

INTRODUÇÃO À ESTATÍSTICA

A origem da estatística remonta às civilizações antigas, em que vários povos coletavam e registravam dados populacionais e econômicos de interesse do Estado. Nessa época, também eram realizadas estimativas das riquezas individuais e familiares, as quais eram utilizadas para determinar o montante de impostos a serem pagos pela população.

O termo estatística se originou da palavra *status*, que significa Estado em latim. O termo era utilizado para designar um conjunto de dados, relativos aos Estados, que os governantes utilizavam com a finalidade de controle fiscal e segurança nacional. O primeiro a utilizar a palavra foi Schmeitzel, ainda no século XVII, em latim. Depois, foi adotada pelo acadêmico alemão Godofredo Achenwall.

A Estatística pode ser definida como a ciência que estuda os processos de coleta, organização, análise e interpretação de dados numéricos variáveis referentes a qualquer fenômeno. Ou ainda, podemos conceituá-la como **um conjunto de técnicas de coleta, organização, análise e interpretação de dados, aplicáveis a várias áreas do conhecimento, que auxiliam no processo de tomada de decisão.**

Os avanços computacionais tornaram a Estatística mais acessível e permitiram aplicações mais sofisticadas em diferentes áreas do conhecimento. Nesse cenário, os softwares estatísticos passaram a disponibilizar ferramentas antes inimagináveis, voltadas para planejamento de experimentos, teste de hipóteses, cálculos de confiabilidade, criação de gráficos complexos e elaboração de modelos preditivos.

A Estatística pode ser dividida em três grandes ramos: Estatística Descritiva (ou dedutiva), Estatística Probabilística e Estatística Inferencial (ou indutiva). Alguns autores, porém, consideram a Estatística Probabilística como parte da Estatística Inferencial.

A Estatística Descritiva (ou Dedutiva) é responsável pela coleta, organização, descrição e resumo dos dados observados. A partir de um determinado conjunto de dados, a Estatística Descritiva busca organizá-los em tabelas (ou gráficos) e estabelecer um sumário por meio de medidas descritivas como a média, os valores mínimo e máximo, o desvio padrão, entre outras.

A Estatística Probabilística é responsável por estabelecer o modelo matemático probabilístico adotado para explicar os fenômenos aleatórios investigados pela Estatística. Os resultados desses fenômenos aleatórios podem variar de uma observação para outra, o que dificulta muito a previsão de um resultado futuro. Por isso, a Teoria da Probabilidade é usada para medir a chance de ocorrência de determinados eventos.

A Estatística Inferencial (ou Indutiva) é responsável pela análise e interpretação dos dados. A partir da análise de dados de uma amostra, a Estatística Indutiva estabelece inferências e previsões sobre a população, auxiliando na tomada decisões. Além disso, busca generalizar conclusões a respeito da população a partir de uma amostra, analisando a representatividade, a significância e a confiabilidade dos resultados obtidos.



ESTATÍSTICA DESCRIPTIVA

É responsável pela coleta, organização, descrição e resumo dos dados observados.

ESTATÍSTICA PROBABILÍSTICA

É responsável por estabelecer o modelo matemático adotado para explicar fenômenos aleatórios.

ESTATÍSTICA INFERENCIAL

É responsável pela análise e interpretação dos dados.



(VUNESP/Câmara Municipal de São Carlos/2013) Trata-se da estatística que tem as atribuições de obtenção, organização, redução e representação de dados estatísticos, bem como a obtenção de algumas informações que auxiliam na descrição de um fenômeno observado. Esta estatística é denominada

- a) Coletora.
- b) Celetista.
- c) Populacional.
- d) Amostral.
- e) Descritiva.

Comentários:

A Estatística está dividida em três grandes ramos, a saber:

- a) **Estatística Descritiva (ou Dedutiva)** - responsável pela coleta (ou obtenção), organização, redução e representação de dados estatísticos;
- b) **Estatística Probabilística** - responsável por estabelecer o modelo matemático probabilístico adotado para explicar os fenômenos aleatórios investigados pela Estatística; e
- c) **Estatística Inferencial (ou Indutiva)** - responsável pela análise e interpretação dos dados.

Portanto, a alternativa correta é a letra E.

Gabarito: E.

CONCEITOS INICIAIS

Neste tópico, apresentaremos alguns conceitos iniciais da estatística que costumam ser abordados em provas de concursos públicos, dentre os quais podemos citar: população, amostra, censo, amostragem, parâmetros e estatísticas.

População

Uma População é um conjunto que contém todos os indivíduos, objetos ou elementos a serem estudados, que apresentam uma ou mais características em comum. A população pode ser finita, quando apresenta um número pequeno ou limitado de observações; ou infinita, quando apresenta um número muito grande ou ilimitado de observações.

Amostra

Uma Amostra é um subconjunto extraído da população para análise, devendo ser representativo daquele grupo. A partir das informações colhidas da amostra, os resultados obtidos podem ser utilizados para generalizar, inferir ou tirar conclusões acerca da população. Como exemplo, podemos citar as pesquisas eleitorais, em que uma amostra de eleitores deve ser extraída de acordo com a proporcionalidade de gênero, idade, grau de instrução e classe social.

Censo

O Censo, ou recenseamento, é um estudo dos dados relativos a todos os elementos de uma população. O censo pode custar muito caro e demandar um tempo considerável, de forma que um estudo considerando apenas uma parcela da população pode ser uma alternativa mais simples, rápida e menos onerosa. Como exemplos, podemos citar a pesquisa sobre o grau de escolaridade dos habitantes brasileiros, o estudo sobre a renda dos brasileiros e a pesquisa de emprego.

Amostragem

A Amostragem é um processo que consiste na seleção criteriosa dos elementos a serem submetidos à investigação. Se forem cometidos erros no processo de seleção da amostra, muito provavelmente, o estudo ficará comprometido e os resultados serão tendenciosos. Portanto, devemos garantir que a amostra seja representativa da população. Isso significa que, com exceção de pequenas discrepâncias inerentes à

aleatoriedade existente no processo de amostragem, uma amostra deve possuir as mesmas características básicas da população, no que diz respeito às variáveis que desejamos pesquisar.

Parâmetros

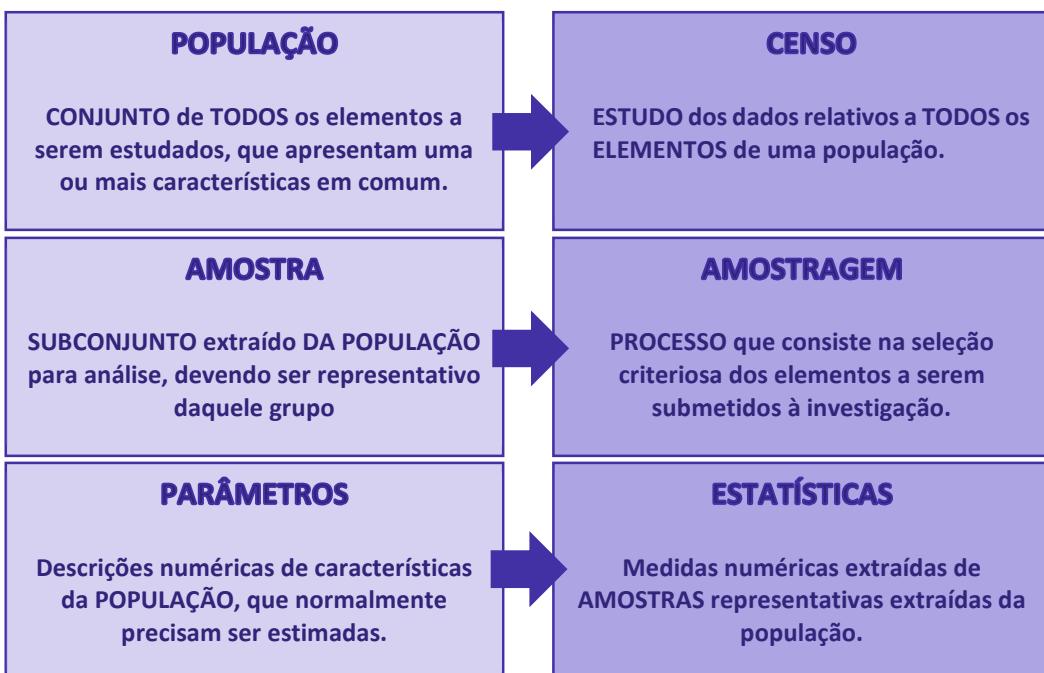
Os parâmetros são descrições numéricas de características populacionais que raramente são conhecidas. Em geral, é muito caro ou demorado obter os dados da população inteira. Assim, **são medidas que normalmente precisam ser estimadas**, a partir de critérios ou métodos definidos pelo pesquisador, **para representar características desconhecidas de uma população (por exemplo, a proporção de homens e mulheres na população brasileira)**. Normalmente, os parâmetros populacionais são constantes para uma população.

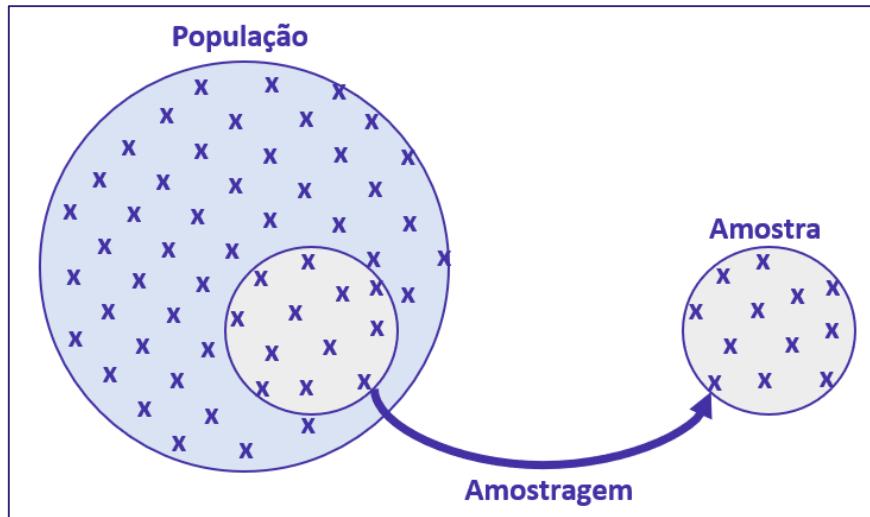
Estatística (ou estimador)

As estatísticas são medidas numéricas obtidas de amostras representativas extraídas da população. A partir das informações colhidas da amostra, as **estatísticas amostrais obtidas podem ser utilizadas para inferir ou tirar conclusões acerca dos parâmetros populacionais**, como a proporção de homens e mulheres na população brasileira. De forma resumida, as estatísticas (ou estimadores) são **descrições numéricas de características amostrais**. Normalmente, as estatísticas amostrais diferem de uma amostra para outra.



ESQUEMATIZANDO





(VUNESP/Pref. Valinhos/2019) O grupo completo de unidades elementares de pessoas, objetos ou coisas é denominado, para a estatística, de

- a) Amostra.
- b) Unidades.
- c) Censo.
- d) População.
- e) Variáveis.

Comentários:

Para resolvemos a questão vamos conceituar cada uma das alternativas:

- Letra A: amostra é um subconjunto de uma população. Os dados coletados e/ou selecionados de uma população estatística por um procedimento definido, denotado de amostragem.
- Letra B: unidades são elementos da amostra (unidades amostrais) ou da população (unidade populacional).

- Letra C: censo consiste na análise de todos os elementos da população.
- Letra D: população é o grupo completo de unidades elementares de pessoas, objetos ou coisas, ou seja, constitui todo o universo de informações de que se necessita.
- Letra E: variável é a característica de relevância que é medida em cada componente da amostra ou população. As variáveis podem ter valores numéricos (variáveis quantitativas) ou não numéricos (variáveis qualitativas).

Gabarito: D.

(CESPE/DEPEN/2015) O diretor de um sistema penitenciário, com o propósito de estimar o percentual de detentos que possuem filhos, entregou a um analista um cadastro com os nomes de 500 detentos da instituição para que esse profissional realizasse entrevistas com os indivíduos selecionados.

A partir dessa situação hipotética e dos múltiplos aspectos a ela relacionados, julgue o item, referente a técnicas de amostragem.

A diferença entre um censo e uma amostra consiste no fato de esta última exigir a realização de um número maior de entrevistas.

Comentários:

Um censo é a análise de todos os elementos de determinada população. Trabalhamos com uma amostra quando obtemos informações de apenas uma parte dessa população. Portanto, o censo requer a realização de um número maior de entrevistas.

Gabarito: Errado.

MÉTODO EXPERIMENTAL X MÉTODO ESTATÍSTICO

Para a investigação de um fenômeno, temos a nossa disposição dois métodos: **experimental** e **estatístico**. De forma resumida, **o método experimental consiste em manter constantes as causas (fatores), com exceção de uma, que é variada para que seus efeitos sejam descobertos**. Contudo, nem sempre poderemos aplicar o método experimental, pois os fatores que afetam um fenômeno podem não permanecer constantes enquanto variamos a causa que nos interessa.

Por exemplo, para analisarmos uma queda nas vendas de uma empresa nacional que produz chocolates finos, teríamos que considerar vários fatores que não necessariamente permanecerão constantes durante toda a investigação do fenômeno, tais como a região, o fluxo de turistas na localidade; a temperatura média; o preço do concorrente; o mês de férias; etc.

Assim, diante da impossibilidade de manter as causas ou fatores constantes, **o método estatístico admite e registra todas as possíveis variações das causas presentes, procurando determinar a influência de cada fator no resultado final**. Dessa forma, o método estatístico busca descobrir relações entre os fatores, como, por exemplo, a influência da temperatura média e do fluxo de turistas na venda de chocolates finos.



MÉTODO EXPERIMENTAL

As CAUSAS são mantidas CONSTANTES, COM EXCEÇÃO DE UMA, que é VARIADA para que seus efeitos sejam descobertos.

MÉTODO ESTATÍSTICO

Admite e REGISTRA TODAS AS POSSÍVEIS VARIAÇÕES DAS CAUSAS PRESENTES, procurando determinar a influência de cada fator no resultado.

DADOS ESTATÍSTICOS

Os **dados estatísticos** constituem os valores resultantes da etapa de coleta de dados. Os dados referem-se a um conjunto de valores, que geralmente são organizados por variáveis (o que está sendo medido) e observações (elementos da amostra/população). É o caso, por exemplo, dos valores obtidos na pesquisa de peso, altura, idade e sexo de uma determinada amostra de indivíduos/população.

Com relação ao número de observações coletadas, os dados são classificados em **univariados**, **bivariados** ou **multivariados**:

- a) **dados univariados**: quando uma única observação de cada indivíduo é registrada. Por exemplo: peso;
- b) **dados bivariados**: quando duas observações de cada indivíduo são registradas. Por exemplo: peso e altura;
- c) **dados multivariados**: quando mais duas observações acerca de cada indivíduo são registradas. Por exemplo: peso, altura, sexo e idade.

Quanto à forma de apresentação, os dados podem ser classificados em **dados brutos** ou **rol**.

Dados Brutos

Os **dados brutos** são aqueles que não foram numericamente organizados em ordem crescente ou decrescente, ou seja, estão na forma como foram coletados. Como exemplo de dados brutos, podemos citar uma relação dos tempos médios de estudo diário, em minutos, de 50 alunos do Estratégia Concursos, em que a escolha dos alunos ocorreu de forma aleatória, não havendo qualquer ordenação de valores.

Aluno	Tempo								
1	143	11	113	21	170	31	124	41	105
2	142	12	143	22	158	32	137	42	154
3	161	13	159	23	123	33	153	43	99
4	126	14	168	24	96	34	129	44	114
5	134	15	123	25	98	35	148	45	161
6	137	16	135	26	135	36	173	46	128
7	171	17	135	27	129	37	126	47	175
8	85	18	175	28	126	38	104	48	137
9	155	19	115	29	103	39	157	49	165
10	171	20	89	30	171	40	127	50	115

Esse tipo de tabela, em que os elementos não aparecem numericamente ordenados, é denominada de **tabela primitiva**. A tabela primitiva, em geral, oferece pouca ou nenhuma informação ao leitor, sendo necessário haver uma organização dos dados, a fim de torná-los mais expressivos.

Rol

O rol é a organização dos dados brutos em ordem de grandeza crescente ou decrescente. Com os dados organizados em rol, podemos saber, com facilidade, qual o menor e o maior elemento de um conjunto de dados. Os dados do nosso exemplo, isto é, os tempos médios de estudo diário, podem ser organizados em ordem crescente ou decrescente:

Rol (em ordem crescente)				
85	115	129	143	161
89	115	129	143	165
96	123	134	148	168
98	123	135	153	170
99	124	135	154	171
103	126	135	155	171
104	126	137	157	171
105	126	137	158	173
113	127	137	159	175
114	128	142	161	175

Rol (em ordem decrescente)				
175	161	142	128	114
175	159	137	127	113
173	158	137	126	105
171	157	137	126	104
171	155	135	126	103
171	154	135	124	99
170	153	135	123	98
168	148	134	123	96
165	143	129	115	89
161	143	129	115	85

VARIÁVEIS ESTATÍSTICAS

A variável estatística consiste no conjunto de características que desejamos averiguar estatisticamente. Ela também pode ser definida como o objeto da pesquisa estatística. Por exemplo, se nosso interesse é conhecer quantas horas os alunos do Estratégia Concursos estudam por dia, então nossa variável é o número de horas estudadas por dia. As variáveis estatísticas podem ser classificadas, inicialmente, em duas categorias: **qualitativas** e **quantitativas**.

Variáveis Qualitativas

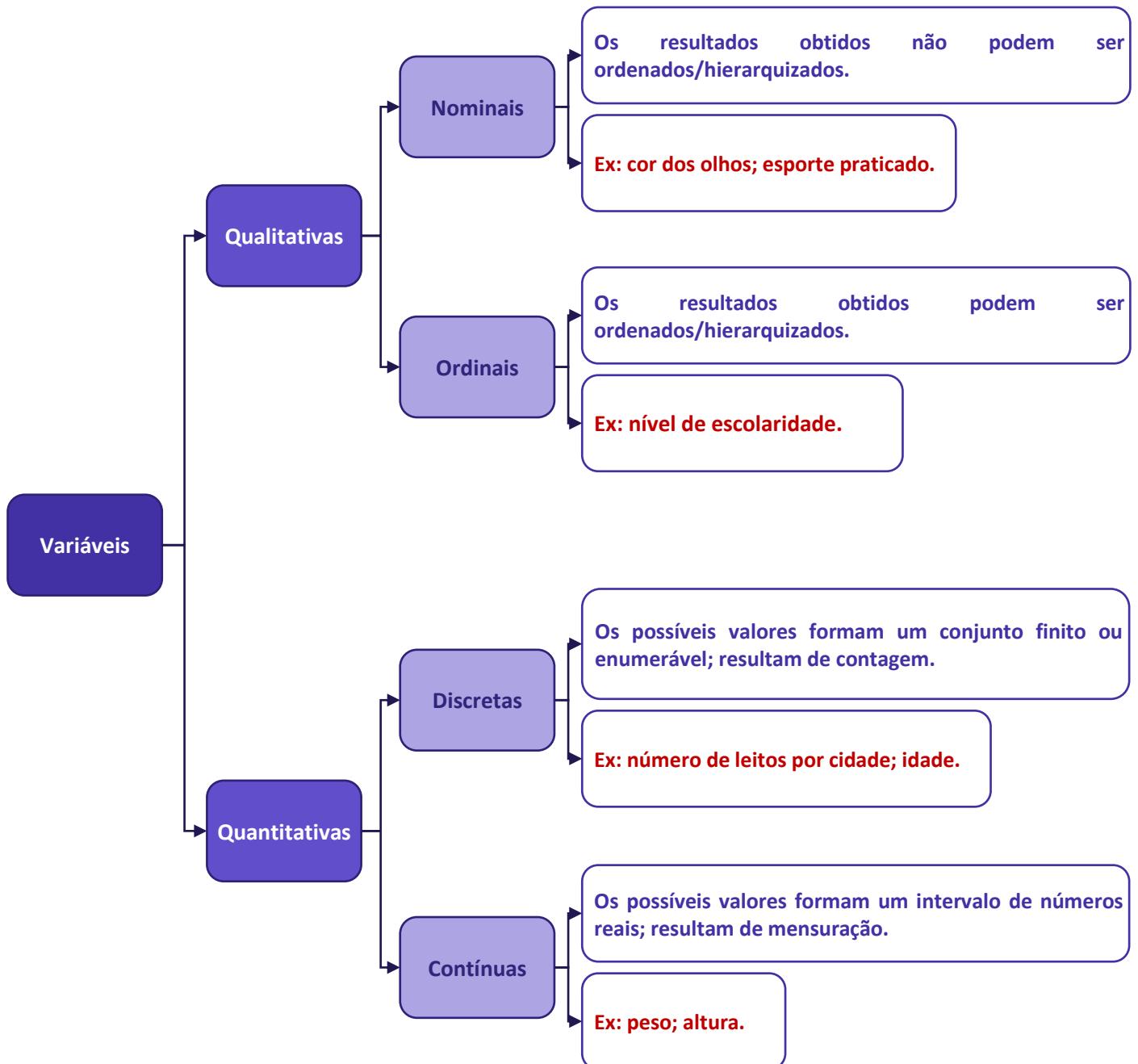
As variáveis qualitativas são as características que não podem ser descritas de forma numérica, mas que podem ser definidas por meio de qualidades (atributos ou categorias) do indivíduo pesquisado. Elas podem ser classificadas em nominais ou ordinais:

- a) variável qualitativa nominal (ou categórica), as possíveis categorias não podem ser ordenadas. Por exemplo, a cor dos olhos dos moradores de uma determinada cidade (pretos, castanhos, azuis e verdes);
- b) variável qualitativa ordinal, as possíveis categorias podem ser ordenadas de alguma forma. Por exemplo, o grau de instrução dos funcionários de um determinado órgão (fundamental, médio, superior).

Variáveis Quantitativas

As variáveis quantitativas são características que podem ser descritas em termos de quantidades (valores numéricos), obtidas por meio de contagem ou mensuração. Elas podem ser classificadas em discretas e contínuas:

- a) variáveis quantitativas discretas, os possíveis valores formam um conjunto finito ou enumerável de números e, geralmente, resultam de um processo de contagem. Por exemplo, o número de leitos abertos em hospitais de uma determinada cidade;
- b) variáveis quantitativas contínuas, os possíveis valores formam um intervalo de números reais e, normalmente, resultam de um processo de mensuração. Por exemplo, a altura dos moradores de uma determinada cidade.





(VUNESP/CM São Joaquim Barra/2018) A Estatística descritiva faz uso de variáveis, que são classificadas como quantitativas ou qualitativas. Assinale a alternativa correta em relação a essas variáveis.

- a) Quantitativas referem-se às variáveis ordinal ou discreta; as qualitativas referem-se às variáveis nominal ou contínua.
- b) Quantitativas referem-se às variáveis ordinal ou contínua; as qualitativas referem-se às variáveis nominal ou discreta.
- c) Quantitativas referem-se às variáveis nominal ou contínua; as qualitativas referem-se às variáveis discreta ou ordinal.
- d) Quantitativas referem-se às variáveis contínua ou discreta; as qualitativas referem-se às variáveis nominal ou ordinal.
- e) Quantitativas referem-se às variáveis nominal ou ordinal; as qualitativas referem-se às variáveis contínua ou discreta.

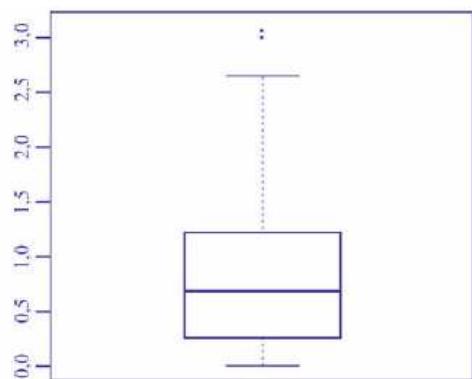
Comentários:

As variáveis podem ser classificadas em:

- a) variáveis quantitativas: são as características que podem ser medidas em uma escala quantitativa, isto é, numérica. Elas podem ser **contínuas** ou **discretas**.
- b) variáveis qualitativas (ou categóricas): são as características que não possuem valores quantitativos. Nesse caso, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Elas podem ser **nominais** ou **ordinais**.

Gabarito: D.

(CESPE/TCE-PA/2016)



média amostral	0,80
desvio padrão amostral	0,70
primeiro quartil	0,25
mediana	0,70
terceiro quartil	1,20
mínimo	0
máximo	3,10

Um indicador de desempenho X permite avaliar a qualidade dos processos de governança de instituições públicas. A figura mostra, esquematicamente, a sua distribuição, obtida mediante estudo amostral feito por determinada agência de pesquisa. A tabela apresenta estatísticas descritivas referentes a essa distribuição.

Com base nessas informações, julgue o item a seguir.

X representa uma variável qualitativa ordinal.

Comentários:

As variáveis podem ser classificadas em:

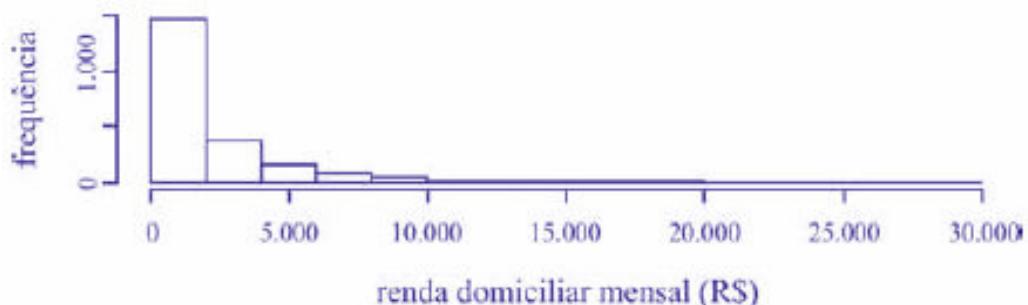
a) variáveis quantitativas: são as características que podem ser representadas numericamente, isto é, podem ser medidas em uma escala quantitativa, podendo ser contínuas ou discretas.

b) variáveis qualitativas (ou categóricas): são as características que não podem ser expressas numericamente, ou seja, não possuem valores quantitativos. Elas são definidas por categorias que classificam os indivíduos, podendo ser nominais ou ordinais.

Analizando as informações dadas na questão, percebemos que X pode adotar valores que vão de 0 (mínimo) a 3,10 (máximo). Logo, como X assume valores numéricos, dizemos que ela é uma variável quantitativa.

Gabarito: Errado.

(CESPE/TELEBRAS/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.

O CEP da localidade dos clientes e o valor da última fatura vencida são variáveis quantitativas.

Comentários:

O CEP, embora possua uma representação numérica, não exprime quantidades e, portanto, não é uma variável quantitativa. Esse código número apenas qualifica as ruas de um determinado endereço, podendo ser classificada como uma variável qualitativa.

O valor da última fatura vencida, por sua vez, é sim uma variável quantitativa.

Gabarito: Errado.

(CESPE/TELEBRAS/2015) Roberto comprou, por R\$ 2.800,00, rodas de liga leve para seu carro, e, ao estacionar no shopping, ficou indeciso sobre onde deixar o carro, pois, caso o coloque no estacionamento público, correrá o risco de lhe roubarem as rodas, ao passo que, caso o coloque no estacionamento privado, terá de pagar R\$ 70,00, com a garantia de que eventuais prejuízos serão resarcidos pela empresa administradora.

Considerando que p seja a probabilidade de as rodas serem roubadas no estacionamento público, que X seja a variável aleatória que representa o prejuízo, em reais, ao deixar o carro no estacionamento público, e que Y seja a variável aleatória que representa o valor, em reais, desembolsado por Roberto ao deixar o carro no estacionamento pago, julgue o item subsequente.

A variável aleatória Y é contínua.

Comentários:

As variáveis quantitativas podem ser classificadas em:

- a) variáveis contínuas: características mensuráveis que assumem valores em uma escala contínua, nesse caso, valores fracionários fazem sentido;
- b) variáveis discretas: características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores e, assim, somente fazem sentido valores inteiros.

No problema em questão, a variável aleatória Y pode assumir o valor R\$ 70,00, se Roberto deixar o carro no estacionamento pago; ou o valor R\$ 0,00, se Roberto não deixar o carro no estacionamento pago. Portanto, Y é uma variável aleatória discreta.

Gabarito: Errado.

SÉRIES ESTATÍSTICAS

Uma série estatística consiste em um conjunto de dados organizado com base em uma característica comum, ou seja, uma mesma variável. Normalmente, uma série estatística é representada por meio de uma tabela ou de um gráfico, conforme ficar melhor representado, a fim de sintetizar os dados estatísticos observados e torná-los mais compreensivos.

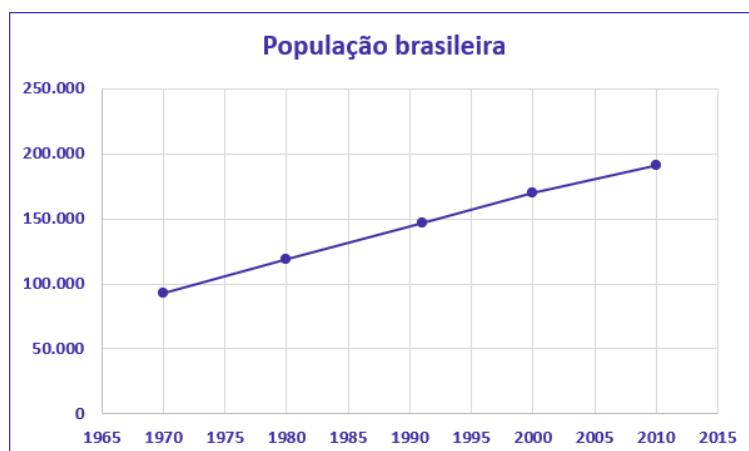
Uma **tabela** é, basicamente, um quadro que resume um conjunto de observações, sendo composta de:

- a) corpo – conjunto de linhas e colunas com as informações sobre a variável em estudo;
- b) cabeçalho – parte superior que especifica o conteúdo das colunas;
- c) coluna indicadora – parte que indica o conteúdo das linhas;
- d) linhas – traços que facilitam a leitura dos dados;
- e) célula – espaço onde os dados são armazenados;
- f) título – identificação da tabela, contendo as informações sobre seu conteúdo.
- g) fonte – referência de onde os dados foram obtidos, localizada no rodapé.

População brasileira no período de 1970 a 2010 (x1000)	
Coluna indicadora	População
1970	93.134
1980	119.011
1991	146.825
2000	169.799
2010	190.755

Título: População brasileira no período de 1970 a 2010 (x1000).
Cabeçalho: Coluna indicadora, População.
Linhas: Anos (1970, 1980, 1991, 2000, 2010).
Coluna numérica: População (93.134, 119.011, 146.825, 169.799, 190.755).
Fonte: Fonte: Censo Demográfico (2010).

Um **gráfico** é uma forma clara e objetiva de apresentar uma série estatística. O objetivo é proporcionar uma compreensão mais rápida do fenômeno em estudo. Para isso, o gráfico deve ser destituído de detalhes sem importância (ser simples); permitir a correta interpretação dos valores representativos do fenômeno (ser claro); e transmitir a verdade sobre o fenômeno (ser verossímil). A série estatística apresentada na tabela anterior pode ser representada graficamente da seguinte forma:



Finalmente, podemos verificar a presença de três elementos nas séries estatística: o tempo, o espaço e a espécie. Conforme os elementos variem, a série pode ser classificada em temporal (ou cronológica), geográfica (ou territorial) e específica.

Séries Temporais (ou Cronológicas)

É a série cujos dados são dispostos segundo a época de ocorrência. Isto é, enquanto o tempo varia, o fato e o local permanecem constantes. Também são chamadas de séries históricas ou evolutivas. A principal característica é o fator cronológico variável.

A seguir temos a série histórica da população residente no Brasil no período de 1970 a 2010, com frequência decenal. Percebam que o tempo varia; contudo, o fato que está sendo analisado (quantidade populacional) e o local alvo da pesquisa (Brasil), continuam iguais.

População brasileira no período de 1970 a 2010 (x1000)

Anos	População
1970	93.134
1980	119.011
1991	146.825
2000	169.799
2010	190.755

Fonte: Censo Demográfico (2010)

Séries Geográficas (ou Territoriais)

É a série cujos dados são dispostos segundo a localidade de ocorrência. Isto é, enquanto o local varia, o fato e o tempo permanecem constantes. Também são chamadas de séries espaciais ou de localização. A principal característica é o fator geográfico variável.

A seguir temos a série geográfica da população urbana residente em cada uma das regiões brasileiras no ano de 2010. Percebam que o local (região) varia; contudo, o fato que está sendo analisado (quantidade populacional) e o tempo (ano de 2010) permanecem constantes.

População Urbana em 2010 (x1000)

Região	População
Norte	11.664
Nordeste	38.821
Sudeste	74.696
Sul	23.260
Centro-Oeste	12.482

Fonte: Censo Demográfico (2010)

Séries Específicas

É a série cujos dados são dispostos segundo a modalidade de ocorrência. Isto é, enquanto o fato varia, a época e o local permanecem constantes. Também são chamadas de séries **categóricas**. A principal característica é o fator especificativo variável.

A seguir temos uma série específica das populações urbana e rural residentes no Brasil no ano de 2010. Percebam que os fatos analisados variam (população urbana x população rural); contudo, o tempo (2010) e o local de análise (Brasil) são constantes.

População Urbana e Rural em 2010 (x1000)

Zona	População
Urbana	93.134
Rural	119.011
Total	190.755

Fonte: Censo Demográfico (2010)

Séries Mistas (ou Compostas)

Muitas vezes, temos a necessidade de apresentar, em uma única tabela, a variação de valores de mais de uma variável, isto é, fazer uma combinação de duas ou mais séries. As séries resultantes desse processo de combinação são chamadas de séries mistas (ou compostas) e apresentadas por meio de **tabelas de dupla entrada**.

O nome da nova série deve levar em consideração pelo menos dois elementos. Assim, **se for uma série mista de fato e tempo, denominaremos de série específico-temporal**. A seguir temos uma série específico-temporal representando as populações de homens e mulheres residentes no Brasil, no período de 1970 a 2010, com variação decenal.

População do Brasil por Sexo de 1970 a 2010 (x1000)

Anos	Sexo	
	Homens	Mulheres
1970	46.327	46.807
1980	59.142	59.868
1991	72.485	74.340
2000	83.602	86.270
2010	93.406	97.348

Fonte: Censo Demográfico (2010)

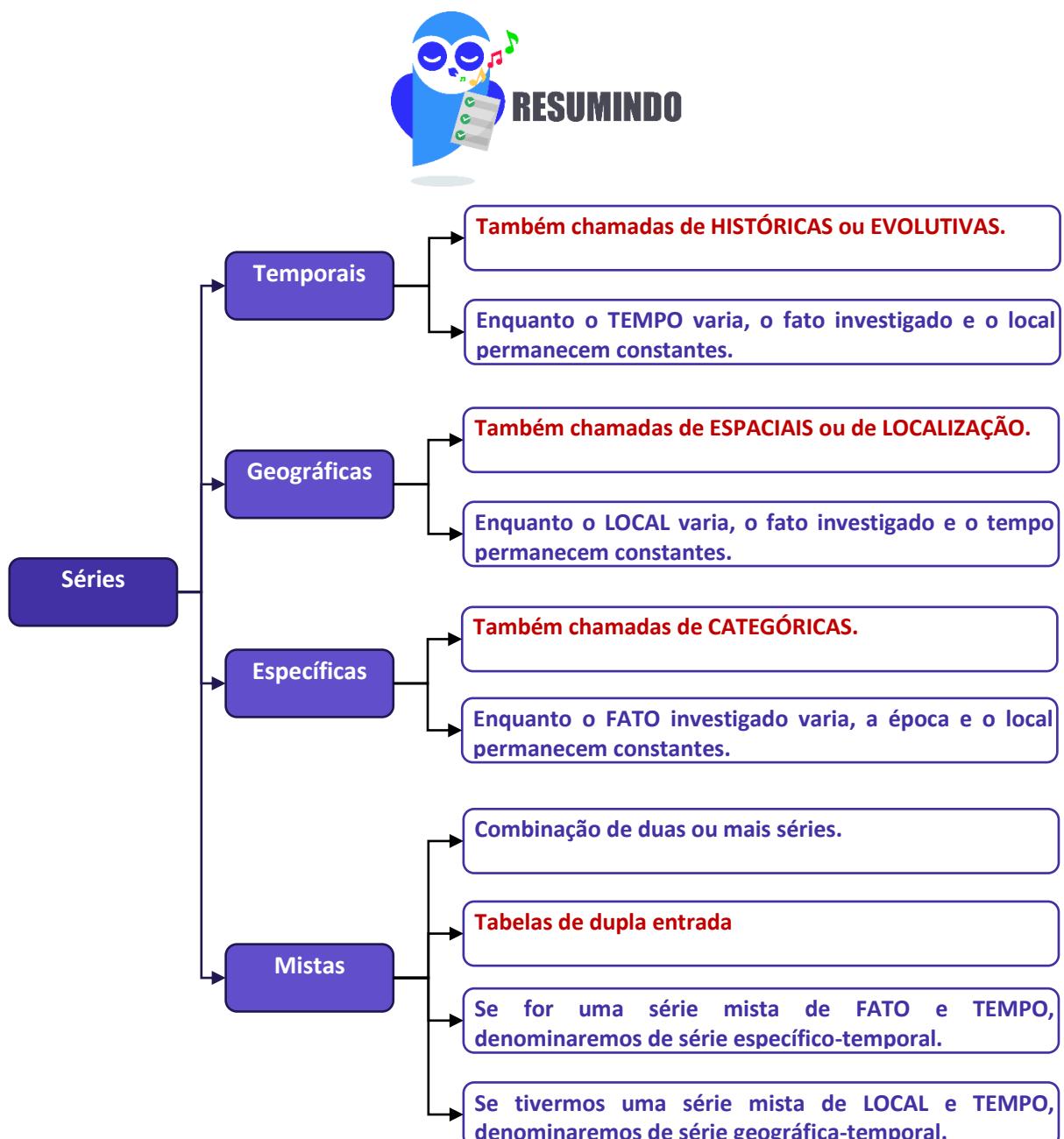
Por sua vez, **se tivermos uma série mista de local e tempo, denominaremos de série geográfica-temporal**. A seguir temos uma série geográfico-temporal representando as populações residentes em cada região brasileira, no período de 1970 a 2010, com variação decenal.

População do Brasil por Região de 1970 a 2010 (x1000)

Anos	Regiões				
	N	NE	SE	S	CO
1970	3.603	28.111	39.850	16.496	5.072
1980	5.880	34.815	51.737	19.031	7.545
1991	10.030	42.497	62.740	22.129	9.427
2000	12.900	47.741	72.412	25.107	11.636
2010	15.864	53.081	80.364	27.386	14.058

Fonte: Censo Demográfico (2010)

Por fim, devemos notar que podem existir séries compostas de três ou mais entradas, embora isso raramente aconteça, por conta da dificuldade de representação.



DISTRIBUIÇÃO DE FREQUÊNCIAS

Vimos anteriormente que, logo após a coleta de dados, temos o que denominamos de **dados brutos**. Como exemplo de dados brutos, citamos uma pesquisa de tempo médio de estudo diário, em minutos, envolvendo 50 alunos do Estratégia Concursos, em que os alunos foram escolhidos de maneira aleatória, não havendo qualquer organização dos valores observados. Por terem sido apresentados na forma em que foram coletados, são denominados de **dados brutos**.

Aluno	Tempo								
1	143	11	113	21	170	31	124	41	105
2	142	12	143	22	158	32	137	42	154
3	161	13	159	23	123	33	153	43	99
4	126	14	168	24	96	34	129	44	114
5	134	15	123	25	98	35	148	45	161
6	137	16	135	26	135	36	173	46	128
7	171	17	135	27	129	37	126	47	175
8	85	18	175	28	126	38	104	48	137
9	155	19	115	29	103	39	157	49	165
10	171	20	89	30	171	40	127	50	115

Normalmente, esse dado fornece pouca informação ao leitor, sendo necessário organizar os dados, com o propósito de aumentar sua capacidade de informação. A simples organização dos dados em um **rol crescente** já aumenta bastante a capacidade de informação. **Com os dados organizados em rol, conseguimos verificar que o menor tempo observado foi 85 minutos, e o maior, 175 minutos, o que nos fornece uma amplitude total (AT = 175 - 85 = 70) de variação da ordem de 70 minutos.**

Rol Crescente				
85	115	129	143	161
89	115	129	143	165
96	123	134	148	168
98	123	135	153	170
99	124	135	154	171
103	126	135	155	171
104	126	137	157	171
105	126	137	158	173
113	127	137	159	175
114	128	142	161	175

Outra informação que conseguimos extrair dos dados organizados em rol crescente é que alguns tempos, como 126 min, 135 min, 137 min e 171 min, foram mais frequentes, ou seja, apareceram mais vezes durante a pesquisa.

Uma maneira mais concisa de mostrar os dados do rol é apresentar cada valor juntamente com o número de vezes em que ocorre, ao invés de repeti-los. **O número de ocorrências de um determinado valor recebe o nome de frequência. A tabela que contém todos os valores com suas respectivas frequências é denominada de distribuição de frequências.**

Uma **distribuição de frequências** também **pode ser definida como uma série estatística na qual permanecem constantes o fato, o local e a época**. Ela pode ser classificada em dois tipos: distribuição de frequências **pontual (ou discreta)** e distribuição de frequências **intervalar (ou contínua)**.

Na distribuição de frequências pontual, são apresentados todos os dados coletados juntamente com suas respectivas frequências, não havendo perda de valores. Contudo, esse processo pode exigir muito espaço, especialmente quando o número de valores da variável tende a aumentar.

Tempo (min)	Freq.						
85	1	114	1	135	3	158	1
89	1	115	2	137	3	159	1
96	1	123	2	142	1	161	2
98	1	124	1	143	2	165	1
99	1	126	3	148	1	168	1
103	1	127	1	153	1	170	1
104	1	128	1	154	1	171	3
105	1	129	2	155	1	173	1
113	1	134	1	157	1	175	2

Nesses casos, quando a variável é contínua, o mais recomendável é agrupar os valores por intervalos de classe. Desse modo, ao invés de listar cada um dos valores que ocorrem, utilizamos uma **distribuição de frequências intervalar**, listando os intervalos de classe e as frequências correspondentes.

Tempo médio (X_i)	Frequência (f_i)
$85 \leq x < 100$	5
$100 \leq x < 115$	5
$115 \leq x < 130$	12
$130 \leq x < 145$	10
$145 \leq x < 160$	7
$160 \leq x < 175$	9
$175 \leq x < 190$	2

Procedendo dessa forma, **perdemos a informação detalhada dos tempos médios**, mas ganhamos em termos de praticidade, o que simplifica o processo de análise de dados. Examinando a tabela acima, percebemos facilmente que a maioria dos alunos estuda diariamente entre 115 e 130 minutos, enquanto uma minoria alcança entre 175 e 190 minutos.

Para identificar uma classe, temos que conhecer os valores dos **limites inferior e superior da classe**, que delimitam um **intervalo de classe**. Desse modo, precisamos definir a natureza do intervalo de classe, se aberto ou fechado. Portanto, temos as seguintes notações para os diferentes tipos de intervalos:



Tipo de Intervalo	Notação matemática	Notação estatística	Significado
Intervalo aberto	$a < x < b$	$a - b$	Engloba todos os elementos entre a e b , mas não engloba a nem b .
Intervalo fechado à esquerda e aberto à direita	$a \leq x < b$	$a \leftarrow b$	Engloba todos os elementos entre a e b , inclusive a mas não b .
Intervalo aberto à esquerda e fechado à direita	$a < x \leq b$	$a \rightarrow b$	Engloba todos os elementos entre a e b , inclusive b mas não a .
Intervalo fechado	$a \leq x \leq b$	$a \mathbin{\text{\texttt{H}}} b$	Engloba todos os elementos entre a e b , inclusive a e b .

Por fim, é importante salientarmos que, em análises estatísticas, constantemente encontramos distribuições de frequências intervalares, pois o objetivo da estatística é justamente fazer o apanhado geral das características de um conjunto de dados, sem adentrar em detalhes de casos particulares.

Elementos de uma Distribuição de Frequências

Agora, analisaremos cada um dos elementos de uma distribuição de frequências de forma mais detalhada. Tomaremos como referência a tabela apresentada anteriormente:

Tempo médio (X_i)	Frequência (f_i)
$85 \leq x < 100$	5
$100 \leq x < 115$	5
$115 \leq x < 130$	12
$130 \leq x < 145$	10
$145 \leq x < 160$	7
$160 \leq x < 175$	9
$175 \leq x < 190$	2

Classe

As classes são os intervalos nos quais o fenômeno é subdividido. Isto é, podemos dizer que as classes são os intervalos ou subdivisões dos elementos que compõem um conjunto de dados. Na tabela anterior, temos as seguintes classes:

Classe	Intervalo	Frequência (f_i)
1ª Classe	$85 \leq x < 100$	5
2ª Classe	$100 \leq x < 115$	5
3ª Classe	$115 \leq x < 130$	12
4ª Classe	$130 \leq x < 145$	10
5ª Classe	$145 \leq x < 160$	7
6ª Classe	$160 \leq x < 175$	9
7ª Classe	$175 \leq x < 190$	2
		$n = 50$

Existem duas maneiras de determinar o número de classes, k , em função do número de dados da tabela, n . A primeira consiste em utilizar a fórmula de Sturges:

$$k = 1 + 3,3 \times \log n$$

Outra abordagem, quando o número de dados é menor ou igual a 50, é por meio da fórmula:

$$k = \sqrt{n}$$

Vamos aproveitar para calcular o número de classes do nosso exemplo:

a) pela fórmula de Sturges:

$$k = 1 + 3,3 \times \log n$$

$$k = 1 + 3,3 \times \log 50$$

$$k = 1 + 3,3 \times 1,7$$

$$k = 1 + 5,61 = 6,61$$

b) pela outra fórmula:

$$k = \sqrt{n}$$

$$k = \sqrt{50} = 7,07$$

Limite de Classe

Cada classe tem um **limite inferior de classe** (l_{inf}), que é o menor número que pode pertencer à classe, e um **limite superior de classe** (l_{sup}), que é o maior número que pode pertencer à classe. **Os limites de uma classe são seus valores extremos.**

Vamos identificar os limites inferiores e superiores do nosso exemplo:

Classes	Limite Inferior (l_{inf})	Limite Superior (l_{sup})
$85 \leq x < 100$	85	100
$100 \leq x < 115$	100	115
$115 \leq x < 130$	115	130
$130 \leq x < 145$	130	145
$145 \leq x < 160$	145	160
$160 \leq x < 175$	160	175
$175 \leq x < 190$	175	190

Amplitude de um Intervalo de Classe

A **amplitude de um intervalo de classe**, ou simplesmente **intervalo de classe**, é a distância entre os limites inferiores (ou superiores) de classes consecutivas. Ela é obtida pela diferença entre dois limites inferiores (ou superiores) consecutivos:

$$h = l_{sup} - l_{inf}$$

em que l_{inf} é o limite inferior do intervalo de classe e l_{sup} é o limite superior do intervalo de classe.

Dando continuidade ao nosso exemplo, vejamos como calcular as amplitudes dos intervalos:

Classes	Limite Inferior (l_{inf})	Limite Superior (l_{sup})	Amplitude ($h = l_{sup} - l_{inf}$)
$85 \leq x < 100$	85	100	$100 - 85 = 15$
$100 \leq x < 115$	100	115	$115 - 100 = 15$
$115 \leq x < 130$	115	130	$130 - 115 = 15$

130 ≤ x < 145	130	145	145 - 130 = 15
145 ≤ x < 160	145	160	160 - 145 = 15
160 ≤ x < 175	160	175	175 - 160 = 15
175 ≤ x < 190	175	190	190 - 175 = 15

Embora seja desejável, a amplitude do intervalo de classe nem sempre será constante ao longo de toda a distribuição de frequências intervalar.

Amplitude Total

A **amplitude total** é a diferença entre o limite superior da **última classe** (limite superior máximo) e o limite inferior da **primeira classe** (limite inferior mínimo). Portanto, corresponde à diferença entre o último e o primeiro elemento de um conjunto de dados ordenado de forma crescente:

$$AT = l_{máx} - l_{mín}$$

Note que, **quando todas as classes possuem a mesma amplitude**, também podemos determinar o valor da amplitude total multiplicando o valor do intervalo de classe (h) pela quantidade de classes da distribuição (k):

$$AT = h \times k.$$

Em nosso exemplo, a amplitude total é calculada da seguinte maneira:

$$AT = l_{máx} - l_{mín}$$

$$AT = 190 - 85 = 105$$

Ponto médio de classe

O **ponto médio** é a média aritmética simples dos valores extremos de uma classe, ou seja, a soma dos limites inferior e superior dividida por dois. Esse ponto divide a classe em duas partes iguais. Também costuma ser chamado de **marca ou representante da classe**.

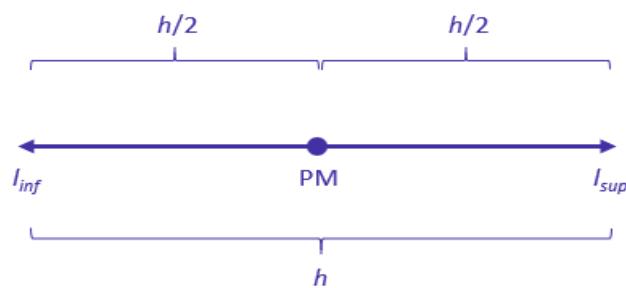
$$PM = \frac{(l_{inf} + l_{sup})}{2}$$

Para praticar, vamos calcular os pontos médios de nossa distribuição de frequências:

Tempo médio (X_i)	Ponto médio (PM_i)	Frequência (f_i)
$85 \leq x < 100$	$\frac{85 + 100}{2} = 92,5$	5
$100 \leq x < 115$	$\frac{100 + 115}{2} = 107,5$	5
$115 \leq x < 130$	$\frac{115 + 130}{2} = 122,5$	12
$130 \leq x < 145$	$\frac{130 + 145}{2} = 137,5$	10
$145 \leq x < 160$	$\frac{145 + 160}{2} = 152,5$	7
$160 \leq x < 175$	$\frac{160 + 175}{2} = 167,5$	9
$175 \leq x < 190$	$\frac{175 + 190}{2} = 182,5$	2

Veja que os pontos médios formaram uma progressão aritmética, vez que a diferença entre dois pontos médios consecutivos foi constante e igual a 15. Isso ocorreu porque o intervalo de classe, h , também foi constante (e igual a 15) em toda a distribuição. Assim, podemos afirmar que, quando o intervalo de classes é constante, a diferença entre os pontos médios também será constante e igual ao intervalo de classe.

Adicionalmente, sabendo que o ponto médio divide a classe em duas partes iguais, podemos derivar outras relações envolvendo o próprio ponto médio, a amplitude de classe e os limites inferior e superior.



Dada a figura anterior, podemos obter os limites de uma classe por meio das seguintes expressões:

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

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

Além disso, podemos encontrar o ponto médio de uma classe a partir das seguintes relações:

$$PM = l_{inf} + \frac{h}{2}$$

$$PM = l_{sup} - \frac{h}{2}$$

Frequência

Ao longo dessa aula, em várias oportunidades abordamos conceitos relacionados à frequência, isto é, ao número de ocorrências de um determinado valor ou de uma certa classe. Esse conceito é de grande relevância para a estatística descritiva e deve ser estudado de forma mais aprofundada. Nesse contexto, é importante sabermos que existem quatro tipos de frequência, os quais serão analisados nas subseções seguintes:

- a) Frequência absoluta simples (f_i);
- b) Frequência absoluta acumulada (f_{ac});
- c) Frequência relativa simples (F_i);
- d) Frequência relativa acumulada (F_{ac}).

Frequência Absoluta Simples

A **frequência absoluta simples** corresponde ao número de observações correspondentes a uma determinada classe ou a um determinado valor.

i	Tempos (min)	Frequência (f_i)
1	85 \leftarrow 100	5
2	100 \leftarrow 115	5
3	115 \leftarrow 130	12
4	130 \leftarrow 145	10
5	145 \leftarrow 160	7
6	160 \leftarrow 175	9
7	175 \leftarrow 190	2

A frequência simples é simbolizada por f_i . No exemplo anterior, temos: $f_1 = 5$, $f_2 = 5$, $f_3 = 12$, $f_4 = 10$, $f_5 = 7$, $f_6 = 9$ e $f_7 = 2$.

A soma de todas as frequências é igual ao número total de dados analisados:

$$\sum_{i=1}^k f_i = n$$

em que a notação $\sum_{i=1}^k f_i$ representa o somatório das frequências de cada uma das k classes.

Para a distribuição em análise, temos:

$$\sum_{i=1}^7 f_i = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 = 5 + 5 + 12 + 10 + 7 + 9 + 2 = 50$$

Agora, podemos incluir essa informação na representação tabular:

i	Tempos (min)	Frequência (f_i)
1	85 \leftarrow 100	5
2	100 \leftarrow 115	5
3	115 \leftarrow 130	12
4	130 \leftarrow 145	10
5	145 \leftarrow 160	7
6	160 \leftarrow 175	9
7	175 \leftarrow 190	2
		$\sum_{i=1}^7 f_i = 50$

Frequência Absoluta Acumulada

A **frequência absoluta acumulada crescente (f_{ac})** é o total das frequências de todos os valores inferiores ao **limite superior do intervalo de uma dada classe**. No exemplo apresentado anteriormente, a frequência acumulada correspondente à quarta classe é: $f_{ac_4} = f_1 + f_2 + f_3 + f_4 = 5 + 5 + 12 + 10 = 32$, o que significa existem 32 alunos que estudam diariamente entre 85 e 145 minutos (limite superior da quarta classe).

$$f_{ac_i} = f_1 + f_2 + f_3 + \cdots + f_i$$

A **frequência absoluta acumulada crescente** é calculada de cima para baixo, da seguinte forma:

- 1) repetimos a frequência absoluta da **primeira** classe;
- 2) os demais valores da frequência absoluta são obtidos a partir da soma da frequência acumulada anterior com a frequência absoluta da classe correspondente;
- 3) a frequência acumulada crescente sempre termina com o valor de n.

i	Tempos (min)	Frequência (f_i)	Frequência Acumulada
1	85 \leftarrow 100	5	5
2	100 \leftarrow 115	5	10
3	115 \leftarrow 130	12	22
4	130 \leftarrow 145	10	32
5	145 \leftarrow 160	7	39
6	160 \leftarrow 175	9	48
7	175 \leftarrow 190	2	50

$$\sum_{i=1}^7 f_i = 50$$

A **frequência absoluta acumulada decrescente** (f_{ad}) é o total das frequências de todos os valores superiores ao limite inferior do intervalo de uma dada classe. No exemplo apresentado anteriormente, a frequência acumulada correspondente à quarta classe é: $f_{ad4} = f_4 + f_5 + f_6 + f_7 = 10 + 7 + 9 + 2 = 28$, o que significa existem 28 alunos que estudam diariamente entre 130 (limite inferior da quarta classe) e 190 minutos.

$$f_{ad_i} = f_i + f_{i+1} + f_{i+2} + \cdots + f_k$$

A **frequência absoluta acumulada decrescente** é calculada de baixo para cima, da seguinte forma:

- 1) repetimos a frequência absoluta da **última** classe;
- 2) os demais valores da frequência acumulada decrescente são obtidos a partir da soma da frequência acumulada anterior com a frequência absoluta da classe correspondente;
- 3) a frequência acumulada decrescente sempre termina com o valor de n.

i	Tempos (min)	Frequência (f_i)	Frequência Acumulada
1	85 \leftarrow 100	5	50
2	100 \leftarrow 115	5	45
3	115 \leftarrow 130	12	40
4	130 \leftarrow 145	10	28
5	145 \leftarrow 160	7	18
6	160 \leftarrow 175	9	11
7	175 \leftarrow 190	2	2
$\sum_{i=1}^7 f_i = 50$			

Frequência Relativa Simples

A **frequência relativa simples** corresponde à proporção de dados existentes em uma determinada classe. Para calcular a frequência relativa de uma classe, dividimos a frequência absoluta simples f_i pela frequência total (isto é, dividimos a parte pelo todo):

$$F_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}$$

Em nosso exemplo, as frequências relativas são:

i	Tempos (min)	Frequência Absoluta (f_i)	Frequência Relativa (F_i)
1	85 \leftarrow 100	5	$F_1 = \frac{5}{50} = 0,10$
2	100 \leftarrow 115	5	$F_2 = \frac{5}{50} = 0,10$
3	115 \leftarrow 130	12	$F_3 = \frac{12}{50} = 0,24$
4	130 \leftarrow 145	10	$F_4 = \frac{10}{50} = 0,20$
5	145 \leftarrow 160	7	$F_5 = \frac{7}{50} = 0,14$
6	160 \leftarrow 175	9	$F_6 = \frac{9}{50} = 0,18$
7	175 \leftarrow 190	2	$F_7 = \frac{2}{50} = 0,04$

Para representar esses valores em termos de porcentagem, basta multiplicarmos por 100%. A tabela ficaria assim:

i	Tempos (min)	Frequência Absoluta (f_i)	Frequência Relativa (F_i)
1	85 \leftarrow 100	5	$F_1 = \frac{5}{50} \times 100\% = 10\%$
2	100 \leftarrow 115	5	$F_2 = \frac{5}{50} \times 100\% = 10\%$
3	115 \leftarrow 130	12	$F_3 = \frac{12}{50} \times 100\% = 24\%$
4	130 \leftarrow 145	10	$F_4 = \frac{10}{50} \times 100\% = 20\%$
5	145 \leftarrow 160	7	$F_5 = \frac{7}{50} \times 100\% = 14\%$
6	160 \leftarrow 175	9	$F_6 = \frac{9}{50} \times 100\% = 18\%$
7	175 \leftarrow 190	2	$F_7 = \frac{2}{50} \times 100\% = 4\%$

O propósito das frequências relativas é facilitar a realização de comparações de classes individuais com o total das observações. Na tabela anterior, por exemplo, conseguimos verificar facilmente que 24% das observações pertencem à quarta classe e que 18% das observações pertencem à sexta classe.

Repare que a soma de todas as frequências relativas deve ser igual a 100%:

$$\sum_{i=1}^k F_i = 100\%$$

Podemos incluir essa informação na representação tabular:

i	Tempos (min)	Frequência Absoluta (f_i)	Frequência Relativa (F_i)
1	85 \leftarrow 100	5	10%
2	100 \leftarrow 115	5	10%
3	115 \leftarrow 130	12	24%
4	130 \leftarrow 145	10	20%
5	145 \leftarrow 160	7	14%
6	160 \leftarrow 175	9	18%
7	175 \leftarrow 190	2	4%
$\sum_{i=1}^7 f_i = 50$		$\sum_{i=1}^7 F_i = 100\%$	

Frequência Relativa Acumulada

A **frequência relativa acumulada crescente** (F_{ac}) é a proporção de valores inferiores ao limite superior do intervalo de uma dada classe. No exemplo apresentado anteriormente, a frequência acumulada correspondente à quarta classe é: $F_{ac4} = F_1 + F_2 + F_3 + F_4 = 10\% + 10\% + 24\% + 20\% = 64\%$, o que significa que 64% dos alunos estudam diariamente entre 85 e 145 minutos (limite superior da quarta classe).

$$F_{ac_i} = F_1 + F_2 + F_3 + \cdots + F_i$$

A **frequência relativa acumulada crescente** é calculada de cima para baixo, da seguinte forma:

- 1) repetimos a frequência relativa da **primeira** classe;
- 2) os demais valores são obtidos a partir da soma da frequência relativa acumulada anterior com a frequência relativa da classe correspondente;
- 3) a frequência relativa acumulada sempre termina com o valor de 100%.

i	Tempos (min)	Frequência Absoluta (f_i)	Frequência Relativa (F_i)	Frequência Acumulada (F_{ac})
1	85 \leftarrow 100	5	10%	10%
2	100 \leftarrow 115	5	10%	20%
3	115 \leftarrow 130	12	24%	44%
4	130 \leftarrow 145	10	20%	64%
5	145 \leftarrow 160	7	14%	78%
6	160 \leftarrow 175	9	18%	96%
7	175 \leftarrow 190	2	4%	100%

$$\sum_{i=1}^7 f_i = 50 \quad \sum_{i=1}^7 F_i = 100\%$$

A **frequência relativa acumulada decrescente** (F_{ad}) a proporção de valores superiores ao limite inferior do intervalo de uma dada classe. No exemplo apresentado anteriormente, a frequência acumulada correspondente à quarta classe é: $F_{ad4} = F_4 + F_5 + F_6 + F_7 = 20\% + 14\% + 18\% + 4\% = 56\%$, o que significa que 56% dos alunos estudam diariamente entre 130 (limite inferior da quarta classe) e 190 minutos.

$$F_{ad_i} = F_i + F_{i+1} + F_{i+2} + \cdots + F_k$$

A **frequência relativa acumulada decrescente** é calculada de baixo para cima, da seguinte forma:

- 1) repetimos a frequência relativa da **última** classe;
- 2) os demais valores são obtidos a partir da soma da frequência acumulada anterior com a frequência relativa da classe correspondente;

3) a frequência acumulada sempre termina com o valor de 100%.

i	Tempos (min)	Frequência Absoluta (f_i)	Frequência Relativa (F_i)	Frequência Acumulada (F_{ad})
1	85 \leftarrow 100	5	10%	100%
2	100 \leftarrow 115	5	10%	90%
3	115 \leftarrow 130	12	24%	80%
4	130 \leftarrow 145	10	20%	56%
5	145 \leftarrow 160	7	14%	36%
6	160 \leftarrow 175	9	18%	22%
7	175 \leftarrow 190	2	4%	4%
$\sum_{i=1}^7 f_i = 50$		$\sum_{i=1}^7 F_i = 100\%$		

Densidade de Frequência

A densidade de frequência de uma classe consiste no quociente entre a frequência da classe (absoluta ou relativa) e sua amplitude:

$$\text{densidade} = \frac{\text{frequência absoluta ou relativa}}{\text{amplitude}}$$

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

Para o nosso exemplo, as densidades de frequência são:

i	Tempos (min)	Frequência Absoluta (f_i)	Densidade de Frequência (d)
1	85 \leftarrow 100	5	$d_1 = \frac{f_1}{h_1} = \frac{5}{15} = 0,33$
2	100 \leftarrow 115	5	$d_2 = \frac{f_2}{h_2} = \frac{5}{15} = 0,33$
3	115 \leftarrow 130	12	$d_3 = \frac{f_3}{h_3} = \frac{12}{15} = 0,80$
4	130 \leftarrow 145	10	$d_4 = \frac{f_4}{h_4} = \frac{10}{15} = 0,66$

5	145 ← 160	7	$d_5 = \frac{f_5}{h_5} = \frac{7}{15} = 0,46$
6	160 ← 175	9	$d_6 = \frac{f_6}{h_6} = \frac{9}{15} = 0,60$
7	175 ← 190	2	$d_7 = \frac{f_7}{h_7} = \frac{2}{15} = 0,13$



Item	Definição	Símbolos e Fórmulas
Número de Classes	As classes são os intervalos nos quais o fenômeno é subdividido.	$k = 1 + 3,3 \times \log n$ $k = \sqrt{n}$
Limites de Classe	Correspondem aos valores extremos.	l_{inf} e l_{sup}
Amplitude de um Intervalo de Classe	Distância entre os limites inferiores (ou superiores) de classes consecutivas.	$h = l_{sup} - l_{inf}$
Amplitude total	Diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo).	$AT = l_{máx} - l_{mín}$ $AT = h \times k$
Ponto Médio	Média aritmética simples dos valores extremos de uma classe.	$PM = \frac{(l_{inf} + l_{sup})}{2}$ $PM = l_{inf} + \frac{h}{2}$ $PM = l_{sup} - \frac{h}{2}$
Frequência Absoluta Simples	Número de observações correspondentes a uma determinada classe ou a um determinado valor.	f_i
Frequência Absoluta Acumulada	Total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe	$f_{ac_i} = f_1 + f_2 + f_3 + \dots + f_i$
Frequência Relativa Simples	Proporção de dados existentes em uma determinada classe.	$F_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}$

Frequência Relativa Acumulada	Proporção de valores inferiores ao limite superior do intervalo de uma dada classe.	$F_{ac_i} = F_1 + F_2 + F_3 + \cdots + F_i$
Densidade de Frequência	Quociente entre a frequência da classe (absoluta ou relativa) e sua amplitude	$d = \frac{f}{h}$



(VUNESP/IPSM-SJC/2018) Considere as informações a seguir para construir uma distribuição de frequência sem intervalo de classe e responder a questão.

Um dado foi lançado 50 vezes e foram registrados os seguintes resultados:

5 4 6 1 2 5 3 1 3 3
 4 4 1 5 5 6 1 2 5 1
 3 4 5 1 1 6 6 2 1 1
 4 4 4 3 4 3 2 2 2 3
 6 6 3 2 4 2 6 6 2 1

A frequência simples relativa do primeiro elemento é

- a) 10%.
- b) 20%.
- c) 50.
- d) 7.
- e) 20.

Comentários:

A frequência absoluta é o número de vezes em que uma determinada variável assume um valor, enquanto a frequência relativa é o quociente entre a frequência absoluta da classe correspondente e a soma das frequências (total observado).

Para os dados apresentados, temos que :

Valor obtido no lançamento	Frequência absoluta	Frequência relativa
1	10	$10/50 = 0,2 = 20\%$
2	9	$9/50 = 0,18 = 18\%$
3	8	$8/50 = 0,16 = 16\%$
4	9	$9/50 = 0,18 = 18\%$
5	6	$6/50 = 0,12 = 12\%$
6	8	$8/50 = 0,16 = 16\%$
Total	50	100%

Portanto, a frequência simples absoluta do primeiro elemento é igual a 20%.

Gabarito: B.

(CESPE/Polícia Federal/2018)

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

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

A tabela em questão descreve a distribuição de frequências da quantidade de drogas apreendidas nos cinco dias que constituem a amostra.

Comentários:

A tabela apenas registra as quantidades de drogas apreendidas por dia. Para que ela configurasse uma distribuição de frequências da quantidade de drogas apreendidas nos cinco dias, era necessário que cada valor estivesse associado a uma frequência, o que não ocorreu. Uma distribuição de frequência dos dados acima seria assim:

Valor	Frequência
10	1
18	1
22	2
28	1

Gabarito: Errado.

(CESPE/BACEN/2013)

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

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

Com base nesses dados, julgue o próximo item.

A distribuição de frequência acumulada para tempo de armazenagem observado na amostra inferior a 8 minutos é igual a 13, o que corresponde a uma frequência relativa superior a 0,60.

Comentários:

Não precisamos construir a distribuição de frequências para responder à questão. Sabemos que o total de observações é 20, pois é esse o número de termos.

Vamos destacar as observações que são inferiores a 8:

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

São 13 números inferiores a 8. Portanto, a frequência acumulada para um tempo menor do que 8 é 13.

Para calcular a frequência relativa, basta dividir a frequência absoluta pelo total de observações:

$$\frac{13}{20} = 0,65$$

Logo, a frequência relativa acumulada é maior do que 0,60

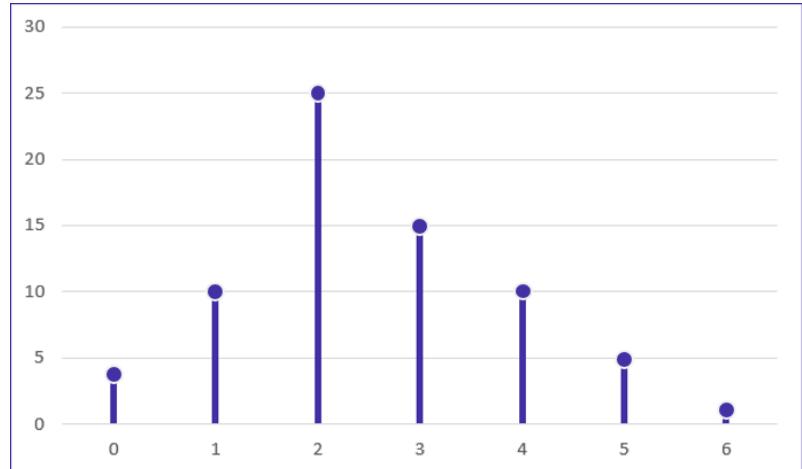
Gabarito: Certo.

REPRESENTAÇÕES GRÁFICAS DAS DISTRIBUIÇÕES DE FREQUÊNCIA

Gráfico de Hastes ou Bastões

O gráfico de hastes ou bastões é muito utilizado para representar dados não agrupados em classes, o que normalmente ocorre com dados discretos. Nesse caso, não há perda de informação pois os valores da variável aparecem individualmente, conforme constam da amostra. Com relação a sua construção, basta representarmos as frequências simples absolutas ou relativas de cada elemento do conjunto de dados.

X_i	Frequência (f_i)
0	4
1	10
2	25
3	15
4	10
5	5
6	1



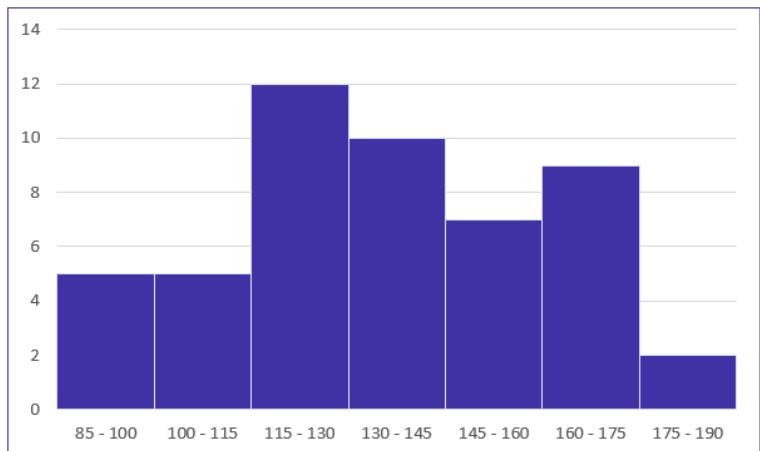
Repare que podemos reconstruir facilmente a tabela de frequências a partir do gráfico de hastes. De igual modo, conhecendo a tabela de frequências, podemos construir rapidamente o gráfico de hastes.

Histogramas

O histograma é um gráfico destinado a representar dados agrupados em classe, sendo composto por um conjunto de retângulos contíguos (justapostos) cujas bases ficam localizadas sobre o eixo horizontal (eixo x), de forma que os seus pontos médios devem coincidir com os pontos médios dos intervalos de classe e seus limites devem coincidir com os limites da classe.

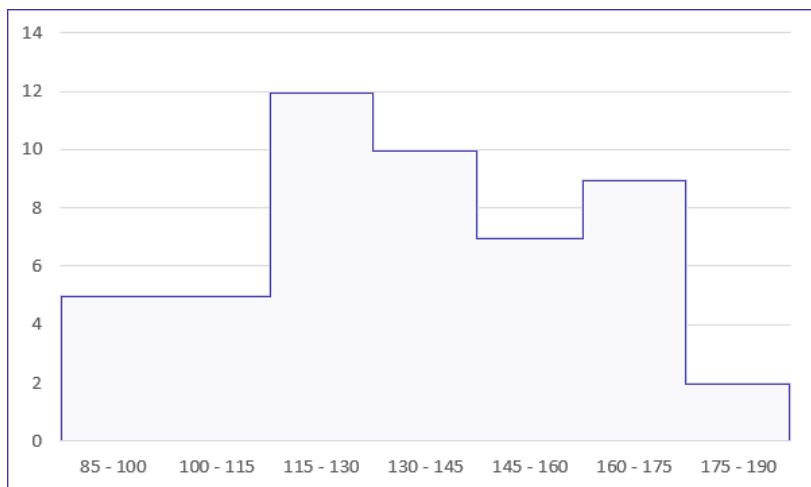
A quantidade de retângulos em um histograma é equivalente ao número de intervalos de classe. A largura de cada retângulo deve ser igual à amplitude do intervalo de classe, enquanto a altura precisa ser proporcional à frequência do intervalo de classe. Além disso, a área do histograma é proporcional ao somatório das frequências.

Tempo médio (X_i)	Frequência (f_i)
$85 \leq x < 100$	5
$100 \leq x < 115$	5
$115 \leq x < 130$	12
$130 \leq x < 145$	10
$145 \leq x < 160$	7
$160 \leq x < 175$	9
$175 \leq x < 190$	2



A diferença básica entre um **histograma** e um **gráfico de colunas** (estudaremos na próxima seção) é a **separação entre os retângulos adjacentes**. Veja que não existe separação entre os retângulos no caso do histograma.

Dito isso, é importante mencionarmos a existência do gráfico denominado de **poligonal característica**, que construímos utilizando apenas os contornos do histograma.



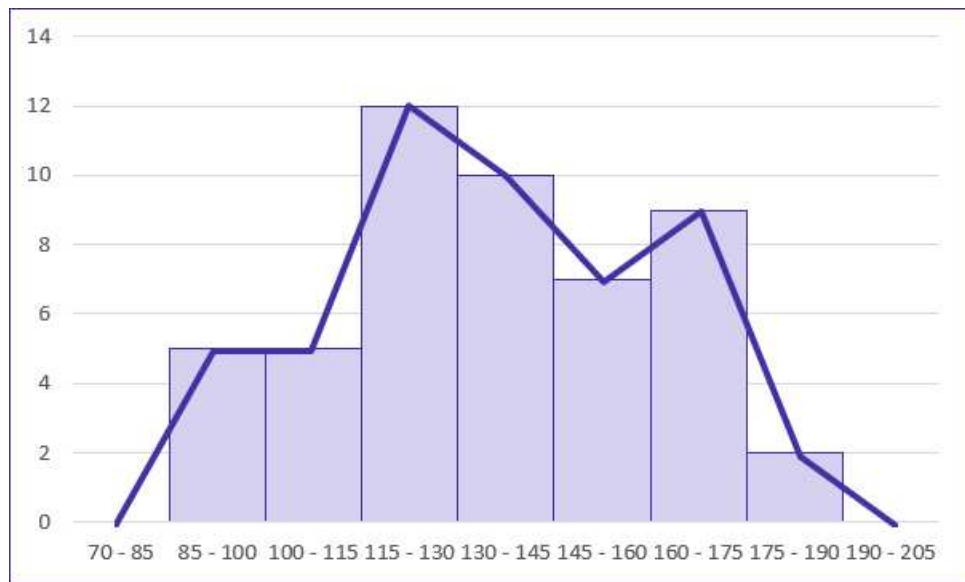
Por fim, o **histograma** pode ocasionar um certo nível de perda de informações, pois os elementos da **distribuição de frequência** não são representados de forma individualizada, mas sim por meio de suas **classes**.

Polígono de Frequências

O **polígono de frequências** é um gráfico em linha obtido por meio da ligação, por segmentos de reta, dos **pontos médios das bases superiores dos retângulos de um histograma**. Também é necessário considerar a existência de uma classe anterior à primeira e outra posterior à última, ambas com a frequência nula.

Assim como o histograma, o polígono de frequências apresenta área proporcional ao somatório das frequências.

Tempo médio (X_i)	Ponto médio (PM_i)	Frequência (f_i)
$70 \leq x < 85$	77,5	0
$85 \leq x < 100$	92,5	5
$100 \leq x < 115$	107,5	5
$115 \leq x < 130$	122,5	12
$130 \leq x < 145$	137,5	10
$145 \leq x < 160$	152,5	7
$160 \leq x < 175$	167,5	9
$175 \leq x < 190$	182,5	2
$190 \leq x < 205$	197,5	0



Curva de Frequências

A curva de frequências é obtida a partir do polimento de um polígono de frequências. Em sentido geométrico, o polimento corresponde à eliminação dos vértices (cantos) da linha poligonal. Esse processo suaviza os contornos do polígono de frequências, o que evidencia a verdadeira natureza dos dados em análise.

O polígono de frequências fornece a imagem real do fenômeno investigado, enquanto a curva de frequência mostra sua tendência. Naturalmente, quando o conjunto de dados é grande, a linha poligonal se torna curva.

Por isso, podemos afirmar que **a curva de frequência** antecipa o comportamento da distribuição para um número maior de dados.

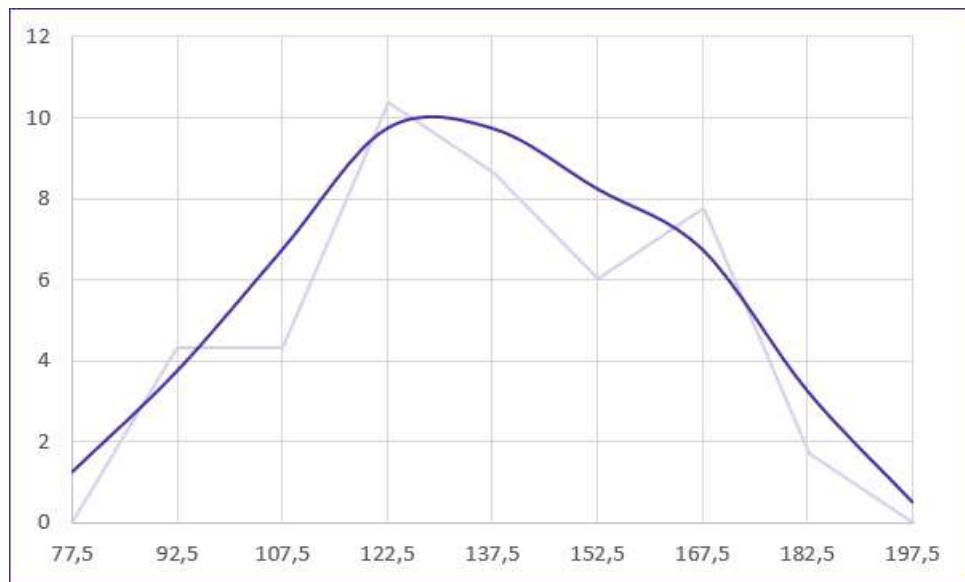
O processo de polimento é realizado por meio da seguinte fórmula:

$$fc_i = \frac{f_{ant} + 2 \times f_i + f_{post}}{4}$$

Em que fc_i é a frequência calculada da classe considerada (freq. polida); f_i é a frequência simples da classe considerada; f_{ant} é a frequência simples da classe anterior à da classe considerada; e f_{post} é a frequência simples da classe posterior à da classe considerada.

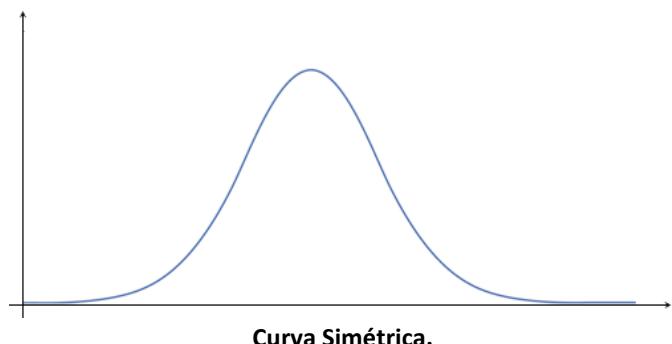
Vejamos como utilizá-la:

Tempo médio (X_i)	Ponto médio (PM_i)	Frequência (f_i)	Frequência Calculada (fc_i)
70 ≤ $x < 85$	77,5	0	$fc_0 = \frac{0 + 2 \times 0 + 5}{4} = 1,25$
85 ≤ $x < 100$	92,5	5	$fc_1 = \frac{0 + 2 \times 5 + 5}{4} = 3,75$
100 ≤ $x < 115$	107,5	5	$fc_2 = \frac{5 + 2 \times 5 + 12}{4} = 6,75$
115 ≤ $x < 130$	122,5	12	$fc_3 = \frac{5 + 2 \times 12 + 10}{4} = 9,75$
130 ≤ $x < 145$	137,5	10	$fc_4 = \frac{12 + 2 \times 10 + 7}{4} = 9,75$
145 ≤ $x < 160$	152,5	7	$fc_5 = \frac{10 + 2 \times 7 + 9}{4} = 8,75$
160 ≤ $x < 175$	167,5	9	$fc_6 = \frac{7 + 2 \times 9 + 2}{4} = 6,75$
175 ≤ $x < 190$	182,5	2	$fc_7 = \frac{9 + 2 \times 2 + 0}{4} = 3,25$
190 ≤ $x < 205$	197,5	0	$fc_8 = \frac{2 + 2 \times 0 + 0}{4} = 0,50$

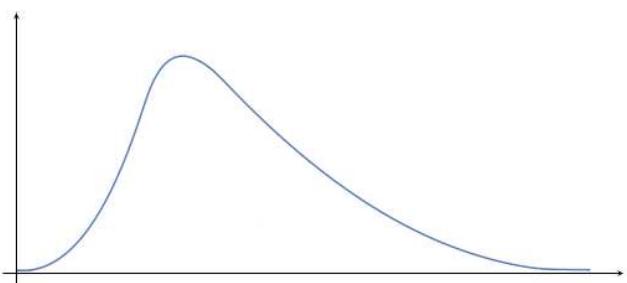


As curvas de frequências podem assumir as seguintes formas características:

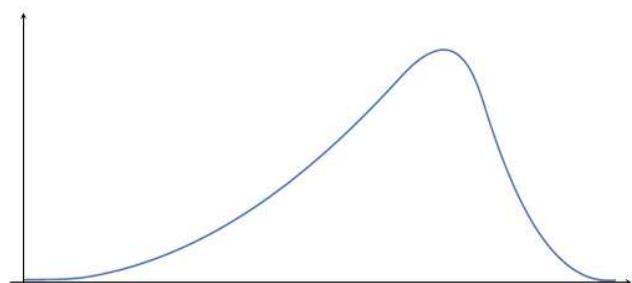
a) **curvas em forma de sino:** são curvas que apresentam concentração de valores em torno da região central da distribuição. Tais curvas podem ser simétricas ou assimétricas. Quando assimétricas, as curvas ainda podem apresentar uma cauda mais alongada à esquerda (assimetria à esquerda) ou mais alongada à direita (assimetria à direita). Assim, as possíveis configurações para as curvas em formas de sino são:



Curva Simétrica.

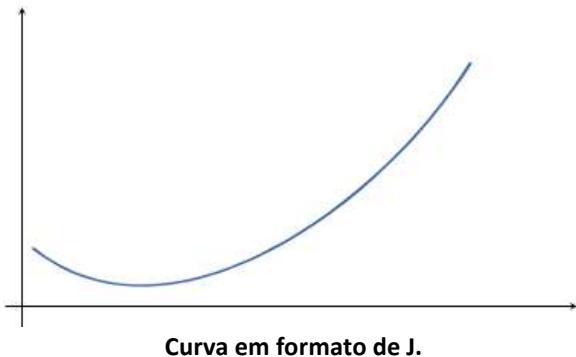


Curva assimétrica à direita.



Curva assimétrica à esquerda.

b) **curvas em forma de jota**: são curvas que apresentam o ponto de ordenada máxima em uma das extremidades, representando distribuições extremamente assimétricas. As possíveis configurações são:

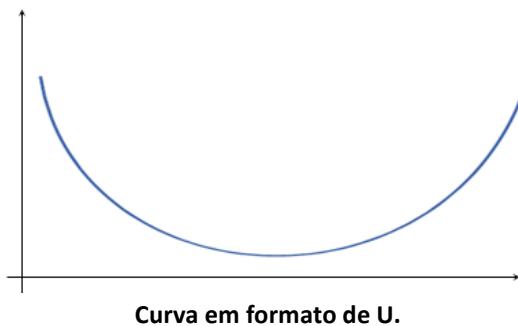


Curva em formato de J.



Curva em formato de J invertido.

c) **curvas em forma de U**: são curvas que apresentam as ordenadas máximas em ambas as extremidades.



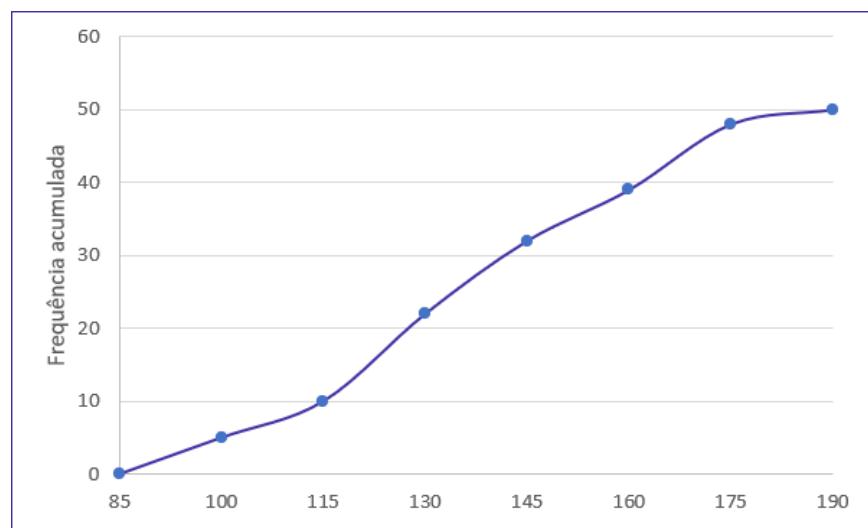
Curva em formato de U.

Ogiva (Polígono de Frequências Acumuladas)

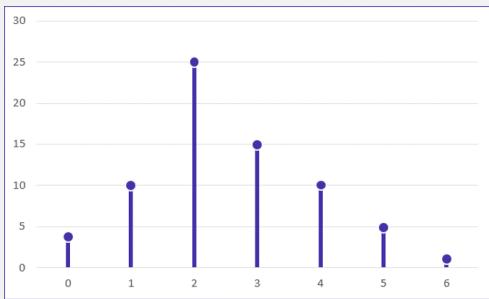
O gráfico de ogiva corresponde a um **polígono de frequências acumuladas**. Esse gráfico é empregado na representação de distribuições de frequências acumuladas, sejam elas crescentes ou decrescentes. No eixo horizontal, colocamos as extremidades de cada classe e no eixo vertical as frequências acumuladas.

Ao contrário do polígono de frequências, a **ogiva** utiliza os pontos extremos das classes, e não os pontos **médios**. Na construção do polígono de frequências acumuladas, devemos considerar a existência de uma classe anterior à primeira, com frequência nula.

Tempo médio (X_i)	Frequência (f_i)	Frequência Acumulada (f_{ac})
$70 \leq x < 85$	0	0
$85 \leq x < 100$	5	5
$100 \leq x < 115$	5	10
$115 \leq x < 130$	12	22
$130 \leq x < 145$	10	32
$145 \leq x < 160$	7	39
$160 \leq x < 175$	9	48
$175 \leq x < 190$	2	50

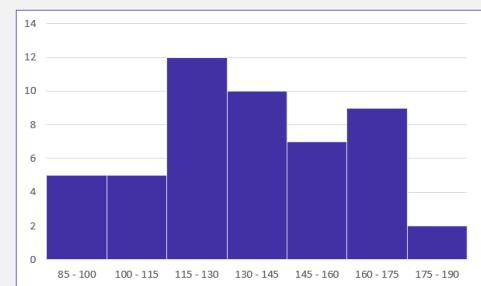


Gráfico

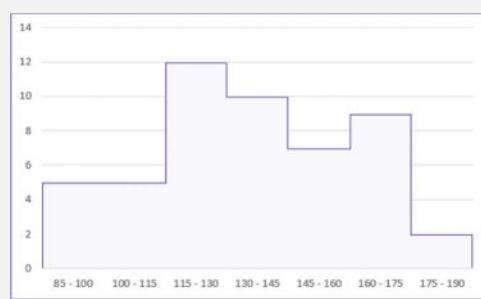


Definição

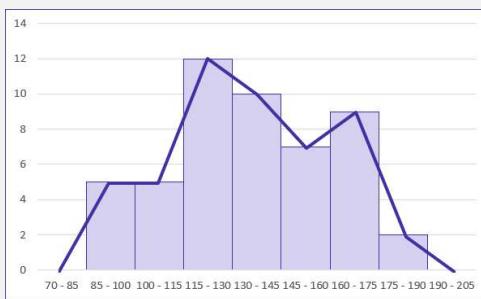
O **gráfico de hastes ou bastões** é muito utilizado para representar dados não agrupados em classes, o que normalmente ocorre com dados discretos.



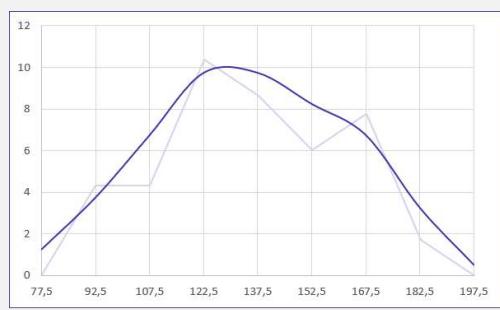
O **histograma** é um gráfico destinado a representar **dados agrupados em classe**, sendo composto por um conjunto de **retângulos contíguos (justapostos)**.



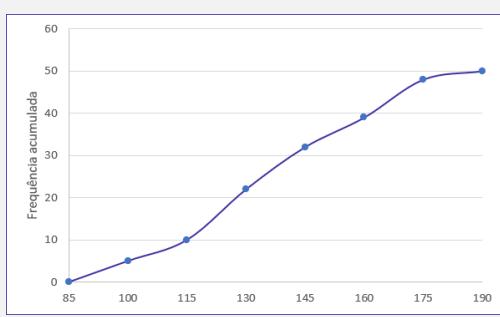
A **poligonal característica** é construída utilizando apenas os contornos do histograma.



O **polígono de frequências** é um gráfico em linha obtido por meio da ligação, por segmentos de reta, dos pontos médios das bases superiores dos retângulos de um histograma.



A **curva de frequências** é obtida a partir do polimento de um polígono de frequências

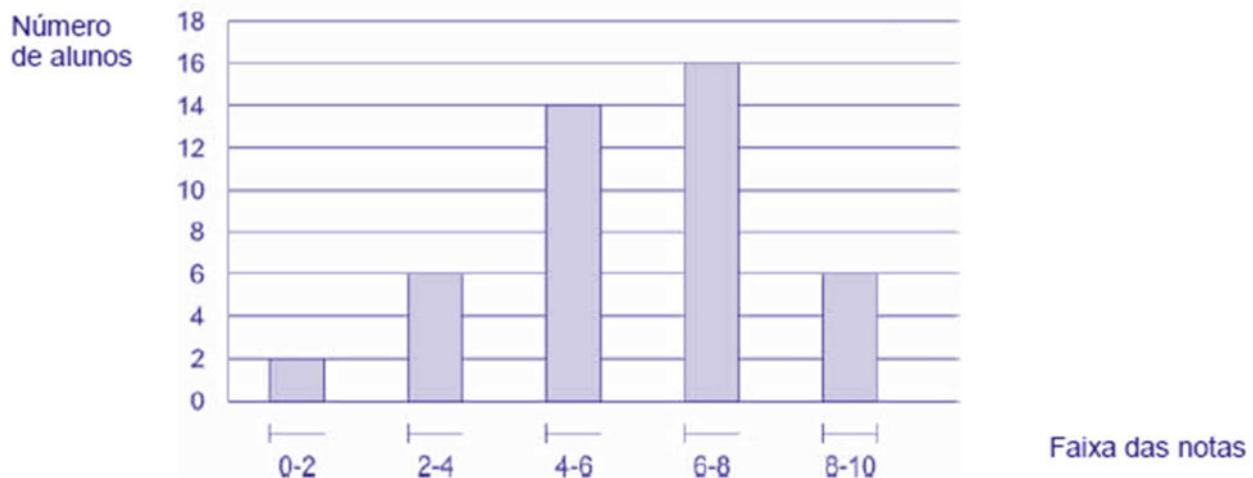


O **gráfico de ogiva** é empregado na representação de distribuições de frequências acumuladas, sejam elas crescentes ou decrescentes



(FCC/SEDU-ES/2018) Depois da aplicação de uma prova para todos os alunos do 7º A, a professora Marli fez um histograma com a distribuição das notas, conforme indicado abaixo.

Distribuição de notas do 7º A



Se a média de notas da sala foi igual a seis, a porcentagem dos alunos que tiraram nota maior ou igual a média da sala nessa prova foi de

- a) 54%.
- b) 50%.
- c) 56%.
- d) 60%.
- e) 52%.

Comentários:

Vamos, inicialmente, calcular o total de alunos:

$$2 + 6 + 14 + 16 + 6 = 44$$

Reparem na notação intervalar adotada para as colunas, vamos relembrar o significado dela:

Tipo de Intervalo	Notação matemática	Notação estatística	Significado
Intervalo fechado à esquerda e aberto à direita	$a \leq x < b$	$a \leftarrow b$	Engloba todos os elementos entre a e b , inclusive a mas não b .
Intervalo fechado	$a \leq x \leq b$	$a \top b$	Engloba todos os elementos entre a e b , inclusive a e b .

Agora, com base nas informações do gráfico, percebemos que existem 22 pessoas com média maior ou igual a 6, representadas pelas duas últimas colunas do gráfico. Assim:

$$16 + 6 = 22$$

Logo, a porcentagem será de:

$$\frac{22}{44} = 0,5 = 50\%$$

Gabarito: B.

(FCC/TRT 14ª Região/2018) De um histograma e uma tabela de frequências absolutas, elaborados para analisar a distribuição dos salários dos empregados em uma empresa, obtém-se a informação que 24 empregados ganham salários com valores pertencentes ao intervalo $(2.000; 4.000]$, em reais, que apresenta uma densidade de frequência de $0,75 \times 10^{-4}(R\$)^{-1}$

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

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

Comentários:

Vamos destacar os dados da questão:

- densidade de frequência: $0,75 \times 10^{-4}(R\$)^{-1}$;
- amplitude da classe: $4000 - 2000 = 2000$;
- número de funcionários na classe: 24; e
- total de funcionários na classe: T.

A frequência relativa é calculada pela fórmula a seguir:

$$\text{Frequência relativa} = \frac{\text{Nº de funcionários na classe}}{\text{Total de funcionários}} = \frac{24}{T}$$

Então, pela fórmula da densidade de frequência, teremos:

$$\text{Densidade de frequência} = \frac{\text{Frequência relativa}}{\text{Amplitude}}$$

$$0,75 \times 10^{-4} = \frac{\frac{24}{T}}{2000}$$

$$0,15 = \frac{24}{T}$$

$$T = 160$$

Agora, vamos aplicar a mesma metodologia para a classe do histograma com amplitude de 3000 e densidade de frequência de 1×10^{-4} . Considerando o número de funcionários na classe igual a N , teremos a seguinte frequência relativa:

$$\text{Frequência relativa} = \frac{\text{Nº de funcionários na classe}}{\text{Total de funcionários}} = \frac{N}{160}$$

Então, pela fórmula da densidade de frequência, teremos:

$$\text{Densidade de frequência} = \frac{\text{Frequência relativa}}{\text{Amplitude}}$$

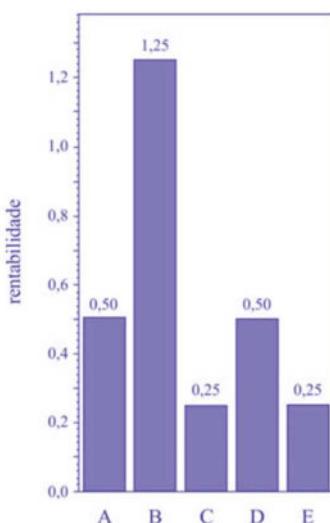
$$10^{-4} = \frac{\frac{N}{160}}{3000}$$

$$0,3 = \frac{N}{160}$$

$$N = 48$$

Gabarito: D.

(CESPE/FUNPRESP/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.

Comentários:

O gráfico não é um histograma por dois motivos: primeiro porque há uma separação entre as colunas, o que não ocorre em um histograma, e sim em um gráfico de colunas; segundo porque um histograma representa dados que estão agrupados em intervalos de classe, e não em categorias, como é o caso.

Gabarito: Errado.

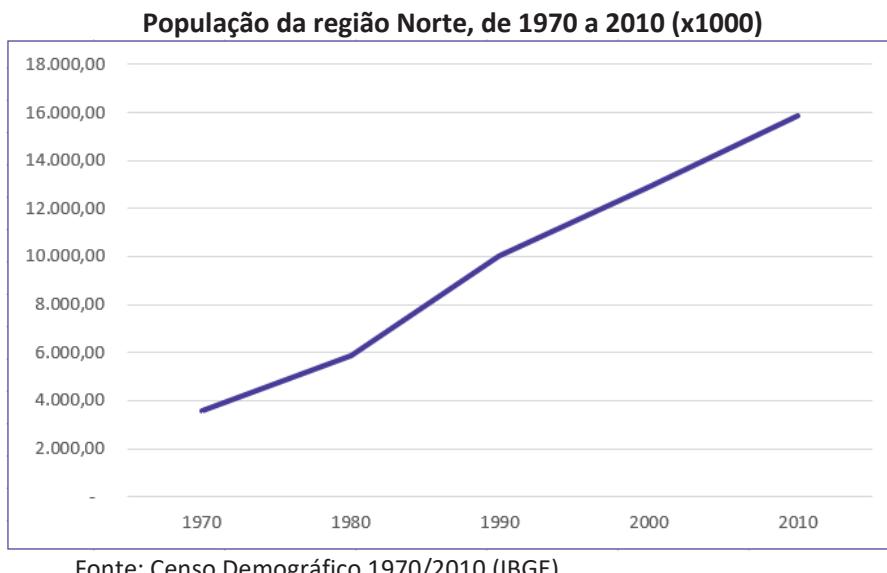
OUTROS GRÁFICOS E REPRESENTAÇÕES

O principal objetivo dos gráficos estatísticos é proporcionar uma visualização mais rápida dos dados estatísticos ou do fenômeno sob investigação. A seguir vamos ver as principais formas de representação de dados estatísticos.

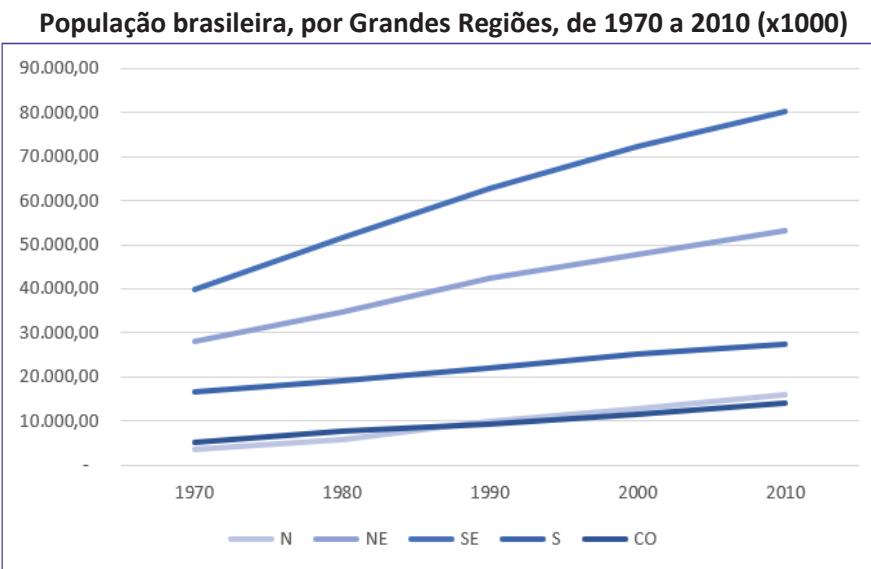
Gráficos em Linhas

Os gráficos em linha normalmente são usados para representar dados de **séries temporais**, com a finalidade de mostrar a variação dos valores de uma variável ao longo do tempo. Esse tipo de gráfico permite-nos comparar duas variáveis: uma é traçada no eixo x (horizontal) e a outra no eixo y (vertical). O eixo y geralmente indica uma quantidade, enquanto o eixo x representa uma unidade de tempo.

Por exemplo, o gráfico a seguir mostra a evolução da população residente na Região Norte do Brasil no período de 1970 a 2010. Veja que o eixo horizontal está indicando o decurso do tempo, enquanto o eixo vertical apresenta a população residente na região.



Também podemos elaborar um **gráfico de linhas múltiplas** para comparar a evolução da população residente nas Grandes Regiões do Brasil em diferentes períodos:

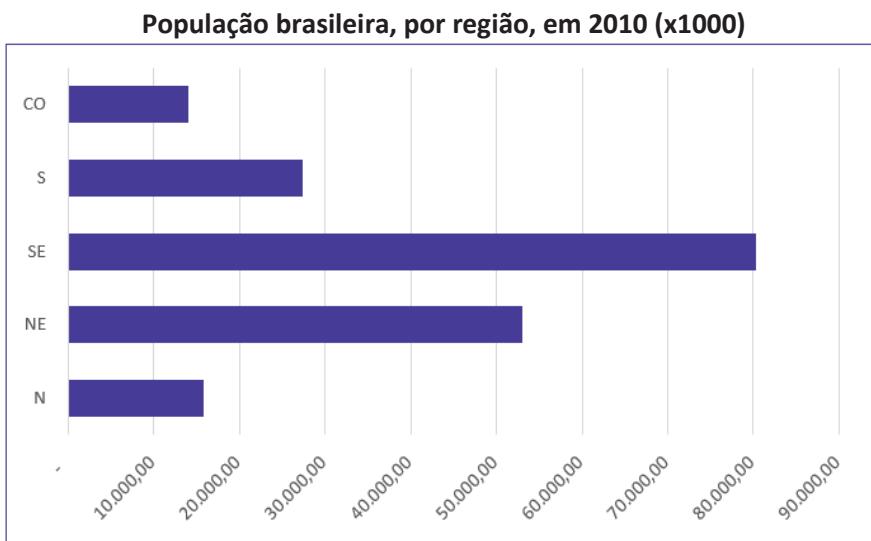


Fonte: Censo Demográfico 1970/2010 (IBGE)

Gráficos em Barras

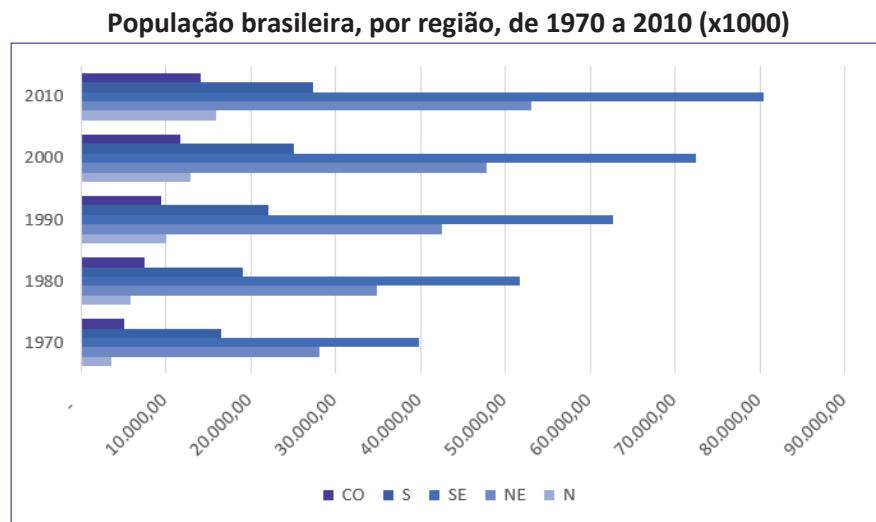
Os gráficos em barra normalmente **são usados para representar distribuições de dados categóricos ou qualitativos**. Uma série estatística é representada por um **conjunto de retângulos dispostos horizontalmente**, cada um indicando uma categoria particular, **os quais possuem a mesma altura e comprimentos proporcionais aos respectivos dados**.

Por exemplo, a distribuição da população residente em cada região brasileira, no ano de 2010, pode ser representada por meio do seguinte gráfico:



Fonte: Censo Demográfico 1970/2010 (IBGE)

Adicionalmente, podemos utilizar um **gráfico de barras justapostas** para representar a evolução da população residente em cada região brasileira, como o mostrado a seguir:

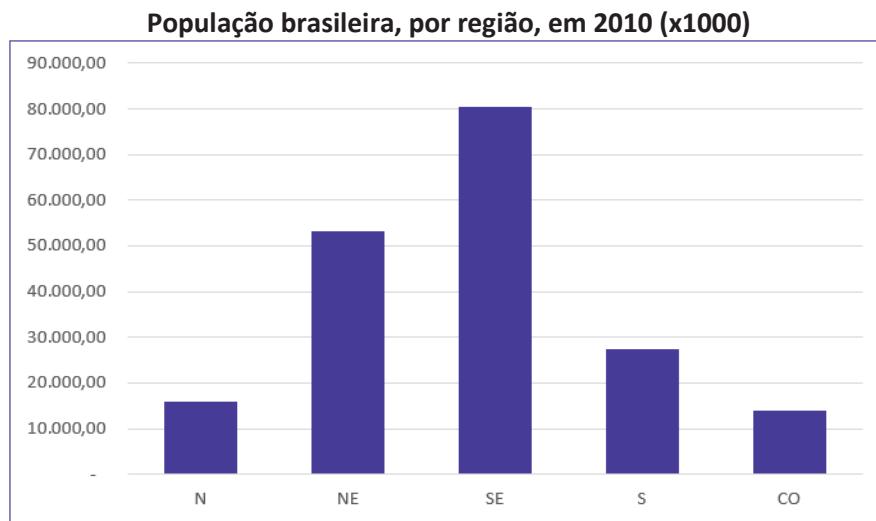


Fonte: Censo Demográfico 1970/2010 (IBGE)

Gráficos em Colunas

Os **gráficos em coluna** também são usados para **distribuições de dados categóricos ou qualitativos**. A diferença básica é que, agora, uma série estatística é representada por um conjunto de retângulos dispostos **verticalmente**, cada um indicando uma categoria particular, **todos com a mesma largura e alturas proporcionais aos respectivos dados**.

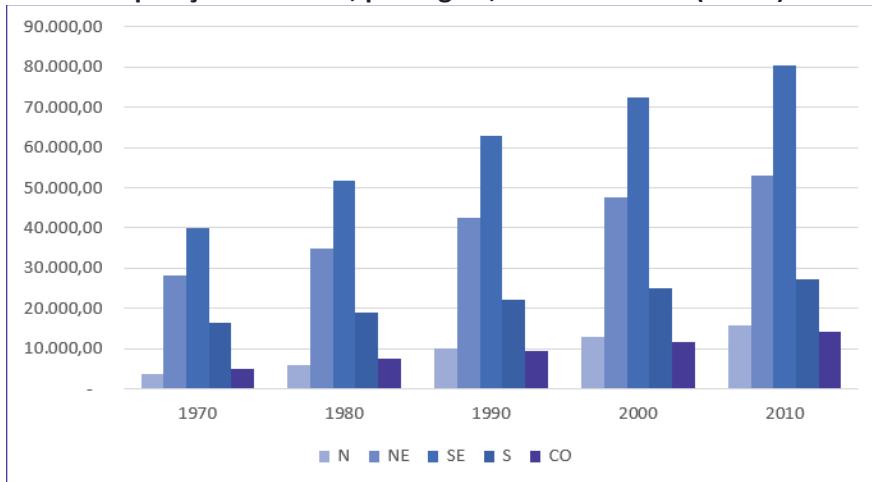
Por exemplo, a distribuição da população residente em cada região brasileira, no ano de 2010, pode ser representada por meio do seguinte gráfico:



Fonte: Censo Demográfico 1970/2010 (IBGE)

Também podemos utilizar um **gráfico de colunas justapostas** para representar a evolução da população residente em cada região brasileira, no período de 1970 a 2010. Dessa maneira, conseguimos apresentar mais informações em um espaço consideravelmente menor. Vejamos o gráfico a seguir:

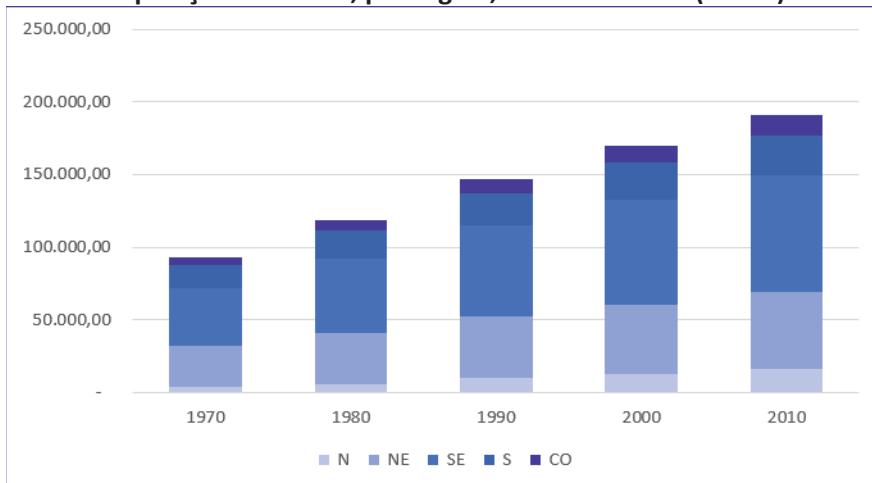
População brasileira, por região, de 1970 a 2010 (x1000)



Fonte: Censo Demográfico 1970/2010 (IBGE)

Adicionalmente, essas informações também podem ser representadas por meio de um **gráfico de colunas sobrepostas** (ou **gráfico de colunas empilhadas**). Esse tipo de gráfico é considerado uma extensão do formato tradicional, pois **permite analisarmos duas dimensões de uma variável categórica**, ao invés de apenas uma. Cada coluna é dividida em várias partes que ficam empilhadas umas sobre as outras, cada uma correspondendo a um nível da segunda variável categórica.

População brasileira, por região, de 1970 a 2010 (x1000)

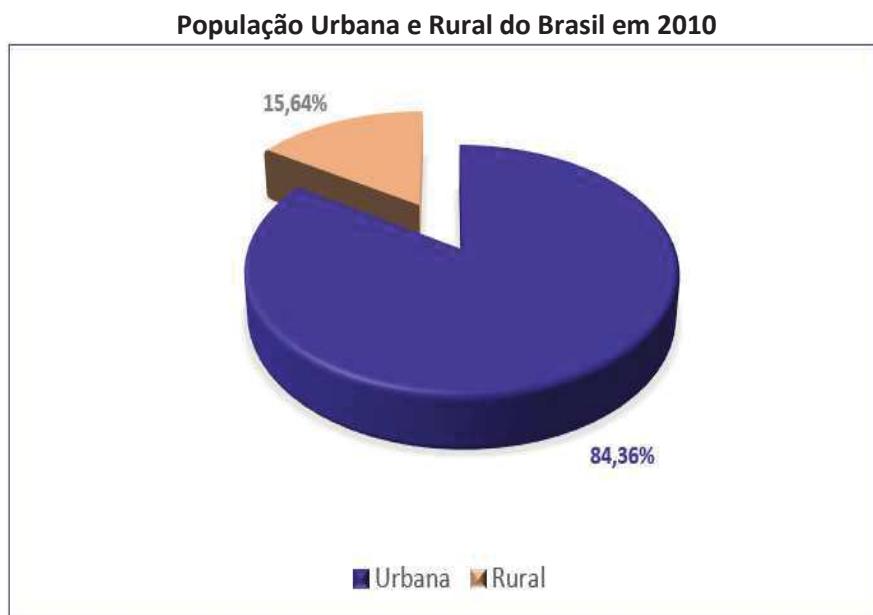


Fonte: Censo Demográfico 1970/2010 (IBGE)

Gráfico em Setores

O gráfico em setores (também conhecido como gráfico de pizza) é usado para representar a frequência relativa (porcentagem) de uma variável categórica. Ele é formado por um círculo dividido em setores circulares, cada um representando uma categoria, cujos ângulos centrais são proporcionais às frequências relativas da categoria.

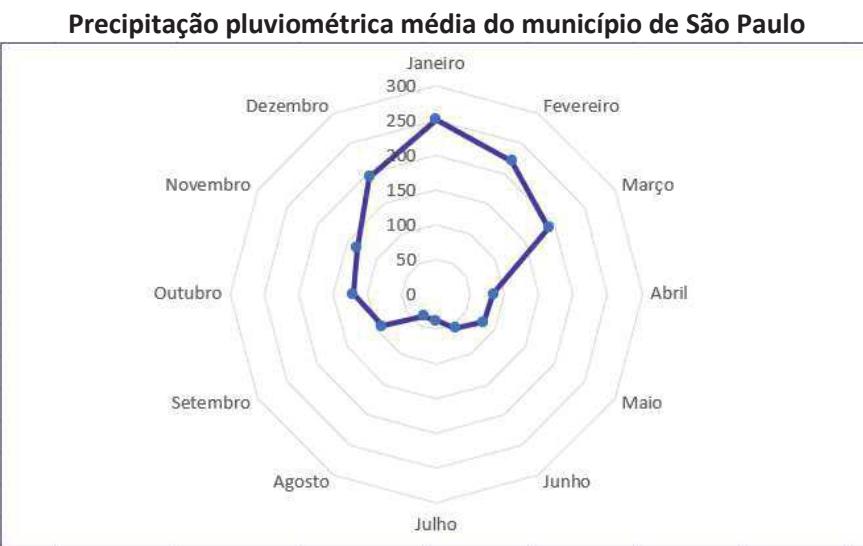
Para a construção do gráfico de setores, utilizaremos uma regra de três simples, em que as frequências relativas de cada categoria correspondem ao ângulo central que desejamos representar em relação à frequência total, que corresponde ao ângulo de 360°.



Fonte: Sinopse do Censo Demográfico de 2010 (IBGE)

Gráfico Polar

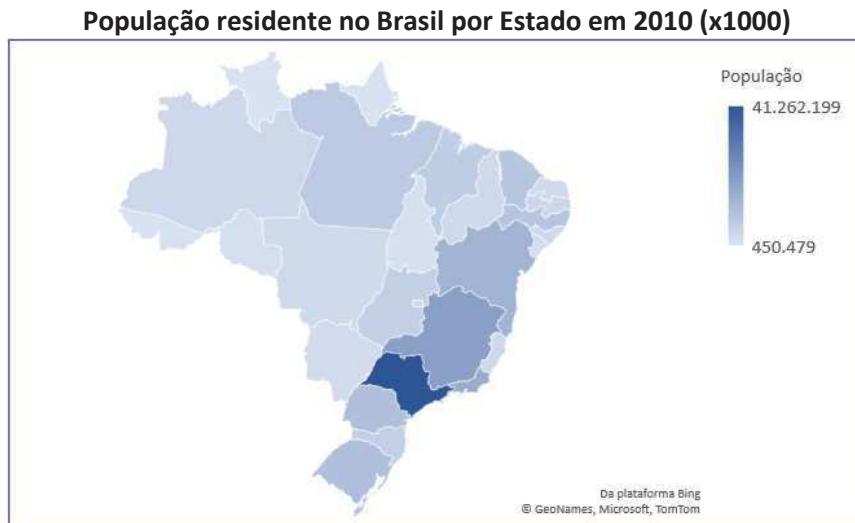
O gráfico polar consiste em uma sequência de eixos igualmente espaçados (ângulos iguais), cada um representando uma das variáveis. Uma linha é desenhada ligando os valores de cada eixo. Esse tipo de gráfico é usado para representar **séries temporais cíclicas**, que apresentam uma determinada periodicidade, como é o caso da precipitação pluviométrica mensal média:



Fonte: Atlas Pluviométrico do Brasil (Serviço Geológico do Brasil)

Cartograma

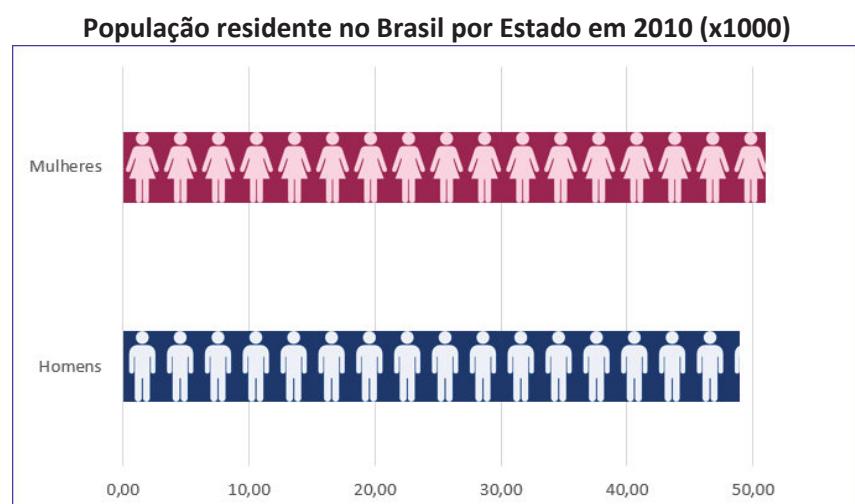
O **cartograma** é empregado com a finalidade de apresentar dados estatísticos diretamente relacionados com **áreas geográficas**. As áreas do cartograma podem ser preenchidas por pontos, hachuras ou cores. O significado do preenchimento será indicado em uma legenda. Vejamos a população residente em cada Estado brasileiro:



Fonte: Sinopse do Censo Demográfico de 2010 (IBGE)

Pictograma

O **pictograma** substitui valores por ícones, tornando os dados mais atraentes e facilitando o entendimento acerca de um determinado fenômeno. Normalmente, uma legenda é utilizada para indicar o que cada ícone representa. Os ícones devem possuir o mesmo tamanho, mas podem aparecer fracionados para mostrar a respectiva fração de uma determinada quantidade. A proporção de homens e mulheres na população brasileira é apresentada no pictograma a seguir:

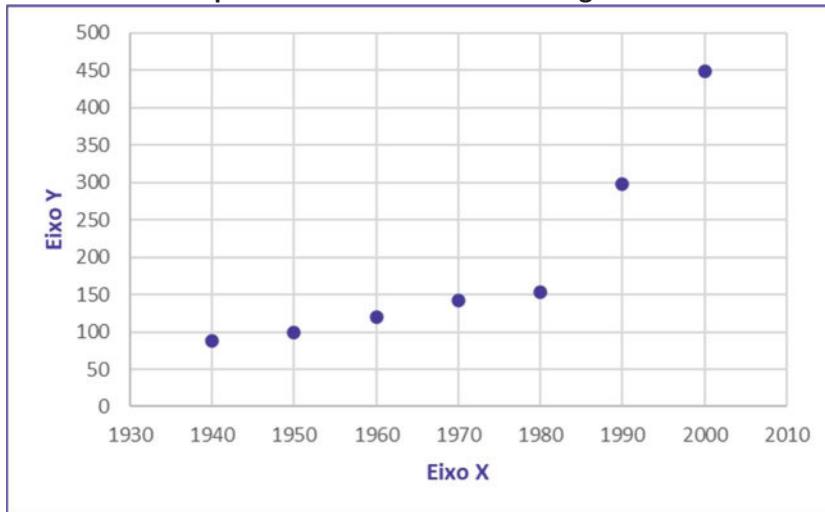


Fonte: Sinopse do Censo Demográfico de 2010 (IBGE)

Gráfico de Dispersão

O gráfico de dispersão é uma representação de pares ordenados em um plano cartesiano, composto por um eixo vertical (ordenada) e um eixo horizontal (abcissa). Os dados são representados como uma coleção de pontos, cada um com o valor de uma variável determinando a posição no eixo horizontal e o valor da outra variável determinando a posição no eixo vertical. É uma ferramenta poderosa para estudar a relação entre duas variáveis.

Municípios criados e instalados na região Norte.



Fonte: Sinopse do Censo Demográfico de 2010 (IBGE)

Diagrama de Ramos e Folhas

O diagrama de ramos e folhas fornece uma maneira rápida de representar graficamente a distribuição dos dados. Nesse diagrama, cada número é separado em duas partes. Em geral, de um lado ficam as unidades do número e do outro lado fica o restante desse número. Consideremos o seguinte rol crescente:

85, 89, 96, 98, 99, 103, 104, 105, 113, 114, 115, 115, 123, 123, 124, 126, 126, 126, 126, 127, 127, 128, 129, 129, 134, 135, 135, 135, 137, 137, 137, 142, 143, 143, 148, 153, 154, 155, 157, 158, 159, 161, 161, 165, 168, 170, 171, 171, 171, 173, 175, 175

A representação utilizando um diagrama de ramos e folhas ficaria assim:

8	5	9	Chave: 8 5 = 85
9	6	8	9
10	3	4	5
11	3	4	5 5
12	3	3	4 6 6 6 7 8 9 9
13	4	5	5 5 7 7 7
14	2	3	3 8
15	3	4	5 7 8 9
16	1	1	5 8
17	0	1	1 1 3 5 5

Repare que, no lado esquerdo, temos as centenas e as dezenas representando os ramos. Por sua vez, no lado direito, temos as unidades representando as folhas. As folhas, portanto, estão vinculadas aos ramos. Dessa maneira, a chave “9 | 6 8 9” significa que, no rol original, havia os números **96, 98 e 99**.

Também é comum encontrarmos diagramas de ramos e folhas em que as unidades são separadas em dois grupos: de 0 a 4 e de 5 a 9. Nesse caso, teríamos o seguinte diagrama:

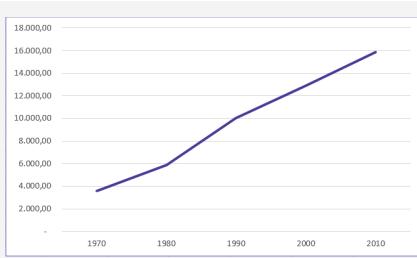
Chave: 8 5 = 85	
8	5 9
9	
9	6 8 9
10	3 4
10	5
11	3 4
11	5 5
12	3 3 4
12	6 6 6 7 8 9 9
13	4
13	5 5 5 7 7 7
14	2 3 3
14	8
15	3 4
15	5 7 8 9
16	1 1
16	5 8
17	0 1 1 1 3
17	5 5

Por fim, é importante observarmos que não existe uma regra única para a construção do diagrama de ramos e folhas. O formato mais comumente encontrado é o que separa o número em duas partes, porém, a depender da chave escolhida, o número pode ser separado em mais partes. Assim, ao resolver questões que envolvem esse tipo de diagrama, devemos observar a chave adotada.

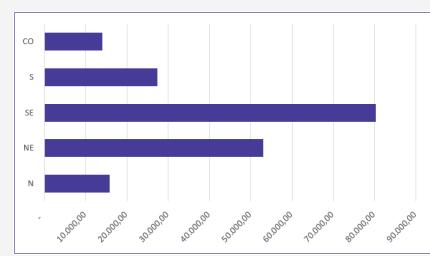


Gráfico

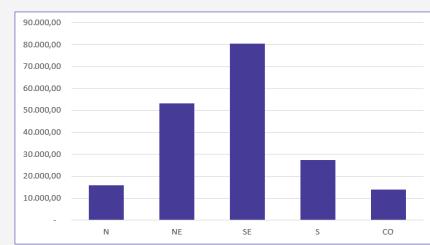
Definição



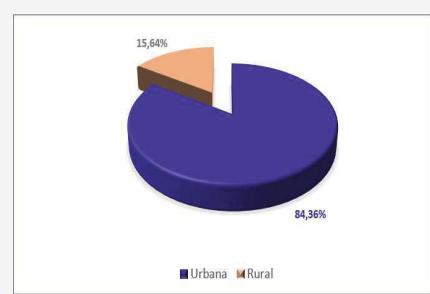
Os **gráficos em linha** normalmente são usados para representar a variação dos valores de uma variável ao longo do tempo. Esse tipo de gráfico permite-nos comparar duas variáveis: uma é traçada no eixo x (horizontal) e a outra no eixo y (vertical).



Os **gráficos em barra** normalmente são usados para representar distribuições de dados categóricos ou qualitativos. Uma série estatística é representada por um conjunto de retângulos dispostos **horizontalmente**, cada um indicando uma categoria particular.



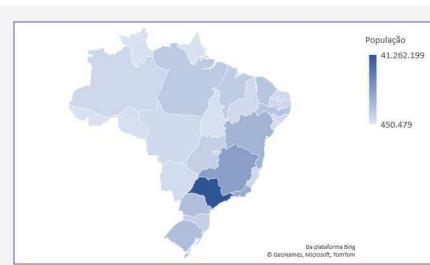
Os **gráficos em coluna** também são usados para distribuições de dados categóricos ou qualitativos. A diferença básica é que, agora, uma série estatística é representada por um conjunto de retângulos dispostos **verticalmente**, cada um indicando uma categoria particular.



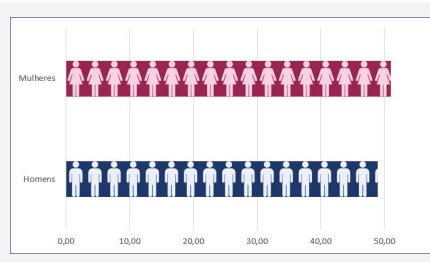
O **gráfico em setores** é usado para representar a frequência relativa (porcentagem) de uma variável categórica, sendo formado por um círculo dividido em setores circulares, cada um representando uma categoria.



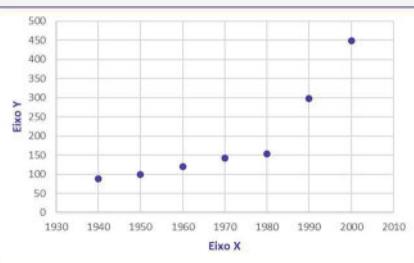
O **gráfico polar** consiste em uma sequência de eixos igualmente espaçados (ângulos iguais), cada um representando uma das variáveis. Uma linha é desenhada ligando os valores de cada eixo.



O **cartograma** é empregado com a finalidade de apresentar dados estatísticos diretamente relacionados com áreas geográficas.



O **pictograma** substitui valores por ícones, tornando os dados mais atraentes e facilitando o entendimento acerca de um determinado fenômeno.



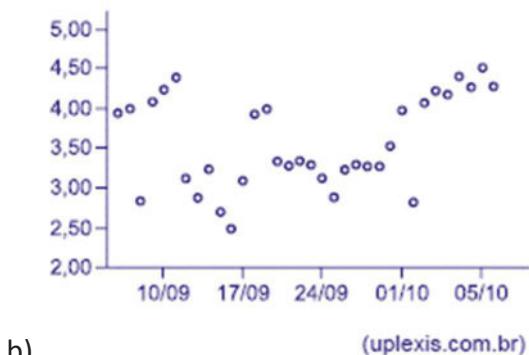
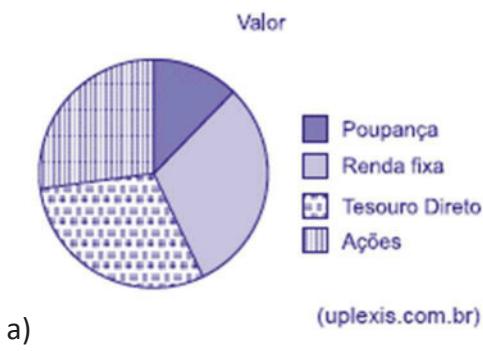
O **gráfico de dispersão** é uma representação de pares ordenados em um plano cartesiano, composto por um eixo vertical (ordenada) e um eixo horizontal (abcissa). Os dados são representados como uma coleção de pontos.

8	5 9	Chave: 8 5 = 85
9	6 8 9	
10	3 4 5	
11	3 4 5 5	
12	3 3 4 6 6 6 7 8 9 9	
13	4 5 5 5 7 7 7	
14	2 3 3 8	
15	3 4 5 7 8 9	

O **diagrama de ramos e folhas** fornece uma maneira rápida de representar graficamente a distribuição dos dados. Nele, cada número é separado em duas partes. Em geral, de um lado ficam as unidades do número e do outro lado fica o restante desse número.



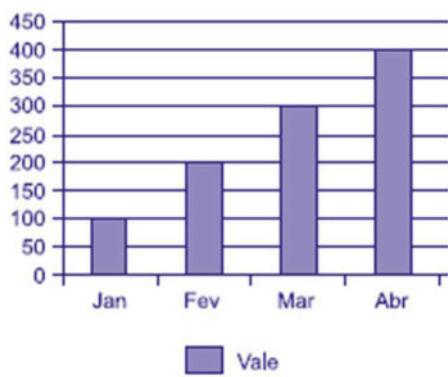
(VUNESP/UNICAMP/2019) Assinale dentre os exemplos a seguir, o gráfico de dispersão.





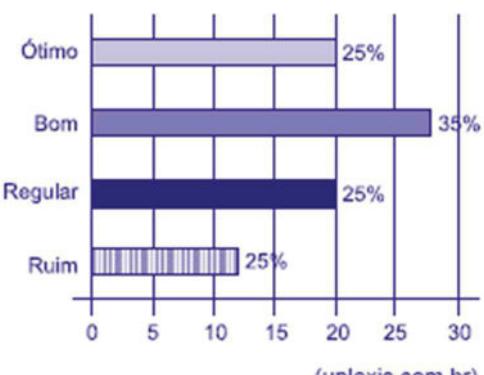
c) (uplexis.com.br)

Valor



d) (uplexis.com.br)

Desempenho em Matemática



e) (uplexis.com.br)

Comentários:

Analizando cada gráfico, temos:

Letra A: Alternativa Errada. O gráfico de setores é formado por um círculo dividido em setores circulares, cada um representando uma categoria, cujos ângulos centrais são proporcionais às frequências relativas da categoria.

Letra B: Alternativa Correta. O gráfico de dispersão é uma representação gráfica que analisa a relação entre duas variáveis quantitativas — uma de causa e outra de efeito, chamadas de variável independente e dependente, respectivamente. Esse tipo de diagrama traz números simultâneos das duas variáveis, deixando visível se o que acontece em uma variável interfere na outra.

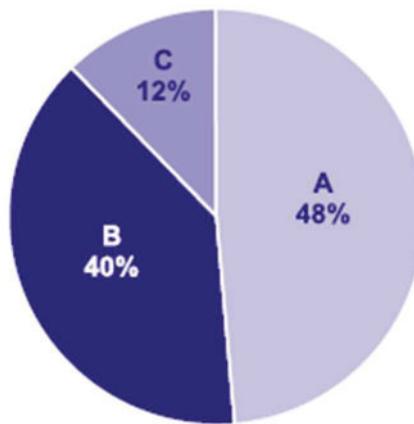
Letra C: Alternativa Errada. O gráfico em linhas é o gráfico em que os pontos são geralmente usados para controlar alterações ao longo do tempo e para facilitar a identificação de tendências ou de anomalias.

Letra D: Alternativa Errada. Os gráficos em colunas assim como os gráficos de barras são usados para distribuições de dados categóricos ou qualitativos. A diferença básica é que, no primeiro, uma série estatística é representada por um conjunto de retângulos dispostos verticalmente, cada um indicando uma categoria particular, todos com a mesma largura e alturas proporcionais aos respectivos.

Letra E: Alternativa Errada. Os gráficos em barras normalmente são usados para representar distribuições de dados categóricos ou qualitativos. Uma série estatística é representada por um conjunto de retângulos dispostos horizontalmente, cada um indicando uma categoria particular, os quais possuem a mesma altura e comprimentos proporcionais aos respectivos dados.

Gabarito: B.

(VUNESP/Pref. Campinas/2019) Uma empresa atua em três segmentos de mercado, A, B e C. O gráfico de setores mostra a distribuição percentual, por segmento, da receita total obtida por essa empresa em 2018.



Sabendo-se que a receita obtida no segmento A superou a receita obtida no segmento B em R\$ 64 milhões, é correto afirmar que a receita obtida no segmento C foi igual a

- a) R\$ 98 milhões.
- b) R\$ 96 milhões.
- c) R\$ 94 milhões.
- d) R\$ 88 milhões.
- e) R\$ 86 milhões.

Comentários:

Consideremos T o total de receitas obtidas pela empresa em 2018. Pelas informações apresentadas no gráfico de pizza, temos:

$$A = 0,48 \times T$$

$$B = 0,4 \times T$$

$$C = 0,12 \times T$$

Sabendo que a receita obtida no segmento A superou a receita obtida no segmento B em R\$ 64 milhões, então:

$$A = B + 64$$

$$0,48 \times T = 0,4 \times T + 64$$

$$0,8 \times T = 64$$

$$T = 800 \text{ milhões}$$

Logo,

$$C = 0,12 \times T$$

$$C = 12 \times 800$$

$$C = 96 \text{ milhões}$$

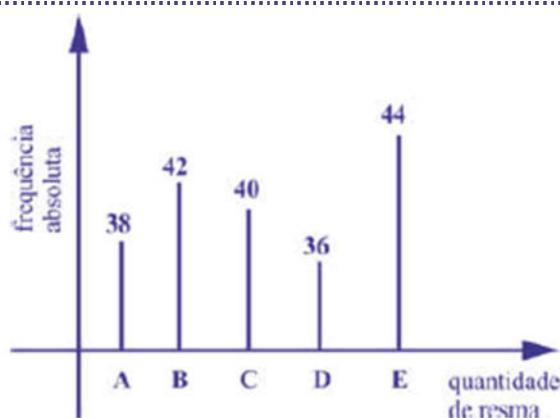
Gabarito: B.

(CESPE/CBM-AL/2017) Na tabela a seguir, A, B, C, D e E são as quantidades de resmas de papel A4 consumidas, em quatro meses, pelas seções administrativas I, II, III, IV e V, respectivamente. Apesar de não mostrar explicitamente essas quantidades, a tabela apresenta as frequências absolutas e (ou) relativas de algumas dessas quantidades.

Seção	Quantidades de Resmas	Frequência Absoluta	Frequência Relativa
I	A	38	19%
II	B		
III	C		20%
IV	D	36	
V	E	44	
	Total		100%

Considerando que cada uma dessas resmas, juntamente com a embalagem, tem forma de um paralelepípedo retângulo reto que mede 5 cm × 21 cm × 30 cm, julgue o item seguinte.

O gráfico de barras verticais a seguir apresenta as frequências absolutas de resmas consumidas pelas cinco seções.



Comentários:

Podemos usar uma regra de três simples para completar a tabela. Primeiro, vamos encontrar o valor que representa 100%:

$$38 - 0,19$$

$$Q - 1,00$$

$$Q = \frac{38 \times 1,00}{0,19} = 200$$

Encontrando o valor de 20%:

$$200 - 1,00$$

$$C - 0,20$$

$$C = \frac{200 \times 0,2}{1,00} = 40$$

Agora, basta sabermos a frequência absoluta na seção II. Como foram consumidas 200 resmas, então o número de resmas consumidas pela seção II foi:

$$200 - 38 - 40 - 36 - 44 = 42$$

Dessa forma, nas seções I, II, III, IV e V foram consumidas, respectivamente 38, 42, 40, 36 e 44 resmas de papel A4, conforme mostra o gráfico.

Gabarito: Certo.

RESUMO DA AULA

INTRODUÇÃO À ESTATÍSTICA

A Estatística pode ser dividida em três grandes ramos: **Estatística Descritiva** (ou dedutiva), **Estatística Probabilística** e **Estatística Inferencial** (ou indutiva).

ESTATÍSTICA DESCRIPTIVA

É responsável pela coleta, organização, descrição e resumo dos dados observados.

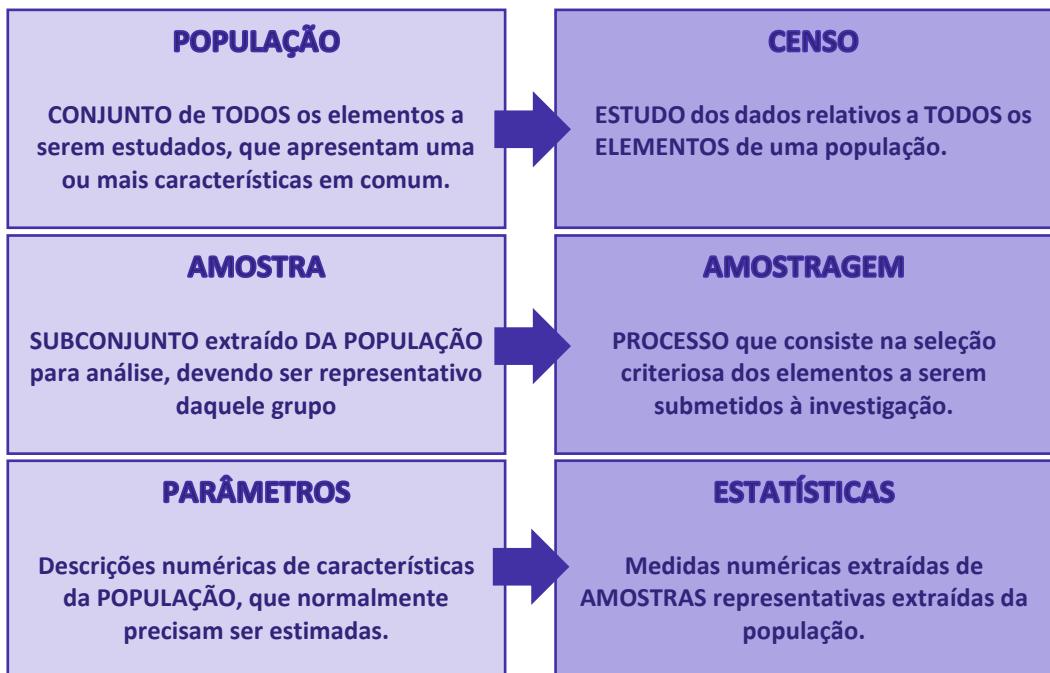
ESTATÍSTICA PROBABILÍSTICA

É responsável por estabelecer o modelo matemático adotado para explicar fenômenos aleatórios.

ESTATÍSTICA INFERENCIAL

É responsável pela análise e interpretação dos dados.

CONCEITOS INICIAIS



MÉTODO EXPERIMENTAL X MÉTODO ESTATÍSTICO

MÉTODO EXPERIMENTAL

As CAUSAS são mantidas **CONSTANTES**, **COM EXCEÇÃO DE UMA**, que é **VARIADA** para que seus efeitos sejam descobertos.

MÉTODO ESTATÍSTICO

Admite e **REGISTRA TODAS AS POSSÍVEIS VARIAÇÕES DAS CAUSAS PRESENTES**, procurando determinar a influência de cada fator no resultado.

DADOS ESTATÍSTICOS

Com relação ao número de observações coletadas, os dados são classificados em **univariados**, **bivariados** ou **multivariados**:

DADOS UNIVARIADOS

É quando uma única observação de cada indivíduo é registrada.

DADOS BIVARIADOS

É quando duas observações de cada indivíduo são registradas.

DADOS MULTIVARIADOS

É quando mais duas observações acerca de cada indivíduo são registradas

Quanto à forma de apresentação, os dados podem ser classificados em **dados brutos** ou **rol**.

DADOS BRUTOS

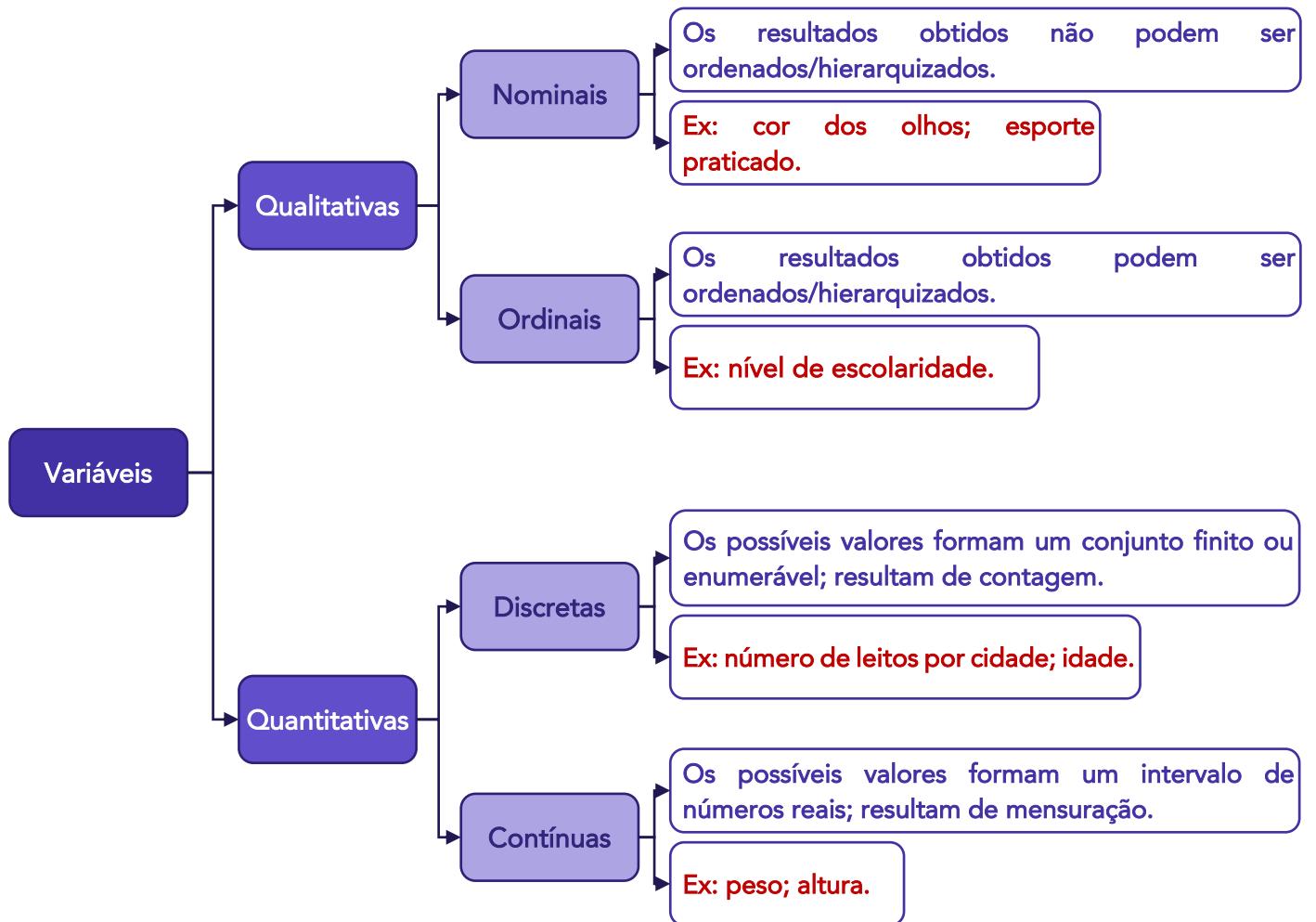
São aqueles que não foram numericamente organizados em ordem crescente ou decrescente, ou seja, estão na forma como foram coletados.

ROL

É a organização dos dados brutos em ordem de grandeza crescente ou decrescente.

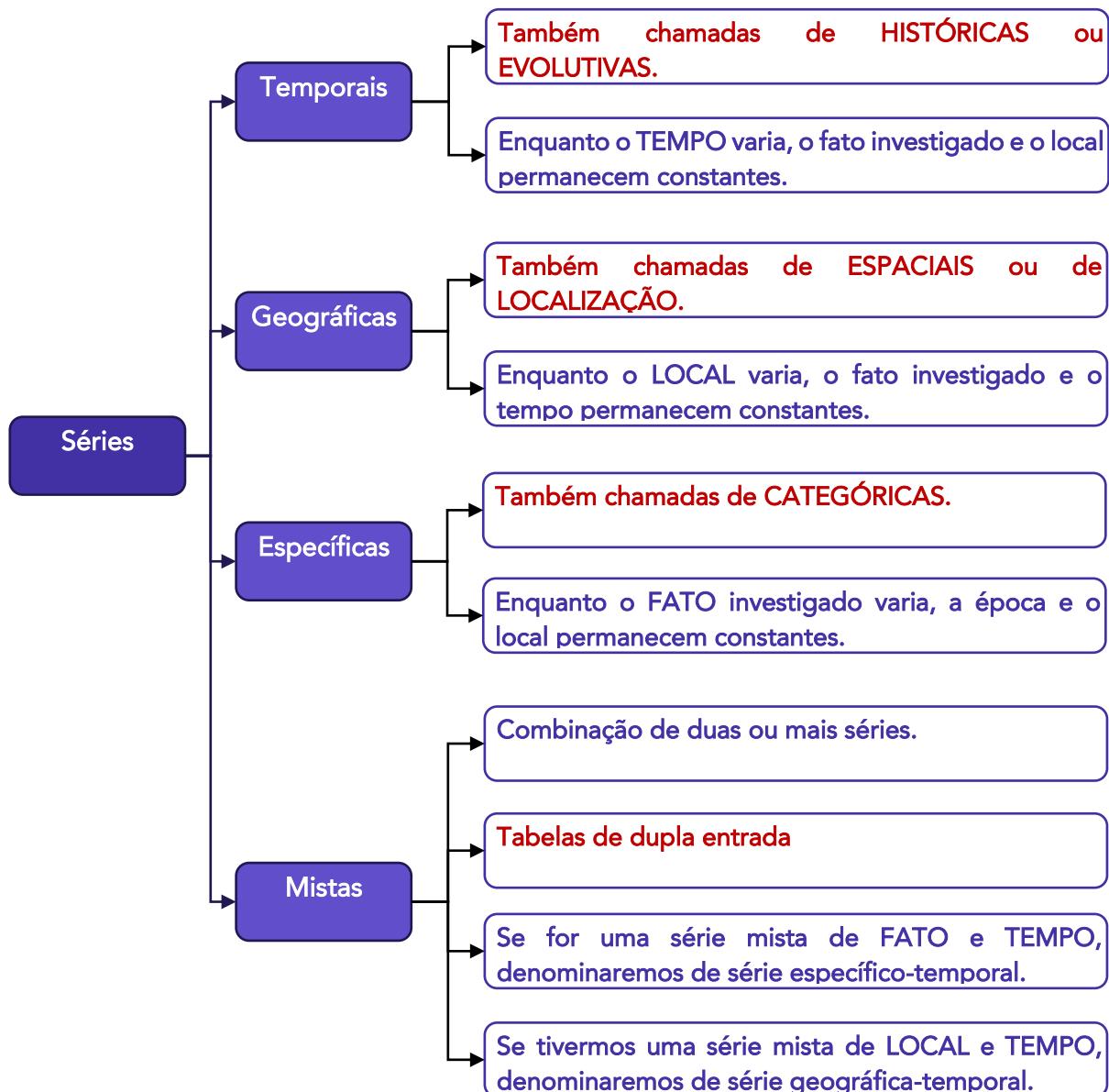
VARIÁVEIS ESTATÍSTICAS

As variáveis estatísticas podem ser classificadas, inicialmente, em duas categorias: **qualitativas** e **quantitativas**.



SÉRIES ESTATÍSTICAS

As **séries estatísticas** podem ser classificadas em: **temporais**, **geográficas**, **específicas** ou **mistas**:



DISTRIBUIÇÃO DE FREQUÊNCIAS

As **distribuições de frequências** podem ser classificadas em dois tipos: distribuição de frequências **pontual** (ou **discreta**) e distribuição de frequências **intervalar** (ou **contínua**).

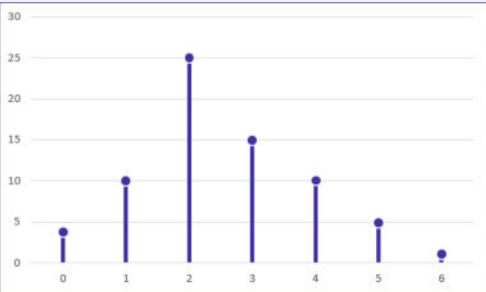
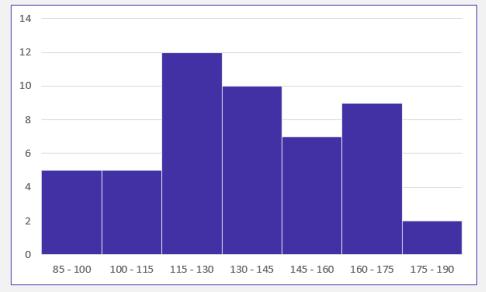
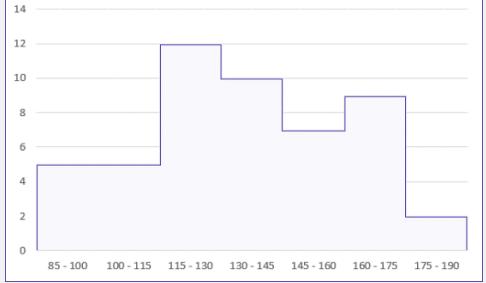
DISTRIBUIÇÃO DE FREQUÊNCIAS PONTUAL	DISTRIBUIÇÃO DE FREQUÊNCIAS INTERVALAR
São apresentados todos os dados coletados juntamente com suas respectivas frequências, não havendo perda de valores.	É agrupamento os valores por intervalos de classe.

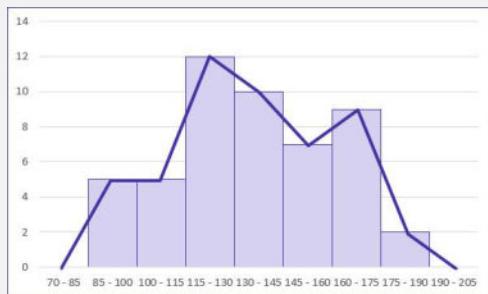
Elementos de uma Distribuição de Frequências

Item	Definição	Símbolos e Fórmulas
Número de Classes	As classes são os intervalos nos quais o fenômeno é subdividido.	$k = 1 + 3,3 \times \log n$ $k = \sqrt{n}$
Limites de Classe	Correspondem aos valores extremos.	l_{inf} e l_{sup}
Amplitude de um Intervalo de Classe	Distância entre os limites inferiores (ou superiores) de classes consecutivas.	$h = l_{sup} - l_{inf}$
Amplitude total	Diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo).	$AT = l_{máx} - l_{mín}$ $AT = h \times k$
Ponto Médio	Média aritmética simples dos valores extremos de uma classe.	$PM = \frac{(l_{inf} + l_{sup})}{2}$ $PM = l_{inf} + \frac{h}{2}$ $PM = l_{sup} - \frac{h}{2}$
Frequência Absoluta Simples	Número de observações correspondentes a uma determinada classe ou a um determinado valor.	f_i

Frequência Absoluta Acumulada	Total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe	$f_{ac_i} = f_1 + f_2 + f_3 + \cdots + f_i$
Frequência Relativa Simples	Proporção de dados existentes em uma determinada classe.	$F_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}$
Frequência Relativa Acumulada	Proporção de valores inferiores ao limite superior do intervalo de uma dada classe.	$F_{ac_i} = F_1 + F_2 + F_3 + \cdots + F_i$
Densidade de Frequência	Quociente entre a frequência da classe (absoluta ou relativa) e sua amplitude	$d = \frac{f}{h}$

REPRESENTAÇÕES GRÁFICAS DAS DISTRIBUIÇÕES DE FREQUÊNCIA

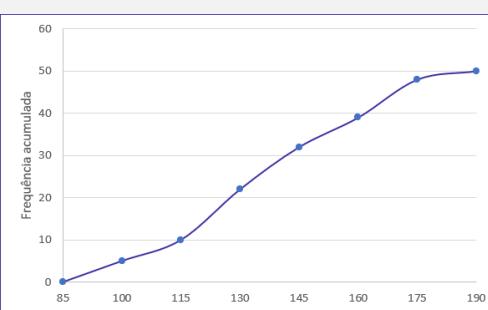
Gráfico	Definição
 A figura mostra um gráfico de hastes ou bastões. O eixo vertical (y) vai de 0 a 30 em incrementos de 5. O eixo horizontal (x) tem rótulos de 0, 1, 2, 3, 4, 5, 6. As hastes correspondem a valores pontuais: 0 (altura ~4), 1 (altura ~10), 2 (altura ~25), 3 (altura ~15), 4 (altura ~10), 5 (altura ~5), 6 (altura ~2).	O gráfico de hastes ou bastões é muito utilizado para representar dados não agrupados em classes, o que normalmente ocorre com dados discretos.
 A figura mostra um histograma. O eixo vertical (y) vai de 0 a 14 em incrementos de 2. O eixo horizontal (x) tem rótulos de classes: 85 - 100, 100 - 115, 115 - 130, 130 - 145, 145 - 160, 160 - 175, 175 - 190. As barras representam frequências: 85-100 (~5), 100-115 (~5), 115-130 (~12), 130-145 (~10), 145-160 (~7), 160-175 (~9), 175-190 (~2).	O histograma é um gráfico destinado a representar dados agrupados em classe , sendo composto por um conjunto de retângulos contíguos (justapostos) .
 A figura mostra uma poligonal característica. O eixo vertical (y) vai de 0 a 14 em incrementos de 2. O eixo horizontal (x) tem rótulos de classes: 85 - 100, 100 - 115, 115 - 130, 130 - 145, 145 - 160, 160 - 175, 175 - 190. A poligonal é formada por segmentos de reta que conectam os pontos (base da classe, altura da classe) para cada classe.	A poligonal característica é construída utilizando apenas os contornos do histograma.



O **polígono de frequências** é um gráfico em linha obtido por meio da ligação, por segmentos de reta, dos pontos médios das bases superiores dos retângulos de um histograma.



A **curva de frequências** é obtida a partir do polimento de um polígono de frequências

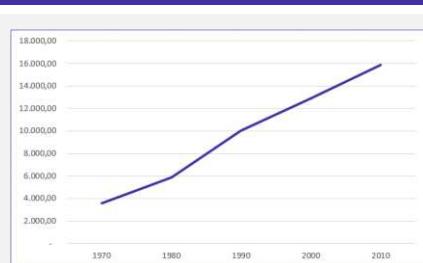


O **gráfico de ogiva** é empregado na representação de distribuições de frequências acumuladas, sejam elas crescentes ou decrescentes

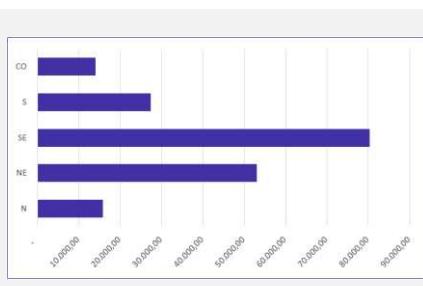
Principais formas de **representação de dados estatísticos**:

Gráfico

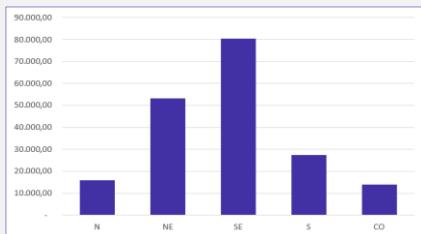
Definição



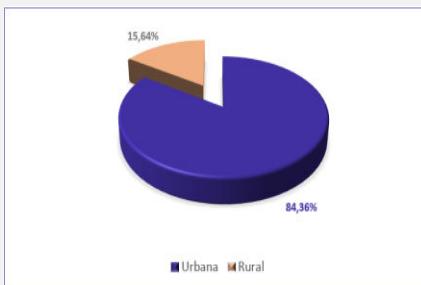
Os **gráficos em linha** normalmente são usados para representar a variação dos valores de uma variável ao longo do tempo. Esse tipo de gráfico permite-nos comparar duas variáveis: uma é traçada no eixo x (horizontal) e a outra no eixo y (vertical).



Os **gráficos em barra** normalmente são usados para representar distribuições de dados categóricos ou qualitativos. Uma série estatística é representada por um conjunto de retângulos dispostos **horizontalmente**, cada um indicando uma categoria particular.



Os **gráficos em coluna** também são usados para distribuições de dados categóricos ou qualitativos. A diferença básica é que, agora, uma série estatística é representada por um conjunto de retângulos dispostos **verticalmente**, cada um indicando uma categoria particular.



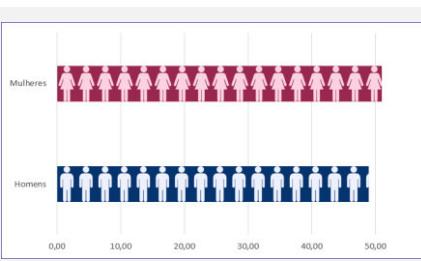
O **gráfico em setores** é usado para representar a frequência relativa (porcentagem) de uma variável categórica, sendo formado por um círculo dividido em setores circulares, cada um representando uma categoria.



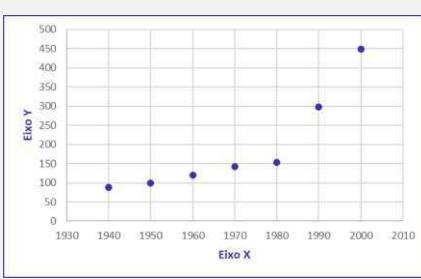
O **gráfico polar** consiste em uma sequência de eixos igualmente espaçados (ângulos iguais), cada um representando uma das variáveis. Uma linha é desenhada ligando os valores de cada eixo.



O **cartograma** é empregado com a finalidade de apresentar dados estatísticos diretamente relacionados com áreas geográficas.



O **pictograma** substitui valores por ícones, tornando os dados mais atraentes e facilitando o entendimento acerca de um determinado fenômeno.



O **gráfico de dispersão** é uma representação de pares ordenados em um plano cartesiano, composto por um eixo vertical (ordenada) e um eixo horizontal (abcissa). Os dados são representados como uma coleção de pontos.

8	5	9	Chave: 8 5 = 85
9	6	8	9
10	3	4	5
11	3	4	5 5
12	3	3	4 6 6 6 7 8 9 9
13	4	5	5 5 7 7 7
14	2	3	3 8
15	3	4	5 7 8 9

O **diagrama de ramos e folhas** fornece uma maneira rápida de representar graficamente a distribuição dos dados. Nele, cada número é separado em duas partes. Em geral, de um lado ficam as unidades do número e do outro lado fica o restante desse número.