Centro Universitário do Instituto de Educação Superior de Brasília – IESB
Graduação em Ciência de Dados e Inteligência Artificial
HMDC065 - Amostragem Aplicada

APOSTILA

Curso: Ciência de Dados e Inteligência Artificial

Disciplina: HMDC065 – Amostragem Aplicada

Período: 6º

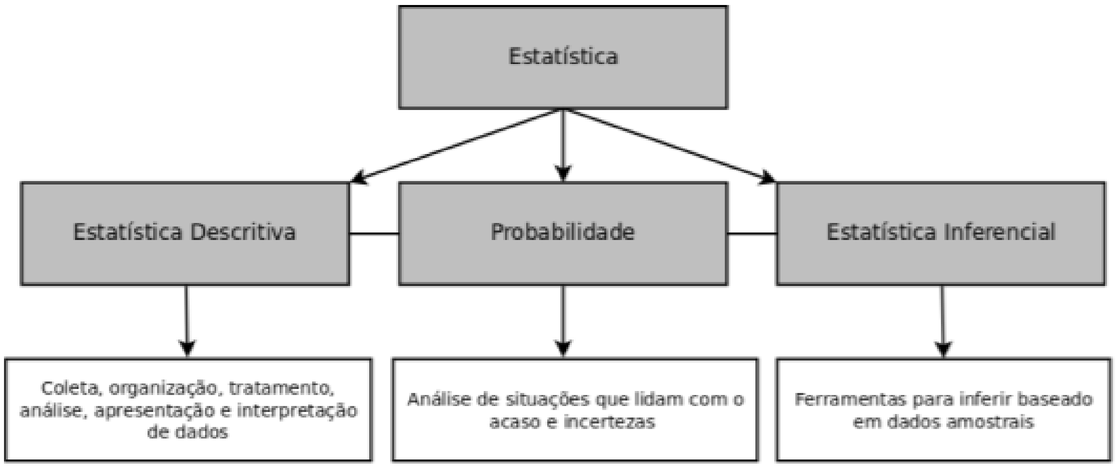
Professora: Natália Ribeiro de Souza Evangelista

E-mail: natalia.evangelista@iesb.br / natalia.evangelista@iesb.edu.br

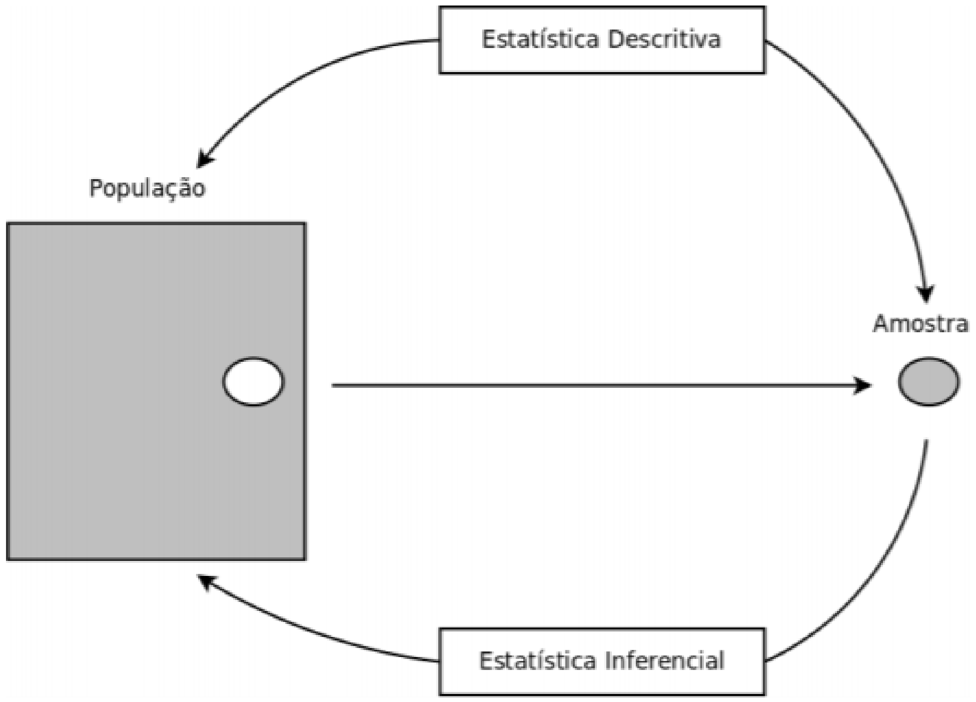
Apoio: Plataforma *Blackboard*

Neste material, será apresentado o tema Amostragem Aplicada, abordando conceitos e práticas de acordo com os tópicos estipulados na ementa do curso. Interessante que os alunos realizem a leitura, também, dos materiais de apoio a serem indicados ao longo do curso e caso tenham dúvidas, as mesmas poderão ser consultadas à professora do presente curso.

Parte I – Introdução à Amostragem - Conceitos



Quando realiza uma pesquisa, ou utiliza algum mecanismo para obter informações, um dos objetivos principais é coletar dados de uma pequena parte de um todo e aprender então alguma coisa sobre esse grupo maior.



População: conjunto de elementos que contém certa característica de interesse. Depende do interesse da pesquisa.

Exemplo:

Característica: peso dos estudantes do IESB

População: todos os estudantes do IESB

Amostra: subconjunto da população que também possui a característica de interesse.

Exemplo:

Característica: peso dos estudantes do IESB

Amostra: 100 estudantes selecionados ao acaso

Por que fazer?

Amostragem significa extrair do todo (população) uma parte representativa (amostra), com o propósito de inferir resultados a respeito do todo. Nas pesquisas científicas é comum trabalhar com amostras e a partir delas obter valores aproximados para as características populacionais de interesse. A seleção dos elementos a serem observados deve ser feita com base em uma metodologia adequada, de tal forma que os resultados sejam informativos. Exemplo típico de amostragem são as pesquisas eleitorais.

O termo inferência estatística refere-se ao uso dos dados da amostra para se ter conhecimento sobre os parâmetros da população.

Parâmetros são características específicas dos elementos da população.

Os valores obtidos a partir dos dados da amostra com a finalidade de avaliar os parâmetros desconhecidos são denominados de estimativas.

Exemplo: percentual de cada candidato divulgado antes da eleição é uma estimativa.

Vantagens da amostragem:

- Custo mais baixo;
- Resultado mais rápido;
- Operacionalidade é mais fácil de controlar.

Desvantagens da amostragem:

- População pequena;
- Característica de fácil mensuração;
- Necessidade de alta precisão.

“NÃO EXISTE NENHUMA TÉCNICA ESTATÍSTICA CAPAZ DE SALVAR UMA AMOSTRA MAL COLETADA!”

Centro Universitário do Instituto de Educação Superior de Brasília – IESB
Graduação em Ciência de Dados e Inteligência Artificial
HMDC065 - Amostragem Aplicada

LISTA DE EXERCÍCIOS nº 01

Questão 1. O gerente de uma empresa, com um total de 150 funcionários, realizou um experimento com o objetivo de verificar o consumo de água dos funcionários durante o turno de trabalho. Foram selecionados, aleatoriamente, 50 funcionários e mensurada a quantidade de litros de água consumida por cada um, no período de 30 dias. Sabe-se, também, que cada funcionário teve a mesma probabilidade de ser incluído na seleção. Com base nestas informações, relacione a segunda coluna de acordo com a primeira:

COLUNA 1	COLUNA 2
(1) Quantidade total de funcionários da empresa.	( ) Variável contínua.
(2) Consumo de litros de água por funcionário.	( ) Amostra.
(3) 50 funcionários selecionados aleatoriamente.	( ) Amostragem aleatória simples.
(4) Técnica utilizada para seleção da amostra.	( ) População.

Questão 2. Para uma população de 10 indivíduos é retirada uma amostra de 3 indivíduos, sem reposição. Assim, o número de amostras possíveis é

a) 80.

b) 120.

c) 240.

d) 720.

Questão 3. Assinale a alternativa correta:

População ou Universo é:

i) Conjunto de pessoas.
ii) Conjunto de indivíduos apresentando uma característica especial.
iii) Conjunto todos os indivíduos apresentando uma característica comum objeto de estudo.

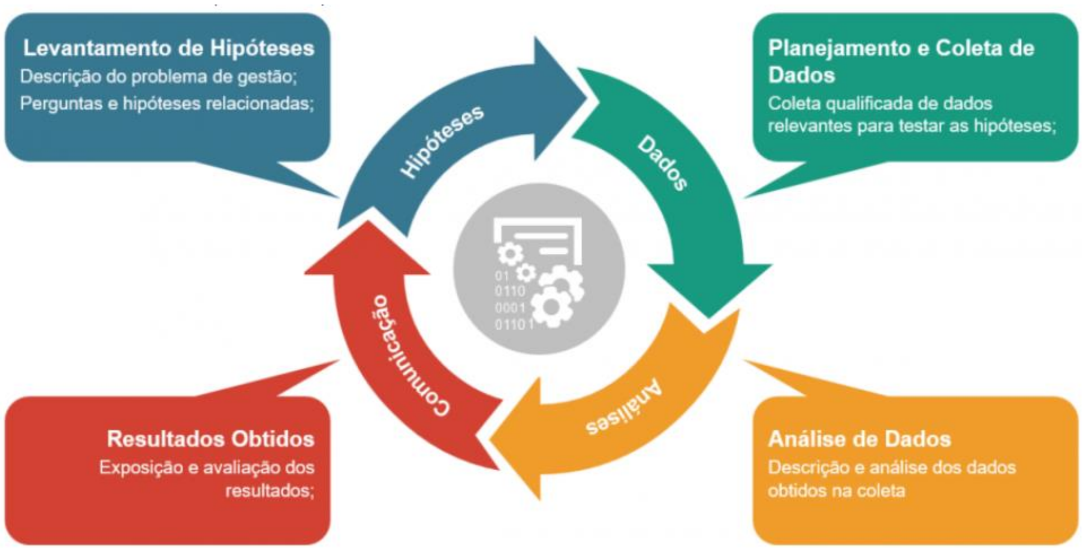
Questão 4. Opine sobre os tipos de problemas que surgiriam no seguinte plano amostral:

Para investigar a proporção dos operários de uma fábrica favorável à mudança do início das atividades das 7h para 7h30, decidiu-se entrevistar os 30 primeiros operários que chegassem à fábrica na quarta-feira.

Questão 5. Mostre que a variância amostral $s^2 = \frac{\sum(y_i - \bar{y})^2}{n-1}$ é equivalente à $s^2 = \frac{1}{n-1}[\sum y_i^2 - n\bar{y}^2]$

Parte II – Introdução à Amostragem – Coleta de Dados

Planejando a coleta de dados: para estudarmos adequadamente uma população através de uma amostra, devemos planejar a coleta de dados.



Perguntas:

Qual a pergunta a ser respondida?

Como comunicar a resposta obtida?

Qual ferramenta de análise pretendemos usar e como utilizar os resultados?

Qual tipo de dado é necessário para utilizar as ferramentas desejadas e responder à pergunta?

Como coletar esses dados com o mínimo de esforço e erro?

Onde acessar estes dados?

Quem pode nos fornecer os dados?

Qual o período em que os dados serão coletados?

Conseguindo as respostas às perguntas anteriores, devemos:

- Construir uma metodologia para nos certificar de que todas as informações estão definidas;
- Coletar os dados de forma consistente e honesta;
- Certificar-se de que existe tempo suficiente para a coleta de dados;

- Definir quais informações adicionais serão necessárias para estudos futuros, referências ou reconhecimento.

Coleta de dados: refere-se à reunião e registro sistemático de dados. É possível distinguir duas espécies de dados: dados primários e dados secundários.

- Dados primários: são publicados ou comunicados pela própria organização que os recolheu.
- Dados secundários: são publicados por outra organização.

A coleta de dados pode ser realizada de forma indireta ou direta.

- Coleta indireta: quando é inferida a partir dos elementos obtidos da coleta direta.
- Coleta direta: quando é obtida diretamente da fonte. Há três tipos de coleta direta:

  1) Coleta contínua: quando os dados são obtidos automaticamente e na vigência de um determinado período.

  Exemplo: registros de nascimento.

  2) Coleta periódica: quando é realizada em períodos determinados.

  Exemplo: recenseamento demográfico.

  3) Coleta ocasional: quando os dados são colhidos esporadicamente. Exemplo: coleta de dados a respeito de um surto epidêmico (tempos atuais COVID-19).

Processos de coleta de dados:

- Observação direta:
  - Neste processo não existe informante o próprio pesquisador é quem fornece a informação;
  - Utilizado para simples contagem;
  - Desvantagem: dependendo do fator a ser observado, pode apresentar problemas por causa do critério subjetivo do observador, do método de observação.
  - Exemplo: pesquisa em laboratórios com ratos; contagem do número de alunos nessa sala.

- Experimental:
  - Processo experimental de demonstração: o objetivo deste processo é provocar reações através de uma experiência com um novo produto. Muito usado nas pesquisas de mercado.

        - Exemplo: testar experimentalmente 2 produtos semelhantes e verificar a preferência dos consumidores após sua utilização.

    - Processo experimental de explicação funcional: o objetivo é pesquisar reações de coletividade à presença ou ausência de certas condições. Neste caso, não existe a demonstração de um ou mais produtos.
        - Exemplo: verificar se a propaganda influi numa campanha de vacinação.

- Mecânico:
    - Não existe informante. O observador é substituído por um aparelho mecânico ou eletrônico.
        - Exemplo: aparelho de contagem de tráfego em uma rodovia.

- Questionário:
    - É o mais usado. Envolve qualquer formulário de registro de informações.
    - Características de um questionário:
        - Devem ser planejados cuidadosamente;
        - Facilita a obtenção de dados;
        - Ajuda a coletar a informação de maneira completa e eficiente;
        - Evita a coleta de dados inúteis.

- Inquérito pessoal:
    - Lista de informações a serem coletadas sem necessidade de entrevista.

Pesquisa-piloto: aplicação efetiva do instrumento em indivíduos com características similares aos indivíduos da população em estudo. Assim, é possível detectar algumas falhas que tenham passado despercebidas, como:

- ambiguidade de perguntas;
- resposta sem previsão;
- não variabilidade de resposta.

LISTA DE EXERCÍCIOS nº 02

Questão 1. Que aspectos devem ser observados quando utilizamos dados secundários?

Questão 2. Cite pelo menos uma situação em que os dados são coletados através de:

a) Levantamentos contínuos.

b) Levantamentos periódicos.

c) Levantamentos ocasionais.

Questão 3. Identifique qual o procedimento de coleta de dados usadas nas situações abaixo e suas vantagens e desvantagens:

a) Amostragem sobre os hábitos de compra de gêneros alimentícios de uma certa área, por telefone.

b) Distribuição de questionários, pelo correio, para estudar hábito de leitura de jornais dos respondentes.

c) Estudo da relação criança hospitalizada e família, acompanhada por um observador.

d) Estudo do nível de poluição atmosférica medida por aparelhos, na Rodoviária.

Questão 4. Que aspectos devem ser observados na confecção de um questionário?

Questão 5. Escolha um assunto/tema de seu interesse, defina as variáveis a serem pesquisadas e elabore um plano de coleta de dados.

Parte III – Introdução à Amostragem

Teoria da amostragem:

Identificar amostra probabilística e não probabilística:

- Amostra probabilística (aleatória): processo de amostragem em que cada elemento da população tem uma chance fixa de ser incluído na amostra. Caracteriza-se por:

• Amostragem de elementos versus amostragem por grupos (cluster);

• Probabilidades iguais de unidades versus probabilidades desiguais;

• Seleção não estratificada versus estratificada;

• Seleção aleatória versus sistemática;

• Técnicas de estágio único versus de estágio múltiplo.

- Amostra não probabilística (não aleatória): técnica de amostragem que não utiliza seleção aleatória. Ao contrário, confia no julgamento pessoal do pesquisador.

Distinguir os dois grandes grupos de amostra:

As amostras podem ser divididas em dois grandes grupos, de acordo com a representatividade dos elementos na população, os objetos de estudos e as possibilidades de generalização de resultado.

- Amostra quantitativa: É utilizada na pesquisa quantitativa. Nesse tipo de amostra, preocupa-se com as características gerais ou comuns de todos os elementos, sendo possível generalizar os resultados objetivos na inferência estatística para toda a população.

Exemplo: pense em uma pesquisa de marketing com os consumidores de uma loja de departamentos específica, sendo essa pesquisa relacionada à disposição dos setores de moda. A característica comum aos elementos da população é frequentar a determinada loja na primeira quinzena do mês de janeiro. Sabemos quem são esses clientes e temos, de certa forma, fácil acesso a eles. Podemos aqui fazer uma pesquisa quantitativa, selecionando de forma aleatória cada um dos elementos da amostra. Nesse caso, poderemos generalizar para todos os frequentadores da loja no período especificado.

- Amostra qualitativa: É utilizada na pesquisa qualitativa. Nesse tipo de amostra, atenta-se para as características individuais de cada elemento, não objetivando, portanto, fazer generalizações a partir dos resultados.

Exemplo: queremos saber sobre hábitos de consumo dos consumidores de chantilly industrializado em uma grande capital brasileira. As características da nossa população são:

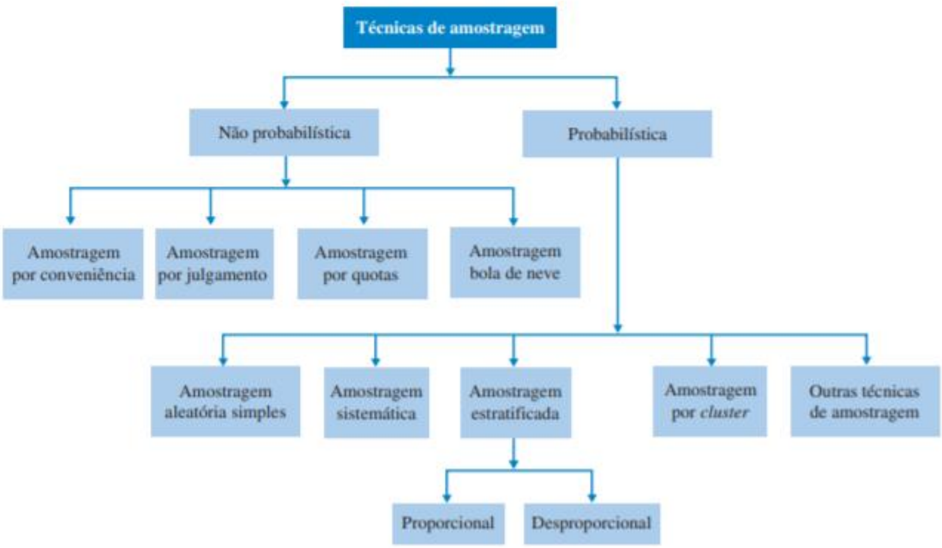
residir na grande capital brasileira escolhida e ter o hábito de adquirir chantilly industrializado. Primeiramente, quantas pessoas que você conhece têm o hábito de comprar chantilly industrializado? Percebe-se que é difícil ter acesso e saber onde encontrar esses consumidores, bem como conhecer o perfil e o comportamento deles. Nesse caso, uma pesquisa quantitativa seria muito difícil e onerosa de ser delineada; o mais apropriado é partir para uma pesquisa qualitativa. Poderíamos, de repente, pedir indicações em redes sociais de pessoas que se disponibilizassem a responder um questionário com perguntas dissertativas sobre o assunto ou, até mesmo, participar de entrevistas pessoais em profundidade. Essa pesquisa de forma alguma permitiria afirmar que toda a população de consumidores da capital tem o perfil encontrado; ou seja, não poderíamos generalizar os resultados, mas para a tomada de decisão de uma empresa de chantilly, as informações poderiam ajudar a nortear a escolha de novas estratégias.

- Como escolher entre quantitativa e qualitativa:

- tempo disponível;
- objetivos da pesquisa;
- resultado esperado;
- análises finais que se deseja realizar;
- população escolhida;
- hipóteses de pesquisa;
- disponibilidade das unidades amostrais;
- valor monetário disponível.

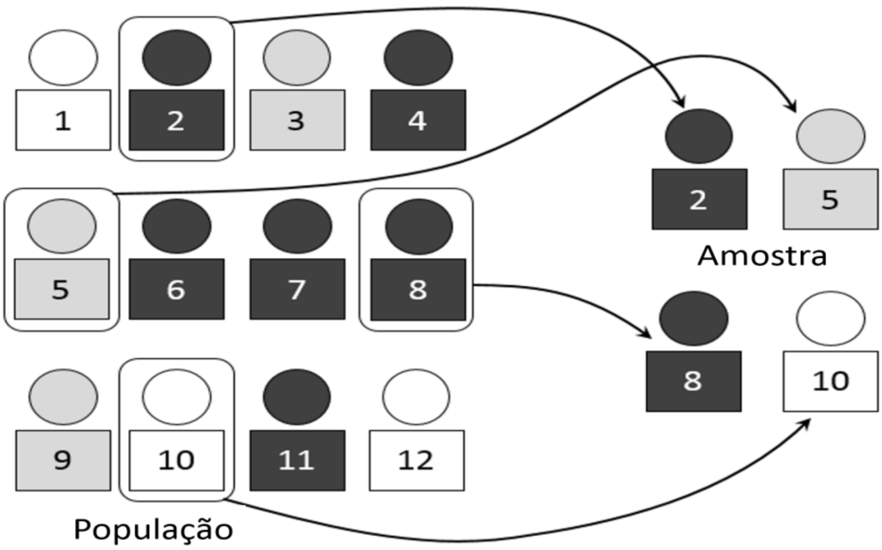
Amostragem	Tamanho da amostra	Forma de coleta	Tipos de resultados
Quantitativa probabilística	Depende do erro e da confiança a serem aceitos. Um número representativo da população.	Escolha feita de forma aleatória (sorteio).	Dados quantitativos, estatísticas, resumos numéricos, tabelas, gráficos. Pode-se fazer inferência para a população. Erro e confiança fixados.
Quantitativa não probabilística	Um número representativo da população.	Escolha feita de forma não aleatória, de acordo com o julgamento do pesquisador (quotas, conveniência, proximidade, etc.).	Dados quantitativos, resumos numéricos, tabelas, gráficos. Não se pode fazer inferência para a população.
Qualitativa	Um pequeno grupo de unidades amostrais.	Escolha feita de forma não aleatória, de acordo com o julgamento do pesquisador.	Dados qualitativos, maior profundidade na análise. Não permite generalizações.

Noções dos tipos de amostragem:



- Amostragem aleatória simples:

- tipo de amostragem probabilística mais utilizada;
- todos os elementos da população têm a mesma probabilidade de pertencerem à amostra;
- população homogênea;
- selecionar uma amostra "n" a partir de uma população "N";
- sorteio / tabela de números aleatórios;
- com / sem reposição.



- Exemplo: um pesquisador deseja comparar os teores médios de proteína de três cultivares de cevada. Para executar o experimento ele dispõe de uma área de terra homogênea (mesma fertilidade, mesma umidade etc.) de tamanho 288 m². Portanto, as três cultivares serão comparadas em igualdade de condições. Um princípio básico da experimentação é o uso de repetições, ou seja, são necessários pelo menos dois valores para cada cultivar. Assim, a área total vai ser dividida em 12 canteiros de tamanhos 6m x 4m, totalizando 24m²/canteiro.

(tabela de números aleatórios) →{03,05,01,08,07,02,12,10,11,04,09,06}.

Canteiro 1 **Cultivar A**	Canteiro 2 **Cultivar B**	Canteiro 3 **Cultivar A**
Canteiro 4 **Cultivar C**	Canteiro 5 **Cultivar A**	Canteiro 6 **Cultivar C**
Canteiro 7 **Cultivar B**	Canteiro 8 **Cultivar A**	Canteiro 9 **Cultivar C**
Canteiro 10 **Cultivar B**	Canteiro 11 **Cultivar C**	Canteiro 12 **Cultivar B**

LISTA DE EXERCÍCIOS nº 03

Questão 1. Defina amostra quantitativa e qualitativa.

Questão 2. Qual amostragem permite realizar inferência estatística?

Questão 3. Exemplifique amostra não probabilística.

Questão 4. Considere os dados da população sobre a "Satisfação do aluno com o Curso" e realize uma amostra aleatória simples de n = 10 questionários. Use a tabela de números aleatórios.

Questão 5. Apresente os dados amostrais correspondentes as variáveis coletadas para esta população (variáveis
3(a) Didática
3(b) Conhec.
3(c) Bibliog
3(d) Labor.
3(e) Discipl.
3(f) Curric.
3(g) Satisf. Geral
4(a) Ponto Posit.
4(b) Ponto Negat.
5 Desemp.)

(vide arquivo Avaliacao_do_curso.xls)

Parte I – Amostragem e Tamanho da Amostra

I – Amostragem Aleatória Simples (AAS): para selecionar uma amostra aleatória simples é necessário, entre outras coisas:

-Conhecer os elementos da população;

-Todos os elementos da população precisam ter a mesma probabilidade de pertencer à amostra.

A amostra pode ser:

-Com reposição: elemento pode estar na amostra mais de 1 vez;

-Sem reposição: elemento só pode estar na amostra 1 vez;

Se a população for infinita (muito grande) então as retiradas com e sem reposição serão equivalentes, o fato de se recolocar o elemento retirado de volta na população não vai afetar em nada a probabilidade de extração do elemento seguinte. Se, no entanto, a população for finita (e pequena) será necessário fazer uma distinção entre os dois procedimentos, pois na extração com reposição as diversas retiradas serão independentes, mas no processo sem reposição haverá dependência entre as retiradas.

Sorteio com a tabela de números aleatórios.

Exemplo: um pesquisador deseja comparar os teores médios de proteína de três cultivares de cevada. Para executar o experimento ele dispõe de uma área de terra homogênea (mesma fertilidade, mesma umidade etc.) de tamanho 288 m². Portanto, as três cultivares serão comparadas em igualdade de condições. Um princípio básico da experimentação é o uso de repetições, ou seja, são necessários pelo menos dois valores para cada cultivar. Assim, a área total vai ser dividida em 12 canteiros de tamanhos 6m x 4m, totalizando 24m²/canteiro.

(tabela de números aleatórios) →{03,05,01,08,07,02,12,10,11,04,09,06}.

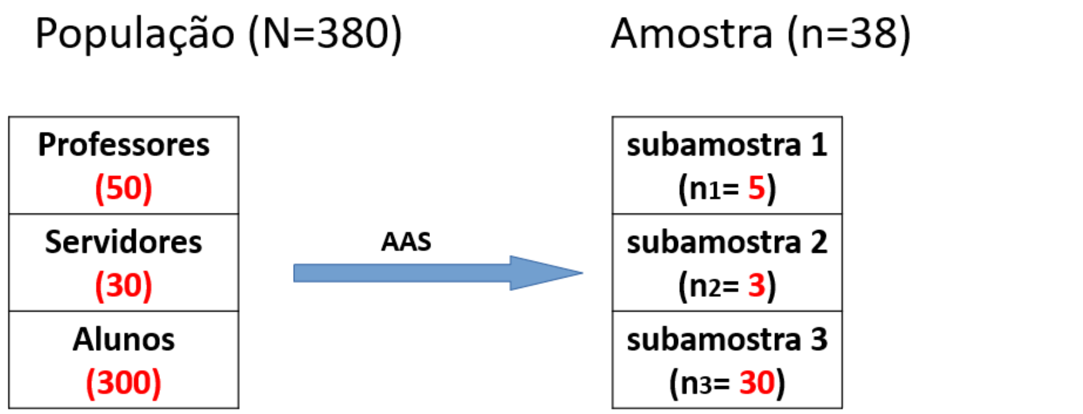
Canteiro 1 **Cultivar A**	Canteiro 2 **Cultivar B**	Canteiro 3 **Cultivar A**
Canteiro 4 **Cultivar C**	Canteiro 5 **Cultivar A**	Canteiro 6 **Cultivar C**
Canteiro 7 **Cultivar B**	Canteiro 8 **Cultivar A**	Canteiro 9 **Cultivar C**
Canteiro 10 **Cultivar B**	Canteiro 11 **Cultivar C**	Canteiro 12 **Cultivar B**

Tamanho da amostra (n):

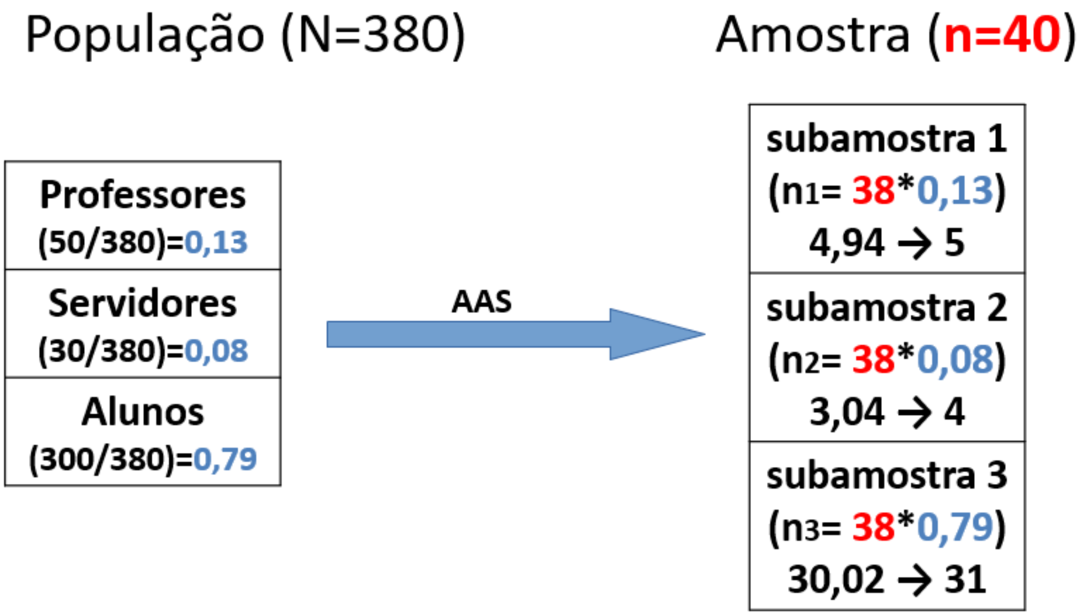
Estimar	n	Erro
μ	$\frac{N\sigma^2}{(N-1)D+\sigma^2}$	$D=\frac{B^2}{4}$
Total	$\frac{N\sigma^2}{(N-1)D+\sigma^2}$	$D=\frac{B^2}{4N^2}$
Proporção q = 1 - p	$\frac{Npq}{(N-1)D+pq}$	$D=\frac{B^2}{4}$

II – Amostragem Estratificada (AE): consiste em dividir a população em subgrupos mais homogêneos e que são denominados de estratos. Em cada um dos estratos formados seleciona-se a amostra utilizando a amostragem aleatória simples.

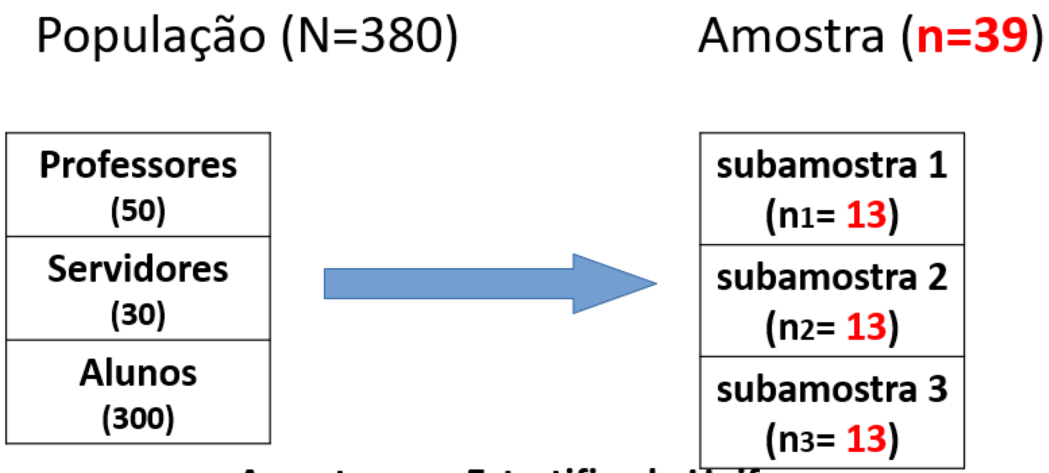
Exemplo: Com o objetivo de estudar o estilo de liderança preferido pela comunidade de uma escola, realizou-se um levantamento por amostragem. A comunidade escolar é composta por 50 professores, 30 servidores técnicos administrativos e 300 alunos (N=380 indivíduos divididos claramente em 3 estratos). Como retirar uma amostra estratificada com tamanho equivalente a 10% do tamanho da população, ou seja, n=38.



F = n/N = 38/380 = 0,10 ou 10% em cada estrato
Cuidado! Representatividade



Amostragem Estratificada Proporcional



Amostragem Estratificada Uniforme

(mesma quantidade de elementos em cada estrato: 38/3≈12,6)

Cuidado! Representatividade

Para pensar e resolver:

1) Em uma empresa com cinco departamentos existem 150 funcionários, sendo: 18 no departamento A; 22 no departamento B; 25 no departamento C; 55 no departamento D e 30 no departamento E. Obtenha uma amostra de 15 funcionários, representativa da população, proporcionalmente a cada departamento.

2) Uma população encontra-se dividida em 4 estratos, com tamanhos, respectivamente: N1 = 50; N2 = 70; N3 = 80 e N4 = 20. Sabendo-se que, ao ser realizada uma amostragem estratificada proporcional, 7 elementos da amostra foram retirados do segundo estrato, determine o número total de elementos da amostra e o número de elementos da amostra de cada estrato.

Tamanho da amostra:

Estimar	n	Erro
μ	$\dfrac{\sum_{i=1}^{k} \frac{N_i^2 \sigma_i^2}{w_i}}{N^2 D + \sum_{i=1}^{k} N_i \sigma_i^2}$	$D = \dfrac{B^2}{4}$
Total	$\dfrac{\sum_{i=1}^{k} \frac{N_i^2 \sigma_i^2}{w_i}}{N^2 D + \sum_{i=1}^{k} N_i \sigma_i^2}$	$D = \dfrac{B^2}{4N^2}$
Proporção q = 1 - p	$\dfrac{\sum_{i=1}^{k} \frac{N_i^2 p_i q_i}{w_i}}{N^2 D + \sum_{i=1}^{k} N_i p_i q_i}$	$D = \dfrac{B^2}{4}$

III - Amostragem Sistemática (AS): conveniente quando a população está ordenada segundo algum critério como fichas, lista telefônica, linhas de plantio.

Na AS devemos considerar uma quantia denominada intervalo de seleção (IS) em que: N é o tamanho da população e n é o tamanho da amostra.

Exemplo: Em uma empresa temos o cadastro de 5.000 clientes (N) dos quais quer sortear 1.000 indivíduos (n) para uma pesquisa.

O primeiro elemento (cliente) que irá pertencer a amostra será sorteado dentre os 5 primeiros cadastros (valor informado pelo IS) e os demais serão obtidos sistematicamente na população da seguinte forma.

1º elemento: 3º cadastro (selecionado ao acaso)

2º elemento: 3º cadastro + IS

3º elemento: 3º cadastro + 2*IS

4º elemento: 3º cadastro + 3*IS

5º elemento: 3º cadastro + 4*IS

...

1.000º elemento: 3º cadastro + (n-1)*IS

Amostra = {3º, 8º, 13º, 18º, 23º, ..., 4.998º}

1) A sistematização proporciona uma boa estimativa da média e do total, devido à distribuição uniforme da amostra em toda população;

2) Uma amostra sistemática é executada com maior rapidez e menor custo que uma aleatória;

3) O deslocamento entre as unidades é mais fácil pelo fato de seguir uma direção fixa e preestabelecida, resultando em tempo gasto menor e, por consequência, um menor custo de amostragem.

Tamanho da amostra:

Estimar	n	Erro
μ	$\frac{N\sigma^2}{(N-1)D+\sigma^2}$	$D = \frac{B^2}{4}$
Total	$\frac{N\sigma^2}{(N-1)D+\sigma^2}$	$D = \frac{B^2}{4N^2}$
Proporção q = 1 - p	$\frac{Npq}{(N-1)D+pq}$	$D = \frac{B^2}{4}$

IV – Amostragem de Conglomerado (AC): estruturalmente chamamos de conglomerado a um agrupamento de elementos da população com características similares.

Exemplo: Em uma população constituída por domicílios residenciais de uma cidade, os quarteirões formam conglomerados de domicílios. Em um 1º estágio selecionam-se os conglomerados (por AAS) e todos os elementos dos conglomerados são avaliados (amostragem de conglomerados em um estágio), ou, como é mais comum, faz-se uma nova seleção dentro de cada conglomerado selecionado no 1º estágio (amostragem de conglomerados em dois estágios).

população é dívida em conglomerados

➔ seleciona-se uma amostra de conglomerados

➔ são observados todos os elementos dos conglomerados selecionados

➔ ou é realizado um segundo estágio (ou são realizados mais estágios) até que no último estágio todos os elementos são observados

➔ deve haver uma lista identificando grupos de elementos (conglomerados) da população

Elementos	Conglomerados
Eleitores	Domicílios
Trabalhadores	Empresas
Alunos	Escolas

Exemplo mais complexo:

Para obter uma amostra de famílias:
selecionar primeiro uma amostra de cidades;
selecionar bairros das cidades sorteadas;
selecionar quarteirões dos bairros sorteados;
selecionar domicílios dos quarteirões sorteados.

Amostragem por conglomerados em 4 estágios

VANTAGENS:

➔ não precisa de uma lista identificando os elementos da população;

➔ em geral, menor custo por elemento, principalmente se o custo por observação cresce quando aumenta a distância entre os elementos.

DESVANTAGEM:

➔ plano amostral é menos eficiente, já que dentro dos conglomerados os elementos tendem a ser mais parecidos.

Tamanho da amostra:

Estimar	n	Erro
μ	$\frac{N\sigma_c^2}{ND + \sigma_c^2}$	$D = \frac{B^2\bar{M}^2}{4}$
Total	$\frac{N\sigma_c^2}{ND + \sigma_c^2}$	$D = \frac{B^2}{4N^2}$
Proporção q = 1 - p	$\frac{N\sigma_c^2}{ND + \sigma_c^2}$	$D = \frac{B^2\bar{M}^2}{4}$

$$\sigma_c^2 é estimado por: Proporção: s_c^2 = \frac{\sum_{i=1}^{k}(a_i - \hat{p}m_i)^2}{(n-1)}; Média e Total: s_c^2 = \frac{\sum_{i=1}^{k}(y_i - \bar{y}m_i)^2}{(n-1)}$$

N = the number of *clusters* in the population
n = the number of clusters selected in a simple random sample
m_i = the number of elements in cluster i, $i = 1, \ldots, N$
$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$ = the average cluster size for the sample
$M = \sum_{i=1}^{N} m_i$ = the number of elements in the population
$\bar{M} = M/N$ = the average cluster size for the population
y_i = the total of all observations in the ith cluster

LISTA DE EXERCÍCIOS nº 04

Questão 1. Numa empresa com 1000 funcionários, deseja-se estimar a percentagem dos favoráveis a certo treinamento. Qual deve ser o tamanho da amostra aleatória simples que garanta um erro amostral não superior a 5%?

Questão 2. Considere 3 estratos. Cada estrato apresenta, respectivamente, $\sigma_1^2 = 25; \sigma_2^2 = 225; \sigma_3^2 = 100$. Deseja-se estimar a média populacional. Escolha o tamanho da amostra considerando um erro (B) = 2 horas. Cada estrato apresenta:$w_1 = w_2 = w_3 = \frac{1}{3}$e $N_1 = 155; N_2 = 62; N_3 = 93$

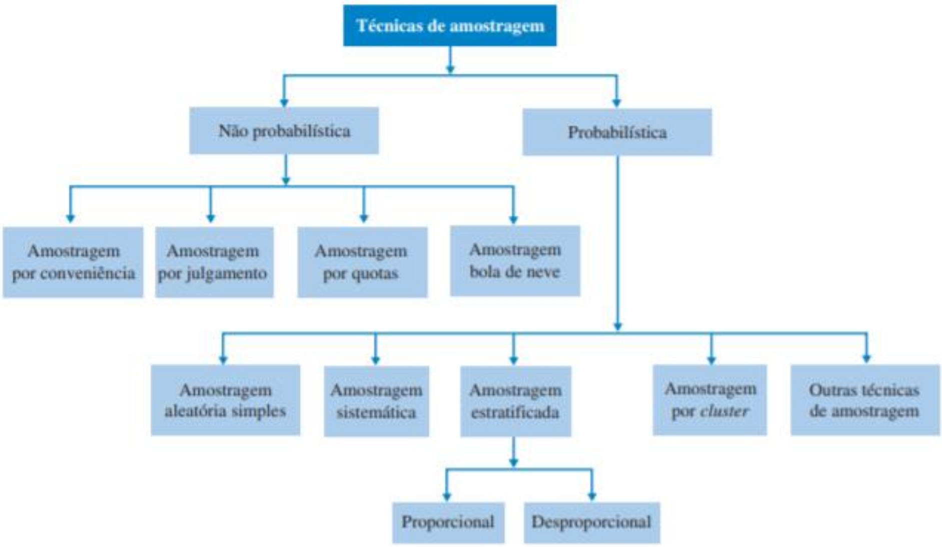
Questão 3. Uma operadora telefônica pretende saber a opinião de seus assinantes comerciais sobre seus serviços na cidade de Vargem Alegre. Supondo que há 25037 assinantes comerciais, e a amostra precisa ter no mínimo 800 elementos. Mostre como seria organizada uma amostragem sistemática para selecionar os respondentes.

Questão 4. Um sociólogo quer estimar a média da renda *per capita* numa pequena cidade. Não há nenhuma lista de residentes. Como delinear a pesquisa amostral nessa situação?

Questão 5. Crie uma situação. Especifique um delineamento amostral para essa situação usando amostragem probabilística.

Questão 6. Pesquise sobre amostragem de Bernoulli.

Parte II – Amostragem e Tamanho da Amostra



https://pesquisa-eaesp.fgv.br/sites/gvpesquisa.fgv.br/files/arquivos/veludo_-_amostragem_nao_probabilistica_adequacao_de_situacoes_para_uso_e_limitacoes_de_amostras_por_conveniencia.pdf

Amostragem não probabilística:

Situações adequadas:

- Quando as probabilidades de seleção de amostras são desconhecidas e não podem ser estimadas, o pesquisador tem diante de si um problema difícil.

- Diante de tal problema, frequentemente, as pessoas fazem inferências sobre a população por critérios arbitrários e por amostras informais.

- Os pesquisadores assumem, implicitamente, que tais itens selecionados são "itens típicos" e que as características importantes estão distribuídas uniformemente ou aleatoriamente na população.

- Uma sugestão para a decisão da escolha do plano de amostragem é avaliar a importância da fidedignidade dos resultados e a tolerância em relação a possíveis erros envolvidos e aos métodos pelos quais os erros possam ser controlados.

- Se as condições se apresentam de tal forma que estimativas razoavelmente "grosseiras" são aceitáveis para a resolução do problema, então o preço a ser pago para uso de uma amostra altamente "precisa" pode não ser justificável. De acordo com essas condições, os vieses decorrentes de um método não probabilístico podem ser considerados menos importante que o custo associado a métodos probabilísticos rigorosos.

- Em algumas situações uma amostra probabilística é praticamente impossível de ser realizada.

- "Uma razão para o uso de amostragem não probabilística pode ser a de não haver uma alternativa viável porque a população não está disponível para ser sorteada. Outra razão é que apesar da amostragem probabilística ser tecnicamente superior na teoria, ocorrem problemas em sua aplicação prática o que enfraquece essa superioridade. O resultado de um processo de amostragem probabilístico *a priori* pode resultar em um estudo não probabilístico devido a erros que os entrevistadores podem cometer quando não seguem corretamente as instruções. Outro motivo pode ser que a obtenção de uma amostra de dados que reflitam precisamente a população não seja o propósito principal da pesquisa. Se não houver intenção de generalizar os dados obtidos na amostra para a população, então não haverá preocupações quanto à amostra ser mais ou menos representativa da população. A última razão para usar amostragem não probabilística se refere às limitações de tempo, recursos financeiros, materiais e pessoas necessários para a realização de uma pesquisa com amostragem probabilística". (Mattar, F. p. 157).

A amostragem não probabilística é usada tipicamente nas seguintes situações exploratórias de um projeto de pesquisa:

1. Pré-teste de questionários;

2. Quando se trata de uma população homogênea;

3.Quando o pesquisador não possui conhecimentos estatísticos suficientes;

4. Quando o fator facilidade operacional é requerido.

Na amostragem não probabilística, os elementos da população não têm a mesma probabilidade de serem selecionados, portanto não há garantias da representatividade da população.

I - Amostragem por Conveniência:

Definição: elementos selecionados por serem imediatamente disponíveis. É adequada e frequentemente utilizada para geração de ideias em pesquisas exploratórias.

A amostra por conveniência é empregada quando se deseja obter informações de maneira rápida e barata.

Uma vez que esse procedimento consiste em simplesmente contatar unidades convenientes da amostragem, é possível recrutar respondentes tais como estudantes em sala de aula, mulheres no shopping, alguns amigos e vizinhos, entre outros. Este método também pode ser empregado em pré-testes de questionários.

Exemplo: uma repórter entrevistando pessoas na rua.

As amostras por conveniência são o tipo de amostragem menos confiável pois o pesquisador seleciona a amostra conforme sua conveniência, havendo pouco rigor na seleção.

O problema de amostras por conveniência é que não há como saber se todas as pessoas incluídas na amostra são representativas da população.

II – Amostragem por Julgamento:

Definição: uma pessoa experiente no assunto escolhe intencionalmente os elementos a serem amostrados. Se for adotado um critério razoável de julgamento, pode-se chegar a resultados favoráveis.

A abordagem da amostragem por julgamento pode ser útil quando é necessário incluir um pequeno número de unidades na amostra. O método de julgamento é muito utilizado para a escolha de uma localidade "representativa" de um país na qual serão realizadas outras pesquisas, sendo algumas vezes até preferida em relação à seleção de uma localidade por métodos aleatórios.

Exemplo: novo produto “testado” entre funcionários. Pode-se identificar grupos específicos que estariam dispostos a dar sua opinião em relação ao produto. Se o pesquisador avaliar que este grupo não gostou do produto, então ele poderia assumir que o resto da população também teria uma percepção negativa em relação à mudança. Se o grupo gostou do produto, então mais pesquisas poderiam ser requeridas para se chegar a uma conclusão a respeito do assunto.

“A característica chave da amostragem por julgamento é que os elementos da população são selecionados intencionalmente. Esta seleção é feita considerando que a amostra poderá oferecer as contribuições solicitadas.” (Churchill, p. 301) Caso isso não ocorra, esse tipo de amostragem terá pouca ou nenhuma validade.

III – Amostragem por Quotas

Definição: A amostra por quotas constitui um tipo especial de amostra intencional em que o pesquisador procura obter uma amostra que seja similar à população sob algum aspecto. Neste caso, são consideradas várias características da população, como sexo, idade e tipo de trabalho as variáveis mais comuns são áreas geográficas, sexo, idade, raça e uma medida qualquer de nível econômico - a amostra pretende incluir proporções similares de pessoas com as mesmas características.

A ideia de amostragem por quotas sugere que se as pessoas são representativas em termos de características, elas também poderão ser representativas em termos da informação procurada pela pesquisa. Depois de serem identificadas as proporções de cada tipo a ser incluído na amostra, o pesquisador estabelece um número ou quota de pessoas que possuem as características determinadas e que serão contatadas pela pesquisa.

Exemplo: as amostras por quotas são muito usadas em pesquisa de opinião eleitoral e pesquisas de mercado.

Para a realização de amostragem por quotas é necessário estabelecer variáveis de controle.

À medida que o número de características e categorias sob controle for sendo elevado, pode-se chegar a uma situação tal que o método não poderá ser empregado, ou pela não disponibilidade das proporções na população, ou pelo exagerado número de células a que se chega, o que permitirá um número elevado de elementos no total da amostra.

IV – Amostragem Bola de Neve

Definição: essa técnica é uma forma de amostra não probabilística utilizada em pesquisas sociais onde os participantes iniciais de um estudo indicam novos participantes que por sua vez indicam novos participantes e assim sucessivamente, até que seja alcançado o objetivo proposto (o "ponto de saturação"). O "ponto de saturação" é atingido quando os novos entrevistados passam a repetir os conteúdos já obtidos em entrevistas anteriores, sem acrescentar novas informações relevantes à pesquisa. É uma técnica de amostragem que utiliza cadeias de referência, uma espécie de rede.

Exemplo: análise de comportamento de crianças após passar por uma situação traumática.

Alcance do "ponto de saturação".

LISTA DE EXERCÍCIOS nº 05

Questão 1. Crie uma situação para amostragem por conveniência.

Questão 2. Crie uma situação para amostragem por julgamento.

Questão 3. Crie uma situação para amostragem por quotas.

Questão 4. Crie uma situação para amostragem bola de neve.

Parte I – Estimação

Distribuição e erro amostral:

O problema da inferência consiste em fazer uma afirmação sobre os parâmetros da população através da amostra.

Suponha que a afirmação deva ser feita sobre um parâmetro Θ da população (média, variância, proporção, dentre outros).

Será usada uma AAS de n elementos sorteados dessa população.

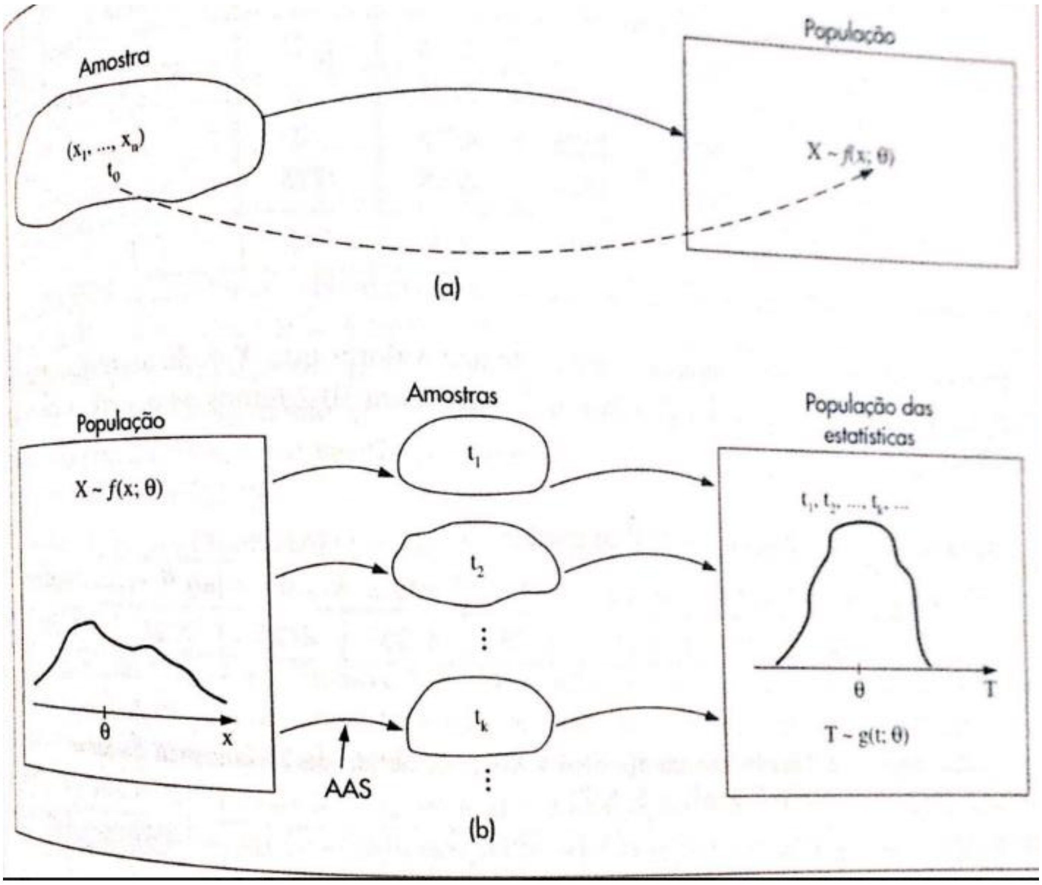
A decisão será baseada na estatística T que é uma função da amostra (X_1, X_2, ..., X_n), ou seja, $T=f(X_1, X_2, ..., X_n)$.

Após a coleta da amostra será observado um valor qualquer de $T(t_0)$ e, baseado nesse valor é que se faz a afirmação sobre Θ.

A resposta sobre essa afirmação seria mais rica se soubéssemos o que aconteceria com a estatística T, quando retiramos todas as amostras de uma população conhecida segundo o plano amostral adotado.

Qual a distribuição de T quando (X_1, X_2, ..., X_n) assume todos os valores possíveis? Essa distribuição é chamada de *distribuição amostral da estatística T* e desempenha importante papel na inferência.

O principal objetivo é identificar um modelo que explique a distribuição amostral de T e, essa irá depender da distribuição de X e do plano amostral adotado.



Exemplo: seja a população {1, 3, 5, 5, 7}. Retire todas as amostras de tamanho 2 com reposição. A distribuição conjunta da variável bidimensional (X_1, X_2) é:

X_2 \ X_1	1	3	5	7	Total
1	1/25	1/25	2/25	1/25	1/5
3	1/25	1/25	2/25	1/25	1/5
5	2/25	2/25	4/25	2/25	2/5
7	1/25	1/25	2/25	1/25	1/5
Total	1/5	1/5	2/5	1/5	1

Qual a distribuição da estatística média amostral?

Média amostral = 3

equivale à A= {(1,5); (3,3); (5,1)}.

P(A) = 1/5.

De forma análoga,

$\bar{x}$	1	2	3	4	5	6	7	Total
$P(\bar{X}=\bar{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25	1,00

Distribuição amostral da média: considere uma população identificada pela variável X, cujos parâmetros: média populacional μ = E(X) e variância populacional σ^2 = Var(X) são supostos conhecidos.

Retire todas as AAS de tamanho n dessa população e, para cada uma calcular a média amostral.

Em seguida vamos considerar a distribuição amostral e suas propriedades.

Exemplo: considere o exemplo anterior. Sabe-se que μ=4,2 e σ^2=4,16. A distribuição amostral da média foi dada anteriormente, da qual obtemos:

$$E(\bar{X}) = \sum_{i}^{k} \bar{x}_i\, p_i = 1x\frac{1}{25} + 2x\frac{2}{25} + 3x\frac{5}{25} + 4x\frac{6}{25} + 5x\frac{6}{25} + 6x\frac{4}{25} + 7x\frac{1}{25} = 4{,}2$$

De modo análogo,

$$Var(\bar{X}) = 2{,}08$$

O que vocês podem observar?

Teorema: seja X uma v.a. com média μ e variância σ^2, e seja (X_1, …, X_n) uma AAS de X. Então:

$$E(\bar{X}) = \mu e Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$E(\bar{X}) = \frac{1}{n}\big(E(X_1)+\ldots+E(X_n)\big) = \frac{1}{n}xn\mu = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2}\big(Var(X_1)+\ldots+Var(X_n)\big) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Determinamos a média e a variância da distribuição amostral da média.

Qual a forma da distribuição dessa estatística?

Para amostras aleatórias simples (X_1, …, X_n) retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média aproxima-se para n grande, de uma distribuição normal com média e variância dada anteriormente (TLC).

Exemplo: considere uma máquina que enche pacotes cujos pesos seguem uma distribuição N(500, 100). Colhendo uma amostra de n=100 pacotes e pesando-os, pelo que foi dito acima, média terá uma distribuição normal com média 500 e variância 1. Logo, se a máquina estiver regulada, a probabilidade de encontrarmos a média de 100 pacotes diferindo de 500g de menos de 2 gramas será?

$$P(|\bar{X} - 500| < 2) = P(498 < \bar{X} < 502) = 95\%$$

Ou seja, dificilmente 100 pacotes terão uma média fora do intervalo (498,502).

Corolário 1: se (X_1, …, X_n) for uma AAS da população X então:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt[2]{n}}} éaproximadamenteN(0,1)$$

Chamemos de ***e*** a v.a. que mede a diferença entre a média amostral e a média populacional. Chamamos de erro amostral da média.

Corolário 2: a distribuição de ***e*** aproxima-se de uma distribuição normal com média 0 e variância σ^2/n (vide distribuição Z do corolário 1).

Distribuição amostral da proporção: considere uma população em que a proporção de elementos com certa característica é p. Seja X uma v.a. da seguinte forma:

X=1, se o indivíduo for portador da característica

X=0, se o indivíduo não for portador da característica

Logo,

E(X)=p

Var(X)=p(1-p)

Retirada uma AAS dessa população e indicando Y_n como o total de indivíduos portadores da característica na amostra,

$$Y_n \sim b(n,p)$$

Vamos definir a proporção de portadores da característica na amostra:

$$\hat{p} = \frac{Y_n}{n}$$

Logo,

$$P(Y_n = k) = P(Y_n/n = k/n) = P(p\hat{}= k/n)$$

Ou seja, a distribuição amostral da proporção amostral é obtida da distribuição de Y_n = n x média amostral.

Sendo E(X) = p e Var(X) = [p(1-p)].

Pelo TLC:

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$Y_n \sim N\big(np, np(1-p)\big)$$

Logo,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Exemplo: suponha que p=30% dos estudantes de uma escola sejam mulheres. Coletamos uma AAS de n=10 estudantes e calculamos a proporção de mulheres na amostra. Qual a probabilidade de que o estimador desvie do parâmetro em menos de 0,01?

$$P(|\hat{p} - p| < 0{,}01) = P(-0{,}01 < \hat{p} - p < 0{,}01)$$

$$\hat{p} - p \sim N\left(0, \frac{p(1-p)}{n}\right)$$

Substituindo p = 0,3.

$$P\left(\frac{-0{,}01}{\sqrt[2]{0{,}021}} < \hat{p} - p < \frac{0{,}01}{\sqrt[2]{0{,}021}}\right) = 0{,}056$$

Assim, como estudamos a distribuição amostral da média e da proporção é possível estudar de qualquer estatística T = f(X_1,..., X_n).

É válido ressaltar que a média amostral e a variância amostral são estimadores não-viesados.

Centro Universitário do Instituto de Educação Superior de Brasília – IESB
Graduação em Ciência de Dados e Inteligência Artificial
HMDC065 - Amostragem Aplicada

LISTA DE EXERCÍCIOS nº 06

Questão 1. Uma v.a. X tem distribuição normal, com média 100 e desvio padrão 10:

a) Qual a P(90<X<110)?

b) Se X for a média de uma amostra de 16 elementos retirados dessa população, calcule P(90< X <110) ?

Questão 2. Um auditor de banco declara que as contas de cartões de crédito são normalmente distribuídas, com uma média de R$2.870,00 e um desvio padrão de R$900,00.

a) Qual é a probabilidade de que um titular de cartão de crédito aleatoriamente selecionado tenha uma conta menor que R$2.500,00?

b) Você selecionou 25 titulares de cartões de crédito de forma aleatória. Qual é a chance de que a média da conta deles seja menor que R$2.500,00?

Questão 3. Um procedimento de controle de qualidade foi planejado para garantir um máximo de 10% de itens defeituosos na produção. A cada 6 horas sorteia-se uma amostra de 20 peças e, havendo mais de 15% de defeituosas, encerra-se a produção para verificação do processo. Qual a probabilidade de uma parada desnecessária?

Questão 4. Supondo que a produção do exemplo anterior esteja sob controle, isto é, p = 10%. E que os itens sejam vendidos em caixas com 100 unidades. Qual a probabilidade de que uma caixa:

a) tenha mais do que 10% de defeituosos?

b) não tenha itens defeituosos?

Questão 5. Um pesquisador deseja estimar a proporção de ratos nos quais se desenvolve um certo tipo de tumor quando submetidos a radiação. Ele deseja que sua estimativa não se desvie da proporção verdadeira por mais de 0,02 com uma probabilidade de pelo menos 90%.

a) Quantos animais ele precisa examinar para satisfazer essa exigência?

b) Como seria possível diminuir o tamanho da amostra utilizando a informação adicional de que em geral esse tipo de radiação não afeta mais que 20% dos ratos?

Parte II – Estimação

Estimação de parâmetros:

Lembrete:

- Parâmetros são funções de valores populacionais;
- Estatísticas são funções de valores amostrais;
- Estimadores pontuais.

Exemplo: uma amostra de n = 500 pessoas de uma cidade é escolhida, e a cada pessoa da amostra é feita uma pergunta a respeito de um problema municipal, para o qual foi apresentada uma solução pela prefeitura. A resposta à pergunta poderá ser SIM (favorável à solução) ou NÃO (contrária à solução). Deseja-se estimar a proporção de pessoas na cidade favoráveis à solução apresentada. Se 300 pessoas responderam SIM à pergunta, então uma estimativa natural para essa proporção seria 300/500 = 60%. Nossa resposta é baseada na suposição de que a amostra é representativa da população. Outra amostra poderia levar a outra estimativa.

Defina as v.a. $X_1,\ldots, X_n$, tais que:

X_i = 1, se a i-ésima pessoa na amostra responder SIM

X_i = 0, se a i-ésima pessoa na amostra responder NÃO

Seja p = P(sucesso) → resposta é igual a SIM

$$Y_n = \sum_{i=1}^{n} X_i \, éaproximadamentebinomial(n; p)$$

Estimador de p:

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{k}{n}$$

Observe que a proporção amostral é uma v.a. e k/n é um valor da v.a., logo é uma estimativa.

Lembrete:

$$E(\hat{p}) = p \; e \; Var(\hat{p}) = \frac{p(1-p)}{n}$$

Como interpretar o valor esperado?

O que acontece se n→ ∞?

Centro Universitário do Instituto de Educação Superior de Brasília – IESB
Graduação em Ciência de Dados e Inteligência Artificial
HMDC065 - Amostragem Aplicada

Propriedades de estimadores:
Problema da estimação é determinar uma função $T=g(X_1,X_2,\ldots,X_n)$ que seja próxima do parâmetro, seguindo alguns critérios:
1) O estimador T é não-viesado para Θ se E(T)=Θ, para todo Θ

Exemplo anterior: E(proporção amostral)=p

O que é estimativa? É o valor assumido pelo estimador em uma particular amostra.
Exemplo anterior: 60%

A média amostral é um estimador viesado ou não-viesado de μ?
(vide distribuição amostral)

Agora considere uma população com N elementos e a variância populacional

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

Possível estimador para variância, baseado numa AAS de tamanho n é:

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

O estimador é viesado ou não-viesado?

2) Uma sequência {Tn} de estimadores de Θ é consistente se
lim n→∞ $E(T_n)$= Θ
lim n→∞ $Var(T_n)$= Θ

3) Se T e To são dois estimadores não-viesados de um parâmetro Θ é eficiente se Var(T) < Var(T°)

4)Erro amostral: e = T – Θ

Métodos para determinar os estimadores:
- Estimadores de momentos;
- Estimadores de mínimos quadrados;
- Estimadores de máxima verossimilhança.

Níveis de confiança:

Vimos os estimadores pontuais, isto é, especificam um único valor para o estimador. Esse procedimento não permite julgar qual a possível magnitude do erro que pode ser cometido.

Surge então os intervalos de confiança que são baseados na distribuição amostral do estimador pontual.

Suponha que queiramos estimar a média de uma população qualquer, e para tanto usamos a média amostral proveniente de uma amostra de tamanho n.

$$\bar{X} \text{ é } N\left(\mu; \frac{\sigma^2}{n}\right)$$

$$e = (\bar{X} - \mu) \text{ é } N\left(0; \frac{\sigma^2}{n}\right)$$

Qual a probabilidade de cometer erros de determinadas magnitudes?

$$\bar{X} \text{ é } N\left(\mu; \frac{\sigma^2}{n}\right)$$

$$P\left(|\bar{X} - \mu| < 1{,}96 \frac{\sigma}{\sqrt[2]{n}}\right) = 0{,}95$$

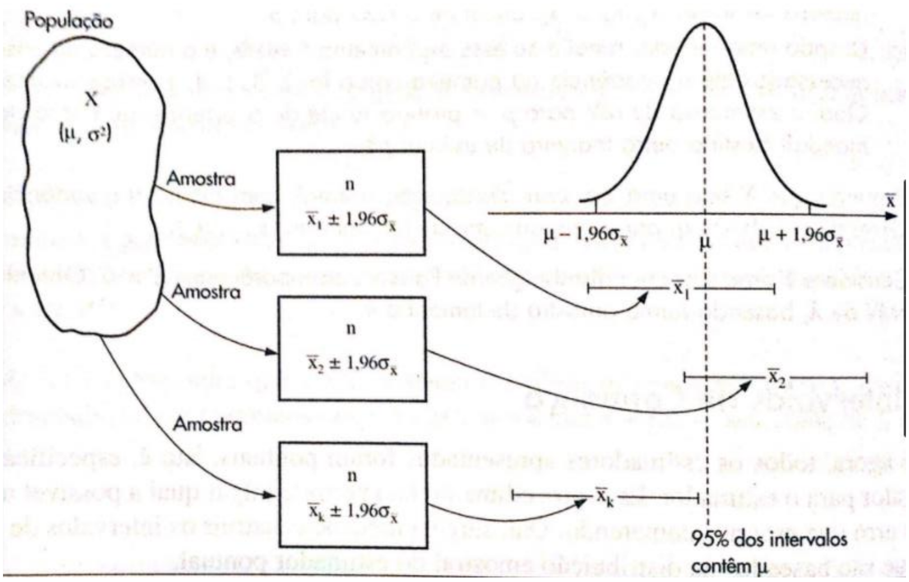
$$P\left(-1{,}96 \frac{\sigma}{\sqrt[2]{n}} < \bar{X} - \mu < 1{,}96 \frac{\sigma}{\sqrt[2]{n}}\right) = 0{,}95$$

$$P\left(\bar{X} - 1{,}96 \frac{\sigma}{\sqrt[2]{n}} < \mu < \bar{X} + 1{,}96 \frac{\sigma}{\sqrt[2]{n}}\right) = 0{,}95$$

Observe que μ é um parâmetro e a interpretação é da seguinte forma: se pudéssemos construir uma quantidade grande de intervalos (aleatórios) da forma

$$\bar{X} - 1{,}96 \frac{\sigma}{\sqrt[2]{n}}; \bar{X} + 1{,}96 \frac{\sigma}{\sqrt[2]{n}}$$

Todos baseados em amostras de tamanho n, 95% (coeficiente de confiança) deles conteriam o parâmetro μ.



Exemplo: uma máquina enche pacotes de café com uma variância igual a 100 g2. Ela estava regulada para encher os pacotes com 500g, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou uma média igual a 485g. Como construir um intervalo de confiança com 95% de confiança para μ.

Suponha que queiramos estimar a proporção de uma população qualquer, e para tanto usamos a proporção amostral proveniente de uma amostra de tamanho n.

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$e = (\hat{p} - p) \sim N\left(0, \frac{p(1-p)}{n}\right)$$

$$P\left(\hat{p} - 1{,}96\frac{\sqrt[2]{p(1-p)}}{\sqrt[2]{n}} < p < \hat{p} + 1{,}96\frac{\sqrt[2]{p(1-p)}}{\sqrt[2]{n}}\right) = 0{,}95$$

Exemplo: suponha que em n=400 provas obtemos k=80 sucessos. Construa um intervalo de confiança para p com coeficiente de confiança de 90%.

Para pensar e resolver:

1) Qual deve ser o tamanho da amostra para que o intervalo com 99,5% de confiança para a média populacional tenha uma semi amplitude não superior a 1,5? Sabe-se que a variância populacional é de 23.

2) Sabe-se que a proporção de animais contaminados com uma determinada doença não é superior a 10%. Qual deve o tamanho da amostra para que a semi amplitude do intervalo com 92% de confiança para a fração populacional não seja superior a 2%?

LISTA DE EXERCÍCIOS nº 07

Elaborem um resumo dos textos AMOSTRA E AMOSTRAGEM e INSPEÇÃO POR AMOSTRAGEM com enfoque nas aplicações práticas.

LISTA DE EXERCÍCIOS nº 08

DIVERSOS

Questão 1. A tabela abaixo apresenta os dados de um levantamento do palmiteiro juçara (Euterpe aedulis, Arecaceae) na região do Vale do Ribeira, estado de São Paulo. O exemplo é composto de 34 arvoredos (1 600 m2), sendo que em cada um dos quatro estratos se realizou uma amostragem aleatória simples. A área basal e os DAP médio e médio quadrático se referem apenas às plantas do palmiteiro juçara. Assume que o tamanho da população é N = 20000.

Estrato	*Parcela*	*Número de Palmiteiros* (ha^{-1})	*Área Basal* ($m^2\ ha^{-1}$)	*DAP*	
				Médio (cm)	*Méd. Quad.* (cm)
I	1004	631.25	8.24	5.91	12.89
I	1006	1025.00	10.32	9.59	11.32
I	1007	1006.25	9.47	10.49	10.95
I	1018	550.00	9.62	6.34	14.92
I	2003	356.25	1.96	7.29	8.38
I	2007	606.25	5.39	9.41	10.64
I	2012	225.00	0.79	6.31	6.67
I	2013	168.75	0.55	6.27	6.42
I	2017	343.75	5.34	13.62	14.07
I	3009	281.25	1.48	7.94	8.19
I	4009	56.25	0.13	5.47	5.40
I	4010	143.75	0.96	3.93	9.22
I	4011	112.50	0.35	6.57	6.26
I	4014	18.75	0.04	5.33	5.36
I	4016	31.25	0.08	5.62	5.69
II	1002	181.25	0.74	6.80	7.19
II	1003	87.50	0.34	7.55	7.03
II	1028	137.50	0.76	8.02	8.41
II	1031	125.00	1.31	2.95	11.57
II	2020	387.50	2.72	9.40	9.46
II	4002	200.00	0.66	6.21	6.49
III	1025	112.50	0.87	4.30	9.91
III	1026	200.00	1.18	5.72	8.67
III	2037	100.00	0.49	7.27	7.89
III	3004	131.25	0.52	6.56	7.10
III	3039	6.25	0.02	5.70	5.70
III	3063	68.75	0.24	6.25	6.64
III	4017	37.50	0.10	5.61	5.70
III	4018	18.75	0.04	5.36	5.37
IV	1029	125.00	1.05	3.43	10.37
IV	2026	93.75	0.29	6.14	6.28
IV	2029	37.50	0.11	5.99	6.05
IV	2035	50.00	0.14	5.86	5.91
IV	3042	218.75	0.98	7.25	7.54

Encontre o intervalo de confiança de 95% dos atributos apresentados, bem como o tamanho de amostra necessário para erro amostral aceitável de 5%.

Questão 2. Considere uma população de empresas de prestação de serviços que pode ser dividida em 3 estratos quanto ao número de trabalhadores que emprega: pequenas – 50 ou menos trabalhadores; médias – entre 51 e 100; grandes – mais de 101 trabalhadores. Supondo que a população de interesse é constituída por 1800 empresas, destas 45% são pequenas, 35% são médias e 20% grandes. Quantas empresas devem ser selecionadas de cada estrato para estimar a proporção? Considere erro de 5%.

Questão 3. Uma população é formada de N = 35 árvores de uma determinada espécie, pertencentes a um parque ecológico, que possuem os seguintes diâmetros a altura do peito em cm (DAP):
25, 20, 35, 21, 22, 22, 24, 25, 30, 38, 24, 20, 21, 25, 20, 15, 25, 23, 20, 24, 28, 24, 24, 22, 28, 26, 23, 19, 22, 27, 25, 23, 28, 27, 42.

a) Com o objetivo de estimar o DAP (diâmetro a altura do peito) médio, como você extrairia uma amostra simples ao acaso, de tamanho n = 10 desta população?

b) Quantas amostras simples ao acaso são possíveis retirar dessa população? Considere "sem reposição".

Questão 4. Verifique se o estimador S^2 é não-viesado.

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