

Elementos de Estatística

Prof. Dr. Clécio da Silva Ferreira
Departamento de Estatística - UFJF

O que é a estatística ?

Para muitos, a estatística não passa de conjuntos de tabelas de dados numéricos. Os estatísticos são pessoas que coletam esses dados.

- A estatística originou-se com a coleta e construção de tabelas de dados para os governos
- A situação evoluiu e esta coleta de dados representa somente um dos aspectos da estatística.

Definição de Estatística

A Estatística é um método científico que tem por objetivo o estudo de uma população a partir de informações sobre a mesma. É um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

População

- **Dois significados:**
- - 1º: refere-se a um conjunto de objetos que tem em comum uma característica denominada *variável* que pode ser classificada, contada ou medida.
 - - **Exemplo:** suponha que um pesquisador deseje estudar as condições dos pacientes de certo hospital num determinado ano.
 - - O conjunto de pacientes do referido hospital no ano considerado constitui a população em estudo e as variáveis podem ser: sexo (masculino ou feminino), classe sócio-econômica (classe baixa, classe média, classe alta), número de consultas e peso.

População

- 2º: refere-se ao conjunto de dados de uma variável em estudo.
- Exemplo: se forem observadas os valores do peso dos paciente acima, obtém-se um conjunto de números que constitui a população de dados do peso dos mesmos.

População

- O número de objetos (ou de dados) de uma população é o tamanho da mesma, sendo geralmente representado por N .
- uma população pode ser finita ou infinita. A população finita possui um número limitado de objetos (ou de dados). Por exemplo, a população definida acima é finita.
- No caso de uma população infinita, o número de objetos (ou de dados) é ilimitado.
 - A título de exemplo, o conjunto de pacientes de um hospital ao longo de um período indeterminado é uma população infinita.

AMOSTRA

- É qualquer parte ou subconjunto de uma população.
- Não se estuda uma população em sua totalidade, pois a população pode ser infinita ou, mesmo sendo finita, é muito grande.
 - Assim sendo, o pesquisador observa apenas uma parte da população, denominada **amostra**.
- Formalmente, denomina-se amostra como um conjunto de objetos (ou de dados) retirados de uma população para fins de estudo da mesma.
 - O número de objetos (ou de dados) de uma amostra é o tamanho da mesma e é representado geralmente por **n** . Uma amostra deve ser **representativa** da população da qual foi extraída.

AMOSTRA

- O termo amostra refere-se também ao conjunto de dados de uma variável numa amostra de objetos.
 - Exemplo: se para estudar o peso dos pacientes da população, o pesquisador observa o peso de 100 pacientes, estes constituem uma amostra de pacientes desta população e o conjunto de dados referentes ao peso destes pacientes constitui uma amostra de tamanho 100 de dados do peso dos pacientes da população.

Censo ou Recenseamento e Amostragem

- Existem dois procedimentos para se obter os dados de uma população: o censo e a amostragem.
 - O censo consiste em obter os dados de uma ou mais variáveis em **todos** os objetos de uma população.
 - pouco utilizado devido ao custo operacional, exceto no caso de populações relativamente pequenas.
 - A amostragem consiste em obter os dados de uma ou mais variáveis numa amostra de objetos retirados aleatoriamente da população.
 - freqüentemente utilizada porque as populações geralmente são muito grandes ou infinitas.
 - os dados são obtidos com rapidez e o custo operacional é bem menor do que no caso do censo.

ÁREAS DA ESTATÍSTICA

■ ESTATÍSTICA DESCRITIVA

- descreve-se os dados das variáveis de uma amostra ou uma população.
- consiste na organização, resumo, análise e interpretação dos dados.

■ INFERÊNCIA ESTATÍSTICA

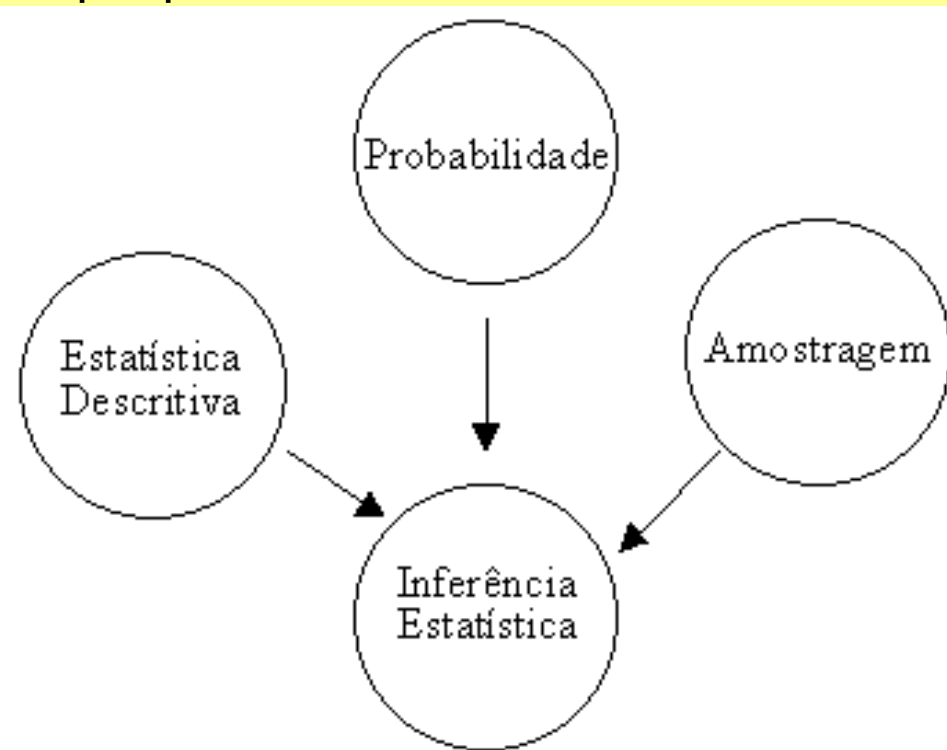
- consiste na análise de uma ou mais variáveis de uma população a partir dos dados de uma amostra extraída desta população.

■ PROBABILIDADE

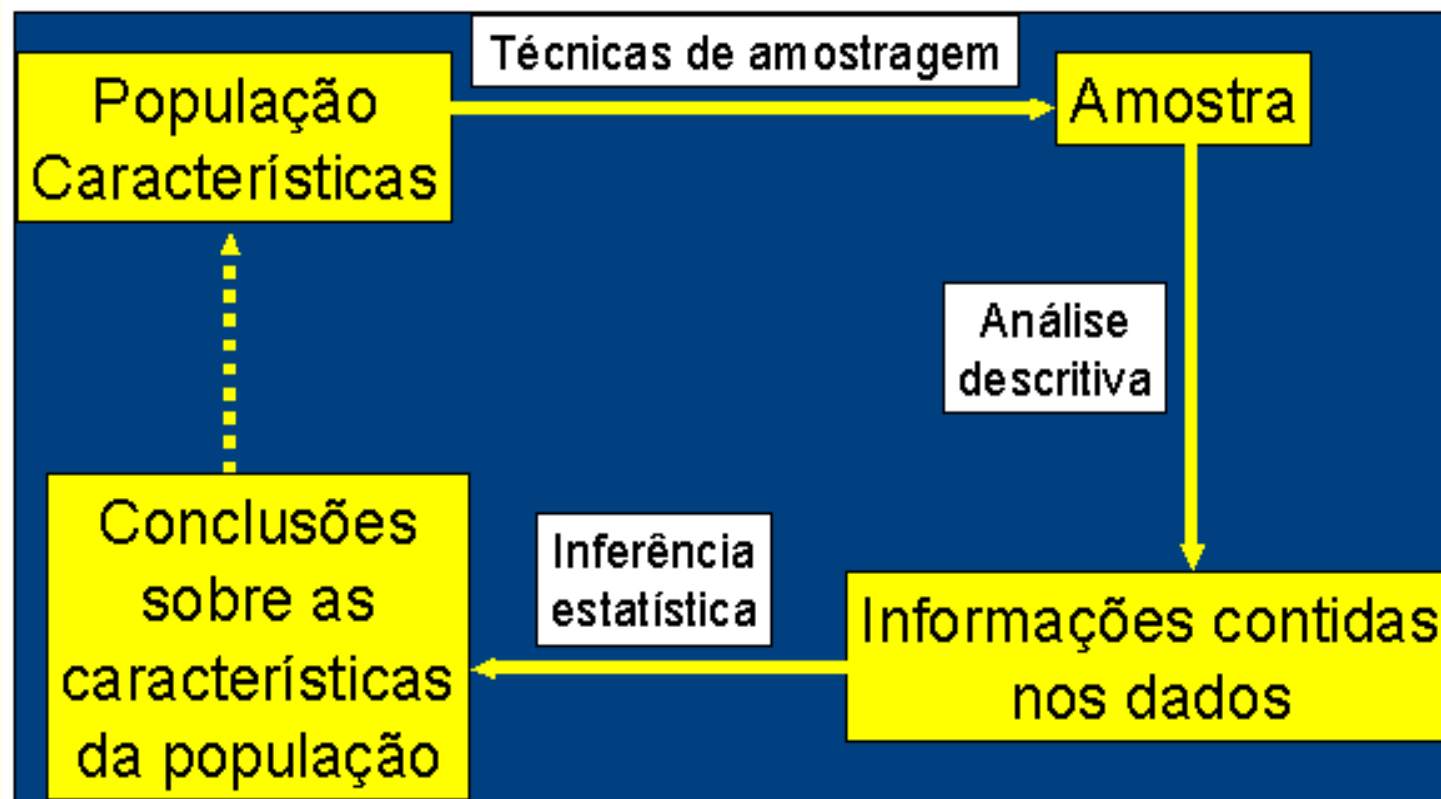
- consiste na medida da incerteza através de números ou funções matemáticas sendo utilizada na inferência estatística.

ÁREAS DA ESTATÍSTICA

- constitui um conjunto de procedimentos para a obtenção de uma amostra representativa da população alvo e no estudo das relações existentes entre uma população e as amostras possíveis que podem ser extraídas da mesma.



Etapas da Análise Estatística



Áreas da Estatística

- 1- Estatística Descritiva
- 2- Probabilidade
- 3- Inferência estatística

ESTATÍSTICA DESCRITIVA

A estatística descritiva é a etapa inicial da análise utilizada para descrever e resumir os dados. A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou esta área da estatística.

PROBABILIDADE

A teoria de probabilidades nos permite descrever os fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza.

INFERÊNCIA ESTATÍSTICA

É o estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir da amostra.

Exemplos

- **Chove ou não chove amanhã;**
- **Número de acidentes que ocorreu numa interseção;**
- **Número de pacientes curados que foram submetidos a um tratamento;**
- **Tempo que decorre até uma lâmpada queimar-se;**
- **Valor das ações na bolsa de valores no fim de um pregão.**

Tipos de Variável

- Dependendo da natureza dos dados (números ou atributos) uma variável pode ser **quantitativa** ou **qualitativa**.
- Se os dados são **números**, a variável é **quantitativa**, podendo ser **discreta** ou **contínua**.
 - na quantitativa discreta, os dados são obtidos por um processo de enumeração ou contagem e como consequência, o número de dados num intervalo finito é finito.
 - exemplo, a variável número de consultas é quantitativa discreta.
 - na quantitativa contínua os dados são obtidos por um processo de medição, sendo portanto números reais. O número de dados num intervalo finito é infinito.
 - a variável peso é quantitativa contínua.

Fases de um trabalho estatístico

■ DEFINIÇÃO DO PROBLEMA

- pesquisador define o problema a ser resolvido, escolhendo as variáveis do seu interesse, planejando a operacionalização do trabalho a ser desenvolvido, formas de obtenção dos dados e outros procedimentos.

■ COLETA DOS DADOS

- podem ser obtidos de duas formas: censo ou amostragem. A escolha da forma de obtenção dos dados depende do tamanho da população e do tempo disponível para a realização do trabalho.

■ ORGANIZAÇÃO DOS DADOS

- Após a coleta, os dados passam por um trabalho de depuração, isto é, observa-se a ocorrência de erros, omissões de dados e outras impropriedades que podem ocorrer durante a obtenção dos dados e podem comprometer os resultados da análise dos mesmos.
 - Em seguida os dados são classificados em tabelas com o objetivo de facilitar a análise dos mesmos.

Fases de um trabalho estatístico

■ APRESENTAÇÃO DOS DADOS

- podem ser apresentados em forma tabular, isto é, por meio de tabelas, e por meio de gráficos.
 - As tabelas são empregadas para apresentar os valores exatos dos dados, enquanto que os gráficos são utilizados para se ter a visão global dos dados.

■ ANÁLISE DOS DADOS

- tarefa fundamental num trabalho estatístico, podendo ser um simples exame de uma tabela ou de um gráfico ou o emprego de recursos sofisticados da matemática.
- A partir da análise obtém-se as informações necessárias à conclusão do trabalho.

■ CONCLUSÃO

- As informações obtidas na análise fornecem ao pesquisador meios de inferir sobre o comportamento dos dados, como médias, variações, tendências, índices, prevalências, etc.
- As conclusões são apresentadas na forma de um relatório.

AMOSTRAGEM

■ Uma área importante em muitas aplicações Estatísticas é a da **Tecnologia de Amostragem**.

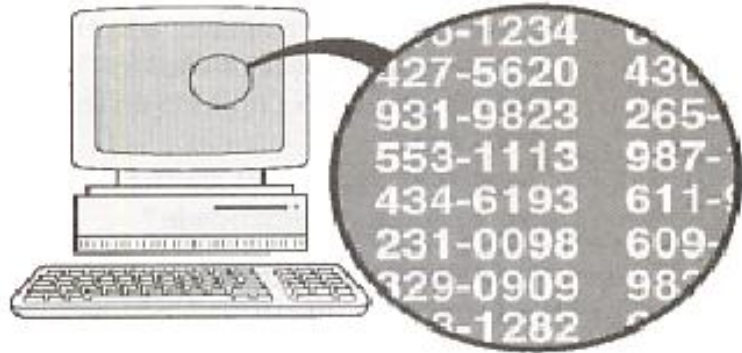
Exemplos de Aplicação:

- Pesquisa de mercado
- Pesquisa de opinião
- Avaliação do processo de produção
- Praticamente em todo experimento.

TÉCNICAS DE AMOSTRAGEM

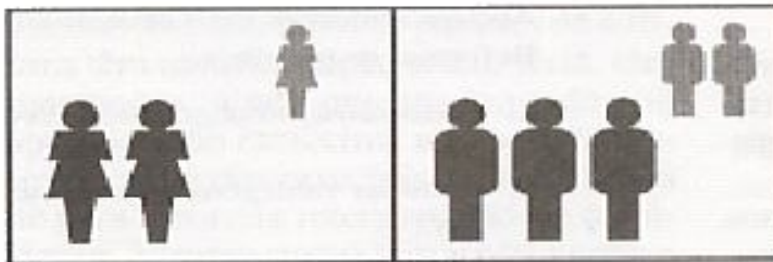
- amostras extraídas devem **representativas** de uma população, isto é, a **variabilidade** da população deve estar presente nas amostras.
 - objetos da população que irão constituir uma amostra devem ser escolhidos aleatoriamente.
- A amostragem pode ser com reposição ou sem reposição.
 - Na **amostragem com reposição** um objeto pode ser considerado mais de uma vez numa amostra. Neste caso, número de amostras possíveis de uma população de tamanho N é N^n .
 - Na **amostragem sem reposição** um objeto é considerado somente uma vez na amostra. O número de amostras possíveis sem reposição de uma população de tamanho N é
- $$A_N^n = \frac{N!}{(N-n)!}$$
- A amostragem pode ser **probabilística** ou **não probabilística**.
 - Na amostragem probabilística pode-se determinar a probabilidade de escolher um objeto da população enquanto que na **amostragem não probabilística** esta probabilidade não pode ser calculada.

Tipos de Amostragem



Amostragem Aleatória

Cada elemento da população tem a mesma chance de ser escolhido.



Amostragem Estratificada

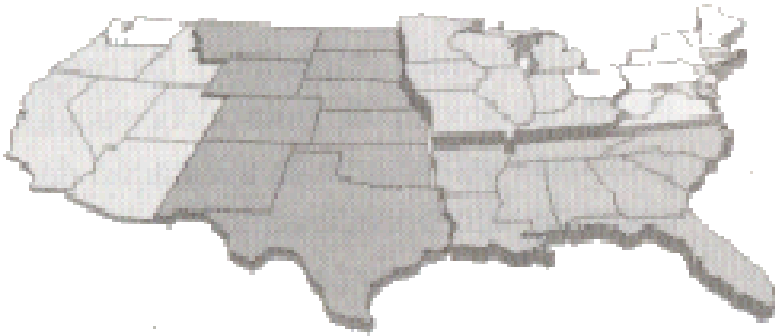
Classificar a população em, ao menos dois estratos e extrair uma amostra de cada um.



Amostragem Sistemática

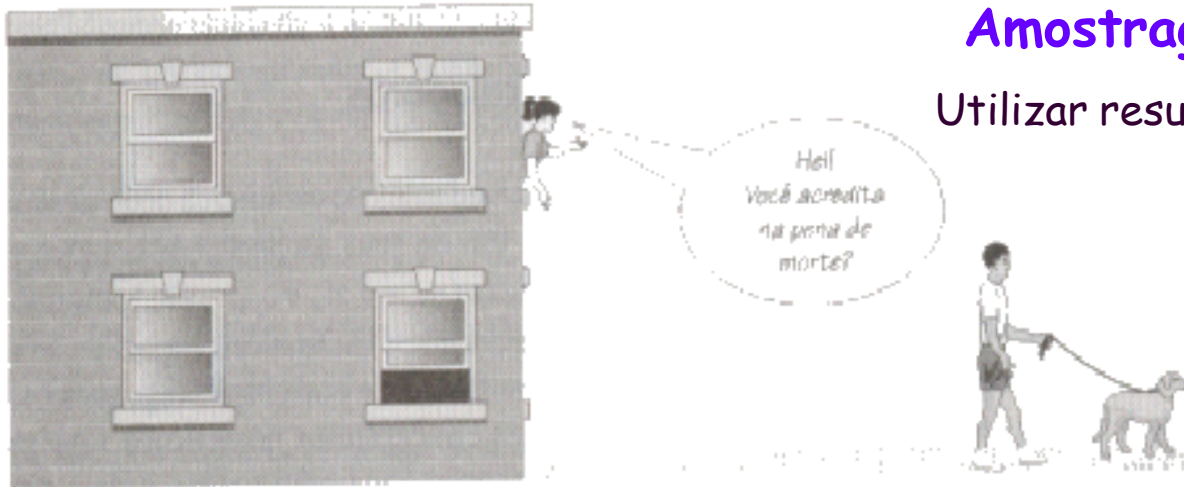
Escolher cada elemento de ordem k .

Tipos de Amostragem



Amostragem por Conglomerados

Dividir em seções a área populacional, selecionar aleatoriamente algumas dessas seções e tomar todos os elementos das mesmas (ou tomar uma amostra aleatória de área sorteada).



Amostragem de Conveniência

Utilizar resultados de fácil acesso.

Amostragem Aleatória Simples

- utilizada quando a população é finita e não é demasiadamente grande. Admite-se também que a população seja homogênea.
- Consiste em enumerar os N objetos de uma população e em seguida extrair uma amostra de n objetos da mesma utilizando bolas numeradas, fichas numeradas, números aleatórios gerados por programas de computador ou tabela de dígitos aleatórios.
- *Exemplo:* Retirar, dentre todos os alunos da disciplina, uma amostra aleatória simples de tamanho $n=10$.

AMOSTRAGEM SISTEMÁTICA

- uma variação da AAS. Consiste em determinar inicialmente a razão $N/n = r$, denominado intervalo de amostragem.
 - O número do primeiro objeto é escolhido entre 0 e $r-1$ (se os objetos na população estiverem numerados de 0 a $N-1$) ou entre 1 e r (se os objetos na população estiverem numerados de 1 a N).
 - Se a divisão não for exata arredonda-se o resultado obtido para o inteiro mais próximo.
- *muito útil quando se deseja estudar uma população que vai se completando ao longo do tempo (exemplos?).*
 - *para calcular o intervalo de amostragem considera-se o número de objetos da população no momento de se extrair a amostra.*
- *Exemplo:* Extraia uma amostra de 50 usuários na fila do RU, supondo que na hora do almoço passem 2.000 pessoas.

AMOSTRAGEM POR CONGLOMERADOS

- Denomina-se conglomerados um conjunto de objetos de uma população.
 - Por exemplo se deseja determinar a proporção de pessoas hipertensas em certo município, pode-se considerar como conglomerados os bairros deste município, as ruas, os quarteirões ou as residências.
 - Neste caso escolhe-se alguns conglomerados e as pessoas destes constituirão a amostra desejada.
- A amostragem por conglomerados tem a seguinte característica: a variação dos dados é grande dentro de cada conglomerado e pequena entre os conglomerados.
- **Exemplos de conglomerados:**
 - **Brasil -> estados -> municípios -> setores -> cidadãos**
 - **Escolas -> turmas -> alunos**

AMOSTRAGEM ESTRATIFICADA

- Quando a população é **heterogênea**, pode-se dividi-la em subpopulações ou estratos homogêneos e aplica-se um processo de amostragem aleatória simples em cada estrato.
 - Por exemplo, os pacientes de um hospital podem ser classificados segundo o sexo, faixa etária e classe sócio-econômica.
 - Um caso importante da amostragem estratificada é aquele em que o pesquisador obtém uma amostra com a mesma proporcionalidade dos estratos que compõem a população.
 - Neste caso a amostragem é denominada **amostragem estratificada com partilha proporcional**.

AMOSTRAGEM ESTRATIFICADA

- se uma população com N objetos é dividida em k estratos, $1, 2, \dots, k$, com N_1, N_2, \dots, N_k objetos, respectivamente, tem-se que
- $\frac{n_i}{n} = \frac{N_i}{N}$ ($i = 1, 2, \dots, k$)
- o número de objetos do i -ésimo estrato que ocorrerá na amostra pelo processo de amostragem estratificada com partilha proporcional é
- $n_i = n \frac{N_i}{N}$ $i = 1, 2, \dots, k$
- **A amostragem estratificada tem a seguinte característica: a variação dos dados é pequena dentro de cada estrato e grande entre os estratos.**
- *Exemplo:* Numa região existem 150.000 pessoas sendo 45.000 com idade inferior 20 anos, 75.000 com idade entre 20 anos (inclusive) e 50 anos (exclusive) e 30.000 com idade igual ou superior a 50 anos. Extrair uma amostra de 30 pessoas pelo processo de amostragem estratificada com partilha proporcional. **FAZER NO QUADRO!**

AMOSTRAGEM NÃO PROBABILÍSTICA

- objetos são escolhidos de acordo com o julgamento pessoal do observador e não na probabilidade de escolher os objetos da população.
 - AMOSTRAGEM ACIDENTAL
 - utilizado em algumas pesquisas de opinião. A amostra é formada por todas as pessoas que vão aparecendo até o tamanho da mesma.
 - Por exemplo, se um pesquisador deseja obter uma amostra de tamanho 30, o mesmo entrevista as primeiras 30 pessoas que aparecerem.
 - AMOSTRAGEM INTENCIONAL
 - amostra é escolhida intencionalmente com base em determinados critérios.
 - Por exemplo, numa pesquisa sobre uma doença infectocontagiosa escolhe-se as pessoas expostas ao risco de tal doença.

AMOSTRAGEM EM MÚLTIPLOS ESTÁGIOS

- Um procedimento de amostragem pode ser realizado em várias etapas e neste caso tem-se uma amostragem em múltiplos estágios.
- O objetivo é combinar os diversos tipos de amostragens utilizando as vantagens de cada tipo.
- Numa amostragem em dois estágios, por exemplo, pode-se no primeiro estágio empregar a amostragem por conglomerados e num segundo estágio empregar a amostragem aleatória simples.