

ESTATÍSTICA

Apresentação das Variáveis Estatísticas



Livro Eletrônico

Presidente: Gabriel Granjeiro

Vice-Presidente: Rodrigo Calado

Diretor Pedagógico: Erico Teixeira

Diretora de Produção Educacional: Vivian Higashi

Gerente de Produção Digital: Bárbara Guerra

Coordenadora Pedagógica: Élica Lopes

Todo o material desta apostila (incluindo textos e imagens) está protegido por direitos autorais do Gran. Será proibida toda forma de plágio, cópia, reprodução ou qualquer outra forma de uso, não autorizada expressamente, seja ela onerosa ou não, sujeitando-se o transgressor às penalidades previstas civil e criminalmente.

CÓDIGO:

240304419179



THIAGO CARDOSO

Engenheiro eletrônico formado pelo ITA com distinção em Matemática, analista-chefe da Múltiplos Investimentos, especialista em mercado de ações. Professor desde os 19 anos e, atualmente, leciona todos os ramos da Matemática para concursos públicos.

GRAN
CONCURSOS

SUMÁRIO

Apresentação	5
Apresentação das Variáveis Estatísticas	6
1. Tabelas Estatísticas	6
1.1. Séries Simples	6
1.2. Séries Mistas	8
1.3. Normas para a Apresentação Tabular	9
2. Gráficos Estatísticos	13
2.1. Frequências	13
2.2. Gráficos de Barras ou Colunas	18
2.3. Gráfico de Setores	21
3. Distribuição de Frequências	25
3.1. Classes	26
3.2. Tamanho Ideal de uma Classe	32
3.3. Frequências Acumuladas Decrescentes	40
3.4. Histogramas com Classes de Amplitude Desiguais	41
3.5. Polígono de Frequências	46
3.6. Histograma x Diagrama de Barras	47
4. Outros Diagramas Estatísticos	50
4.1. Diagrama de Ramos e Folhas	50
4.2. Linha de Tendência	53
4.3. Histogramas ou Diagramas de Colunas Multivariados	53
4.4. Gráficos de Linha e Dispersão	54
5. Separatrizes	57
5.1. Cálculo de Quantis em Dados Categorizados	60
5.2. Diagrama Box Plot	66
Resumo	71

Questões Comentadas em Aula	74
Questões de Concurso.....	83
Gabarito	109
Gabarito Comentado.....	110

APRESENTAÇÃO

Olá, seja bem-vindo(a) a mais uma aula do nosso curso de Estatística.

Nesta aula, falaremos sobre a apresentação das variáveis estatísticas, por meio de gráficos e tabelas. O uso de gráficos e tabelas é amplo na Estatística. E, por isso, considero esse um tema de base.

Vale notar que as provas de Estatística têm ficado de um nível mais elevado nos últimos cinco anos de concursos públicos.

Antigamente, eram comuns as provas em que o aluno tinha que meramente interpretar um gráfico ou tabela. Hoje em dia, as provas pedem mais.

Você precisará ter um conhecimento mais detalhado e aprofundado sobre os principais gráficos e tabelas estatísticos não só para resolver as questões diretamente sobre a apresentação de variáveis, mas também para entender e resolver as questões da parte de Estatística Inferencial.

Por isso, o presente capítulo é muito importante para o seu progresso na matéria.

Como sempre, gostaria de te passar meus contatos: thiagofernando.pe@gmail.com

Feitas essas orientações iniciais, vamos juntos ao topo da montanha?

APRESENTAÇÃO DAS VARIÁVEIS ESTATÍSTICAS

1. TABELAS ESTATÍSTICAS

Também conhecidas como séries estatísticas, as tabelas são muito famosas e utilizadas na Estatística. Elas são muito importantes, pois permitem **a inspeção rigorosa e uma análise detalhada dos dados apresentados.**

As tabelas são muito utilizadas na **apuração dos dados**. Por outro lado, **os gráficos são mais visuais**, portanto, mais simples de entender à primeira vista. Por isso, os gráficos serão mais utilizados na **apresentação dos dados**, principalmente quando a apresentação é rápida ou destinada a um público leigo.

Certamente, podemos apresentar os dados na forma de uma tabela. **Isso é muito útil quando o público** a que se destina o gráfico **será bastante criterioso com aqueles dados.**

Por exemplo, se você está fazendo uma apresentação para o diretor do seu órgão sobre a quantidade de erros encontrados em licitações, ele se interessará por examinar a fundo esses dados. Portanto, preferirá uma tabela.

Por outro lado, se você vai fazer a mesma apresentação para a sociedade, **será mais fácil de entender por meio de um gráfico.**

De maneira geral, as tabelas exibem o comportamento de uma variável estatística em função de outra grandeza de interesse, como o tempo e o espaço geográfico.

As séries estatísticas resumem um conjunto de observações por meio de três fatores fundamentais:

- tempo;
- espaço;
- espécie.

As séries são classificadas de acordo com o fator que varia ao longo da tabela.

1.1. SÉRIES SIMPLES

As séries simples são aquelas em que apenas um fator varia. São elas:

- **Série Histórica:** é estudada apenas a influência do **fator TEMPO** sobre a variável observada.

Figura 1: Exemplo de Série Histórica

Tabela 2.1. Casos de sarampo notificados no Brasil de 1987 a 1992.

Ano	Número de casos
1987	65.459
1988	26.173
1989	55.556
1990	61.435
1991	45.532
1992	7.934

Fonte: Anuários estatísticos – IBGE.

Nessa série estatística, foi estudado o número de casos notificados de sarampo em função do ano.

Por exemplo, com a inspeção dos dados, podemos verificar que os anos de 1988 e 1992 foram bem atípicos em relação a essa variável estatística. Um pesquisador mais interessado nesses dados poderia buscar o que aconteceu nesses dois anos a fim de entender as causas da queda e propor mudanças para os anos seguintes no combate à doença.

Algumas perguntas que ele poderia se fazer: Houve alguma campanha de vacinação bem-sucedida? Ou algum fator climático?

Algo que podemos aproveitar este momento para comentar é que a Estatística é, de maneira geral, uma ferramenta, e não o objetivo em si de uma pesquisa.

- **Séries Geográficas:** é estudada apenas a influência do **fator ESPAÇO** sobre a variável observada.

De maneira geral, essas séries são usadas para comparar países diferentes ou cidades diferentes. Vejamos um exemplo:

Figura 2: Exemplo de Série Geográfica

Tabela 2.2. Necessidades médias de energia em alguns países, em 1973.

País	kcal/per capita/dia
Brasil	2.174
Estados Unidos	2.397
Etiópia	2.120
Japão	1.125
México	2.114

Fonte: Necessidades Humanas de Energia – IBGE.

A série geográfica apresentada mostra que o país em que mora a pessoa exerce influência no seu consumo de calorias diário.

- **Séries Específicas:** é estudado apenas o efeito do **fator ESPÉCIE** sobre a variável estatística em estudo.

Figura 3: Exemplo de Série Específica

Tabela 2.3. Abate de animais, por espécie, no Brasil, em 1993.

Espécie	Número de cabeças
Aves	1.232.978.796
Bovinos	14.951.359
Suínos	13.305.932
Ovinos	926.818
Caprinos	803.188
Equinos	165.691

Fonte: Anuário Estatístico do Brasil (1994).

Nesse caso, a tabela mostra como varia o número de cabeças abatidas em função do tipo de animal: ave, bovino, suíno etc.

1.2. SÉRIES MISTAS

Nas séries mistas, **vários fatores variam simultaneamente**, produzindo um conjunto de dados mais amplo sobre a variável estatística.

É preciso tomar bastante **cuidado** com as séries mistas, **pois elas dificultam bastante o entendimento dos dados**. Porém, de maneira geral, permitem uma quantidade maior de dados, o que pode ser útil para o pesquisador entender melhor a variável em estudo.

Por exemplo, a Tabela 2.4 estudou a taxa de atividade feminina em função do tempo (série histórica) e da região do Brasil (série geográfica).

Figura 4: Exemplo de Série Mista Geográfica e Histórica

Tabela 2.4. Taxa de atividade feminina urbana (em percentual) em três regiões do Brasil, 1981/90.

Região	Ano			
	1981	1984	1986	1990
Norte	28,9	30,3	34,0	37,1
Nordeste	30,2	32,6	34,3	37,8
Sudeste	34,9	37,2	40,1	40,7

Fonte: Anuário Estatístico do Brasil (1992).

Nesse caso, podemos tirar algumas conclusões. Por exemplo, a taxa de atividade feminina cresceu ao longo do tempo em todas as regiões, porém, com mais intensidade no Norte do que no Nordeste.

A próxima Tabela 2.7 é interessante, pois envolve os três tipos de séries. Ela estuda o número de vítimas de acidentes. Leva em consideração a variação no tempo (1991 e 1992), portanto, é uma série histórica. Mas também considera a região do Brasil (série geográfica) e o tipo de vítimas (vítimas fatais e não-fatais).

Figura 5: Exemplo de Série Mista Histórica, Especificativa e Geográfica

Tabela 2.7. Número de vítimas em acidentes, segundo as grandes regiões do Brasil, nos anos de 1991 e 1992.

Região	Vítimas fatais		Vítimas não fatais	
	1991	1992	1991	1992
Norte	1.188	1.165	10.229	9.739
Nordeste	3.857	3.843	23.774	23.942
Sudeste	11.555	10.217	130.938	159.669
Sul	4.402	4.213	61.797	58.832
Centro-Oeste	2.220	1.949	22.147	22.086
Brasil	23.222	21.387	248.885	274.268

Fonte: Anuário Estatístico do Brasil (1994).

1.3. NORMAS PARA A APRESENTAÇÃO TABULAR

Com o intuito de garantir a clareza das informações apresentadas, o IBGE criou um conjunto de normas que fixam conceitos e procedimentos aplicáveis à elaboração de tabelas de dados estatísticos.

Considero um assunto com pouca probabilidade de cair, porém, vale a pena você ler essas normas superficialmente para se precaver.

1.3.1. ELEMENTOS DE UMA TABELA ESTATÍSTICA

Uma tabela é uma forma de apresentar informações, na qual **o dado numérico se destaca como informação central**.

Os elementos fundamentais de uma tabela estatística são:

- **Título:** é a indicação contida na parte superior da tabela, onde se define o fato observado e as condições da observação (local e época).
- **Corpo:** são as linhas e colunas que fornecem o conteúdo pretendido pela tabela.
- **Cabeçalho:** indica a natureza da informação de cada coluna.
- **Coluna Indicadora:** geralmente, é a primeira coluna da tabela, sendo a que determina a natureza do conteúdo das linhas.

Vamos observar como esses elementos estão dispostos em uma tabela estatística real:

Figura 6: Elementos Fundamentais de uma Tabela Estatística

título

Tabela 2.7. Número de vítimas em acidentes, segundo as grandes regiões do Brasil, nos anos de 1991 e 1992.						
coluna indicadora	Região	Vítimas fatais		Vítimas não fatais		cabeçalho
		1991	1992	1991	1992	
	Norte	1.188	1.165	10.229	9.739	
	Nordeste	3.857	3.843	23.774	23.942	
	Sudeste	11.555	10.217	130.938	159.669	
	Sul	4.402	4.213	61.797	58.832	
	Centro-Oeste	2.220	1.949	22.147	22.086	
	Brasil	23.222	21.387	248.885	274.268	
Fonte: Anuário Estatístico do Brasil (1994).					corpo	

Por outro lado, **os elementos complementares de uma tabela estatística são:**

- **Fonte:** designa a entidade que forneceu os dados estatísticos. É muito útil não só para conferir credibilidade aos dados apresentados, mas também uma forma de reconhecer o trabalho da entidade. Temos também um exemplo na Figura 6.
- **Notas:** são esclarecimentos adicionais, de natureza geral.
- **Chamadas:** são esclarecimentos de natureza específica.

As notas e as chamadas devem ser posicionadas **no rodapé da tabela**.

Também é interessante conhecermos alguns sinais e convenções bastante utilizados nas tabelas estatísticas:

- **Três Pontos (...):** quando o dado existe, mas seu valor é desconhecido.
- **Ponto de Interrogação (?):** quando há dúvidas sobre o dado informado.
- **Traço Horizontal (–):** quando o valor do dado é igual a zero.
- **Letra (Z):** quando o valor do dado for rigorosamente igual a zero.
- **Zero (0):** quando o valor do dado é muito pequeno para ser expresso na unidade adotada na tabela.

Esse é um detalhe muito interessante. O “0” em uma tabela estatística não significa um valor exatamente nulo, mas sim apenas um valor pequeno.

Porém, vale notar que nem sempre essa convenção é seguida à risca. Principalmente quando a variável for expressa apenas por número inteiros, veremos muitas tabelas com “0” em lugar de “Z”.

DIRETO DO CONCURSO

001. (CESPE/ANTAQ/2009/TÉCNICO EM REGULAÇÃO) Considerando a tabela acima, que apresenta a movimentação anual de cargas no porto de Santos de 2003 a 2007, em milhões de toneladas/ano e associa as quantidades de carga movimentadas para exportação e importação às variáveis X e Y, respectivamente, julgue o item subsequente.

	variável	2003	2004	2005	2006	2007
exportação	X	40	46	50	52	54
importação	Y	20	21	22	24	27
total	X + Y	60	67	72	76	81

As séries estatísticas apresentadas na tabela formam três séries temporais.



Observe que são mostradas o comportamento de três variáveis: “exportação”, “importação” e “total” em função do tempo, marcado em anos.

Tem-se, portanto, três séries temporais.

Certo.

002. (CESPE ABIN/2018/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4) Com base nos dados da tabela anterior, extraídos do Relatório das Notas Estatísticas do Censo Escolar de 2017, do INEP, julgue o item a seguir.

evolução da quantidade de docentes por etapa de ensino Brasil 2013 – 2017				
ano	educação infantil	anos iniciais do ensino fundamental	anos finais do ensino fundamental	ensino médio
2013	478.811	750.366	802.902	507.617
2014	502.445	757.950	797.577	522.426
2015	518.308	758.840	786.140	522.826
2016	540.567	763.927	778.561	519.883
2017	557.541	761.737	764.731	509.814
soma total das quantidades de docentes no período	2.597.672	3.792.820	3.929.911	2.582.566

O quantitativo de professores dos anos finais do ensino fundamental sofreu queda superior a 1,3% no ano de 2015 em relação ao ano anterior.



Observe que foi fornecida uma tabela estatística. Vamos destacar os pontos referidos no enunciado:

evolução da quantidade de docentes por etapa de ensino Brasil 2013 – 2017				
ano	educação infantil	anos iniciais do ensino fundamental	anos finais do ensino fundamental	ensino médio
2013	478.811	750.366	802.902	507.617
2014	502.445	757.950	797.577	522.426
2015	518.308	758.840	786.140	522.826
2016	540.567	763.927	778.561	519.883
2017	557.541	761.737	764.731	509.814
soma total das quantidades de docentes no período	2.597.672	3.792.820	3.929.911	2.582.566

A queda observada no ano de 2015 em relação ao ano de 2014 pode ser calculada como a razão da diferença entre as observações nesse ano e no ano anterior dividida pela observação no ano de 2014.

$$\% = \frac{\text{Variação}}{\text{Observação no Ano de 2014}}$$

Substituindo os valores fornecidos no enunciado, temos:

$$\% = \frac{786140 - 797577}{797577} = -\frac{11437}{797577} \cong -0,014 = -1,4\%$$

Portanto, foi registrada uma queda de aproximadamente 1,4%, que é superior a 1,3%.

Certo.

003. (CESPE/ABIN/2018/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4) A menor diferença entre as quantidades de professores dos anos iniciais do ensino fundamental e do ensino médio ocorreu em 2014.



Vamos olhar a evolução da série histórica da diferença da quantidade de professores dos anos iniciais e finais do ensino fundamental.

Ano	Anos iniciais do ensino fundamental	Anos finais do ensino fundamental	Diferença
2013	750.366	507.617	242.749
2014	757.950	522.426	235.524
2015	758.840	522.826	236.014
2016	763.927	519.883	244.044
2017	761.737	509.814	251.923

Portanto, a diferença máxima aconteceu no ano de 2014, como mostrado na tabela.

Certo.

2. GRÁFICOS ESTATÍSTICOS

O ser humano é bastante visual. E os gráficos se comunicam muito bem com essa necessidade.

Os gráficos transmitem a informação de maneira muito simples de entender. Por isso, eles podem ser vistos nas mais variadas fontes de informação.

Nesta seção, vamos estudar os principais tipos de gráficos.

Vale ressaltar que o objetivo desta seção não é ensiná-lo a desenhar gráficos. Tudo o que você precisa aprender aqui é interpretá-los.

2.1. FREQUÊNCIAS

Para estudar o histograma, precisamos ver primeiramente alguns conceitos.

- **Frequência Absoluta:** é a **contagem** de todas as vezes que uma variável estatística assume determinado valor.
- **Frequência Relativa:** indica a **proporção** de vezes que uma variável estatística assume determinado valor. Por proporção, entendemos uma proporção percentual.
- **Frequência Acumulada:** indica a **proporção** de vezes que uma variável estatística **assume um valor igual ou menor** ao valor de referência.

Para exemplificar a diferença entre esses três conceitos, vamos tomar como exemplo um conjunto de dados sobre algumas pessoas.

Tabela 1: Conjunto de Dados de Referência

N.	Estado Civil	Grau de Instrução	N. de Filhos	Salário	Idade (anos)	Idade (meses)	Região de Procedência
1	solteiro	fundamental	...	4,00	26	3	interior
2	casado	fundamental	1	4,56	32	10	capital
3	casado	fundamental	2	5,25	36	5	capital
4	solteiro	Médio	...	5,73	20	10	outra
5	solteiro	fundamental	...	6,26	40	7	outra
6	casado	fundamental	0	6,86	28	0	interior
7	solteiro	Fundamental	...	6,86	41	0	interior
8	solteiro	Fundamental	...	7,39	43	4	capital
9	casado	Médio	1	7,59	34	10	capital
10	solteiro	Médio	...	7,44	23	6	outra
11	casado	Médio	2	8,12	33	6	interior
12	solteiro	Fundamental	...	8,46	27	11	capital
13	solteiro	Médio	...	8,74	37	5	outra
14	casado	Fundamental	3	8,95	44	2	outra
15	casado	médio	0	9,13	30	5	interior
16	solteiro	médio	...	9,35	38	8	outra
17	casado	médio	1	9,77	31	7	capital
18	casado	fundamental	2	9,80	39	7	outra
19	solteiro	superior	...	10,53	25	8	interior
20	solteiro	Médio	...	10,76	37	4	interior
21	casado	médio	1	11,06	30	9	outra

N.	Estado Civil	Grau de Instrução	N. de Filhos	Salário	Idade (anos)	Idade (meses)	Região de Procedência
22	solteiro	médio	...	11,59	34	2	capital
23	solteiro	fundamental	...	12,00	41	0	outra
24	casado	superior	0	12,79	26	1	outra
25	casado	médio	2	13,23	32	5	interior
26	casado	médio	2	13,60	35	0	outra
27	solteiro	fundamental	...	13,85	46	7	outra
28	casado	médio	0	14,69	29	8	interior
29	casado	médio	5	14,71	40	6	interior
30	casado	médio	2	15,99	35	10	capital
31	solteiro	superior	...	16,22	31	5	outra
32	casado	médio	1	16,61	36	4	interior
33	casado	superior	3	17,26	43	7	capital
34	solteiro	superior	...	18,75	33	7	capital
35	casado	médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	2	interior

Quando se deseja fazer um gráfico de colunas ou barras sobre **uma variável qualitativa**, deve-se fazer **a contagem direta das vezes em que a variável assume cada valor possível**. Vejamos como exemplo o grau de instrução.

Tabela 2: Exemplo de Distribuição de Frequências para uma Variável Qualitativa

Grau de Instrução	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
Fundamental	12		
Médio	18		
Superior	6		
	36		

Se você voltar à Tabela 1, poderá constatar que realmente existem 12 pessoas com ensino fundamental, 18 pessoas com ensino médio e 6 pessoas com ensino superior na pesquisa.

De posse da frequência absoluta, podemos calcular a frequência relativa de cada categoria como a razão entre a frequência absoluta da categoria e o total de elementos na pesquisa (36). Geralmente, a frequência relativa de cada classe é representada por uma letra minúscula.

$$f_1 = \frac{12}{36} = \frac{1}{3} = 0,3333 = 33,33\%$$

$$f_2 = \frac{18}{36} = \frac{1}{2} = 0,5 = 50\%$$

$$f_3 = \frac{6}{36} = \frac{1}{6} = 0,1667 = 16,67\%$$

Tabela 3: Cálculo da Frequência Relativa

Grau de Instrução	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
Fundamental	12	33,33%	
Médio	18	50,00%	
Superior	6	16,67%	
	36	100,00%	

Agora, vamos passar ao cálculo da frequência acumulada, que pode ser obtida como a soma de todas as frequências relativas menores ou iguais ao grau de instrução em apreço.

Por exemplo, para o grau de instrução fundamental, o único grau que devemos considerar é o próprio grau fundamental.

$$F_1 = f_1 = 33,33\%$$

Para o grau de instrução médio, devemos considerar tanto o grau fundamental como o médio.

$$F_2 = f_1 + f_2 = 33,33\% + 50\% = 83,33\%$$

Por fim, podemos calcular a frequência acumulada para o último grau de instrução, que é o superior. Ela leva em conta todos os graus iguais ou anteriores ao grau superior.

$$F_3 = f_1 + f_2 + f_3 = 83,33\% + 16,67\% = 100\%$$

Note que a frequência acumulada para a última classe deve ser sempre igual a 100%.

Tabela 4: Cálculo da Frequência Relativa e Acumulada

Grau de Instrução	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
Fundamental	12	33,33%	33,33%
Médio	18	50,00%	83,33%
Superior	6	16,67%	100,00%
	36	100,00%	100,00%

O cálculo da frequência acumulada só faz sentido **quando a variável é quantitativa ou ordinativa**, ou seja, existe o primeiro, o segundo grau de instrução etc.

Quando a variável não comporta ordem, não há que se falar em primeiro e segundo valor.

DIRETO DO CONCURSO

004. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte mostra as quantidades de patrimônios históricos cadastrados nos estados A e B.

patrimônios	estado A	estado B
estátuas	40	10
museus	60	20
templos	50	20
total	150	50

A partir dessa tabela, julgue o seguinte item.

As estátuas cadastradas nos estados A e B correspondem a mais de 20% dos patrimônios históricos cadastrados nesses estados.



Vamos calcular as frequências relativas referentes às estátuas nos estados A e B.

$$f_{estátuas}^A = \frac{40}{150} = 0,2667 = 26,67\%$$

$$f_{estátuas}^B = \frac{10}{50} = 0,20 = 20\%$$

Portanto, a frequência relativa das estátuas no estado B não é maior que 20%. É apenas igual.

Errado.

2.2. GRÁFICOS DE BARRAS OU COLUNAS

Os gráficos de barras (ou colunas) são muito utilizados para representar séries estatísticas. A opção por esses gráficos geralmente é uma questão apenas visual. Eles costumam trazer informações muito semelhantes.

São gráficos bastante versáteis, podendo ser utilizados facilmente com variáveis qualitativas e também com variáveis quantitativas discretas. Mas **não podem ser utilizados com variáveis quantitativas contínuas**, como veremos ainda.

São muito úteis para notar tendências estatísticas nos dados.

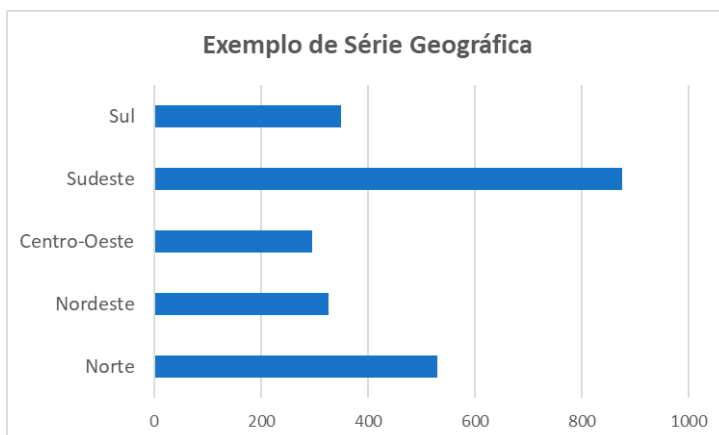
Figura 7: Exemplo de Gráfico de Colunas



Fonte: G1 e FBSP

No gráfico mostrado na Figura 7, podemos notar que a tendência de número de assassinatos no Brasil é relativamente estável. Houve um pequeno crescimento entre os anos 2012 e 2017, porém, foi normalizado retornando à faixa de 40 mil a 50 mil assassinatos por ano nos anos de 2018 e 2019.

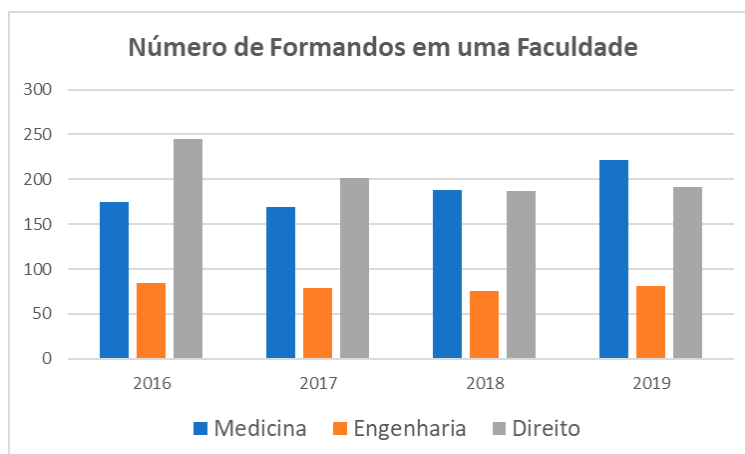
Figura 8: Exemplo de Gráfico de Barras



Nessa série geográfica, notamos uma distribuição praticamente uniforme entre as regiões do Brasil, porém, com prevalência para a região Sudeste. Portanto, memorize que os gráficos de barras são muito úteis para estudar tendências de comportamento em uma variável estatística, pois tornam essas tendências mais intuitivas.

No caso de uma série mista, podemos usar também colunas e barras mistas. Vejamos exemplos.

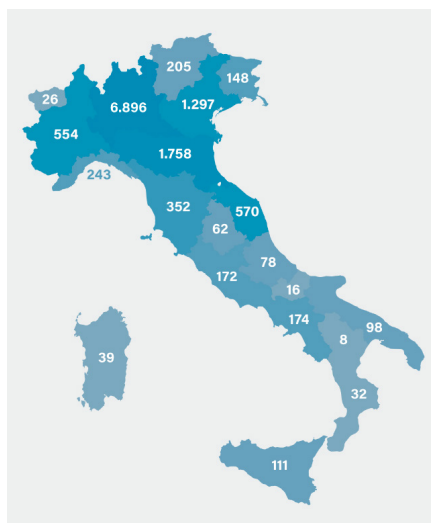
Figura 9: Exemplo de um Gráfico para uma Série Mista



No caso de séries geográficas, é também bastante comum o uso de gráficos em cima do próprio mapa da região. Esse tipo de gráfico é muito útil para investigar se regiões próximas têm comportamentos semelhantes ou se a variável em questão está amplamente distribuída.

Como exemplo, mostramos uma série geográfica distribuída pelos estados da Itália.

Figura 10: Exemplo de Gráfico para uma Série Geográfica



No caso de séries temporais, é também possível construir um diagrama de linha. Esse diagrama é especialmente útil quando a série é relativamente longa (são feitas mais de dez observações). Nesse caso, um gráfico de colunas pode ficar muito cheio. No entanto, não há nenhuma obrigatoriedade nisso.

Figura 11: Exemplo de Gráfico em Linha para uma Série Histórica

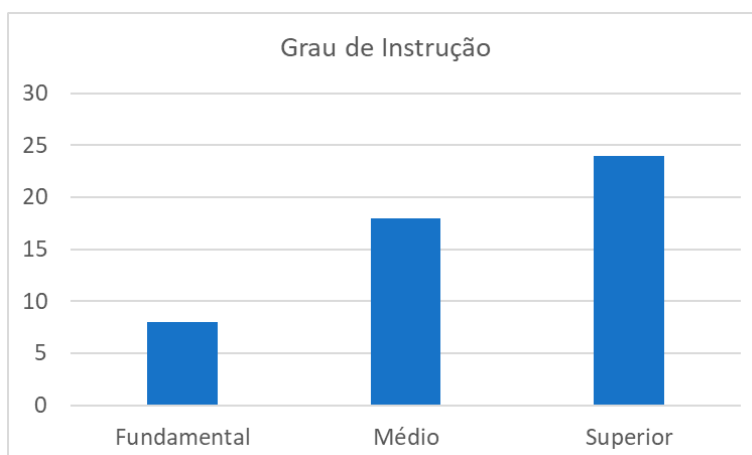


DIRETO DO CONCURSO

005. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) O gráfico de barras é adequado para a análise de variáveis qualitativas ordinais ou quantitativas discretas, pois permite investigar a presença de tendência nos dados.

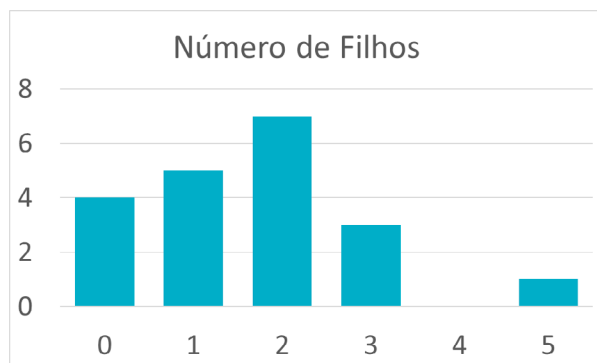


É isso mesmo. O gráfico de barras é muito útil para investigar tendências de dados. Vejamos um exemplo com uma variável qualitativa ordinal, como o grau de instrução dos trabalhadores de uma empresa.



Nesse gráfico, notamos claramente que a empresa tende a contratar mais trabalhadores com ensino superior.

Vejamos também um exemplo com uma variável discreta, como o número de filhos dos trabalhadores de uma empresa, em que podemos notar que a maioria dos funcionários tende a ter até dois filhos.



Certo.

2.3. GRÁFICO DE SETORES

O gráfico de setores, também conhecido como pizza, é muito utilizado para **comparar proporções**. São muito úteis para representar séries especificativas e geográficas, mas não há nada que os impeça de serem utilizados para séries temporais.

Nesses gráficos, geralmente, **o círculo fechado (ou a pizza completa) corresponde ao ângulo de 360° e à proporção de 100%.**

Vejamos um exemplo:

Curso	Alunos	Porcentagem
Medicina	200	40%
Engenharia	50	10%
Direito	250	50%
Total	500	

O ângulo ocupado pelas instâncias é proporcional à sua porcentagem ocupada na amostra. Basta multiplicar essa porcentagem pelo ângulo de 360°, que corresponde ao círculo completo.

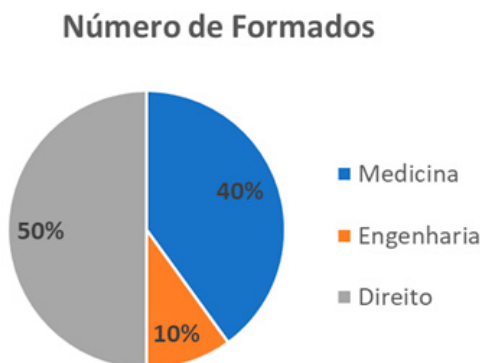
$$\theta_{Medicina} = 0,40 \cdot 360 = 144^\circ$$

$$\theta_{Engenharia} = 0,10 \cdot 360 = 36^\circ$$

$$\theta_{Direito} = 0,50 \cdot 360 = 180^\circ$$

Dessa forma, podemos traçar o seguinte gráfico de pizza.

Figura 12: Exemplo de Gráfico de Setores



De maneira geral, você pode ter em mente o seguinte esquema.

ATENÇÃO

O **gráfico de pizza** é usado preferencialmente para representar a **frequência relativa**.

O diagrama de barras (ou colunas) é usado preferencialmente para representar a frequência absoluta.

DIRETO DO CONCURSO

006. (CESPE/CEBRASPE/FUB/ESTATÍSTICO/2022/Q2444768) A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

disciplinas estudantes	2	3	4	5	6	7	8
	10	15	40	35	28	10	4

Considerando essas informações, julgue o item seguinte.

O gráfico do tipo pizza é o mais apropriado para representar os dados apresentados na tabela, visto que a variável analisada é qualitativa ordinal.



O **GRÁFICO DE BARRAS** é adequado para análise de variáveis qualitativas ORDINAIS ou quantitativas DISCRETAS, pois permite investigar a presença de tendência dos dados.

Qualitativas ORDINAIS = ordenação entre categorias. Ex.: Ensino fundamental, ensino médio etc.

Quantitativas DISCRETAS = números inteiros.

A variável da questão é quantitativa discreta.

Errado.

007. (VUNESP/UNICAMP/2022/ENGENHEIRO CIVIL/Q2491143) Em uma pesquisa realizada, 450 pessoas responderam “sim” à pergunta feita, 320 responderam “não”, e 40 pessoas não responderam à pergunta. Na construção de um gráfico de setor (também conhecido como gráfico de pizza) para representar os dados obtidos nessa pergunta, o setor circular correspondente às respostas “sim” deverá ter ângulo medindo

- a) 170.
- b) 180.
- c) 190.
- d) 200.
- e) 210.



Veja que 320 pessoas responderam “não”, 40 pessoas não responderam. Então, o número de total de pessoas que responderam à pesquisa foi:

$$\#Total = 450 + 320 + 40 = 450 + 360 = 810$$

O total de pessoas que responderam à pesquisa corresponde ao ângulo total de 360° . Por outro lado, o ângulo que deve ser ocupado pelo “Sim” é proporcional. Montemos a proporcionalidade.

$$\alpha = \frac{450}{810} \cdot 360^\circ = 200^\circ$$

Letra d.

008. (CESPE/FUB/2022/ESTATÍSTICO) A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

disciplinas	2	3	4	5	6	7	8
estudantes	10	15	40	35	28	10	4

Considerando essas informações, julgue o item seguinte.

O gráfico do tipo pizza é o mais apropriado para representar os dados apresentados na tabela, visto que a variável analisada é qualitativa ordinal.



O número de disciplinas é uma variável quantitativa discreta, tendo em vista que ela é medida como um valor numérico.

As variáveis qualitativas ordinais não são expressas por números.

Errado.

009. (CESPE/AL/CE/2011/ANALISTA LEGISLATIVO) Um levantamento foi realizado para se avaliar, por município, a quantidade X de obras que estão sob suspeita de irregularidade. Com base em uma amostra de municípios, foi obtida a distribuição de frequências mostrada na tabela acima. Com base nessas informações, julgue o item.

X	0	1	2	3	4	5
frequência absoluta	80	47	30	20	6	1

O diagrama de setores (ou pizza) é um gráfico apropriado para se representar a distribuição de probabilidade da variável X .



O gráfico de pizza é mais adequado para representar a frequência relativa, e não para a frequência absoluta. Deve ser sempre utilizado para representar proporções.

Como foram fornecidos dados sobre a frequência absoluta, o diagrama de setores não é adequado.

Errado.

010. (CESPE/TRE/ES/2011/ANALISTA) Com base na tabela acima, referente às eleições de 2010, que apresenta a quantidade de candidatos para os cargos de presidente da República, governador de estado, senador, deputado federal e deputado estadual/distrital, bem como a quantidade de candidatos considerados aptos pela justiça eleitoral e o total de eleitos para cada cargo pretendido, julgue o item a seguir.

cargo	candidatos	candidatos aptos	eleitos
presidente da República	9	9	1
governador de estado	170	156	27
senador	272	234	54
deputado federal	6.021	5.058	513
deputado estadual/distrital	15.268	13.076	1.059
total	21.640	18.533	1.658

Internet: <www.tse.gov.br> (com adaptações)

Considerando-se a representação das quantidades de eleitos para cada cargo em um gráfico de pizza, a fatia desse gráfico correspondente ao cargo de deputado federal terá ângulo superior a 120° .



Os deputados federais eleitos são 513, enquanto o total de candidatos eleitos é igual a 1.658. Devemos lembrar que o ângulo ocupado pela variável é proporcional à frequência absoluta. Se os deputados federais correspondessem a 1.658, eles ocupariam toda a pizza (360°). Portanto, o ângulo ocupado está para 513 assim como o ângulo total da pizza (360°) está para 1658.

$$\frac{\theta}{513} = \frac{360^\circ}{1658}$$

$$\therefore \theta = \frac{513}{1658} \cdot 360^\circ \cong 0,31 \cdot 360^\circ = 111^\circ$$

Observe que a conta $513/1650$ é inferior a $1/3$, portanto, realmente o ângulo ocupado pelos deputados federais eleitos é inferior a 120° .

Errado.

3. DISTRIBUIÇÃO DE FREQUÊNCIAS

Uma distribuição de frequência é uma organização da variável em **classes de valores**, e não mais em valores únicos. Nesse caso, são muito utilizados os gráficos na forma de **histogramas**.

As variáveis contínuas podem ser agrupadas em faixas. Somos acostumados a ouvir “faixa salarial de quatro a oito salários mínimos”.

A razão para isso é que ficaria muito complicado e seria uma informação muito vaga, se criássemos um diagrama de barras, sem categorizar a variável aleatória contínua. Olhe, por exemplo, a Tabela 1. Todas as pessoas possuem salários diferentes.

Que sentido faria um gráfico estatístico, mostrando que cada valor de salário é recebido por uma única pessoa?

Figura 13: Exemplo de um Diagrama de Barras para uma Variável Contínua sem o Arranjo em Classes



Qual sentido e utilidade tem o gráfico desenhado na Figura 13? Provavelmente nenhum, não acha? É justamente por isso que costumamos, na Estatística, **separar os valores de uma variável contínua em classes**.

3.1. CLASSES

Um exemplo de possíveis classes que podem ser utilizadas para essa distribuição é:

Tabela 5: Exemplo de Distribuição de Frequências em Classes

Classe de Salários
4 – 8
8 – 12
12 – 16
16 – 20
20 – 24

O símbolo |– deve ser lido como “a”. Assim, a classe 4 |– 8 é a classe das pessoas que ganham de quatro a oito salários mínimos. Se uma pessoa X ganha 6,5 salários mínimos, ela deve ser incluída nessa classe.

Para obter a frequência absoluta de cada uma das classes, precisamos retornar à Tabela 1, que contém os dados originais, e associar cada elemento a uma classe, como mostrado na Tabela 6.

Tabela 6: Contagem dos Elementos pertencentes a cada Classe

N.	Salário	Classe Correspondente		N.	Salário	Classe Correspondente
1	4,00	4,00 – 8,00		19	10,53	8,00 – 12,00
2	4,56	4,00 – 8,00		20	10,76	8,00 – 12,00
3	5,25	4,00 – 8,00		21	11,06	8,00 – 12,00
4	5,73	4,00 – 8,00		22	11,59	8,00 – 12,00
5	6,26	4,00 – 8,00		23	12,00	12,00 – 16,00
6	6,86	4,00 – 8,00		24	12,79	12,00 – 16,00
7	6,86	4,00 – 8,00		25	13,23	12,00 – 16,00
8	7,39	4,00 – 8,00		26	13,60	12,00 – 16,00
9	7,59	4,00 – 8,00		27	13,85	12,00 – 16,00
10	7,44	4,00 – 8,00		28	14,69	12,00 – 16,00
11	8,12	8,00 – 12,00		29	14,71	12,00 – 16,00
12	8,46	8,00 – 12,00		30	15,99	12,00 – 16,00

N.	Salário	Classe Correspondente		N.	Salário	Classe Correspondente
13	8,74	8,00 – 12,00		31	16,22	16,00 – 20,00
14	8,95	8,00 – 12,00		32	16,61	16,00 – 20,00
15	9,13	8,00 – 12,00		33	17,26	16,00 – 20,00
16	9,35	8,00 – 12,00		34	18,75	16,00 – 20,00
17	9,77	8,00 – 12,00		35	19,40	16,00 – 20,00
18	9,80	8,00 – 12,00		36	23,30	20,00 – 24,00

Agora, basta contar a quantidade de elementos de cada classe para chegar às frequências absolutas.

Tabela 7: Exemplo de Distribuição de Frequências em Classes

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
4 – 8	10		
8 – 12	12		
12 – 16	8		
16 – 20	5		
20 – 24	1		
Total	36		

Façamos também as contas para as frequências relativas, que podem ser obtidas como a razão entre a frequência absoluta de cada classe e o total de elementos pertencentes à amostra.

$$f_1 = \frac{10}{36} \cong 0,2778 = 27,78\%$$

$$f_2 = \frac{12}{36} \cong 0,3333 = 33,33\%$$

$$f_3 = \frac{8}{36} \cong 0,2222 = 22,22\%$$

$$f_4 = \frac{5}{36} \cong 0,1389 = 13,89\%$$

$$f_5 = \frac{1}{36} \cong 0,0278 = 2,78\%$$

Tabela 8: Exemplo de Distribuição de Frequências em Classes: Cálculo da Frequência Relativa

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
4 – 8	10	27,78%	
8 – 12	12	33,33%	
12 – 16	8	22,22%	
16 – 20	5	13,89%	
20 – 24	1	2,78%	
Total	36	100,00%	

E, por fim, vamos calcular as frequências acumuladas. Para a primeira classe, temos que a frequência acumulada corresponde somente à frequência relativa da primeira classe.

$$F_1 = f_1 = 27,78\%$$

Para a segunda classe, podemos obter a frequência acumulada como a frequência acumulada da classe anterior somada à frequência relativa da segunda classe.

$$F_2 = F_1 + f_2 = 27,78\% + 33,33\% = 61,11\%$$

Façamos o mesmo procedimento para calcular as frequências acumuladas referentes às demais classes.

$$F_3 = F_2 + f_3 = 61,11\% + 22,22\% = 83,33\%$$

$$F_4 = F_3 + f_4 = 83,33\% + 13,89\% = 97,22\%$$

$$F_5 = F_4 + f_5 = 97,22\% + 2,78\% = 100\%$$

Tabela 9: Exemplo de Distribuição de Frequências em Classes: Cálculo da Frequência Acumulada

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
4 – 8	10	27,78%	27,78%
8 – 12	12	33,33%	61,11%
12 – 16	8	22,22%	83,33%
16 – 20	5	13,89%	97,22%
20 – 24	1	2,78%	100,00%
Total	36	100,00%	100,00%

Podemos trazer algumas definições importantes sobre as classes:

Intervalo da Classe: corresponde ao conjunto de elementos contidos entre o limite inferior e superior da classe.

Um ponto interessante, que pode despertar dúvidas, **é a respeito dos extremos**. Por exemplo, uma pessoa que ganha exatamente 20 salários mínimos deve ser alocada na classe 16 |– 20 ou na classe 20 |– 24?

Para saber a resposta dessa pergunta, devemos observar as pontuações encontradas na tabela. O sinal “|” é utilizado para afirmar que um determinado elemento está incluso na classe. Vejamos exemplos:

16 |– 20: é um intervalo fechado à esquerda e aberto à direita. Assim, ele inclui o extremo inferior, mas não inclui o extremo superior. Logo, ele inclui o 16, mas não inclui o 20.

16 –|20: é um intervalo aberto à esquerda e fechado à direita. Assim, ele não inclui o extremo inferior, mas inclui o extremo superior. Logo, ele não inclui o 16, mas inclui o 20.

16 |–|20: é um intervalo fechado tanto à esquerda como à direita. Assim, ele inclui tanto o extremo inferior como o extremo superior. Logo, ele inclui tanto o 16 como o 20.

16 – 20: é um intervalo aberto tanto à esquerda como à direita. Assim, ele não inclui nem o extremo inferior nem o extremo superior. Logo, ele não inclui nem o 16 nem o 20.

No caso do exemplo em apreço, a classe 16 |– 20 não inclui o valor 20. Portanto, uma pessoa que ganha 20 salários mínimos não deve ser incluída nessa classe. Por outro lado, a classe 20 |– 24 inclui o valor 20. Portanto, uma pessoa que ganha 20 salários mínimos deve ser incluída nessa classe.

É importante destacar que as classes **devem ser complementares**. Isso significa que **um mesmo elemento não pode ser incluído em duas classes diferentes**. Portanto, a distribuição a seguir está incorreta e não pode ser utilizada em uma distribuição de frequências estatística:

Tabela 10: Exemplo de Erro e uma Distribuição de Frequências em Classes

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada
4 – 8	10	27,78%	27,78%
8 – 12	12	33,33%	61,11%
12 – 16	8	22,22%	83,33%
16 – 20	5	13,89%	97,22%
20 – 24	1	2,78%	100,00%
Total	36	100,00%	100,00%

O erro da distribuição mostrada na Tabela 10 é que uma pessoa que ganha oito salários mínimos seria incluída em duas classes diferentes: na primeira e na segunda.

Além disso, as classes podem ser organizadas em ordem crescente ou em ordem decrescente. No exemplo mostrado na Tabela 9, escolheu-se a ordem crescente, que é notadamente a mais utilizada.

Vejamos mais alguns parâmetros importantes em uma classe:

- **Amplitude (h):** corresponde à diferença entre o limite inferior e superior da classe.
- **Ponto Médio (PM):** corresponde à média aritmética entre os extremos de uma classe.

Agora, façamos as contas correspondentes à amplitude e o ponto médio das classes referidas na Tabela 9.

Tabela 11: Cálculo da Amplitude e do Ponto Médio de Diversas Classes

Classe de Salários	Ponto Médio	Amplitude
4 – 8	$(4 + 8)/2 = 6$	$8 - 4 = 4$
8 – 12	$(8 + 12)/2 = 10$	$12 - 8 = 4$
12 – 16	$(12 + 16)/2 = 14$	$16 - 12 = 4$
16 – 20	$(16 + 20)/2 = 18$	$20 - 16 = 4$
20 – 24	$(20 + 24)/2 = 22$	$24 - 20 = 4$

Podemos relacionar a amplitude e o ponto médio aos limites inferior e superior de cada classe. Basta observar que cada um deles dista exatamente de $h/2$ do ponto médio.

$$L_{sup} = PM + \frac{h}{2}$$

$$L_{inf} = PM - \frac{h}{2}$$

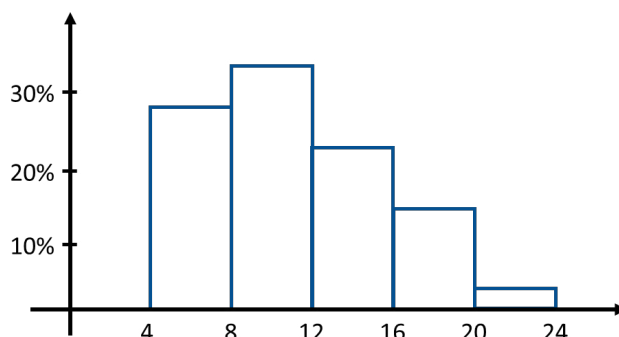
Vejamos um exemplo de como isso realmente funciona. Por exemplo, na classe 16 |– 20, temos que $PM = 18$ e $h = 4$.

$$L_{sup} = PM + \frac{h}{2} = 18 + \frac{4}{2} = 18 + 2 = 20$$

$$L_{inf} = PM - \frac{h}{2} = 18 - \frac{4}{2} = 18 - 2 = 16$$

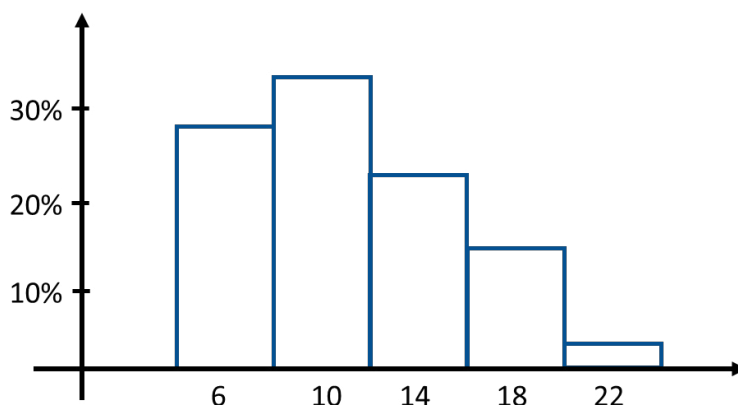
As classes podem ser representadas em histogramas **tanto pelo seu intervalo completo como por seu ponto médio**. Vejamos exemplos das duas representações:

Figura 14: Representação de um Histograma em Classes representadas por seus intervalos



Causa muita estranheza quando as classes são representadas por seus pontos médios. Mas saiba que isso pode acontecer em algumas situações. Então, esteja preparado para isso.

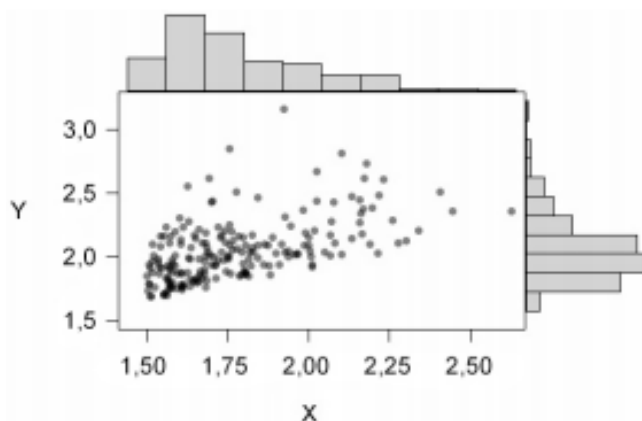
Figura 15: Representação de um Histograma em Classes representadas por seus pontos médios



É possível também que as classes sejam representadas por suas próprias amplitudes. Aliás, essa é a situação prática mais comum. Nesse caso, em vez de 6, teríamos 4 a 8 e, assim, por diante.

DIRETO DO CONCURSO

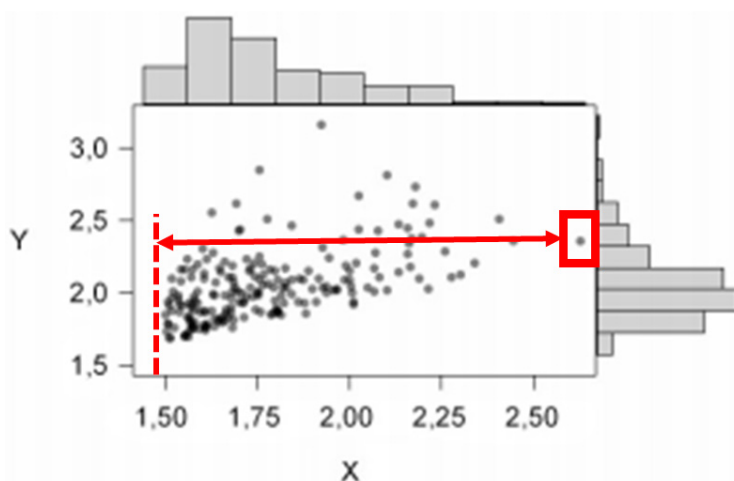
011. (CESPE/MPU/2013/ANALISTA/ESTATÍSTICA) A figura acima mostra a dispersão dos valores previstos (X, em R\$ milhões) e dos valores efetivamente gastos (Y, em R\$ milhões) em 200 obras de pavimentação em determinado estado, e os respectivos histogramas das distribuições dos valores X e Y. Com base nessas informações, julgue o item a seguir.



A amplitude total da distribuição de X foi igual ou superior a R\$ 2 milhões.



A amplitude deve ser obtida como a diferença entre o valor máximo encontrado e o valor mínimo encontrado na amostra.



Observe que o valor máximo da amostra é muito próximo (porém, menor) a 3 milhões e o valor mínimo é próximo a 1,5 milhão.

Portanto, a amplitude deve corresponder a aproximadamente 1,5 milhões – não 2 milhões como afirmado no enunciado.

Errado.

3.2. TAMANHO IDEAL DE UMA CLASSE

Embora não faça sentido que você construa uma distribuição de frequências na hora da prova, é bastante possível que você seja questionado a respeito do procedimento de construção de um esquema ótimo para a construção de classes.

De maneira geral, o número de classes e a sua amplitude podem ser adaptados, de acordo com os critérios do próprio pesquisador ou estatístico. Logo, a própria pessoa que constrói o gráfico, de maneira geral, tem liberdade para projetar a distribuição em classes que considerar mais adequada.

Porém, existem algumas regras que podem ser utilizadas para otimizar o projeto de organização em classes.

- **Definição do número de classes (K):** existem muitas regras, porém, utilizaremos no nosso curso a mais simples: a Regra do Quadrado, que estabelece que o número de classes deve ser igual à raiz quadrada do número de elementos na amostra.

$$K = \sqrt{N}$$

Por exemplo, no caso da distribuição estudada na Tabela 1, que é estudada desde o início deste capítulo, temos 36 elementos. Essa tabela será reproduzida logo a seguir na Tabela 13.

Vamos calcular o número ideal de classes ideal para esse conjunto de dados, utilizando a Regra do Quadrado.

$$K_{\text{quadrado}} = \sqrt{N} = \sqrt{36} = 6$$

Um ponto importante é observar que o número de classes deve ser necessariamente inteiro. Portanto, **se a raiz quadrada não for exata, precisamos arredondar para o número inteiro mais próximo.**

- **Cálculo da Amplitude Total (AT):** a amplitude total da amostra corresponde à diferença entre o maior valor observado (Lmax) e o menor valor observado (Lmin) na amostra inteira.

Lmax = 4,00	Lmax = 23,30
-------------	--------------

Portanto, a amplitude total da amostra pode ser obtida como a diferença entre esses dois valores.

$$AT = L_{\text{max}} - L_{\text{min}} = 23,30 - 4,00 = 19,30$$

- **Cálculo da Amplitude de Classe (h):** é obtida como a razão entre a amplitude total da amostra e o número de classes.

$$h = \frac{AT}{K}$$

Agora, façamos as contas com os valores encontrados anteriormente.

$$h = \frac{AT}{N} = \frac{19,30}{6} \cong 3,217$$

Se o número de elementos na amostra for um quadrado perfeito, podemos também utilizar diretamente a expressão:

$$h = \frac{L_{max} - L_{min}}{\sqrt{N}}$$

- **Limites das Classes:** o limite inferior da primeira classe coincide com o menor valor observado. Para obter o limite superior da classe, basta somarmos a amplitude.

$$L_{sup}^1 = 4,00 + 3,217 = 7,217$$

O limite inferior da segunda classe é igual ao limite superior da primeira classe. E, da mesma forma, o limite superior da segunda classe pode ser obtido somando-se:

$$L_{sup}^2 = 7,217 + 3,217 = 10,433$$

Observe que, na verdade, para calcular o limite superior da segunda classe, nós utilizamos o valor exato da conta $19,3/6 = 3,21666666...$

Façamos o mesmo procedimento para a terceira classe. Basta somar a amplitude de classe ao limite superior da segunda classe.

$$L_{sup}^3 = 10,433 + 3,217 = 13,65$$

Façamos a mesma conta para a quarta classe:

$$L_{sup}^4 = 13,65 + 3,217 = 16,867$$

Também para a quinta classe:

$$L_{sup}^5 = 16,867 + 3,217 = 20,083$$

E, finalmente, façamos a conta para a sexta e última classe:

$$L_{sup}^6 = 20,083 + 3,217 = 23,300$$

Vamos, agora, montar as classes calculadas pelas regras que estudamos:

Tabela 12: Obtenção das Classes de Salários pela Regra do Quadrado

Classe de Salários	Frequência Absoluta
4,00 – 7,217	
7,217 – 10,433	
10,433 – 13,65	
13,65 – 16,867	
16,867 – 20,083	
20,083 – 23,30	

Para obter as frequências absolutas, precisamos contar o número de elementos que pertencem a cada uma das classes. Para isso, devemos voltar à Tabela 1, que será reproduzida na Tabela 13.

Tabela 13: Contagem dos Elementos pertencentes a cada Classe

N.	Salário	Classe Correspondente		N.	Salário	Classe Correspondente
1	4,00	4,00 – 7,217		19	10,53	10,433 – 13,65
2	4,56	4,00 – 7,217		20	10,76	10,433 – 13,65
3	5,25	4,00 – 7,217		21	11,06	10,433 – 13,65
4	5,73	4,00 – 7,217		22	11,59	10,433 – 13,65
5	6,26	4,00 – 7,217		23	12,00	10,433 – 13,65
6	6,86	4,00 – 7,217		24	12,79	10,433 – 13,65
7	6,86	4,00 – 7,217		25	13,23	10,433 – 13,65
8	7,39	7,217 – 10,433		26	13,60	10,433 – 13,65
9	7,59	7,217 – 10,433		27	13,85	13,65 – 16,867
10	7,44	7,217 – 10,433		28	14,69	13,65 – 16,867
11	8,12	7,217 – 10,433		29	14,71	13,65 – 16,867
12	8,46	7,217 – 10,433		30	15,99	13,65 – 16,867
13	8,74	7,217 – 10,433		31	16,22	13,65 – 16,867
14	8,95	7,217 – 10,433		32	16,61	13,65 – 16,867
15	9,13	7,217 – 10,433		33	17,26	16,867 – 20,083
16	9,35	7,217 – 10,433		34	18,75	16,867 – 20,083
17	9,77	7,217 – 10,433		35	19,40	16,867 – 20,083
18	9,80	7,217 – 10,433		36	23,30	20,083 – 23,30

Agora, basta fazer a contagem do número de elementos em cada classe.

Tabela 14: Obtenção das Classes de Salários pela Regra do Quadrado

Classe de Salários	Frequência Absoluta
4,00 – 7,217	7
7,217 – 10,433	11
10,433 – 13,65	8
13,65 – 16,867	6
16,867 – 20,083	3
20,083 – 23,30	1

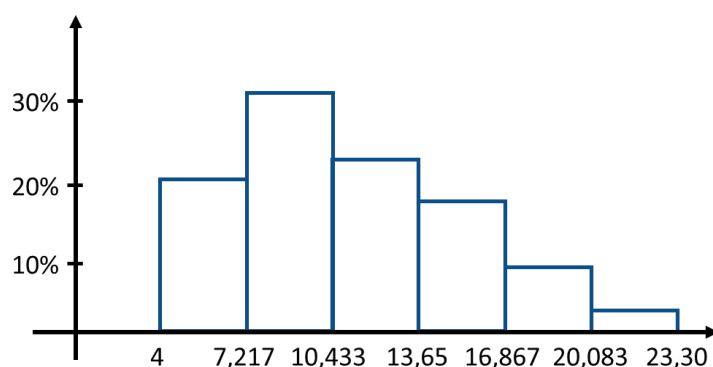
Agora, basta obter as frequências relativas, observando que o total de elementos na amostra é igual a 36.

Tabela 15: Obtenção das Classes de Salários pela Regra do Quadrado

Classe de Salários	Frequência Absoluta	Frequência Relativa
4,00 – 7,217	7	$7/36 = 0,1944 = 19,44\%$
7,217 – 10,433	11	$11/36 = 0,3056 = 30,56\%$
10,433 – 13,65	8	$8/36 = 0,2222 = 22,22\%$
13,65 – 16,867	6	$6/36 = 0,1667 = 16,67\%$
16,867 – 20,083	3	$3/36 = 0,0833 = 8,33\%$
20,083 – 23,30	1	$1/36 = 0,0278 = 2,78\%$

Com base nas frequências relativas obtidas na Tabela 15, podemos desenhar o diagrama de frequências ilustrado na Figura 16.

Figura 16: Distribuição de Frequências com o uso da Regra do Quadrado



O grande problema desse diagrama é que, embora ele tenha sido construído com conceitos matemáticos, ele não é agradavelmente visualmente para o ser humano.

Observe que o fato de as classes serem números não inteiros torna o diagrama mais carregado e de compreensão um pouco mais difícil.

Por isso, é muito comum que o pesquisador faça adaptações a esse método. Mostraremos a seguir uma técnica muito utilizada.

Ela consiste em alterar a amplitude total e/ou o número de classes, de modo que ambos sejam inteiros e que a amplitude total seja divisível pelo número de classes.

É importante também **que a amplitude total seja sempre alterada para cima. Não é possível reduzi-la**, tendo em vista que isso faria que alguns elementos não fossem incluídos em nenhuma das classes.

No caso em apreço, vimos que a amplitude total obtida foi igual a 19,30 e que o número de classes calculado foi igual a 6.

Temos duas opções:

- Podemos arredondar a amplitude total para 20,00 e alterar o número de classes para 5. Dessa forma, teríamos uma amplitude de classe dada por:

$$h = \frac{AT}{K} = \frac{20}{5} = 4$$

- Podemos arredondar a amplitude total para 24,00 e manter o número de classes igual a 6. Dessa forma, a amplitude total será:

$$h = \frac{AT}{K} = \frac{24}{6} = 4$$

Ambas as ideias são válidas. Eu, particularmente, prefiro a primeira, pois altera menos a amplitude total da classe. E foi assim que construí a distribuição de frequências que usamos neste capítulo.

O próximo passo é adaptar os limites mínimo e máximo da amostra para adaptá-los ao novo valor de amplitude de classe. É conveniente utilizar sempre números inteiros.

Como, no exemplo em estudo, o menor valor possível para a amostra já é um número inteiro (4,00), podemos construir assim os limites de classe. Começamos com o primeiro limite inferior igual a 4.

Já vimos que o limite superior da primeira classe pode ser obtido somando-se a amplitude ao limite inferior.

$$L_{sup}^1 = 4,00 + 4,00 = 8,00$$

O limite inferior da segunda classe é igual ao limite superior da primeira classe. E, da mesma forma, o limite superior da segunda classe pode ser obtido somando-se:

$$L_{sup}^2 = 8,00 + 4,00 = 12,00$$

Façamos o mesmo procedimento para a terceira classe. Basta somar a amplitude de classe ao limite superior da segunda classe:

$$L_{sup}^3 = 12,00 + 4,00 = 16,00$$

E façamos a mesma conta para a quarta classe:

$$L_{sup}^4 = 16,00 + 4,00 = 20,00$$

E, finalmente, para a quinta e última classe:

$$L_{sup}^5 = 20,00 + 4,00 = 24,00$$

E, assim, chegamos às classes que já haviam sido trabalhadas anteriormente.

Tabela 16: Exemplo de Distribuição de Frequências em Classes

Classe de Salários	Frequência Absoluta
4 – 8	10
8 – 12	12
12 – 16	8
16 – 20	5
20 – 24	1
Total	36

012. (IFTO/2023) Dados os valores de uma população: $X_1 = 4$; $X_2 = 6$; $X_3 = 8$ e $X_4 = 10$. A amplitude total dessa população será:

- a) 3.
- b) 8.
- c) 4.
- d) 2.
- e) 6.



A amplitude corresponde à diferença entre o valor máximo e o mínimo entre os observados na amostra.

$$A = 10 - 4 = 6$$

Letra e.

013. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela a seguir mostra as quantidades de bibliotecas públicas presentes em 20 microrregiões brasileiras.

90	66	78	82
77	60	64	90
87	85	67	91
82	70	81	80
69	78	90	67

A partir desses dados, pretende-se construir um gráfico de distribuição de frequências com quatro classes de igual amplitude. Os valores mínimo e máximo de cada classe devem ser números inteiros.

Considerando essas informações, julgue o item subsequente, relativo ao gráfico de distribuição a ser apresentado.

A amplitude de cada classe deverá ser superior a 6.



Por força do enunciado, os valores mínimo e máximo de cada classe devem ser **números inteiros**, portanto, devemos utilizar as aproximações que aprendemos nessa última seção. A quantidade de classes pode ser obtida pela Regra do Quadrado. Como temos 20 elementos na amostra, temos:

$$K = \sqrt{20}$$

O número de classes é necessariamente um número inteiro. Portanto, devemos aproximar para o inteiro mais próximo, que é $4 = \sqrt{16}$.

Logo, a distribuição em frequência terá 4 classes.

90	66	78	82
77	60	64	90
87	85	67	91
82	70	81	80
69	78	90	67

Logo, a amplitude total da amostra é:

$$AT = 91 - 60 = 31$$

Por força do enunciado, devemos modificar a amplitude total, de modo que ela seja divisível pelo número de classes.

$$AT = 32$$

Agora, podemos calcular a amplitude das classes.

$$h = \frac{AT}{K} = \frac{32}{4} = 8$$

Realmente, é superior a 6.

Certo.

3.3. FREQUÊNCIAS ACUMULADAS DECRESCENTES

Já vimos diversos exemplos de frequências acumuladas crescentes, em que a frequência acumulada da última classe é sempre igual a 100%.

No caso das frequências acumuladas decrescentes, a primeira classe terá frequência acumulada igual a 100%. Para a segunda classe, devemos subtrair a frequência relativa da primeira classe.

$$F_2 = F_1 - f_1 = 100\% - 27,78\% = 72,22\%$$

Para as próximas classes, devemos sempre usar que a frequência acumulada decrescente da classe posterior é igual à frequência acumulada da classe anterior menos a frequência relativa da classe anterior.

$$F_3 = F_2 - f_2 = 72,22\% - 33,33\% = 38,89\%$$

$$F_4 = F_3 - f_3 = 38,89\% - 22,22\% = 16,67\%$$

$$F_5 = F_4 - f_4 = 16,67\% - 13,89\% = 2,78\%$$

Tabela 17: Exemplo de Distribuição de Frequências em Classes: Cálculo da Frequência Acumulada

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada Decrescente
4 – 8	10	27,78%	100,00%
8 – 12	12	33,33%	72,22%
12 – 16	8	22,22%	38,89%

Classe de Salários	Frequência Absoluta	Frequência Relativa	Frequência Acumulada Decrescente
16 – 20	5	13,89%	16,67%
20 – 24	1	2,78%	2,78%
Total	36	100,00%	100,00%

Nessa situação, a frequência acumulada decrescente da última classe será sempre igual à sua própria frequência relativa.

3.4. HISTOGRAMAS COM CLASSES DE AMPLITUDE DESIGUAIS

Outro tipo de histograma que precisamos ficar de olho é quando **os intervalos são desiguais**. Mostraremos como exemplo uma entrevista com 250 empregados para se descobrir o tamanho das empresas em que eles trabalhavam. Os resultados dessa pesquisa foram anotados na tabela a seguir.

Tabela 18: Pesquisa sobre o Tamanho das Empresas em que diversas pessoas trabalham

Número de Empregados	Frequência Absoluta	Proporção
0 a 10	5	2%
10 a 20	20	8%
20 a 30	35	14%
30 a 40	40	16%
40 a 60	50	20%
60 a 80	30	12%
80 a 100	20	8%
100 a 140	20	8%
140 a 180	15	6%
180 a 260	15	6%
Total	250	–

Além dessas colunas já conhecidas, precisamos introduzir outras duas: a amplitude de cada uma das categorias. A amplitude corresponde à diferença entre o maior e o menor valor dentro dessas categorias.

Por exemplo, na categoria de 180 a 260, a amplitude é igual a $260 - 180 = 80$. Já a amplitude da primeira categoria de 0 a 10 é igual a 10.

Tabela 19: Tratamento dos Dados para a Construção do Histograma

Número de Empregados	Frequência Absoluta	Amplitude
0 a 10	5	10 – 0 = 10
10 a 20	20	20 – 10 = 10
20 a 30	35	30 – 20 = 10
30 a 40	40	40 – 30 = 10
40 a 60	50	60 – 40 = 20
60 a 80	30	80 – 60 = 20
80 a 100	20	100 – 80 = 20
100 a 140	20	140 – 100 = 40
140 a 180	15	180 – 140 = 40
180 a 260	15	260 – 180 = 80
Total	250	–

Nesse tipo de distribuição, é muito interessante calcular as densidades de frequência. São elas:

- **Densidade de Frequência Absoluta:** corresponde à razão entre a frequência absoluta de cada classe e a sua amplitude.
- **Densidade de Frequência:** essa é a mais utilizada, corresponde à razão entre a frequência da classe relativa da classe e a sua amplitude. Podemos escrever matematicamente:

$$d_i = \frac{f_i}{A_i}$$

Nessa expressão, temos que d_i , f_i e A_i são, respectivamente, a densidade de frequência, a frequência relativa e a amplitude de cada categoria.

Após feito esse tratamento de dados, podemos construir a tabela com mais colunas.

Tabela 20: Tratamento dos Dados para a Construção do Histograma

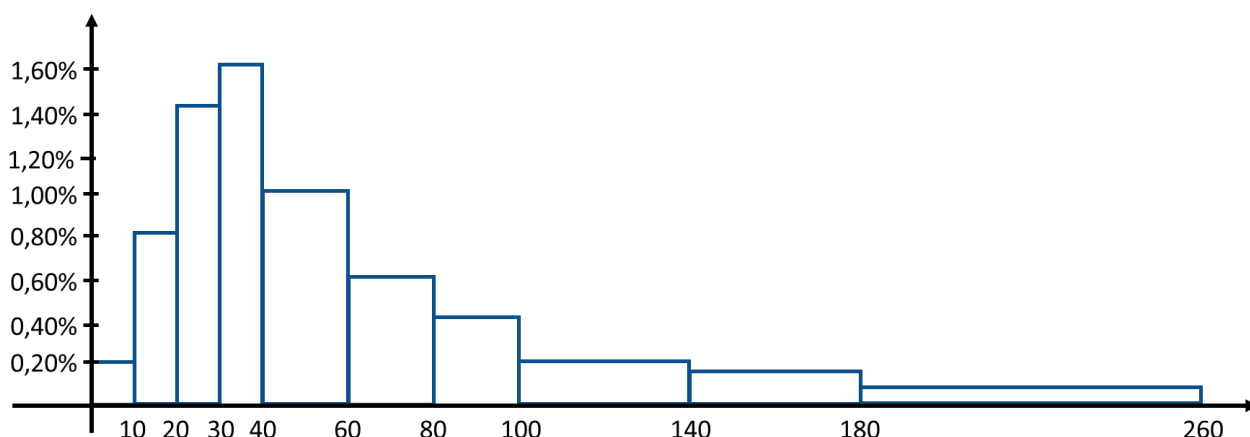
Número de Empregados	Frequência Absoluta	Amplitude	Densidade Absoluta	Frequência Relativa	Densidade de Frequência
0 a 10	5	10	5/10 = 0,50	2%	2%/10 = 0,20%
10 a 20	20	10	20/10 = 2,00	8%	8%/10 = 0,80%
20 a 30	35	10	35/10 = 3,50	14%	14%/10 = 1,40%
30 a 40	40	10	40/10 = 4,00	16%	16%/10 = 1,60%
40 a 60	50	20	50/20 = 2,50	20%	20%/20 = 1,00%
60 a 80	30	20	30/20 = 1,50	12%	12%/20 = 0,60%

Número de Empregados	Frequência Absoluta	Amplitude	Densidade Absoluta	Frequência Relativa	Densidade de Frequência
80 a 100	20	20	$20/20 = 1,00$	8%	$8\%/20 = 0,40\%$
100 a 140	20	40	$20/40 = 0,50$	8%	$8\%/40 = 0,20\%$
140 a 180	15	40	$15/40 = 0,38$	6%	$6\%/40 = 0,15\%$
180 a 260	15	80	$15/80 = 0,19$	6%	$6\%/80 = 0,075\%$
Total	250	-	-	-	

Em geral, plota-se o gráfico da densidade de frequência. Nesse gráfico, a amplitude de cada categoria fica bem evidenciada pela largura da barra correspondente a ela.

Para fins didáticos, colocamos também a frequência relativa de cada categoria como legenda no gráfico.

Figura 17: Exemplo de Histograma com Distribuições de Frequência com Classes de Amplitudes Desiguais



Perceba que a altura da barra nada tem a ver com a frequência relativa, mas sim com a densidade de frequência.

A frequência relativa da classe pode ser obtida **como o produto entre a densidade de frequência e a amplitude da classe**.

$$f_i = d_i \cdot A_i$$

Ambas as informações podem ser extraídas diretamente do gráfico, observando a seguinte correspondência:

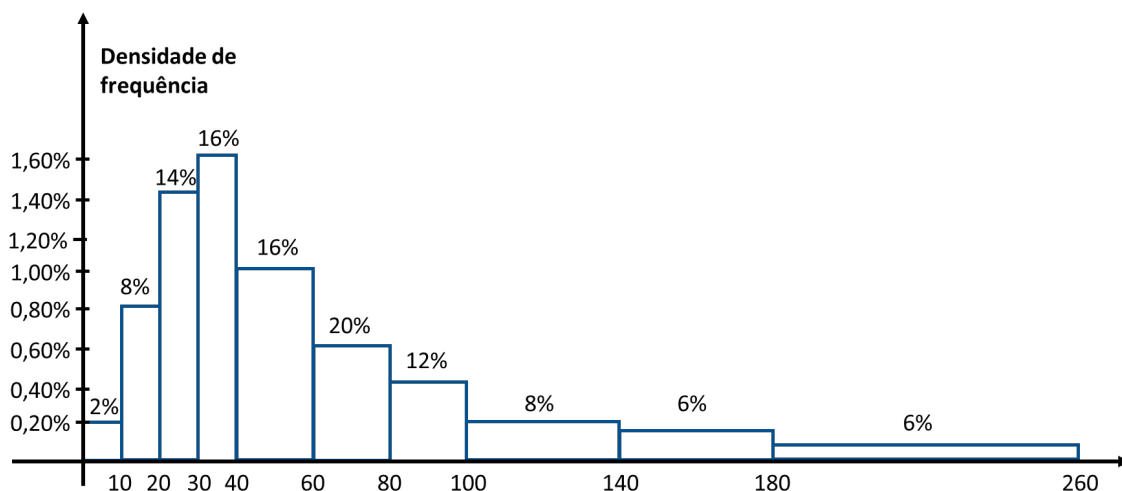
- **Altura das Barras:** representa a densidade de frequência da classe.
- **Largura das Barras:** representa a amplitude da classe.

Vejamos exemplos de como isso se aplica para as classes mostradas na Figura 17.

0 – 10	$f_1 = d_1 \cdot A_1 = 0,20 \cdot 10 = 2\%$
10 – 20	$f_2 = d_2 \cdot A_2 = 0,80 \cdot 10 = 8\%$
20 – 30	$f_3 = d_3 \cdot A_3 = 1,40 \cdot 10 = 14\%$
30 – 40	$f_4 = d_4 \cdot A_4 = 1,60 \cdot 10 = 16\%$
40 – 60	$f_5 = d_5 \cdot A_5 = 1,00 \cdot 20 = 20\%$
60 – 80	$f_6 = d_6 \cdot A_6 = 0,60 \cdot 20 = 12\%$
80 – 100	$f_7 = d_7 \cdot A_7 = 0,40 \cdot 20 = 8\%$
100 – 140	$f_8 = d_8 \cdot A_8 = 0,20 \cdot 40 = 8\%$
140 – 180	$f_9 = d_9 \cdot A_9 = 0,015 \cdot 40 = 6\%$
180 – 260	$f_{10} = d_{10} \cdot A_{10} = 0,075 \cdot 80 = 6\%$

Olha só, que interessante! A classe de 40 a 60 é mais numerosa, tendo 20% dos empregados, porém, a sua barra tem altura menor que outras classes, como as classes de 30 a 40 e 40 a 60.

Figura 18: Cálculo das Frequências Relativas a partir da Densidade de Frequência



Portanto, tenha em mente que **a altura do histograma corresponde à densidade de frequência**, e **não à frequência relativa da classe**.

A utilidade do conceito de **densidade de frequência** é que, muitas vezes, uma classe só contém um grande número de elementos porque é muito ampla.

É o caso da classe 40 a 60. Essa classe engloba 20% dos funcionários somente porque é muito ampla. A classe de 30 a 40 consegue englobar 16%, mesmo tendo metade da amplitude da classe anterior. Por isso, a sua barra ficou bem maior no gráfico.

Por isso, o diagrama de densidade de frequência traz uma **ideia bem mais realista** a respeito do papel de uma classe para a distribuição da variável estatística.

DIRETO DO CONCURSO

014. (FCC/TRT/14ª REGIÃO (RO E AC)/2012/ANALISTA JUDICIÁRIO/ESTATÍSTICA) De um histograma e uma tabela de frequências absolutas, elaborados para analisar a distribuição dos salários dos empregados em uma empresa, obtém-se a informação que 24 empregados ganham salários com valores pertencentes ao intervalo (2.000; 4.000], em reais, que apresenta uma densidade de frequência de $0,75 \times 10^{-4}(\text{R\$})^{-1}$.

Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela amplitude deste intervalo. Em um intervalo do histograma que está sendo analisado, com uma amplitude de R\$ 3.000,00 e uma densidade de frequência de $1 \times 10^{-4}(\text{R\$})^{-1}$, tem-se que o correspondente número de empregados é igual a:

- a) 40.
- b) 36.
- c) 30.
- d) 48.
- e) 42.



Na classe com 24 empregados, podemos calcular a sua frequência relativa como o produto entre a densidade de frequência e a amplitude da classe.

$$f = d \cdot A = 0,75 \cdot 10^{-4} \cdot 2000 = 0,15 = 15\%$$

Como sabemos o número de empregados na classe igual a 24 e a sua frequência relativa de 15%, podemos dizer que 15% do total do número de empregados da empresa (N) é igual a 24 empregados. Dessa forma, temos:

$$0,15 \cdot N = 24$$

$$\therefore N = \frac{24}{0,15} = \frac{2400}{15} = 160 \text{ empregados}$$

Portanto, a empresa tem um total de 160 empregados.

A frequência relativa da outra classe também deve ser obtida como o produto entre a densidade de frequência e a amplitude da classe.

$$f = d \cdot A = 1 \cdot 10^{-4} \cdot 3000 = 0,30 = 30\%$$

Por fim, a frequência absoluta dessa classe é igual ao produto da frequência relativa pelo total de empregados da empresa.

$$n = 0,30 \cdot 160 = 48$$

Letra d.

3.5. POLÍGONO DE FREQUÊNCIAS

O polígono de frequências é constituído ligando os pontos médios de um histograma.

Figura 19: Polígono de Frequências



É muito comum vermos também uma extrapolação que pode ser construída no Polígono de Frequências. Nessa extrapolação, consideramos dois pontos adicionais, com frequência nula. Esses pontos correspondem aos pontos médios de uma classe anterior à primeira classe e uma classe posterior à última classe, ambas vazias.

No exemplo em apreço, o ponto médio da primeira classe é 6, o ponto médio da última classe é 22, e a amplitude das classes é igual a 4. Assim, podemos adicionar dois pontos:

$$P_0 = 6 - 4 = 2$$

$$P_f = 22 + 4 = 26$$

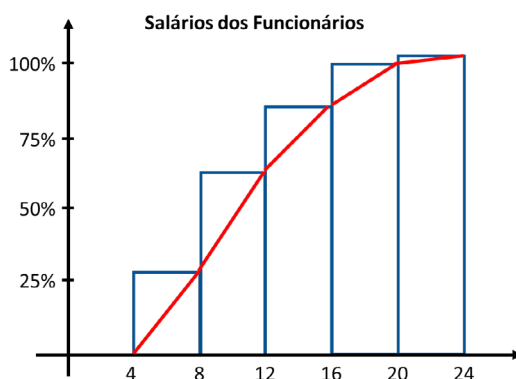
Figura 20: Polígono de Frequências Extrapolado



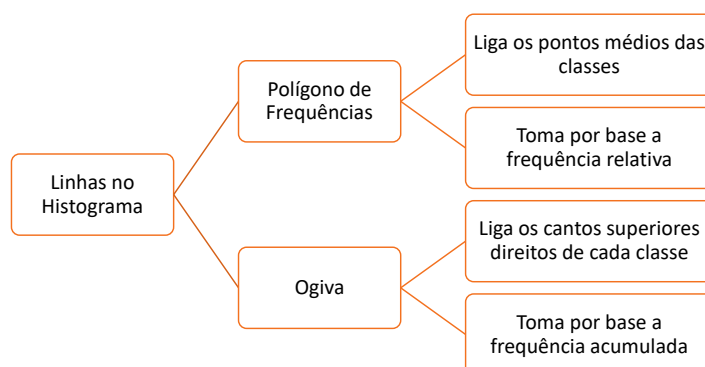
A ogiva é uma linha construída com base no histograma de frequência acumulada. A linha de ogiva é construída ligando-se os cantos superiores direitos (que correspondem ao ponto de máximo) de cada classe do histograma.

A linha de ogiva se inicia no eixo das abscissas, a partir do menor valor da primeira classe da distribuição de frequências. Perceba que o primeiro segmento corresponde exatamente à diagonal do primeiro retângulo do histograma.

Figura 21: Exemplo de Ogiva



Perceba as diferenças entre a ogiva e o polígono de frequências.



A ogiva também é chamada de **Ogiva de Galton**, em homenagem a seu criador. Como curiosidade, ele também descobriu nossas diferenças nas impressões digitais e inventou seu teste. Era primo de Charles Darwin.

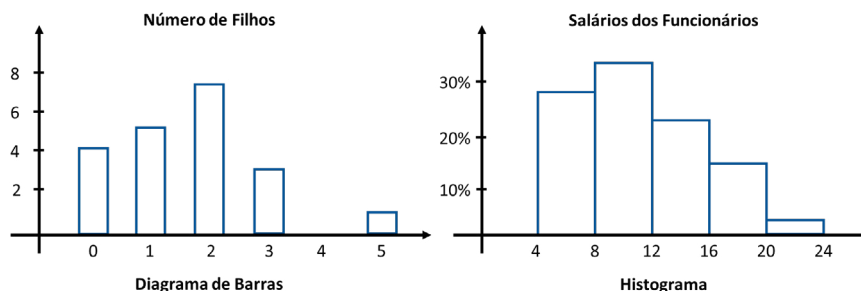
3.6. HISTOGRAMA X DIAGRAMA DE BARRAS

Devemos observar que existe uma importante diferença entre o histograma e o diagrama de barras:

- O histograma é utilizado para distribuições de frequências em classes, o que normalmente acontece com variáveis contínuas.
- O diagrama de barras só pode ser utilizado para variáveis discretas, pois só se aplica a frequências propriamente ditas, e não a distribuições em classes.

É comum se dizer também que o histograma **não tem espaços vazios**. Para entender essa diferença, vamos comparar um diagrama de barras sobre o Número de Filhos (variável discreta) e os Salários dos Funcionários (variável contínua) de uma empresa.

Figura 22: Comparação entre um Diagrama de Barras e um Histograma



O diagrama de barras que representa a variável discreta tem espaços vazios, porque não é possível que uma pessoa tenha um número fracionário de filhos. Não é possível que ela tenha 0,5 filho. Ou ela tem 0 filhos ou ela tem 1 filho.

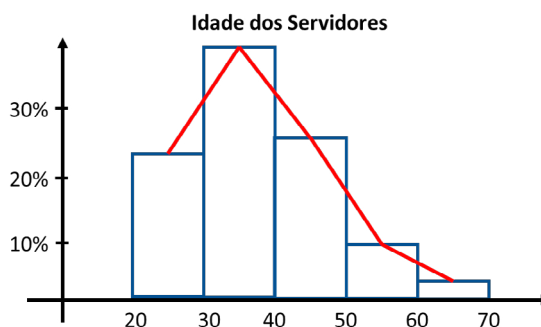
Por outro lado, o histograma que representa a variável contínua não tem espaços vazios, porque as classes em que foram divididas a variável aleatória realmente abrangem todos os valores. Dessa forma, podemos localizar no histograma as pessoas que ganham 5,67 salários mínimos – elas estão na classe de 4 a 8.

Vale ressaltar que **isso não significa que não seja possível construir um histograma com uma variável discreta**. É sempre possível construir o histograma para uma variável que possa ser **separada em classes**. Normalmente, isso acontece quando, embora a variável seja discreta, ela pode assumir uma quantidade muito grande de valores.

Por exemplo, a idade dos funcionários de uma empresa é uma variável discreta – uma pessoa pode ter 18 anos ou 27 anos ou 58 anos. Porém, nesse caso, faz sentido criar faixas etárias para facilitar a classificação.

Na Figura 23, mostramos um histograma com as idades dos funcionários de uma repartição militar. Notamos uma sensível queda a partir dos 50 anos, indicando o início das aposentadorias.

Figura 23: Idade dos Funcionários de um Órgão Público



DIRETO DO CONCURSO

015. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4) A tabela mostrada apresenta a quantidade de detentos no sistema penitenciário brasileiro por região em 2013. Nesse ano, o déficit relativo de vagas — que se define pela razão entre o déficit de vagas no sistema penitenciário e a quantidade de detentos no sistema penitenciário — registrado em todo o Brasil foi superior a 38,7%, e, na média nacional, havia 277,5 detentos por 100 mil habitantes.

região	quantidade de detentos no sistema penitenciário brasileiro (mil pessoas)	déficit de vagas no sistema penitenciário (mil vagas)	população brasileira (milhões de habitantes)
Norte	37	13	17
Centro-oeste	51	24	15
Nordeste	94	42	55
Sudeste	306	120	85
Sul	67	16	28
total	555	215	200

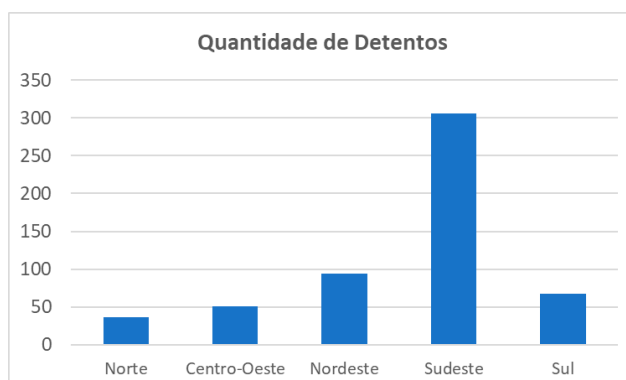
Ministério da Justiça — Departamento Penitenciário Nacional — Sistema Integrado de Informações Penitenciárias – InfoPen, Relatório Estatístico Sintético do Sistema Prisional Brasileiro, dez./2013 Internet: <www.justica.gov.br> (com adaptações)

Com base nessas informações e na tabela apresentada, julgue o item a seguir.

Na análise exploratória, o histograma é um gráfico adequado para descrever a distribuição da quantidade de detentos por região em 2013.



Observe que a variável aleatória é descrita em função da região. Portanto, não se trata de uma variável categorizada, mas sim uma variável nominativa. Desta forma, o gráfico mais adequado é o diagrama de barras.



Errado.

4. OUTROS DIAGRAMAS ESTATÍSTICOS

4.1. DIAGRAMA DE RAMOS E FOLHAS

O diagrama de ramos e folhas é, na verdade, um tipo interessante de histograma.

Consideremos um exemplo de construção desse diagrama.

Um biólogo anotou as idades dos machos e fêmeas dos habitantes de uma espécie animal:

- **Idades dos Machos:** 9 9 8 10 10 10 11 15 15 17 18 19 19 21 21 21 23 32 32 33 34 34 34 35 45 45 51 55 56 58 68 70 81 83 115 177
- **Idades das Fêmeas:** 8 9 11 17 17 19 20 44 45 37 37 37 38 81 82 83 100 104

O diagrama será construído da seguinte forma. Agruparemos os animais pela sua idade. As categorias de idades serão contadas de 10 em 10. Sendo assim, teremos as categorias 0 a 9, 10 a 19 e, assim, por diante.

Em cada categoria, destacaremos as unidades e colocaremos os demais algarismos numa coluna no meio.

À esquerda, colocaremos os machos e, à direita, as fêmeas.

Com isso, temos o diagrama a seguir:

Figura 24: Exemplo de Diagrama de Ramos e Folhas

Machos		Fêmeas
9 9 8	0	8 9
9 9 8 7 5 5 1 0 0 0	1	1 7 7 9
3 1 1 1	2	0
5 4 4 4 3 2 2	3	7 7 7 8
5 5	4	4 5
8 6 5 1	5	
8	6	
0	7	
3 1	8	1 2 3
	9	
	10	0 4
5	11	
	12	
	13	
	14	
	...	
7	17	

Ao contrário dos histogramas, o diagrama de Ramos e Folhas **retém os dados originais** até, no mínimo, dois algarismos, e põe os dados em ordem, facilitando, assim, a inferência.

DIRETO DO CONCURSO



016. (FCC/TJ/AP/2009) O diagrama de ramo e folhas a seguir corresponde às idades dos 40 funcionários de um setor de um órgão público em uma determinada data.

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

A soma da mediana e da moda destas idades é igual a

- a) 67,0.
- b) 66,5.
- c) 66,0.
- d) 65,5.
- e) 65,0.



No diagrama de ramos e folhas, a coluna da esquerda representa o algarismo das dezenas, enquanto à direita, temos o algarismo das unidades das idades dos funcionários.

A moda corresponde à observação mais frequente. Nesse caso, trata-se da observação $x = 33$, que aparece 4 vezes.

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

Observe que, embora o número 3 apareça outras vezes na coluna da direita, ele representa outras idades. Por exemplo:

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

Para a mediana, devemos notar que, como a amostra tem 40 elementos, o ponto médio a ser tomado é:

$$n = \frac{1 + 40}{2} = \frac{41}{2} = 20,5$$

Portanto, devemos tomar os elementos x_{20} e x_{21} . Uma das vantagens do diagrama de ramos e folhas é que ele já traz os elementos ordenados em ordem crescente. Por isso, podemos riscar os primeiros elementos do diagrama até encontrar o x_{20} e o x_{21} .

Na primeira linha, riscamos os 3 elementos; na segunda linha, riscamos 10 elementos; na terceira linha, riscamos mais 6 elementos, totalizando 19 elementos riscados. Os dois elementos seguintes são os que nos interessam.

1	8 8 9 → 3
2	0 1 1 2 2 2 7 8 8 9 → 10
3	1 3 3 3 3 4 4 5 6 7 8 8 8
4	0 1 2 2 3 4 8 9
5	1 5 8
6	2 5

$$Md = \frac{34 + 34}{2} = \frac{68}{2} = 34$$

Por fim, a soma da média com a mediana é igual a:

$$Me + Md = 33 + 34 = 67$$

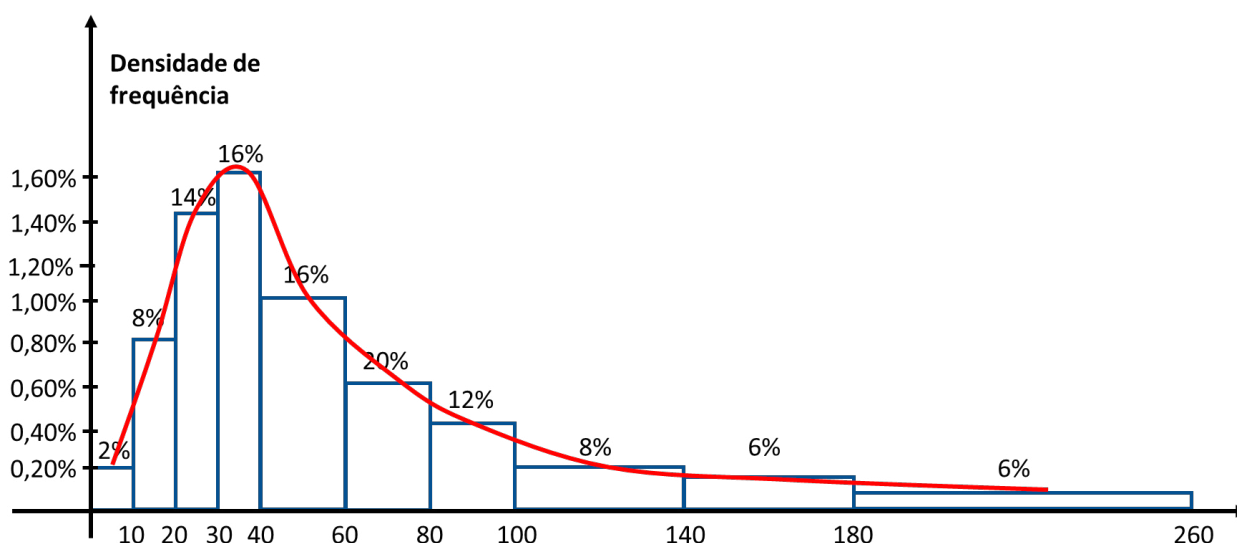
Letra a.

4.2. LINHA DE TENDÊNCIA

Uma linha de tendência mostra o comportamento aproximado do histograma por uma linha. Existem vários métodos de obtenção dessa linha de tendência, mas que não serão discutidos aqui, pois são temas mais avançados da Estatística.

Essa linha também é conhecida como **histograma alisado**.

Figura 25: Linha de Tendência no Histograma



O grande interesse pela linha de tendência é que ela evidencia o comportamento do histograma. Na Figura 25, a linha de tendência deixa claro que a densidade de frequência tende a diminuir a partir da classe 40 |– 60.

A linha de tendência torna o histograma mais suave e, com isso, o seu comportamento geral é mais facilmente visualizado.

Nas provas de concursos, o que é mais interessante é você entender a linha de tendência e para que ela serve. Não se preocupe em resolver equações para obtê-la.

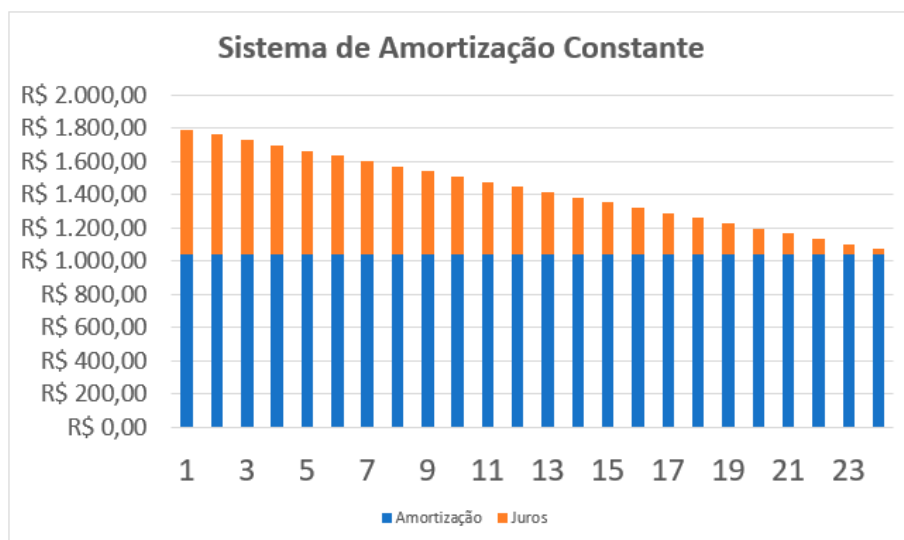
4.3. HISTOGRAMAS OU DIAGRAMAS DE COLUNAS MULTIVARIADOS

É muito comum agrupar vários dados em um mesmo histograma ou diagrama de colunas. Por exemplo, no nosso curso de Matemática Financeira, eu faço bastante isso para ilustrar um financiamento.

Em um financiamento, as parcelas são compostas por duas partes: juros e amortizações.

Existem duas formas de empilhar. Por exemplo, no caso das parcelas do Sistema de Amortização Constante, o foco foi o total da classe. Perceba que nesse gráfico da Figura 26 ficou bem visível a soma dos juros com a amortização.

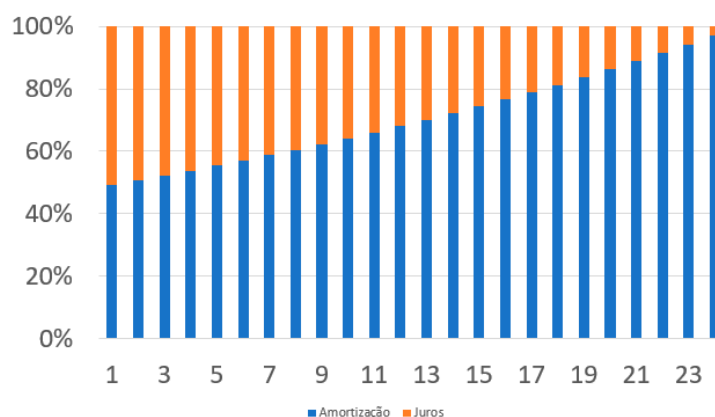
Figura 26: Diagrama de Colunas Multivariado



O interessante desse gráfico empilhado é que torna fácil a visualização da soma das duas variáveis – no caso, os juros e a amortização somando as parcelas.

Outra forma de empilhar os gráficos é por meio de uma proporção.

Figura 27: Diagrama de Colunas Multivariado com Proporções



4.4. GRÁFICOS DE LINHA E DISPERSÃO

O gráfico de linha é o gráfico mais utilizado na Matemática para representar funções.

Esse gráfico pode ser utilizado sempre que a variável estatística em estudo **somente puder assumir um único valor** para cada categoria da variável independente.

Por exemplo, o comportamento da taxa de juros do Brasil (Taxa Selic) em função do tempo. Para cada momento, o Brasil só pode ter uma única de taxa de juros. Por isso, o gráfico mais adequado para representá-lo é o gráfico de linha.

Figura 28: Exemplo de Gráfico de Linha

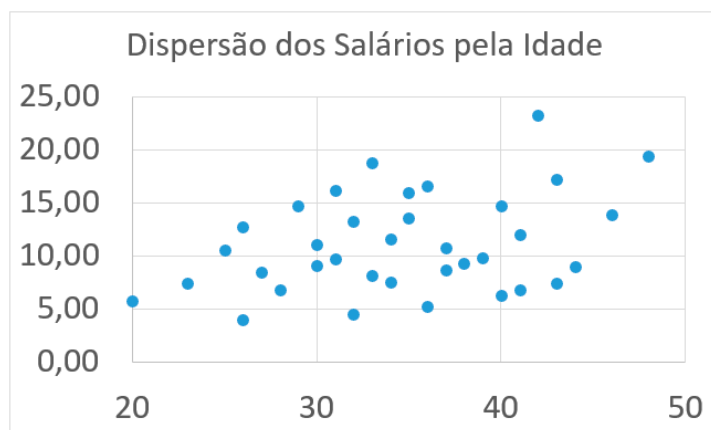


Por outro lado, deve-se usar um gráfico de dispersão quando a variável estatística puder assumir **diversos valores** para a mesma variável independente.

Para ilustrar a diferença entre os dois gráficos, vamos voltar aos dados apresentados na Tabela 1 e montar um gráfico sobre o salário das pessoas em função da idade.

É fácil perceber que duas pessoas com a mesma idade podem ter salários diferentes. Por isso, o gráfico adequado é o gráfico de dispersão.

Figura 29: Gráfico de Dispersão de Salários pela Idade

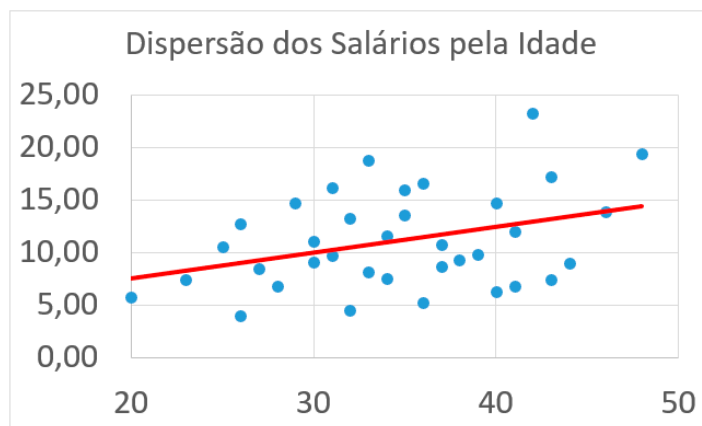


Nesse gráfico, temos a representação de todas as pessoas da amostra. Cada ponto representa uma pessoa por sua idade e por seu salário.

O interessante do gráfico de dispersão é que ele já nos dá várias ideias sobre a relação entre as duas variáveis.

Por exemplo, você pode perceber que há uma ligeira **tendência** de que o salário cresça com a idade. Para averiguá-la melhor, podemos traçar uma linha de tendência:

Figura 30: Linha de Tendência para um Gráfico de Dispersão



Cuidado para não confundir a linha de tendência com um gráfico de linha. A linha de tendência pode ser feita para um histograma ou para um gráfico de dispersão. Já o gráfico de linha é um gráfico em si que pode ser feito quando a variável dependente só pode assumir um único valor em função da variável independente.

Fiquemos, portanto, com as diferenças de aplicação entre os gráficos de linha e de dispersão.

ATENÇÃO

O gráfico de linha é usado quando a variável estudada só pode assumir um único valor em função da variável independente.

O gráfico de dispersão é usado quando a variável estudada pode assumir diversos valores em função da variável independente.

DIRETO DO CONCURSO

017. (CESPE/BANCO DA AMAZÔNIA/2012) A tabela acima mostra a distribuição de frequências do número de meses em atraso nos pagamentos das prestações dos financiamentos de crédito em um grupo de 80 clientes de certa empresa. Considerando que esses clientes formam uma amostra aleatória simples e que atraso é considerado quando $X > 0$, julgue o item que se segue, com base nessas informações.

clientes em atraso (N)	45	20	10	3	2
meses em atraso (X)	0	1	2	3	4

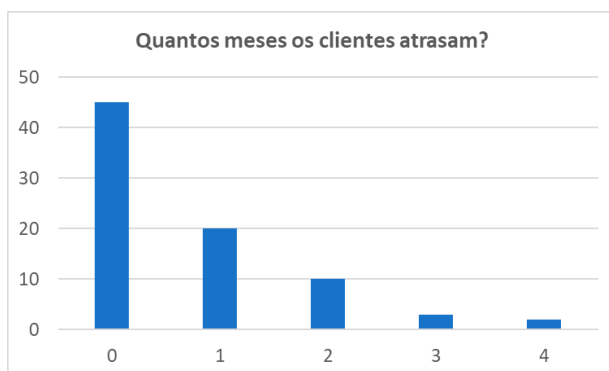
O diagrama de dispersão permite representar corretamente a distribuição de frequências da variável X.



Observe que temos uma função em que o número de clientes em atraso assume um único valor para cada observação da quantidade de meses em atraso.

Não é possível que, para três meses em atraso, existam dois valores diferentes para a contagem de clientes em atraso. Ou ela é igual a três ou é igual a dez, não é possível assumir os dois valores ao mesmo tempo.

Por isso, o diagrama de dispersão não é adequado para esse estudo. Nesse caso, o gráfico mais adequado seria o gráfico de barras.



Errado.

5. SEPARATRIZES

Já estudamos a mediana, que divide uma amostra em duas partes de igual tamanho, sendo uma delas formada por elementos maiores que ela e a outra parte por elementos menores.

A mediana é uma medida de posição, classificada como **medida de tendência central**, tendo em vista que ela tem por objetivo dividir a amostra em duas partes de igual tamanho.

Porém, em muitos casos, não nos interessa simplesmente encontrar um ponto central na amostra. Em alguns casos, pode ser interessante o estudo de um certo grupo extremo.

Por exemplo, suponha que você queira trazer um carro de super luxo para o Brasil e gostaria de saber se haverá público. Para você, pode ser interessante conhecer o patrimônio médio dos 10% mais ricos do país.

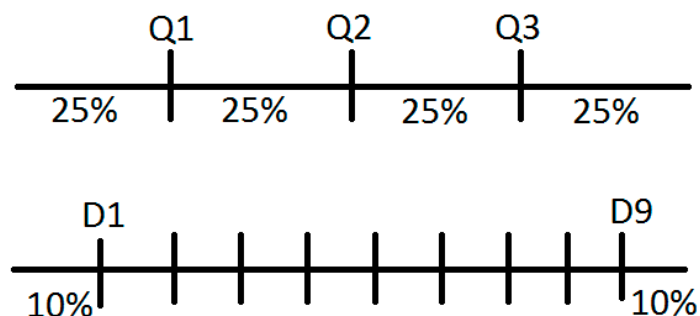
Outra situação é se você for um produtor de arroz e feijão e você precisa saber a renda dos 30% mais pobres do país para avaliar qual seria o custo ideal do seu produto.

E é justamente aí que entram as separatrizes. Com o auxílio delas, podemos determinar um valor que separa as 10% maiores observações da amostra ou as 30% menores.

Para calcular as separatrizes, primeiramente, devemos organizar as observações da variável estatística estudada em ordem crescente.

As separatrizes mais comuns são os quartis e os decis. Os quartis são três marcos que dividem a amostra em quatro partes de igual tamanho. Já os decis são nove marcos.

Figura 31: Quartis e Decis



A respeito dos quartis de uma amostra, pode-se dizer que:

- **Q1** é o número que deixa 25% das observações abaixo e 75% acima;
- **Q2** é a mediana da amostra e deixa 50% das observações abaixo e 50% das observações acima;
- **Q3** é o número que deixa 75% das observações abaixo e 25% acima.

Analogamente, pode-se dizer para os decis que:

- **D1** é o número que deixa 10% das observações abaixo e 90% acima.

D2 é o número que deixa 20% das observações abaixo e 80% acima.

D3 é o número que deixa 30% das observações abaixo e 70% acima.

E, assim por diante, até chegarmos ao **D9**, que é o número que deixa 90% das observações abaixo e 10% acima.

Os quartis e os decis fazem parte de um grupo de medidas de posição, conhecidas genericamente como **quantis**. Um quantil é qualquer tipo de separatriz. Também existem outros tipos de quantis. Por exemplo, os percentis: podemos separar o 1% de maiores valores observados. Existem os percentis $P_1, P_2, \dots, P_{99}$. E, analogamente, é possível também criar outros quantis. Observe que quantil é qualquer divisão proporcional que podemos fazer. Os mais comuns são os quartis, decis e percentis. Logo, não confunda quantil com quartil, que é um tipo de quantil.

Na obtenção das separatrizes, assim como é feito para a mediana, é necessário realizar a correção $N + 1$. Dessa maneira, se uma amostra é composta por N elementos, os quartis e decis serão:

Tabela 21: Posição associada aos Quartis e Decis

Quartil	Elemento	Frequência Acumulada
Q1	$(N+1)/4$	25%
Q2	$(N+1)/2$	50%
Q3	$3(N+1)/4$	75%
D1	$(N+1)/10$	10%
D9	$9(N+1)/10$	90%

Como exemplo, temos a seguir a população das 19 cidades mais populosas do Brasil:

Posição	Município	População	Posição	Município	População
1	São Paulo	12.038.175	11	Goiânia	1.448.639
2	Rio de Janeiro	6.498.837	12	Belém	1.446.042
3	Brasília	2.977.216	13	Guarulhos	1.337.087
4	Salvador	2.938.092	14	Campinas	1.173.370
5	Fortaleza	2.609.716	15	São Luís	1.082.935
6	Belo Horizonte	2.513.451	16	São Gonçalo	1.044.058
7	Manaus	2.094.391	17	Maceió	1.021.709
8	Curitiba	1.893.997	18	Duque de Caxias	886.917
9	Recife	1.625.583	19	Natal	877.662
10	Porto Alegre	1.481.019			

No caso do cálculo de quartis, apenas peço atenção para o fato de que as cidades foram organizadas em ordem decrescente de população, como é comum em tabelas de dados desse gênero. Porém, a medida de quartis pede a organização por ordem crescente.

Sendo assim, a primeira cidade (x_1) seria Natal, a quinta cidade seria São Luís – é só contar de baixo para cima – e a décima quinta cidade seria Fortaleza.

Como a amostra é formada por 19 cidades, podemos calcular os quartis da seguinte maneira:

$$\frac{1 + 19}{4} = \frac{20}{4} = 5$$

$$Q_1 = x_5 = 1.082.935 \text{ (São Luís)}$$

$$Q_3 = x_{15} = 2.609.716 \text{ (Fortaleza)}$$

Nesse caso, o segundo quartil (mediana) seria a população de Porto Alegre.

Por fim, gostaria de chamar a atenção para o caso de que sejam fornecidos dados categorizados. Nesse caso, os quartis devem ser calculados por **interpolação linear**.

5.1. CÁLCULO DE QUANTIS EM DADOS CATEGORIZADOS

Para mostrar como devemos proceder no cálculo dos quantis no caso de dados categorizados, vamos retornar a uma distribuição já estudada.

É importante destacar que, assim como no caso do cálculo da mediana, a frequência acumulada também auxilia bastante.

Tabela 22: Exemplo de Distribuição de Frequências em Classes: Cálculo da Frequência Acumulada

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 – 8	40%	40%
8 – 12	25%	65%
12 – 16	20%	85%
16 – 20	10%	95%
20 – 24	5%	100%
Total	100,00%	100,00%

Suponha que precisamos calcular o primeiro quartil (Q_1), o terceiro quartil (Q_3), o primeiro decil (D_1) e o nono decil (D_9).

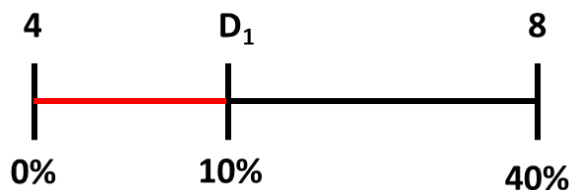
O primeiro quartil será aquele em que a frequência acumulada atingirá $1/4$ ou 25%. Já o terceiro quartil é registrado quando a frequência acumulada atinge 75%.

Analogamente, para o primeiro decil, que divide a amostra em 10 partes, temos que o primeiro decil acontece quando a frequência acumulada atinge 10% e o nono decil quando atinge 90%.

Vamos localizar as classes em que os quantis desejados estão localizados. Tanto o primeiro quartil (Q_1) como o primeiro decil (D_1) estão localizados na primeira classe de 4 a 8 salários mínimos.

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 – 8	40%	40%
8 – 12	25%	65%
12 – 16	20%	85%
16 – 20	10%	95%
20 – 24	5%	100%
Total	100,00%	100,00%

No caso específico da primeira classe, a interpolação linear deve ser feita considerando que o menor valor possível (4) corresponde à frequência acumulada de 0% e o maior valor possível (8) corresponde à frequência relativa da classe (40%).

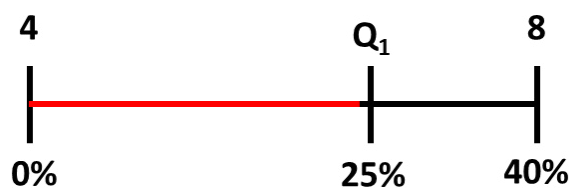


Pela interpolação linear, o pedaço de segmento à esquerda é proporcional ao segmento inteiro. Isso significa que devemos pegar a razão entre as diferenças da porção superior ($D_1 - 4$) e inferior ($10\% - 0\%$) da porção esquerda do segmento.

$$\begin{aligned}\frac{D_1 - 4}{10\% - 0\%} &= \frac{8 - 4}{40\% - 0\%} \\ \frac{D_1 - 4}{10\%} &= \frac{4}{40\%} \\ \therefore D_1 - 4 &= \frac{10\%}{40\%} \cdot 4 = \frac{1}{4} \cdot 4 = 1 \\ \therefore D_1 &= 4 + 1 = 5\end{aligned}$$

Devemos igualar essa razão à mesma razão tomada do segmento inteiro. Assim, devemos tomar a diferença na porção superior ($8 - 4$) e dividir pela diferença parte inferior ($30\% - 0\%$) referentes ao segmento inteiro.

E, agora, fazemos o mesmo procedimento para o primeiro quartil (Q_1).

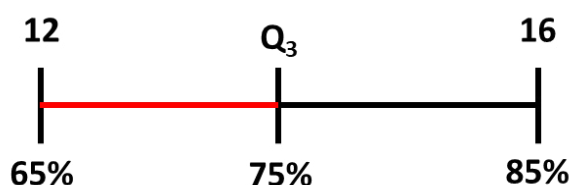


$$\begin{aligned}\frac{Q_1 - 4}{25\% - 0\%} &= \frac{8 - 4}{40\% - 0\%} \\ \frac{Q_1 - 4}{25\%} &= \frac{4}{40\%} \\ \therefore Q_1 - 4 &= \frac{25\%}{40\%} \cdot 4 = \frac{100}{40} = 2,5 \\ \therefore Q_1 &= 4 + 2,5 = 6,5\end{aligned}$$

Agora, vamos calcular o terceiro quartil (Q_3). Para isso, devemos localizar a classe em que a frequência acumulada atinge 75%.

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 – 8	40%	30%
8 – 12	25%	65%
12 – 16	20%	85%
16 – 20	10%	95%
20 – 24	5%	100%
Total	100,00%	100,00%

Para a interpolação linear, devemos considerar o extremo da classe anterior (12) correspondendo à marcação de 65% e o extremo da classe do quantil (16) correspondente à marcação de 85%.



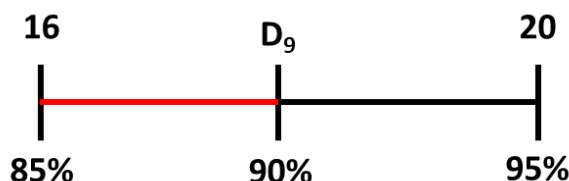
Fazendo a interpolação linear, temos:

$$\begin{aligned}\frac{Q_3 - 12}{75\% - 65\%} &= \frac{16 - 12}{85\% - 65\%} \\ \frac{Q_3 - 12}{10\%} &= \frac{4}{20\%} \\ \therefore Q_3 - 12 &= \frac{10\%}{20\%} \cdot 4 = 2 \\ \therefore Q_3 &= 12 + 2 = 14\end{aligned}$$

Por fim, vamos calcular o nono decil.

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 – 8	40%	40%
8 – 12	25%	65%
12 – 16	20%	85%
16 – 20	10%	95%
20 – 24	5%	100%
Total	100,00%	100,00%

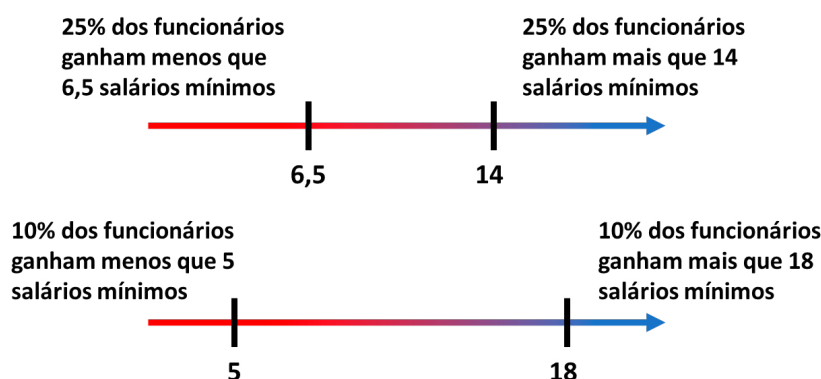
Vamos montar o esquema da interpolação linear:



E, agora, façamos as contas.

$$\begin{aligned}\frac{D_9 - 16}{90\% - 85\%} &= \frac{20 - 16}{95\% - 85\%} \\ \frac{D_9 - 16}{5\%} &= \frac{4}{10\%} \\ \therefore D_9 - 16 &= \frac{5\%}{10\%} \cdot 4 = 2 \\ \therefore D_9 &= 16 + 2 = 18\end{aligned}$$

Agora, vamos esquematizar para entendermos melhor o significado dos quantis calculados.



No caso de dados categorizados, lembre-se de que, mesmo quando eles são fornecidos na forma de frequência absoluta, não devemos efetuar a correção $(N+1)/4$ para o cálculo dos quartis nem $(N+1)/2$ para o cálculo da mediana.

Quando os dados são fornecidos em classes, tomamos o quartil como $N/4$ e a mediana como $N/2$.

seleção para uma vaga de emprego. A última fase do processo de seleção consiste em uma entrevista, que será realizada apenas com os candidatos que obtiveram desempenho acima da nota de corte, calculada com base no valor do sexto decil, cujo valor é 51,8:

Distribuição das Notas dos Candidatos				
NOTAS	F_i	F_{ac}	F_i	F_{ac}
0 – 20	5	5	0,07	0,07
20 – 40	24	29	0,32	0,39
40 – 60	27	56	0,36	0,75
60 – 80	14	70	0,18	0,93
80 – 100	5	75	0,07	1,00
Total	75		1,00	
Legenda: F_i — frequência absoluta simples F_{ac} — frequência absoluta acumulada F_i — frequência relativa simples F_{ac} — frequência relativa acumulada				

“De acordo com os dados apresentados, é correto afirmar que 60% dos candidatados obtiveram notas _____ de 51,8; portanto, a empresa realizará _____ 40 entrevistas.” Assinale a alternativa que completa correta e sequencialmente a afirmativa anterior.

- a) acima / mais.
- b) abaixo / mais.
- c) acima / menos.
- d) abaixo / menos.
- e) acima / exatamente.



Por definição, o sexto decil (60%) é o limiar em que 60% dos candidatos se situam abaixo dessa nota. Desse modo, podemos concluir que 60% dos candidatos obtiveram notas abaixo de 51,8.

Os demais, portanto, menos de 40, serão chamados para a entrevista.

Letra d.

019. (CESGRANRIO/TRANSPETRO/2012/ADMINISTRADOR JÚNIOR) A tabela apresenta uma distribuição hipotética. Não há observações coincidentes com os limites das classes. A melhor estimativa para o terceiro quartil da distribuição é, aproximadamente, de:

Classes	Frequência absoluta
de 0 a 10	4
de 10 a 20	10
de 20 a 30	50
de 30 a 40	100
Total	164

- a) 34,75.
- b) 34,9.
- c) 35.
- d) 35,75.
- e) 35,9.



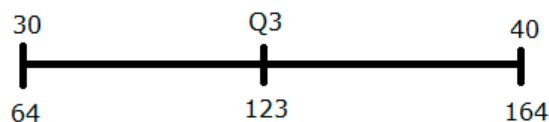
Uma questão bastante ilustrativa sobre quartis em dados categorizados. Como a amostra é formada por 164 elementos, temos que o terceiro quartil será associado à posição:

$$\frac{3}{4} \cdot 164 = 123 \therefore Q_3 = x_{123}$$

O primeiro passo para calcular esse quartil é identificar em que categoria ele se localiza.

Classes	Frequência Absoluta	Elementos
0 a 10	4	1 a 4
10 a 20	10	5 a 14
20 a 30	50	15 a 64
30 a 40	100	65 a 164

O terceiro quartil está localizado na última categoria entre os dois elementos destacados em negrito. Sendo assim, devemos proceder à interpolação linear.



Usando a proporção entre a menor parte e todo, pode-se escrever:

$$\frac{Q_3 - 30}{123 - 64} = \frac{40 - 30}{164 - 64} \therefore \frac{Q_3 - 30}{59} = \frac{10}{100} = 0,1$$

$$\therefore Q_3 - 30 = 0,1 \cdot 59 = 5,9 \therefore Q_3 = 30 + 5,9 = 35,9$$

Letra e.

5.2. DIAGRAMA BOX PLOT

Os diagramas Box Plot são utilizados para representar várias observações de uma mesma variável em função quantitativa, normalmente contínua, em função de outra variável independente.

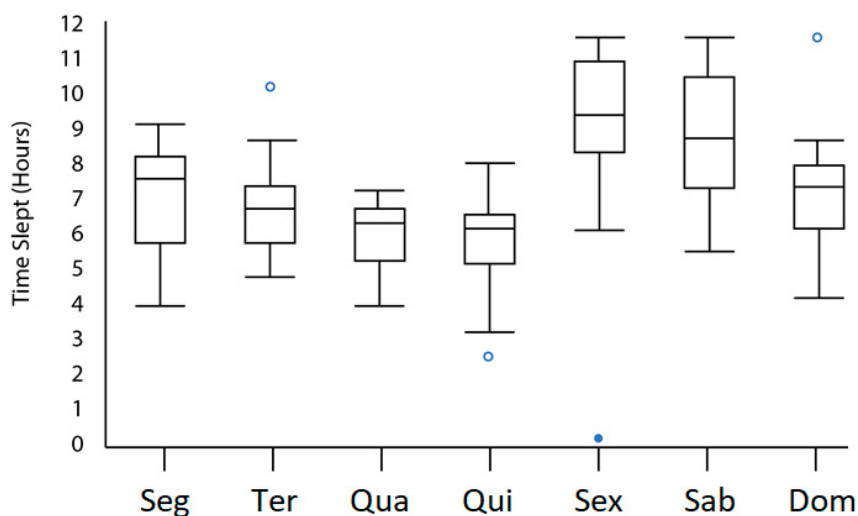
Esses diagramas são amplamente usados no meio científico, principalmente em pesquisas relacionadas à área da saúde. No mercado de investimentos, é utilizado o chamado *diagrama candle stick* (ou diagrama de velas), que é uma ligeira alteração do Box Plot.

Sendo assim, eu os considero um dos mais importantes no meio profissional e, por isso, as chances de serem cobrados em uma prova de concurso público são razoáveis.

Os diagramas Box Plot resumem uma grande quantidade de informação em suas linhas.

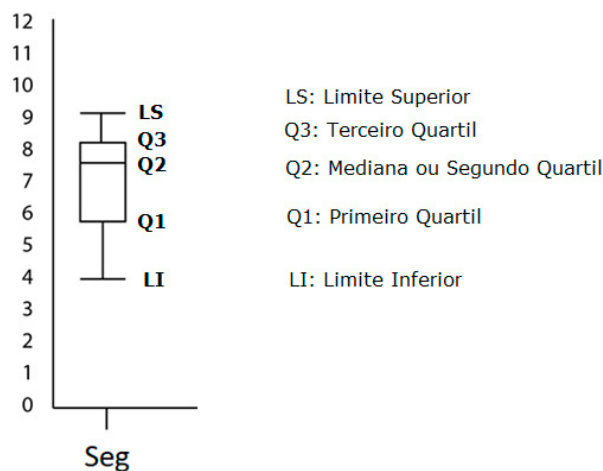
Por exemplo, suponha que monitoramos o número de horas que um grupo de pessoas dormiu durante uma semana. O resultado obtido pode ser resumido a partir do seguinte gráfico:

Figura 32: Diagrama Box Plot sobre o número de horas dormidas por um grupo de pessoas



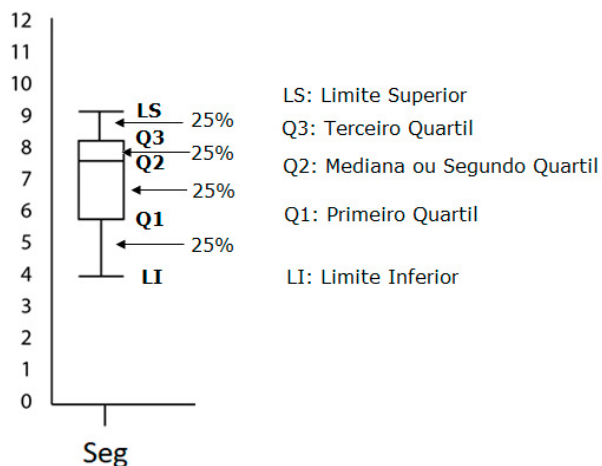
O diagrama Box Plot recebe esse nome porque é formado por uma série de caixas. As caixas são compostas por 5 elementos:

Figura 33: Composição da Caixa do Box Plot



Por ser composta pelos quartis, a caixa do Box Plot forma quatro regiões em que cada uma delas contém 25% das observações da amostra.

Figura 34: Box Plot evidenciando os quartis



Sendo assim, podemos afirmar, com segurança, que o interior da caixa do Box Plot contém 50% das observações da amostra. Isso acontece porque 50% das observações se situam entre o primeiro e o terceiro quartis.

ATENÇÃO

A faixa no interior da caixa do Box Plot representa a mediana. As questões vão tentar lhe enganar, afirmando que é a média. Porém, o Box Plot somente evidencia só os quartis da amostra, incluindo a mediana (Q_2), ele não mostra a média.

Agora, vamos falar sobre os limites. Os limites superior e inferior são os **limites para a detecção de outliers**. Eles são calculados a partir do chamado desvio interquartílico (D) ou intervalo interquartílico, que é dado pela diferença entre o terceiro e o primeiro quartis.

$$D = Q_3 - Q_1$$

O limite superior se situa 1,5 desvio interquartílico acima do terceiro quartil e o limite inferior se situa 1,5 desvio interquartílico abaixo do primeiro quartil. Assim, podemos escrever:

$$LS = Q_3 + 1,5 \cdot D = Q_3 + 1,5(Q_3 - Q_1)$$

$$LI = Q_1 - 1,5 \cdot D = Q_1 - 1,5(Q_3 - Q_1)$$

Se uma observação qualquer da amostra está fora dos limites superior e inferior, essa observação é chamada de **outlier**.

Os *outliers* são observações atípicas de uma variável aleatória. No diagrama Box Plot, eles são destacados. Na Figura 32, pode-se ver vários *outliers* tanto para cima como para baixo.

Por exemplo, na terça-feira, houve um *outlier* – uma pessoa que dormiu muito mais que todos os outros. Ela conseguiu dormir mais de dez horas, enquanto todas as demais pessoas observadas dormiram menos de nove horas – o limite superior é um pouco inferior a nove horas.

Por outro lado, na sexta-feira, houve uma pessoa que dormiu 0 horas. Essa também é um *outlier*, pois todas as demais dormiram entre 6 e 12 horas.

É importante citar que, quando não é encontrado nenhum *outlier*, os limites inferior e/ou superior devem ser modificados para **o valor máximo e mínimo da amostra**.

Por exemplo, na segunda-feira, não houve nenhum *outlier*. Portanto, podemos garantir que uma pessoa dormiu quatro horas e outra pessoa dormiu nove horas, pois esses são os limites inferior e superior.

5.2.1. AVALIAÇÃO DE OUTLIERS

Um dos grandes objetivos do Box Plot é evidenciar a existência ou não de *outliers*.

Em alguns casos, os *outliers* podem indicar a existência de alguma falha no procedimento de medida.

Por exemplo, imagine que você esteja avaliando um radar de trânsito e descobre que houve um carro que passou por ele a mais de 500km/h. Essa medida é um *outlier*.

Porém, é um *outlier* que não faz nenhum sentido, portanto, é um sinal de que houve uma falha no radar. Pode ser interessante pedir para um técnico monitorar o equipamento para descobrir se a falha foi pontual ou se outras medidas podem estar comprometidas.

Em outros casos, como na Figura 32, a existência de *outliers* é normal. Porém, o excesso deles é que deve chamar a atenção do pesquisador.

Por exemplo, imagine que houve um dia da semana em que 5% da população de uma cidade dormiu abaixo de três horas, aparecendo vários *outliers*.

Esse fato, com certeza, é estranho e o pesquisador pode se interessar por entender o que causou esse comportamento.

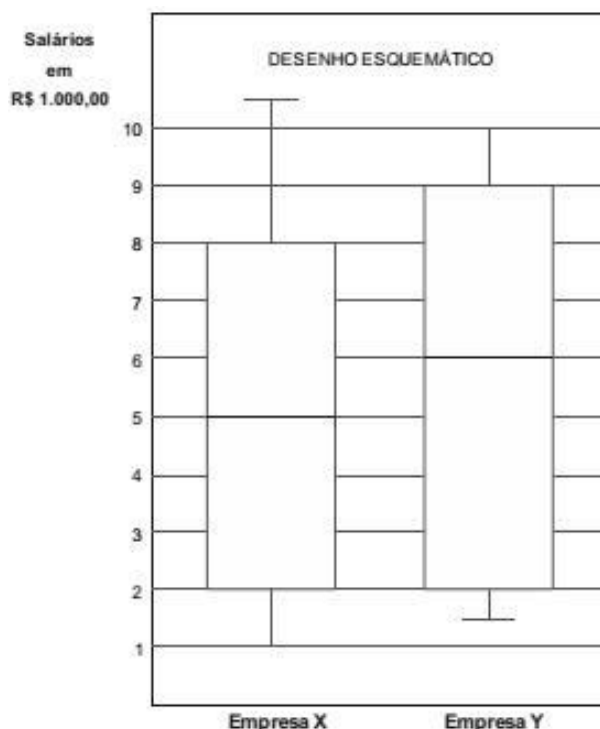
Em outros casos, a presença de *outliers* pode ser, por si só, de grande interesse. Por exemplo, numa pesquisa sobre a margem de lucro de empresas do mesmo setor, se você encontrar um *outlier* positivo, essa empresa mais eficiente pode ser um alvo de estudos.

O que faz que ela tenha um desempenho tão acima da média das outras? É possível que outras empresas do setor copiem as suas práticas para melhorar seus resultados?

Dessa forma, podemos concluir que a avaliação dos *outliers* depende muito do tipo de pesquisa que está sendo feita. Vale lembrar que a Estatística é uma ferramenta para outras áreas do conhecimento e não gera, por si só, conclusões sobre a matéria em análise.

DIRETO DO CONCURSO

020. (FCC/TRT/19ª REGIÃO/AL/2014/ANALISTA JUDICIÁRIO/ESTATÍSTICA) 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.



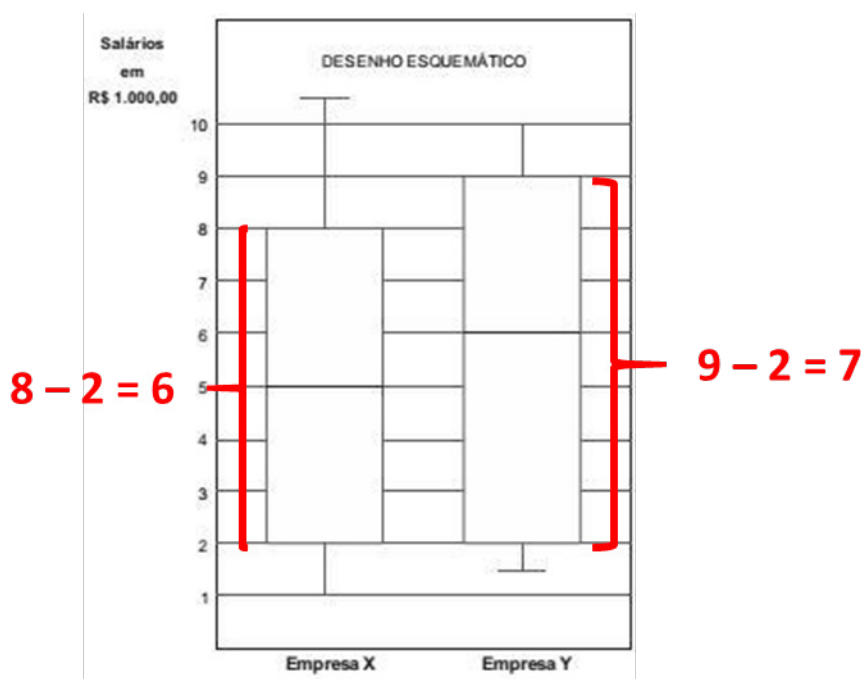
De acordo com o desenho esquemático apresentado, é correto afirmar que

- 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.



Vamos analisar as afirmações sobre o Diagrama Box Plot em apreço.

- a) Errada. Não há como afirmar isso olhando somente para o gráfico. O diagrama box plot não fornece nenhuma informação sobre o número de empregados da empresa X.
- b) Certa. A distância interquartil corresponde ao comprimento da caixa do box plot.



Notemos que, realmente, a empresa Y apresenta uma distância interquartilica maior.

- c) Errada. O menor valor de salário na empresa pode ser obtido olhando os whiskers inferiores. No caso da empresa X, é de R\$1.000,00 e, no caso da empresa Y, é superior a R\$1.000,00 e inferior a R\$2.000,00. De qualquer modo, a afirmação está incorreta.
- d) Errada. A mediana corresponde à barra dentro da caixa do box plot. Na empresa X, é igual a R\$5.000,00 e, na empresa Y, é igual a R\$6.000,00. Portanto, na verdade, a mediana de Y excede a de X em 20%.
- e) Errada. O maior salário da empresa X excede os R\$10.000, pois é contado pelo *whisker* superior. Na empresa Y, o maior salário é de R\$10.000,00.

Letra b.

RESUMO

- **Tabelas Estatísticas**
 - As tabelas facilitam a **inspeção** dos dados. Os gráficos facilitam a **visualização**.
- **Séries Estatísticas:** estudam o efeito do tempo (séries temporais), do espaço (série geográfica) ou do tipo de elemento (séries especificativas) sobre o comportamento da variável.
- **Gráficos de Setores e de Barras**
 - São utilizados para variáveis qualitativas ou discretas, desde que não estejam distribuídas em classes.
 - O gráfico de setores é particularmente útil para representar a frequência relativa.
- **Histogramas**
 - A variável é organizada em classes.
 - **Cuidado:** o gráfico de barras é utilizado quando a variável não é organizada em classes; o histograma é utilizado quando a variável é organizada em classes.
 - **Número de Classes:** é definido pela Regra do Quadrado

$$K = \sqrt{N}$$

- **Amplitude de Classe:** é dado pela razão entre a amplitude total da amostra e o número de classes

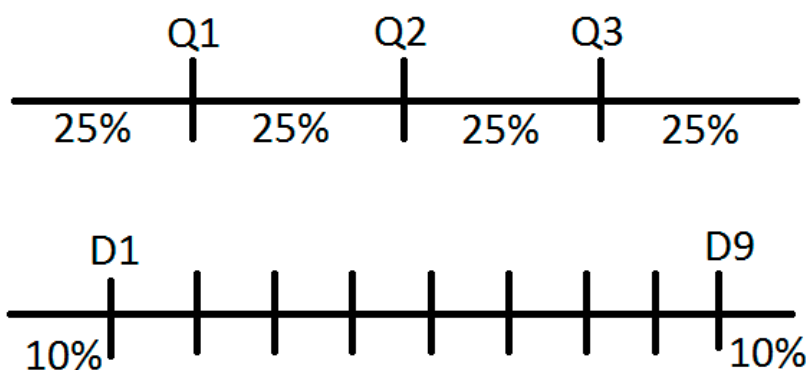
$$h = \frac{AT}{K}$$

- O número de classes e a amplitude podem ser adaptados para criar um histograma mais agradável aos olhos humanos. Geralmente, se prefere que as classes tenham amplitudes iguais a um número inteiro.
- **Polígono de Frequências:** une os pontos médios de cada barra do histograma.
- **Ogiva:** une os cantos superiores direitos do histograma de frequências acumuladas.
- **Histogramas com Densidades Diferentes**
 - A altura das barras representa a densidade de frequência, enquanto a largura representa a amplitude da classe.
 - A frequência relativa da classe é igual à área das barras

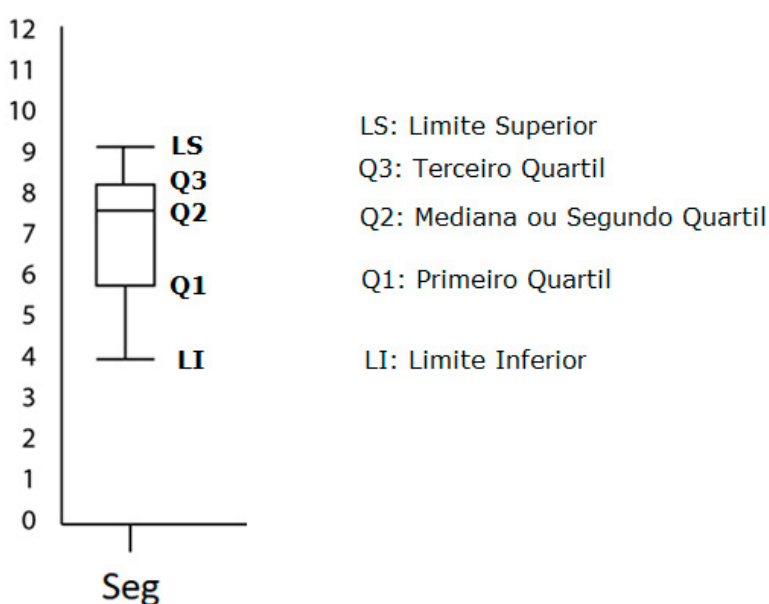
$$f_i = d_i A_i$$

- **Outros Gráficos e Diagramas**
 - **Ramos e Folhas:** é um caso especial de histograma que conserva os dados originais.

- **Gráfico de Linha:** é utilizado quando uma variável só pode assumir um valor em função da outra. Por exemplo, renda per capita do Brasil e tempo. Em um determinado ano, a renda per capita do Brasil só pode ter um único valor.
- **Gráfico de Dispersão:** é utilizado quando uma variável pode assumir diversos valores conjuntamente com a outra. Por exemplo, idade e salário. Duas pessoas com 25 anos podem ter salários diferentes. Portanto, utiliza-se o gráfico de dispersão nesse caso.
- **Separatrizes:**
 - Dividem a amostra em duas partes. Por exemplo, os 10% maiores valores ou os 25% menores valores.
 - A mediana é uma separatriz correspondente ao quinto decil (D_5) e ao segundo quartil (Q_2).



- **Box Plot:**
 - O diagrama box plot é construído com base nos quartis (Q_1 , Q_2 e Q_3).



- Os limites inferiores e superiores distam 1,5 intervalos interquartílicos do primeiro quartil (Q_1) e do terceiro quartil (Q_3), respectivamente.

$$LS = Q_3 + 1,5.D = Q_3 + 1,5(Q_3 - Q_1)$$

$$LI = Q_1 - 1,5.D = Q_1 - 1,5(Q_3 - Q_1)$$

QUESTÕES COMENTADAS EM AULA

001. (CESPE/ANTAQ/2009/TÉCNICO EM REGULAÇÃO) Considerando a tabela acima, que apresenta a movimentação anual de cargas no porto de Santos de 2003 a 2007, em milhões de toneladas/ano e associa as quantidades de carga movimentadas para exportação e importação às variáveis X e Y, respectivamente, julgue o item subsequente.

	variável	2003	2004	2005	2006	2007
exportação	X	40	46	50	52	54
importação	Y	20	21	22	24	27
total	X + Y	60	67	72	76	81

As séries estatísticas apresentadas na tabela formam três séries temporais.

002. (CESPE ABIN/2018/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4) Com base nos dados da tabela anterior, extraídos do Relatório das Notas Estatísticas do Censo Escolar de 2017, do INEP, julgue o item a seguir.

evolução da quantidade de docentes por etapa de ensino Brasil 2013 – 2017				
ano	educação infantil	anos iniciais do ensino fundamental	anos finais do ensino fundamental	ensino médio
2013	478.811	750.366	802.902	507.617
2014	502.445	757.950	797.577	522.426
2015	518.308	758.840	786.140	522.826
2016	540.567	763.927	778.561	519.883
2017	557.541	761.737	764.731	509.814
soma total das quantidades de docentes no período	2.597.672	3.792.820	3.929.911	2.582.566

O quantitativo de professores dos anos finais do ensino fundamental sofreu queda superior a 1,3% no ano de 2015 em relação ao ano anterior.

003. (CESPE/ABIN/2018/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4) A menor diferença entre as quantidades de professores dos anos iniciais do ensino fundamental e do ensino médio ocorreu em 2014.

004. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte mostra as quantidades de patrimônios históricos cadastrados nos estados A e B.

patrimônios	estado A	estado B
estátuas	40	10
museus	60	20
templos	50	20
total	150	50

A partir dessa tabela, julgue o seguinte item.

As estátuas cadastradas nos estados A e B correspondem a mais de 20% dos patrimônios históricos cadastrados nesses estados.

005. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) O gráfico de barras é adequado para a análise de variáveis qualitativas ordinais ou quantitativas discretas, pois permite investigar a presença de tendência nos dados.

006. (CESPE/CEBRASPE/FUB/ESTATÍSTICO/2022/Q2444768) A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

disciplinas	2	3	4	5	6	7	8
estudantes	10	15	40	35	28	10	4

Considerando essas informações, julgue o item seguinte.

O gráfico do tipo pizza é o mais apropriado para representar os dados apresentados na tabela, visto que a variável analisada é qualitativa ordinal.

007. (VUNESP/UNICAMP/2022/ENGENHEIRO CIVIL/Q2491143) Em uma pesquisa realizada, 450 pessoas responderam “sim” à pergunta feita, 320 responderam “não”, e 40 pessoas não responderam à pergunta. Na construção de um gráfico de setor (também conhecido como gráfico de pizza) para representar os dados obtidos nessa pergunta, o setor circular correspondente às respostas “sim” deverá ter ângulo medindo

- a) 170.
- b) 180.
- c) 190.
- d) 200.
- e) 210.

008. (CESPE/FUB/2022/ESTATÍSTICO) A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

disciplinas	2	3	4	5	6	7	8
estudantes	10	15	40	35	28	10	4

Considerando essas informações, julgue o item seguinte.

O gráfico do tipo pizza é o mais apropriado para representar os dados apresentados na tabela, visto que a variável analisada é qualitativa ordinal.

009. (CESPE/AL/CE/2011/ANALISTA LEGISLATIVO) Um levantamento foi realizado para se avaliar, por município, a quantidade X de obras que estão sob suspeita de irregularidade. Com base em uma amostra de municípios, foi obtida a distribuição de frequências mostrada na tabela acima. Com base nessas informações, julgue o item.

X	0	1	2	3	4	5
frequência absoluta	80	47	30	20	6	1

O diagrama de setores (ou pizza) é um gráfico apropriado para se representar a distribuição de probabilidade da variável X.

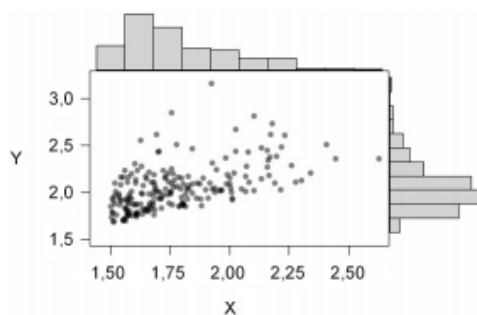
010. (CESPE/TRE/ES/2011/ANALISTA) Com base na tabela acima, referente às eleições de 2010, que apresenta a quantidade de candidatos para os cargos de presidente da República, governador de estado, senador, deputado federal e deputado estadual/distrital, bem como a quantidade de candidatos considerados aptos pela justiça eleitoral e o total de eleitos para cada cargo pretendido, julgue o item a seguir.

cargo	candidatos	candidatos aptos	eleitos
presidente da República	9	9	1
governador de estado	170	156	27
senador	272	234	54
deputado federal	6.021	5.058	513
deputado estadual/distrital	15.268	13.076	1.059
total	21.640	18.533	1.658

Internet: <www.tse.gov.br> (com adaptações)

Considerando-se a representação das quantidades de eleitos para cada cargo em um gráfico de pizza, a fatia desse gráfico correspondente ao cargo de deputado federal terá ângulo superior a 120° .

011. (CESPE/MPU/2013/ANALISTA/ESTATÍSTICA) A figura acima mostra a dispersão dos valores previstos (X, em R\$ milhões) e dos valores efetivamente gastos (Y, em R\$ milhões) em 200 obras de pavimentação em determinado estado, e os respectivos histogramas das distribuições dos valores X e Y. Com base nessas informações, julgue o item a seguir.



A amplitude total da distribuição de X foi igual ou superior a R\$ 2 milhões.

012. (IFTO/2023) Dados os valores de uma população: $X_1 = 4$; $X_2 = 6$; $X_3 = 8$ e $X_4 = 10$. A amplitude total dessa população será:

- a) 3.
- b) 8.
- c) 4.
- d) 2.
- e) 6.

013. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela a seguir mostra as quantidades de bibliotecas públicas presentes em 20 microrregiões brasileiras.

90	66	78	82
77	60	64	90
87	85	67	91
82	70	81	80
69	78	90	67

A partir desses dados, pretende-se construir um gráfico de distribuição de frequências com quatro classes de igual amplitude. Os valores mínimo e máximo de cada classe devem ser números inteiros.

Considerando essas informações, julgue o item subsequente, relativo ao gráfico de distribuição a ser apresentado.

A amplitude de cada classe deverá ser superior a 6.

014. (FCC/TRT/14ª REGIÃO (RO E AC)/2012/ANALISTA JUDICIÁRIO/ESTATÍSTICA) De um histograma e uma tabela de frequências absolutas, elaborados para analisar a distribuição dos salários dos empregados em uma empresa, obtém-se a informação que 24 empregados ganham salários com valores pertencentes ao intervalo (2.000; 4.000], em reais, que apresenta uma densidade de frequência de $0,75 \times 10^{-4}(\text{R\$})^{-1}$.

Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela amplitude deste intervalo. Em um intervalo do histograma que está sendo analisado, com uma amplitude de R\$ 3.000,00 e uma densidade de frequência de $1 \times 10^{-4}(\text{R\$})^{-1}$, tem-se que o correspondente número de empregados é igual a:

- a) 40.
- b) 36.
- c) 30.
- d) 48.
- e) 42.

015. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4) A tabela mostrada apresenta a quantidade de detentos no sistema penitenciário brasileiro por região em 2013. Nesse ano, o déficit relativo de vagas — que se define pela razão entre o déficit de vagas no sistema penitenciário e a quantidade de detentos no sistema penitenciário — registrado em todo o Brasil foi superior a 38,7%, e, na média nacional, havia 277,5 detentos por 100 mil habitantes.

região	quantidade de detentos no sistema penitenciário brasileiro (mil pessoas)	déficit de vagas no sistema penitenciário (mil vagas)	população brasileira (milhões de habitantes)
Norte	37	13	17
Centro-oeste	51	24	15
Nordeste	94	42	55
Sudeste	306	120	85
Sul	67	16	28
total	555	215	200

Ministério da Justiça — Departamento Penitenciário Nacional — Sistema Integrado de Informações Penitenciárias – InfoPen, Relatório Estatístico Sintético do Sistema Prisional Brasileiro, dez./2013 Internet: <-
www.justica.gov.br> (com adaptações)

Com base nessas informações e na tabela apresentada, julgue o item a seguir.

Na análise exploratória, o histograma é um gráfico adequado para descrever a distribuição da quantidade de detentos por região em 2013.

016. (FCC/TJ/AP/2009) O diagrama de ramo e folhas a seguir corresponde às idades dos 40 funcionários de um setor de um órgão público em uma determinada data.

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

A soma da mediana e da moda destas idades é igual a

- a) 67,0.
- b) 66,5.
- c) 66,0.
- d) 65,5.
- e) 65,0.

017. (CESPE/BANCO DA AMAZÔNIA/2012) A tabela acima mostra a distribuição de frequências do número de meses em atraso nos pagamentos das prestações dos financiamentos de crédito em um grupo de 80 clientes de certa empresa. Considerando que esses clientes formam uma amostra aleatória simples e que atraso é considerado quando $X > 0$, julgue o item que se segue, com base nessas informações.

clientes em atraso (N)	45	20	10	3	2
meses em atraso (X)	0	1	2	3	4

O diagrama de dispersão permite representar corretamente a distribuição de frequências da variável X.

018. (INSTITUTO CONSULPLAN/MPE/BA/2023/ANALISTA DE SISTEMAS/Q2852433) A tabela representa a distribuição das notas de desempenho dos candidatos em um processo de seleção para uma vaga de emprego. A última fase do processo de seleção consiste em uma entrevista, que será realizada apenas com os candidatos que obtiveram desempenho acima da nota de corte, calculada com base no valor do sexto decil, cujo valor é 51,8:

Distribuição das Notas dos Candidatos				
NOTAS	F_i	F_{ac}	F_i	F_{ac}
0 – 20	5	5	0,07	0,07
20 – 40	24	29	0,32	0,39
40 – 60	27	56	0,36	0,75
60 – 80	14	70	0,18	0,93
80 – 100	5	75	0,07	1,00
Total	75		1,00	
Legenda: F_i — frequência absoluta simples F_{ac} — frequência absoluta acumulada F_i — frequência relativa simples F_{ac} — frequência relativa acumulada				

“De acordo com os dados apresentados, é correto afirmar que 60% dos candidatados obtiveram notas _____ de 51,8; portanto, a empresa realizará _____ 40 entrevistas.” Assinale a alternativa que completa correta e sequencialmente a afirmativa anterior.

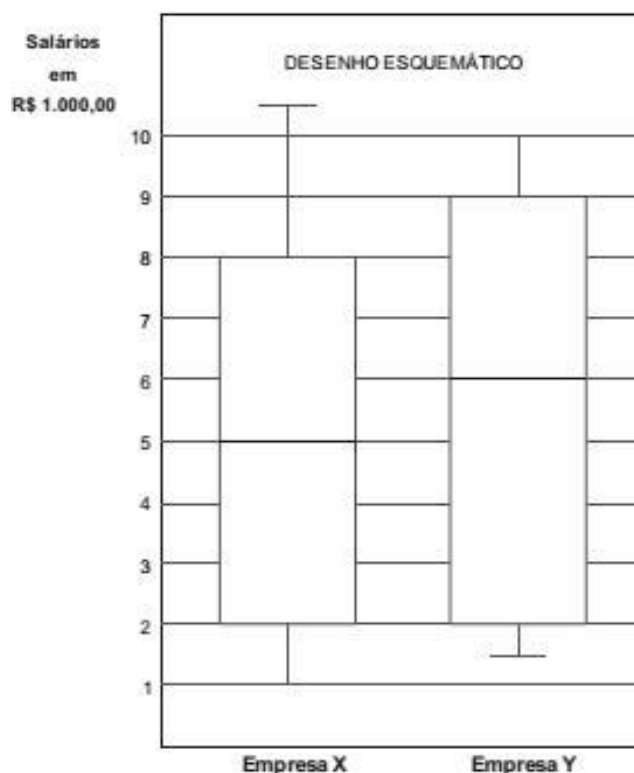
- a) acima / mais.
- b) abaixo / mais.
- c) acima / menos.
- d) abaixo / menos.
- e) acima / exatamente.

019. (CESGRANRIO/TRANSPETRO/2012/ADMINISTRADOR JÚNIOR) A tabela apresenta uma distribuição hipotética. Não há observações coincidentes com os limites das classes. A melhor estimativa para o terceiro quartil da distribuição é, aproximadamente, de:

Classes	Frequência absoluta
de 0 a 10	4
de 10 a 20	10
de 20 a 30	50
de 30 a 40	100
Total	164

- a) 34,75.
- b) 34,9.
- c) 35.
- d) 35,75.
- e) 35,9.

020. (FCC/TRT/19ª REGIÃO/AL/2014/ANALISTA JUDICIÁRIO/ESTATÍSTICA) 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.

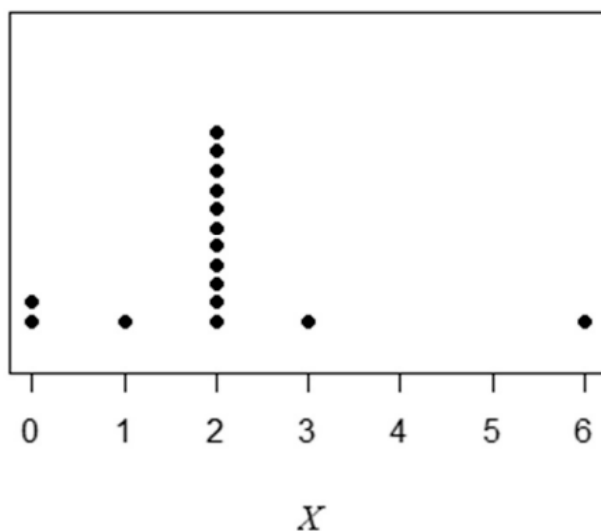


De acordo com o desenho esquemático apresentado, é correto afirmar que

- 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.

QUESTÕES DE CONCURSO

021. (CESPE/CEBRASPE/CGE/RJ/AUDITOR DO ESTADO/2024/Q3142790)



Considerando a figura precedente, que mostra a distribuição de frequências de uma variável X em uma amostra aleatória simples de tamanho igual a 16, julgue o próximo item.
A média, a mediana e a moda do conjunto de dados em questão são iguais a 2.

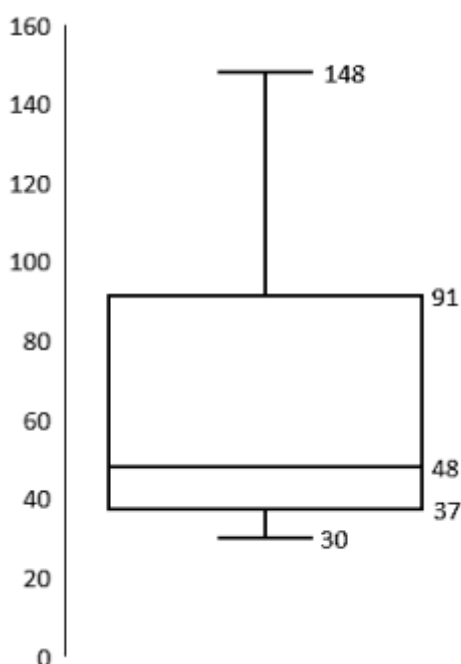
022. (CESPE/CEBRASPE/ANP/ATIVIDADES DE REGULAÇÃO DE NOVAS ATRIBUIÇÕES/2022/Q2726118)

ocorrência	comportamento	registrado	educação	idade
1	agressivo	sim	fundamental	15-24
3	calmo	sim	fundamental	25-39
2	agressivo	sim	fundamental	40-
2	calmo	sim	médio	15-24
8	agressivo	sim	médio	25-39
1	calmo	sim	fundamental	15-24
3	agressivo	sim	fundamental	25-39
4	calmo	sim	fundamental	40-
2	calmo	sim	médio	15-24
2	agressivo	sim	médio	25-39
5	agressivo	sim	médio	25-39
7	calmo	sim	infantil	5-14
1	agressivo	não	infantil	5-14
3	agressivo	sim	infantil	5-14

Considerando os dados da tabela anterior, representados na forma de um dataframe hipotético em relação a determinado conjunto de registros escolares a respeito da violência e do comportamento de discentes, julgue o item a seguir.

O número de ocorrências com discentes de comportamento agressivo e com educação de nível médio é, no total, inferior a 16.

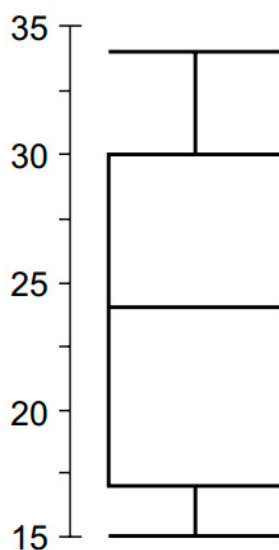
023. (FUMARC/TRT 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2022/Q2631573) O diagrama a seguir é chamado Box Plot (Diagrama de Caixa). Ele está representando uma amostra de 130 valores.



Baseado nas informações fornecidas no diagrama, é CORRETO afirmar:

- a) A Amplitude Interquartílica é de 128.
- b) Caso existisse na amostra um valor igual a 10, ele seria considerado um Outlier.
- c) Caso existisse na amostra um valor superior a 172, ele seria considerado um Outlier.
- d) No intervalo de 30 a 91, existem 50% do total de valores da amostra.
- e) O valor do 1º Quartil é 30, o do 2º Quartil é 37, o do 3º Quartil é 48 e o do 4º Quartil é 91.

024. (VUNESP/UNICAMP/PROFISSIONAL/ÁREA: DESENHISTA DE PÁGINAS DA INTERNET/2023/Q2924437) A seguinte figura ilustra um conjunto de medições de temperatura de uma determinada localidade conforme a técnica box plot, apresentando uma escala em graus Celsius. Não há ocorrências de outliers.



É possível afirmar que

- a) a temperatura máxima é igual a 35 °C.
- b) pelo menos 75% das medições são iguais ou inferiores a 30 °C.
- c) mais que 50% das medições são maiores que 25 °C.
- d) a temperatura máxima é igual a 30 °C.
- e) até 75% das medições são iguais ou superiores a 15 °C.

025. (IBFC/TRF 5/TÉCNICO JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: TECNOLOGIA DA INFORMAÇÃO/2024/Q3170134) Com base nos dados apresentados na Tabela 1 – Distribuição das idades (em anos) dos sócios do Clube Juvenil de Recreação e Bem-Estar, assinale a alternativa incorreta.

Tabela 1 – Distribuição das idades (em anos) dos sócios do Clube Juvenil de Recreação e Bem-Estar.

Idade (em anos)	Número de Sócios
10 – 12	92
12 – 14	107
14 – 16	231
16 – 18	189
Total	619

- a) Um jovem com 12 anos completos é alocado na primeira classe, 10 |– 12.
- b) A mediana assume um valor dentro da terceira classe, 14 |– 16
- c) Por ser a classe de maior frequência, a terceira classe, 14 |– 16, contém o valor da moda.
- d) A idade média pode ser calculada usando o ponto médio de cada classe, uma vez que não temos a informação dos valores observados.

026. (FGV/TJDFT/2022/ANALISTA JUDICIÁRIO/ESTATÍSTICA) O resultado de uma pesquisa sobre a produtividade dos magistrados em uma determinada região foi publicado em uma revista científica e está sintetizado na tabela a seguir.

Classes de produtividade	Frequência relativa acumulada
0 --- 2	0,25
2 --- 4	0,65
4 --- 6	0,85
6 --- 9	1

Não existem observações coincidentes com os extremos das classes.

É sabido que, quanto maior a classe de produtividade, maior é a produtividade do magistrado. Um estatístico precisa estimar a produtividade a partir da qual se encontram os 10% mais eficientes, isto é, o 9º decil dessa distribuição.

A melhor estimativa é, aproximadamente:

- a) 6,5.
- b) 7.
- c) 7,5.
- d) 8.
- e) 8,5.

027. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte, com alguns valores não-identificados, mostra os resultados de uma inspeção visual no campo, relativos ao estado de conservação de 200 centros históricos de determinada região.

categoria	frequência	percentual (%)
ruim	50	C
regular	A	10
bom	100	D
excelente	B	E

Acerca dessa tabela, julgue o item subsequente.

Na tabela, a letra C corresponde a 20%.

028. (CESPE/ABIN/2018/OFICIAL TÉCNICO DE INTELIGÊNCIA/ÁREA 4) O diagrama de dispersão é adequado para se descrever o comportamento conjunto de duas variáveis quantitativas. Cada ponto do gráfico representa um par de valores observados.

029. (INSTITUTO VERBENA/MPE AC/ANALISTA MINISTERIAL/ÁREA ESTATÍSTICA/2023/Q3090436) Em um gráfico do tipo box-plot são representados: a mediana (Md), o 1º quartil ($Q1$), o 3º quartil ($Q3$) e os possíveis pontos atípicos (outliers). O intervalo interquartil (IQR) é a diferença entre $Q3$ e $Q1$. Para um valor x ser considerado atípico em um gráfico box-plot, ele deve satisfazer

- a) $x < Q1 - IQR$ ou $x > Q3 + IQR$
- b) $x > Q1 - IQR$ ou $x < Q3 + IQR$
- c) $x < Q1 - 1,5 \cdot IQR$ ou $x > Q3 + 1,5 \cdot IQR$
- d) $x > Q1 - 1,5 \cdot IQR$ ou $x < Q3 + 1,5 \cdot IQR$

030. (FGV/TJ SE/ANALISTA JUDICIÁRIO/ÁREA ESTATÍSTICA/2023/Q3033108) O boxplot é um recurso gráfico usado para analisar a distribuição dos dados e visualizar a presença de valores discrepantes (outlier) com base nas medidas estatísticas descritivas, como mínimo, máximo, primeiro quartil, segundo quartil ou mediana e terceiro quartil.

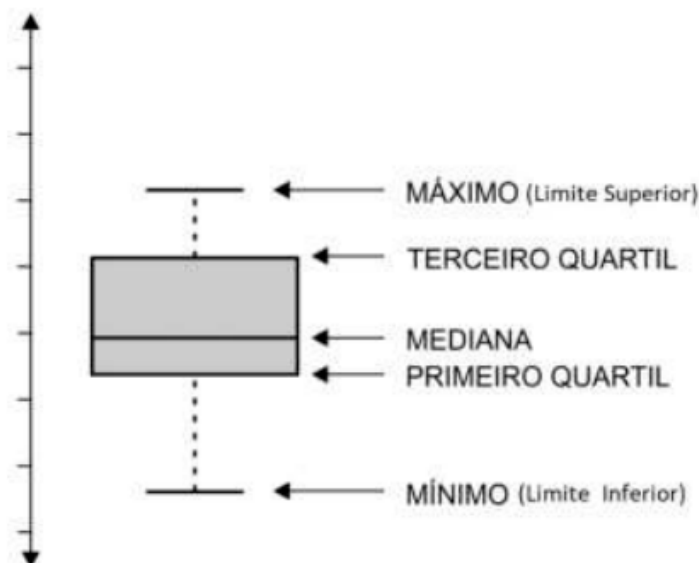
Nesse diagrama, o Intervalo Interquartil (IQR) é utilizado para calcular os limites superior e inferior, como:

$IQR = (\text{Terceiro Quartil} - \text{Primeiro Quartil})$

Limite Superior = Terceiro Quartil + $1,5 \cdot IQR$

Limite Inferior = Primeiro Quartil - $1,5 \cdot IQR$

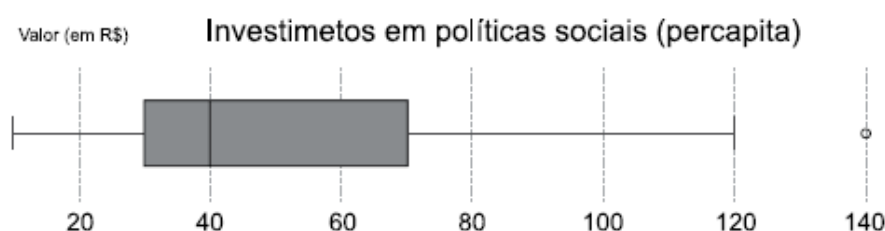
Observe o *boxplot* a seguir.



Dado um conjunto de valores, um valor é considerado outlier se estiver:

- a) entre o terceiro quartil e a mediana.
- b) entre a mediana e o primeiro quartil.
- c) entre o primeiro quartil e o limite inferior.
- d) entre o terceiro quartil e o limite superior.
- e) acima do limite superior ou abaixo do limite inferior.

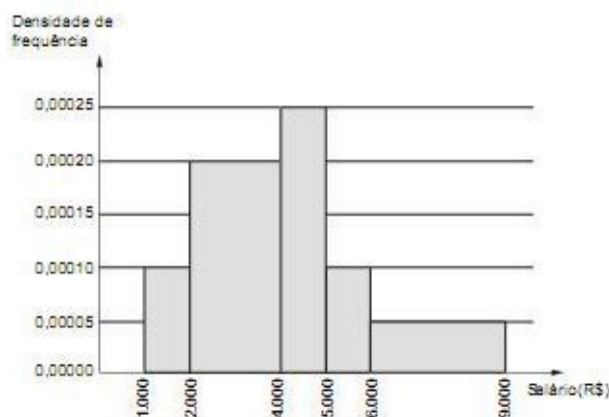
031. (VUNESP/PREFEITURA DE SÃO PAULO/ANALISTA DE POLÍTICAS PÚBLICAS/ÁREA GESTÃO GOVERNAMENTAL/2023/Q3059735) Analise o box plot a seguir, que traz informações sobre o valor de investimentos per capita em política social durante um mês em um determinado município.



Em relação ao gráfico apresentado, é possível afirmar que

- a) 25% das pessoas analisadas estão entre os valores de R\$40,00 e R\$60,00.
- b) não há outlier no gráfico analisado.
- c) a moda deste conjunto é de R\$40,00.
- d) a mediana do conjunto de dados apresentado é de R\$40,00.
- e) o primeiro quartil possui o valor de R\$40.

032. (FCC/TRF (2ª REGIÃO)/2012/ANALISTA JUDICIÁRIO) Considere que a distribuição dos salários dos funcionários em um setor público está representada por um histograma conforme abaixo, em que no eixo vertical constam as densidades de frequências, em $(R\$)^{-1}$. Densidade de frequência de um intervalo de classe é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.



Considerando que todos os intervalos classe são fechados à esquerda e abertos à direita, a porcentagem P dos funcionários que ganham no mínimo R\$ 2.000,00 e menos que R\$ 6.000,00 é tal que

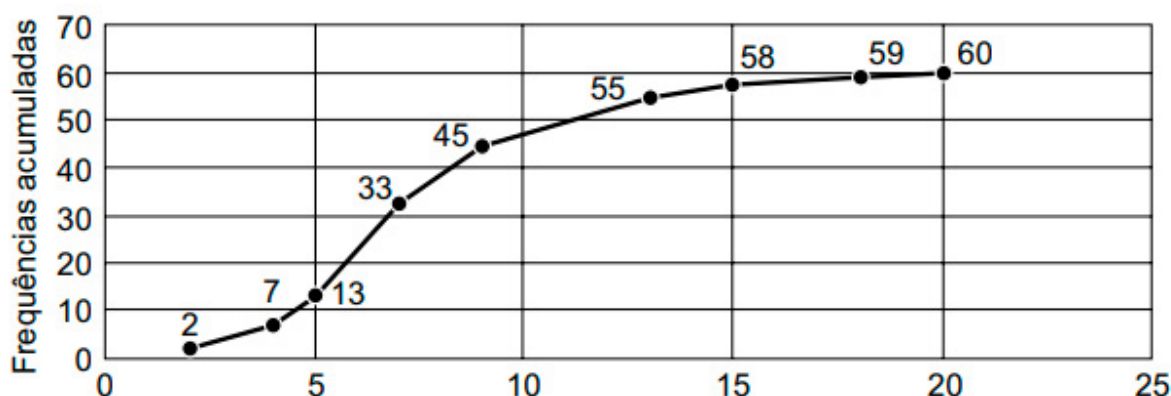
- a) $P < 65\%$.
- b) $65\% < P < 70\%$.
- c) $70\% < P < 75\%$.
- d) $75\% < P < 80\%$.
- e) $P > 80\%$.

033. (VUNESP/TRF 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE/ ESTATÍSTICA/2023/Q3160772)

O gráfico a seguir mostra os valores da frequência acumulada da distribuição de frequências sem perda de informação de certa variável x .

GRÁFICO 1

Frequências acumuladas da variável x



Variável x

Do gráfico, podemos inferir que a frequência relativa de $x = 5$ é:

- a) 7,3%;
- b) 12%;
- c) 15,8%.
- d) 10%.
- e) 15%.

034. (FGV/TJ/SE/2023/ANALISTA JUDICIÁRIO/ESTATÍSTICA) A análise dos dados multivariados, como a correlação linear entre dois atributos do conjunto de dados X , pode ser facilitada pelo uso do gráfico de:

- a) pizza.
- b) boxplot.

- c) densidade.
- d) histograma.
- e) scatter plot.

035. (CEBRASPE/CGDF/2023/AUDITOR DE CONTROLE INTERNO) Considerando que a tabela precedente represente a distribuição de frequências absolutas de uma variável quantitativa x , é correto afirmar que, em relação a x , a média será

x	frequência absoluta
0	1
1	2
2	4
3	0
4	2
total	9

- a) igual à mediana e igual à moda.
- b) igual à mediana, que é menor que a moda.
- c) menor que a mediana, que é menor que a moda.
- d) maior que a mediana, que é maior que a moda.

036. (INSTITUTO CONSULPLAN/MPE BA/ANALISTA TÉCNICO/ÁREA ANALISTA DE SISTEMAS/2023/Q2765770) A tabela representa a distribuição das notas de desempenho dos candidatos em um processo de seleção para uma vaga de emprego. A última fase do processo de seleção consiste em uma entrevista, que será realizada apenas com os candidatos que obtiveram desempenho acima da nota de corte, calculada com base no valor do sexto decil, cujo valor é 51,8:

Distribuição das Notas dos Candidatos					
NOTAS	F_i	F_{ac}	f_i	f_{ac}	
0 — 20	5	5	0,07	0,07	
20 — 40	24	29	0,32	0,39	
40 — 60	27	56	0,36	0,75	
60 — 80	14	70	0,18	0,93	
80 — 100	5	75	0,07	1,00	
TOTAL	75		1,00		

Legenda:
 F_i – frequência absoluta simples
 F_{ac} – frequência absoluta acumulada
 f_i – frequência relativa simples
 f_{ac} – frequência relativa acumulada

“De acordo com os dados apresentados, é correto afirmar que 60% dos candidatados obtiveram notas _____ de 51,8; portanto, a empresa realizará _____ 40 entrevistas.” Assinale a alternativa que completa correta e sequencialmente a afirmativa anterior.

- a) acima / mais.
- b) abaixo / mais.
- c) acima / menos.
- d) abaixo / menos.
- e) acima / exatamente.

037. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte, com alguns valores não-identificados, mostra os resultados de uma inspeção visual no campo, relativos ao estado de conservação de 200 centros históricos de determinada região.

categoria	frequência	percentual (%)
ruim	50	C
regular	A	10
bom	100	D
excelente	B	E

Acerca dessa tabela, julgue o item subsequente.

A letra B, na tabela, representa 25 centros.

038. (CESPE/MPU/2013) A tabela acima mostra algumas estatísticas descritivas produzidas por um estudo acerca da quantidade de acidentes de trabalho (N), ocorridos em 2012, a partir de uma amostra aleatória simples de 200 indústrias de pequeno porte. Com base nessas informações, julgue o próximo item.

estatística	quantidade de acidentes (N)
mínimo	0
primeiro quartil	2
segundo quartil	4
terceiro quartil	10
máximo	30

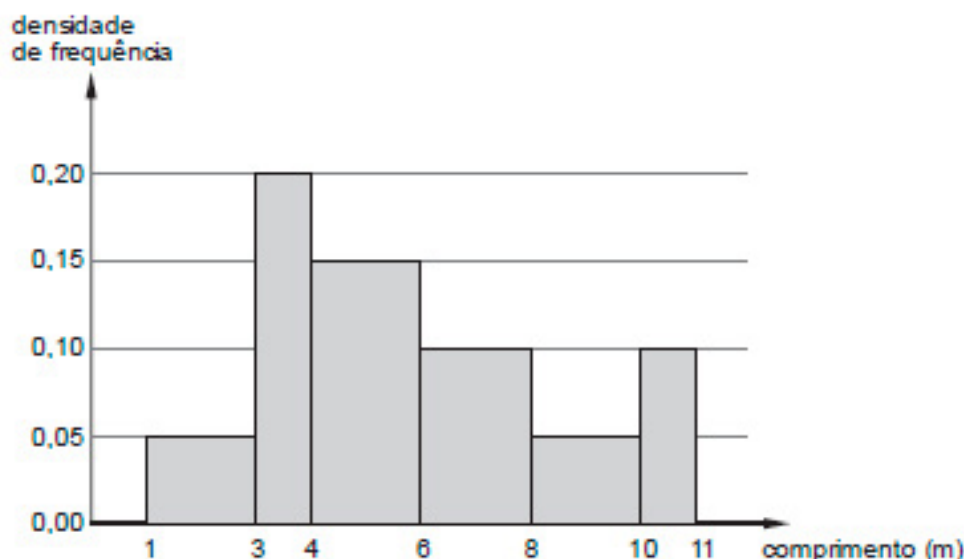
O intervalo interquartil da variável N foi igual a 8.

039. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) 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 ($Q_1/4$) é igual a 3.

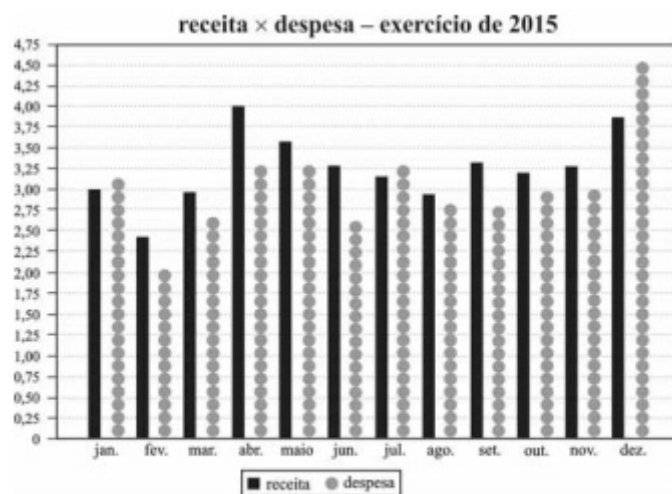
040. (FCC/TRT/5ª REGIÃO (BA)/2018/ANALISTA JUDICIÁRIO/ESTATÍSTICA) A distribuição das medidas em metros (m) dos comprimentos dos cabos no estoque de uma fábrica está representada pelo histograma mostrado abaixo, em que no eixo vertical constam as densidades de frequências, em $(m)^{-1}$, e no eixo horizontal os intervalos de classe. Define-se densidade de frequência de um intervalo de classe como sendo o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.



Sabendo-se que todos os intervalos de classe são fechados à esquerda e abertos à direita, então a porcentagem dos cabos que apresentam uma medida de comprimento de pelo menos igual a 4 m e inferior a 10 m é de

- a) 50%.
- b) 60%.
- c) 70%.
- d) 80%.
- e) 90%.

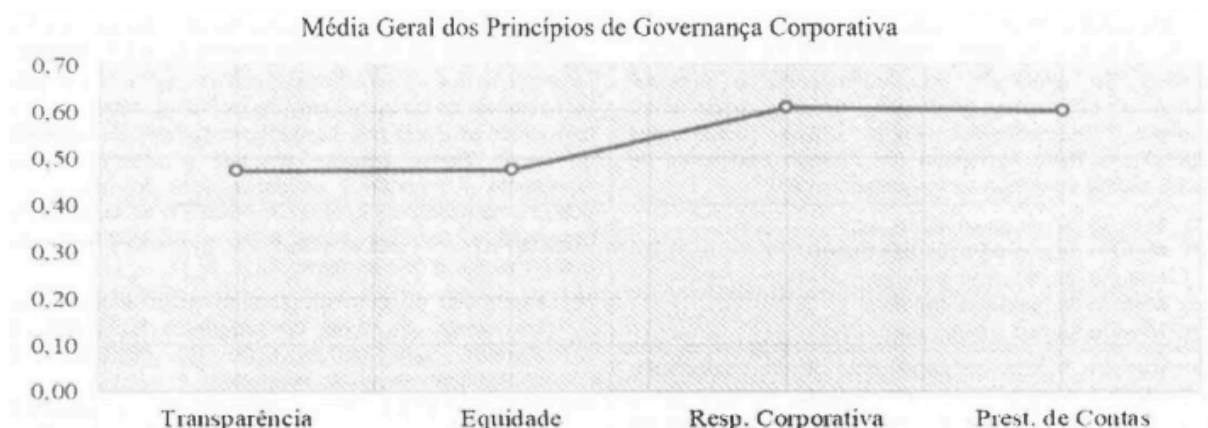
041. (CESPE/TCE/PR/2016) Tendo como referência o gráfico precedente, que mostra os valores, em bilhões de reais, relativos à arrecadação de receitas e aos gastos com despesas do estado do Paraná nos doze meses do ano de 2015, assinale a opção correta.



Internet: <www.gestaozinhopublico.pr.gov.br> (com adaptações).

- No ano considerado, o segundo trimestre caracterizou-se por uma queda contínua na arrecadação de receitas, situação que se repetiu no trimestre seguinte.
- No primeiro quadrimestre de 2015, houve um período de queda simultânea dos gastos com despesas e da arrecadação de receitas e dois períodos de aumento simultâneo de gastos e de arrecadação.
- No último bimestre do ano de 2015, foram registrados tanto o maior gasto com despesas quanto a maior arrecadação de receitas.
- No ano em questão, janeiro e dezembro foram os únicos meses em que a arrecadação de receitas foi ultrapassada por gastos com despesas.
- A menor arrecadação mensal de receitas e o menor gasto mensal com despesas foram verificados, respectivamente, no primeiro e no segundo semestre do ano de 2015.

042. (LEGALLE CONCURSOS/DPE PA/ANALISTA/ÁREA: ADMINISTRAÇÃO/2023/Q3011129)
Analisar o gráfico a seguir:



Da análise do gráfico, verifica-se que o princípio de Responsabilidade Corporativa (Resp. Corporativa) é aquele que rege a evidenciação de atributos de governança corporativa e apresenta a média:

- a) 0,47.
- b) 0,70.
- c) 0,61.
- d) 0,35.
- e) 0,48.

043. (CEBRASPE BANESE 2021 TÉCNICO BANCÁRIO I) O primeiro quartil do conjunto de dados em tela é igual ou superior a 33.

044. (CEBRASPE/PETROBRÁS/2021) A amplitude total dos dados em tela é inferior a 6.

045. (CEBRASPE/PETROBRÁS/2021) Nesse diagrama, a porção da distribuição dos dados representada pela parte inferior do diagrama mostrada a seguir representa exatamente 25% dos dados em questão.



046. (CEBRASPE/PETROBRÁS/2021) O terceiro quartil é inferior a 11 e superior a 10.

047. (VUNESP/TRF 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2023/Q3160731) Para as questões de números 21 a 25, considere a seguinte tabela de distribuição de frequência com perda de informação relativa ao consumo mensal de água, em m³, do quarteirão Q da cidade C, obtida de uma amostra aleatória de residências do local.

TABELA 1

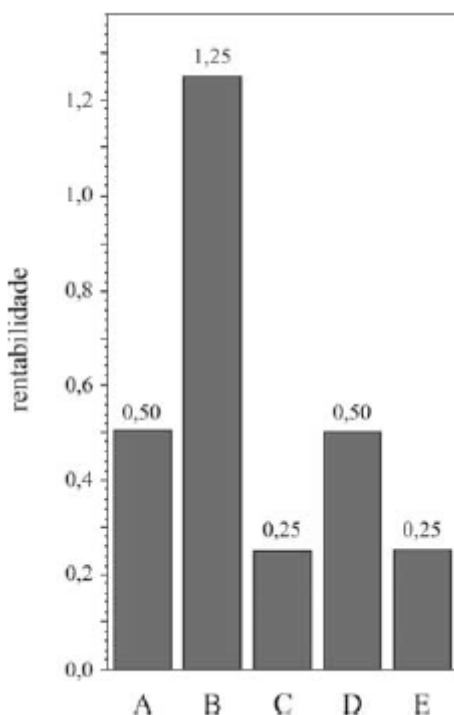
Consumo mensal de água do quarteirão Q da cidade C

Consumo em m ³	Número de residências (f_i)
10 ----20	4
20 ----30	6
30 ----40	10
40 ----50	12
50 ----60	8

A estimativa de residências que consomem entre o valor do limite inferior da 2ª classe e o valor do limite superior da 4ª classe é:

- a) 84%.
- b) 80%.
- c) 65%.
- d) 70%.
- e) 74%.

048. (CESPE/FUNPRESP/EXE/2016/ANALISTA/ÁREA DE INVESTIMENTOS) O gráfico ilustra cinco possibilidades de fundos de investimento com suas respectivas rentabilidades. Considerando que as probabilidades de investimento para os fundos A, B, C e D sejam, respectivamente, $P(A) = 0,182$; $P(B) = 0,454$; $P(C) = 0,091$; e $P(D) = 0,182$, julgue o item subsequente.

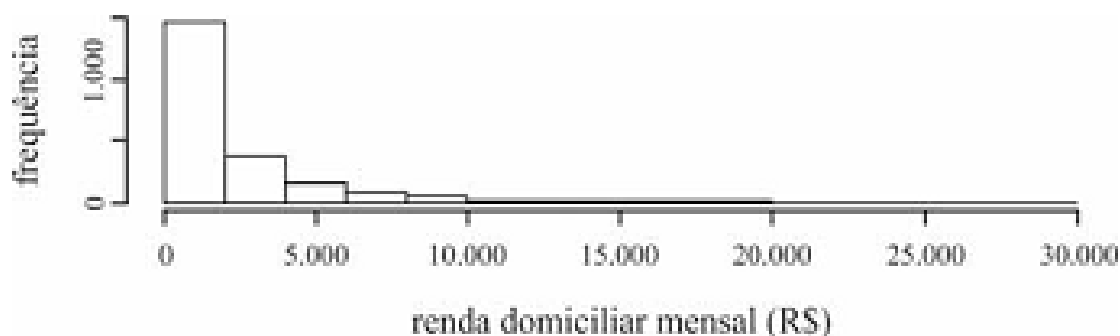


O gráfico apresentado é um histograma.

049. (COMDN/MARINHA/QUADRO TÉCNICO DO CORPO AUXILIAR DA MARINHA/ÁREA: ESTATÍSTICA/2022/Q2634582) A representação gráfica das séries estatísticas tem por finalidade representar os resultados obtidos, permitindo que se chegue a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Assinale a opção que corresponde à representação de uma série por meio de um polígono que, para ser construído, divide-se uma circunferência em tantos arcos iguais forem os dados a representar.

- a) Gráfico em colunas.
- b) Gráfico em barras.
- c) Gráfico em setores.
- d) Gráfico polar.
- e) Gráfico em curvas.

050. (CESPE/TELEBRÁS/2015/ANALISTA SUPERIOR/AUDITORIA) Uma empresa coletou e armazenou em um banco de dados diversas informações sobre seus clientes, entre as quais estavam o valor da última fatura vencida e o pagamento ou não dessa fatura. Analisando essas informações, a empresa concluiu que 15% de seus clientes estavam inadimplentes. A empresa recolheu ainda dados como a unidade da Federação (UF) e o CEP da localidade em que estão os clientes. Do conjunto de todos os clientes, uma amostra aleatória simples constituída por 2.175 indivíduos prestou também informações sobre sua renda domiciliar mensal, o que gerou o histograma apresentado.



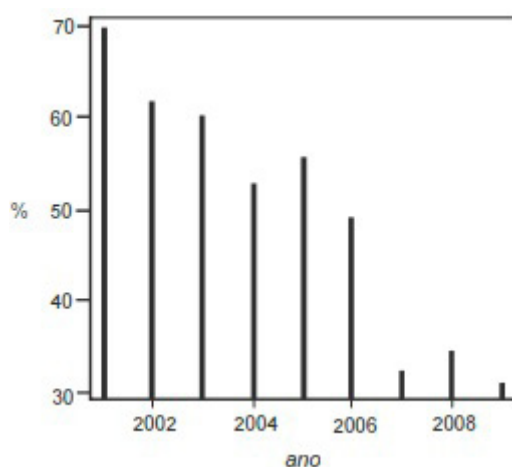
Com base nessas informações e no histograma, julgue o item a seguir.

Se for elaborado um histograma com classes de larguras variáveis para representar a distribuição dos valores das últimas faturas vencidas, então a classe com maior altura no histograma será, necessariamente, aquela com maior frequência no banco de dados.

051. (FCC/TRT/3ª REGIÃO (MG)/2015/ANALISTA JUDICIÁRIO/ESTATÍSTICA) Em um histograma representando os preços unitários de microcomputadores em estoque, observa-se que no eixo das abscissas constam os intervalos de classe em R\$ e no eixo das ordenadas as respectivas densidades de frequências em (R\$) -1 . Densidade de frequência de um intervalo de classe é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo. Um determinado intervalo de classe com amplitude igual a R\$ 2.500,00 apresenta uma densidade de frequência, em (R\$) -1 , igual a $12,8 \times 10^{-5}$. Se o número de microcomputadores deste intervalo é igual a 48, então o número total de microcomputadores em estoque é igual a

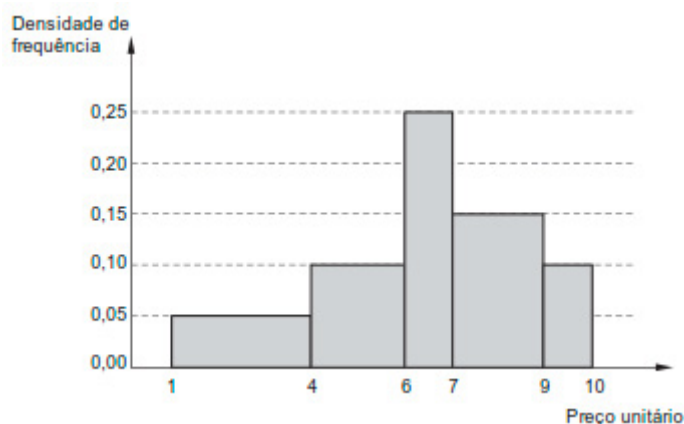
- a) 150.
- b) 120.
- c) 240.
- d) 160.
- e) 96.

052. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4) Se os percentuais forem representados por barras verticais, conforme o gráfico a seguir, então o resultado será denominado histograma.



053. (FCC/TRE/RR/2015/ANALISTA JUDICIÁRIO) 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.



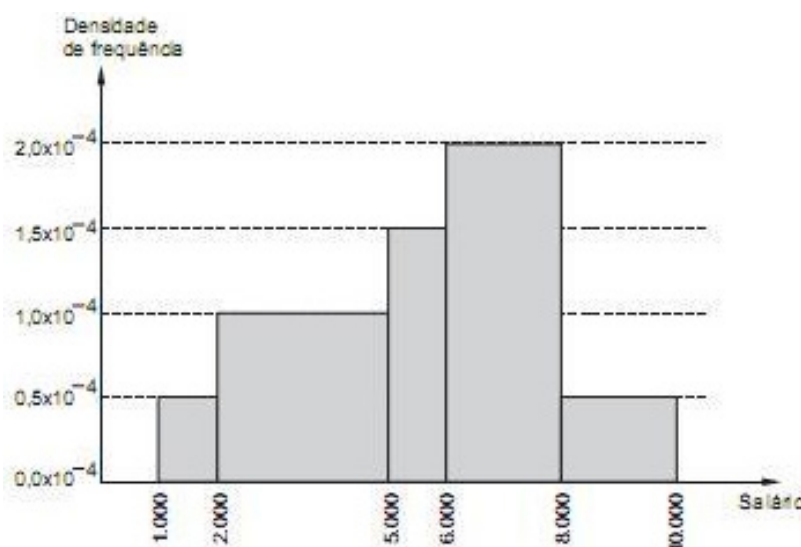
Considerando os intervalos de classe fechados à esquerda e abertos à direita, se 105 preços apresentam valores menores que R\$ 6,00, então o número de preços que apresentam valores iguais ou superiores a R\$ 4,00 é:

- a) 240.
- b) 195.
- c) 215.
- d) 230.
- e) 255.

054. (FCC/SEFAZ/PI/2015/AUDITOR FISCAL DA FAZENDA ESTADUAL/ADAPTADA) O histograma é um gráfico apropriado para representar dados de variáveis quantitativas contínuas.

055. (CESPE/SEE/AL/2013/SECRETÁRIO ESCOLAR) Considere que, para classificar as escolas de uma cidade, segundo as notas médias em matemática e português, tenha sido obtida uma amostra das notas de cinquenta alunos de cada uma das dez escolas do município. Com base nessas informações, julgue o item subsequente.
Considerando-se que a variável nota em matemática seja quantitativa contínua, então o gráfico que representa esse tipo de dado é o histograma.

056. (FCC/TRT/13ª REGIÃO (PB)/2014/ANALISTA JUDICIÁRIO/ESTATÍSTICA) O histograma, abaixo, refere-se à distribuição dos salários dos funcionários lotados em um setor de um órgão público. No eixo das abscissas constam os intervalos de classe em R\$ (todos fechados à esquerda e abertos à direita) e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)^{-1}$. Define-se densidade de frequência de um intervalo como sendo o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.



Se o número de funcionários que tem um salário inferior a R\$ 5.000,00 é igual a 56, então verifica-se que o número de funcionários que tem um salário igual ou superior a R\$ 2.000,00 e inferior a R\$ 8.000,00 é igual a

- a) 104.
- b) 152.
- c) 136.
- d) 144.
- e) 120.

057. (FCC/TRT 23/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2022/Q2593239) A tabela de frequências absolutas abaixo refere-se à distribuição dos salários (S) dos empregados que não possuem nível superior em uma empresa, sendo que não foram fornecidas as frequências da 2a e 3a classes dos empregados homens, denotadas na tabela por x e y, respectivamente.

Classes de Salários R\$	(Frequências) Homens	(Frequências) Mulheres	Total
$1.500 < S \leq 2.500$	20	25	45
$2.500 < S \leq 3.500$	x	40	$40 + x$
$3.500 < S \leq 4.500$	y	10	$10 + y$
$4.500 < S \leq 5.500$	15	5	20
Total	80	80	160

Foram calculadas separadamente, pelo método da interpolação linear, as medianas dos empregados homens e das mulheres, sendo que o valor da mediana referente ao dos homens superou em R\$ 425,00 o valor referente ao das mulheres. O valor médio dos salários dos homens, calculado como se todos os valores de uma classe coincidam com o ponto médio da respectiva classe, é igual a

- a) R\$ 3.925,00.
- b) R\$ 3.525,00.
- c) R\$ 3.875,00.
- d) R\$ 3.375,00.
- e) R\$ 3.625,00.

058. (FGV/EPE/ANALISTA DE PESQUISA ENERGÉTICA/ÁREA PLANEJAMENTO DA GERAÇÃO DE ENERGIA/2022/Q2358240) Os dados abaixo representam a distribuição de 100 domicílios residenciais, por classe de consumo de energia elétrica mensal, medidos em 2016.

Faixa de Consumo	Frequência Absoluta
0 --- 100	12
100 --- 200	25
200 --- 300	30
300 --- 400	25
400 --- 800	8

O consumo médio mensal, em kWh, pode ser estimado em

- a) 180.
- b) 254.
- c) 275.
- d) 300.
- e) 336.

059. (INSTITUTO CONSULPLAN/IFPA/ESTATÍSTICO/2023/Q2868120) O histograma, também conhecido como distribuição de frequências, é a representação gráfica em colunas ou em barras (retângulos) contíguas de um conjunto de dados previamente tabulado e dividido em classes equidistantes ou não. O eixo vertical do histograma pode ser construído a partir da frequência absoluta ou da densidade, definida como a frequência relativa da classe dividida pela amplitude da classe. Sobre a construção de um histograma, analise as afirmativas a seguir.

I – Se as classes são equidistantes, as formas dos histogramas construídos a partir da frequência absoluta e da densidade são equivalentes.

II – Em qualquer histograma, a classe modal é representada pela barra de maior altura.

III – No histograma construído a partir da densidade, a área do gráfico é igual a 1.

Está correto o que se afirma em

- a) I, II e III.
- b) II, apenas.
- c) I e III, apenas.
- d) II e III, apenas.

060. (IBFC/PREFEITURA DE CUIABÁ/ESTATÍSTICO/2023/Q2738868) A tabela a seguir fornece a distribuição de frequências das notas de 45 alunos de um curso de estatística básica:

Intervalo de classe	Frequência simples	Frequência relativa
0 [— 3	4	9
3 [— 5	A	22
5 [— 7	25	C
7 [— 10	B	D

Os valores de A, B, C e D são, aproximadamente:

- a) 10, 6, 40, 13.
- b) 6, 10, 50, 13.
- c) 10, 6, 13, 40.
- d) 10, 6, 40, 14.

061. (EPL CONCURSOS/TRE RN/ESTATÍSTICA/2023/Q3103792) Na tabela de distribuição de frequência apresentada a seguir alguns valores foram omitidos e outros foram substituídos pelas letras A, B, C, D e E.

Percentual de redução	f_i	$f_r(\%)$	F_{AC}
2 – 15	12	30,0	12
15 – 28	11	D	23
28 – 41	B		35
41 – 54	3		C
54 – 67	1	2,50	
67 – 80		E	
Total	A	100	-

Considerando os valores numéricos de A, B, C, D e E, assinale a alternativa correta.

- a) A = 39.
- b) B = 13.
- c) C = 40.
- d) D = 27.
- e) E = 2,5.

062. (CESGRANRIO/IBGE/2010/ANALISTA DE SISTEMAS) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

Julgue o item a seguir.

Classes (em anos)	f_i
0 – 2	5
2 – 4	2
4 – 6	4
6 – 8	2
8 – 10	7

A média das idades dessas crianças, em anos, é

- a) 5,0.
- b) 5,2.
- c) 5,4.
- d) 5,6.
- e) 5,8.

063. (CESGRANRIO/IBGE/2010/ANALISTA DE SISTEMAS) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

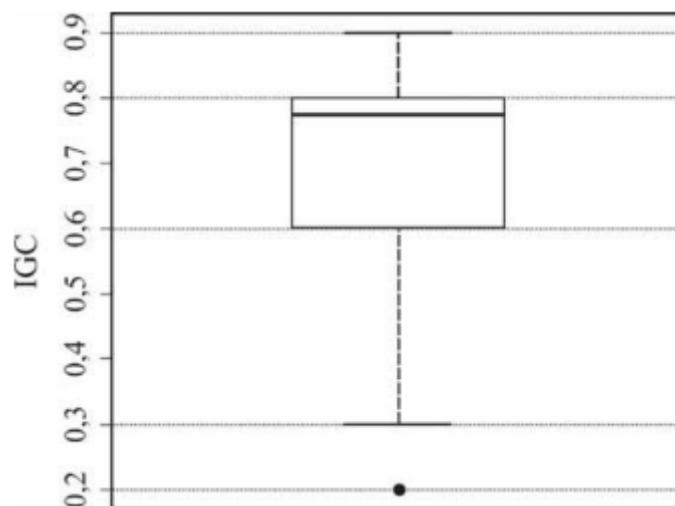
Julgue o item a seguir.

Classes (em anos)	f_i
0 – 2	5
2 – 4	2
4 – 6	4
6 – 8	2
8 – 10	7

A mediana das idades dessas crianças, em anos, é

- a) 5,5.
- b) 5,6.
- c) 5,7.
- d) 5,8.
- e) 5,9.

(CESPE/TCE/PR/2016/ADAPTADA) 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, julgue os seguintes itens.



064. O menor IGC observado na amostra foi 0,3.

065. O maior IGC observado na amostra foi 0,9.

066. O diagrama mostrado na figura em questão é denominado curva de frequência.

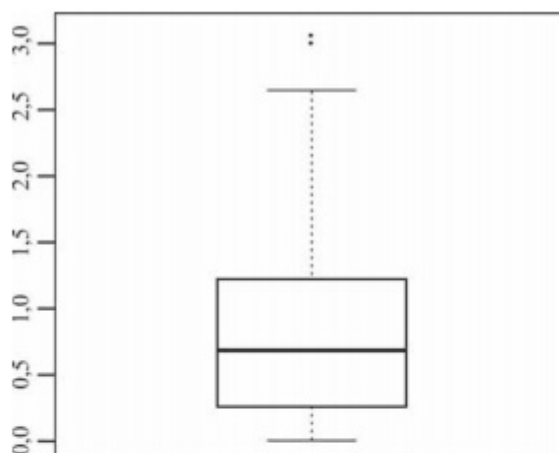
067. O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

068. Na amostra considerada, a média dos indicadores observados foi superior a 0,7 e inferior a 0,8.

069. A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

070. O desvio quartílico para a amostra em questão é de 0,3.

(CESPE/TCE/PA/2016) 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.



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

Com base nessas informações, julgue os itens a seguir.

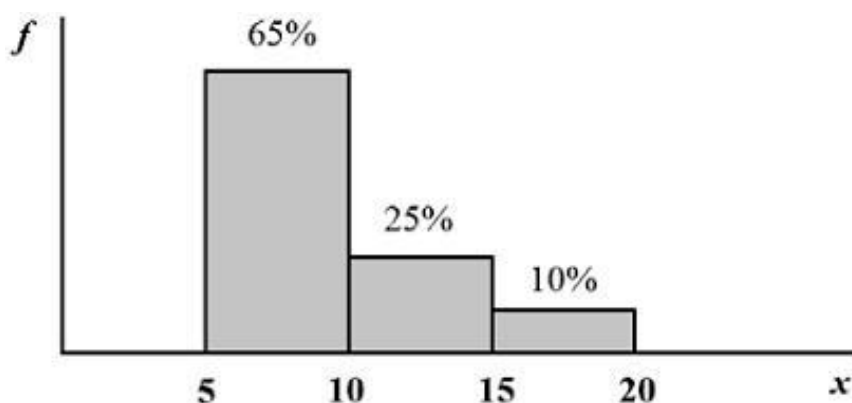
071. O diagrama box-plot mostrado na figura sugere a existência de pelo menos duas observações atípicas.

072. X representa uma variável qualitativa ordinal.

073. A amplitude total da amostra é inferior a 3.

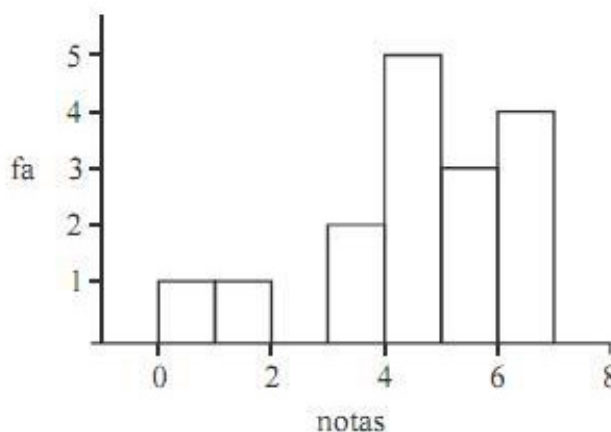
074. Pelo menos 50% das observações de X são menores que 1,5.

075. (CESPE/STF/2013/ANALISTA JUDICIÁRIO/ESTATÍSTICA) Com referência à figura acima, que mostra a distribuição da renda mensal — x , em quantidades de salários mínimos (sm) — das pessoas que residem em determinada região, julgue o item subsequente.



A variável x , por possuir quatro níveis de respostas, é do tipo qualitativa ordinal.

076. (CESPE/FUB/2013/ESTATÍSTICO) Considerando que o histograma acima apresenta a distribuição das notas finais dos estudantes matriculados em determinada disciplina, e que f_a representa a frequência absoluta, julgue o seguinte item acerca de estatística descritiva.



Nessa disciplina, o percentual de estudantes com notas inferiores a 2,5 é igual a 12,5%.

077. (CESPE/FUB/2013/ESTATÍSTICO) A amplitude total da distribuição das notas é igual a 6.

078. (CESPE/BACEN/2013/ANALISTA DE INFRAESTRUTURA E LOGÍSTICA)

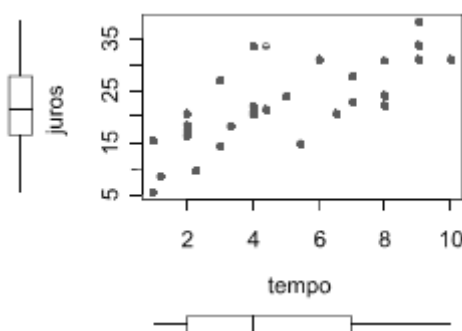
2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

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.

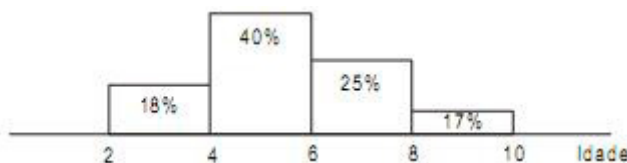
É inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos; dessa forma, apenas por meio de um gráfico de barras pode ser realizada a representação gráfica.

079. (CESPE/BANCO DA AMAZÔNIA/TÉCNICO CIENTÍFICO/ESTATÍSTICA) Em um estudo sobre o valor de juros (R\$) e encargos pagos versus tempo de atraso (dias) considerando o pagamento da fatura de cartão de crédito de 30 clientes, foi construído um diagrama de dispersão e ajustado um modelo de regressão linear simples, seguido de quatro gráficos diagnósticos, apresentados a seguir.



Os gráficos projetados nas margens do diagrama de dispersão são chamados de desenho esquemático ou *boxplot*, que resume graficamente a tendência central e a dispersão dos dados, sendo construído utilizando-se mínimo, máximo, média e quartis.

080. (FCC/ANS/2007/ESPECIALISTA EM REGULAÇÃO) O histograma abaixo representa a distribuição das idades dos pacientes atendidos no ano de 2000 em uma clínica infantil, expressa em anos.



A idade que separa os 30% mais jovens é

- a) 3,5.
- b) 4,2.
- c) 4,4.
- d) 4,6.
- e) 5,0.

081. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL) 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.

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

O segundo quartil da distribuição das quantidades diárias de incidentes registradas nessa penitenciária é igual a 2.

082. (CESPE/MEC/2014/ANALISTA DE POLÍTICA REGULATÓRIA) A tabela acima, resultado de um estudo socioeconômico, mostra a distribuição percentual da renda familiar mensal dos estudantes do ensino médio em determinado município brasileiro. Considerando essas informações e a tabela acima, julgue o item seguinte.

renda familiar (R) (em salários mínimos)	percentual (%)
$0 < R \leq 1$	40
$1 < R \leq 3$	50
$R > 3$	10
total	100

O terceiro quartil dessa distribuição de renda é superior a três salários mínimos.

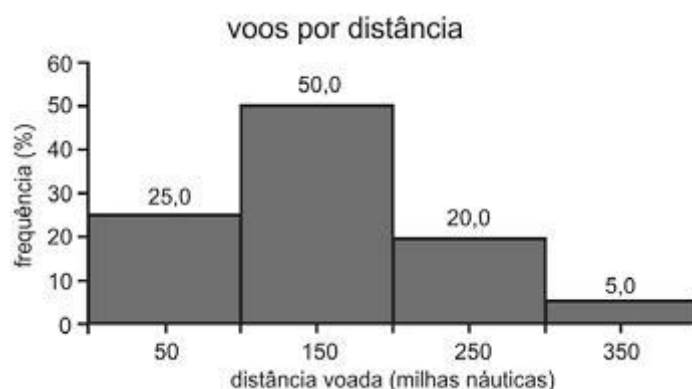
083. (CESPE/CEHAP/PB/2009/ADMINISTRADOR) O custo médio nacional para a construção de habitação com padrão de acabamento normal, segundo levantamento realizado em novembro de 2008, foi de R\$ 670,00 por metro quadrado, sendo R\$ 400,00/m² relativos às despesas com materiais de construção e R\$ 270,00/m² com mão-de-obra. Nessa mesma

pesquisa, os custos médios regionais apontaram para os seguintes valores por metro quadrado: R\$ 700,00 (Sudeste), R\$ 660,00 (Sul), R\$ 670,00 (Norte), R\$ 640,00 (Centro-Oeste) e R\$ 630,00 (Nordeste).

Considerando o par de variáveis (X, Y), em que X representa o custo médio por metro quadrado (em R\$) e Y representa os atributos Sudeste, Norte, Sul, Centro-Oeste ou Nordeste, assinale a opção correta.

- a) A correlação linear de Pearson entre X e Y é positiva.
- b) As variáveis X e Y são qualitativas.
- c) As variáveis X e Y formam uma série geográfica.
- d) O gráfico de X versus Y é um histograma.

084. (CESPE/ANAC/2012/ESPECIALISTA EM REGULAÇÃO DE AVIAÇÃO CIVIL) Com base no histograma acima apresentado, julgue o item a seguir.



Os quartis das distâncias voadas (em milhas náuticas) obtidos a partir do gráfico são: 1º quartil = 100,0; 2º quartil (mediana) = 150,0; e 3º quartil = 200,0.

085. (FCC/TRT/3ª REGIÃO (MG)/2015) A tabela de frequências relativas abaixo refere-se à distribuição dos salários dos empregados de uma empresa no mês de maio de 2015.

CLASSE DE SALÁRIOS (R\$)	FREQUÊNCIA RELATIVA
a — b	0,10
b — c	0,20
c — d	0,25
d — e	0,40
e — f	0,05
TOTAL	1,00

Observação: Todos os intervalos de classe apresentam a mesma amplitude igual a R\$ 1.000,00 e utilizou-se o método da interpolação linear para calcular a mediana (M_d) e o terceiro quartil (Q_3) da distribuição.

Se $Md = R\$ 4.200,00$, então $Q3$ é, em $R\$$, igual a

- a) 4.500,00.
- b) 4.900,00.
- c) 4.640,00.
- d) 4.800,00.
- e) 4.720,00.

086. (FGV/IBGE/2017/ANALISTA CENSITÁRIO/MÉTODOS QUANTITATIVOS) Uma vez concluída a etapa de críticas de dados, relativa a um conjunto de registros obtido através de uma pesquisa de campo, inicia-se o trabalho de tabulação e elaboração de gráficos. Durante as análises, algumas variáveis surgem com destaque.

(i) número de indivíduos por faixa etária; (ii) percentuais do nível de escolaridade; e (iii) pares de valores de consumo e renda.

Portanto, os tipos de gráficos considerados adequados a serem empregados em cada caso são, respectivamente:

- a) colunas, linha e setores.
- b) setores, cartograma e barras.
- c) barras, setores e linha.
- d) linha, colunas e cartograma.
- e) colunas, setores e linha.

087. (VUNESP/PREFEITURA DE ILHABELA/SP/2020/ANALISTA DE GESTÃO PÚBLICA/DESAFIO)

Um histograma foi elaborado com o objetivo de se analisar a distribuição de todos os salários dos funcionários em uma empresa. Sabe-se que 108 funcionários dessa empresa ganham salários com valores pertencentes ao segundo intervalo do histograma $[4, 7)$, em salários-mínimos (SM), que apresenta uma densidade de frequência igual a $0,15 (SM)^{-1}$. Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela amplitude deste intervalo. Se o primeiro intervalo do histograma $[2, 4)$, em SM, apresenta uma densidade de frequência igual a $0,125 (SM)^{-1}$, então o número de empregados do segundo intervalo supera o número de empregados do primeiro intervalo em

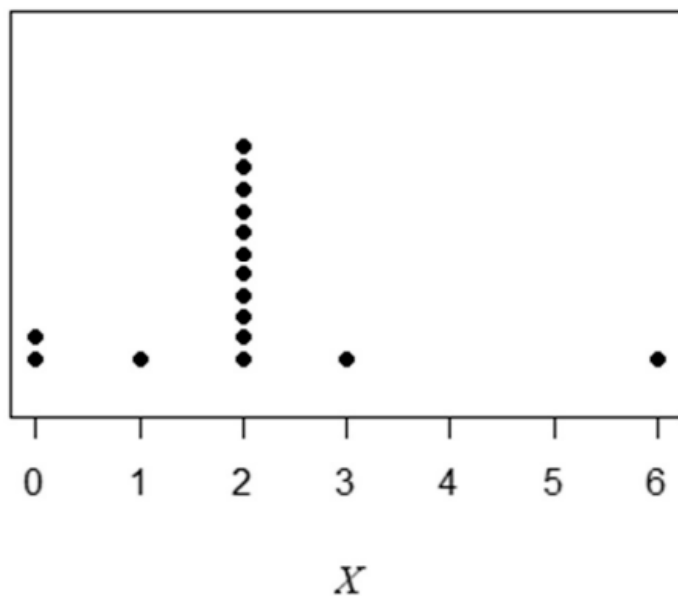
- a) 48.
- b) 60.
- c) 72.
- d) 78.
- e) 84.

GABARITO

- | | | |
|-------|-------|-------|
| 1. C | 36. d | 71. C |
| 2. C | 37. E | 72. E |
| 3. C | 38. C | 73. E |
| 4. E | 39. E | 74. C |
| 5. C | 40. b | 75. E |
| 6. E | 41. b | 76. C |
| 7. d | 42. c | 77. E |
| 8. E | 43. E | 78. E |
| 9. E | 44. E | 79. E |
| 10. E | 45. E | 80. d |
| 11. E | 46. C | 81. C |
| 12. e | 47. d | 82. E |
| 13. C | 48. E | 83. c |
| 14. d | 49. d | 84. C |
| 15. E | 50. E | 85. b |
| 16. a | 51. a | 86. e |
| 17. E | 52. E | 87. a |
| 18. d | 53. e | |
| 19. e | 54. C | |
| 20. b | 55. C | |
| 21. C | 56. c | |
| 22. C | 57. d | |
| 23. c | 58. b | |
| 24. b | 59. c | |
| 25. a | 60. a | |
| 26. b | 61. e | |
| 27. E | 62. c | |
| 28. C | 63. a | |
| 29. c | 64. E | |
| 30. e | 65. C | |
| 31. c | 66. E | |
| 32. c | 67. E | |
| 33. d | 68. E | |
| 34. e | 69. C | |
| 35. a | 70. E | |

GABARITO COMENTADO

021. (CESPE/CEBRASPE/CGE/RJ/AUDITOR DO ESTADO/2024/Q3142790)



Considerando a figura precedente, que mostra a distribuição de frequências de uma variável X em uma amostra aleatória simples de tamanho igual a 16, julgue o próximo item.

A média, a mediana e a moda do conjunto de dados em questão são iguais a 2.



Temos que $n=16$

Vamos organizar os dados em uma tabela

X	Frequência
0	2
1	1
2	11
3	1
4	0
5	0
6	1

Se $n=16$, a mediana será a média dos valores nas posições 8 e 9, que estão na 3ª classe de Var $x = 2$, isso nos dá:

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

Moda é o valor que mais se repetiu, que será 2 (repetiu-se 11 vezes)

Média será:

$$Média = \frac{(0.2) + (1.1) + (2.11) + (3.1) + (4.0) + (5.0) + (6.1)}{16} = \frac{32}{16} = 2$$

Certo.

022. (CESPE/CEBRASPE/ANP/ATIVIDADES DE REGULAÇÃO DE NOVAS ATRIBUIÇÕES/2022/Q2726118)

ocorrência	comportamento	registrado	educação	idade
1	agressivo	sim	fundamental	15-24
3	calmo	sim	fundamental	25-39
2	agressivo	sim	fundamental	40-
2	calmo	sim	médio	15-24
8	agressivo	sim	médio	25-39
1	calmo	sim	fundamental	15-24
3	agressivo	sim	fundamental	25-39
4	calmo	sim	fundamental	40-
2	calmo	sim	médio	15-24
2	agressivo	sim	médio	25-39
5	agressivo	sim	médio	25-39
7	calmo	sim	infantil	5-14
1	agressivo	não	infantil	5-14
3	agressivo	sim	infantil	5-14

Considerando os dados da tabela anterior, representados na forma de um dataframe hipotético em relação a determinado conjunto de registros escolares a respeito da violência e do comportamento de discentes, julgue o item a seguir.

O número de ocorrências com discentes de comportamento agressivo e com educação de nível médio é, no total, inferior a 16.

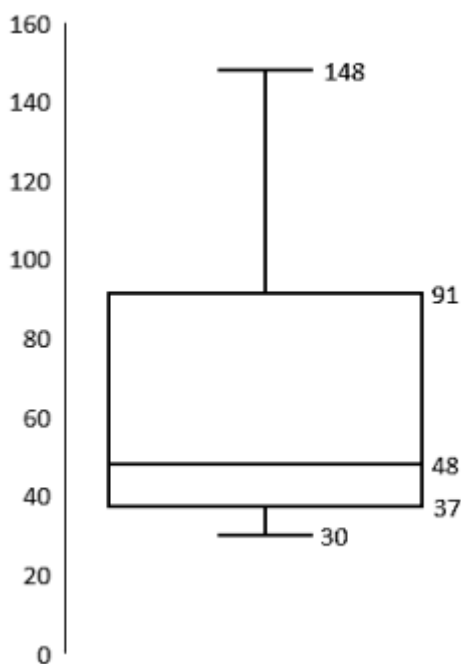


Observando a tabela identificamos 3 classes de alunos do ensino médio com comportamento agressivo, que nos dá um total de 15 ocorrências.

ocorrência	comportamento	registrado	educação	idade
1	agressivo	sim	fundamental	15-24
3	calmo	sim	fundamental	25-39
2	agressivo	sim	fundamental	40-
2	calmo	sim	médio	15-24
8	agressivo	sim	médio	25-39
1	calmo	sim	fundamental	15-24
3	agressivo	sim	fundamental	25-39
4	calmo	sim	fundamental	40-
2	calmo	sim	médio	15-24
2	agressivo	sim	médio	25-39
5	agressivo	sim	médio	25-39
7	calmo	sim	infantil	5-14
1	agressivo	não	infantil	5-14
3	agressivo	sim	infantil	5-14

Certo.

023. (FUMARC/TRT 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2022/Q2631573) O diagrama a seguir é chamado Box Plot (Diagrama de Caixa). Ele está representando uma amostra de 130 valores.



Baseado nas informações fornecidas no diagrama, é CORRETO afirmar:

- a) A Amplitude Interquartílica é de 128.
- b) Caso existisse na amostra um valor igual a 10, ele seria considerado um Outlier.
- c) Caso existisse na amostra um valor superior a 172, ele seria considerado um Outlier.
- d) No intervalo de 30 a 91, existem 50% do total de valores da amostra.
- e) O valor do 1º Quartil é 30, o do 2º Quartil é 37, o do 3º Quartil é 48 e o do 4º Quartil é 91.



Vamos resolver cada alternativa:

- a) **Errada.** A amplitude interquartílica é igual a $Q3 - Q1 = 91 - 37 = 54$.
- b) **Errada.** Os valores que são chamados de outliers são aqueles que estão fora dos limites inferior e superior.

Os limites inferior (Li) e superior (Ls) são calculados da seguinte forma:

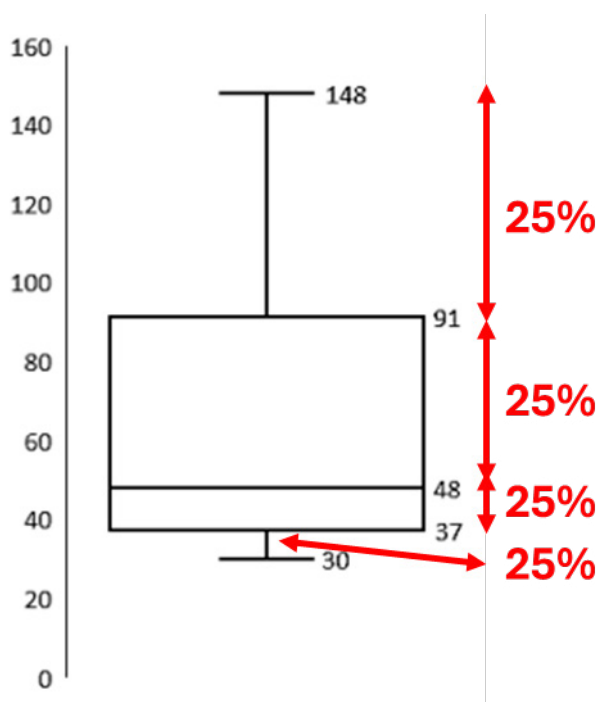
$$Li = Q1 - 1,5(Q3 - Q1) = 37 - 1,5(91 - 37) = -50$$

$$Ls = Q3 + 1,5(Q3 - Q1) = 91 + 1,5(91 - 37) = 172$$

Portanto, se o valor da amostra fosse igual a 10, não seria considerado um outlier.

c) Certa. O Limite Superior calculado no item anterior é igual a 172.

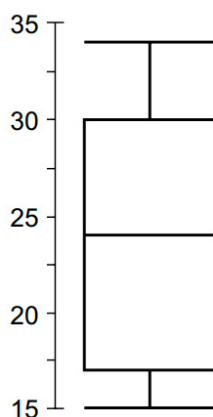
d) **Errada.** No intervalo de 30 a 91, existem 75% do total de valores da amostra, pois essa é uma característica geral dos diagramas box plot.



e) **Errada.** O valor do 1º Quartil é 37, o do 2º Quartil é 48 e o do 3º Quartil é 91. Não existe 4º Quartil.

Letra c.

024. (VUNESP/UNICAMP/PROFISSIONAL/ÁREA: DESENHISTA DE PÁGINAS DA INTERNET/2023/Q2924437) A seguinte figura ilustra um conjunto de medições de temperatura de uma determinada localidade conforme a técnica box plot, apresentando uma escala em graus Celsius. Não há ocorrências de outliers.

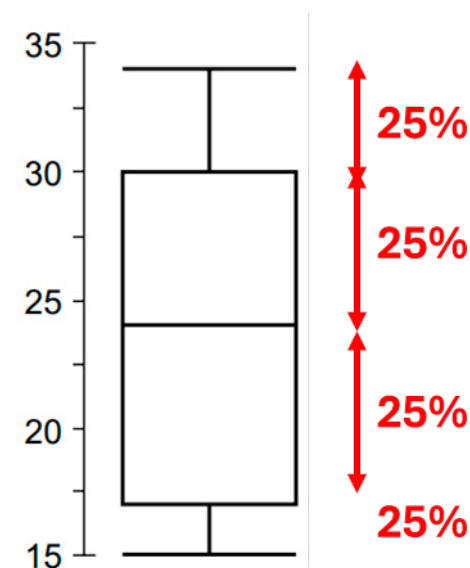


É possível afirmar que

- a) a temperatura máxima é igual a 35 °C.
- b) pelo menos 75% das medições são iguais ou inferiores a 30 °C.
- c) mais que 50% das medições são maiores que 25 °C.
- d) a temperatura máxima é igual a 30 °C.
- e) até 75% das medições são iguais ou superiores a 15 °C.



Observe que o boxplot divide as observações em quartis. Abaixo do Q1, temos as 25% menores; entre o Q1 e o Q2, temos as 25% intermediárias; e, assim, por diante, como mostrado a seguir.



- a) **Errada.** A temperatura máxima é menor que 35 °C, pois o topo do diagrama boxplot está ligeiramente abaixo dessa temperatura.
- b) **Certa** – pelo menos 75% das medições são iguais ou inferiores a 30 °C, pois essa temperatura é o terceiro quartil (Q3).

$$Q_1 + Q_2 + Q_3 = 75\%$$

- c) **Errada.** **Menos** que 50% das medições são maiores que 25 °C.
- d) **Errada.** A temperatura máxima é **superior** a 30 °C.
- e) **Errada.** **Mais de** 75% das medições são iguais ou superiores a 15 °C.

Letra b.

025. (IBFC/TRF 5/TÉCNICO JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: TECNOLOGIA DA INFORMAÇÃO/2024/Q3170134) Com base nos dados apresentados na Tabela 1 – Distribuição das idades (em anos) dos sócios do Clube Juvenil de Recreação e Bem-Estar, assinale a alternativa incorreta.

Tabela 1 – Distribuição das idades (em anos) dos sócios do Clube Juvenil de Recreação e Bem-Estar.

Idade (em anos)	Número de Sócios
10 – 12	92
12 – 14	107
14 – 16	231
16 – 18	189
Total	619

- a) Um jovem com 12 anos completos é alocado na primeira classe, 10 |– 12.
- b) A mediana assume um valor dentro da terceira classe, 14 |– 16
- c) Por ser a classe de maior frequência, a terceira classe, 14 |– 16, contém o valor da moda.
- d) A idade média pode ser calculada usando o ponto médio de cada classe, uma vez que não temos a informação dos valores observados.



a) Errada. Achamos o erro logo na primeira alternativa, pois o valor de 12 anos completos pertence à 2ª classe. Na 1ª classe o intervalo 10 |– 12 significa que o intervalo inclui os jovens com menos de 12, e não com exatamente 12 anos completos.

b) Certa. Como os dados foram fornecidos na forma de dados categorizados, podemos calcular a mediana pela posição:

$$\frac{619}{2} = 309,5$$

Para saber em que classe se encontra esse termo, precisamos calcular as frequências acumuladas nas classes.

Idade (em anos)	Número de Sócios	Frequência Acumulada
10 – 12	92	x_1 a x_{92}
12 – 14	107	x_{93} a x_{199}
14 – 16	231	x_{200} a x_{430}
16 – 18	189	x_{431} a x_{619}
Total	619	

De fato, a classe mediana é a classe 14 |– 16.

c) Certa. A moda deve estar na classe modal, que é aquela de maior frequência relativa.

d) Certa. Corretíssimo! A média é calculada pela técnica do ponto médio.

Letra a.

026. (FGV/TJDFT/2022/ANALISTA JUDICIÁRIO/ESTATÍSTICA) O resultado de uma pesquisa sobre a produtividade dos magistrados em uma determinada região foi publicado em uma revista científica e está sintetizado na tabela a seguir.

Classes de produtividade	Frequência relativa acumulada
0 --- 2	0,25
2 --- 4	0,65
4 --- 6	0,85
6 --- 9	1

Não existem observações coincidentes com os extremos das classes.

É sabido que, quanto maior a classe de produtividade, maior é a produtividade do magistrado.

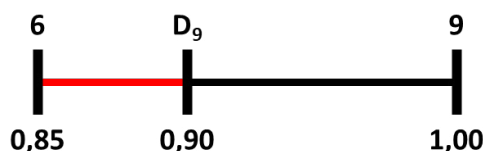
Um estatístico precisa estimar a produtividade a partir da qual se encontram os 10% mais eficientes, isto é, o 9º decil dessa distribuição.

A melhor estimativa é, aproximadamente:

- a) 6,5.
- b) 7.
- c) 7,5.
- d) 8.
- e) 8,5.



Para calcular o nono decil, devemos recorrer à técnica da interpolação linear.



Vamos usar a proporcionalidade entre a parte e o todo.

$$\frac{D_9 - 6}{0,90 - 0,85} = \frac{9 - 6}{1 - 0,85}$$

$$\frac{D_9 - 6}{0,05} = \frac{3}{0,15}$$

Vamos fazer o meio pelos extremos.

$$\therefore D_9 - 6 = \frac{3 \cdot 0,05}{0,15} = 1$$

Então, podemos passar o 6 para o outro lado.

$$\therefore D_9 = 6 + 1 = 7$$

Letra b.

027. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte, com alguns valores não-identificados, mostra os resultados de uma inspeção visual no campo, relativos ao estado de conservação de 200 centros históricos de determinada região.

categoria	frequência	percentual (%)
ruim	50	C
regular	A	10
bom	100	D
excelente	B	E

Acerca dessa tabela, julgue o item subsequente.

Na tabela, a letra C corresponde a 20%.



Como vimos, a letra C corresponde ao percentual ocupado pelos centros de categoria ruim. Para obtê-lo, basta dividir o número de centros nessa categoria pelo total de centros na pesquisa e multiplicar por 100%.

$$C = \frac{50}{200} \cdot 100\% = 0,25 \cdot 100\% = 25\%$$

Podemos, ainda, completar a tabela.

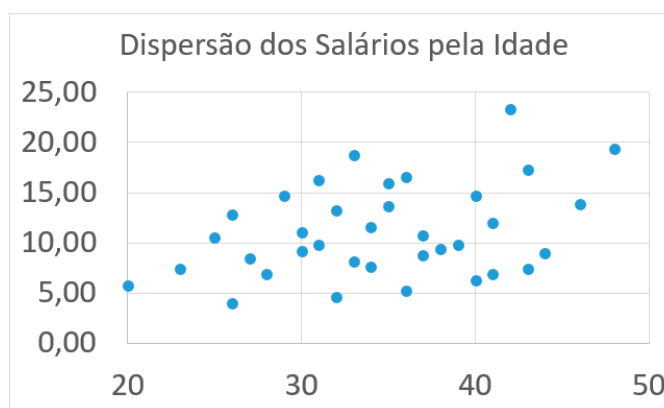
Categoria	Frequência	Percentual
Ruim	50	25%
Regular	20	10%
Bom	100	50%
Excelente	30	15%
Total	200	

Errado.

028. (CESPE/ABIN/2018/OFICIAL TÉCNICO DE INTELIGÊNCIA/ÁREA 4) O diagrama de dispersão é adequado para se descrever o comportamento conjunto de duas variáveis quantitativas. Cada ponto do gráfico representa um par de valores observados.



Vejam um exemplo de gráfico de dispersão.



Nesse gráfico, podemos ver em um mesmo ponto informações sobre a idade e o salário das pessoas. Portanto, cada ponto realmente representa um par de observações.

Certo.

029. (INSTITUTO VERBENA/MPE AC/ANALISTA MINISTERIAL/ÁREA ESTATÍSTICA/2023/Q3090436) Em um gráfico do tipo box-plot são representados: a mediana (Md), o 1º quartil (Q_1), o 3º quartil (Q_3) e os possíveis pontos atípicos (outliers). O intervalo interquartil (IQR) é a diferença entre Q_3 e Q_1 . Para um valor x ser considerado atípico em um gráfico box-plot, ele deve satisfazer

- a) $x < Q_1 - IQR$ ou $x > Q_3 + IQR$
- b) $x > Q_1 - IQR$ ou $x < Q_3 + IQR$
- c) $x < Q_1 - 1,5 \cdot IQR$ ou $x > Q_3 + 1,5 \cdot IQR$
- d) $x > Q_1 - 1,5 \cdot IQR$ ou $x < Q_3 + 1,5 \cdot IQR$



Lembramos que valor atípico(outlier) são valores que extrapolam o intervalo, ou seja, que passam de ou inferior a. Temos então:

Valor Atípico $x < Q_1 - 1,5 \cdot IQR$ ou $Q_3 - 1,5 \cdot IQR$

Letra c.

030. (FGV/TJ SE/ANALISTA JUDICIÁRIO/ÁREA ESTATÍSTICA/2023/Q3033108) O boxplot é um recurso gráfico usado para analisar a distribuição dos dados e visualizar a presença de valores discrepantes (outlier) com base nas medidas estatísticas descritivas, como mínimo, máximo, primeiro quartil, segundo quartil ou mediana e terceiro quartil.

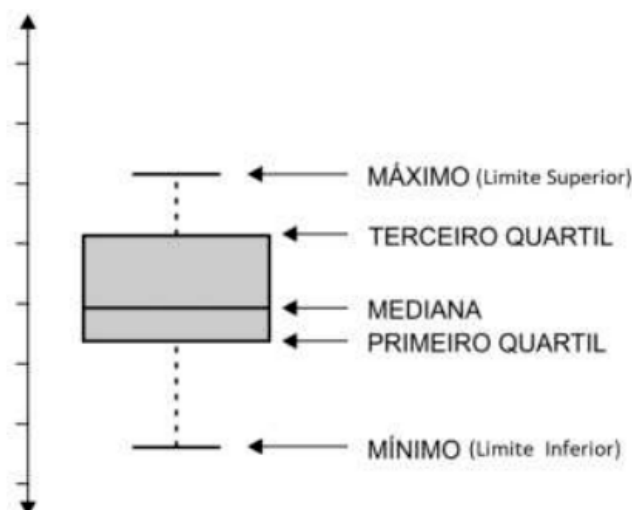
Nesse diagrama, o Intervalo Interquartil (IQR) é utilizado para calcular os limites superior e inferior, como:

IQR = (Terceiro Quartil – Primeiro Quartil)

Limite Superior = Terceiro Quartil + $1,5 \cdot IQR$

Limite Inferior = Primeiro Quartil – $1,5 \cdot IQR$

Observe o *boxplot* a seguir.



Dado um conjunto de valores, um valor é considerado outlier se estiver:

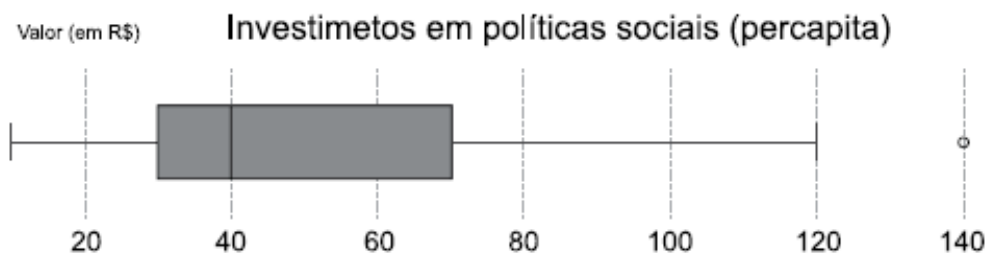
- a) entre o terceiro quartil e a mediana.
- b) entre a mediana e o primeiro quartil.
- c) entre o primeiro quartil e o limite inferior.
- d) entre o terceiro quartil e o limite superior.
- e) acima do limite superior ou abaixo do limite inferior.



Um *outlier* é um valor que destoa do conjunto de dados, ou seja, um valor que ultrapassa o limite superior = $Q_3 - 1,5IQR$ ou abaixo do limite inferior = $Q_1 - 1,5IQR$.

Letra e.

031. (VUNESP/PREFEITURA DE SÃO PAULO/ANALISTA DE POLÍTICAS PÚBLICAS/ÁREA GESTÃO GOVERNAMENTAL/2023/Q3059735) Analise o box plot a seguir, que traz informações sobre o valor de investimentos per capita em política social durante um mês em um determinado município.



Em relação ao gráfico apresentado, é possível afirmar que

- a) 25% das pessoas analisadas estão entre os valores de R\$40,00 e R\$60,00.
- b) não há outlier no gráfico analisado.
- c) a moda deste conjunto é de R\$40,00.
- d) a mediana do conjunto de dados apresentado é de R\$40,00.
- e) o primeiro quartil possui o valor de R\$40.



a) Errada. A mediana é o valor de R\$40,00, porém, o terceiro quartil (Q_3) é um valor maior que R\$60,00. Poderíamos dizer que 25% das pessoas estão entre R\$40,00 e o terceiro quartil, mas, como R\$60,00 está abaixo do terceiro quartil, podemos concluir que menos de 25% das pessoas estão na faixa de R\$40,00 a R\$60,00.

b) Errada. Existe um outlier em 140.

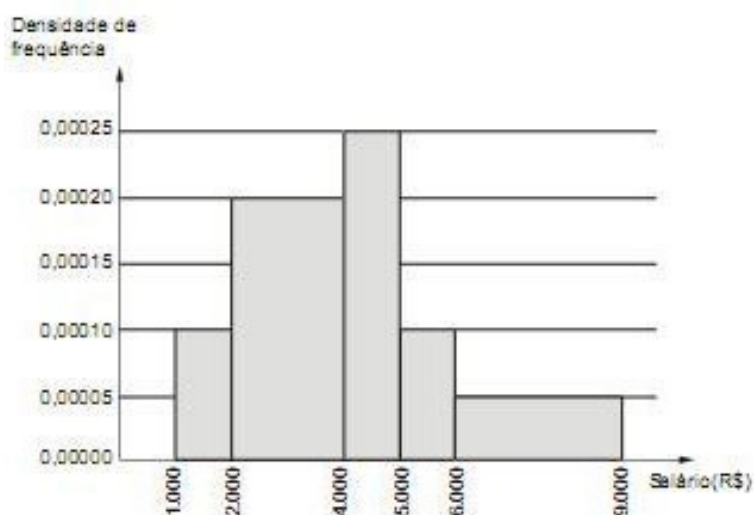
No diagrama boxplot, não é possível visualizar a moda, caso ela exista. O valor de R\$40,00 é a mediana. Afirmação incorreta.

c) **Certa**. A mediana do conjunto de dados apresentado é de R\$40,00, pois esse valor está na barra do diagrama.

d) **Errada**. O primeiro quartil é o extremo inferior da caixa e possui um valor menor que R\$40.

Letra c.

032. (FCC/TRF (2ª REGIÃO)/2012/ANALISTA JUDICIÁRIO) Considere que a distribuição dos salários dos funcionários em um setor público está representada por um histograma conforme abaixo, em que no eixo vertical constam as densidades de frequências, em (R\$)⁻¹. Densidade de frequência de um intervalo de classe é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.

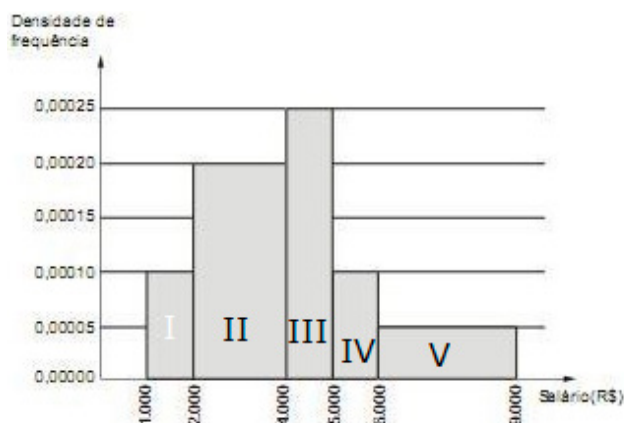


Considerando que todos os intervalos classe são fechados à esquerda e abertos à direita, a porcentagem P dos funcionários que ganham no mínimo R\$ 2.000,00 e menos que R\$ 6.000,00 é tal que

- a) $P < 65\%$.
- b) $65\% < P < 70\%$.
- c) $70\% < P < 75\%$.
- d) $75\% < P < 80\%$.
- e) $P > 80\%$.



Tem-se um clássico gráfico de densidade de frequências. Os funcionários que foram pedidos correspondem à soma das regiões II, III e IV.



Agora, vamos calcular as probabilidades associadas a cada região. Devemos nos lembrar que a frequência relativa é igual ao produto da densidade de frequência pela amplitude da classe.

$$p_{II} = d_{II}A_{II} = 0,00020 \cdot (4000 - 2000) = 0,0002 \cdot 2000 = 0,40$$

$$p_{III} = d_{III}A_{III} = 0,00025 \cdot (5000 - 4000) = 0,00025 \cdot 1000 = 0,25$$

$$p_{IV} = d_{IV}A_{IV} = 0,00010 \cdot (6000 - 5000) = 0,00010 \cdot 1000 = 0,10$$

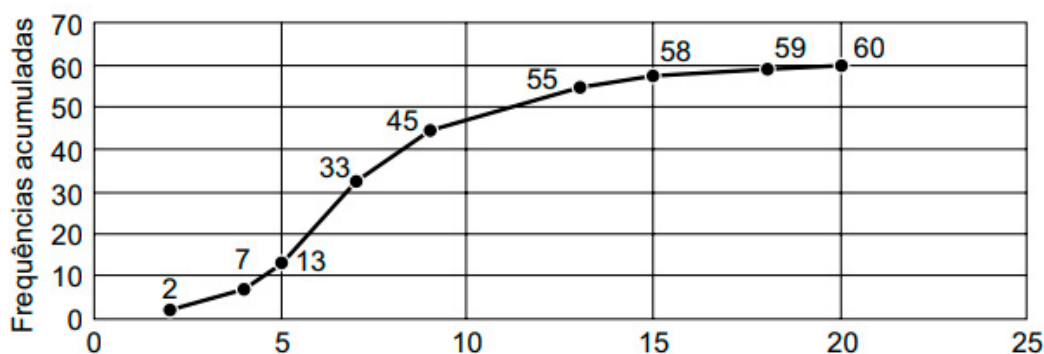
$$\therefore P = p_{II} + p_{III} + p_{IV} = 0,40 + 0,25 + 0,10 = 0,75 = 75\%$$

Letra c.

033. (VUNESP/TRF 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE/ ESTATÍSTICA/2023/Q3160772) O gráfico a seguir mostra os valores da frequência acumulada da distribuição de frequências sem perda de informação de certa variável x.

GRÁFICO 1

Frequências acumuladas da variável x



Variável x

Do gráfico, podemos inferir que a frequência relativa de $x = 5$ é:

- a) 7,3%;
- b) 12%;
- c) 15,8%.
- d) 10%.
- e) 15%.



Precisamos encontrar as frequências absolutas (F_i).

No primeiro valor assumido, a frequência acumulada é a própria frequência absoluta, ou seja, $F_i = 2$

Já o segundo valor acumulado, 7 é a soma da primeira frequência absoluta com a segunda, totalizando 7. Assim:

$$2 + F_i = 7 \therefore F_i = 5$$

Analogamente, para cada uma das frequências acumuladas do gráfico temos:

Fac	F_i
2	2(repete)
7	$(7-2) = 5$
13	$(13-7) = 6$
33	$(33-13) = 20$
45	$(45-33) = 12$
55	$(55-45) = 10$
58	$(58-55) = 3$
59	$(59-58) = 1$
60	$(60-59) = 1$

Totalizando $n = 60$.

Para $x=5$, a frequência absoluta $f_i = 6$, e do total $n = 60$

$$F_i = \frac{6}{60} \times 100\% = 10\%$$

Letra d.

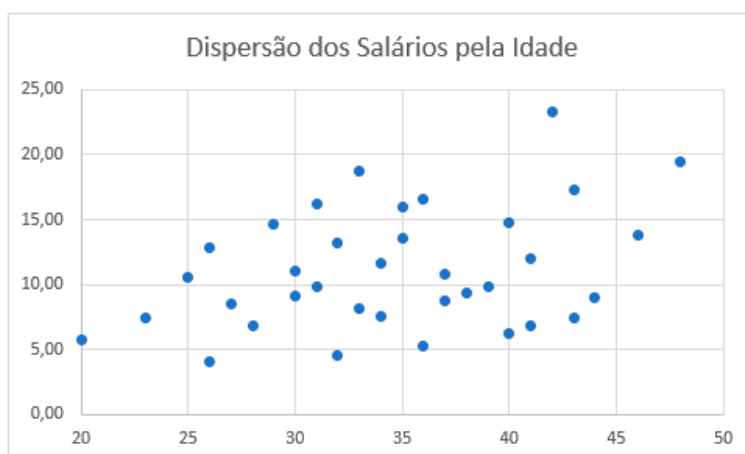
034. (FGV/TJ/SE/2023/ANALISTA JUDICIÁRIO/ESTATÍSTICA) A análise dos dados multivariados, como a correlação linear entre dois atributos do conjunto de dados X, pode ser facilitada pelo uso do gráfico de:

- a) pizza.
- b) boxplot.

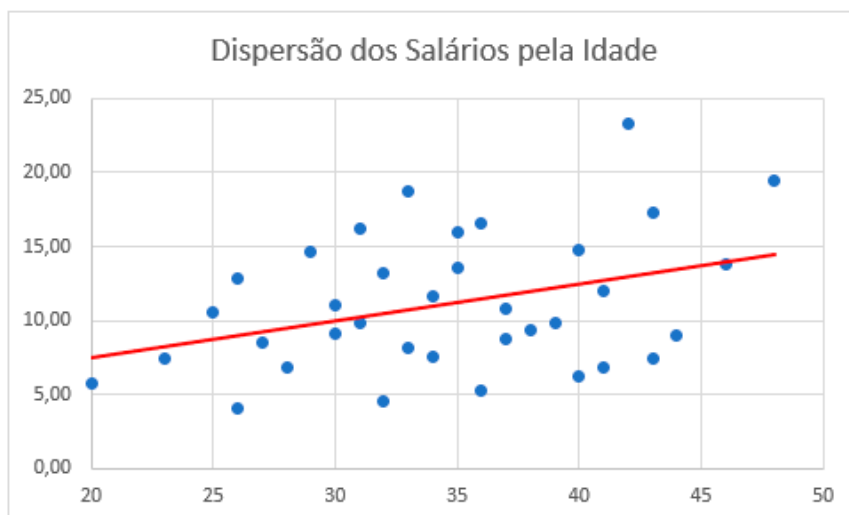
- c) densidade.
- d) histograma.
- e) scatter plot.



O diagrama de dispersão (ou o “scatter plot”) serve para visualizar todas as observações de duas variáveis aleatórias ao mesmo tempo. Esse diagrama é muito utilizado para verificar relações de causa e efeito. Como exemplo, veja a dispersão dos salários pela idade.



Nesse exemplo, é possível construir uma linha de tendência que facilita visualizar a existência de correlação positiva ou negativa entre as variáveis.



Letra e.

035. (CEBRASPE/CGDF/2023/AUDITOR DE CONTROLE INTERNO) Considerando que a tabela precedente represente a distribuição de frequências absolutas de uma variável quantitativa x , é correto afirmar que, em relação a x , a média será

x	frequência absoluta
0	1
1	2
2	4
3	0
4	2
total	9

- a) igual à mediana e igual à moda.
b) igual à mediana, que é menor que a moda.
c) menor que a mediana, que é menor que a moda.
d) maior que a mediana, que é maior que a moda.



A média pode ser calculada usando as frequências absolutas como pesos:

$$\mu = \frac{1 \cdot 0 + 2 \cdot 1 + 4 \cdot 2 + 0 \cdot 3 + 2 \cdot 4}{1 + 2 + 4 + 0 + 2} = \frac{0 + 2 + 8 + 0 + 8}{9} = \frac{18}{9} = 2$$

Por outro lado, mediana pode ser obtida pelo termo:

$$\frac{1 + 9}{2} = \frac{10}{2} = 5 \rightarrow Md = x_5$$

Para obter o quinto termo, precisamos observar as frequências acumuladas em cada classe.

x	Frequência Absoluta	Frequência Acumulada
0	1	x_1
1	2	x_2 a x_3
2	4	x_4 a x_7
3	0	
4	2	x_8 a x_9

Portanto, a mediana é $Md = x_5 = 2$.

A moda, por sua vez, corresponde ao valor de "x" observado com mais frequência, que, no caso, é $x = 2$, cuja frequência absoluta é igual a 4.

Letra a.

036. (INSTITUTO CONSULPLAN/MPE BA/ANALISTA TÉCNICO/ÁREA ANALISTA DE SISTEMAS/2023/Q2765770) A tabela representa a distribuição das notas de desempenho dos candidatos em um processo de seleção para uma vaga de emprego. A última fase do

processo de seleção consiste em uma entrevista, que será realizada apenas com os candidatos que obtiveram desempenho acima da nota de corte, calculada com base no valor do sexto decil, cujo valor é 51,8:

Distribuição das Notas dos Candidatos					
NOTAS		F_i	F_{ac}	f_i	f_{ac}
0 — 20		5	5	0,07	0,07
20 — 40		24	29	0,32	0,39
40 — 60		27	56	0,36	0,75
60 — 80		14	70	0,18	0,93
80 — 100		5	75	0,07	1,00
TOTAL		75		1,00	

Legenda:
 F_i – frequência absoluta simples
 F_{ac} – frequência absoluta acumulada
 f_i – frequência relativa simples
 f_{ac} – frequência relativa acumulada

“De acordo com os dados apresentados, é correto afirmar que 60% dos candidatados obtiveram notas _____ de 51,8; portanto, a empresa realizará _____ 40 entrevistas.” Assinale a alternativa que completa correta e sequencialmente a afirmativa anterior.

- a) acima / mais.
- b) abaixo / mais.
- c) acima / menos.
- d) abaixo / menos.
- e) acima / exatamente.



Decil é uma medida estatística utilizada para dividir uma amostra em 10 partes iguais, cada uma representando 10% dos dados. Isso significa que cada decil contém uma décima parte dos dados da amostra.

Os decis são calculados ordenando os dados da amostra do menor para o maior e, em seguida, dividindo-os em 10 grupos iguais. O primeiro decil é o valor abaixo do qual 10% dos dados da amostra estão localizados, enquanto o décimo decil é o valor abaixo do qual 90% dos dados estão localizados.

Na questão se fala sobre que o 6º decil é igual a 51,8, ou seja:

De acordo com os dados apresentados, é correto afirmar que 60% dos candidatados obtiveram notas **abaixo** de 51,8.

Além disso, uma vez que 75 candidatos foram considerados, então o 6º decil ocupa a posição:

$$60\% \cdot 75 = 45$$

Portanto, 45 candidatos foram eliminados, de modo que 30 realizarão a entrevista.

Portanto a empresa realizará **menos** de 40 entrevistas.

Letra d.

037. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) A tabela seguinte, com alguns valores não-identificados, mostra os resultados de uma inspeção visual no campo, relativos ao estado de conservação de 200 centros históricos de determinada região.

categoria	frequência	percentual (%)
ruim	50	C
regular	A	10
bom	100	D
excelente	B	E

Acerca dessa tabela, julgue o item subsequente.

A letra B, na tabela, representa 25 centros.



Notemos que os centros com frequência regular correspondem a 10% do total. Portanto, a frequência absoluta correspondente é:

$$A = 0,10 \cdot 200 = 20$$

Por fim, devemos impor que a soma das frequências absolutas referentes a cada avaliação deve ser igual ao total de 200 centros.

$$50 + A + 100 + B = 150 + A + B = 200$$

$$150 + 20 + B = 200$$

$$170 + B = 200$$

$$\therefore B = 200 - 170 = 30$$

Portanto, a letra B representa 30 centros. Afirmação errada.

Errado.

038. (CESPE/MPU/2013) A tabela acima mostra algumas estatísticas descritivas produzidas por um estudo acerca da quantidade de acidentes de trabalho (N), ocorridos em 2012, a

partir de uma amostra aleatória simples de 200 indústrias de pequeno porte. Com base nessas informações, julgue o próximo item.

estatística	quantidade de acidentes (N)
mínimo	0
primeiro quartil	2
segundo quartil	4
terceiro quartil	10
máximo	30

O intervalo interquartílico da variável N foi igual a 8.



O intervalo interquartílico corresponde à diferença entre o terceiro e o primeiro quartis.

$$D = Q_3 - Q_1 = 10 - 2 = 8$$

Certo.

039. (CESPE/IPHAN/2018/ANALISTA I/ÁREA 2) 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 ($Q_{1/4}$) é igual a 3.



Considerando que a amostra tem 11 cidades, o primeiro quartil deve ser calculado de forma semelhante à mediana.

$$\frac{1 + 11}{4} = \frac{12}{4} = 3$$

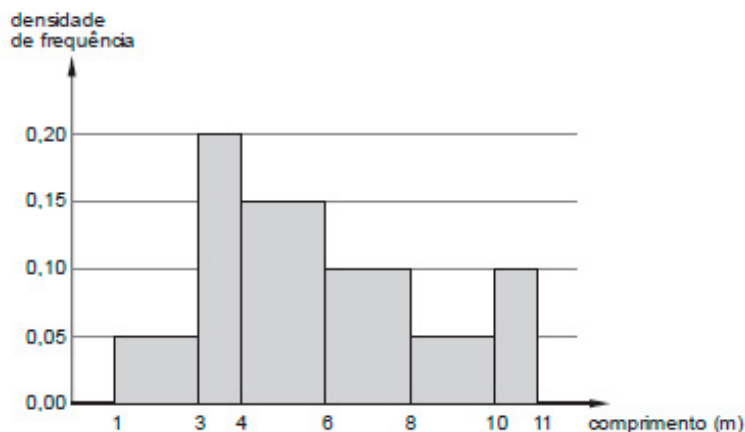
Portanto, o primeiro quartil corresponde ao elemento:

$$Q_1 = x_3 = 2$$

Errado.

040. (FCC/TRT/5ª REGIÃO (BA)/2018/ANALISTA JUDICIÁRIO/ESTATÍSTICA) A distribuição das medidas em metros (m) dos comprimentos dos cabos no estoque de uma fábrica está representada pelo histograma mostrado abaixo, em que no eixo vertical constam as densidades de frequências, em $(m)^{-1}$, e no eixo horizontal os intervalos de classe. Define-se

densidade de frequência de um intervalo de classe como sendo o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.

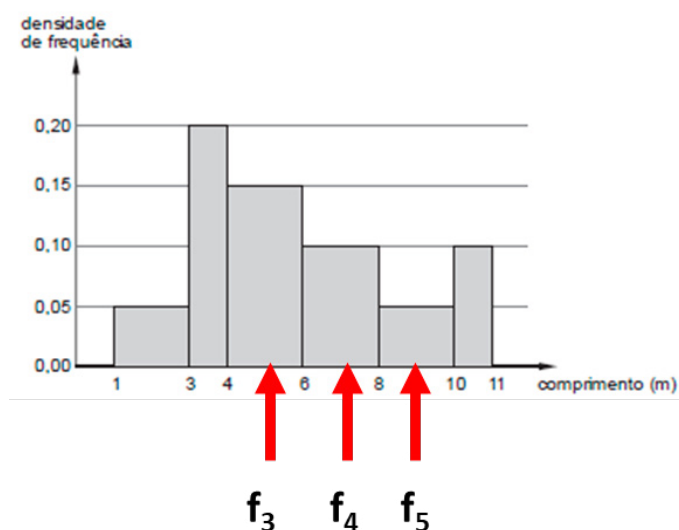


Sabendo-se que todos os intervalos de classe são fechados à esquerda e abertos à direita, então a porcentagem dos cabos que apresentam uma medida de comprimento de pelo menos igual a 4 m e inferior a 10 m é de

- a) 50%.
- b) 60%.
- c) 70%.
- d) 80%.
- e) 90%.



Vamos marcar no gráfico as classes que nos interessam.



Como o gráfico foi fornecido em densidade de frequência, podemos obter as frequências relativas de cada classe como o produto da densidade de frequência pela amplitude de classe.

$$f_3 = d_3 A_3 = 0,15 \cdot (6 - 4) = 0,15 \cdot 2 = 0,30$$

$$f_4 = d_4 A_4 = 0,10 \cdot (8 - 6) = 0,10 \cdot 2 = 0,20$$

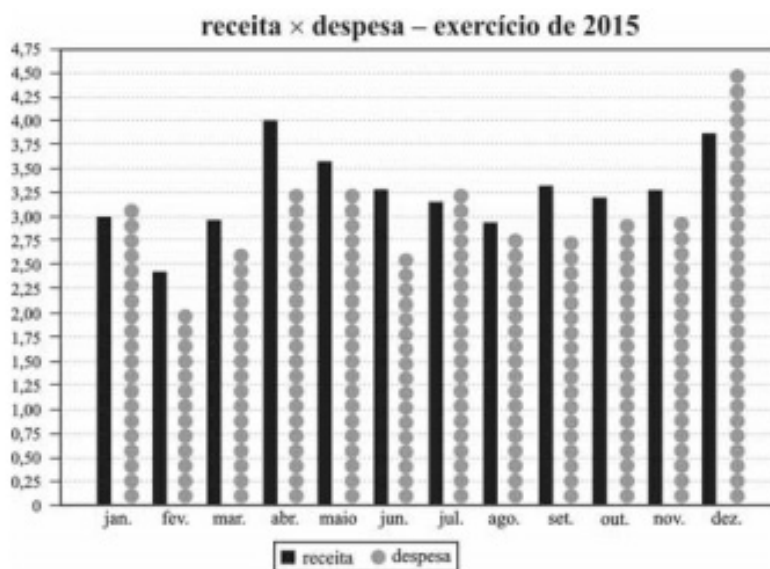
$$f_5 = d_5 A_5 = 0,05 \cdot (10 - 8) = 0,05 \cdot 2 = 0,10$$

Portanto, a frequência acumulada total das três classes desejadas é:

$$F = f_3 + f_4 + f_5 = 0,30 + 0,20 + 0,10 = 0,60$$

Letra b.

041. (CESPE/TCE/PR/2016) Tendo como referência o gráfico precedente, que mostra os valores, em bilhões de reais, relativos à arrecadação de receitas e aos gastos com despesas do estado do Paraná nos doze meses do ano de 2015, assinale a opção correta.



Internet: <www.gestaoedinheiropublico.pr.gov.br> (com adaptações).

- a) No ano considerado, o segundo trimestre caracterizou-se por uma queda contínua na arrecadação de receitas, situação que se repetiu no trimestre seguinte.
- b) No primeiro quadrimestre de 2015, houve um período de queda simultânea dos gastos com despesas e da arrecadação de receitas e dois períodos de aumento simultâneo de gastos e de arrecadação.
- c) No último bimestre do ano de 2015, foram registrados tanto o maior gasto com despesas quanto a maior arrecadação de receitas.
- d) No ano em questão, janeiro e dezembro foram os únicos meses em que a arrecadação de receitas foi ultrapassada por gastos com despesas.

e) A menor arrecadação mensal de receitas e o menor gasto mensal com despesas foram verificados, respectivamente, no primeiro e no segundo semestre do ano de 2015.



Vamos analisar as afirmações.

a) Errada. O segundo trimestre é composto pelos meses abril, maio e junho, enquanto o terceiro trimestre é composto pelos meses julho, agosto e setembro. Dessa forma, podemos desenhar as ogivas referentes à arrecadação dos dois trimestres.



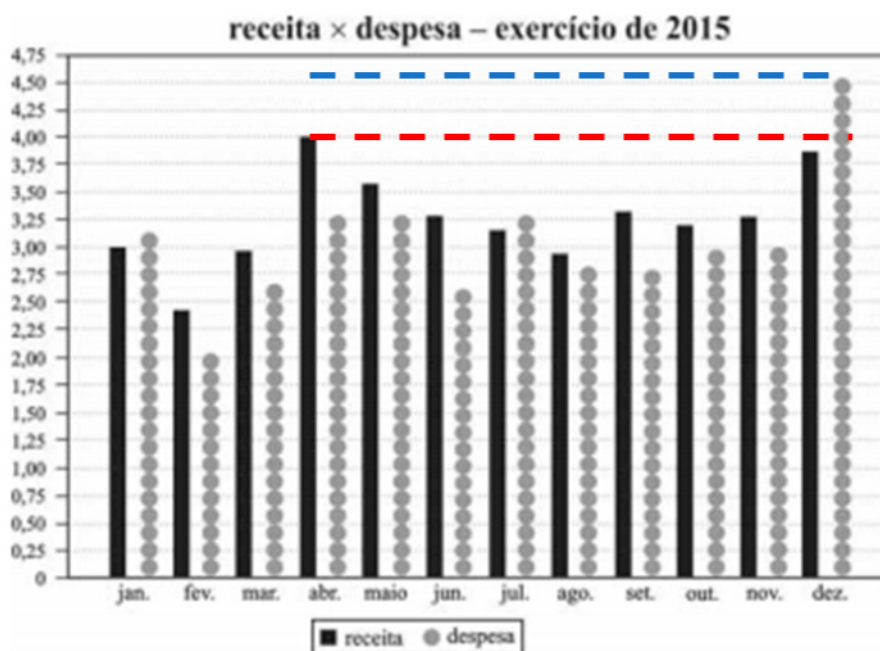
Podemos observar, portanto, que no terceiro trimestre não houve queda contínua na arrecadação de receitas, porque, no mês de agosto, houve aumento em relação ao mês anterior.

b) Certa. O primeiro quadrimestre é composto pelos meses de janeiro a abril. Vamos desenhar as ogivas referentes à arrecadação de receitas e de despesas.



Portanto, vemos um período de redução de janeiro para fevereiro e dois períodos de crescimento tanto nas receitas como nas despesas.

c) Errada. Podemos observar que o auge das despesas aconteceu realmente em dezembro de 2015. Porém, o maior valor da arrecadação de receitas aconteceu em abril de 2015, que não se localiza no último bimestre do ano.



d) Errada. Essa situação aconteceu também no mês de julho.



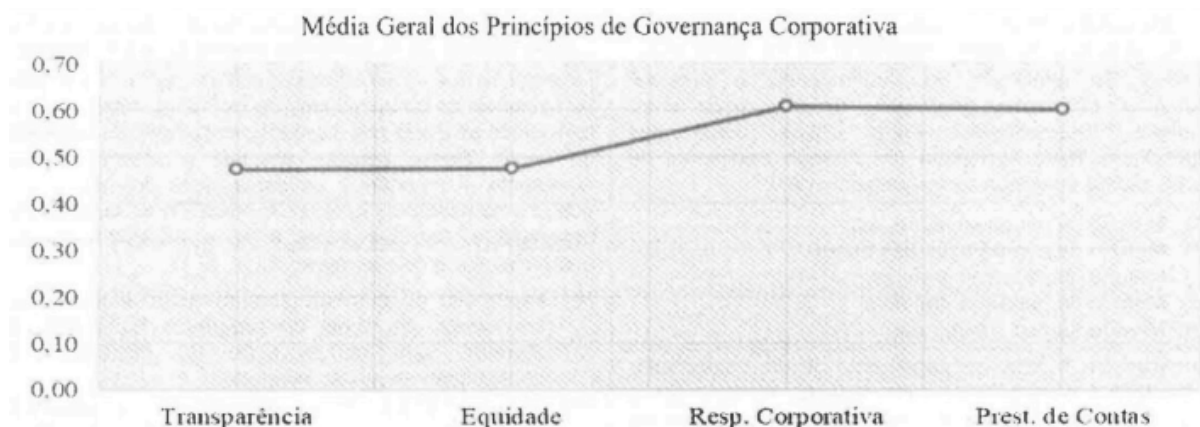
e) Errada. Vamos identificar o mínimo da receita mensal e da despesa mensal.



Portanto, os valores mínimos tanto da receita como da despesa aconteceram no mês de fevereiro, que pertence ao primeiro semestre.

Letra b.

042. (LEGALLE CONCURSOS/DPE PA/ANALISTA/ÁREA: ADMINISTRAÇÃO/2023/Q3011129)
Analisar o gráfico a seguir:



Da análise do gráfico, verifica-se que o princípio de Responsabilidade Corporativa (Resp. Corporativa) é aquele que rege a evidenciação de atributos de governança corporativa e apresenta a média:

- a) 0,47.
- b) 0,70.
- c) 0,61.

d) 0,35.

e) 0,48.



Observe que o princípio de responsabilidade corporativo está no terceiro ponto do gráfico, e tem o valor ligeiramente acima de 0,60.

Desse modo, o único valor compatível entre as alternativas propostas pelo enunciado é 0,61.

Letra c.

A respeito do conjunto de dados {11, 6, 28, 51, 49, 32, 33}, julgue os itens a seguir.

043. (CEBRASPE BANESE 2021 TÉCNICO BANCÁRIO I) O primeiro quartil do conjunto de dados em tela é igual ou superior a 33.



Para o primeiro quartil, devemos utilizar a posição:

$$\frac{n + 1}{4} = \frac{7 + 1}{4} = \frac{8}{4} = 2$$

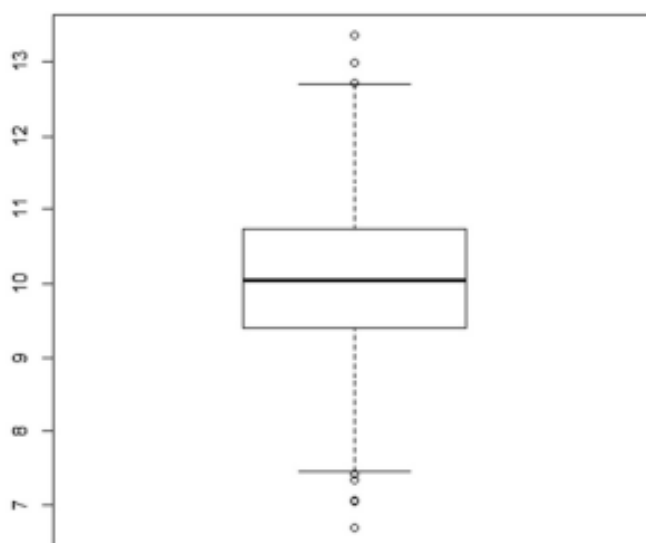
Assim, vamos procurar o elemento x_2 dessa sequência por ordem crescente.

11	6	28	51	49	32	33
x2	x1	x3	x7	x6	x4	x5

Portanto, o primeiro quartil é:

$$Q_1 = x_2 = 11$$

Errado.



Com relação aos dados que resultaram no diagrama mostrado na figura precedente, julgue os itens a seguir.

044. (CEBRASPE/PETROBRÁS/2021) A amplitude total dos dados em tela é inferior a 6.



Observe que existe pelo menos uma observação maior que 13 e uma observação menor que 7. Então, a amplitude, que corresponde à diferença entre a maior e a menor observação da amostra, é

$$Amp > 13 - 7 = 6$$

Portanto, a amplitude é maior que 6.

Errado.

045. (CEBRASPE/PETROBRÁS/2021) Nesse diagrama, a porção da distribuição dos dados representada pela parte inferior do diagrama mostrada a seguir representa exatamente 25% dos dados em questão.



Abaixo da caixa do box plot, encontram-se 25% dos dados. Portanto, na região citada, encontram-se 25% dos dados menos os 4 outliers.

Errado.

046. (CEBRASPE/PETROBRÁS/2021) O terceiro quartil é inferior a 11 e superior a 10.



O terceiro quartil é o limite superior da caixa, que realmente se encontra entre 10 e 11.

Certo.

047. (VUNESP/TRF 3/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2023/Q3160731) Para as questões de números 21 a 25, considere a seguinte tabela de distribuição de frequência com perda de informação relativa ao consumo mensal de água, em m³, do quarteirão Q da cidade C, obtida de uma amostra aleatória de residências do local.

TABELA 1

Consumo mensal de água do quarteirão Q da cidade C

Consumo em m³	Número de residências (f_i)
10 ----20	4
20 ----30	6
30 ----40	10
40 ----50	12
50 ----60	8

A estimativa de residências que consomem entre o valor do limite inferior da 2ª classe e o valor do limite superior da 4ª classe é:

- a) 84%.
- b) 80%.
- c) 65%.
- d) 70%.
- e) 74%.



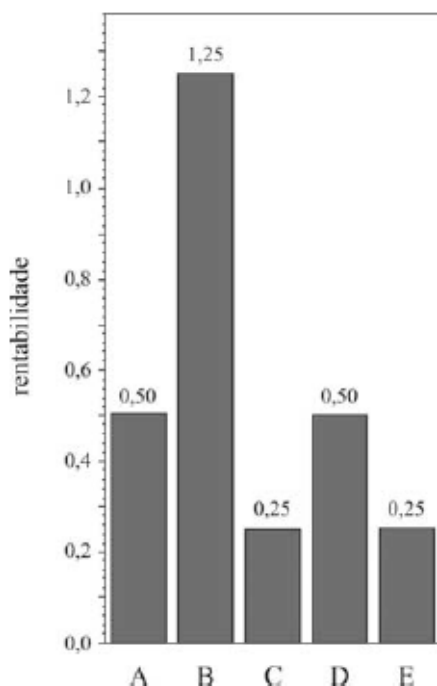
A questão nos pede a frequência relativa ou F_i (porcentagem%) da 2ª, 3ª e 4ª classe juntas em relação ao total.

A estimativa ser a soma das F_i das 3 classes dividido pelo total de residências (40), temos:

$$F_i = \frac{6 + 10 + 12}{40} = 0,7 = 70\%$$

Letra d.

048. (CESPE/FUNPRESF/EXE/2016/ANALISTA/ÁREA DE INVESTIMENTOS) O gráfico ilustra cinco possibilidades de fundos de investimento com suas respectivas rentabilidades. Considerando que as probabilidades de investimento para os fundos A, B, C e D sejam, respectivamente, $P(A) = 0,182$; $P(B) = 0,454$; $P(C) = 0,091$; e $P(D) = 0,182$, julgue o item subsequente.



O gráfico apresentado é um histograma.



O histograma se refere a uma distribuição de frequências de uma variável quantitativa contínua. Esse gráfico não mostra uma distribuição de frequências, mas sim a rentabilidade dos fundos.

Portanto, esse não pode ser um histograma, mas sim um gráfico de barras genérico.

Errado.

049. (COMDN/MARINHA/QUADRO TÉCNICO DO CORPO AUXILIAR DA MARINHA/ÁREA: ESTATÍSTICA/2022/Q2634582) A representação gráfica das séries estatísticas tem por finalidade representar os resultados obtidos, permitindo que se chegue a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Assinale a opção que corresponde à representação de uma série por meio de um polígono que, para ser construído, divide-se uma circunferência em tantos arcos iguais forem os dados a representar.

- a) Gráfico em colunas.
- b) Gráfico em barras.
- c) Gráfico em setores.
- d) Gráfico polar.
- e) Gráfico em curvas.



Questão conceitual sobre tipos de Gráficos

Gráficos em barras horizontais: Sua finalidade é comparar grandezas, através de retângulos de igual largura e alturas proporcionais às respectivas grandezas. No eixo das abscissas encontramos os valores da variável, e no eixo das ordenadas encontramos as frequências ou porcentagens. Cada barra representa a intensidade de uma modalidade do atributo

Gráficos em colunas ou gráficos em barras verticais: só diferem pela direção dos retângulos. Possuem as mesmas características dos gráficos em barras. Para construir este tipo de gráfico devemos primeiro traçar os eixos x e y. Em seguida colocamos, no eixo das abscissas, as categorias da variável e suas respectivas barras retangulares com altura igual a frequência ou a porcentagem da respectiva categoria.

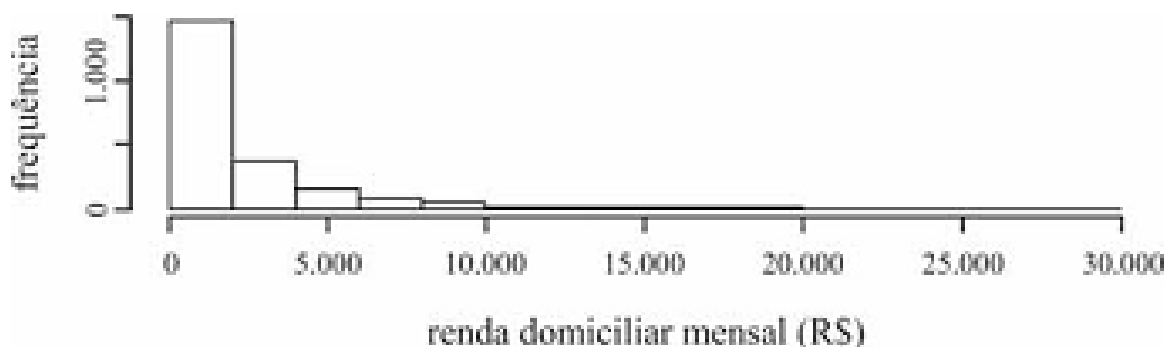
Gráficos em Setores: São usados para representar valores absolutos ou porcentagens complementares. Para construí-lo, parte-se do fato de que o número total de graus de um arco de circunferência é 360° . Cada uma das parcelas componentes do total dos valores poderá ser expressa em graus, e a correspondência se fará através de regra de três simples. A legenda pode ser dispensada inscrevendo-se no interior de cada setor a porcentagem ou a quantidade correspondente de cada um. Pode-se usar um semicírculo ou um quarto de circunferência, desde que seja usada para a representação completa do fenômeno.

Gráficos em Coordenadas Polares: É a representação de uma série por meio de um polígono. Geralmente presta-se para apresentação de séries temporais. Para construí-lo, divide-se uma circunferência em tantos arcos iguais quantos forem os dados a representar. Pelos pontos de divisas traçam-se raios. Em cada raio é representado um valor da série, marcando-se um ponto cuja distância ao centro é diretamente proporcional a esse valor; a seguir unem-se os pontos.

Letra d.

050. (CESPE/TELEBRÁS/2015/ANALISTA SUPERIOR/AUDITORIA) Uma empresa coletou e armazenou em um banco de dados diversas informações sobre seus clientes, entre as quais estavam o valor da última fatura vencida e o pagamento ou não dessa fatura. Analisando essas informações, a empresa concluiu que 15% de seus clientes estavam inadimplentes. A empresa recolheu ainda dados como a unidade da Federação (UF) e o CEP da localidade

em que estão os clientes. Do conjunto de todos os clientes, uma amostra aleatória simples constituída por 2.175 indivíduos prestou também informações sobre sua renda domiciliar mensal, o que gerou o histograma apresentado.



Com base nessas informações e no histograma, julgue o item a seguir.

Se for elaborado um histograma com classes de larguras variáveis para representar a distribuição dos valores das últimas faturas vencidas, então a classe com maior altura no histograma será, necessariamente, aquela com maior frequência no banco de dados.



Quando o histograma é elaborado com classes de larguras, ou seja, amplitudes variáveis, a altura será fornecida em densidade de frequência.

Nesse caso, a frequência relativa de cada classe será igual ao produto entre a densidade de frequência e a amplitude. Logo, a classe mais alta (maior densidade de frequência) não é necessariamente a classe que tem a maior frequência relativa. Afirmação incorreta.

Errado.

051. (FCC/TRT/3ª REGIÃO (MG)/2015/ANALISTA JUDICIÁRIO/ESTATÍSTICA) Em um histograma representando os preços unitários de microcomputadores em estoque, observa-se que no eixo das abscissas constam os intervalos de classe em R\$ e no eixo das ordenadas as respectivas densidades de frequências em (R\$) -1 . Densidade de frequência de um intervalo de classe é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo. Um determinado intervalo de classe com amplitude igual a R\$ 2.500,00 apresenta uma densidade de frequência, em (R\$) -1 , igual a $12,8 \times 10^{-5}$. Se o número de microcomputadores deste intervalo é igual a 48, então o número total de microcomputadores em estoque é igual a

- a) 150.
- b) 120.
- c) 240.

d) 160.

e) 96.



Como vimos, a frequência relativa de uma classe é igual ao produto da densidade de frequência pela amplitude da classe.

$$f = dA = 12,8 \cdot 10^{-5} \cdot 2500 = 32000 \cdot 10^{-5}$$

Multiplicar por 10^{-5} é a mesma coisa que dividir por 10^5 , que corresponde a 1 seguido por 5 zeros, ou seja, 100.000.

$$f = \frac{32000}{100000} = 0,32$$

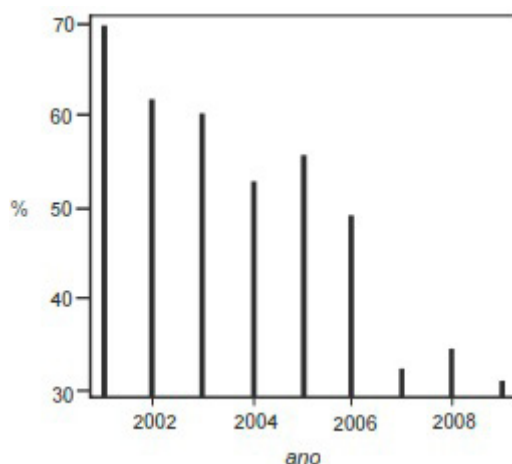
Considerando que a classe em apreço tem 48 computadores e que essa quantidade corresponde a uma frequência relativa de 0,32 ou 32%, podemos impor que 32% do total de N microcomputadores na empresa é igual a 48.

$$0,32 \cdot N = 48$$

$$\therefore N = \frac{48}{0,32} = \frac{4800}{32} = 150$$

Letra a.

052. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4) Se os percentuais forem representados por barras verticais, conforme o gráfico a seguir, então o resultado será denominado histograma.



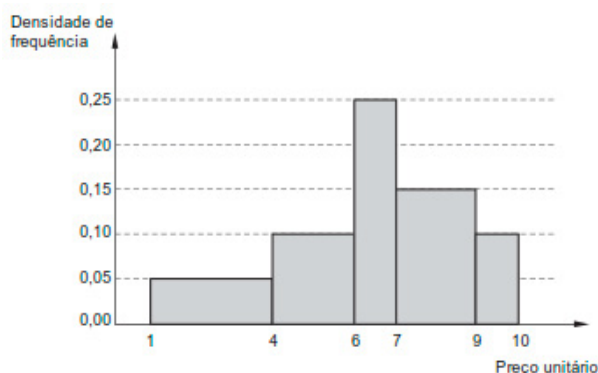


O histograma é utilizado quando ambas as variáveis aleatórias são contínuas. Nesse caso, o ano é uma variável discreta, portanto, o gráfico apresentado é um diagrama de barras.

Errado.

053. (FCC/TRE/RR/2015/ANALISTA JUDICIÁRIO) 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.

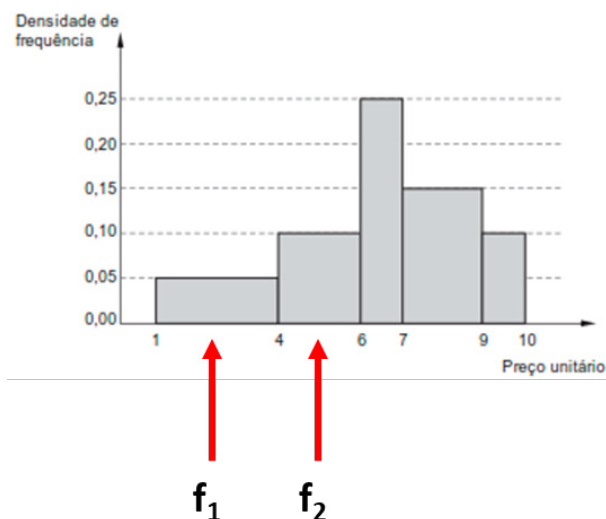


Considerando os intervalos de classe fechados à esquerda e abertos à direita, se 105 preços apresentam valores menores que R\$ 6,00, então o número de preços que apresentam valores iguais ou superiores a R\$ 4,00 é:

- a) 240.
- b) 195.
- c) 215.
- d) 230.
- e) 255.



Vamos marcar no gráfico as classes que correspondem aos preços menores que R\$6,00.



A frequência relativa de cada uma das classes pode ser obtida como o produto da densidade de frequência pela amplitude do gráfico.

$$f_1 = d_1 A_1 = 0,05 \cdot (4 - 1) = 0,05 \cdot 3 = 0,15$$

$$f_2 = d_2 A_2 = 0,10 \cdot (6 - 4) = 0,10 \cdot 2 = 0,20$$

Portanto, o percentual de valores menores que R\$6,00 corresponde à frequência acumulada pelas duas classes.

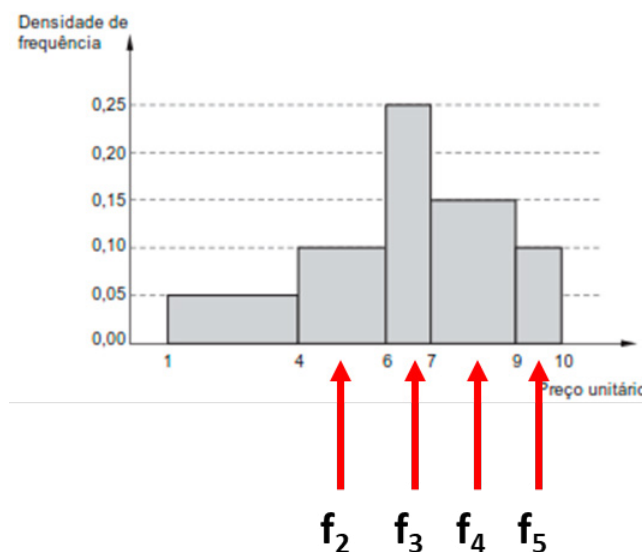
$$F_2 = f_1 + f_2 = 0,15 + 0,20 = 0,35$$

Como sabemos que são 105 preços que se enquadram nessa condição, sendo N o total de elementos, temos que 35% de N é igual a 105. Logo, podemos escrever:

$$0,35 \cdot N = 105$$

$$\therefore N = \frac{105}{0,35} = \frac{10500}{35} = 300$$

Para os valores iguais ou superiores a R\$4,00, temos as seguintes classes:



As frequências relativas de cada uma das classes podem ser calculadas como o produto da densidade de frequência pela amplitude de classe.

$$f_2 = d_2 A_2 = 0,10 \cdot (6 - 4) = 0,10 \cdot 2 = 0,20$$

$$f_3 = d_3 A_3 = 0,25 \cdot (7 - 6) = 0,25 \cdot 1 = 0,25$$

$$f_4 = d_4 A_4 = 0,15 \cdot (9 - 7) = 0,15 \cdot 2 = 0,30$$

$$f_5 = d_5 A_5 = 0,10 \cdot (10 - 9) = 0,10 \cdot 1 = 0,10$$

Dessa forma, o percentual total de preços superiores a R\$4,00 corresponde à frequência acumulada por essas quatro classes.

$$F = f_2 + f_3 + f_4 + f_5 = 0,20 + 0,25 + 0,30 + 0,10 = 0,85$$

Considerando que a amostra tinha um total de 300 elementos, concluímos que o número de preços iguais ou superiores a R\$4,00 é 85% desse total, ou seja:

$$N = 0,85 \cdot 300 = 255$$

Letra e.

054. (FCC/SEFAZ/PI/2015/AUDITOR FISCAL DA FAZENDA ESTADUAL/ADAPTADA) O histograma é um gráfico apropriado para representar dados de variáveis quantitativas contínuas.



É isso mesmo. O histograma é utilizado para estabelecer o comportamento de uma variável contínua em função de outra variável contínua.

Utilizamos o diagrama de barras para traçar o gráfico entre variáveis discretas ou quando uma delas é quantitativa.

Certo.

055. (CESPE/SEE/AL/2013/SECRETÁRIO ESCOLAR) Considere que, para classificar as escolas de uma cidade, segundo as notas médias em matemática e português, tenha sido obtida uma amostra das notas de cinquenta alunos de cada uma das dez escolas do município. Com base nessas informações, julgue o item subsequente.

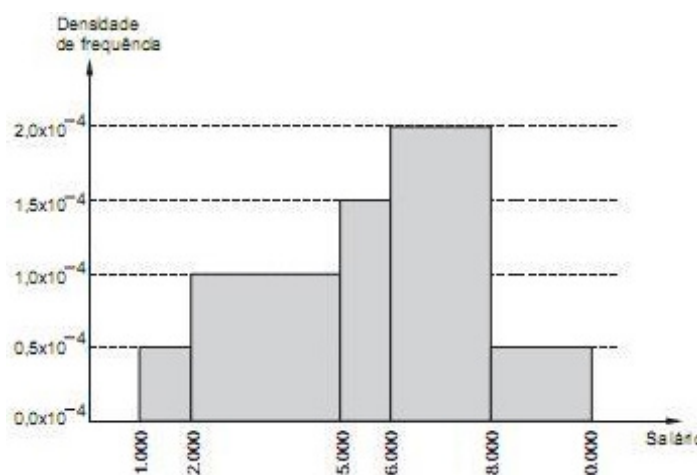
Considerando-se que a variável nota em matemática seja quantitativa contínua, então o gráfico que representa esse tipo de dado é o histograma.



Nunca é demais nos lembrar de que o histograma é utilizado para variáveis contínuas e o diagrama de barras é utilizado para variáveis discretas. Afirmação correta.

Certo.

056. (FCC/TRT/13ª REGIÃO (PB)/2014/ANALISTA JUDICIÁRIO/ESTATÍSTICA) O histograma, abaixo, refere-se à distribuição dos salários dos funcionários lotados em um setor de um órgão público. No eixo das abscissas constam os intervalos de classe em R\$ (todos fechados à esquerda e abertos à direita) e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)-1$. Define-se densidade de frequência de um intervalo como sendo o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.

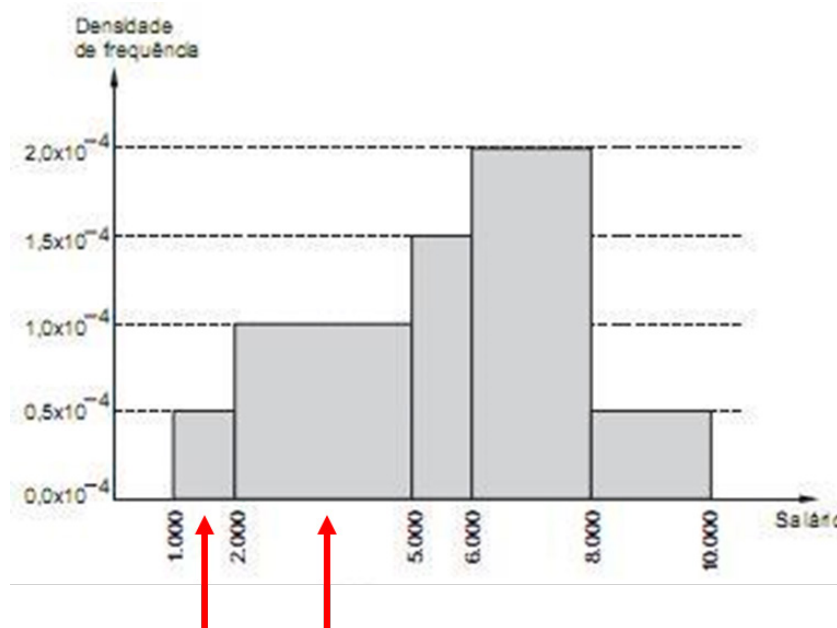


Se o número de funcionários que tem um salário inferior a R\$ 5.000,00 é igual a 56, então verifica-se que o número de funcionários que tem um salário igual ou superior a R\$ 2.000,00 e inferior a R\$ 8.000,00 é igual a

- a) 104.
- b) 152.
- c) 136.
- d) 144.
- e) 120.



A primeira informação fornecida foi o número de funcionários que possuem salário inferior a R\$5.000,00. Esses funcionários somam duas classes.



As frequências relativas de cada classe podem ser calculadas como o produto entre a densidade de frequência (altura) e a amplitude (largura) da classe.

$$f_1 = d_1 A_1 = 0,5 \cdot 10^{-4} \cdot (2000 - 1000) = 0,5 \cdot 10^{-4} \cdot 1000 = \frac{500}{10000} = 0,05$$

$$f_2 = d_2 A_2 = 1,0 \cdot 10^{-4} \cdot (5000 - 2000) = 1,0 \cdot 10^{-4} \cdot 3000 = \frac{3000}{10000} = 0,30$$

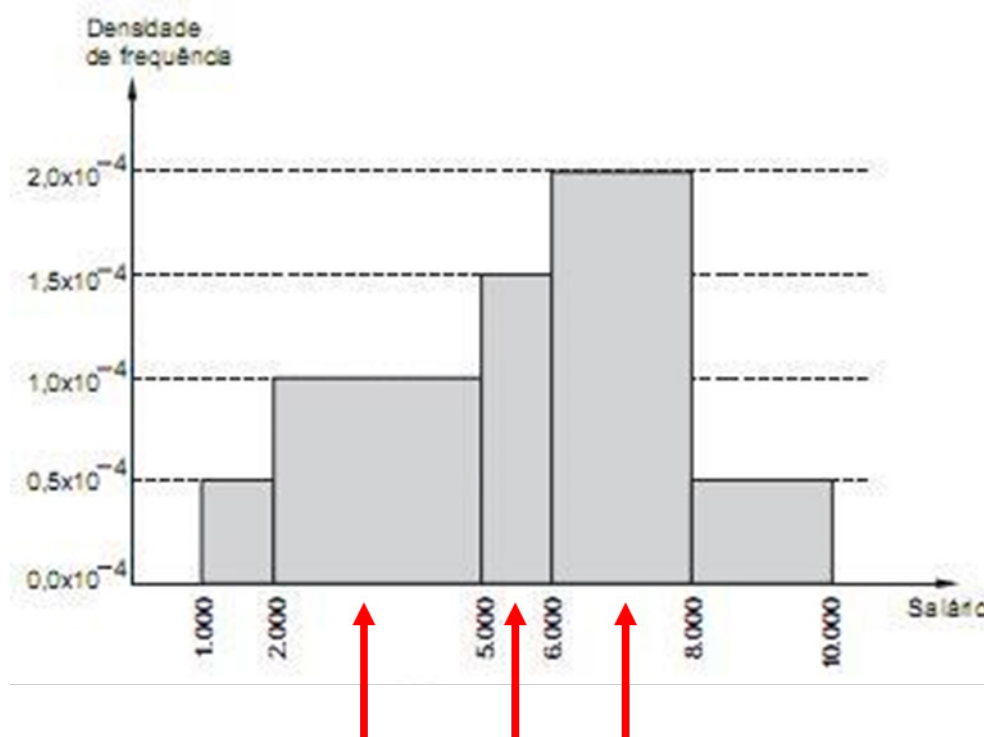
Portanto, a frequência acumulada de pessoas que ganham menos de R\$5.000,00 é igual à soma das duas classes.

$$F_2 = f_1 + f_2 = 0,05 + 0,30 = 0,35 = 35\%$$

Como sabemos que existem 56 pessoas nessa condição, sendo N o total de funcionários, temos que 35% desse total é igual a 56.

$$0,35 \cdot N = 56 \therefore N = \frac{56}{0,35} = \frac{5600}{35} = 160$$

Portanto, o total de funcionários da empresa analisados no gráfico é igual a 160. Podemos, agora, calcular o número de funcionários com o salário entre R\$2.000,00 e R\$8.000,00 como pedido. Para isso, basta tomar as seguintes classes:



As frequências relativas de cada classe são também iguais ao produto entre a densidade de frequência e a amplitude da classe.

$$f_2 = d_2 A_2 = 1,0 \cdot 10^{-4} \cdot (5000 - 2000) = 1,0 \cdot 10^{-4} \cdot 3000 = \frac{3000}{10000} = 0,30$$

$$f_3 = d_3 A_3 = 1,5 \cdot 10^{-4} \cdot (6000 - 4000) = 1,5 \cdot 10^{-4} \cdot 2000 = \frac{3000}{10000} = 0,15$$

$$f_4 = d_4 A_4 = 2,0 \cdot 10^{-4} \cdot (8000 - 6000) = 2,0 \cdot 10^{-4} \cdot 2000 = \frac{4000}{10000} = 0,40$$

Portanto, a soma das frequências associadas às três classes é:

$$F = f_2 + f_3 + f_4 = 0,30 + 0,15 + 0,40 = 0,85$$

Por fim, o total de funcionários que se enquadram nessa faixa salarial é igual ao produto da frequência acumulada pelo total de funcionários.

$$n = 0,85.160 = 136$$

Letra c.

057. (FCC/TRT 23/ANALISTA JUDICIÁRIO/ÁREA APOIO ESPECIALIZADO/ESPECIALIDADE: ESTATÍSTICA/2022/Q2593239) A tabela de frequências absolutas abaixo refere-se à distribuição dos salários (S) dos empregados que não possuem nível superior em uma empresa, sendo que não foram fornecidas as frequências da 2ª e 3ª classes dos empregados homens, denotadas na tabela por x e y , respectivamente.

Classes de Salários R\$	(Frequências) Homens	(Frequências) Mulheres	Total
$1.500 < S \leq 2.500$	20	25	45
$2.500 < S \leq 3.500$	x	40	$40 + x$
$3.500 < S \leq 4.500$	y	10	$10 + y$
$4.500 < S \leq 5.500$	15	5	20
Total	80	80	160

Foram calculadas separadamente, pelo método da interpolação linear, as medianas dos empregados homens e das mulheres, sendo que o valor da mediana referente aos homens superou em R\$ 425,00 o valor referente ao das mulheres. O valor médio dos salários dos homens, calculado como se todos os valores de uma classe coincidam com o ponto médio da respectiva classe, é igual a

- a) R\$ 3.925,00.
- b) R\$ 3.525,00.
- c) R\$ 3.875,00.
- d) R\$ 3.375,00.
- e) R\$ 3.625,00.



Primeiramente, vamos encontrar a mediana das mulheres

Observamos que a **mediana dos salários das mulheres $M(m)$ ocorre na segunda classe**, isto é, **R\$ 2.500,00 < S ≤ R\$ 3.500,00**.

Nesse intervalo, temos uma amplitude de 1000, pois $(3500 - 2500) = 1000$.

Como há **25** mulheres com salário **inferior a R\$ 2.500,00** e **25 + 40 = 65** mulheres com salário **inferior a R\$ 3.500,00**.

Podemos concluir que **existe algum valor $M(m)$ entre R\$ 2.500,00 e R\$ 3.500,00** tal que, no máximo **40 mulheres ganham o valor máximo $M(m)$** e mais do que **40 mulheres ganham pelo menos o valor mínimo $M(m)$** .

A questão pede uma estimativa para **essa mediana pelo método da interpolação linear**, isto é, sabendo-se que 25 mulheres ganham no máximo R\$ 2.500,00 e 65 mulheres ganham até R\$ 3.500,00, **devemos interpolar linearmente os valores** que encontraremos **$M(m)$** , que corresponde **ao valor para o qual no máximo recebido por 40 mulheres**.

A reta de interpolação f passa pelos pontos (25, 2.500) e (65, 3.500)

$$F(t) = 2500 + \left(\frac{3500 - 2500}{65 - 25} \right) \times (t - 25) = 2500 + 25 \times (t - 25) = 25t + 1875$$

Para $t = 40$, temos:

$$M(m) = 25 \times 40 + 1875 = 2875$$

Seja **$M(h)$ a mediana do salário dos homens**, a questão afirma que **$M(h)$ supera em R\$ 425,00 $M(m)$** , isto é:

$$M(h) = M(m) + 425 = 2875 + 425 = 3300$$

Portanto a **mediana dos homens** também **ocorre na segunda classe**, isto é, **R\$ 2.500,00 < S ≤ R\$ 3.500,00**.

Vamos **utilizar a interpolação linear** G entre os pontos (20, 2.500) e (40, 3.300) para determinar a abscissa do ponto (x + 20, 3.500)

$$G(t) = 2500 + \left(\frac{3300 - 2500}{40 - 20} \right) \times (t - 20) = 2500 + 40 \times (t - 20) = 40t + 1700$$

Substituindo o ponto (x + 20, 3.500), temos:

$$3500 = 40(x + 20) + 1700 = 40x + 800 + 1700 = 40x + 2500 \therefore x = 25$$

Como 20 + x + y + 15 = 80, temos que x + y = 45 e, portanto, **y = 20**.

Agora podemos calcular **o salário médio dos homens (Me)** tomando por base **o ponto médio dos intervalos**:

$$Me = \left(\frac{20 \times 2000 + 25 \times 3000 + 20 \times 4000 + 15 \times 5000}{80} \right) = \frac{27000}{8} = 3375$$

Letra d.

058. (FGV/EPE/ANALISTA DE PESQUISA ENERGÉTICA/ÁREA PLANEJAMENTO DA GERAÇÃO DE ENERGIA/2022/Q2358240) Os dados abaixo representam a distribuição de 100 domicílios residenciais, por classe de consumo de energia elétrica mensal, medidos em 2016.

Faixa de Consumo	Frequência Absoluta
0 --- 100	12
100 --- 200	25
200 --- 300	30
300 --- 400	25
400 --- 800	8

O consumo médio mensal, em kWh, pode ser estimado em

- a) 180.
- b) 254.
- c) 275.
- d) 300.
- e) 336.



Vamos encontrar o ponto médio de cada classe:

$$\frac{0 + 100}{2} = 50$$

$$\frac{200 + 100}{2} = 150$$

$$\frac{300 + 200}{2} = 250$$

$$\frac{400 + 300}{2} = 350$$

$$\frac{800 + 400}{2} = 600$$

Somando os números da frequência absoluta:

$$12 + 25 + 30 + 25 + 8 = 100 \text{ elementos}$$

Para o consumo médio, temos:

$$\mu = \frac{(50 \times 12) + (150 \times 25) + (250 \times 30) + (350 \times 25) + (600 \times 8)}{100}$$

$$\mu = \frac{25400}{100} = 254 \text{ kWh}$$

Letra b.

059. (INSTITUTO CONSULPLAN/IFPA/ESTATÍSTICO/2023/Q2868120) O histograma, também conhecido como distribuição de frequências, é a representação gráfica em colunas ou em barras (retângulos) contíguas de um conjunto de dados previamente tabulado e dividido em classes equidistantes ou não. O eixo vertical do histograma pode ser construído a partir da frequência absoluta ou da densidade, definida como a frequência relativa da classe dividida pela amplitude da classe. Sobre a construção de um histograma, analise as afirmativas a seguir.

I – Se as classes são equidistantes, as formas dos histogramas construídos a partir da frequência absoluta e da densidade são equivalentes.

II – Em qualquer histograma, a classe modal é representada pela barra de maior altura.

III – No histograma construído a partir da densidade, a área do gráfico é igual a 1.

Está correto o que se afirma em

- a) I, II e III.
- b) II, apenas.
- c) I e III, apenas.
- d) II e III, apenas.



Vamos analisar cada uma das alternativas da questão.

I – Certa. Uma vez que a densidade é a razão entre a frequência absoluta e a amplitude da classe, se as classes são equidistantes (à mesma amplitude), o gráfico da densidade terá a mesma forma da frequência absoluta.

Densidade = Frequência Relativa / Amplitude da Classe

II – Errada. Isso só é verdade para classes equidistantes. Se tivermos classes com diferentes amplitudes, pode haver uma classe modal (mais ocorrências na classe) mesmo com barra de menor altura. Há ainda os casos multimodais.

III – Certa. Quando um histograma é construído usando densidade, a área total sob o histograma é igual a 1, pois representa a distribuição de probabilidade dos dados. Matematicamente, temos:

$$\text{Área Total} = \sum (\text{Densidade} \times \text{Amplitude da Classe}) = 1$$

Observe que o produto da densidade pela amplitude de uma classe qualquer é igual à frequência relativa daquela classe, e a soma das frequências relativas de todas as classes deve ser sempre igual a 1.

Letra c.

060. (IBFC/PREFEITURA DE CUIABÁ/ESTATÍSTICO/2023/Q2738868) A tabela a seguir fornece a distribuição de frequências das notas de 45 alunos de um curso de estatística básica:

Intervalo de classe	Frequência simples	Frequência relativa
0 [— 3	4	9
3 [— 5	A	22
5 [— 7	25	C
7 [— 10	B	D

Os valores de A, B, C e D são, aproximadamente:

- a) 10, 6, 40, 13.
- b) 6, 10, 50, 13.
- c) 10, 6, 13, 40.
- d) 10, 6, 40, 14.



No início da questão destacamos uma imprecisão da banca. A rigor, **frequência relativa**, dada pela razão entre a quantidade de elementos de uma classe e o total de elementos (N), é um valor entre 0 e 1.

$$\text{fre relativa}(x) = \frac{\text{freq absoluta}(x)}{N}$$

Possivelmente a banca se referia a uma frequência relativa percentual (%) quando indicou valores maiores que 1. Vamos assim considerar para chegar à resposta da questão.

Podemos resolver a questão utilizando regra de três simples:

A frequência absoluta 4 está para a relativa 9% assim como A está para 22%. Temos:

$$\frac{4}{9} = \frac{A}{22} \therefore A \cong \frac{4 \cdot 22}{9} = \frac{88}{9} \cong 10$$

A frequência absoluta 4 está para a relativa 9% assim como 25 está para C%. Temos:

$$\frac{4}{9} = \frac{25}{C} \therefore C \cong \frac{9 \cdot 25}{4} = \frac{225}{4} \cong 56$$

Sabendo que C = 56, vamos determinar o valor de D, pois o somatório das frequências relativas é 100%, logo:

$$9\% + 22\% + 56\% + D = 100\%$$

$$\therefore D = 100\% - 56\% - 22\% - 9\% = 13\%$$

Por fim, a quantidade B também pode ser determinada por regra de três simples.

A frequência absoluta B está para a relativa 13% assim como 25 está para C%. Temos:

$$\frac{B}{13} = \frac{4}{9} \therefore B = \frac{13 \cdot 4}{9} \cong 6$$

Portanto, os valores de A, B, C e D são, aproximadamente: **10, 6, 56 e 13**.

A **alternativa A**, é de fato, a que melhor responde à questão, mas não é razoável a aproximação de 56 para 40, motivo pelo qual a questão deveria ser **anulada**.

Letra a.

061. (EPL CONCURSOS/TRE RN/ESTATÍSTICA/2023/Q3103792) Na tabela de distribuição de frequência apresentada a seguir alguns valores foram omitidos e outros foram substituídos pelas letras A, B, C, D e E.

Percentual de redução	f_i	$f_r(\%)$	F_{AC}
2 – 15	12	30,0	12
15 – 28	11	D	23
28 – 41	B		35
41 – 54	3		C
54 – 67	1	2,50	
67 – 80		E	
Total	A	100	-

Considerando os valores numéricos de A, B, C, D e E, assinale a alternativa correta.

- a) A = 39.
- b) B = 13.
- c) C = 40.
- d) D = 27.
- e) E = 2,5.



Vamos achar letra por letra e ir preenchendo a tabela

Sabemos que a frequência relativa(F_r) é a frequência absoluta dividida pelo total de elementos. Sabendo disso vamos calcular o valor de A:

$$\frac{12}{A} = 30\% \therefore A = 12 \cdot \frac{30}{100} = 40$$

Para achar o valor de B, vamos calcular a F_{AC} , a 1ª se repete e em seguida somamos a F_i (frequência relativa) da próxima classe.

Com isso, achamos B e C:

$$23 + B = 35 \therefore B = 12$$

$$C = 35 + 3 \therefore C = 38$$

Usando o mesmo raciocínio que achamos o valor de A, vamos calcular D e E.

$$D = \frac{11}{40} = 0,275 \therefore D = 27,5\%$$

Percentual de redução	f_i	$f_r(\%)$	F_{AC}
2 – 15	12	30,0	12 (se repete)
15 – 28	11	D=27,5	23(12+11)
28 – 41	B		35(23+B) B=12
41 – 54	3		C=35+3=38
54 – 67	1	2,50	39
67 – 80	1	E=2,5	
Total	A=40	100	–

Sabendo o valor de B, completamos a f_i da última classe, que é 1.

Sabendo o valor de C, completamos a F_{AC} da penúltima classe, que é 39.

Após isso, temos informações para descobrir o valor de E:

$$E = \frac{1}{40} = 0,025 \therefore E = 2,5\%$$

Letra e.

062. (CESGRANRIO/IBGE/2010/ANALISTA DE SISTEMAS) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

Julgue o item a seguir.

Classes (em anos)	f_i
0 – 2	5
2 – 4	2
4 – 6	4
6 – 8	2
8 – 10	7

A média das idades dessas crianças, em anos, é

- a) 5,0.
- b) 5,2.
- c) 5,4.

- d) 5,6.
e) 5,8.



Quando se tem dados categorizados, a média pode ser calculada usando o ponto médio das classes.

$$\mu = \frac{1.5 + 2.3 + 4.5 + 2.7 + 7.9}{5 + 2 + 4 + 2 + 7} = \frac{5 + 6 + 20 + 14 + 63}{20} = \frac{108}{20} = 5,4$$

Letra c.

063. (CESGRANRIO/IBGE/2010/ANALISTA DE SISTEMAS) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças. Julgue o item a seguir.

Classes (em anos)	f_i
0 – 2	5
2 – 4	2
4 – 6	4
6 – 8	2
8 – 10	7

A mediana das idades dessas crianças, em anos, é

- a) 5,5.
b) 5,6.
c) 5,7.
d) 5,8.
e) 5,9.



Por outro lado, para a mediana, quando se tem dados categorizados, devemos usar que:

$$\frac{20}{2} = 10 \therefore Md = x_{10}$$

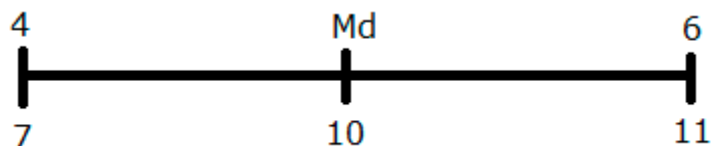
No caso de dados categorizados, não fazemos aquela correção de que a mediana seria calculada pela média aritmética de dois termos da amostra.

A mediana só pode ser calculada por interpolação linear.

Classes	Frequência Absoluta	Elementos
0 a 2	5	1 a 5
2 a 4	2	6 a 7

Classes	Frequência Absoluta	Elementos
4 a 6	4	8 a 11
6 a 8	2	12 a 13
8 a 10	7	14 a 20

Fazendo o esquema que já conhecemos para a interpolação linear, temos:

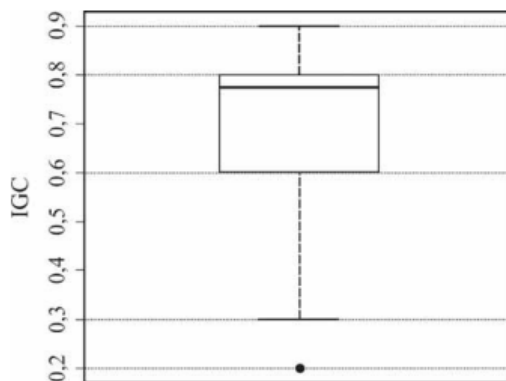


$$\frac{Md - 4}{10 - 7} = \frac{6 - 4}{11 - 7} = \frac{2}{4} \therefore \frac{Md - 4}{3} = \frac{2}{4}$$

$$\therefore Md - 4 = \frac{2 \cdot 3}{4} = 1,5 \therefore Md = 4 + 1,5 = 5,5$$

Letra a.

(CESPE/TCE/PR/2016/ADAPTADA) 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, julgue os seguintes itens.



064. O menor IGC observado na amostra foi 0,3.

065. O maior IGC observado na amostra foi 0,9.

066. O diagrama mostrado na figura em questão é denominado curva de frequência.

067. O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

068. Na amostra considerada, a média dos indicadores observados foi superior a 0,7 e inferior a 0,8.

069. A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

070. O desvio quartílico para a amostra em questão é de 0,3.



Essa questão caiu no TCE/PR como abcde, mas eu resolvi desmembrá-la e adicionar alguns itens para torná-la mais interessante.

Tem-se um diagrama box plot, não uma curva de frequência, com um *outlier* inferior (IGC = 0,2) e limite superior (IGC = 0,9). Portanto, o menor IGC da amostra foi 0,2 e o maior foi 0,9. Lembre que, quando não existem *outliers*, o limite superior ou inferior são ajustados para o valor máximo ou mínimo da amostra, respectivamente.

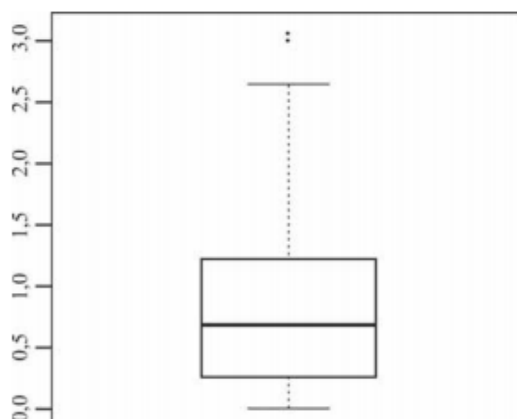
A caixa do box plot evidencia os quartis. Portanto, o primeiro quartil foi 0,6 e o terceiro foi 0,8. Além disso, a mediana (e não a média) que corresponde ao segundo quartil foi superior a 0,7 e inferior a 0,8.

Podemos calcular o desvio quartílico pela expressão:

$$D = \frac{Q_3 - Q_1}{2} = \frac{0,8 - 0,6}{2} = \frac{0,2}{2} = 0,1$$

Errado/Certo/Errado/Errado/Errado/Certo/Errado.

(CESPE/TCE/PA/2016) 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.



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

Com base nessas informações, julgue os itens a seguir.

071. O diagrama box-plot mostrado na figura sugere a existência de pelo menos duas observações atípicas.

072. X representa uma variável qualitativa ordinal.

073. A amplitude total da amostra é inferior a 3.

074. Pelo menos 50% das observações de X são menores que 1,5.



No topo do diagrama, temos dois *outliers*, ou seja, duas observações atípicas.

X é uma variável quantitativa, portanto, não pode ser ordinal, que é uma classificação para variáveis qualitativas.

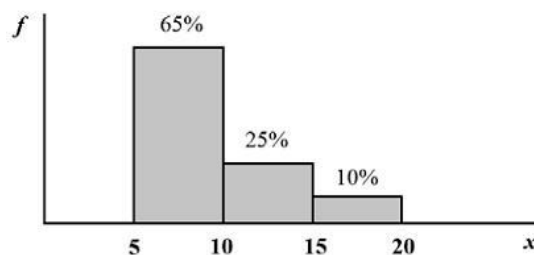
A amplitude da amostra é:

$$A = \max - \min = 3,1 - 0 = 3,1 > 3$$

Observe que o terceiro quartil é superior a 1,5, portanto, 75% dos elementos da amostra são inferiores a 1,5. Então, é realmente possível garantir que 50% das observações de X são menores que 1,5.

Certo/Errado/Errado/Certo.

075. (CESPE/STF/2013/ANALISTA JUDICIÁRIO/ESTATÍSTICA) Com referência à figura acima, que mostra a distribuição da renda mensal — x, em quantidades de salários mínimos (sm) — das pessoas que residem em determinada região, julgue o item subsequente.



A variável x, por possuir quatro níveis de respostas, é do tipo qualitativa ordinal.



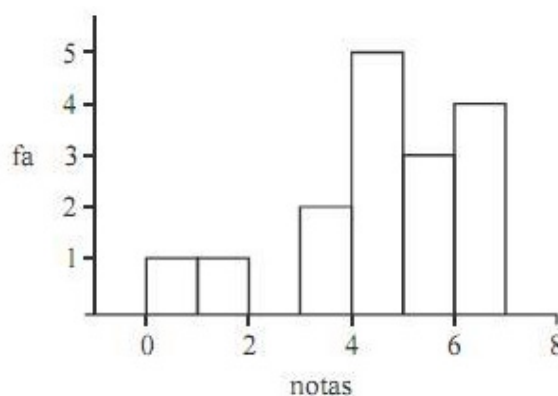
Observe que X não tem apenas quatro níveis de respostas. Na verdade, o diagrama representado é um histograma, o que mostra que as quatro marcações indicam as classes a que pertence a variável X.

Portanto, X pode se distribuir em três classes: cinco a dez salários mínimos; dez a quinze salários mínimos e quinze a vinte salários mínimos.

Dessa forma, X é uma variável quantitativa contínua.

Errado.

076. (CESPE/FUB/2013/ESTATÍSTICO) Considerando que o histograma acima apresenta a distribuição das notas finais dos estudantes matriculados em determinada disciplina, e que fa representa a frequência absoluta, julgue o seguinte item acerca de estatística descritiva.



Nessa disciplina, o percentual de estudantes com notas inferiores a 2,5 é igual a 12,5%.



Como o gráfico foi fornecido na forma de frequências acumuladas, primeiramente, precisamos somar todas elas para obter o total de elementos e, assim, obter as frequências relativas pedidas.

$$N = 1 + 1 + 2 + 5 + 3 + 4 = 16$$

O percentual de estudantes com notas inferiores a 2,5 pode ser obtido observando-se as duas primeiras classes, que são as únicas que tratam notas inferiores ao valor pedido. O total de estudantes que satisfazem à essa condição e, portanto, pertencem às duas primeiras classes é:

$$n = 1 + 1 = 2$$

Com base nisso, podemos calcular a frequência relativa.

$$f = \frac{2}{16} = \frac{1}{8} = 0,125 = 12,5\%$$

Certo.

077. (CESPE/FUB/2013/ESTATÍSTICO) A amplitude total da distribuição das notas é igual a 6.



A amplitude corresponde à diferença entre a maior e a menor nota. Pelo diagrama, a maior nota é igual a 7 e a menor nota é igual a 0. Portanto, a amplitude das notas é:

$$A = 7 - 0 = 7$$

Errado.

078. (CESPE/BACEN/2013/ANALISTA DE INFRAESTRUTURA E LOGÍSTICA)

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.

É inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos; dessa forma, apenas por meio de um gráfico de barras pode ser realizada a representação gráfica.

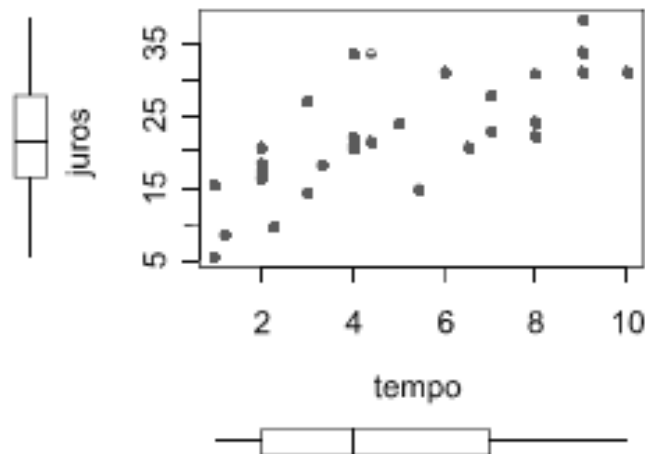


O histograma pode ser utilizado sempre que os dados podem ser separados em classes. Como podemos também separar dados discretos em classes em algumas situações, é possível também fazer um histograma com dados discretos.

Errado.

079. (CESPE/BANCO DA AMAZÔNIA/TÉCNICO CIENTÍFICO/ESTATÍSTICA) Em um estudo sobre o valor de juros (R\$) e encargos pagos versus tempo de atraso (dias) considerando o pagamento da fatura de cartão de crédito de 30 clientes, foi construído um diagrama de

dispersão e ajustado um modelo de regressão linear simples, seguido de quatro gráficos diagnósticos, apresentados a seguir.



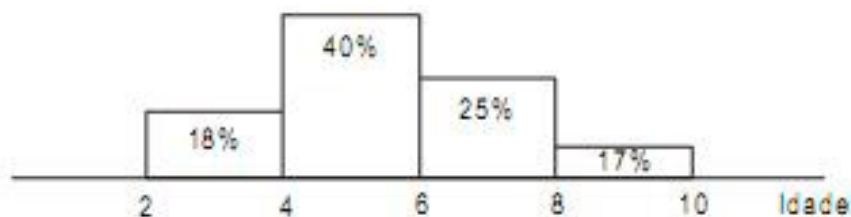
Os gráficos projetados nas margens do diagrama de dispersão são chamados de desenho esquemático ou *boxplot*, que resume graficamente a tendência central e a dispersão dos dados, sendo construído utilizando-se mínimo, máximo, média e quartis.



Realmente, o gráfico apresentado é um diagrama Box Plot. Porém, esse diagrama é construído com a mediana, e não com a média. Afirmiação incorreta.

Errado.

080. (FCC/ANS/2007/ESPECIALISTA EM REGULAÇÃO) O histograma abaixo representa a distribuição das idades dos pacientes atendidos no ano de 2000 em uma clínica infantil, expressa em anos.



A idade que separa os 30% mais jovens é

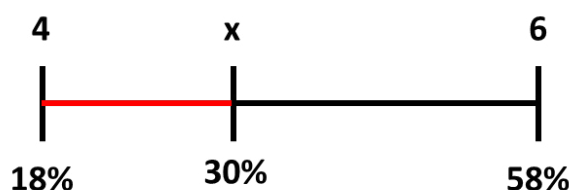
- a) 3,5.
- b) 4,2.
- c) 4,4.
- d) 4,6.
- e) 5,0.



A forma mais simples de obter o $x_{30\%}$, que é a idade que separa os 30% mais jovens, é pelo método de interpolação linear das frequências acumuladas. Para isso, primeiramente, devemos identificar a classe em que se verifica o acúmulo de frequência em 30%.

Idade	Frequência Relativa	Frequência Acumulada
2 – 4	18%	18%
4 – 6	40%	58%

A classe do quantil está identificada. Para usar a técnica da interpolação linear, devemos considerar os pontos extremos da classe do quantil e da classe anterior coincidentes com o nível da frequência acumulada, conforme esquematizado a seguir:



Agora, usemos a ideia de que o pedaço à esquerda é proporcional ao segmento inteiro.

$$\frac{x - 4}{30\% - 18\%} = \frac{6 - 4}{58\% - 18\%}$$

$$\frac{x - 4}{12\%} = \frac{2}{40\%}$$

$$\therefore x - 4 = \frac{12\%}{40\%} \cdot 2 = 0,3 \cdot 2 = 0,6$$

$$\therefore x = 4 + 0,6 = 4,6$$

Letra d.

081. (CESPE/DEPEN/2015/AGENTE PENITENCIÁRIO FEDERAL) 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.

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

O segundo quartil da distribuição das quantidades diárias de incidentes registradas nessa penitenciária é igual a 2.



O segundo quartil corresponde exatamente à mediana. Vamos destacar a classe mediana.

Idade	Frequência Relativa	Frequência Acumulada
0	0,10	0,10
1	0,20	0,30
2	0,50	0,70

Portanto, o segundo quartil é igual a 2.

Note que não utilizamos a interpolação linear, porque os dados não foram categorizados.

Certo.

082. (CESPE/MEC/2014/ANALISTA DE POLÍTICA REGULATÓRIA) A tabela acima, resultado de um estudo socioeconômico, mostra a distribuição percentual da renda familiar mensal dos estudantes do ensino médio em determinado município brasileiro. Considerando essas informações e a tabela acima, julgue o item seguinte.

renda familiar (R) (em salários mínimos)	percentual (%)
$0 < R \leq 1$	40
$1 < R \leq 3$	50
$R > 3$	10
total	100

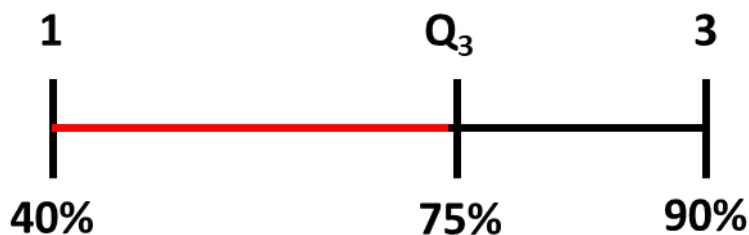
O terceiro quartil dessa distribuição de renda é superior a três salários mínimos.



Como temos dados organizados em categorias, podemos calcular o terceiro quartil, que corresponde à posição em que a frequência acumulada é igual a 75%.

Renda	Frequência Relativa	Frequência Acumulada
0 – 1	40%	40%
1 – 3	50%	90%
> 3	10%	100%

Agora, façamos a interpolação linear.



Note que nem precisamos terminar a interpolação, pois é fácil ver que $Q_3 < 3$. Portanto, a afirmação está incorreta. Porém, para fins de treino, podemos calcular o valor estimado de Q_3 .

$$\frac{Q_3 - 1}{75\% - 40\%} = \frac{3 - 1}{90\% - 40\%}$$

$$\frac{Q_3 - 1}{35\%} = \frac{2}{50\%}$$

$$\therefore Q_3 - 1 = \frac{35\%}{50\%} \cdot 2 = 1,4$$

$$\therefore Q_3 - 1 = 1,4 \therefore Q_3 = 1 + 1,4 = 2,4$$

Errado.

083. (CESPE/CEHAP/PB/2009/ADMINISTRADOR) O custo médio nacional para a construção de habitação com padrão de acabamento normal, segundo levantamento realizado em novembro de 2008, foi de R\$ 670,00 por metro quadrado, sendo R\$ 400,00/m² relativos às despesas com materiais de construção e R\$ 270,00/m² com mão-de-obra. Nessa mesma pesquisa, os custos médios regionais apontaram para os seguintes valores por metro quadrado: R\$ 700,00 (Sudeste), R\$ 660,00 (Sul), R\$ 670,00 (Norte), R\$ 640,00 (Centro-Oeste) e R\$ 630,00 (Nordeste).

Considerando o par de variáveis (X, Y), em que X representa o custo médio por metro quadrado (em R\$) e Y representa os atributos Sudeste, Norte, Sul, Centro-Oeste ou Nordeste, assinale a opção correta.

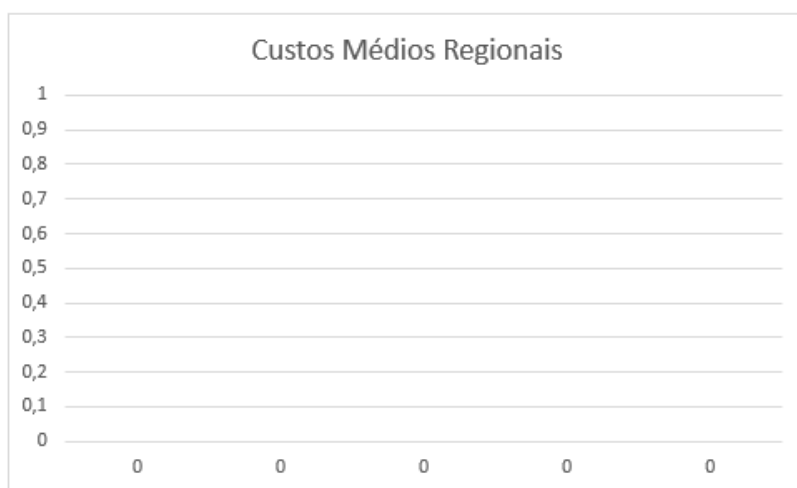
- a) A correlação linear de Pearson entre X e Y é positiva.
- b) As variáveis X e Y são qualitativas.
- c) As variáveis X e Y formam uma série geográfica.
- d) O gráfico de X versus Y é um histograma.



A variável Y é qualitativa. Portanto, em primeiro lugar, é preciso ter em mente que não é possível calcular correlação quando uma das variáveis é qualitativa.

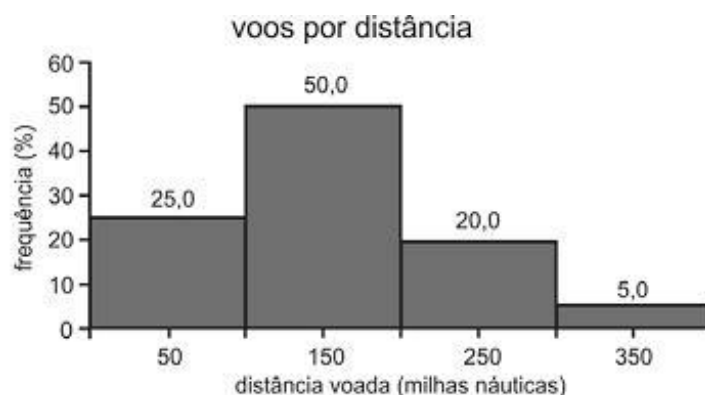
Como Y é a região do espaço, as variáveis X e Y realmente formam uma série geográfica.

Por fim, o erro da letra d) apontado pela banca é que o diagrama que pode ser construído não é um histograma, mas sim um gráfico de barras.



Letra c.

084. (CESPE/ANAC/2012/ESPECIALISTA EM REGULAÇÃO DE AVIAÇÃO CIVIL) Com base no histograma acima apresentado, julgue o item a seguir.



Os quartis das distâncias voadas (em milhas náuticas) obtidos a partir do gráfico são: 1º quartil = 100,0; 2º quartil (mediana) = 150,0; e 3º quartil = 200,0.



Como foi fornecido um histograma, temos a distribuição das distâncias voadas (milhas náuticas) organizadas em classes, cuja amplitude é exatamente igual à distância entre os pontos médios, ou seja:

$$A = 150 - 50 = 100$$

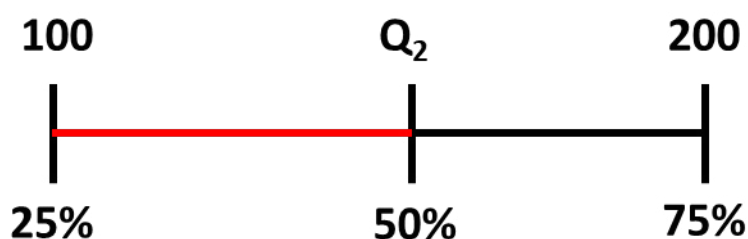
É interessante observar que, como a primeira classe tem frequência relativa igual a 25%, o primeiro quartil corresponde exatamente ao seu extremo. Logo, podemos escrever:

$$Q_1 = PM + \frac{h}{2} = 50 + \frac{100}{2} = 50 + 50 = 100$$

Como a segunda classe tem frequência relativa igual a 50%, a frequência acumulada entre as duas primeiras classes é igual a 75%, logo, o extremo da segunda classe corresponde ao terceiro quartil.

$$Q_3 = PM + \frac{h}{2} = 150 + \frac{100}{2} = 150 + 50 = 200$$

Por fim, a mediana pode ser obtida por interpolação linear entre as duas classes.



$$\frac{Q_2 - 100}{50\% - 25\%} = \frac{200 - 100}{75\% - 25\%}$$

$$\frac{Q_2 - 100}{25\%} = \frac{100}{50\%}$$

$$\therefore Q_2 - 100 = \frac{25\%}{50\%} \cdot 100$$

$$\therefore Q_2 - 100 = 50 \therefore Q_2 = 100 + 50 = 150$$

Certo.

085. (FCC/TRT/3ª REGIÃO (MG)/2015) A tabela de frequências relativas abaixo refere-se à distribuição dos salários dos empregados de uma empresa no mês de maio de 2015.

CLASSE DE SALÁRIOS (R\$)	FREQUÊNCIA RELATIVA
a — b	0,10
b — c	0,20
c — d	0,25
d — e	0,40
e — f	0,05
TOTAL	1,00

Observação: Todos os intervalos de classe apresentam a mesma amplitude igual a R\$ 1.000,00 e utilizou-se o método da interpolação linear para calcular a mediana (M_d) e o terceiro quartil (Q_3) da distribuição.

Se $M_d = \text{R\$ } 4.200,00$, então Q_3 é, em R\$, igual a

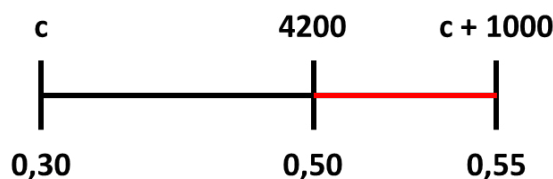
- a) 4.500,00.
- b) 4.900,00.
- c) 4.640,00.
- d) 4.800,00.
- e) 4.720,00.



Primeiramente, vamos localizar a classe mediana, que é aquela em que a frequência acumulada atinge 0,50.

Classe de Salários	Frequência Relativa	Frequência Acumulada
a — b	0,10	0,10
b — c	0,20	0,30
c — d	0,25	0,55

Para obter a mediana por interpolação linear, devemos considerar que o final da classe anterior (c) se encerra exatamente na frequência acumulada 0,30 e que o final da classe mediana (d) se encerra exatamente na frequência acumulada 0,55.



Como a amplitude de classe é igual a 1.000, sabemos que $d = c + 1.000$. Por interpolação linear, podemos dizer que a porção à direita da linha de interpolação é proporcional ao comprimento inteiro do segmento.

$$\frac{4200 - c}{0,50 - 0,30} = \frac{c + 1000 - c}{0,55 - 0,30}$$

$$\frac{4200 - c}{0,20} = \frac{1000}{0,25}$$

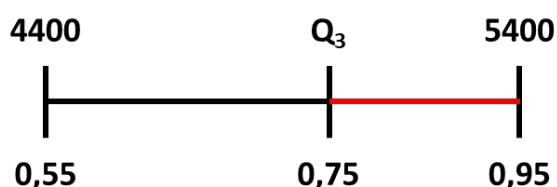
$$\therefore 4200 - c = \frac{0,20}{0,25} \cdot 1000 = 0,20 \cdot 4000 = 800$$

$$\therefore c = 4200 - 800 = 3400$$

Agora que fomos capazes de calcular os limites de uma classe, como sabemos a amplitude, somos capazes de calcular todas as classes. Como queremos encontrar o terceiro quartil (Q_3), buscamos o ponto em que a frequência acumulada é igual a 0,75.

Classe de Salários	Frequência Relativa	Frequência Acumulada
1400 – 2400	0,10	0,10
2400 – 3400	0,20	0,30
3400 – 4400	0,25	0,55
4400 – 5400	0,40	0,95
5400 – 6400	0,05	1,00

Usemos, novamente, a técnica de interpolação linear com os valores destacados.



Pela técnica da interpolação linear, temos que a parte à direita é proporcional ao segmento inteiro.

$$\frac{5400 - 4400}{0,95 - 0,55} = \frac{Q_3 - 4400}{0,75 - 0,55}$$

$$\frac{1000}{0,40} = \frac{Q_3 - 4400}{0,20}$$

$$\therefore Q_3 - 4400 = \frac{0,20}{0,40} \cdot 1000$$

$$\therefore Q_3 - 4400 = 500 \therefore Q_3 = 4400 + 500 = 4900$$

Letra b.

086. (FGV/IBGE/2017/ANALISTA CENSITÁRIO/MÉTODOS QUANTITATIVOS) Uma vez concluída a etapa de críticas de dados, relativa a um conjunto de registros obtido através de uma pesquisa de campo, inicia-se o trabalho de tabulação e elaboração de gráficos. Durante as análises, algumas variáveis surgem com destaque.

(i) número de indivíduos por faixa etária; (ii) percentuais do nível de escolaridade; e (iii) pares de valores de consumo e renda.

Portanto, os tipos de gráficos considerados adequados a serem empregados em cada caso são, respectivamente:

- a) colunas, linha e setores.
- b) setores, cartograma e barras.
- c) barras, setores e linha.
- d) linha, colunas e cartograma.
- e) colunas, setores e linha.



É preciso conhecer o tipo de cada variável para determinar o gráfico mais recomendado para elas. Vejamos:

- **número de indivíduos por faixa etária:** é uma variável aleatória contínua, portanto, pode ser representada adequadamente por um histograma;
- **percentuais do nível de escolaridade:** os percentuais são melhor representados pelo gráfico de setores (ou de pizza);
- **pares de valores de consumo e renda:** o consumo pode ser expresso como uma função da renda, então, é possível usar um gráfico de linha.

Nessa questão, o examinador confundiu o diagrama de colunas com o histograma. Embora o histograma seja formado por colunas, este se aplica a variáveis discretas e aquele para variáveis contínuas.

Já o diagrama de colunas é semelhante ao de barras – a única diferença é que as barras são horizontais e as colunas são verticais.

Letra e.

087. (VUNESP/PREFEITURA DE ILHABELA/SP/2020/ANALISTA DE GESTÃO PÚBLICA/DESAFIO)

Um histograma foi elaborado com o objetivo de se analisar a distribuição de todos os salários dos funcionários em uma empresa. Sabe-se que 108 funcionários dessa empresa ganham salários com valores pertencentes ao segundo intervalo do histograma $[4, 7)$, em salários-mínimos (SM), que apresenta uma densidade de frequência igual a $0,15 \text{ (SM)}^{-1}$. Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela amplitude deste intervalo. Se o primeiro intervalo do histograma $[2, 4)$, em SM, apresenta

uma densidade de frequência igual a 0,125 (SM)–1, então o número de empregados do segundo intervalo supera o número de empregados do primeiro intervalo em

- a) 48.
- b) 60.
- c) 72.
- d) 78.
- e) 84.



A frequência de uma categoria pode ser obtida como o produto da amplitude pela densidade de frequência. Então, a frequência da classe $[4, 7)$ é:

$$F_1(4:7) = (7 - 4) \cdot 0,15 = 3 \cdot 0,15 = 0,45$$

Como a frequência absoluta da classe é igual a 108 e sabemos que ela corresponde a 45% do total da população, podemos usar essas informações para calcular o número de elementos totais da população.

$$0,45 \cdot N = 108$$

$$\therefore N = \frac{108}{0,45} = 240$$

Portanto, temos uma população de 240 elementos.

A frequência relativa da classe $[2, 4)$ é, por sua vez, a amplitude da classe multiplicada pela sua densidade de frequência:

$$f_2(2:4) = (4 - 2) \cdot 0,125 = 2 \cdot 0,125 = 0,25$$

Então, essa classe corresponde a uma frequência relativa de 25%; a frequência absoluta da classe pode ser obtida como o produto do número de elementos da população pela sua frequência relativa. Vejamos:

$$F_2(2:4) = 0,25 \cdot 240 = 60$$

Portanto, o primeiro intervalo tem 108 funcionários e o segundo intervalo tem 60 funcionários. Então, o primeiro intervalo supera o segundo em:

$$\Delta F = 108 - 60 = 48$$

Letra a.

Chegamos ao final de mais uma aula.

Forte abraço!

Thiago Cardoso.

Abra



caminhos



crie

futuros

gran.com.br

