

Amostragem e Delineamento


Amostrando uma população estatística

Fabio Cop (fabiocopf@gmail.com)

Instituto do Mar - UNIFESP

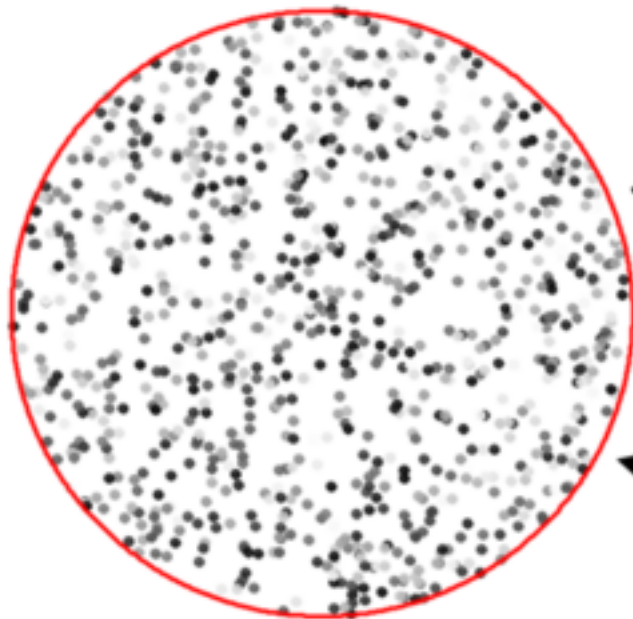
Última atualização em 25 de abril de 2022

Conteúdo da aula

- 
- The background of the slide is a close-up photograph of several pushpins with black and orange heads pinned to a white surface. The pushpins are scattered across the frame, with some in sharp focus and others blurred in the background.
1. Amostragem aleatória simples
 2. Amostragem aleatória estratificada
 3. Amostragem sistemática
 4. Erro amostral, acurácia e precisão
 5. Efeitos dos tipos de amostragem sobre as estimativas

O processo de amostragem e inferência sobre uma população estatística

População estatística



Amostragem

Amostra



Inferência

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$\mu = ?$ (média Populacional)

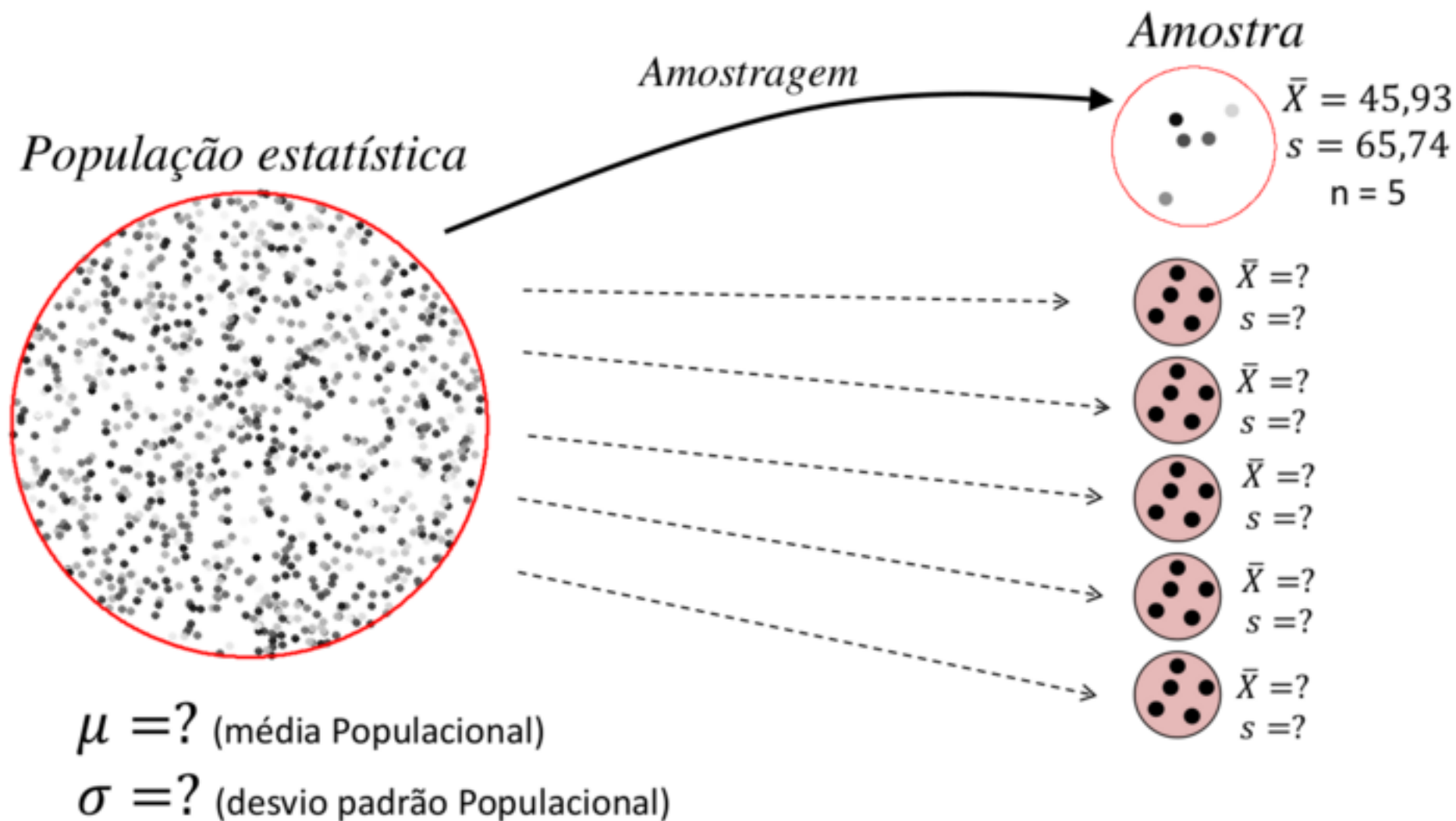
$\sigma = ?$ (desvio padrão Populacional)

Parâmetros populacionais

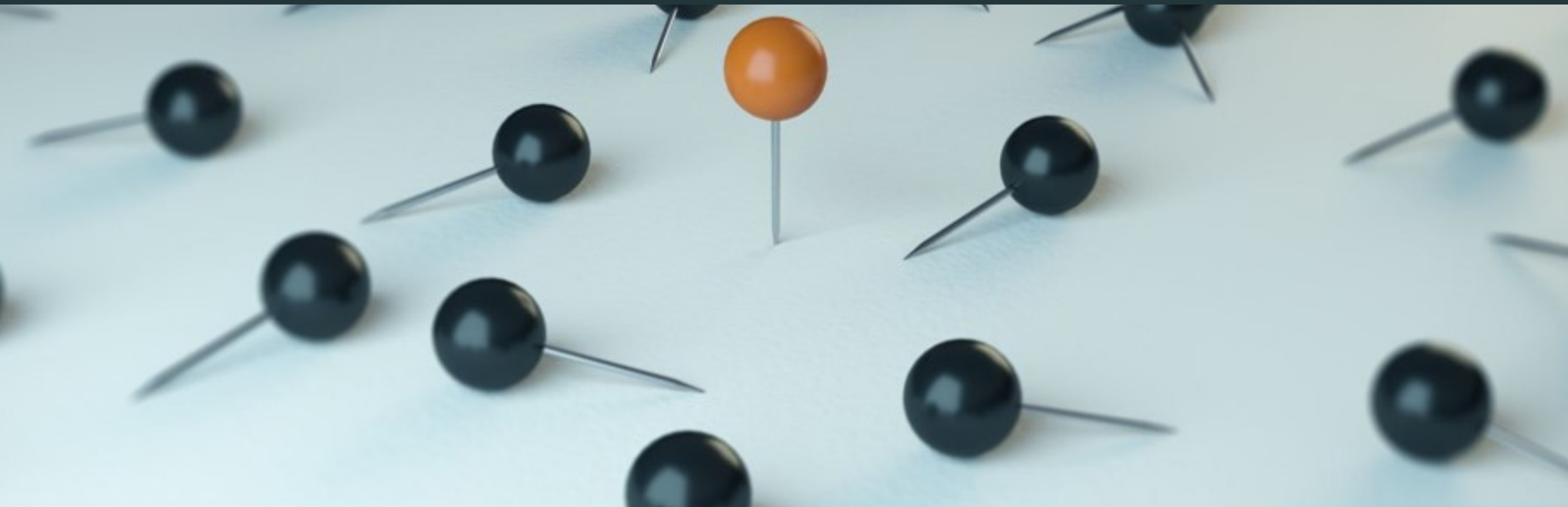
$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

$n = 5$

O processo de amostragem e inferência sobre uma população estatística



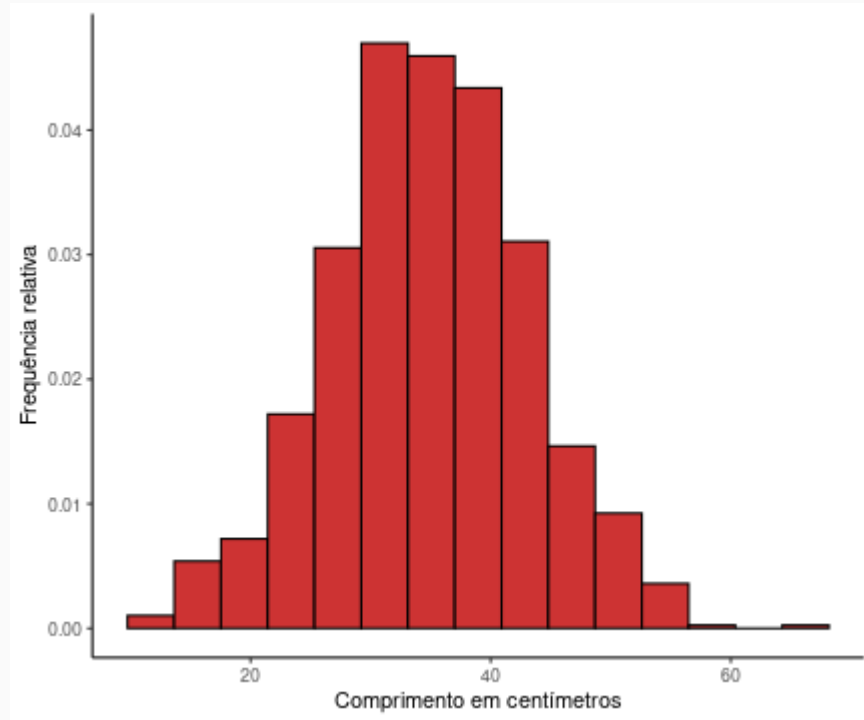
1. Amostragem aleatória simples



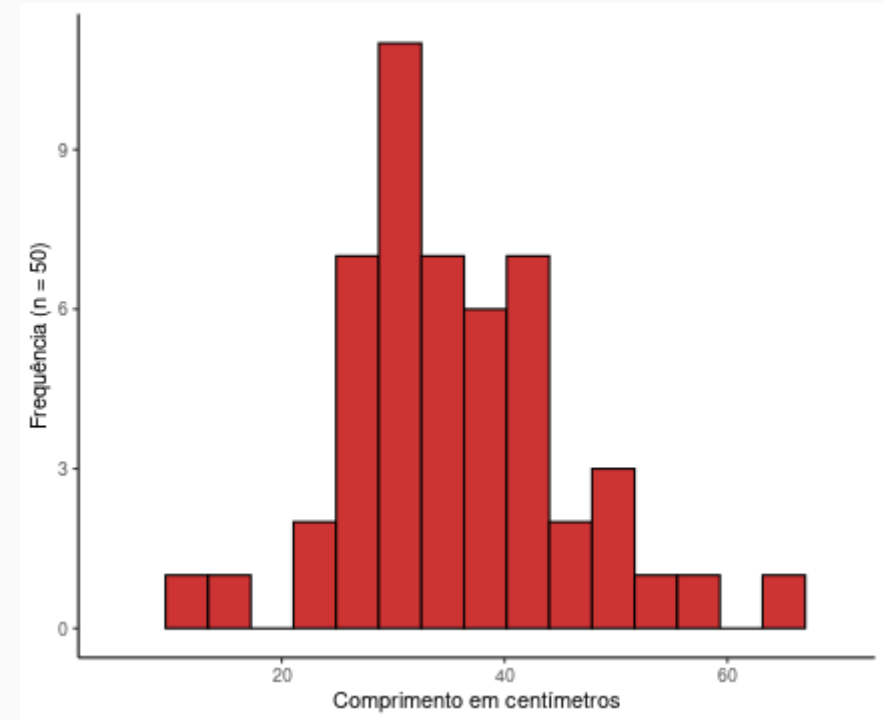
Na amostragem aleatória simples, cada elemento da população tem a **mesma probabilidade** de compor uma amostra. Se a população tem N elementos, cada um tem probabilidade $\frac{1}{N}$ de ser selecionado.

1. Amostragem aleatória simples

População estatística



Amostra



Se não sabemos nada sobre a distribuição de frequências da população estatística, a amostra aleatória simples é a melhor forma de **aproximar** esta distribuição.

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: $n = 5$

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: $n = 5$

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

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

Ainda que nos dois casos acima tenhamos conduzido um *sorteio aleatório*, 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

POPULAÇÃO ESTATÍSTICA



A amostragem sistemática tende a gerar **os mesmos** resultados da amostragem aleatória, exceto em situações muito particulares.

3. Amostragem sistemática

POPULAÇÃO ESTATÍSTICA



Se houver uma periodicidade que coincida com o intervalo escolhido, a média amostral **não será** igual à média populacional e a variância amostral s^2 irá **subestimar** a variância populacional σ^2 .

4. Erro amostral, acurácia e precisão

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

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

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

$$\mu_{\bar{X}} - \mu = 0$$

- **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**.

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

4. Erro amostral

Voltemos à nossa população fictícia com $N = 10$ elementos e $\mu = 25.6$:

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 $n = 5$ **sem reposição** de nossa população de tamanho $N = 10$.

Se tomadas 8 destas amostras veremos que as médias amostrais \bar{X} diferem entre 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

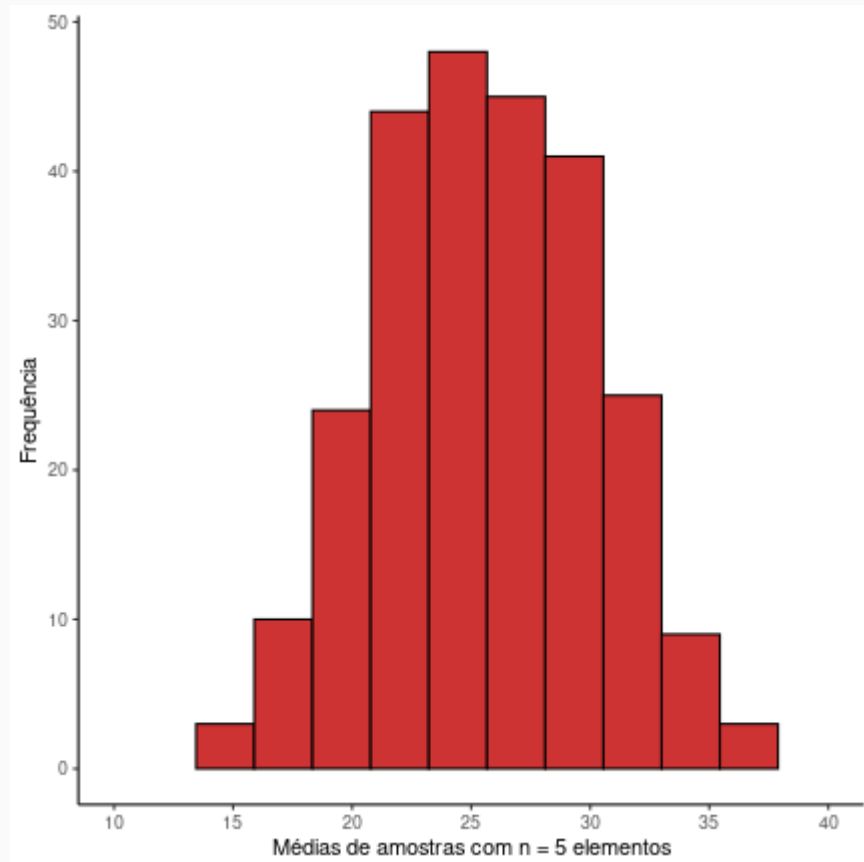
4. Acurácia

Se tomarmos TODAS as **252** amostras possíveis e calcularmos as respectivas médias teremos os respectivos \overline{X} :

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

4. Acurácia

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 $n = 5$

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

Verificamos que o estimador é acurado pois:

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

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

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:

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

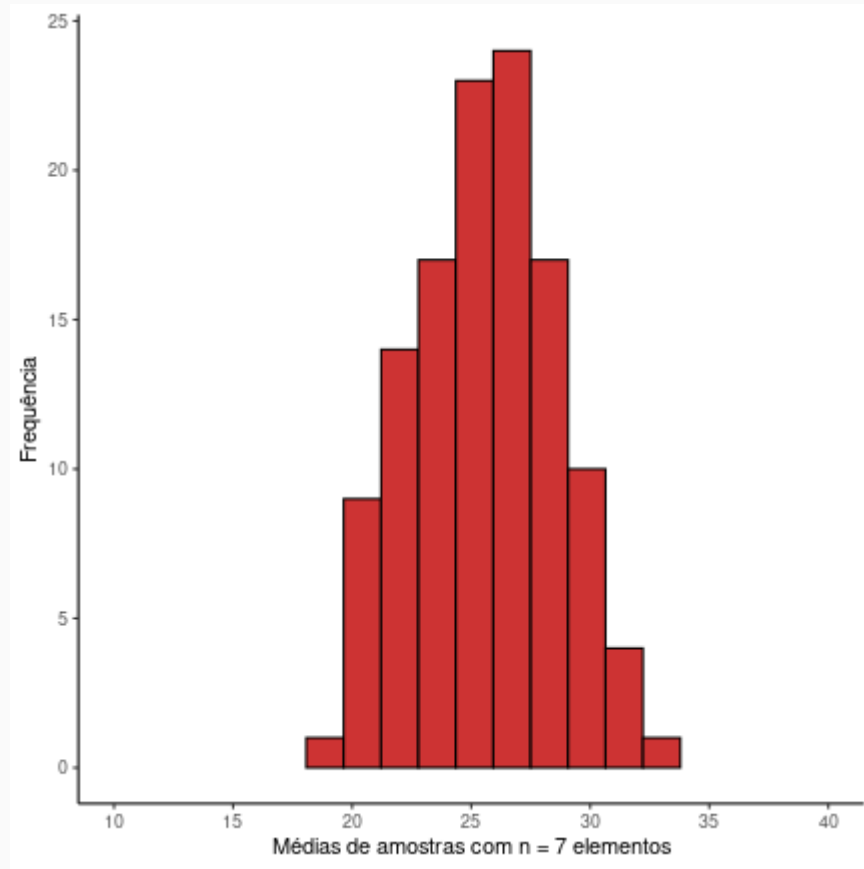
amostras diferentes de tamanho $n = 7$ que podem ser retiradas **sem reposição** de uma população de tamanho $N = 10$.

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

Se tomarmos estas **120** amostras e calcularmos suas respectivas médias amostrais, teremos os respectivos \bar{X} :

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

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



- 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 $n = 7$

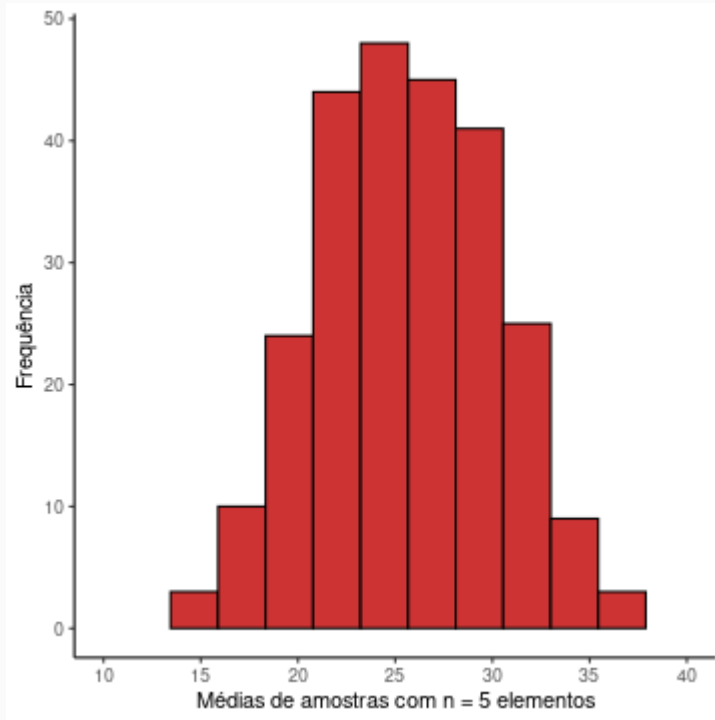
$$\mu_{\bar{X}} = \sum_{i=1}^{120} \frac{\bar{X}_i}{120} = 25.6$$

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

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

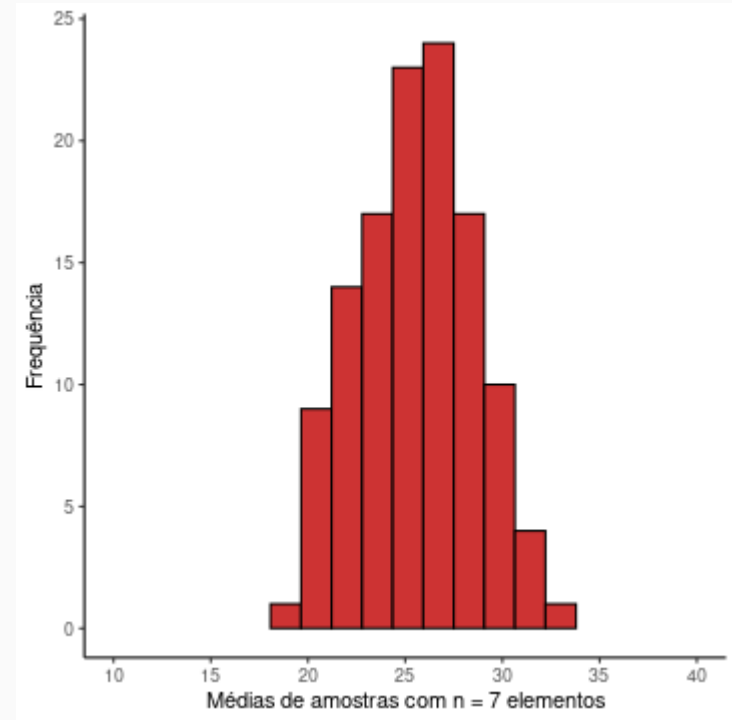
4. Precisão: o erro padrão da média - $\sigma_{\bar{X}}$

Distribuição das médias amostrais para n = 5



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

Distribuição das médias amostrais para n = 7



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

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

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

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

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

5. Efeitos dos tipos de amostragem sobre as estimativas

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

Se os estratos são identificados corretamente, a amostragem estratificada irá gerar estimativas **mais precisas**, uma vez que evita casos extremos como por exemplo, a seleção puramente ao acaso de um único estrato.

Amostragem Aleatória Estratificada: $n = 4$

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

Amostragem Aleatória Simples: $n = 4$

Amostra_1	Amostra_2	Amostra_3	Amostra_4
14.4	14.8	16.1	7.3
10.6	6.6	20	10.6
16.1	13.6	9.8	13.3
20	7.3	13.6	6.6

5. Efeitos dos tipos de amostragem sobre as estimativas

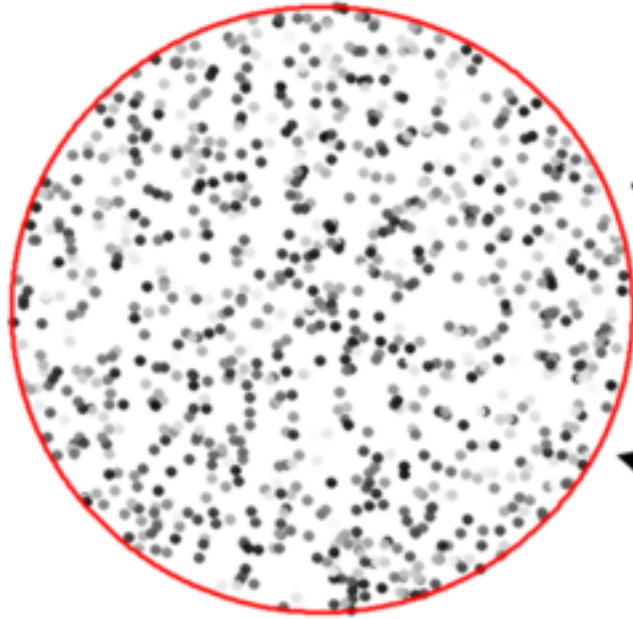
POPULAÇÃO



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

5. Efeitos dos tipos de amostragem sobre as estimativas

População estatística



Amostragem

Amostra



Inferência

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$\mu = ?$ (média Populacional)

$\sigma = ?$ (desvio padrão Populacional)

Parâmetros populacionais

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

$n = 5$