

População e Amostra

Ana Amelia Benedito Silva
aamelia@usp.br

disciplina: EAH5002

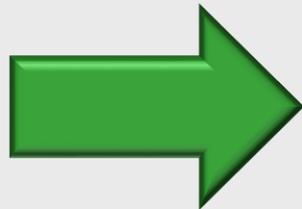
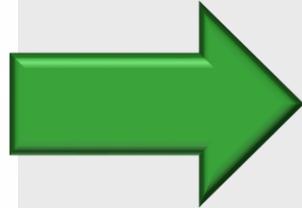
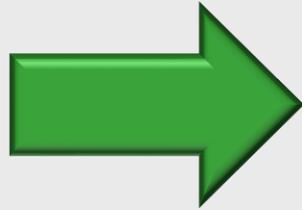
População e Amostra

Profa Ana Amélia Benedito-Silva
aamelia@usp.br

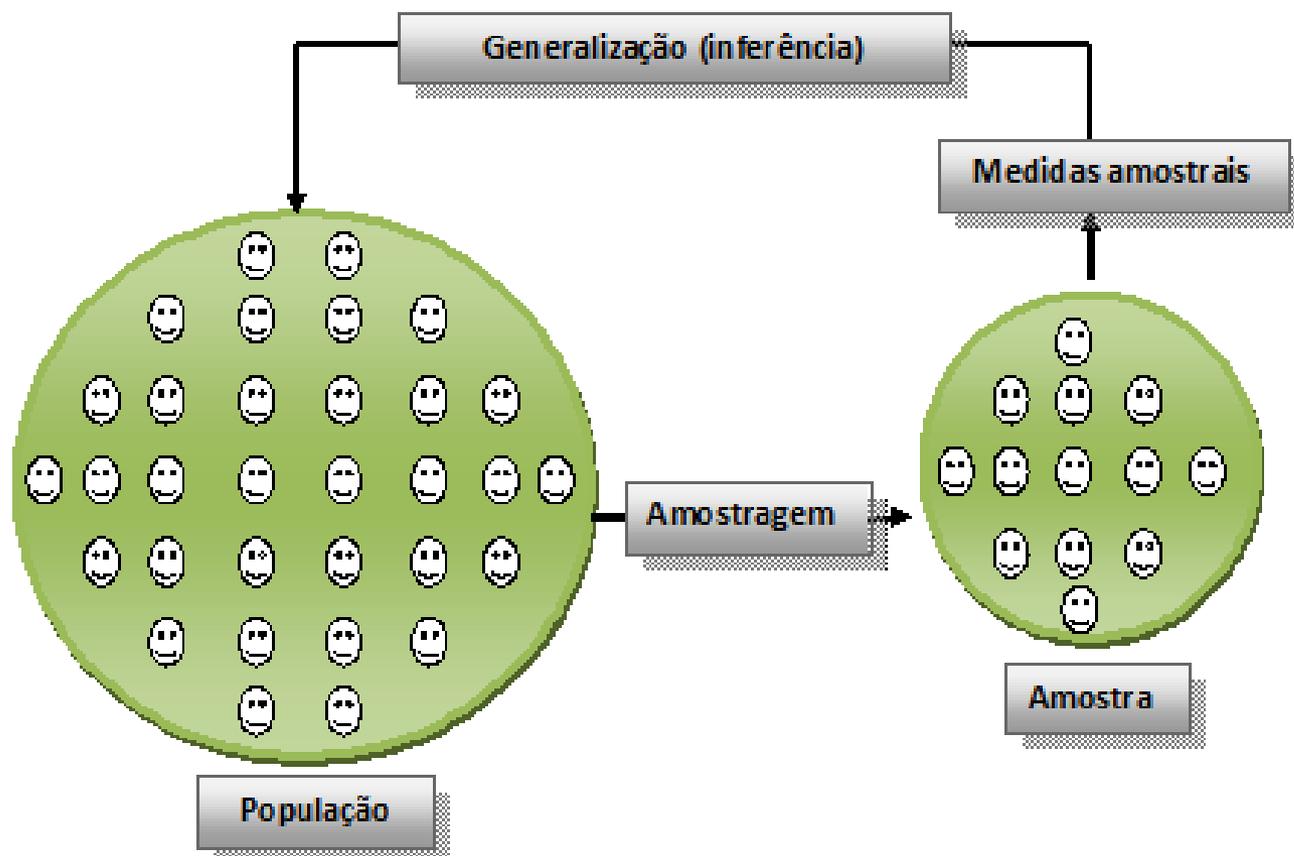
Como funcionam as pesquisas de opinião

- [https://www.youtube.com/watch?v=mV7FE
EUUZbI&t=3s](https://www.youtube.com/watch?v=mV7FE
EUUZbI&t=3s)

Conceito de amostragem

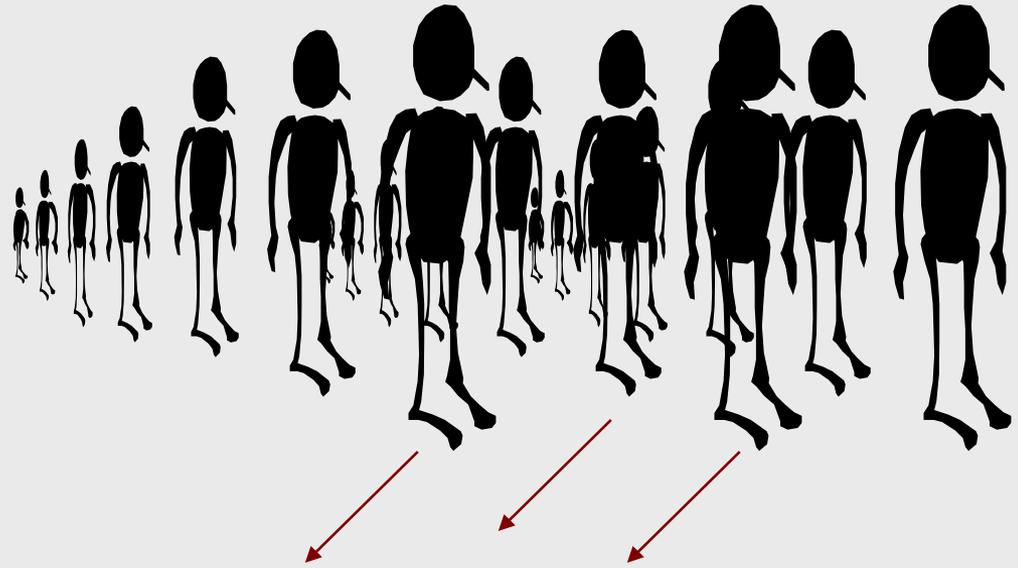


O processo intuitivo de amostragem faz parte do nosso dia-a-dia!



População e a variável a ser observada

POPULAÇÃO



Característica **X** observável:

X_1 X_2 X_3 ...

Pesquisa eleitoral: um exemplo de levantamento por amostragem

POPULAÇÃO:
eleitores brasileiros



AMOSTRA:
uma parte dos
eleitores



Voto do eleitor: X_1 X_2 X_3

POPULAÇÃO e AMOSTRA

POPULAÇÃO

Conjunto de elementos (objetos, pessoas, eventos) com características em comum

- Total de alunos da EACH no ano de 2022.
- Total de supermercados em SP
- Total de escolas de ensino fundamental em SP

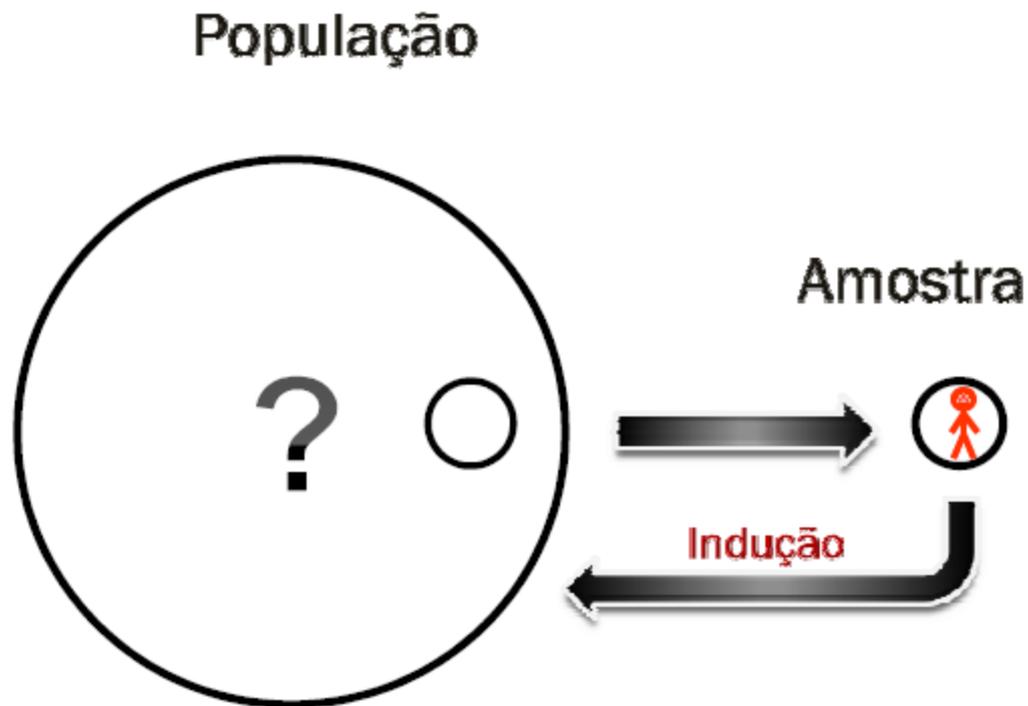
AMOSTRA

Subconjunto de elementos que pertencem à mesma população

- 30% de alunos da EACH no ano de 2022
- Supermercados de SP que ficam dentro de shoppings
- Escolas de ensino fundamental na Zona Leste de SP

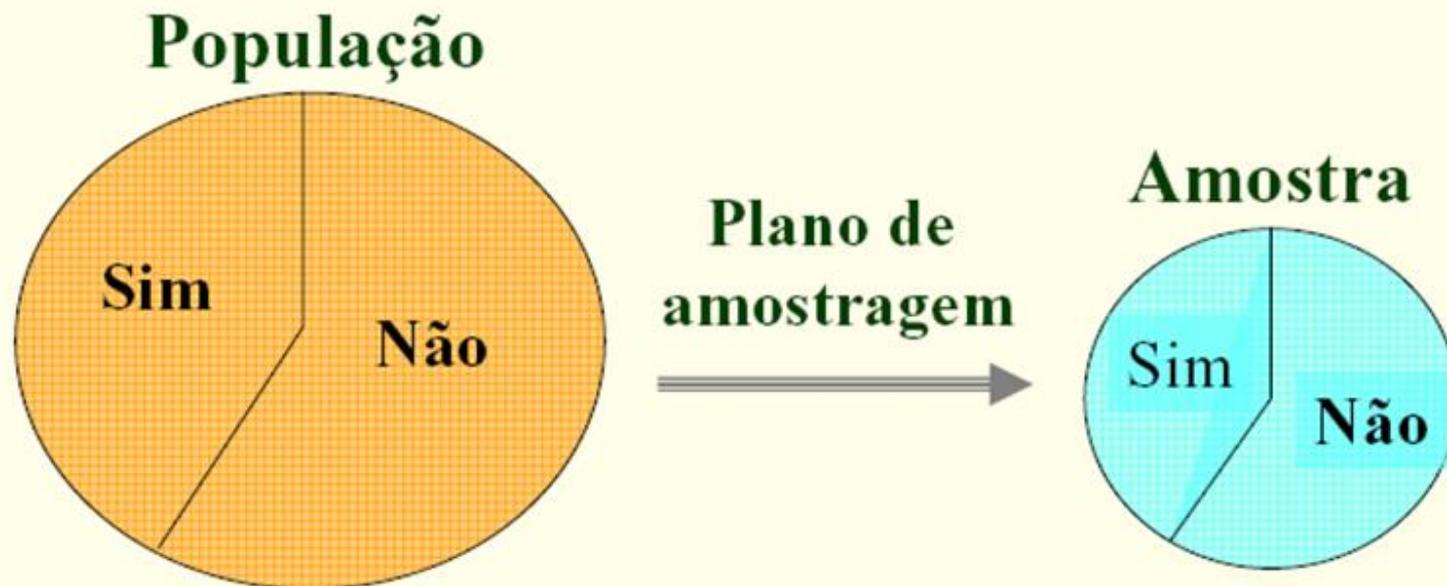
Amostra

Com a finalidade de estudar a população retiramos desta população uma parte representativa que chamamos de amostra e coletamos dados apenas desta amostra.

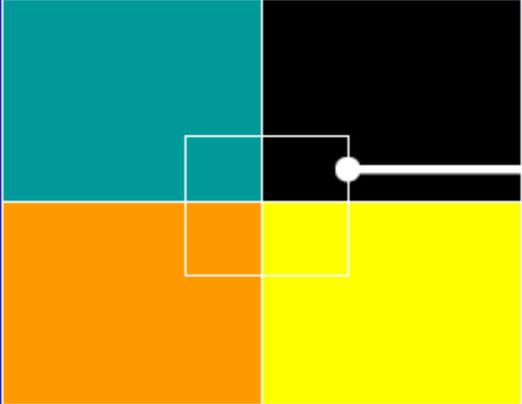


Amostragem

- A amostra deve ser *representativa*!

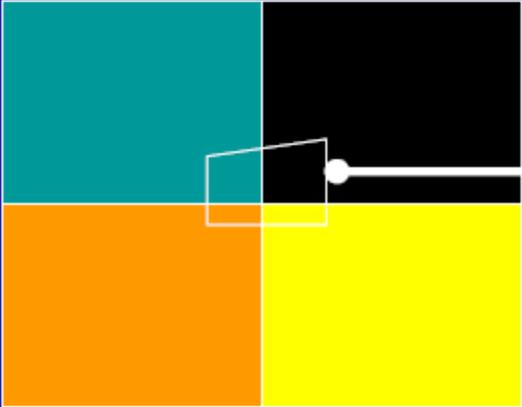


A amostra deve ser representativa da população



A diagram illustrating a representative sample. It features a large square divided into four quadrants: top-left is teal, top-right is black, bottom-left is orange, and bottom-right is yellow. A smaller white square, representing a sample, is positioned in the center, overlapping all four quadrants. A white arrow points from the center of the sample square to the right.

A idéia de aleatorização é conseguir uma amostra com características bem próximas da população



A diagram illustrating a non-representative sample. It features a large square divided into four quadrants: top-left is teal, top-right is black, bottom-left is orange, and bottom-right is yellow. A smaller white square, representing a sample, is positioned in the top-right quadrant, overlapping only the black and yellow quadrants. A white arrow points from the center of the sample square to the right.

Essa amostra não pode ser considerada representativa da população

Amostra

Para representar bem uma população a amostra deve ter:



QUALIDADE



QUANTIDADE

Qualidade da amostra

- Condição de saúde bucal das crianças de 12 anos de um município
- Se a pesquisa for domiciliar, quais domicílios: sorteio de bairros ou de residências?
- Se a pesquisa for nas escolas: sorteio escola ou sorteio classe ou sorteio alunos?

Quantidade

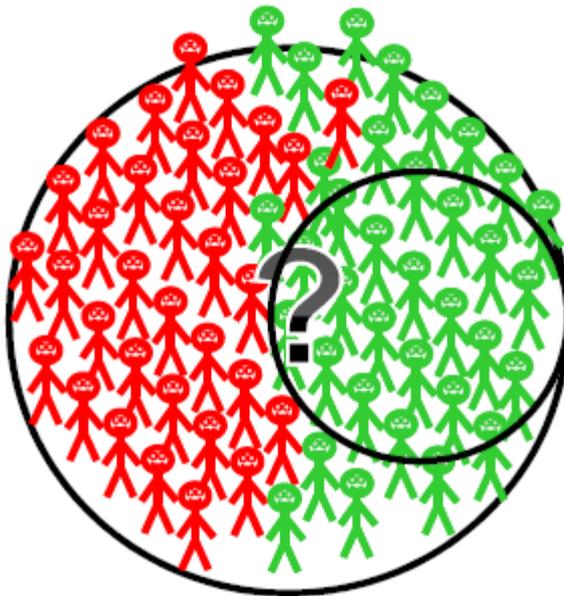
- É dependente da variabilidade da variável na população
 - Como determinar a variabilidade da população antes de estudar a população?
 - Estudo piloto
 - Literatura
 - Utilizar a maior variabilidade possível

Relação

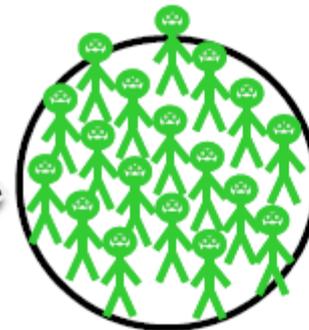
Quantidade e Qualidade

Boa quantidade sem qualidade

População



Amostra



~~Conclusão: a população
é verde~~

Boa qualidade sem quantidade

População



Amostra



~~Conclusão: a população
é predominantemente
verde~~

Boa qualidade e boa quantidade

População



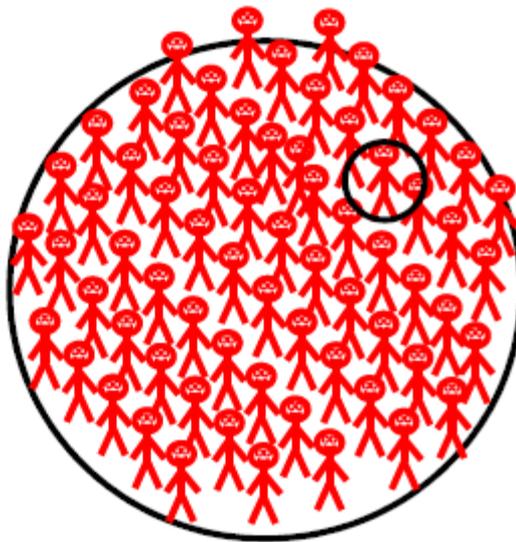
Amostra



Conclusão: a população é verde e vermelha

População homogênea

População



Amostra

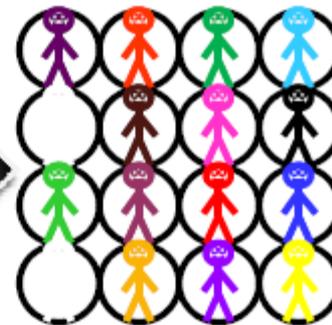


População heterogênea
(grande variabilidade)

População



Amostra



Viés de amostragem

- quando os indivíduos selecionados para a amostra não representam de forma adequada a população.
- Em 1936, nos EUA
 - um Instituto de opinião previu a partir de uma amostra de 2 milhões de eleitores, escolhidos em listas de proprietários de carros e catálogos telefônicos que Landon receberia 57% dos votos contra Roosevelt. Entretanto.....

Censo e Amostragem

Censo

- estudo através do exame de todos os elementos da **população**

Amostragem

- estudo por meio do exame de uma **amostra**

Censo ou Amostragem?

Censo

- quando população é pequena (amostra tem tamanho da população)
- quando resultado deve ser exato

Amostragem

- mais economia
- menos tempo
- adequado para população infinita

Exercício

Analise as situações descritas abaixo e decida se a pesquisa deve ser feita por amostragem ou por censo, justificando sua resposta.

- a) Numa linha de produção de empacotamento de café, observar o peso dos pacotes produzidos.
- b) Em uma sala de aula composta por 40 alunos, analisar suas idades.
- c) Observar se a água de uma lagoa está contaminada.
- d) Verificar a carga horária diária de trabalho dos funcionários da cozinha de um restaurante.
- e) Num lote de cabos de aço, verificar a resistência dos mesmos à tração.
- f) Pesquisa de opinião eleitoral para governador do Estado de São Paulo.

Métodos para a composição de uma amostra

Probabilísticos

1. aleatória simples
2. sistemática
3. estratificada proporcional
4. por conglomerados

Não-probabilísticos

5. quotas ou proporcional
6. acidental ou conveniência
7. Intencional

Métodos Probabilísticos

- Aleatória simples #09 - Amostragem Aleatória Simples

https://www.youtube.com/watch?v=Tk-7fTtUP_Y

- #10 - Números Aleatórios

<https://www.youtube.com/watch?v=5xWb9Nyttcc>

- #11 - Amostragem Sistemática

<https://www.youtube.com/watch?v=H1jq8bEvl34>

- #12 - Amostragem Estratificada

<https://www.youtube.com/watch?v=PNdMANe96cU>

- #13 - Amostragem por Conglomerado

<https://www.youtube.com/watch?v=EI8x6vucRs>

Métodos Não-probabilísticos

- Acidental

https://www.youtube.com/watch?v=olHF_ICKKpM

- Intencional

https://www.youtube.com/watch?v=xi7CNuYgZ_0

- Proporcional

<https://www.youtube.com/watch?v=ojoDnOTfvSo>

<https://www.youtube.com/watch?v=pgdh3wapzJY>

<https://www.youtube.com/watch?v=36ulultiNoM>

- Desproporcional #08 - Amostragem Desproporcional

<https://www.youtube.com/watch?v=HMcGa0X9nEQ>

Métodos probabilísticos

- cada elemento da população tem **uma chance conhecida e diferente de zero** de ser escolhido para compor a amostra.
- a partir do conhecimento da amostra podem-se fazer generalizações sobre a população
- mais corretos, mais confiáveis

Métodos probabilísticos

1. Amostragem aleatória simples

- É equivalente a um bingo, um sorteio de loteria.
- Numera-se a população de 1 a n
- Sorteiam-se x números dessa seqüência, que corresponderão aos elementos pertencentes à amostra

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo 1

- Pesquisa sobre prática de atividade física de famílias de funcionários de uma certa empresa.

População: 32 funcionários

Amostra: 5 funcionários

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo 1

População

01. Aristóteles	02. Anastácia	03. Arnaldo	04. Bartolomeu	05. Bernardino
06. Cardoso	07. Carlito	08. Cláudio	09. Ermílio	10. Hercílio
11. Ernestino	12. Endevaldo	13. Francisco	14. Felício	15. Fabrício
16. Geraldo	17. Gabriel	18. Getúlio	19. Hiraldo	20. João da Silva
21. Joana	22. Joaquim	23. Joaquina	24. José da Silva	25. José de Souza
26. Josefa	27. Josefina	28. Maria José	29. M ^a Cristina	30. Mauro
31. Paula	32. Paulo César			

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo 1

População

01. Aristóteles	02. Anastácia	03. Arnaldo	04. Bartolomeu	05. Bernardino
06. Cardoso	07. Carlito	08. Cláudio	09. Ermílio	10. Hercílio
11. Ernestino	12. Endevaldo	13. Francisco	14. Felício	15. Fabrício
16. Geraldo	17. Gabriel	18. Getúlio	19. Hiraldo	20. João da Silva
21. Joana	22. Joaquim	23. Joaquina	24. José da Silva	25. José de Souza
26. Josefa	27. Josefina	28. Maria José	29. M ^a Cristina	30. Mauro
31. Paula	32. Paulo César			

Sorteio de uma amostra de 5 elementos

População e amostra

01. Aristóteles	02. Anastácia	03. Arnaldo	04. Bartolomeu	05. Bernardino
06. Cardoso	07. Carlito	08. Cláudio	09. Ermílio	10. Hercílio
11. Ernestino	12. Endevaldo	13. Francisco	14. Felício	15. Fabrício
16. Geraldo	17. Gabriel	18. Getúlio	19. Hiraldo	20. João da Silva
21. Joana	22. Joaquim	23. Joaquina	24. José da Silva	25. José de Souza
26. Josefa	27. Josefina	28. Maria José	29. M ^a Cristina	30. Mauro
31. Paula	32. Paulo César			

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo2

- Pesquisa de estatura de crianças do 6º ano do ensino fundamental de uma escola
- População: 1000 alunos
- Amostra: 10% da população

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo2

- 1º passo - numeramos os alunos de 1 a 1000.
- 2º passo - escrevemos os números dos alunos, de 1 a 1000, em pedacinhos de papel
- 3º passo - retiramos 100 números.
- Procedimento **MUITO** trabalhoso!!!!
- **Alternativa:** tabela de números aleatórios no Excel ALEATORIOENTRE (0 e 1000)

Métodos probabilísticos

1. Amostragem aleatória simples

Exemplo 3

Pesquisa sobre satisfação com os serviços de uma agência de viagens

População: 200 pessoas retiradas de um banco de dados de clientes

Amostra: 10 clientes

Métodos probabilísticos

2. Amostragem sistemática

- Utilizada quando os elementos da população já se acham ordenados.

Exemplos

casas de uma rua

prontuários médicos de um hospital

alunos de uma classe em ordem alfabética

- **Vantagem:** facilidade, um único sorteio

Métodos probabilísticos

2. Amostragem sistemática

Exemplo4

- Suponhamos uma rua com 900 casas, das quais desejamos obter uma amostra formada por 50 casas para uma pesquisa de opinião.

População 900 casas

Amostra 50 casas

Métodos probabilísticos

2. Amostragem sistemática

Exemplo4

- 1º passo - calcular o salto = $900/50 = 18$
- 2º passo - sortear um número de 1 a 18, que indicará o primeiro elemento da amostra
- 3º passo - determinar os próximos elementos, somando-se o salto ao valor sorteado, até que as 50 casas sejam obtidas

Métodos probabilísticos

2. Amostragem sistemática

Exemplo4

- se o número sorteado for o 4, a amostra seria formada pelas seguintes casas:
4^a casa, 22^a casa, 40^a casa, 58^a casa,...
- se o número sorteado for o 3, a amostra seria formada pelas seguintes casas:
3^a casa, 21^a casa, 39^a casa, 57^a casa,...

Métodos probabilísticos

2. Amostragem sistemática

Exemplo5

Pesquisa sobre satisfação com os serviços de uma agência de viagens

População 200 clientes

Amostra 10 clientes

Métodos probabilísticos

2. Amostragem sistemática

Exemplo5

- 1º passo - calcular o salto = $200/10 = 20$
- 2º passo - sortear um número de 1 a 20, que indicará o primeiro elemento da amostra
- 3º passo - determinar os próximos elementos, somando-se o salto ao valor sorteado, até que 10 clientes sejam obtidos

Métodos probabilísticos

2. Amostragem sistemática

Exemplo 5

- se o número sorteado for o 4, a amostra seria formada pelos seguintes clientes:

4^a cliente, 24^a cliente, 44^a cliente,
64^a cliente,...

- se o número sorteado for o 3, a amostra seria formada pelos seguintes clientes:

3^a cliente, 23^a cliente, 43^a cliente,
63^a cliente,...

Métodos probabilísticos

3. Amostragem estratificada proporcional

POPULAÇÃO:
comunidade da escola

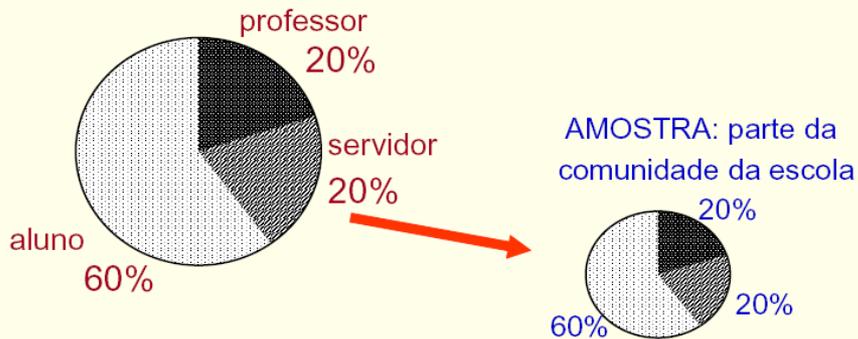
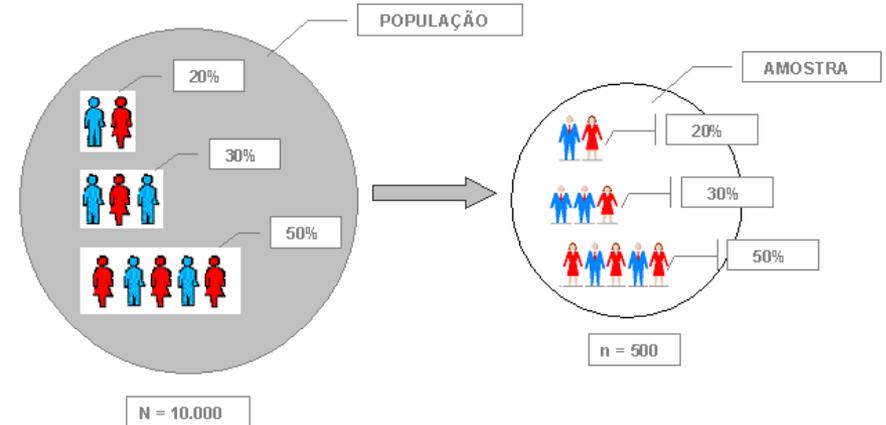


FIGURA 3 – AMOSTRAGEM ALEATÓRIA ESTRATIFICADA PROPORCIONAL



- Utilizada quando a população se divide em **estratos** (subpopulações)
- O sorteio é feito dentro de cada **estrato**

Métodos probabilísticos

3. Amostragem estratificada proporcional

- Pesquisas de mercado:
 - homens e mulheres
 - faixas etárias
- Pesquisas eleitorais:
 - região demográfica
 - cidades pequenas médias e grandes
 - área urbana e rural

Métodos probabilísticos

3. Amostragem estratificada proporcional

Exemplo 6

Pesquisa sobre prática de atividade física da comunidade de uma universidade

Deseja-se uma amostra de 10% da população.

População: 4000 pessoas

3700 alunos

200 professores

100 funcionários

Amostra: 400 pessoas

370 alunos

20 professores

10 funcionários

Os elementos da amostra devem ser sorteados nos estratos:

- 370 alunos dentre os 3700
- 20 professores dentre os 200
- 10 funcionários dentre os 100.

Métodos probabilísticos

3. Amostragem estratificada proporcional

Exemplo7

- Pesquisa sobre satisfação com os serviços de uma agência de viagens

População: 200 pessoas

60 mulheres (30%), 140 homens (70%)

- **Amostra: 10 pessoas**
3 mulheres, 7 homens

Métodos probabilísticos

4. Amostragem por conglomerados

Apenas os subgrupos (conglomerados) são identificados.

Pesquisa com alunos

- apenas as escolas são identificadas

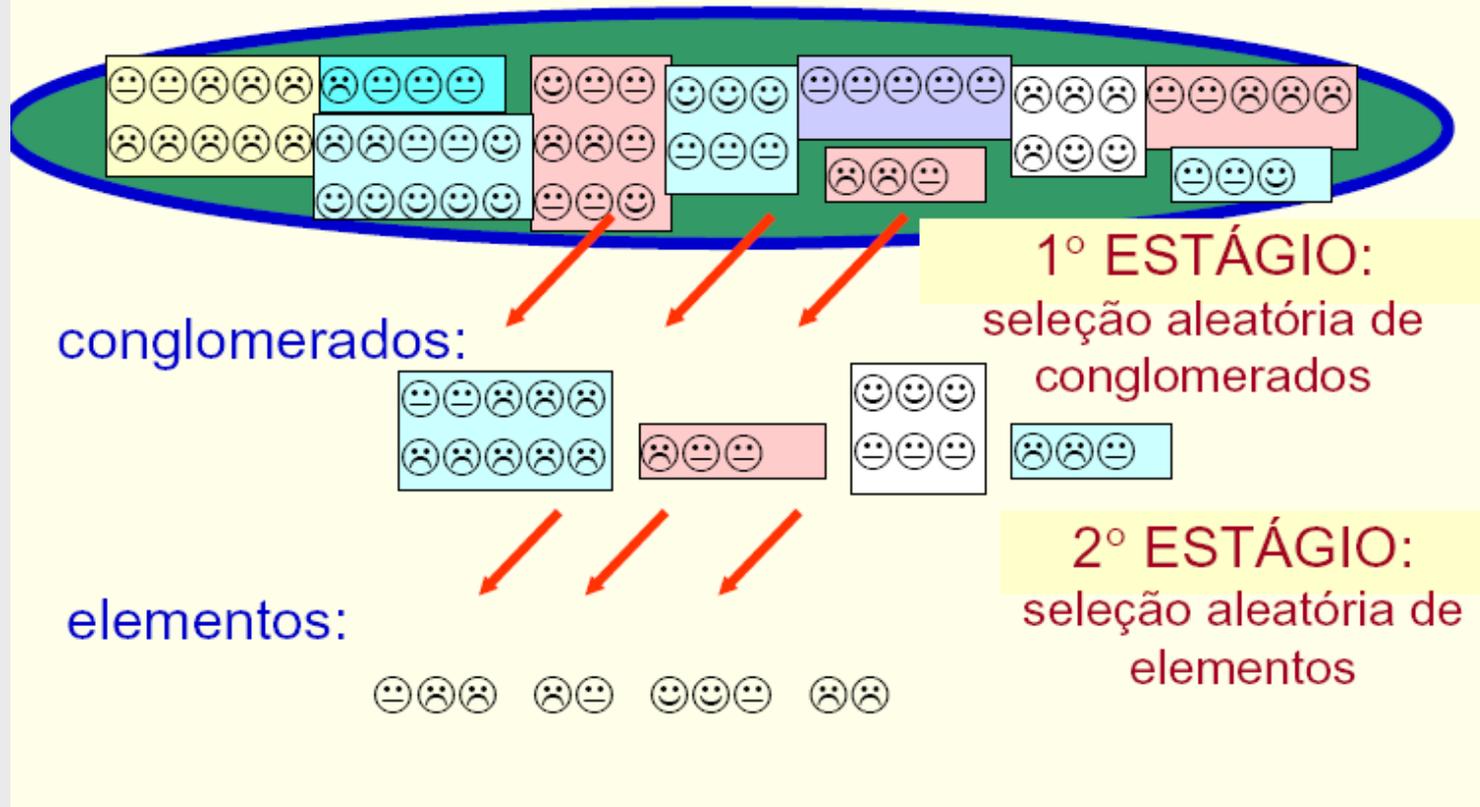
Pesquisa com moradores

- apenas os quarteirões são identificados

Métodos probabilísticos

4. Amostragem por conglomerados

Amostragem por Conglomerados



Métodos probabilísticos

4. Amostragem por conglomerados

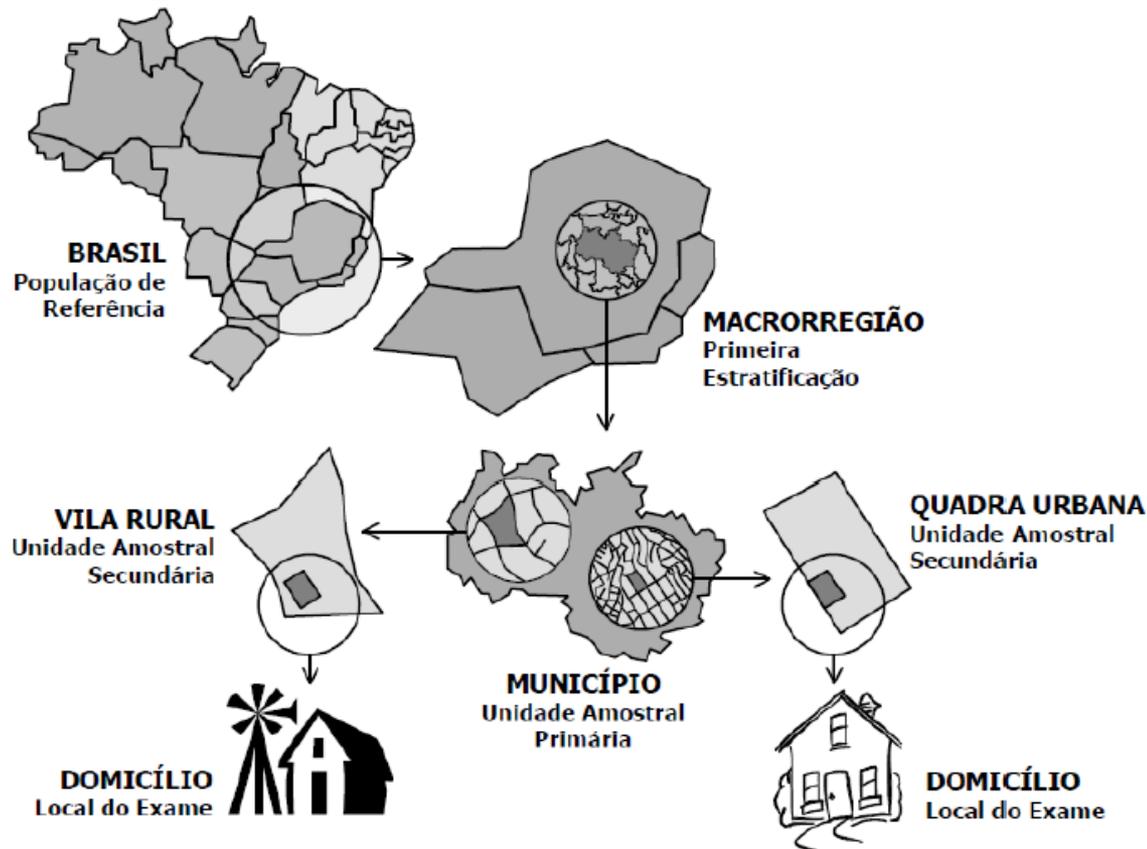


Figura 3. Esquema do plano amostral para o levantamento em domicílios em municípios até 50 mil habitantes (1º, 2º e 3º estratos).

Métodos probabilísticos

4. Amostragem por conglomerados

Exemplos de conglomerados

Elementos	Conglomerados
Eleitores	Domicílios
Trabalhadores	Empresas
Alunos	Escolas

Exemplo mais complexo:

Para obter uma amostra de famílias:

selecionar primeiro uma amostra de cidades;

selecionar bairros das cidades sorteadas;

selecionar bairros dos bairros sorteados;

selecionar domicílios dos bairros sorteados.

Amostragem por
conglomerados em 4 estágios

Métodos probabilísticos

4. Amostragem por conglomerados

Qual é a porcentagem de moradores da cidade de SP com queixas de sono?

1a etapa

Sorteio de bairros

2a etapa

Sorteio de quarteirões

3a etapa

Sorteio de residências

4a etapa

Aplicar o questionário em todos os moradores da residências sorteadas

Métodos não-Probabilísticos

- cada elemento da população **não tem uma chance conhecida** de ser escolhido para compor a amostra.
- não é possível generalizar os resultados das pesquisas para a população
- as amostras não-probabilísticas não garantem a representatividade da população.

Razões para o uso de amostragens não probabilísticas

- Quando os dados sobre a população não existem ou não estão disponíveis para serem sorteados.
- Quando não há intenção de generalizar os dados obtidos na amostra para toda a população
- Quando não há disponibilidade de tempo e/ou de recursos para a realização de pesquisa com amostragem probabilística

Teoricamente a amostragem probabilística é superior, mas na prática, as dificuldades de aplicação enfraquecem essa superioridade.

Métodos não-probabilísticos

5. Amostragem por cotas

usado em levantamentos de mercado e prévias eleitorais.

Exemplo

- Suponha que deseja-se fazer uma pesquisa sobre a intenção de voto numa cidade de 50 mil eleitores.
- O tamanho da amostra foi dimensionado em 400 eleitores.

Métodos não-probabilísticos

5. Amostragem por cotas

Primeira fase

Classifica-se a população de acordo com características consideradas relevantes para o estudo.

- Na pesquisa de intenção de votos as variáveis a serem consideradas serão:
sexo e nível de escolaridade

Métodos não-probabilísticos

5. Amostragem por cotas

Segunda fase

Com base na constituição conhecida da população (cartório eleitoral) determina-se a proporção para cada característica

Sexo

Masculino 55%

Feminino 45%

Escolaridade

Analfabeto 15%

Fundamental 35%

Médio 40%

Superior 10%

Métodos não-probabilísticos

5. Amostragem por cotas

Terceira fase: fixam-se as cotas

	Analfabetos	Fundamental	Médio	Superior
masculino 55% de 400 220	15% de 220 33	35% de 220 77	40% de 220 88	10% de 220 22
feminino 45% de 400 180	15% de 180 27	35% de 180 63	40% de 180 72	10% de 180 18

Tamanho da amostra = **400** eleitores da cidade

Métodos não-probabilísticos

5. Amostragem por cotas

Terceira fase: fixam-se as cotas

	Masculino 55% de 400 = 220	Feminino 45% de 400 = 180
Analfabeto (15%)	15% de 220 = 33	15% de 180 = 27
Fundamental (35%)	35% de 220 = 77	35% de 180 = 63
Médio (40%)	40% de 220 = 88	40% de 180 = 72
Superior (10%)	10% de 220 = 22	10% de 180 = 18

Métodos não-probabilísticos

6. Amostragem por conveniência

- Elementos da amostra são escolhidos por conveniência - estão no lugar certo, na hora certa.
- Utilizada em pesquisas exploratórias para se testar ou obter idéias sobre um assunto.
- Utilizada em pesquisas de opinião em praças, ruas, supermercados, etc.
- Método menos rigoroso de compor uma amostra

Métodos não-probabilísticos

6. Amostragem por conveniência

Exemplo

- Pesquisa sobre o uso de coleiras em cachorros em praças públicas

A amostra é colhida através de entrevistas com frequentadores de praças públicas

Métodos não-probabilísticos

7. Amostragem intencional

- A amostra é formada por elementos de acordo com um determinado critério.
- Utilizado em pesquisas para determinar o potencial de um novo produto
- Os elementos a serem incluídos na amostra são selecionados com base no julgamento do pesquisador.

Métodos não-probabilísticos

7. Amostragem intencional

Pesquisa sobre preferência por um novo shampoo:

pesquisador se dirige a um salão de beleza e entrevista as pessoas que ali se encontram.

Pesquisa sobre hábitos alimentares dos alunos de um colégio:

pesquisador escolhe uma determinada semana e entrevista os alunos, na lanchonete daquele colégio, durante o intervalo.

Sites interessantes

- <https://profasoniavieira.wixsite.com/estatistica>
- <https://pt.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/v/techniques-for-random-sampling-and-avoiding-bias>
- <https://pt.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/e/sampling-methods>
- <https://pt.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/a/sampling-methods-review>
- <https://pt.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/a/worksheet-samples-and-surveys>

Cálculo do tamanho da amostra para amostragem aleatória

- **Parâmetro** característica que está sendo estudada na população
- **Estatística** medida descritiva de elementos de uma amostra (média, proporção, desvio-padrão...)
- **Estimativa** valor avaliado por uma estatística que estima o valor de um parâmetro populacional
- **Erro amostral** diferença entre o valor que a estatística pode acusar e o verdadeiro valor do parâmetro que se deseja estimar
- **Erro amostral tolerável** quanto o pesquisador admite errar na avaliação dos parâmetros de interesse numa população

Fórmula para cálculo do tamanho da amostra

- N = Tamanho da população
- E_0 = erro amostral tolerável
- n_0 = primeira aproximação do tamanho da amostra

$$n_0 = \frac{1}{E_0^2}$$

- n = tamanho da amostra

$$n = \frac{N.n_0}{N + n_0}$$

Exemplo de cálculo de tamanho da amostra

N = 200 famílias

E_0 = erro amostral tolerável = 4% ($E_0 = 0,04$)

n_0 (1ª aproximação do tamanho da amostra)

$$n_0 = 1/(0,04)^2 = 625 \text{ famílias}$$

n (tamanho da amostra corrigido pelo tamanho da população)

$$n = (N \cdot n_0) / (N + n_0) = (200 \cdot 625) / (200 + 625)$$

Portanto, $n = 152$ famílias

Exemplo de cálculo de tamanho da amostra

$N = 200.000$ famílias

$E_0 =$ erro amostral tolerável = 4% ($E_0 = 0,04$)

n_0 (1ª aproximação do tamanho da amostra)

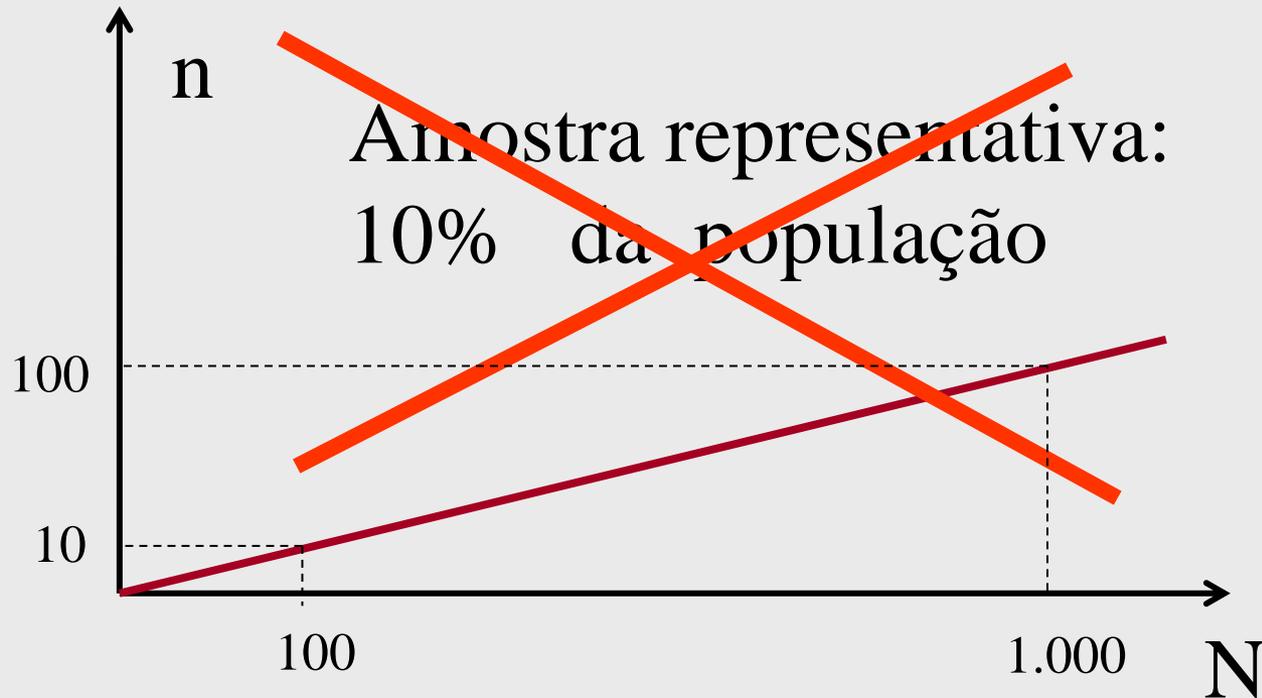
$$n_0 = 1/(0,04)^2 = 625 \text{ famílias}$$

n (tamanho da amostra corrigido pelo tamanho da população)

$$n = (N * n_0) / (N + n_0) = (200.000 * 625) / (200.000 + 625)$$

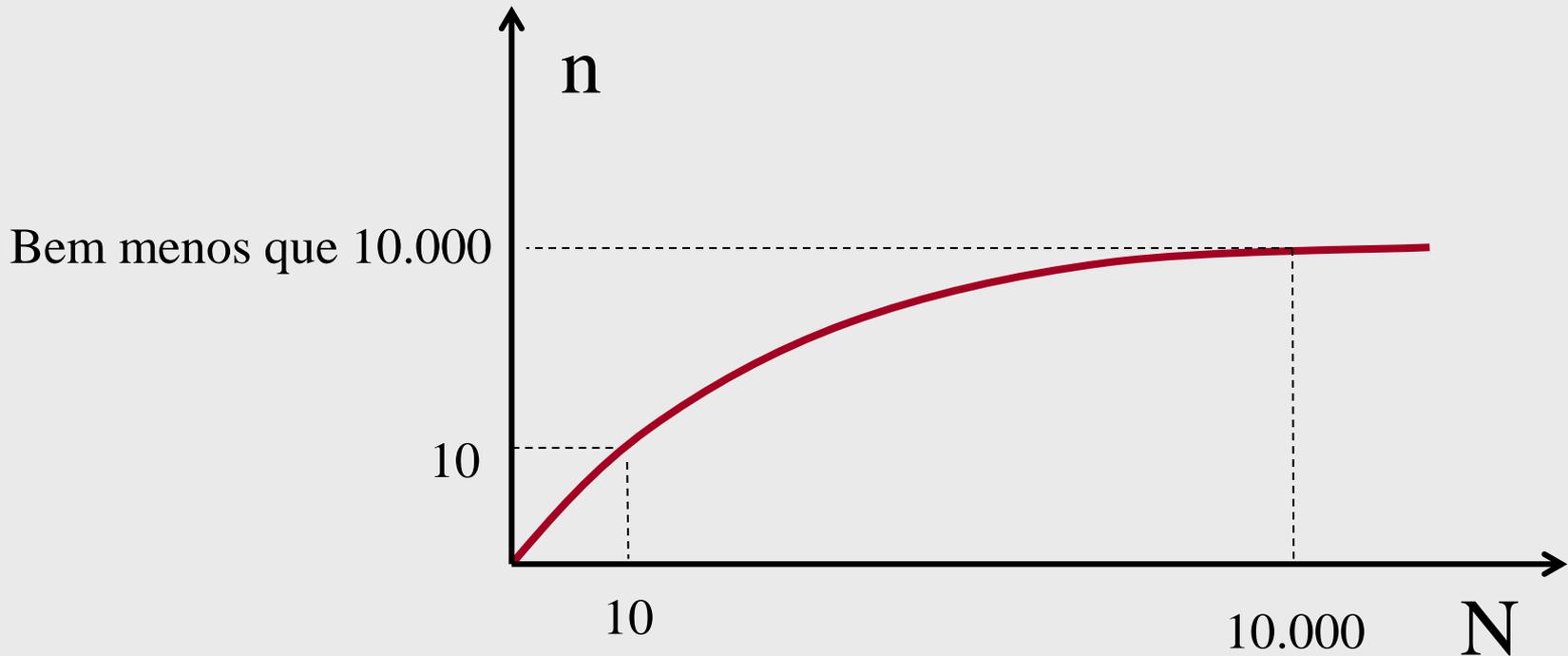
Portanto, $n = 623$ famílias

Tamanho da amostra (n) e tamanho da população (N)



A relação não é linear

Tamanho da amostra (n) e tamanho da população (N)



IMPORTANTE: forma de seleção da amostra

Tamanho da amostra ...

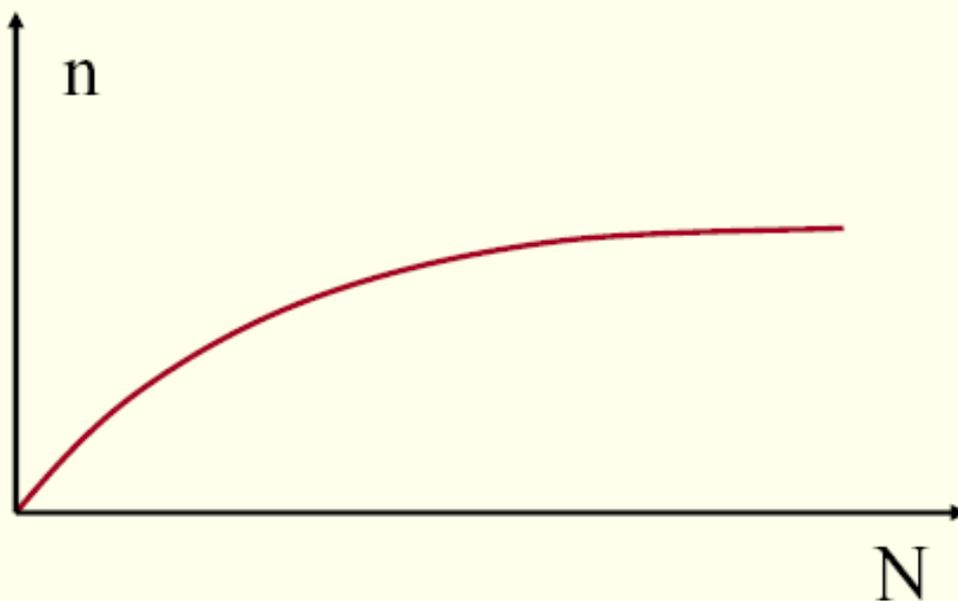
Observe que: $N = 200$ famílias, $E_0 = 4\%$

$n = 152$ famílias → 76% da população

Observe que: $N = 200.000$ famílias, $E_0 = 4\%$

$n = 623$ famílias → 0,3% da população

Logo, é errôneo pensar que o tamanho da amostra deve ser tomado como um percentual do tamanho da população para ser representativa



fim