

Métodos de amostragem probabilística

Prof. Wagner Hugo Bonat

Departamento de Estatística
Universidade Federal do Paraná



Motivação

- Vivemos em uma era na qual a disponibilidade e o acesso a **dados** não têm precedentes na história.

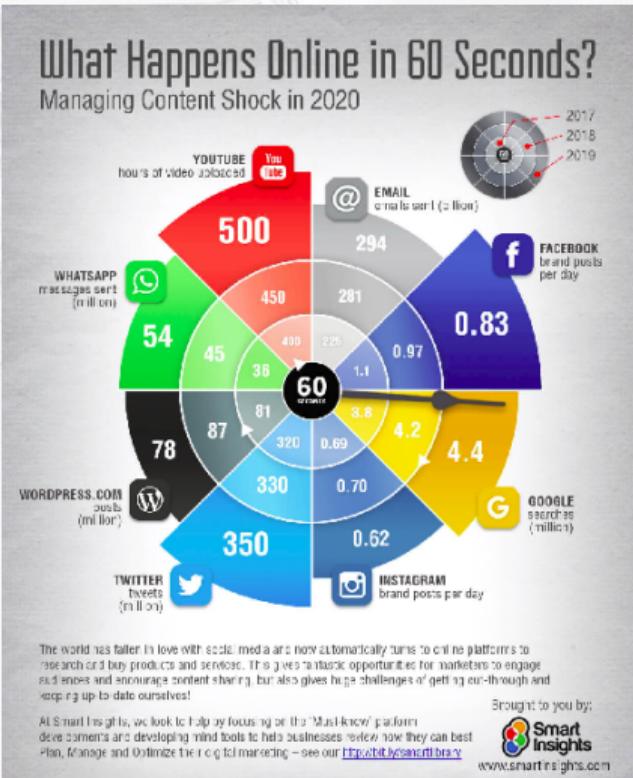


Figura 1. Extraído de smart insights.

De onde vêm os dados?

- ▶ Fonte de dados primárias:
 - ▶ Estudo de caso.
 - ▶ Experimentos.
 - ▶ Pesquisas.
- ▶ Fonte de dados secundários (já coletados).
 - ▶ Registros administrativos.
 - ▶ Pesquisa bibliográfica.
 - ▶ *Big Data* (banco de dados corporativos, etc).



Figura 2. Photo by Markus Spiske from Pexels.

Planejamento da coleta de dados

- ▶ Coleta de dados observacionais.
 - ▶ Presença de seres vivos num ambiente.
 - ▶ Fenômenos climáticos.
 - ▶ Fluxo de usuários em um website.
 - ▶ Transações bancárias.
- ▶ Definição do experimento.
 - ▶ Variáveis respostas/interesse.
 - ▶ Variáveis de controle (o que afeta a resposta?).
 - ▶ Desenho do experimento e randomização.

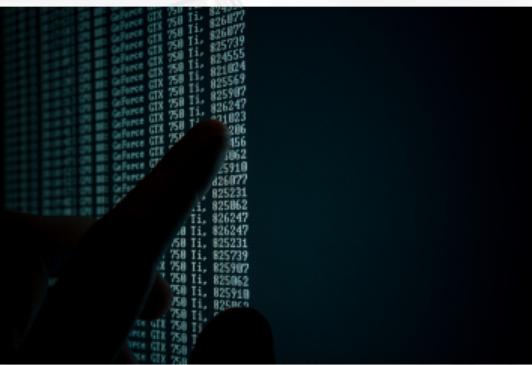


Figura 3. Foto de Vitaly Vlasov do Pexels.



Figura 4. Foto de martin Lopez do Pexels.

Dados experimentais

- ▶ Comum em ciências agrárias, biológicas, da saúde e da terra, etc.
- ▶ Coleta é feita por meio de **experimentos**.
- ▶ Características típicas:
 - ▶ Intervenção na realidade.
 - ▶ Condições controladas.
 - ▶ Observação dos efeitos das intervenções.
 - ▶ Aleatorização usada para alocar unidades em grupos: tratamento(s) × Figura 5. Photo by Chokniti Khongchum from Pexels. controle(s).
 - ▶ Custo de coleta elevado → amostras pequenas.



Figura 5. Photo by Chokniti Khongchum from Pexels.

Dados observacionais

- ▶ Comuns em ciências humanas e sociais.
- ▶ Coleta é feita por meio de **estudos observacionais**.
- ▶ Características típicas:
 - ▶ Observação “passiva” da realidade.
 - ▶ Amostragem (aleatória) é usada para selecionar unidades para observar.
 - ▶ Custo de coleta menor → amostras **grandes**.



Figura 6. Photo from Pexels.

Dados, dados ...

- ▶ Fonte de dados típicas:
 - ▶ Censo.
 - ▶ **Pesquisas amostrais.**
 - ▶ Registros administrativos.
- ▶ Característica comum → **dados planejados.**
- ▶ Novas fontes de dados (Big Data):
 - ▶ Transações (nota fiscal eletrônica).
 - ▶ Comunicações e mensagens.
 - ▶ Imagens.
 - ▶ Buscas, web-scraping, etc.
- ▶ Características comuns:
 - ▶ Grande **volume, variedade e velocidade.**
 - ▶ Dados não planejados (dados orgânicos).

Conceitos importantes

- ▶ **Pesquisa:** Coleta de informações sobre uma característica de interesse de unidades de uma população, usando métodos e procedimentos bem definidos. Acompanhado da **compilação** dessas informações em uma forma resumida útil.
- ▶ **Amostragem:** Consiste em selecionar **parte** de uma população para observar, de modo que seja possível estimar alguma coisa sobre toda a população (Thompson, 2012).
- ▶ Características desejáveis da amostra:
 - ▶ Capacidade de generalização.
 - ▶ Imparcialidade e representatividade.
 - ▶ Capacidade de **medir a precisão** das estimativas.
- ▶ Como podemos obter amostras adequadas? → Métodos de amostragem.

Métodos de amostragem

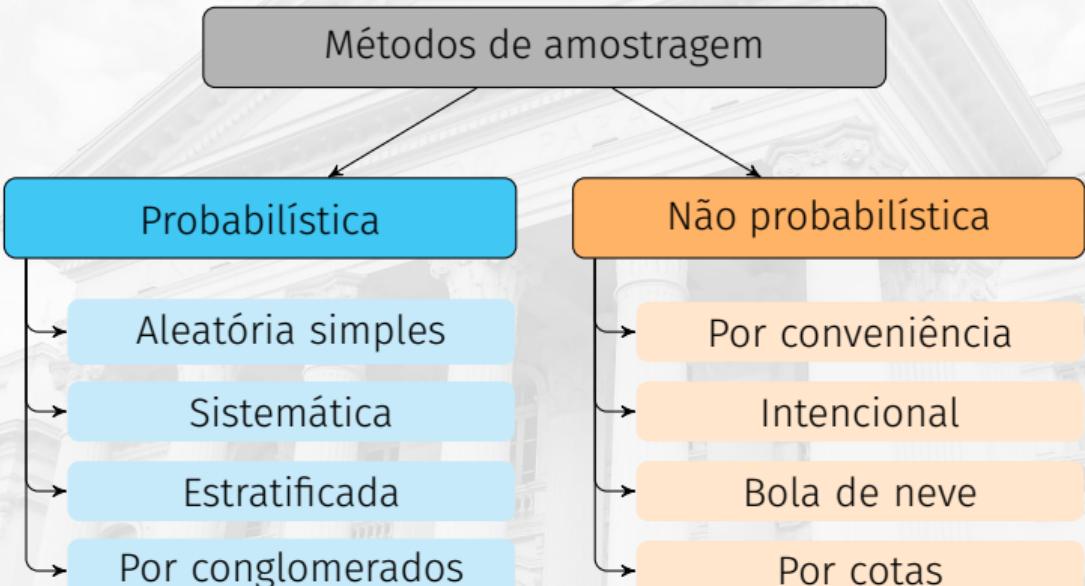


Figura 7. Métodos de amostragem mais comuns.

Amostragem aleatória simples (com ou sem reposição)

- ▶ Procedimento do método:
 1. Selecione uma unidade de U (população) com probabilidade $1/N$.
 2. Repita o passo (1) $n - 1$ vezes, sendo cada seleção independente das anteriores.



Figura 8. Amostragem aleatória simples.

Amostragem aleatória simples

- ▶ Vantagens
 - ▶ Simplicidade.
 - ▶ Permite medir a precisão das estimativas.
 - ▶ Propriedades conhecidas e testadas.
- ▶ Desvantagens
 - ▶ Precisa de cadastro da população.
 - ▶ Custo elevado (amostra espalhada).
 - ▶ Não usa informações auxiliares.

Aplicações

- ▶ Qualquer situação em que se tenha um cadastro da população.
- ▶ Opinião dos alunos UFPR sobre alguma política estudantil.
- ▶ Clima organizacional de uma empresa.
- ▶ Satisfação dos alunos de Estatística básica com o ensino remoto.
- ▶ Satisfação dos clientes de uma rede bancária.
- ▶ Produtos retirados de um lote para inspeção.
- ▶ Sorteio de pessoas para serem mesários nas eleições.



Figura 9. Retirado do Pexels.

Amostragem sistemática

► Procedimento do método:

1. Selecione uma unidade de partida r ao acaso entre 1 e K , com probabilidade $1/K$.
2. Selecione cada K -ésima unidade do cadastro a partir da primeira selecionada, isto é, $r + K$, $r + 2K$, etc.

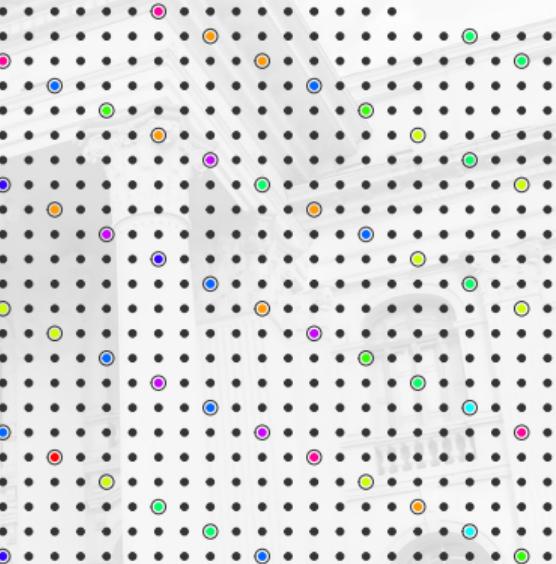


Figura 10. Amostragem sistemática.

Amostragem sistemática

► Vantagens

- ▶ Simplicidade.
- ▶ Cadastro pode ser construído junto com a amostra.
- ▶ Fácil de estimar quantidades populacionais (exceto variância).

► Desvantagens

- ▶ Difícil para estimar a precisão da estimação.
- ▶ Custo elevado (amostra espalhada).
- ▶ Periodicidade no cadastro pode impactar nas estimativas.

Aplicações

- ▶ Avaliação da qualidade de peças em uma linha de produção.
- ▶ Avaliação de lotes de produtos que chegam em sequência.
- ▶ Pesquisas de boca de urna.
- ▶ Pode substituir a amostragem aleatória simples.
- ▶ Plantas em um pomar ou lavoura.
- ▶ Animais que passam por um corredor.
- ▶ Pessoas em um teatro/cinema.



Figura 11. Foto de Alejandro Barrón no Pexels.

Amostragem estratificada

- ▶ Procedimento do método:
 1. Particione a população U em H grupos disjuntos e homogêneos, chamados de **estratos**.
 2. Selecione uma amostra **dentro** de cada um dos estratos, independentemente.
 3. A amostra a ser pesquisada é a **união** das amostras selecionadas nos estratos.
- ▶ Dentro de cada estrato pode ser necessário usar diferentes métodos de coleta e/ou amostragem.

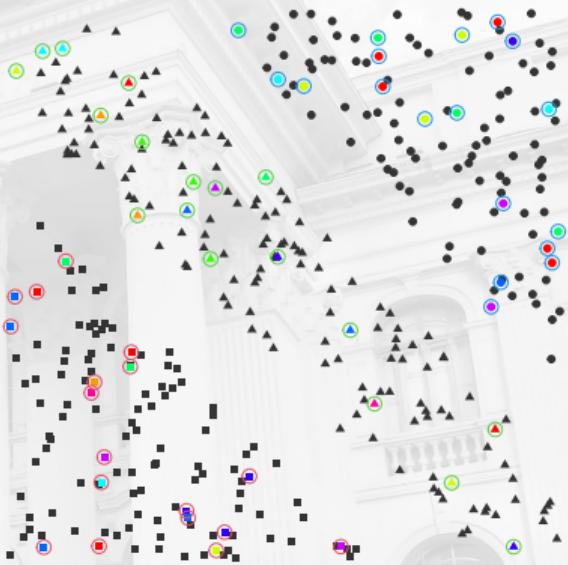


Figura 12. Amostragem estratificada.

Amostragem estratificada

► Vantagens

- ▶ Plano intuitivo.
- ▶ Estratos são divisões naturais da população. Ex. regiões geográficas, gênero, etc.
- ▶ Melhora a eficiência amostral (reduz a variância).
- ▶ Permite medir a precisão das estimativas.

► Desvantagens

- ▶ Precisa de informação prévia sobre os estratos (tamanho).
- ▶ Pode ser necessário usar métodos de coleta diferentes em cada estrato.
- ▶ Pelo menos duas etapas para coleta de dados.

Aplicações

- ▶ Sempre que a população puder ser estratificada.
- ▶ Taxa de desemprego nas capitais e RMC.
- ▶ Pesquisas eleitorais.
- ▶ Clima organizacional em diferentes setores de uma empresa.
- ▶ Pesquisas de opinião em populações fechadas e estratificadas. Ex. UFPR estratos diferentes setores ou cursos.
- ▶ Amostragem de árvores por talhão na floresta.
- ▶ Amostragem de animais por rebanho.



Figura 13. Retirado de iStock.



Figura 14. Retirado de iStock.

Amostragem por conglomerados

► Procedimento do método:

1. Particione a população U em M grupos disjuntos, chamados de **conglomerados**.
2. Selecione uma amostra de m conglomerados.
3. Se amostragem em apenas **1 etapa**, inclua **todas** as unidades da população U encontradas nos m conglomerados na amostra.
4. Se amostragem em **2 etapas**, selecione uma amostra de unidades elementares em cada um dos conglomerados selecionados no item 1.

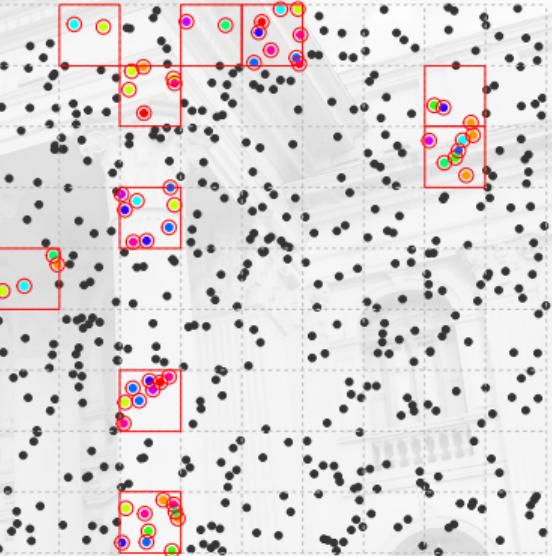


Figura 15. Amostragem por conglomerado.

Amostragem por conglomerados

► Vantagens

- ▶ Pode facilitar a coleta dos dados.
- ▶ Conglomerados são grupos naturais dentro da população (família, departamentos, etc).
- ▶ Permite medir a precisão das estimativas.
- ▶ Mais barato (menos deslocamento).

► Desvantagens

- ▶ Precisa de informação prévia sobre os conglomerados.
- ▶ Pode ser necessário usar métodos de coleta diferentes em cada conglomerado.
- ▶ Pelo menos duas etapas para coleta de dados.
- ▶ Mais complicado para estimar parâmetros de interesse.

Aplicações

- ▶ Sempre que a população estiver conglomerada.
- ▶ Renda das famílias de uma cidade.
- ▶ Pesquisas eleitorais (casa em casa).
- ▶ Qualidade de vida dos trabalhadores da indústria.
- ▶ Pesquisas por telefone (DDD e código de regiões, bairros).
- ▶ Diferença conglomerado e estratificada: amostra concentrada apenas nos conglomerados selecionados na etapa 1.



Figura 16. Foto de veeterzy no Pexels.

Outros procedimentos

- ▶ Existem diversos outros tipos de planos amostrais.
- ▶ Planos amostrais complexos (combinação dos anteriores).
- ▶ O DEST oferece um curso de métodos de amostragem.
- ▶ Para cada plano a distribuição de uma estatística deve ser obtida.
- ▶ Trabalhamos apenas com AASc.



Figura 17. Foto de Juhasz Imre no Pexels.