

1. Introdução

O que é amostragem ?

Amostragem é a área da estatística que estuda técnicas e procedimentos para retirar e analisar uma amostra com o objetivo de fazer inferência a respeito da população de onde essa foi retirada.

Qual a importância da área ?

- Amostragem enumerativa
- Amostragem analítica

Exemplos de áreas de aplicação em populações humanas:

Marketing - Pesquisas de mercado
 - Pesquisas de satisfação do cliente

Sociologia - Pesquisas de opinião

Ciência Política - Pesquisas eleitorais

Economia - Levantamentos econômicos

Psicologia - Pesquisas de comportamento
 - Pesquisas de qualidade de vida

Exemplos de áreas de aplicação em populações não-humanas:

Agronomia - Levantamentos de produção agrícola

Ecologia - Levantamentos de impacto ambiental
- Levantamentos de recursos naturais

Engenharia - Implementação de planos para monitoramento de qualidade
- “acceptance sampling”

Biologia - Levantamentos de abundância de espécies

Importância social e econômica

Em vários países existe pelo menos um órgão do governo encarregado de produzir informações “oficiais” sobre a nação. Esses órgãos fazem uso intensivo de técnicas estatísticas de amostragem.

Brasil: IBGE
www.ibge.gov.br

Canadá: Statistics Canada
www.statcan.ca

EUA: US Census Bureau
www.census.gov

Bureau of Labor Statistics
www.bls.gov e outros...

Colômbia: DANE
www.dane.gov.co

Chile: INE
www.ine.gov.cl

Índia: MOSPI
www.mospi.nic.in

Gastos estimados com levantamentos amostrais nos EUA: 5 bilhões de dólares por ano.

Importância acadêmica...
gerar técnicas que atendam às
necessidades da sociedade.

Importância prática...
mercado de trabalho.

2. Levantamentos amostrais

Qualquer procedimento que faça uso de uma amostra para obter informação sobre uma população, de onde essa foi retirada, poderia ser chamado de um levantamento amostral.

No entanto, nem todo procedimento usado para a composição de uma amostra é suficiente para justificar inferência estatística.

É preciso que o procedimento tenha como base um método probabilístico de amostragem.

Vantagens de um levantamento amostral (comparando com um censo)

- *Baixo custo*
- *Rapidez*
- *Qualidade da informação*

Em algumas situações, utilizar uma amostra é a única alternativa:

- *Quando a obtenção da informação desejada implica na destruição da unidade pesquisada*
- *Quando o custo de realização de um censo é proibitivo*

Talvez a justificativa menos natural dentre as descritas acima seja a de que os resultados obtidos via amostragem

podem ser de melhor qualidade do que os obtidos via pesquisa censitária. Como isso é possível ? Para ilustrar a idéia, o seguinte experimento é proposto:

Leia o texto seguinte, de autoria de Chico Science, apenas uma vez. Conte e escreva quantas vezes aparece a letra “N”: ____.

EU VIM COM A NAÇÃO ZUMBI / AO SEU
OUVIDO FALAR / QUERO VER A
POEIRA SUBIR / E MUITA FUMAÇA NO AR /
CHEGUEI COM O MEU UNIVERSO / E
ATERRISO NO SEU PENSAMENTO / TRAGO AS
LUZES DOS POSTES NOS OLHOS /
RIOS E PONTES NO CORAÇÃO / PERNAMBUCO
EM BAIXO DOS PÉS / E MINHA MENTE NA
IMENSIDÃO.

Conceitos básicos iniciais

População-alvo

É o conjunto de todos os elementos sobre os quais alguma informação é procurada.

Cadastro

É o meio pelo qual os elementos de uma população alvo são identificados.

População pesquisada

É o conjunto de todos os elementos da população alvo que podem ser selecionados para participar da amostra.

Amostra

É um subconjunto da população pesquisada, onde a informação desejada é observada.

***Variável
de
interesse***

Uma característica de interesse, relativa a cada elemento da população pesquisada, mas que é observada apenas na amostra.

***Unidade
amostral
(u.a.)***

É a unidade que é de fato selecionada para compor a amostra.

Dentre os conceitos básicos citados acima, o de cadastro é fundamental para a implementação de um plano amostral probabilístico.

Cadastros

No contexto de amostragem, *cadastro* é a designação dada a qualquer material ou instrumento usado para identificar e obter acesso aos elementos que compõem a população-alvo.

O processo físico de aleatorização relativo a cada plano amostral é aplicado às unidades que compõem o cadastro (unidades amostrais).

Uma unidade amostral pode ser um elemento ou um conjunto (conglomerado) de elementos da população. Nesse sentido, é possível identificar dois tipos de cadastros:

Tipos de cadastro:

- a) Cadastros compostos por uma listagem de elementos da população-alvo;
- b) Cadastros compostos por uma listagem de conjuntos de elementos da população-alvo.

A disponibilidade de um cadastro do tipo a) permite a utilização de planos amostrais com seleção direta de elementos.

É comum, porém, encontrar situações em que apenas um cadastro do tipo b) está disponível. Nesse caso, para observar um elemento, é preciso fazer uso de planos amostrais em estágios. É possível ainda, selecionar uma unidade do cadastro e observar todos os elementos que a compõem.

Exemplo:

*Levantamento Amostral do Sistema de
Produção de Gado de Leite do Agreste
Meridional de Pernambuco*

População-alvo:

Cadastro:

Amostra:

População pesquisada:

Variáveis de interesse:

Elemento da população:

Unidade amostral:

Etapas gerais de um levantamento amostral

Um levantamento amostral pode ser dividido em várias etapas.

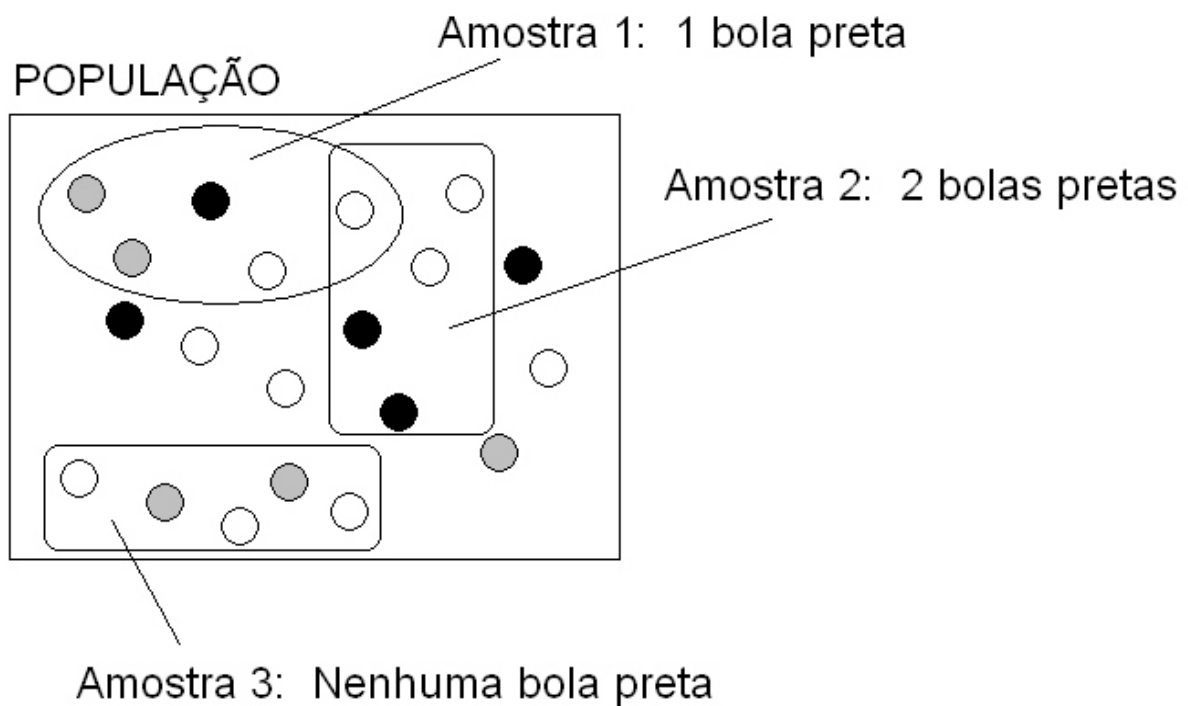
Cochran (1977) às descreve de acordo com os seguintes tópicos gerais:

- i. Identificação dos objetivos do levantamento;
- ii. Definição da população-alvo;
- iii. Definição das variáveis de interesse e dos dados a serem colhidos;
- iv. Identificação do grau de precisão desejado;

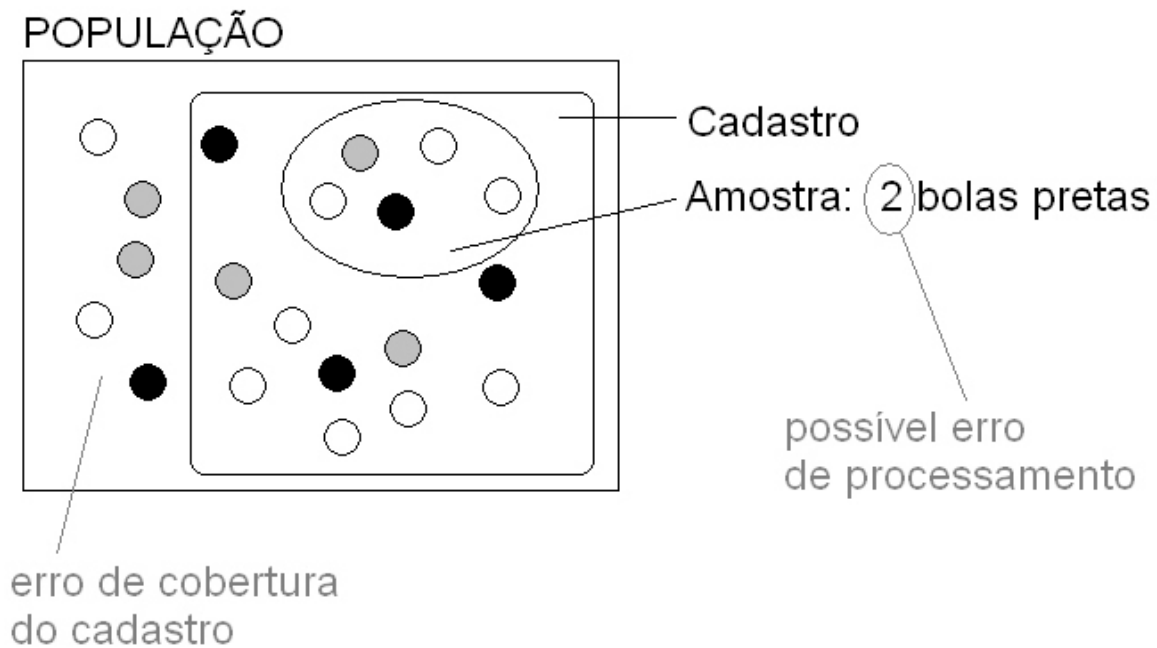
- v. Escolha do instrumento de coleta de dados;
- vi. Identificação de um cadastro;
- vii. Planejamento da amostra;
- viii. Pré-teste;
- iv. Seleção da amostra e coleta de dados / Organização de trabalho de campo;
- x. Descrição e análise dos dados;
- xi. Resumo da informação adquirida e recomendações para futuros levantamentos.

Tipos de erros em levantamentos amostrais

Erro amostral \Rightarrow É o erro associado ao fato de que a variável de interesse é observada apenas na amostra.



Erro não-amostrai ⇒ É o erro associado ao processo operacional envolvido em um levantamento amostral.



$$\text{ERRO TOTAL} = \text{Erro amostral} + \text{Erro não-amostrai}$$

Biemer e Lyberg (2003) subdividem os erros não-amostrais nas seguintes categorias:

a) Erros de especificação;

b) Erros de cobertura do cadastro;

c) Erros de processamento;

d) Erros de mensuração;

e) Erros de não-observação.

Etapas operacionais de um levantamento amostral e tipos de erro associados

Do ponto de vista operacional, Särndal, Swensson e Wretman (1992) identificam as seguintes etapas de um levantamento amostral (os tipos de erro associados a cada etapa estão entre parêntesis):

- i. Seleção da amostra;
(erro amostral e de cobertura do cadastro)
- ii. Coleta de dados;
(erro de mensuração e de não-observação)

- iii. Processamento de dados;
(erro de processamento)
- iv. Estimaco e anlise;
(todos os erros citados nos itens anteriores)
- v. Disseminao dos resultados e avaliao ps-levantamento.

3. O método probabilístico de amostragem

Um método de amostragem é considerado probabilístico se as seguintes condições gerais são satisfeitas:

1. *É possível definir o conjunto de todas as amostras possíveis de ser retiradas (\mathfrak{S}) de acordo com o procedimento de seleção da amostra empregado:*

$$\mathfrak{S} = \{s_1, s_2, \dots, s_M\} .$$

2. *A amostra é selecionada por um processo físico de aleatorização que associa a cada amostra possível s uma probabilidade exata de seleção dada por $p(s)$.*

3. *O procedimento de seleção da amostra dá uma probabilidade de seleção positiva para cada elemento da população.*

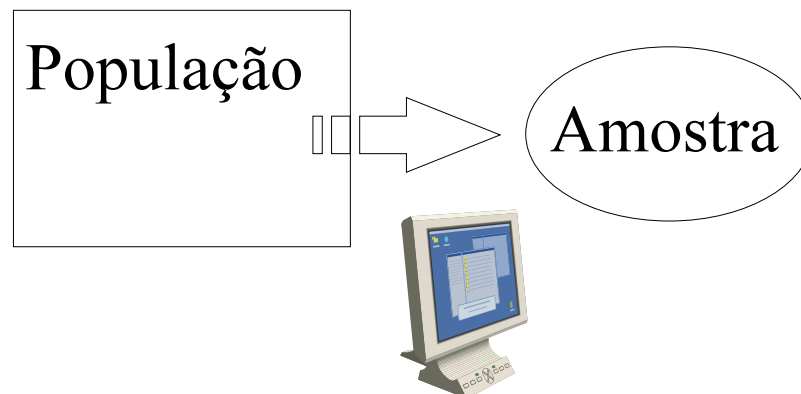
Esquemas e planos amostrais

Plano amostral:

É a estratégia de associação de probabilidades às amostras possíveis de serem retiradas.

Esquema amostral:

É o conjunto de procedimentos físicos/computacionais realizados para compor a amostra de acordo com o plano amostral.



Esquema amostral:
Como retirar a amostra ?

Exemplo: Selecionar uma amostra de tamanho n de uma população de tamanho N com cadastro tipo a) disponível.

Plano amostral:

Amostra aleatória simples sem reposição.

Escolher esse plano significa adotar a estratégia de associar a cada amostra s , probabilidade de seleção

$$p(s) = \left(\frac{N}{n} \right)^{-1}.$$

Um esquema amostral:

Passo 1: Selecione, com probabilidade igual para todos os elementos, $1/N$, o primeiro elemento da amostra, sem reposição;

Passo 2: Selecione, com probabilidade igual para todos os elementos restantes, $1/(N-1)$, o segundo elemento da amostra, sem reposição;

⋮

Passo n : Selecione, com probabilidade igual para todos os elementos restantes, $1/(N-n+1)$, o n -ésimo elemento da amostra.

Parâmetros e estimadores

Denote por U , o conjunto dos índices que identificam os elementos que compõem a população finita, de tamanho N , sob estudo:

População: $U = \{1, 2, \dots, N\}$

Denote por y uma variável de interesse e, por y_k , o valor dessa variável associado ao elemento k da população.

Parâmetro: Característica numérica da população.

$$\theta = \theta(\{y_k : k \in U\})$$

Exemplos de parâmetros de interesse:

- Total populacional ($\theta = t$)

$$t = \sum_{k \in U} y_k = \sum_U y_k$$

- Média populacional ($\theta = \bar{y}_U$)

$$\bar{y}_U = N^{-1} \sum_{k \in U} y_k = N^{-1} \sum_U y_k = N^{-1} t$$

- Mediana populacional ($\theta = M$)

$$M = F^{-1}(0.5)$$

$$F(y) = \#A_y / N$$

$\#A_y$ = Número de elementos de A_y

$$A_y = \{k : k \in U, y_k \leq y\}$$

Amostra: Variável aleatória: S
Realização: s

$$Prob(S=s) = p(s)$$

Estatística: Característica numérica
da amostra.

$$Q(S) = Q[(k, \mathbf{y}_k): k \in S]$$

Estimador: É uma estatística em que,
para grande parte das
amostras possíveis,
assume valores próximos
ao do parâmetro de
interesse.

$$\hat{\theta} = \hat{\theta}(S)$$

Propriedades estatísticas básicas de um estimador

Esperança:

$$E_p(\hat{\theta}) = \sum_{s \in \mathfrak{S}} p(s) \hat{\theta}(s)$$

Variância:

$$V_p(\hat{\theta}) = \sum_{s \in \mathfrak{S}} p(s) \{ \hat{\theta}(s) - E_p[\hat{\theta}(s)] \}^2$$

Erro-padrão:

$$EP_p(\hat{\theta}) = V_p(\hat{\theta})^{1/2}$$

Erro-padrão relativo:

$$CV_p(\hat{\theta}) = EP_p(\hat{\theta})/E_p(\hat{\theta})$$

Tendência:

$$Tend_p(\hat{\theta}) = E_p(\hat{\theta} - \theta)$$

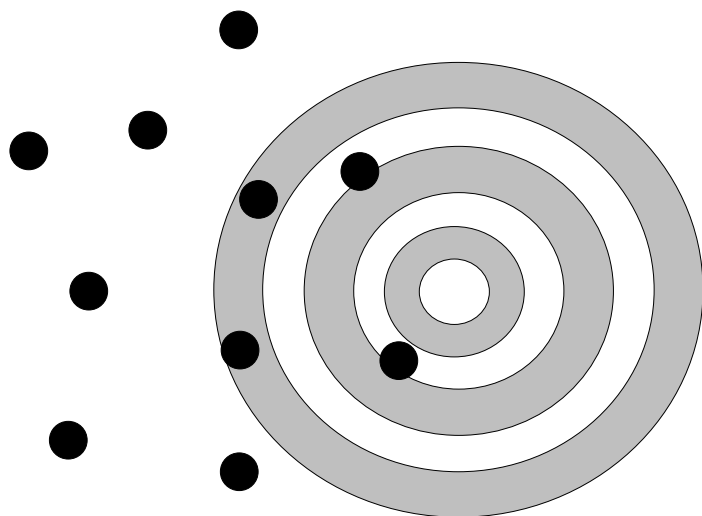
Erro Quadrático Médio:

$$EQM_p(\hat{\theta}) = \sum_{s \in \mathfrak{S}} p(s) [\hat{\theta}(s) - \theta]^2$$

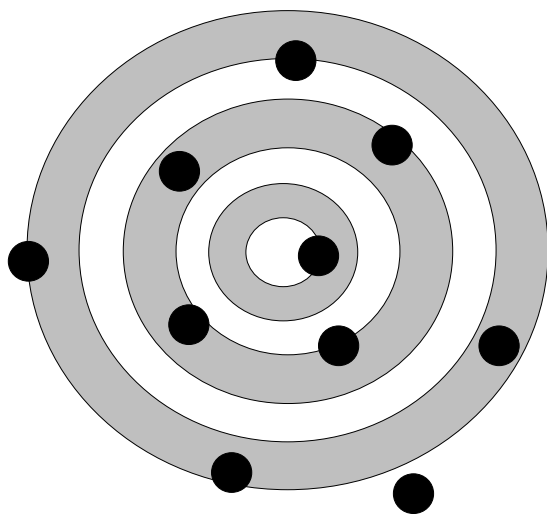
O erro quadrático médio é a soma da variância com o quadrado da tendência do estimador:

$$\begin{aligned}EQM_p(\hat{\theta}) &= E_p(\hat{\theta} - \theta)^2 \\&= E_p[\hat{\theta} - E_p(\hat{\theta}) + E_p(\hat{\theta}) - \theta]^2 \\&= E_p\{[\hat{\theta} - E_p(\hat{\theta})]^2 + [E_p(\hat{\theta}) - \theta]^2 \\&\quad + 2[\hat{\theta} - E_p(\hat{\theta})][E_p(\hat{\theta}) - \theta]\} \\&= E_p[\hat{\theta} - E_p(\hat{\theta})]^2 + [E_p(\hat{\theta}) - \theta]^2 \\&= Var_p(\hat{\theta}) + Tend_p^2.\end{aligned}$$

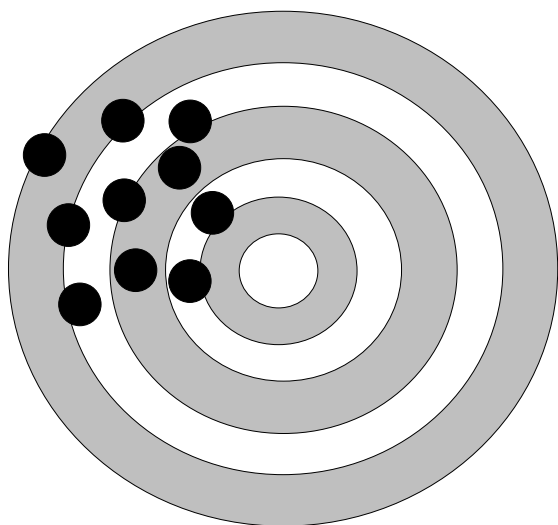
Analogia ao tiro-ao-alvo



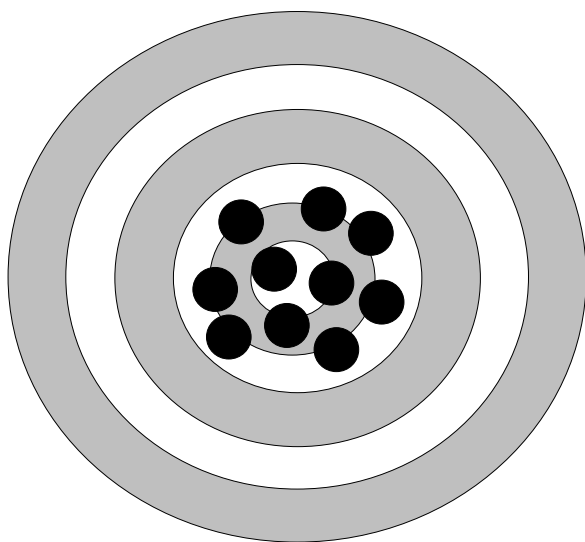
(i)



(ii)



(iii)



(iv)

Daqui por diante, a menos que seja dito o contrário, serão apresentados os passos para o desenvolvimento de uma teoria unificada de amostragem, visando contruir estimadores CENTRADOS.