

1. Introdução

Nos dias de hoje, a demanda por informações que ajudem em processos de tomadas de decisão é considerável. Frequentemente essas informações são de caráter quantitativo, como índice de inflação, taxa de desemprego, percentual de clientes satisfeitos com um produto ou serviço, percentual de intenção de votos em favor de certo candidato, etc... Em todos esses casos, ter acesso à informação desejada seria impraticável sem recorrer à análise de uma amostra. Por outro lado, não é qualquer amostra que fornecerá a informação procurada.

☞ O que é Amostragem ?

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

☞ Qual o objetivo de um curso em Amostragem ?

Em linhas gerais, um curso em amostragem estatística tem dois objetivos: capacitar o aluno para obter amostras que permitam acesso à informação desejada (planejamento amostral) e fornecer o embasamento necessário para, de posse de tal

amostra, conseguir tal informação (estimação).

☞ Qual a importância de um curso em Amostragem ?

Cursos em métodos estatísticos assumem, em geral, a disponibilidade de um conjunto de dados que satisfaz critérios estatísticos para análise. Como conseguir tal conjunto de dados ? Ou ainda, para um certo conjunto de dados disponível, como avaliar se este satisfaz os critérios necessários para análise ? A resposta para essas questões fundamentais está no escopo de um curso em amostragem estatística.

☞ O curso de Amostragem Estatística da UFPE.

O curso de graduação em Estatística da UFPE oferece capacitação em amostragem através de duas disciplinas: Amostragem 1 e 2. Na disciplina Amostragem 1, os princípios estatísticos de amostragem são ensinados, englobando-se os planos amostrais básicos (amostra aleatória simples, amostra estratificada, amostra em conglomerados, em múltiplos estágios, etc...). A disciplina Amostragem 2 aborda métodos de estimação eficientes através do uso de informações auxiliares. Além disso, tópicos especiais relativos ao planejamento amostral e estimação de parâmetros são discutidos.

2. 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 Amostrada

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

Amostra

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

***Variável
de
Interesse***

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

***Unidade
Observacional
(u.o.)***

É o elemento no qual a medição da(s) variável(eis) de interesse é feita.

***Unidade
Amostral
(u.a.)***

É o elemento que é de fato selecionado para compor a amostra.

Exemplo:

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

População alvo:

Cadastro:

Amostra:

População Amostrada:

Variáveis de Interesse:

Unidade Observacional:

Unidade Amostral:

3. Levantamentos Amostrais

Qualquer procedimento que faça uso de uma amostra *probabilística* para obter informação sobre uma população (de onde essa amostra foi retirada), pode ser chamado de um levantamento amostral. Pesquisas de opinião, pesquisas de mercado, econômicas, pesquisas de avaliação de impacto ambiental, são possíveis exemplos de levantamentos amostrais.

Vantagens de um Levantamento Amostral (comparando com um censo)

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

Etapas de um Levantamento Amostral

Um levantamento amostral consiste de várias etapas. Cochran (1977) descreve essas etapas de acordo com os seguintes tópicos gerais:

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

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

Planos Amostrais Não-Probabilísticos

Os planos amostrais não-probabilísticos caracterizam-se por assumirem critérios com bases não aleatórias para a seleção de uma amostra. São exemplos dessa categoria:

Amostragem por Quotas

Aqueles que fazem uso de um plano amostral por quotas justificam suas inferências pelo fato de que a amostra é composta como uma “mini população”. O termo “amostra representativa” é freqüentemente usado.

Amostragem Proposital

Planos Amostrais Probabilísticos: a importância da aleatorização

Uma amostra é dita probabilística ou aleatória se ela for selecionada por um processo físico de aleatorização que permita conhecer qual a probabilidade de seleção de cada elemento da população. Como consequência do processo de aleatorização empregado, uma distribuição de referência é gerada, servindo de base para inferência. O seguinte exemplo ilustra a idéia.

Considere uma micro população composta por cinco elementos, como apresentado no Quadro 1 a seguir.

Quadro 1.

Micro população de 5 elementos

y_1	y_2	y_3	y_4	y_5
2	8	3	1	6

No Quadro 1, y_k representa o valor de uma variável de interesse Y associado ao elemento k da população ($k = 1, \dots, 5$).

O objetivo é estimar a média populacional ($\mu_y = 4$) dessa população com base em uma amostra aleatória simples¹ de tamanho 3. Ao todo,

$\binom{5}{3} = 10$ amostras diferentes são

possíveis:

¹ Nesse curso, todos os planos amostrais considerados são sem reposição, a menos que seja dito o contrário.

Elementos da Amostra (S)	$\sum_{k \in S} y_k$
1 2 3	13
1 2 4	11
1 2 5	16
1 3 4	6
1 3 5	11
1 4 5	
2 3 4	
2 3 5	
2 4 5	
3 4 5	

EXERCÍCIO EM CLASSE:

1. Complete as informações do exemplo anterior.
2. Qual a probabilidade de se obter qualquer uma das amostras possíveis ?
3. Considere o uso da média amostral \bar{y}_s para estimar a média populacional μ_y . Calcule o valor esperado e a variância desse estimador usando as definições fundamentais a seguir:

$$E_p(\bar{y}_S) = \sum_{S \in \mathfrak{S}} p(S) \bar{y}_S$$

$$Var_p(\bar{y}_S) = \sum_{S \in \mathfrak{S}} p(S) [\bar{y}_S - E(\bar{y}_S)]^2$$

4. Com base na questão anterior, defina o que vem a ser um estimador centrado.

Erro Quadrático Médio (EQM)

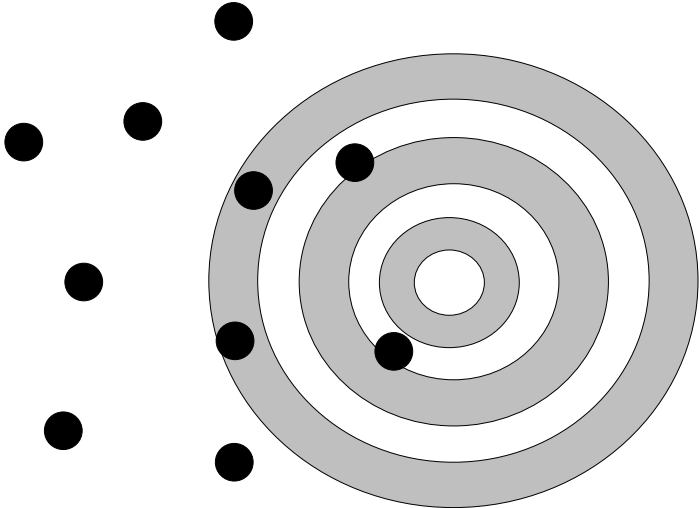
O erro quadrático médio é uma medida de desempenho de um estimador em relação ao objetivo de revelar o valor do parâmetro de interesse.

Considere θ como o parâmetro de interesse e $\hat{\theta}$, o estimador amostral em questão. O erro quadrático médio de $\hat{\theta}$ é definido como segue:

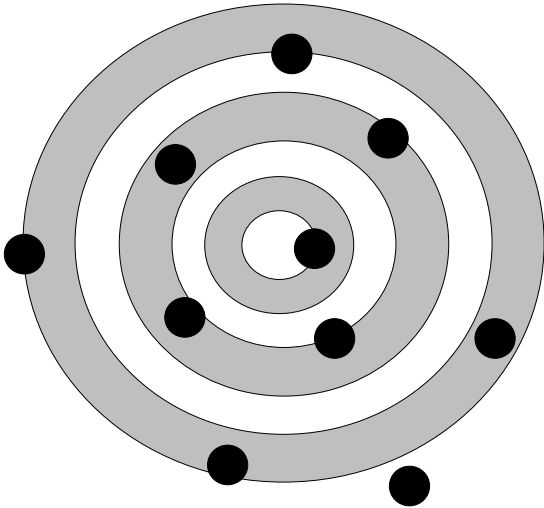
$$\begin{aligned}EQM(\hat{\theta}) &= E_p(\hat{\theta} - \theta)^2 \\&= E_p[\hat{\theta} - E_p(\hat{\theta}) + E_p(\hat{\theta}) - \theta]^2 \\&= E_p[\hat{\theta} - E_p(\hat{\theta})]^2 + [E_p(\hat{\theta}) - \theta]^2 \\&= \text{Var}_p(\hat{\theta}) + \text{Tendência}_p^2.\end{aligned}$$

Se $\hat{\theta}$ for um estimador centrado, sua tendência é nula e o erro quadrático médio reduz-se à variância de $\hat{\theta}$.

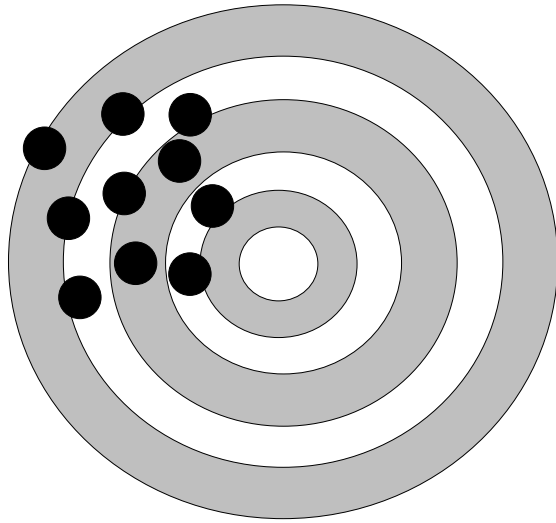
Analogia ao Tiro-ao-alvo



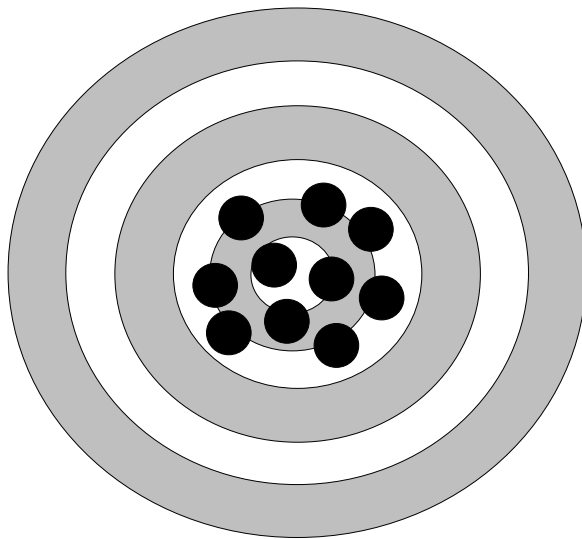
(i)



(ii)



(iii)



(iv)

Tipos de Erros em Levantamentos Amostrais

$$\mathbf{ERRO\ TOTAL} = \mathbf{Erro\ Amostral} + \mathbf{Erro\ Não-Amostral}$$

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

Erro não-amostral \Rightarrow É o erro associado ao processo operacional envolvido em um levantamento amostral. As fontes de tendências discutidas anteriormente geram erros não-amostrais.