



PROFISSÃO
POLICIAL

Estatística

Professor Rodolfo Schmit

ESTATÍSTICA

Professor Rodolfo Schmit

Sumário

APRESENTAÇÃO	3
1 CONCEITO DE ESTATÍSTICA	9
2 RAMOS DA ESTATÍSTICA.....	10
2.1 ESTATÍSTICA DESCRITIVA.....	11
2.2 ESTATÍSTICA INFERENCIAL.....	13
3 OBJETO DE ESTUDO.....	14
3.1 ELEMENTOS AVALIADOS: POPULAÇÃO	15
3.1.1 <i>População x Amostra</i>	16
3.1.2 <i>Censo x Amostragem</i>	17
3.1.3 <i>Parâmetro x Estimativa</i>	18
3.2 CARACTERÍSTICA AVALIADA: VARIÁVEL.....	20
3.2.1 <i>Variáveis Qualitativas</i>	21
3.2.1.1 Variáveis Qualitativas Nominais	21
3.2.1.2 Variáveis Qualitativas Ordinais.....	0
3.2.2 <i>Variáveis Quantitativas</i>	21
3.2.2.1 Variáveis Quantitativas Discretas	21
3.2.2.2 Variáveis Quantitativas Contínuas	1
4 CONTEXTUALIZAÇÃO DE QUESTÕES	24
5 APRESENTAÇÃO DE DADOS	33
5.1 DADOS BRUTOS	34
5.2 DADOS PONDERADOS (TABELA DE FREQUÊNCIA POR VALOR).....	36
5.3 DADOS AGRUPADOS (TABELA DE FREQUÊNCIA COM INTERVALOS)	39
5.4 GRÁFICOS DE FREQUÊNCIA	42
5.4.1 <i>Colunas de Frequência</i>	42
5.4.2 <i>Histograma</i>	43

5.4.3	Curva de frequência	45
5.4.4	Ogiva.....	46
6	TABELAS	47
6.1	TABELA SIMPLES DE ASSOCIAÇÃO DE VARIÁVEIS.....	47
6.2	TABELA DE DUPLA ENTRADA	48
6.3	SÉRIES ESTATÍSTICAS	49
6.3.1	Séries Históricas	50
6.3.2	Séries Territoriais	50
6.3.3	Séries Categóricas.....	51
6.3.4	Séries Mistas	52
7	GRÁFICOS	53
7.1	DIAGRAMA DE PONTOS.....	54
7.2	DIAGRAMA DE RAMOS E FOLHAS	55
7.3	GRÁFICO DE COLUNAS	56
7.4	GRÁFICO DE BARRAS.....	57
7.5	GRÁFICO DE COLUNAS MÚLTIPLAS	59
7.6	GRÁFICO DE SETORES (PIZZA)	58
7.7	GRÁFICO DE DISPERSÃO	59
7.8	GRÁFICO DE LINHAS (POLÍGONOS)	61
	QUESTÕES DE RENDIMENTO.....	63

APRESENTAÇÃO



Fala, meu caro aluno! Aqui é o Professor Rodolfo Schmit. Com grande satisfação e honra aceitei o desafio de lançar um material de Estatística focado para carreiras policiais.

Inicialmente, irei me apresentar! Tenho 29 anos, sou natural de Bombinhas, Santa Catarina. Sou formado em Agronomia e me aprofundi na área de Estatística quando desenvolvi pesquisas na área Genética e Melhoramento Vegetal. Desenvolvi muitos trabalhos científicos que exigiam muito conhecimento e domínio da Estatística, e acabei desenvolvendo gosto na didática para ensinar essa disciplina.

Após essa fase acadêmica, percebi que precisava de mais ação na minha vida. Foi aí que me dediquei aos estudos para concurso da Polícia Federal e fui aprovado em 2021. Hoje, sou Agente de Polícia Federal e estou completamente realizado com a minha profissão.

Durante minha jornada de estudos para ingressar na PF, acabei me tornando professor de Estatística em cursos preparatórios para concurso público. Foi uma sinergia fantástica, estudar e passar meu conhecimento ajudando outras pessoas a alcançar a aprovação também. Analisei muito o estilo e o padrão das questões de Estatísticas e como elas são elaborados para os certames. E meu objetivo é apresentar para você toda essência de como essa disciplina é cobrada nos concursos.

É fundamental salientar que este curso será elaborado com base nos últimos editais lançados para as carreiras policiais, sendo destinado aos alunos que querem se preparar com antecedência. Iremos passar por todo conteúdo programático do último certame e analisaremos centenas questões de concursos passados. Elaborei o curso para qualquer tipo de candidato: os que têm base e para os que não sabem absolutamente nada da matéria!

Comentarei todos os pontos que costumam cair, bem como a tendência para futuros certames. Vamos, portanto, estudar com bastante vontade, perseverança, alegria e entusiasmo esta disciplina tão interessante chamada de Estatística, “a ciência da investigação!”

O curso será composto por 80 videoaulas + materiais em PDF. Enfrentaremos todos os assuntos de Estatística que são cobrados em carreiras policiais. A referência base desse curso serão os últimos editais da Polícia Federal (PF) e da Polícia Civil do Distrito Federal (PCDF), pois são aqueles que cobram um maior conteúdo programático de Estatística. Veja detalhadamente:

AULA	NOME DA AULA	CONTEÚDO NO EDITAL
1	Introdução a Estatística: Conceitos Básicos	Introdução a disciplina
2	Introdução a Estatística: Interpretação de Questões	Introdução a disciplina
3	Estatística Descritiva: Apresentação de Dados	1 Estatística descritiva e análise exploratória de dados: gráficos, diagramas, tabelas
4	Estatística Descritiva: Tabelas e Gráficos	1 Estatística descritiva e análise exploratória de dados: gráficos, diagramas, tabelas
5	Estatística Descritiva: Exercícios - Dados, Tabelas e Gráficos	1 Estatística descritiva e análise exploratória de dados: gráficos, diagramas, tabelas
6	Estatística Descritiva: Média	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (posição)
7	Estatística Descritiva: Mediana	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (posição)
8	Estatística Descritiva: Moda	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (posição)
9	Estatística Descritiva: Exercícios - Medidas Tendência Central	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (posição)
10	Estatística Descritiva: Separatrizes	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (posição)
11	Estatística Descritiva: Dispersão de Dados	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (dispersão)
12	Estatística Descritiva: Variância, Desvio Padrão	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (dispersão)
13	Estatística Descritiva: Exercícios - Medidas Separatrizes e Dispersão	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (dispersão)
14	Estatística Descritiva: Coeficientes de Variação e Momentos	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (dispersão)
15	Estatística Descritiva: Medidas de Forma	1 Estatística descritiva e análise exploratória de dados: medidas descritivas (assimetria e curtose)

16	Estatística Descritiva: Transformação de Dados	1 Estatística descritiva e análise exploratória de dados: medidas descritivas
17	Estatística Descritiva: Exercícios - CV, Forma e Transformação de dados	1 Estatística descritiva e análise exploratória de dados: medidas descritivas
18	Teoria da Probabilidade: Conceitos e Axiomas	2 Probabilidade. 2.1 Definições básicas e axiomas.
19	Teoria da Probabilidade: Interação de Eventos Probabilísticos	2.2 Probabilidade condicional e independência
20	Teoria da Probabilidade: Probabilidade Condicional	2.2 Probabilidade condicional e independência
21	Teoria da Probabilidade: Exercícios - Probabilidade	2 Probabilidade. 2.1 Definições básicas e axiomas. 2.2 Probabilidade condicional e independência
22	Variáveis Aleatórias Discretas: Conceito e Funções de Probabilidade	2.3 Variáveis aleatórias discretas 2.4 Distribuição de probabilidades. 2.5 Função de probabilidade.
23	Variáveis Aleatórias Discretas: Medidas Descritivas	2.3 Variáveis aleatórias discretas 2.7 Esperança e momentos.
24	Variáveis Aleatórias Discretas: Distribuição Bernoulli e Binomial	2.3 Variáveis aleatórias discretas 2.8 Distribuições especiais.
25	Variáveis Aleatórias Discretas: Exercícios - Valor Esperado, Variância, Bernoulli e Binomial	2.3 Variáveis aleatórias discretas 2.4 Distribuição de probabilidades. 2.5 Função de probabilidade. 2.7 Esperança e momentos. 2.8 Distribuições especiais.
26	Variáveis Aleatórias Discretas: Distribuição Poisson	2.3 Variáveis aleatórias discretas 2.8 Distribuições especiais.
27	Variáveis Aleatórias Discretas: Distribuição Geométrica, Binomial Negativa	2.3 Variáveis aleatórias discretas 2.8 Distribuições especiais.
28	Variáveis Aleatórias Discretas: Distribuição Hipergeométrica	2.3 Variáveis aleatórias discretas 2.8 Distribuições especiais.
29	Variáveis Aleatórias Discretas: Exercícios - Poisson, Geométrica e Hipergeométrica	2.3 Variáveis aleatórias discretas 2.8 Distribuições especiais.
30	Variáveis Aleatórias Contínuas: Função Densidade de Probabilidade	2.3 Variáveis aleatórias contínuas 2.4 Distribuição de probabilidades 2.6 Função densidade de probabilidade. 2.7 Esperança e momentos.
31	Variáveis Aleatórias Contínuas: Medidas Descritivas - Sem cálculo de Integral	2.3 Variáveis aleatórias contínuas 2.4 Distribuição de probabilidades 2.6 Função densidade de probabilidade. 2.7 Esperança e momentos.
32	Variáveis Aleatórias Contínuas: Medidas Descritivas - Com cálculo de Integral	2.3 Variáveis aleatórias contínuas 2.4 Distribuição de probabilidades 2.6 Função densidade de probabilidade. 2.7 Esperança e momentos.
33	Variáveis Aleatórias Contínuas: Exercícios - Função Densidade e Medida Descritivas	2.3 Variáveis aleatórias contínuas 2.4 Distribuição de probabilidades 2.6 Função

		densidade de probabilidade. 2.7 Esperança e momentos.
34	Variáveis Aleatórias Contínuas: Distribuição Uniforme e Exponencial	2.3 Variáveis aleatórias contínuas 2.8 Distribuições especiais.
35	Variáveis Aleatórias Contínuas: Distribuição Normal - Conceitos	2.3 Variáveis aleatórias contínuas 2.8 Distribuições especiais.
36	Variáveis Aleatórias Contínuas: Distribuição Normal - Cálculos	2.3 Variáveis aleatórias contínuas 2.8 Distribuições especiais.
37	Variáveis Aleatórias Contínuas: Exercícios - Uniforme, Exponencial e Normal	2.3 Variáveis aleatórias contínuas 2.8 Distribuições especiais.
38	Variáveis Aleatórias Bidimensionais: Transformação de Variáveis	2.10 Transformação de variáveis.
39	Variáveis Aleatórias Bidimensionais: Discreta x Discreta - Independentes	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
40	Variáveis Aleatórias Bidimensionais: Discreta x Discreta - Condicional	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
41	Variáveis Aleatórias Bidimensionais: Exercícios - Transformação e Variáveis Bidimensionais Discretas	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência 2.10 Transformação de variáveis.
42	Variáveis Aleatórias Bidimensionais: Contínua x Contínua - Independentes	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
43	Variáveis Aleatórias Bidimensionais: Contínua x Contínua - Condicional	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
44	Variáveis Aleatórias Bidimensionais: Mista (Contínua x Discreta)	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
45	Variáveis Aleatórias Bidimensionais: Exercícios - Variáveis Bidimensionais Contínuas e mistas	2.3 Variáveis aleatórias discretas e contínuas 2.9 Distribuições condicionais e independência
46	Amostragem: Técnicas de Amostragem	5 Técnicas de amostragem: amostragem aleatória simples, estratificada, sistemática e por conglomerados.
47	Amostragem: Medidas descritivas em Amostras	5 Técnicas de amostragem: amostragem aleatória simples, estratificada, sistemática e por conglomerados.
48	Amostragem: Exercícios - Amostragem	5 Técnicas de amostragem: amostragem aleatória simples, estratificada, sistemática e por conglomerados.
49	Estatística Inferencial: Distribuição Amostral	2.13 Amostras aleatórias. 2.14 Distribuições amostrais.

50	Estatística Inferencial: Estimadores e Propriedades	3 Inferência estatística. 3.1 Estimação pontual: métodos de estimação, propriedades dos estimadores, suficiência.
51	Estatística Inferencial: Teorema Limite Central e Lei Grandes Números	2.11 Leis dos grandes números. 2.12 Teorema central do limite
52	Estatística Inferencial: Exercícios - Distribuição Amostral, Estimadores, TLC e LGN	3 Inferência estatística
53	Estatística Inferencial: Intervalo de Confiança	3.2 Estimação intervalar: intervalos de confiança
54	Estatística Inferencial: Erro Amostral e Intervalo de Credibilidade	3.2 Estimação intervalar: intervalos de confiança, intervalos de credibilidade
55	Estatística Inferencial: Tamanho Amostral	5.1 Tamanho amostral.
56	Estatística Inferencial: Exercícios - Intervalo de Confiança, Erro e Tamanho Amostral	3.2 Estimação intervalar: intervalos de confiança, intervalos de credibilidade 5.1 Tamanho amostral.
57	Estatística Inferencial: Formulação de Hipóteses	3.3 Testes de hipóteses: hipóteses simples e compostas
58	Estatística Inferencial: Teste de Hipótese para Média	3.3 Testes de hipóteses: teste t de Student, teste Z
59	Estatística Inferencial: Teste de Hipótese para Proporção	3.3 Testes de hipóteses: teste Z
60	Estatística Inferencial: Exercícios - Teste de Hipóteses	3.3 Testes de hipóteses
61	Estatística Inferencial: Valor-P e Nível de Significância	3.3 Testes de hipóteses: níveis de significância
62	Estatística Inferencial: Tipos de Erro	3.3 Testes de hipóteses: níveis de significância e potência de um teste
63	Estatística Inferencial: Potência do teste	3.3 Testes de hipóteses: níveis de significância e potência de um teste
64	Estatística Inferencial: Exercícios - Valor-p, tipos de erro e potência do teste	3.3 Testes de hipóteses: níveis de significância e potência de um teste
65	Estatística Inferencial: Teste Qui-quadrado Ajustamento	3.3 Testes de hipóteses: teste qui-quadrado.
66	Estatística Inferencial: Teste Qui-quadrado Independência	3.3 Testes de hipóteses: teste qui-quadrado.
67	Estatística Inferencial: Teste Qui-quadrado Homogeneidade	3.3 Testes de hipóteses: teste qui-quadrado.
68	Estatística Inferencial: Exercícios - Teste qui-quadrado	3.3 Testes de hipóteses: teste qui-quadrado.
69	Regressão Linear: Covariância	4 Análise de regressão linear
70	Regressão Linear: Correlação Linear	4 Análise de regressão linear
71	Regressão Linear: Teste do Coeficiente de Correlação	4.1 Critérios de mínimos quadrados e de máxima verossimilhança. 4.2 Modelos de regressão linear.
72	Regressão Linear: Exercícios - Covariância, Correlação	4 Análise de regressão linear. 4.1 Critérios de mínimos quadrados e de máxima verossimilhança. 4.2 Modelos de regressão linear.

73	Regressão Linear: Modelo de Regressão Linear Simples	4.1 Critérios de mínimos quadrados e de máxima verossimilhança. 4.2 Modelos de regressão linear.
74	Regressão Linear: Modelo de Regressão Linear Múltipla	4.3 Inferência sobre os parâmetros do modelo.
75	Regressão Linear: Inferência sobre os Parâmetros do modelo	4.3 Inferência sobre os parâmetros do modelo.
76	Regressão Linear: Exercícios - Modelo de regressão linear	4.1 Critérios de mínimos quadrados e de máxima verossimilhança. 4.2 Modelos de regressão linear. 4.3 Inferência sobre os parâmetros do modelo.
77	Regressão Linear: Análise de Variância para modelo simples	4.4 Análise de variância
78	Regressão Linear: Análise de Variância para modelo múltiplo	4.4 Análise de variância
79	Regressão Linear: Análise de Resíduos	4.5 Análise de resíduos
80	Regressão Linear: Exercícios - Análise de variância e Resíduos	4.4 Análise de variância 4.5 Análise de resíduos

Veja que, no decorrer das videoaulas, teremos uma pausa do conteúdo teórico e iremos realizar resoluções de exercícios para maximizar o seu conhecimento e depois seguiremos avançando o conteúdo programático. Afinal, a melhor forma de aprender estatística é aplicando-a na prática, com muitos exercícios!

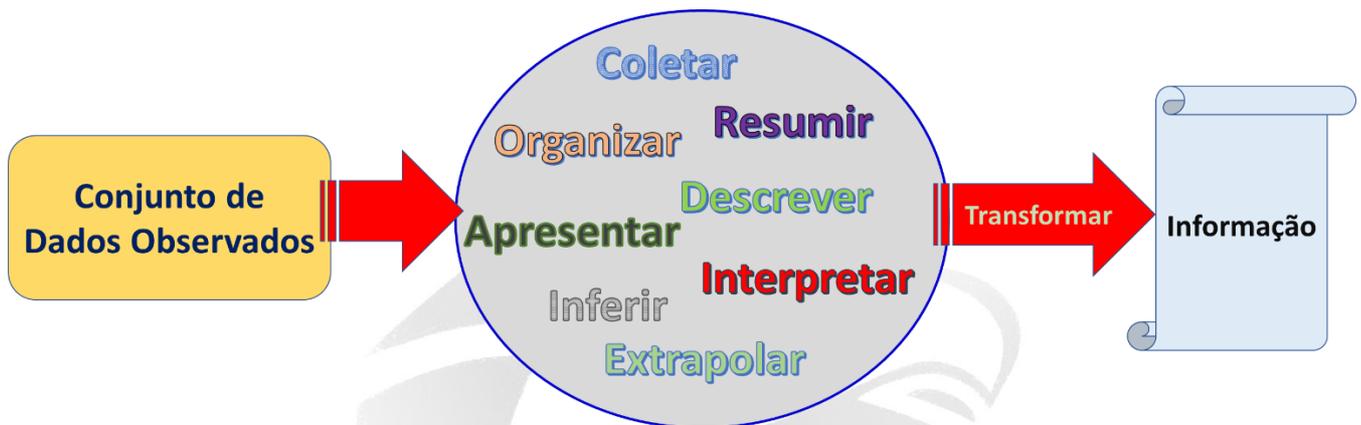
Destaco que a maioria das questões das últimas provas cobraram diversos assuntos do edital, com questões de vários níveis de dificuldade. Por isso, vamos enfrentar esse desafio juntos e da forma mais dinâmica possível. Tenho a certeza de que após a realização deste curso você estará apto para fazer uma ótima prova.

Caso você tenha algumas dúvidas no decorrer dos seus estudos, podemos conversar também pelas minhas redes sociais: [@rodolfooschmit09](https://www.instagram.com/rodolfooschmit09).

INTRODUÇÃO À ESTATÍSTICA

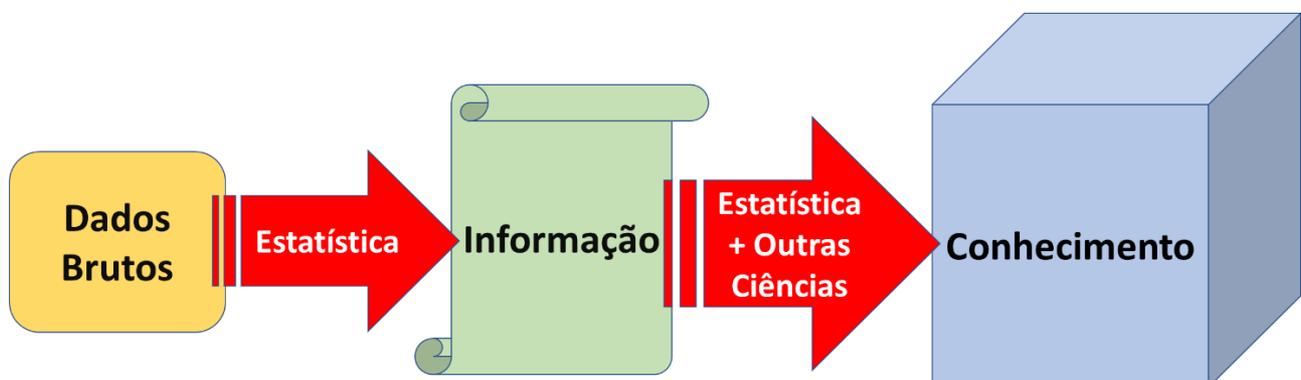
1 CONCEITO DE ESTATÍSTICA

A Estatística é uma ciência que utiliza um conjunto de metodologias especialmente apropriadas com a finalidade de gerar uma informação.



O principal objetivo da Estatística é a compreensão de uma realidade/fenômeno, a partir do que é observado no mundo real. Dessa forma, a utilização apropriada da Estatística extrai informações úteis para **investigação** e influencia diretamente na **tomada de decisões** sobre qualquer coisa do interesse humano.

Conceitualmente, **o dado bruto** é qualquer caractere com o potencial de se transformar em informação. De modo geral, a Estatística é uma ciência especializada em **transformar dados brutos em informação**. Isso se aplica para qualquer área de estudo. Em adição, combinado com outras ciências, é capaz de gerar **um conhecimento**, evoluindo para produção de inteligência e de novas tecnologias.



A Estatística aplica a matemática para fazer a transformação dos dados brutos em conhecimento. Todavia, por mais que a doutrina classifique a Estatística como um ramo da matemática, é muito importante que o aluno compreenda, primeiramente, a essência da disciplina e veja que todas as deduções e cálculos matemáticos são apenas instrumentos para compreender um fenômeno que ocorre no mundo real, de interesse do ser humano.

Exemplo Prático:

*O ser humano tem o interesse compreender sobre **a estatura das pessoas em uma região...***

*Para isso, o ser humano irá coletar uma série de **dados brutos** sobre a altura das pessoas naquela região, e vai aplicar as ferramentas da estatística para gerar uma **informação sobre esse fenômeno.***

2 RAMOS DA ESTATÍSTICA

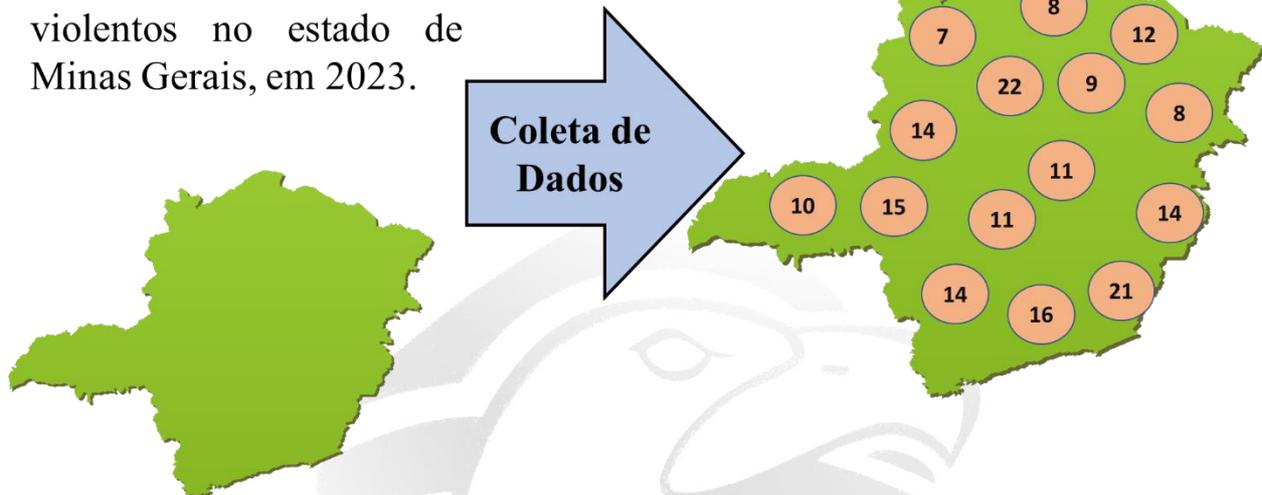
Para iniciar os estudos, é interessante compreender alguns conceitos e classificações que serão utilizados durante todo o raciocínio desenvolvido pela disciplina. Deve-se entender que a Estatística é dividida em dois grandes ramos conforme o objetivo das técnicas utilizadas:



Para explicar sobre esses dois ramos da Estatística, vamos além de conceitos exemplificar sua aplicação e começar se habituar com esquemas da Estatística. Veja a Ilustração a seguir que representará um fenômeno a ser estudado estatisticamente:

Fenômeno em estudo:

Ocorrências de homicídios violentos no estado de Minas Gerais, em 2023.



Dados Brutos

2.1 Estatística descritiva

A Estatística Descritiva consiste em análises que resumem concisamente um conjunto de dados, sintetizam em informações mais simples, organizam em tabelas, e ilustram por meio de gráficos.

**Estatística
Descritiva**

Resumir Conj. Dados

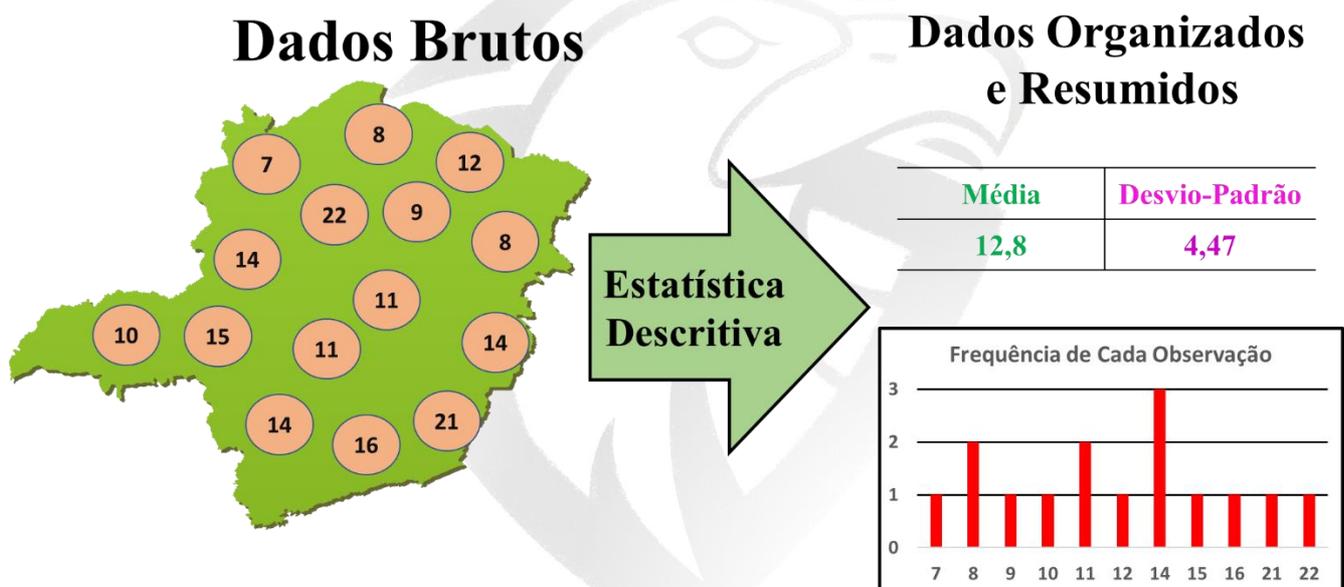
Organizar em Tabelas

Ilustrar em Gráficos

Esse campo da Estatística também é aplicado para efetuar **análises exploratórias** sobre um conjunto de dados, pois é possível compreender uma tendência geral dos dados observados. Com uso de medidas descritivas, gráficos e tabelas, pode ser observado um comportamento padrão, intuitivo e representativo do fenômeno em estudo.

Por exemplo, com o desempenho médio de uma determinada característica obtida a partir de um conjunto de dados brutos, é possível levantar hipóteses sobre o real desempenho desse objeto de estudo, ou até mesmo, se sua performance pode ser superior a um outro fenômeno estudado de interesse.

Um conjunto de dados brutos pode ser descritos a partir de medidas como média e desvio-padrão, ou organizados em tabela de frequência, e ilustrado em gráfico. Veja demonstração a seguir:



As informações geradas pela Estatística Descritiva, no exemplo acima, resumem o valor da tendência central dos dados pela média, fornece uma ideia do quanto os dados se dispersam pelo desvio-padrão, como também, apresenta a distribuição dos dados pelo gráfico de barras que apresenta a frequência.

A Estatística Descritiva somente é completa e autossuficiente quando é possível obter dados, com exatidão, de todos os elementos que compõem seu objeto de estudo (população). Nessa situação, apenas descrever o conjunto de dados já bastaria para obter conclusões sobre o fenômeno em estudo.

No entanto, a Estatística Descritiva, a partir de uma amostra (uma parte do conjunto de dados), apenas descreve os dados dessa amostra e **não permite inferir** sobre toda a população. Seria uma afirmação equivocada comparar duas médias estimadas simplesmente pela magnitude do valor. Para isso, é necessário o uso de métodos da Estatística Inferencial.

2.2 Estatística Inferencial

A Estatística Inferencial utiliza de um conjunto de dados incompletos (amostra) para generalizar/extrapolar informações sobre o fenômeno em estudo, e, com isso, tirar conclusões satisfatórias e tomar decisões. Além disso, a Estatística Inferencial quantifica o erro dessas inferências.

**Estatística
Inferencial**

**Generalizar/Extrapolar
informações**

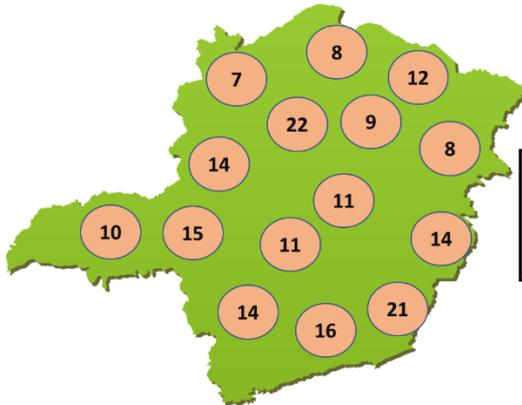
Tomar decisões

Quantificar o Erro

As principais técnicas empregadas na Estatística Inferencial são a estimação pontual, a estimação intervalar e o teste de hipóteses. O uso da Estatística Inferencial é valioso quando não é conveniente ou possível examinar todos os elementos de uma população (que costuma a realidade de quase de todos objeto de estudo na Estatística).

Quando se obtêm informações a partir de conjunto de dados incompletos (uma amostra), é necessário entender que essas informações foram obtidas por um subconjunto da população e que possuem uma **probabilidade de corresponder coerentemente ao fenômeno estudado**. Por isso, essas informações estão sujeitas a erros e imprecisão. Desse modo, cabe à Estatística Inferencial quantificar essas incertezas e avaliar a confiança da informação gerada. Exemplo:

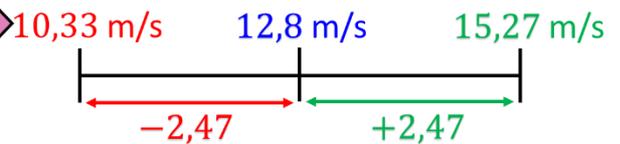
Dados Brutos



Estatística Inferencial

Estimativa dos Dados

Estimativa da Média	Estimativa do Erro
12,8	±2,47



Nível de Confiança: 95%

O valor da média estimada por um conjunto de dados de uma amostra é composto por um possível erro de estimativa. Baseado nisso, calcula-se um intervalo de confiança em que o valor estimado pode oscilar, isto é, a média é 12,8, mas pode oscilar entre 10,33 até 15,27. De forma bem genérica, essa é a ideia da Estatística Inferencial que trabalha com informações incompletas e usa matemática aplicada para generalizar a informação obtida.

3 OBJETO DE ESTUDO

O objeto de estudo na disciplina de Estatística compreende a um fenômeno do mundo real, uma realidade específica. Em outras palavras, tudo aquilo que ocorre ao redor do homem e que ele tem interesse de compreender suas características, seus comportamentos, e seus desempenhos.

Desse modo, o objeto de estudo pode ser: fenômenos biológicos e naturais; comportamentos sociais; aspectos políticos; avaliação de equipamentos; entre outros. Praticamente, o objeto de estudo da disciplina é um **“coringa”**, ou seja, a Estatística pode ser aplicada em qualquer fenômeno/realidade em que é possível extrair dados.

O fenômeno estudado em cada análise estatística é caracterizado basicamente por dois componentes essenciais: os elementos que serão estudados (População) e a característica que será observada (Variável). Segue a ilustração:

Objeto de Estudo

De quem será coletado dados?

População: elementos avaliados

O que será observado?

Variável: característica observada

Dessa forma, em toda questão de Estatística, o aluno deve primeiramente analisar o cenário apresentado em cada questão. Com isso, deve ser identificado quem são os elementos avaliados ao qual pretende-se fazer inferências e o que será avaliado em cada elemento (qual característica, atributo será observado, contabilizado ou mensurado). Essa é uma leitura essencial para **iniciar a resolução de uma questão de Estatística**.

Dica Essencial:

Compreender e delimitar o objeto de estudo é ponto mais importante para dar início a uma questão de Estatística!

3.1 Elementos avaliados: população

Sempre que a Estatística é aplicada para estudar um fenômeno qualquer, é preciso saber quem são os elementos que compõem o objeto de estudo, em cada situação particular. De forma prática, os elementos avaliados podem ser compreendidos ao responder as seguintes perguntas:

- “De quem” ou “de que” coletamos dados para obter informações?
- “Sobre quem” ou “Sobre que” queremos gerar uma informação?

A população corresponde ao conjunto total de elementos do fenômeno estudado, enquanto o termo elemento corresponde a **unidades de avaliação (ou**

unidades de observação) “de quem” ou “de que” serão extraídos dados. Segue com exemplos práticos:

População	Elementos
População humana do Brasil	Cada pessoa brasileira
População de onças-pintadas do Pantanal	Cada onça-pintada do Pantanal
Armamento do Exército Brasileiro	Cada arma vinculada ao Exército Brasileiro
Automóveis da empresa Ford	Cada veículo produzido pela empresa Ford
Criminalidade em 2023	Cada crime ocorrido em 2023

Veja que os elementos podem ser compostos por indivíduos (pessoas, animais), objetos (armas, veículos, equipamentos etc.) ou até mesmo eventos (crimes, jogos etc.), tudo dependerá do objeto de estudo de interesse.

A população deve ser criteriosamente especificada e delimitada. Todas as inferências obtidas com uso de técnicas estatísticas serão válidas apenas para essa população, portanto a tomada de decisão não pode gerar conclusões incoerentes a população de estudo.

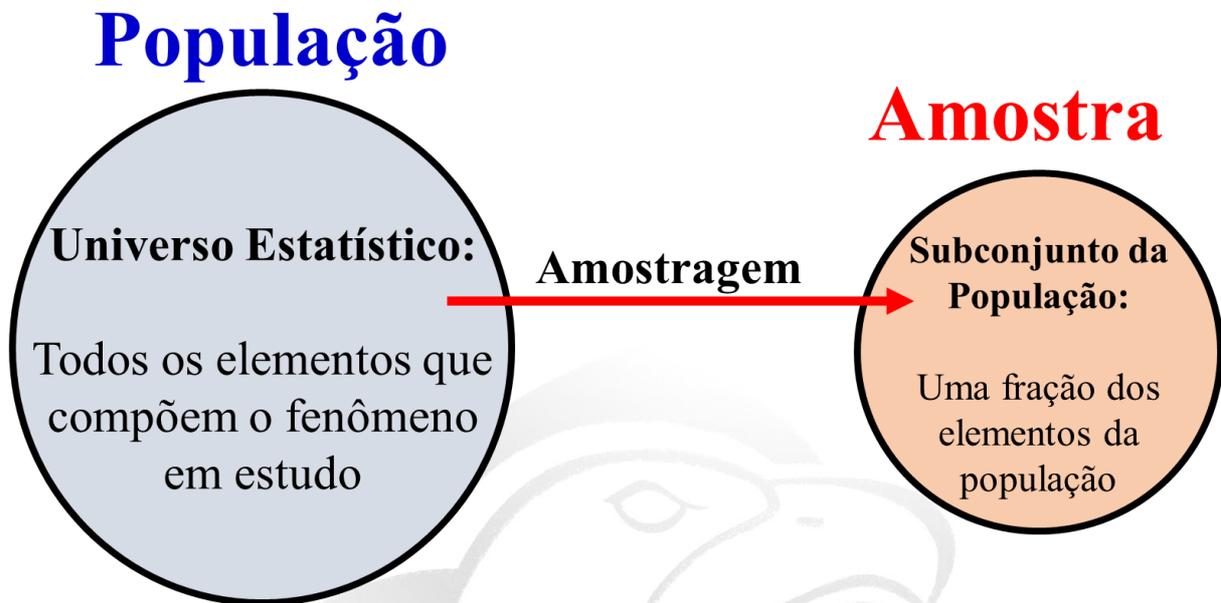
3.1.1 População x Amostra

Como já vimos, o conjunto de todos os elementos estudados representa o nosso **universo estatístico**, também denominado de **população**.

Então, população corresponde ao conjunto total de elementos sobre os quais se desejam obter dados de características que são comuns a todos. Portanto, população é todo o universo de um objeto de estudo de interesse.

Entretanto, acontece que muitas vezes a população torna-se infinitamente grande e fica muito oneroso, impraticável, inviável de se avaliar todos os elementos. Nesse momento, uma alternativa para estudar um fenômeno é avaliar parte dessa

população e utilizar métodos, como a Estatística Inferencial, para generalizar as informações dessa parte para toda população. Veja a ilustração a seguir:



Conforme apresentado na ilustração, a fração dos elementos observados que é retirada de uma população corresponde a **amostra**.

Desse modo, a amostra é um subconjunto de elementos da população que são coletados e analisados com objetivo de representar fidedignamente toda a população de estudo.

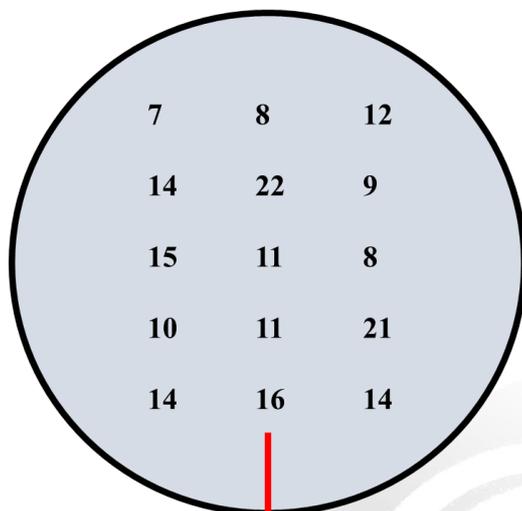
3.1.2 Censo x Amostragem

A análise estatística mais valiosa será sempre obter dados sobre toda a população, o que denominamos de **censo**. *Dessa forma, denomina-se censo a coleta exaustiva de dados referente a todos os elementos que compõem a população, ou seja, quando se trata de **informações completas** da população em estudo.*

No entanto, sabe-se que, muitas vezes, não será possível efetuar um censo. Em contraponto, trabalha-se com o conjunto de dados proveniente da técnica de amostragem. *Assim, a amostragem consiste na coleta de um **conjunto de dados incompletos** de uma população.*

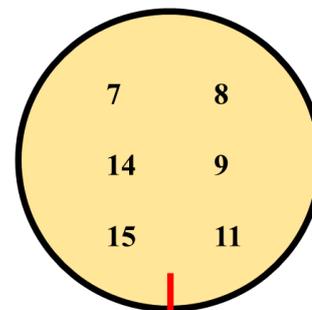
Essa técnica deve seguir um método criterioso e adequado para que os dados extraídos representem o fenômeno de estudo presente na população. Mais à frente, teremos um tópico somente estudando sobre a amostragem e suas técnicas.

População



Censo

Amostra



Amostragem

3.1.3 Parâmetro x Estimativa

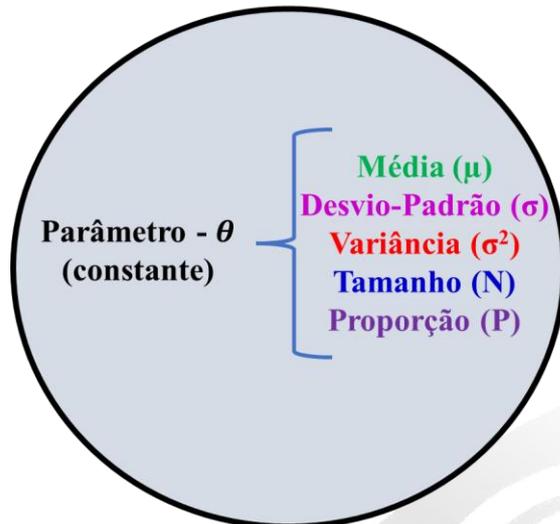
Quando coletamos dados referentes ao nosso objeto de estudo, as informações geradas (por exemplo, média, desvio-padrão, tamanho), a partir de uma população ou de uma amostra, possuem características diferentes e específicas.

As informações obtidas de uma população são **constantes** que se referem ao verdadeiro valor do fenômeno estudado. Isso porque, se todos os elementos de um objeto de estudo forem analisados, será obtido um valor único e uma informação absoluta sobre o fenômeno.

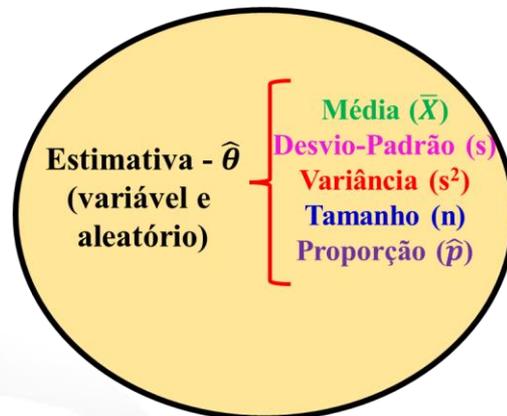
Portanto, definimos que as informações provenientes de uma população são **parâmetros**, ou seja, constantes, muitas vezes desconhecidas, de um valor representativo que permite modelar a realidade.

Por outro lado, quando se obtêm valores provenientes da amostra, sabe-se que as informações geradas representam uma **fração** do fenômeno estudado e, portanto, são valores **variáveis e aleatórios**. Com isso, tem-se a ideia de que as informações obtidas da amostra são estimativas de um parâmetro ao qual se deseja conhecer. Então, denomina-se **estimativa** as informações provenientes de dados incompletos com objetivo de generalizar um parâmetro populacional. Assim, é possível esquematizar:

População



Amostra



Sobretudo, as principais informações estudadas – média, desvio-padrão, variância, tamanho e proporção – são simbolizadas por letras gregas ou maiúsculas, quando se referem a parâmetros da população, e por letras do alfabeto comum ou minúsculas ou letras gregas acompanhadas do símbolo “^” quando se referem a estimativas da amostra.

Em síntese a todos esses conceitos abordados, pode ser aplicado o seguinte esquema conceitual:

	Conjunto Completo	Conjunto Incompleto
Quanto aos elementos avaliados	População	Amostra
Quanto aos dados coletados	Censo	Amostragem
Quanto à característica das informações obtidas	Parâmetro	Estimativa

3.2 Característica Avaliada: Variável

Após identificar os elementos em que serão extraídos os dados, é preciso determinar qual característica/atributo será avaliada, o que denominamos na Estatística como **variável**. Logo, o termo variável é a atribuição dos valores correspondentes aos dados observados e sua respectiva representação do fenômeno estudado no mundo real, isto é, a natureza dos dados, a grandeza.

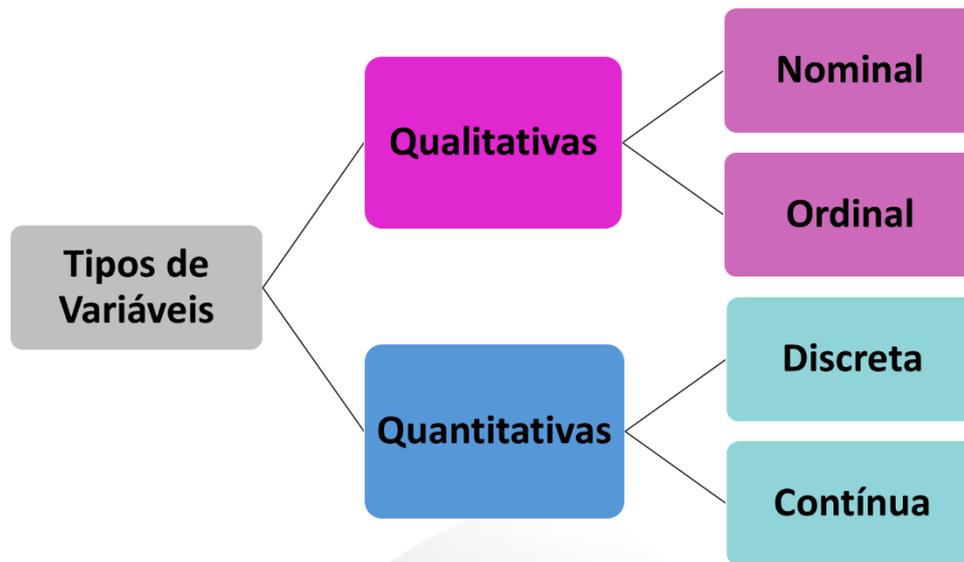
Sendo assim, ao se identificar a variável analisada, deve ser compreendido:

- A natureza ou o tipo dessa variável
- Os valores que ela pode assumir
- Como os valores foram obtidos
- Sua unidade de medida (quando tiver)
- A distribuição dessa variável; e
- Qual é o objetivo, a informação, que pretende se obter com essa variável.

Todas essas perguntas devem ser elucidadas na leitura de uma questão de Estatística, pois, com isso, obtém-se um panorama geral do contexto apresentado na questão e facilita absurdamente a sua resolução.

Os dados extraídos são uma variável que pode assumir diversos valores de acordo com fenômeno em estudo. Por isso, normalmente as variáveis são representadas por letras, como “X”, “Y”, “W”, entre qualquer outra letra. Isso é uma simbologia comum na matemática.

É importante ressaltar que os dados em questão não são necessariamente numéricos, uma vez que podem dizer a respeito de atributos qualitativos observados na população. Portanto, quanto a natureza/tipo das variáveis, elas podem ser qualitativas (nominais ou ordinais) ou quantitativas (discretas ou contínuas). Segue a ilustração:

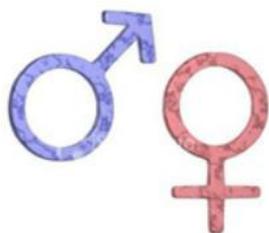


3.2.1 Variáveis Qualitativas

As variáveis de natureza qualitativa são representadas por dados que não tem informação numérica e não é possível efetuar cálculos matemáticos com esses dados. Dessa forma, os dados extraídos correspondem a categorias, classes definidas, ou qualidades. Podem também ser denominadas de **variáveis categóricas**. Esta se subdivide em duas outras classificações: nominais e ordinais.

3.2.1.1 Variáveis Qualitativas Nominais

São compostas por dados que representam qualidade, mas que não possuem nenhuma ordem ou hierarquia entre eles, apenas identificam as categoriais distintas. Não existe ordenação ou relação hierárquica entre os valores que a variável qualitativa nominal pode assumir. Exemplos:



Sexo (masculino e feminino)



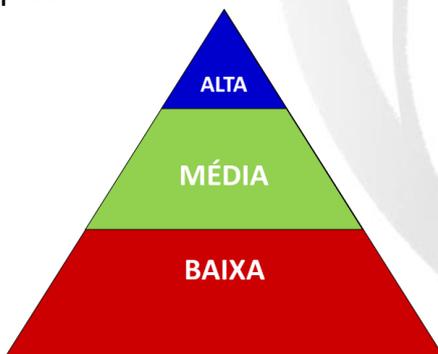
Cor (azul, preto, amarelo etc.)



País (Brasil, Espanha, Suécia etc.).

3.2.1.2 Variáveis Qualitativas Ordinais

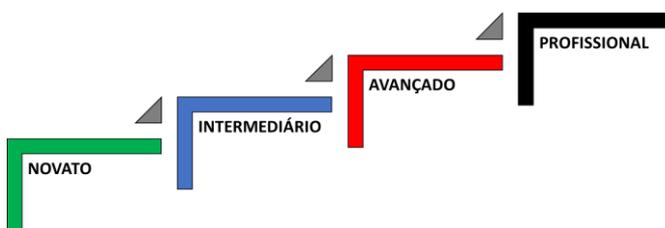
Em contrapartida, as variáveis ordinais são aquelas representadas por dados categóricos que possuem uma determinada ordenação ou hierarquia entre os valores que a variável pode assumir. Nesse contexto, é possível estabelecer uma ordem em cada categoria como baixa, média e alta, ou como em novato, intermediário e veterano. Exemplos:



Classe de renda (baixa, média, alta)



Grau de escolaridade (fundamental, médio, graduação, mestrado, doutorado).



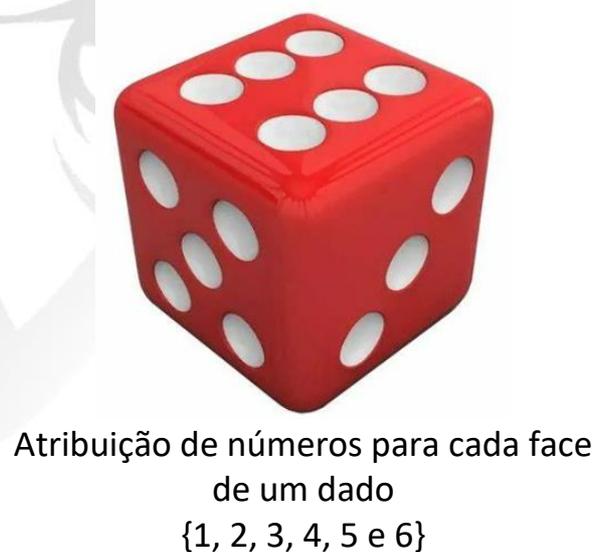
Experiência no trabalho (novato, intermediário, veterano)

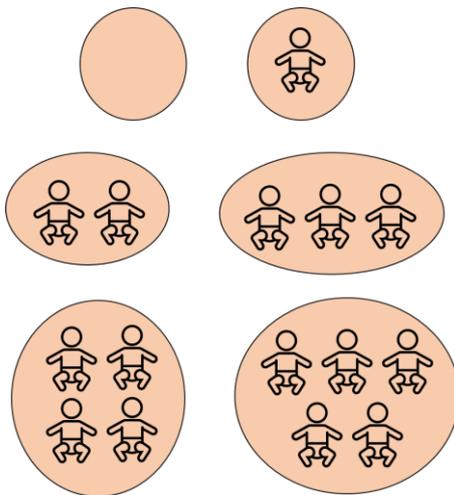
3.2.2 Variáveis Quantitativas

As variáveis de natureza quantitativa são representadas por dados que possuem informação numérica e neles podem ser efetuados cálculos matemáticos. Dessa forma, os dados são basicamente representados por valores numéricos atribuídos, contabilizados ou mensurados. Esta variável se subdivide em duas classificações: discretas e contínuas.

3.2.2.1 Variáveis Quantitativas Discretas

São aquelas variáveis representadas por valores numéricos inteiros e definidos, não possuem um intervalo entre um valor e outro (não possuem valores decimais). Geralmente, essas variáveis correspondem a **atribuição numérica** de eventos de um fenômeno específico, ou então, a **contagem** de algum fenômeno. Veja que não é coerente quantificar valores não inteiros (decimais), afinal, não existe metade de um filho, ou meio homicídio. Para esses fenômenos os resultados são taxativos, ou ocorre ou não ocorre. Exemplos:





Número de filhos por família
{0, 1, 2, 3, 4, 5...+∞}



Ocorrências de prisões por dia
{0, 1, 2, 3, 4, 5 ...+∞}

3.2.2.2 Variáveis Quantitativas Contínuas

Contudo, as variáveis contínuas são aquelas representadas por valores numéricos que podem ser inteiros e decimais. Existe um intervalo infinito entre um valor e outro. Essas variáveis estão associadas a fenômenos que no mundo real **são mensurados (medidos)** por algum instrumento e convenção criado pelo homem. Geralmente, essas variáveis estão associadas a uma grandeza e são acompanhadas por uma unidade de medida. Exemplos:



Massa (em quilogramas, toneladas etc.)



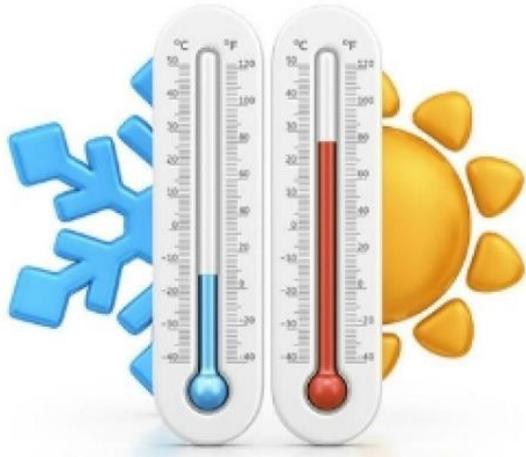
Altura (em metros, centímetros etc.)



Tempo (em horas, dias etc.)



Velocidade (em quilômetro por hora, metro por segundo etc.)



Temperatura (em graus Celsius, Kelvin etc.)

Veja que existe um intervalo infinito entre 20 kg e 30 kg, por exemplo, que pode ser 21 kg; 20,5 kg; 20,01 kg; 20,0001 kg e assim infinitamente representado.

O Quadro a seguir resume toda a classificação de variáveis e fornece exemplos:

Variável	Tipo	Definição	Exemplos	Valores
Qualitativa (Categórica)	Nominal	Classes (nomes), sem ordenação ou hierarquia	Gênero Cor País	{M, F} {Azul, Preto,...} {Brasil, China,...}
	Ordinal	Classes (nomes), com alguma ordenação ou hierarquia	Renda Nível Profissional Escolaridade	{Baixa, Média, Alta} {Novato, Interm., Veterano} {Fund., Médio,...}
Quantitativa	Discreta	Número inteiros, sem intervalo; Atribuição numérica ou contagem	Sucesso e Fracasso Fases do dado Nº de filhos Reg. De homicídios	{0, 1} {1, 2, 3, 4, 5, 6} {0, 1, 2, 3, ... +∞} {0, 1, 2, 3, ... +∞}
	Contínua	Números inteiros e fracionados, com intervalo; Mensurações	Peso Altura Tempo Velocidade	{∞, 10kg, ∞, 20kg, ∞} {∞, 70cm, ∞, 75kg, ∞} {∞, 30s, ∞, 60s, ∞} {∞, 50km/h, ∞, 80km/h, ∞}

4 CONTEXTUALIZAÇÃO DE QUESTÕES

Com base nos conhecimentos adquiridos até agora, vamos exercitar a contextualização do objeto de estudo estatístico, por meio de trechos de questões de concurso público. No decorrer dessa prática, o aluno deve observar que a essência gira em torno de algum fenômeno que o ser humano quer compreender, para isso, ele irá definir sua população de estudo, os elementos de que irá coletar os dados, qual característica será observada, como vai coletar esses dados, e assim por diante.

O objetivo desse exercício é analisar uma questão de Estatística e tentar buscar os **termos básicos** dessa disciplina em cada questão. Não se preocupe, nesse momento, com números, tabelas, gráficos e cálculos. A ideia aqui é bem objetiva: **definir os elementos da população** e **identificar a característica analisada**.

Vamos lá:

01. A tabela a seguir apresenta os resultados de uma enquete com 1000 pessoas a respeito de satisfação no trabalho.

Conceito	Frequência
Não sabe ou não opina	100
Muito insatisfeito	62
Moderadamente insatisfeito	108
Moderadamente satisfeito	320
Muito satisfeito	410

Análise da Questão:

Nessa questão, pretende-se obter a informação sobre a satisfação das pessoas em seu respectivo trabalho. Essa questão não se limitou em circunstâncias de tempo e de espaço (não especificou um ano de análise ou uma região, por exemplo).

ENTÃO, QUAL SERIA O UNIVERSO ESTATÍSTICO NESSE CASO?

Podemos afirmar que a população de estudo seriam **todas as pessoas no mundo** (até as que não trabalham, pois temos a opção de “não opina”). E os elementos do qual foi coletado os dados é **cada pessoa**. Com esse objetivo, o pesquisador coletou uma amostra de **mil pessoas**, uma vez que seria impossível interrogar todas as pessoas no mundo.

E COMO FOI OBSERVADA A VARIÁVEL SATISFAÇÃO NO TRABALHO?

Foi atribuído valores **qualitativos** para a satisfação do trabalho, são eles:

{Não sabe ou não opina; Muito insatisfeito; Moderadamente insatisfeito; Moderadamente satisfeito e Muito satisfeito}

Observe que existe uma ordenação entre os valores, desde não saber, ao aumento do grau de satisfação. Assim, podemos classificar como variável qualitativa **ordinal**.

Portanto, o dado foi coletado entrevistando cada uma das mil pessoas e se registrou a resposta que ela fornecia. Esse é objeto de estudo estatístico apresentado nessa questão.

02. 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).

Análise da Questão:

A questão 02 tem o objetivo de analisar a quantidade de imóveis ofertados e vendidos em determinado município por ano. Nessa ideia, foi contabilizado a quantidade de imóveis no decorrer dos anos 2005, 2006 e 2007, apresentados na tabela.

DE QUE OS DADOS FORAM COLETADOS?

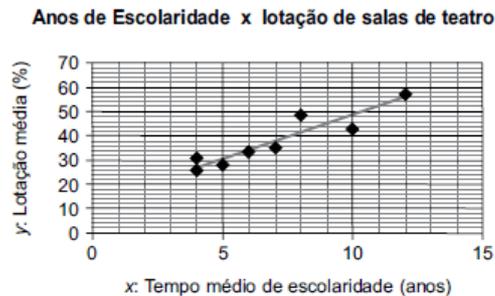
Os dados foram coletados de **cada imóvel**, desse município, nesses anos. Logo, a população de estudo, nesse contexto, são todos os **imóveis existentes nesse município e nesse intervalo de tempo (2005 a 2007)**. A questão não detalhou se foi coletada uma amostra, contudo, tranquilamente é possível raciocinar que:

- Ou foi realizado um censo observando todos os imóveis existentes desse município, nesses anos;
- Ou foi realizado uma amostragem observando apenas uma parte dos imóveis existentes desse município, nesses anos.

QUAL É A VARIÁVEL OBSERVADA NESSE CONTEXTO?

Na verdade, nessa questão, existem duas variáveis que foram quantificadas: o **número de imóveis ofertados e número de imóveis vendidos**. As duas são contagens, isto é, uma variável **quantitativa discreta**. Assim, para cada imóvel analisado foi registrado se ele era ou não ofertado e vendido, em caso positivo atribuía-se valor 1 e em cada negativo valor 0. No final, foi somado os casos positivos e obteve a contagem de imóveis ofertados (X) e vendidos (Y).

03. O diagrama de dispersão a seguir resulta de uma pesquisa com amostras aleatórias para comparar níveis de escolaridade (em anos de estudo) com a lotação média (em percentual) das salas de teatro das localidades de onde as amostras aleatórias foram colhidas.



Análise da Questão:

Para a questão 03, temos o objetivo de obter informações sobre a relação entre dois fenômenos: o nível de escolaridade e a lotação média das salas de teatro.

COMO É POSSÍVEL COLETAR E ANALISAR DADOS DE DOIS FENÔMENOS CONJUNTAMENTE?

É uma técnica Estatística que veremos com o avançar dos estudos. Muitas vezes é de interesse querer analisar a associação de dois fenômenos, desde que estejam em um mesmo contexto, nesse caso, é o **fato de visitar o teatro**. Assim, foi observado o nível de escolaridade de pessoas **que frequentavam as salas de teatro**. Ao mesmo tempo, foi observado **em cada sala de teatro**, a lotação média.

QUAIS FORAM OS ELEMENTOS, A POPULAÇÃO, E OS TIPOS DE VARIÁVEIS?

Para nível de escolaridade, foi observado em **cada pessoa que frequentava as salas de teatro**, o tempo de estudo, em anos. A população desse objeto corresponde **a todas as pessoas que frequentam o teatro**. A questão comenta sobre localidade, mas não especifica muito. E essa variável se refere ao tempo, logo, é do tipo **quantitativa contínua**.

Para a lotação média, a unidade observada foi **as salas de teatro**, e a população refere-se a **todas as salas de teatro** (possivelmente de uma região não especificada). E a variável lotação média corresponde a uma proporção da quantidade de pessoas presentes nas salas de teatro sob a quantidade de assentos disponíveis. Essa proporção foi quantificada em porcentagem (%), sem uma unidade de medida de específica. Como pode assumir infinitos valores entre 0% até 100%, consiste em uma variável **quantitativa contínua**.

04. Um funcionário de um banco foi incumbido de acompanhar o perfil dos clientes de um determinado produto por meio da Análise de Dados, de forma a aprimorar as atividades de marketing relativas a esse produto. Para isso, ele utilizou a variável classe social desses clientes, coletada pelo banco, que tem os valores A, B, C, D e E, sem referência a valores contínuos.



Análise da Questão:

Na questão 04, o objetivo é obter informações sobre o perfil dos clientes de um determinado produto. Esse perfil, refere-se a classe social do cliente, que pode estar relacionado ao poder aquisitivo, renda mensal, entre outras características que o banco determinou. Assim, nossa população em estudo são **todas as pessoas que são clientes desse determinado produto**, e **cada cliente desse produto** foi a unidade de observação.

E COMO DEFINIR ESSA VARIÁVEL CLASSE SOCIAL?

É uma convenção humana. São atribuições qualitativas sobre uma série de característica de que o banco achou necessário para definir as classes sociais. Os valores que essa variável pode assumir são A, B, C, D e E. Normalmente, como trata-se de classes e estão representadas por letras do alfabeto, entende-se que existe uma relação de ordem, como exemplo:

- A** → Muito alto nível de renda
- B** → Alto nível de renda
- C** → Médio nível de renda
- D** → Baixo nível de renda
- E** → Muito Baixo nível de renda

Logo, podemos classificar como uma variável **qualitativa ordinal**.

05. O tempo de atendimento em um ambulatório público é de grande interesse para os gestores de um município. Uma amostra com 25 pessoas indicou tempo médio de atendimento $\bar{X}=5,5$ minutos e desvio-padrão $s=0,36$. Supondo que a variável tempo de atendimento seja distribuída conforme uma distribuição normal de média μ e variância igual σ^2 .

**Análise da Questão:**

Nessa próxima questão, a informação de interesse é sobre o tempo de atendimento em um ambulatório público de um determinado município.

DE QUEM FORAM COLETADOS OS DADOS?

Já podemos concluir que ao elemento avaliado foi **cada paciente desse ambulatório**. A população desse objeto de estudo são **todos os pacientes que são atendidos nesse ambulatório público**. Nesse sentido, foi coletado uma amostra de **25 pacientes**, da qual foi a base para o estudo estatístico.

QUAL O TIPO DA VARIÁVEL ANALISADA?

Por fim, podemos definir que a característica analisada é o **tempo de atendimento**, isto é, o quanto cada paciente levava para ser atendido nesse ambulatório. A variável foi mensurada em minutos. Dessa forma, temos uma variável **quantitativa contínua**.

06. Foi realizado um experimento com uma perfuradora hidráulica com o objetivo de conhecer sua capacidade de perfuração em estruturas rochosas. Para isso foi observada a profundidade, em polegadas, de perfuração em 10 locais, cujos dados estão apresentados na Tabela a seguir.

Locais	Profundidade em polegadas
1	10,4
2	10,7
3	9,4
4	10,9
5	10,8
6	11,0
7	10,5
8	10,6
9	10,9
10	9,8



Análise da Questão:

Na questão 06, o objetivo da análise estatística é expressamente colocado: conhecer a capacidade de perfuração em estruturas rochosas de uma determinada perfuradora hidráulica.

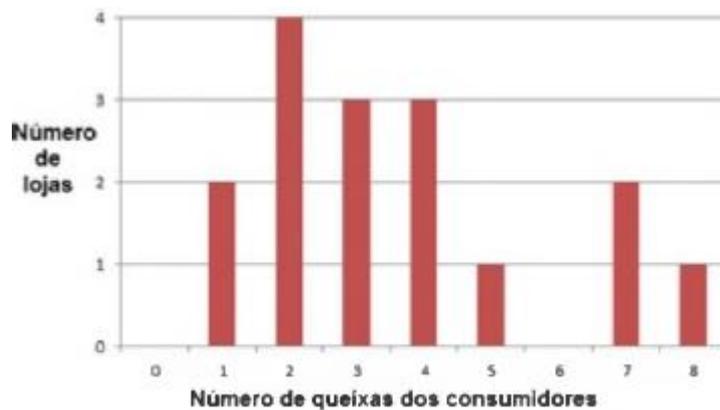
MAS DE QUE FORAM COLETADOS OS DADOS?

Nesse estudo, a unidade de observação foi **cada local perfurado pelo equipamento**. E o universo estatístico corresponde a **todos locais de estruturas rochosas que essa perfuradora hidráulica poderia perfurar**. Com esse propósito, uma amostra com **10 locais perfurados** foi coletada.

COMO DEFINIR ESSA VARIÁVEL?

Sobretudo, em cada local de perfuração foi mensurado **a profundidade que o equipamento conseguiu alcançar**, em polegadas. Facilmente, podemos observar que se trata de uma variável **quantitativa contínua**, pois o comprimento pode assumir infinitos valores.

07. Uma rede de lojas fez um levantamento da quantidade de queixas apresentadas por seus clientes ao longo de uma semana, nas 16 lojas da rede em uma região. O resultado é apresentado no gráfico abaixo. Acerca do levantamento realizado, em relação ao número de queixas por loja, analise:



Análise da Questão:

Claramente nessa questão, tem-se o objetivo de avaliar o desempenho de uma determinada rede de lojas de uma região. Esse desempenho é quantificado por meio da quantidade de queixas recebidas.

COMO DEFINIR ESSE UNIVERSO ESTATÍSTICO?

Podemos definir que o elemento avaliado foi **cada uma das lojas dessa determinada região e específica a essa rede de lojas**, enquanto a população desse estudo corresponde a **rede de lojas dessa região**. Veja que aqui a questão delimitou que essa rede de lojas é de uma determinada região, e que foi avaliado no decorrer de uma semana. Com essa especificação, podemos concluir que as **16 lojas consistem em todo o universo estatístico**, logo, analisar todas essas 16 lojas corresponderá a um **censo**.

COMO IDENTIFICAR A CARACTERÍSTICA OBSERVADA?

Para cada loja, foi quantificado o número de reclamações de seus clientes. Podemos perceber que se trata de uma contagem, logo, temos uma variável **quantitativa discreta**. Isto é, ou houve queixa ou não houve queixa, não existe a possibilidade de existir meia queixa, nesse contexto analisado.

08. O quadro, a seguir, representa uma pesquisa sobre o instrumento avaliativo de maior preferência de um grupo de alunos de um colégio.

Instrumento avaliativo	Frequência absoluta	Frequência relativa
Lista de exercícios	40	
Prova	30	
Seminário	16	
Trabalho em classe	74	
Total	160	



Análise da Questão:

Na questão 08, o objetivo da análise estatística é gerar informações sobre a preferência de um instrumento avaliativo para um grupo de alunos de um colégio.

QUAL É A UNIDADE OBSERVADA E O UNIVERSO ESTATÍSTICO?

Nesse raciocínio, podemos determinar que **cada aluno desse colégio** foi avaliado quanto a sua preferência ao instrumento avaliativo. Ao mesmo tempo, nossa população refere-se a **todos os alunos desse colégio**. A questão não especifica se foi coletada uma amostra.

QUAL A VARIÁVEL DESSA QUESTÃO? E QUE VALORES ELA PODE ASSUMIR?

A característica analisada é o tipo de instrumento avaliativo, que nessa pesquisa enquadra-se nas seguintes opções:

Lista de Exercícios
Prova
Seminário
Trabalho em Classe

Essas eram as opções que os alunos optavam quanto a sua preferência. Veja que a variável assume informação não numérica, do tipo **qualitativa nominal**, uma vez que não existe ordem ou hierarquia entre os instrumentos avaliativos.

ESTATÍSTICA DESCRITIVA

A Estatística Descritiva consiste em análises com objetivo de descrever, organizar, resumir, simplificar e sintetizar um conjunto de dados coletados sobre algum fenômeno em estudo. Além disso, ela tem utilidade como análise exploratória, uma vez que resume o conjunto de dados. Com isso, a Estatística Descritiva permite apontar tendências e levantar hipóteses sobre os possíveis resultados do fenômeno estudado.

Como principais ferramentas para esse propósito, a Estatística Descritiva trabalha com formas sintetizadas de apresentação de dados como tabelas e gráficos. Além disso, utiliza medidas descritivas que são informações provenientes de cálculos que tentam descrever informações geradas pelo conjunto de dados.

Estatística
Descritiva

Tabelas

Gráficos

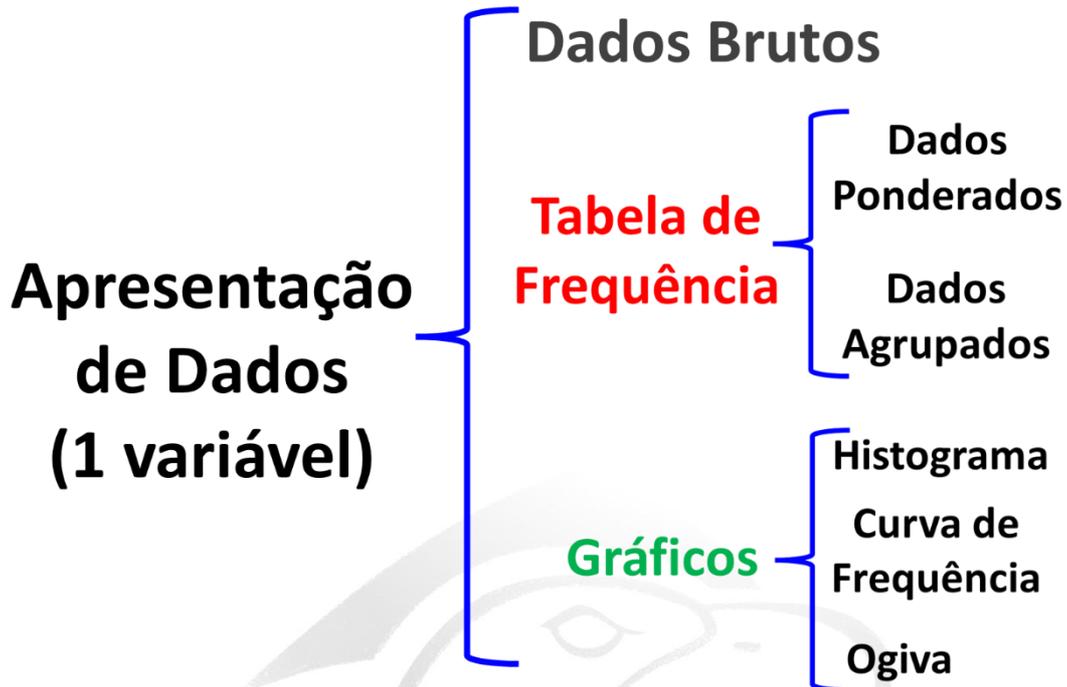
Medidas Descritivas

5 APRESENTAÇÃO DE DADOS

Os dados coletados de uma variável podem ser apresentados de muitas formas. Basicamente, o aluno deve entender que poderá se deparar com qualquer umas das formas de apresentação de dados em uma prova de Estatística e, com isso, é necessário entender como extrair as informações em cada situação.

Os dados referentes a uma única variável podem ser encontrados na forma de **dados brutos, em tabela de frequência (como dados ponderados ou dados agrupados)**, ou então, por meio de gráficos, como **histograma, curva de frequência, ogiva**, entre outros. Essas citadas são as principais formas encontradas.

Para exemplificar a apresentação dos dados referentes a uma variável, vamos utilizar um exemplo e comparar como os dados serão simplificados no decorrer de cada forma de apresentação.



Objeto de estudo:

Para analisar o nível de criminalidade de um estado, registra-se o número de ocorrências de crimes por dia desse estado. Essa variável foi observada no decorrer de 16 dias.

Variável X: *ocorrências de crimes por dia (crimes/dia);*

5.1 Dados Brutos

Os dados brutos correspondem à listagem dos dados conforme eles foram coletados em um estudo qualquer. Cada dado é uma **observação** feita pelo homem, no qual ele atribui valor referente a uma característica para um elemento da população.

Conforme o exemplo abordado, os dados brutos referentes à quantidade de crimes registrados por dia podem ser representados da seguinte maneira:

$$X = \{12, 14, 16, 18, 22, 14, 16, 18, 20, 15, 16, 10, 12, 16, 18, 20\}$$

Nesse exemplo, cada valor numérico coletado representa a quantidade de crimes registrado **por dia nesse determinado estado**. Assim, como foi analisado no decorrer de 16 dias, temos 16 dados coletados. Assim, temos nossa primeira informação:

$$n = 16$$

n : quantidade de dados coletados; ou número de elementos; ou tamanho do conjunto de dados; ou tamanho da amostra etc.

Tudo isso pode ser concluído pela análise da unidade variável que consiste em crimes/dia, ou seja, cada coleta corresponde a um dia transcorrido. Com isso, outra forma em que os dados brutos podem ser encontrados é:

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_i	12	14	16	18	22	14	16	18	20	15	16	10	12	16	18	20

Nessa tabela, é indicada o dia e seu respectivo registro de quantidade de crimes. Assim, podemos fazer seguintes interpretações:

Ocorreu 12 crimes no 1º dia analisado

Ocorreu 14 crimes no 2º dia analisado

...

Ocorreu 22 crimes no 5º dia analisado

...

Ocorreu 15 crimes no 10º dia analisado

...

Ocorreu 20 crimes no 16º dia analisado

Observe que, nessa forma de representação, ainda tratamos de dados brutos.

Outra característica que pode ser observada nos dados brutos é a **repetição** de dados observados com mesmo valor, isso é uma distinção muito importante comparada as demais formas de apresentação de dados.

Os dados brutos muitas vezes precisam ser ordenados para estudar a posição e distribuição dos dados. Quando os dados estão ordenados, é denominado de **rol**. Segue exemplo dos dados observados em rol:

Dados em Rol

$$X = \{10, 12, 12, 14, 14, 15, 16, 16, 16, 16, 18, 18, 18, 20, 20, 22\}$$

5.2 Dados Ponderados (Tabela de Frequência por valor)

Também indicada como tabela de frequência sem intervalo. Os dados são ponderados quando suas repetições são sintetizadas em uma informação que contabiliza sua repetição: **a frequência**. Essa informação denominada de frequência consiste na contagem de dados com mesmo valor (repetidos). Assim, cada valor observado que uma variável pode assumir é **ponderado pelas suas repetições**.

Os dados são representados por uma tabela de frequência, em que cada dado da variável “ocorrência de crime por dia” é colocado em uma linha com sua respectiva frequência (quantidade de vezes que aquele dado se repete). Segue ilustração a seguir:

Ocorrências de crimes por dia (X)	Frequência Absoluta (f_i)	Frequência Acumulada (F_i)
10	1	1
12	2	3
14	2	5
15	1	6
16	4	10
18	3	13
20	2	15
22	1	16
Total (Σ)	16	-

Na tabela de frequência de dados ponderados, cada linha corresponde às frequências de uma observação dos dados brutos. A primeira coluna corresponde aos valores observados da variável X e nas demais colunas temos as frequências que podem ser na forma absoluta simples ou acumulada.

- **Frequência Absoluta (f_i):** contagem de repetições de cada valor observado; Portanto, pela frequência absoluta, podemos fazer a seguinte leitura desses dados:

Ocorreu 10 crimes em 1 dia analisado

Ocorreu 12 crimes em 2 dias analisados

Ocorreu 14 crimes em 2 dias analisados

Ocorreu 15 crimes em 1 dia analisado

Ocorreu 16 crimes em 4 dias analisados

Ocorreu 18 crimes em 3 dias analisados

Ocorreu 20 crimes em 2 dias analisados

Ocorreu 22 crimes em 1 dia analisado

- **Frequência Acumulada (F_i):** número de repetições de cada valor observado, somando (acumulando) os valores das linhas anteriores (valores já registrados);

Portanto, pela frequência absoluta, podemos fazer a seguinte leitura desses dados:

Em 1 dia, ocorreu até 10 crimes

Em 3 dias, ocorreu até 12 crimes

Em 5 dias, ocorreu até 14 crimes

Em 6 dias, ocorreu até 15 crimes

Em 10 dias, ocorreu até 16 crimes

Em 13 dias, ocorreu até 18 crimes

Em 15 dias, ocorreu até 20 crimes

Em 16 dias, ocorreu até 22 crimes

Outras formas de representar a frequência é, ao invés de representar a contagem de repetições de cada dado (frequência absoluta), aplicar a proporção (ou porcentagem) de vezes que se repetiu. Veja:

Ocorrências de crimes por dia (X)	Frequência Relativa (fr_i)	Frequência Acumulada Relativa (Fr_i)
10	1/16 ou 0,0625 ou 6,25%	6,25%
12	2/16 ou 0,125 ou 12,5%	18,75%
14	2/16 ou 0,125 ou 12,5%	31,25%
15	1/16 ou 0,0625 ou 6,25%	37,5%
16	4/16 ou 0,25 ou 25%	62,5%
18	3/16 ou 0,1875 ou 18,75%	81,25%
20	2/16 ou 0,125 ou 12,5%	93,75%
22	1/16 ou 0,0625 ou 6,25%	100%
Total (Σ)	16/16 ou 1 ou 100%	-

- **Frequência Relativa (fr_i):** proporção da frequência absoluta sobre o número de elementos, $fr_i = \frac{f_i}{n}$ (conceito muito associado a probabilidade);
- **Frequência Acumulada Relativa (Fr_i):** proporção da frequência acumulada sobre o número de elementos, $Fr_i = \frac{F_i}{n}$.

Desse modo, não temos a simples contagem das repetições, mas sim a proporção do quanto se repete. Transformamos esse dado de frequência de **absoluto para relativo**. Quando afirmamos que dados são relativos se entende que o valor observado está sendo relativizado em relação a um total de referência, que nesse caso é o total de elementos observados (n). Logo, dados relativos são dados em fração ou em porcentagem.

Esses quatro tipos de frequências podem aparecer em sua prova de Estatística, dificilmente todos os quatro tipos juntos em uma questão. Por isso, é necessário identificar as características de cada uma das frequências, como também diferenciar suas simbologias. Observe que quando a frequência for acumulada o F é representado por letra maiúscula e f minúscula para absoluta, sem acumular.

Alguns pontos úteis para se observar:

- A soma total da frequência absoluta deve sempre ser o n – total de dados observados

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

- A soma da frequência relativa deve sempre ser 1 ou 100%.

$$\sum_{i=1}^n f_{ri} = 1 \text{ ou } 100\%$$

- A última linha da frequência acumulada deve ser o total de elementos (n).
- A última linha da frequência acumulada relativa deve ser 1 ou 100%.

5.3 Dados Agrupados (Tabela de Frequência com intervalos)

Quando o conjunto de dados fica cada vez maior, ocorrem muitas observações diferentes e representá-las por meio de dados ponderados pode ficar muito extensivo e pouco claro. Para isso, outra forma de apresentação de dados sugere que as observações da variável X sejam agrupadas em intervalos predefinidos. Um detalhe importante é que o agrupamento de dados somente é coerente para dados quantitativos, sendo mais frequentemente aplicado em dados quantitativos contínuos.

Para agrupar os dados, é necessário definir duas informações:

- **Número de classes (n_c)**
- **Amplitude/intervalo da classe (h).**

Para definir o número de classes, podem ser utilizadas diferentes metodologias, entre as principais está a regra de Sturges e o critério da raiz quadrada.

Regra de Sturges:

$$n_c = 1 + 3,3 \text{Log } n$$

Critério da Raiz Quadrada:

$$n_c = \sqrt{n}$$

As questões de Estatística raramente pedem para estabelecer o número de classes, no entanto conhecer os critérios para formação de classes é útil para entender uma tabela de frequência com dados agrupados. Por praticidade nos cálculos e quando não for fornecida uma informação específica na questão, recomenda-se usar o critério da raiz quadrada. Para esse exemplo, será utilizado esse critério, logo:

$$n_c = \sqrt{16} = 4 \text{ classes}$$

Portanto, quatro classes são suficientes para agrupar todo o conjunto de dados. Quando o valor do número de classes não for inteiro, pode ser feito um arredondamento. Após isso, resta saber qual é o intervalo da classe. Essa informação pode ser calculada a partir do seguinte cálculo:

$$h = \frac{(X_{Máx} - X_{Mín})}{n_c}$$

Dessa forma, calcula-se a diferença entre o valor máximo e o mínimo observado no conjunto de dados em análise, em seguida, essa diferença é dividida pelo número de classes. Assim, tem-se o seguinte valor para h:

$$h = \frac{(22 - 12)}{4} = 2,5$$

O valor $h = 2,5$ corresponde que a cada 2,5 unidades da variável X há uma classe definida, e assim sucessivamente até estabelecer as quatro classes. Contudo, é coerente **arredondar esse valor para 3** e não deixar valor fracionado nesse exemplo, pois estamos tratando de ocorrência de crime, uma variável quantitativa discreta. Se fosse uma variável quantitativa contínua o raciocínio manter em 2,5.

Desse modo, a tabela de frequência para dados agrupados fica da seguinte forma:

Ocorrências de Crime por dia (X_i)	Frequência Absoluta (f_i)	Frequência Acumulada (F_i)
10 ┆ 13	3	3
13 ┆ 16	3	6
16 ┆ 19	7	13
19 ┆ 22	3	16
Total (Σ_i)	9	-

A tabela de frequência com intervalos agrupa os dados observados da variável X e simplifica as informações. No entanto, a partir do momento que os dados são agrupados, a precisão de algumas informações é perdida. Por exemplo, é possível constatar que no intervalo de 10 ┆ 13 existem três observações, entretanto, não é possível afirmar precisamente quais são os dados que foram observados nesse intervalo, que nesse caso seria {10, 12, 12}. Mesmo perdendo a precisão de algumas informações, a apresentação de dados agrupados é interessante por reduzir e simplificar um conjunto grande de dados.

Cada intervalo de classe é caracterizado pelo seu limite inferior e superior. Exemplo, a primeira classe é caracterizada por 10 ┆ 13, que engloba observações de 10 crimes/dia até valores < 13 crimes/dia. Observe que o limite superior não é enquadrado no primeiro intervalo, pois faz parte do limite inferior da segunda classe. O símbolo ┆ indica que o limite inferior está contido na classe e o superior não está. Já na terceira e última classe, note que o intervalo foi representado por 19 ┆ 22, que engloba tanto limite inferior como superior, para poder englobar todos os dados na tabela de frequência.

As simbologias de intervalos podem ser as seguintes:

- ┆ Intervalo que **inclui** o limite inferior e **inclui** o limite superior;
- ┆ Intervalo que **inclui** o limite inferior e **exclui** o limite superior;
- ┆ Intervalo que **exclui** o limite inferior e **inclui** o limite superior;
- ┆ Intervalo que **exclui** limite inferior e **exclui** o limite superior.

A tabela de frequência para dados agrupados também pode ser expressa em dados **relativos**:

Ocorrências de crimes por dia (X)	Frequência Relativa (f_{r_i})	Frequência Acumulada Relativa (Fr_i)
10 – 13	18,75%	18,75%
13 – 16	18,75%	37,5%
16 – 19	43,75%	81,25%
19 – 22	18,75%	100%
Total (Σ)	100%	-

Importante: com essa forma resumida, não se sabe mais a frequência específica de cada ocorrência de crime observada (valor no ponto). Tem-se apenas a frequência naquele intervalo. Por isso, a expressão mais adequada é **densidade de frequência** D_f (frequência distribuída no intervalo de valores observados da classe).

$$D_f = \frac{f_i}{\Delta X_i} = \frac{\text{Frequência absoluta}}{\text{Intervalo de valores da classe}}$$

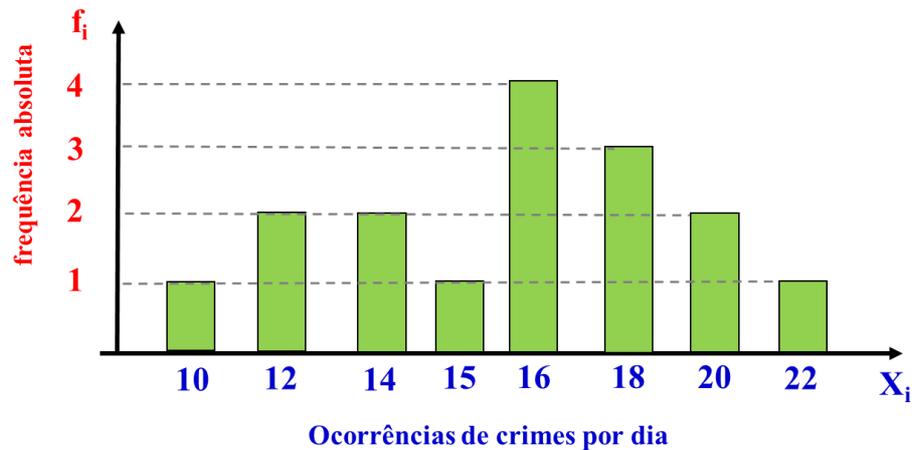
5.4 Gráficos de Frequência

Seguindo com as formas de apresentação de dados da variável X, a frequência pode ser também ilustrada por meio de gráficos, tanto para os dados ponderados quanto os dados agrupados. As principais representações gráficas associadas à frequência são: colunas de frequência, histograma, curva de frequência, e ogiva.

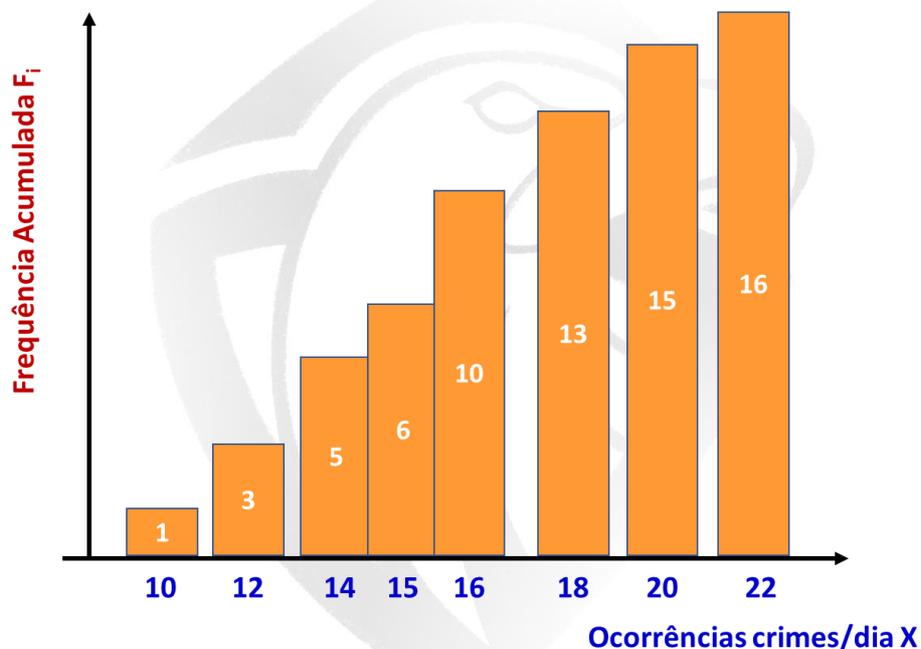
5.4.1 Colunas de Frequência

Ao invés de tabela, o conjunto de **dados ponderados** pode ser representado por um gráfico de colunas, em que cada valor da variável X possui uma coluna com altura proporcional a sua frequência.

➤ Gráfico de Colunas com frequência absoluta:



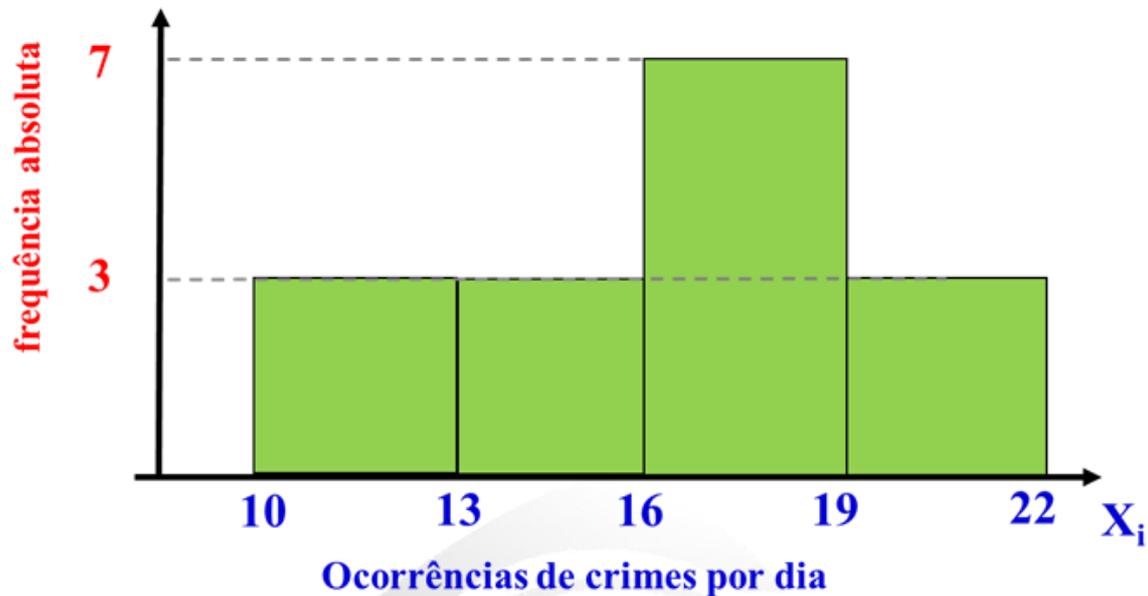
➤ Gráfico de Colunas com frequência acumulada:



5.4.2 Histograma

Histograma é uma representação gráfica em retângulos (como um gráfico de barras verticais) da distribuição de frequências de um conjunto de **dados agrupados**. Os histogramas podem ser apresentados das seguintes formas:

➤ Histograma da Frequência:



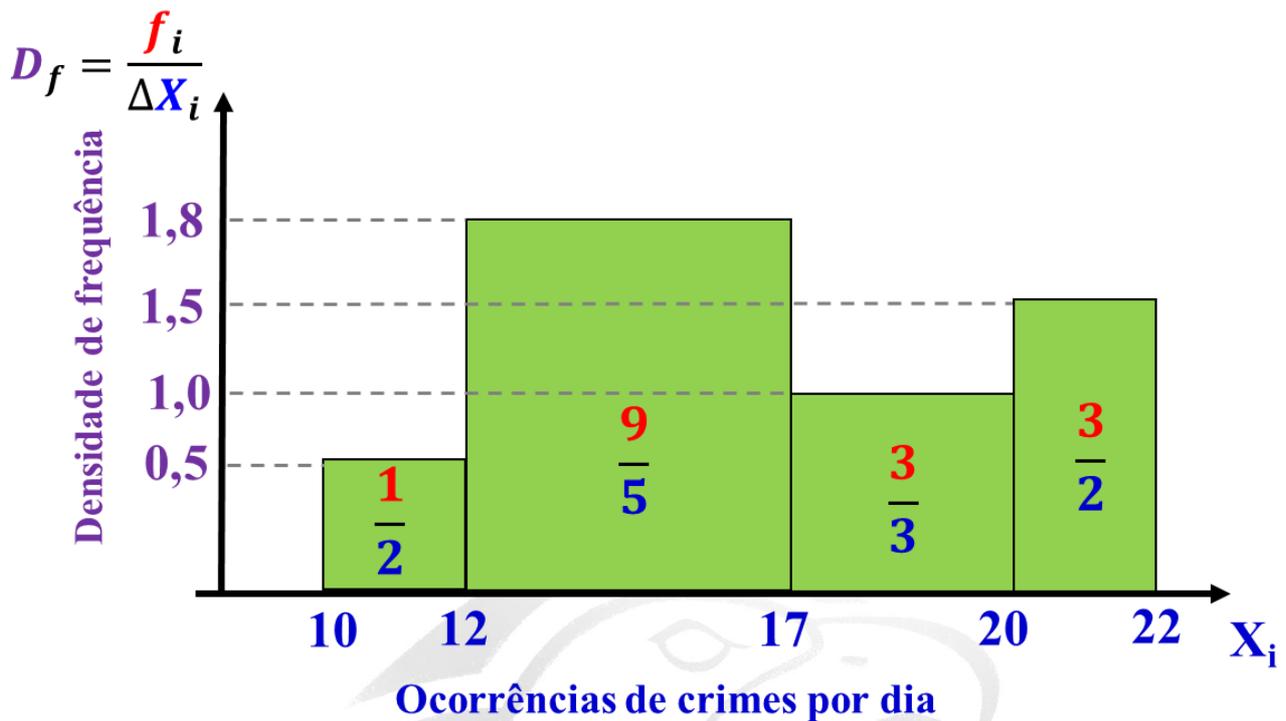
O histograma, assim como as colunas de frequências, pode ser representado na forma vertical ou horizontal, porém sempre será composto pela relação entre a variável estudada e sua frequência, cada uma representando um eixo do plano cartesiano. O histograma estuda a frequência de apenas uma variável, assim não deve ser confundido com outros gráficos de barras ou colunas que costumam associar duas variáveis distintas.

Os histogramas da frequência absoluta irão assumir uma forma variável (distribuição) conforme a repetição de cada observação. Em contrapartida, o histograma da frequência acumulada sempre será crescente com última observação ou classe com o valor total de observações acumulado.

Como o histograma é uma representação gráfica de dados agrupados é mais comum o eixo da frequência ser representado pela **densidade de frequência**. Mais ainda quando, por alguma razão, é de interesse agrupar os dados **em classes de amplitude variadas**. Nessa situação, é utilizado a densidade de frequência para ilustrar o histograma.

Exemplo:

Por alguma circunstância, é mais interessante agrupar as ocorrências de crimes pelas seguintes classes: **10 a 12**; **12 a 17**; **17 a 20**; e **20 a 22**. Com isso, teremos o seguinte histograma:



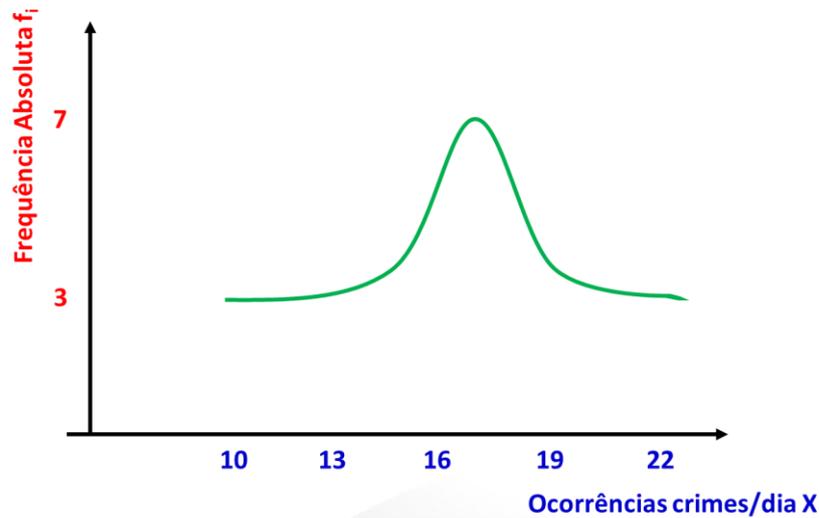
Veja que a classe “20 a 22” possui uma coluna com maior altura que a classe “17 a 20”, apesar de possuírem a mesma frequência absoluta. Isso ocorre porque o intervalo da classe “20 a 22” é menor (de 2 unidades), o que gera uma densidade de frequência maior ($D_f = 1,5$) que a classe “17 a 20” com intervalo de 3 unidades ($D_f = 1$).

5.4.3 Curva de frequência

Nesse gráfico, basicamente, ele representa a frequência por linha curva, onde com o aumento da frequência a linha é puxada para cima e vice-versa. Há representações gráficas também com linha reta (comumente chamado de polígono).

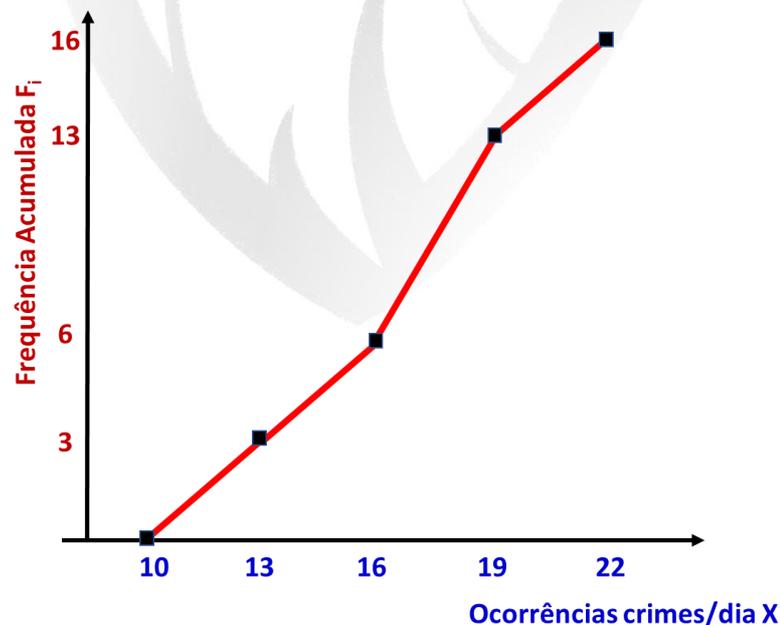
A curva de frequência é de interessante quando o objetivo é evidenciar uma tendência de o suposto desempenho que o fenômeno teria com maior número de dados coletados. Esse gráfico é muito usado para estudar a distribuição e o formato do conjunto de dados. A seguir, a representação gráfica das curvas de frequência:

- Curva de Frequência:

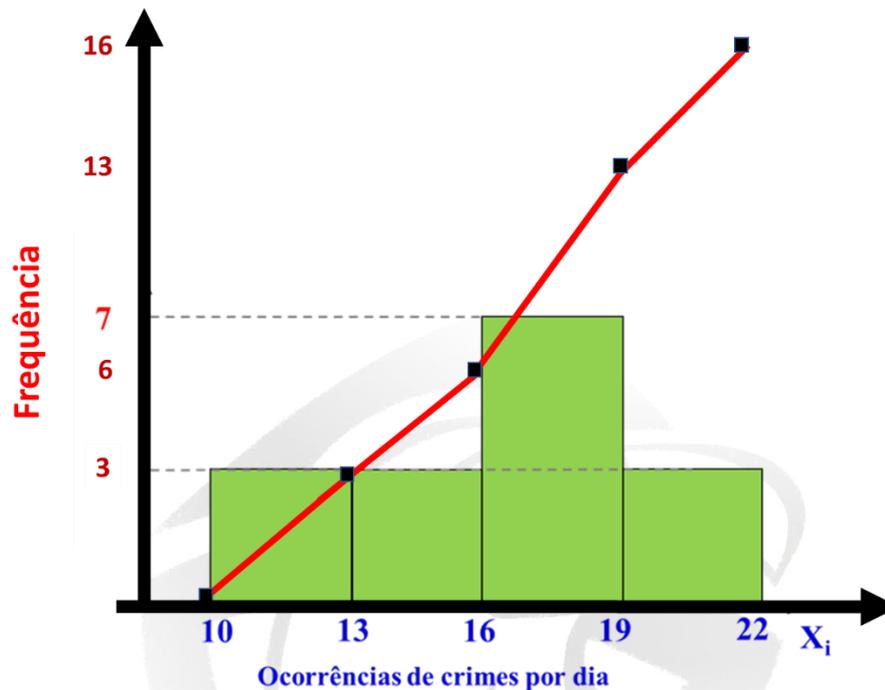


5.4.4 Ogiva

O gráfico de Ogiva (ou Ogiva de Galton) consiste em uma representação focada na **frequência acumulada** (com dados absolutos ou relativos). Ele é representado por linhas retas ou curvas, e mostra os valores observados da variável de acordo com sua frequência acumulada. Assim, a menor observação (limite inferior) sempre iniciará com frequência acumulada zero e a maior observação (limite superior) terá frequência acumulada igual a quantidade de dados observados, isto é, igual a n .



Em seus estudos, o aluno pode encontrar a junção dessas representações gráficas de frequência, como o histograma e a ogiva. Conhecendo as características de cada uma, é facilmente possível extrair as informações que for de interesse. Veja o exemplo:



6 TABELAS

As tabelas de modo geral servem para organizar e apresentar os dados coletados no sentido de facilitar a sua interpretação. A tabela de frequência, que vimos nos tópicos anteriores, geralmente apresenta apenas informações referentes a uma variável e sua respectiva frequência, já as demais tabelas costumam associar dados de diferentes variáveis.

6.1 Tabela Simples de Associação de Variáveis

A organização estruturada dos dados observados em tabelas faz com os valores de cada variável fiquem alinhados. Desse modo, cada coluna é representada por uma variável, e cada linha é uma associação de valores dessas variáveis que estão ligados a um **mesmo contexto**. Essa organização, transforma os dados dessas variáveis em informação. É muito utilizada em **banco de dados**, onde se tem grandes quantidades de dados e variáveis. Vamos ver isso através de um exemplo:

Objeto de estudo:

Foram coletados dados sobre os concursos de carreiras policiais da esfera federal. As informações obtidas de cada concurso foram o número de vagas, quantidade de inscritos, a remuneração do concurso e nível de dificuldade do concurso.

Com todas essas características obtidas, a melhor forma de ilustrar é por meio de uma tabela simples que associa todas essas informações.

Concurso	Vagas	Inscritos	Remuneração	Dificuldade
Polícia Federal	600	85.000	R\$ 12.600,00	Alta
Polícia Penal Federal	150	21.000	R\$ 8.900,00	Média
Polícia Rodoviária Federal	400	60.000	R\$ 10.800,00	Alta
Polícia Legislativa	80	55.000	R\$ 19.700,00	Baixa

Assim, é possível afirmar que o concurso da Polícia Federal, oferecerá 650 vagas, tem 85 mil inscritos, remuneração de R\$ 12.600,00, e avaliado por algum critério qualquer com dificuldade alta na prova. Veja que variáveis de diferentes tipos (qualitativa e quantitativa) podem ser associadas sem problema algum em uma tabela. É a forma mais ideal de organizar análises multivariáveis (quando se observa mais de uma característica da população).

6.2 Tabela de Dupla Entrada

A tabela de dupla entrada apresenta os valores de duas variáveis conjuntamente, coletadas sobre um mesmo contexto e que seja possível essa associação. São ideais para apresentar as frequências de duas variáveis **qualitativas (nominais e ordinais)**, também, ao invés da frequência pode ser apresentado uma variável quantitativa associada a duas variáveis qualitativas. Vejam nos exemplos a seguir:

Objeto de estudo:

Em uma empresa de automóveis, foi realizada uma entrevista com 200 pessoas para saber qual marca de automóvel a pessoa utilizava. Era de interesse também distinguir o sexo do cliente e associada com a marca de automóvel.

Sexo do Cliente	Marca de Automóvel			Total
	A	B	C	
Masculino	36	51	33	120
Feminino	24	34	22	80
Total	60	85	55	200

Nesse exemplo, temos a associação de duas variáveis qualitativas nominais (sexo e marca do automóvel). Destarte, é possível observar nessa tabela que os números em seu centro correspondem a um dado associado as duas características. Assim, podemos afirmar que existem 36 homens que utilizam a marca A; 24 mulheres que utilizam a marca A; 22 mulheres que utilizam a marca C, e assim por diante. Nas extremidades da tabela, temos os subtotais de cada classe (os valores da variável qualitativa). Com isso, podemos afirmar que foram entrevistados 120 homens; e que 85 pessoas utilizam a marca de automóvel B. Por fim, na extremidade inferior direito, temos o total de observações.

6.3 Séries Estatísticas

Séries estatísticas é um conceito atribuído para qualquer tabela que apresenta a distribuição de um conjunto de dados analisados em **função do tempo, do local ou da espécie (ou fato)**. Assim, conforme o fator que tem sua variação apresentada na tabela, classificamos as séries estatísticas em: históricas, territoriais, categóricas ou mistas.

6.3.1 Séries Históricas

Também podem ser chamadas de séries **cronológicas ou temporais**. São tabelas que apresentam os valores de uma variável em estudo, em função do **tempo** (que pode ser em meses, semanas, anos etc.). Nessa série, o local e a categoria são fixos e determinados. Veja Exemplo:

Objeto de estudo:

Estudo sobre o percentual de inquéritos policiais solucionados na circunscrição de Três Lagoas/MS no decorrer dos anos de 2016 a 2020.

Anos	Inquéritos Solucionados
2016	21%
2017	26%
2018	27%
2019	30%
2020	33%

Veja que nesse exemplo temos uma região determinada (Três Lagoas/MS) e sem variações de qualquer categoria ou especificação.

6.3.2 Séries Territoriais

Também podem ser chamadas de séries **geográficas ou espaciais**. São tabelas que apresentam os valores de uma variável em estudo, em função do **espaço** (que pode ser cidades, regiões, estados, países etc.). Nessa série, o tempo e a categoria são fixos e determinados. Veja Exemplo:

Objeto de estudo:

Estudo sobre o índice de segurança pública no ano de 2022 avaliados nas capitais dos estados de São Paulo, Santa Catarina, Rio Grande do Sul, Paraná e Mato Grosso do Sul.

Capitais	Índice de Segurança Pública
São Paulo	61%
Florianópolis	79%
Porto Alegre	66%
Curitiba	73%
Campo Grande	64%

Veja que nesse exemplo temos uma época determinada (ano de 2022) e sem variações de qualquer categoria ou especificação.

6.3.3 Séries Categóricas

Também pode ser chamada de séries **específicas**. São tabelas que apresentam os valores de uma variável em estudo, em função de uma **categoria** ou **especificações** (que pode ser fatos, eventos, classes, espécies biológicas etc.). Nessa série, o tempo e o espaço são fixos e determinados. Veja Exemplo:

Objeto de estudo:

Estudo sobre o risco de extinção de espécies animais, no ano de 2021, que vivem no bioma do Pantanal.

Espécie Animal	Risco de Extinção
Onça-pintada	84%
Jacaré	56%
Pica-pau	49%
Capivara	38%

Veja que nesse exemplo temos uma época determinada (ano de 2021) e um espaço determinado (bioma do pantanal).

6.3.4 Séries Mistas

São tabelas que apresentam variações de mais de uma variável, conjugando duas ou mais séries estatísticas. Dessa forma, são comumente apresentadas em uma tabela de dupla entrada combinando dois diferentes tipos de séries estatísticas. Assim, as séries mistas podem ser: categórica-cronológica, categórica-territorial, cronológica-territorial. Pode também combinar as três séries estatísticas, utilizando uma tabela de maior complexidade.

Objeto de estudo:

Estudo sobre o número de homicídios violentos registrados nas unidades federativas da região centro-oeste do Brasil, no decorrer dos últimos três anos completos.

Anos	Centro-Oeste			
	MS	MS	GO	DF
2020	467	562	415	657
2021	400	341	447	603
2022	312	222	198	498

Nesse exemplo, temos a variável número de homicídios violentos em função do tempo (2020, 2021 e 2022) e em função do espaço (MS, MT, GO e DF). Não há variação em função de alguma especificação. Logo, temos série mista cronológica-territorial.

7 GRÁFICOS

Os gráficos, de modo geral, são representações ilustrativas do conjunto de dados que proporcionam um maior impacto visual. Basicamente, os gráficos devem apresentar simplicidade, clareza na leitura dos valores e veracidade nas informações sobre o fenômeno estudado. Existem inúmeros formas de representação gráfica para duas ou mais variáveis, nesse material serão abordados os principais gráficos cobrados nas provas de Estatística, entre eles, diagramas, gráficos de colunas, barras, setor (pizzas), dispersão e linhas.

De acordo com o tipo de variável analisada, existirá uma representação gráfica mais adequada. Contudo, a escolha de qual gráfico usar é subjetiva, cabe ao pesquisador/investigador decidir. No entanto, é comum provas de Estatística cobrarem o gráfico mais apropriado de acordo com o fenômeno estudado. Vamos esclarecer isso em cada tópico, mas de modo geral segue um esquema resumindo essa aplicação:

Qual Gráfico Utilizar?

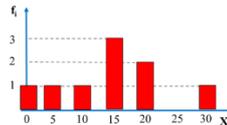
➤ **Dados Qualitativos Nominais**

↳ **Gráfico de setores (Pizzas)**



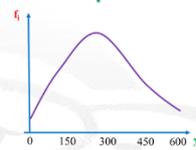
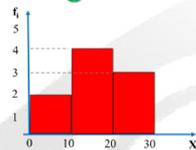
➤ **Dados Qualitativos Ordinais e Quantitativos Discretos (não agrupados)**

↳ **Gráfico de Colunas ou de Barras**



➤ **Dados Quantitativos Discretos (agrupados) e Contínuos**

↳ **Histograma ou Curva de Frequência**

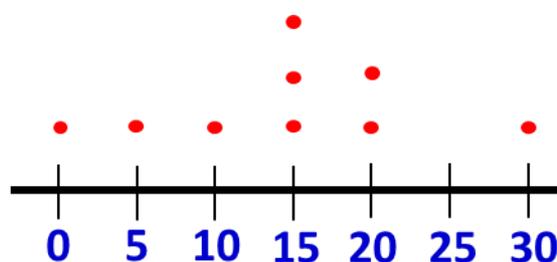


7.1 Diagrama de Pontos

Os dados de uma variável podem ser representados por um diagrama de pontos que apresenta a frequência de cada observação com uso de pontos, isto é, basicamente um histograma representado por pontos ao invés de colunas ou barras. Veja como fica representado o diagrama de pontos, no mesmo exemplo abordado da aula anterior:

Objeto de estudo:

Dados referentes a quantidade de drogas (X), em quilogramas, apreendidas por semana em uma delegacia de polícia.



Quantidade de drogas apreendidas por semana (kg/semana)

Essa ilustração corresponde ao seguinte conjunto de dados brutos:

$$X = \{0, 5, 10, 15, 15, 15, 20, 20, 30\}$$

7.2 Diagrama de ramos e folhas

O diagrama de ramos e folhas trata-se de outra representação gráfica para expor um conjunto de dados referentes a uma variável. Para apresentá-lo, será utilizado um exemplo que desenvolve melhor sua aplicação.

Objeto de estudo:

Altura de alunos (Y), em centímetros, da turma de Academia de Polícia Federal.

Ramos	Folhas
16	1226
17	445778
18	1134555667
19	14677
20	0236

Esta representação gráfica aglomera os dados brutos a partir de um esquema que associa a ideia de uma árvore com ramos e folhas. Cada observação é composta pela **concatenação do ramo com suas respectivas folhas**. Assim, o ramo 16 concatenado com cada folha tem as seguintes observações: 161cm, 162cm, 162cm, 166cm. Para obter o total de observações, basta contabilizar a quantidade de números em cada folha. Esse gráfico é indicado para conjunto de dados com poucas observações (o suficiente para não perder a simplicidade e clareza).

Em adição, o diagrama de ramos e folhas evidencia a distribuição dos dados, bem como a frequência das observações. Isso pode ser constatado pelo comprimento das linhas de cada folha, quanto maior a linha (mais números) maior é a quantidade de dados naquele intervalo. Também pode ser compreendido que cada ramo consiste em um intervalo de dados agrupados com amplitude, nesse exemplo, de uma dezena 16 + 17, 17 + 18, e assim consecutivamente.

O conjunto de dados representado no diagrama de ramos e folhas pode ser expresso em dados brutos da seguinte forma:

$Y = \{161, 162, 162, 166, 174, 174, 175, 177, 177, 178, 181, 181, 183, 184, 185, 185, 185, 186, 186, 187, 191, 194, 196, 197, 197, 200, 202, 203, 206\}$

7.3 Gráfico de Colunas

O objeto de estudo exemplificado a seguir será aplicado para ilustrar os gráficos de colunas, barras e setores.

Objeto de estudo:

Quantidade de prisões efetuadas por mês nos estados do Sudeste brasileiro.

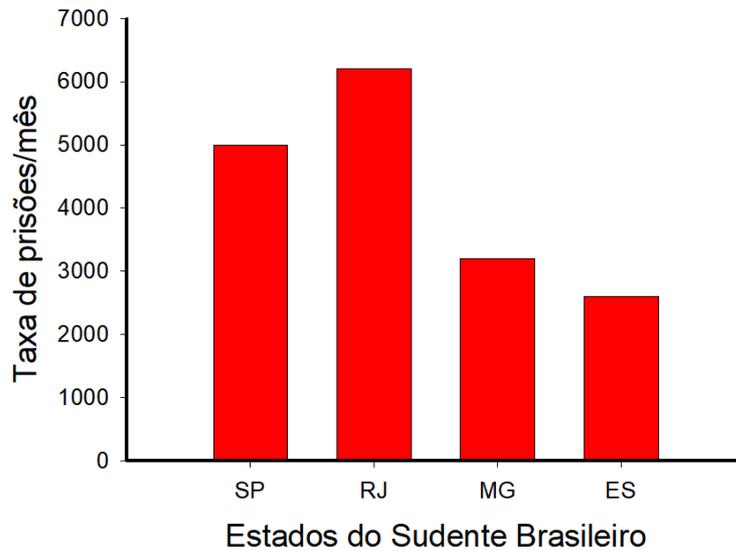
Rio de Janeiro: 6200

São Paulo: 5000

Minas Gerais: 3200

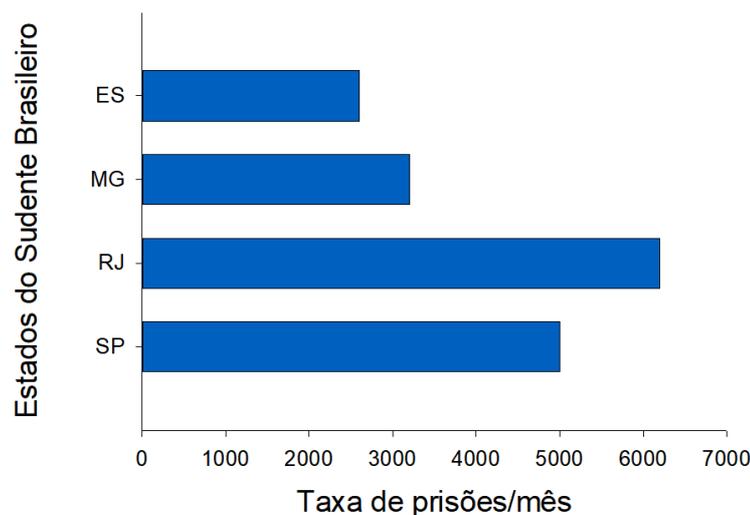
Espírito Santo: 2600

Os gráficos de colunas apresentam uma associação entre duas variáveis diferentes, geralmente se trata de uma variável qualitativa (categórica), representado por cada coluna, com uma variável quantitativa, representado pela altura das colunas. Esse gráfico não pode ser confundido com o histograma (a banca costuma muito induzir o aluno a esse erro). O detalhe para diferenciá-los está pelo fato que o gráfico de colunas não contempla a frequência de uma variável e apresenta duas variáveis em cada eixo.



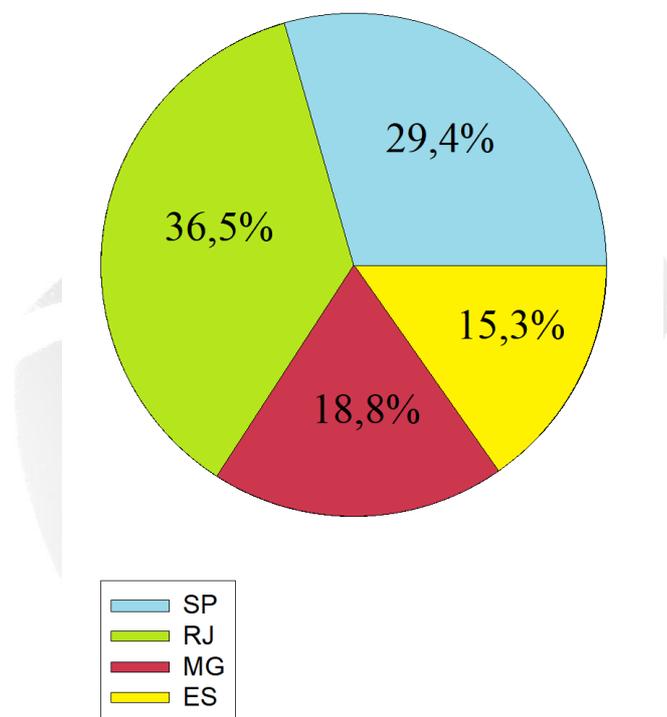
7.4 Gráfico de barras

Os gráficos de barras praticamente invertem a relação dos eixos comparado aos gráficos de colunas. Mas ainda permanece a associação de duas variáveis em que as barras representam uma variável qualitativa e o comprimento das barras representa a variável quantitativa. O mesmo exemplo abordado para o gráfico de colunas pode ser abordado no gráfico de barras.



7.5 Gráfico de Setores (Pizza)

Os gráficos de setores focam na proporção em que cada variável qualitativa evidencia comparativamente ao total observado. Desse modo, abordando o mesmo exemplo que os demais gráficos, é possível verificar que a variável quantitativa perde sua ênfase nessa representação gráfica. Para esclarecer ao que se refere a quantidade em cada classe, é necessário especificar no título junto ao gráfico que o quantitativo representa o número de prisões por mês.



Ainda, os gráficos de setores estabelecem uma relação do ângulo de cada setor com o quantitativo de cada classe. Desse modo, é possível inferir que o total de prisões por mês observado nesse exemplo corresponde a um ângulo de 360° . Assim, em simples cálculos de proporção (regra de três) é possível calcular o ângulo do setor de cada classe. Exemplo:

$$360^\circ = 100\%$$

Para o Estado de SP com 29,4%, tem-se:

$$\frac{X}{360^\circ} = \frac{29,4\%}{100\%}$$

$$360 \times 29,4 = 100X$$

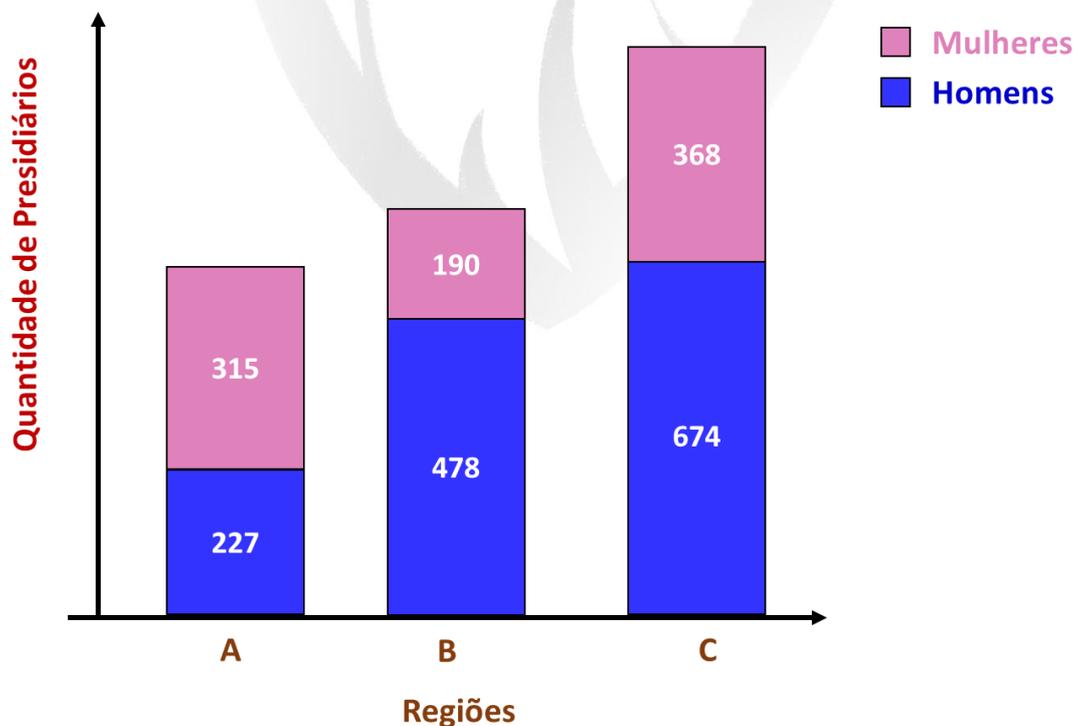
$$X = \frac{10584}{100} = 105,84^\circ$$

7.6 Gráfico de colunas múltiplas

Esse gráfico discrimina cada coluna em função de outra variável **qualitativa**. Assim, é possível representar até três variáveis nesse gráfico, uma vez que: cada coluna representa uma variável; a fragmentação da coluna representa outra variável; e altura das colunas indicam uma terceira variável no eixo da ordenada. Veja no exemplo:

Objeto de estudo:

Um estudo mostra a quantidade de presidiários homens e mulheres, em três regiões distintas do país (A, B e C).



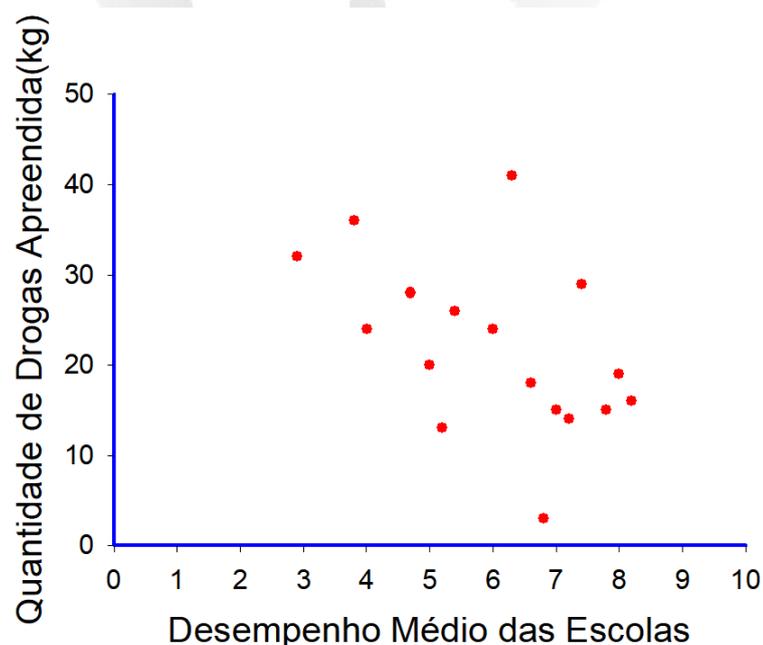
Geralmente, o gráfico vem associado com uma legenda indicando o que se refere cada partição da coluna. Esse gráfico também pode ser apresentado na forma de barras (na horizontal).

7.7 Gráfico de Dispersão

O gráfico de dispersão também é conhecido como gráfico de correlação. Isso porque é possível identificar visualmente pelo gráfico uma tendência de correlação entre as variáveis, mais a frente, no decorrer do conteúdo, iremos trabalhar profundamente com conceito de correlação. Sobretudo, entenda que o importante desse gráfico é identificar o que ocorre com valores de uma variável quando a outra variável aumenta ou diminui. Para exemplificar a aplicação desse gráfico, é necessário trabalhar com outro exemplo.

Objeto de estudo:

Uma investigação tem objetivo de estudar a associação da quantidade de drogas apreendidas, em kg, pela Polícia Civil em relação ao desempenho escolar nos municípios do estado de Mato Grosso. O desempenho escolar foi avaliado pela média das notas dos alunos de cada município.



Cada ponto presente no gráfico indica uma coordenada (associação) do valor da variável desempenho médio das escolas com a variável quantidade de drogas apreendidas. Nesse exemplo, é possível identificar uma tendência em que quanto maior o desempenho médio das escolas menor é quantidade de drogas apreendidas no município.

O gráfico de dispersão é utilizado para associar duas variáveis quantitativas, não é recomendado para variáveis qualitativa.

7.8 Gráfico de Linhas (Polígonos)

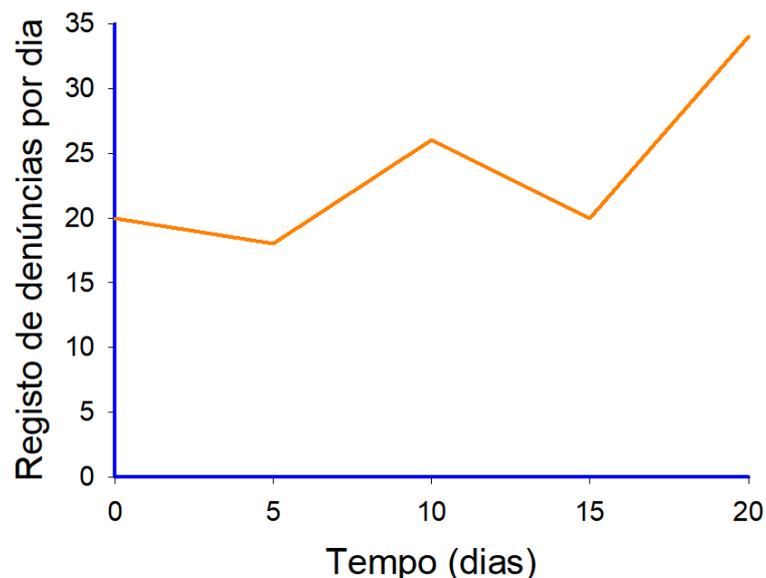
O gráfico de linhas é semelhante ao polígono de frequência, a diferença é que o gráfico associa duas variáveis diferentes e não trabalha com a frequência.

Objeto de estudo:

Registro de denúncias na delegacia Y no decorrer de 20 dias, após fatos que levaram a calamidade pública do município.

X = Tempo, em dias {0, 5, 10, 15, 20}

Y = Registro de Denúncias {20, 18, 26, 20, 34}



Basicamente, o gráfico ilustra a quantidade de denúncias no decorrer de uma série temporal. Esse gráfico é indicado para associar variáveis **quantitativas** (principalmente as contínuas) uma vez que a linha gera a ideia de transitividade e progressividade entre um valor e outro, ou seja, existem valores entre o intervalo do dia inicial da contagem (dia 0) até o 5º dia. Não seria indicado para variáveis qualitativas, uma vez que as classes são bem definidas sem transição entre um valor e outro, nesse caso um gráfico de colunas ou barras é o mais apropriado.



QUESTÕES DE RENDIMENTO

01 (CESPE | 2022 | DPE RO | ANALISTA)

Variável	Valores
Estado civil	Casado, solteiro, divorciado
Quantidade de filhos	0, 1, 2, 3...
Salário	6.510,25; 7.915,68
Idade	22, 23, 27

Com relação às variáveis apresentadas na tabela anterior, julgue os itens a seguir.

- I. A variável estado civil é qualitativa nominal.
- II. A variável quantidade de filhos é quantitativa discreta.
- III. As variáveis salário e estado civil são quantitativas discretas.
- IV. As variáveis idade e quantidade de filhos são qualitativas nominais.

Estão certos apenas os itens

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

Resolução

Vamos analisar cada variável citada na questão!

- O estado civil assume os valores {casado, solteiro, divorciado}, isto é, valores não numéricos que atribuem qualidade e não é possível estabelecer uma ordem entre eles. Portanto tem-se uma variável **Qualitativa Nominal**;
- Quantidade de filhos é tipicamente uma contagem, logo, trata-se de uma variável **Quantitativa Discreta**, pois só assume valores inteiros.

- Salário e Idade são valores quantitativos mensurados por uma convenção humana, por monetário e pela escala de tempo, podem fracionados em infinitos valores. Assim, são tipicamente variáveis **Quantitativas Contínuas**.

Com esse conhecimento vamos analisar as assertivas:

- I. A variável estado civil é qualitativa nominal. **Certo!**
- II. A variável quantidade de filhos é quantitativa discreta. **Certo!**
- III. As variáveis salário e estado civil são ~~quantitativas discretas~~.
- IV. As variáveis idade e quantidade de filhos são ~~qualitativas nominais~~.

ALTERNATIVA CERTA: LETRA A.

02 (CESPE | 2022 | DPE RO | ANALISTA)

O valor de um atributo de um dado objeto é uma medida da quantidade daquele atributo, à qual pode ser numérica ou categórica.

Nesse caso, estado civil e sexo são classificados como atributo

- a) binário.
- b) nominal.
- c) ordinal.
- d) ausente.
- e) razão.

Resolução

O estado civil pode assumir valores como {casado, separado, solteiro}, ao mesmo tempo o sexo assume valores {masculino e feminino}. Portanto, trata-se de variáveis **Qualitativas Nominais**, pois não existe nenhuma relação de ordem ou hierarquia entre esses valores.

ALTERNATIVA CERTA: LETRA B.

03 (IBFC|2022|DETRAN AM|ESTATÍSTICO)

Variável, em Estatística, pode ser definida como sendo o conjunto de resultados possíveis de um fenômeno, podendo ser qualitativa ou quantitativa. Nesse sentido, assinale a alternativa que apresenta somente variáveis contínuas.

- a) Valor numérico do salário, número de filhos, altura de uma pessoa (em metros), comprimento de um parafuso
- b) Altura de uma pessoa (em metros), total de parafusos produzidos, diâmetro de um círculo, valor numérico do salário.
- c) Peso de um pacote de farinha, comprimento de um prego (em cm), diâmetro de uma circunferência, Ph da água contida em um recipiente
- d) Comprimento de um segmento de reta, índice pluviométrico em uma região, quantidade de indústrias de bebidas, índice de inflação em um determinado mês

Resolução

As variáveis contínuas são aquelas apresentam valores numéricos com ideia de quantidade e que são mensurados por equipamentos ou por convenção humana. Pode assumir valores fracionados, e existe infinitas possibilidades entre um valor e outro. Nessa ideia, vamos identificar a questão que apresenta somente variáveis contínuas.

- a) Valor numérico do salário ✓, ~~número de filhos~~ (**Quantitativa Discreta**), altura de uma pessoa (em metros) ✓, comprimento de um parafuso ✓
- b) Altura de uma pessoa (em metros) ✓, ~~total de parafusos produzidos~~ (**Quantitativa Discreta**), diâmetro de um círculo ✓, valor numérico do salário ✓.
- c) Peso de um pacote de farinha ✓, comprimento de um prego (em cm) ✓, diâmetro de uma circunferência ✓, Ph da água contida em um recipiente ✓

- d) Comprimento de um segmento de reta ✓, índice pluviométrico em uma região ✓, ~~quantidade de indústrias de bebidas~~ (Quantitativa Discreta), índice de inflação em um determinado mês ✓

Essa questão foi mais específica quanto a algumas variáveis e como elas são medidas. O aluno precisa ter um bom conhecimento de outras disciplinas que utilizam essas variáveis. Vamos detalhar um pouco mais:

Salário é um valor monetário, quantificado pelo dinheiro e medido em escala de diferentes tipos de moeda. No caso do Brasil, em reais. Pode assumir infinitos valores.

Peso, altura, comprimento, diâmetro, são típicas variáveis quantitativas contínuas.

O índice pluviométrico consiste na altura da lâmina d'água formada após uma chuva, geralmente mensurada em milímetros. Logo, quantitativa contínua.

O pH é um índice que quantifica a acidez ou alcalinidade de uma solução, é adimensional e sua escala de valores é de 0 a 14. Contudo, pode ser particionado em infinitos valores entre 0 a 14, logo, é quantitativa contínua.

ALTERNATIVA CERTA: LETRA C.

04 (INSTITUTO AOCP | 2022 | CBM PA | SOLDADO BOMBIERO MILITAR)

De acordo com o disposto a seguir, preencha as lacunas e assinale a alternativa com a sequência **correta**.

No universo estatístico, muitas vezes, há o interesse em classificar as variáveis entre qualitativas ou quantitativas, considerando a possibilidade de subdividi-las em variáveis contínuas ou discretas. A seguir, são listadas algumas variáveis:

- Número de bombeiros em um batalhão.
- Cor predominante dos cabelos.
- Idade em anos.
- Tempo gasto entre dois referenciais.
- Gostar de azeitona.
- Altura.
- Temperatura.

Entre as variáveis citadas, _____ podem ser classificadas como variáveis contínuas e _____ como variáveis discretas.

- a) 5 / 2
- b) 2 / 5
- c) 2 / 3
- d) 3 / 2
- e) 4 / 3

Resolução

Vamos analisar cada variável e identificar seu tipo.

- Número de bombeiros em um batalhão. → **Quantitativa Discreta**
- Cor predominante dos cabelos. → **Qualitativa Nominal**
- Idade em anos. → **Quantitativa Contínua** ou **Quantitativa Discreta**
- Tempo gasto entre dois referenciais. → **Quantitativa Contínua**
- Gostar de azeitona. → **Qualitativa Nominal**
- Altura. → **Quantitativa Contínua**
- Temperatura. → **Quantitativa Contínua**

Para alcançar a alternativa correta, precisamos classificar a Idade, em anos, como variável quantitativa discreta. É uma interpretação subjetiva, se fosse usado a expressão “anos completos” ficaria mais compreensível. Isso porque a idade de uma pessoa é medida pela escala do tempo, comumente medida em anos, mas pode ser particionado em meses, dias. Exemplo claro disso são os bebês recém-nascidos em que normalmente sua idade é medida em meses. Contudo, parece ser esse o entendimento das bancas de concursos público, que idade é uma variável quantitativa discreta. Porém, essa interpretação é facilmente passível de apresentar em contraponto.

ALTERNATIVA CERTA: LETRA D.

05 (CESPE | 2015 | DEPEN | AGENTE PENITENCIÁRIO)

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

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

Certo () Errado ()

 **Resolução**

A questão tenta confundir os termos censo e amostra! O **elemento** desse objeto de estudo são os detentos de um sistema penitenciário. Eles serão entrevistados quanto a característica “possuir filhos”. A questão afirma que a amostra exige uma realização de um número maior de entrevistas comparativamente a um censo. O que é algo totalmente invertido.

O censo consiste na coleta exaustiva de dados sobre todos os elementos da população, enquanto a amostra é um subconjunto de uma parte dos elementos da população. Assim, a coleta de dados da amostra (denominado de amostragem) necessitará de **menos elementos**, isto é, uma menor quantidade de entrevistas.

ERRADA.

06 e 07 (CESPE | 2022 | FUB | ESTATÍSTICO)

A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

Disciplinas	2	3	4	5	6	7	8
Estudantes	10	15	40	35	28	10	4

Considerando essas informações, julgue o item seguinte.

Na pesquisa foram entrevistados 142 alunos.

() Certo () Errado

A proporção de alunos que cursam mais de 6 disciplinas é maior que a proporção de alunos que cursam 3 disciplinas.

() Certo () Errado

Resolução Questão 06

A questão apresenta uma **tabela de frequência** para a variável **quantidade de disciplinas cursadas por semestre**. Na tabela, para cada valor dessa variável, temos sua respectiva frequência, que corresponde a quantidade de alunos. Assim, temos as seguintes interpretações:

10 alunos cursam 2 disciplinas por semestre; 15 alunos cursam 3 disciplinas por semestre...

Dessa forma, para saber a quantidade total de alunos, basta somar todos os valores da linha estudante, isto é, somar todas as frequências absolutas. Logo:

$$\text{Qtd alunos} = 10 + 15 + 40 + 35 + 28 + 10 + 4 = 142 \text{ Alunos}$$

CERTA.

Resolução Questão 07

Cuidado com essa questão. Tem um detalhe de interpretação que é muito comum o aluno errar por uma pequena falta de atenção. Veja só:

A proporção de alunos que estudam **mais de 6** disciplinas por curso, quando se usa a expressão mais de (>), **não inclui** alunos que estudam 6 disciplinas. Você deve incluir nessa soma apenas alunos que estudam 7 e 8 disciplinas. Assim, temos:

$$f_{>6} = 10 + 4 = 14 \text{ Alunos}$$

Enquanto isso, a proporção de alunos que estudam 3 disciplinas por semestre é exatamente a frequência apresentada na tabela.

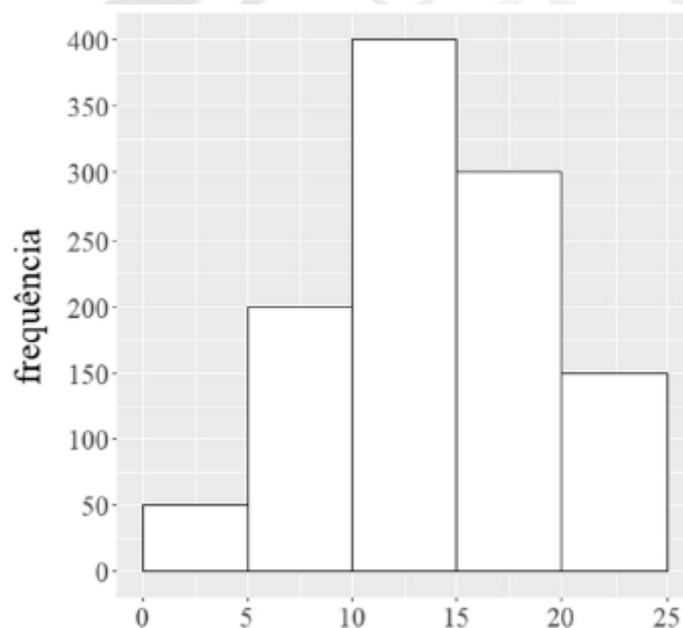
$$f_3 = 15 \text{ Alunos}$$

Logo, a questão está errada, pois a proporção de alunos que cursam mais de 6 disciplinas é **menor** que a proporção de alunos que cursam 3 disciplinas.

$$f_{>6} < f_3$$

ERRADA.

08 (CESPE | 2022 | TELEBRÁS | GESTÃO EM COMUNICAÇÃO)



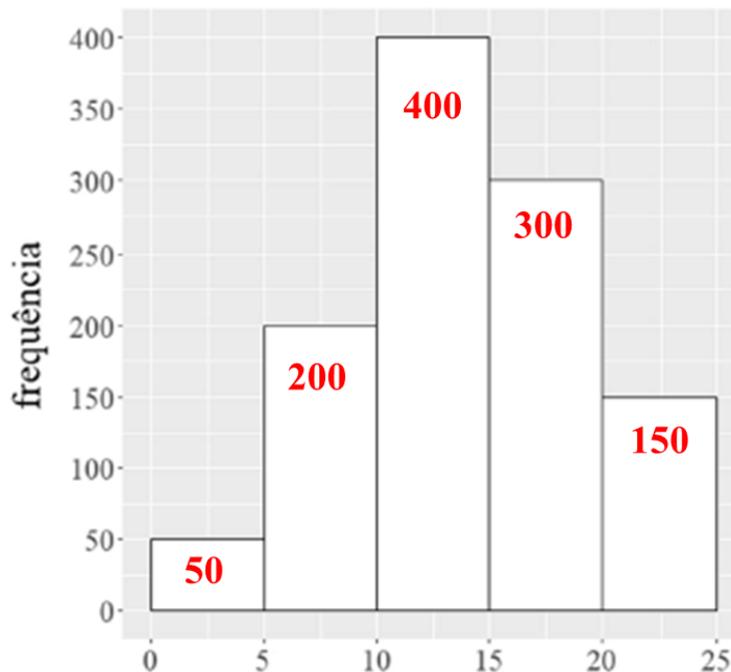
Considerando que o histograma apresentado descreve a distribuição de uma variável quantitativa X por meio de frequências absolutas, julgue o item que se segue.

O número de observações que constituem a variável X é igual a 1.000.

() Certo () Errado

 **Resolução**

Nessa questão, temos um histograma que apresenta a distribuição em frequência absoluta de uma variável X. Ao analisar a altura das colunas de cada intervalo de classe da variável X, podemos verificar os seguintes valores de frequência:



Assim, para obter o número total de observações basta somar todas as frequências absolutas. Logo:

$$n = 50 + 200 + 400 + 300 + 150 = 1100$$

Temos 1100 observações, e não 1000 como afirmado na questão.

ERRADA.

09 (IBFC|2022|EBSERH|ADMINISTRATIVO)

Na fabricação de camisas em uma indústria de confecção, a tabela a seguir indica a distribuição de frequência da variável tamanho:

Tamanho	f_i
P	3
M	4
G	7
GG	6

De acordo com os dados da tabela é correto afirmar que a diferença entre a frequência relativa da terceira classe e a frequência relativa da quarta classe é igual a:

- a) 4,5%
- b) 6%
- c) 5%
- d) 5,5%
- e) 10%

 **Resolução**

Com base na frequência absoluta (f_i) da variável tamanho da camiseta, vamos obter as frequências relativas (fr_i) para cada tamanho. Lembre-se que frequência relativa consiste na proporção da frequência absoluta sob o total de elementos $\frac{f_i}{n}$. Inicialmente vamos determinar o total de camisetas (n):

$$n = 3 + 4 + 7 + 6 = 20$$

Com isso, as frequências relativas de cada classe serão:

Tamanho	f_i	fr_i
P	3	$\frac{3}{20} = 15\%$
M	4	$\frac{4}{20} = 20\%$
G	7	$\frac{7}{20} = 35\%$
GG	6	$\frac{6}{20} = 30\%$
Total	20	100%

Por fim, a questão pede a diferença entre a frequência relativa da terceira classe (G) em relação a quarta classe (GG). Então:

$$fr_3 - fr_4 = 35\% - 30\% = 5\%$$

ALTERNATIVA CERTA: LETRA C.

10 (IADES | 2022 | UnDF | PROFESSOR)

Nº de filhos dependentes (x)	Frequência observada (f)
0	9
1	12
2	39
3	54
4	24
5 ou mais	12

Suponha que uma pesquisa para determinar o número de filhos dependentes tenha sido realizada com 150 famílias residentes em uma comunidade carente e que tenha resultado em uma distribuição conforme o quadro.

Considerando a distribuição de frequência, apresentada no quadro, na classe de maior frequência relativa (fr), a frequência acumulada (fa) em porcentagem é

- a) 100%.
- b) 76%.
- c) 36%.
- d) 12%.
- e) 8%.


Resolução

Nessa questão, temos as frequências absolutas (f_i) da variável número de filhos dependentes. A questão quer saber a frequência acumulada relativa (Fr_i em porcentagem) da classe com valor de frequência (seja absoluta ou relativa, será a mesma). Nesse sentido, vamos obter as frequências relativas e depois fazer a tabela da acumulada, que é simplesmente ir somando as frequências relativas da classe com a da classe anterior. A questão já nos forneceu o total de observações que é de 150 famílias, logo, temos que $n = 150$. Assim, vamos desenvolver toda a tabela:

Nº de filhos dependentes	f_i	fr_i	Fr_i
0	9	$\frac{9}{150} = 6\%$	6%
1	12	$\frac{12}{150} = 8\%$	6 + 8 = 14%
2	39	$\frac{39}{150} = 26\%$	14 + 26 = 40%
3	54	$\frac{54}{150} = 36\%$	40 + 36 = 76%
4	24	$\frac{24}{150} = 16\%$	76 + 16 = 92%
5 ou mais	12	$\frac{12}{150} = 8\%$	92 + 8 = 100%
Total	150	100%	-

Com o desenvolver de toda a tabela das frequências relativas, observamos que a classe com maior frequência relativa é a classe de **três filhos independentes**, com 36% de frequência relativa. Logo, a frequência relativa acumulada nessa respectiva classe é de 76%.

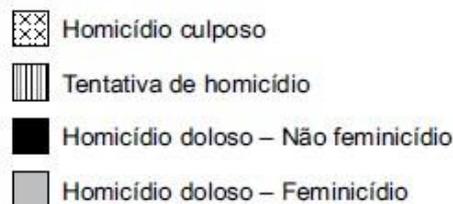
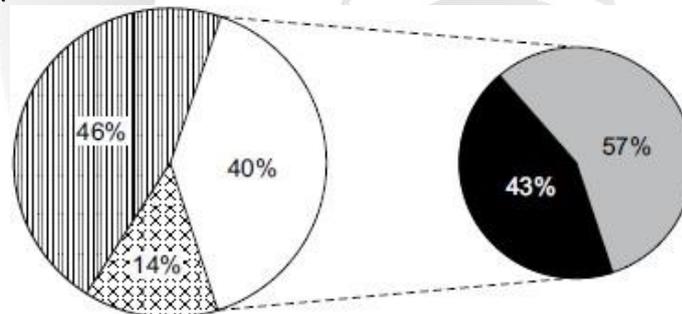
ALERTIVA CERTA: LETRA B.

11 (VUNESP | 2023 | PM-SP | SOLDADO)

O gráfico, elaborado com dados divulgados no *site* da Secretaria de Segurança Pública do Estado de São Paulo, apresenta a distribuição dos números de homicídios doloso e culposo e o número de tentativas de homicídio, associados à violência contra mulheres, no mês de outubro de 2022.

Distribuição dos números de homicídios e tentativa de homicídios

(Violência contra mulheres – Outubro de 2022)



Se adicionados os números de homicídios e de tentativas de homicídios tem-se 92 casos, então, o número de homicídios dolosos classificados como feminicídios foi maior ou igual a

- a) 14 e menor que 17.
- b) 17 e menor que 20.
- c) 26 e menor que 29.
- d) 23 e menor que 26.
- e) 20 e menor que 23.

 **Resolução**

A questão apresenta dois gráficos de setores com a finalidade de informar sobre a violência contra a mulher. Nesse sentido, o primeiro gráfico distingue os registros de homicídios nas seguintes modalidades: tentativa de homicídio (14%); homicídio culposo (46%); e homicídio doloso (40%). Vamos transformar isso em números absolutos. A questão afirma que os números de homicídios e de tentativas de homicídios somados são de 92 casos. Logo, temos que o total de casos analisados foi 92, isto é:

$$n = 92$$

Desse total, sabemos que 40% correspondem aos homicídios dolosos, então:

$$Hd = 92 \times 40\% \cong 37$$

Por fim, dessa quantidade de homicídios dolosos, sabemos que 57% correspondem a feminicídio, logo:

$$Hd_{fem} = 37 \times 57\% \cong 21$$

Então, o número de feminicídios é **maior ou igual a 20 e menor que 23**.

ALTERNATIVA CERTA: LETRA E.

12 (MARINHA|2022|MARINHA|AUXILIAR)

A representação gráfica das séries estatísticas tem por finalidade representar os resultados obtidos, permitindo que se chegue a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Assinale a opção que corresponde à representação de uma série por meio de um polígono que, para ser construído, divide-se uma circunferência em tantos arcos iguais forem os dados a representar.

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

**Resolução**

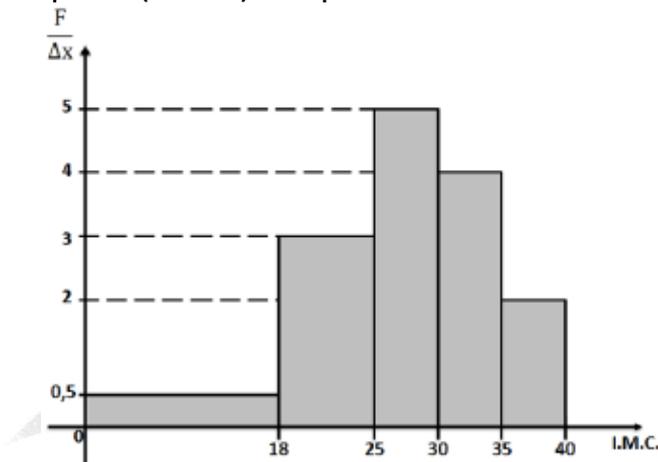
A questão apresenta sobre o gráfico de pizza ou também chamados de setores. Ele utiliza uma circunferência e a divide em tantos setores se acordo com a quantidade de classes de uma variável. É ideal para representar variáveis de classes qualitativas nominal, onde não existe nenhuma relação de ordem.

ALTERNATIVA CERTA: LETRA C.

13 (INSTITUTO AOCP|2019|IBGE|ANALISTA CENSITÁRIO)

O histograma a seguir indica o Índice de Massa Corporal de um grupo de pessoas de um bairro M de uma cidade Z, sendo que F é número de pessoas e Δx é o comprimento de cada intervalo de classe.

Índice de Massa Corporal (I.M.C.) das pessoas do bairro M da cidade Z



Com base nesse gráfico, é correto afirmar que:

- 5 pessoas desse grupo possuem o I.M.C. variando de 25 a 30.
- 2 pessoas desse grupo possuem o I.M.C. variando de 0 a 18.
- o gráfico apresenta no total a informação do I.M.C. de 95 pessoas.
- os comprimentos dos intervalos de classe são todos iguais.
- 76 pessoas desse grupo possuem o I.M.C. variando de 18 a 40.



Resolução

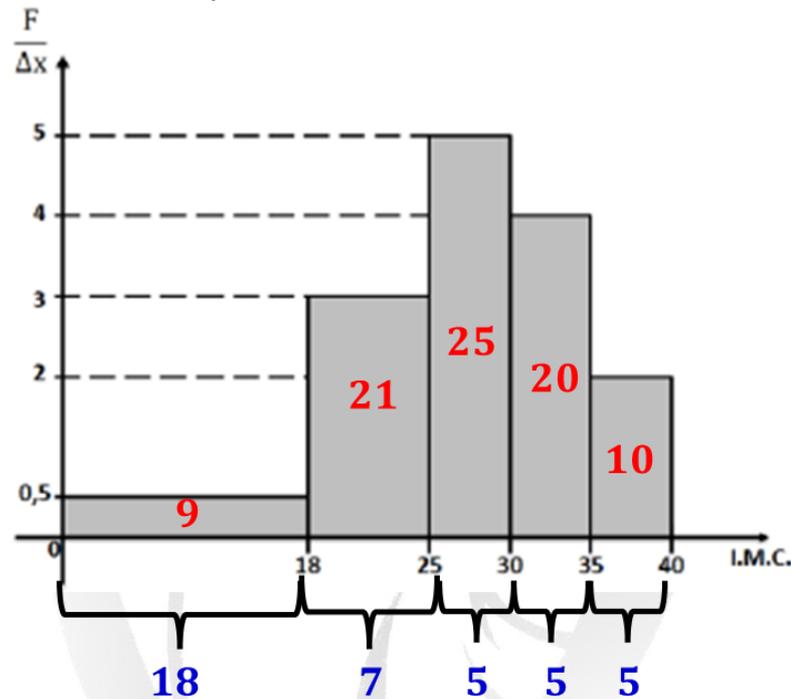
Nessa questão temos que ficar atentos, pois o histograma não possui classes equidistantes (com mesma amplitude). Nesse sentido, ao invés de utilizar a frequência, utiliza-se a **densidade de frequência**, que consiste na frequência relativizada conforme a amplitude de cada classe. Veja:

$$D_f = \frac{f_i}{\Delta X_i} = \frac{\text{Frequência absoluta}}{\text{Intervalo de valores da classe}}$$

Com isso, para responder os itens dessa questão com facilidade, vamos obter as frequências absolutas em cada classe. Para isso, basta analisar o histograma e realizar o produto da densidade de frequência com a amplitude de cada classe.

$$f_i = D_f \times \Delta X_i$$

Agora, vamos calcular a frequência de cada classe:



Agora, podemos analisar cada assertiva:

a) ~~5 pessoas~~ desse grupo possuem o I.M.C. variando de 25 a 30.

A densidade de frequência que é 5, mas na verdade, nesse grupo temos 25 pessoas.

b) ~~2 pessoas~~ desse grupo possuem o I.M.C. variando de 0 a 18.

Errado, existem 9 pessoas no grupo de 0 a 18 I.M.C.

c) o gráfico apresenta no total a informação do I.M.C. de ~~95 pessoas~~.

Ao todo, são 85 pessoas analisadas nesse estudo quanto ao I.M.C.

$$n = 9 + 21 + 25 + 20 + 10 = 85$$

d) os comprimentos dos intervalos de classe **são todos iguais**.

Errado, as classes têm intervalos diferentes. A primeira classe tem intervalo 0 a 18; a segunda de 18 a 25; e em seguida tem-se: 25 a 30; 30 a 35; 35 a 40.

e) 76 pessoas desse grupo possuem o I.M.C. variando de 18 a 40. ✓

Ao somar todas as frequências no intervalo de 18 a 40, temos 76 pessoas.

$$n_{18 \leq x \leq 40} = 21 + 25 + 20 + 10 = 76$$

ALTERNATIVA CERTA: LETRA E.

14 (CESPE | 2018 | POLÍCIA FEDERAL | PERITO CRIMINAL)

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

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

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

Certo () Errado ()



Resolução

A tabela em questão **não evidencia** uma distribuição de frequência da quantidade de drogas. Na verdade, essa forma de apresentação de dados consiste em **dados brutos**, isto é, as observações da variável quantidade de drogas apreendidas, em kg, estão dispostas conforme os dados foram coletados diariamente. Assim, a coleta do 1º dia

corresponde a 10 kg, 2º dia corresponde a 22 kg e assim por diante. A questão tenta confundir o candidato com tabela de frequência (dados ponderados). Porém, veja que **não foi contabilizado suas repetições**. Os dados da variável X são associados com o dia coletado e não a sua frequência. Caso os dados fossem apresentados em tabela de frequência, a observação de 22 kg/dia seria exposta com uma frequência de 2 repetições. Os dados apresentados em tabela de frequências seria a seguinte:

Drogas Apreendidas	Frequência
10	1
18	1
22	2
28	1

ERRADA.



CONCURSEIRO QUE PRETENDE SER POLICIAL NÃO FAZ RATEIO

Todo o material desta apostila (textos e imagens) está protegido por direitos autorais do Profissão Policial Concursos de acordo com a Lei 9.610/1998. Será proibida toda forma de cópia, plágio, reprodução ou qualquer outra forma de uso, não autorizada expressamente, seja ela onerosa ou não, sujeitando-se o transgressor às penalidades previstas civil e criminalmente.