PROBABILIDADE E ESTATÍSTICA

Apresentação das Variáveis Estatísticas

Livro Eletrônico

GRAN CURSOS ONLINE

SUMÁRIO

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, esse é um tema de base.

Vale notar que as provas de Estatística têm ficado em 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.

Desse modo, o presente capítulo é muito importante para o seu progresso na matéria.

Segue meu contato:

E-mail: 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 a 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 é 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 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.

Série histórica: é estudada apenas a influência do **fator tempo** sobre a variável observada.

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.

Figura 1: Exemplo de Série Histórica

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 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 o 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:

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.

Figura 2: Exemplo de Série Geográfica

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

Séries especificativas: é estudado apenas o efeito do fator espécie sobre a variável estatística em estudo.

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).

Figura 3: Exemplo de Série Especificativa

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 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).

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).

Figura 4: Exemplo de Série Mista Geográfica e Histórica

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, levando 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).

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).

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

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 definem 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:

título

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

coluna indicadora | cabeçalho

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).

corpo

Figura 6: Elementos Fundamentais de uma Tabela Estatística

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 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" no lugar de "Z".

DIRETO DO CONCURSO

Questão 1 (CESPE/ANTAQ/TÉCNICO EM REGULAÇÃO/2009) 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.

COMENTÁRIO

Certo.

Observe que é mostrado o comportamento de três variáveis: "exportação", "importação" e "total" em função do tempo, marcado em anos. Há, portanto, três séries temporais.

Questão 2 (CESPE/ABIN/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4/2018) 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.

COMENTÁRIO

Certo.

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

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

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%.

QUESTÃO 3 (CESPE/ABIN/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4/2018) A menor diferença entre as quantidades de professores dos anos iniciais do ensino fundamental e do ensino médio ocorreu em 2014.

COMENTÁRIO

Certo.

Vamos olhar a evolução da série história 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 2014, como mostrado na tabela.

---

2. Gráficos Estatísticos

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

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

N.	Estado Civil	Grau de Instrução	N. de Filhos	Salário	Idade (anos)	Idade (meses)	Região de Procedência
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
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

N.	Estado Civil	Grau de Instrução	N. de Filhos	Salário	Idade (anos)	Idade (meses)	Região de Procedência
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

Tabela 1: Conjunto de Dados de Referência

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.

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

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

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.

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

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

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

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

Tabela 3: Cálculo da Frequência Relativa

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%.

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%

Tabela 4: Cálculo da Frequência Relativa e Acumulada

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

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

DIRETO DO CONCURSO

Questão 4 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

COMENTÁRIO

Errado.

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.

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 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.

Fonte: G1 e FBSP

Figura 7: Exemplo de Gráfico de Colunas

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 2018 e 2019.

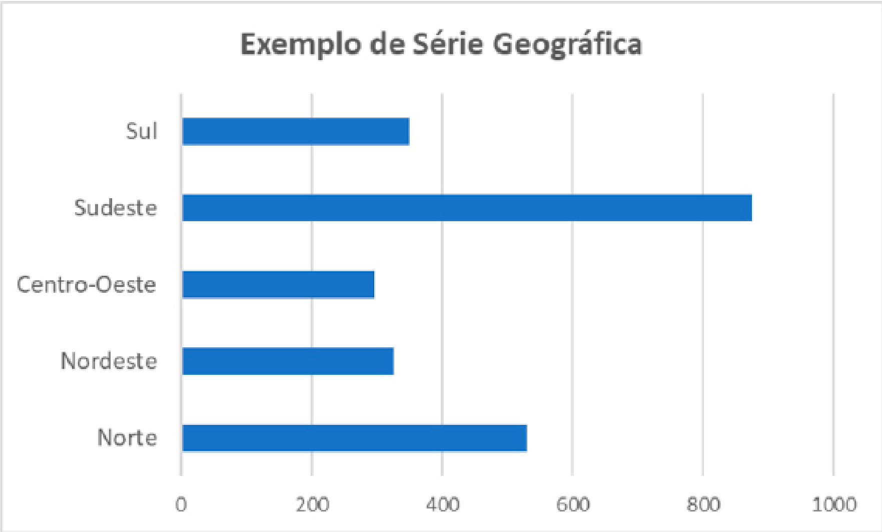


Figura 8: Exemplo de Gráfico de Barras

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:

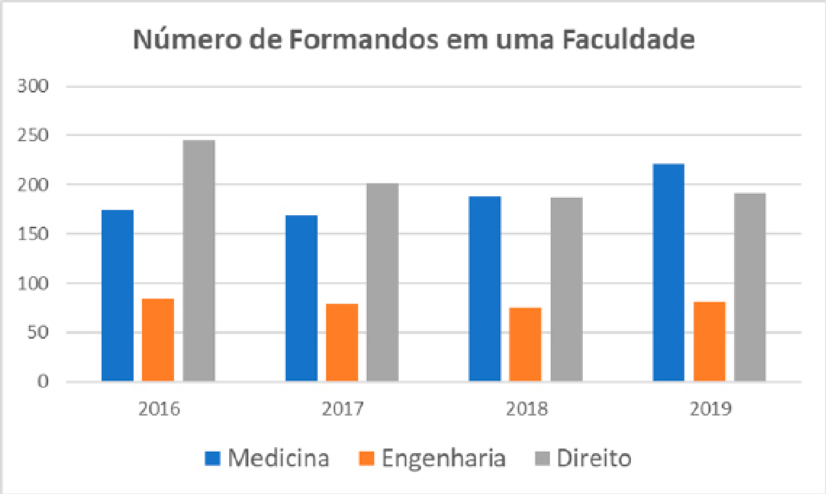


Figura 9: Exemplo de um Gráfico para uma Série Mista

No caso de séries geográficas, é 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.

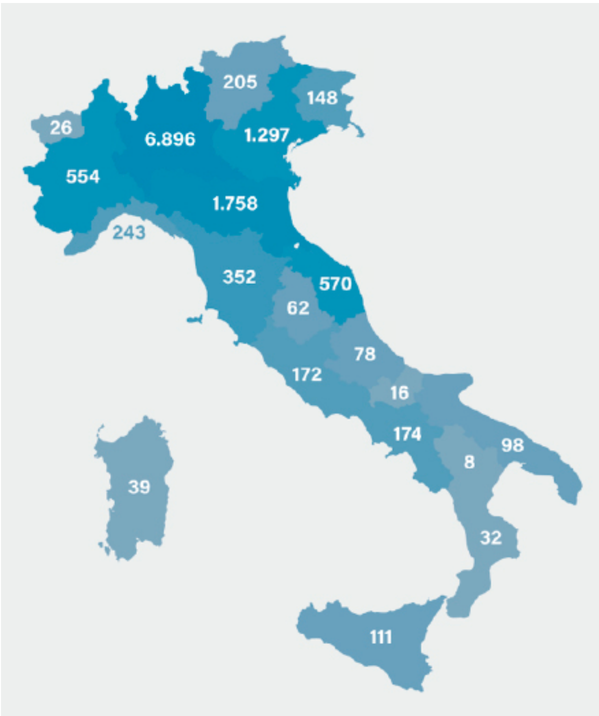


Figura 10: Exemplo de Gráfico para uma Série Geográfica

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. Mas não há nenhuma obrigatoriedade nisso.

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

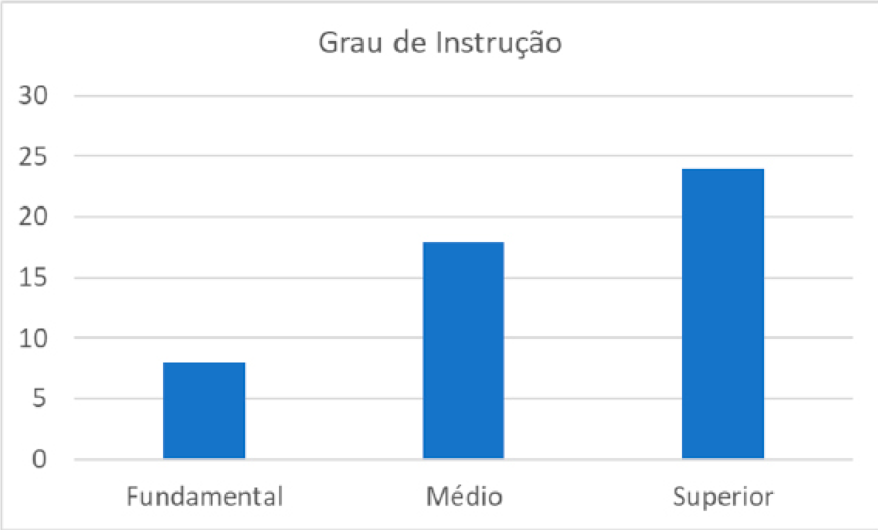
DIRETO DO CONCURSO

Questão 5 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

COMENTÁRIO

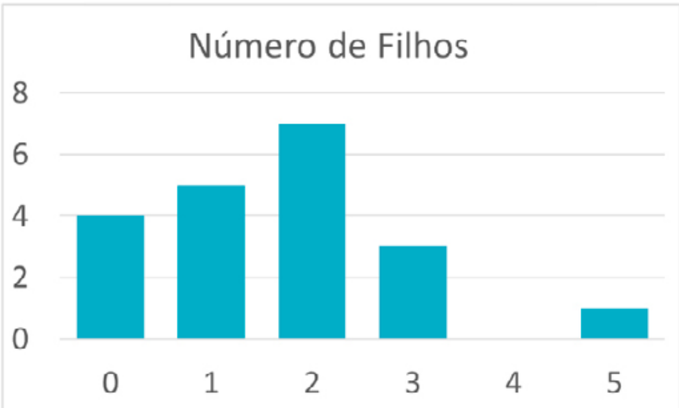
Certo.

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.



2.3. Gráfico de Setores

O gráfico de setores, também conhecido como pizza, é muito utilizado para **comparar proporções.** É muito útil 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 360º 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 360º, que corresponde ao círculo completo.

$$\theta_{Medicina} = 0{,}40.360 = 144°$$

$$\theta_{Engenharia} = 0{,}10.360 = 36°$$

$$\theta_{Direito} = 0{,}50.360 = 180°$$

Dessa forma, podemos traçar o seguinte gráfico de pizza:

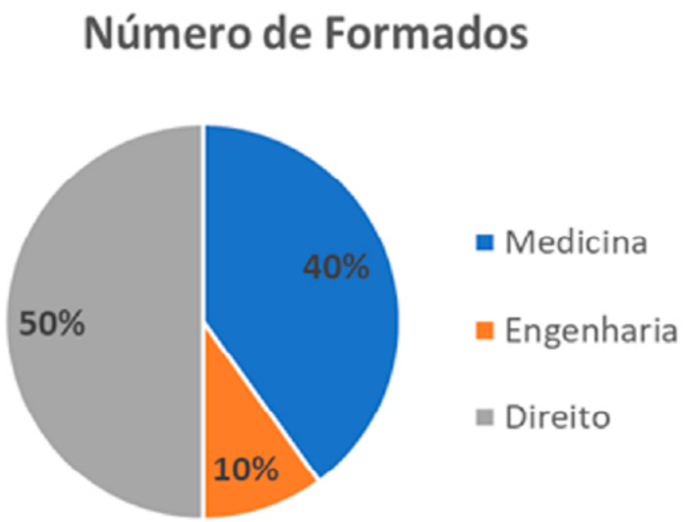


Figura 12: Exemplo de Gráfico de Setores

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

Questão 6 (CESPE/AL-CE/ANALISTA LEGISLATIVO/2011) 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.

COMENTÁRIO

Errado.

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.

Questão 7 (CESPE/TER-ES/ANALISTA/2011) 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 os itens 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> (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 120º.

COMENTÁRIO

Errado.

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 (360º). Portanto, o ângulo ocupado está para 513 assim como o ângulo total da pizza (360º) está para 1658.

$$\frac{\theta}{513} = \frac{360°}{1658}$$

$$\therefore \theta = \frac{513}{1658} \cdot 360° \cong 0{,}31.360° = 111°$$

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

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?

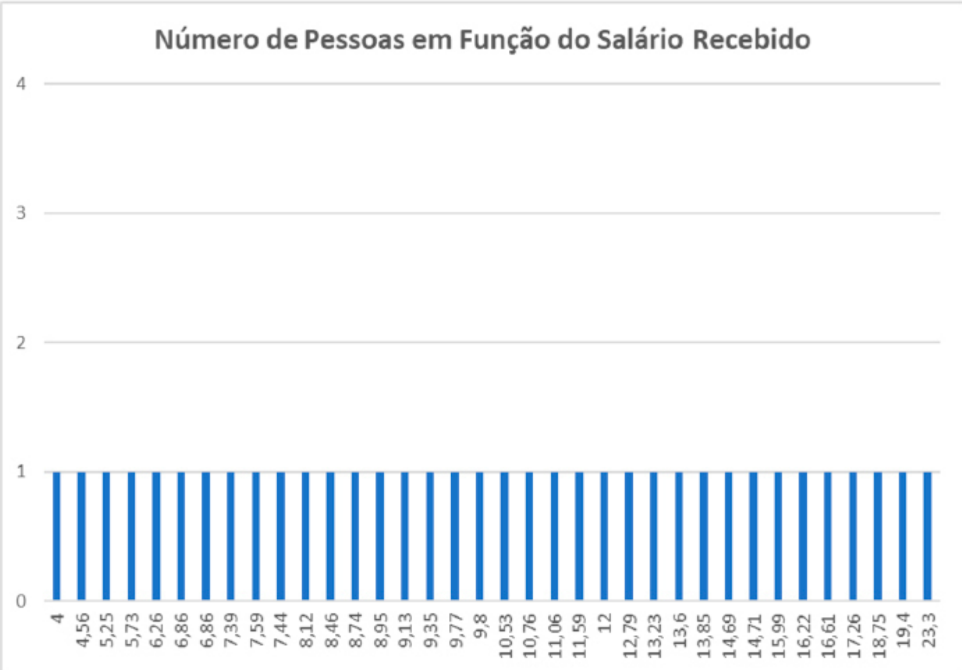


Figura 13: Exemplo de um Diagrama de Barras para uma Variável Contínua sem o Arranjo em Classes

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:

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

Tabela 5: Exemplo de Distribuição de Frequências em Classes

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.

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
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

Tabela 6: Contagem dos Elementos pertencentes a cada Classe

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

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**		

Tabela 7: Exemplo de Distribuição de Frequências em Classes

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.

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

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

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

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

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

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%**	

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

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.

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

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

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

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%**

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

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 a 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 a 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 estatísticas:

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%

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

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 ao ponto médio das classes referidas na Tabela 9.

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

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

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.

$$Lsup = PM + \frac{h}{2}$$

$$Linf = PM - \frac{h}{2}$$

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

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

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

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

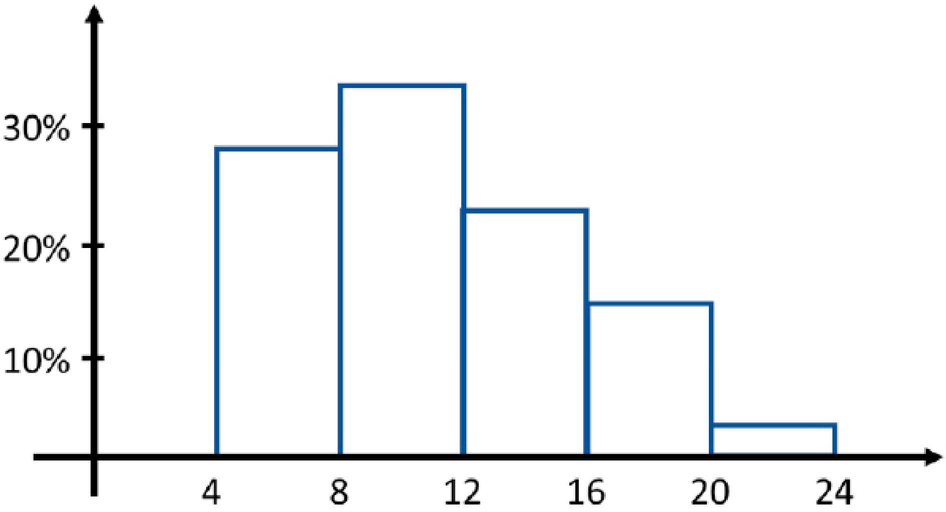


Figura 14: Representação de um Histograma em Classes representadas por seus intervalos

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.

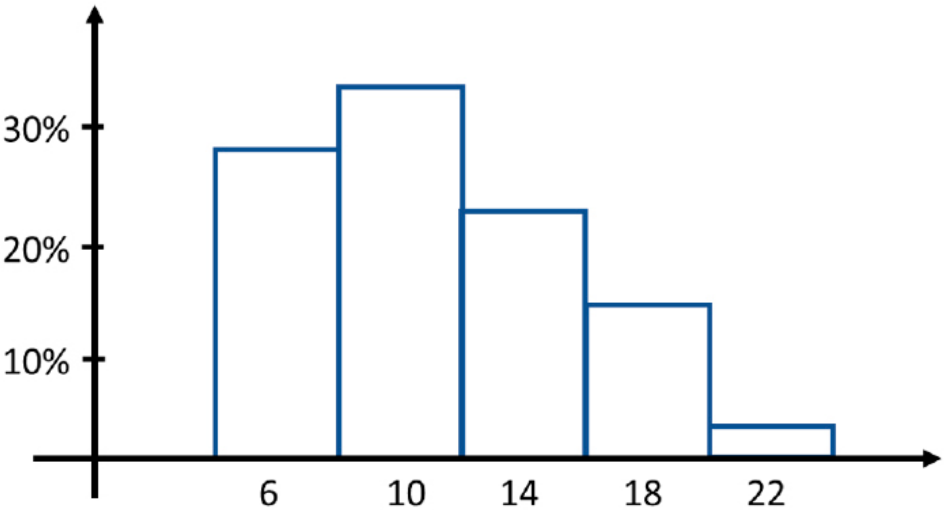
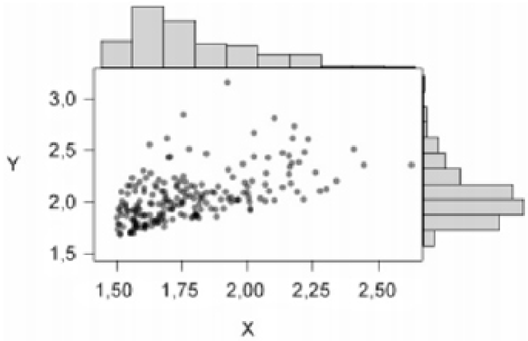


Figura 15: Representação de um Histograma em Classes representadas por seus pontos médios

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

Questão 8 (CESPE/MPU201/ANALISTA/ESTATÍSTICA/2013) 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.

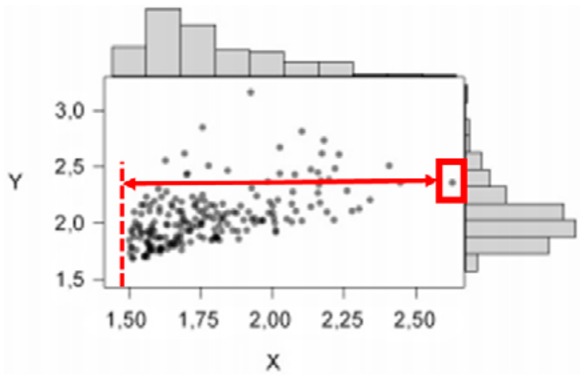


A amplitude total da distribuição de X foi igual ou superior a R$ 2 milhões.

COMENTÁRIO

Errado.

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.

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 bem estruturado 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.

Obs.: $K = \sqrt{N}$

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

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

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

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.

$$\boldsymbol{Lmax = 4,00 \quad Lmin = 23,30}$$

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

$$AT = Lmax - Lmin = 23{,}30 - 4{,}00 = 19{,}30$$

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

Obs.: $$h = \frac{AT}{K}$$

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:

Obs.: $$h = \frac{Lmax - Lmin}{\sqrt{N}}$$

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:

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	

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

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.

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

N.	Salário	Classe Correspondente		N.	Salário	Classe Correspondente
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

Tabela 13: Contagem dos Elementos pertencentes a cada Classe

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

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

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

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

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%

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

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

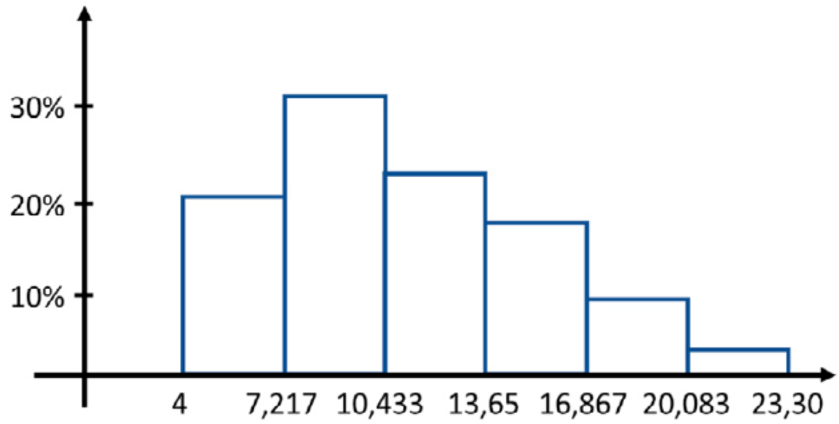


Figura 16: Distribuição de Frequências com o uso da Regra do Quadrado

O grande problema desse diagrama é que, embora ele tenha sido construído com conceitos matemáticos, ele não é agradável visualmente. 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, que 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 com 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. Particularmente, prefiro a primeira, pois altera menos a amplitude total da classe. E foi assim que construí a distribuição de frequências que usamos nesse 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. Comecemos 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:

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

Façamos a mesma conta para a quarta classe:

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

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.

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

Tabela 16: Exemplo de Distribuição de Frequências em Classes

DIRETO DO CONCURSO

Questão 9 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

COMENTÁRIO

Certo.

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 são 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 = √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

E a amplitude total da amostra será:

$$AT = 91 - 60 = 31$$

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.

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.

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

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

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

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%
16 |– 20	5	13,89%	16,67%
20 |– 24	1	2,78%	2,78%
Total	**36**	**100,00%**	**100,00%**

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

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.

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	**260**	**-**

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

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.

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	**260**	**-**

Tabela 19: Tratamento dos Dados para a Construção do Histograma

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:

Obs.:

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

Nessa expressão, 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.

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%
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	**260**	-	-	-	

Tabela 20: Tratamento dos Dados para a Construção do Histograma

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.

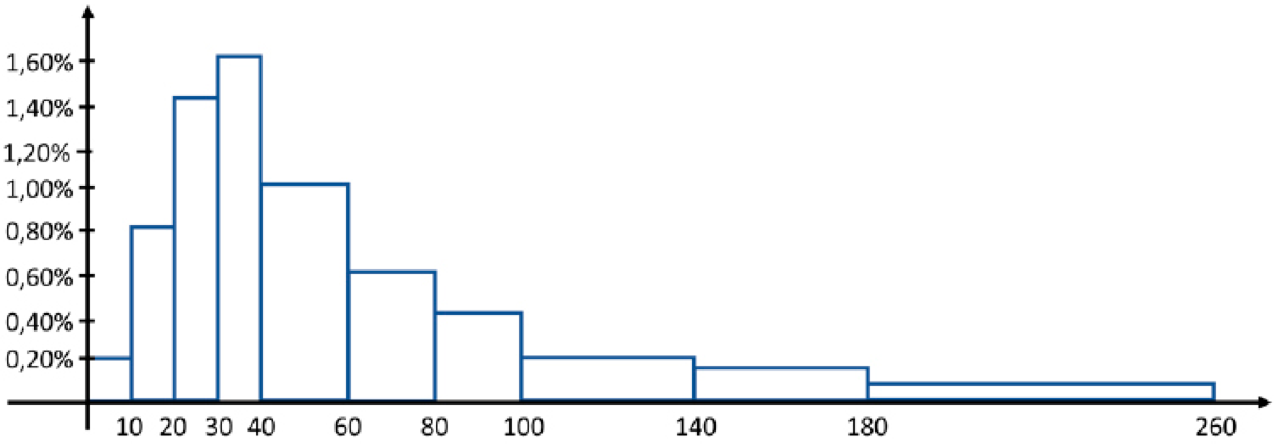


Figura 17: Exemplo de Histograma com Distribuições de Frequência com Classes de Amplitudes Desiguais

Perceba que a altura da barra nada tem a ver com a frequência relativa, mas 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.**

Obs.: $f_i = d_i . A_i$

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.

$$\mathbf{0| - 10} \qquad \boldsymbol{f_1 = d_1 . A_1 = 0{,}20.10 = 2\%}$$

$$\mathbf{10| - 20} \qquad f_2 = d_2 . A_2 = 0{,}80.10 = 8\%$$

$$\mathbf{20| - 30} \qquad f_3 = d_3 . A_3 = 1{,}40.10 = 14\%$$

$$\mathbf{30| - 40} \qquad f_4 = d_4 . A_4 = 1{,}60.10 = 16\%$$

$$\mathbf{40|-60} \qquad f_5 = d_5.A_5 = 1{,}00.20 = 20\%$$

$$\mathbf{60|-80} \qquad f_6 = d_6.A_6 = 0{,}60.20 = 12\%$$

$$\mathbf{80|-100} \qquad f_7 = d_7.A_7 = 0{,}40.20 = 8\%$$

$$\mathbf{100|-140} \qquad f_8 = d_8.A_8 = 0{,}20.40 = 8\%$$

$$\mathbf{140|-180} \qquad f_9 = d_9.A_9 = 0{,}015.40 = 6\%$$

$$\mathbf{180\,|-260} \quad f_{10} = d_{10}.A_{10} = 0{,}075.80 = 6\%$$

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.

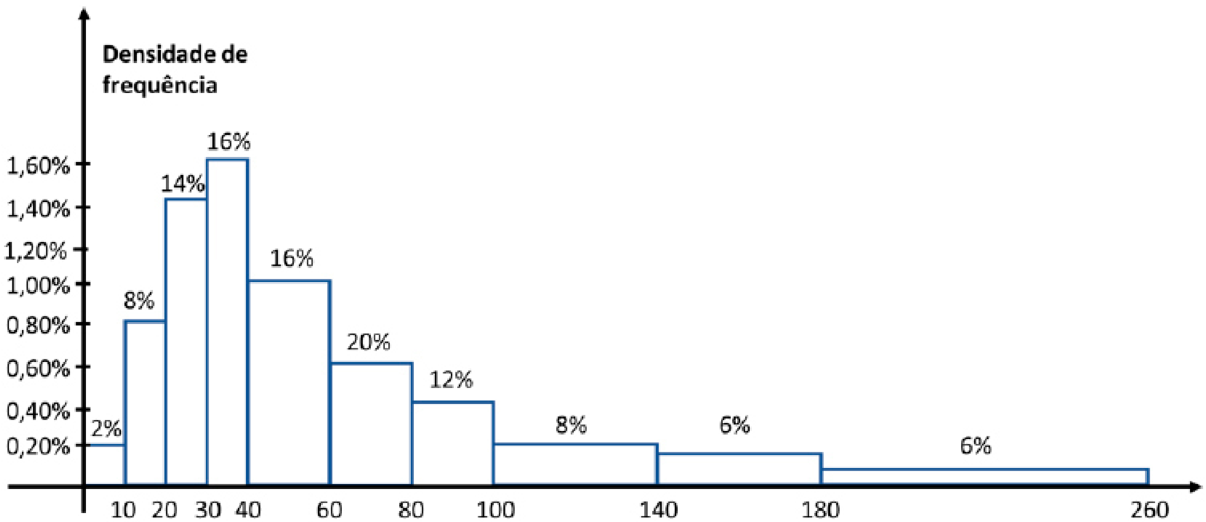


Figura 18: Cálculo das Frequências Relativas a partir da Densidade de Frequência

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 muitos 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

Questão 10 (FCC/TRT-14ª REGIÃO-RO-AC/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 × 10-4(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 × 10^{-4}(R$)$^{-1}$, tem-se que o correspondente número de empregados é igual a:

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

COMENTÁRIO

Letra d.

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

$$f = d.A = 0{,}75.10^{-4}.2000 = 0{,}15 = 15\%$$

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

$$0{,}15.N = 24$$

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

Portanto, a empresa tem o 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.

$$f = d.A = 1.10^{-4}.3000 = 0{,}30 = 30\%$$

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

$$n = 0{,}30.160 = 48$$

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 segunda 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 adicionais dois pontos:

$$P_0 = 6 - 4 = 2 \qquad\qquad P_f = 22 + 4 = 26$$

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.

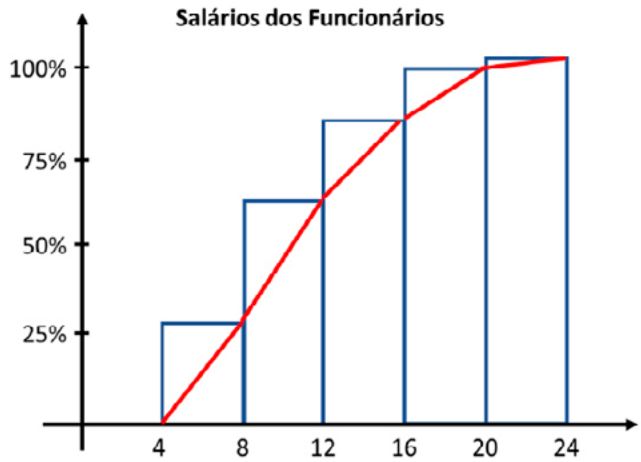
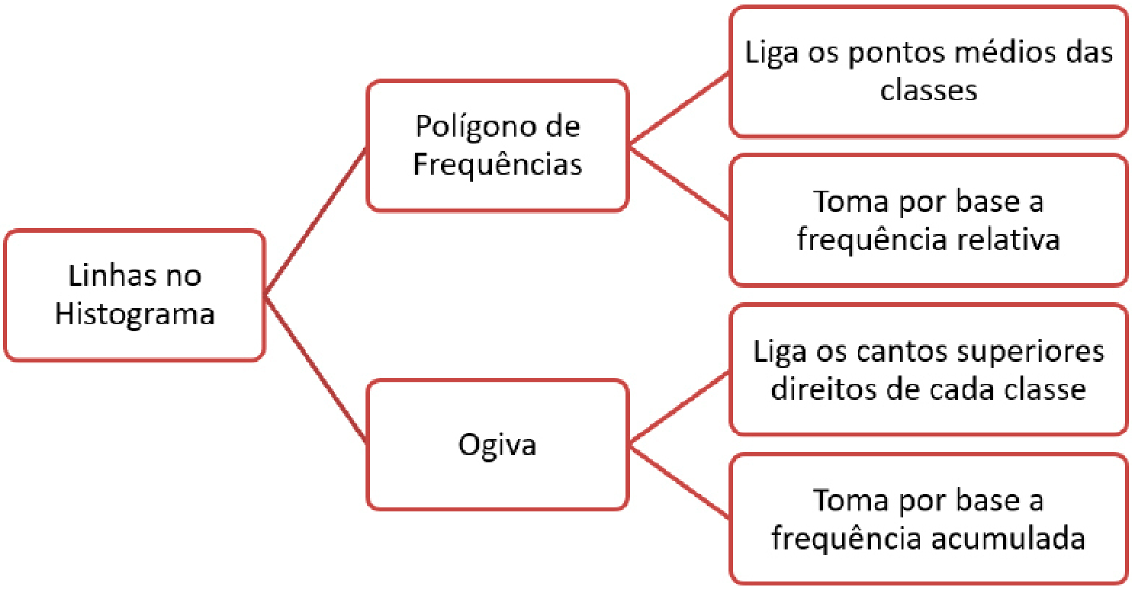


Figura 21: Exemplo de Ogiva

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:

Obs.: 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 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.

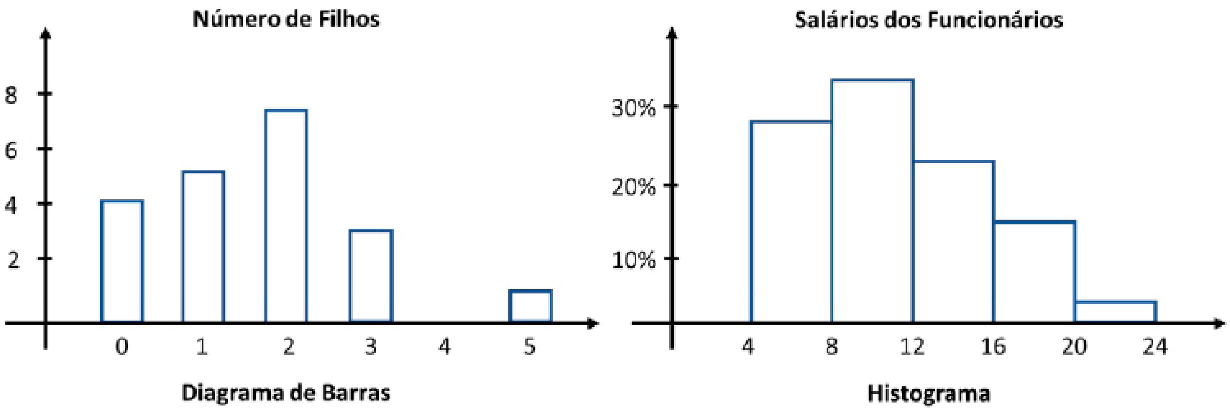


Figura 22: Comparação entre um Diagrama de Barras e um Histograma

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 tem 0 filho ou 1 filho).

Por outro lado, o histograma que representa a variável contínua não tem espaços vazios, porque as classes em que foi dividida 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.

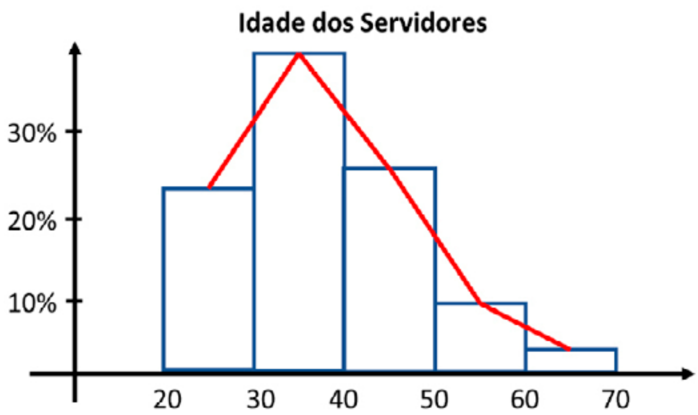


Figura 23: Idade dos Funcionários de um Órgão Público

DIRETO DO CONCURSO

Questão 11 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4/2015) 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

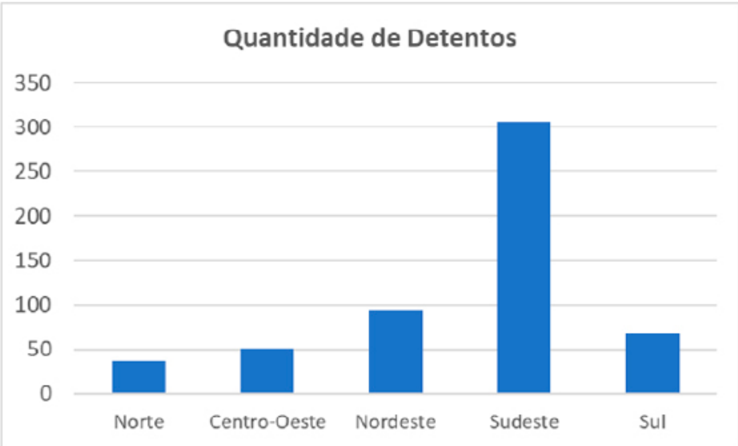
Ministério da Justiça — Departamento Penitenciário Nacional — Sistema Integrado de Informações Penitenciárias – InfoPen, Relatório Estatístico Sintético do Sistema Prisional Brasileiro, *dez./2013 Internet:<www.justica.gov.br> (com adaptações)*

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.

COMENTÁRIO

Errado.

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



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:

Obs.: **Idade 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

Idade 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, haverá 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, o diagrama a seguir:

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

Figura 24: Exemplo de Diagrama de Ramos e Folhas

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

QUESTÃO 12 (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 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

COMENTÁRIO

Letra a.

Devemos nos lembrar que, à esquerda, há o algarismo das dezenas das idades dos funcionários e, à direita, o algarismo das unidades. Então, podemos calcular a soma de todas as idades de todos os funcionários do setor ou órgão.

$$S = 18 + 18 + 19 + 20 + 21 + 21 + 3.22 + 27 + 28 + 28 + 29 + 31 + 4.33 + 3.34 + 35 + 36 + 37 + 3.38 + 40 + 41 + 2.42 + 43 + 44 + 48 + 49 + 51 + 55 + 58 + 62 + 65 = 1320$$

A soma é um pouco trabalhosa, porém encontramos o valor de 1.320. A média pode ser obtida como a razão entre o valor total somado e o número de funcionários do setor.

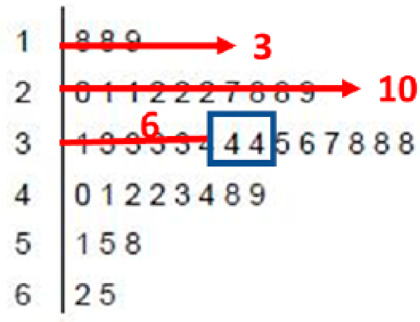
$$Me = \frac{1320}{40} = 33$$

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



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

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

$$Me + Md = 33 + 34 = 6$$

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.**

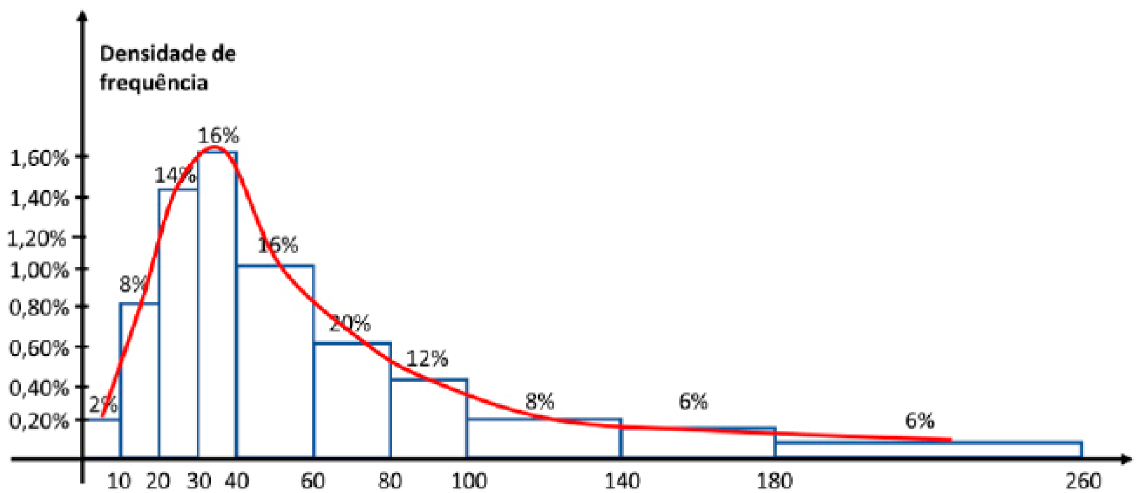


Figura 25: Linha de Tendência no Histograma

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 Multivariados

É muito comum agrupar vários dados em um mesmo histograma. Por exemplo, no nosso curso de Matemática Financeira, eu faço 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.

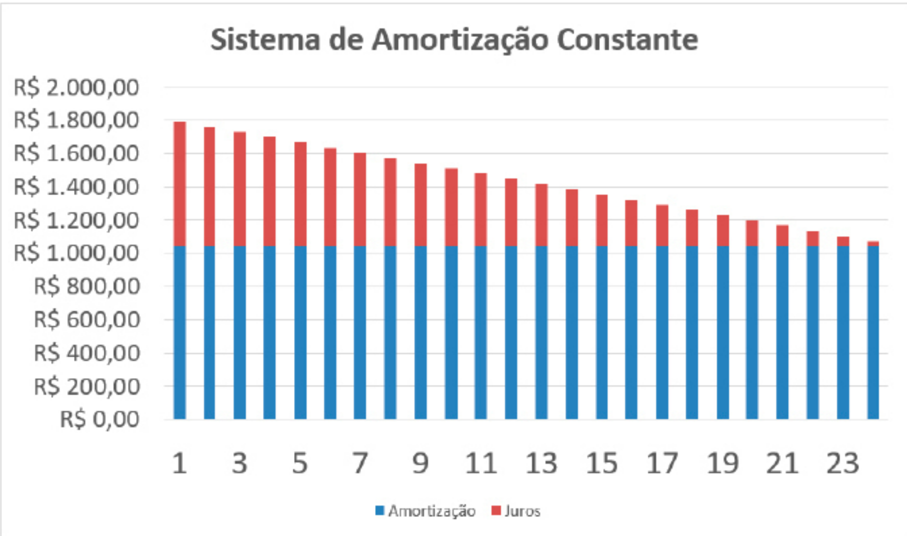


Figura 26: Histograma Multivariado

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.

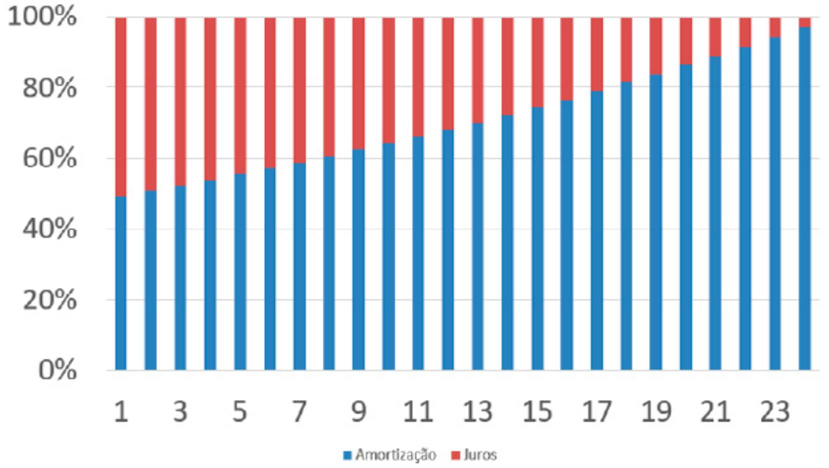


Figura 27: Histograma Multivariado com Proporções

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.

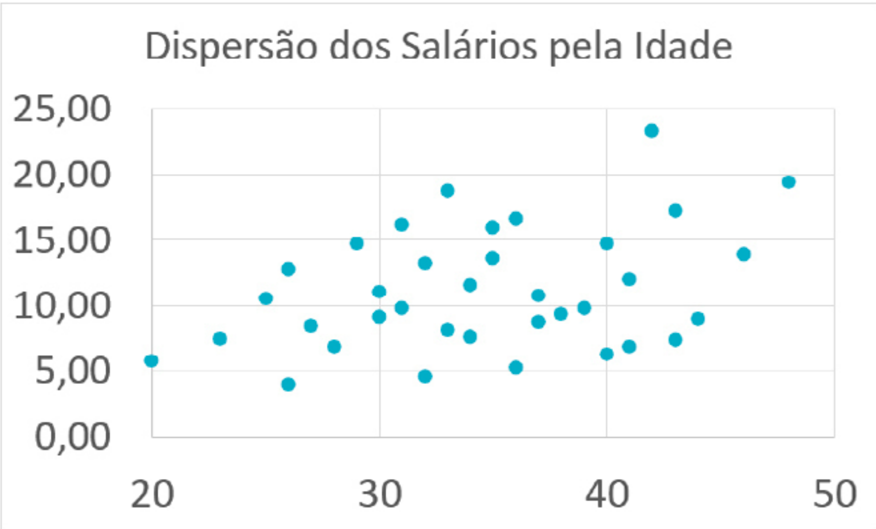


Figura 29: Gráfico de Dispersão de Salários pela Idade

Nesse gráfico, aparece 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:

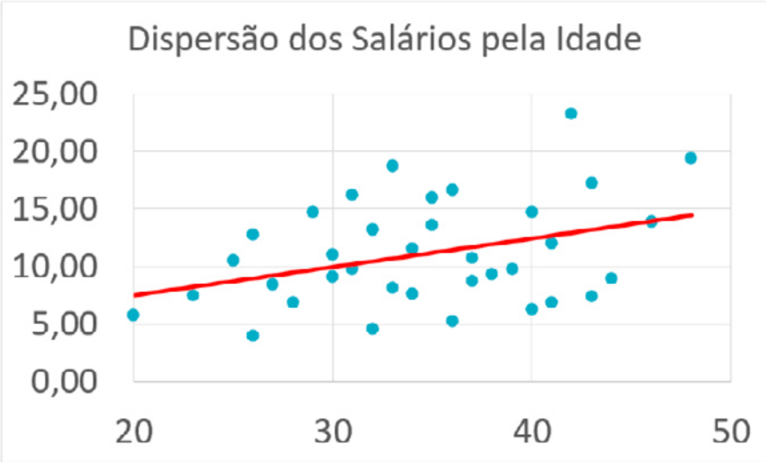


Figura 30: Linha de Tendência para um Gráfico de Dispersão

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

Questão 13 (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 os itens que se seguem, 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.

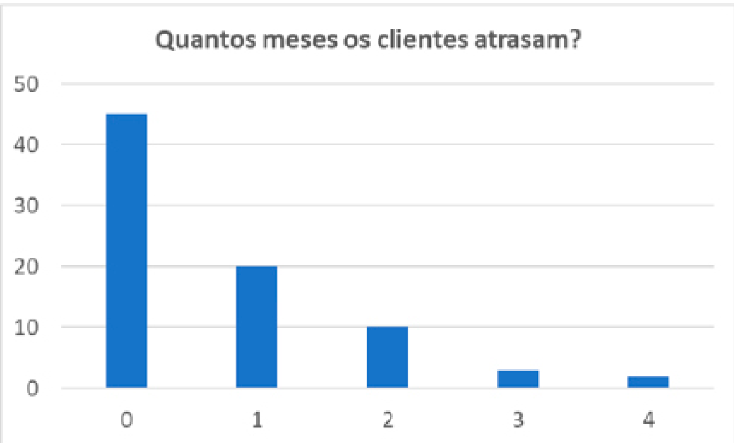
COMENTÁRIO

Errado.

Observe que existe 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 sendo 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.



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, 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.

Exemplos: suponha que você queira trazer um carro de luxo para o Brasil e gostaria de saber se haverá consumidores interessados. Para isso, 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 precise saber a renda dos 30% mais pobres do país para avaliar qual seria o custo ideal do seu produto.

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.

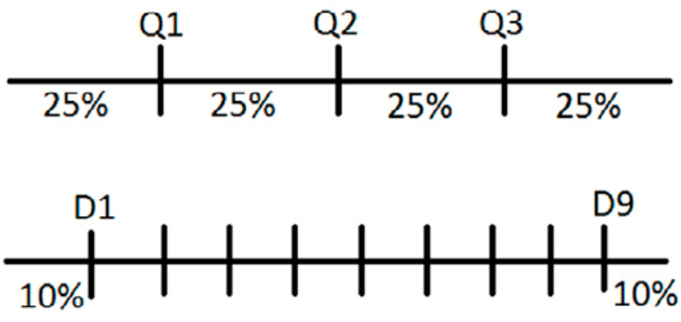


Figura 31: Quartis e Decis

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 80% 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. Há os percentis P_1, P_2, ..., 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:

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%

Tabela 21: Posição associada aos Quartis e Decis

Como exemplo, note 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:

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

$$Q_1 = x_5 = 1.082.935\ (São\ Luís)$$

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

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

Por fim, chamo 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.

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 |– 8	30%	30%
8 |– 12	35%	65%
12 |– 16	20%	85%
16 |– 20	10%	95%
20 |– 24	5%	100%
Total	**100,00%**	**100,00%**

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

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

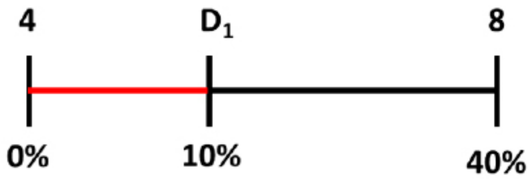
O primeiro quartil será aquele em que a frequência acumulada atingirá 1/4 ou 25%. Já o terceiro quartil será registrado quando a frequência acumulada atingir 75%.

Analogamente, para o primeiro decil, que divide a amostra em 10 partes, nota-se o que 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 (30%).



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.

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.

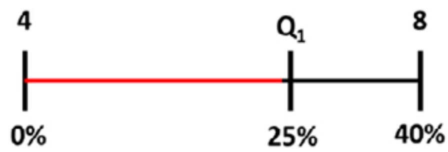
$$\frac{D_1 - 4}{10\% - 0\%} = \frac{8 - 4}{40\% - 0\%}$$

$$\frac{D_1 - 4}{10\%} = \frac{4}{40\%}$$

$$\therefore D_1 - 4 = \frac{10\%}{40\%} \cdot 4 = \frac{1}{4} \cdot 4 = 1$$

$$\therefore D_1 = 4 + 1 = 5$$

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



$$\frac{Q_1 - 4}{25\% - 0\%} = \frac{8 - 4}{40\% - 0\%}$$

$$\frac{Q_1 - 4}{10\%} = \frac{4}{40\%}$$

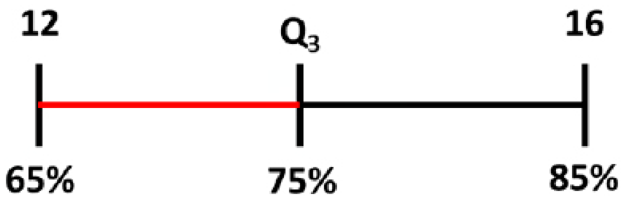
$$\therefore Q_1 - 4 = \frac{25\%}{40\%} \cdot 4 = \frac{100}{40} = 2{,}5$$

$$\therefore Q_1 = 4 + 2{,}5 = 6{,}5$$

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	30%	30%
8 |– **12**	35%	**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 extrema da classe do quantil (16) correspondente à marcação de 85%.



Fazendo a interpolação linear, há:

$$\frac{Q_3 - 12}{75\% - 65\%} = \frac{16 - 12}{85\% - 65\%}$$

$$\frac{Q_3 - 12}{10\%} = \frac{4}{20\%}$$

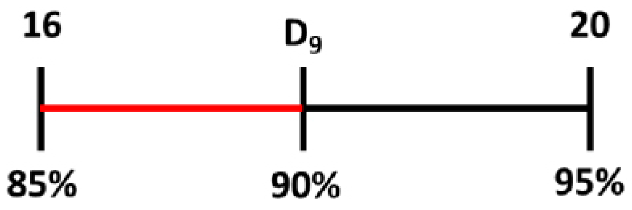
$$\therefore Q_3 - 12 = \frac{10\%}{20\%} \cdot 4 = 2$$

$$\therefore Q_3 = 12 + 2 = 14$$

Por fim, vamos calcular o nono decil.

Classe de Salários	Frequência Relativa	Frequência Acumulada
4 |– 8	30%	30%
8 |– 12	35%	65%
12 |– **16**	20%	**85%**
16 |– 20	**10%**	**95%**
20 |– 24	5%	100%
Total	**100,00%**	**100,00%**

Faremos o esquema da interpolação linear:



E realizaremos as contas:

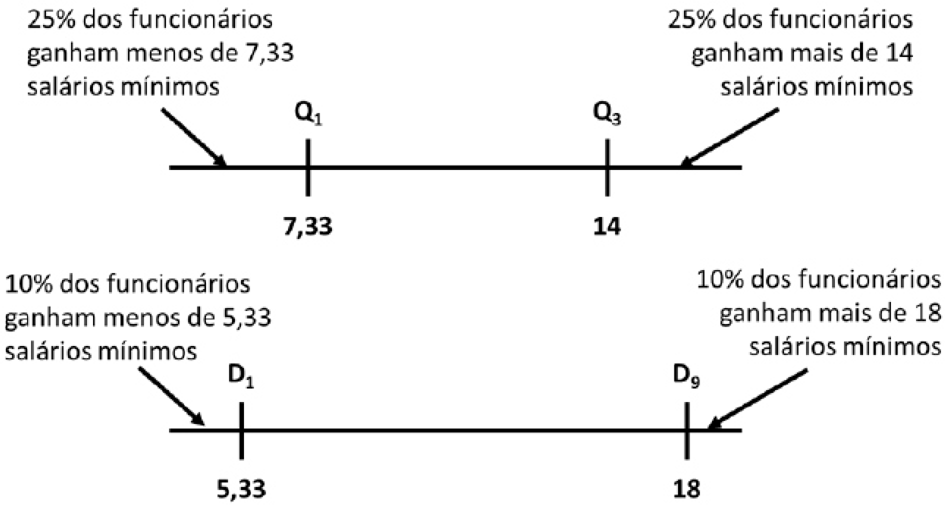
$$\frac{D_9 - 16}{90\% - 85\%} = \frac{20 - 16}{95\% - 85\%}$$

$$\frac{D_9 - 16}{5\%} = \frac{4}{10\%}$$

$$\therefore D_9 - 16 = \frac{5\%}{10\%} \cdot 4 = 2$$

$$\therefore D_9 = 16 + 2 = 18$$

Para entendermos melhor o significado dos quantis calculados, criamos um esquema:



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.

DIRETO DO CONCURSO

QUESTÃO 14 (CESGRANRIO/TRANSPETRO/ADMINISTRADOR JÚNIOR/2012) 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

COMENTÁRIO

Letra e.

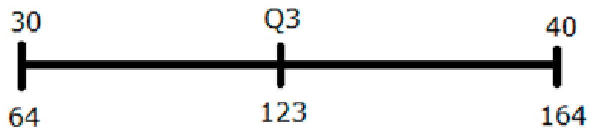
Uma questão bastante ilustrativa sobre quartis em dados categorizados. Como a amostra é formada por 164 elementos, o terceiro quartil será associado à posição:

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

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:

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

$$\therefore Q_3 - 30 = 0{,}1.59 = 5{,}9 \therefore Q_3 = 30 + 5{,}9 = 35{,}9$$

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, considero um dos mais importantes no meio profissional e, por isso, as chances de ser cobrado 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:

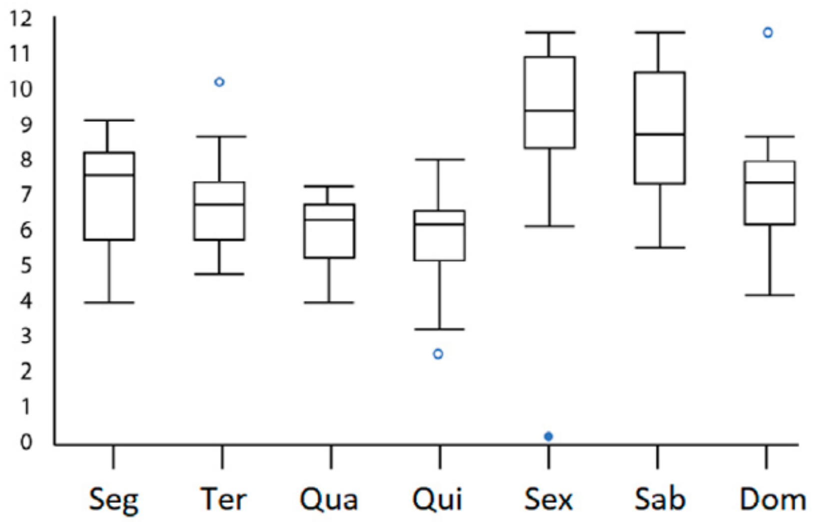


Figura 32: Diagrama Box Plot sobre o número de horas dormidas por um grupo de pessoas

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

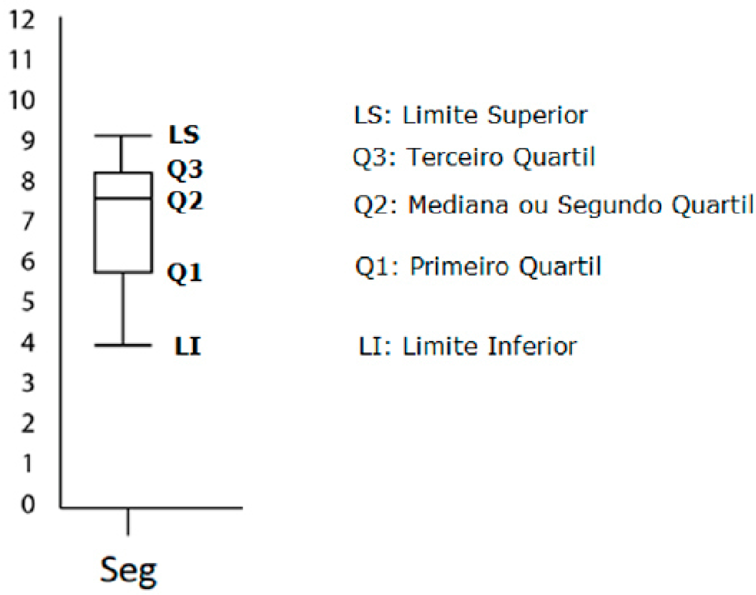


Figura 33: Composição da Caixa do Box Plot

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

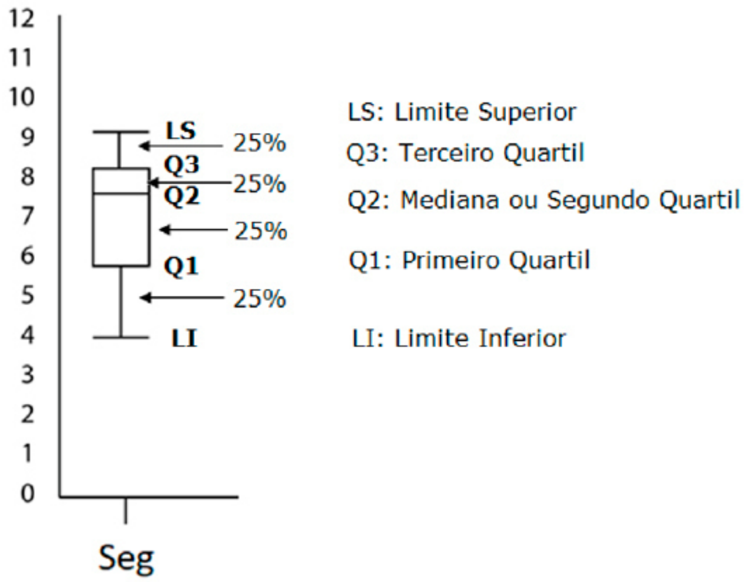


Figura 34: Box Plot evidenciando os quartis

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 os quartis da amostra, incluindo a mediana (Q_2), que não mostra a média.

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.

Obs.: $D = Q_3 - Q_1$

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:

Obs.:

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

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

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, podem-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 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 hora. 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, porque 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*.

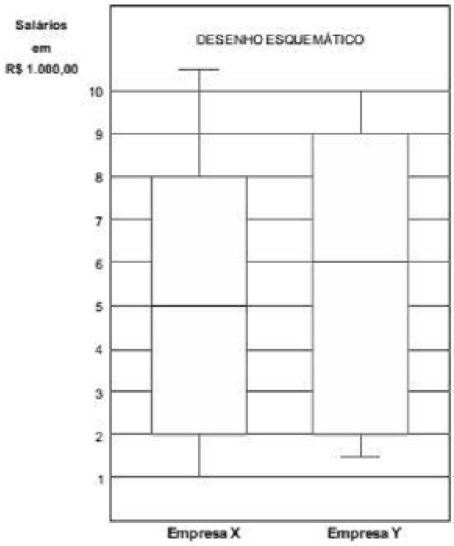
Esse fato, com certeza, é estranho, e o pesquisador pode se interessar por entender o que causou esse comportamento. Em outros casos, a presença de *outliers* pode ser, por si só, de grande interesse. Por exemplo, numa pesquisa sobre a margem de lucro de empresas do mesmo setor, se você encontrar um *outlier* positivo, essa empresa mais eficiente pode ser um alvo de estudos.

O que faz com 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

Questão 15 (FCC/TRT-19ª REGIÃO-AL/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2014) Para comparar os salários dos empregados de duas empresas X e Y, considerou-se o desenho esquemático abaixo com os valores dos salários em R$ 1.000,00.



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

a) o número de empregados da empresa X é maior que o número de empregados da empresa Y.

b) a distância interquartil da empresa Y é superior à distância interquartil da empresa X.

c) o menor salário verificado tanto na empresa X como na empresa Y é igual a R$ 2.000,00.

d) o valor da mediana da empresa Y supera o da empresa X em 10%.

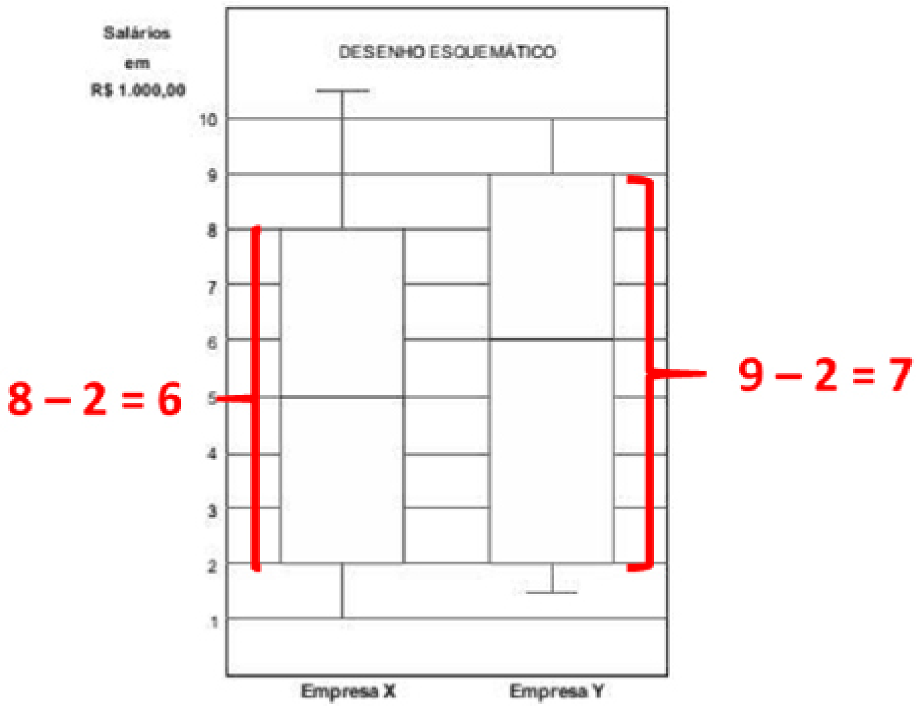
e) o maior salário verificado na empresa X é inferior ao maior salário verificado na empresa Y.

COMENTÁRIO

Letra b.

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

A distância interquartil corresponde ao comprimento da caixa do *Box Plot*.



Notemos que realmente a empresa Y apresenta uma distância interquartílica maior. Afirmação correta.

a) Errada. Não há como afirmar olhando somente para o gráfico. O diagrama *box plot* não fornece nenhuma informação sobre o número de empregados da empresa X.

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.

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.

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, prefere-se 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

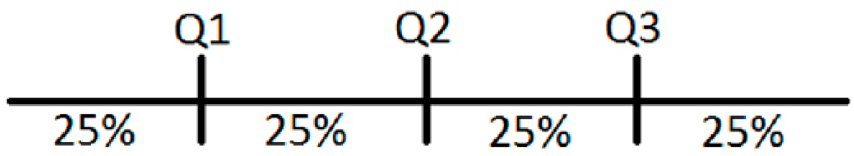
Ramos e Folhas: é um caso especial de histograma que conserva os dados originais.

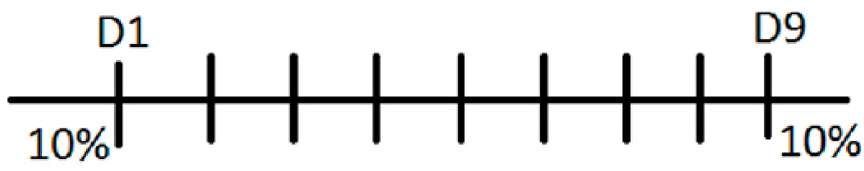
Gráfico de Linha: é utilizado quando uma variável só pode assumir um valor em função da outra. Por exemplo, renda *per capita* do Brasil e tempo. Em um determinado ano, a renda *per capita* do Brasil só pode ter um único valor.

Gráfico de Dispersão: é utilizado quando uma variável pode assumir diversos valores 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.

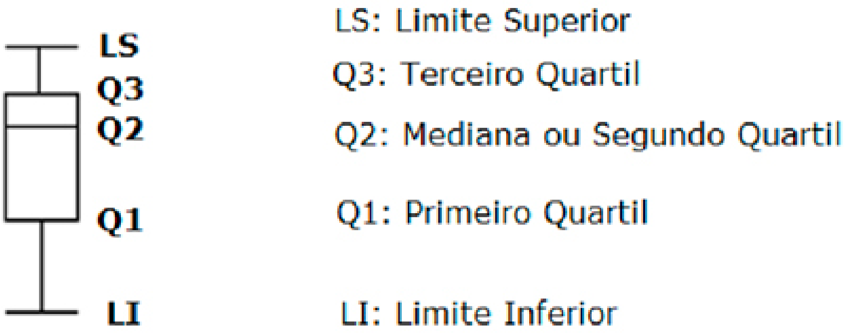
Separatrizes:

- 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).

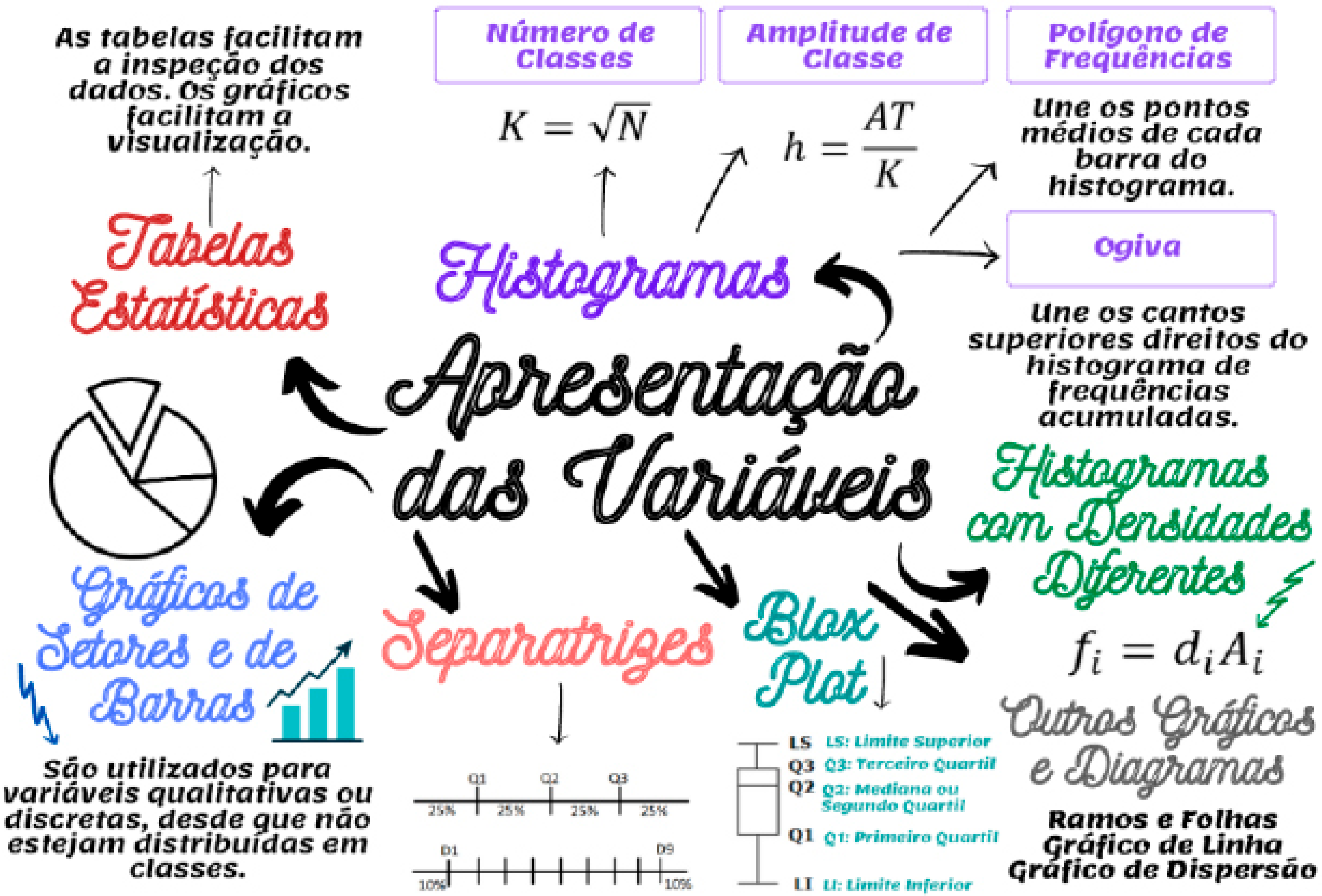


Os limites inferiores e superiores distam 1,5 intervalos interquartílicos do primeiro quartil (Q_1) e do terceiro quartil (Q_3) respectivamente.

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

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

MAPA MENTAL


As tabelas facilitam a inspeção dos dados. Os gráficos facilitam a visualização.
Tabelas Estatísticas
Número de Classes
$K = \sqrt{N}$
Amplitude de Classe
$h = \frac{AT}{K}$
Polígono de Frequências
Une os pontos médios de cada barra do histograma.
Ogiva
Une os cantos superiores direitos do histograma de frequências acumuladas.
Histogramas
Apresentação das Variáveis
Histogramas com Densidades Diferentes
$f_i = d_i A_i$
Gráficos de Setores e de Barras
São utilizados para variáveis qualitativas ou discretas, desde que não estejam distribuídas em classes.
Separatrizes
Q1 Q2 Q3
25% 25% 25% 25%
D1 D9
10% 10%
Blox Plot
LS LS: Limite Superior
Q3 Q3: Terceiro Quartil
Q2 Q2: Mediana ou Segundo Quartil
Q1 Q1: Primeiro Quartil
LI LI: Limite Inferior
Outros Gráficos e Diagramas
Ramos e Folhas
Gráfico de Linha
Gráfico de Dispersão

QUESTÕES COMENTADAS EM AULA

Questão 1 (CESPE/ANTAQ/TÉCNICO EM REGULAÇÃO/2009) 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.

Questão 2 (CESPE/ABIN/OFICIAL TÉCNICO DA INTELIGÊNCIA – ÁREA 4/2018) 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.

Questão 3 (CESPE/ABIN/OFICIAL TÉCNICO DA INTELIGÊNCIA/ÁREA 4/2018) A menor diferença entre as quantidades de professores dos anos iniciais do ensino fundamental e do ensino médio ocorreu em 2014.

Questão 4 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

Questão 5 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

Questão 6 (CESPE/AL-CE/ANALISTA LEGISLATIVO/2011) 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.

Questão 7 (CESPE/TRE-ES/ANALISTA/2011) Com base na tabela acima, referente às eleições de 2010, que apresenta a quantidade de candidatos para os cargos de presidente da

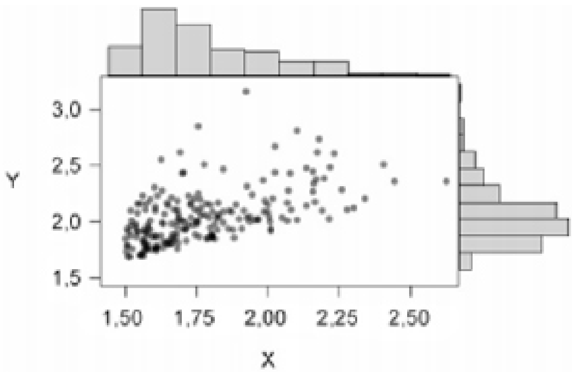
República, governador de estado, senador, deputado federal e deputado estadual/distrital, bem como a quantidade de candidatos considerados aptos pela justiça eleitoral e o total de eleitos para cada cargo pretendido, julgue os itens a seguir.

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> (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 120º.

Questão 8 (CESPE/MPU201/ANALISTA/ESTATÍSTICA/2013) 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.

Questão 9 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

Questão 10 (FCC/TRT-14ª REGIÃO-RO/AC/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 × 10-4(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 × 10^{-4}(R$)$^{-1}$, tem-se que o correspondente número de empregados é igual a:

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

Questão 11 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4/2015) 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.

Questão 12 (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 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

Questão 13 (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 os itens que se seguem, com base nessas informações.

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

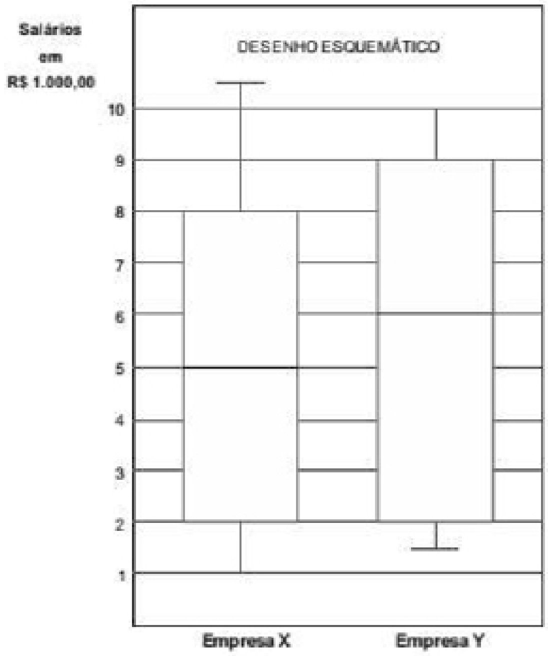
O diagrama de dispersão permite representar corretamente a distribuição de frequências da variável X.

Questão 14 (CESGRANRIO/TRANSPETRO/ADMINISTRADOR JÚNIOR/2012) 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

Questão 15 (FCC/TRT-19ª REGIÃO-AL/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2014) Para comparar os salários dos empregados de duas empresas X e Y, considerou-se o desenho esquemático abaixo com os valores dos salários em R$ 1.000,00.



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

a) o número de empregados da empresa X é maior que o número de empregados da empresa Y.

b) a distância interquartil da empresa Y é superior à distância interquartil da empresa X.

c) o menor salário verificado tanto na empresa X como na empresa Y é igual a R$ 2.000,00.

d) o valor da mediana da empresa Y supera o da empresa X em 10%.

e) o maior salário verificado na empresa X é inferior ao maior salário verificado na empresa Y.

QUESTÕES DE CONCURSO

Questão 16 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

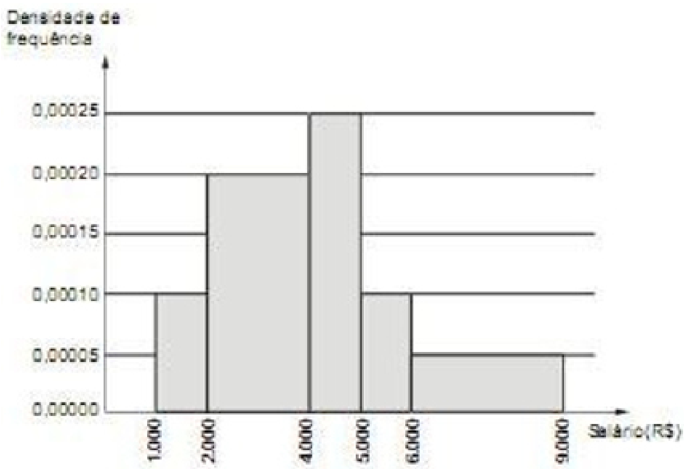
categoria	frequência	percentual (%)
ruim	50	C
regular	A	10
bom	100	D
excelente	B	E

Acerca dessa tabela, julgue o item subsequente.

Na tabela, a letra C corresponde a 20%.

Questão 17 (CESPE/ABIN/OFICIAL TÉCNICO DE INTELIGÊNCIA/ÁREA 4/2018) 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.

Questão 18 (FCC/TRF-2ª REGIÃO/ANALISTA JUDICIÁRIO/2012) 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%

Questão 19 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

Questão 20 (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.

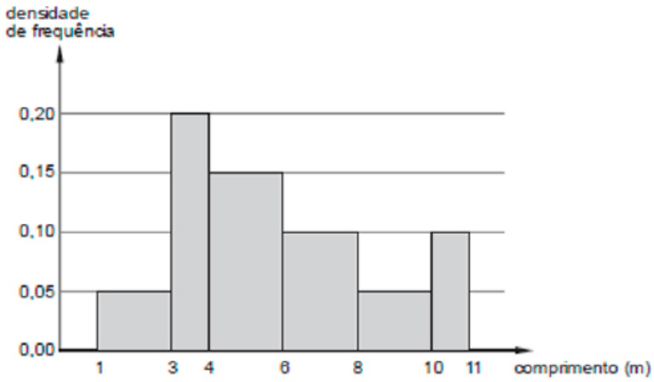
estatística	quantidade de acidentes (N)
mínimo	0
primeiro quartil	2
segundo quartil	4
terceiro quartil	10
máximo	30

O intervalo interquartílico da variável N foi igual a 8.

Questão 21 (CESPE/IPHAN/ANALISTA I/ÁREA 2/2018) Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.
O valor do primeiro quartil do conjunto de dados (Q1/4) é igual a 3.

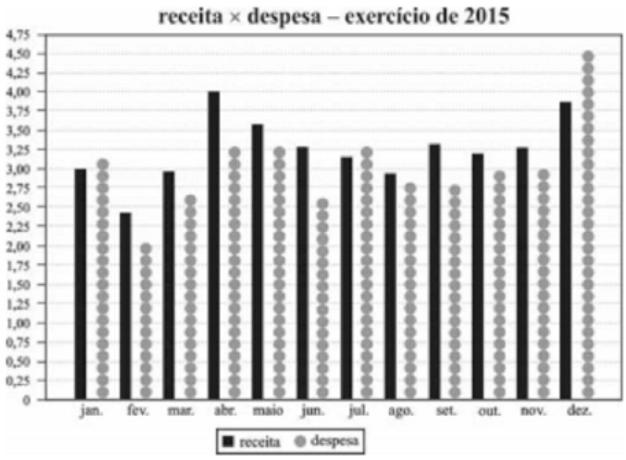
Questão 22 (FCC/TRT-5ª REGIÃO/BA/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2018) A distribuição das medidas em metros (m) dos comprimentos dos cabos no estoque de uma fábrica está representada pelo histograma mostrado abaixo, em que no eixo vertical constam as densidades de frequências, em $(m)^{-1}$, e no eixo 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%.

Questão 23 (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.gestaodinheiropublico.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.

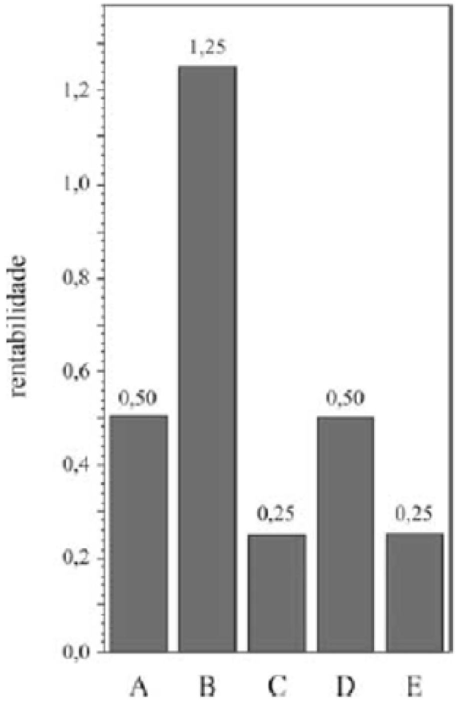
b) No primeiro quadrimestre de 2015, houve um período de queda simultânea dos gastos com despesas e da arrecadação de receitas e dois períodos de aumento simultâneo de gastos e de arrecadação.

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.

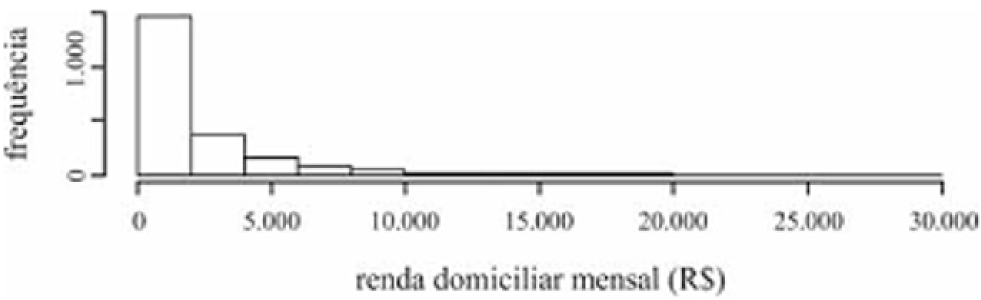
e) A menor arrecadação mensal de receitas e o menor gasto mensal com despesas foram verificados, respectivamente, no primeiro e no segundo semestre do ano de 2015.

Questão 24 (CESPE/FUNPRESP-EXE/ANALISTA/ÁREA DE INVESTIMENTOS/2016) 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.

Questão 25 (CESPE/TELEBRAS/ANALISTA SUPERIOR/AUDITORIA/2015) 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.



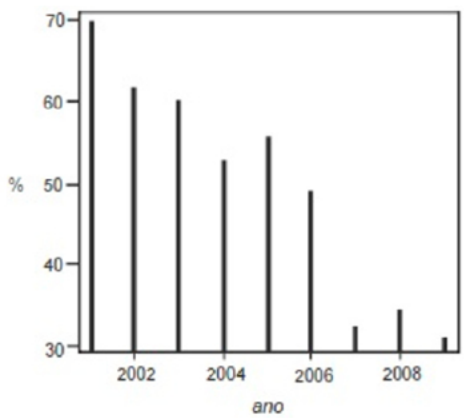
Com base nessas informações e no histograma, julgue o item a seguir.

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.

Questão 26 (FCC/TRT-3ª REGIÃO-MG/ANALISTA JUDICIÁRIO ESTATÍSTICA/2015) 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

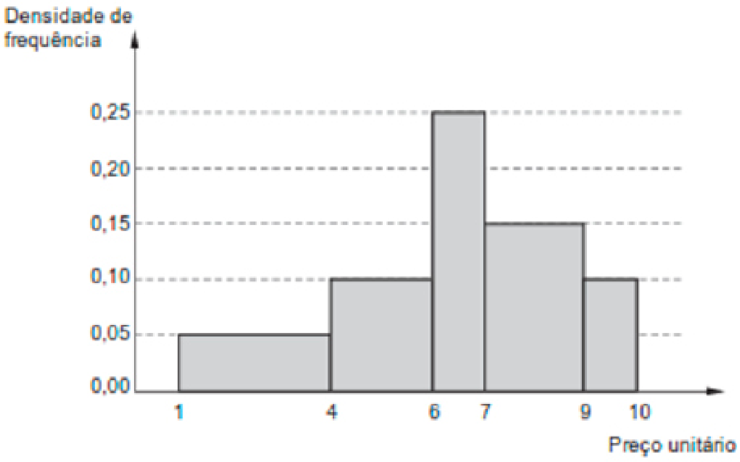
a) 150.
b) 120.
c) 240.
d) 160.
e) 96.

Questão 27 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4/2015) Se os percentuais forem representados por barras verticais, conforme o gráfico a seguir, então o resultado será denominado histograma.



QUESTÃO 28 (FCC/TRE-RR/ANALISTA JUDICIÁRIO/2015) O histograma abaixo representa a distribuição dos preços unitários de custo, em R$, de determinado equipamento de informática no mercado. No eixo das abscissas constam os intervalos de classe, em R$, e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)^{-1}$.

Observação: Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.



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

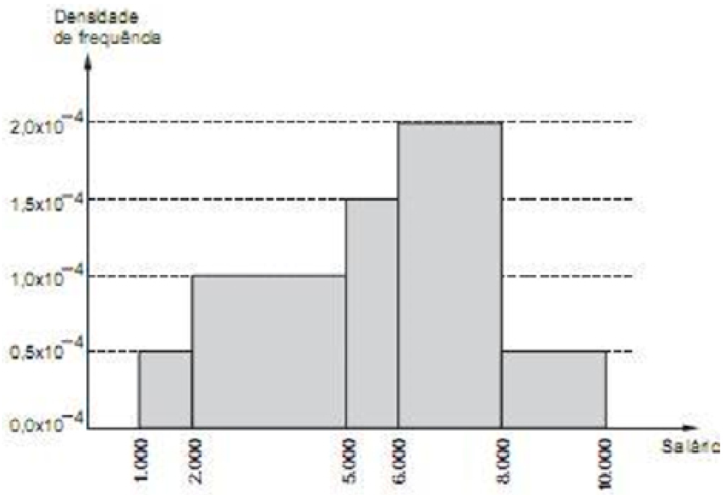
a) 240.
b) 195.
c) 215.
d) 230.
e) 255.

QUESTÃO 29 (FCC/SEFAZ-PI/AUDITOR-FISCAL DA FAZENDA ESTADUAL/2015/ADAPTADA) O histograma é um gráfico apropriado para representar dados de variáveis quantitativas contínuas.

QUESTÃO 30 (CESPE/SEE-AL/SECRETÁRIO ESCOLAR/2013) 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.

Questão 31 (FCC/TRT-13ª REGIÃO-PB/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2014) 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.

Questão 32 (CESGRANRIO/IBGE/ANALISTA DE SISTEMAS/2010) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças. Julgue os itens 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

Questão 33 (CESGRANRIO/IBGE/ANALISTA DE SISTEMAS/2010) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

Julgue os itens 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

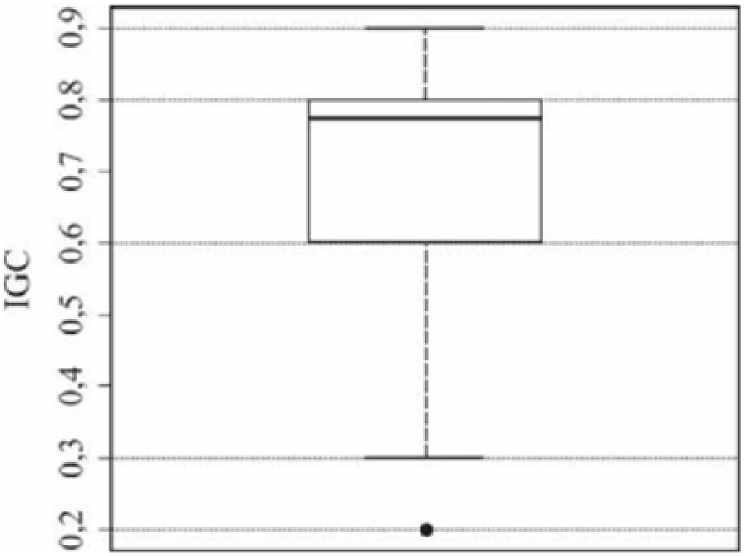
b) 5,6

c) 5,7

d) 5,8

e) 5,9

(CESPE/TCE-PR/201/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.



Questão 34 O menor IGC observado na amostra foi 0,3.

Questão 35 O maior IGC observado na amostra foi 0,9.

Questão 36 O diagrama mostrado na figura em questão é denominado curva de frequência.

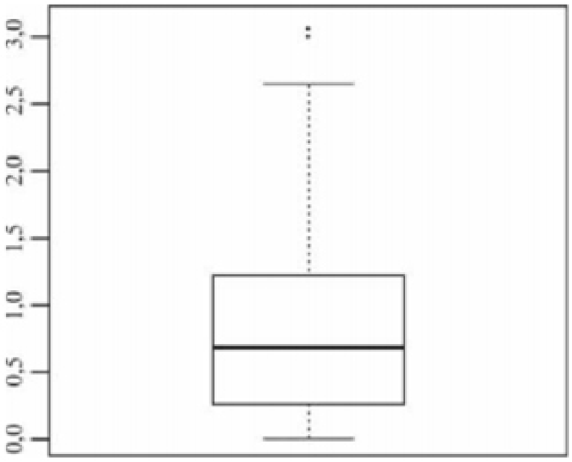
Questão 37 O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

Questão 38 Na amostra considerada, a média dos indicadores observados foi superior a 0,7 e inferior a 0,8.

Questão 39 A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

Questão 40 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.

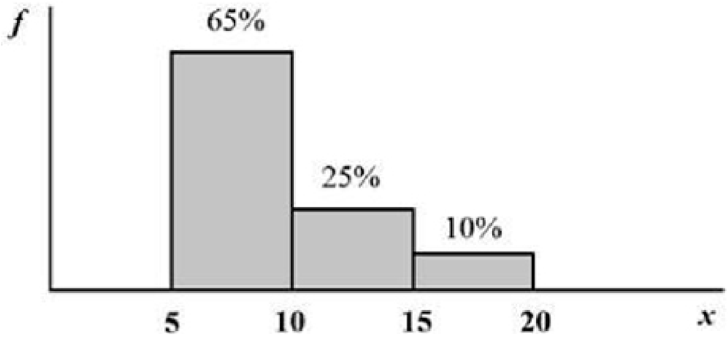
Questão 41 O diagrama *box plot* mostrado na figura sugere a existência de pelo menos duas observações atípicas.

Questão 42 *X* representa uma variável qualitativa ordinal.

Questão 43 A amplitude total da amostra é inferior a 3.

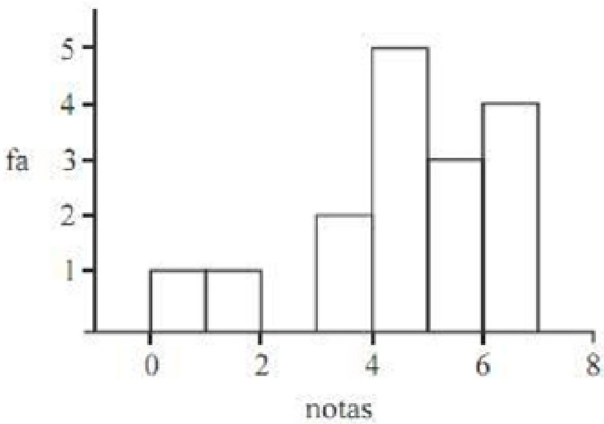
Questão 44 Pelo menos 50% das observações de X são menores que 1,5.

Questão 45 (CESPE/STF/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2013) 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.

Questão 46 (CESPE/FUB/ESTATÍSTICO/2013) 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 os seguintes itens acerca de estatística descritiva.



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

Questão 47 (CESPE/FUB/ESTATÍSTICO/2013) A amplitude total da distribuição das notas é igual a 6.

Questão 48 (CESPE/BACEN/ANALISTA DE INFRAESTRUTURA E LOGÍSTICA/2013)

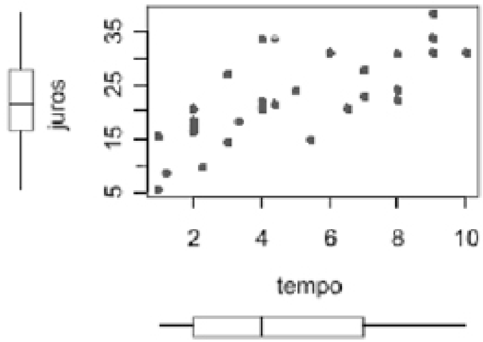
2 4 8 4 8 1 2 32 12 1 5 7 5 5 3 4 24 19 4 14

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários.

Com base nesses dados, julgue os próximos itens.

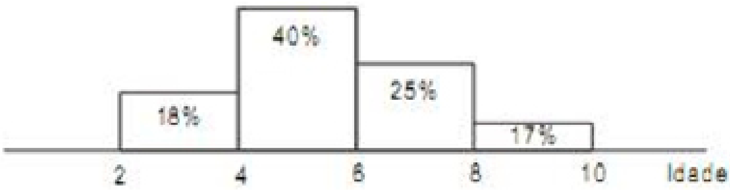
É inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos; dessa forma, apenas por meio de um gráfico de barras pode ser realizada a representação gráfica.

Questão 49 (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.

Questão 50 (FCC/ANS/ESPECIALISTA EM REGULAÇÃO/2007) 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

Questão 51 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/2015) 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.

Questão 52 (CESPE/MEC/ANALISTA DE POLÍTICA REGULATÓRIA/2014) 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.

Questão 53 (CESPE/CEHAP-PB/ADMINISTRADOR/2009) 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.

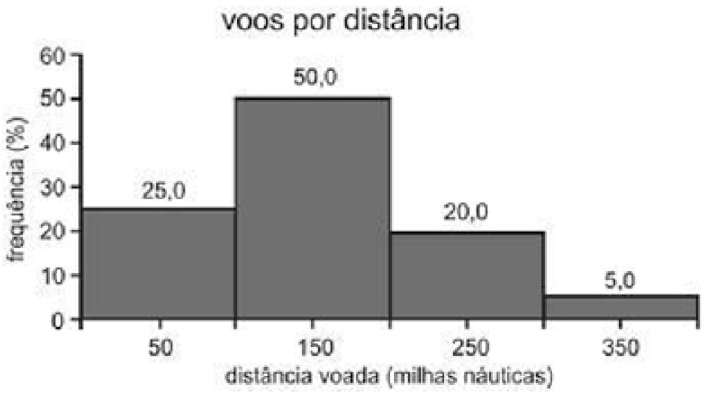
a) A correlação linear de Pearson entre X e Y é positiva.

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.

Questão 54 (CESPE/ANAC/ESPECIALISTA EM REGULAÇÃO DE AVIAÇÃO CIVIL/2012) Com base no histograma acima apresentado, julgue os itens 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.

Questão 55 (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 (Md) e o terceiro quartil (Q_3) da distribuição.

Se Md = 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.

GABARITO

1. C
2. C
3. C
4. E
5. C
6. E
7. E
8. E
9. C
10. d
11. E
12. a
13. E
14. e
15. b
16. E
17. C
18. c
19. E
20. C
21. E
22. b
23. b
24. E
25. E
26. a
27. E
28. e
29. C
30. C
31. c
32. C
33. a
34. E
35. C
36. E
37. E
38. E
39. C
40. E
41. C
42. E
43. E
44. C
45. E
46. C
47. E
48. E
49. E
50. d
51. C
52. E
53. c
54. C
55. b

GABARITO COMENTADO

QUESTÃO 16 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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%.

Errado.

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%.

$$C = \frac{50}{200} \cdot 100\% = 0{,}25.100\% = 25\%$$

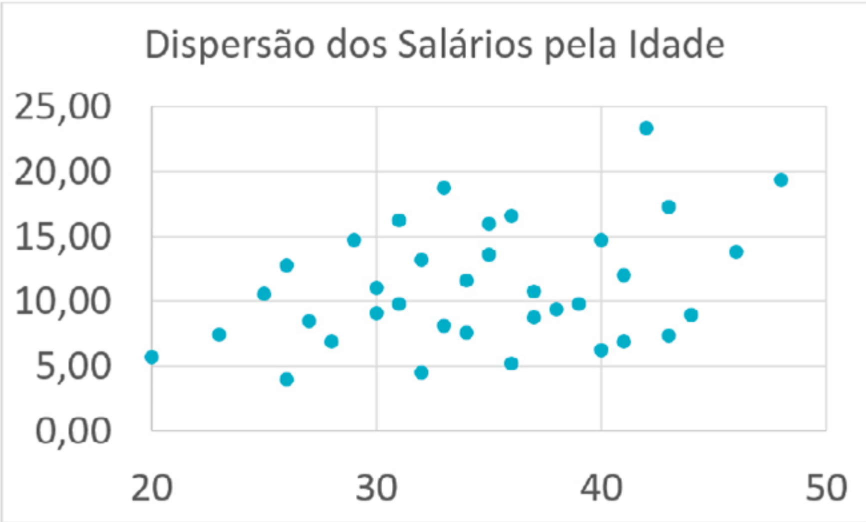
Podemos, ainda, completar a tabela.

Categoria	Frequência	Percentual
Ruim	50	25%
Regular	20	10%
Bom	100	50%
Excelente	30	15%
Total	200	

QUESTÃO 17 (CESPE/ABIN/OFICIAL TÉCNICO DE INTELIGÊNCIA/ÁREA 4/2018) 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.

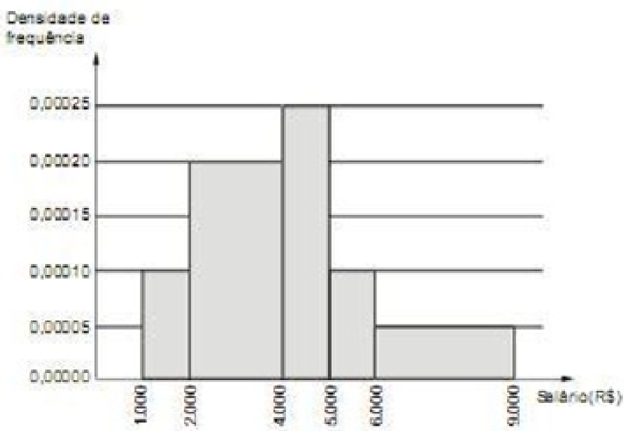
Certo.

Vejamos 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.

Questão 18 (FCC/TRF-2ª REGIÃO/ANALISTA JUDICIÁRIO/2012) 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%

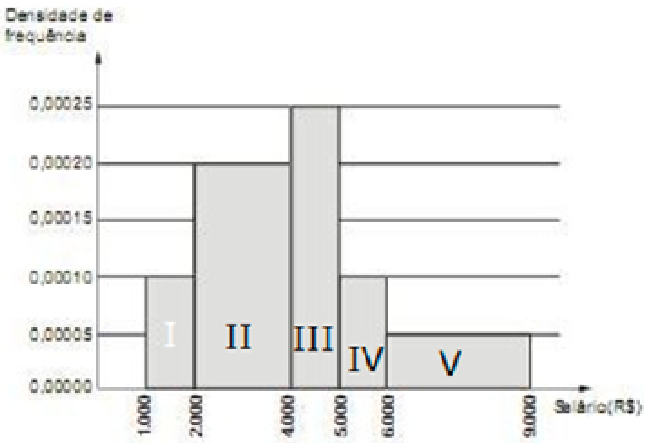
c) 70% < P ≤ 75%

d) 75% < P ≤ 80%

e) P > 80%

Letra c.

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.

$$p_{II} = d_{II}A_{II} = 0{,}00020.(4000 - 2000) = 0{,}0002.2000 = 0{,}40$$

$$p_{III} = d_{III}A_{III} = 0{,}00025.(5000 - 3000) = 0{,}00025.1000 = 0{,}25$$

$$p_{IV} = d_{IV}A_{IV} = 0{,}00010.(6000 - 5000) = 0{,}00010.1000 = 0{,}10$$

$$\therefore P = p_{ii} + p_{III} + p_{IV} = 0{,}40 + 0{,}25 + 0{,}10 = 0{,}75 = 75\%$$

Questão 19 (CESPE/IPHAN/ANALISTA 1/ÁREA 2/2018) 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.

Errado.

Notemos que os centros com frequência regular correspondem a 10% do total. Portanto, a frequência absoluta correspondente é:

$$A = 0{,}10.200 = 20$$

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.

Questão 20 (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.

Certo.

O intervalo interquartílico corresponde à diferença entre o terceiro e o primeiro quartis.

$$D = Q_3 - Q_1 = 10 - 2 = 8$$

Questão 21 (CESPE/IPHAN/ANALISTA I/ÁREA 2/2018) Uma pesquisa a respeito das quantidades de teatros em cada uma de 11 cidades brasileiras selecionadas apresentou o seguinte resultado: {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4}.

Com referência a esses dados, julgue o item seguinte.

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

Errado.

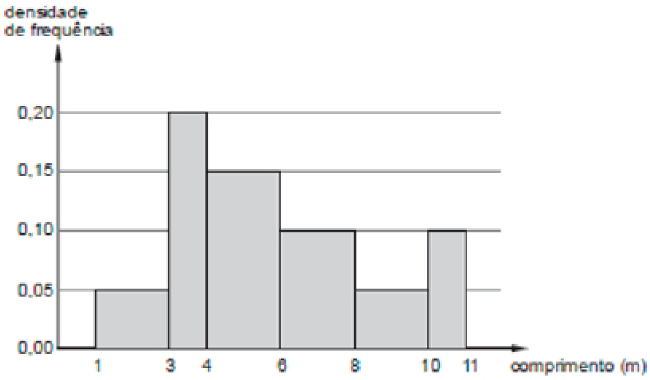
Considerando que a amostra tem 11 cidades, o primeiro quartil deve ser calculado de forma semelhante à mediana.

$$\frac{1+11}{4} = \frac{12}{4} = 3$$

Portanto, o primeiro quartil corresponde ao elemento:

$$Q_1 = x_3 = 2$$

Questão 22 (FCC/TRT-5ª REGIÃO/BA/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2018) 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.

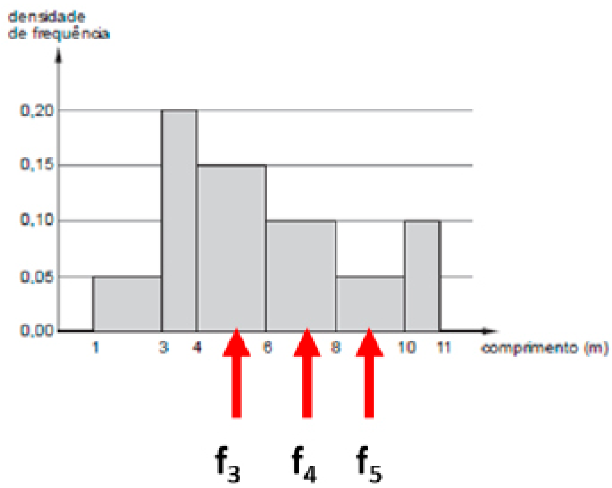


Sabendo-se que todos os intervalos de classe são fechados à esquerda e abertos à direita, então a porcentagem dos cabos que apresentam uma medida de comprimento de pelo menos igual a 4 m e inferior a 10 m é de

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

Letra b.

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.

$$f_3 = d_3 A_3 = 0{,}15.(6-4) = 0{,}15.2 = 0{,}30$$

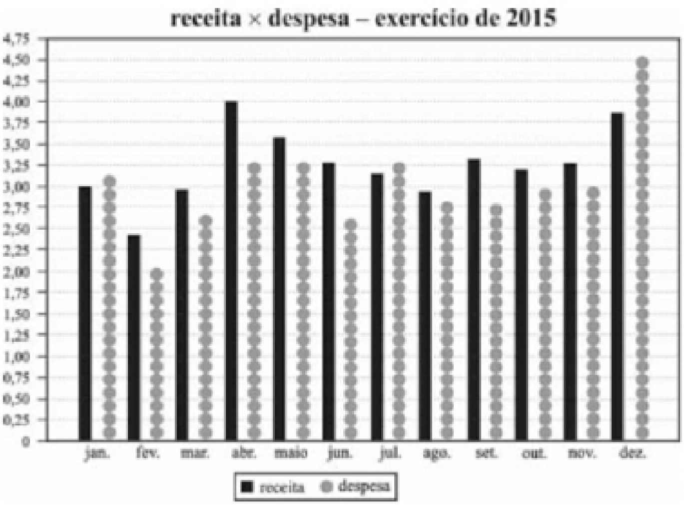
$$f_4 = d_4 A_4 = 0{,}10.(8-6) = 0{,}10.2 = 0{,}20$$

$$f_5 = d_5 A_5 = 0{,}05.(10-8) = 0{,}05.2 = 0{,}10$$

Portanto, a frequência acumulada total das três classes desejadas é:

$$F = f_3 + f_4 + f_5 = 0{,}30 + 0{,}20 + 0{,}10 = 0{,}60$$

Questão 23 (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.gestaodinheiropublico.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.

Letra b.

O primeiro quadrimestre é composto pelos meses de janeiro a abril. Vamos desenhar as ogivas referentes à arrecadação de receitas e de despesas.

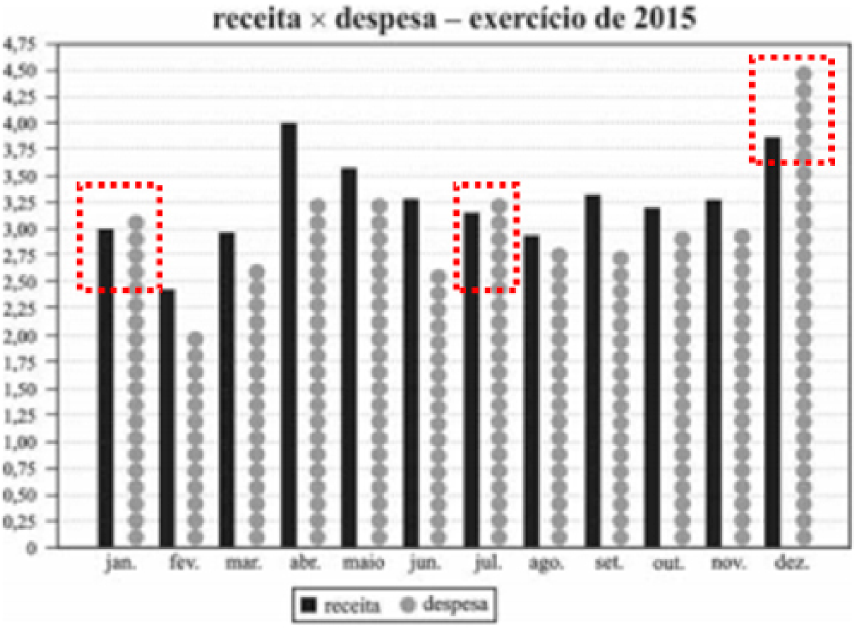
Portanto, vemos um período de redução de janeiro para fevereiro e dois períodos de crescimento tanto nas receitas como nas despesas.

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.

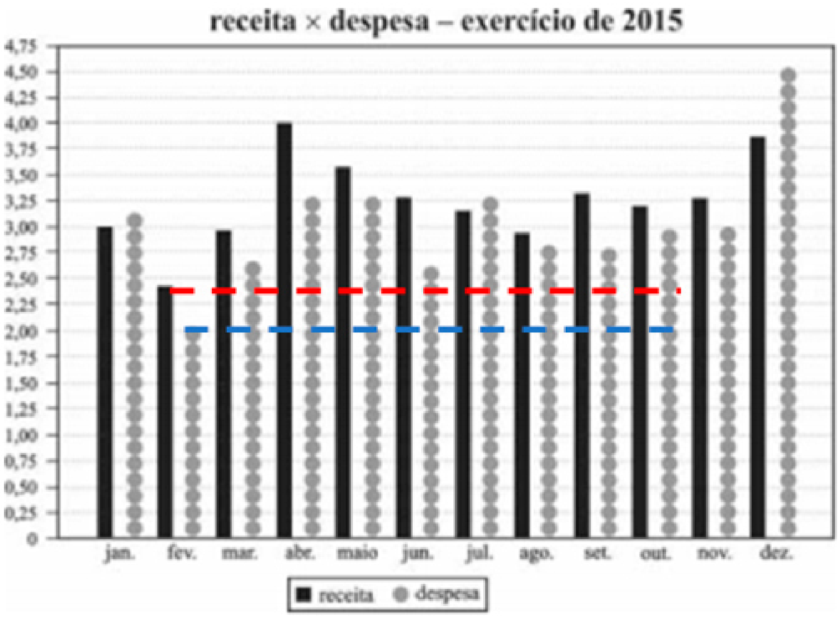
Podemos observar, portanto, que no terceiro trimestre não houve queda contínua na arrecadação de receitas, porque, no mês de agosto, houve aumento em relação ao mês anterior.

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.



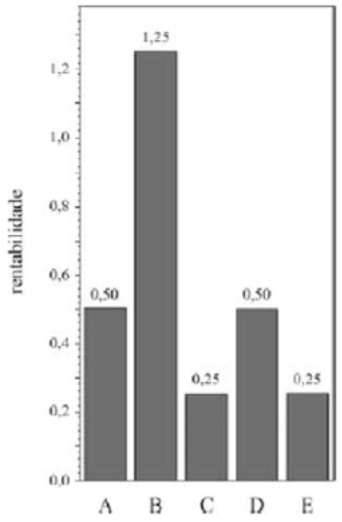
e) Errada. Vamos identificar o mínimo da receita mensal e da despesa mensal.



Portanto, os valores mínimos tanto da receita como da despesa aconteceram no mês de fevereiro, que pertence ao primeiro semestre.

Questão 24 (CESPE/FUNPRESP-EXE/ANALISTA/ÁREA DE INVESTIMENTOS/2016) 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.

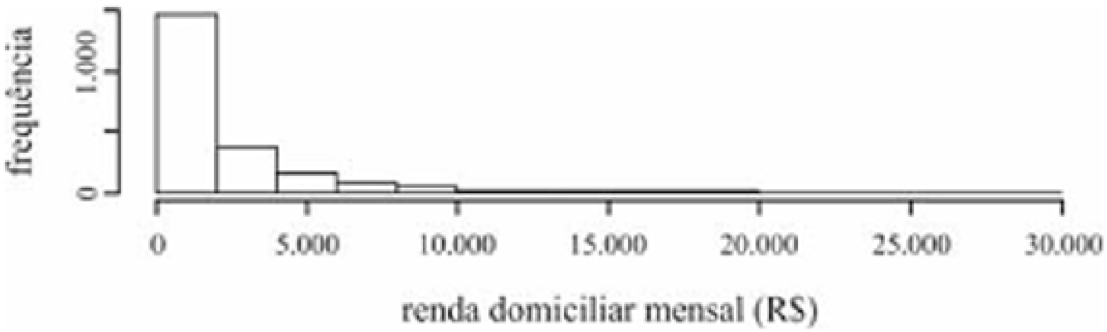


O gráfico apresentado é um histograma.

Errado.

O histograma se refere a uma distribuição de frequências de uma variável quantitativa contínua. Esse gráfico não mostra uma distribuição de frequências, mas a rentabilidade dos fundos. Portanto, esse não pode ser um histograma, mas um gráfico de barras genérico.

Questão 25 (CESPE/TELEBRAS/ANALISTA SUPERIOR/AUDITORIA/2015) 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.

Errado.

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.

Questão 26 (FCC/TRT-3ª REGIÃO/MG/ANALISTA JUDICIÁRIO ESTATÍSTICA/2015) 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.

Letra a.

Como vimos, a frequência relativa de uma classe é igual ao produto da densidade de frequência pela amplitude da classe.

$$f = dA = 12{,}8.10^{-5}.2500 = 32000.10^{-5}$$

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.

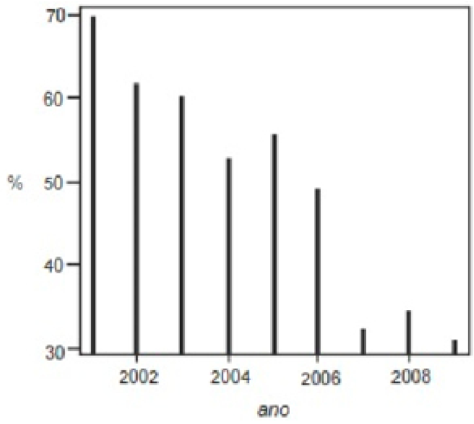
$$f = \frac{32000}{100000} = 0{,}32$$

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.

$$0{,}32.N = 48$$

$$\therefore N = \frac{48}{0{,}32} = \frac{4800}{32} = 150$$

Questão 27 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/ÁREA 4/2015) Se os percentuais forem representados por barras verticais, conforme o gráfico a seguir, então o resultado será denominado histograma.

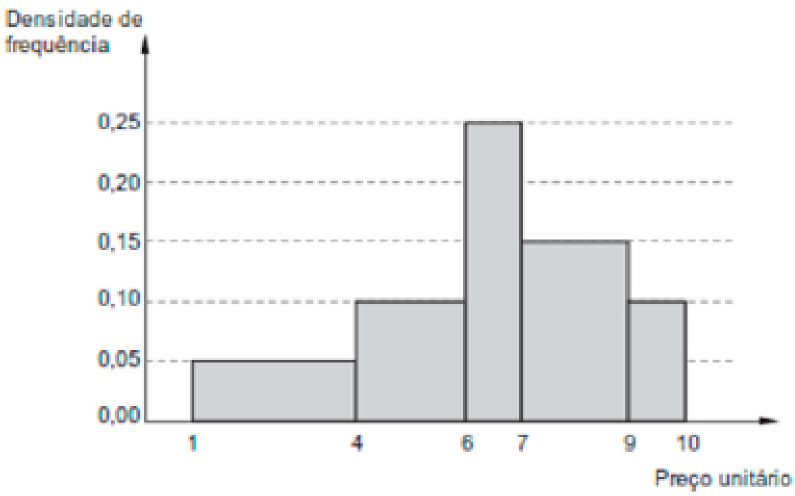


Errado.

O histograma é utilizado quando ambas as variáveis aleatórias são contínuas. Neste caso, o ano é uma variável discreta, portanto, o gráfico apresentado é um diagrama de barras.

QUESTÃO 28 (FCC/TRE-RR/ANALISTA JUDICIÁRIO/2015) O histograma abaixo representa a distribuição dos preços unitários de custo, em R$, de determinado equipamento de informática no mercado. No eixo das abscissas constam os intervalos de classe, em R$, e no eixo das ordenadas as respectivas densidades de frequências em $(R\$)^{-1}$.

Observação: Densidade de frequência de um intervalo é o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo.



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

a) 240.

b) 195.

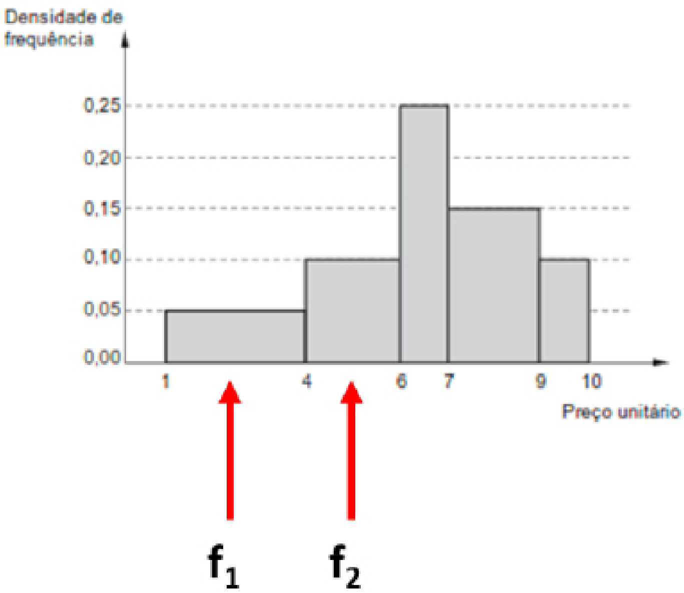
c) 215.

d) 230.

e) 255.

Letra e.

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.

$$f_1 = d_1 A_1 = 0{,}05.(4-1) = 0{,}05.3 = 0{,}15$$

$$f_2 = d_2 A_2 = 0{,}10.(6-4) = 0{,}10.2 = 0{,}20$$

Portanto, o percentual de valores menores que R$6,00 corresponde à frequência acumulada pelas duas classes.

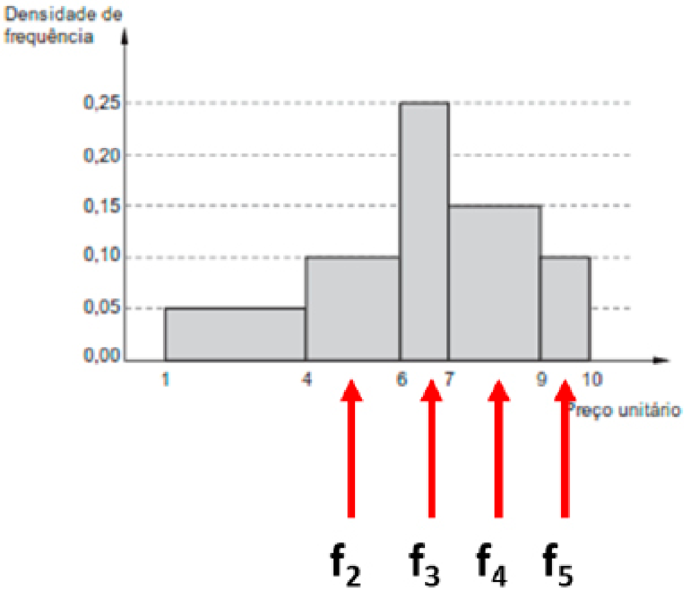
$$F_2 = f_1 + f_2 = 0{,}15 + 0{,}20 = 0{,}35$$

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:

$$0{,}35.N = 105$$

$$\therefore N = \frac{105}{0{,}35} = \frac{10500}{35} = 300$$

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.

$$f_2 = d_2A_2 = 0{,}10.(6-4) = 0{,}10.2 = 0{,}20$$

$$f_3 = d_3A_3 = 0{,}25.(7-6) = 0{,}25.1 = 0{,}25$$

$$f_4 = d_4A_4 = 0{,}15.(9-7) = 0{,}15.2 = 0{,}30$$

$$f_5 = d_5A_5 = 0{,}10.(10-9) = 0{,}10.1 = 0{,}10$$

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:

$$N = 0{,}85.300 = 255$$

Questão 29 (FCC/SEFAZ-PI/AUDITOR-FISCAL DA FAZENDA ESTADUAL/2015/ADAPTADA) O histograma é um gráfico apropriado para representar dados de variáveis quantitativas contínuas.

Certo.

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.

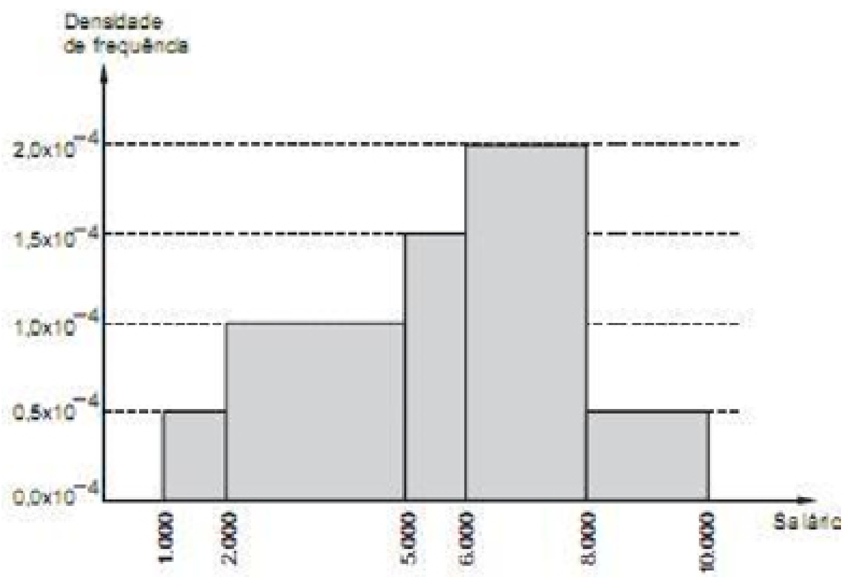
Questão 30 (CESPE/SEE-AL/SECRETÁRIO ESCOLAR/2013) Considere que, para classificar as escolas de uma cidade, segundo as notas médias em matemática e português, tenha sido obtida uma amostra das notas de cinquenta alunos de cada uma das dez escolas do município. Com base nessas informações, julgue o item subsequente.

Considerando-se que a variável nota em matemática seja quantitativa contínua, então o gráfico que representa esse tipo de dado é o histograma.

Certo.

Nunca é demais nos lembrarmos de que o histograma é utilizado para variáveis contínuas, e o diagrama de barras é utilizado para variáveis discretas.

Questão 31 (FCC/TRT-13ª REGIÃO/PB/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2014) 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.

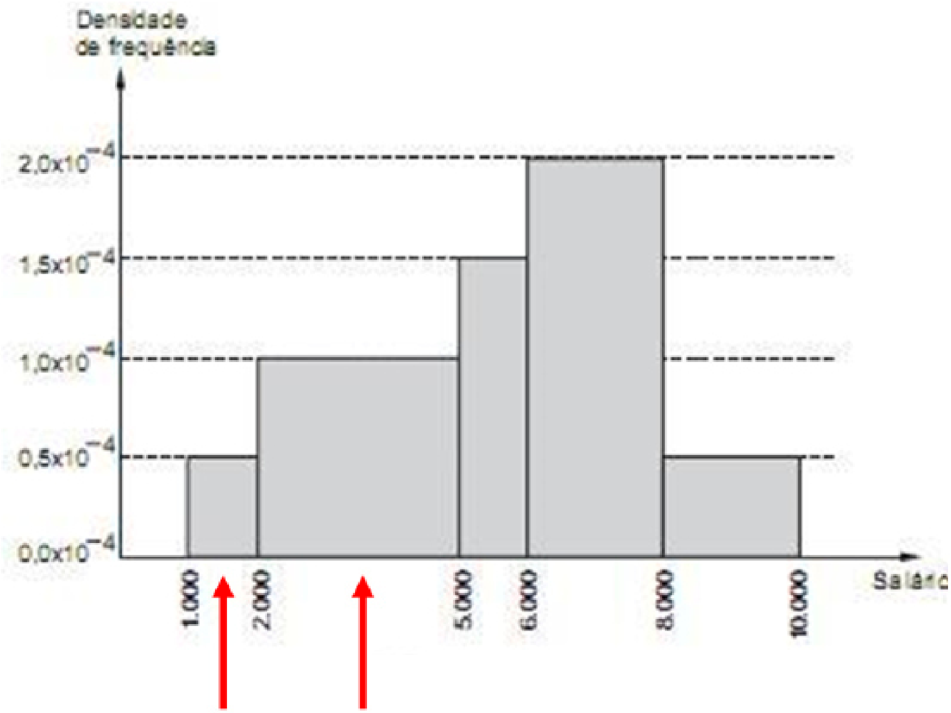
c) 136.

d) 144.

e) 120.

Letra c.

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.

$$f_1 = d_1A_1 = 0{,}5.10^{-4}.(2000 - 1000) = 0{,}5.10^{-4}.1000 = \frac{500}{10000} = 0{,}05$$

$$f_2 = d_2A_2 = 1{,}0.10^{-4}.(5000 - 2000) = 1{,}0.10^{-4}.3000 = \frac{3000}{10000} = 0{,}30$$

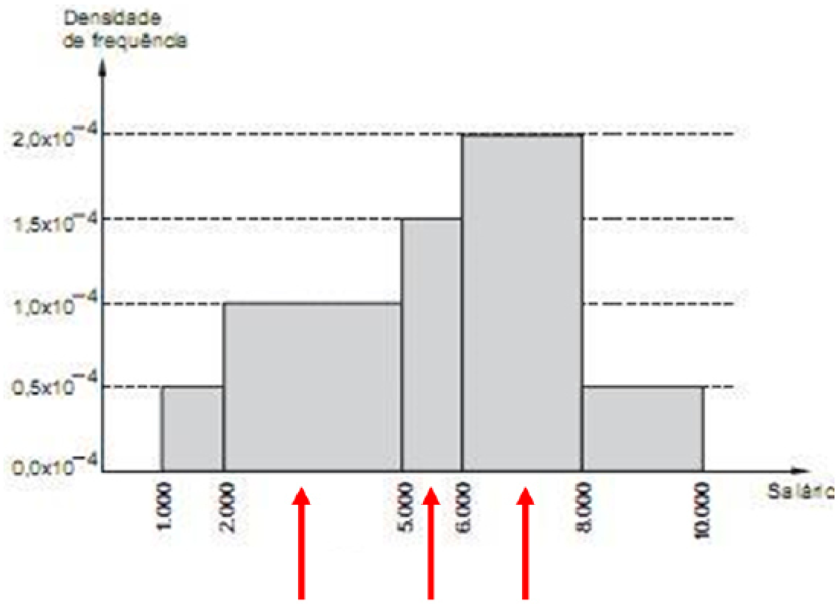
Portanto, a frequência acumulada de pessoas que ganham menos de R$5.000,00 é igual à soma das duas classes.

$$F_2 = f_1 + f_2 = 0{,}05 + 0{,}30 = 0{,}35 = 35\%$$

Como sabemos que existem 56 pessoas nessa condição, sendo N o total de funcionários, temos que 35% desse total é igual a 56.

$$0{,}35.N = 56 \therefore N = \frac{56}{0{,}35} = \frac{5600}{35} = 160$$

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.

$$f_2 = d_2A_2 = 1{,}0.10^{-4}.(5000 - 2000) = 1{,}0.10^{-4}.3000 = \frac{3000}{10000} = 0{,}30$$

$$f_3 = d_3A_3 = 1{,}5.10^{-4}.(6000 - 4000) = 1{,}5.10^{-4}.1000 = \frac{1500}{10000} = 0{,}15$$

$$f_4 = d_4A_4 = 2{,}0.10^{-4}.(8000 - 6000) = 2{,}0.10^{-4}.2000 = \frac{4000}{10000} = 0{,}40$$

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.

$$n = 0{,}85.160 = 136$$

Questão 32 (CESGRANRIO/IBGE/ANALISTA DE SISTEMAS/2010) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

Julgue os itens 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

Certo.

Quando há dados categorizados, a média pode ser calculada usando o ponto médio das classes.

$$\mu = \frac{1.5 + 2.3 + 4.5 + 2.7 + 7.9}{5 + 2 + 4 + 2 + 7} = \frac{5 + 6 + 20 + 14 + 63}{20} = \frac{108}{20} = 5{,}4$$

QUESTÃO 33 (CESGRANRIO/IBGE/ANALISTA DE SISTEMAS/2010) A tabela abaixo apresenta a distribuição de frequências das idades de um grupo de crianças.

Julgue os itens 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

Letra a.

Por outro lado, para a mediana, quando há dados categorizados, devemos usar que:

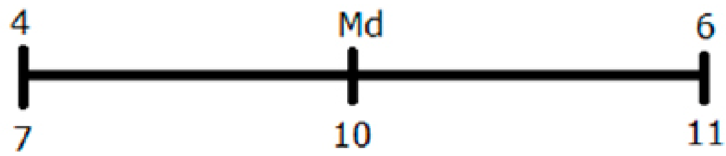
$$\frac{20}{2} = 10 \therefore Md = x_{10}$$

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**
4 a 6	4	8 a **11**
6 a 8	2	12 a 13
8 a 10	7	14 a 20

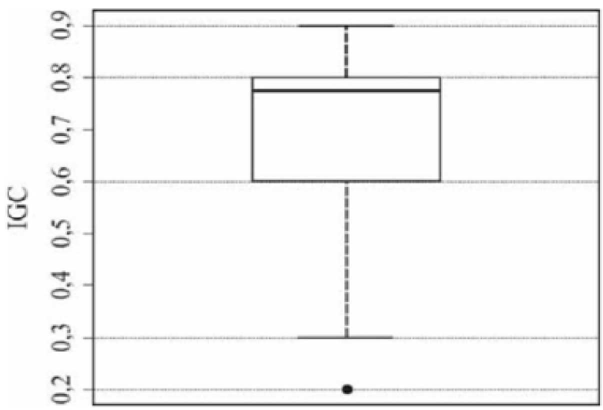
Fazendo o esquema que já conhecemos para a interpolação linear, temos:



$$\frac{Md-4}{10-7}=\frac{6-4}{11-7}=\frac{2}{4}\therefore\frac{Md-4}{3}=\frac{2}{4}$$

$$\therefore\ Md-4=\frac{2.3}{4}=1{,}5\ \therefore\ Md=4+1{,}5=5{,}5$$

(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.



Questão 34 O menor IGC observado na amostra foi 0,3.

Errado.

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.

Questão 35 O maior IGC observado na amostra foi 0,9.

Certo.

O maior IGC da amostra foi 0,9. Lembre-se de que, quando não existem *outliers*, o limite superior ou inferior é ajustado para o valor máximo ou mínimo da amostra, respectivamente.

Questão 36 O diagrama mostrado na figura em questão é denominado curva de frequência.

Errado.

Tem-se um diagrama *box plot*, não uma curva de frequência.

Questão 37 O primeiro quartil da distribuição dos indicadores foi igual a 0,3.

Errado.

A caixa do *box plot* evidencia os quartis. Portanto, o primeiro quartil foi 0,6.

Questão 38 Na amostra considerada, a média dos indicadores observados foi superior a 0,7 e inferior a 0,8.

Errado.

A mediana (e não a média) que corresponde ao segundo quartil foi superior a 0,7 e inferior a 0,8.

QUESTÃO 39 A figura em apreço sugere a existência de, pelo menos, uma observação destoante das demais.

Certo.

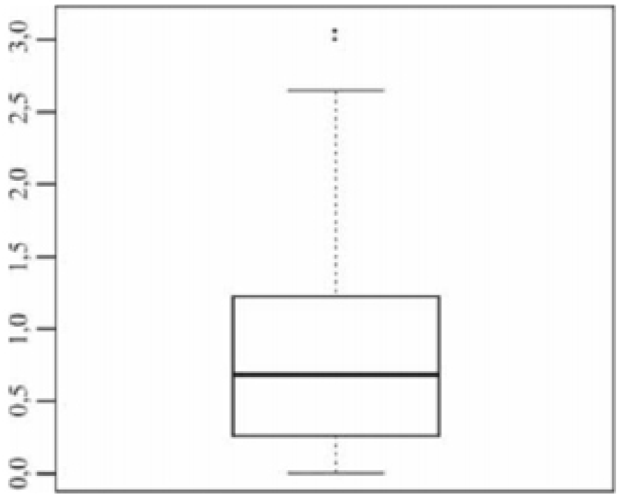
QUESTÃO 40 O desvio quartílico para a amostra em questão é de 0,3.

Errado.

Podemos calcular o desvio quartílico pela expressão:

$$D = \frac{Q_3 - Q_1}{2} = \frac{0{,}8 - 0{,}6}{2} = \frac{0{,}2}{2} = 0{,}1$$

(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.

Questão 41 O diagrama *box plot* mostrado na figura sugere a existência de pelo menos duas observações atípicas.

Certo.

No topo do diagrama, temos dois *outliers*, ou seja, duas observações atípicas.

QUESTÃO 42 *X* representa uma variável qualitativa ordinal.

Errado.

X – é uma variável quantitativa, portanto, não pode ser ordinal, que é uma classificação para variáveis qualitativas.

QUESTÃO 43 A amplitude total da amostra é inferior a 3.

Errado.

A amplitude da amostra é:

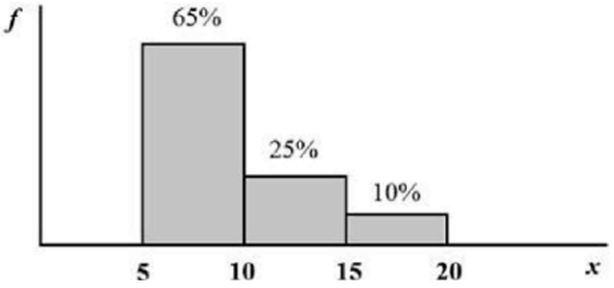
$$A = max - min = 3{,}1 - 0 = 3{,}1 > 3$$

QUESTÃO 44 Pelo menos 50% das observações de X são menores que 1,5.

Certo.

Observe que o terceiro quartil é 1,2, portanto, 75% dos elementos da amostra são inferiores a 1,2. Então, é realmente possível garantir que 50% das observações de X são menores que 1,5.

QUESTÃO 45 (CESPE/STF/ANALISTA JUDICIÁRIO/ESTATÍSTICA/2013) 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.



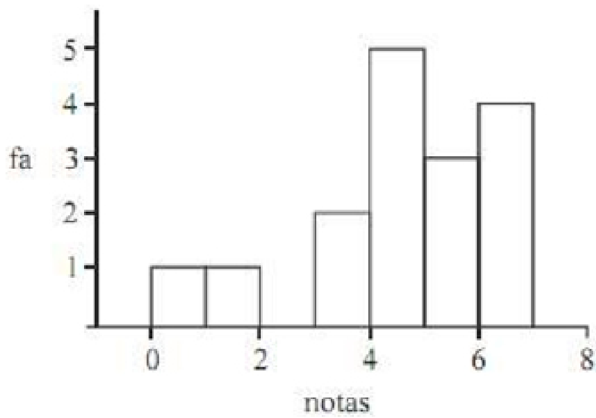
A variável x, por possuir quatro níveis de respostas, é do tipo qualitativa ordinal.

Errado.

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 ser distribuída 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.

QUESTÃO 46 (CESPE/FUB/ESTATÍSTICO/2013) 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 os seguintes itens acerca de estatística descritiva.



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

Certo.

Como o gráfico foi fornecido na forma de frequências acumuladas, primeiramente, precisamos somar todas elas para obter o total de elementos e, assim, 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.

$$f = \frac{2}{16} = \frac{1}{8} = 0{,}125 = 12{,}5\%$$

Questão 47 (CESPE/FUB/ESTATÍSTICO/2013) A amplitude total da distribuição das notas é igual a 6.

Errado.

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:

$$A = 7 - 0 = 7$$

Questão 48 (CESPE/BACEN/ANALISTA DE INFRAESTRUTURA E LOGÍSTICA/2013)

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

Os dados mostrados acima representam uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários.

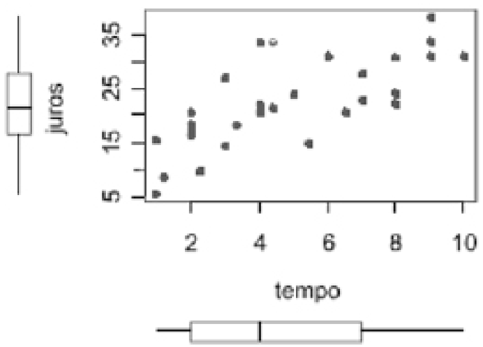
Com base nesses dados, julgue os próximos itens.

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.

Errado.

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.

QUESTÃO 49 (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.

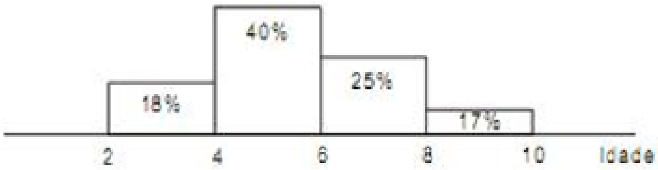


Os gráficos projetados nas margens do diagrama de dispersão são chamados de desenho esquemático ou *boxplot*, que resume graficamente a tendência central e a dispersão dos dados, sendo construído utilizando-se mínimo, máximo, média e quartis.

Errado.

Realmente, o gráfico apresentado é um diagrama *Box Plot*. Porém, esse diagrama é construído com a mediana, e não com a média.

QUESTÃO 50 (FCC/ANS/ESPECIALISTA EM REGULAÇÃO/2007) 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

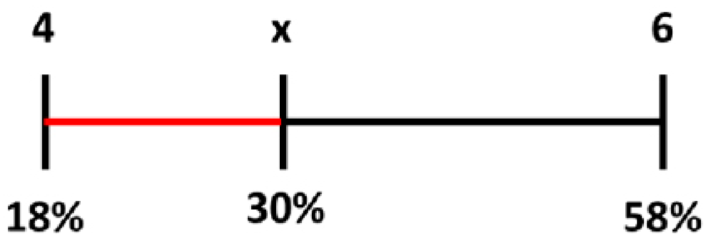
e) 5,0

Letra d.

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, usamos a ideia de que o pedaço à esquerda é proporcional ao segmento inteiro.

$$\frac{x-4}{30\%-18\%}=\frac{6-4}{58\%-18\%}$$

$$\frac{x-4}{12\%}=\frac{2}{40\%}$$

$$\therefore x-4=\frac{12\%}{40\%}\cdot 2=0{,}3.2=0{,}6$$

$$\therefore x=4+0{,}6=4{,}6$$

Questão 51 (CESPE/DEPEN/AGENTE PENITENCIÁRIO FEDERAL/2015) 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.

Certo.

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.

Questão 52 (CESPE/MEC/ANALISTA DE POLÍTICA REGULATÓRIA/2014) 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

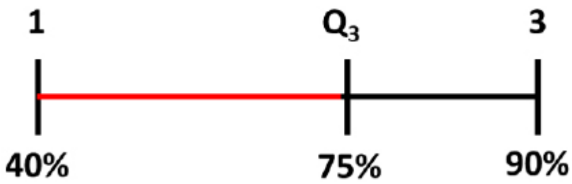
O terceiro quartil dessa distribuição de renda é superior a três salários mínimos.

Errado.

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.

$$\frac{Q_3 - 1}{75\% - 40\%} = \frac{3 - 1}{90\% - 40\%}$$

$$\frac{Q_3 - 1}{35\%} = \frac{2}{50\%}$$

$$\therefore Q_3 - 1 = \frac{35\%}{50\%} \cdot 2 = 1{,}4$$

$$\therefore Q_3 - 1 = 1{,}4 \therefore Q_3 = 1 + 1{,}4 = 2{,}4$$

Questão 53 (CESPE/CEHAP-PB/ADMINISTRADOR/2009) 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.

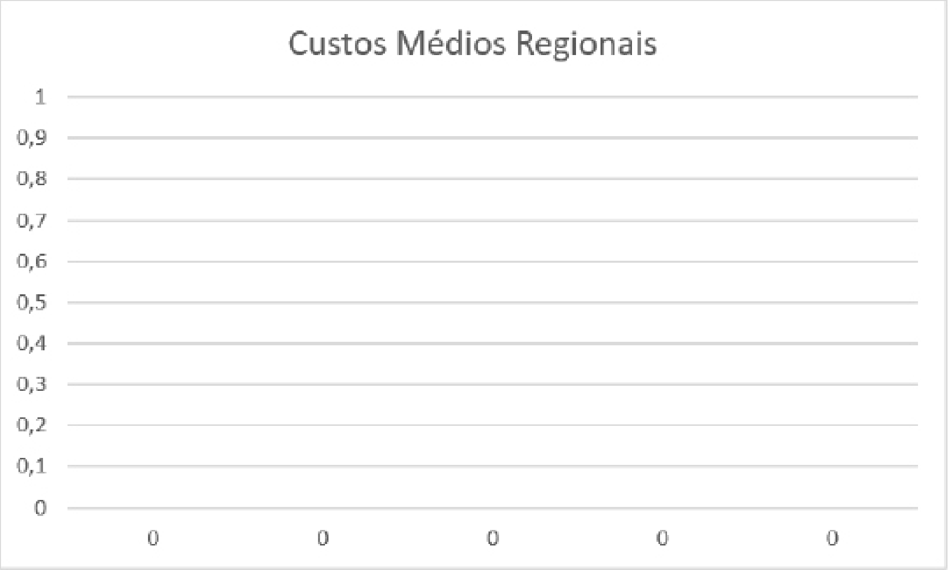
d) O gráfico de X *versus* Y é um histograma.

Letra c.

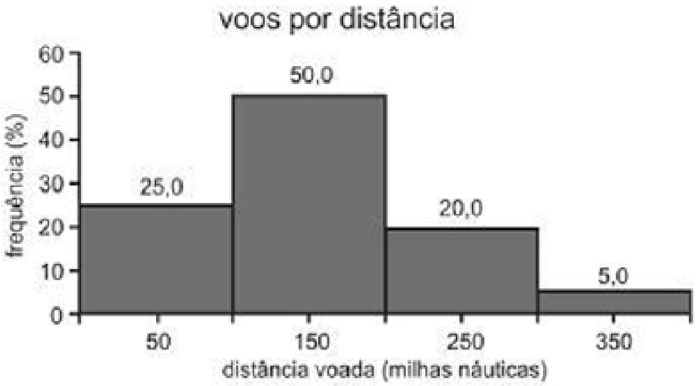
A variável Y é qualitativa. Portanto, em primeiro lugar, é preciso ter em mente que não é possível calcular correlação quando uma das variáveis é qualitativa.

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 um gráfico de barras.



QUESTÃO 54 (CESPE/ANAC/ESPECIALISTA EM REGULAÇÃO DE AVIAÇÃO CIVIL/2012) Com base no histograma acima apresentado, julgue os itens 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.

Certo.

Como foi fornecido um histograma, há a distribuição das distâncias voadas (milhas náuticas) organizadas em classes, cuja amplitude é exatamente igual à distância entre os pontos médios, ou seja:

$$A = 150 - 50 = 100$$

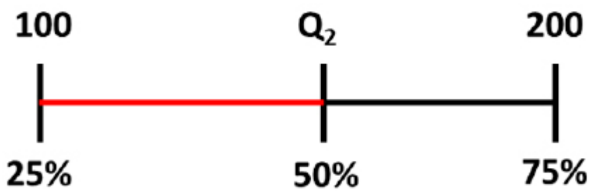
É interessante observar que, como a primeira classe tem frequência relativa igual a 25%, o primeiro quartil corresponde exatamente ao seu extremo. Logo, podemos escrever:

$$Q_1 = PM + \frac{h}{2} = 50 + \frac{100}{2} = 50 + 50 = 100$$

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.



$$\frac{Q_2 - 100}{50\% - 25\%} = \frac{200 - 100}{75\% - 25\%}$$

$$\frac{Q_2 - 100}{25\%} = \frac{100}{50\%}$$

$$\therefore Q_2 - 100 = \frac{25\%}{50\%} \cdot 100$$

$$\therefore Q_2 - 100 = 50 \therefore Q_2 = 100 + 50 = 150$$

QUESTÃO 55 (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 (Md) e o terceiro quartil (Q_3) da distribuição.

Se Md = 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.

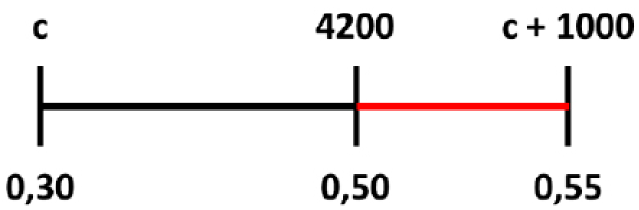
e) 4.720,00.

Letra b.

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.

$$\frac{4200 - c}{0{,}50 - 0{,}30} = \frac{c + 1000 - c}{0{,}55 - 0{,}30}$$

$$\frac{4200 - c}{0{,}20} = \frac{1000}{0{,}25}$$

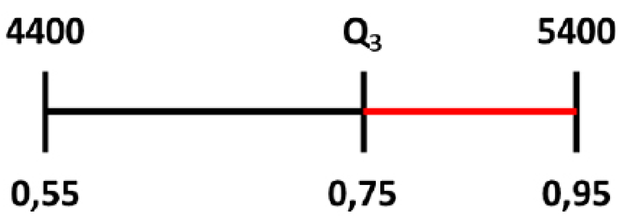
$$\therefore 4200 - c = \frac{0{,}20}{0{,}25} \cdot 1000 = 0{,}20.4000 = 800$$

$$\therefore c = 4200 - 800 = 3400$$

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

Usamos novamente a técnica de interpolação linear com os valores destacados.



Pela técnica da interpolação linear, a parte à direita é proporcional ao segmento inteiro.

$$\frac{5400 - 4400}{0,95 - 0,55} = \frac{Q_3 - 4400}{0,75 - 0,55}$$

$$\frac{1000}{0,40} = \frac{Q_3 - 4400}{0,20}$$

$$\therefore Q_3 - 4400 = \frac{0,20}{0,40} \cdot 1000$$

$$\therefore Q_3 - 4400 = 500 \therefore Q_3 = 4400 + 500 = 4900$$

Chegamos ao final de mais uma aula.

Forte abraço,

Thiago Cardoso.

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.

ANOTAÇÕES

ANOTAÇÕES