

Introdução

- Aplicação da Estatística
- Conceitos básicos
- Síntese histórica
- Classificação de variáveis
- Método científico e a Estatística

Parte de perguntas/desafios do mundo real

- Um político quer saber qual é o percentual de eleitores que pretende votar nele nas próximas eleições.
- Cientistas querem verificar se uma nova vacina contra a dengue faz efeito.
- Uma montadora de automóveis quer verificar a qualidade de um lote inteiro de peças fornecidas através de uma pequena amostra.

Matéria-prima da Estatística → variabilidade

É “difícil” encontrar duas coisas exatamente iguais

Objetivo da Estatística → A Estatística desenvolve métodos para descobrir e expor os padrões de comportamento (regularidade) que estão escondidos nos dados.

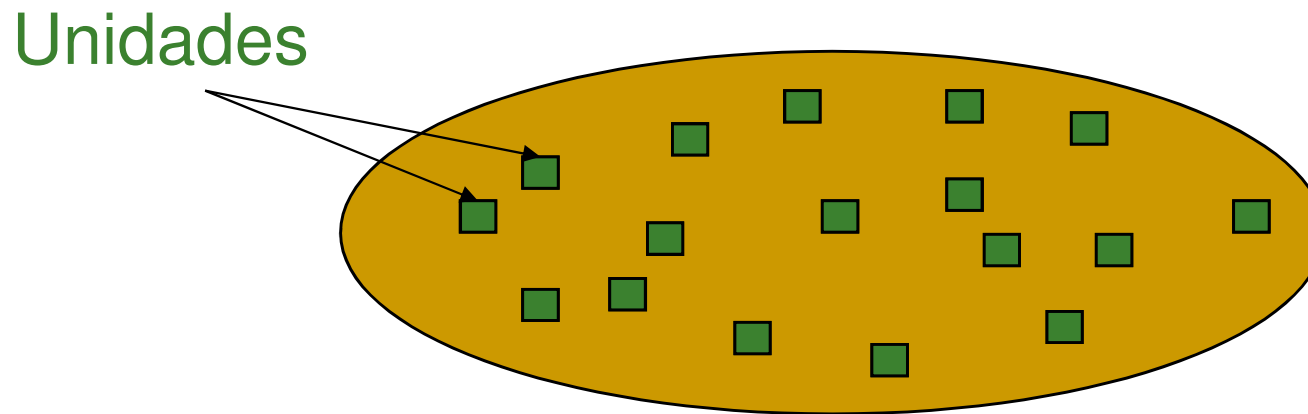
Definição de Estatística

- A **estatística** engloba um conjunto de métodos científicos para:
 - coleta, organização, resumo e análise de dados.
 - obtenção de conclusões (suporte à tomada de decisão)

Alguns conceitos básicos:

- **População**
- **Amostra**
- **Amostragem**

População: é o conjunto de todas as unidades (elementos) de interesse que têm pelo menos uma característica em comum. Pode ser finita ou infinita.



Exemplo: Pesquisas eleitorais no RS

Característica definidora → votar no RS

População → conjunto de todos os eleitores que votam no RS

Unidade → o eleitor

Amostra é parte de uma população, convenientemente escolhida, que tem a finalidade de representá-la.

Deve apresentar as mesmas características da população.

Amostragem é a metodologia de obtenção das amostras.

Exemplos

- Antes da **eleição** diversos órgãos de pesquisa e imprensa ouvem um conjunto selecionado de eleitores para ter uma ideia do desempenho dos vários candidatos nas futuras eleições.
- Uma empresa metal-mecânica toma uma amostra do produto fabricado em intervalos de tempo especificados para verificar se o **processo está sob controle** e evitar a fabricação de itens defeituosos.
- O IBGE faz **levantamentos periódicos** sobre emprego, desemprego, inflação, etc.
- Redes de rádio e TV se utilizam constantemente dos **índices de popularidade** dos programas para fixar valores da propaganda ou então modificar ou eliminar programas com audiência insatisfatória.

Por que fazer amostragem?

- população infinita
- diminuir custo
- aumentar velocidade na caracterização (medidas que variam no tempo)
- minimizar perdas por medidas destrutivas

Quanto amostrar? depende:

- da variabilidade original dos dados (maior variabilidade → maior n)
- da precisão requerida na estimação (maior precisão → maior n)
- do tempo disponível (menor o tempo → menor n)
- do custo da amostragem (maior o custo → menor n)

Amostragem

- Probabilística:** Todos os elementos da população têm probabilidade conhecida e diferente de zero de participar da amostra. A realização deste tipo de amostragem só é possível se a população for finita e totalmente acessível.
- Não probabilística:** Presença dos elementos na amostra deve-se a outros critérios. Por exemplo, quando somos obrigados a colher a amostra na parte da população a que temos acesso.

A amostragem **probabilística** é a mais recomendável porque garante a **imparcialidade** da amostra.

Amostragem probabilística:

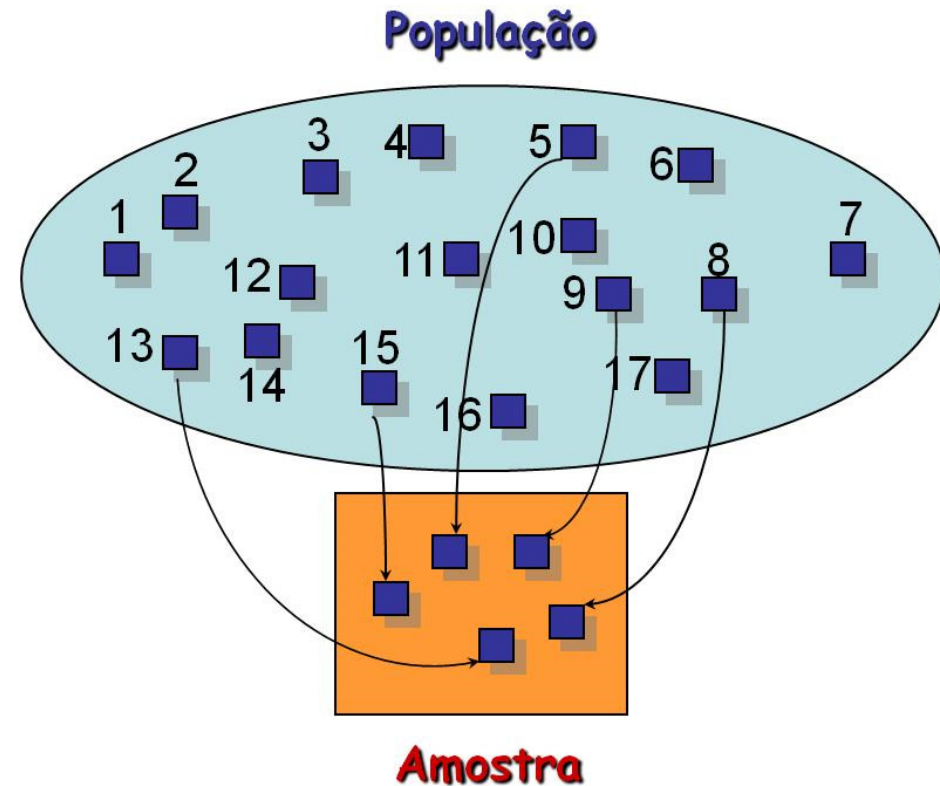
- Amostragem aleatória simples
- Amostragem aleatória estratificada
- Amostragem aleatória por conglomerados
- Amostragem aleatória sistemática

Amostragem não probabilística:

- Amostragem intencional
- Amostragem por quota
- Amostragem a esmo

Amostragem probabilística – aleatória simples

- É equivalente a um sorteio de loteria;
- Considera a população homogênea;
- Cada elemento da população tem a mesma oportunidade de ser escolhido;
- Utilizam-se números aleatórios, programas computacionais, calculadoras, bolinhas numeradas, etc.



Exemplo: Imagine que você queira amostrar um número de pessoas que estão fazendo um determinado concurso com N inscritos. Devemos enumerar cada um dos N candidatos e sortear n deles.

Amostragem probabilística – **estratificada**

- Quando a variável de interesse apresenta heterogeneidade na população e esta heterogeneidade permite a identificação de grupos homogêneos, divide-se a população em grupos (estratos) e faz-se uma amostragem dentro de cada estrato, garantindo, assim, a representatividade de cada estrato na amostra.



Exemplo: Podemos verificar que pesquisas eleitorais apresentam uma grande heterogeneidade em relação à intenção de votos, quando consideramos, por exemplo, a faixa salarial ou o nível de escolaridade.

Amostragem probabilística – por conglomerados

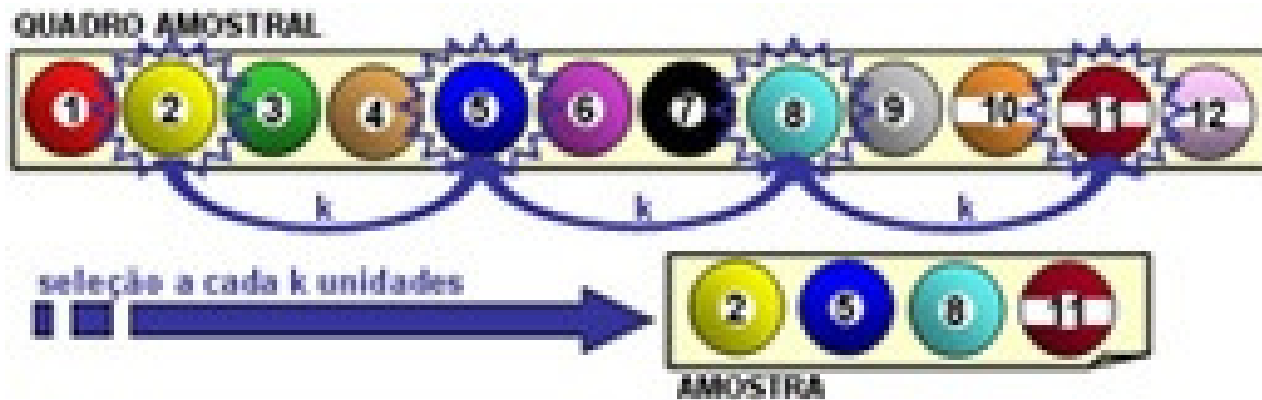
- A população já é dividida em diferentes conglomerados (grupos), extraíndo-se uma amostra apenas dos conglomerados selecionados, e não de toda a população. O ideal é que cada conglomerado represente tanto quanto possível o total da população.
- Os conglomerados são definidos em função da experiência do pesquisador. Geralmente, podemos definir os conglomerados por fatores geográficos, como por exemplo, bairros e quarteirões.



Exemplo: este tipo de amostragem é muito útil quando a população é grande, por exemplo, no caso de uma pesquisa em nível nacional.

Amostragem probabilística – sistemática

- Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, temos uma amostragem sistemática.
- É de fundamental importância que a variável de interesse não apresente ciclos de variação coincidente com os ciclos de retirada, pois este fato tornará a amostragem não aleatória.



Exemplo: em uma linha de produção, podemos, a cada dez itens produzidos, retirar um para avaliar a qualidade da produção.

Amostragem não probabilística – **intencional**

- A amostra pesquisada muitas vezes está disponível no local e no momento onde a pesquisa está sendo realizada.
- A seleção das unidades amostrais é deixada a cargo do pesquisador. Com base em seu julgamento, o pesquisador seleciona os elementos que julga mais representativos da população.

Exemplos:

- Para saber a preferência por determinado cosmético, o pesquisador entrevista os frequentadores de um grande salão de beleza.
- Para saber a aceitação em relação a uma nova marca de whisky a ser inserida no mercado, farão parte da amostra pessoas que façam uso da bebida e que tenham condições financeiras de comprar esta nova marca.

Amostragem não probabilística – **por quota**

- A população é dividida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, e sim os elementos são procurados até que a cota de cada grupo seja cumprida.
- As quotas asseguram que a composição da amostra seja a mesma da população com relação às características escolhidas.

Exemplo: Em pesquisas eleitorais, a divisão de uma população em grupos (ex. sexo, escolaridade, idade e renda) pode servir de base para a definição dos grupos, partindo da suposição de que estas variáveis definem grupos com comportamentos diferenciados no processo eleitoral. Para saber o tamanho destes grupos, pode-se recorrer a pesquisas feitas anteriormente pelo IBGE.

Amostragem não probabilística – a esmo

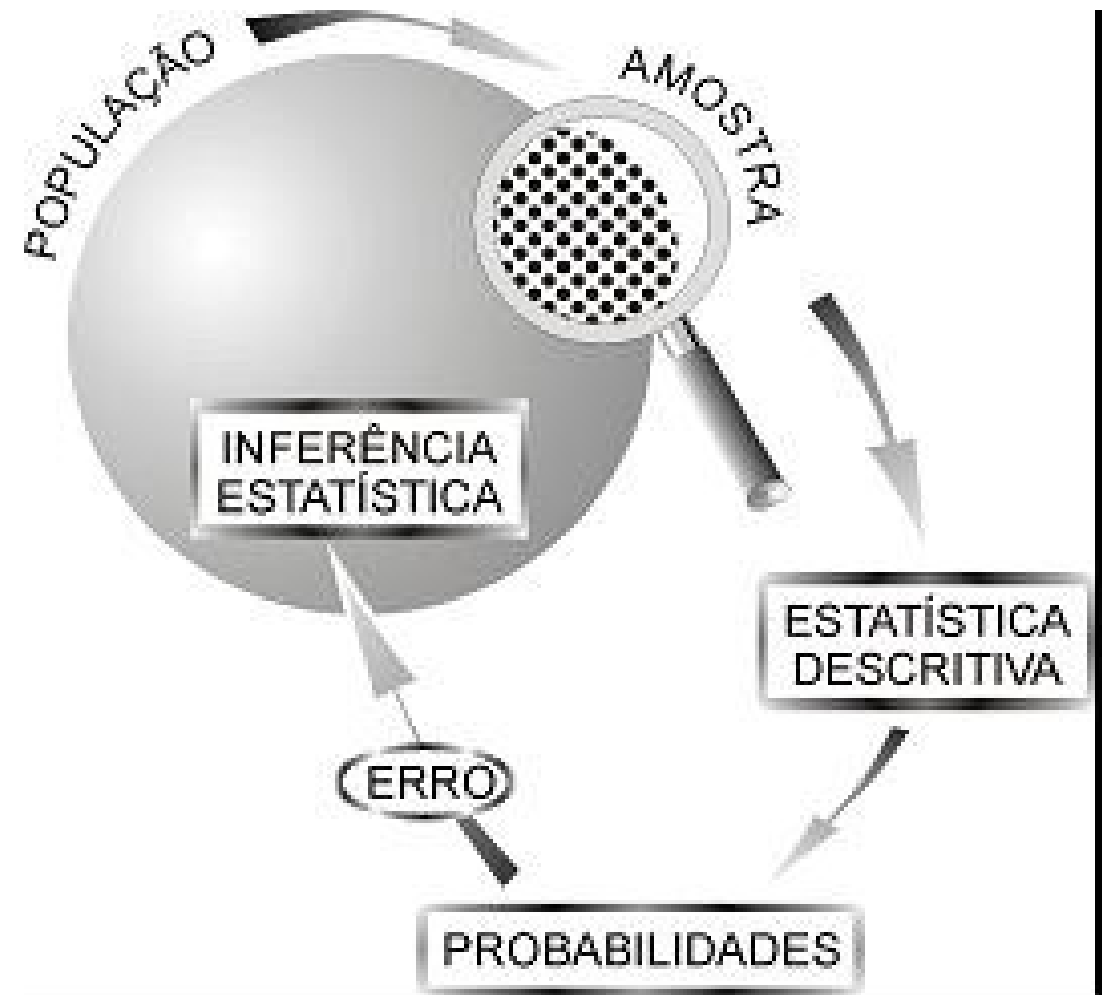
- É a amostragem em que o pesquisador, para simplificar o processo, procura ser aleatório sem, no entanto, realizar propriamente o sorteio usando algum dispositivo aleatório confiável.
- Os resultados da amostragem a esmo são, em geral, equivalentes aos da amostragem probabilística se não existir a possibilidade de o pesquisador ser inconscientemente influenciado por alguma característica da população.

Exemplo: Se desejarmos retirar uma amostra de 100 parafusos de uma caixa contendo 10.000, evidentemente não faremos uma AAS, pois seria muito trabalhosa a tarefa de enumeração destes parafusos. Então retiramos simplesmente a esmo (sem regra) nossa amostra.

- **Riscos da amostragem:** o processo de amostragem envolve riscos, pois toma-se decisões sobre toda a população com base em apenas uma parte dela.
- A **teoria da probabilidade** pode ser utilizada para fornecer uma ideia do risco envolvido, ou seja, do erro que se comete ao utilizar uma amostra ao invés de toda a população.

Divisão da Estatística

- **Descritiva**
- **Inferência**



Estatística Descritiva

- É a parte da estatística que cuida:
 - da apresentação de dados através de **tabelas e gráficos**
 - do resumo ou descrição de dados através de **medidas descritivas**
- Em geral, não tem por objetivos tirar conclusões.



Inferência Estatística

- Métodos que propiciem a realização das inferências sobre populações a partir de amostras delas retiradas, tendo por base o cálculo das probabilidades
 - **estimação de parâmetros**
 - **testes de hipóteses**
 - **modelos de regressão**

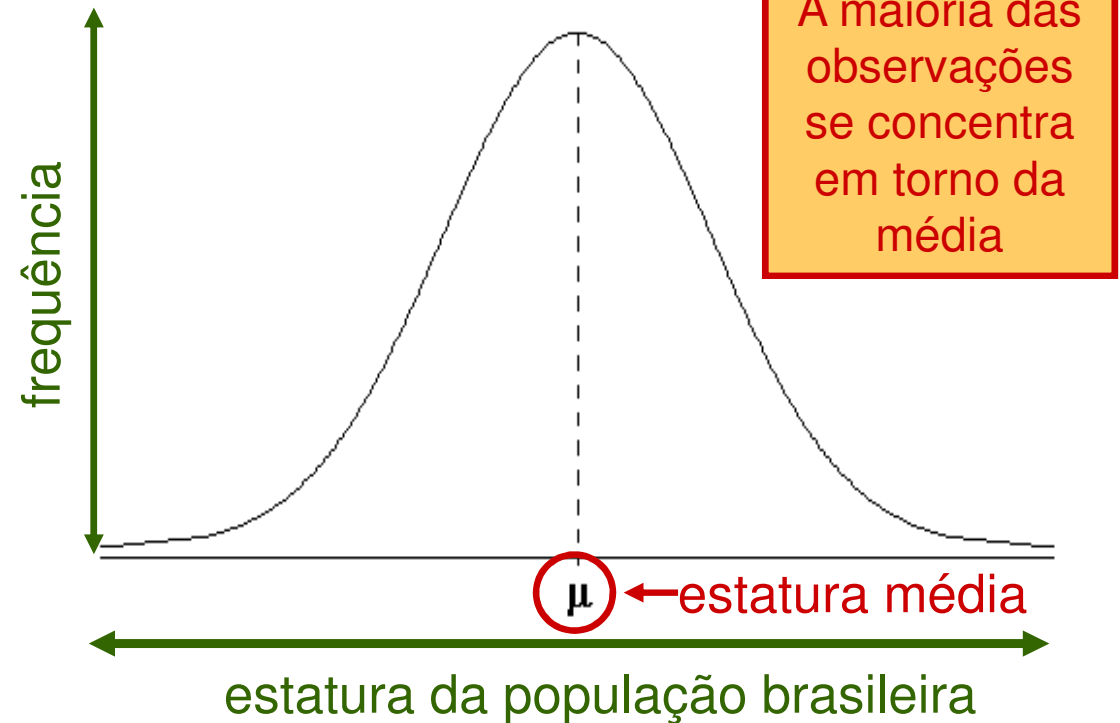


Informações históricas

- ⇒ Existem indícios de que há mais de 2000 anos a.C. já se faziam **censos** na Babilônia, na China e no Egito.
- ⇒ O objetivo do censo era saber o número de **pessoas disponíveis para fazer a guerra** e para a **cobrança de impostos**.
- ⇒ A Estatística teve origem na necessidade do Estado Político conhecer os seus domínios.
- ⇒ Sob a palavra Estatística, provavelmente derivada da palavra “**status**” (estado, em latim), acumularam-se descrições e **dados relativos ao Estado**. Nas mãos dos governantes, a Estatística passou a constituir-se verdadeira ferramenta administrativa.

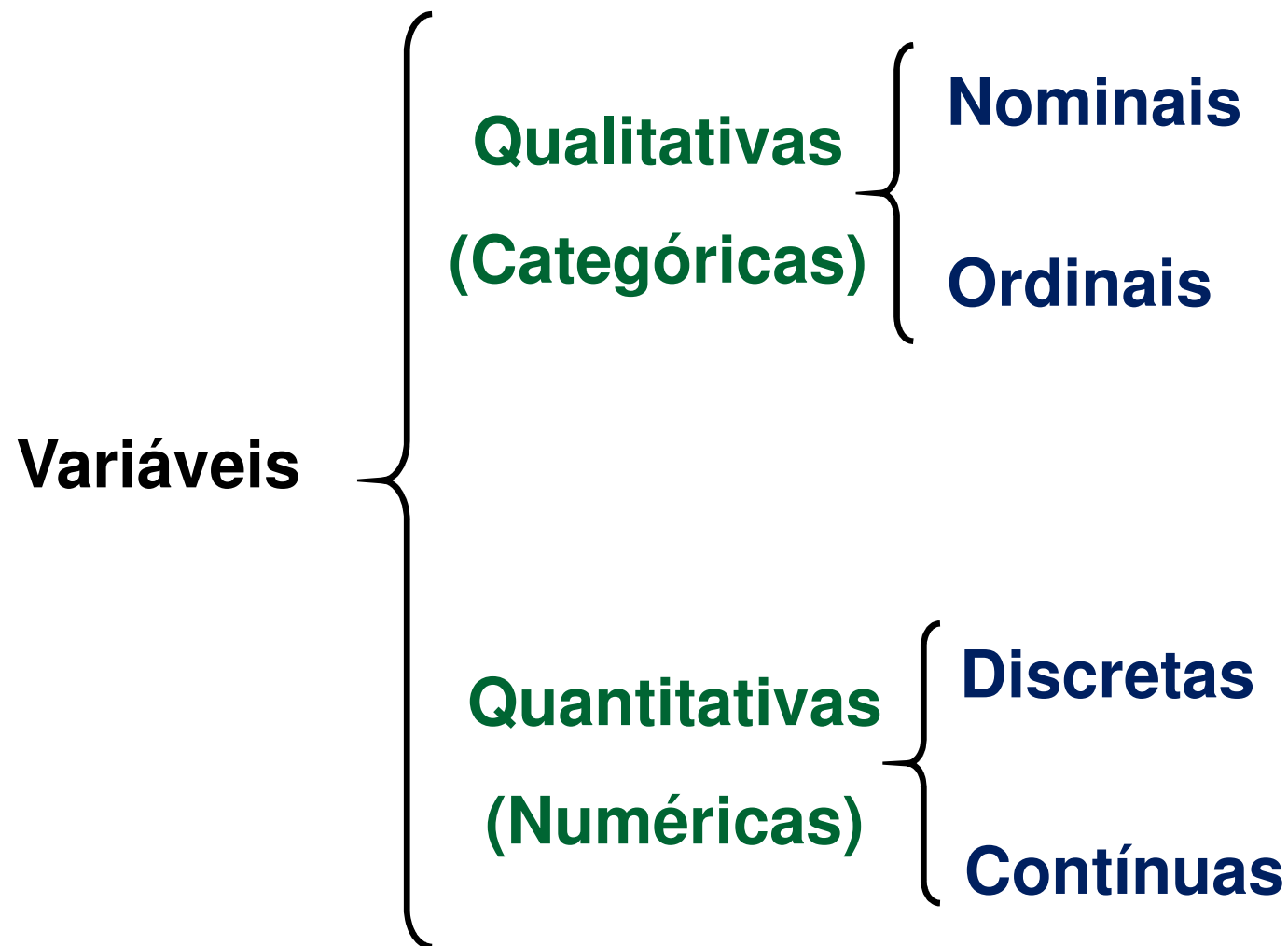


⇒ Um grande avanço na teoria das probabilidades se deu, no início do século XIX, através dos estudos de Laplace e Gauss que descreveram a famosa **Distribuição Normal**



- ⇒ A partir das **distribuições de probabilidade** foi possível a criação de técnicas de amostragem mais adequadas e de formas de relacionar as amostras com as populações.
- ⇒ Outro marco decisivo no desenvolvimento da Estatística foi o advento da **computação**, ferramenta que permitiu que a Estatística ampliasse seus horizontes.

Classificação de variáveis



Variáveis qualitativas (categóricas) → descrevem qualidades (categorias ou classes)

Nominais → não há um sentido de ordem entre seus níveis

Exemplos: sexo (masculino e feminino)
profissão (engenheiro, professor, médico, etc.)
região geográfica (norte, sul, sudeste, etc.)

Ordinais → há um sentido de ordem entre seus níveis

Exemplos: faixas de idade (criança, adolescente, adulto, idoso)
intensidade de cor (claro, médio, escuro)
nível de instrução (fundamental, médio, superior)

Variáveis quantitativas (numéricas) → seus valores são números reais (observados)

Discretas

- descrevem dados discretos ou de enumeração (geralmente obtidos por **processo de contagem**)
- assumem valores inteiros não negativos (0, 1, 2, 3, ...)

Exemplos: número de carros sinistrados

número de pacientes que se recuperam

número de filhos de um casal

Contínuas

- descrevem dados contínuos ou de mensuração (geralmente obtidos por **processo de medição**)
- podem assumir qualquer valor dos reais (-10, 0, π)

Exemplos: peso, altura, tempo, velocidade, temperatura

Porque classificar as variáveis

Metodologia
Estatística



A complexidade e a
informação
aumentam.

É de fundamental importância saber classificar corretamente uma variável porque esta discriminação é que irá indicar a possibilidade e a forma de utilização dos procedimentos estatísticos disponíveis.

Exercício proposto: Classifique as variáveis abaixo.

- a) Religião
- b) N° de vendas diárias de uma empresa
- c) Distância entre duas cidades
- d) Consumo mensal de energia elétrica
- e) Estado civil
- f) Satisfação com o salário
- g) Temperatura de uma mistura
- h) Estado de nascimento
- i) Idade
- j) N° de transações financeiras
- k) Velocidade de um carro
- l) Postos em um exército
- m) N° de peças com defeito em um lote
- n) Altura de uma pessoa
- o) Classe social
- p) Valor de venda diária de uma empresa

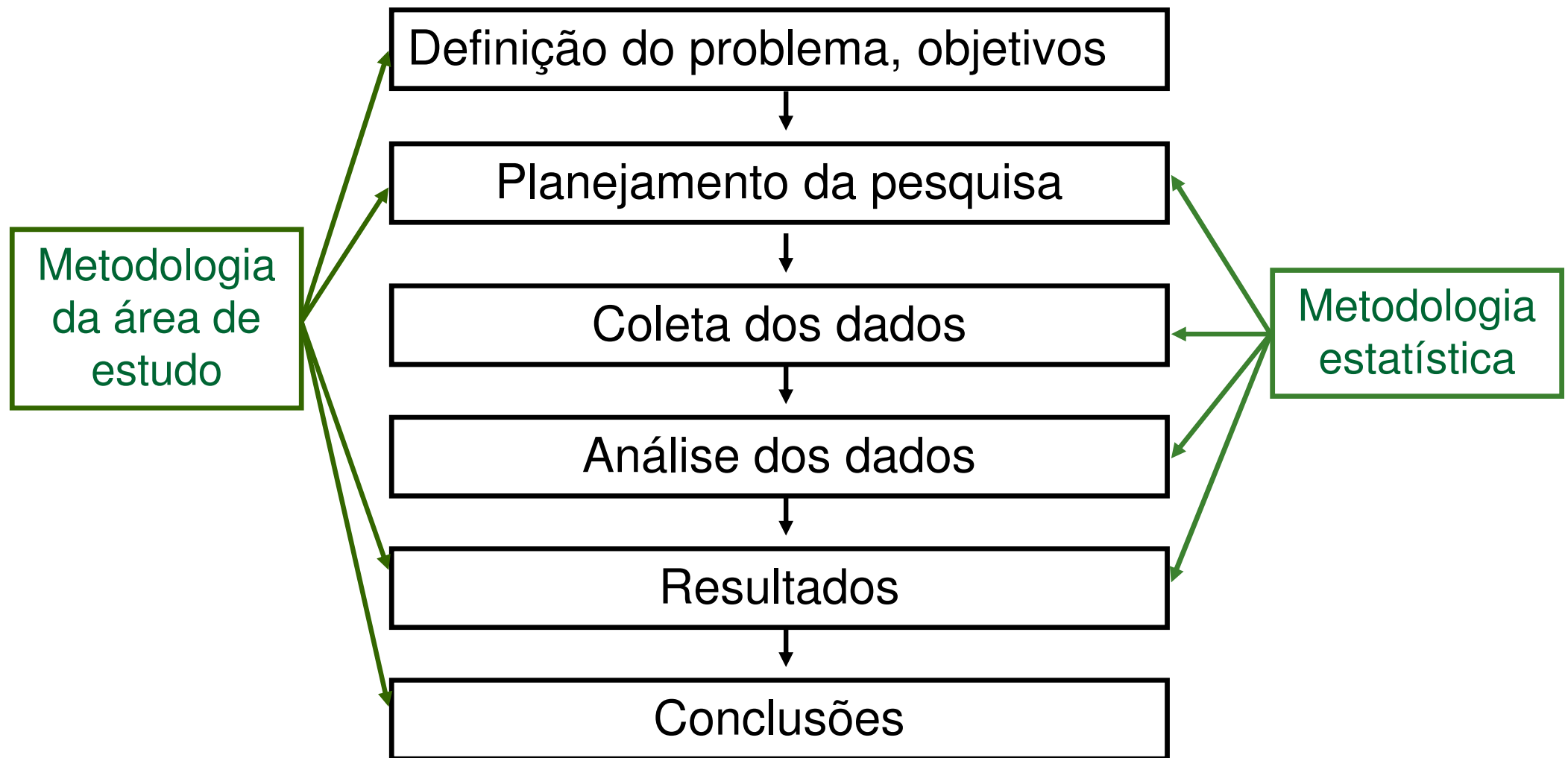
Categórica {
Nominal
Ordinal

Numérica {
Discreta
Contínua

A estatística na pesquisa científica

- Nas pesquisas científicas precisamos coletar dados que possam fornecer informações que respondam nossas indagações
- Para que os resultados da pesquisa sejam confiáveis, tanto a coleta como a análise dos dados devem ser feitas de forma criteriosa e objetiva
- A metodologia estatística deve ser aplicada nas diversas etapas da pesquisa

Principais etapas da pesquisa científica



Arredondamento

➤ Quando o primeiro algarismo a ser abandonado for 0, 1, 2, 3 ou 4, **fica inalterado** o último algarismo a permanecer.

Exemplo: 48,23 → 48,2

➤ Quando o primeiro algarismo a ser abandonado for 5, 6, 7, 8 ou 9, **aumenta-se** de uma unidade o último algarismo a permanecer.

Exemplo: 23,87 → 23,9