

MAT02025 - Amostragem 1

Principais fases de um levantamento por amostragem

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2021



Introdução

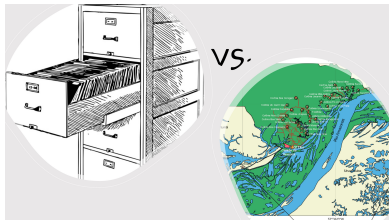
Introdução

- ▶ Como introdução à discussão do papel que a teoria representa em um levantamento por amostragem, é útil descrever, resumidamente, as fases abrangidas pelo planejamento e pela execução de um levantamento¹.

¹Veja uma discussão alternativa em <https://amostragemcomr.github.io/livro/conceitos.html#planejamento-e-execu%C3%A7%C3%A3o-de-pesquisas-por-amostragem>

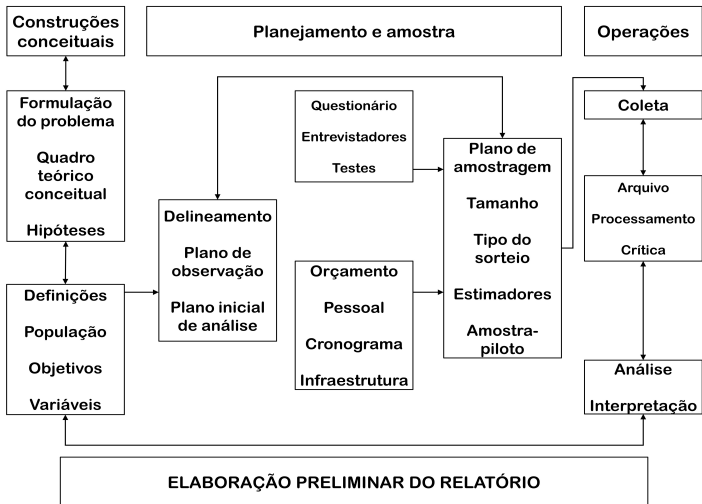
Introdução

A complexidade de levantamentos pode variar bastante.



- ▶ Tirar uma amostra entre 5000 cartões perfeitamente arrumado e numerados em um arquivo é uma tarefa fácil.
- ▶ Outra coisa é coletar amostras de habitantes de uma região onde os transportes se fazem sobre água e através de florestas, no qual falam-se 15 línguas e dialetos diferentes, em que forasteiros “perguntadores” nem sempre são bem-vindos.

Atividades típicas em um levantamento



Construções conceituais

Formulação do problema

- ▶ Construção do **quadro conceitual**.
- ▶ **Hipóteses** e expectativas sobre os resultados.

População de estudo, objetivos e variáveis

- ▶ A **população de estudo** é definida como a parte identificável e acessível do conjunto (**população-alvo**) para o qual idealmente se deseja a informação.
 - ▶ A distância entre ambas é decorrente quer das dificuldades da realidade, quer das limitações impostas pelo delineamento.

População de estudo, objetivos e variáveis

Exemplo (pop.-alvo vs. pop. estudo)

- ▶ Um estudo desejou avaliar o impacto de programas incrementais sobre a prevalência de cárie dental em crianças em idade escolar.
 - ▶ A **população-alvo** era composta de **todas as crianças do município de São José dos Campos (SP) que frequentavam escolas públicas dos ensinos fundamental e médio.**
- ▶ Entretanto, a necessidade de controlar um possível efeito de “confundimento” da idade e do tempo de exposição aos programas, associada à realidade dos recursos disponíveis, indicou ser adequado e oportuno restringir a **população de estudo** aos **escolares de 12 anos que participaram dos programas incrementais por no mínimo dois anos consecutivos.**

População de estudo, objetivos e variáveis

- ▶ Definir os **objetivos** significa explicitar, com precisão, as evidências pretendidas e o uso do conhecimento adquirido sobre a realidade estudada.
- ▶ Estabelecer as **variáveis** é identificar as características observadas e analisadas.
- ▶ Reconhecer a natureza das observações facilita a determinação do tipo de variável (quantitativa ou qualitativa) e as decisões técnicas durante o planejamento.

Comentário

Essas atividades marcam a passagem do enfoque abstrato para a representação concreta e observável do objeto da pesquisa.

Planejamento e amostra

Plano inicial

Monta a estrutura lógica do estudo.

- ▶ O **plano de coleta** determina se o estudo é **descritivo** ou **analítico**, define as **variáveis prioritárias** a serem observadas, os fatores que vão ser controlados e os **grupos populacionais** a serem separados para a comparação (sexo, faixa etária, raça/cor, escolaridade, região).
- ▶ O **plano de análise** é preliminar, não consistindo, porém, na descrição detalhada de métodos e técnicas utilizados na análise estatística.
 - ▶ Decide a forma de organização dos resultados e as medidas empíricas (índices) para o cálculo dos valores populacionais pesquisados e também escolhe as estratégias para verificação de hipóteses ou dos resultados esperados.

Questionário

- ▶ A **elaboração** e o **teste** do **instrumento de coleta** de dados, acompanhado do levantamento de **recursos necessários** à realização do estudo, podem ocorrer durante a construção do plano inicial.
 - ▶ **recursos necessários**: contratação da equipe de campo; supervisores; treinamento (a entrevista será presencial ou por telefone?); passagens e diárias serão necessárias?
- ▶ O **tempo** necessário para coleta e processamento de dados e o **quadro de pessoal** disponível influenciam o tipo de amostragem adotado.

Plano de amostragem

- ▶ Esta etapa compreende a definição do **tamanho** e do **desenho da amostra**, consistindo também na escolha de procedimentos para cálculos das estimativas.
- ▶ O plano inicial e os recursos disponíveis fundamentam as decisões sobre os critérios de custo e precisão dos resultados.

Desenho da amostra

- ▶ O **desenho** (esquema) da amostra consiste em decidir o modelo de **sorteio** ou **tipo de amostra**:
 - ▶ Sistemática;
 - ▶ Casual (aleatória/acidental) simples;
 - ▶ Estratificada;
 - ▶ Por conglomerados.

Desenho da amostra

- ▶ Nessa etapa, a população de estudo é delimitada em termos **espaciais** e **temporais**, o que viabiliza a identificação, mesmo que indireta, do total de elementos que a compõem.
- ▶ A listagem desses elementos forma o **sistema** ou **cadastros**², do qual a amostra é sorteada.
- ▶ O tamanho da amostra resulta da conciliação do atendimento do plano de análise proposto com as limitações dos recursos disponíveis.

²Veja os principais cadastros utilizados em diversas pesquisas:
<https://amostragemcomr.github.io/livro/conceitos.html#cadastros>

Estudo-piloto

- ▶ O **estudo-piloto** impõe-se como teste para a aplicação e o ajuste de todas as decisões tomadas na composição do processo.
 - ▶ Sorteio de amostra proposta, coleta e no processamento preliminar de dados.
- ▶ Deve ser realizado em populações semelhantes à de estudo.
- ▶ Também pode ser usado para determinação do tamanho da amostra (“final”).

Operações

Coleta, consistência e processamento

- ▶ **Coletar**, **conferir** e **processar** os dados são atividades que podem e devem ser realizadas conjunta e progressivamente.
- ▶ A **crítica** e **tratamento** dos dados coletados é uma etapa indispensável para permitir a eliminação de erros grosseiros na massa de dados coletados (idade > 200 anos; indivíduo do sexo masculino com tempo de gestação > 0), os quais podem distorcer significativamente os resultados da pesquisa.
 - ▶ É preciso ter formas de detectar inconsistências e definir o tratamento para a correção dos dados individuais.
- ▶ **Codificação**, **criação de variáveis derivadas** e **cálculo de pesos amostrais** fazem parte da elaboração dos **arquivos de dados** para análises.
 - ▶ **Dicionários de variáveis** também são produzidos nesta etapa.

Análises de dados

- ▶ Esta compreende a **análise estatística** e a **interpretação dos resultados**.
- ▶ O plano preliminar de análise é aprofundado e os métodos são escolhidos para a análise estatística dos dados.
 - ▶ As **estimativas por intervalo** são comuns aos estudos descritivos, enquanto **testes de significância** são aplicados para verificar diferenças entre grupos.
- ▶ A **interpretação dos resultados** obtidos deve ser realizada à luz do contexto construído na primeira fase do processo.
 - ▶ Alcançar respostas para o problema tal como o formulado e atender aos objetivos devem ser os critérios e nortear a última atividade do processo da pesquisa.

Elaboração do relatório

- ▶ Esta não deve ser considerada uma atividade final, mas precisa ser estabelecida como uma dinâmica de registro que acompanha todas as fases do trabalho.
 - ▶ Relatório científico como **repositório do conhecimento** produzido no decorrer processo da pesquisa.

Etapas e considerações complementares

Etapas complementares

- ▶ Divulgação dos resultados:
 - ▶ Relatório(s);
 - ▶ Artigos em periódicos científicos;
 - ▶ Imprensa e público especializado.
- ▶ Gerenciamento de repositório (público?) dos arquivos de dados (e documentação necessária para a análise).

Fontes de erros

Erros amostrais

- ▶ Boas práticas de levantamentos incluem o cálculo de erros amostrais (variação), que é possível se métodos probabilísticos são usados na seleção da amostra.
 - ▶ Intervalos de confiança, erros padrões, entre outros, devem ser divulgados em conjunto com as estimativas pontuais.

Fontes de erros

Erros não-amostrais

- ▶ Incluem o efeito de recusas e ausência no momento da visita do entrevistador, respondentes fornecendo informação incorreta, codificação ou outros erros de processamento.
- ▶ Não é possível estimar o tamanho dos erros não-amostrais.
- ▶ No entanto, podemos aprender com diferentes levantamentos amostrais (prévios) a respeito dos tipos de erros não-amostrais que costumam surgir, e tentar minimizá-los.

Fontes de erros

Os principais erros que podem causar viéses nos resultados são:

1. **Operações de amostragem:** seleção da amostra, definição inadequada da população (parte da população é omitida).
2. **Não-entrevistados:** diferenças entre os não-entrevistados (dados ausentes) e os entrevistados.
3. **Adequação do respondente:** em alguns casos os respondentes não podem ser entrevistados e a informação é obtida por outros, porém o respondente “substituto” nem sempre é tão conhecedor dos fatos.
4. **Entendimento dos conceitos e falta de conhecimento:** alguns respondentes podem não entender o que é desejado, ou não saber a informação requisitada.
5. **Questões mal formuladas:** acabam induzindo o respondente a uma certa resposta.
6. **Erros de processamento:** erros de codificação, chaves-de-identificação erradas, erros de programação, etc.
7. **Erros de entrevistadores:** entrevistadores podem errar na leitura da questão ou embaralhar as respostas e assim introduzir viés.

Para casa

- ▶ Lista de exercícios 1 (no Moodle).

Próxima aula

- ▶ Amostragem probabilística.

Por hoje é só!

Bons estudos!

