



Engenharia de Dados Experimentais

Conceitos Introdutórios

Raquel Guiné

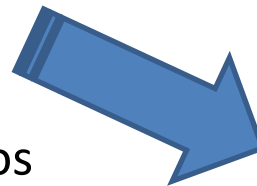


Conteúdo

1. Dados experimentais – como obter e o que fazer?
2. Colheita de dados
3. Tratamento de dados
4. Análise dos dados
5. Previsão
6. Interação entre estatística e pesquisa científica
7. Unidade estatística ou elemento estatístico
8. Variáveis
9. Amostras

1. DADOS EXPERIMENTAIS – COMO OBTER E O QUE FAZER?

- Planeamento da colheita de dados (Delineamento experimental)
- Colheita dos dados
- Classificação e organização dos dados
- Análise dos dados
- Interpretação dos resultados



Fase descritiva
Fase analítica
Fase de previsão



2. COLHEITA DE DADOS

a) Classificação segundo a forma como se realiza a observação dos factos

Observação direta

- Quando a observação incide propositadamente sobre os dados para o trabalho a realizar
- Em Ciências Naturais ou de Engenharia (trabalho de natureza experimental) fazem-se análises laboratoriais, por exemplo
- Em Ciências Sociais, faz-se aplicação de questionário



Observação indireta

- Quando a recolha de dados estatísticos acerca de um determinado fenómeno se baseia em informações constantes de registos já existentes e na divulgação de organismos oficiais (ex. dados do INE, FAO, outro tipo de registos, etc...)

b) Classificação segundo a forma como no tempo é efetuada a observação

Permanente ou contínua

- Observação efetuada através de registos permanentemente atualizados (ex. registos de nascimentos, de óbitos, entrada e saída de doentes, etc.)

Periódica

- Recolha de dados efetuada com periodicidade (ex. balanço comercial de uma empresa; recenseamento decenal de um país, etc.)



Ocasional

- Recolha de dados efetuada em função das necessidades que se manifestam num determinado momento (ex. sondagem sobre hábitos tabágicos, programas de TV preferidos - estudo transversal ...)

3. TRATAMENTO DE DADOS

a) Crítica de dados:

- Fase de correção de erros.
- Procede-se à retificação e depuração de dados de modo a reduzir ao mínimo os erros, a fim de não alterarem os resultados finais da análise

p. ex.

- ✓ Se uma variável tiver 10% de não-respostas deve ser eliminada
- ✓ Alguém a preencher os dados de altura (que está em metros) escreveu 175 (porque considerou cm) em vez de 1,75



b) Codificação de dados:

- No caso de dados recolhidos por questionário, os dados devem ser codificados, traduzir numericamente uma informação de carácter qualitativo.

Exemplo:

1. Idade: _____ anos

2. Género

Feminino ☐₁ Masculino ☐₂

3. Habilitações literárias

Primário ☐₁ Básico (2.º ciclo) ☐₂ Básico (3.º ciclo) ☐₃
Secundário ☐₄ CET ☐₅ Superior ☐₆

4. Estado civil

Solteiro (a) ☐₁ Casado (a)/União de facto ☐₂ Divorciado/Separado (a) ☐₃ Viúvo (a) ☐₄

5. Profissão

Estudante ☐₁ Funcionário público(a) ☐₂ Empregado(a) por conta de outrem ☐₃
Profissão liberal ☐₄ Empresário(a) ☐₅ Desempregado(a) ☐₆ Outro(a) ☐₇
Qual? _____ (5.a)

- No caso de dados recolhidos por observação experimental há também por vezes necessidade de codificar as amostras.

Exemplo:

Descrição da amostra	Codificação da amostra
Pera secada a 60 °C em estufa convectiva	P60EC
Maçã secada a 50 °C em secador de túnel	M50ST
Pera secada ao sol	PSol

As codificações são importantes para depois apresentar os dados na forma gráfica, já que não seria possível incluir neles a descrição completa da amostra.

c) Redução de dados:

- Os dados apresentam-se por vezes em grande n.º e torna-se impossível a sua compreensão (particularmente para dados recolhidos por questionário)
- É necessário condensá-los, por exemplo, numa tabela de distribuição para os tornar mais compreensíveis e calcular as frequências absolutas, relativas, etc.

Exemplo:

Alturas	Frequências		Relativa Percentual
	Absoluta	Relativa	
1,69 \mapsto 1,74	6	$6/20 = 0,30$	30%
1,74 \mapsto 1,79	3	$3/20 = 0,15$	15%
1,79 \mapsto 1,84	2	$2/20 = 0,10$	10%
1,84 \mapsto 1,89	4	$4/20 = 0,20$	20%
1,89 \mapsto 1,94	5	$5/20 = 0,25$	25%
Total	20		100%

4. ANÁLISE DOS DADOS

Nesta fase é possível fazer comparações de dados e investigar o grau de dependência entre eles

Compreende:

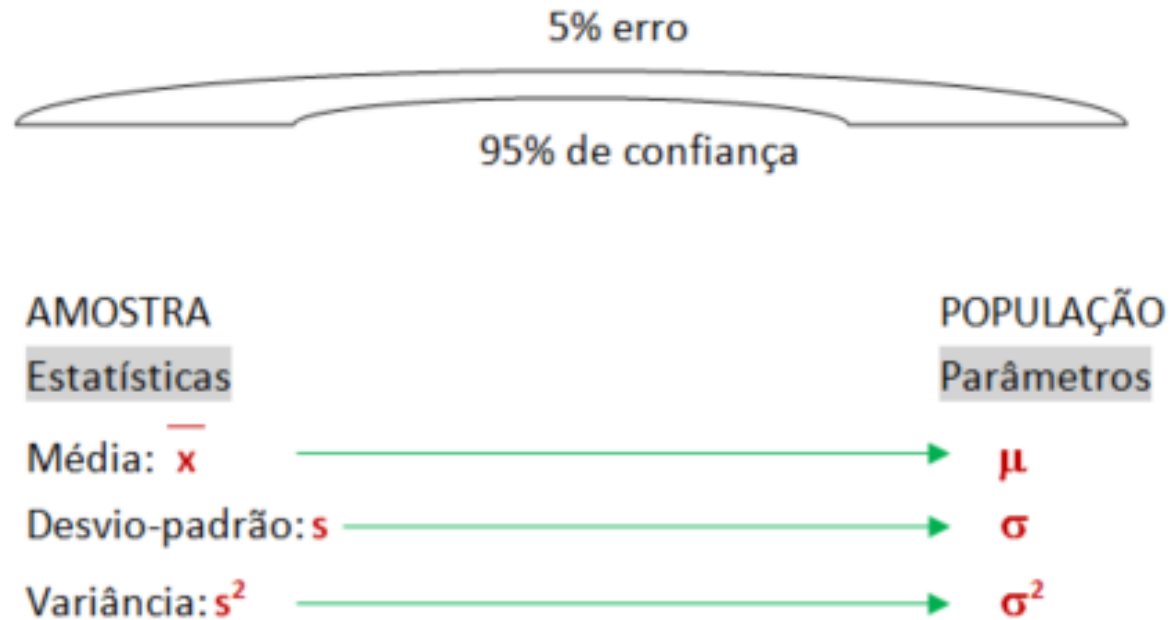
- a) Parametrização de dados
- b) Testes de hipóteses
- c) Outras medidas ou inferências



a) Parametrização de dados

- Da “massa” inicial de dados extrai-se um pequeno conjunto de medidas, denominados por parâmetros, que caracterizam e descrevem muito sinteticamente a forma como os dados se distribuem

- Medidas estatísticas mais comuns:
 - ✓ Medidas de tendência central
 - ✓ Medidas de dispersão
 - ✓ Medidas de forma



b) Testes de hipóteses

- Permite a inferência, estudar a relação entre variáveis
- Testamos sempre a hipótese H_0 (na amostra)

H_0 : Não há relação entre as variáveis

ou

H_0 : não há diferenças entre as variáveis

- Se aceitarmos a H_0



Não há diferenças entre as variáveis

- Se rejeitarmos a H_0



Aceita-se a H_1 ou hipótese de investigação



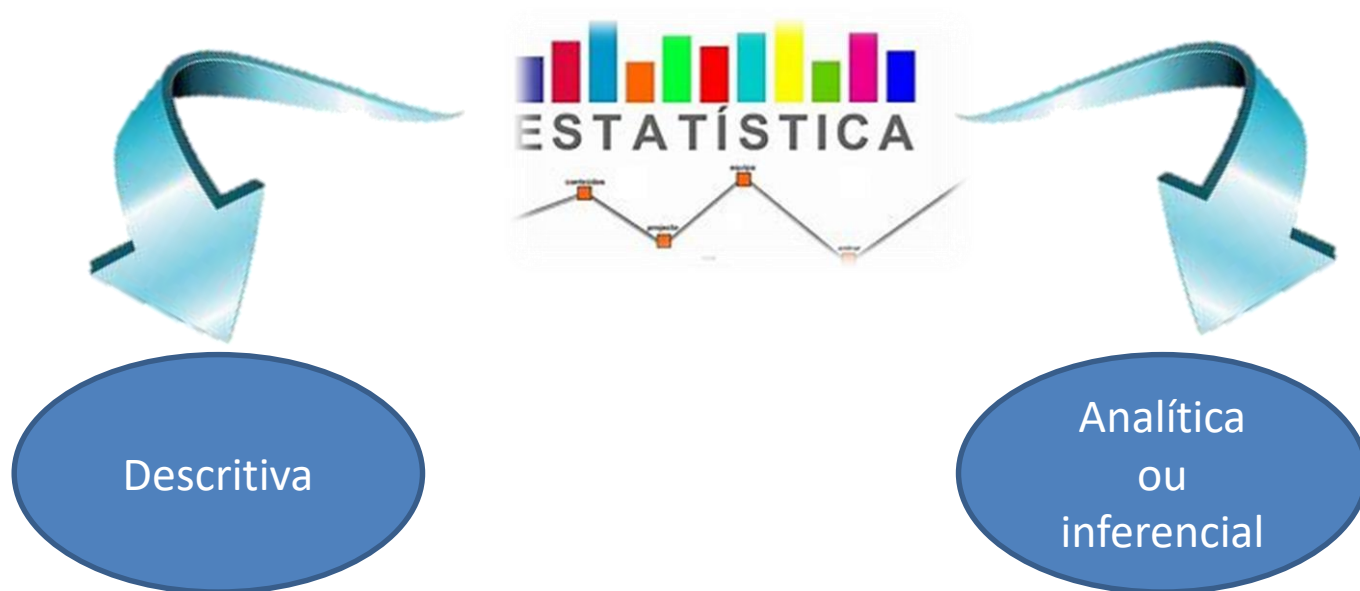
Há diferenças entre as variáveis



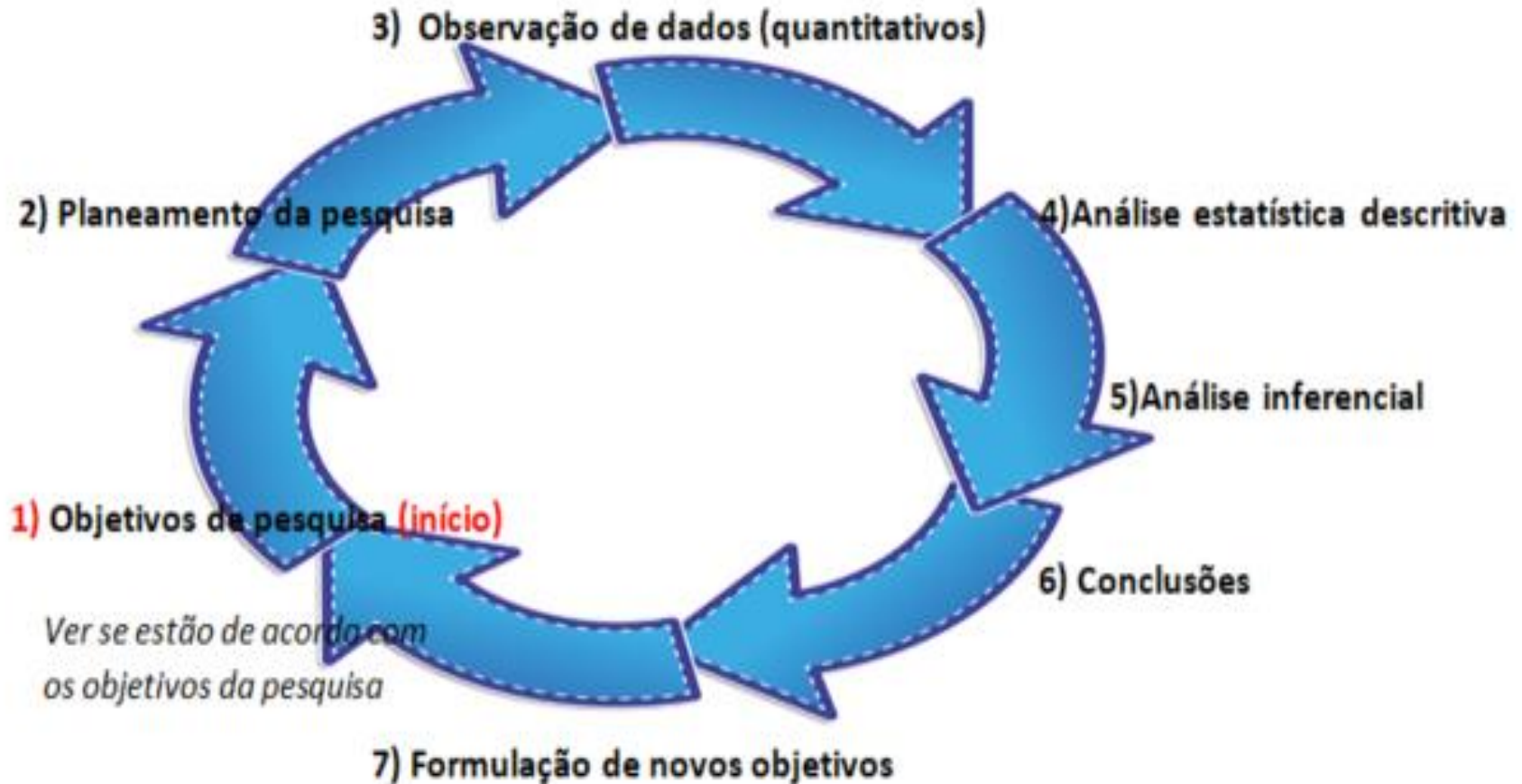
5. PREVISÃO

- Fase em que se aceitam as leis encontradas como certezas estatísticas permitindo, assim, prever novos acontecimentos (se a potência do teste for muito elevada (+90%), posso aceitar os dados como “certezas” estatísticas, prevendo assim novos acontecimentos).

Nota: A fase de previsão na prática é incluída na estatística analítica!

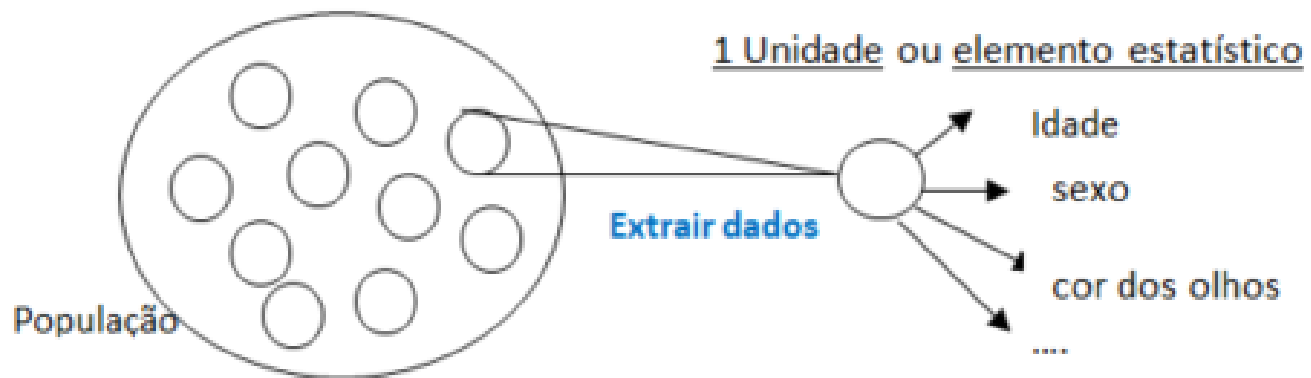


6. INTERAÇÃO ENTRE ESTATÍSTICA E PESQUISA CIENTÍFICA



7. UNIDADE ESTATÍSTICA OU ELEMENTO ESTATÍSTICO

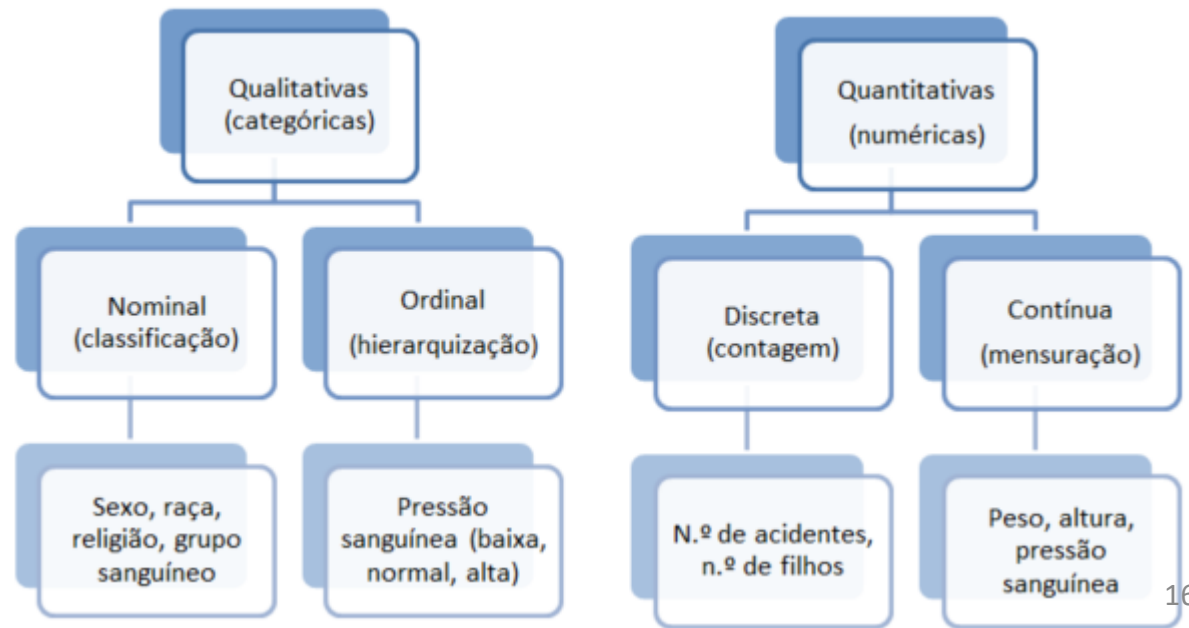
- É cada elemento de um conjunto de dados da população
- Unidade de amostragem – cada um dos elementos da população
- Uma unidade concreta sobre o qual se recolhem informações designa-se por CASO
- Número de identificação (ID) – número sequencial de 1 a N atribuídos aos casos que compõem a amostra (particularmente da recolha de dados por questionário)



8. VARIÁVEIS

- Características observadas em relação a cada unidade estatística
- É representada por símbolos (como x , y , a , b ...) e pode assumir qualquer um dos valores que lhe são atribuídos
- O conjunto de valores que assume é chamado de campo de variação ou domínio da variável
- A classificação das variáveis pode ser feita segundo vários critérios

Tipos de variáveis



9. AMOSTRAS

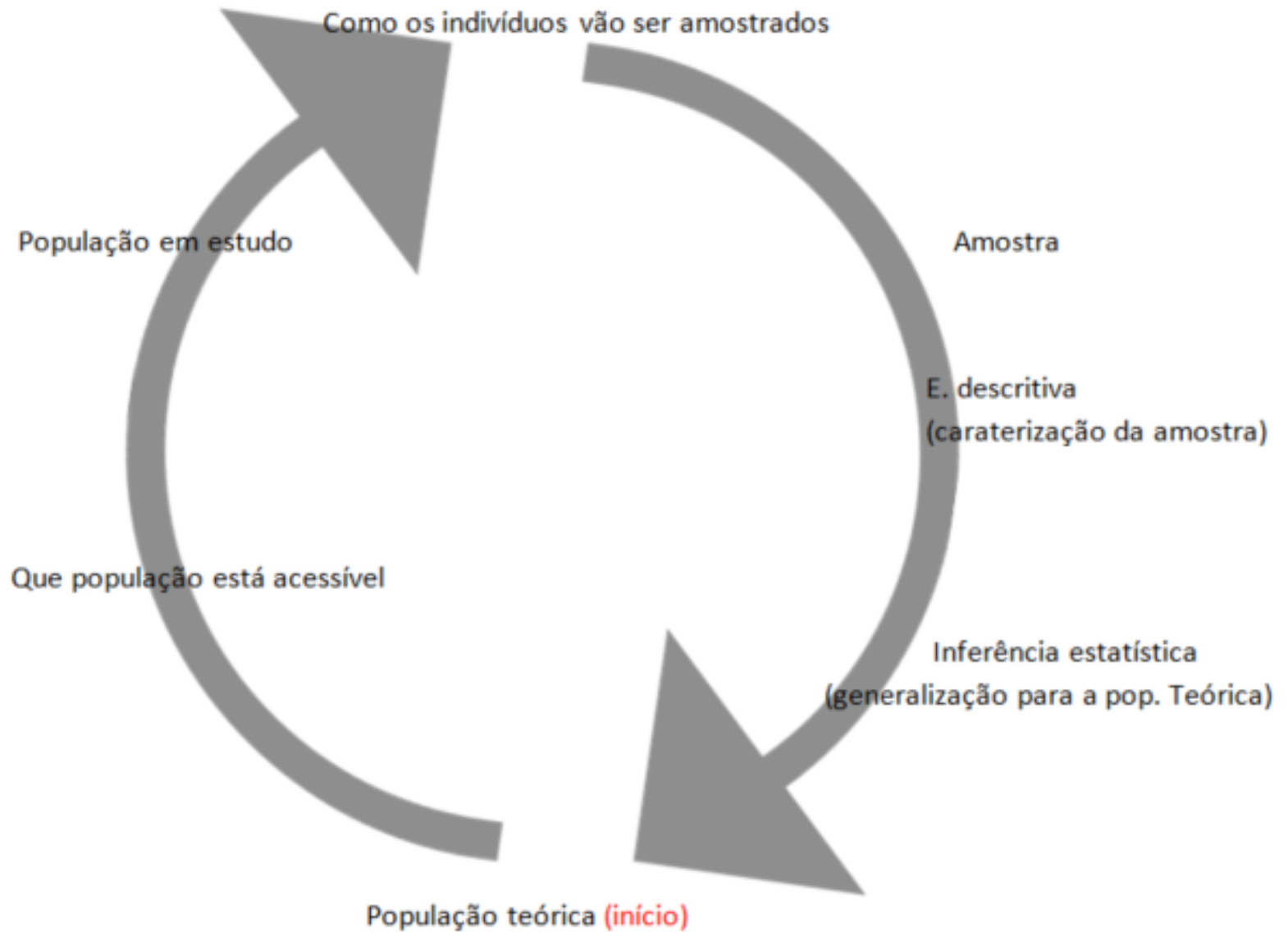
- Amostra é uma parcela convenientemente selecionada do universo.
- A amostra deve ser representativa (ter a mesma diversidade de elementos da população) e proporcional (só desta forma é que é também representativa).
- É a partir da amostra que o pesquisador pode generalizar os resultados para toda a população de onde foi extraída



Tipos de amostras

- Amostra teórica
 - A totalidade da unidade população-grelha de amostragem, que foi efetivamente selecionada para o inquérito
 - Totalidade dos casos escolhidos para análise
- Amostra obtida
 - As unidades contactadas e que responderam ao inquérito
 - Casos em que as análises deram resultados válidos

Relação entre a população e a amostra



Amostragem

- A amostragem é o ato de selecionar ou recolher uma amostra da população.

a) Amostragem Probabilística: procedimento em que todos os elementos da população têm uma probabilidade conhecida e superior a zero de integrar a amostra

b) Amostragem não probabilística: procedimento em que nem todos os elementos da população têm uma probabilidade conhecida e superior a zero de integrar a amostra



a) Amostragem probabilística

- Amostragem aleatória simples
 - Sorteio sem reposição
 - Sorteio com reposição
 - Números aleatórios
- Amostragem aleatória sistemática
- Amostragem aleatória estratificada
 - Com partilha proporcional
 - Sem partilha proporcional
- Amostragem aleatória por clusters
- Amostragem multi-etapas (multistage sampling)
- Amostragem multifásica (multiphase sampling)



a) Amostragem NÃO probabilística

- Amostragens de conveniência: utilizam-se os resultados de fácil acesso, isto é, o que se tem disponível.
- Amostragens por quotas: Cada quota representa um estrato.
- Amostragem intencional: escolhe-se, intencionalmente, um grupo de elementos que irão compor a amostra, dos quais se deseja saber a opinião.
- Amostragem bola de neve: Uma pessoa tem uma doença muito grave e procura-se na Internet se há outras pessoas iguais; essa depois fala de outra, e esta de uma outra, e, assim, o grupo vai crescendo como uma bola de neve.



FIM