



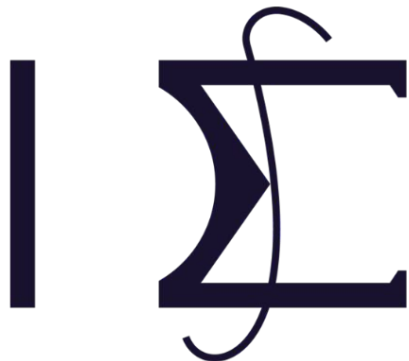


ANA CAROLINA

Cientista de Dados - TakeBlip

Formulação e Levantamento de Hipóteses

T Referências



DESCOMPLICA ESTATÍSTICA

[@descomplicaestatistica](https://www.instagram.com/descomplicaestatistica)



descomplicaestatistica

205 publicações

2.992 seguidores

188 seguindo

Ana Carolina | Estatística

📊 Estatística e Ciência de Dados de forma simples e descomplicada.

😂 Se você gosta de memes como eu aqui é seu lugar hehe.

🔻 Links úteis

bit.ly/DescomplicaEstatistica



Comece aq...



Storytelling



Lives 📺



Materiais E...



Mapas me...

PUBLICAÇÕES

GUIAS

REELS

VÍDEOS

SALVOS

MARCADOS



AGENDA

- **Bloco 1: Pensamento Científico e Incerteza**
- **Bloco 2: Introdução ao teste de Hipótese**
- **Bloco 3: Construção de Hipóteses**
+ Intervalo - 10 min
- **Bloco 4: Tipos de Amostragem**
- **Bloco 5: Cálculo de Amostragem**



**O que é teste de Hipótese
para você?**

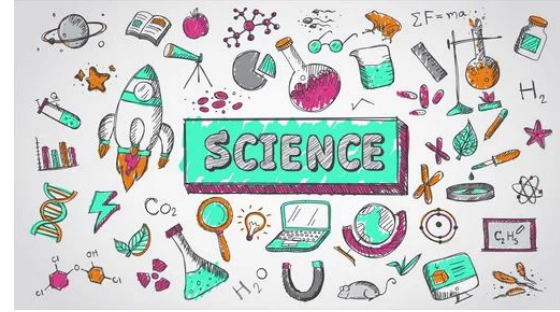


Pensamento Científico e Incerteza

T Por que estamos aqui?



Perguntas, dúvidas e incertezas.



- Ciência representa todo o conhecimento adquirido através do estudo, pesquisa ou da prática, baseado em princípios certos. Esta palavra deriva do latim scientia, cujo significado é "**conhecimento**" ou "**saber**".
- **Ciência é o conhecimento que explica os fenômenos** obedecendo a leis que foram verificadas por métodos experimentais.
- A ciência baseia-se na regularidade, na previsão e no controle de fenômenos que podem ser observados.

T Ciência

O objetivo da ciência é explicar, descrever e prever os fenômenos a partir do **desenvolvimento de procedimentos metodológicos** que possam ser constantemente verificados e reproduzidos.

A metodologia é essencial na ciência, assim como a ausência de preconceitos e juízos de valor.

Características da ciência:

- **Objetiva:** a ciência é um conhecimento que pressupõe uma objetividade, ou seja, busca ser imparcial, com a mínima influência dos interesses pessoais, com linguagem clara, rigorosa e precisa para evitar ambiguidades.
- **Verificável:** todas as teorias científicas são postas à prova. Teorias que não resistem à verificação são descartadas. A verificação deve ser realizada pelo próprio cientista ou por qualquer pessoa e também passa pelo julgamento da comunidade científica.
- **Controlada:** todos os elementos das ciências devem ser controlados para possibilitar a sua verificação e reprodução.
- **Lógica:** a ciência é um conhecimento baseado na lógica, não aceita contradições.

ESTATÍSTICA

É a **ciência** que corresponde a coletar, analisar, apresentar e interpretar dados, bem como tomar decisões com base nessas análises”.

- **Prem S. Mann**

T Método científico

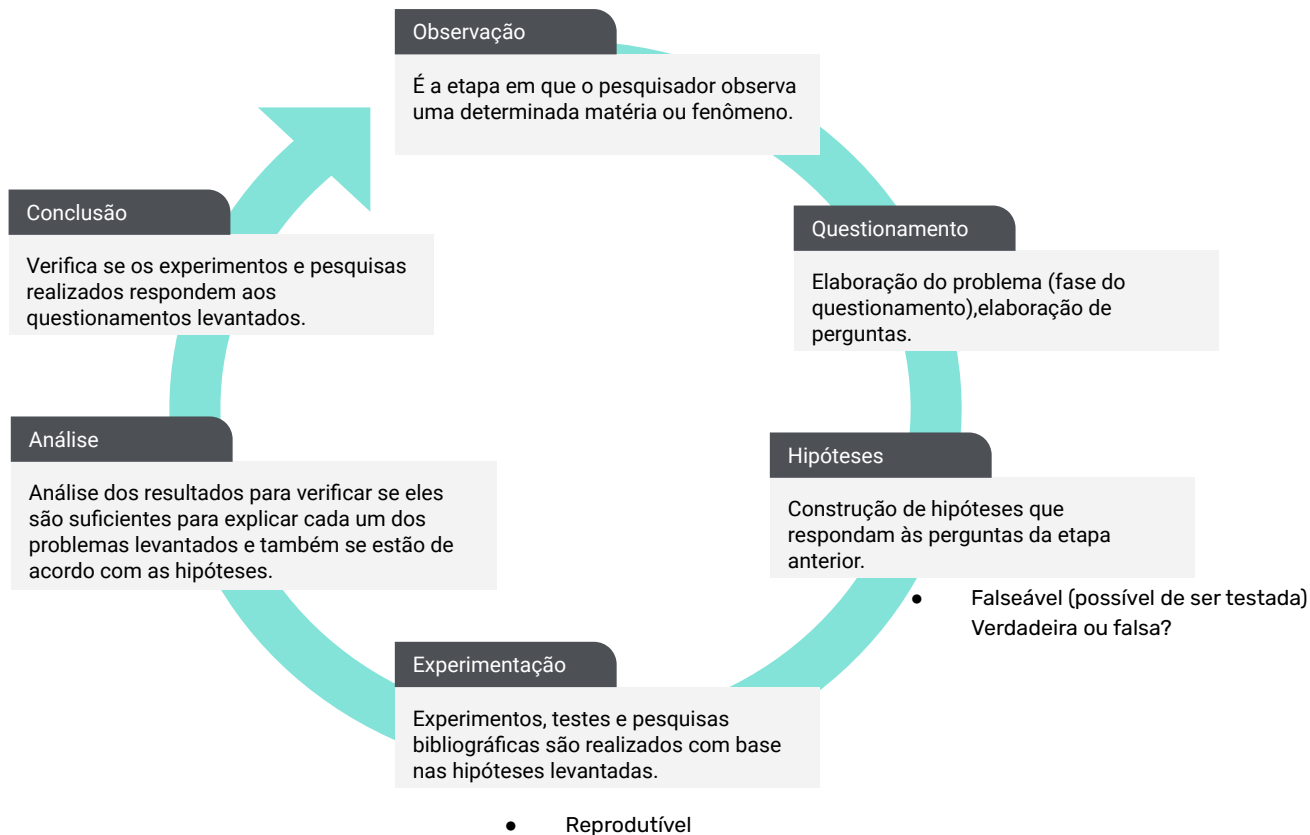
O método científico funciona como um manual que guia a pesquisa científica.

Ele determina um conjunto de passos ou um caminho que deverão ser seguidos até **que os objetivos da pesquisa sejam atingidos.**

Essas etapas devem ser organizadas para que, sempre que necessário, o pesquisador consiga repetir o processo, passando pelas mesmas etapas e chegando a resultados esperados.

- **Objetiva**
- **Verificável (reprodutível)**
- **Controlada**
- **Lógica**

T Método científico





INCERTEZAS

O motor da ciência não é o conhecimento, e sim a falta dele. - Stuart Firestein

- Será que vai chover amanhã?

Uma das maneiras de lidar com a incerteza acerca de um evento é expressá-la em números. Ou seja, quantificando a incerteza.

Probabilidade é uma medida da incerteza acerca de um evento.

- Esta medida é um número que vai de 0 a 1. Quanto maior a probabilidade de evento, menor a incerteza acerca dele.

T

TESTES DE HIPÓTESE

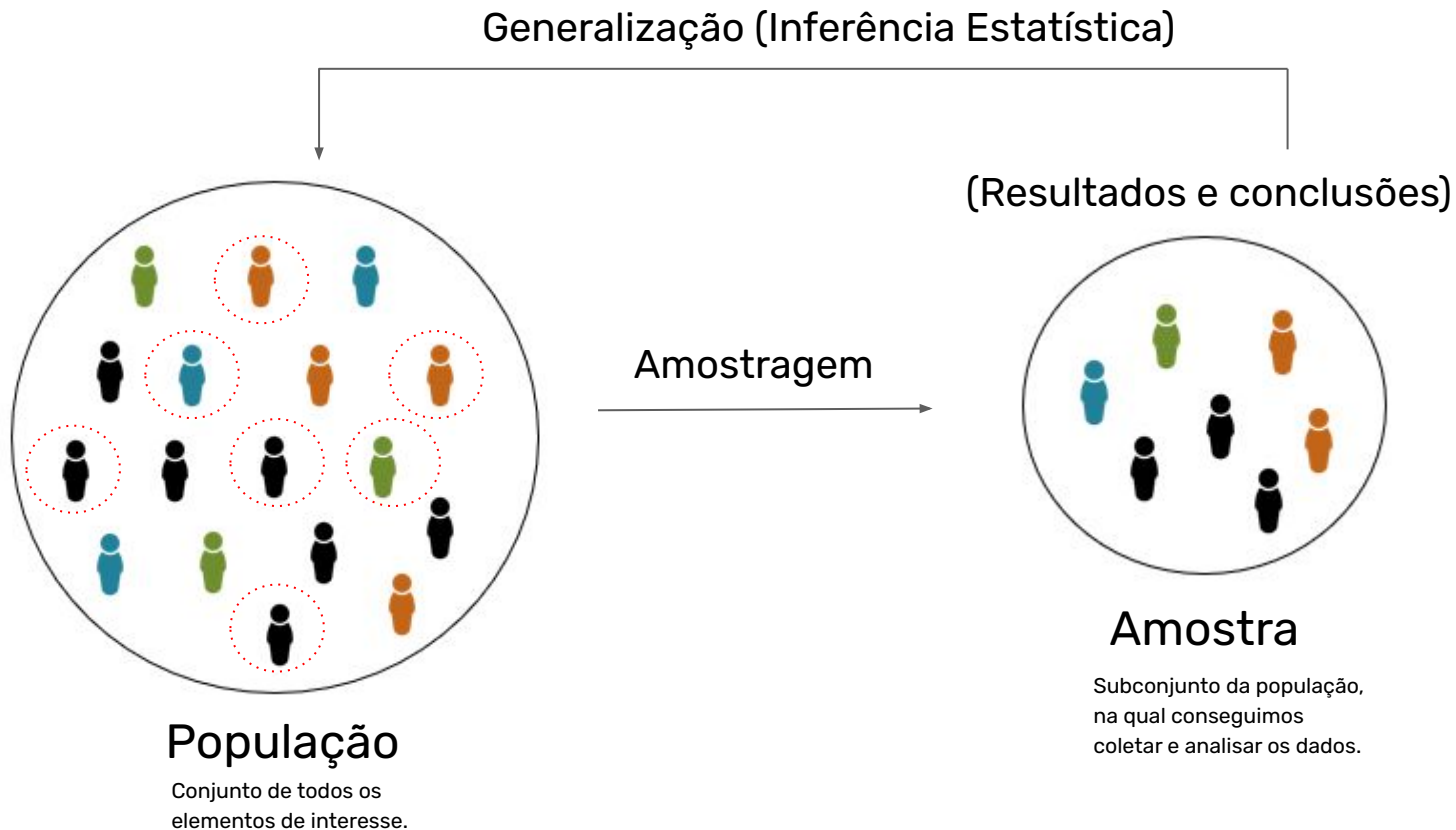
T Hipótese

Hipótese é uma suposição a respeito de algo. É parte integrante do método científico que forma a base de experimentos científicos.

Uma hipótese testável é uma hipótese que pode ser provada ou refutada como resultado da experimentação.



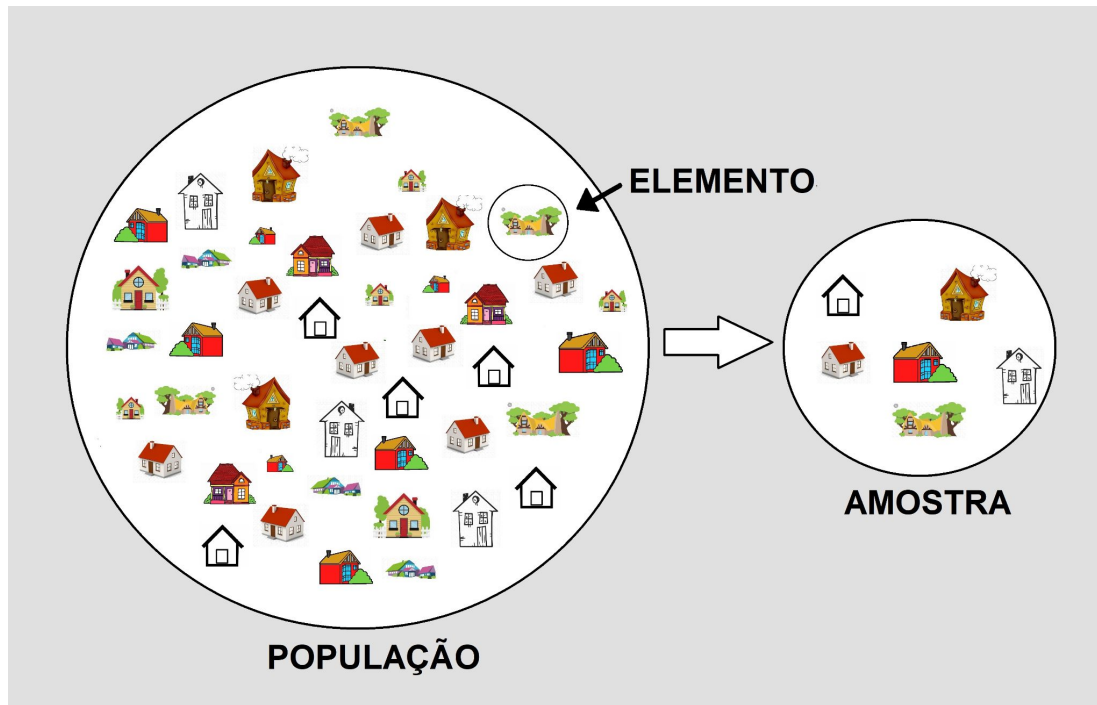
T População e amostra





População

O conceito de população não se restringe à pessoas.





Exemplo:

Uma pesquisa foi realizada com 1000 adolescentes brasileiros sobre o hábito de fumar:

- 280 responderam que “fumam”.
- 720 responderam que “não fumam”.

- Quem é a população e quem é amostra?

População: consiste em todos os adolescentes brasileiros.

Amostra: consiste das 1000 respostas coletadas na pesquisa.

INFERÊNCIA ESTATÍSTICA

O objetivo é produzir afirmações sobre dada característica de uma população, na qual estamos interessados, a partir de informações colhidas de uma parte dessa população (amostra).

- Estimar o parâmetro populacional - Intervalo de Confiança
- Testar hipóteses sobre o parâmetro. - Teste de Hipótese

T Estatística

Uma estatística é uma **característica da amostra**, ou seja, uma estatística é uma função (um valor) calculado através dos valores obtidos da amostra.

As estatísticas mais comuns são:

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n} \longrightarrow \text{Média amostral}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n - 1} \longrightarrow \text{Variância amostral}$$

$$X_{(1)} = \min\{X_1, X_2, X_3, \dots, X_n\} \longrightarrow \text{Menor valor da amostra}$$

$$X_{(n)} = \max\{X_1, X_2, X_3, \dots, X_n\} \longrightarrow \text{Maior valor da amostra}$$

$$W = X_{(n)} - X_{(1)} \longrightarrow \text{Amplitude amostral}$$

T Variância

A variância é uma medida de dispersão que mostra o quão distante cada valor observado está do valor central (média).

Quanto **menor** é a variância, mais próximos os valores estão da média; mas quanto maior ela é, mais distantes os valores estão da média.

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Variância **populacional**

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n - 1)}$$

Variância **amostral**

T Parâmetro

Um parâmetro é uma medida usada para descrever **uma característica da população**.

Ex.: média, variância, proporção etc., Muitas vezes é representado por letras gregas

Denominação	Parâmetro
Média	μ
Variância	σ^2
Desvio padrão	σ
Proporção	p
Tamanho	N

T Estimador

Função/combinção dos elementos da amostra usado para estimar o parâmetro populacional desconhecido.

Denominação	Parâmetro (População)	Estatística (Amostra)
Média	μ	\bar{x}
Variância	σ^2	s^2
Desvio padrão	σ	s
Proporção	p	\hat{p}
Tamanho	N	n

Ex.:

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Estimativa: Valor numérico do estimador quando aplicado aos dados observados na amostra.

Ex: $\bar{x} = 34.5$



Exemplo:

Defina se temos estamos falando de um parâmetro populacional ou de uma estatística:

- Proporção de adolescentes brasileiros que fumam

Parâmetro.

→ **valor desconhecido, (não há pesquisa com todos os adolescentes do Brasil).**

- Proporção de adolescentes na amostra que responderam “fumo”:
 $280/1000 = 0,28$ (ou 28%)

Estatística - Valor calculado através da amostra

O campeonato nacional de futebol do ano passado teve um número médio de gols por partida igual a 2.

→ **Parâmetro, pois foram contabilizados os gols de todas as partidas.**

T Teste de Hipótese

Teste de Hipóteses é o processo de decisão entre duas hipóteses sobre um parâmetro da população.

$$\begin{cases} H_0 : \text{Hipótese nula} \\ H_1 : \text{Hipótese alternativa} \end{cases}$$

Rejeitamos a Hipótese Nula.

Não rejeitamos a Hipótese Nula.

- **Hipótese nula:** é a afirmação que estamos tentando rejeitar.
- **Hipótese alternativa:** é a afirmação verdadeira para o caso de a hipótese nula ser falsa.

A hipótese nula é o estado atual das coisas (ponto de partida), enquanto a alternativa é a mudança ou inovação (hipótese do pesquisador).

As hipóteses são mutuamente exclusivas (complementares) elas somam 100% das opções possíveis.

Hipótese é uma afirmação sobre um parâmetro da população.

- Qual **proporção** de adolescentes do Brasil que fumam?
- Altura **média** da população Brasileira é 1,70?
- Qual o tamanho **médio** das baleias azuis?

Denominação	Parâmetro
Média	μ
Variância	σ^2
Desvio padrão	σ
Proporção	p
Tamanho	N

$$H_0 : p \leq 68\%$$

$$H_1 : p > 68\%$$

$$H_0 : p \geq 68\%$$

$$H_1 : p < 68\%$$

$$H_0 : p = 68\%$$

$$H_1 : p \neq 68\%$$

p_0 é um valor que definimos.

- Qual **proporção** de adolescentes do Brasil que fumam?
- Estado atual (A **proporção** de adolescentes do Brasil que fumam é de 68%)

Hipóteses:

- A **proporção** de adolescentes do Brasil que fumam é maior do que 68%?
- A **proporção** de adolescentes do Brasil que fumam é menor do que 68%?
- A **proporção** de adolescentes do Brasil que fumam é diferente de 68%?

T Exemplo:

Uma máquina automática é programada para encher pacotes de café.

Ela estava regulada para encher os pacotes com 500 g, em média. **Agora, você precisa testar para ver se ela desregulou. Quais seriam as nossas hipóteses nula e alternativa?**

E as hipóteses que nos interessam são:

$$\begin{cases} H_0 : \text{A máquina está regulada} & H_0: \mu = 500 \text{ g,} \\ H_1 : \text{A máquina está desregulada} & H_1: \mu \neq 500 \text{ g,} \end{cases}$$

T Exemplo:


Um fabricante afirma que seus cigarros contêm não mais que 30 mg de nicotina em média. Qual a nossa hipótese para verificar se os dados refutam ou não a afirmação do fabricante?

E as hipóteses que nos interessam são:

$$\begin{cases} H_0 : \text{Os cigarros contem 30 mg de nicotina ou menos} \\ H_1 : \text{Os cigarros contem mais 30 mg de nicotina} \end{cases}$$

$$H_0 : \mu \leq 30$$

$$H_1 : \mu > 30$$



**ALGUMA DÚVIDA
ATÉ AQUI?**

T



ELABORANDO HIPÓTESES

DINÂMICA DE HOJE

15 min

- **Escreva uma hipótese para o tema do seu projeto.**
- **O que você mudaria na sua Hipótese do projeto agora que introduzimos o conteúdo?**

T Exercício:

A empresa **Suco Bom** produz **sucos de frutas em embalagens de 500 ml**. Seu processo de produção é quase todo automatizado e as embalagens de sucos são preenchidas por uma máquina que às vezes apresenta um certo desajuste, levando a erros no preenchimento das embalagens para mais ou menos conteúdo.

Quando o volume médio cai abaixo de 500 ml, a empresa se preocupa em perder vendas e ter problemas com os órgãos fiscalizadores. Quando o volume passa de 500 ml, a empresa começa a se preocupar com prejuízos no processo de produção.

O setor de controle de qualidade da empresa **Suco Bom** extrai, periodicamente, **amostras de 50 embalagens** para monitorar o processo de produção. Para cada amostra, é realizado um **teste de hipóteses**.

Escreva quais são as hipóteses testadas para este problema.



T Exercício Resposta:

Quando o volume médio cai abaixo de 500 ml, a empresa se preocupa em perder vendas e ter problemas com os órgãos fiscalizadores. Quando o volume passa de 500 ml, a empresa começa a se preocupar com prejuízos no processo de produção.

Como nesse caso a empresa terá problemas se o volume for maior ou menor que o especificado, usamos o teste bilateral.

$$H_0 : \mu = 500$$

$$H_1 : \mu \neq 500$$



Se rejeitamos a hipótese nula, concluímos que a máquina está desregulada.

T Exercício:

Um famoso fabricante de refrigerantes alega que uma garrafinha de 350 ml de seu principal produto contém, **no máximo, 37 gramas de açúcar**.

Um consumidor desconfiado e com conhecimentos em inferência estatística resolve testar a alegação do fabricante e seleciona, aleatoriamente, em um conjunto de estabelecimentos distintos, **uma amostra de 25 latas** do refrigerante em questão. Utilizando o equipamento correto o consumidor obteve as quantidades de açúcar em todas as 25 latas de sua amostra.

Quais são as hipóteses testadas pelo consumidor para este problema?

$$H_0 : \mu \leq 37$$

$$H_1 : \mu > 37$$



INTERVALO 10 MIN



APROVEITE PARA:

- Fazer anotações do que viu até agora (aprendizados, insights, dúvidas)
- Levantar-se, esticar os braços e as pernas, relaxar por mais tempo
- Comer algo para voltar com energia renovada
- Ir ao toalete



Métodos de Amostragem

T Amostra

A amostra é a parte da população que é selecionada para o estudo, ou seja é um subconjunto de elementos de uma população maior que você coleta e analisa para fazer inferências.

Mas por quê usar uma amostra e não a população toda?

- **Custo alto** e pouca verba para obter informação da população toda.
- O **Tempo** necessário para obter informação da população toda é muito longo.
- **Impossibilidade** da coleta de todos os elementos (exemplo: um estudo com espécies marinhas).

A amostra deve ser confiável e portanto os elementos devem ser escolhidos de forma que eles representem adequadamente a população.

T Erro amostral

Por mais que o processo de coleta dos dados seja bem planejado e executado,

SEMPRE vai existir um erro amostral nos resultados, ou seja na inferência sobre o valor que queremos estimar da população.

Erro amostral é a diferença entre o resultado amostral e o verdadeiro resultado da população, esses erros resultam das flutuações amostrais devido a aleatoriedade das seleções.

Erro não amostral

Erro não amostral ocorre quando os dados amostrais são coletados, registrados ou analisados de forma incorreta (como por exemplo, seleção de uma amostra tendenciosa, uso de um instrumento de medida defeituoso, confundimento nas aplicações de tratamento dos pacientes, etc.).

T Tipos de amostragem





Amostragens Probabilísticas

T Amostra Aleatória Simples

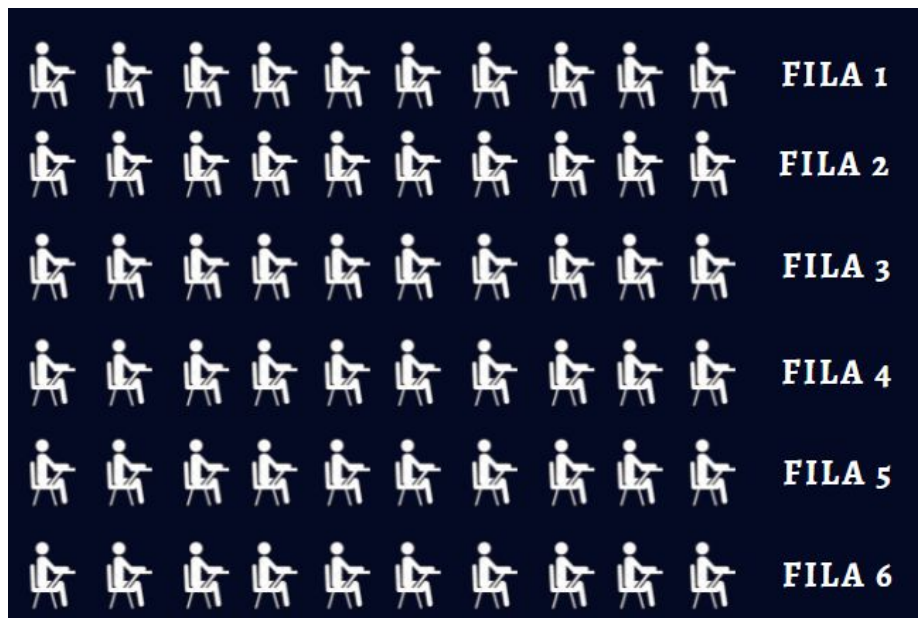
- Na amostragem aleatória, **os membros de uma população são selecionados através de um sorteio.**
- Neste sorteio todos os indivíduos têm a mesma chance de serem selecionados para fazerem parte da amostra, dessa forma, **não há critério ou filtro.**
- *Pode gerar qualquer combinação de elementos presentes em um universo, isso pode ser bom ou ruim.*
- Para que seja feito o sorteio, é necessário ter uma lista identificando cada membro da população. Pesquisas que contêm um universo muito grande, se torna quase impossível obter uma listagem atualizada de todos.

T Exemplo

Imagine uma sala de aula com 60 alunos arrumados em seis filas de 10 alunos cada.

Suponha que o professor selecione uma amostra de 10 alunos jogando um dado e selecionando a fila correspondente ao resultado da jogada.

- É uma amostra probabilística?
- É uma amostra aleatória simples?



Exemplo

- É uma amostra probabilística?

Sim! É uma amostra probabilística porque cada estudante tem uma chance conhecida de ser selecionado e a amostra é feita de forma aleatória.

- É uma amostra aleatória simples?

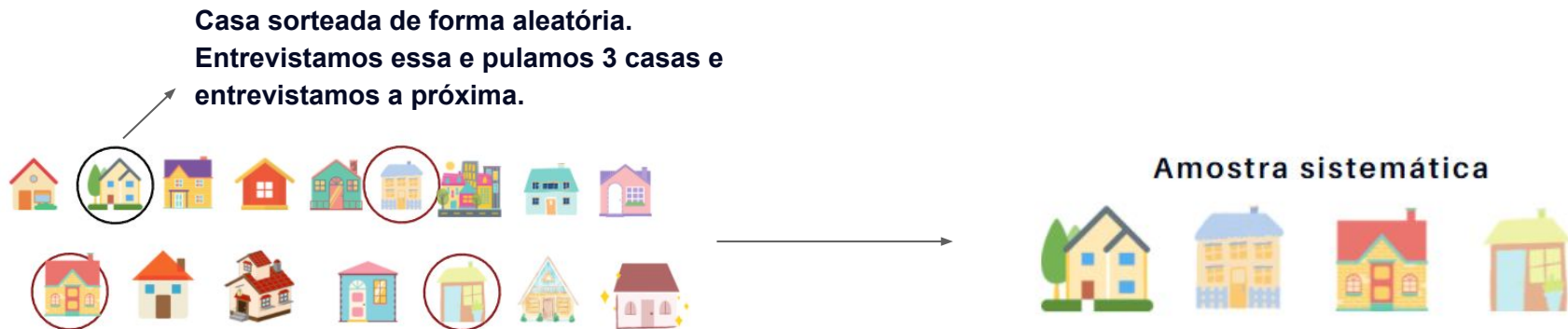
Sim! A amostra é uma amostra aleatória simples porque cada fila de estudantes tem a mesma chance (uma em seis) de serem escolhidos.

T Amostra Sistemática

A amostra sistemática é um processo muito simples e que só requer a seleção de um indivíduo aleatório. O restante é um processo rápido e simples.

- Na amostragem sistemática selecionamos um ponto inicial através de uma lista ordenada e *em seguida selecionamos os elementos da população a um intervalo fixo.*

Por exemplo, ao entrevistar um bairro, ordenamos a lista de casas pelos números, selecionamos o primeiro número a ser entrevistado e em seguida, fazemos a entrevista pulando 3 casas.



Amostra Sistemática

Os processos para determinar uma amostra sistemática são:

1. Elaborar uma lista ordenada dos N indivíduos da população.
2. Calcular o intervalo de coleta. O intervalo é obtido dividindo o total de elementos da população pelo número de elementos que queremos na nossa amostra.
3. Sorteamos um número aleatório inteiro, menor ou igual ao tamanho do intervalo para ser o ponto de início da nossa amostra.
4. Selecionamos indivíduos restantes de acordo com o tamanho do intervalo definido.

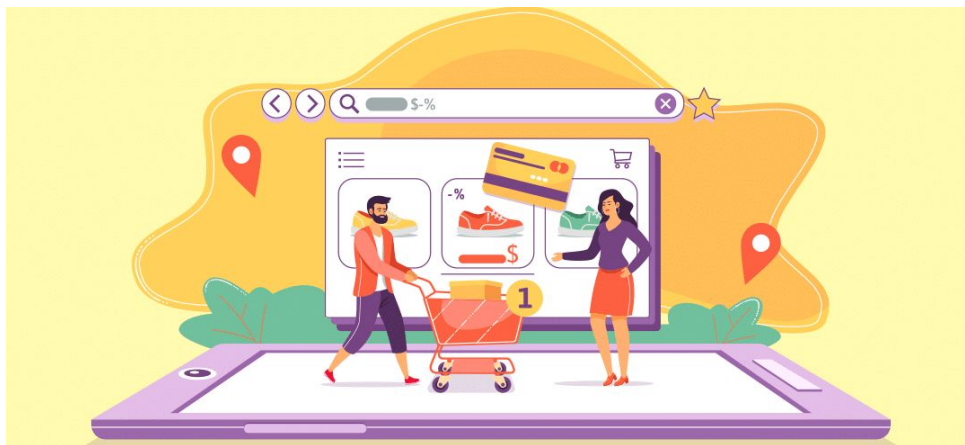
T Exemplo:

- Se tivermos uma população de 600 casas quisermos selecionar uma amostra de 100 casas: o Intervalo de Amostragem é de: **$600/100 = 6$. Isso significa que de 6 em 6 casas escolhemos uma.**
- Para determinar o ponto de início da amostragem fazemos o seguinte: como no nosso exemplo o intervalo de amostragem é de 6, teremos que escolher aleatoriamente um número entre 1 e 6.
- Supondo que em um sorteio aleatório o número escolhido seja o 4, isso quer dizer que selecionaremos uma casa de 6 em 6, começando da 4ª casa da lista.

Então a lista de casas entrevistadas seriam: (4, 10, 16, 22, 28, 34, 40, 46...).

A amostragem aleatória sistemática é útil quando se tem apenas uma lista de telefone ou uma lista de ruas com uma contagem de casas existentes em cada uma delas, por exemplo.

Quando existem **evidências sobre a diferença significativa de comportamento ou preferência entre subgrupos da população** que pretendemos estudar, é importante fazer uma amostragem que garanta que esses subgrupos (estratos) vão estar representados na nossa amostra de forma proporcional ao seu peso na população.



Ex: A literatura científica diz que existe diferença significativa entre o consumo da população feminina e masculina e entre a população rural e urbana.

Quando fizermos um estudo onde o consumo é a variável relevante, é importante que a amostra tenha homens e mulheres residentes em áreas urbanas e rurais de forma proporcional ao que existe na população em estudo.

Amostra aleatória estratificada

Amostra aleatória estratificada

Para garantir essa representação proporcional, utilizamos a amostragem aleatória estratificada que consiste em:

1. Identificar os subgrupos (estratos).
2. Calcular o peso relativo (%) de cada um dos estratos na população.
3. Utilizar em cada um dos estratos, um procedimento de amostragem aleatória simples para escolher os indivíduos de cada estrato que irão integrar a amostra (escolhemos na mesma proporção em que estão representados na população).

População:

- **Mulheres (57%):**

63% Área urbana

37% Área rural

- **Homens (43%):**

60% Área urbana

40% Área rural



Amostra (300)

- **Mulheres (57% - 171):**

63% Área urbana (108)

37% Área rural (63)

- **Homens (43% - 129):**

60% Área urbana (78)

40% Área rural (51)

- **Cada elemento da população deve pertencer a um único estrato.**

■ Amostra aleatória por Conglomerado

Na amostragem por conglomerados (clusters) o que selecionamos de forma aleatória não são os indivíduos mas os grupos.

Suponha que você deseja entrevistar os alunos de escolas públicas por exemplo.

Na amostragem por conglomerado você tem que seguir as seguintes etapas:

- 1) Levantar a lista de todas as escolas públicas existentes.
- 2) Fazer um sorteio aleatório usando a lista de escolas para definir quais escolas serão entrevistadas.
- 3) Entrevistar **todos** os alunos que estudam nas escolas sorteadas.

■ Amostra aleatória por Conglomerado

Lista de escolas:



Escolas selecionadas aleatoriamente.



150 alunos



80 alunos



95 alunos

Pesquisa é feita com todos os alunos.

■ Amostra aleatória por Conglomerado

O objetivo **principal da amostragem por conglomerado é reduzir os custos**. Usar clusters geográficos por exemplo, podem representar uma economia significativa no deslocamento.

Contudo, quanto maior a concentração de entrevistas por localidade, maior é a chance de se obter respostas homogêneas nos resultados, podendo haver uma perda de representatividade devido a exclusão de muitas regiões. Dessa forma é importante ter cuidado na escolha da área de pesquisa.

T Estratificada x Conglomerado

A principal diferença é que a amostragem por conglomerado usa todos os membros do conglomerado selecionado, enquanto a amostragem estratificada usa uma amostra proporcional dos membros de cada estrato.

Se quiséssemos, por exemplo, fazer uma amostragem considerando hospitais diferentes teríamos:

- **Na amostragem por conglomerado**

Selecionar aleatoriamente alguns destes hospitais e entrevistar todos os doentes hospitalizados.

- **Na amostragem estratificada**

Que saber quantos doentes estão hospitalizados em cada um dos hospitais e depois selecionar aleatoriamente doentes de todos os hospitais mantendo a proporção entre eles.



Amostragens não Probabilísticas

Amostragem Bola de Neve

- As pessoas entrevistadas indicam ou convidam outras pessoas para participar do questionário, fazendo com que a amostragem se comporte como uma bola de neve.
- Boa técnica para encontrar subgrupos ou segmentos de uma população que são desconhecidos ou dificilmente encontrados.

Exemplo; Usuários de drogas que indicam outros usuários que aceitam participar do estudo.

T Amostragem por Conveniência

A amostra por conveniência consiste em selecionar uma amostra da população que seja acessível.

- Ou seja, os indivíduos nessa pesquisa são selecionados porque eles estão prontamente disponíveis e não porque eles foram selecionados por meio de um critério estatístico.

Exemplo: Estudantes de uma turma em que o pesquisador é o professor, pacientes de um determinado hospital.

- O uso desse tipo de amostra é muito frequente em vários campos. Por exemplo, estudos clínicos com voluntários (vacina do Covid-19). Nestes estudos, pessoas com certas características participam voluntariamente da pesquisa ou tratamento.
- O pesquisador que usa essa técnica nestes casos, considera que qualquer indivíduo que tenha a característica (como ser hipertenso, por exemplo) será representativo.

T Amostragem por Cotas

Na amostra por quotas dividimos a população do estudo em grupos de forma que nenhum indivíduo fique de fora e que cada indivíduo só pode estar em um único grupo.

Sendo que geralmente segmentamos os indivíduos de acordo com variáveis sócio-demográficas, como: sexo, idade, classe social ou região.

Em seguida estabelecemos a quantidade de indivíduos a serem entrevistados para cada um desses grupos. Essa quantidade geralmente é escolhida de forma proporcional ao tamanho do grupo populacional.

Por exemplo, se nós definimos os grupos por sexo numa população em que 60% são homens e 40% são mulheres, e queremos obter uma amostra de 1.000 pessoas, definimos uma meta de 600 homens e 400 mulheres.

Para finalizar, buscamos participantes para cobrir todas as quotas definidas. É neste ponto que esse método deixa de ser uma amostra probabilística.

T Amostragem por Julgamento

- A escolha dos respondentes é feita partir do julgamento do pesquisador.
- O pesquisador busca por indivíduos que possuem características definidas previamente para sua amostra, ou seja, pessoas que têm comportamentos que se encaixam às características pré selecionadas, características visuais, por frequentar algum lugar que interessa a amostra.
- Tem uma função mais exploratória em uma pesquisa de opinião ou mercado.
- Pode ser utilizada para pesquisas menores, ou como um pré-pesquisa para outras que buscarão dados mais aprofundados.

T Amostragem Desproporcional

- É utilizada quando há grupos e subgrupos que geram resultados com pesos desiguais em uma pesquisa , nesse caso não há a preocupação em ter uma exata proporcionalidade da população estudada.
- O importante na amostra desproporcional é quanto um grupo dessa população é importante para o estudo.
- Grupos minoritários são priorizados em um estudo, para que pequenos grupos obtenham um mínimo de representatividade nos resultados da pesquisa.

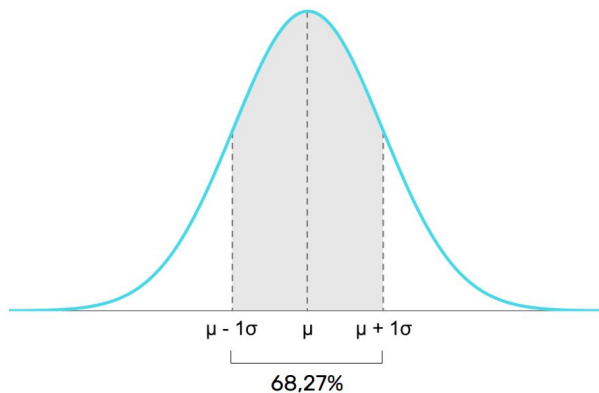
Municípios	População da região em %	Nº de amostras proporcionais	Nº de amostras desproporcionais. (Reorganização)
Município 1	61%	610	580
Município 2	34%	340	320
Município 3	5%	50	100

T

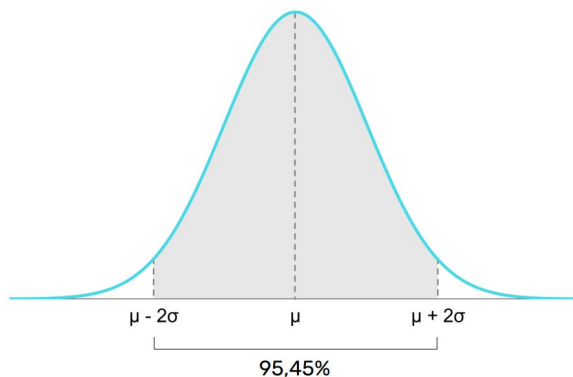
A vertical bar with a gradient from light green at the top to light blue at the bottom.

TAMANHO AMOSTRAL

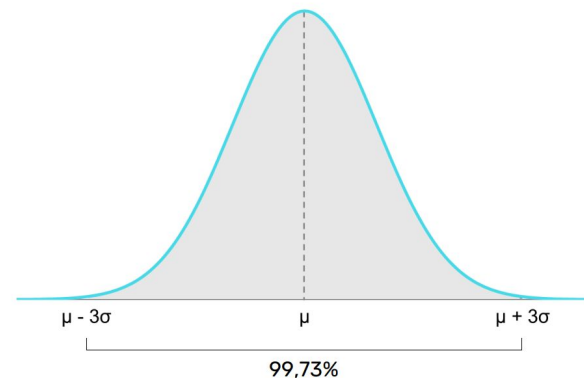
T Distribuição normal



Cerca de 68,27% de todos os valores ficam a um desvio padrão da média.



Cerca de 95,45% de todos os valores ficam a dois desvios padrão da média.



Cerca de 99,73% de todos os valores ficam a três desvios padrão da média.

T Nível de confiança



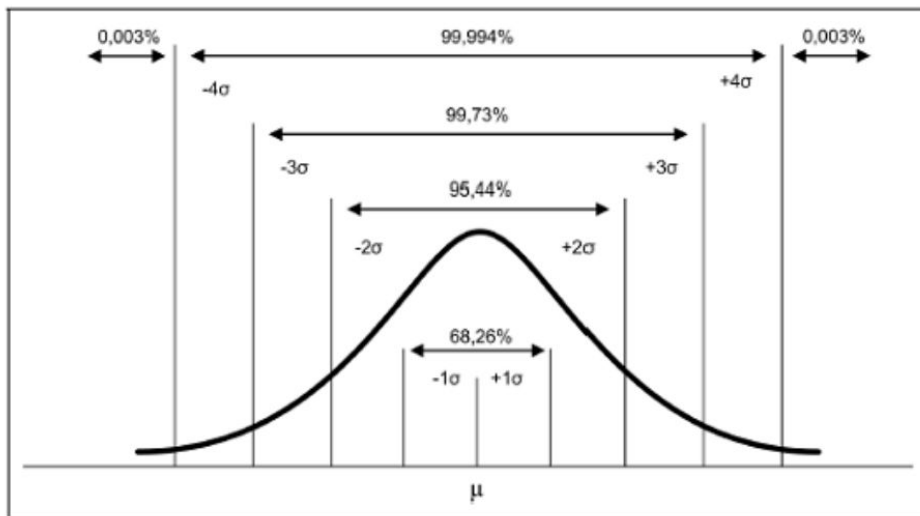
O nível de confiança é, muitas vezes, expresso como uma probabilidade $(1 - \alpha)$, onde α é o complemento do nível de confiança (**chamado de nível de significância**).

Se $\alpha = 0.05$, $(1 - \alpha) = 0,95$ ou 95%.

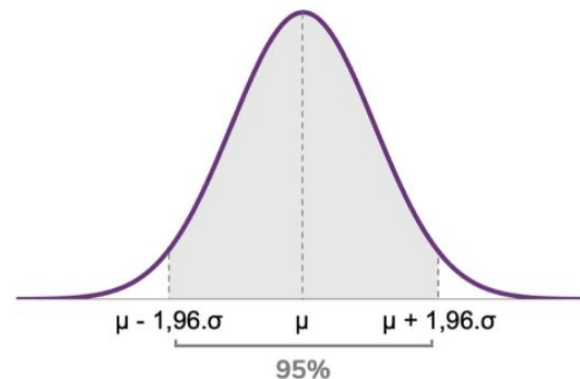
Se $\alpha = 0.01$, $(1 - \alpha) = 0,99$ ou 99%.

Se $\alpha = 0.10$, $(1 - \alpha) = 0,90$ ou 90%.

T Distribuição Normal



Como em geral estabelecemos como nível de significância (α) o valor de 5%, o ponto de corte mais usado na curva normal é o de **1,96** desvios-padrão a partir da média, o qual corresponde à **área de 95%**:





Margem de Erro

- A margem de erro é uma porcentagem que indica a proximidade dos resultados obtidos da amostra do valor real para a população total do estudo.
- Margens de erro menores oferecem resultados mais precisos, mas também exigem amostras maiores
- Na apresentação dos resultados de pesquisas eleitorais, a margem de erro geralmente é mostrada em pontos percentuais.

Por exemplo: "35% das pessoas concordam votariam no candidato X, com uma margem de erro de dois pontos percentuais para mais ou para menos".

Como definir uma amostra

- Para uma população desconhecida:

$$n = \frac{Z^2 \sigma^2 (1 - \sigma)}{\epsilon^2}$$

desvio-padrão populacional

margem de erro

Valor da tabela Z (tabela da distribuição normal baseado no nível de confiança).

Margem de Erro (α)		Nível de confiança ($1 - \alpha$)		Z
	em %		em %	
0,01	1%	0,99	99%	2,576
0,02	2%	0,98	98%	2,33
0,05	5%	0,95	95%	1,96
0,1	10%	0,9	90%	1,645

O nível de confiança é, muitas vezes, expresso como uma probabilidade ($1 - \alpha$), onde α é o complemento do nível de confiança (chamado de nível de significância).

Se $\alpha = 0.05$, ($1 - \alpha$) = 0,95 ou 95%.

Se $\alpha = 0.01$, ($1 - \alpha$) = 0,99 ou 99%.

Se $\alpha = 0.10$, ($1 - \alpha$) = 0,90 ou 90%.

Exemplo: Determine o tamanho da amostra necessário para uma população desconhecida considerando um nível de confiança de 90% , um desvio de 50% e uma margem de erro de 3%.

$$\alpha = 0.10, Z = 1.645$$

$$\varepsilon = 0.03$$

$$\sigma = 0.5$$

$$n = \frac{Z^2 \sigma (1 - \sigma)}{\varepsilon^2}$$

$$= \frac{1.645^2 0.5 (1 - 0.5)}{0.03^2}$$

$$= 756,22 = 757$$

Margem de Erro (α)		Nível de confiança ($1 - \alpha$)		z
	em %		em %	
0,01	1%	0,99	99%	2,576
0,02	2%	0,98	98%	2,33
0,05	5%	0,95	95%	1,96
0,1	10%	0,9	90%	1,645

Como definir uma amostra

- Para uma **população com tamanho conhecido**:

$$n = \frac{\frac{Z^2 \sigma(1-\sigma)}{\varepsilon^2}}{1 + \left(\frac{Z^2 \sigma(1-\sigma)}{\varepsilon^2 N} \right)}$$

- Tamanho da população

[Calculadora de tamanho de amostra](#)

Exemplo: Determine o tamanho da amostra para uma população de 425 pessoas, considerando um nível de confiança de 99% , um desvio de 50% e uma margem de erro de 5%.

$$\alpha = 0.01, Z = 2.58$$

$$\varepsilon = 0.05$$

$$\sigma = 0.5$$

Margem de Erro (α)		Nível de confiança ($1 - \alpha$)		z
	em %		em %	
0,01	1%	0,99	99%	2,576
0,02	2%	0,98	98%	2,33
0,05	5%	0,95	95%	1,96
0,1	10%	0,9	90%	1,645

$$n = \frac{\frac{Z^2 \sigma(1-\sigma)}{\varepsilon^2}}{1 + \left(\frac{Z^2 \sigma(1-\sigma)}{\varepsilon^2 N} \right)}$$

$$= \frac{\frac{2.58^2 0.5(1-0.5)}{0.05^2}}{1 + \left(\frac{2.58^2 0.5(1-0.5)}{0.05^2 425} \right)}$$

$$= 259,39 = 260$$

T Fórmula de Slovin

- é uma equação geral que é utilizada quando precisamos estimar uma população mas não temos ideia do comportamento dela.

$$n = \frac{N}{1 + N\varepsilon^2}$$

- margem de erro
- Tamanho da população

- essa é a fórmula mais imprecisa e menos recomendável de todas. Ela só deve ser utilizada em casos onde é impossível determinar um desvio padrão e um nível de confiança apropriados (o que também impede a definição de um escore z).

Exemplo: Determine o tamanho da amostra para uma população de 240 pessoas, considerando uma margem de erro de 4%.

$$n = \frac{N}{1 + N\varepsilon^2} = \frac{240}{1 + 240(0.04)^2} = 173.41 = 174$$



Fórmula genérica

Regra de bolso:

$$n = \frac{1}{\varepsilon^2} \quad \bullet \quad \text{margem de erro}$$

- Regra de bolso que fornece o número de entrevistados para a margem de erro desejada (X% de margem de erro)



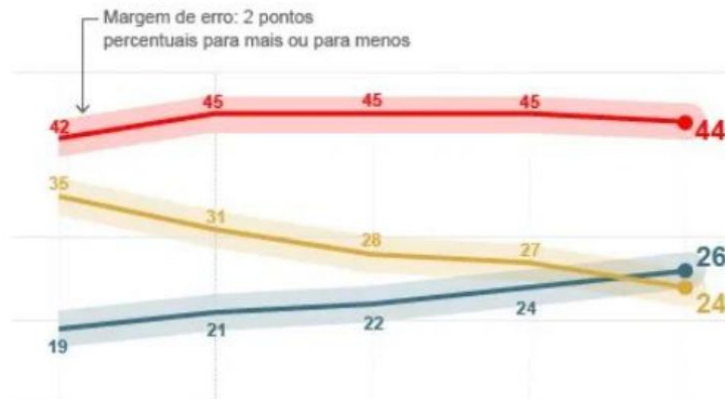
Dilma
Rousseff



Marina
Silva



Aécio
Neves



Exemplo: Determine o tamanho da amostra para um estudo que tenha margem de erro igual a 2%

$$n = \frac{1}{\varepsilon^2} = \frac{1}{0.02^2} = 2500$$

T



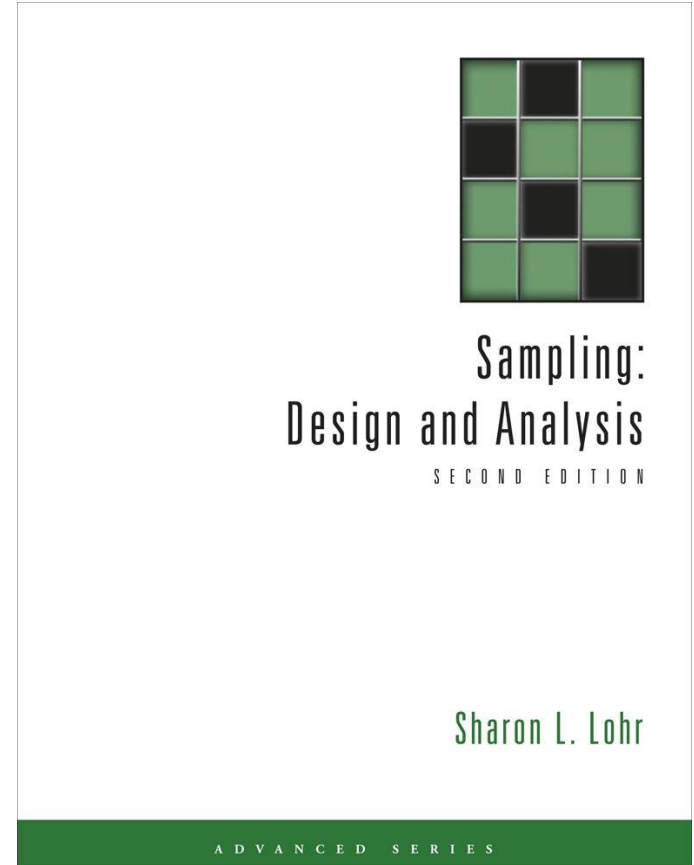
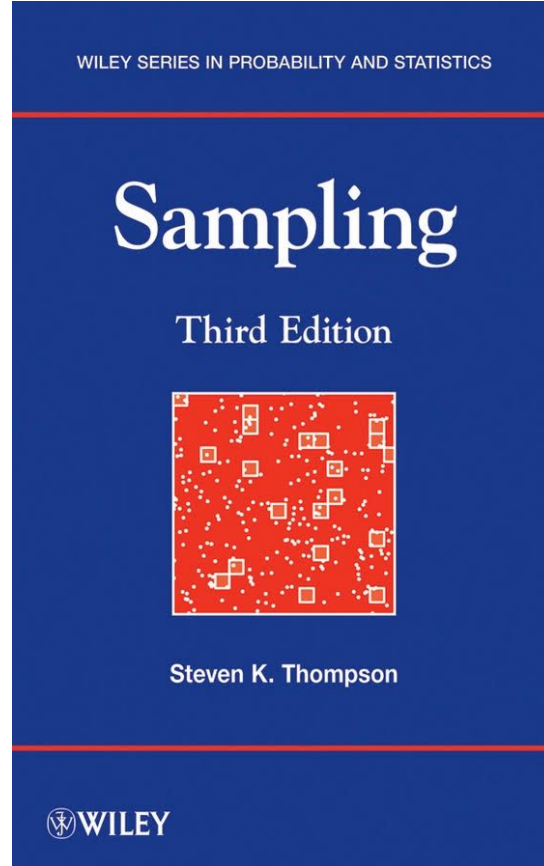
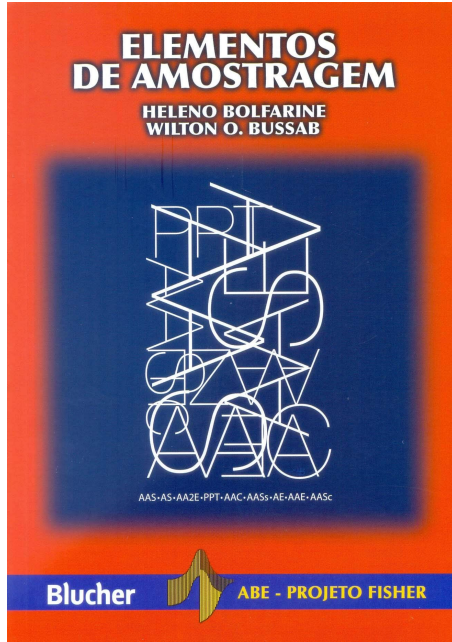
DÚVIDAS FINAIS

T

COMO FOI?



T Referências



**Boa noite
E ATÉ A PRÓXIMA!**

