



# Probabilidade e Estatística

## Introdução

# Nesta aula

- Estatística: o quê? Para quê?
- População × amostra.
- Parâmetro × estatística.
- Planejamento de um estudo estatístico.
- Técnicas de amostragem.

# Estatística: o quê? Para quê?

- Variação : *qual a altura de um homem brasileiro?*
- Incerteza : *qual o naipe da carta que está virada para baixo na mesa?*
- Correlação : *o nível de instrução está ligado ao consumo de **chocolate**?*
- Causação : *o nível de instrução é causado pelo consumo de chocolate?*
- Observação : *pessoas que fumam mais vivem menos?*
- Experimentação : *ratos expostos a mais nicotina vivem menos?*
- Descrição e visualização : *altura média do homem brasileiro? Mediana? Desvio padrão?*
- Exploração : *análise exploratória de dados (EDA)*
- Previsão (inferência) : *que conclusões podemos tirar a respeito de possíveis novos casos?*



<http://xkcd.com/552/>

# População × amostra

- População: *todos* os indivíduos, resultados, respostas ou medições.
- Amostra: um *subconjunto* da população.
  - Representativa: boa para tirar conclusões sobre a população.
  - Aleatória: escolhida ao acaso.

# Parâmetro × estatística

- Parâmetro: um número característico da *população*.
- Estatística: um número característico da *amostra*.

# Planejamento de um estudo estatístico

1. Identificar a população e as variáveis de interesse.
2. Planejar a coleta de dados (inclusive escolha da amostra).
3. Coletar os dados.
4. Descrever e explorar os dados coletados; formular hipóteses.
5. Interpretar os dados, testar hipóteses, inferir conclusões sobre a população.
6. Identificar possíveis erros.

# Planejamento de um estudo estatístico

## Estudo observacional

- Observar.
- Medir.
- Não modificar as condições existentes.

## Estudo experimental (experimento)

- Indivíduos são *unidades experimentais*.
- *Grupo de tratamento* sofre uma intervenção.
- *Grupo de controle* não sofre intervenção (ou recebe um *placebo*).
- Comparar respostas entre os grupos.

# Planejamento de um estudo estatístico

## Coleta de dados

- Observação: contagem, medição, etc.
- Simulação: usar modelo físico, matemático ou computacional.
  - Exemplo: simular lançamento de dados em um computador.
  - Exemplo: simular acidente de carros com bonecos.
- Pesquisa: perguntas, questionários, telefonemas.
  - Cuidado com perguntas tendenciosas.
  - Cuidado com seleção tendenciosa de indivíduos.
  - Exemplo: telefonemas só atingem quem tem telefone!

# Planejamento de um estudo estatístico

## Possíveis problemas

- *Tamanho da amostra*: deve ser grande o bastante.
- *Efeito placebo*: indivíduo reage a tratamento que não recebeu.
- *Efeito Hawthorne*: indivíduo muda comportamento só porque está participando de experimento.
- *Variáveis de confusão*: fatores ocultos provocam correlação entre variáveis observadas.
  - Exemplo: o *aumento na quantidade de assaltos* na praia está correlacionada com o *aumento na venda de sorvetes*.
  - Na verdade, o *aumento da temperatura* causa as duas coisas. A temperatura é uma variável de confusão.

# Planejamento de um estudo estatístico

## Técnicas

- *Cegamento*: indivíduo não sabe se está recebendo tratamento ou placebo.
- *Duplo cegamento*: indivíduo e pesquisador não sabem quem está recebendo tratamento e quem está recebendo placebo.
- *Aleatorização*: os indivíduos de cada grupo (controle e tratamento) são escolhidos ao acaso.
  - A idéia é que os grupos só vão diferir no tratamento.
  - A aleatorização pode não ser total.
  - Exemplo: mesma quantidade de homens e mulheres em cada grupo.

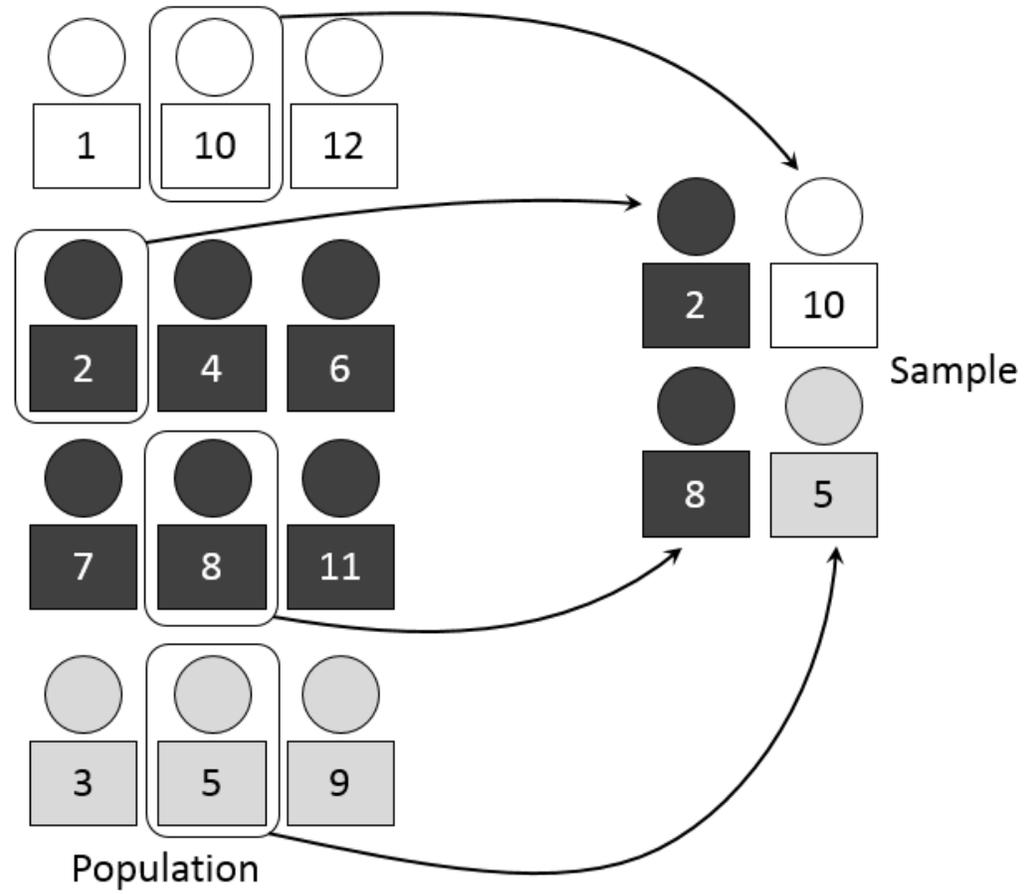
# Técnicas de amostragem

- *Censo*: examina toda a população.
- *Amostragem*: examina parte da população.
- Amostra *representativa* × amostra *tendenciosa*.
- Amostragem *aleatória*: todos os elementos da população têm a mesma chance de ser escolhidos.
- Amostragem *com reposição*: o mesmo elemento da população pode aparecer mais de uma vez na amostra.
- Amostragem *sem reposição*: não.

# Técnicas de amostragem

## Amostragem estratificada

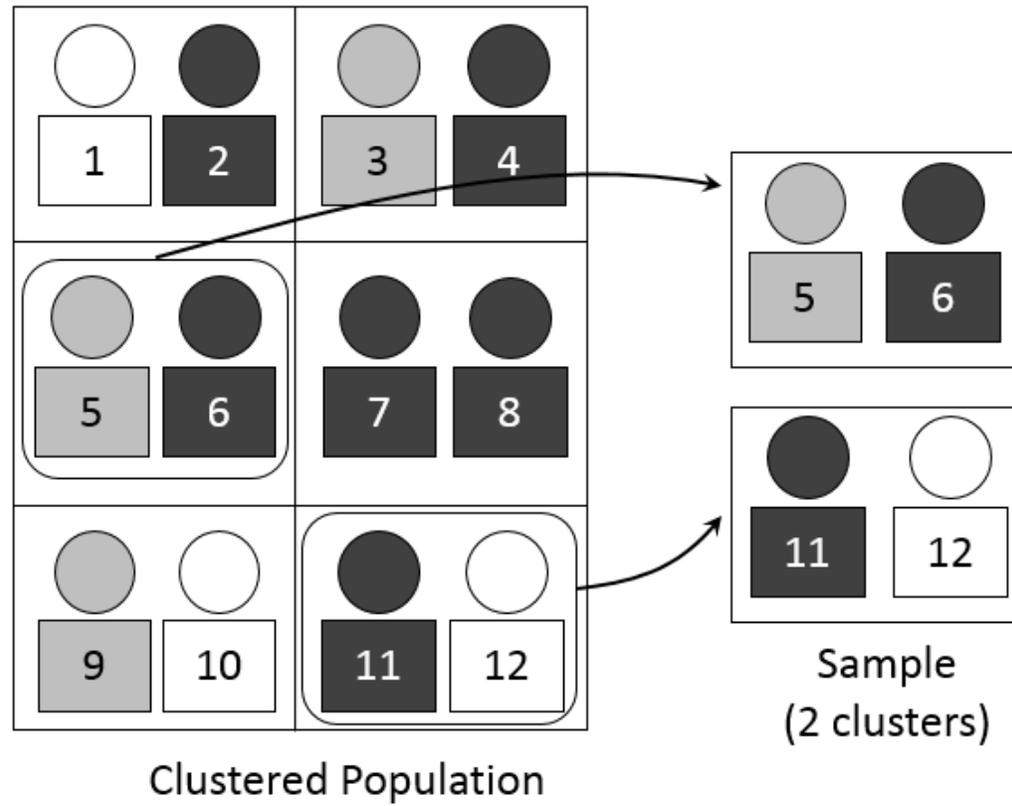
- Cada segmento (estrato) da população é representado na amostra.
- De cada estrato, é selecionada uma amostra aleatória.
- As amostras devem ter tamanhos proporcionais aos tamanhos dos estratos.



# Técnicas de amostragem

## Amostragem por conglomerado

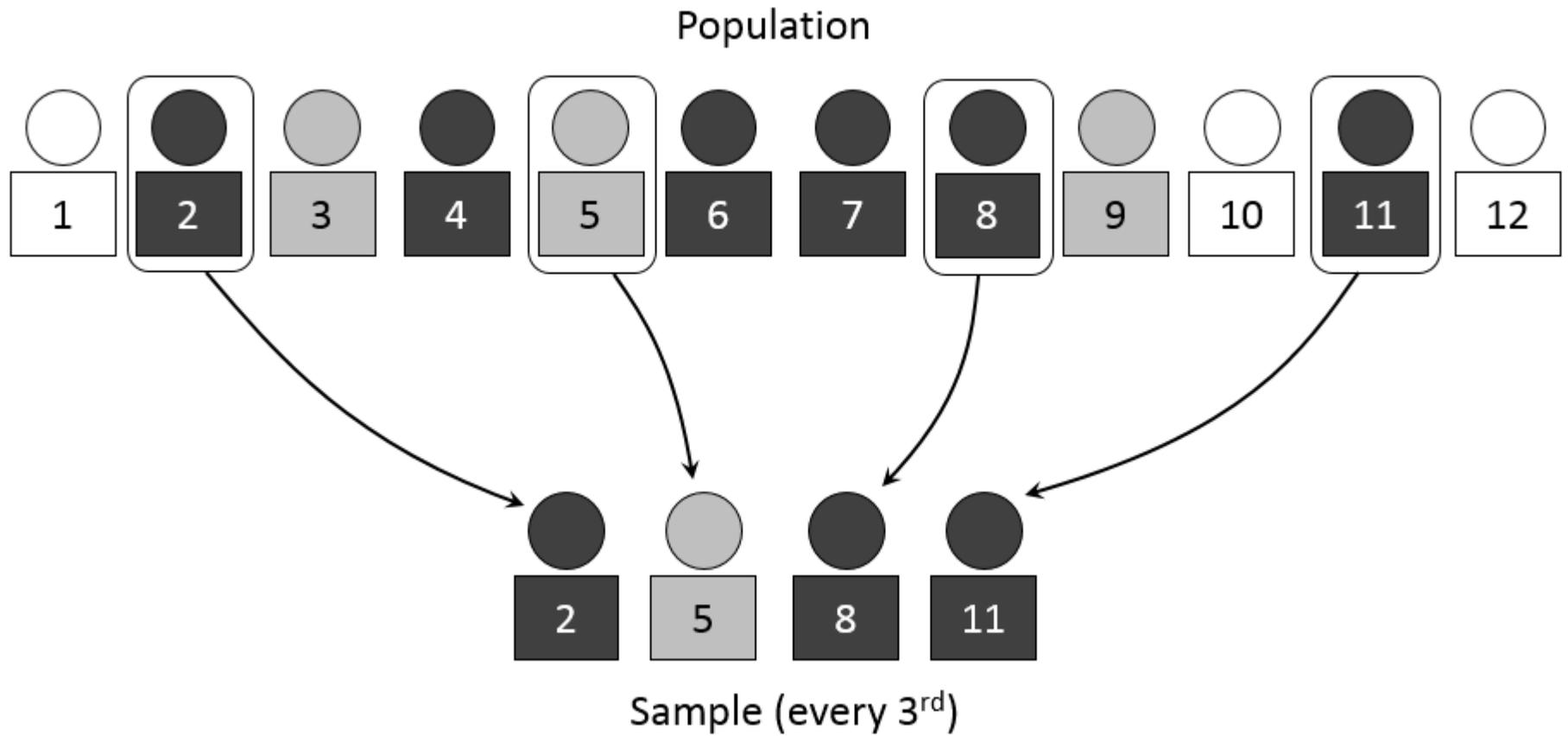
- Existem subgrupos na população.
- Os subgrupos são semelhantes.
- Escolhemos um ou mais subgrupos para representar a população.



# Técnicas de amostragem

## Amostragem sistemática

- Dividir a população em grupos.
- Quantidade de grupos = tamanho da amostra.
- Sortear um número aleatório  $n \leq$  tamanho do grupo.
- Escolher o  $n$ -ésimo elemento de cada grupo.



# Técnicas de amostragem

## Amostragem por conveniência

- Escolher elementos que são mais acessíveis.
- Costuma gerar amostras tendenciosas.

# Técnicas de amostragem

## Mais informações

[https://en.wikipedia.org/wiki/Sampling\\_\(statistics\)](https://en.wikipedia.org/wiki/Sampling_(statistics))