quantidade de dias
0	9
1	q_1
2	q_2
3	5
4	1
Total	**40**

Sabendo-se que a mediana correspondente foi igual 1,5, tem-se que a soma da moda e da média aritmética (número de pessoas atendidas por dia) foi igual a

a) 3,00.

b) 2,80.

c) 3,45.

d) 3,20.

e) 2,95.

28. (FUNDATEC/ESE/2019) A taxa de desemprego na grande São Paulo nos últimos 10 anos, em percentual da população, é dada pela tabela abaixo:

ANO	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
TOTAL	13,8	11,9	10,5	10,9	10,4	10,8	13,2	16,8	18,0	17,3

Nesse caso, podemos dizer que a mediana é:

a) 10,60.

b) 11,09.

c) 10,40.

d) 12,55.

e) 13,36.

29. (VUNESP/MP-SP/2019) Considere o seguinte conjunto de dados numéricos para estatística.

8	30	5	6	4	8	12	6	13	8

Então, a soma da moda com a mediana e a média é igual a:

a) 22.

b) 24.

c) 26.

d) 28.

e) 30.

30. (VUNESP/PM-SP/2019) Na tabela, são apresentadas informações sobre o número de armas apreendidas pela Polícia Militar do Estado de São Paulo, no segundo semestre de 2018.

MÊS	NÚMERO DE ARMAS APREENDIDAS
Julho	688
Agosto	702
Setembro	680
Outubro	638
Novembro	695
Dezembro	629

(http://www.policiamilitar.sp.gov.br)

Se uma pesquisa utilizar a média aritmética simples do número de armas apreendidas, mensalmente, no segundo semestre, pela Polícia Militar do Estado de São Paulo, e outra pesquisa utilizar a mediana do número de armas apreendidas no segundo semestre, a diferença entre a mediana e a média será de

a) 30 armas.

b) 21 armas.

c) 12 armas.

d) 10 armas.

e) 08 armas.

31. (VUNESP/Pref. Cerquilho/2019) A tabela apresenta informações sobre o número de funcionários em um escritório e os salários que são pagos a eles. Utilize as informações para responder a questão seguinte.

NÚMERO DE FUNCIONÁRIOS	SALÁRIOS
4	R$ 1.200,00
2	R$ 1.500,00
3	R$ 1.800,00
1	R$ 2.100,00

A mediana dos salários dos funcionários desse escritório é igual a

a) R$ 2.100,00.

b) R$ 1.800,00.

c) R$ 1.700,00.

d) R$ 1.500,00.

e) R$ 1.200,00.

32. (VUNESP/TRANSERP/2019) Considerando a taxa de desemprego hipotética de 11,7%(maio); 10,3% (junho); 8,9% (julho); 7,7% (agosto); 6,3% (setembro) e 4,9% (outubro) a mediana do período é

a) 8,3%

b) 7,9%

c) 7,7%

d) 7,5%

e) 7,2%

33. (VUNESP/TJ-SP/2019) Durante um período, decidiu-se analisar o comportamento do número de processos especiais autuados por dia em uma repartição pública. A tabela a seguir apresenta os resultados obtidos, sendo k a quantidade de dias em que não foram autuados processos.

NÚMERO DE PROCESSOS	0	1	2	3	4	5	TOTAL
QUANTIDADE DE DIAS	k	14	18	24	14	2	10k

Com relação a esta tabela, foram obtidos os respectivos valores da moda (Mo), mediana (Md) e média aritmética (Me), em número de processos por dia. Verifica-se então que (Mo + Md + Me) é igual a

a) 6,30

b) 7,85

c) 6,80

d) 6,85

e) 7,35

34. (CESPE/Polícia Federal/2018)

	DIA				
	1	2	3	4	5
X (quantidade diária de drogas apreendidas, em kg)	10	22	18	22	28

Tendo em vista que, diariamente, a Polícia Federal apreende uma quantidade X, em kg, de drogas em determinado aeroporto do Brasil, e considerando os dados hipotéticos da tabela precedente, que apresenta os valores observados da variável X em uma amostra aleatória de 5 dias de apreensões no citado aeroporto, julgue o item.

A mediana das quantidades X observadas na amostra em questão foi igual a 18 kg.

35. (FCC/ALESE/2018) Em um grupo de pessoas encontramos as seguintes idades: 20, 30, 50, 39, 20, 25, 41, 47, 36, 45, 41, 52, 18, 41. A mediana e a moda são, respectivamente,

a) 36 e 45.

b) 40 e 41.

c) 41 e 20.

d) 42 e 39.

e) 39 e 42.

36. (FCC/ISS-São Luís/2018) Um levantamento foi realizado com 40 instituições financeiras, localizadas em uma região, com relação às taxas mensais de juros aplicadas para financiamento de veículos. Verificou-se que cinco instituições aplicam a taxa de 0,80% ao mês, duas aplicam a taxa de 1,20% ao mês, oito aplicam a taxa de 1,25% ao mês, x aplicam a taxa de 1,12% ao mês e y aplicam a taxa de 0,96% ao mês. Se a média aritmética destas taxas foi igual a 1,05%, então a soma da mediana e a moda correspondentes foi de

a) 2,00%

b) 2,24%

c) 2,08%

d) 2,16%

e) 1,92%

37. (FCC/SEDU-ES/2018) As notas dos dez alunos de uma sala foram: 1, 2, 4, 6, 6, 7, 8, 8, 8, 10. A diferença entre a moda e a mediana dessas notas é

a) 1,5.

b) 2,5.

c) 0,5.

d) 2,0.

e) 1,0.

38. (FCC/TRT 2ª Região/2018) Considerando na tabela abaixo a distribuição de frequências absolutas, referente aos salários dos n empregados de uma empresa, em R$ 1.000,00, observa-se que além do total dos empregados (n) não é fornecida também a frequência correspondente ao intervalo da 4ª classe (f_4).

Classes de salários (R$ 1.000,00)	$1 \vdash 3$	$3 \vdash 5$	$5 \vdash 7$	$7 \vdash 9$	$9 \vdash 11$	Total
Frequências	4	8	10	f_4	2	n

O valor da média aritmética destes salários, obtido considerando que todos os valores incluídos num certo intervalo de classe são coincidentes com o ponto médio deste intervalo, é igual a R$ 6.200,00. O valor da mediana em R$, obtido pelo método da interpolação linear, é igual a

a) 400,0 f_4

b) 412,5 f_4

c) 387,5 f_4

d) 350,0 f_4

e) 375,0 f_4

39. (FGV/ALE-RO/2018) Sejam x, y e z, respectivamente, a média, a mediana e a moda dos sete valores 9, 10, 6, 5, 20, 9 e 4. É correto concluir que

a) x < y < z .

b) x < y = z

c) x = y < z

d) y < z = x

e) x = y = z

40. (FGV/ALE-RO/2018) A tabela a seguir mostra o número de gols sofridos por um time de futebol nas dez primeiras partidas de um campeonato:

Jogo	1	2	3	4	5	6	7	8	9	10
Gols Sofridos	0	1	2	0	1	2	1	0	3	2

A média e a mediana do número de gols sofridos nesses jogos são respectivamente

a) 1,2 e 1,0.

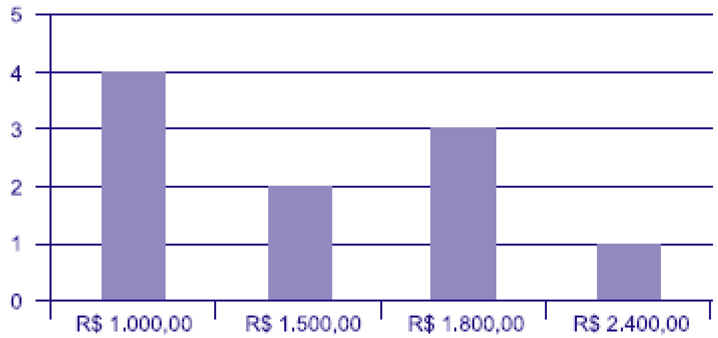
b) 1,2 e 1,5.

c) 1,1 e 1,0.

d) 1,0 e 1,0.

e) 1,0 e 1,5.

41. (VUNESP/Pref. Serrana/2018) O gráfico apresenta informações sobre o número de funcionários em um pequeno comércio e os salários que são pagos a eles. Utilize as informações para responder à questão.



A mediana dos salários dos funcionários desse comércio é igual a

a) R$ 2.400,00.

b) R$ 2.000,00.

c) R$ 1.800,00.

d) R$ 1.500,00.

e) R$ 1.000,00.

42. (VUNESP/Pref. Sertãozinho/2018) Para responder à questão seguinte, considere os números 1,5; 1,5; 1,6; 1,6; 1,6; 1,7; 1,7; 1,8; 1,9 e 1,9 como as alturas, em metros, de 10 pessoas.

A mediana das alturas dessas pessoas

a) É 1,8 metro.

b) Está entre 1,7 e 1,8 metro.

c) É 1,7 metro.

d) Está entre 1,6 e 1,7 metro.

e) É 1,6 metro.

43. (FCC/SEFAZ-MA/2016) Os registros da temperatura máxima diária dos primeiros 6 dias de uma semana foram: 25 °C; 26 °C, 28,5 °C; 26,8 °C; 25 °C; 25,6 °C. Incluindo também o registro da temperatura máxima diária do 7º dia dessa semana, o conjunto dos sete dados numéricos será unimodal com moda igual a 25 °C, e terá mediana igual a 26 °C. De acordo com os dados, é correto afirmar que, necessariamente, a temperatura máxima diária do 7º dia foi

a) Inferior a 25 °C.

b) Superior a 26,8 °C.

c) Igual a 26 °C.

d) Inferior a 25,6 °C.

e) Superior a 26 °C.

44. (FGV/CODEBA/2016) Uma das características principais da mediana é

a) A invariância à unidade de medida utilizada.

b) A robustez à presença de outliers.

c) A identificação da observação mais frequente.

d) O fato de, em seu cálculo, dar mais peso às observações mais frequentes.

e) A normalização pelos desvios em relação à média.

45. (FGV/IBGE/2016) Após a extração de uma amostra, as observações obtidas são tabuladas, gerando a seguinte distribuição de frequências:

Valor	3	5	9	13
Frequência	5	9	10	3

Considerando que E(X) = Média de X, Mo(X) = Moda de X e Me(X) = Mediana de X, é correto afirmar que:

a) E(X) = 7 e Mo(X) = 10;

b) Me(X) = 5 e E(X) = 6,3;

c) Mo(X) = 9 e Me(X) = 9;

d) Me(X) = 9 e E(X) = 6,3;

e) Mo(X) = 9 e E(X) = 7.

46. (VUNESP/CM Registro/2016) Foi construída uma amostra com 10 funcionários de uma determinada repartição para a verificação do consumo de papel A4 por mês naquela área. A verificação do consumo desses funcionários mostrou-se conforme a tabela a seguir.

FUNCIONÁRIO	CONSUMO EM FOLHAS DE PAPEL POR MÊS
A	15
B	20
C	18
D	10
E	16
F	22
G	24
H	30
I	8
J	12

Assinale a alternativa que contém, correta e respectivamente, os valores do consumo médio e do consumo mediano de papel A4 dessa repartição.

a) 15 e 16.

b) 17 e 17,5.

c) 17,5 e 17.

d) 18 e 19.

e) 19,5 e 18.

47. (CESPE/TELEBRAS/2015) Considerando que os possíveis valores de um indicador X, elaborado para monitorar a qualidade de um serviço de cabeamento residencial para a comunicação de dados, sejam elementos do conjunto {0, 1, 2, 3, 4, 5} e que uma amostra aleatória de 5 residências tenha apontado os seguintes indicadores: 4, 4, 5, 4 e 3, julgue o próximo item.

A mediana e a moda dos indicadores registrados na amostra foram iguais a 4.

48. (CESPE/DEPEN/2015)

Quantidade diária de incidentes (N)	Frequência relativa
0	0,1
1	0,2
2	0,5
3	0,0
4	0,2
total	1

Considerando os dados da tabela mostrada, que apresenta a distribuição populacional da quantidade diária de incidentes (N) em determinada penitenciária, julgue o item que se segue.

A distribuição de N não é simétrica em torno da média, apesar de a média e a mediana serem iguais.

49. (FCC/DPE SP/2015) A tabela abaixo corresponde às frequências absolutas dos salários de todos os homens e de todas as mulheres que são empregados de uma empresa.

Classe de Salários (R$)	Frequências Absolutas	
	Homens	Mulheres
$1.500 \vdash 2.500$	10	5
$2.500 \vdash 3.500$	40	20
$3.500 \vdash 4.500$	$2k$	$(K + 10)$
$4.500 \vdash 5.500$	15	10
$5.500 \vdash 6.500$	5	5

Total	$70 + 2K$	$50 + K$

Utilizando o método da interpolação linear para o cálculo da mediana, tem-se que o valor da mediana dos homens é igual a R$ 3.750,00 e o das mulheres é igual a

a) R$ 3.875,00.

b) R$ 4.025,00.

c) R$ 3.925,00.

d) R$ 3.825,00.

e) R$ 4.000,00.

50. (CESPE/ANTAQ/2014)

Nota atribuída pelo passageiro	Frequência
0	15
1	30
2	45
3	50
4	35
5	5

A tabela acima apresenta os resultados de uma pesquisa de satisfação realizada em uma amostra de usuários dos serviços de transporte fluvial prestados por uma empresa. Com base nessas informações e na tabela, julgue o próximo item.

A mediana da série de notas obtidas pela empresa é 3.

51. (FCC/SEFAZ RJ/2014) O Departamento de Pessoal de certo órgão público fez um levantamento dos salários, em número de salários mínimos (SM), dos seus 400 funcionários, obtendo os seguintes resultados:

Salários (em número de SM)	Frequência absoluta
$4 \vdash 6$	48
$6 \vdash 8$	100

$8 \vdash 10$	x
$10 \vdash 12$	y
$12 \vdash 16$	40
Total	**400**

Sabe-se que a mediana dos salários desses funcionários calculada por meio dessa tabela pelo método da interpolação linear é igual a 8,8 SM. Nessas condições, o salário médio desses 400 funcionários, em número de salários mínimos, considerando que todos os valores incluídos em um intervalo de classe são coincidentes com o ponto médio do intervalo, é igual a

a) 8,54

b) 8,83

c) 8,62

d) 8,93

e) 8,72

52. (FGV/AL-BA/2014) Os dados a seguir são uma amostra de 11 salários mensais (aproximados) em reais:

2.080 1.830 2.480 3.010 1.450 1.650 2.500 1.740 3.600 1.900 2.840

A mediana desses salários, em reais, é

a) 1.990.

b) 2.080.

c) 1.650.

d) 2.000.

e) 2.220.

53. (FGV/CGE-MA/2014) Sobre uma amostra com uma quantidade ímpar de valores, todos diferentes de uma variável aleatória, sabe-se que a média é maior que a mediana.

Com relação aos valores dessa amostra é necessariamente verdade que

a) Há mais valores acima da média do que abaixo da média.

b) Há mais valores abaixo da média do que acima da média.

c) Há mais valores acima da média do que abaixo da mediana.

d) Há mais valores acima da mediana do que abaixo da média.

e) A quantidade de valores acima da média é igual à quantidade de valores abaixo da média.

54. (FGV/Pref. Recife/2014) A seguinte amostra de idades foi obtida:

19; 25; 39; 20; 16; 27; 40; 38; 28; 32; 30.

Assinale a opção que indica a mediana dessas idades.

a) 27

b) 28

c) 29

d) 30

e) 31

55. (CESPE/PRF/2013)

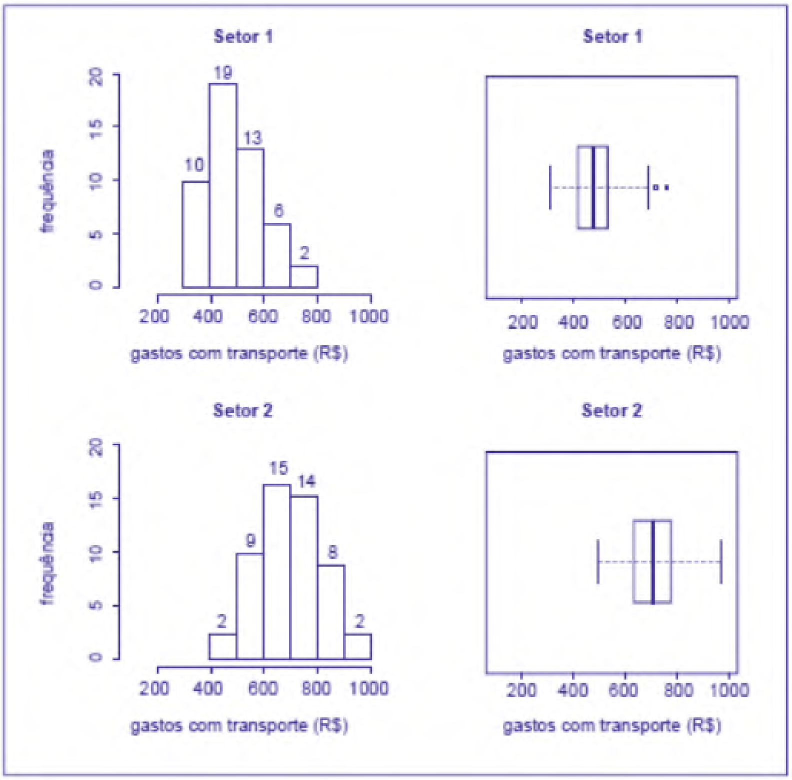
Considerando os dados apresentados no gráfico, julgue o item seguinte.

A média do número de acidentes ocorridos no período de 2007 a 2010 é inferior à mediana da sequência de dados apresentada no gráfico.

56. (CESPE/Polícia Federal/2012) Com relação a estatística, julgue o item seguinte.

Ao contrário da mediana amostral, a média aritmética é menos sensível à presença de valores extremos (ou valores atípicos ou *outliers*).

57. (CESPE/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
723,79	767,53				

Setor 2

488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56
663,81	670,12	671,90	673,78	684,69	685,98
693,35	698,58	708,78	719,80	721,16	734,84

735,94 746,34 754,83 756,10 756,96 760,80

762,29 766,24 770,11 797,73 804,06 805,97

807,29 832,83 844,00 866,77 878,27 897,09

943,10 963,25

Considerando essas informações, julgue o item.

A mediana das despesas registradas pelos servidores do setor 2 é igual a R$ 693,35.

58. (FUNDATEC/SEFAZ-RS/2009) A tabela a seguir representa a distribuição de frequências da idade de uma amostra de moradores de um asilo. Utilize para resolver a questão.

X_i	f_i
70 ⊢ 74	7
74 ⊢ 78	19
78 ⊢ 82	13
82 ⊢ 86	11
86 ⊢ 90	6
90 ⊢ 94	4
Total	**60**

A idade aproximada da mediana é:

a) 78,22.

b) 80,00.

c) 79,38.

d) 78,55.

e) 79,23.

Gabarito

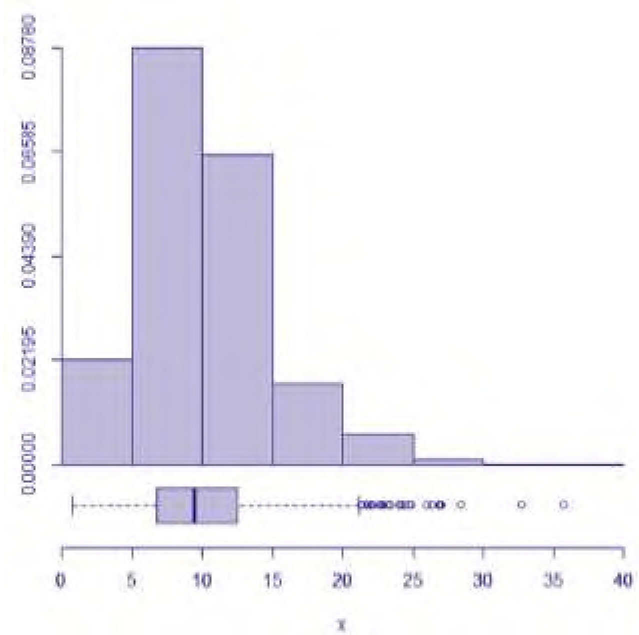
Mediana

1. ERRADO
2. CERTO
3. CERTO
4. LETRA B
5. LETRA C
6. LETRA D
7. LETRA A
8. LETRA B
9. CERTO
10. ERRADO
11. ERRADO
12. LETRA B
13. LETRA E
14. LETRA C
15. LETRA D
16. CERTO
17. ERRADO
18. LETRA E
19. LETRA D
20. LETRA E
21. LETRA E
22. ERRADO
23. LETRA C
24. LETRA D
25. LETRA B
26. LETRA D
27. LETRA C
28. LETRA D
29. LETRA C
30. LETRA C
31. LETRA D
32. LETRA A
33. LETRA B
34. ERRADO
35. LETRA B
36. LETRA A
37. LETRA A
38. LETRA B
39. LETRA E
40. LETRA A
41. LETRA D
42. LETRA D
43. LETRA E
44. LETRA B
45. LETRA E
46. LETRA C
47. CERTO
48. CERTO
49. LETRA A
50. ERRADO
51. LETRA D
52. LETRA B
53. LETRA B
54. LETRA B
55. ERRADO
56. ERRADO
57. ERRADO
58. LETRA E

LISTA DE QUESTÕES

Quartil, Decil e Percentil

1. (CESPE/PETROBRAS/2022) A figura seguinte mostra o histograma como uma estimativa da função de densidade de uma distribuição X, juntamente com o diagrama boxplot correspondente a esse conjunto de dados.

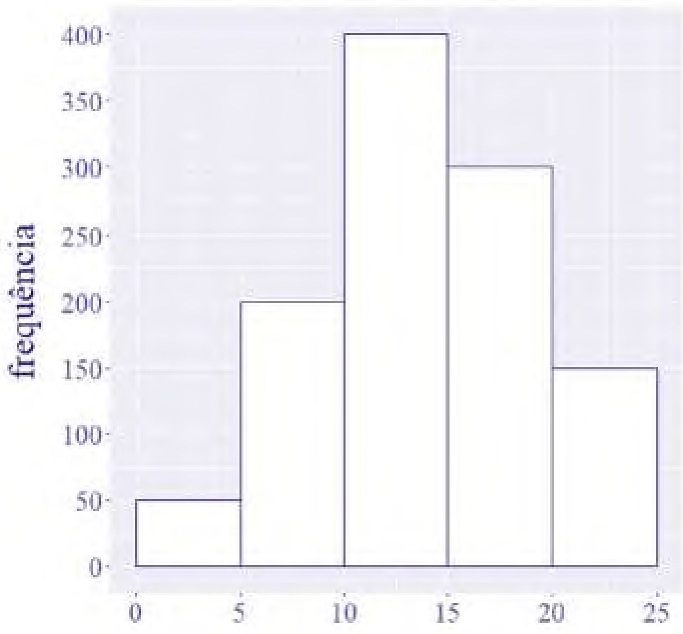


Tamanho da amostra	**1.000**
Média amostral	**10**
Desvio padrão amostral	**4,7**

Considerando a figura e as informações apresentadas no quadro, julgue o item que se segue.

O primeiro decil da distribuição do conjunto de dados em tela é igual ou inferior a 5.

2. (CESPE/TELEBRAS/2022)



Considerando que o histograma apresentado descreve a distribuição de uma variável quantitativa X por meio de frequências absolutas, julgue o item que se segue.

O segundo decil da distribuição da variável X é igual a 20.

3. (CESPE/BANESE/2021) A respeito do conjunto de dados {11, 6, 28, 51, 49, 32, 33}, julgue o item a seguir.

O primeiro quartil do conjunto de dados em tela é igual ou superior a 33.

4. (IBFC/IBGE/2021) Uma pesquisa foi realizada com 40 pessoas e o terceiro quartil da distribuição de frequência é igual a 21,25. Sabendo que a amplitude da classe é 2, a frequência acumulada da classe anterior é 20, então a frequência da classe referente ao terceiro quartil é igual a:

a) 12

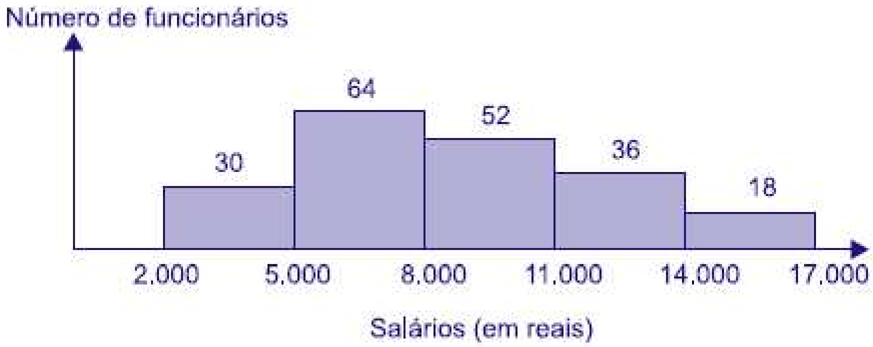
b) 16

c) 20

d) 8

e) 10

5. (VUNESP/EsFCEx/2021) O gráfico apresenta informações sobre a distribuição dos funcionários em relação aos salários pagos em uma empresa.



Com base nas informações apresentadas, é correto afirmar que um funcionário do grupo dos 25% dos maiores salários nessa empresa não ganha menos do que

a) R$ 11.333,33.

b) R$ 11.666,66.

c) R$ 11.527,00.

d) R$ 11.478,25.

e) R$ 11.745,75.

6. (CESPE/ME/2020)

Mínimo	**5**
Moda	**9**
Mediana	**9**
Média	**10**
Máximo	**15**

Um levantamento amostral proporcionou as estatísticas precedentes, referentes a determinada variável quantitativa X.

Considerando essas informações e que a variável X é composta por 1240 observações, julgue o item subsequente.

O terceiro quartil da variável X foi inferior a 9.

7. (MARINHA DO BRASIL/2020) Sabe-se que o valor da amplitude semi-interquartílica de um conjunto de dados é 180 e que o valor do 3° quartil é 4 vezes o valor do 1° quartil. Assinale a opção que apresenta o valor do 1 ° e do 3° quartil, respectivamente.

a) 120 e 480

b) 90 e 360

c) 180 e 480

d) 120 e 360

e) 180 e 720

8. (MARINHA DO BRASIL/2020) Dada a distribuição amostral abaixo, determine o 4° Decil e 66º Percentil e assinale a opção correta.

Classes	Frequência
2 ⊢ 6	6
6 ⊢ 10	12
10 ⊢ 14	8
14 ⊢ 18	14
18 ⊢ 22	10

a) 8 e 18

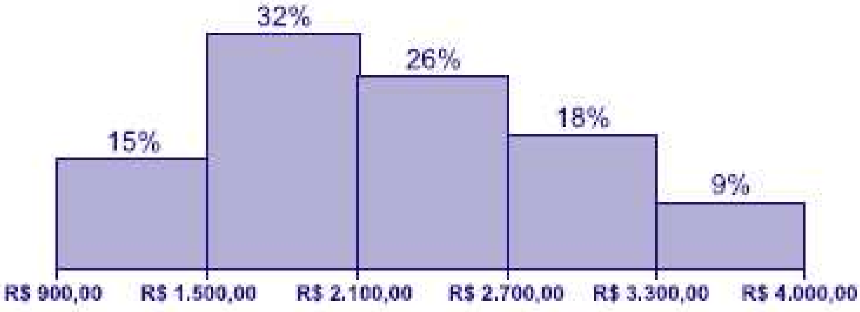
b) 8 e 20

c) 9 e 20

d) 11 e 16

e) 11 e 14

9. (VUNESP/UNICAMP/2019) O gráfico apresenta a distribuição dos salários dos funcionários de um escritório.



Sabendo-se que, em cada classe a distribuição de salários é uniforme, 30% dos salários mais baixos desse grupo variam de R$ 900,00 a

a) R$ 1.768,25.

b) R$ 1.779,25.

c) R$ 1.781,25.

d) R$ 1.795,25.

e) R$ 1.801,25.

10. (FGV/TJ-AL/2018) Para avaliar a produtividade de um dado conjunto de varas da justiça, é extraída uma amostra do número de audiências efetivamente realizadas durante um determinado período.

Os dados foram tratados, obtendo-se as seguintes estatísticas:

Me (A.) = 22, Q_1=19 e Q_3 =27

Essas estatísticas representam os Quartis da distribuição.

Adotando a técnica de Box-Plot para fins da identificação de outliers, sobre os valores A1 = 6, A2 = 11 e A3 = 40 tem-se que:

a) Todos são outliers;

b) Os dois primeiros são outliers;

c) Apenas A3 é um outlier;

d) A1 e A3 são outliers;

e) Nenhum deles é outlier.

11. (CESPE/SEE-DF/2017) Um levantamento estatístico, feito em determinada região do país, mostrou que jovens com idades entre 4 e 17 anos assistem à televisão, em média, durante 6 horas por dia. A tabela a seguir apresenta outras estatísticas produzidas por esse levantamento.

	Distribuição dos tempos gastos assistindo televisão (*T, em horas*)
1° quartil	2
2° quartil	4
3° quartil	8
1° decil	1
9° decil	10

Tendo como referência essas informações, julgue o seguinte item.

O desvio quartílico dos tempos *T* foi igual a 3.

12. (CESPE/SEE-DF/2017) Um levantamento estatístico, feito em determinada região do país, mostrou que jovens com idades entre 4 e 17 anos assistem à televisão, em média, durante 6 horas por dia. A tabela a seguir apresenta outras estatísticas produzidas por esse levantamento.

	Distribuição dos tempos gastos assistindo televisão (*T, em horas*)
1° quartil	2
2° quartil	4
3° quartil	8
1° decil	1
9° decil	10

Segundo esse levantamento, metade dos jovens com idades entre 4 e 17 anos assistem à televisão durante 8 horas por dia.

13. (FGV/IBGE/2016) Adotando-se para as estatísticas de posição de uma dada distribuição de frequências as convenções, Q_k = Quartil de ordem k, D_k = Decil de ordem k, Qt_k = Quintil de ordem k e P_k = Pencentil de ordem k, é correto afirmar que:

a) $Q_3 \geq D_6 \geq Qt_4 = P_{80}$;

b) $Qt_2 \leq P_{55} \leq D_6 \leq Q_3$;

c) $D_9 \geq P_{85} \geq Q_3 = Qt_3$;

d) $Q_1 \geq Qt_2 = P_{20} \leq D_3$;

e) $D_6 \leq Q_3 = P_{75} \leq Qt_3$.

14. (FCC/TRE RR/2015) A distribuição dos valores dos salários, em dezembro de 2014, dos 200 funcionários em um órgão público é representada por uma tabela de frequências absolutas, com todos os intervalos de classe apresentando a mesma amplitude, sendo fechados à esquerda e abertos à direita. O valor da mediana, obtido pelo método da interpolação linear, foi igual a R$ 5.600,00 e pertencente ao intervalo de classe, em reais, [5.000,00 ; 6.500,00). Se 80 funcionários possuem um salário inferior a R$ 5.000,00, então a porcentagem dos funcionários que apresentam um salário igual ou superior a R$ 6.500,00 é, em %, igual a

a) 45.

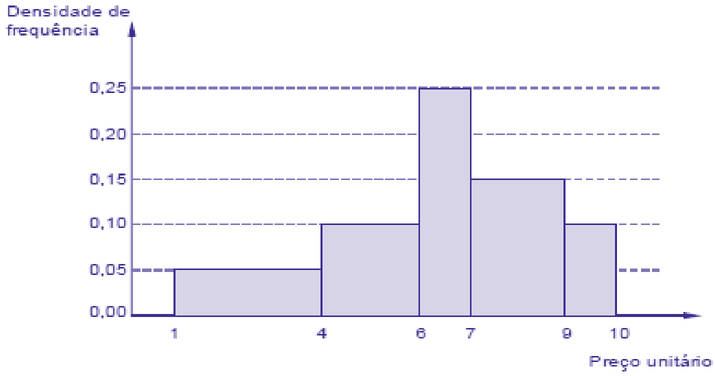
b) 30.

c) 50.

d) 25.

e) 35.

15. (FCC/TRE RR/2015) O histograma abaixo representa a distribuição dos preços unitários de custo, em R$, de determinado equipamento de informática no mercado. No eixo das abscissas constam os intervalos de classe, em R$, e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)^{-1}$.



Observação:
Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.

Considerando os intervalos de classe fechados à esquerda e abertos à direita, se 105 preços apresentam valores menores que R$ 6,00, então o número de preços que apresentam valores iguais ou superiores a R$ 4,00 é

a) 240.

b) 195.

c) 215.

d) 230.

e) 255.

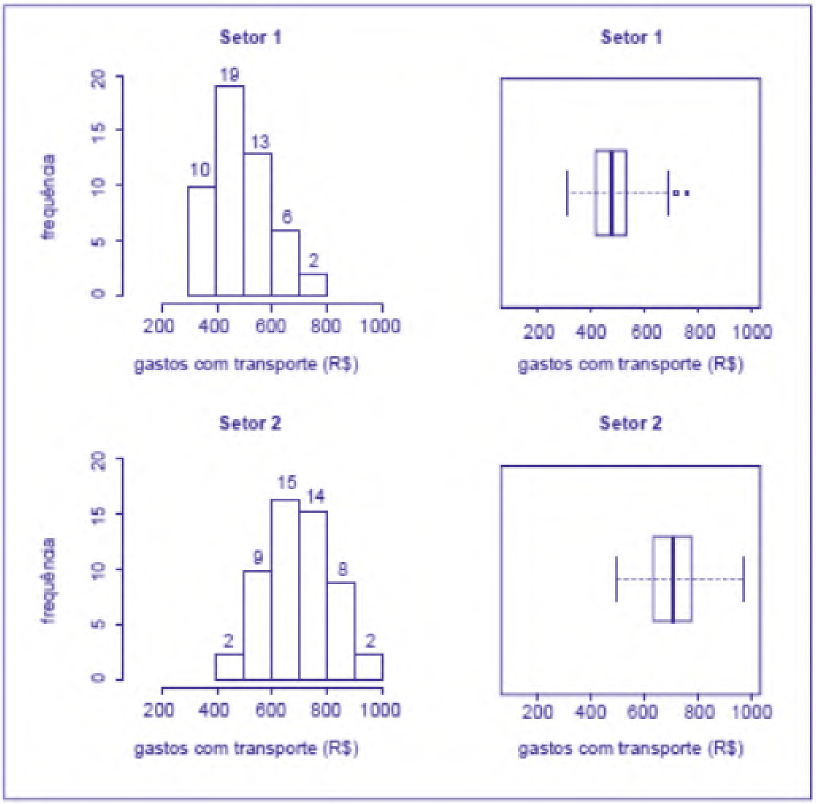
16. (CESPE/BACEN/2013)

2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários. Com base nesses dados, julgue o próximo item.

A mediana é maior que o 50º percentil.

17. (CESPE/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores 1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
723,79	767,53				

Setor 2

488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56
663,81	670,12	671,90	673,78	684,69	685,98

693,35	**698,58**	**708,78**	**719,80**	**721,16**	**734,84**
735,94	**746,34**	**754,83**	**756,10**	**756,96**	**760,80**
762,29	**766,24**	**770,11**	**797,73**	**804,06**	**805,97**
807,29	**832,83**	**844,00**	**866,77**	**878,27**	**897,09**
943,10	**963,25**				

Considerando essas informações, julgue o item.

Trinta por cento das despesas dos servidores do setor 1 correspondem a um valor superior a R$ 525,00.

Gabarito

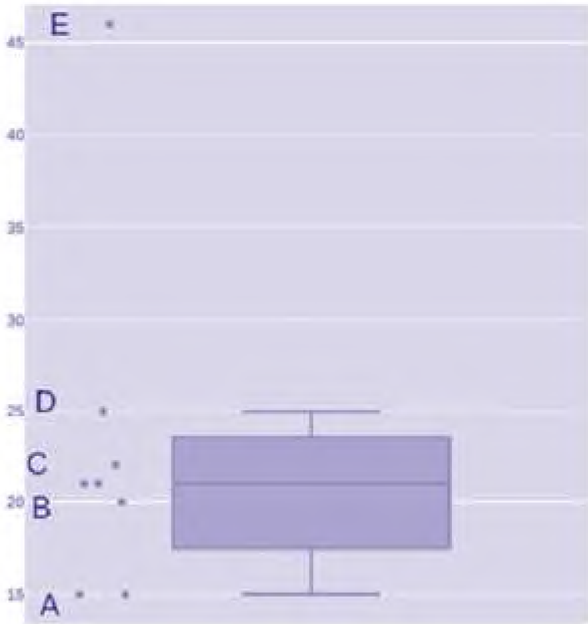
Quartil, Decil e Percentil

1. CERTO
2. ERRADO
3. ERRADO
4. LETRA B
5. LETRA A
6. ERRADO
7. LETRA A
8. LETRA D
9. LETRA C
10. LETRA D
11. CERTO
12. ERRADO
13. LETRA B
14. LETRA E
15. LETRA E
16. ERRADO
17. CERTO

LISTA DE QUESTÕES

Box Plot

1. (CESPE/DPE RO/2022)



No gráfico boxplot anteriormente apresentado, o outlier do conjunto de dados é representado pelo ponto

a) A.

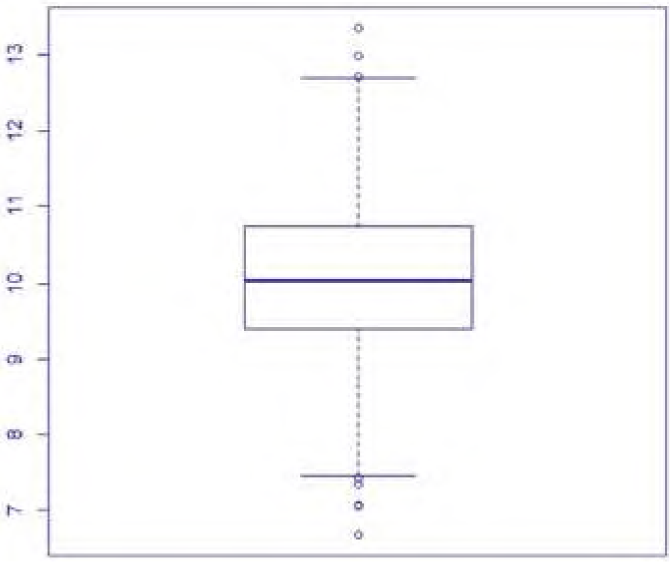
b) E.

c) B.

d) C.

e) D.

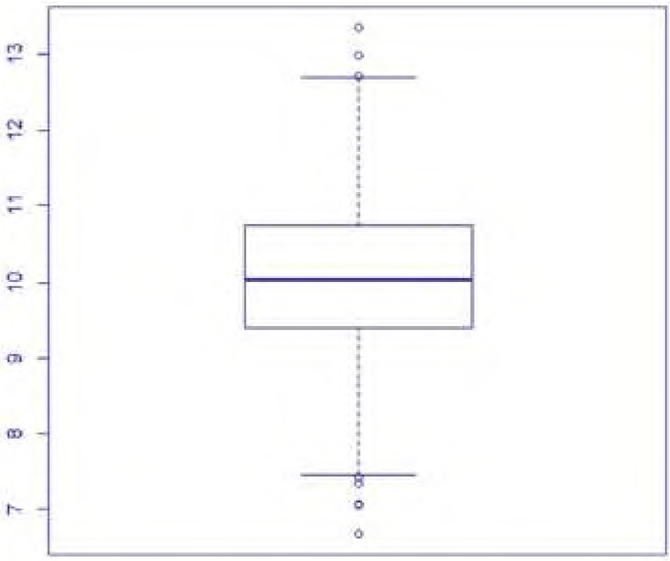
2. (CESPE/PETROBRAS/2022)



Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

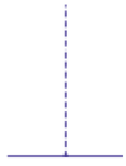
A amplitude total dos dados em tela é inferior a 6.

3. (CESPE/PETROBRAS/2022)

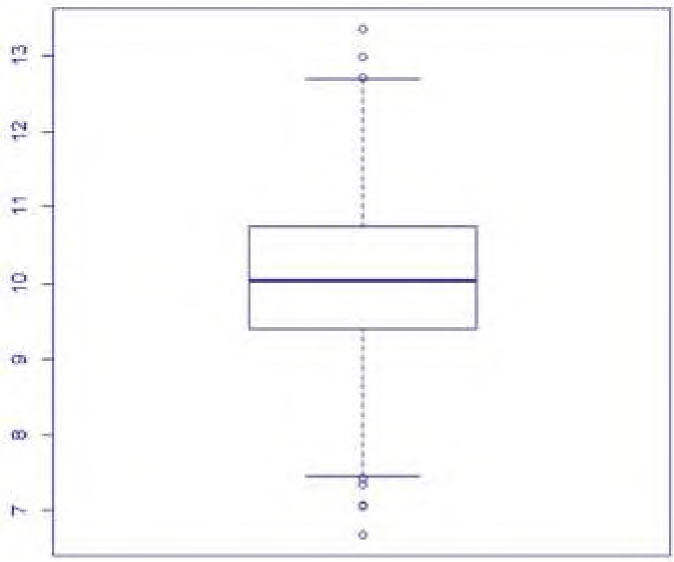


Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

Nesse diagrama, a porção da distribuição dos dados representada pela parte inferior do diagrama mostrada a seguir representa exatamente 25% dos dados em questão.



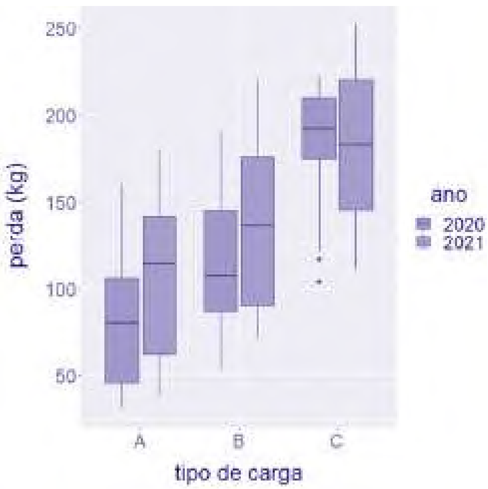
4. (CESPE/PETROBRAS/2022)



Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue o item a seguir.

O terceiro quartil é inferior a 11 e superior a 10.

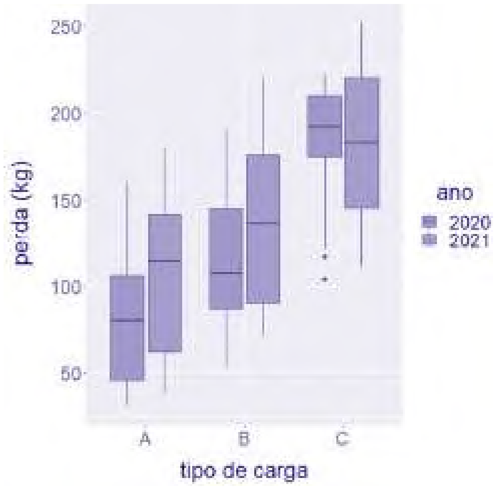
5. (CESPE/PETROBRAS/2022)



Considerando a figura precedente, que mostra desenhos esquemáticos das distribuições das quantidades de cargas perdidas nos anos de 2020 e 2021, segundo o tipo de carga transportada por uma mineradora, julgue o item que se segue.

No que se refere à distribuição da quantidade de carga do tipo B perdida em 2021, observa-se que o valor da perda mínima foi superior a Q_1 - 1,5Dq.

6. (CESPE/PETROBRAS/2022)



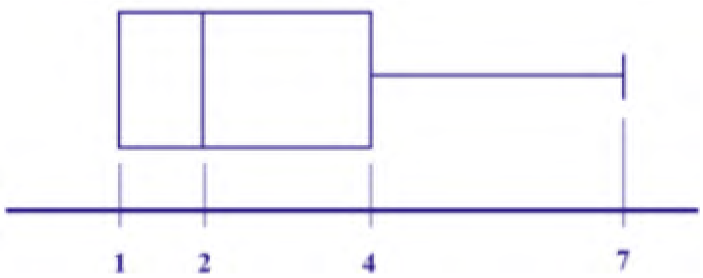
Considerando a figura precedente, que mostra desenhos esquemáticos das distribuições das quantidades de cargas perdidas nos anos de 2020 e 2021, segundo o tipo de carga transportada por uma mineradora, julgue o item que se segue.

Na distribuição da quantidade de carga do tipo A perdida em 2020, observa-se que o primeiro quartil foi superior a 100 kg, enquanto o terceiro quartil foi inferior a 50 kg.

7. (CESPE/TELEBRAS/2022) Com respeito ao conjunto de dados {0, 0, 1, 1, 1, 3}, julgue o item que se segue.

Se esse conjunto de dados fosse representado por um diagrama de box-plot, então os valores 0 e 3 seriam chamados valores exteriores, ou, ainda, discrepantes, atípicos ou outliers.

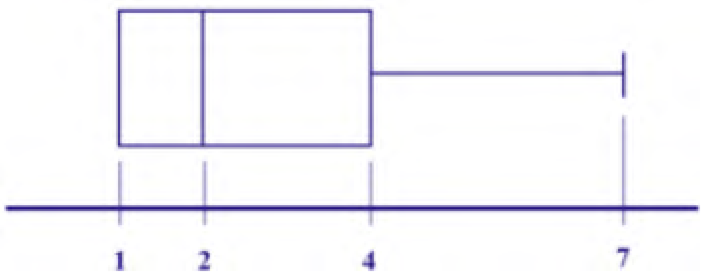
8. (CESPE/TCE-RJ/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

A mediana da variável X é igual a 4.

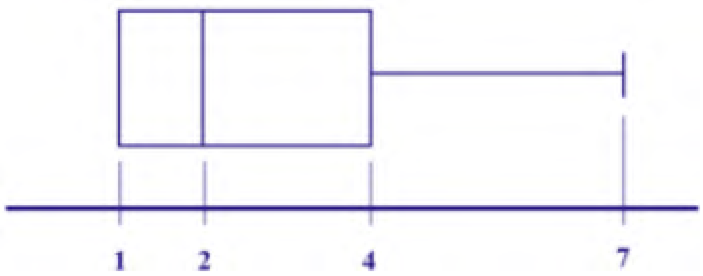
9. (CESPE/TCE-RJ/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

Se f denota a frequência relativa do valor 1, então $0{,}25 \leq f < 0{,}50$.

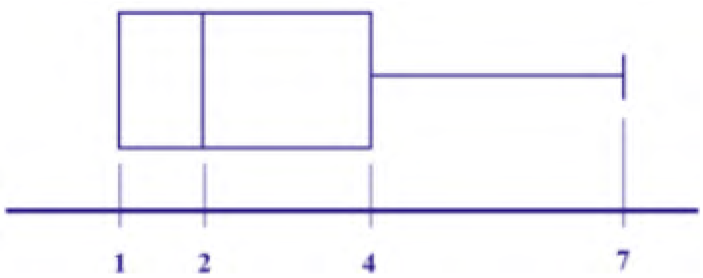
10. (CESPE/TCE-RJ/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

$\frac{1}{3}$ das observações da variável X são iguais a 7.

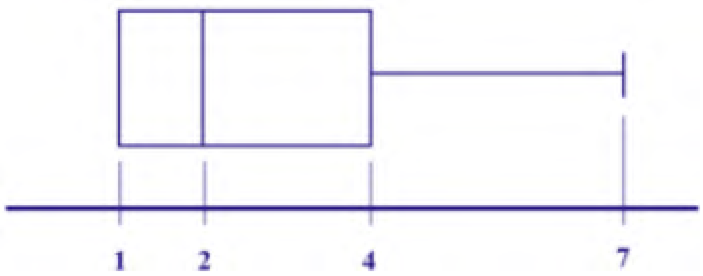
11. (CESPE/TCE-RJ/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

O diagrama *boxplot* indica que o intervalo interquartil (ou interquartílico) da distribuição da variável X é igual a 3.

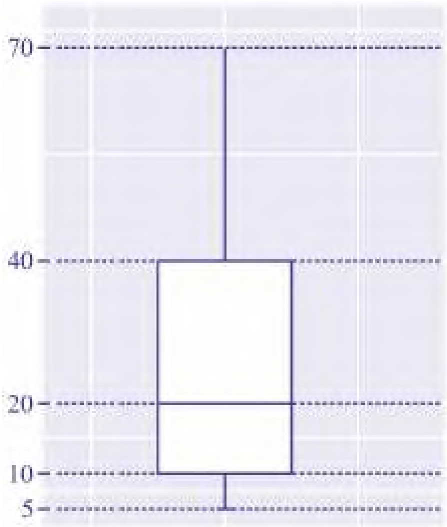
12. (CESPE/TCE-RJ/2021)



Considerando que uma variável quantitativa discreta X se distribui conforme o diagrama boxplot anterior, julgue o item seguinte.

As observações da variável X que assumem valores iguais a 7, com base nesse diagrama boxplot, são considerados *outliers.*

13. (CESPE/TJ PA/2020)



Considerando que o desenho esquemático (boxplot) antecedente se refere a uma variável quantitativa X, assinale a opção correta.

a) O intervalo interquartil é igual a 65.

b) Metade da distribuição da variável X se encontra entre os valores 20 e 40.

c) Os valores da variável X que se encontram no intervalo [5;10] representam 5% da distribuição de X.

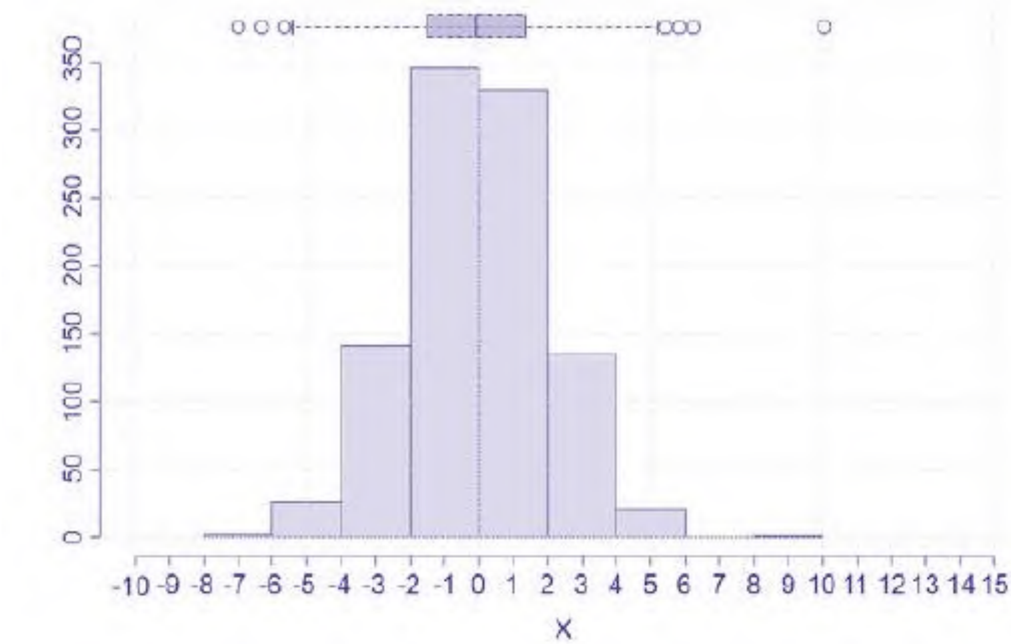
d) A mediana de X é igual a 25.

e) O primeiro quartil da distribuição de X é igual a 10.

14. (CESPE/ME/2020) Acerca de visualização e análise exploratória de dados, julgue o item seguinte.

Outlier ou anomalias são padrões nos dados que não estão de acordo com uma noção bem definida de comportamento normal.

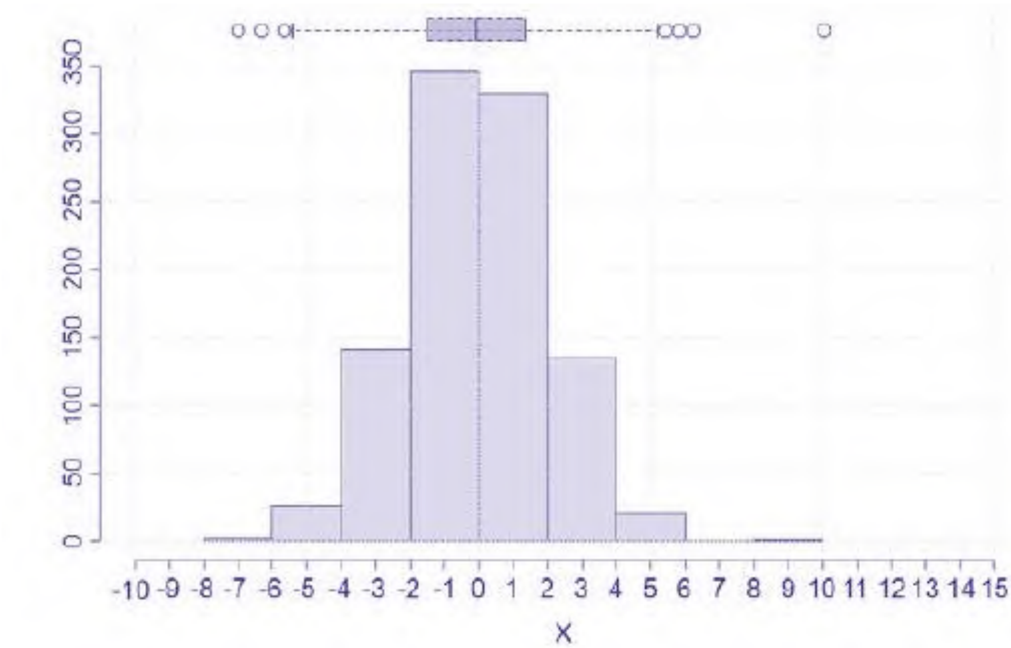
15. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

O primeiro quartil da distribuição de X é inferior a −2.

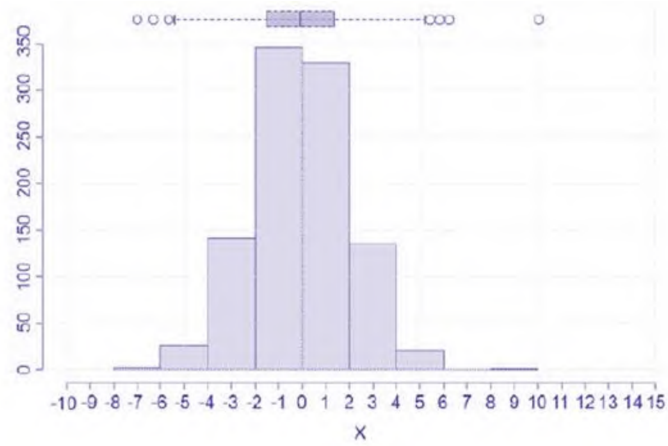
16. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

No diagrama de *boxplot*, os pontos indicados pelo símbolo ○ representam *outliers.*

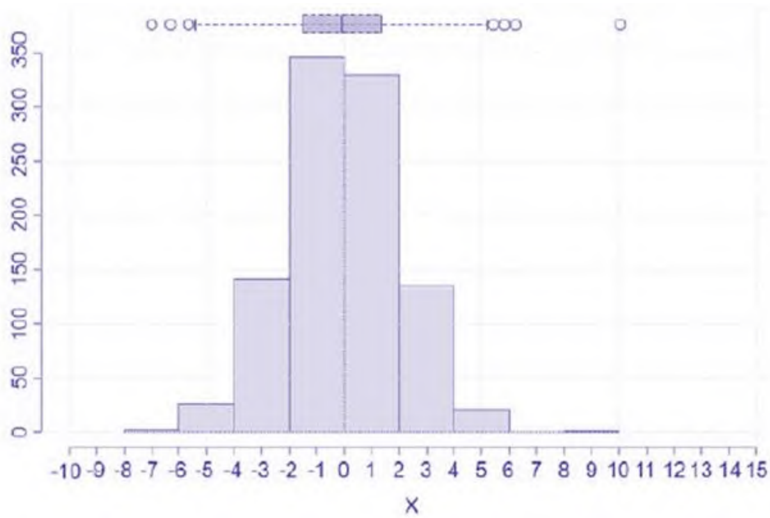
17. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

O percentual das observações da variável X que se encontram no intervalo [−2,+2] é igual ou superior a 50%.

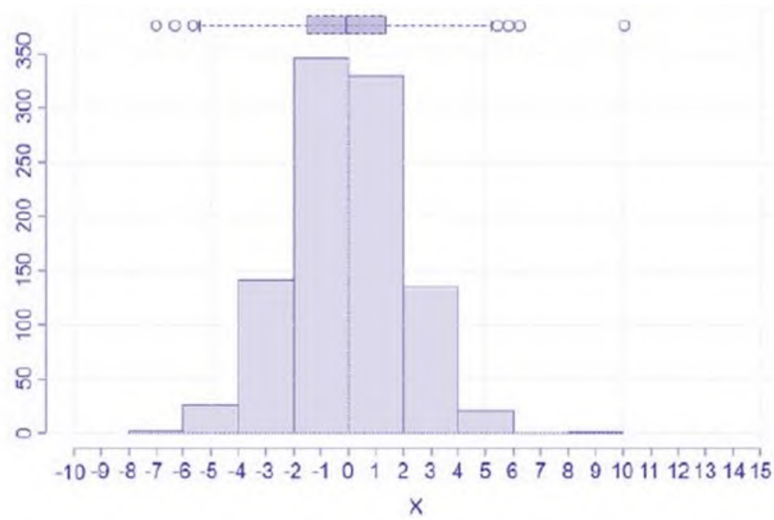
18. (CESPE/ME/2020)



Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

O comprimento da caixa do diagrama de *boxplot*, representado pela figura , é obtido multiplicando-se o desvio padrão de X por 2.

19. (CESPE/ME/2020)



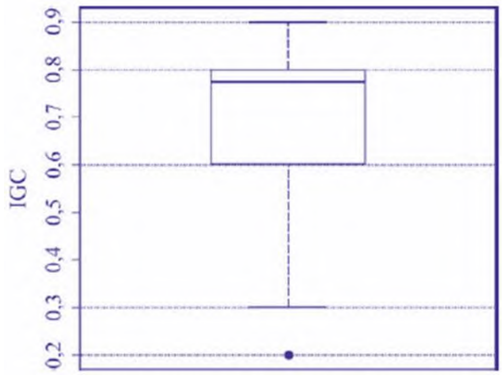
Considerando o histograma e o diagrama boxplot mostrados anteriormente, julgue o item a seguir.

A mediana de X se encontra no intervalo [−1,+1].

20. (CESPE/ME/2020) Acerca de visualização e análise exploratória de dados, julgue o item seguinte.

O gráfico apresentado a seguir é denominado caixa de barra.

21. (CESPE/TCE-PR/2016)



Com base na figura antecedente, que apresenta a distribuição dos indicadores de governança corporativa (IGC) observados em uma amostra de empresas prestadoras de serviços terceirizados, assinale a opção correta.

a) O menor e o maior IGC observados na amostra foram, respectivamente, iguais a 0,3 e 0,9.

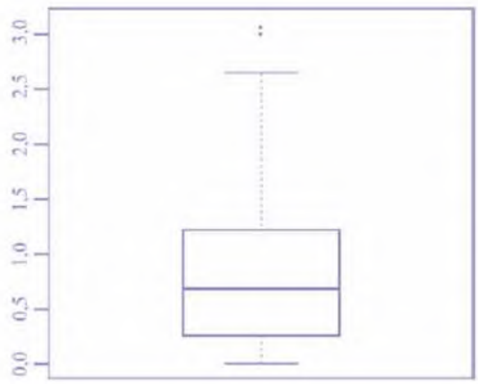
b) O diagrama mostrado na figura em questão é denominado curva de frequência.

c) O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

d) Na amostra considerada, a mediana dos indicadores observados foi inferior a 0,7.

e) A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

22. (CESPE/TCE-PA/2016)



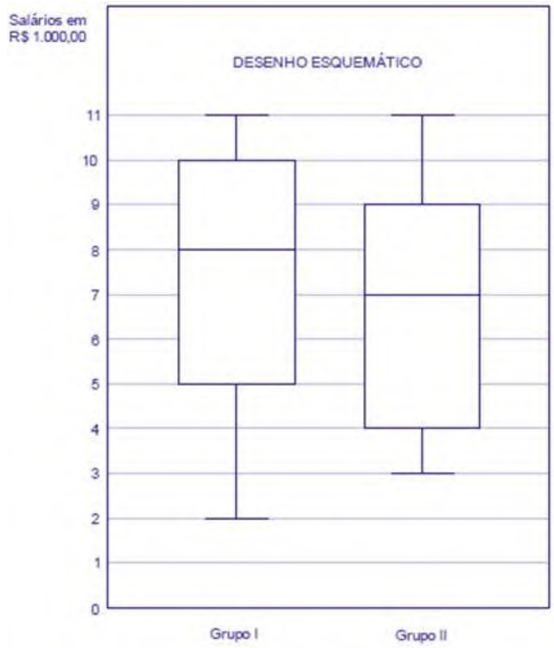
média amostral	0,80
desvio padrão amostral	0,70
primeiro quartil	0,25
mediana	0,70
terceiro quartil	1,20
mínimo	0
máximo	3,10

Um indicador de desempenho X permite avaliar a qualidade dos processos de governança de instituições públicas. A figura mostra, esquematicamente, a sua distribuição, obtida mediante estudo amostral feito por determinada agência de pesquisa. A tabela apresenta estatísticas descritivas referentes a essa distribuição.

Com base nessas informações, julgue o item a seguir.

O diagrama boxplot mostrado na figura sugere a existência de pelo menos duas observações atípicas.

23. (FCC/TRT 3ª Região/2015) Seja uma representação gráfica de dados de acordo com o desenho esquemático abaixo (box-plot) que foi preparado para comparar todos os salários dos funcionários do sexo masculino (Grupo I) com todos os salários dos funcionários do sexo feminino (Grupo II) lotados em um órgão público.



Neste desenho esquemático

a) O número de elementos do Grupo I é superior ao número de elementos do Grupo II.

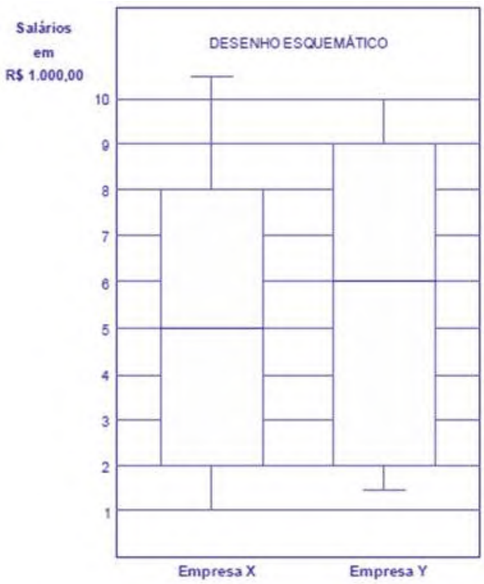
b) O módulo da diferença entre as medianas dos 2 grupos é igual a 25% do menor salário deste órgão público.

c) Mais da metade dos elementos do Grupo I possui um salário inferior a R$ 5.000,00 ou superior a R$ 10.000,00.

d) O valor do menor salário do Grupo II corresponde a 37,5% do valor da mediana do Grupo I.

e) A diferença interquartil do Grupo I é superior à diferença interquartil do Grupo II.

24. (FCC/TRT 19ª Região/2014) Para comparar os salários dos empregados de duas empresas X e Y, considerou-se o desenho esquemático abaixo com os valores dos salários em R$ 1.000,00.



De acordo com o desenho esquemático apresentado, é correto afirmar que

a) O número de empregados da empresa X é maior que o número de empregados da empresa Y.

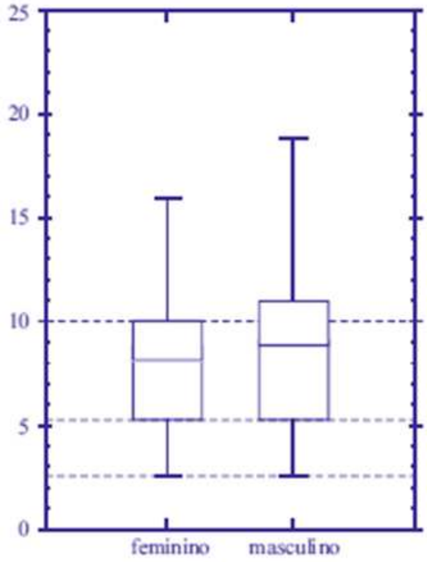
b) A distância interquartil da empresa Y é superior à distância interquartil da empresa X.

c) O menor salário verificado tanto na empresa X como na empresa Y é igual a R$ 2.000,00.

d) O valor da mediana da empresa Y supera o da empresa X em 10%.

e) O maior salário verificado na empresa X é inferior ao maior salário verificado na empresa Y.

25. (CESPE/PRF/2012)



M. M. B. Paolielo et al.. In: ***Saúde Pública****, 1997 (com adaptações).*

Em decorrência do desenvolvimento urbano e tecnológico, tem-se a preocupação de monitorar os efeitos nocivos da poluição ambiental sobre a saúde da população urbana. A figura acima mostra o diagrama de caixa (box-plot) que descreve a distribuição da concentração de chumbo no sangue, em µg.dL-1, obtida com base em uma amostra aleatória de 200 pessoas do sexo masculino e de 100 pessoas do sexo feminino que trabalham em postos de combustível localizados em determinado município brasileiro.

Com base nessas informações, julgue o item.

Com base nas linhas horizontais que cortam as caixas do diagrama apresentado, conclui-se corretamente que a média das concentrações de chumbo encontradas no sangue das pessoas do sexo feminino que trabalham em postos de combustível do referido município brasileiro é inferior à média das concentrações dessa mesma substância no sangue das pessoas do sexo masculino que trabalham nesses postos de combustível.

Gabarito

Box Plot

1. LETRA B
2. ERRADO
3. ERRADO
4. CORRETO
5. ERRADO
6. ERRADO
7. ERRADO
8. ERRADO
9. ERRADO
10. ERRADO
11. CERTO
12. ERRADO
13. LETRA E
14. CERTO
15. ERRADO
16. CERTO
17. CERTO
18. ERRADO
19. CERTO
20. ERRADO
21. LETRA E
22. CERTO
23. LETRA D
24. LETRA B
25. ERRADO