Aula 04

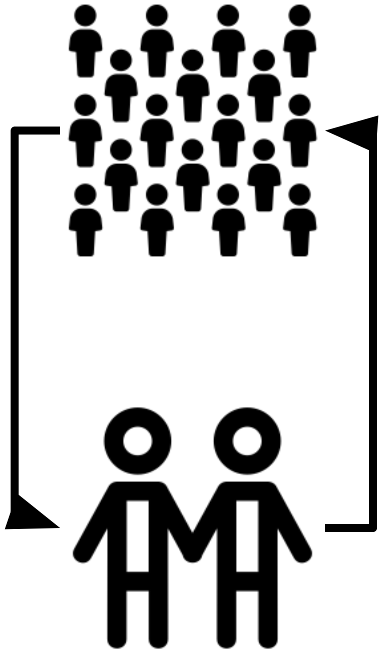
INTRODUÇÃO À ESTATÍSTICA INFERENCIAL

Módulo 6 - Estatística I

Romero Florentino de Carvalho

Estatística inferencial

Examinar uma parte



Formular julgamentos sobre o todo

Estimação de parâmetros

Teste de hipóteses

Tirar conclusões sobre uma população a partir de uma amostra.

Situação 1: Como saber se uma sopa está corretamente salgada?

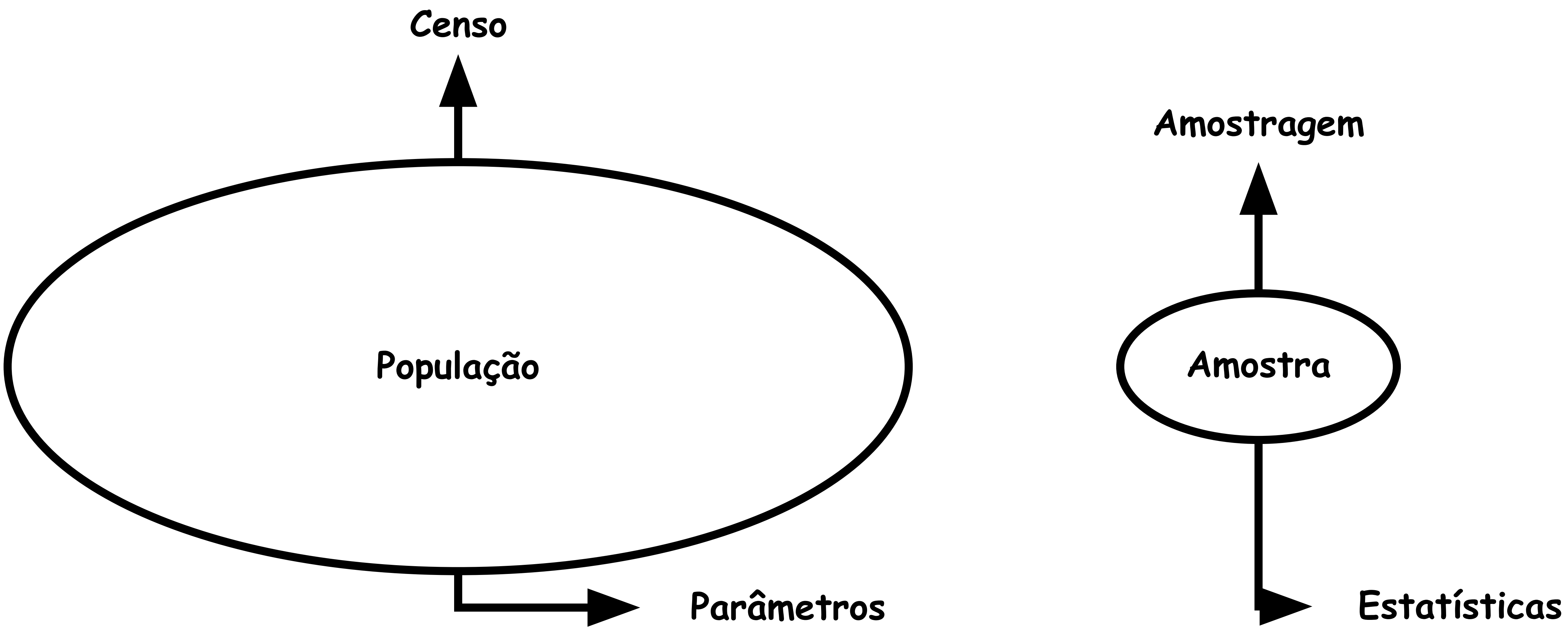
Situação 2: Quantos mL de sangue são necessários para um exame?

Situação 3: Como saber a % da população com doença emocional?

Situação 4: Como saber a popularidade de um político?

Situação 5: Como saber a média de idade dos alunos de psicologia do ES?

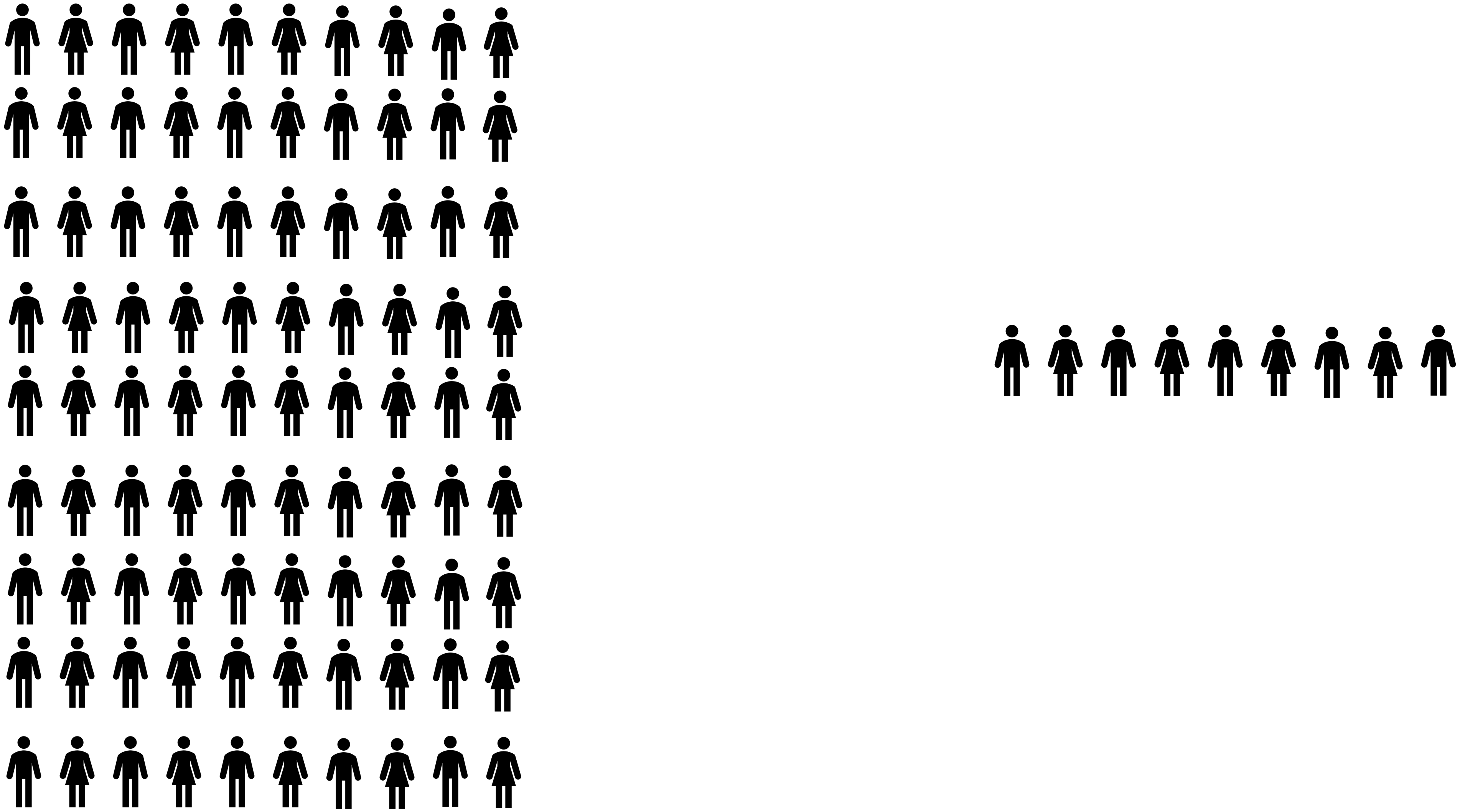
Situação 6: Como avaliar a relação entre a renda média e nível educacional?



Pergunta: Como garantir a melhor amostra possível?

- **Amostragem probabilística:**
  - Todos os elementos da população apresentam probabilidade maior que zero de serem selecionados. (aleatória simples, estratificada, sistemática e por conglomerados);
- **Amostragem não probabilística:**
  - Quando não há probabilidade clara/conhecida de seleção dos elementos. Os elementos são escolhidos subjetivo ou por julgamento (acidental, intencional, por cotas).

Amostragem Aleatória Simples



Amostragem Estratificada

Sexo	População	10%	Amostra
M	54	5,4	5
F	36	3,6	4
Total	90	9	9

Amostragem Sistemática

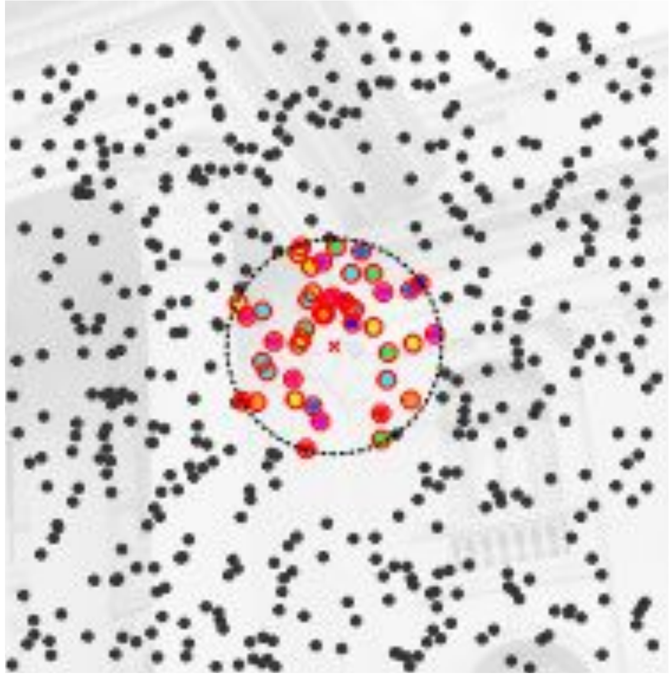
Amostra de 9 casas:

Como encontrar?

Amostragem por conglomerados

Vitória
Cachoeiro de Itapemirim
Brejetuba
Colatina
Aracruz
Marataízes
Nova Venécia
Serra

Amostragem por conveniência

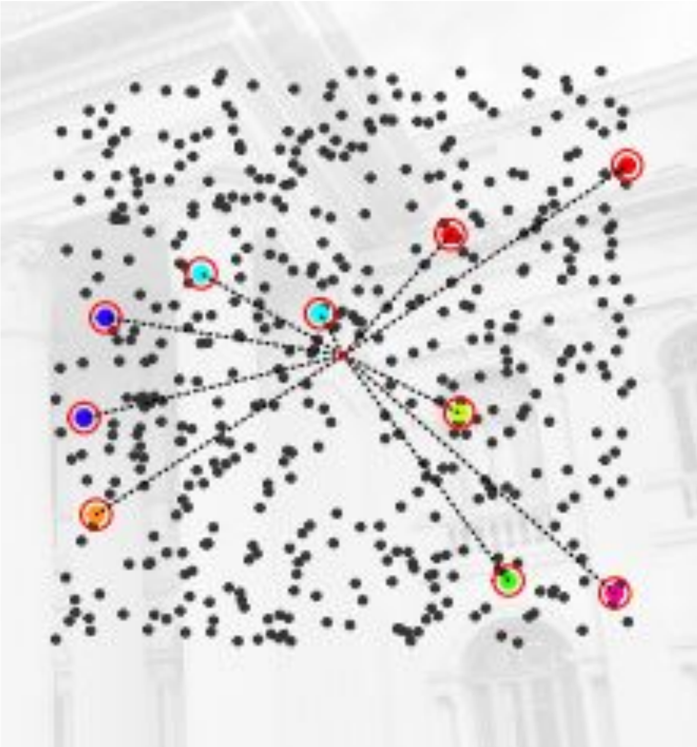


Cada unidade é escolhida sem probabilidade especificada ou conhecida.

Ex.: Perguntar a 100 pessoas na rua o que acham sobre uma celebridade concorrendo à presidência.

Aplicações: pesquisa com voluntário, animais selvagens capturados, paciente, enquete em rede social.

Amostragem intencional ou julgamento

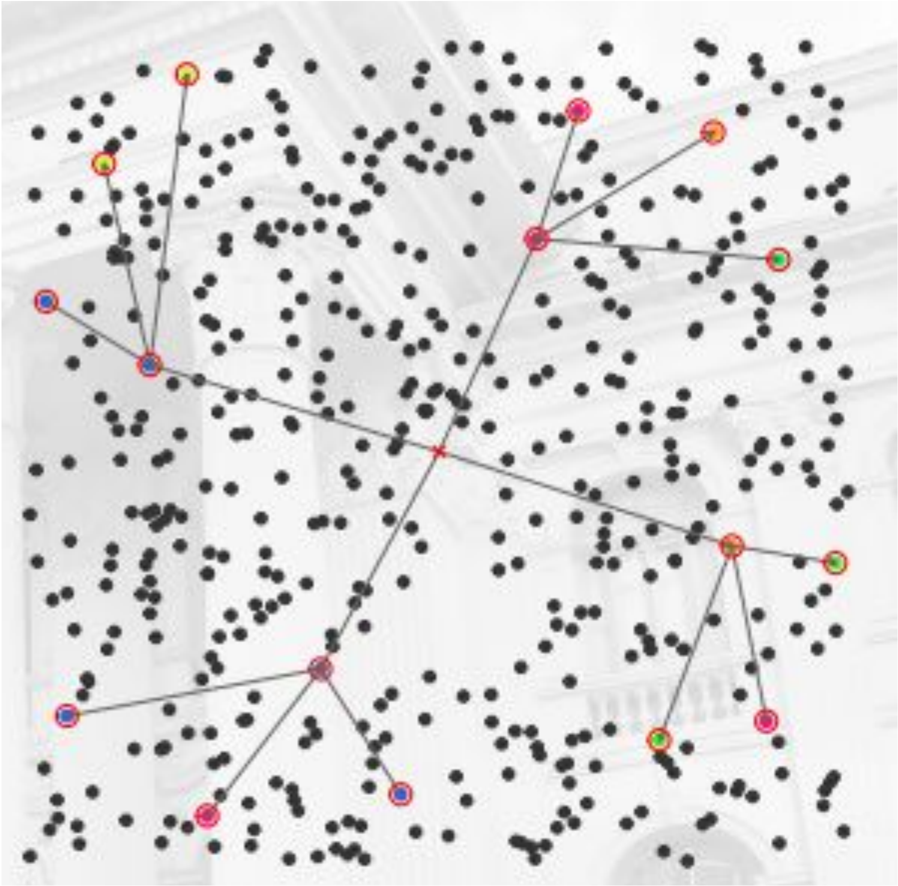


Escolha feita por um **expert**.

Ex.: Escolher uma cidade para representar o universo urbano e rural do país.

Aplicações: escolher 3 senadores para saber a opinião sobre um tema, estudo de uma doença genética rara.

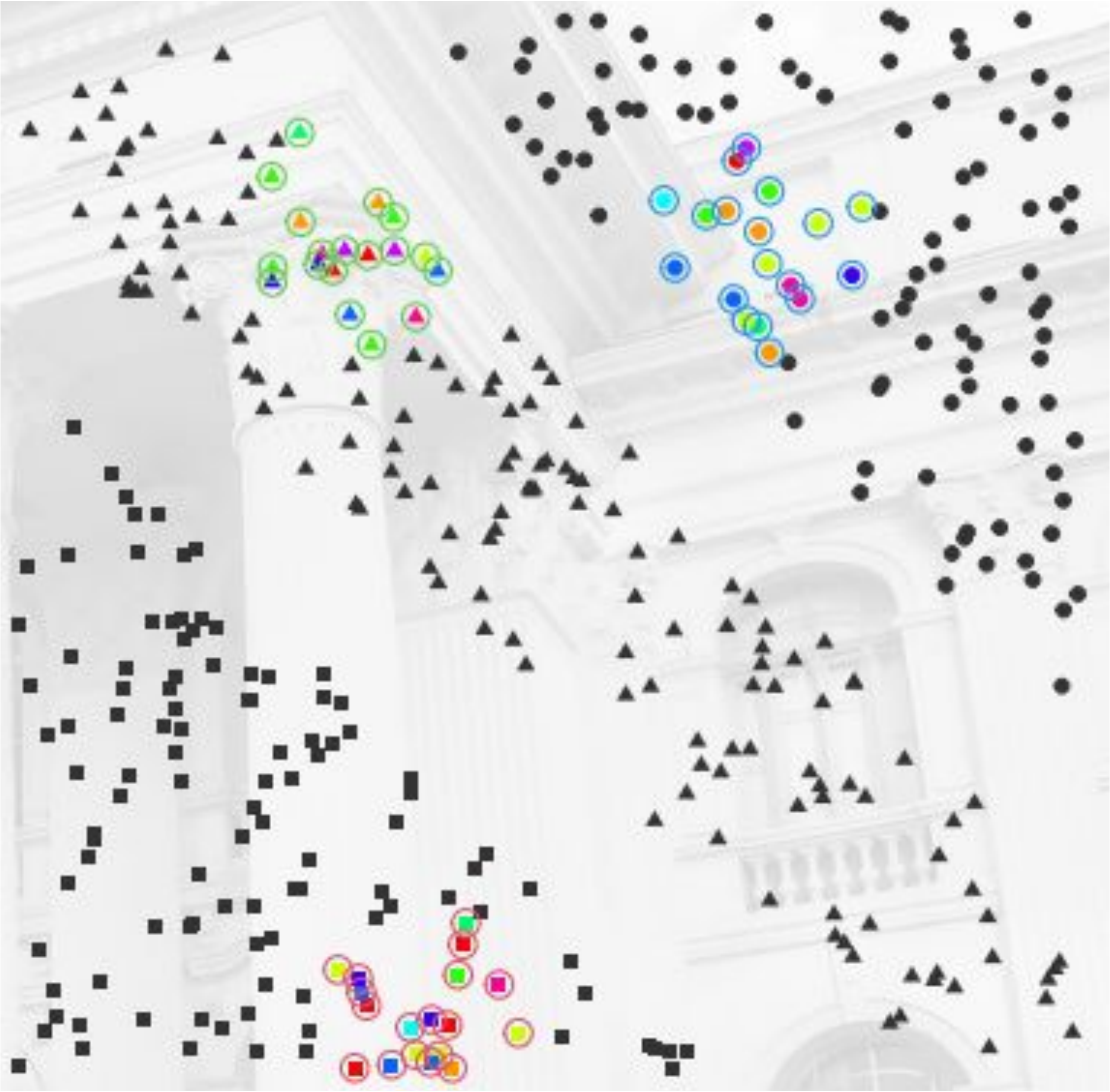
Amostragem por bola de neve



Pesquisador identifica as unidades amostrais e elas indicam similares

Aplicações: Condições de saúde de imigrantes, pacientes com HIV, condições psicológicas de famílias com filhos PNE.

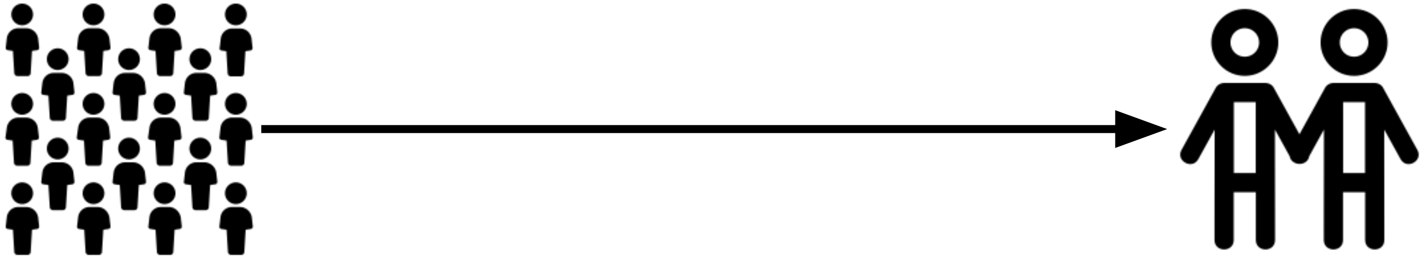
Amostragem por cotas



Populações em subgrupos, avalia-se a proporção e seleciona uma amostra em cada grupo respeitando-a.

Ex.: 50 homens e 50 mulheres.

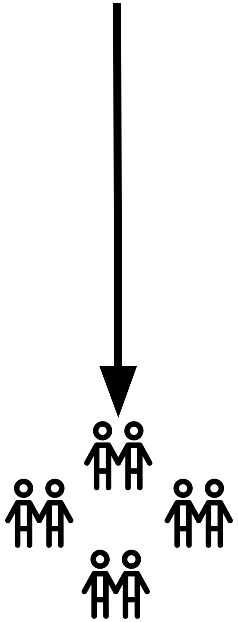
Aplicações: pesquisa eleitoral, mercadológica, coleta de água em diferentes pontos de um rio.



Estatísticas amostrais:

$\bar{x}, \sigma_{\bar{x}}$

Variabilidade amostral



Distribuições amostrais

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

- O valor esperado para a média da distribuição amostral é igual à média da população:

$$\mu_{\bar{x}} = \mu_x$$

- O desvio-padrão amostral da média é:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

- X1,X2,X3 -> (X1+X2+X3)/3
  - X1,X2 -> (X1+X2)/2
  - X1,X3 -> (X1+X3)/2
  - X2,X3 -> (X2+X3)/2

  ((X1+X2)/2+(X1+X3)/2+(X2+X3)/2)3 = (X1+X2+X3)/3

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

- O valor esperado para a média da distribuição amostral é à proporção da população:

$$\bar{p} = p$$

- O desvio-padrão amostral da proporção é:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Capacidade de inferir pela amostra?
Depende do conhecimento da distribuição amostral.

Tendência geral:
Distribuições de médias e proporções se apresentarem aproximadamente normais.

Distribuições não-normal:
Distribuições amostrais aproxima da normal para amostras grandes.

Resultado notável:
Não é necessário conhecer a distribuição da população para fazer inferência sobre ela a partir de dados amostrais.

ÚNICA RESTRIÇÃO:

Tamanho da amostra grande (n>=30).

TEOREMA DO LIMITE CENTRAL

Se X1, X2,...,Xn, for uma amostra aleatória retirada de uma população normal com média μ e variância σ^2, a distribuição amostral da média $\bar{X}$, terá forma dada por:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

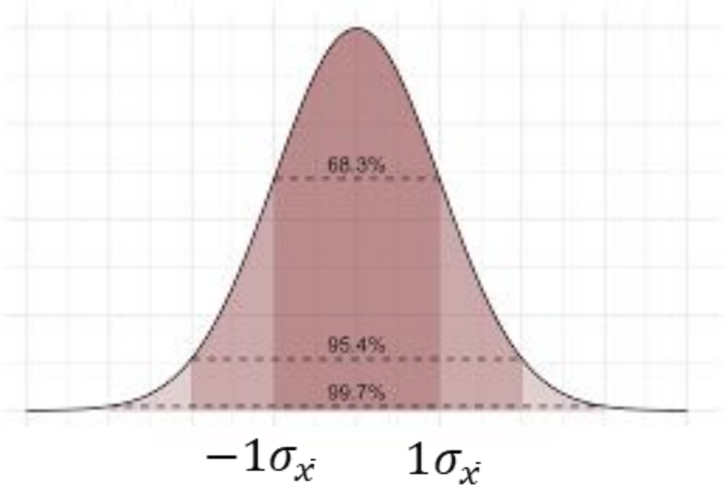
APLICAÇÃO DO TLC

- Ex.: Uma população muito grande tem μ = 20 e desvio-padrão σ=1,4. Extrai-se uma amostra de 49 observações. Responda:
  - A média da distribuição amostral é: $\mu_{\bar{x}} = \mu_x = 20;$
  - O desvio-padrão da distribuição amostral: $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1.4}{\sqrt{49}} = 0{,}2;$
  - A percentagem de dados que diferirão por mais de 0,2 da média.

APLICAÇÃO DO TLC

- Ex.: Uma população muito grande tem μ = 20 e variância σ=1,4. Extrai-se uma amostra de 49 observações. Responda:
  - *A percentagem de dados que diferirão por mais de 0,2 da média.*

$$\frac{19,8-20}{0,2} = -1\sigma_{\bar{x}}$$



$$\frac{20,2-20}{0,2} = 1\sigma_{\bar{x}}$$

Resposta: 31,7% diferirão da média por mais de 0,2.