Amostragem e Delineamento

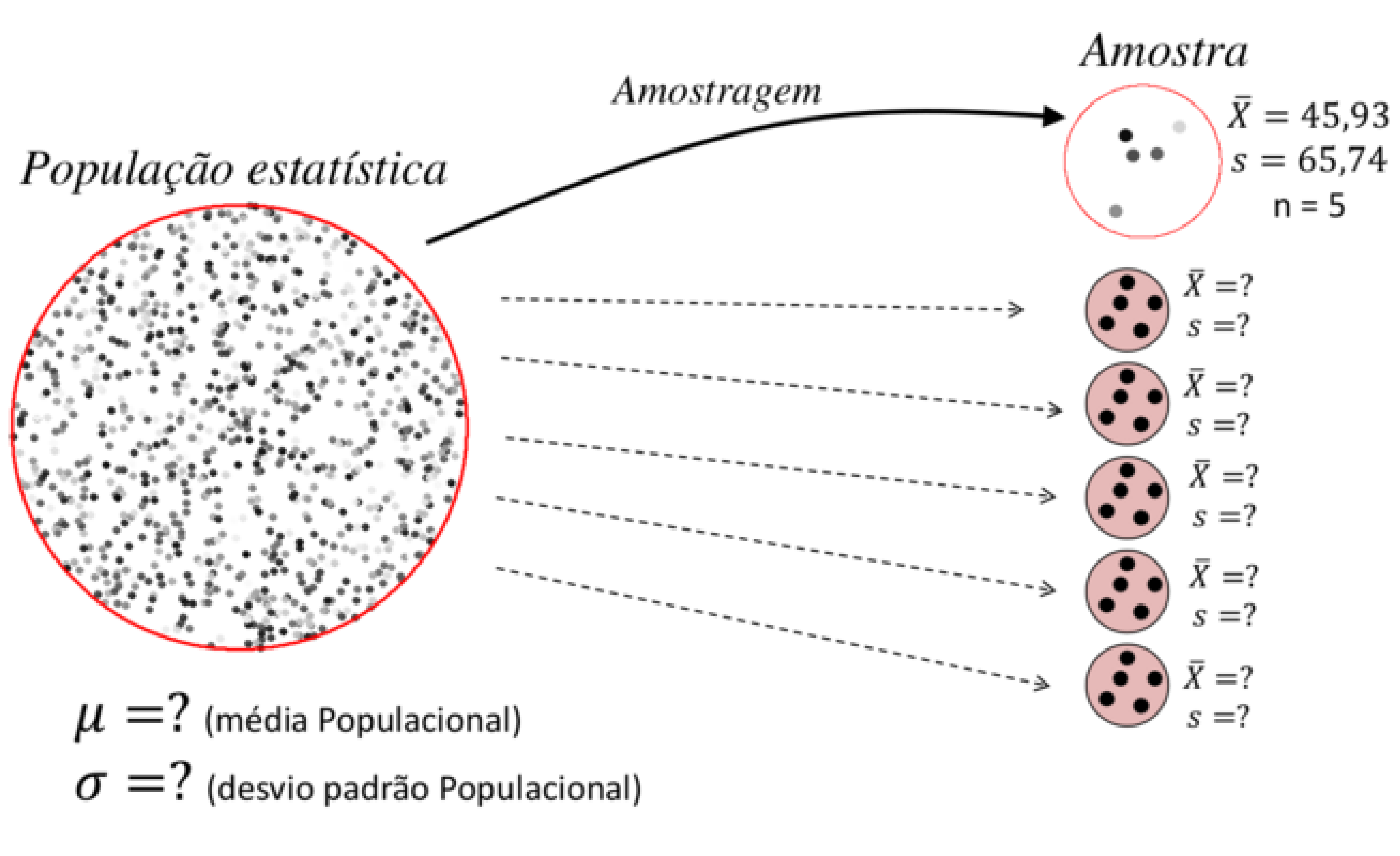
Amostrando uma população estatística

Fabio Cop (fabiocopf@gmail.com)
Instituto do Mar - UNIFESP
Última atualização em 28 de março de 2022

Conteúdo da aula

1. Amostragem aleatória simples
2. Amostragem aleatória estratificada
3. Amostragem sistemática
4. Erro amostral, acurácia e precisão

O processo de amostragem e inferência sobre uma população estatística



1. Amostragem aleatória simples

Suponha uma população hipotética de somente 10 elementos:

População: 3, 10, 14, 19, 27, 28, 29, 41, 42, 43

Exemplos de amostras aleatórias simples

Tamanho amostral: $n = 10$

Amostra 1: 42, 19, 29, 3, 10

Amostra 2: 27, 28, 42, 3, 43

Se nos dois casos, os elementos foram *sorteados* a partir da população estatística, a amostra 1 é tão aleatória e válida do ponto de vista estatístico quanto a amostra 2.

Na amostragem aleatória simples, cada elemento da população tem a **mesma probabilidade** de compor uma amostra. Se a população tem N elementos, cada um tem probabilidade $\frac{1}{N}$ de ser selecionado.

1. Amostragem aleatória simples

Suponha uma população hipotética de somente 10 elementos:

População: 3, 10, 14, 19, 27, 28, 29, 41, 42, 43

Exemplos de amostras aleatórias simples

Tamanho amostral: $n = 10$

Amostra 3: 3, 10, 14, 19, 27

Amostra 4: 43, 42, 41, 29, 28

Ainda que nos dois casos acima tenhamos conduzido um *sorteio aleratório*, as amostras 3 e 4 resultaram respectivamente, nos 5 **menores** ou nos 5 **maiores** valores da população estatística.

Um resultado de uma amostragem aleatória simples pode ser válido do ponto de vista estatístico, mas ainda assim **não representativo** da população estatística.

2. Amostragem aleatória estratificada

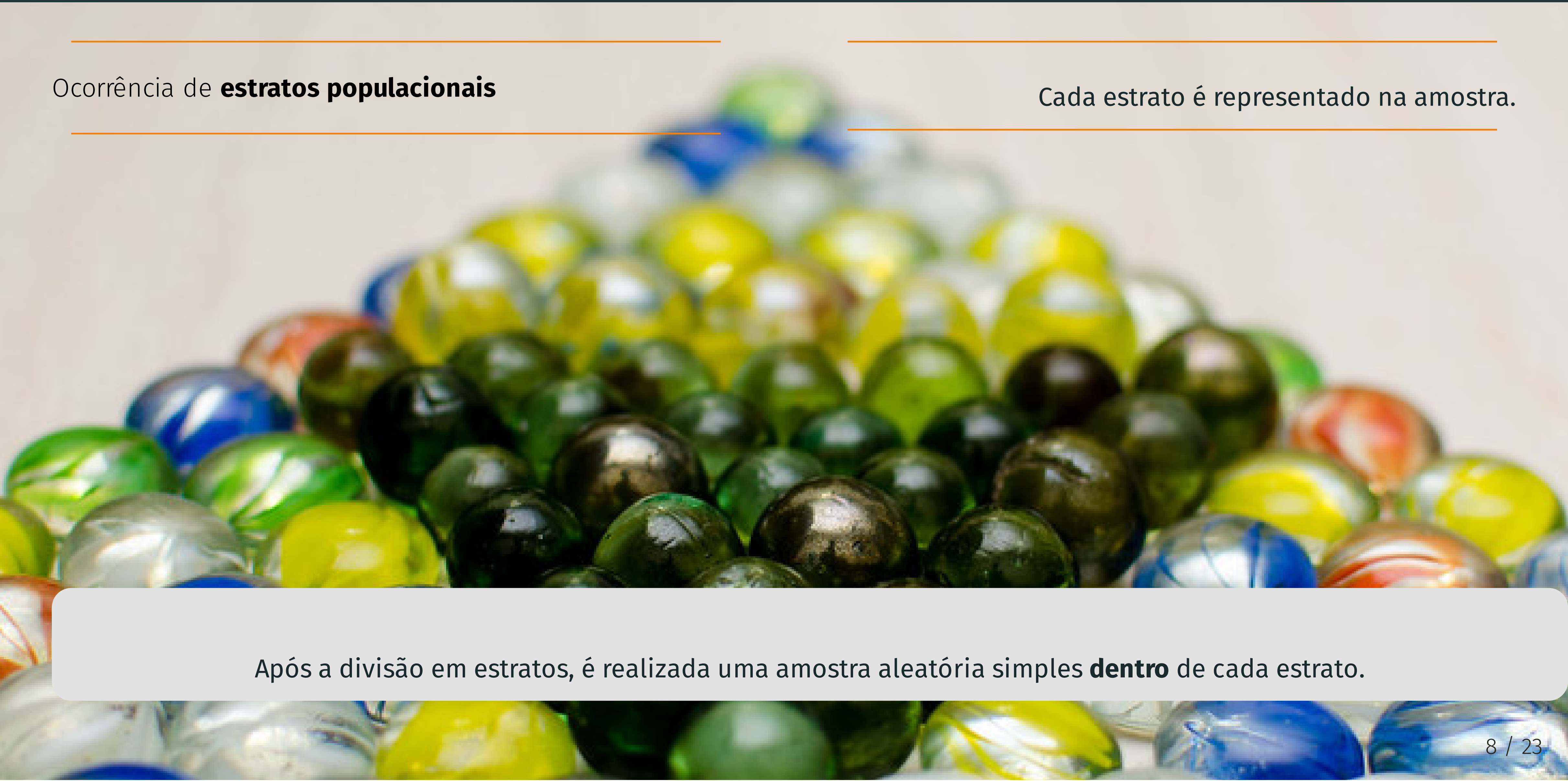
Ocorrência de **estratos populacionais**

Cada estrato é representado na amostra.

2. Amostragem aleatória estratificada

Ocorrência de **estratos populacionais**

Cada estrato é representado na amostra.



Após a divisão em estratos, é realizada uma amostra aleatória simples **dentro** de cada estrato.

2. Amostragem aleatória estratificada

Suponha uma população de 10 elementos composta de dois estratos (A ou B).

X	Estrato
7.3	A
10.6	A
14.8	A
6.6	A
9.8	A
14.4	B
16.1	B
13.3	B
20.0	B
13.6	B

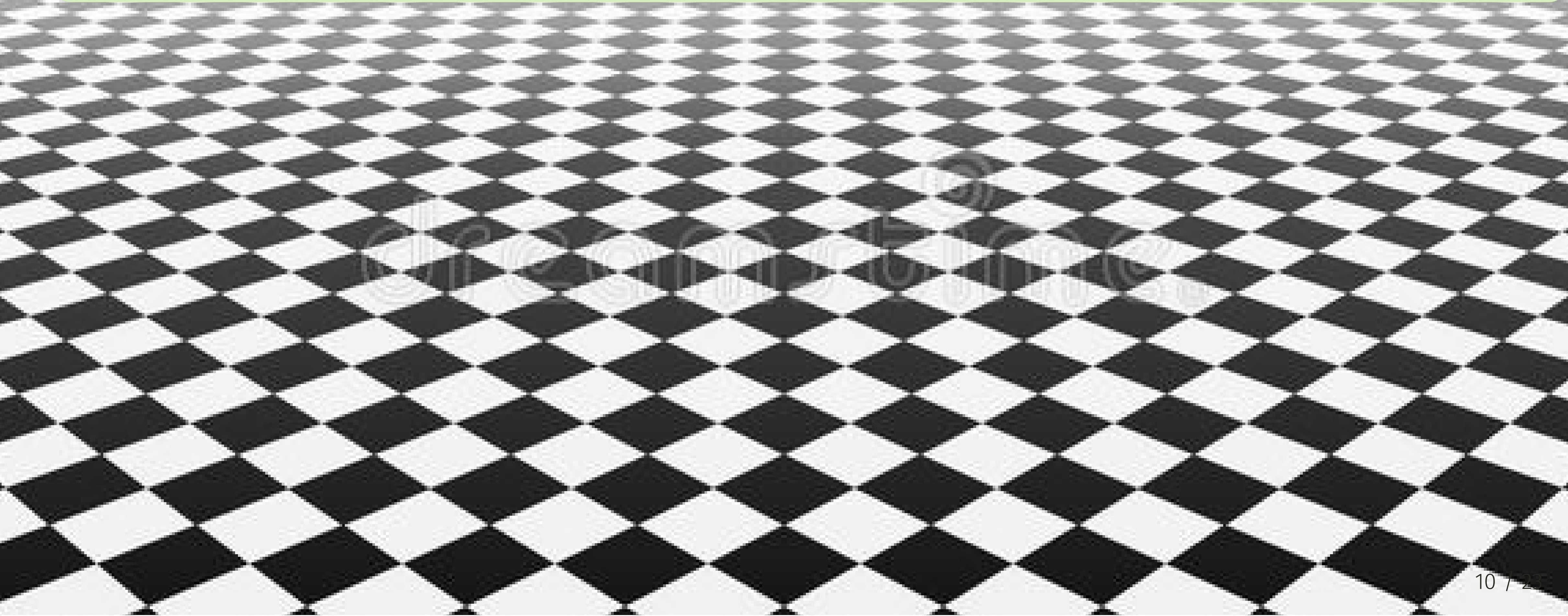
Amostras estratificadas com n = 4 seriam.

Amostra_1	Amostra_2	Amostra_3	Amostra_4
14.8	7.3	7.3	14.8
9.8	6.6	10.6	7.3
14.4	20.0	13.6	13.6
16.1	13.3	16.1	14.4

Ainda que dentro dos blocos ocorra um sorteio aleatório, sempre são sorteados **exatamente** 2 elementos de cada estrato.

3. Amostragem sistemática

Em uma amostragem sistemática as unidades amostrais são **ordenadas** seguindo determinado critério e os elementos amostrados em intervalos regulares.



3. Amostragem sistemática

Implicações da amostragem sistemática na estimativa dos parâmetros populacionais.

POPULAÇÃO

Quando a variável de interesse **não tem relação** com a sequência escolhida, a amostragem sistemática tende a gerar **os mesmos** resultados da amostragem aleatória.

3. Amostragem sistemática

Implicações da amostragem sistemática na estimativa dos parâmetros populacionais.

POPULAÇÃO

Se houver um gradiente justamente na direção do transecto, a variância amostral s^2 irá **superestimar** a variância populacional σ^2.

3. Amostragem sistemática

Implicações da amostragem sistemática na estimativa dos parâmetros populacionais.

POPULAÇÃO

Se houver uma periodicidade que coincida com o intervalo escolhido, a variância amostral s^2 irá **subsetimar** a variância populacional σ^2.

4. Erro amostral, acurácia e precisão

- **Erro amostral (E)**: diferença entre uma estimativa em particular e a média populacional.

$$E = \overline{X} - \mu$$

- **Acurácia**: se refere à proximidade entre o parâmetro e o estimador. Um estimador acurado é, em média, igual ao parâmetro populacional.

$$\mu_{\overline{X}} - \mu$$

- **Precisão**: tem relação com a variabilidade do estimador. Estimadores que geram estimativas similares entre si são mais precisos. A precisão é medida pelo **erro padrão da média**.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

4. Erro amostral

Voltemos à nossa população fictícia com $N = 10$ elementos:

População: 3, 10, 14, 19, 27, 28, 29, 41, 42, 43

A amostra aleatória de tamanho $n = 5$:

Amostra 1: $41, 14, 42, 29, 19$

Tem média:

$\overline{X}_1 = \frac{41+14+42+29+19}{5} = 29$

E **erro amostral**:

$E_1 = 29 - 25.6 = 3.4$

Outra amostra aleatória de tamanho $n = 5$:

Amostra 2: $27, 29, 19, 10, 14$

Tem média:

$\overline{X}_2 = 19.8$

E **erro amostral**:

$E_2 = 19.8 - 25.6 = -5.8$

O erro amostral mede diferença entre a estimativa obtida de uma amostra *particular* e a média populacional.

4. Acurácia

Existem:

$$\binom{10}{5} = \frac{10!}{(10-5)! \times 5!} = 252$$

formas diferentes de tomarmos uma amostra de tamanho $n = 5$ de nossa população de tamanho $N = 10$.

Se tomadas 8 destas amostras veremos que as médias amostrais $\overline{X}$ diferementre si.

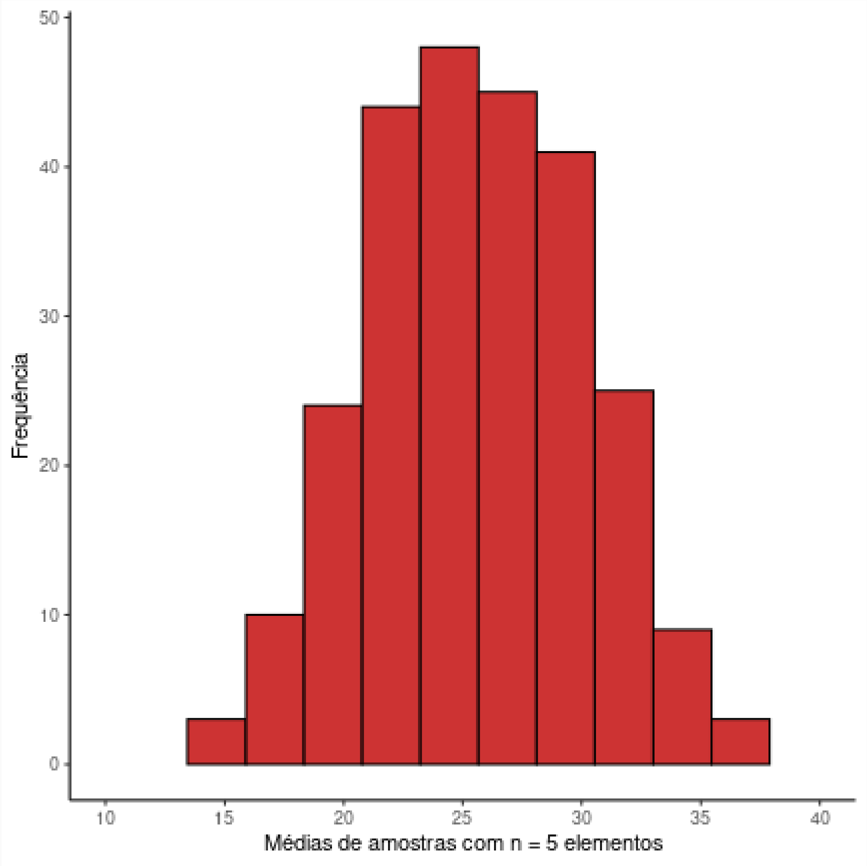
	A 1	A 2	A 3	A 4	A 5	A 6	A 7	A 8
	19.0	29	10.0	19.0	42.0	29.0	28	43.0
	29.0	43	14.0	41.0	29.0	10.0	42	28.0
	10.0	28	42.0	43.0	41.0	27.0	10	42.0
	42.0	3	28.0	28.0	14.0	42.0	43	27.0
	43.0	27	29.0	3.0	28.0	43.0	27	19.0
Medias	28.6	26	24.6	26.8	30.8	30.2	30	31.8

4. Acurácia

Se tomarmos TODAS as 252 amostras possíveis e calcularmos as respectivas médias teremos:

14.6	14.8	15.0	17.4	17.6	17.8	16.4	16.6	19.0	19.2	19.4	16.8	19.2	19.4	19.6	19.4	19.6	19.8	22.0	22.2	22.4
17.4	17.6	20.0	20.2	20.4	17.8	20.2	20.4	20.6	20.4	20.6	20.8	23.0	23.2	23.4	19.4	21.8	22.0	22.2	22.0	22.2
22.4	24.6	24.8	25.0	22.2	22.4	22.6	24.8	25.0	25.2	25.0	25.2	25.4	27.8	18.2	18.4	20.8	21.0	21.2	18.6	21.0
21.2	21.4	21.2	21.4	21.6	23.8	24.0	24.2	20.2	22.6	22.8	23.0	22.8	23.0	23.2	25.4	25.6	25.8	23.0	23.2	23.4
25.6	25.8	26.0	25.8	26.0	26.2	28.6	21.2	23.6	23.8	24.0	23.8	24.0	24.2	26.4	26.6	26.8	24.0	24.2	24.4	26.6
26.8	27.0	26.8	27.0	27.2	29.6	25.6	25.8	26.0	28.2	28.4	28.6	28.4	28.6	28.8	31.2	28.6	28.8	29.0	31.4	31.6
19.6	19.8	22.2	22.4	22.6	20.0	22.4	22.6	22.8	22.6	22.8	23.0	25.2	25.4	25.6	21.6	24.0	24.2	24.4	24.2	24.4
24.6	26.8	27.0	27.2	24.4	24.6	24.8	27.0	27.2	27.4	27.2	27.4	27.6	30.0	22.6	25.0	25.2	25.4	25.2	25.4	25.6
27.8	28.0	28.2	25.4	25.6	25.8	28.0	28.2	28.4	28.2	28.4	28.6	31.0	27.0	27.2	27.4	29.6	29.8	30.0	29.8	30.0
30.2	32.6	30.0	30.2	30.4	32.8	33.0	23.4	25.8	26.0	26.2	26.0	26.2	26.4	28.6	28.8	29.0	26.2	26.4	26.6	28.8
29.0	29.2	29.0	29.2	29.4	31.8	27.8	28.0	28.2	30.4	30.6	30.8	30.6	30.8	31.0	33.4	30.8	31.0	31.2	33.6	33.8
28.8	29.0	29.2	31.4	31.6	31.8	31.6	31.8	32.0	34.4	31.8	32.0	32.2	34.6	34.8	33.4	33.6	33.8	36.2	36.4	36.6

Que descrevem o seguinte padrão:



- Média populacional

$$\mu = \sum_{i=1}^{10} \frac{3, 10, 14, 19, 27, 28, 29, 41, 42, 43}{10} = 25.6$$

- Média das $N = 252$ médias amostrais com $n = 5$

$$\mu_{\overline{X}} = \sum_{i=1}^{N} \frac{\overline{X}_i}{252} = 25.6$$

Verificamos que o estimador é acurado pois:

$$\mu_{\overline{X}} - \mu = 25.6 - 25.6 = 0$$

4. Precisão: o erro padrão da média - $\sigma_{\overline{X}}$

População: 3, 10, 14, 19, 27, 28, 29, 41, 42, 43

Suponha agora que tomemos ao acaso amostras com $n = 7$ desta mesma população.

Existem ao todo:

$$\binom{10}{7} = \frac{10!}{(10-7)! \times 7!} = 120$$

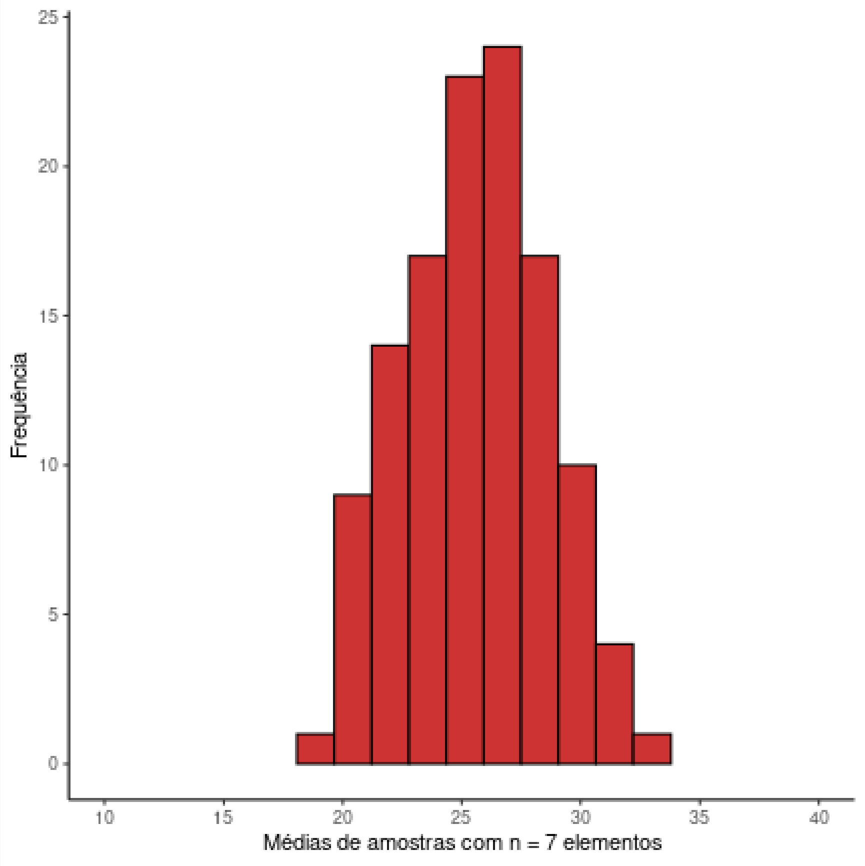
amostras diferentes de tamanho $n = 7$ que podem ser retiradas de uma população de tamanho $N = 10$.

4. Precisão: o erro padrão da média - $\sigma_{\overline{X}}$

Se tomarmos estas 120 amostras e calcularmos suas respectivas médias amostrais, teremos:

18.6	20.3	20.4	20.6	20.4	20.6	20.7	22.3	22.4	22.6	20.6	20.7
20.9	22.4	22.6	22.7	22.6	22.7	22.9	24.6	21.7	21.9	22.0	23.6
23.7	23.9	23.7	23.9	24.0	25.7	23.9	24.0	24.1	25.9	26.0	22.4
22.6	22.7	24.3	24.4	24.6	24.4	24.6	24.7	26.4	24.6	24.7	24.9
26.6	26.7	25.7	25.9	26.0	27.7	27.9	28.0	23.0	23.1	23.3	24.9
25.0	25.1	25.0	25.1	25.3	27.0	25.1	25.3	25.4	27.1	27.3	26.3
26.4	26.6	28.3	28.4	28.6	27.0	27.1	27.3	29.0	29.1	29.3	30.4
24.0	24.1	24.3	25.9	26.0	26.1	26.0	26.1	26.3	28.0	26.1	26.3
26.4	28.1	28.3	27.3	27.4	27.6	29.3	29.4	29.6	28.0	28.1	28.3
30.0	30.1	30.3	31.4	28.6	28.7	28.9	30.6	30.7	30.9	32.0	32.7

4. Precisão: o erro padrão da média - $\sigma_{\overline{X}}$



- Média populacional

$$\mu = \sum_{i=1}^{10} \frac{3, 10, 14, 19, 27, 28, 29, 41, 42, 43}{10} = 25.6$$

- Média das 120 médias amostrais com $n = 7$

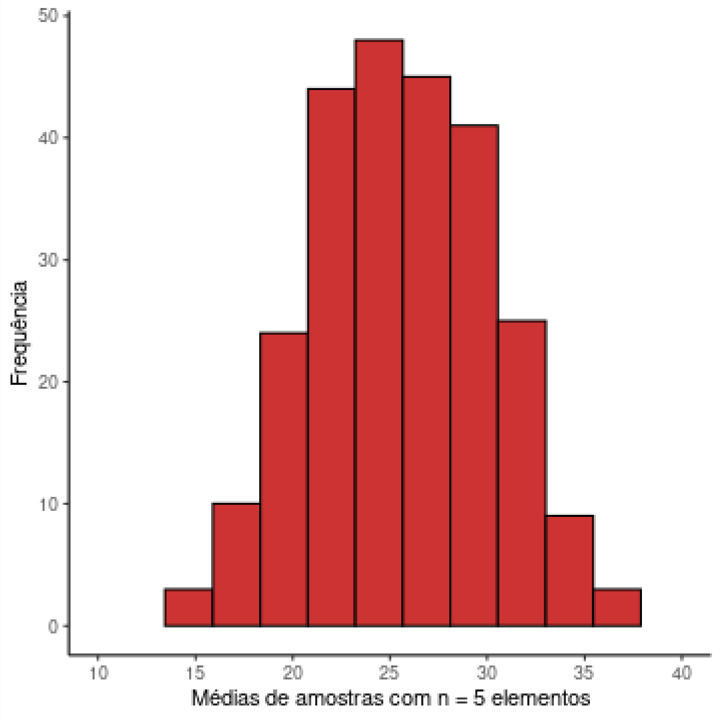
$$\mu_{\overline{X}} = \sum_{i=1}^{N} \frac{\overline{X}_i}{120} = 25.6$$

Como no exemplo anterior, verificamos que o estimador é acurado pois:

$$\mu_{\overline{X}} - \mu = 25.6 - 25.6 = 0$$

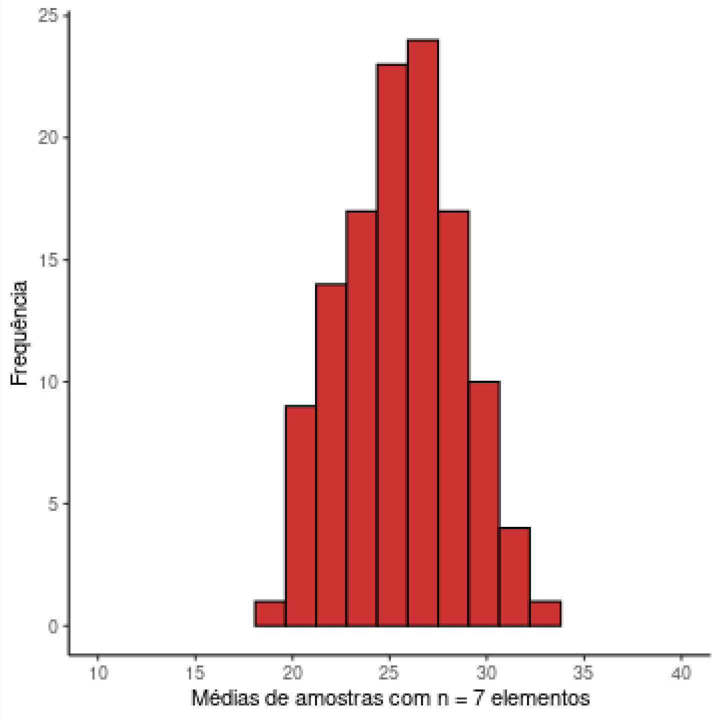
4. Precisão: o erro padrão da média - $\sigma_{\overline{X}}$

Distribuição das médias amostrais para n = 5



$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{13.27}{\sqrt{5}} = 5.934$$

Distribuição das médias amostrais para n = 7



$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{13.27}{\sqrt{7}} = 5.015$$

4. Precisão: o erro padrão da média - $\sigma_{\overline{X}}$

Na prática científica não conhecemos o desvio padrão populacional σ e, consequentemente, não temos obter o erro padrão populacional $\sigma_{\overline{X}}$. No entanto, dado que temos uma amostra particular, podemos **estimá-lo** a partir do desvio padrão amostral s pela expressão:

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

em que $s_{\overline{X}}$ é denominado de **erro padrão amostral**