Estratégia Concursos

Aula 00

Estatística p/ SEFAZ-RJ (Oficial da Fazenda) - 2021 - Pré-Edital

Autor:
Equipe Exatas Estratégia Concursos

05 de Abril de 2021

Sumário

Apresentação do Curso 3

Cronograma de Aulas 4

1. Conceito e Classificação 5

2. População, Amostra e Censo 7

3. Dados Estatísticos 9

3.1. Dados Brutos 9

3.2. Rol 10

3.2.1. Diagrama de Ramos e Folhas 10

4. Variáveis Estatísticas 12

4.1. Tipos de Variáveis e Cálculo de Medidas Estatísticas 13

5. Séries Estatísticas 15

6. Distribuições de Frequências 18

6.1. Dados Não-Agrupados (Tabela de Frequências) 19

6.2. Dados Agrupados (Classes) 20

6.2.1. Elementos de uma Distribuição de Frequências 21

6.2.1.1. Classes 21

6.2.1.2. Limites de Classe 22

6.2.1.3. Amplitude de um Intervalo de Classe 22

6.2.1.4. Ponto Médio de uma Classe 23

6.2.1.5. Amplitude Total da Distribuição 24

6.2.2. Frequência 25

6.2.2.1. Frequências Absolutas 25

6.2.2.2. Frequências Relativas 27

7. Representação Gráfica de Dados 29
7.1. Gráfico em Barras 29
7.2. Gráfico de Colunas 30
7.3. Gráfico em Linhas 30
7.4. Gráfico em Setores 31
7.5. Gráfico de Hastes ou Bastões 32
7.6. Histograma 33
7.7. Polígono de Frequências 34
7.8. Curva de Frequências 35
7.8.1. Formas das Curvas de Frequências 36
7.9. Ogiva 37
7.10. Box Plot (Gráfico de Caixa) 38
Questões Comentadas 39
Lista de Questões 65
Gabarito 79

APRESENTAÇÃO DO CURSO

Olá, pessoal! Tudo bem?

É com enorme alegria que damos início ao nosso **Curso de Estatística p/ SEFAZ-RJ (Oficial da Fazenda).**

O curso contemplará toda a abordagem teórica da disciplina, bem como a parte prática com a resolução de muitas questões, visando uma preparação eficiente para concurso público.

Assim, procure realizar o estudo das aulas em PDF, realizando as **marcações** do material para otimizar as suas futuras **revisões**. Além disso, não deixe de realizar as **questões**. Elas serão essenciais para lhe auxiliar na fixação do conteúdo.

Além do livro digital, você também terá acesso a videoaulas, esquemas, slides e dicas de preparação no estudo da **Estatística**. Ademais, você poderá fazer perguntas sobre as aulas em nosso **fórum de dúvidas**.

Quanto à **metodologia de estudo**, vale dizer que as aulas em PDF têm por característica essencial a **didática**. O curso todo se desenvolverá com uma leitura de fácil compreensão e assimilação. Isso, contudo, não significa superficialidade. Pelo contrário, sempre que necessário e importante os assuntos serão aprofundados.

Com essa estrutura e proposta pretendemos conferir segurança e tranquilidade para uma **preparação completa**, **sem necessidade de recurso a outros materiais didáticos**. Fique tranquilo que abordaremos todos os tópicos exigidos para o seu concurso.

Bons estudos!

Equipe Estratégia Concursos

CRONOGRAMA DE AULAS

Distribuição das aulas:

AULAS	TÓPICOS ABORDADOS	DATA
Aula 00	Estatística Descritiva	05/04
Aula 01	Medidas de Dispersão	08/04
Aula 02	Probabilidade	12/04
Aula 03	Distribuição de Probabilidades	15/04
Aula 04	Teste de Hipótese	19/04

1. Conceito e Classificação

Antes de tudo, é fundamental que você entenda o conceito de Estatística e a classificação que essa disciplina possui.

Diariamente estamos expostos a informações estatísticas. Para exemplificar, considere as seguintes manchetes de jornais e revistas:

> "*Produção industrial sobe 0,8% e tem melhor agosto em 5 anos, diz IBGE*". (Revista Exame, em 1º de outubro de 2019)
> "*Nas famílias mais pobres, 24% da renda vem de aposentadorias e benefícios*" (Revista Veja, em 4 de outubro de 2019)
> "*Desemprego sobe para 12,7% em março e atinge 13,4 milhões de brasileiros*" (site g1.globo.com, em 30 de abril de 2019)

As três frases que acabou de ler são baseadas em levantamentos de **dados**, que são **informações oriundas de observações, contagens, medidas ou respostas**. Todavia, a estatística hoje envolve mais do que levantamento e apresentação de dados, cálculo de médias e construção de gráficos.

Então, afinal de contas, o que é estatística? As múltiplas facetas dessa área do conhecimento dificultam defini-la de modo satisfatoriamente preciso. Porém, o conceito apresentado a seguir é amplamente aceito nas principais obras, de maneira que podemos adotá-lo:

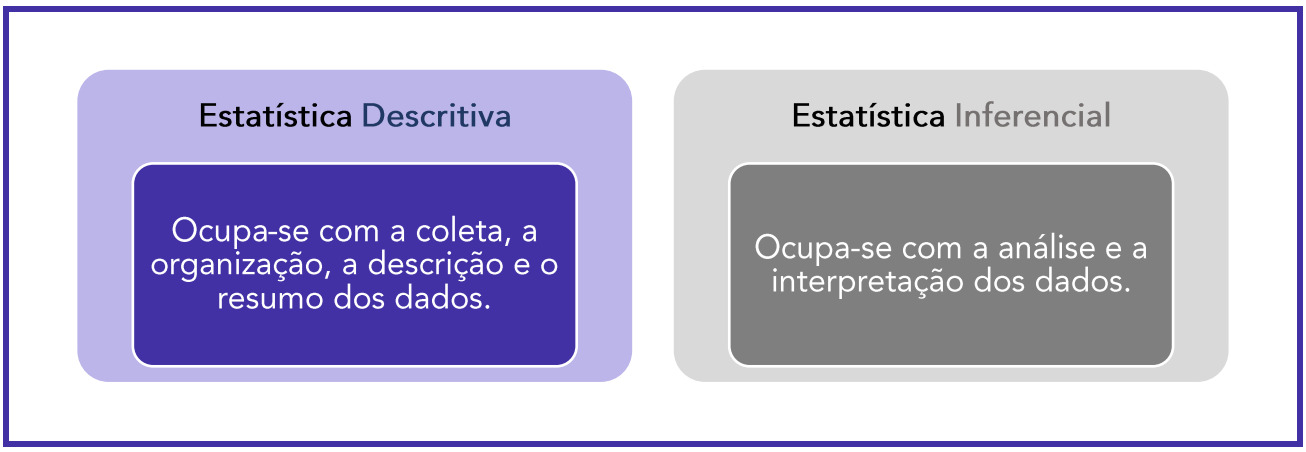
A **Estatística** é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados para sua utilização na tomada de decisões.

A depender do objetivo que se tem em mente ao utilizar os dados, a Estatística pode ser classificada em Descritiva, Inferencial e Probabilística.

A **Estatística Descritiva ou Dedutiva**, também conhecida como **Estatística Básica**, consiste na aplicação de métodos para **coletar, organizar, descrever e resumir dados**, de modo que é responsável pelas primeiras etapas do processo estatístico.

A **Estatística Probabilística** consiste na aplicação de técnicas e métodos apropriados para o **estudo do processo de generalização de informações**, o que envolve certa medida de risco ou incerteza e constitui os fundamentos da **Teoria da Probabilidade**. O objetivo da Estatística Probabilística é **calcular as chances de ocorrer determinados eventos**, principalmente quando envolvem o acaso, como jogos de dados e de cartas ou o lançamento de uma moeda para o ar.

A **Estatística Inferencial ou Indutiva**, também conhecida pelos concurseiros como **Estatística Avançada**, consiste na aplicação de métodos para **analisar e interpretar dados com vistas à tomada de decisões**, de modo que é responsável pelas últimas etapas do processo estatístico. O objetivo da Estatística Inferencial é **fazer generalizações a respeito do todo (população) por meio da análise de partes (amostra)**. Além disso, com seus métodos, é possível analisar a representatividade dos resultados obtidos, a significância e a confiabilidades dos estudos realizados.



2. POPULAÇÃO, AMOSTRA E CENSO

Em seguida, é importante que você saiba distinguir os conceitos de População, Amostra e Censo. Nesse sentido, apresento a você uma forma prática para atingir esse objetivo.

Suponhamos que fomos incumbidos da missão de obter as idades de 20 alunos de uma determinada turma de um curso para concursos. A seguir, é apresentado o resultado da nossa pesquisa:

25, 21, 20, 20, 21, 19, 41, 33, 41, 35, 35, 33, 19, 19, 51, 51, 25, 21, 19, 20

O conjunto formado pelas idades de todos os alunos do curso é a nossa **população**.

População é o conjunto de **todos** os **elementos** que possuem uma determinada **característica em comum**.

No nosso caso, estamos interessados nos dados que representam idades de alunos que estudam no mesmo curso. Essa é a característica em comum.

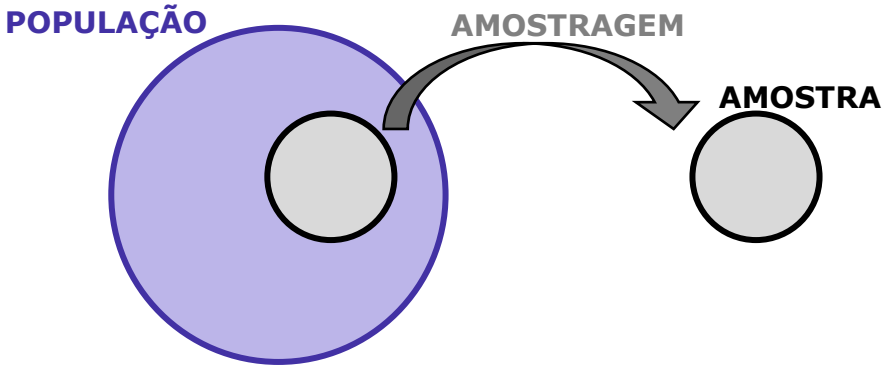
Uma população pode ser finita ou infinita. Por exemplo, a população constituída por todos os parafusos produzidos numa fábrica em certo dia é **finita**, enquanto a população composta por todos os resultados (cara ou coroa) em sucessivos lances de uma moeda é **infinita**.

Além disso, se a mesma pesquisa fosse realizada com todas as pessoas de um determinado bairro, estaríamos realizando um **censo**.

Censo é a contagem de **todos** os elementos de uma **população**.

Agora, dependendo da população, fica inviável entrevistar todo mundo. Imagine se for uma cidade muito grande. Possivelmente, não se tem tempo suficiente para esperar que todo mundo seja entrevistado, ou não se tem dinheiro para pagar toda a quantidade de pessoal que seria necessária para coletar tais dados e, mesmo assim, se fizéssemos um censo, em alguns casos os resultados serão falhos. Nesses casos, em vez de entrevistarmos todo mundo, escolhemos uma **amostra**, que, basicamente, constitui uma **redução da população a dimensões menores**, sem perda das características essenciais. É, portanto, uma **parte da população**.

Amostra é qualquer subconjunto não vazio **da** população**.**



O exemplo de considerarmos apenas 20 alunos do sexo masculino de uma determinada turma de um curso constitui um caso de aplicação do método estatístico de **amostragem** para obter uma parcela adequada da população "todos os alunos de um curso para concursos". Salientamos que estudaremos o assunto amostragem detalhadamente em capítulo específico.

É importante destacar que uma amostra, para ser **boa**, tem que ser **representativa**, ou seja, deve conter em proporção tudo o que a população possui qualitativa e quantitativamente. Além disso, deve ser **imparcial**, isto é, todos os elementos da população devem ter igual oportunidade de fazer parte da amostra!

POPULAÇÃO	CENSO	AMOSTRA
É qualquer conjunto de informações que tenham, entre si, uma **característica comum**.	É a contagem de **todos** os elementos de uma POPULAÇÃO.	É uma parcela representativa da população.

3. DADOS ESTATÍSTICOS

Os **dados estatísticos** são **observações coletadas**, como é o caso de medidas, sexo, altura e outras informações obtidas por meio de uma pesquisa.

Um conjunto de dados estatístico que consiste em **apenas uma única característica**, ou atributo, é classificado como **univariado**. Já se medimos, ou registramos, **duas observações** de cada indivíduo ou objeto, o conjunto de dados é **bivariado**. Por fim, se houver **mais de duas observações** sobre o mesmo indivíduo ou objeto, o conjunto de dados é dito **multivariado**.

Quanto à forma de realização da coleta dos dados estatísticos, ela pode ser classificada como direta ou indireta.

A coleta é **direta** quando realizada sobre elementos informativos de registro obrigatório (nascimentos, casamentos e óbitos) ou quando os dados são obtidos pelo próprio pesquisador por meio de questionários, como é o caso do censo demográfico.

É possível classificar a coleta direta de dados levando em conta o fator temporal em:

a) **contínua**: **realizada continuamente**, tal como a de nascimentos e óbitos e a de frequência dos alunos às aulas;

b) **periódica**: **feita em intervalos constantes de tempo**, como os censos (geralmente ocorrem de 10 em 10 anos) e as avaliações mensais dos alunos;

c) **ocasional**: **executada por demanda ou extemporaneamente** para atender uma conjuntura ou uma situação emergencial, a exemplo de epidemias que afetam rebanhos inteiros.

Já a coleta é **indireta** quando os dados são obtidos a partir de elementos conhecidos (coleta direta) ou por meio da análise de outros fenômenos relacionados com o fenômeno estudado. Para exemplificar, podemos citar a pesquisa sobre a mortalidade infantil, que é feita com o uso de dados colhidos por uma coleta direta.

Podemos classificar os dados quanto à **organização da sua apresentação** em brutos ou rol.

3.1. Dados Brutos

Os **dados brutos** são aqueles que **não estão numericamente organizados**. Podemos citar como exemplo uma relação dos pesos de 50 alunos de uma determinada academia, feita em ordem alfabética, de modo que não há organização de valores em ordem.

A esse tipo de informação, cujos elementos não foram numericamente ordenados, denominamos **tabela primitiva**.

3.2. Rol

Após realizar a coleta dos dados, é necessário organizar esses dados em ordem crescente ou decrescente, formando um **rol**.

> **Rol** é a organização dos dados brutos em ordem crescente ou decrescente.

Assim, tenha em mente que, para ser um rol, os dados em análise devem estar ordenados, de forma crescente ou decrescente. Dessa maneira, organizando os dados brutos do nosso exemplo das idades dos alunos de uma sala de um curso para concursos, teremos:

> **19, 19, 19, 19, 20, 20, 20, 21, 21, 21, 25, 25, 33, 33, 35, 35, 41, 41, 51, 51**

Por fim, mencionamos que a **diferença entre o maior e o menor número** do rol é denominada **amplitude total** dos dados.

No caso em consideração, a menor idade (19 anos) e a maior (51 anos) resultam numa amplitude total de variação de **32** anos.

3.2.1. Diagrama de Ramos e Folhas

O **Diagrama de Ramos e Folhas** é uma forma alternativa de resumir um conjunto de valores e que nos dá uma ideia de como se dá a sua distribuição, tratando-se de um **caso particular do rol**.

No diagrama de ramos e folhas, cada número deve ser separado em duas partes. Os diagramas que mais caem em prova **separam a unidade de um lado e o resto do número do outro lado**. Assim, considere o seguinte rol:

> 10, 11, 13, 14, 15, 15, 16, 18, 18, 19, 20, 22, 25, 26, 29.

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

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

Na coluna da esquerda temos as dezenas, que correspondem aos ramos. Do lado direito, temos as unidades, que seriam as folhas. As folhas se prendem aos ramos.

Assim, “1” – espaço – “0 1 3 4”, num diagrama de ramos e folhas, significa que, no rol original, nós temos os números 10, 11, 13, 14.

Além disso, é muito comum que os diagramas de ramos e folhas separem as unidades em dois grupos: de 0 a 4 e de 5 a 9.

Resumindo, os diagramas que mais caem em concursos adotam as seguintes regras:

- ✓ Separam as unidades do resto do número (a unidade seria a folha);
- ✓ Para cada dezena geralmente são necessárias duas linhas: uma para as unidades de 0 a 4; outra para as unidades de 5 a 9. Contudo, isso não é obrigatório, pois todos os elementos da dezena poderiam estar em uma linha só.

Enfatizamos que não existe uma regra fixa para construção do diagrama de ramos e folhas. A ideia é apenas esta: dividir os números em duas partes.

Por fim, destacamos que a maior parte da questão que aborda essa temática é relacionada a assuntos que veremos adiante. Por exemplo, o enunciado apresenta um diagrama de ramos e folhas e pergunta o valor de determinada medida estatística, como a média ou a mediana dos dados apresentados. Então, repare que dificilmente o candidato conseguiria solucionar a questão sem antes conhecer a estrutura do diagrama de ramos e folhas.

4. VARIÁVEIS ESTATÍSTICAS

Nas mais diversas questões de estatística, você será confrontado com a palavra “variável” ou “variáveis”. Assim, é importante conhecer o seu significado e algumas características relacionadas.

A **variável estatística** consiste no **objeto da pesquisa**, ou seja, é aquilo que estamos investigando. Sendo assim, na pesquisa que fizemos com os alunos de um curso para concursos, a variável é a idade dos alunos que compõem a turma considerada.

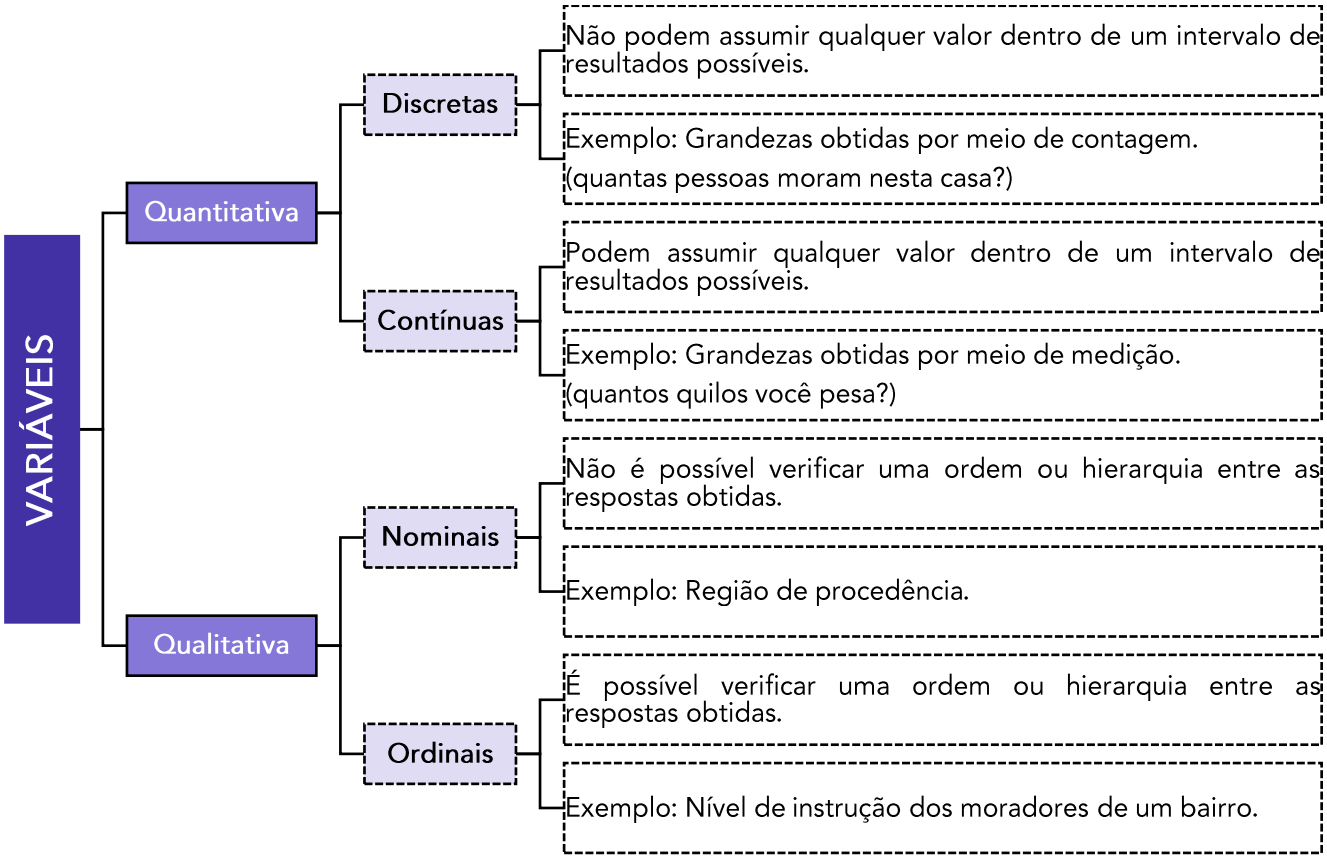
> Variável é o objeto da pesquisa, podendo assumir diferentes valores a cada experimento.

Após entendermos o conceito de variável, vamos direto ao que mais interessa no assunto: classificar as variáveis. Isso realmente é o que possui maior incidência nas provas no âmbito deste tópico.

Você precisa ter em mente a classificação geral de uma variável em qualitativa ou quantitativa. Os próprios nomes já dão uma boa noção de como funciona essa classificação. De fato, **quantitativa** lembra **quantidade**, que, por sua vez, lembra **número**, de modo que estamos falando das variáveis numéricas. Por outro lado, temos que **qualitativa** lembra qualidade, a qual lembra **atributo**, de forma que estamos diante de variáveis não-numéricas, que indicam atributos.

Portanto, fica claro que a principal diferença entre elas é que à variável **quantitativa** pode ser designado um **valor numérico**, enquanto para a variável **qualitativa** isso não é possível, por se tratar de **atributos**, não trazendo em si a ideia de mensuração.

> Perceba a diferença pelo prefixo de cada variável:
>
> **QUANT**itativa → **Quanto?**
>
> **QUAL**itativa → **Qual?**



4.1. Tipos de Variáveis e Cálculo de Medidas Estatísticas

Mais à frente estudaremos diversas medidas. Porém, vamos antecipar algumas apenas para entendermos como elas se comportam perante os tipos de variáveis. Saliento que isto pode ser cobrado na sua prova!

Por exemplo, suponha três pessoas, com as seguintes informações:

VARIÁVEL	A	B	C	D	E
Idade	14	22	24	25	25
Sexo	feminino	masculino	feminino	masculino	feminino
Escolaridade	fundamental	médio	superior	superior	superior
UF	RJ	RJ	SP	MG	GO

Ainda não estudamos a **média**, mas é possível que você se recorde que seu cálculo envolve **somar todos os valores e dividir pelo número de observações**. Assim, para determinar a idade média desse grupo de pessoas, basta somar as idades e dividir por 5, pois são 5 pessoas: (14 + 22 + 24 + 25 + 25) ÷ 5 = 22.

A que conclusão chegamos? Para calcular a média, foi preciso somar os valores, o que só é possível para dados numéricos, isto é, não conseguimos somar atributos. Por exemplo, como calcularíamos a média da variável "sexo"? Simplesmente não dá para fazer, concorda? Portanto, **só é possível calcular a média para variáveis quantitativas**.

Outra medida que estudaremos é a **moda**, que corresponde ao **termo que mais se repete**. No nosso exemplo, conseguimos calcular a moda para todas as variáveis. De fato, nas idades, dizemos que a moda vale 25, pois foi a idade que mais ocorreu. Já na variável “sexo”, a moda é "feminino", pois foi o sexo que mais ocorreu. Portanto, concluímos que **conseguimos determinar a moda para qualquer tipo de variável**, já que basta verificarmos qual termo ocorreu mais vezes.

Também teremos oportunidade de estudar a **mediana**, que é o **termo central da série**. Para determiná-la, precisamos inicialmente ordenar os valores, para só depois determinar quem é o central. No nosso exemplo, na variável idade, conseguimos ordenar os valores assim: 14, 22, 24, 25, 25. O termo central é 24. Logo, a mediana vale 24. Na variável "escolaridade", conseguimos ordenar assim: fundamental / médio /superior / superior / superior. O termo central é "superior", que é nossa mediana. Porém, na variável "UF", sequer conseguimos ordenar as observações, pois trata-se de uma variável qualitativa nominal, de modo que não é possível determinar a mediana. Dessa forma, para determinar a mediana precisamos que a variável, no mínimo, possa ser ordenada. Logo, **não conseguimos indicar a mediana no caso de variável qualitativa nominal**.

VARIÁVEL	MÉDIA	MEDIANA	MODA
QUANTITATIVA DISCRETA	OK	OK	OK
QUANTITATIVA CONTÍNUA	OK	OK	OK
QUALITATIVA ORDINAL	-	OK	OK
QUALITATIVA NOMINAL	-	-	OK

5. SÉRIES ESTATÍSTICAS

As **séries estatísticas** consistem na apresentação das informações ou variáveis estatísticas em formas de tabelas, objetivando sintetizar os dados estatísticos observados e tornando-os mais compreensíveis.

Os **elementos** presentes numa série estatística são os seguintes:

✓ **Fato**: é o **fenômeno investigado**, cujos valores são apresentados na tabela.
✓ **Local**: indica a região **onde o fato aconteceu**.
✓ **Época**: refere-se à **data**, ao **período** ou ao **tempo** em que o fato foi observado.

A maior incidência das questões de concursos relativas ao presente assunto está na classificação das séries. Então, vamos estudá-la!

Inicialmente, temos a **série histórica** (temporal, cronológica ou evolutiva) em que os dados são observados segundo a época de ocorrência. Nesse caso, **o tempo é variável** e o fato e o local são fixos. Um bom exemplo disso é o Índice de Preços ao Consumidor (IPC) para os meses de março a dezembro de 2014:

Índice de Preços ao Consumidor - Brasil

Período	Inflação (%)
mar/14	6,154
abr/14	6,28
mai/14	6,374
jun/14	6,523
jul/14	6,502
ago/14	6,513
set/14	6,746
out/14	6,587
nov/14	6,555
dez/14	6,409

Repare que o fato é sempre o mesmo (inflação), o local é fixo (Brasil) e o período varia (de março a dezembro de 2014).

Já na **série geográfica** (ou de localização) os dados são observados segundo a localidade de ocorrência. Nessa situação, **o local varia** e a época e o fato são fixos. Para exemplificar, temos o PIB dos países da América do Sul em 2012:

Produto Interno Bruto da América do Sul
Por Países, 2012

(Em Valores Correntes - US$ Milhões)

Países	2012
Argentina	475.502
Bolívia	27.035
Brasil	2.252.664
Chile	269.869
Colômbia	369.606
Equador	84.040
Guiana e Guiana Francesa	2.851
Paraguai	25.502
Peru	203.790
Suriname	5.012
Uruguai	49.920
Venezuela	381.286

Fonte: Banco Mundial (favereiro 2014)
OBS: Em 2012, a Venezuela tornou-se o quinto membro a integrar o Mercosul.

Nesse caso, observamos que o fato é sempre o mesmo (PIB em milhões de dólares), o período é fixo (2012) e o local é que muda (Argentina, Bolívia, Brasil etc.).

Por sua vez, na **série específica** (ou categórica) os dados são agrupados segundo a modalidade de ocorrência, em que **o fato é variável**, enquanto a época e o local são fixos. Por exemplo, considere a tabela a seguir:

Número de Técnicos Administrativos - USP 2013

Nível superior	4.174
Nível técnico	8.104
Nível básico	5.149
Outros	24
Total	**17.451**

Vemos que o local é fixo (USP), o tempo é fixo (2013) e o fato varia (na primeira linha quantificamos técnicos administrativos de nível superior; na segunda, de nível técnico; e assim por diante). É interessante notar, ainda, que tivemos na verdade um fato maior (técnico administrativo), que por sua vez é decomposto em categorias.

E, para finalizar, é possível encontrarmos a **série mista (ou conjugada)**, que corresponde a uma **combinação de duas ou mais dos três tipos** de séries anteriores. Veja na tabela a seguir um bom exemplo:

Produção Agrícola, por região, segundo os principais produtos — 2009 (valores em milhões de ton.)

Região	Produtos				Total
	Grãos	Frutas	Legumes	Outros	
Norte	1,5	45,2	5,4	0,5	52,6
Nordeste	2,4	120,0	1,2	4,5	128,1
Centro-Oeste	10,5	10,5	4,2	1,5	26,7
Sudeste	50,7	100,1	10,2	5,9	166,9
Sul	12,5	75,0	15,2	2,5	105,2

Fonte: Dados fictícios

Note que o ano é fixo (2009), mas há variação no local (região do país) e na categorização dos dados (tipos de produtos), de modo que está presente uma conjugação entre a série geográfica e a série específica.

6. DISTRIBUIÇÕES DE FREQUÊNCIAS

Chegamos num momento vital da nossa aula de Estatística Descritiva, em que colocaremos a mão na massa e faremos cálculos. A análise das distribuições de frequências será uma marca presenta ao longo do seu estudo na disciplina.

Denominamos **frequência** o **número de vezes que cada valor de uma variável aparece repetido** num levantamento de dados.

Anteriormente, consideramos o levantamento das idades de 20 alunos de um curso, e chegamos ao seguinte rol de dados:

19, 19, 19, 19, 20, 20, 20, 21, 21, 21, 25, 25, 33, 33, 35, 35, 41, 41, 51, 51

Nesse exemplo, a variável em questão, idade, será observada e estudada mais facilmente se dispusermos os valores ordenados em uma coluna e colocarmos, ao lado de cada valor, o número de vezes que aparece repetido, isto é, a frequência com que os dados são observados na amostra.

Denominamos **frequência** o número de vezes que cada valor de uma variável aparece repetido num levantamento de dados.

A partir das idades observadas no levantamento das idades de 20 alunos de um curso e a frequência que aparecem, podemos construir uma tabela conhecida como **distribuição de frequência**:

Idade (anos)	Frequência
19	4
20	3
21	3
25	2
33	2
35	2
41	2
51	2
Total	**20**

Dá-se o nome de **distribuição de frequência** à **tabela** que relaciona os valores de uma variável com as suas respectivas frequências.

Note que, para ser distribuição de frequência, a tabela precisa apresentar os dados levantados junto com as respectivas quantidades de vezes que aparecem repetidos. Ou seja, não é qualquer tabela que será definida como uma distribuição de frequência!

Basicamente, existem dois **tipos**[1] de distribuições de frequências que podem ser construídas:

1) quando as observações são ordenadas em valores **individuais**, na forma da tabela anterior. Nesse caso, o resultado é conhecido como distribuição de frequências para dados **não agrupados**, ou simplesmente **tabela de frequências**;

2) quando as observações são ordenadas em **classes** de mais de um valor. Nessa situação, a tabela resultante é conhecida como distribuição de frequências para dados **agrupados**, ou simplesmente **tabela de classes**.

6.1. Dados Não-Agrupados (Tabela de Frequências)

Chamamos de **dados não agrupados** as observações correspondentes a valores **individuais**. Um bom exemplo disso é a tabela a seguir contendo as frequências relativas às estaturas dos 20 alunos daquele curso para concursos:

Estatura (cm)	Frequência
150	1
152	1
154	2
156	1
158	3
160	5
162	3
164	2
Total	**20**

É importante entendermos o significado dessa coluna de frequências. Na primeira linha (altura igual a 150 cm), a frequência é 1, o que significa que para tal valor há 1 aluno com essa estatura. Em outras palavras, dizemos que a altura 150 cm ocorre uma vez. Já no caso da terceira linha (altura igual a 154 cm), percebemos que a frequência é 2, indicando a existência de 2 alunos na turma com a estatura de 154 cm. Além disso, repare que a altura de 160 cm é aquela com mais alunos, pois é a que apresenta maior frequência (5) dentre

[1] Para vários autores, enquadram-se no conceito de distribuição de frequências apenas o segundo caso, ou seja, quando estamos diante de uma tabela em que os dados são ordenados em classes.

as consideradas. E, com um exame mais detalhado, notamos que há uma concentração das alturas entre os valores 158 cm e 164 cm, ao passo que há valores abaixo de 154 cm.

6.2. Dados Agrupados (Classes)

Suponha que foi realizada uma pesquisa sobre a estatura de 40 moradores de um bairro, obtendo os seguintes dados, ordenados na forma de **rol**:

142	149	155	162	165	168	171	175
147	150	156	163	165	168	171	176
147	152	159	164	166	168	172	176
148	153	159	165	167	170	173	177
148	155	159	165	167	171	175	179

Repare que agora estamos lidando com uma quantidade de dados bem maior, o que torna inconveniente a utilização de uma simples tabela de frequências.

Ocorre que a amplitude dos dados tornaria a tabela de difícil manipulação em função de existirem muitos valores possíveis entre a menor e a maior observação.

Adicionalmente, é possível que haja considerável perda de informação diante da necessidade de exibir apenas parcialmente as alturas de todos os moradores. De fato, seria praticamente inviável alistar todas as observações. Veja:

Estatura (cm)	Frequência
142	1
147	2
148	2
149	1
(...)	
165	4
166	1
167	2
(...)	
176	2
177	1
179	1
Total	**40**

Em situações como essa, em que há mais de 20 valores possíveis, a solução recomendada é o agrupamento dos dados em vários intervalos. Desse modo, se um dos intervalos for 140 ⊢ 150[2], em vez de afirmarmos que a estatura de 1 morador é de 142 cm; de 2 moradores, 147 cm; de 2 moradores, 148 cm; e de 1 morador, 149 cm, diremos que 6 moradores têm estaturas entre 140, inclusive, e 150 cm.

Perceba que esse procedimento implica agrupar os valores da variável em intervalos, os quais são chamados de **classes**. Fazendo o mesmo para as demais estaturas, teremos a tabela a seguir, denominada **distribuição de frequências para dados agrupados**:

Estatura (cm)	Frequência
140 ⊢ 150	6
150 ⊢ 160	9
160 ⊢ 170	13
170 ⊢ 180	12
Total	**40**

6.2.1. Elementos de uma Distribuição de Frequências

Analisemos, agora, detalhadamente, cada um dos elementos de uma distribuição de frequências. Para isso, vamos usar como referência o exemplo apresentado no tópico anterior.

6.2.1.1. Classes

A apresentação dos dados de uma pesquisa por meio de **classes** é a mais comum, e será a que você provavelmente encontrará em sua prova.

> **Classes** são **intervalos** ou subdivisões dos elementos do conjunto de dados.

Assim, a primeira classe vai de 140 a 150; a segunda classe vai de 150 a 160 e assim por diante.

[2] A notação 140 ⊢ 150 será muito utilizada, trata-se de um intervalo fechado à esquerda e aberto à direita, tal que: $140 \leq x < 150$. Isso significa que o intervalo abrange valores desde o 140 (incluindo-o) até valores menores que 150 (excluindo-o). Em outras palavras, o símbolo ⊢ indica que incluímos o limite inferior do intervalo e excluímos o limite superior do intervalo.

6.2.1.2. Limites de Classe

Note que cada classe é composta por dois valores, chamados de **limites da classe**.

> **Limites de uma classe** são os seus **extremos**. O menor número é o limite inferior da classe (l_{inf}) e o maior número, o limite superior da classe (l_{sup}).

Por exemplo, na terceira classe temos que:

- $l_{inf} = 160$
- $l_{sup} = 170$

6.2.1.3. Amplitude de um Intervalo de Classe

Além disso, cada classe possui uma amplitude ou tamanho, denominada **intervalo de classe**, obtido pela **diferença entre os limites superior e inferior** dessa classe e designamos por **h**. Logo:

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

Uma dúvida que pode surgir é a qual classe determinada observação pertencerá. Por exemplo, um aluno que tenha uma altura de 160cm integrará a primeira ou a segunda classe? Note que os alunos que possuem tal estatura foram incluídos na terceira classe. Por quê? Tudo depende do símbolo que fica entre os limites inferior e superior da classe. Esquematizando as possibilidades, teremos:

Intervalo	Símbolo
Fechado à esquerda e à direita	|---|
Aberto à esquerda e fechado à direita	---|
Aberto à direita e fechado à esquerda	|---
Aberto à esquerda e à direita	---

Assim, como a nossa distribuição de frequências fez uso do símbolo |--- (Aberto à direita e fechado à esquerda), isso significa que **o valor à esquerda está incluído na classe, enquanto o valor à direita está excluído**. Logo, na terceira classe (35 |--- 45), um aluno que tivesse 35 anos de idade está incluído ao passo que um aluno com 45 anos de idade estaria fora desta classe.

6.2.1.4. Ponto Médio de uma Classe

Outro conceito importante é o de **Ponto Médio de uma Classe**, mas de facílimo entendimento.

> **Ponto Médio de uma classe** é o ponto que divide a classe em duas partes iguais.

Para obtermos o ponto médio de uma classe, calculamos a **média aritmética** dos limites da classe:

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

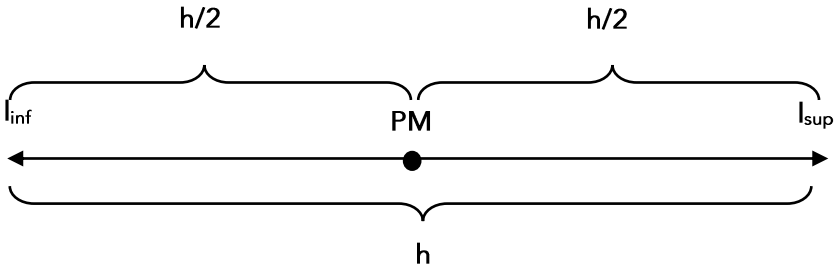
Como treinamento, vamos calcular a coluna dos Pontos Médios da nossa distribuição de frequências:

Idade	PM
140 ⊢ 150	**145** (= (140 + 150)/2
150 ⊢ 160	**155** (= (150 + 160)/2
160 ⊢ 170	**165** (= (160 + 170)/2
170 ⊢ 180	**175** (= (170 + 180)/2

Você deve ter notado que os pontos médios estão dispostos numa progressão aritmética, já que a diferença entre dois pontos médios consecutivos foi constante (= 10). Isto aconteceu porque as todas as classes da distribuição de frequência possuem a mesma amplitude.

> Nas distribuições de frequência em que as classes possuem a **mesma amplitude**, a **diferença entre dois pontos médios consecutivos será igual à amplitude** das classes (h).

Além disso, é interessante termos em mente duas relações envolvendo o ponto médio, a amplitude das classes e os limites inferior e superior.



De acordo com o desenho acima, obtemos as seguintes formas para encontrar os **limites de uma classe**:

$$l_{sup} = PM + \left(\frac{h}{2}\right)$$

$$l_{inf} = PM - \left(\frac{h}{2}\right)$$

Além disso, encontramos outra maneira para calcularmos o **ponto médio de uma classe**:

$$PM = l_{inf} + \left(\frac{h}{2}\right)$$

$$PM = l_{sup} - \left(\frac{h}{2}\right)$$

6.2.1.5. Amplitude Total da Distribuição

Finalizando este tópico, apresento a seguir uma boa definição para **Amplitude Total da Distribuição**:

> A **amplitude total da distribuição** é o **tamanho de todo o conjunto** observado, correspondendo à diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo):
>
> $$AT = l_{máx} - l_{mín}$$

No caso específico de as classes apresentarem a **mesma amplitude**, para determinar o valor da Amplitude Total da Distribuição, basta multiplicarmos o valor da amplitude da classe (h) pela quantidade de classes da distribuição:

$$\boxed{AT = h\ x\ k}$$

Em que **k** é a quantidade de classes da distribuição.

6.2.2. Frequência

Do que estudamos até aqui, fica claro que o conceito de frequência é de grande importância no âmbito da estatística descritiva, pois revela a quantidade de vezes que uma determinada característica observada na pesquisa aparece.

Nesse momento, é importante que você saiba que há quatro **tipos de frequência**, os quais analisaremos a seguir:

- Frequência absoluta simples (**fi**);
- Frequência absoluta acumulada (**fac**);
- Frequência relativa simples (**Fi**);
- Frequência relativa acumulada (**Fac**).

Letra	Tipo de Frequência
f	Frequência Absoluta
F	Frequência Relativa

6.2.2.1. Frequências Absolutas

A **frequência absoluta simples (fi)** indica o **número de elementos** que faz parte da classe correspondente.

Por que recebe o nome de "absoluta"?

Pelo fato de esta medida não levar em conta o quanto cada valor assumido pelas variáveis representa do total, que é o caso da frequência relativa, que estudaremos a seguir.

Dessa forma, na tabela das estaturas dos moradores, apresentada anteriormente, a frequência absoluta simples do número 147 é 2, indicando que a altura 147 ocorre duas vezes.

Repare que o total das frequências absolutas simples é 40. E 40 é justamente o número de moradores pesquisados. Isso não é coincidência. Na tabela anterior, indicamos a quantidade de moradores que possui determinada estatura. Se são 40 pessoas, é natural esperar que, somando todas as frequências, obtenhamos justamente 40.

Como regra geral, se tivermos **n** elementos, podemos dizer que:

$$\boxed{\sum fi = n}$$

Por sua vez, a **frequência absoluta acumulada (fac)** nos dá quantas observações são **menores ou iguais** ao valor considerado. Para a nossa sequência de dados, podemos construir a seguinte tabela:

Estatura (cm)	fi	fac
142	1	1
147	2	3
148	2	**5** (= 3 + 2)
149	1	6
(...)	(...)	(...)
Total	**40**	**40**

Iguais na primeira classe

Note que **a fac é construída de cima para baixo**, obedecendo ao seguinte:

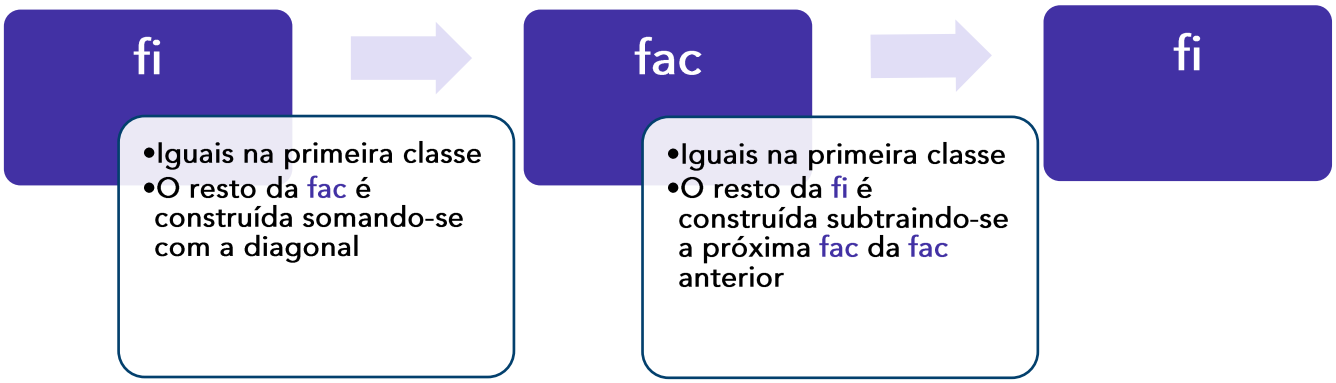
1º) É construída a partir da **fi** (as duas são "frequências irmãs");

2º) A **fac** e a **fi** possuem o mesmo valor na classe em que a **fac** começa a ser construída;

3º) Os demais valores da **fac** serão obtidos somando-se o valor da **fac** anterior com a **fi** da diagonal;

4º) A **fac** termina sempre com o mesmo valor de **n**.

Além disso, é importante sabermos realizar a **conversão de fi para fac e vice-versa**:



6.2.2.2. Frequências Relativas

As **frequências relativas** são muito parecidas com as absolutas. A única diferença é que, em vez de estarmos interessados em valores absolutos, queremos saber valores relativos.

Pois bem, as frequências relativas serão obtidas a partir de uma **divisão**, em que o denominador é o número de dados.

A **frequência relativa simples (Fi)** é dada pela **frequência absoluta simples dividida pelo número de dados**:

$$\boldsymbol{Fi = \frac{fi}{n}}$$

Na nossa pesquisa de estaturas, temos 40 valores (n = 40). Vamos, a título de exemplo, calcular a frequência relativa simples do número 147.

O número 147 ocorre duas vezes (a frequência absoluta simples do número 147 é dois; isto porque há duas pessoas com altura 147cm).

Para obter a frequência relativa simples do número 147, basta dividir 2 por 40:

$$\boldsymbol{Fi} = \frac{2}{40} = \frac{5}{100} = \mathbf{5}\%$$

O que isto significa? Significa que cinco por cento dos moradores pesquisados têm 147cm de altura.

A tabela abaixo nos mostra as frequências relativas simples para os dados:

Estatura (cm)	fi	Fi
142	1	2,5%
147	2	5%
148	2	5%
149	1	2,5%
(...)	(...)	(...)
Total	**40**	**100%**

Note que a **Fi** é construída obedecendo ao seguinte:

1º.É construída a partir da **fi**;
2º.A soma da **Fi** é sempre **100%**;
3º.O valor resultante da divisão entre 100% e **n** deve ser dividido/multiplicado por cada valor da **fi**.

a. Se **n** é menor do que 100, então **multiplica**;
b. Se **n** é maior do que 100, então **divide**.

Já a **Frequência Relativa Acumulada (Fac)** tem basicamente o mesmo significado da **fac**, servindo para indicar a **porcentagem** de observações que são menores ou iguais ao valor considerado. Para a nossa sequência de dados, podemos construir a seguinte tabela:

Estatura (cm)	fi	Fi	fac
142	1	2,5%	2,5%
147	2	5%	7,5%
148	2	5%	12,5%
149	1	2,5%	15%
(...)	(...)	(...)	(...)
Total	**40**	**100%**	**100%**

Iguais na primeira classe

Note que a **Fac** é construída, obedecendo ao seguinte:

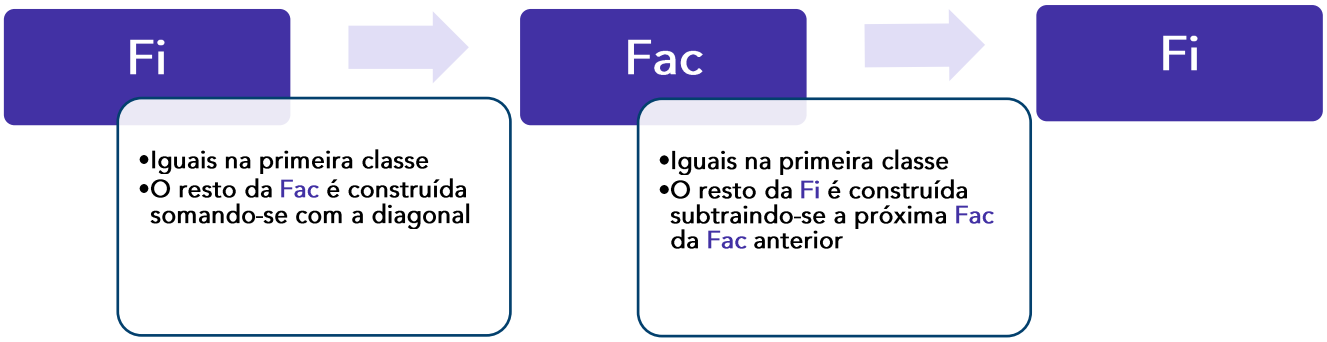
1º) É construída a partir da **Fi** (as duas são "frequências irmãs");

2º) A **Fac** e a **Fi** possuem o mesmo valor na classe em que a **Fac** começa a ser construída;

3º) Os demais valores da **Fac** serão obtidos somando-se o valor da **Fac** anterior com a **Fi** da diagonal;

4º) A **Fac** termina sempre com 100%.

Por fim, é muito importante sabermos realizar a **conversão de Fi para Fac e vice-versa**. Daí teremos:



7. Representação Gráfica de Dados

Chegamos a um ponto muito importante da nossa aula. Os gráficos estatísticos têm como objetivo transmitir uma informação direta e objetiva da ocorrência em análise, permitindo uma melhor visualização e também uma análise mais detalhada dos dados apresentados.

Professor, como as provas de concursos cobram este tópico?

Dificilmente encontraremos uma questão exigindo que você simplesmente desenhe um gráfico. Na verdade, percebemos que a tendência de cobrança nas provas está concentrada em duas formas:

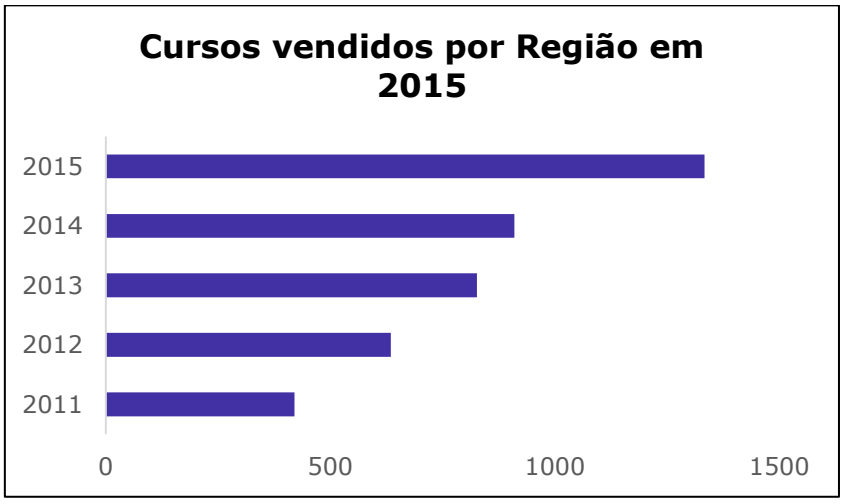
1) A questão cobra o seu conhecimento sobre os tipos de gráficos, de modo que você precisa ter em mente a finalidade e principais características de cada um deles; e

2) O enunciado apresenta um gráfico e solicita que você calcule uma determinada **medida estatística** com base nas informações fornecidas por meio dele. Ou seja, além de saber a forma adequada de calcular média, moda, mediana, etc, você precisa inicialmente interpretar de modo correto as informações trazidas no gráfico!

Assim, vamos começar a análise dos principais tipos de gráficos estatísticos.

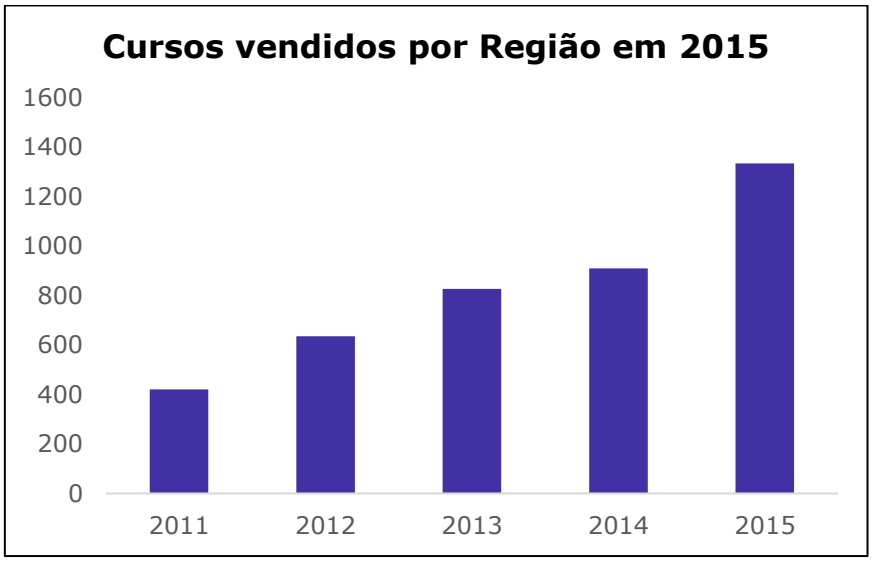
7.1. Gráfico em Barras

O **gráfico em Barras** é utilizado normalmente para representar uma série estatística (geralmente *séries geográficas*) através de **retângulos dispostos horizontalmente**, os quais possuem mesma altura e comprimentos proporcionais aos respectivos dados.



7.2. Gráfico de Colunas

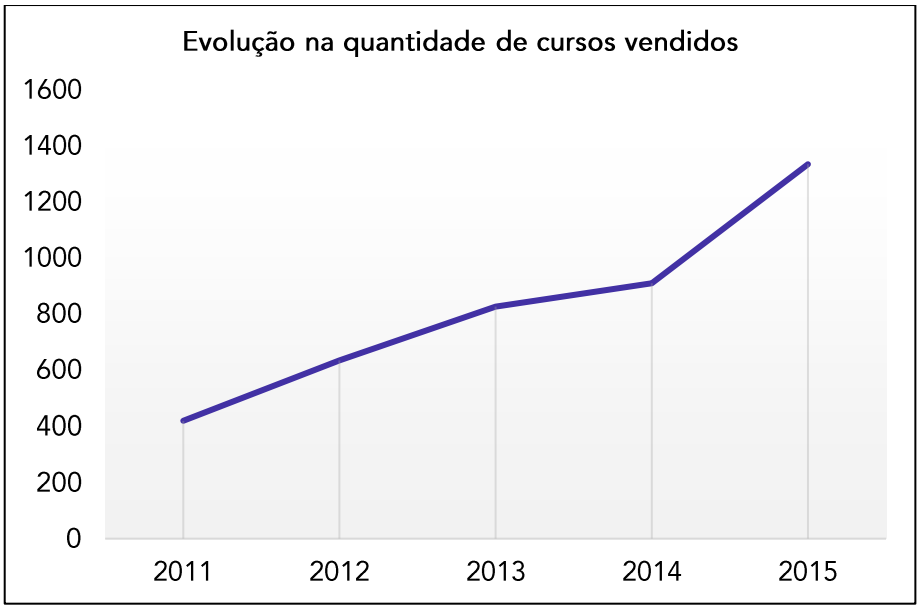
Similar ao gráfico em barras, o **Gráfico de Colunas** pode usado na representação de uma série estatística por meio de **retângulos dispostos verticalmente**, com a mesma base e alturas proporcionais aos respectivos dados.



7.3. Gráfico em Linhas

O **Gráfico em Linhas** é usado basicamente na representação de dados em *séries temporais*, com a finalidade de demonstrar a **evolução dos valores de uma variável no decorrer do tempo**.

Evolução na quantidade de cursos vendidos	
2011	420
2012	635
2013	827
2014	910
2015	1334

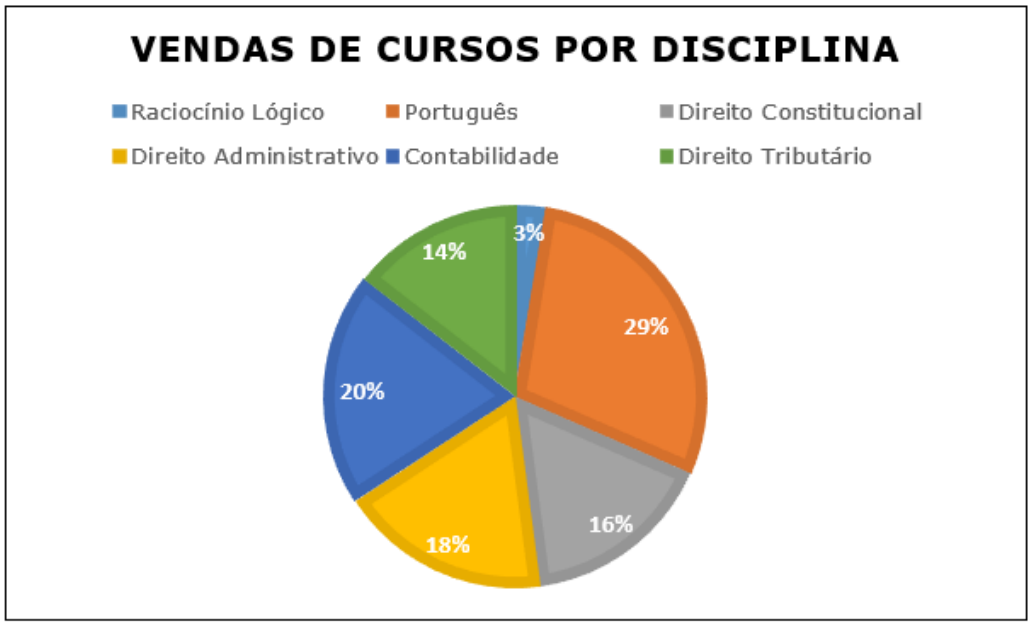


7.4. Gráfico em Setores

O **Gráfico em Setores** é utilizado com a finalidade de **exibir as partes de um todo**, isto é, quando o objetivo consiste em comparar proporções.

Quanto a sua estrutura, o gráfico em setores é formado por um **círculo**, em que cada classe de dados é representada através de um **setor circular**, cujo ângulo é proporcional ao tamanho da amostra e são obtidas por meio de uma regra de três simples, na qual considera-se que a área total da circunferência (100%) corresponde a 360º.

VENDAS DE CURSOS POR DISCIPLINA		
Disciplina	**Quantidade**	**Porcentagem do total**
Raciocínio Lógico	10	**2,63%**
Português	110	**28,95%**
Direito Constitucional	62	**16,32%**
Direito Administrativo	68	**17,89%**
Contabilidade	75	**19,74%**
Direito Tributário	55	**14,47%**
Total	**380**	**100,00%**



7.5. Gráfico de Hastes ou Bastões

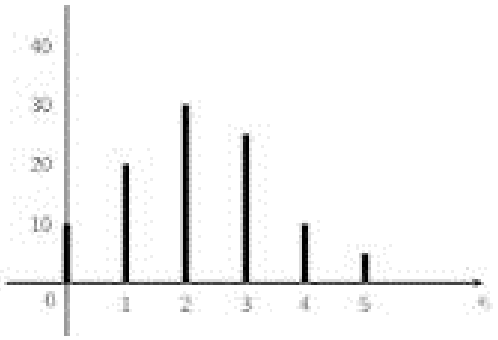
A partir deste tipo de gráfico precisaremos recordar as informações estudadas anteriormente, em especial quanto às distribuições de frequências.

O principal procedimento que espero fazê-los compreender e executar adequadamente diz respeito a montar a tabela de frequências com base no gráfico apresentado. Trata-se de conhecimento extremamente importante para o seu sucesso nas provas de concursos. Portanto, atenção total!

O **Gráfico de Hastes ou Bastões** é frequentemente usado para **representar dados não agrupados em classes** (normalmente *dados discretos*), em que não há perda de informação.

Em relação a sua construção, basta utilizar as **frequências simples absolutas ou relativas** de cada elemento.

x_i	f_i
0	10
1	20
2	30
3	25
4	10
5	5



Ressalto que é fundamental que você saiba realizar os dois caminhos, ou seja, a partir da tabela montar o gráfico e vice-versa.

7.6. Histograma

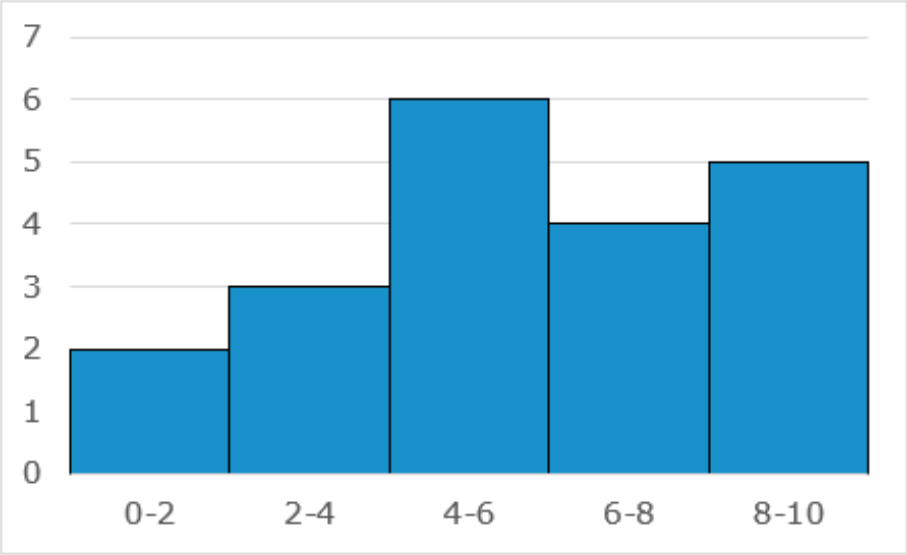
Preparem-se, estamos diante do gráfico mais frequente em provas de concursos públicos!

O **Histograma** é o gráfico destinado a representar **dados agrupados em classe**, de modo que haverá uma perda de informação, pois o trabalho realizado nas distribuições de frequências é com classes de elementos, e não com elementos dispostos individualmente.

Relacione sempre o Histograma com a Distribuição de Frequências em classe!

Quanto a sua formação, o Histograma pode ser construído com base em frequências simples absolutas ou relativas, indistintamente. Nesse sentido, as classes são representadas por **retângulos dispostos vertical e contiguamente** (sem espaço entre eles), com bases margeadas pelos limites destas classes e alturas determinadas pelas frequências de cada classe, em que o eixo das ordenadas apresenta a densidade de frequência e o das abscissas, as classes dos dados.

Notas de 20 alunos numa prova	
Xi	**fi**
0-2	2
2-4	3
4-6	6
6-8	4
8-10	5
Total	**20**



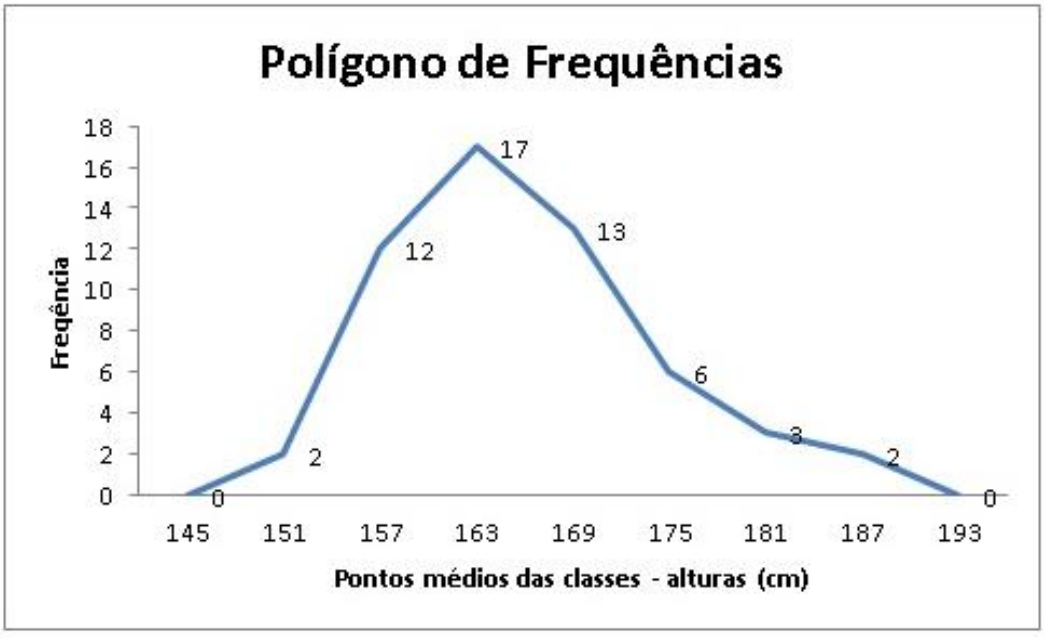
Por fim, vale a pena mencionar a existência de um gráfico chamado **Poligonal Característica**, o qual é construído com base apenas nos **contornos do histograma**.

O histograma é semelhante ao gráfico de colunas, mas nesse tipo de gráfico, não existem espaços em branco entre as colunas.

7.7. Polígono de Frequências

Polígono de Frequências é o **gráfico em linhas** obtido através da união dos **pontos médios** das bases superiores dos retângulos de um **Histograma** (que utiliza colunas verticais para mostrar as frequências).

Por sua vez, os polígonos de frequência para dados agrupados constroem-se a partir da marca de classe que coincide com o ponto médio de cada coluna do histograma.



O ponto mais alto num polígono de frequência representa a **maior frequência**, ao passo que a **área abaixo da curva** inclui a totalidade dos dados existentes.

7.8. Curva de Frequências

Curva de Frequências é o gráfico construído a partir do **Polígono de Frequências**, por meio de ***contornos mais suaves*** (polígono de frequência polido), deixando evidente a verdadeira natureza da distribuição da população em análise.

Dessa maneira, podemos concluir que, enquanto o polígono de frequência nos dá a **imagem real** do fenômeno estudado, a curva de frequência nos dá a **imagem tendenciosa**.

Nesse sentido, após o traçado de um polígono de frequência, é desejável, muitas vezes, que se faça um **polimento** (geometricamente, corresponde à eliminação dos vértices da linha poligonal), de modo a mostrar o que seria tal polígono com um número maior de dados. Isso é realizado por meio do emprego de uma fórmula bastante simples:

$$f_{ci} = \frac{(f_{ant} + 2fi + f_{post})}{4}$$

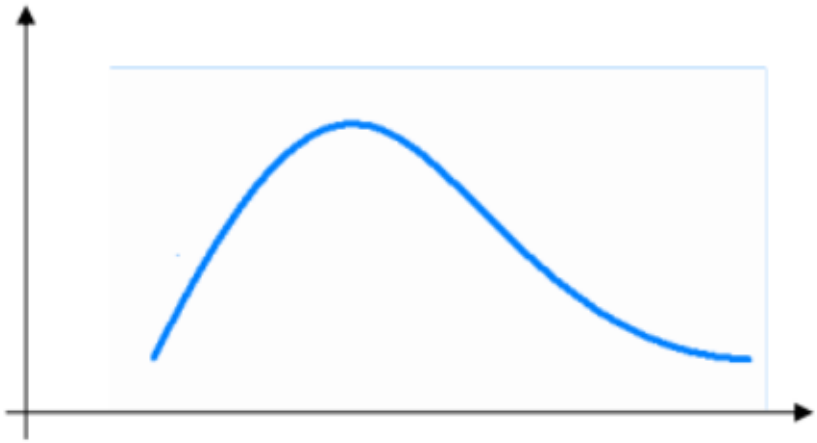
Em que:

$\mathbf{f_{ci}}$ = frequência calculada da classe considerada (freq. polida)

fi = frequência simples da classe considerada

$\mathbf{f_{ant}}$ = frequência simples da classe anterior à da classe considerada

$\mathbf{f_{post}}$ = frequência simples da classe posterior à da classe considerada

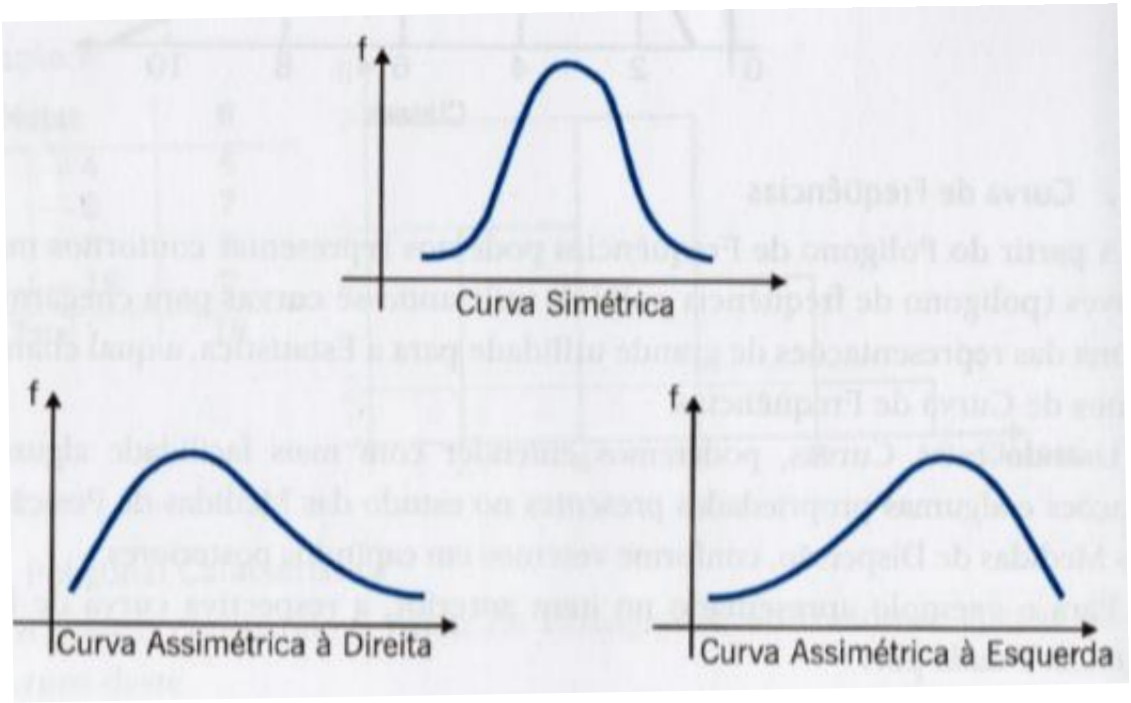


Por fim, mencionamos que a grande importância das curvas de frequências reside na melhor compreensão que elas proporcionam de algumas relações e propriedades presentes no estudo das Medidas de Posição e das Medidas de Dispersão.

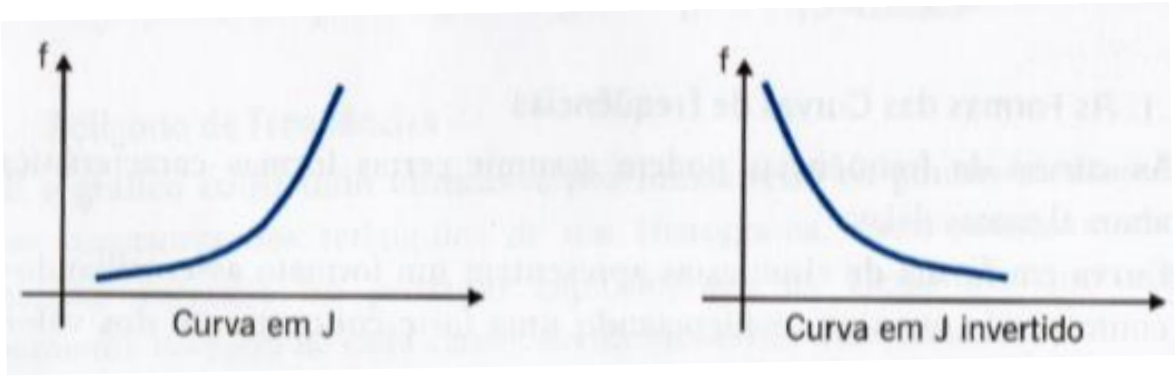
7.8.1. Formas das Curvas de Frequências

Vejamos algumas **formas** específicas podem ser assumidas pelas curvas de frequências:

❖ **Curvas em forma de sino**: as curvas evidenciam uma **notável concentração dos valores em torno da região central** da distribuição, podendo apresentar uma certa assimetria à esquerda (cauda mais alongada à esquerda) ou à direita (cauda mais alongada à direita). Assim, três **configurações** para as curvas em formas de sina podem surgir:



❖ **Curvas em forma de jota**: as curvas apresentam ponto de ordenada (eixo vertical do gráfico) máxima *em uma das extremidades*.

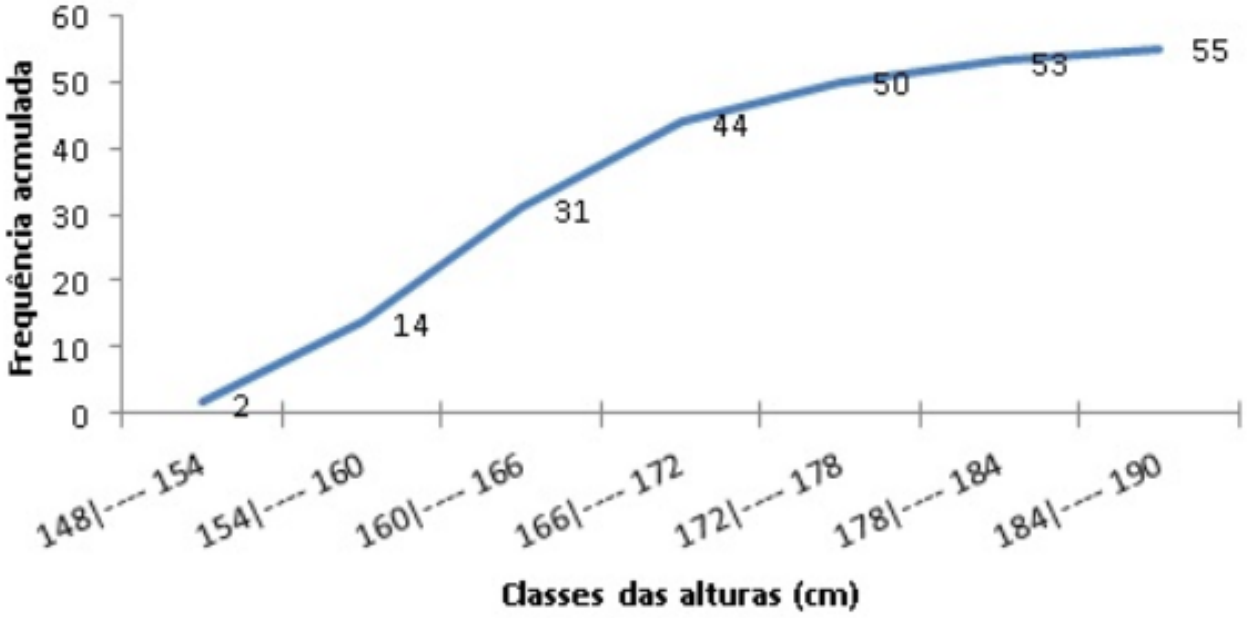


❖ **Curvas em forma de U**: possuem ponto de ordenada máxima *em ambas as extremidades*.

7.9. Ogiva

O gráfico estatístico chamado **Ogiva** é obtido a partir das **frequências acumuladas**, razão pela qual também é conhecido como **Polígono de Frequências Acumuladas**, possibilitando informações quanto às medidas separatrizes.

Como é construída a Ogiva? Bem, inicialmente deve-se desenhar, para cada classe, os retângulos com alturas correspondentes ao valor da **frequência acumulada** (crescente ou decrescente). Em seguida, é realizada propriamente a formação da Ogiva por **interligar os pontos das bases superiores dos retângulos**, que equivalem aos limites superiores de cada classe.



7.10. Box Plot (Gráfico de Caixa)

O **Box plot** é um gráfico que apresenta a informação contida nos **quartis** e nos pontos extremos dos dados, permitindo, caso se tenha mais de um caso (por exemplo, dados associados a diferentes valores de uma variável qualitativa), fazer a comparação entre os casos de forma visualmente fácil.

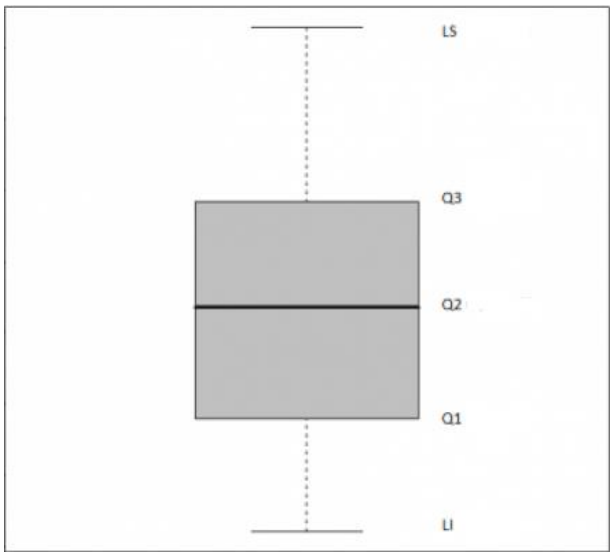
Com relação à construção do Gráfico de Caixa, é realizada por meio de um **retângulo** alinhado verticalmente (ou horizontalmente) com duas **semirretas**, uma em cada um dos lados opostos do retângulo. A altura do retângulo é definida pelos quartis Q1 e Q3. Uma linha secciona o retângulo no valor da mediana (ou Q2). As semirretas ligam respectivamente os quartis Q1 e Q3 ao valor mínimo e ao máximo do conjunto de dados. Assim, os seguintes passos devem ser seguidos:

1º) calcular os **quartis**, de forma a desenhar a caixa.

2º) Obter a **distância entre o primeiro e o terceiro quartil** (amplitude interquartil): Q3 - Q1

3º) Os **bigodes** tem como tamanho máximo 1,5 vezes a amplitude interquartil, mas vão apenas até o último dado dentro desta distância.

4º) Identificar com **asteriscos** (*outliers*) os dados que estiverem além.



QUESTÕES COMENTADAS

1. (CESPE/ACE/TC-DF/2002) Por Estatística Descritiva entende-se um conjunto de ferramentas, tais como gráficos e tabelas, cujo objetivo é apresentar, de forma resumida, um conjunto de observações.

Comentários:

É exatamente isso, pessoal! A **Estatística Descritiva** tem como foco a **apresentação de conjunto de dados** de forma sucinta. Para isso, faz uso de tabelas, gráficos e algumas medidas, como medidas de posição, dispersão etc.

Gabarito: Certo.

2. (CESPE - Agente Penitenciário Federal/2015) A diferença entre um censo e uma amostra consiste no fato de esta última exigir a realização de um número maior de entrevistas.

Comentários:

A questão trata da diferença conceitual entre censo e amostra.

O **censo** consiste em entrevistar **todos os elementos da população**.

Já a **amostra** consiste em entrevistar um **subconjunto da população**.

Dessa forma, fica claro que **o censo exige mais entrevistas**.

Gabarito: Errado.

3. (CESPE/Fiscal de Tributos Estaduais/Sefaz-AL/2002) Como a realização de um censo tipicamente é muito onerosa e(ou) demorada, muitas vezes é conveniente estudar um subconjunto próprio da população, denominado amostra.

Comentários:

A afirmação está correta. Imagine que se deseja obter informações sobre intenções de votos para governador de determinado estado brasileiro. Não há dúvidas quanto à expressiva onerosidade e ao significativo esforço a serem empregados em uma pesquisa de toda a população de eleitores do estado, isto é, em um **censo**.

Como alternativa, temos a **amostra**, ou seja, pesquisa-se parcela da população que contenha os seus elementos essenciais, de modo a se ter uma pesquisa viável e com um nível aceitável de confiança.

Gabarito: Certo.

4. (CESPE/ Estatístico/SEDUC-AM/ 2011) O ramo e folhas é uma ferramenta exploratória que, além de servir para ordenar um pequeno conjunto de dados, permite destacar a forma da sua distribuição.

Comentários:

Isso mesmo, o **diagrama de ramos e folhas** é utilizado para **explorar dados** estatísticos. Seu uso permite criar uma **organização** ou ordenação das informações, assim como ocorre em relação ao rol. Além disso, ele **evidencia a distribuição dos dados**, dando destaque àqueles com maior relevância no âmbito do conjunto em análise.

Gabarito: Certo.

5. (CESPE/AUFC/TCU/2008)

tipo	2006	2007
funcionários	70	74
estagiários	16	16
prestadores de serviço	14	10
total	**100**	**100**

Demonstrações contábeis da CAIXA – Exercício de 2007.
Internet: <www.caixa.gov.br> (com adaptações).

Considerando a tabela acima, que apresenta a distribuição do quadro de colaboradores da CAIXA, em mil pessoas, no final dos anos de 2006 e 2007, julgue os itens seguintes.

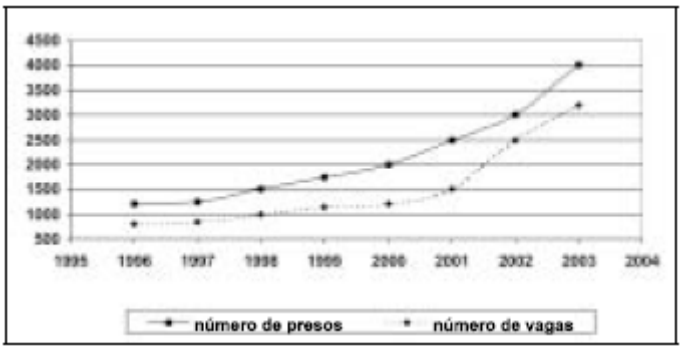
Se uma variável X registra a classificação de cada pessoa do quadro como funcionária, estagiária ou prestadora de serviço, então X é uma variável qualitativa.

Comentários:

Perfeito. As possíveis realizações da variável **X** são atributos, e **não números**. Logo, trata-se de uma **variável qualitativa**.

Gabarito: Certo.

6. (CESPE/PPF/PF/2004)



Nos últimos oito anos, a população carcerária em uma unidade da Federação cresceu de 1.200 presos (1996) para 4.000 presos (2003). Essa população carcerária é formada por presos nas casas penais, seccionais e delegacias. Por causa desse crescimento, foram construídas novas cadeias públicas, penitenciárias e novos blocos carcerários. Mesmo assim, não foi possível resolver o problema de superlotação. Em 1996, a capacidade total de lotação das casas penais, seccionais e delegacias era de apenas 800 vagas. Após a inauguração das novas instalações em 2003, o número de vagas aumentou para 3.200, o que resulta em déficit de 800 vagas. O gráfico acima apresenta a evolução temporal da população carcerária (linha contínua) e do número de vagas (linha pontilhada) de 1996 a 2003.

A capacidade total de lotação das casas penais, seccionais e delegacias (número de vagas) em 2000 é uma variável aleatória contínua.

Comentários:

Note que a capacidade total de lotação **não** **pode assumir qualquer valor** em um intervalo real, pois assume apenas valores inteiros. Assim, trata-se de uma **variável discreta**, as quais indicam características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores.

Ademais, considerando apenas o caso do ano 2000, temos uma observação fixa. Logo, a capacidade de lotação para o ano 2000 é uma constante, sem variação.

Gabarito: Errado.

7. (CESPE/TRE-ES/2011)

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

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

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

A variável "cargo" classifica-se como uma variável qualitativa ordinal.

Comentários:

Aprendemos que variável **qualitativa ordinal** é aquela em que **é possível verificar uma ordem** ou hierarquia entre as respostas obtidas.

E é isso que ocorre com a variável “cargo”, já que além de se tratar de atributos (e não números), podemos ordenar as observações, como é o caso do número de cargos disponíveis.

Gabarito: Certo.

8. (CESPE/TCU/AUFC/2008) Uma agência de desenvolvimento urbano divulgou os dados apresentados na tabela a seguir, acerca dos números de imóveis ofertados (X) e vendidos (Y) em determinado município, nos anos de 2005 a 2007.

ano	número de imóveis	
	ofertados (X)	vendidos (Y)
2005	1.500	100
2006	1.750	400
2007	2.000	700

Correio Braziliense, 29/4/2008, p. 17 (com adaptações).

A variável *X* forma uma série estatística denominada série temporal.

Comentários:

A **série estatística** apresentada no enunciado possui os seguintes elementos:

- **Fato**: é o **mesmo** (número de imóveis ofertados).
- **Local**: é **fixo** (determinado município).
- **Época**: é **variável** (anos de 2005 a 2006).

Assim, estamos diante de uma **série histórica** (cronológica ou temporal).

Gabarito: Certo.

9. (CESPE/CEHAP-PB/Administrador/2009 - Adaptada) O custo médio nacional para a construção de habitação com padrão de acabamento normal, segundo levantamento realizado em novembro de 2008, foi de R$ 670,00 por metro quadrado, sendo R$ 400,00/m2 relativos às despesas com materiais de construção e R$ 270,00/m2 com mão-de-obra. Nessa mesma pesquisa, os custos médios regionais apontaram para os seguintes valores por metro quadrado: R$ 700,00 (Sudeste), R$ 660,00 (Sul), R$ 670,00 (Norte), R$ 640,00 (Centro-Oeste) e R$ 630,00 (Nordeste).

Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil. SINAPI/IBGE, nov./2008.

Considerando o par de variáveis (X,Y), em que X representa o custo médio por metro quadrado (em R$) e Y representa os atributos Sudeste, Norte, Sul, Centro-Oeste ou Nordeste, é correto afirmar que as variáveis X e Y formam uma série geográfica.

Comentários:

A **série geográfica** (ou de localização) é aquela em que os dados são observados segundo a **localidade de ocorrência**, cujo **local varia** e a época e o fato são fixos.

No caso da série estatística apresentada, temos que o custo médio por metro quadrado (em R$) varia conforme o atributo relacionado às regiões do país. Assim, estamos diante de uma **série geográfica**.

Gabarito: Certo.

10. (CESPE/ Auditor de Controle Externo/TCE-PA/ 2016)

número diário de denúncias registradas (X)	frequência relativa
0	0,3
1	0,1
2	0,2
3	0,1
4	0,3
total	**1,0**

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

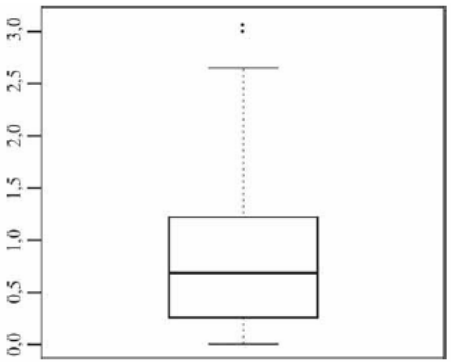
A variável X é do tipo qualitativo nominal.

Comentários:

Note que a variável X diz respeito à **quantidade** de denúncias diárias registradas, portanto, são atribuídos valores numéricos a ela. Logo, concluímos que X é uma variável **quantitativa** e não qualitativa.

Gabarito: Errado.

11. (CESPE/ Auditor de Controle Externo/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.

X representa uma variável qualitativa ordinal.

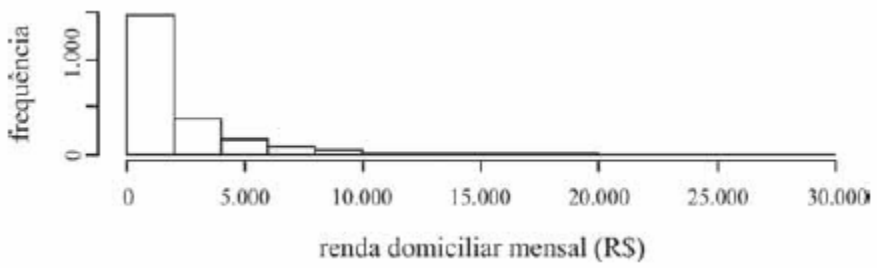
Comentários:

Segundo a tabela do enunciado, podemos observar que a variável X possui uma distribuição amostral cujo valor **mínimo** é **0** e **máximo** é **3,10**. Assim, percebemos que são atribuídos valores numéricos à variável X, o que nos leva a classificá-la como uma variável **quantitativa**.

O diagrama apresentado é chamado de box-plot, o qual veremos melhor futuramente.

Gabarito: Errado.

12. (CESPE/Ana Sup/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.

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

Comentários:

Não obstante o **CEP** ser representado por números, ele não pode ser **contado** e **nem medido**. Com isso, concluímos que o CEP não é uma variável **quantitativa**. Assim, já podemos avaliar que a afirmação da questão está **errada**.

Por sua vez, é possível contar os **valores monetários** da última fatura vencida, de modo que tal variável é **quantitativa**.

Gabarito: Errado.

13. (CESPE/TST/Ana Jud/2008) Considere que, em um ambiente de trabalho industrial, as seguintes medições acerca da poluição do ar tenham sido observadas:

1, 6, 4, 3, 2, 3, 1, 5, 1, 4.

A terceira e a oitava estatísticas de ordem são respectivamente iguais a 4 e 5.

Comentários:

Perceba que as medições acerca da poluição do ar constituem **dados brutos**, pois não estão organizados. Daí, precisamos colocar os valores observados em ordem crescente, formando um **rol**, a fim de verificar quais posições ocupam a terceira e a oitava posições. Daí, teremos:

1, 1, **1**, 2, 3, 3, 4, **4**, 5, 6

Logo, o valor que ocupa a terceira posição é o **1**; e o da oitava posição é o **4**. Portanto, o item está **errado** por afirmar que a terceira e a oitava posições são respectivamente iguais a 4 e 5.

Gabarito: Errado.

14. (CESPE/SEFAZ-AL/2002) Em uma distribuição de frequências para um conjunto de n indivíduos, pode-se calcular as frequências relativas, dividindo-se cada frequência absoluta pela amplitude da classe ou intervalo.

Comentários:

Acabamos de aprender que a **frequência relativa simples (Fi)** é dada pela **frequência absoluta simples dividida pelo número de dados**:

$$Fi = \frac{fi}{n}$$

Gabarito: Errado.

15. (CESPE/BACEN/Analista/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.

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:

O enunciado da questão apresenta os dados brutos de uma amostra. A tabela seguir organiza as informações:

Valor (X)	Frequência (f)	Frequência acumulada (F)	Memória de cálculo
1	2	2	=2
2	2	4	=2+2
3	1	5	=4+1
4	4	9	=5+4
5	3	12	=9+3
7	1	13	=12+1
8	2	15	=13+2
12	1	16	=15+1
14	1	17	=16+1
19	1	18	=17+1
24	1	19	=18+1
32	1	20	=19+1

Como é possível observar, **13 valores são inferiores a 8**. Nesse sentido, basta notar a frequência acumulada para X = 7. Assim, dividindo 13 pelo total de observações, temos:

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

Portanto, trata-se de uma **frequência relativa maior que 60%**.

Gabarito: Certo.

16. (CESPE - Analista/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.

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

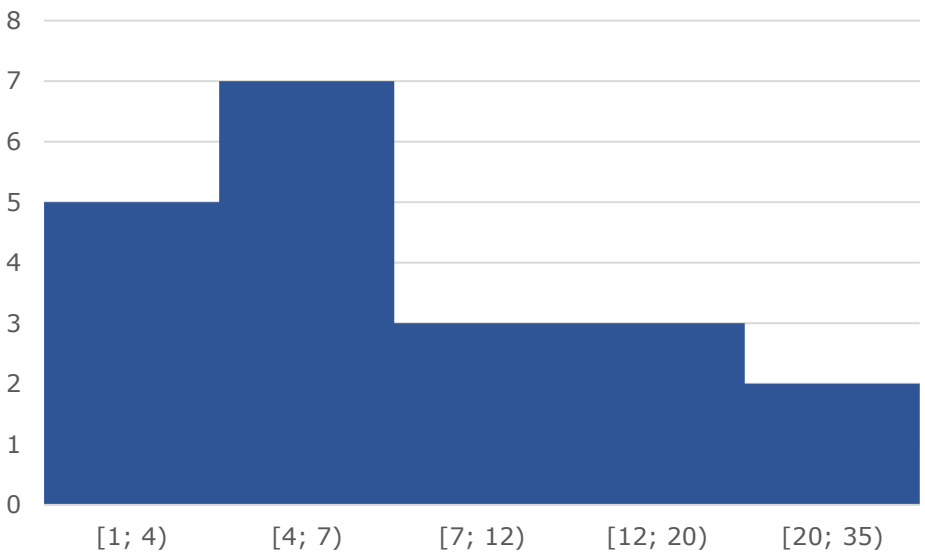
Comentários:

O enunciado apresenta dados relativos a uma amostra, em minutos, do tempo utilizado na armazenagem de formulários no almoxarifado central de certa instituição por diversos funcionários.

A tabela a seguir agrupa os dados em **classe**:

Classe	Frequência
[1; 4)	5
[4; 7)	7
[7; 12)	3
[12; 20)	3
[20; 35)	2

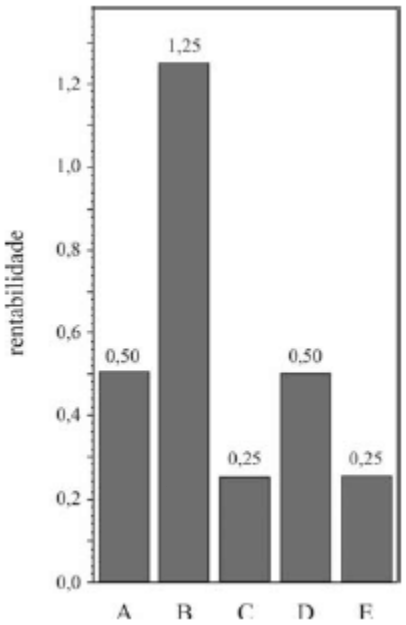
Dessa maneira, podemos construir o **histograma** abaixo:



Portanto, o item está **errado** ao afirmar que seria inviável a elaboração de um histograma em decorrência do fato de ser este um conjunto de dados quantitativos discretos.

Gabarito: Errado.

17. (CESPE/FUNPRESP/Analista de Investimentos/2016)



O gráfico ilustra cinco possibilidades de fundos de investimento com suas respectivas rentabilidades. Considerando que as probabilidades de investimento para os fundos A, B, C e D sejam, respectivamente, P(A) = 0,182; P(B) = 0,454; P(C)= 0,091; e P(D)= 0,182, julgue o item subsequente.

O gráfico apresentado é um histograma.

Comentários:

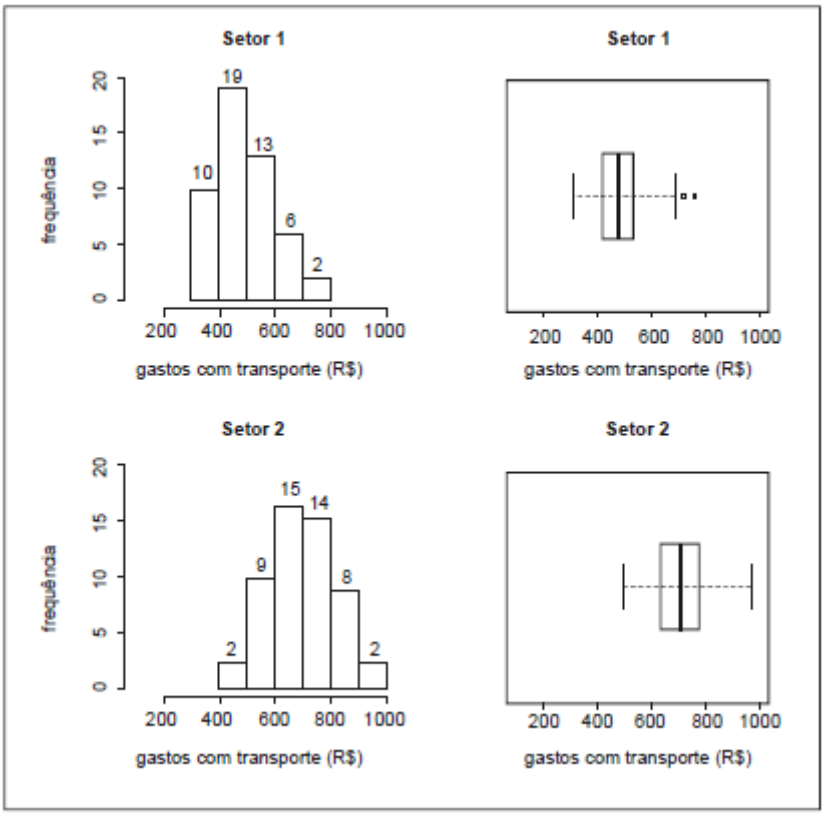
O **Histograma** é o gráfico destinado a representar **dados agrupados em classe**, que pode ser construído com base em frequências simples absolutas ou relativas, indistintamente.

Nesse sentido, as classes são representadas por **retângulos dispostos vertical e contiguamente** (sem espaço entre eles), com bases margeadas pelos limites destas classes e alturas determinadas pelas frequências de cada classe, em que o eixo das ordenadas apresenta a densidade de frequência e o das abscissas, as classes dos dados.

Assim, o gráfico apresentado no enunciado **não é um histograma**, mas sim um simples **gráfico em barras**.

Gabarito: Errado.

18. (CESPE – Analista/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
		723,79	767,53		

Setor 2

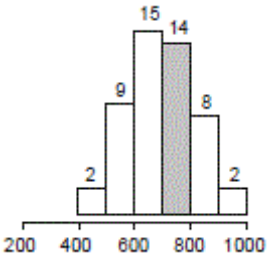
488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56

663,81	670,12	671,90	673,78	684,69	685,98
693,35	698,58	708,78	719,80	721,16	734,84
735,94	746,34	754,83	756,10	756,96	760,80
762,29	766,24	770,11	797,73	804,06	805,97
807,29	832,83	844,00	866,77	878,27	897,09
		943,10	963,25		

Na amostra do setor 2, a frequência relativa das despesas entre R$ 700,00 e R$ 799,99 é superior a 20%.

Comentários:

Destacamos abaixo a barra do **histograma** relativa ao intervalo de 700 a 799,99:



Percebe-se que a frequência dessa classe é de 14.

Por sua vez, o cálculo da **frequência relativa** é feito dividindo-se 14 pelo total das frequências (50):

$$\frac{14}{50} = 0{,}28 = \mathbf{28}\%$$

Assim, a frequência relativa das despesas entre R$ 700,00 e R$ 799,99 é, de fato, **superior a 20%**.

Gabarito: Certo.

19. (FCC/ARCE/2006) O processo estatístico que consiste em avaliação direta de um parâmetro, utilizando-se todos os componentes da população chama-se:

a) Amostragem

b) Estimação

c) Censo

d) Parametrização

e) Correlação

Comentários:

Acabamos de ver que, se uma pesquisa é realizada com **todos os indivíduos de uma população**, estaríamos realizando um **CENSO**.

Gabarito: C.

20. (FCC/MPU/Analista/Estatística/2007) Uma empresa procurou estudar a ocorrência de acidentes com seus empregados e realizou um levantamento por um período de 36 meses. As informações apuradas estão na tabela a seguir:

Número de empregados acidentados	Número de meses
1	1
2	2
3	4
4	5
5	7
6	6
7	5
8	3
9	2
10	1

A porcentagem de meses em que houve menos de 5 empregados acidentados é

a) 50%

b) 45%

c) 35%

d) 33%

e) 30%

Comentários:

O enunciado da questão apresentou a seguinte **tabela**:

Nº de empregados acidentados	Nº de meses
1	1
2	2
3	4
4	5
5	7
6	6

7	5
8	3
9	2
10	1

Estamos interessados nas frequências associadas a observações **menores que 5**, pois a questão quer a porcentagem de meses em que houve menos de 5 empregados acidentados. Vamos destacar isso na tabela:

Nº de empregados acidentados	Nº de meses
1	**1**
2	**2**
3	**4**
4	**5**
5	**7**
6	**6**
7	**5**
8	**3**
9	**2**
10	**1**

Fazendo o somatório dessas frequências:

$$1 + 2 + 4 + 5 = \mathbf{12}$$

Isto significa que em 12 dos 36 meses considerados tivemos menos de 5 empregados acidentados. Percentualmente, temos:

$$\frac{12}{36} = \mathbf{33,33\%}$$

Gabarito: D.

21. (FCC/TRT 6ª Região/Analista Judiciário/2012) A distribuição dos 500 preços unitários de um equipamento é representada por um histograma em que no eixo das abscissas constam os intervalos de classe e no eixo das ordenadas estão assinaladas as respectivas densidades de frequências, em (R$)−1. Define-se densidade de frequência de um intervalo de classe como sendo o resultado da divisão da

respectiva frequência relativa pela correspondente amplitude do intervalo. Um intervalo de classe no histograma apresenta uma amplitude de R$ 2,50 com uma densidade de frequência igual a 0,096. A quantidade de preços unitários referente a este intervalo é

a) 96

b) 120.

c) 144.

d) 150.

e) 192.

Comentários:

Conforme descrito no enunciado, a **densidade de frequência (d_f)** corresponde à divisão entre a frequência relativa (Fi) e a amplitude da classe (h):

$$\boldsymbol{d_f = \frac{Fi}{h}}$$

$$0{,}096 = \frac{Fi}{2{,}50}$$

$$\boldsymbol{Fi} = 0{,}096 \,.\, 2{,}50 = 0{,}24 = \mathbf{24}\%$$

O que significa este valor que encontramos? Bem, temos que 24% dos preços estão neste intervalo em consideração. Visto que existem 500 preços unitários no total, a quantidade correspondente será:

$$24\% \; de \; 500 = 0{,}24 \,.\, 500 = \mathbf{120}$$

Gabarito: B.

22. (FCC - ATARH/SERGAS/2013 - Adaptada) O gráfico de setores não é adequado para representar variáveis quantitativas.

Comentários:

O **Gráfico de Setores** é sim adequado para representar **variáveis quantitativas**, bastando para isso agrupar os dados considerados em intervalos de valores.

Gabarito: Errado.

23. (FCC - ATARH/SERGAS/2013 - Adaptada) Histograma é um gráfico que apresenta a distribuição de frequências de uma variável por meio de retângulos justapostos, feitos sobre as classes dessa variável, sendo que a área de cada retângulo é proporcional à frequência observada da correspondente classe.

Comentários:

Perfeito, meus amigos! A Fundação Carlos Chagas apresentou fielmente neste item as **características básicas de um Histograma**:

✓ Apresenta distribuição de frequências relativas a **classes** de dados;
✓ Utilização de **retângulos justapostos**;
✓ A **área de cada retângulo é proporcional à frequência** observada da correspondente classe.

Gabarito: Certo.

24. (FCC – Analista Judiciário/TRT 1ª Região/2011) Um histograma representa a distribuição dos preços unitários de venda de um determinado equipamento no mercado. No eixo das ordenadas estão assinaladas as respectivas densidades de frequência para cada intervalo em (R$)−1. Define-se densidade de frequência de um intervalo de classe como sendo o quociente da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo. Um intervalo de classe do histograma corresponde aos preços unitários maiores ou iguais a R$ 32,00 e inferiores a R$ 44,50 com uma densidade de frequência igual a 1,6 × 10−2 (R$)−1. Se todos os intervalos de classe do histograma têm a mesma frequência relativa, então um intervalo de classe com densidade de frequência igual a 5,0 × 10−3 (R$)−1 apresenta uma amplitude de

a) R$ 64,00

b) R$ 48,00

c) R$ 40,00

d) R$ 32,00

e) R$ 24,00

Comentários:

O enunciado trata de dois intervalos distintos, de modo que iremos analisá-los separadamente.

- **Intervalo de R$ 32,00 a R$ 44,50**:

A **amplitude** desse intervalo é de:

$$44{,}50 - 32 = \mathbf{12{,}50}$$

Foi dito que a **densidade** desse intervalo é de 1,6 . 10^{-2}. Bem, a densidade de frequência de um intervalo de classe (d_f) é definida como sendo a razão entre a frequência relativa (Fi) pela correspondente amplitude do intervalo (h). Daí, temos:

$$\boldsymbol{d_f = \frac{Fi}{h}}$$

$$1{,}6 \,.\, 10^{-2} = \frac{Fi}{12{,}50}$$

$$\boldsymbol{Fi} = 1{,}6 \,.\, 10^{-2} \,.\, 12{,}50 = 20 \,.\, 10^{-2} = 0{,}2 = \mathbf{20}\%$$

Desse modo, visto que foi informado que todos os intervalos possuem a mesma frequência, podemos concluir que **todos eles têm frequência de 20%**.

- **Intervalo com densidade de 5,0 . 10^{-3} (R$)$$^{-1}$**:

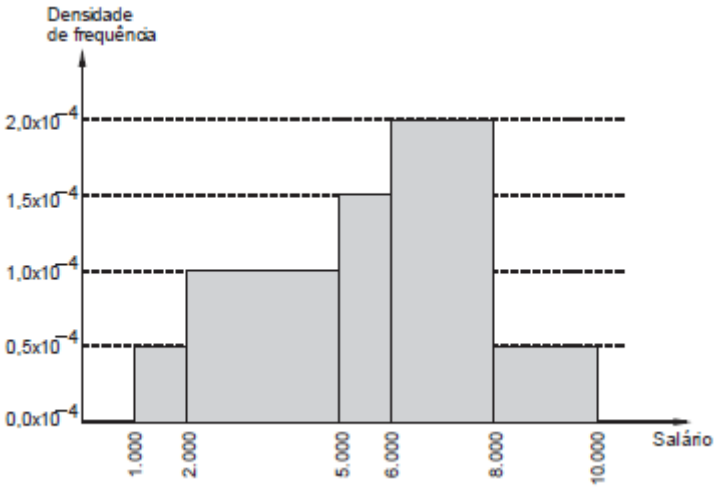
A **frequência relativa** desse intervalo é de **20%**, pois sabemos que todos os intervalos têm a mesma frequência. Logo, podemos obter a **amplitude**:

$$\boldsymbol{d_f = \frac{Fi}{h} \rightarrow h = \frac{Fi}{d_f}}$$

$$\boldsymbol{h = \frac{0{,}2}{5 \,.\, 10^{-3}} = 40}$$

Gabarito: C.

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



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

a) 104.

b) 152.

c) 136.

d) 144.

e) 120.

Comentários:

Inicialmente vamos encontrar a **frequência relativa** de cada classe, utilizando a seguinte fórmula:

$$\boldsymbol{Fi = d_f \,.\, h}$$

- **1ª classe**: $(0{,}0 \,.\, 10^{-4}) \,.\, (1.000 - 0) = \mathbf{0}$
- **2ª classe**: $(0{,}5 \,.\, 10^{-4}) \,.\, (2.000 - 1.000) = \mathbf{0{,}05}$
- **3ª classe**: $(1{,}0 \,.\, 10^{-4}) \,.\, (5.000 - 2.000) = \mathbf{0{,}3}$
- **4ª classe**: $(1{,}5 \,.\, 10^{-4}) \,.\, (6.000 - 5.000) = \mathbf{0{,}15}$
- **5ª classe**: $(2{,}0 \,.\, 10^{-4}) \,.\, (8.000 - 6.000) = \mathbf{0{,}4}$
- **6ª classe**: $(0{,}5 \,.\, 10^{-4}) \,.\, (10.000 - 8.000) = \mathbf{0{,}1}$

Em seguida, a tabela de frequências abaixo apresenta os dados do histograma:

Classes	Fi	Fac
[0-1.000)	0	0
[1.000-2.000)	0,05	0,05
[2.000-5.000)	**0,3**	**0,35**
[5.000-6.000)	**0,15**	0,50
[6.000-8.000)	**0,4**	0,90

[8.000-10.000)	0,1	1,00
	1,00	

Precisamos saber que o número de empregados com salários inferior a R$ 5.000,00 corresponde às três primeiras classes. Por sua vez, o número de funcionários com salário igual ou superior a R$ 2.000,00 e inferior a R$ 8.000 está relacionado às 3ª, 4ª e 5ª classes.

Além disso, os 56 empregados com salários inferiores a R$ 5.000,00 correspondem a Fac de 0,35 (colorido de verde). Assim, para encontrarmos a frequência acumulada correspondente ao **número de funcionários com salário igual ou superior a R$ 2.000,00 e inferior a R$ 8.000,00**, devemos somar as frequências relativas da 3ª, 4ª e 5ª classes (colorido de amarelo):

$$0{,}3 + 0{,}15 + 0{,}4 = \mathbf{0,85}$$

Por fim, resolvemos a seguinte regra de três simples:

$$0{,}35 \rightarrow 56$$

$$0{,}85 \rightarrow x$$

$$\boldsymbol{x = \frac{0{,}85 \,.\, 56}{0{,}35} = 136}$$

Gabarito: C.

26. (FJG - FR/Pref RJ/2002) Os dados de um determinado estudo representam muitas variáveis para cada uma das pessoas que se submeteram ao estudo. Uma variável considerada qualitativa é a seguinte:

a) idade

b) altura

c) sexo

d) peso

Comentários:

Aprendemos que uma variável pode ser qualitativa ou quantitativa. A principal diferença entre elas é que à **variável quantitativa** pode ser atribuído um **valor numérico**, enquanto para a **variável qualitativa** isso não é possível, por se tratar de **atributos**.

Nesse sentido, as variáveis "idade", "altura" e "peso" correspondem a quantidades. Logo, tratam-se de **variáveis quantitativas**.

Por outro lado, o mesmo não ocorre com a variável “sexo”. Ela tem duas realizações: masculino e feminino. Logo, trata-se de **variáveis qualitativas.**

Gabarito: C.

27. (VUNESP/Pref-SP/Biblioteconomia/2015 - Adaptada) Uma variável de natureza quantitativa pode ser classificada como ordinal ou nominal.

Comentários:

As **variáveis de natureza quantitativa** podem ser classificadas em **discretas** ou **contínuas**, ao passo que as **varáveis qualitativas** são classificadas em **nominais** ou **ordinais**.

Gabarito: Errado.

28. (CESGRANRIO/FINEP/Assistente/2014) A Tabela a seguir apresenta a frequência absoluta das faixas salariais mensais dos 20 funcionários de uma pequena empresa.

Faixa salarial (R$)	Frequência Absoluta
Menor que 1.000,00	6
Maior ou igual a 1.000,00 e menor que 2.000,00	7
Maior ou igual a 2.000,00 e menor que 3.000,00	5
Maior ou igual a 3.000,00	2
Total	20

A frequência relativa de funcionários que ganham mensalmente menos de R$ 2.000,00 é de

a) 0,07

b) 0,13

c) 0,35

d) 0,65

e) 0,70

Comentários:

Vimos que a **frequência relativa simples (Fi)** é dada pela **frequência absoluta simples dividida pelo número de dados**:

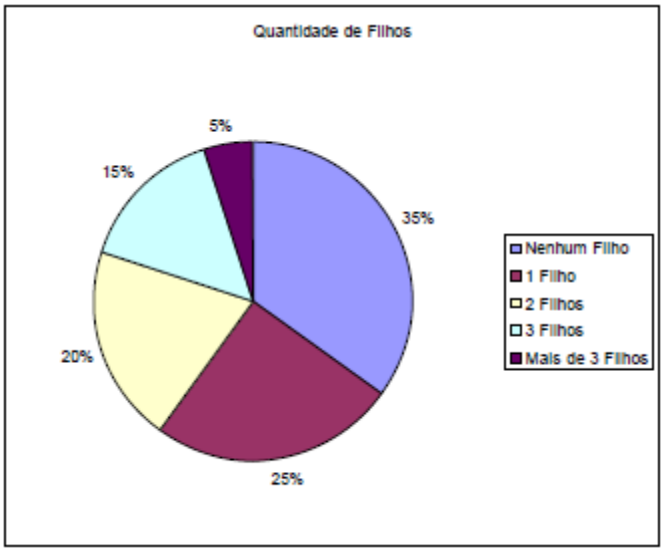
$$Fi = \frac{fi}{n}$$

Por sua vez, a **frequência absoluta (fi)** é dada pela soma das frequências das duas primeiras classes, já que é nelas que estão presentes os salários menores que R$ 2.000,00. Por sua vez, o número total de elementos (**n**) é de 20. Logo:

$$\mathbf{Fi} = \frac{\mathbf{fi}}{\boldsymbol{n}} = \frac{6+7}{20} = \mathbf{0,65}$$

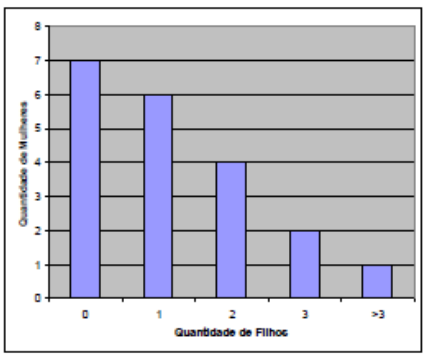
Gabarito: D.

29. (FGV – Assistente Técnico/CODEBA/2010) O gráfico de pizza apresentado abaixo fornece os resultados de uma pesquisa feita com 20 mulheres acerca da quantidade de filhos que cada uma possui.

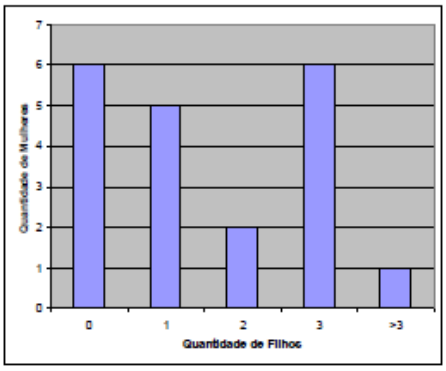


Assinale o gráfico de colunas que melhor representa os dados apresentados acima.

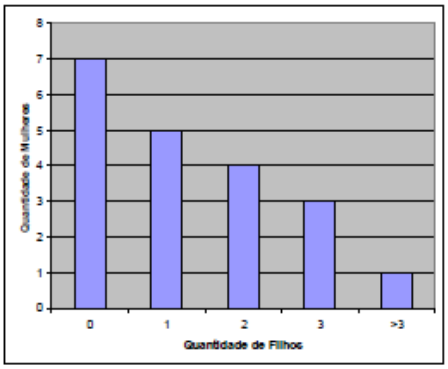
a)



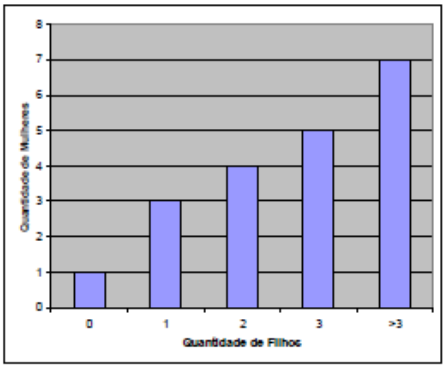
b)


Quantidade de Mulheres
0
1
2
3
4
5
6
7
0
1
2
3
>3
Quantidade de Filhos

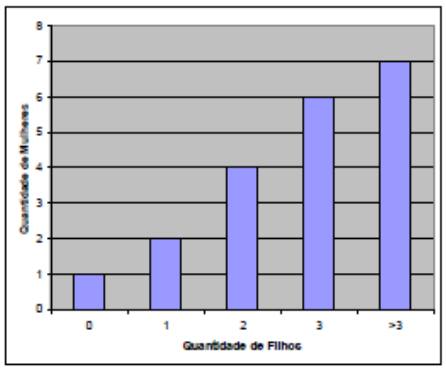
c)


Quantidade de Mulheres
0
1
2
3
4
5
6
7
8
0
1
2
3
>3
Quantidade de Filhos

d)


Quantidade de Mulheres
0
1
2
3
4
5
6
7
8
0
1
2
3
>3
Quantidade de Filhos

e)

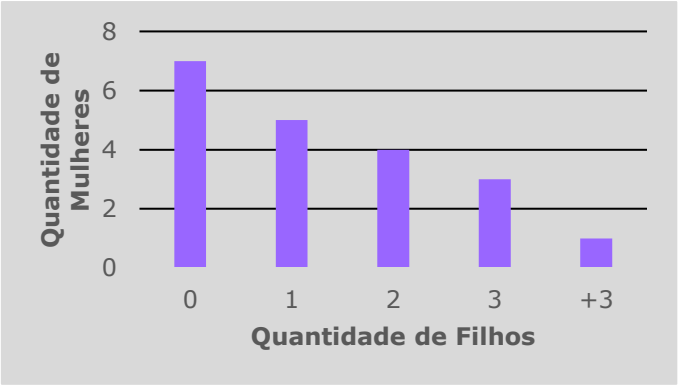


Comentários:

Sabemos que a pesquisa foi feita com **20 mulheres**. Daí, com base no gráfico de setores apresentado, temos condições de descobrir a quantidade de mulheres pertences a cada grupo:

- **Nenhum filho**: 35% de 20 = 0,35 . 20 = **7**
- **1 filho**: 25% de 20 = 0,25 . 20 = **5**
- **2 filhos**: 20% de 20 = 0,20 . 20 = **4**
- **3 filhos**: 15% de 20 = 0,15 . 20 = **3**
- **Mais de 3 filhos**: 5% de 20 = 0,05 . 20 = **1**

Em seguida, podemos construir o gráfico de colunas exigido:



Gabarito: C.

30. (Cesgranrio/Petrobras/Analista/2010) Histogramas e polígonos de frequências são duas representações gráficas de distribuições

a) uniformes.

b) de frequências.

c) de acumulações.

d) não uniformes.

e) assimétricas.

Comentários:

Vimos que **histogramas** e **polígonos de frequência** são formas gráficas de representarmos **dados em classes**.

Ao conjunto formado pelas classes, associadas às suas respectivas frequências, é comum darmos o nome de distribuição de frequências.

Assim, o histograma e o polígono de frequências representariam graficamente uma **distribuição de frequências**.

Gabarito: B.

LISTA DE QUESTÕES

1. (CESPE/ACE/TC-DF/2002) Por Estatística Descritiva entende-se um conjunto de ferramentas, tais como gráficos e tabelas, cujo objetivo é apresentar, de forma resumida, um conjunto de observações.

2. (CESPE - Agente Penitenciário Federal/2015) A diferença entre um censo e uma amostra consiste no fato de esta última exigir a realização de um número maior de entrevistas.

3. (CESPE/Fiscal de Tributos Estaduais/Sefaz-AL/2002) Como a realização de um censo tipicamente é muito onerosa e(ou) demorada, muitas vezes é conveniente estudar um subconjunto próprio da população, denominado amostra.

4. (CESPE/ Estatístico/SEDUC-AM/ 2011) O ramo e folhas é uma ferramenta exploratória que, além de servir para ordenar um pequeno conjunto de dados, permite destacar a forma da sua distribuição.

5. (CESPE/AUFC/TCU/2008)

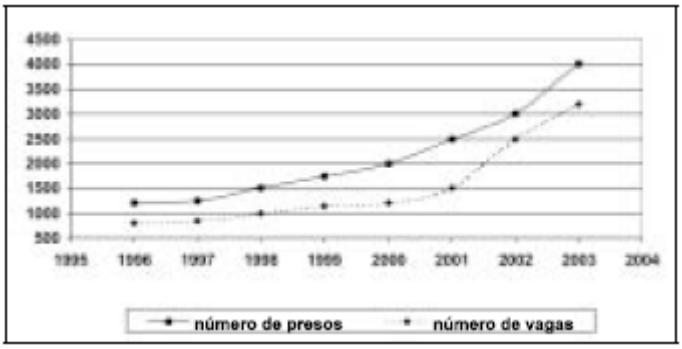
tipo	2006	2007
funcionários	70	74
estagiários	16	16
prestadores de serviço	14	10
total	**100**	**100**

Demonstrações contábeis da CAIXA – Exercício de 2007.
Internet: <www.caixa.gov.br> (com adaptações).

Considerando a tabela acima, que apresenta a distribuição do quadro de colaboradores da CAIXA, em mil pessoas, no final dos anos de 2006 e 2007, julgue os itens seguintes.

Se uma variável X registra a classificação de cada pessoa do quadro como funcionária, estagiária ou prestadora de serviço, então X é uma variável qualitativa.

6. (CESPE/PPF/PF/2004)



Nos últimos oito anos, a população carcerária em uma unidade da Federação cresceu de 1.200 presos (1996) para 4.000 presos (2003). Essa população carcerária é formada por presos nas casas penais, seccionais e delegacias. Por causa desse crescimento, foram construídas novas cadeias públicas, penitenciárias e novos blocos carcerários. Mesmo assim, não foi possível resolver o problema de superlotação. Em 1996, a capacidade total de lotação das casas penais, seccionais e delegacias era de apenas 800 vagas. Após a inauguração das novas instalações em 2003, o número de vagas aumentou para 3.200, o que resulta em déficit de 800 vagas. O gráfico acima apresenta a evolução temporal da população carcerária (linha contínua) e do número de vagas (linha pontilhada) de 1996 a 2003.

A capacidade total de lotação das casas penais, seccionais e delegacias (número de vagas) em 2000 é uma variável aleatória contínua.

7. (CESPE/TRE-ES/2011)

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

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

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

A variável "cargo" classifica-se como uma variável qualitativa ordinal.

8. (CESPE/TCU/AUFC/2008) Uma agência de desenvolvimento urbano divulgou os dados apresentados na tabela a seguir, acerca dos números de imóveis ofertados (X) e vendidos (Y) em determinado município, nos anos de 2005 a 2007.

ano	número de imóveis	
	ofertados (*X*)	vendidos (*Y*)
2005	1.500	100
2006	1.750	400
2007	2.000	700

Correio Braziliense, 29/4/2008, p. 17 (com adaptações).

A variável *X* forma uma série estatística denominada série temporal.

9. (CESPE/CEHAP-PB/Administrador/2009 - Adaptada) O custo médio nacional para a construção de habitação com padrão de acabamento normal, segundo levantamento realizado em novembro de 2008, foi de R$ 670,00 por metro quadrado, sendo R$ 400,00/m2 relativos às despesas com materiais de construção e R$ 270,00/m2 com mão-de-obra. Nessa mesma pesquisa, os custos médios regionais apontaram para os seguintes valores por metro quadrado: R$ 700,00 (Sudeste), R$ 660,00 (Sul), R$ 670,00 (Norte), R$ 640,00 (Centro-Oeste) e R$ 630,00 (Nordeste).

Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil. SINAPI/IBGE, nov./2008.

Considerando o par de variáveis (X,Y), em que X representa o custo médio por metro quadrado (em R$) e Y representa os atributos Sudeste, Norte, Sul, Centro-Oeste ou Nordeste, é correto afirmar que as variáveis X e Y formam uma série geográfica.

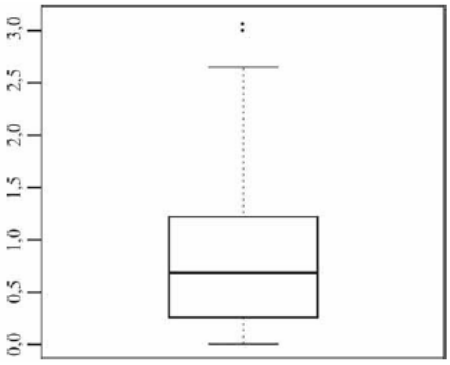
10. (CESPE/ Auditor de Controle Externo/TCE-PA/ 2016)

número diário de denúncias registradas (X)	frequência relativa
0	0,3
1	0,1
2	0,2
3	0,1
4	0,3
total	**1,0**

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

A variável X é do tipo qualitativo nominal.

11. (CESPE/ Auditor de Controle Externo/TCE-PA/ 2016)

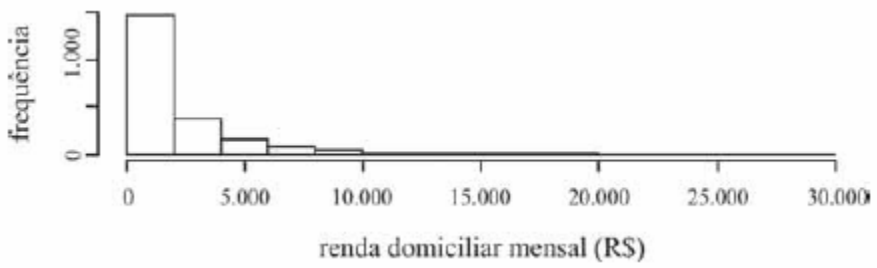


média amostral	0,80
desvio padrão amostral	0,70
primeiro quartil	0,25
mediana	0,70
terceiro quartil	1,20
mínimo	0
máximo	3,10

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.

X representa uma variável qualitativa ordinal.

12. (CESPE/Ana Sup/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.

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

13. (CESPE/TST/Ana Jud/2008) Considere que, em um ambiente de trabalho industrial, as seguintes medições acerca da poluição do ar tenham sido observadas:

1, 6, 4, 3, 2, 3, 1, 5, 1, 4.

A terceira e a oitava estatísticas de ordem são respectivamente iguais a 4 e 5.

14. (CESPE/SEFAZ-AL/2002) Em uma distribuição de frequências para um conjunto de n indivíduos, pode-se calcular as frequências relativas, dividindo-se cada frequência absoluta pela amplitude da classe ou intervalo.

15. (CESPE/BACEN/Analista/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.

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.

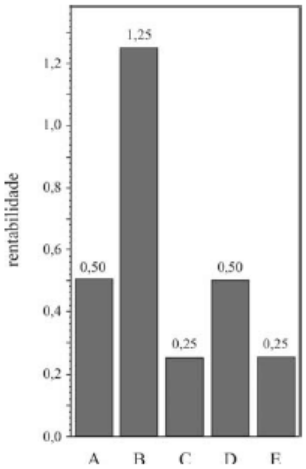
16. (CESPE - Analista/BACEN/2013)

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.

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

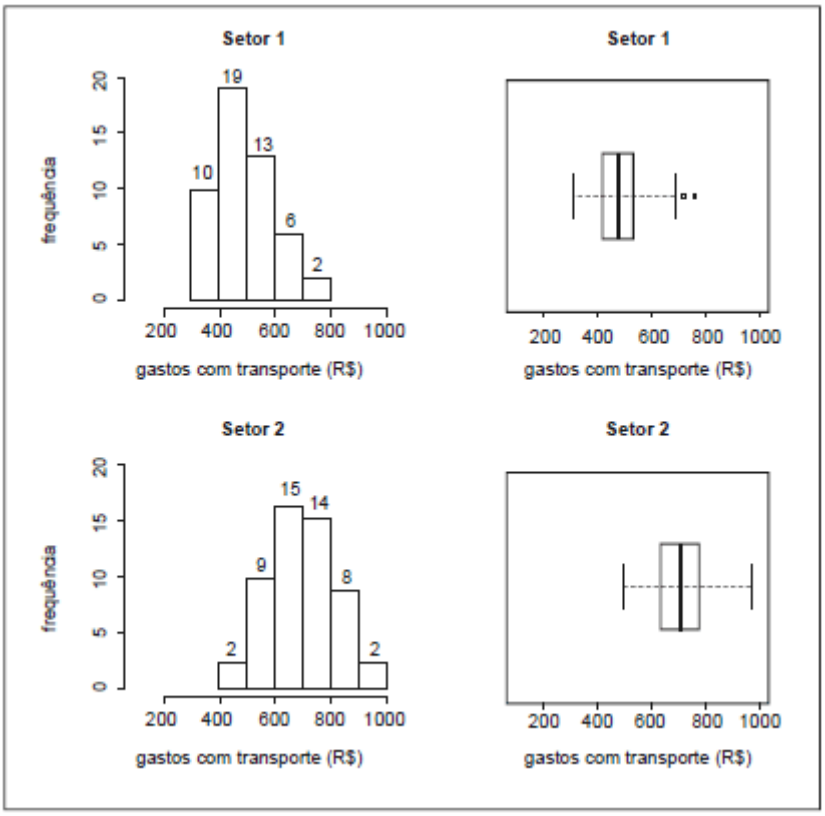
17. (CESPE/FUNPRESP/Analista de Investimentos/2016)



O gráfico ilustra cinco possibilidades de fundos de investimento com suas respectivas rentabilidades. Considerando que as probabilidades de investimento para os fundos A, B, C e D sejam, respectivamente, P(A) = 0,182; P(B) = 0,454; P(C)= 0,091; e P(D)= 0,182, julgue o item subsequente.

O gráfico apresentado é um histograma.

18. (CESPE – Analista/Câmara dos Deputados/2012)



Para avaliar os gastos com transporte de determinada diretoria, um analista coletou amostras de despesas com transportes (em R$) registradas por servidores dos setores1 e 2. Para cada setor, a amostra é constituída por 50 registros. Essas amostras foram organizadas graficamente, e os resultados são mostrados na figura acima. Nesta figura, as frequências absolutas estão indicadas nos histogramas correspondentes. Os dados foram os seguintes:

Setor 1

308,73	311,80	358,33	359,89	371,53	379,82
383,76	388,66	391,53	394,65	414,60	416,38
418,34	419,42	427,85	428,58	432,06	436,61
442,49	450,53	450,98	452,35	471,70	473,11
476,76	481,46	484,89	490,07	499,87	500,52
502,06	513,80	514,39	521,96	522,18	526,42
528,76	531,53	547,91	572,66	591,43	596,99
609,44	632,15	639,71	677,48	683,76	688,76
		723,79	767,53		

Setor 2

488,37	493,73	547,72	552,66	567,94	571,49
572,26	582,00	583,63	594,77	598,46	619,25
624,20	631,03	634,51	637,21	655,70	657,56

663,81	670,12	671,90	673,78	684,69	685,98
693,35	698,58	708,78	719,80	721,16	734,84
735,94	746,34	754,83	756,10	756,96	760,80
762,29	766,24	770,11	797,73	804,06	805,97
807,29	832,83	844,00	866,77	878,27	897,09
		943,10	963,25		

Na amostra do setor 2, a frequência relativa das despesas entre R$ 700,00 e R$ 799,99 é superior a 20%.

19. (FCC/ARCE/2006) O processo estatístico que consiste em avaliação direta de um parâmetro, utilizando-se todos os componentes da população chama-se:

a) Amostragem

b) Estimação

c) Censo

d) Parametrização

e) Correlação

20. (FCC/MPU/Analista/Estatística/2007) Uma empresa procurou estudar a ocorrência de acidentes com seus empregados e realizou um levantamento por um período de 36 meses. As informações apuradas estão na tabela a seguir:

Número de empregados acidentados	Número de meses
1	1
2	2
3	4
4	5
5	7
6	6
7	5
8	3
9	2
10	1

A porcentagem de meses em que houve menos de 5 empregados acidentados é

a) 50%

b) 45%

c) 35%

d) 33%

e) 30%

21. (FCC/TRT 6ª Região/Analista Judiciário/2012) A distribuição dos 500 preços unitários de um equipamento é representada por um histograma em que no eixo das abscissas constam os intervalos de classe e no eixo das ordenadas estão assinaladas as respectivas densidades de frequências, em (R$)−1. Define-se densidade de frequência de um intervalo de classe como sendo o resultado da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo. Um intervalo de classe no histograma apresenta uma amplitude de R$ 2,50 com uma densidade de frequência igual a 0,096. A quantidade de preços unitários referente a este intervalo é

a) 96

b) 120.

c) 144.

d) 150.

e) 192.

22. (FCC - ATARH/SERGAS/2013 - Adaptada) O gráfico de setores não é adequado para representar variáveis quantitativas.

23. (FCC - ATARH/SERGAS/2013 - Adaptada) Histograma é um gráfico que apresenta a distribuição de frequências de uma variável por meio de retângulos justapostos, feitos sobre as classes dessa variável, sendo que a área de cada retângulo é proporcional à frequência observada da correspondente classe.

24. (FCC – Analista Judiciário/TRT 1ª Região/2011) Um histograma representa a distribuição dos preços unitários de venda de um determinado equipamento no mercado. No eixo das ordenadas estão assinaladas as respectivas densidades de frequência para cada intervalo em (R$)−1. Define-se densidade de frequência de um intervalo de classe como sendo o quociente da divisão da respectiva frequência relativa pela correspondente amplitude do intervalo. Um intervalo de classe do histograma corresponde aos preços unitários maiores ou iguais a R$ 32,00 e inferiores a R$ 44,50 com uma densidade de frequência igual a 1,6 × 10−2 (R$)−1. Se todos os intervalos de classe do histograma têm a mesma frequência relativa, então um intervalo de classe com densidade de frequência igual a 5,0 × 10−3 (R$)−1 apresenta uma amplitude de

a) R$ 64,00

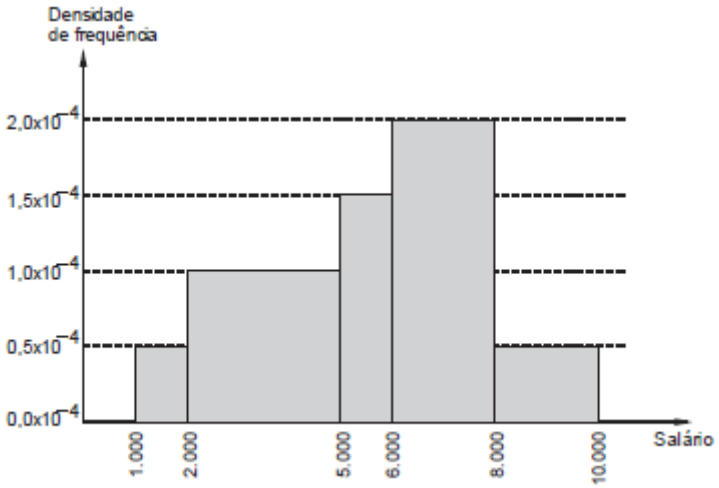
b) R$ 48,00

c) R$ 40,00

d) R$ 32,00

e) R$ 24,00

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



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

a) 104.

b) 152.

c) 136.

d) 144.

e) 120.

26. (FJG - FR/Pref RJ/2002) Os dados de um determinado estudo representam muitas variáveis para cada uma das pessoas que se submeteram ao estudo. Uma variável considerada qualitativa é a seguinte:

a) idade b) altura c) sexo d) peso

27. (VUNESP/Pref-SP/Biblioteconomia/2015 - Adaptada) Uma variável de natureza quantitativa pode ser classificada como ordinal ou nominal.

28. (CESGRANRIO/FINEP/Assistente/2014) A Tabela a seguir apresenta a frequência absoluta das faixas salariais mensais dos 20 funcionários de uma pequena empresa.

Faixa salarial (R$)	Frequência Absoluta
Menor que 1.000,00	6
Maior ou igual a 1.000,00 e menor que 2.000,00	7
Maior ou igual a 2.000,00 e menor que 3.000,00	5
Maior ou igual a 3.000,00	2
Total	20

A frequência relativa de funcionários que ganham mensalmente menos de R$ 2.000,00 é de

a) 0,07

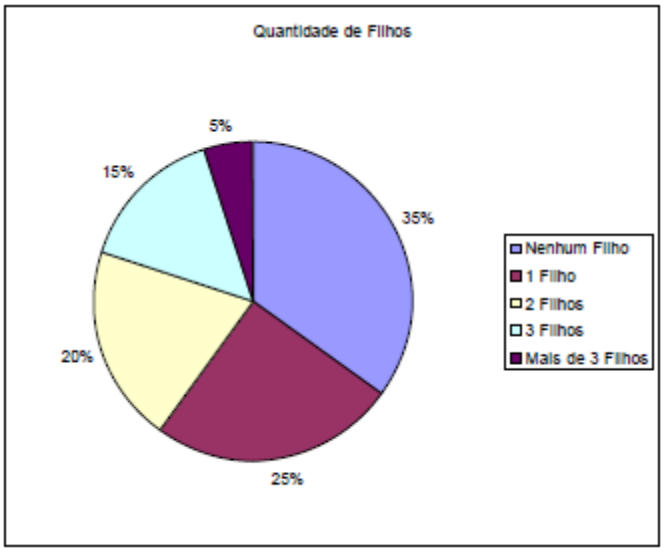
b) 0,13

c) 0,35

d) 0,65

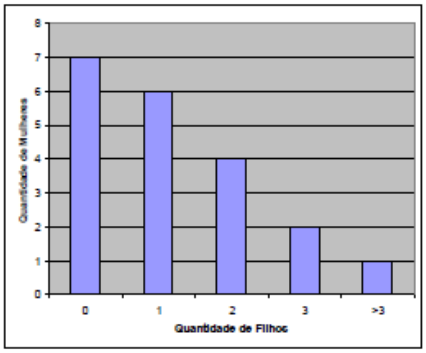
e) 0,70

29. (FGV – Assistente Técnico/CODEBA/2010) O gráfico de pizza apresentado abaixo fornece os resultados de uma pesquisa feita com 20 mulheres acerca da quantidade de filhos que cada uma possui.

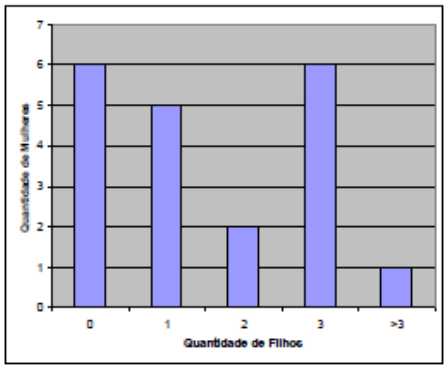


Assinale o gráfico de colunas que melhor representa os dados apresentados acima.

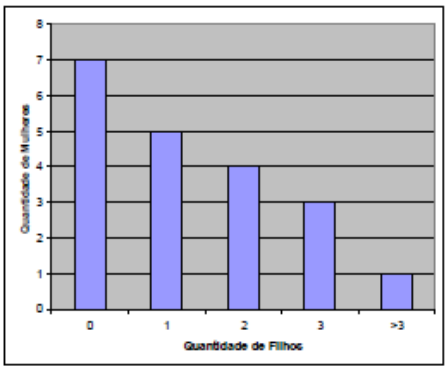
a)



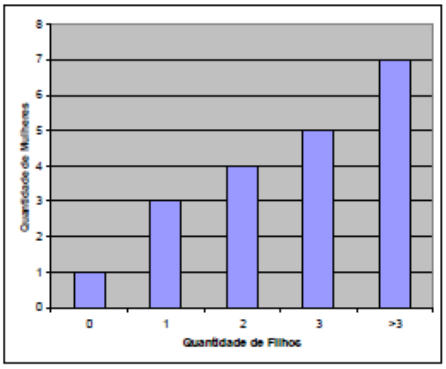
b)



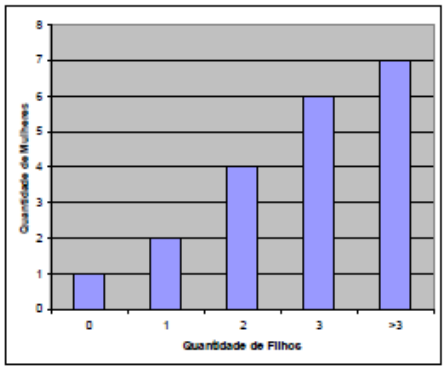
c)



d)



e)



30. (Cesgranrio/Petrobras/Analista/2010) Histogramas e polígonos de frequências são duas representações gráficas de distribuições

a) uniformes.

b) de frequências.

c) de acumulações.

d) não uniformes.

e) assimétricas.

GABARITO

1. CERTO
2. ERRADO
3. CERTO
4. CERTO
5. CERTO
6. ERRADO
7. CERTO
8. CERTO
9. CERTO
10. ERRADO
11. ERRADO
12. ERRADO
13. ERRADO
14. ERRADO
15. CERTO
16. ERRADO
17. ERRADO
18. CERTO
19. LETRA C
20. LETRA D
21. LETRA B
22. ERRADO
23. CERTO
24. LETRA C
25. LETRA C
26. LETRA C
27. ERRADO
28. LETRA D
29. LETRA C
30. LETRA B

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.

1.

Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.

2.

Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).

3.

Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).

4.

Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).

5.

Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).

6.

Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).

7.

Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.

8.

O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.

Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.