

CIÊNCIAS AMBIENTAIS / BIOLÓGICAS / DA
NATUREZA

(BIO)ESTATÍSTICA

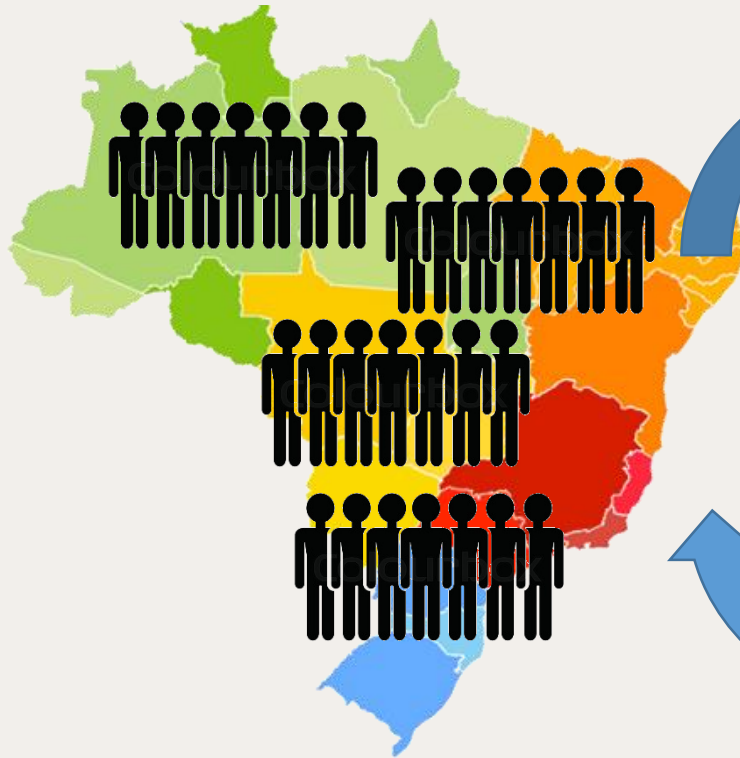
Prof^a. Letícia Raposo
profleticiaraposo@gmail.com



INFERÊNCIA ESTATÍSTICA

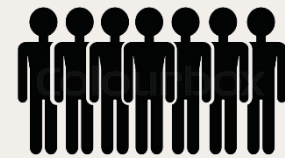
Inferência Estatística

População:
eleitores brasileiros



Amostragem

Amostra: uma parte dos
eleitores brasileiros



Inferência

OBJETIVO DA AULA

- **Aprender as diferentes técnicas de amostragem;**
- Aprender a generalizar resultados de uma amostra para a população onde ela foi extraída;
- Aprender a testar hipóteses com base em amostras.





TÉCNICAS DE AMOSTRAGEM

INTRODUÇÃO

A amostragem é naturalmente usada em nossa vida diária.
Exemplo: verificar o tempero de um alimento.

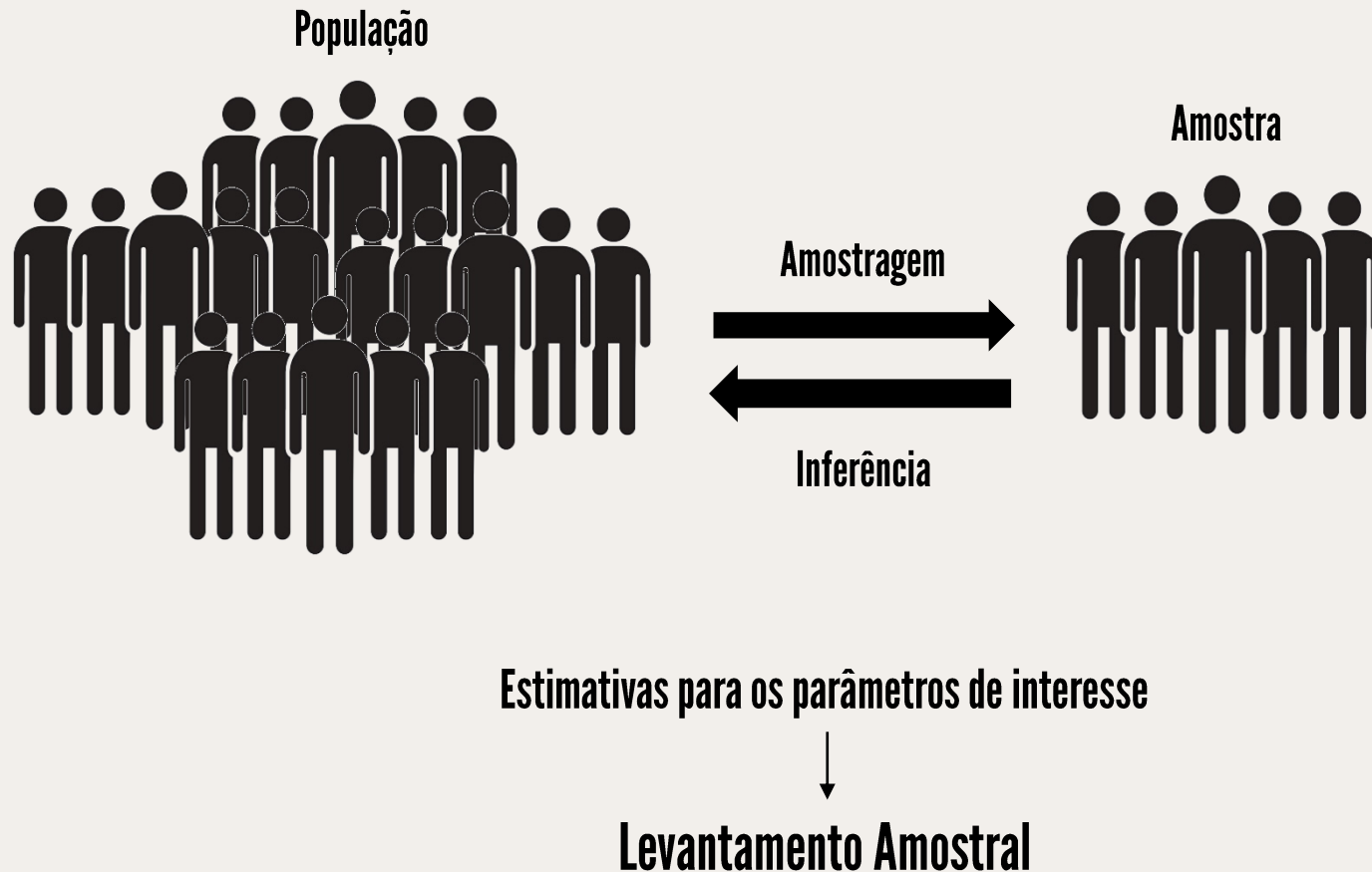
Provar (observar) um pequena porção → amostragem.



Ter uma ideia (inferir) sobre a qualidade do tempero de todo o alimento.



INTRODUÇÃO



ALGUNS CONCEITOS E EXEMPLOS

- **População:** conjunto de elementos para os quais desejamos que as conclusões da pesquisa sejam válidas, esses elementos podem ser observados ou mensurados sobre as mesmas condições.
 - Podem ser finitas ou infinitas (pode ser contado, porém é muito grande).
- **Parâmetro:** medida que descreve certa característica dos elementos da população.

Exemplo: numa pesquisa epidemiológica, a população pode ser definida como todas as pessoas da região em estudo, no momento da pesquisa. O principal parâmetro a ser avaliado deve ser a porcentagem de pessoas contaminadas.

Quando é difícil pesquisar toda a população, recomenda-se usar a amostragem.

ALGUNS CONCEITOS E EXEMPLOS

- **Inferência estatística:** refere-se ao uso apropriado dos dados de uma amostra para se ter conhecimento sobre parâmetros da população de onde foi extraída a amostra.
- **Amostra:** parte dos elementos de uma população.
- **Amostragem:** processo de seleção da amostra.
- **Estimativa:** valor calculado com base na amostra e usado com a finalidade de avaliar aproximadamente um parâmetro.

POR QUE AMOSTRAGEM?



Economia



**Confiabilidade
das respostas**



Tempo



Operacionalidade

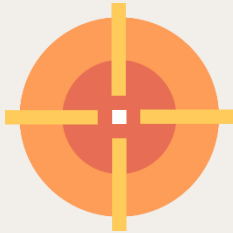
QUANDO O USO DE AMOSTRAGEM NÃO É INTERESSANTE?



População pequena



Característica de fácil mensuração



Necessidade de alta precisão

PLANO DE AMOSTRAGEM

- Para elaborar um plano de amostragem, devemos ter bem definidos:
 - Os objetivos da pesquisa;
 - A população a ser amostrada,
 - Os parâmetros a ser estimados.
- Precisamos também definir:
 - A unidade de amostragem;
 - A forma de seleção dos elementos da população;
 - O tamanho da amostra.

PLANO DE AMOSTRAGEM

- Unidade de amostragem: unidade a ser selecionada para se chegar aos elementos da população.



- A seleção dos elementos que farão parte da amostra pode ser feita sobre alguma forma de sorteio. São as chamadas amostragens aleatórias.

TIPOS DE AMOSTRAGEM

Amostragem Probabilística ou Aleatória



Amostragem Não Probabilística ou Não Aleatória



AMOSTRAGEM PROBABILÍSTICA OU ALEATÓRIA

Amostras obtidas aleatoriamente – probabilidade de cada elemento da população fazer parte da amostra é igual.

Vantagens:

- Critério de seleção dos elementos rigorosamente definido (sem subjetividade dos investigadores/entrevistadores);
- Determinação matemática da dimensão da amostra.

Desvantagens:

- Dificuldade em obter listagens ou regiões atuais e completas da população;
- Pode originar uma amostra muito dispersa geograficamente, aumentando os custos, o tempo envolvido no estudo e a dificuldade de coleta de dados.

AMOSTRAGEM NÃO PROBABILÍSTICA OU NÃO ALEATÓRIA

Probabilidade de alguns ou de todos os elementos da população pertencer à amostra é desconhecida.

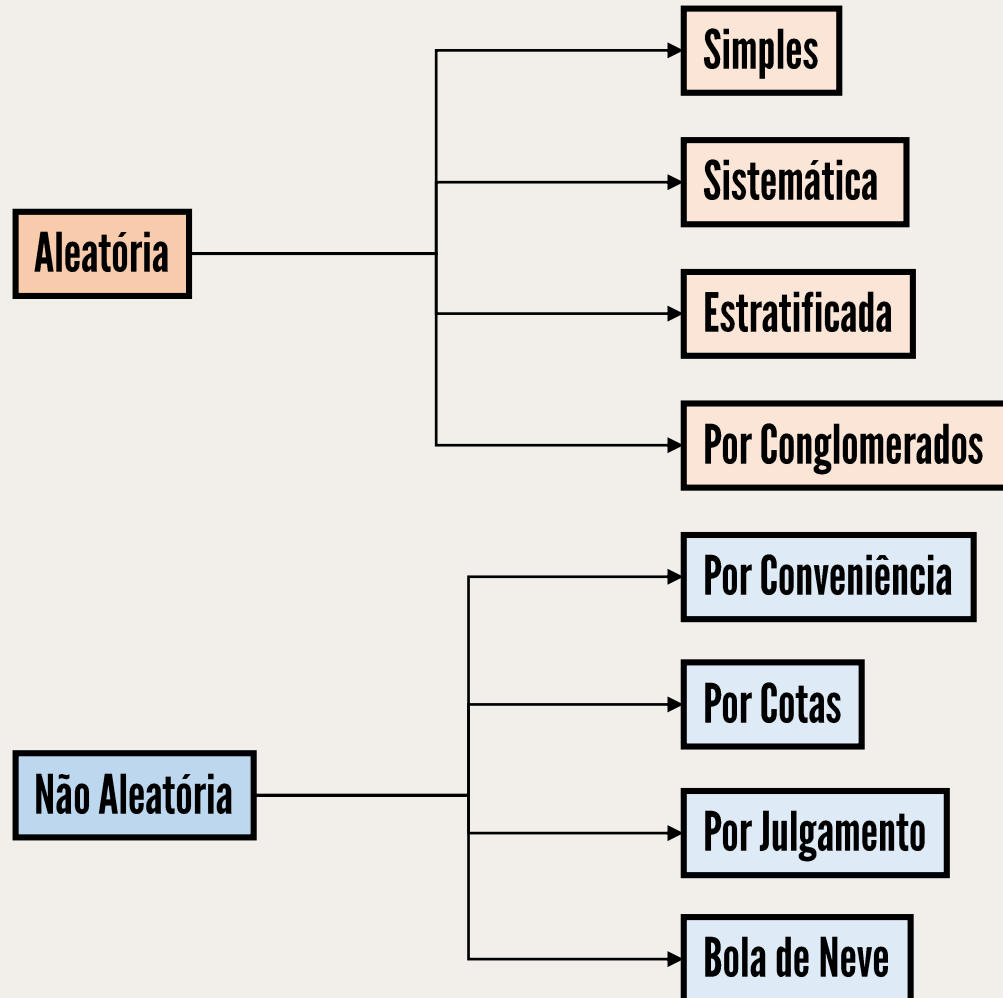
Vantagens:

- Menor custo, menor tempo de estudo, e menor necessidade de mão-de-obra.

Desvantagem:

- Pode ocorrer um viés de opinião pessoal;
- Não se sabe com que grau de confiança as conclusões obtidas podem ser inferidas para a população - não há garantia de que a amostra selecionada seja representativa da população.

TIPOS DE AMOSTRAGEM



AMOSTRAGEM ALEATÓRIA SIMPLES (AAS)

- Consiste em selecionar n elementos da população (N) através de um sorteio.
- Pode ser com ou sem reposição.
- Qualquer subconjunto da população com o mesmo número de elementos têm a mesma probabilidade de fazer parte da amostra.
 - Temos que cada elemento da população tem a mesma probabilidade (dada por n/N) de pertencer à amostra.

AMOSTRAGEM ALEATÓRIA SIMPLES (AAS)



AMOSTRAGEM ALEATÓRIA SIMPLES (AAS)

Simple random sample

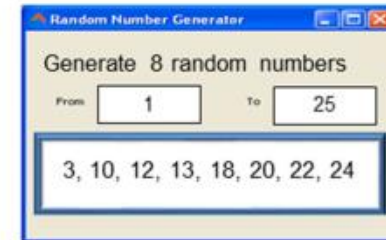
$n = 8$

1 2 3 4

5 6 7 8 9 10 11 12

13 14 15 16 17 18 19

20 21 22 23 24 25



AMOSTRAGEM SISTEMÁTICA

Os elementos da população estão ordenados e são retirados periodicamente.

- Ex: em determinada linha de produção de medicamentos, podemos retirar um elemento a cada 50 itens produzidos.

Vantagens:

- Comparada a AAS: executada com maior rapidez e menor custo.

Desvantagem:

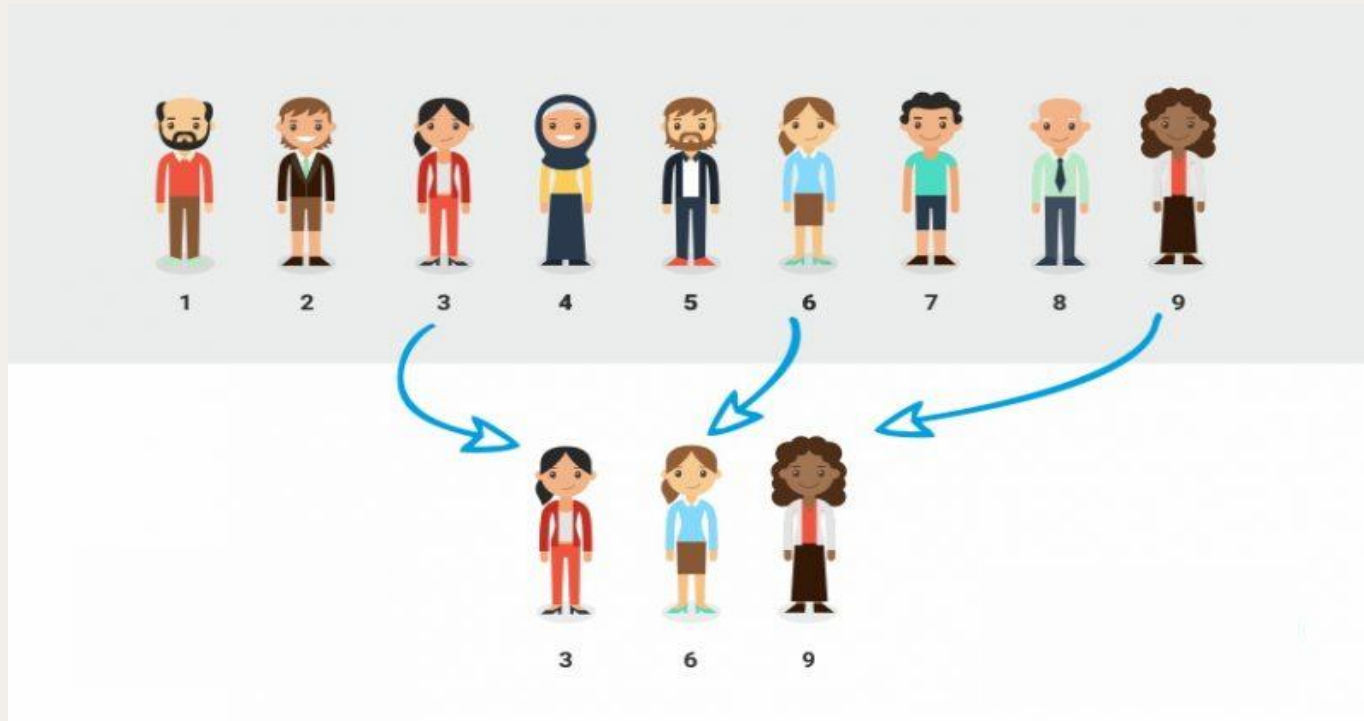
- Possibilidade de existirem ciclos de variação, especialmente se o período de ciclos coincide com o período de retirada dos elementos da amostra.

AMOSTRAGEM SISTEMÁTICA

Suponha que os elementos da população estejam ordenados de 1 a N e que já conhecemos o tamanho da amostra n , a amostragem sistemática opera da seguinte forma:

1. Devemos determinar o intervalo de amostragem $k = \frac{N}{n}$. Esse valor deve ser arredondado para o inteiro mais próximo;
2. Nesta fase, introduzimos um elemento de aleatoriedade escolhendo a unidade de partida. O primeiro elemento escolhido pode ser um elemento qualquer entre 1 e k ;
3. Escolhido o primeiro elemento, a cada k elementos, um novo elemento é retirado da população. O processo é repetido até atingir o tamanho da amostra.

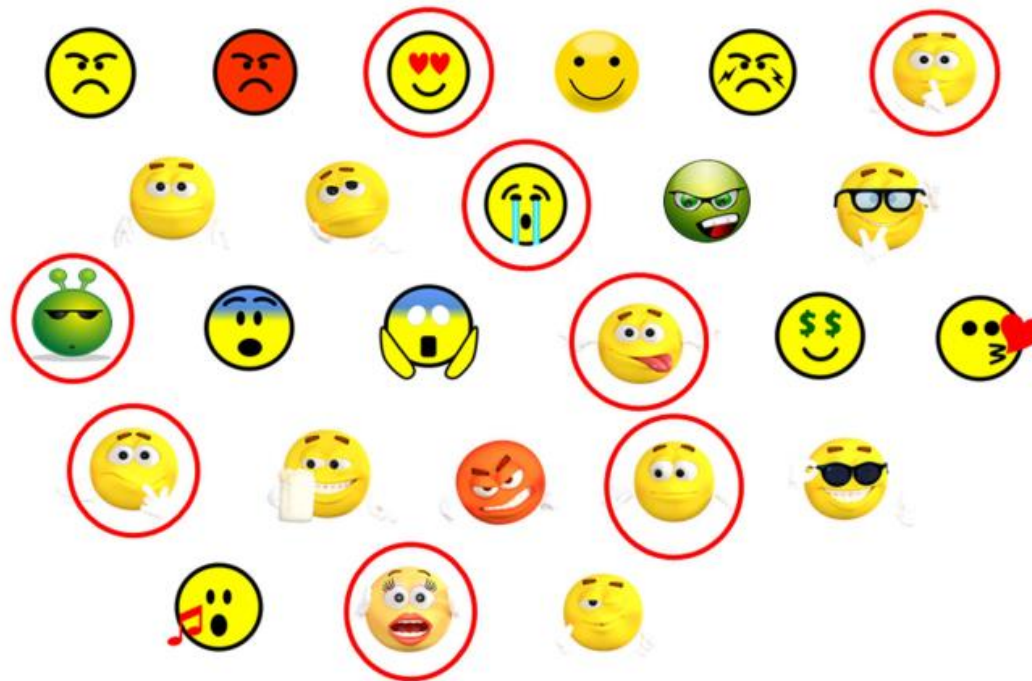
AMOSTRAGEM SISTEMÁTICA



AMOSTRAGEM SISTEMÁTICA

Systematic sample

n = 8 (every third emoji from left to right, top to bottom)



AMOSTRAGEM ESTRATIFICADA

- Consiste em dividir a população em subgrupos (estratos).
- Deve-se escolher um critério de estratificação que forneça estratos bem homogêneos com respeito ao que se está estudando → necessidade de conhecimento prévio sobre a população.
- Seleções aleatórias de forma independente são realizadas em cada estrato. A amostra é obtida através da agregação das amostras de cada estrato.
 - Ex: estratificação de uma cidade em bairros, de uma população por sexo ou faixa etária, de alunos por escola.

AMOSTRAGEM ESTRATIFICADA

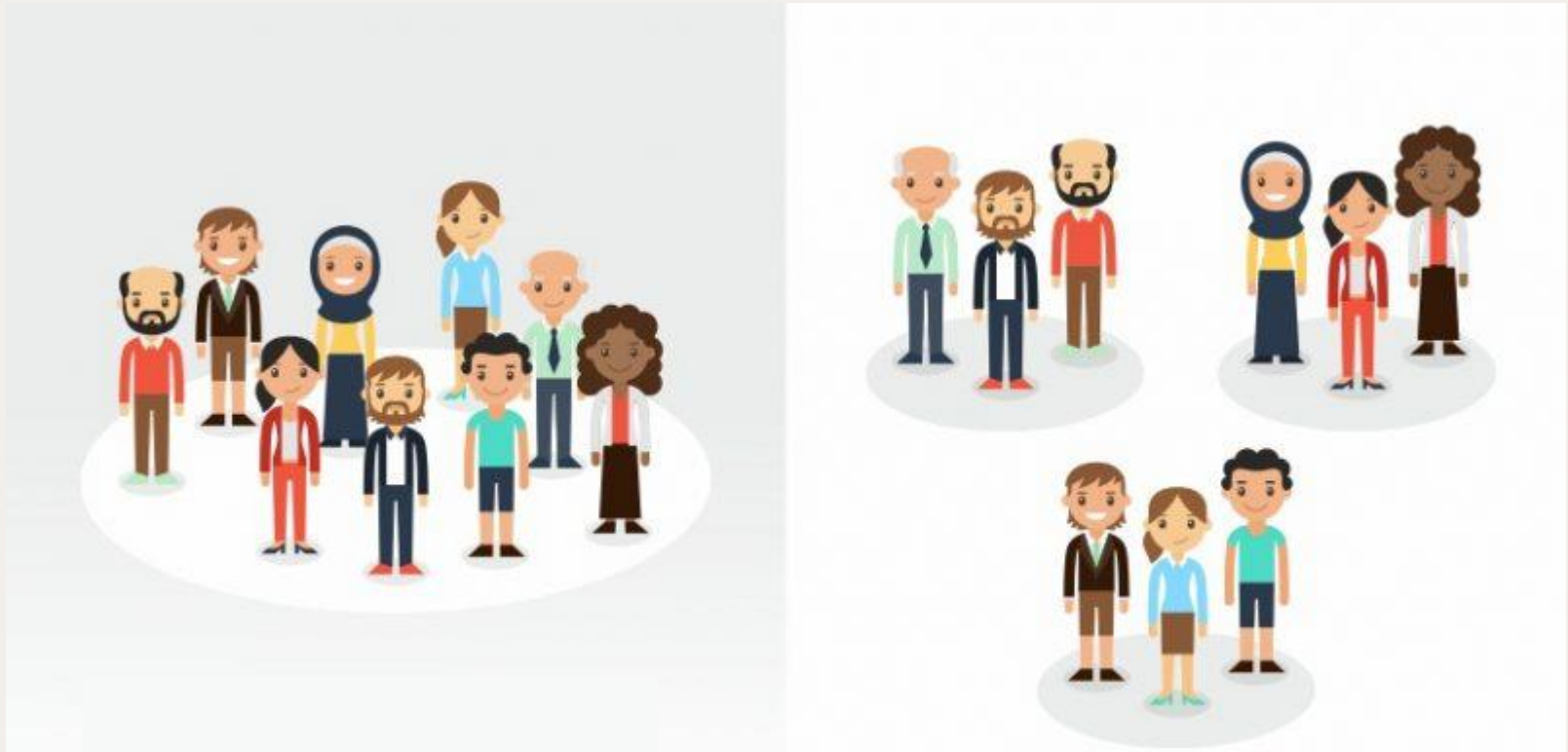
Amostragem Estratificada Uniforme

- ✓ Sorteia-se um nº igual de elementos em cada estrato ($n_i = n/k$);
- ✓ Recomendada quando os estratos são aproximadamente do mesmo tamanho;
- ✓ Mais usada para obter estimativas separadas para cada estrato, ou quando se deseja comparar os diversos estratos.

Amostragem Estratificada Proporcional

- ✓ O nº de elementos em cada estrato ($n_i = N_i/N$) $\times n$ é proporcional ao nº de elementos existentes no estrato.

AMOSTRAGEM ESTRATIFICADA



AMOSTRAGEM ESTRATIFICADA

Stratified random sample

n = 8 (4 randomly selected happy emojis and 4 randomly selected unhappy emojis)



Unhappy stratum



AMOSTRAGEM POR CONGLOMERADOS

- Conglomerados: agrupamento de elementos da população.
 - Ex: numa população de domicílios residenciais de uma cidade, os quarteirões formam conglomerados de domicílios.
- AC em um estágio: seleciona-se alguns conglomerados e depois se observa todos os elementos dos conglomerados selecionados.
- AC em dois estágios: (mais comum) faz-se nova seleção, tomando amostras de elementos dos conglomerados extraídos no primeiro estágio.
- Todas as seleções devem ser aleatórias.
- Em pesquisas de grande escala, a amostragem pode ser feita em mais estágios.

AMOSTRAGEM POR CONGLOMERADOS

- Fração de amostragem (n/N): proporção da população que será efetivamente observada.
- As estimativas de uma amostra de conglomerados tendem a gerar resultados mais distantes dos parâmetros populacionais quando comparada com uma AAS de mesmo tamanho.
- Os elementos dentro de um mesmo conglomerado tendem a apresentar características similares.

AMOSTRAGEM POR CONGLOMERADOS

Vantagens:

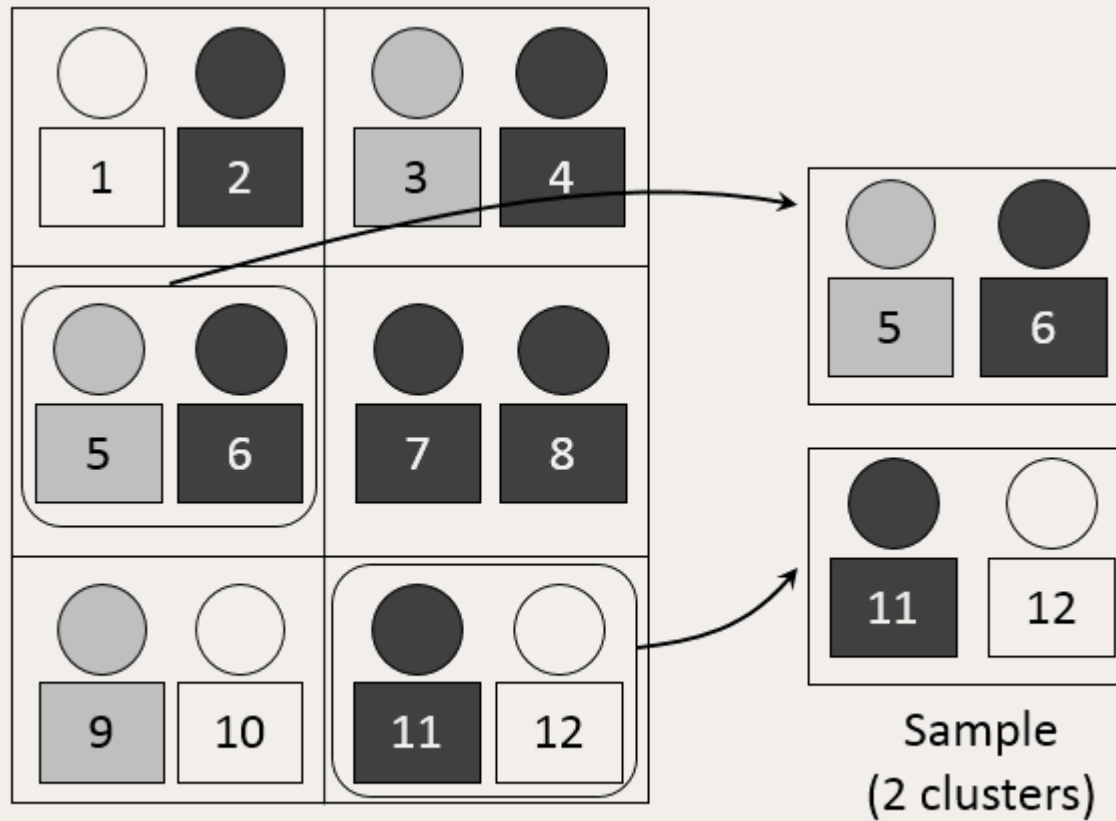
- Muitas populações já estão agrupadas em subgrupos naturais ou geográficos;
- Permite uma redução substancial nos custos de obtenção da amostra, sem comprometer sua precisão;
- É rápida, barata e eficiente.

Desvantagens:

- Os conglomerados raramente são do mesmo tamanho, dificultando o controle da amplitude da amostra.

Exs: produção de uma fábrica dividida em linhas de montagem, trabalhadores de uma empresa divididos por área, médicos de um município divididos por hospitais, população de um município dividida em distritos.

AMOSTRAGEM POR CONGLOMERADOS



Clustered Population

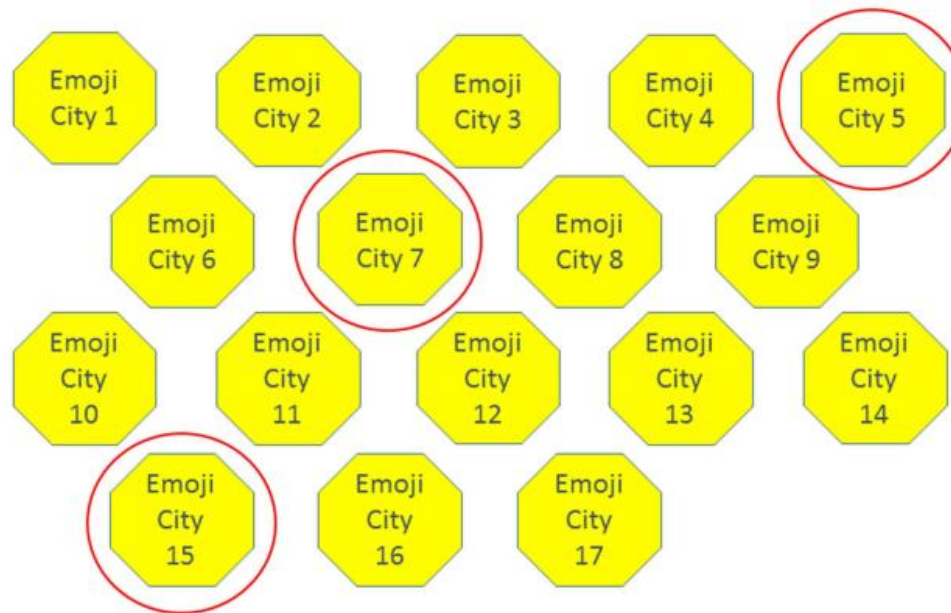
Sample
(2 clusters)

AMOSTRAGEM POR CONGLOMERADOS

Cluster sampling: stage 1

$n = 3$ clusters (cities), randomly selected

Include all emoji houses in each of the 3 cities sampled.

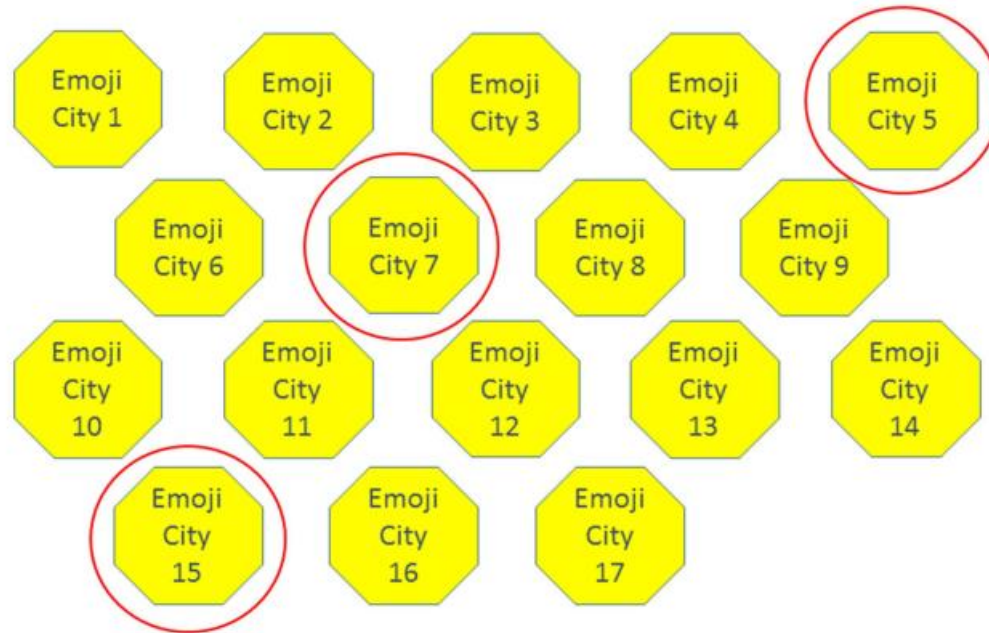


AMOSTRAGEM POR CONGLOMERADOS

Cluster sampling: stage 2

$n = 3$ clusters (cities), randomly selected

Then, use another sampling method to select a subset of emoji houses in each of the 3 cities sampled.



AMOSTRAGENS NÃO ALEATÓRIAS

- Esse tipo de amostragem é muitas vezes empregado pela simplicidade ou impossibilidade de obtermos amostras probabilísticas.
- Ela é subjetiva, baseada nos critérios e julgamentos do pesquisador, e a variabilidade amostral não pode ser estabelecida com precisão.

AMOSTRAGEM POR CONVENIÊNCIA

- É empregada quando há participação voluntária ou os elementos da amostra são escolhidos por uma questão de conveniência ou simplicidade, por exemplo, amigos, vizinhos ou estudantes.

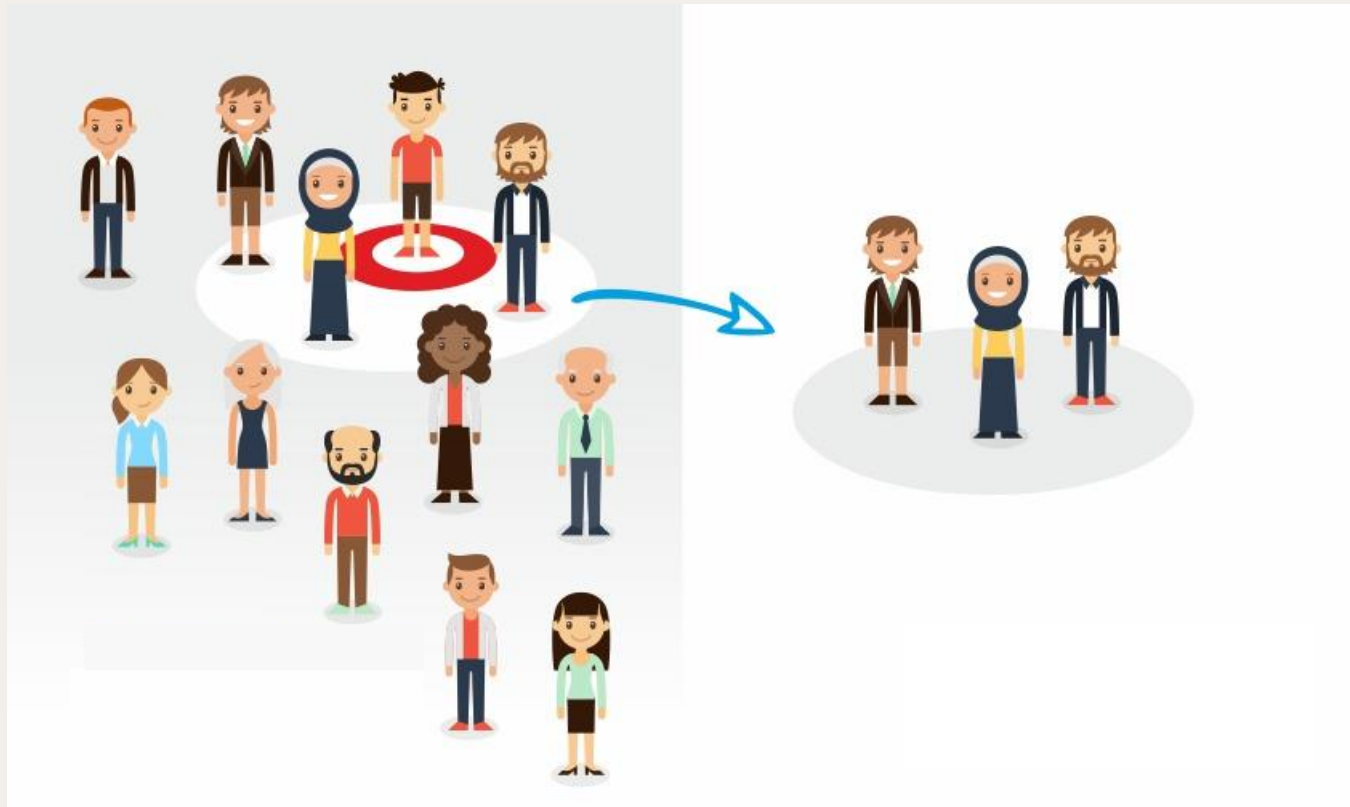
Vantagem:

- Obter informações de maneira rápida e barata.

Desvantagem:

- O processo amostral não garante que a amostra seja representativa da população, devendo ser empregada apenas em situações extremas e em casos especiais que justifiquem a sua utilização.

AMOSTRAGEM POR CONVENIÊNCIA



AMOSTRAGEM POR COTAS

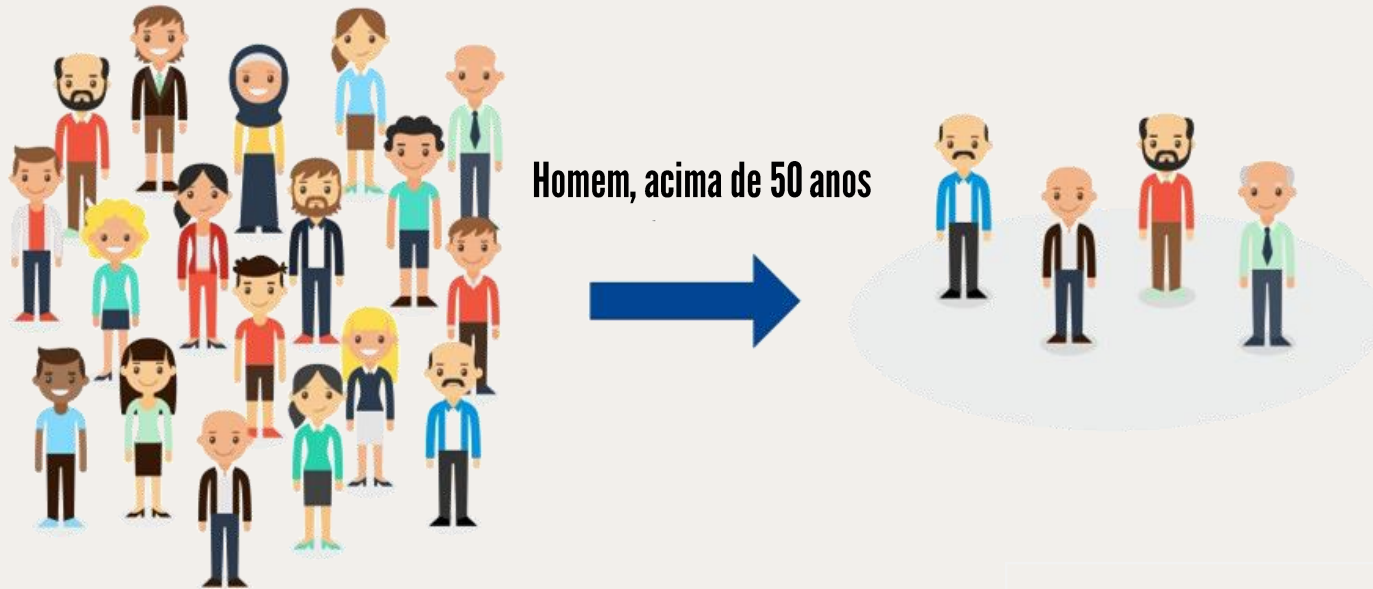
- A população é vista de forma segregada, dividida em diversos subgrupos.
- Seleciona-se uma cota de cada subgrupo proporcional ao seu tamanho.
 - A seleção não precisa ser aleatória.
- Para compensar a falta de aleatoriedade na seleção, costuma-se dividir a população num grande número de subgrupos.
 - Ex: em uma pesquisa socioeconômica, a população pode ser dividida por localidade, por nível de instrução, por faixas de renda etc.
- Um dos mais utilizados em pesquisa de mercado e de opinião eleitoral.

AMOSTRAGEM POR COTAS

1. Seleccionamos as variáveis de controle ou as características da população consideradas relevantes para o estudo em questão;
2. Determinamos a proporção da população (%) para cada uma das categorias das variáveis relevantes;
3. Dimensionamos as cotas (número de pessoas a serem entrevistadas que possuem as características determinadas) para cada entrevistador, de modo que a amostra tenha proporções iguais a da população.

- **Vantagens:** baixo custo, rapidez, conveniência ou facilidade para o entrevistador selecionar elementos.
- **Desvantagens:** como a seleção não é aleatória, não há garantia de que a amostra seja representativa da população.

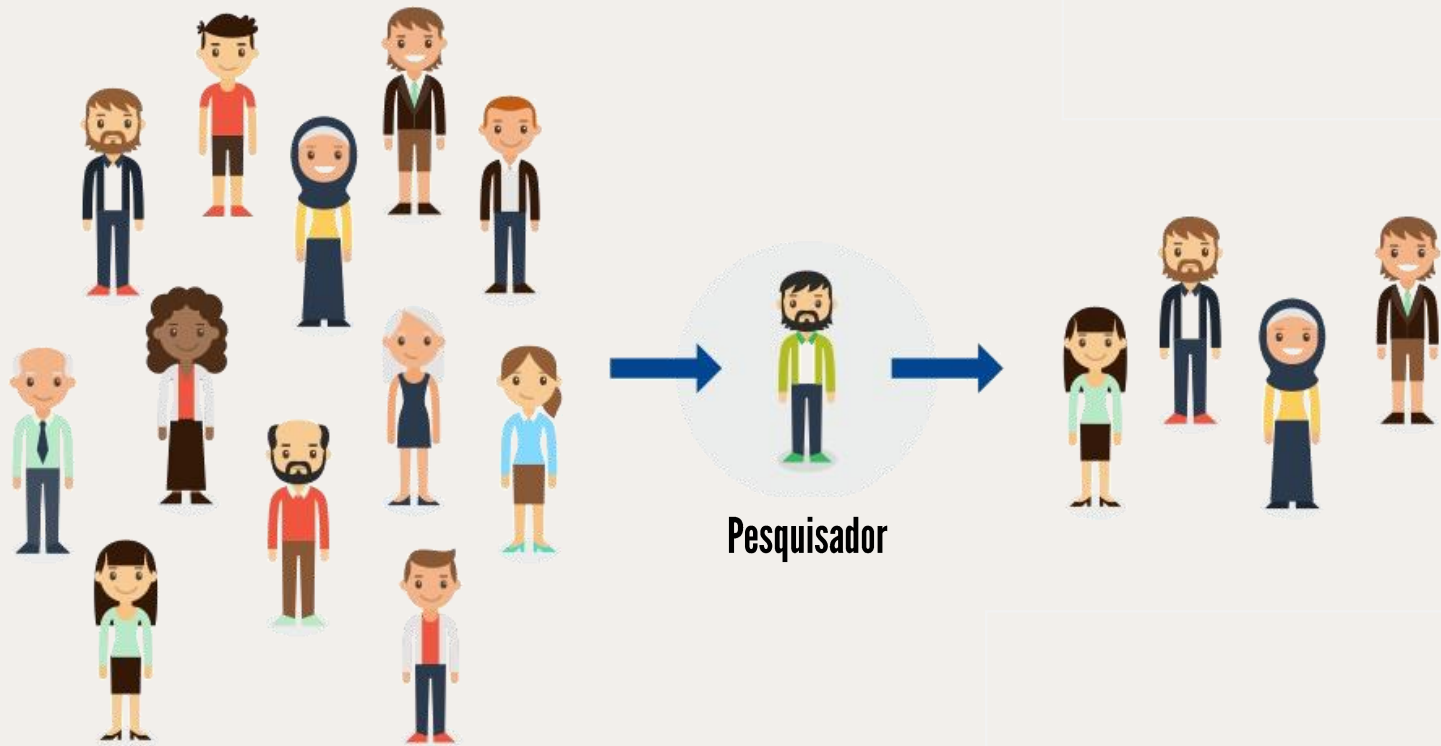
AMOSTRAGEM POR COTAS



AMOSTRAGEM POR JULGAMENTO

- Os elementos escolhidos são aqueles julgados como típicos da população que se deseja estudar.
 - Por exemplo, no estudo sobre a produção científica dos departamentos de ensino de uma universidade, um estudioso sobre o assunto pode escolher os departamentos que ele considera serem aqueles que melhor representam a universidade em estudo.
- Requer conhecimento da população e dos elementos selecionados.

AMOSTRAGEM POR JULGAMENTO

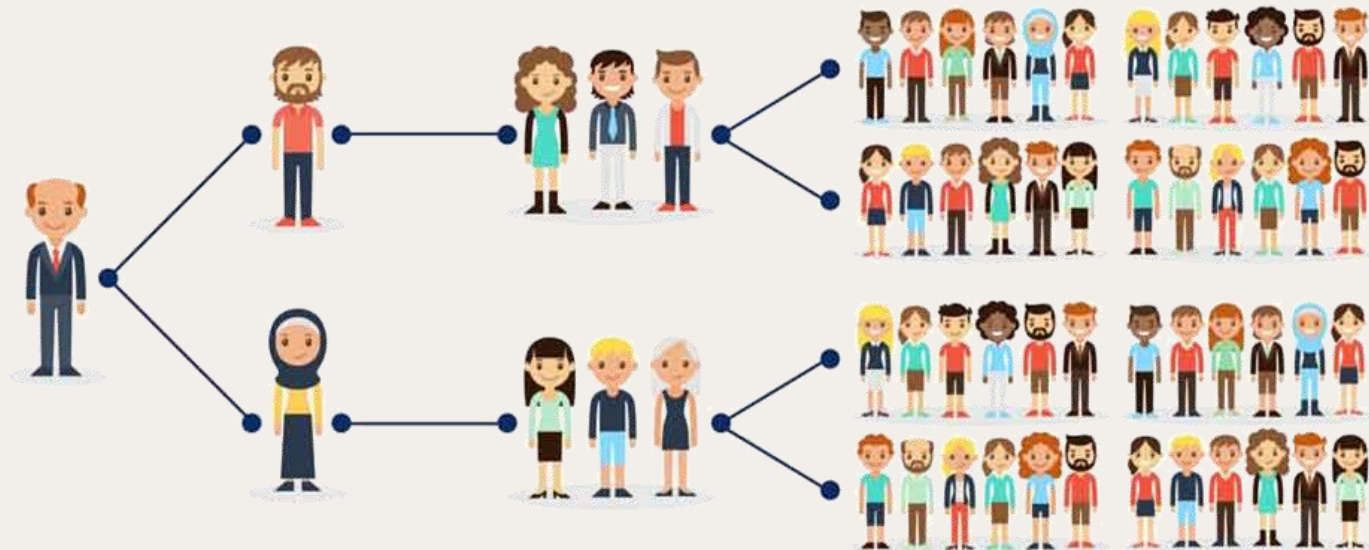


AMOSTRAGEM POR BOLA DE NEVE

- Muito utilizada quando os elementos da população são raros, de difícil acesso ou desconhecidos.
- Devemos identificar um ou mais indivíduos da população-alvo e eles identificarão outros indivíduos pertencentes à mesma população. O processo é repetido até que seja alcançado o objetivo proposto.
 - Exemplo: determinada empresa está recrutando profissionais com perfil específico. O grupo contratado inicialmente indica outros profissionais com o mesmo perfil. Processo se repete até que seja contratada o número necessário de funcionários.

Vantagens: permite ao pesquisador localizar a característica desejada da população, facilidade de aplicação, baixo custo, eficiente em penetrar em populações de difícil acesso.

AMOSTRAGEM POR BOLA DE NEVE



CONCEITO DE ERRO AMOSTRAL

Para a determinação do tamanho da amostra, o pesquisador precisa especificar o erro amostral tolerável, ou seja, o quanto ele admite errar na avaliação dos parâmetros de interesse.

Por exemplo, na divulgação de pesquisas eleitorais, é comum encontrarmos: a presente pesquisa tolera um erro de 2%.

Quando a pesquisa aponta determinado candidato com 20% de preferência do eleitorado → 18% a 22%, ou seja, $20\% \pm 2\%$.

Consideraremos sempre o erro amostral sob 95% de probabilidade. Assim, se fixarmos o erro amostral tolerável em 2%, estaremos afirmando que uma estatística, calculada com base na amostra a ser selecionada, não deve diferir do parâmetro em mais que 2%, com 95% de probabilidade.

FÓRMULA PARA O TAMANHO MÍNIMO DA AMOSTRA

Um primeiro cálculo do tamanho da amostra pode ser feito, mesmo sem se conhecer o tamanho da população, através da seguinte expressão:

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

N : tamanho (número de elementos) da população;

n : tamanho (número de elementos) da amostra;

n_0 : uma primeira aproximação para o tamanho da amostra;

E_0 : erro amostral tolerável.

FÓRMULA PARA O TAMANHO MÍNIMO DA AMOSTRA

- Se a população for muito grande, digamos mais que 20 vezes o valor calculado em n_0 , então n_0 já pode ser adotado como o tamanho da amostra $n = n_0$.
- Caso contrário é sugerido a seguinte correção:

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

EXEMPLO

Planeja-se um levantamento por amostragem para avaliar diversas características (parâmetros) da população das $N = 200$ famílias moradoras de um certo bairro.

- Os principais parâmetros são proporções, tais como porcentagem de famílias que usam programas de alimentação popular, porcentagem de famílias que moram em casas próprias, etc.
- *Qual deve ser o tamanho mínimo de uma amostra aleatória simples para que possamos admitir, com 95% de probabilidade, que os erros amostrais não ultrapassem 4%?*

EXEMPLO

População de 200 famílias:

$$n_0 = \frac{1}{(0,04^2)} = 625$$

A população é 20 vezes maior que 625? Não!

$$n = \frac{200 \cdot 625}{200 + 625} = 152$$

EXEMPLO

A população é 20 vezes maior que 625? Sim!

- E se a população agora fosse de $N = 200.000$ famílias?

$$n = \frac{200000 \cdot 625}{200000 + 625} = 623$$

**Perceba que não
precisaríamos fazer a
correção!**

- Muito próximo do n_0 .

Para garantir um erro amostral não superior a 4%, foi necessário uma amostra abrangendo 76% (152/200) da população quando o N era de 200, e quando o N era de 200.000 foi suficiente apenas uma amostra de 0,3% (623/200000) da população.

TAMANHO DA AMOSTRA EM SUBGRUPOS DA POPULAÇÃO

- Quando precisamos efetuar estimativas sobre partes (subgrupos) da população, é necessário calcular o tamanho da amostra para cada uma dessas partes.
- O tamanho total da amostra vai corresponder a soma dos tamanhos das amostras dos subgrupos.

TAMANHO DA AMOSTRA EM SUBGRUPOS DA POPULAÇÃO

Vamos supor agora que se queira fazer estimativas isoladas para os seguintes estratos: centro da cidade, bairros e periferia, mantendo-se a mesma precisão para cada estrato ($E_0 = 0,04$).

$$n_0 = \frac{1}{(0,04^2)} = 625$$

$$n_{Total} = 3 \cdot (625) = 1875$$

ATIVIDADE

Para cada uma das amostras abaixo, informar o tipo do processo de amostragem:

P - Amostragem probabilística

NP - Amostragem não probabilística

Para uma pesquisa sobre os hábitos dos estudantes, construí uma amostra com o seguinte procedimento:

- () Todos os meus colegas da faculdade (tenho telefone e e-mail de todos eles).
- () Fiquei na única porta de entrada da escola abordando todos os meus conhecidos.
- () Fiquei na única porta de entrada da escola e cada 12 pessoas que entravam, eu abordava uma.
- () Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista.
- () Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados.

ATIVIDADE

Para cada uma das amostras abaixo, informar o tipo do processo de amostragem:

P - Amostragem probabilística

NP - Amostragem não probabilística

Para uma pesquisa sobre os hábitos dos estudantes, construí uma amostra com o seguinte procedimento:

(NP) Todos os meus colegas da faculdade (tenho telefone e e-mail de todos eles).

(NP) Fiquei na única porta de entrada da escola abordando todos os meus conhecidos.

(P) Fiquei na única porta de entrada da escola e cada 12 pessoas que entravam, eu abordava uma.

(P) Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista.

(P) Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados.

ATIVIDADE - PLICKER

Um estudo sobre o desempenho dos vendedores de uma grande cadeia de lojas de varejo está sendo planejado. Para tanto, deve ser colhida uma amostra probabilística dos vendedores. Classifique cada uma das amostras abaixo conforme a seguinte tipologia:

(AAS) Amostragem aleatória simples (AS) Amostragem Sistemática (AE) Amostragem estratificada (AC) Amostragem por conglomerados.

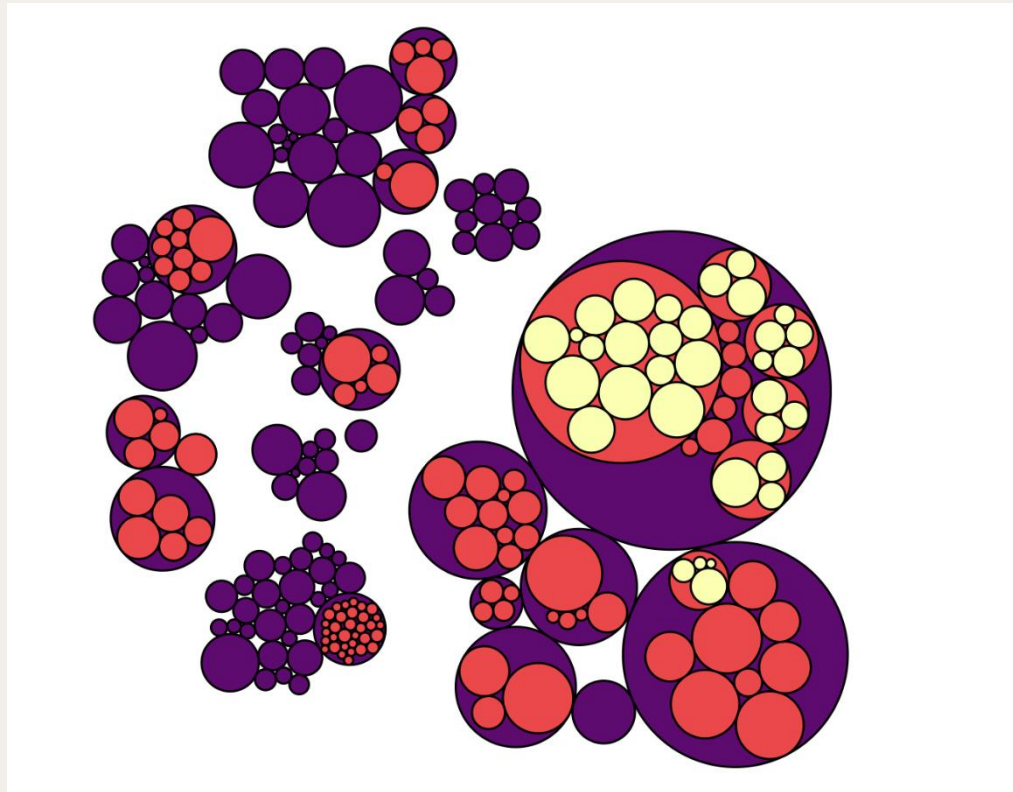
() Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei todos vendedores que ocupavam posições múltiplas de 15 (15^a posição, 30^a posição, 45^a posição, 60^a posição, 75^a posição, 90^a posição, 105^a posição, etc.)

() Escolhi casualmente 3 lojas da rede. A amostra foi composta de todos os vendedores que atuam em cada uma destas 3 lojas.

() Em cada uma das lojas, identifiquei todos os vendedores (lista de vendedores por loja). Selecionei aleatoriamente k vendedores da loja. Onde k é um número inteiro proporcional à quantidade de vendedores da loja.

() Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei aleatoriamente N vendedores.

ARTE DO DIA FEITA EM R



<https://www.r-graph-gallery.com/315-hide-first-level-in-circle-packing.html>

REFERÊNCIAS BIBLIOGRÁFICAS

- BARBETTA, Pedro Alberto. Estatística aplicada às ciências sociais. Ed. UFSC, 2008.
- DANCEY, Christine P.; REIDY, John G.; ROWE, Richard. Estatística Sem Matemática para as Ciências da Saúde. Penso Editora, 2017.
- MAGNUSSON, Willian E. Estatística [sem] matemática: a ligação entre as questões e a análise. Planta, 2003.