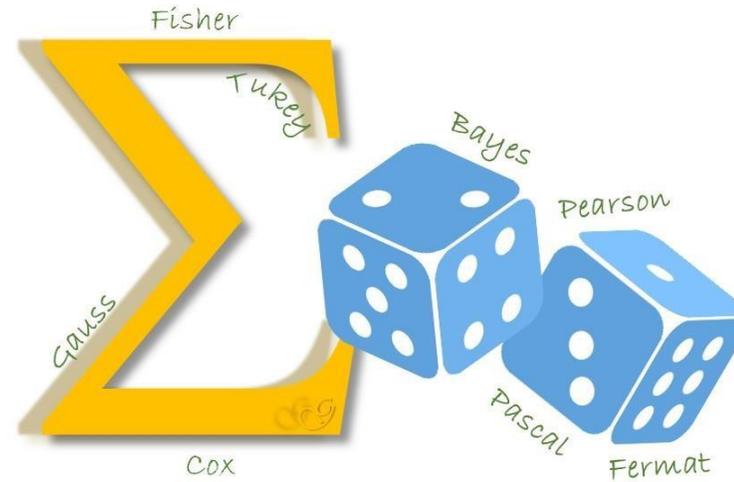


BIOESTATÍSTICA

# HEP - BIOESTATÍSTICA I

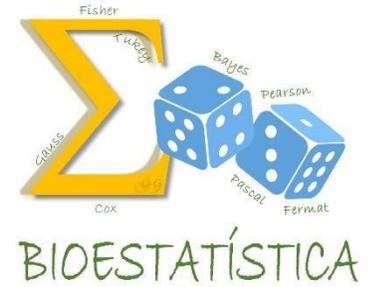
GLEICE M S CONCEIÇÃO  
MARIA DO ROSÁRIO D O LATORRE  
FSP USP



BIOESTATÍSTICA

AMOSTRAGEM

# Definição de Amostragem

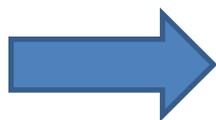
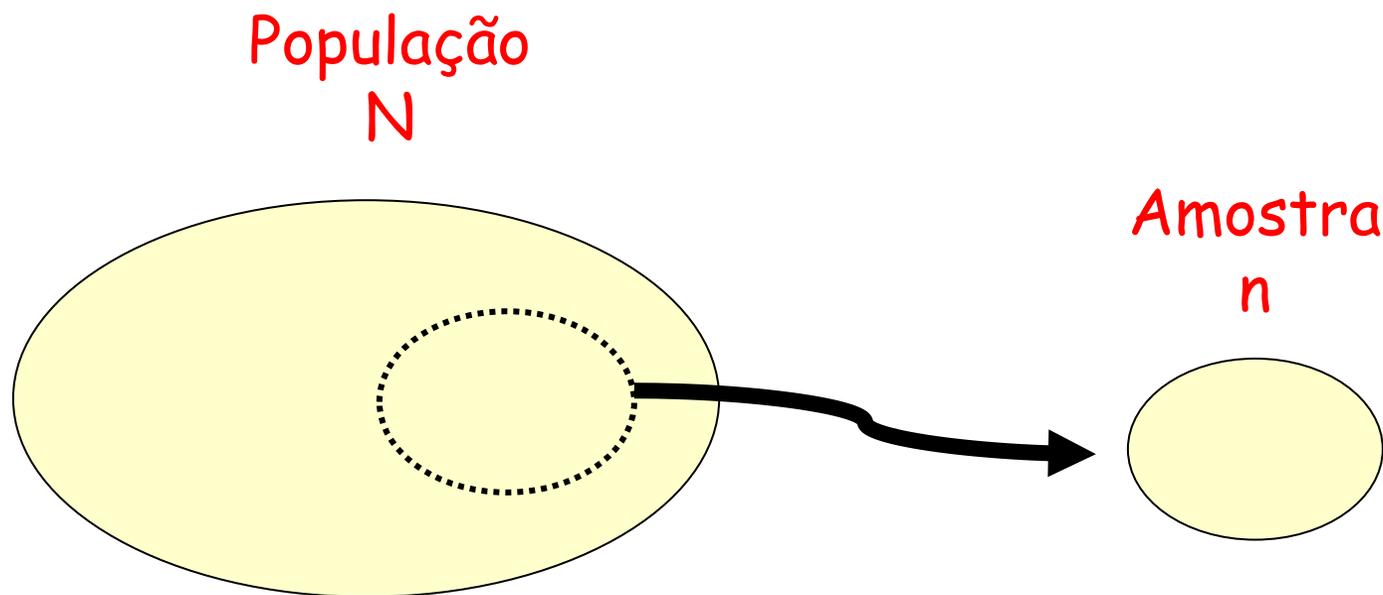
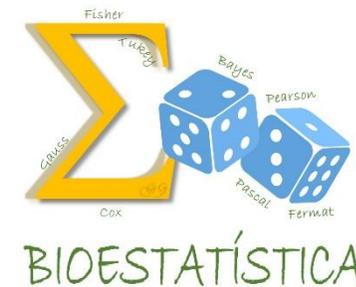


- ✓ É a obtenção de um subconjunto selecionado de uma população.
  
- ✓ **População** é um conjunto de elementos em que cada um dos elementos apresentam uma ou mais características comuns.
  
- ✓ **Amostra** é simplesmente parte de uma população, escolhida para estimar uma característica de interesse a respeito desta população.
  
- ✓ Ex.: Qual(is) a(s) característica(s) comum(ns) de vocês na sala de aula?
  - ✓ Estudantes universitários
  - ✓ Estudantes da USP
  - ✓ Estudantes da FSP

# População e amostra

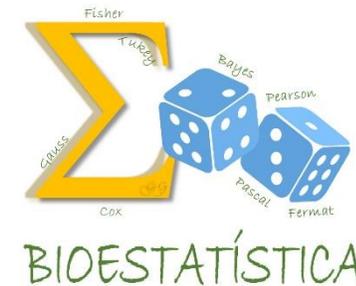
Medidas populacionais são quase sempre desconhecidas.

Na maioria das vezes não é possível estudar a população toda.



A Estatística permite tirar conclusões sobre a população a partir do estudo de alguns de seus elementos (amostra).

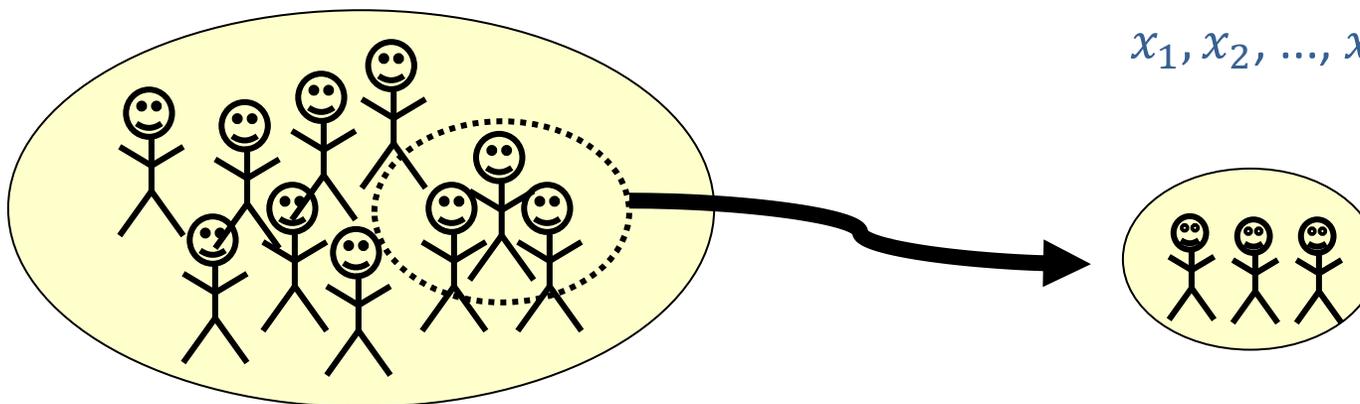
# População e amostra



Seja  $X$  uma variável aleatória (quantitativa) de interesse

Na população (tamanho  $N$ ):

$$X_1, X_2, \dots, X_N$$



Na amostra (tamanho  $n$ ):

$$x_1, x_2, \dots, x_n$$

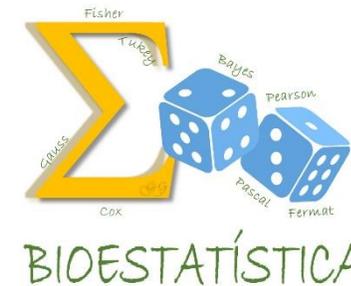
$\mu$  é a média de  $X$  na população =  $E(X)$

$\sigma^2$  é a variância de  $X$  na população =  $VAR(X)$

$\bar{X}$  é a média de  $X$  na amostra

$S^2$  é a variância de  $X$  na amostra  
GLEICEM S CONCEIÇÃO  
MARIA DO ROSÁRIO D D LATORRE  
FSP - USP

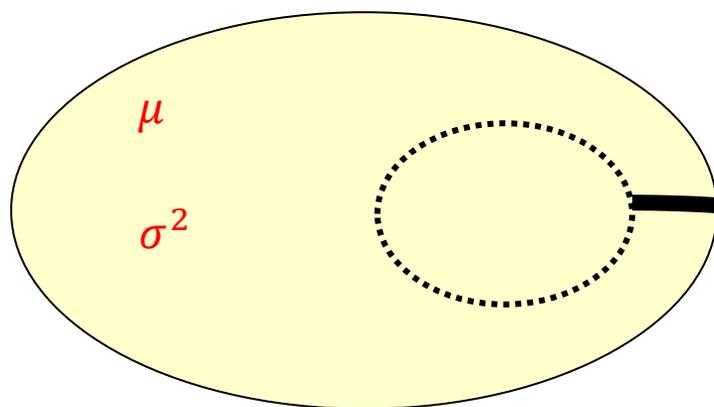
# População e amostra



Seja  $X$  uma variável aleatória (quantitativa) de interesse

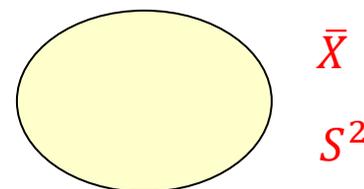
Na população (tamanho  $N$ ):

$$X_1, X_2, \dots, X_N$$



Na amostra (tamanho  $n$ ):

$$x_1, x_2, \dots, x_n$$

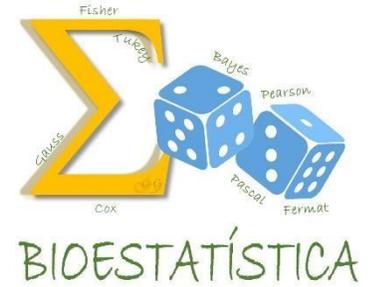


$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad \sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

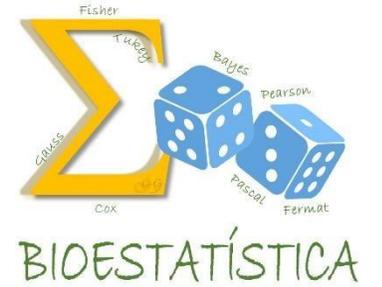
$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$

# Vantagens da Amostragem

- ✓ Custo menor, em relação ao levantamento total
- ✓ Resultado em menor tempo
- ✓ Objetivos mais amplos
- ✓ Dados mais fidedignos

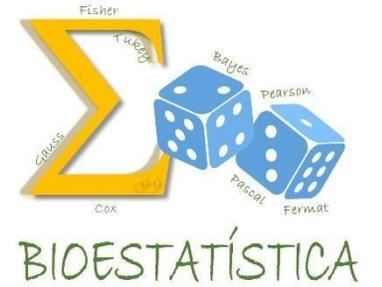


# Razões para a Amostragem



- ✓ População de estudo muito grande (impraticável o levantamento total)
- ✓ Processo de investigação das características de cada elemento for destrutivo (ex.: resistência dos materiais)
- ✓ Desenvolvimento de vacinas, medicamentos e técnicas cirúrgicas
- ✓ Quando a população do estudo é hipotética (ex.: infinitas possibilidades de caras e coroas ao jogar uma moeda para cima infinitas vezes)

# Etapas de um levantamento por amostragem



- ✓ Seja
  - ✓ dentro de um laboratório (ex.: escolha de ratos para um experimento sobre uma nova droga)
  - ✓ para uma população geral (ex.: fertilidade das mulheres moradoras de São Paulo)
  
- ✓ Existem etapas que devem ser seguidas, indistintamente
  - ✓ Explicitação dos objetivos
  - ✓ Definição da população a ser amostrada
  - ✓ Escolha das variáveis a serem observadas
  - ✓ Especificação do grau de precisão desejado
  - ✓ Escolha do instrumentos de medida e da forma de abordagem
  - ✓ Escolha da **unidade amostral** (menor parte distinta e identificável)
  - ✓ Execução de prova experimental (prova-piloto ou pré-teste)
  - ✓ Seleção da amostra

**População alvo** ⇒ **população acessível** ⇒ **população em estudo**

especificação

amostragem

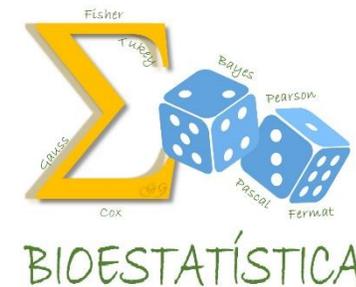


**População alvo**: é definida pelas características clínicas e demográficas  
p.ex: pacientes com câncer de esôfago

**População acessível**: características geográficas e temporais  
p.ex: todos os pacientes com câncer de esôfago atendidos no Hospital do Câncer, em 1997.

**População em estudo**: representativa da população acessível e fácil de trabalhar

## VALIDADE:



**externa:** a população acessível é representativa da população alvo, em relação ao fenômeno de interesse?

**interna:** a população de estudo é representativa da população acessível?

**Em resumo, o quanto os resultados da população de estudo podem ser inferidos para outras populações**

# CRITÉRIOS



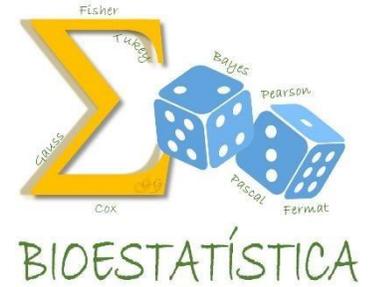
## de inclusão

caracterização da população alvo. Definição (no tempo e no espaço) da população acessível .

## de exclusão

eliminação de indivíduos cuja inclusão diminuiria a qualidade dos dados e/ou da interpretação dos resultados.

# Tipos de Amostragem Probabilística



## ✓ Causal (aleatória) Simples

- ✓ Sem reposição
- ✓ Com reposição

## ✓ Estratificada

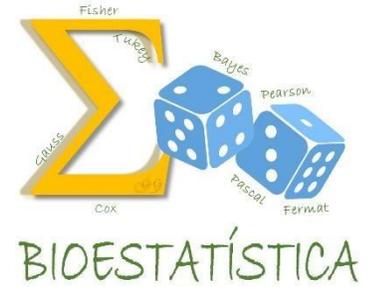
- ✓ Partilha Proporcional

## ✓ Sistemática

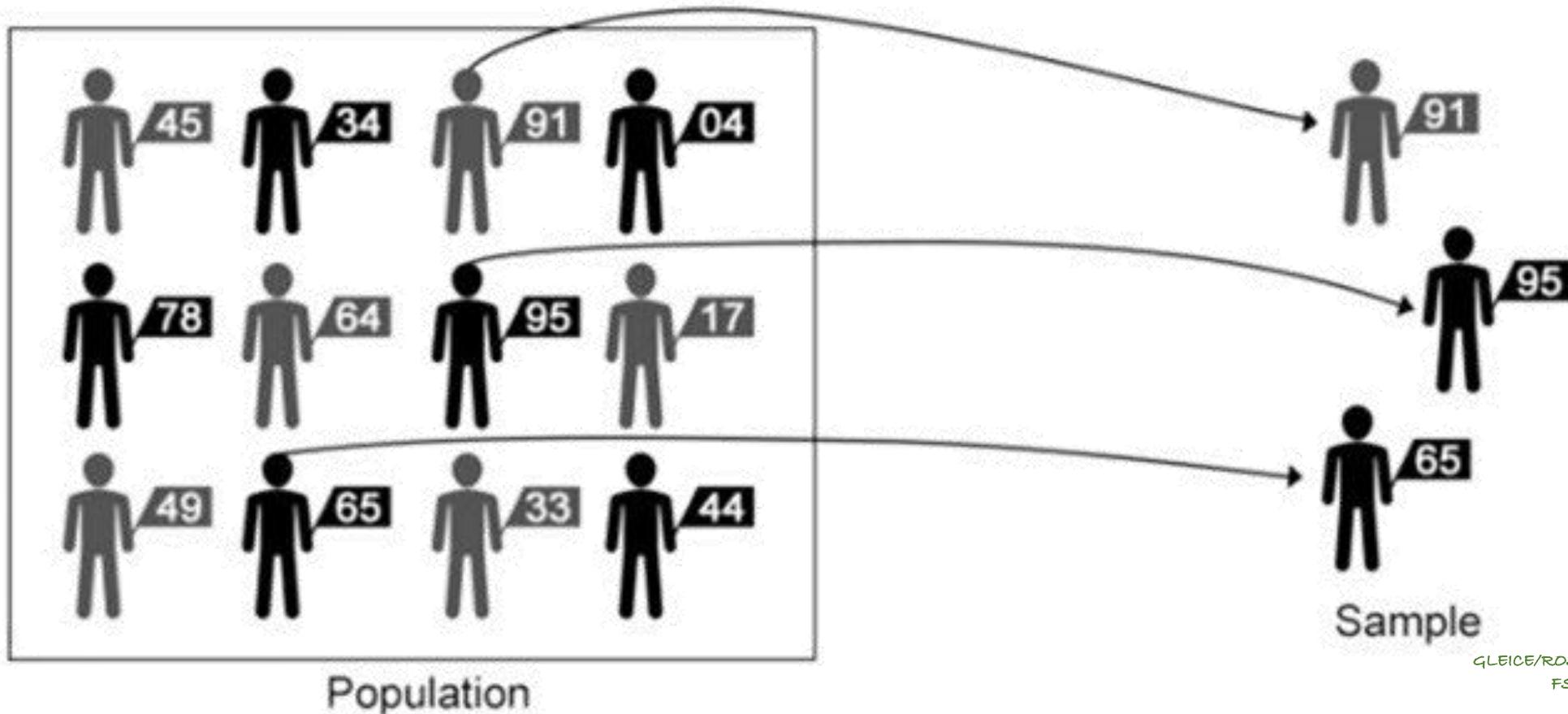
## ✓ Conglomerados

## ✓ Etapa Dupla

# Tipos de Amostragem



- ✓ **Probabilística:** quando cada *unidade amostral* na população tem uma probabilidade conhecida e diferente de zero de pertencer à amostra



# Amostragem Probabilística

## Amostragem Casual (Aleatória) Simples sem Reposição



- ✓ Muito utilizado em função de sua simplicidade
- ✓ Cada unidade amostral, antes da tomada da amostra, em igual probabilidade de pertencer à amostra
- ✓ Exemplo:
  - ✓ Uma população composta de N unidades amostrais
  - ✓ Deseja-se uma amostra de tamanho n
  - ✓ As unidades amostrais são numeradas em sequência (1, 2, ..., N)
  - ✓ Entre esses N números, sorteiam-se n, cujas unidades amostrais correspondentes comporão a amostra
  - ✓ Antes do sorteio, cada unidade tem a probabilidade  $\frac{n}{N}$  de pertencer à amostra

# Amostragem Probabilística

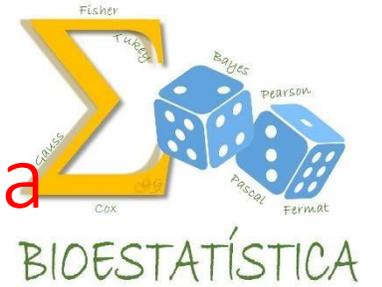
## Amostragem Casual (Aleatória) Simples com Reposição



- ✓ Se unidades amostrais já sorteadas puderem ser novamente sorteadas
- ✓ Serão representadas uma, duas ou mais vezes na amostra
  
- ✓ Exemplo:
  - ✓ De uma população de 7 estudantes, deseja-se obter uma amostra casual simples de 2 estudantes
  - ✓ Sejam: João, Maria, Pedro, Paula, Roberta, Renato e André
  - ✓ Escrevemos o nome de cada um num papel e colocamos num saco de veludo preto (para que não possamos ver os nomes)
  - ✓ Se for **sem** reposição, quando um dos nomes for sorteado ele é retirado do saco
  - ✓ Se for **com** reposição, o nome sorteado é devolvido ao saco de veludo

# Amostragem Probabilística

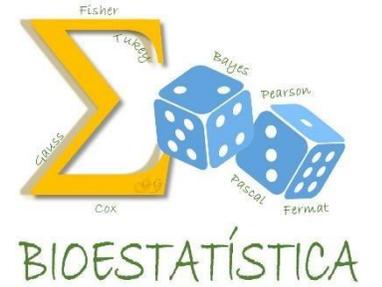
## Amostragem Casual (Aleatória) Simples Estratificada



- ✓ Muitas vezes uma população é composta de subpopulações (ou estratos) bem definidos
- ✓ Encontramos maior homogeneidade entre as unidades amostrais de cada estrato do que entre unidades amostrais de estratos diferentes
- ✓ Exemplo: idade, condições socioeconômicas, etc
- ✓ Nestas condições, o sorteio da amostra é feito dentro de cada um, independentemente
- ✓ Daí o nome amostragem estratificada

# Amostragem Probabilística

## Amostragem com Partilha Proporcional



- ✓ Um caso muito importante da amostragem estratificada
- ✓ O pesquisador deseja que as subpopulações sejam representadas na amostra com a mesma proporcionalidade com que compõem a população total
- ✓ Exemplo para uma população composta de  $h$  estratos

$N$  = tamanho da população

$N_h$  = tamanho de cada estrato populacional

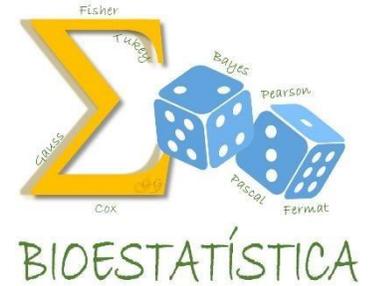
$n$  = tamanho da amostra

$n_h$  = tamanho da amostra do estrato  $h$

- ✓ Temos a partilha proporcional ao multiplicar o tamanho total da amostra  $n$  pelo fator de proporcionalidade  $\frac{N_h}{N}$  com que a subpopulação é representada

# Amostragem Probabilística

## Amostragem com Partilha Proporcional

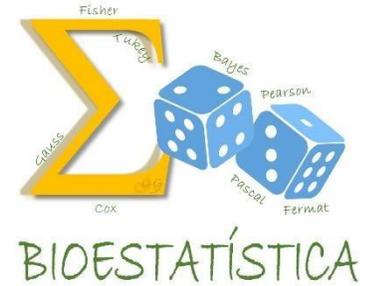


✓ Exemplo: uma população composta de  $N = 500$  unidades amostrais, com três estratos (1, 2 e 3), para obtenção de uma amostra de tamanho total  $n = 40$

$h =$ estrato	$N_h =$ tamanho do estrato $h$ na população	$n_h =$ tamanho da amostra do estrato $h$	Relação $\frac{N_h}{N} = \frac{n_h}{n}$
1	$N_1 = 50$	$n_1 =$	
2	$N_2 = 150$	$n_2 =$	
3	$N_3 = 300$	$n_3 =$	
Total	$N = 500$	$n = 40$	

# Amostragem Probabilística

## Amostragem com Partilha Proporcional

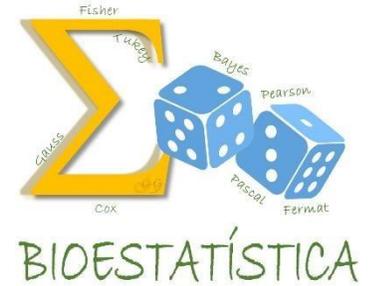


✓ Exemplo: uma população composta de  $N = 500$  unidades amostrais, com três estratos (1, 2 e 3), para obtenção de uma amostra de tamanho total  $n = 40$

$h =$ estrato	$N_h =$ tamanho do estrato $h$ na população	$n_h =$ tamanho da amostra do estrato $h$	Relação $\frac{N_h}{N} = \frac{n_h}{n}$
1	$N_1 = 50$	$n_1 =$	<b>0,1</b>
2	$N_2 = 150$	$n_2 =$	<b>0,3</b>
3	$N_3 = 300$	$n_3 =$	<b>0,6</b>
Total	$N = 500$	$n = 40$	

# Amostragem Probabilística

## Amostragem com Partilha Proporcional

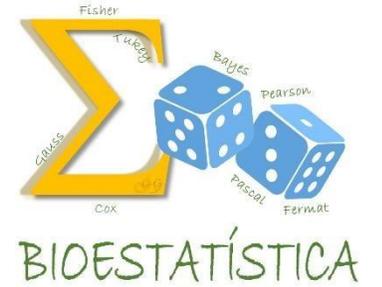


✓ Exemplo: uma população composta de  $N = 500$  unidades amostrais, com três estratos (1, 2 e 3), para obtenção de uma amostra de tamanho total  $n = 40$

$h =$ estrato	$N_h =$ tamanho do estrato $h$ na população	$n_h =$ tamanho da amostra do estrato $h$	Relação $\frac{N_h}{N} = \frac{n_h}{n}$
1	$N_1 = 50$	$n_1 = 4$	0,1
2	$N_2 = 150$	$n_2 = 12$	0,3
3	$N_3 = 300$	$n_3 = 24$	0,6
Total	$N = 500$	$n = 40$	$\frac{n}{N} = 0,08$

# Amostragem Probabilística

## Amostragem Sistemática



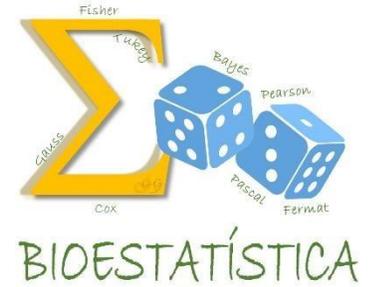
- ✓ Utiliza-se uma ordenação natural de indivíduos, prontuários, quarteirões de uma cidade para este outro tipo de amostragem
- ✓ Sendo  $N$  o total de unidades amostrais e  $n$  o tamanho da amostra desejada, define-se  $k$  como intervalo da amostragem

$$\frac{N}{n} = k$$

- ✓ Sendo  $k$  um número inteiro, sorteia-se entre os números 1, 2, ...,  $k$ , e o valor obtido é chamado de **início causal (i)**
- ✓ Com este único sorteio toda a amostra fica selecionada:  $i; i + k; i + 2k; \dots i + (n-1)k$

# Amostragem Probabilística

## Amostragem Sistemática

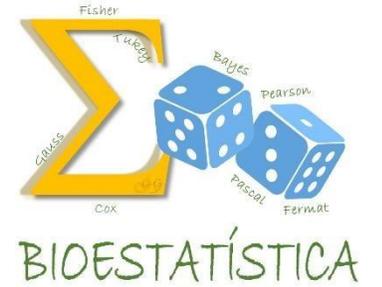


- ✓ Exemplo: uma população de tamanho  $N = 32$ , numerada sequencialmente de 1 a 32, e amostra de tamanho  $n = 8$ , tem-se  $k = 4$
- ✓ O início causal ( $i$ ) é sorteado entre 1, 2, 3 e 4
- ✓ Admitamos que  $i$  sorteado foi 3
- ✓ Assim nossa amostragem sistemática fica constituída

$i$	$i + k$	$i + 2k$	$i + 3k$
3	7	11	15
$i + 4k$	$i + 5k$	$i + 6k$	$i + 7k$
19	23	27	31

# Amostragem Probabilística

## Amostragem Sistemática



- ✓ Exemplo: uma população de tamanho  $N = 32$ , numerada sequencialmente de 1 a 32, e amostra de tamanho  $n = 7$ , tem-se  $k = 4,5714$
- ✓ Com  $k$  não inteiro, considera-se as casas decimais de forma cumulativa, mas despreza-se na definição da unidade sorteada
- ✓ O início causal ( $i$ ) deve ser sorteado entre 0,0001 e 4,5714
- ✓ Supondo que tenha sido  $i = 1,2587$ , o processo é o mesmo

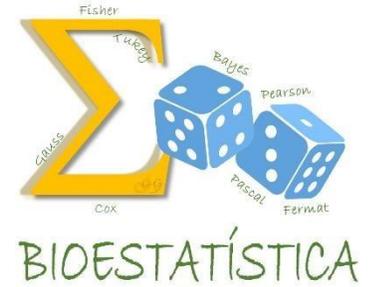
$i = 1,2587 \rightarrow 1$  (unidade sorteada)

$i + k = 1,2587 + 4,5714 = 5,8301 \rightarrow 5$  (unidade sorteada)

- ✓ E assim sucessivamente

# Amostragem Probabilística

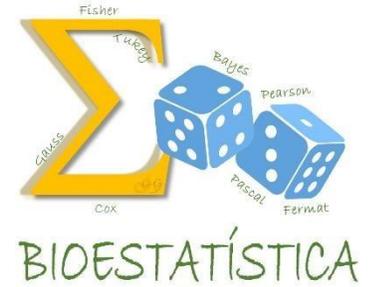
## Amostragem por Conglomerados



- ✓ **Conglomerado:** um conjunto de unidades elementares da população
- ✓ Exemplo: estimação do consumo médio diário de proteína por pessoa residente em uma área urbana
  - ✓ População: conjunto de pessoas que residem naquela área urbana
  - ✓ Unidade elementar ou de análise: cada pessoa
- ✓ Pode-se decidir a amostragem por conglomerado utilizando cada quarteirão do bairro como um conglomerado
- ✓ O sorteio é idêntico aos processos anteriores, após a numeração de cada conglomerado

# Amostragem Probabilística

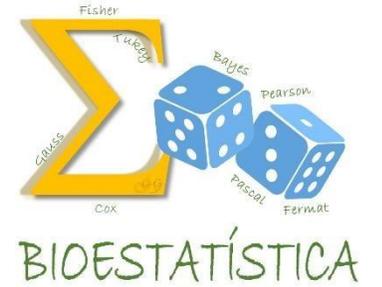
## Amostragem por Etapa Dupla (ou estágio duplo)



- ✓ Uma modificação da amostragem por conglomerados
  
- ✓ Primeira etapa
  - ✓ Selecionam-se os conglomerados que passam a se chamar unidades amostrais de primeiro estágio (ou unidades primárias)
  - ✓ Não se desejando usar todos os elementos que constituem cada conglomerado, passa-se à segunda etapa
  
- ✓ Segunda etapa
  - ✓ Sortear dentro de cada conglomerado já selecionado as unidades amostrais de segundo estágio (ou unidades secundárias), que vão constituir a amostra
  
- ✓ Este processo poder ser generalizado para vários estágios

# Amostragem Probabilística

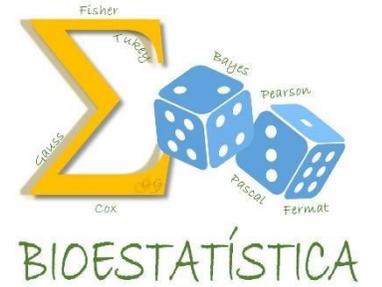
## Precisão



- ✓ Um processo de amostragem pode gerar várias possíveis amostras, mas somente uma é utilizada
- ✓ Cada uma destas amostras forneceria uma determinada estimativa do valor médio da população (ou flutuação)
- ✓ A amostragem probabilística permite estimar a magnitude esta flutuação em torno da média
- ✓ Aumentando-se o tamanho da amostra, aumenta-se a precisão
- ✓ Estimativas obtidas por amostragem sem reposição são mais precisas do que as obtidas com reposição

# Amostragem Probabilística

## Vício, viés ou tendenciosidade

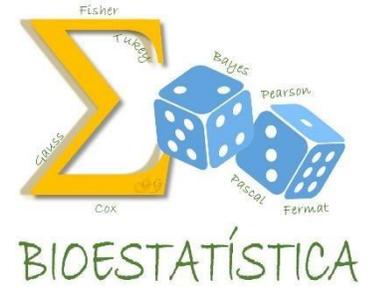


- ✓ É a diferença entre a média que seria obtida de todas as possíveis amostras (precisão) e o verdadeiro valor do parâmetro populacional em estudo



# Amostragem Probabilística

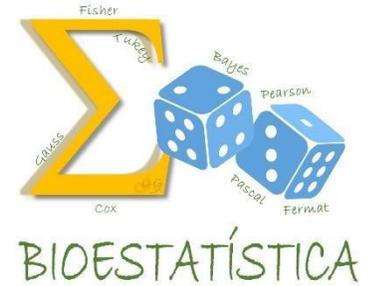
## Vício, viés ou tendenciosidade



- ✓ A amostragem é viciada quando o processo de escolha é intencional
- ✓ A escolha das unidades amostrais estará inevitavelmente influenciado por tendências, preferência e fatores subjetivos pessoais diversos
- ✓ Exemplo: perguntar aos participantes de uma convenção *Star Trek* qual o gênero literário preferido deles
- ✓ Existem vícios que são alheios ao processo de amostragem

# Amostragem Probabilística

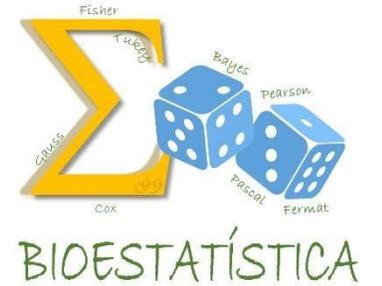
## Ausência de resposta



- ✓ É a falha na obtenção de dados de alguma unidade da amostra selecionada
- ✓ Na coleta de dados de populações humanas com o uso de questionários, temos algumas causas da ausência de resposta:
  - ✓ Unidade não encontrada (ex.: a pessoa não estava em casa)
  - ✓ A pessoa entrevistada não sabe responder
  - ✓ A pessoa entrevistada é fisicamente incapacitada para responder
  - ✓ A pessoa entrevistada se recusa inflexivelmente a responder
  - ✓ A pessoa a ser entrevistada fica ausente durante todo o tempo disponível para o trabalho de campo
- ✓ Quando as características da parte da população que não responde não diferem daquelas da parte que responde, não haverá vício
- ✓ Quando há diferenças, o fato de amostra apresentar ausência de respostas introduzirá vício

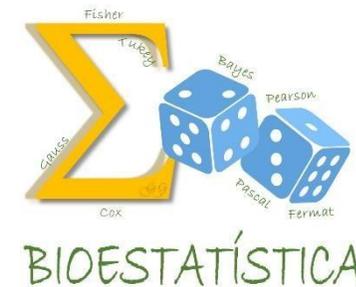
# Amostragem Probabilística

## Ausência de resposta



- ✓ Maneiras de solucionar a ausência de resposta
- ✓ **Repasse:** estabelecer o número de repasses (ou retornos) feitos às unidades que não responderam, antes de declarar “impossibilidade de estabelecer contato”
  - ✓ Pode-se empregar a mesma técnica de coleta de dados ou uma técnica melhorada para obtenção de dados
  - ✓ Quando não é possível o repasse em todas as unidades não respondentes, aplica-se uma subamostra
- ✓ **Reposição:** acrescentar aos endereços do levantamento atual alguns endereços de levantamento semelhantes
- ✓ **Substituição:** substituir os casos de não resposta por vizinhos ou por nova amostra de tamanho igual ao número de ausência de respostas. Este procedimento não garante a eliminação ou diminuição do vício, pois as unidades substituídas podem se assemelhar mais às unidades que já responderam

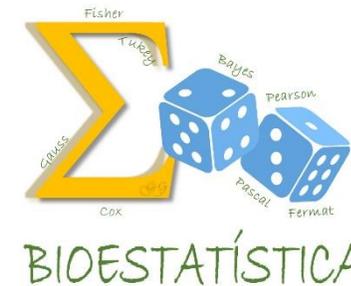
# Tipos de Amostragem



## ✓ Não-Probabilística

- ✓ Risco de escolher uma unidade amostral em função de sua representatividade em relação à população
- ✓ Situações em que é utilizada:
  - ✓ Voluntários em ensaios clínicos para testes de novos medicamentos
  - ✓ Intencional: quando as unidades são escolhidas pelo pesquisador (ex.: verificação de poluição de praias)
  - ✓ Acesso mais fácil: quando as unidades amostrais são escolhidas por estarem em melhores condições de acessibilidade (ex.: o conjunto de habitantes de uma zona rural, tendo como unidade amostral o domicílio)

# Amostragem não probabilística (geralmente utilizada em estudos analíticos)



## amostragem consecutiva (cumulativa)

São selecionados os indivíduos que preenchem o(s) critério(s) de inclusão, em um específico intervalo de tempo ou tamanho de amostra.

problema: se o tempo de coleta é muito curto, posso não ter tempo de coletar todos os indivíduos que seriam representativos da população acessível.

## amostragem por conveniência

São selecionados os indivíduos que estão facilmente disponíveis.

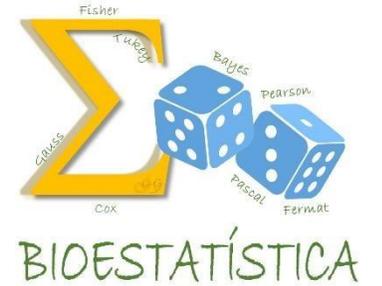
problema: os voluntários poderão não ser representativos da população acessível (p. ex.: só são coletados dados de indivíduos mais saudáveis ou dos mais doentes, ou dos mais expostos, etc).

## amostragem intencional

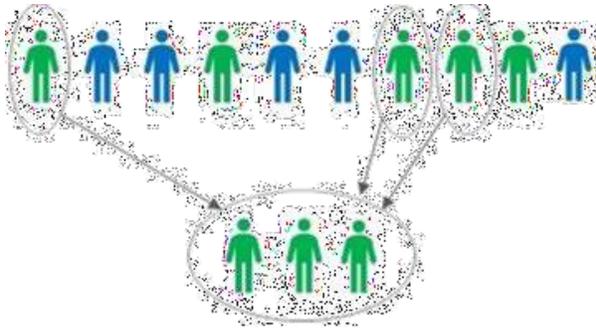
Da população acessível são **escolhidos** aqueles indivíduos **que se julga** serem mais apropriados para o estudo.

problema: é arriscado inferir as conclusões para outras populações.

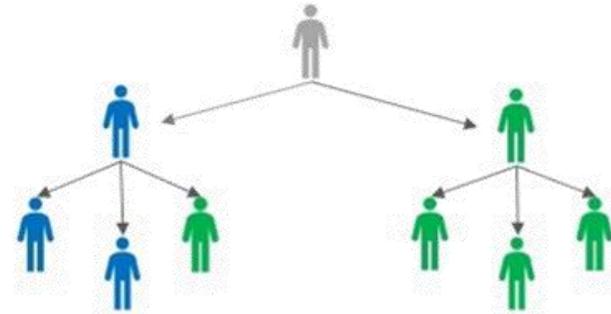
# Tipos de Amostragem Não-Probabilística



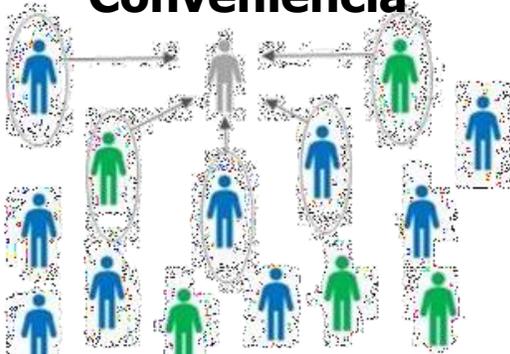
## Quotas



## Bola de Neve



## Conveniência



## Julgamento/Intencional

