

Universidade Federal de São Paulo

Processos de amostragem



Professor Julio Cezar



REVISÃO

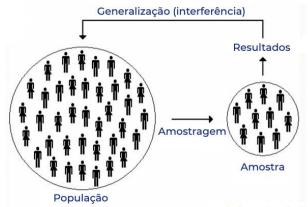
Em aulas anteriores trabalhamos com a ideia de probabilidade considerando uma população como um todo, ou seja, tínhamos um razoável conhecimento sobre a população e estávamos calculando a probabilidade de ocorrência de determinados eventos, normalmente representados por variáveis aleatórias. Nesse momento a gente começa a chegar em um novo campo da estatística, onde o foco é, a partir de uma determinada amostra compreender quais são as principais características de uma população.



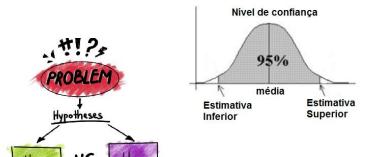
REVISÃO

Nessa aula vamos passar do campo da probabilidade para o campo da estatística.





Intervalo de Confiança



Análise da variância (ANOVA)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
Regressão 1		SQReg	SQReg = QMReg	$\frac{QMReg}{S_e^2}$
Resíduo n — 2		SQRes	$\frac{SQRes}{n-2} = S_e^2$	
Total n-1		$\frac{SQTot}{n-1} = S^2$		



A Estatística é a parte da matemática aplicada que fornece métodos para:

- Coleta;
- Organização;
- Descrição (apresentação);
- Análise;
- Interpretação de dados.

34.1% 34.1% -30 -20 -10 0 10 20 30 5AS 55 70 85 100 115 12 145

Definição da American Statistical Association

A estatística é a aplicação científica de princípios matemáticos para a coleta, análise e apresentação de dados numéricos.

Outros significados

Estatística é um conjunto de métodos usados para se analisar dados. Pode ser aplicada em diferentes áreas do conhecimento:

- Ciências Biológicas;
- Ciências Médicas;
- Ciências Agrárias;
- Marketing;
- Mercado Financeiro...



Estatística: mesma raiz latina da palavra Estado (organização política): status.

Originalmente, as estatísticas eram obtidas para as finalidades relacionadas com o Estado (com objetivos militares, tributários, recenseamentos, entre outros).

Antiguidade: os povos já registravam o número de habitantes, nascimentos, óbitos.

China (2238 AC): censo de nascimento de meninos e meninas.

Bíblia: Referências do censo dos Hebreus.

Egípcios: Devido às inundações do Nilo, se efetuavam anualmente trabalhos cadastrais para a repartição de terras férteis no Egito.....

Século 17 e 18: estudos demográficos.





Por que usar estatística?

Por que a natureza apresenta **VARIABILIDADE**:

- Variações de indivíduo para indivíduo (características dos objetos de estudo);
- Variações sobre um mesmo indivíduo (objeto de estudo).

Exemplos:

- 1. Tome todos os alunos desta sala com mesma idade: existe variabilidade de altura, de peso,...
- 2. condições não são totalmente controladas: planta-se num mesmo solo sementes do mesmo tipo de uma cultura, aplica-se a mesma quantidade de água, sol,...e mesmo assim existe variabilidade no tamanho das plantas, no tempo de germinação...

AMOSTRAGEM: OBJETIVOS ESPECÍFICOS

- Mostrar os conceitos de amostragem e amostra;
- Mostrar algumas técnicas de amostragem probabilística e não probabilística.







População (N)

Conjunto de todos os elementos relativos a um determinado fenômeno que possuem pelo menos uma característica em comum, a população é o conjunto Universo, podendo ser finita (apresenta um número limitado de observações, que é passível de contagem) ou infinita (apresenta um número ilimitado de observações que é impossível de contar e geralmente esta associada a processos).

Exemplos:

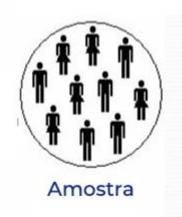
- Todos os alunos da UNIFESP;
- Todos os computadores produzidos de uma determinada fábrica.



Amostra (n)

É um subconjunto retirado da população, que se supõe ser representativo de todas as características da mesma, sobre o qual será feito o estudo, com o objetivo de serem tiradas conclusões válidas sobre a população.







Obs: À medida que o tamanho de uma amostra for crescendo, as informações relativas à amostra vão se tornando cada vez mais verdadeiras.

Por que considerar uma amostra ao invés da população?



Diversos fatores justificam o fato de trabalhar-se com amostras, no lugar de estudar a respectiva população, entre os quais, destacam-se: Custo, Velocidade e Praticabilidade (às vezes, a dimensão da população torna as pesquisas impraticáveis).



Pesquisa estatística

É qualquer informação retirada de uma população ou amostra, podendo ser através de Censo ou Amostragem.

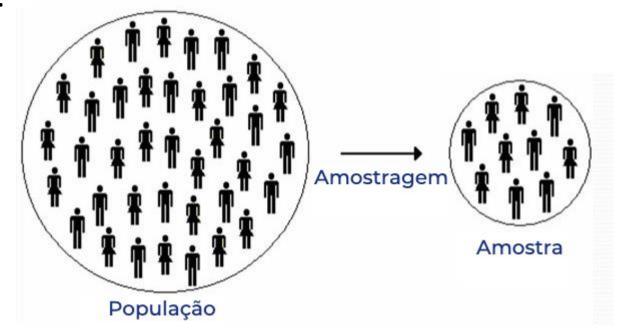
Censo

O censo consiste em obter os dados de uma ou mais variáveis em **todos** os objetos mediante ao **recenseamento** (processo de se coletar dados de uma população).

Exemplo: Censo IBGE.

Amostragem

É o procedimento pelo qual um subconjunto de uma população é escolhido com o intuito de obter informações relacionadas com um fenômeno, com a intenção de inferir o comportamento da população. Resumindo, amostragem é o processo de escolha da amostra.





O que vem a ser dados brutos?



Dados brutos são uma sequência de valores, não organizados, obtidos diretamente da observação de um fenômeno.

Variáveis

É aquilo que se deseja observar para se tirar algum tipo de conclusão.

- Geralmente as variáveis são selecionadas por amostragem.
- São representadas por letras maiúsculas do alfabeto, tais como X,Y, Z... que pode assumir qualquer valor do conjunto de dados.

 UNIFESI

Parâmetro

Característica que descreve a população.

Obs:

Parâmetro (População)		VS	Estatística (amostra)	
Valor médio	μ		Média	$\overline{\mathbf{X}}$
Desvio padrão	σ		Desvio padrão	S
Proporção	p		Proporção	$\widehat{m{p}}$
Correlação	ρ		Correlação	r

Se você está falando de um **parâmetro**, está falando de toda a população. Uma estatística descreve apenas uma amostra da população, ou seja, a Teoria Estatística define uma **estatística** como uma função de uma amostra em que a função por si mesma é independente da distribuição que gerou a amostra.

Estimativa

É o valor numérico associado ao parâmetro, obtido por meio da amostra, ou seja, é o valor obtido pelo estimador numa amostra.

Obs: Oh professor, estimativa é a mesma coisa que estimador?



Estimador é um parâmetro amostral que é usado como aproximação para o valor de um parâmetro populacional. Por exemplo, o desvio padrão amostral s é um estimador para o desvio padrão populacional s. **Estimativa** é o valor do nosso estimador, é o número propriamente dito. Por exemplo, calculamos que a variância amostral da nossa

amostra vale 45 isso quer dizer que 45 é a estimativa para o valor da

variância populacional σ^2 .

Definição do problema

Consiste na:

- Formulação correta do problema;
- Examinar outros levantamentos realizados no mesmo campo (revisão da literatura);
- Saber exatamente o que se pretende pesquisar de definindo o problema corretamente (variáveis, população, hipóteses, etc.)



Planejamento

Determinar o procedimento necessário para resolver o problema:

- Como levantar informações;
- Tipos de levantamentos:
- 1. Por Censo (completo);
- 2. Por Amostragem (parcial);
- Cronograma, Custos, etc.



Coleta de dados

Consiste na obtenção dos dados referentes ao trabalho que desejamos fazer:

- A coleta pode ser:
- 1. Direta diretamente da fonte ou
- 2. Indireta feita através de outras fontes.
- Os dados podem ser obtidos pela própria pessoa (primários) ou se baseia no registro de terceiros (secundários).



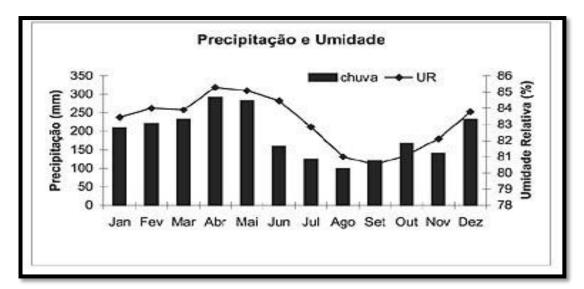
Coleta de dados

Pode ser feito o **levantamento** da observação do fenômeno na natureza.

Exemplos:

Dados climáticos:

- Precipitação, velocidade do vento, temperatura, umidade....
- Efeito da poluição sobre a taxa de doenças respiratórias em um determinado período.





Coleta de dados

Experimento:

Nestes casos as **observações** são geradas e feitas comumente sob condições controladas pelo pesquisador, e os fatos, eventos ou fenômenos são forçados a sofrer variações sistemáticas mediante aplicação de tratamentos. Pretende-se determinar relações de causa e efeito.

Exemplos:

- produção de milho mediante a aplicação de diferentes doses de nitrogênio;
- mortalidade de insetos mediante aplicação de diferentes inseticidas;







Apuração dos dados

Consiste em resumir os dados, através de uma contagem e agrupamento. É um trabalho de coordenação e de tabulação.

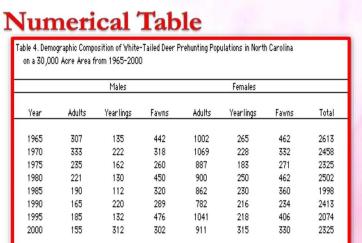




Apresentação dos dados

É a fase em que vamos mostrar os resultados obtidos na coleta e na organização. Esta apresentação pode ser:

- Tabular (apresentação numérica);
- Gráfica (apresentação geométrica).







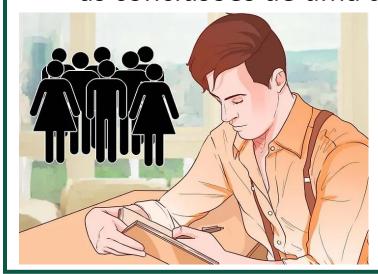
Análise e interpretação dos dados

É a fase mais importante e também a mais delicada. Tira conclusões que auxiliam o pesquisador a resolver seu problema.



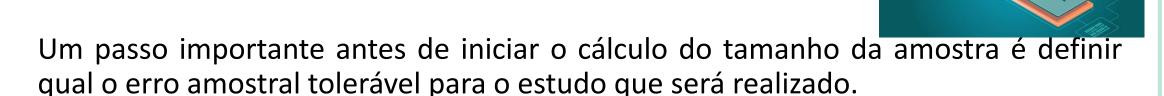


- Na realização de qualquer estudo quase nunca é possível examinar todos os elementos da população de interesse. Temos usualmente que trabalhar com uma amostra da população.
- A inferência estatística nos dá elementos para generalizar, de maneira segura, as conclusões obtidas da amostra para a população.
- Os erros de coleta e manuseio de um grande número de dados são maiores do que as imprecisões a que estamos sujeitos quando generalizamos, via inferência, as conclusões de uma amostra bem selecionada.





Erro amostral tolerável



Erro amostral tolerável: Margem de erro aceitável em um estudo estatístico.

Exemplo:

Quando o apresentador do telejornal, em ano de eleições, anuncia: "O candidato Fulano de Tal tem 42% das intenções de voto, 2 para mais, 2 para menos." O que ele fala de "2 para mais, 2 para menos", ele se refere ao erro amostral tolerável para aquela pesquisa de intenções de voto.



Tamanho da

e Erro Amostra

Erro amostral tolerável

É razoável pensar que, quanto menor o erro amostral toleráve será o tamanho da amostra necessário para obtê-lo. Isso fica claro ao ver a formula para a obtenção da primeira estimativa do tamanho da amostra.

$$n_0 = \frac{1}{E_0^2}$$

 E_0 : é o erro amostral tolerável;

 n_0 : primeira estimativa do tamanho da amostra.

Se o tamanho da população, N, for conhecido podemos corrigir a primeira estimativa.

$$n = \frac{N \times n_0}{N + n_0}$$



Tamanho da

e Erro Amostra

Tamanho da Amostra

Em se tratando de amostra, a preocupação central é que ela seja representativa. Assim que decidimos obter informações por meio de um levantamento amostral, temos imediatamente dois problemas:

- Definir cuidadosamente a população de interesse;
- Selecionar a característica que iremos pesquisar.

São várias as formulas que permitem calcular o tamanho de uma amostra. A escolha depende do fenômeno em estudo ou dos parâmetros disponíveis.

É conveniente planejar o tamanho da amostra para que se possa ter amostras grandes o suficiente para detectar diferenças importantes.



Tamanho da Amostra

Procedimentos para determinar o tamanho da amostra

- Analisar o questionário, ou roteiro da entrevista e escolher uma variável que julgue mais importante para o estudo. Se possível mais do que uma;
- Verificar o nível de mensuração da variável: nominal, ordinal ou intervalar;
- Considerar o tamanho da população: infinita ou finita;

Nota: Se $\frac{n}{N} \ge 0.05$ população finita; se $\frac{n}{N} < 0.05$ população infinita

Se a variável escolhida for:



Tamanho da Amostra

- Intervalar e a população considerada infinita, pode-se determinar o tamanho da amostra pela fórmula

$$n = \left(\frac{Z \cdot \sigma}{d}\right)^2$$

nde: Z = abscissa da curva normal padrão, fixado um nível de confiança $(1 - \alpha)$

$$Z = 1,65 \rightarrow (1 - \alpha) = 90\%$$

$$Z = 1.96 \rightarrow (1 - \alpha) = 95\%$$

$$Z = 2.0 \rightarrow (1 - \alpha) = 95.5\%$$

$$Z = 2,57 \rightarrow (1 - \alpha) = 99\%$$

Geralmente usa-se Z = 2.



Tamanho da Amostra

σ: desvio padrão da população, expresso na unidade variável, onde poderá ser determinado por:

- Especificações Técnicas
- Resgatar o valor de estudos semelhantes
- Fazer conjeturas sobre possíveis valores

d: erro amostral, expresso na unidade da variável. O erro amostral é a máxima diferença que o investigador admite suportar entre μ e \bar{x} , isto é: $|\mu - \bar{x}| < d$.



Tamanho da Amostra

- Intervalar e a população considerada finita, pode-se determinar o tamanho da amostra pela fórmula:

$$n = \frac{Z^{2}.\sigma^{2}.N}{d^{2}(N-1) + Z^{2}.\sigma^{2}}$$

em que:

Z: abscissa da normal padrão;

 σ^2 : variância populacional;

N: tamanho da população;

d: erro amostral.



Tamanho da Amostra

- Nominal ou ordinal, e a população considerada infinita, pode-se determinar o tamanho da amostra pela fórmula:

$$n = \frac{Z^2 \cdot \hat{p} \cdot \hat{q}}{d^2}$$

em que:

Z: abscissa da normal padrão;

p̂: estimativa da verdadeira proporção de um dos níveis da variável escolhida. Por exemplo, se a variável escolhida for parte da empresa, p̂ poderá ser a estimativa da verdadeira proporção de grandes empresas do setor que está sendo estudado;

 \hat{p} : será expresso em decimais ($\hat{p}=30\%$ -> $\hat{p}=0.30$).

Tamanho da Amostra

$$\hat{q} = 1 - \hat{p}$$
;

d: erro amostral, expresso em decimais. O erro amostral neste caso será a máxima diferença que o investigador admite suportar entre \hat{p} e π , isto é: $|\pi - \hat{p}| < d$, em que π é a verdadeira proporção (frequência relativa do evento a ser calculado a partir da amostra).



Tamanho da Amostra

- Nominal ou ordinal, e a população considerada finita, pode-se determinar o tamanho da amostra pela fórmula:

$$n = \frac{Z^2.\,\hat{p}.\,\hat{q}.\,N}{d^2(N-1) + Z^2.\,\hat{p}.\,\hat{q}}$$

em que:

Z: abscissa da normal padrão;

p: estimativa da proporção;

$$\hat{\mathbf{q}} = 1 - \hat{\mathbf{p}};$$

d: erro amostral.



Relação entre o erro e o tamanho da amostra

Erro amostral (ou margem de erro), nível de confiança e tamanho da amostra sempre estão relacionadas entre si. Modificar qualquer uma destas 3 medidas, alterará os restantes:

- 1. Reduzir a margem de erro obriga a aumentar o tamanho da amostra.
- 2. Aumentar o nível de confiança obriga a aumentar o tamanho da amostra.
- 3. Se eu aumentar o tamanho da minha amostra, posso reduzir a margem de erro ou incrementar o nível de confiança.



Há duas grandes divisões no processo de amostragem

- Probabilística: É aquela em que todos os elementos da população têm probabilidade conhecida, diferente de zero, de ser incluídos na amostra, o que garante a representatividade da amostra em relação à população.
- Não probabilística: A escolha dos elementos da amostra é feita de forma não aleatória.



Probabilística



Amostragem aleatória simples

Objetivo - Obter uma amostra representativa quando os elementos da população são todos homogêneos. Neste processo de amostragem, todos os elementos da população têm a mesma probabilidade de serem amostrados.

Procedimento - Na prática, a amostragem aleatória simples pode ser realizada da seguinte forma:

- 1) Enumerar a população de 1 a N;
- 2) Sortear por meio de um dispositivo aleatório qualquer n números dessa sequência;
- 3) Os elementos correspondentes aos números escolhidos formarão uma amostra de n elementos.



Probabilística

Amostragem aleatória simples

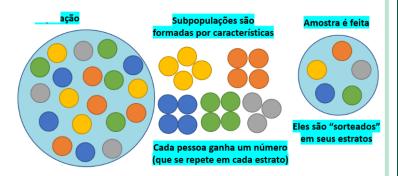
Exemplo:

Vamos obter uma amostra representativa de 10% dos valores, para obtermos a estatura média de noventa alunos de uma escola:

- Numeramos os alunos de 01 a 90;
- Sorteamos os números, de 01 a 90, um a um, nove números que formarão a amostra.



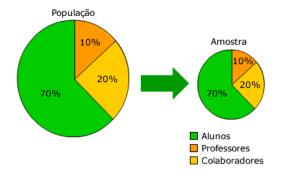
Probabilística



Amostragem estratificada

Objetivo - Melhorar a representatividade da amostra quando os elementos da população são heterogêneos, porém, podem ser agrupados em subpopulações (ESTRATOS) contendo elementos homogêneos.

Procedimento - A população é dividida em grupos ou estratos contendo elementos homogêneos e as amostras são retiradas separadamente de cada um desses grupos.





Probabilística

Amostragem estratificada

Exemplo:

Dada a população de 50.000 operários da indústria, selecionar uma amostra proporcional estratificada de 5% de operários para estimar seu salário médio. Usando a variável critério "cargo" para estratificar essa população, e considerando amostras de 5% de cada estrato obtido, chegamos a seguinte tabela:

Tabela: Operários de uma determinada indústria conforme o cargo.

Cargo	População	5%	Amostra
Chefes de seção	5000	5(5000)/100=250	250
Operários especializados	15000	5(15000)/100=750	750
Operários não especializados	30000	5(30000)/100=1500	1500
Total	50000	5(50000)/100=2500	2500

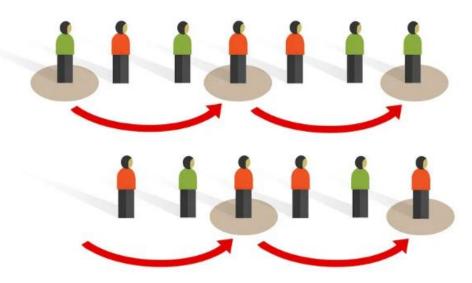
Probabilística

População Cada pessoa ganha um número Eles são sistematicamente acessados Amostra é feita 2 11 5 8 11 8

Amostragem sistemática

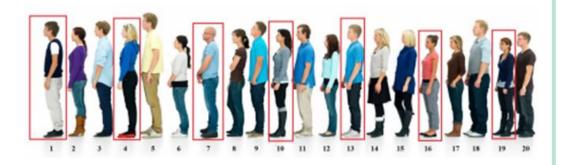
Objetivo - Aumentar a representatividade da amostra dando maior cobertura à população. É usada quando todos os elementos são homogêneos.

Systematic sampling





Probabilística



Amostragem sistemática

Procedimento - Consideremos uma população, com elementos ordenados, de tamanho N e dela tiramos uma amostra de tamanho n, por meio de uma amostragem sistemática, da seguinte maneira:

- 1) Definimos FS como fator de sistematização, dado por FS = N/n;
- 2) Sorteamos um número entre 1 e FS. Esse número é simbolizado por m, que será o primeiro elemento da amostra;
- 3) O segundo elemento da amostra é o de número FS + m;
- 4) O terceiro elemento da amostra é o de número 2FS + m;
- 5) O k-ésimo elemento da amostra é o número (k 1)FS + m.



Probabilística

Amostragem sistemática

Exemplo:

Uma rua contém 1000 prédios, dos quais desejamos obter uma amostra sistemática formada por 100 deles.

em que m será um número entre 1 a 10. Vamos supor que m = 7. Então temos:

1º elemento da amostra = (1 - 1)10 + 7 = 7, então 7º elemento da população;

100º elemento da amostra = (100 - 1)10 + 7 = 997, então 997° elemento da

população.

Probabilística



Amostragem por Conglomerados

Esta técnica é usada quando a identificação dos elementos da população é extremamente difícil, porém pode ser relativamente fácil dividir a população em conglomerados (subgrupos) heterogêneos representativos da população global. Assim,

- 1) Seleciona uma amostra aleatória simples dos conglomerados existentes;
- 2) Realizar o estudo sobre todos os elementos do conglomerado selecionado.



Probabilística

Amostragem por Conglomerados

Exemplo:

Estudar a população de uma cidade, dispondo apenas do mapa dos quarteirões da cidade. Como não temos a relação dos moradores da cidade, para realizar o estudo estatístico sobre a cidade, realizaremos os seguintes procedimentos:

- 1) Numerar os quarteirões de 1 a n (subgrupos heterogêneos, ou conglomerados);
- 2) Escrever os números de 1 a n em pedaços de papel e colocá-los em uma urna;
- 3) Retirar um pedaço de papel da urna e realizar o estudo sobre os elementos do conglomerado selecionado.

Não probabilística

Alguns dos usos habituais da amostragem não probabilística são os seguintes:

- Como etapa preliminar em projetos de pesquisa;
- Em projetos de pesquisa qualitativa;
- Em casos onde a população de trabalho não pode ser enumerada.



Não probabilística

Amostragem Acidental (amostra tomada a esmo)

Trata-se da formação de amostras por aqueles elementos que vão aparecendo. Este método é utilizado, geralmente, em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos.

Exemplo:

Pesquisas de opinião em praças públicas, ruas movimentadas de grandes cidades; Parar pessoas no supermercado e colher opiniões; Programa de TV ao vivo, registrando automaticamente opiniões contra ou favor de uma situação. etc.



Não probabilística

Amostragem Intencional

De acordo com determinado critério, é escolhido intencionalmente um grupo de elementos que comporão a amostra. O pesquisador se dirige intencionalmente a grupos de elementos dos quais deseja saber a opinião.

Exemplo:

Em uma pesquisa sobre preferência por determinado cosmético, o pesquisador entrevista os frequentadores de um grande salão de beleza.



Não probabilística

Amostragem por quotas

As amostras são obtidas dividindo a população por categorias ou estratos e selecionando um certo número (quota) de elementos de cada categoria de modo não aleatório, ou seja, entrevistam-se sujeitos de cada um dos subgrupos requeridos à medida que os mesmos vão sendo contatados, até se atingir a percentagem que foi previamente identificada para cada um deles.

Exemplo:

Em uma pesquisa de opinião eleitoral em uma cidade, poderíamos dividir a população de eleitores por sexo, nível de instrução, faixas de renda entre outros aspectos, e obter cotas proporcionais ao tamanho dos grupos (que poderia ser obtido através das informações do IBGE). Na prática UNIFESP pesquisas são realizadas utilizando amostragem por cotas.

Não probabilística

Observação: Nos estágios iniciais de um estudo, os pesquisadores podem coletar dados representativos de uma amostra formada usando o método de amostragem por quota (método Não probabilístico), que é muito semelhante à amostragem estratificada (método probabilístico), porém a principal diferença entre essas duas técnicas é que, na amostragem por quota, os elementos da amostra não são selecionados aleatoriamente de cada estrato, como é feito na amostragem estratificada.



Observação:

- A amostragem **probabilística** é a mais indicada quando se deseja fazer generalizações dos resultados.
- Há situações em que apenas técnicas não probabilísticas são possíveis.



CLASS FINISHED



