

Inferência Estatística – Distribuições Amostrais e TCL

- Introdução e conceitos fundamentais
- Distribuição amostral
- Teorema central do limite

Divisão da Estatística

- **Descritiva**
- **Inferência**



Inferência Estatística

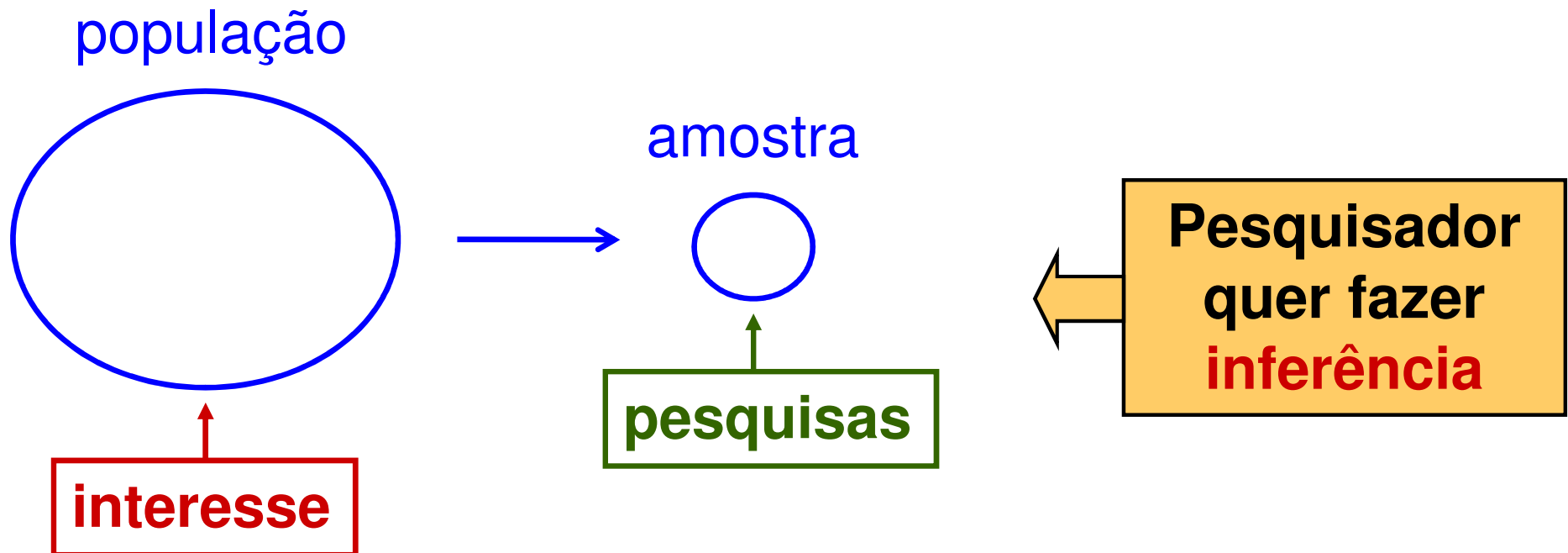
Idéia fundamental



Dar informação sobre o **todo com
base no conhecimento da **parte****

**“Não é preciso beber toda a garrafa
para saber se o vinho é bom.”**

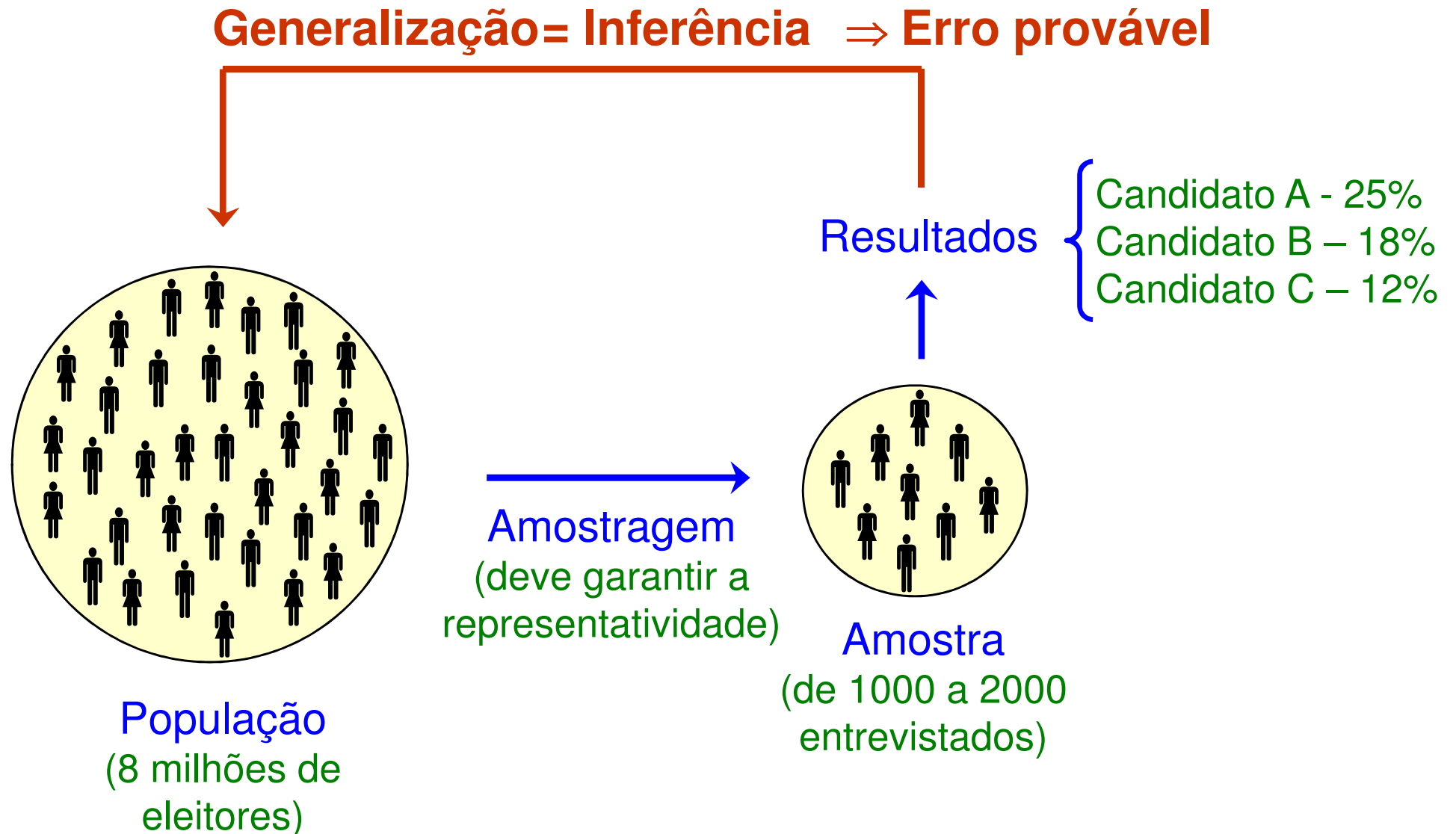
⇒ Pesquisas são feitas com **amostras**, mas o pesquisador quer **estender os resultados** que obteve para toda a **população**.



Inferência estatística

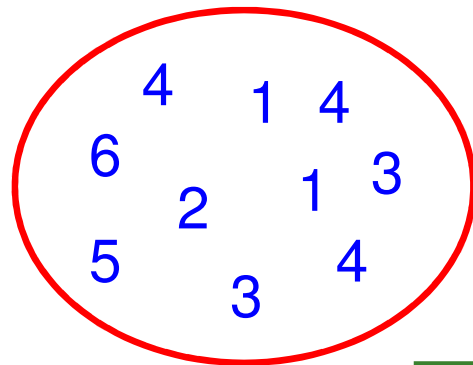
Conceito: É o conjunto de procedimentos estatísticos que têm por finalidade **generalizar conclusões** de uma amostra para uma população.

Exemplo: Pesquisas eleitorais no Rio Grande do Sul



Exemplo

N = tamanho da população



$N=10$

n = tamanho da amostra

$[X_1, X_2]$

$n=2$

com reposição

$\left\{ \begin{array}{l} (1, 1) \\ (1, 2) \\ (1, 3) \\ \dots \\ (6, 6) \end{array} \right\}$

A amostra é uma variável aleatória

$[X_1, X_2]$

$(1, 1)$

$(1, 2)$

$(1, 3)$

\dots

$(6, 6)$

\bar{X}

$\bar{x}_1 = 1$

$\bar{x}_2 = 1,5$

$\bar{x}_3 = 2$

\dots

$\bar{x}_k = 6$

S^2

$s_1^2 = 0$

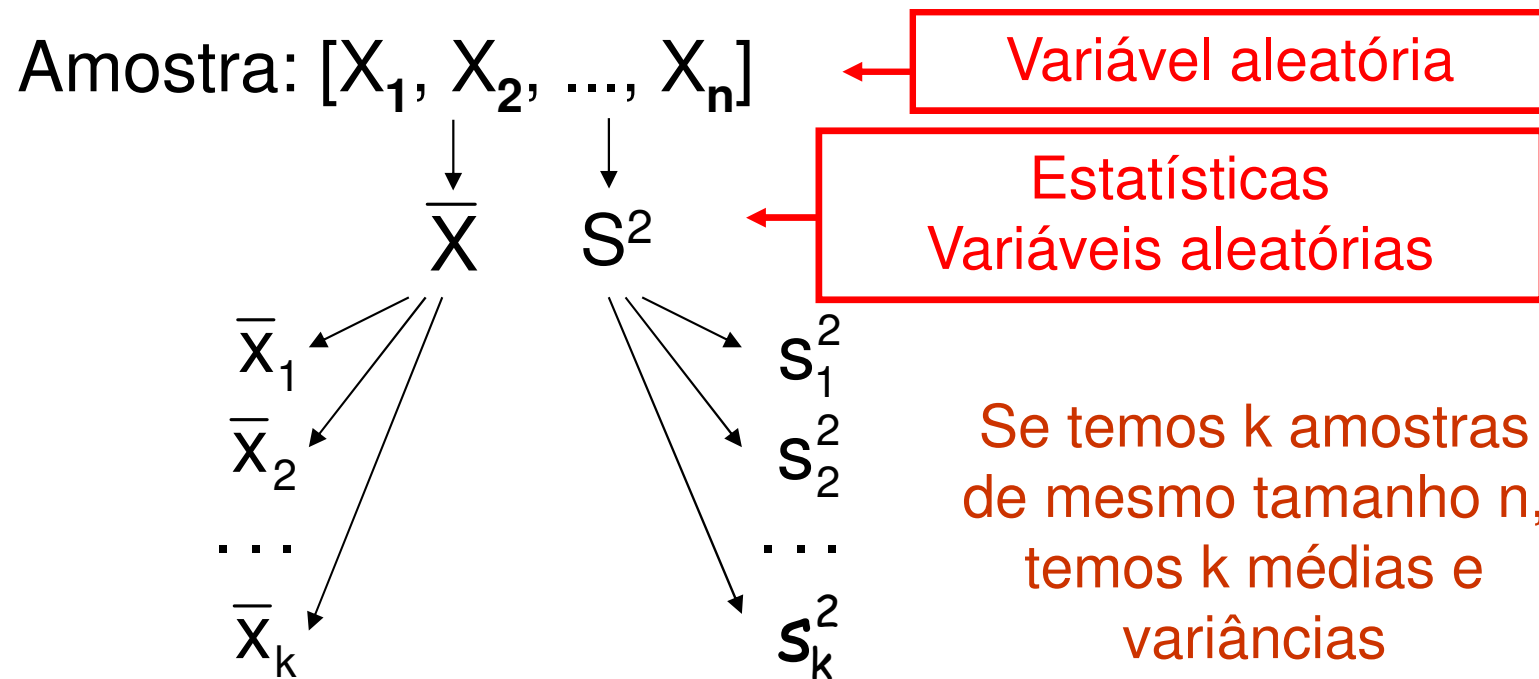
$s_2^2 = 0,5$

$s_3^2 = 2$

\dots

$s_k^2 = 0$

Como a amostra é uma variável aleatória, qualquer função da amostra (soma, média, variância) também será uma variável aleatória.



Estatística é qualquer **valor obtido em função da amostra**.

⇒ Como as estatísticas são variáveis aleatórias, também terão alguma distribuição de probabilidade com média, variância, etc.

Distribuição amostral é a **distribuição de probabilidade de uma estatística**.

Exemplo:

O mecânico de uma oficina de regulagem para carros 1.0, 1.4, 2.0, cobra pelo serviço 40, 45 e 50 reais, respectivamente. Seja a variável X = valor cobrado pelo mecânico, com a seguinte distribuição de probabilidade:

$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

a) Determine a **média** e a **variância da população**.

$$E(X) = \mu = \sum_{x \in S_X} x p(x) = 40 \times 0,2 + 45 \times 0,3 + 50 \times 0,5 = 46,5$$

$$\begin{aligned} V(X) = \sigma^2 &= E(X^2) - \mu^2 \\ &= (40^2 \times 0,2 + 45^2 \times 0,3 + 50^2 \times 0,5) - 46,5^2 = 15,25 \end{aligned}$$

Exemplo:

O mecânico de uma oficina de regulagem para carros 1.0, 1.4, 2.0, cobra pelo serviço 40, 45 e 50 reais, respectivamente. Seja a variável X = valor cobrado pelo mecânico, com a seguinte distribuição de probabilidade:

$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

b) Suponha a retirada de uma amostra de tamanho **$n = 2$** , com reposição.

Quantas e quais são as possíveis amostras retiradas da população e qual a **probabilidade** associada a cada uma?

Determine a **média e a variância da distribuição amostral da média**.

$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

$$n = 2 \rightarrow [X_1, X_2] \rightarrow k = N^n = 3^2 = 9 \leftarrow \text{quantas}$$

onde:

k = número de amostras possíveis

N = tamanho da população

n = tamanho da amostra

Amostra	$[X_1, X_2]$	$P [X_1, X_2]$	\bar{X}
1	(40, 40)	$0,2 \cdot 0,2 = 0,04$	40
2	(40, 45)	$0,2 \cdot 0,3 = 0,06$	42,5
3	(40, 50)	$0,2 \cdot 0,5 = 0,10$	45
4	(45, 40)	0,06	42,5
5	(45, 45)	0,09	45
6	(45, 50)	0,15	47,5
7	(50, 40)	0,10	45
8	(50, 45)	0,15	47,5
9	(50, 50)	0,25	50

quais

Distribuição amostral da média das amostras de tamanho 2

$\bar{X} = \bar{x}$	40	42,5	45	47,5	50	Σ
$P(\bar{X} = \bar{x})$	0,04	0,12	0,29	0,3	0,25	1

$$\begin{aligned} E(\bar{X}) &= \mu_{\bar{X}} = \sum_{\bar{x} \in S_{\bar{X}}} \bar{x} p(\bar{x}) \\ &= 40 \times 0,04 + 42,5 \times 0,12 + 45 \times 0,29 + 47,5 \times 0,3 + 50 \times 0,25 \\ &= 46,5 \leftarrow = \mu \text{ (média da população)} \end{aligned}$$

$$\begin{aligned} V(\bar{X}) &= \sigma_{\bar{X}}^2 = E(\bar{X}^2) - \mu_{\bar{X}}^2 \\ &= (40^2 \times 0,04 + 42,5^2 \times 0,12 + \dots + 50^2 \times 0,25) - 46,5^2 \\ &= 7,625 \leftarrow = \text{metade de } \sigma^2 \text{ (variância da população)} \end{aligned}$$

Exemplo:

O mecânico de uma oficina de regulagem para carros 1.0, 1.4, 2.0, cobra pelo serviço 40, 45 e 50 reais, respectivamente. Seja a variável X = valor cobrado pelo mecânico, com a seguinte distribuição de probabilidade:

$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

c) Suponha a retirada de uma amostra de tamanho $n = 3$, com reposição.

Quantas e quais são as possíveis amostras retiradas da população e qual a **probabilidade** associada a cada uma?

Determine a **média e a variância da distribuição amostral da média**.

Quantas?

$$n = 3 \rightarrow [X_1, X_2, X_3] \rightarrow k = N^n = 3^3 = 27$$

Quais?

Amostra	$[X_1, X_2, X_3]$	$P[X_1, X_2, X_3]$	\bar{X}	Amostra	$[X_1, X_2, X_3]$	$P[X_1, X_2, X_3]$	\bar{X}
1	(40, 40, 40)	0,008	40	15	(45, 45, 50)	0,045	46,7
2	(40, 40, 45)	0,012	41,7	16	(45, 50, 40)	0,030	45
3	(40, 40, 50)	0,020	43,3	17	(45, 50, 45)	0,045	46,7
4	(40, 45, 40)	0,012	41,7	18	(45, 50, 50)	0,075	48,3
5	(40, 45, 45)	0,018	43,3	19	(50, 40, 40)	0,020	43,3
6	(40, 45, 50)	0,030	45	20	(50, 40, 45)	0,030	45
7	(40, 50, 40)	0,020	43,3	21	(50, 40, 50)	0,050	46,7
8	(40, 50, 45)	0,030	45	22	(50, 45, 40)	0,030	45
9	(40, 50, 50)	0,050	46,7	23	(50, 45, 45)	0,045	46,7
10	(45, 40, 40)	0,012	41,7	24	(50, 45, 50)	0,075	48,3
11	(45, 40, 45)	0,018	43,3	25	(50, 50, 40)	0,050	46,7
12	(45, 40, 50)	0,030	45	26	(50, 50, 45)	0,075	48,3
13	(45, 45, 40)	0,018	43,3	27	(50, 50, 50)	0,125	50
14	(45, 45, 45)	0,027	45				

Distribuição amostral da média das amostras de tamanho 3

$\bar{X} = \bar{x}$	40	41,7	43,3	45	46,7	48,3	50	Σ
$P(\bar{X} = \bar{x})$	0,008	0,036	0,114	0,207	0,285	0,225	0,125	1

$$E(\bar{X}) = \mu_{\bar{X}} = \sum_{\bar{x} \in S_{\bar{X}}} \bar{x} p(\bar{x})$$

$$= 40 \times 0,008 + 41,7 \times 0,036 + \dots + 50 \times 0,125$$

$$= 46,5 \quad \leftarrow = \mu \text{ (média da população)}$$

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = E(\bar{X}^2) - \mu_{\bar{X}}^2$$

$$= (40^2 \times 0,008 + 41,7^2 \times 0,036 + \dots + 50^2 \times 0,125) - 46,5^2$$

$$= 5,083 \quad \leftarrow = \text{um terço de } \sigma^2 \text{ (variância da população)}$$

População

$$E(X) = \mu = 46,5$$

$$V(X) = \sigma^2 = 15,25$$

Amostras de tamanho $n = 2$

$$E(\bar{X}) = \mu_{\bar{X}} = 46,5 \leftarrow = \mu$$

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = 7,625 \leftarrow = \sigma^2/2$$

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Amostras de tamanho $n = 3$

$$E(\bar{X}) = \mu_{\bar{X}} = 46,5 \leftarrow = \mu$$

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = 5,083 \leftarrow = \sigma^2/3$$

$$\sigma_{\bar{X}} = \sqrt{V(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Erro padrão
da média

Distribuição de probabilidade da população

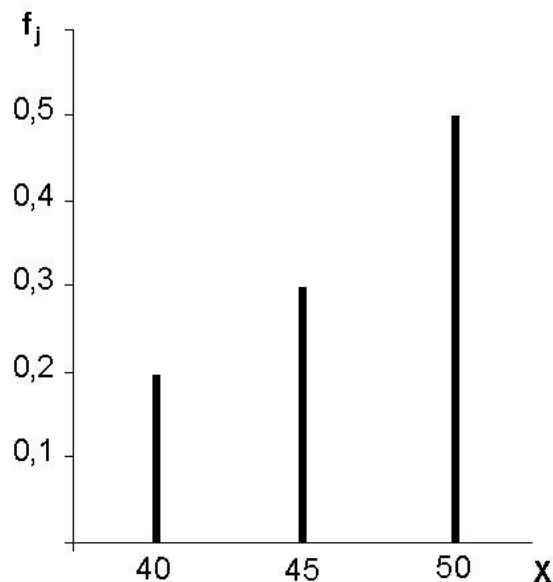
$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

Distribuição amostral da média das amostras de tamanho 2

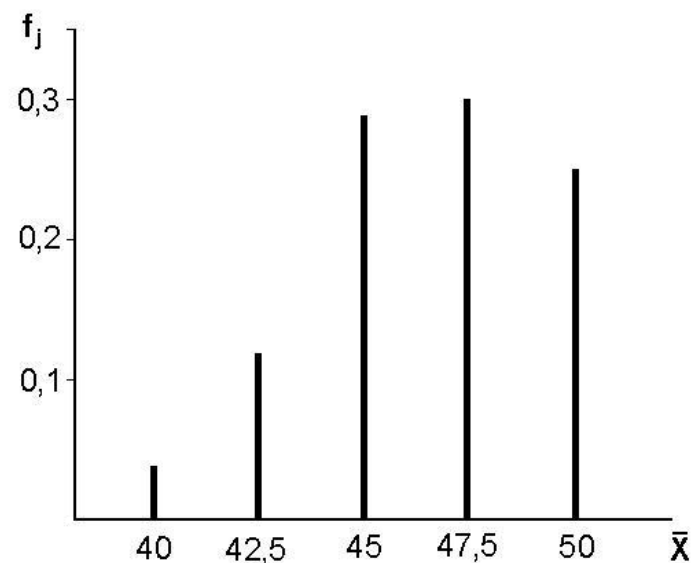
$\bar{X} = \bar{x}$	40	42,5	45	47,5	50	Σ
$P(\bar{X} = \bar{x})$	0,04	0,12	0,29	0,3	0,25	1

Distribuição amostral da média das amostras de tamanho 3

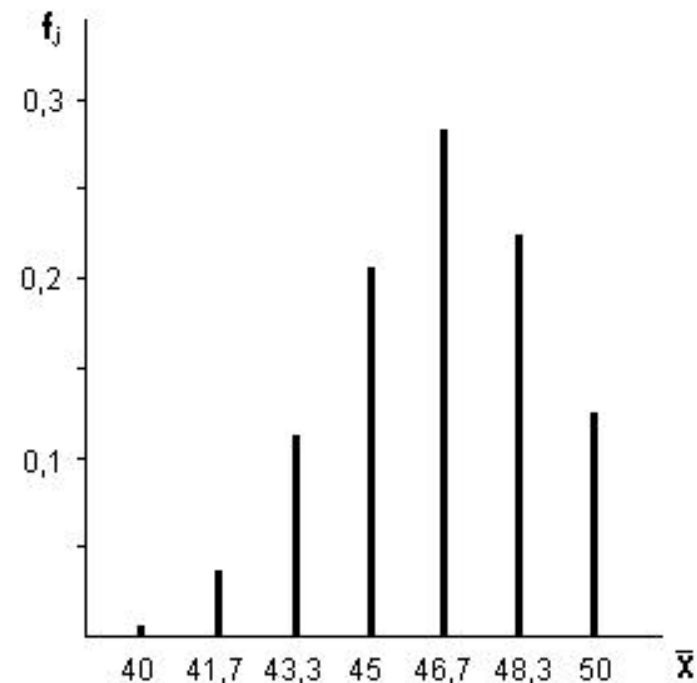
$\bar{X} = \bar{x}$	40	41,7	43,3	45	46,7	48,3	50	Σ
$P(\bar{X} = \bar{x})$	0,008	0,036	0,114	0,207	0,285	0,225	0,125	1



Distribuição da população



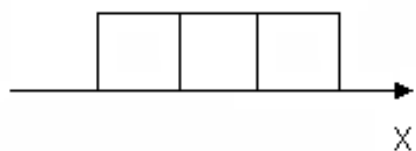
Distribuição da média das amostras de tamanho 2



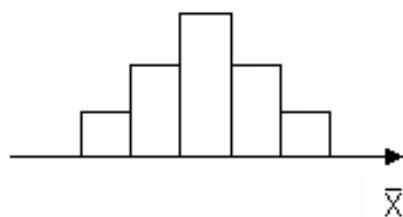
Distribuição da média das amostras de tamanho 3

Comparando o histograma da população X com os histogramas da média para as amostras de tamanhos 2 e 3, observamos que, **mesmo a distribuição da população não sendo simétrica**, a distribuição amostral da média **tende para a simetria** à medida que **o tamanho da amostra aumenta**.

População



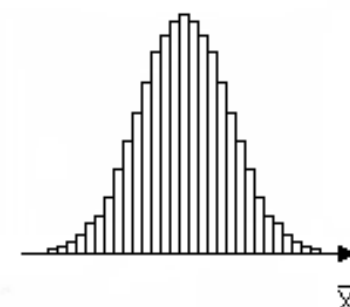
$n=2$



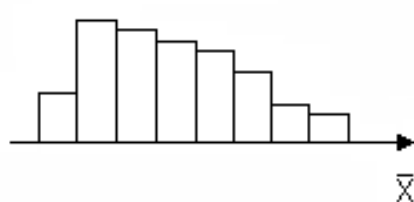
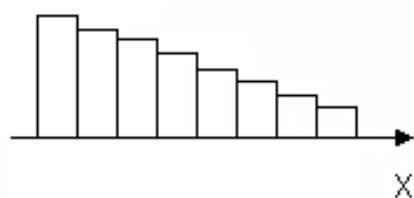
$n=5$



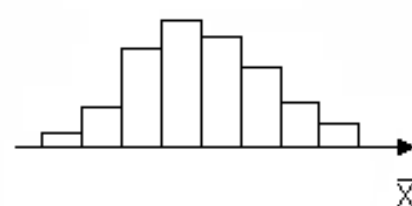
$n=25$



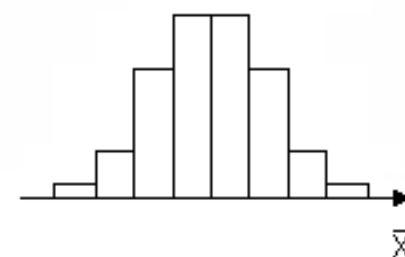
$n=2$



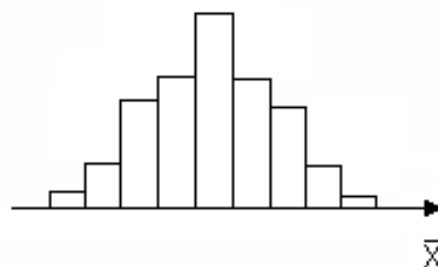
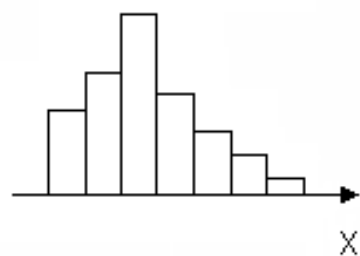
$n=10$



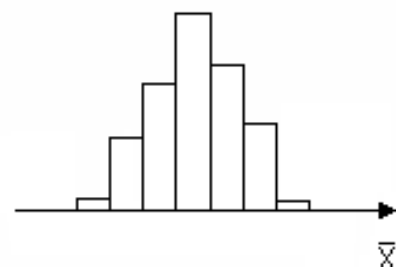
$n=30$



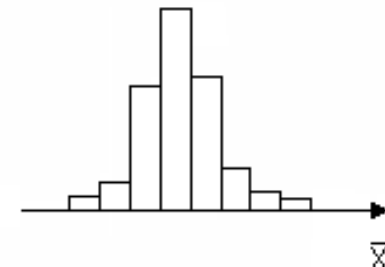
$n=3$



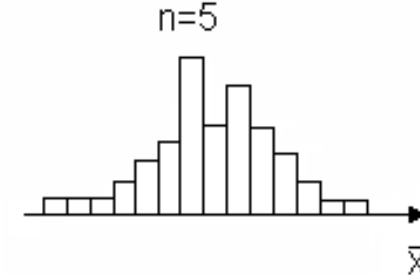
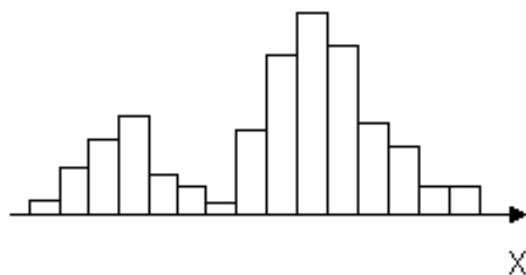
$n=5$



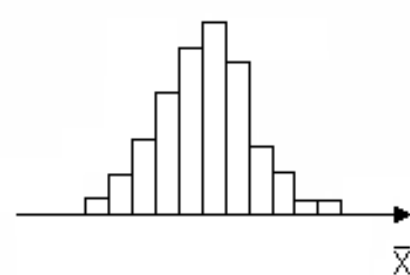
$n=10$



$n=5$



$n=15$



$n=25$

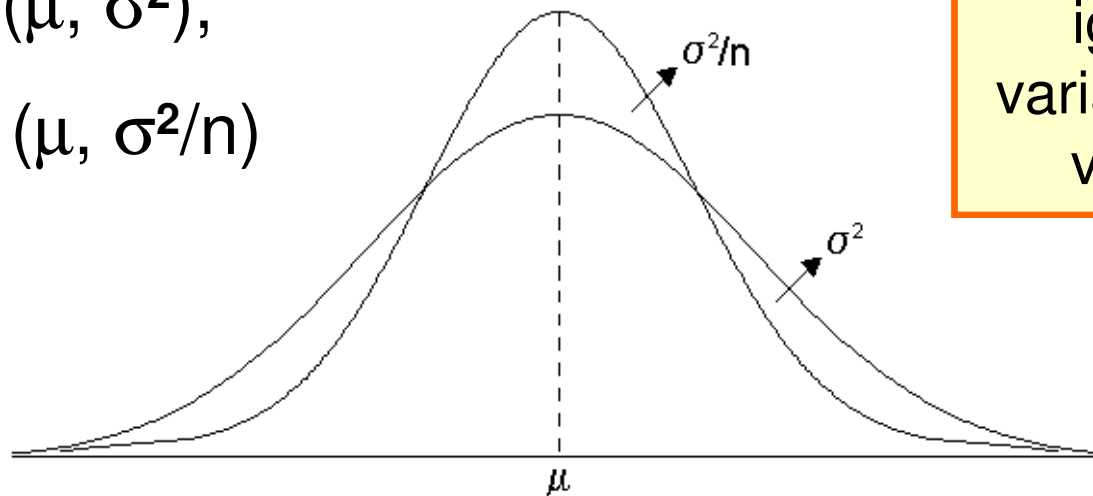


Qual é a distribuição da média?

⇒ Se a população (X) de onde foi extraída a amostra aleatória não tiver distribuição normal, então a distribuição amostral da média se **aproximará da normal** à medida que o tamanho da amostra (n) **cresce** (**Teorema Central do Limite**).

⇒ Se a população (X) de onde foi extraída a amostra aleatória tiver distribuição normal, então a distribuição amostral da média será normal.

se $X \sim N(\mu, \sigma^2)$,
então, $\bar{X} \sim N(\mu, \sigma^2/n)$



As médias são iguais, mas a variância de \bar{X} é n vezes menor.

Exercício proposto: Sabe-se a variável X (peso do queijo ralado contido em pacotinhos de 100g vendidos em supermercados) tem distribuição normal de média 100 e desvio padrão 10.

(a) Qual a probabilidade de encontrar um pacote de queijo ralado com peso entre 95g e 105g?

(b) Se 16 pacotes são escolhidos ao acaso em um supermercado, qual a probabilidade do peso médio estar entre 95g e 105g?

Solução:

(a) Como X é uma $N(100, 10^2)$ vem:

$$P(95 < X < 105) = P(-0,5 < Z < 0,5) = 2 \cdot 0,1915 = 38,30\%$$

(b) Neste caso é uma $N(100; 2,5^2)$, então:

$$P(95 < \bar{X} < 105) = P(-2,0 < Z < 2,0) = 2 \cdot 0,4772 = 95,44\%$$