

População

é o conjunto de indivíduos, ou objetos, tendo pelo menos uma variável em comum observável.

Unidade Amostral

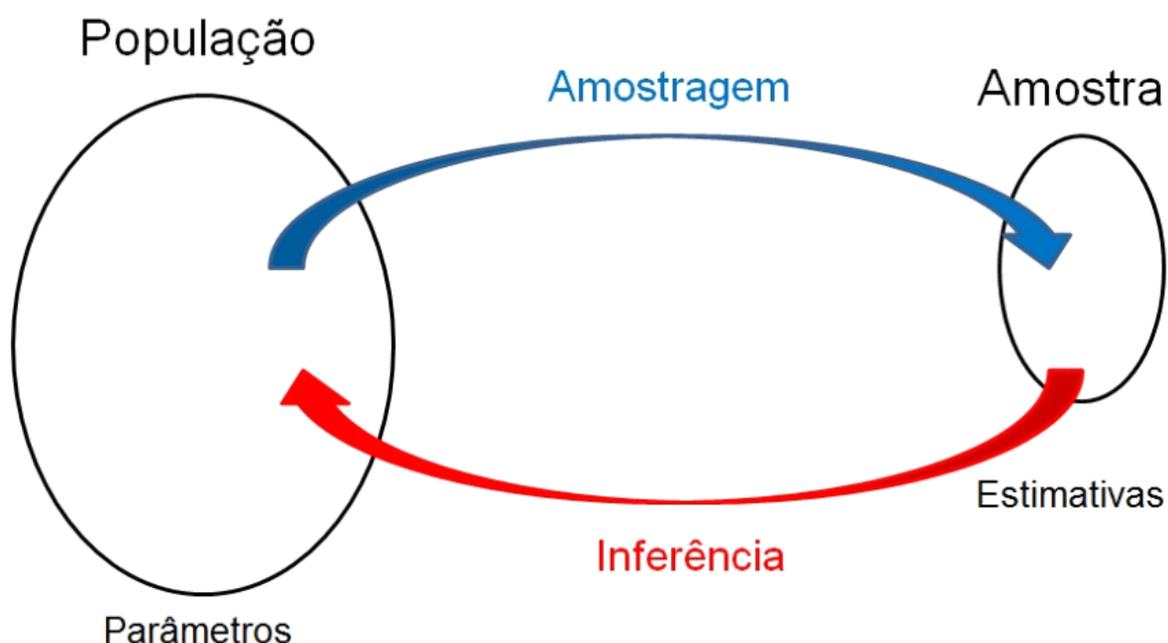
um item individual da população.

Amostra

é qualquer subconjunto de elementos da população. Em geral, no entanto, trabalhamos apenas com amostras aleatórias, onde todos os elementos da população têm a mesma probabilidade de serem escolhidos.

Variável

é a característica de interesse que é medida/observada em cada unidade amostral.



Exemplo: O objetivo de um estudo é determinar a taxa média de ocorrência de determinada praga numa fazenda produtora de laranjas.

População: *plantação de laranjas da fazenda*

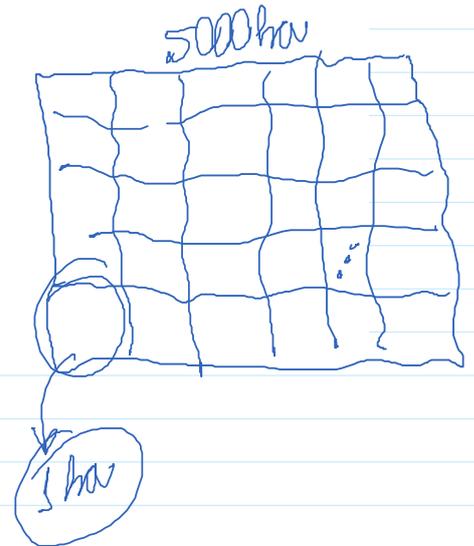
Unidade Amostral: *cada árvore de laranja / 1 ha da fazenda*

Variável: *modo de estar doente / número de frutos danificados / ...*

População: plantas em um campo no jardim
 Unidade Amostral: cada pé de laranja / 1 ha da fazenda
 Variável: nº de frutos de cada pé / proporção de árvores de cada
 Amostra: grupo de árvores escolhida aleatoriamente / unidades de 1 ha cada selecionado aleatoriamente
 Parâmetro: proporção (π), média (μ)
 Estatística: p, \bar{x}

Exemplo: Para saber a produção (madeira para serraria) de uma Floresta Nativa de 5000 ha um Engenheiro Agrônomo mediu o volume de madeira (m^3/ha) em 52 parcelas de 1ha.

População: floresta de 5000 ha
 Unidade Amostral: 1 parcela de 1 ha
 Variável: volume madeira (m^3/ha)
 Amostra: 52 parcelas de 1 ha
 Parâmetro: π, μ
 Estatística: p, \bar{x}



Exemplo: Estudar o crescimento de microrganismos em um alimento.

População: as refeições preparadas em um restaurante
Unidade amostral: o alimento
Variável: taxa de crescimento de microrganismos
Amostra: uma porção de cada alimento preparado
Parâmetro: proporção (π), quantidade de microrganismos encontrados
Estatística: proporção, média, quantidade total

Amostragem

é o processo da escolha da amostra. Consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.

Quando uma amostra é boa?

- Viés...
- Precisão...

Viés: tendenciosidade presente no processo de amostragem que impede que a amostra seja perfeitamente representativa da população, independentemente do tamanho da amostra.

- **Viés de seleção:** Tendência sistemática de subrepresentar ou excluir da amostra um certo tipo de unidade;
- **Viés de não resposta:** Numa pesquisa de opinião, o resultado de pessoas não responderem às questões do entrevistador;
- **Viés de resposta:** Numa pesquisa de opinião, a forma de apresentação das questões e a atitude do entrevistador podem influenciar a resposta.

Exemplo: Estudando a influência do sistema de amostragem na estrutura de tamanho das árvores de uma floresta tropical, um Ecologista tomou 10 Amostras compostas de 5 árvores, 10 amostras compostas de 20 árvores e 10 amostras compostas por 40 árvores.

- Qual o efeito do tamanho da amostra (número de árvores) nos resultados obtidos?
- Qual tipo de amostra se mostra menos variável?

Tabela: Média do diâmetro de árvores amostradas numa floresta nativa

Amostras	Tamanho das amostras (número de árvores)		
	5	20	40
1	26,6	34,1	30,5
2	32,9	31,2	29,8
3	39,2	40,1	37,3
4	30,3	30,4	29,4
5	31,8	29,7	32,7
6	48,0	38,8	31,5
7	37,7	37,3	29,3
8	20,0	35,6	30,6
9	31,8	37,7	31,2
10	33,2	26,7	29,3
Desvio padrão	7,48	4,48	2,42
CV(%)	22,55	13,13	7,75

As amostras de tamanho maior

Precisão: capacidade dos sistemas de amostragem de gerar Estatísticas com pequena variabilidade entre as amostras (de mesmo tamanho) tomadas numa mesma população.



Sem Viés
Sem Precisão



Com Viés
Com Precisão



Sem Viés
Com Precisão

Exemplo: Uma equipe de cientistas deseja monitorar as mudanças ecológicas numa floresta após a fragmentação desta devido a conversão da área florestal para área agrícola. A equipe se encontra dividida entre duas alternativas de amostragem:

- Monitorar um grande número de fragmentos que sejam de fácil acesso e que representem o atual estado da região;
- Monitorar um número pequeno de fragmentos que sejam uma amostra representativa da floresta original, independentemente da facilidade de acesso.

Quais as vantagens e desvantagens das duas alternativas?

Exemplo: Um Engenheiro Agrônomo dispõe de dois instrumentos para medir a altura das árvores. O instrumento A tem alta precisão, mas apresenta um viés sistemático que independe da altura da árvore sendo medida. O instrumento B tem precisão inferior ao instrumento A, mas não apresenta qualquer tipo de viés. Este Engenheiro foi chamado para realizar medições em duas situações:

- Experimento que compara diferentes híbridos de *Pinus caribaea* - *Pinus elliottii*;
- Levantamento do volume de madeira em pé numa floresta a ser vendida.

Qual dos dois instrumentos deve ser utilizado em cada situação? Por quê?

B - apesar da precisão menor estamos medindo o

B - apesar da precisão menor estamos usando o mesmo instrumento para comparar os diferentes híbridos.

Alguns métodos mais comuns de extração da amostra:

- Amostragem aleatória simples
- Amostragem estratificada
- Amostragem sistemática
- Amostragem por conglomerados

Amostragem Aleatória Simples

Amostragem aleatória simples (AAS)

é um método de selecionar, sem reposição, n elementos de uma população de tamanho N , conhecido e finito, onde cada elemento tem a mesma chance de ser selecionado.

Procedimento para a obtenção de uma amostra aleatória simples de tamanho n :

- 1 Enumerar os N elementos da população de 1 a N .
- 2 Sortear, sem reposição, n números compreendidos de 1 a N .
- 3 Os elementos correspondentes aos números escolhidos formarão a amostra.

Utilizando a tabela de números aleatórios numa amostragem aleatória simples

- 1 Enumerar os N elementos da população da seguinte forma:
 - Se o número de dígitos de N for igual a 2, utilize: 01, 02, 03, ..., N
 - Se o número de dígitos de N for igual a 3, utilize: 001, 002, 003, ..., N
 - Se o número de dígitos de N for igual a 4, utilize: 0001, 0002, 0003, ..., N
 - ...
- 2 Iniciar em um ponto qualquer da tabela de números aleatórios
- 3 Ler os números sistematicamente seguindo as linhas ou colunas da Tabela
- 4 Se um número já lido surgir novamente ou se aparecer um número que não está presente na enumeração dos elementos da população, ignorá-lo e seguir para o próximo
- 5 Continuar até que n unidades sejam selecionadas.

Exemplo:

Selecionar uma amostra de tamanho 10, de 50 unidades da população.

Tabela 1 Números aleatórios

59 58 48 36 47	92 85 05 08 65	47 49 10 41 05	10 75 59 75 99	17 28 97 99 75
53 26 21 50 21	37 93 85 52 86	86 22 75 34 37	69 85 25 03 78	50 26 18 25 10
07 02 16 58 67	05 32 93 87 84	31 30 62 78 60	59 90 24 22 07	74 43 43 56 91
92 87 67 56 36	58 58 16 88 16	17 83 52 09 99	86 17 20 95 93	01 46 77 18 11
90 57 05 58 96	84 33 68 15 87	28 18 08 76 89	94 60 94 48 76	92 93 49 13 91
24 26 56 02 33	33 21 75 54 04	96 28 85 78 11	54 01 92 86 36	65 19 45 97 79
20 09 49 50 27	33 86 85 59 39	02 25 60 56 26	01 11 24 44 15	58 00 54 54 09
22 74 50 39 12	83 91 03 38 78	85 56 78 41 44	26 04 12 13 50	38 15 61 02 51
10 45 36 09 86	07 68 31 98 41	98 17 56 93 84	16 01 48 99 36	44 61 71 69 67
09 82 11 18 29	96 19 12 47 26	26 01 14 78 55	33 11 13 56 95	68 66 57 90 33
04 63 02 45 50	61 91 02 14 07	57 36 29 12 74	89 47 84 89 69	13 85 22 66 83
55 93 05 63 30	40 05 51 03 31	68 15 33 85 87	94 80 24 96 62	31 38 95 35 38
66 15 07 64 38	16 44 52 26 42	34 65 99 71 63	87 22 04 62 15	76 94 00 00 77
96 31 72 41 94	47 03 44 73 77	96 17 02 97 50	26 67 60 63 57	66 81 92 03 20
07 10 58 83 63	35 47 34 05 38	92 26 05 33 40	91 23 43 68 72	29 74 60 67 01
04 47 64 02 49	10 52 21 00 80	40 56 68 97 32	43 46 70 65 08	96 52 25 29 44
56 24 53 31 96	65 42 53 27 78	23 30 61 34 18	56 59 23 69 27	83 66 50 03 12
98 15 27 91 71	24 15 28 61 91	83 49 05 82 54	53 59 30 25 19	36 31 31 56 58
36 96 23 77 26	79 74 28 12 16	08 88 07 28 71	45 43 40 07 66	11 26 38 51 87
66 01 53 03 67	92 27 27 17 54	31 23 30 42 83	85 78 21 68 34	86 33 77 84 40
48 07 09 48 65	92 33 41 97 63	48 97 19 86 81	10 85 42 84 49	03 82 01 82 88
95 44 86 84 32	09 03 56 46 96	64 51 33 75 10	29 00 99 23 82	92 31 77 08 17
91 73 15 42 46	72 21 07 34 11	92 70 89 58 54	11 30 93 38 29	00 53 93 14 09
08 35 79 86 83	06 89 37 82 12	81 14 08 82 04	91 88 04 86 36	18 10 09 78 99
37 20 97 09 96	86 34 77 09 31	04 38 18 79 61	68 66 47 40 35	40 16 50 22 54

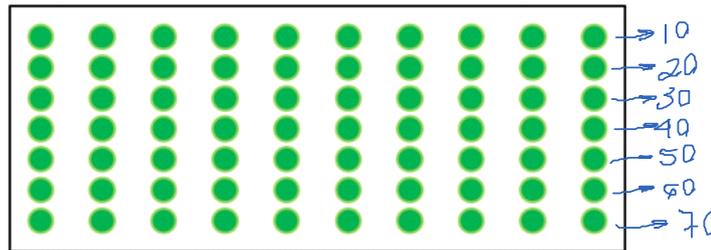
A amostra será composta pelos indivíduos:

05 09 10 21 22 25 29 45 48

Seja uma pesquisa sobre a ocorrência de uma determinada doença em uma plantação de laranjas. Obter uma amostra aleatória simples de tamanho

Seja uma pesquisa sobre a ocorrência de uma determinada doença em uma plantação de laranjas. Obter uma amostra aleatória simples de tamanho $n = 10$ usando a Tabela de Números aleatórios.

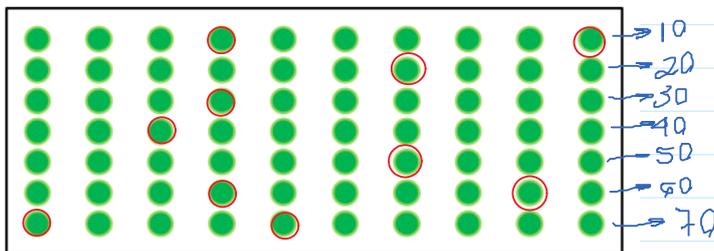
$N = 1 \dots 70$



1ª nº de verde selecionado

Tabela 1 Números aleatórios

59	58	48	36	47	92	85	05	08	65	47	49	10	41	05	10	75	59	75	99	17	28	97	99	75
53	26	21	50	21	37	93	85	52	86	86	22	75	34	37	69	85	25	03	78	50	26	18	25	10
07	02	16	58	67	05	32	93	87	84	31	30	62	78	60	59	90	24	22	07	74	43	43	56	91
92	87	67	56	36	58	58	16	88	16	17	83	52	09	99	86	17	20	95	93	01	46	77	18	11
90	57	05	58	96	84	33	68	15	87	28	18	08	76	89	94	60	94	48	76	92	93	49	13	91
24	26	56	02	33	33	21	75	54	04	96	28	85	78	11	54	01	92	86	36	65	19	45	97	79
20	09	49	50	27	33	86	85	59	39	02	25	60	56	26	01	11	24	44	15	58	00	54	54	09
22	74	50	39	12	83	91	03	38	78	85	56	78	41	44	26	04	12	13	50	38	15	61	02	51
10	45	36	09	86	07	68	31	98	41	98	17	56	93	84	16	01	48	99	36	44	61	71	69	67
09	82	11	18	29	96	19	12	47	26	26	01	14	78	55	33	11	13	56	95	68	66	57	90	33
04	63	02	45	50	61	91	02	14	07	57	36	29	12	74	89	47	84	89	69	13	85	22	66	83
55	93	05	63	30	40	05	51	03	31	68	15	33	85	87	94	80	24	96	62	31	38	95	35	38
66	15	07	64	38	16	44	52	26	42	34	65	99	71	63	87	22	04	62	15	76	94	00	00	77
96	31	72	41	94	47	03	44	73	77	96	17	02	97	50	26	67	60	63	57	66	81	92	03	20
07	10	58	83	63	35	47	34	05	38	92	26	05	33	40	91	23	43	68	72	29	74	60	67	01
04	47	64	02	49	10	52	21	00	80	40	56	68	97	32	43	46	70	65	08	96	52	25	29	44
56	24	53	31	96	65	42	53	27	78	23	30	61	34	18	56	59	23	69	27	83	66	60	03	12
98	15	27	91	71	24	15	28	61	91	83	49	05	82	54	53	59	30	25	19	36	31	31	56	58
36	96	23	77	26	79	74	28	12	16	08	88	07	28	71	45	43	40	07	66	11	26	38	51	87
66	01	53	03	67	92	27	27	17	54	31	23	30	42	83	85	78	21	68	34	86	33	77	84	40
48	07	09	48	65	92	33	41	97	63	48	97	19	86	81	10	85	42	84	49	03	82	01	82	88
95	44	86	84	32	09	03	56	46	96	64	51	33	75	10	29	00	99	23	82	92	31	77	08	17
91	73	15	42	46	72	21	07	34	11	92	70	89	58	54	11	30	93	38	29	00	53	93	14	09
08	35	79	86	83	06	89	37	82	12	81	14	08	82	04	91	88	04	86	36	18	10	09	78	99
37	20	97	09	96	86	34	77	09	31	04	38	18	79	61	68	66	47	40	35	40	16	50	22	54



Código em R:

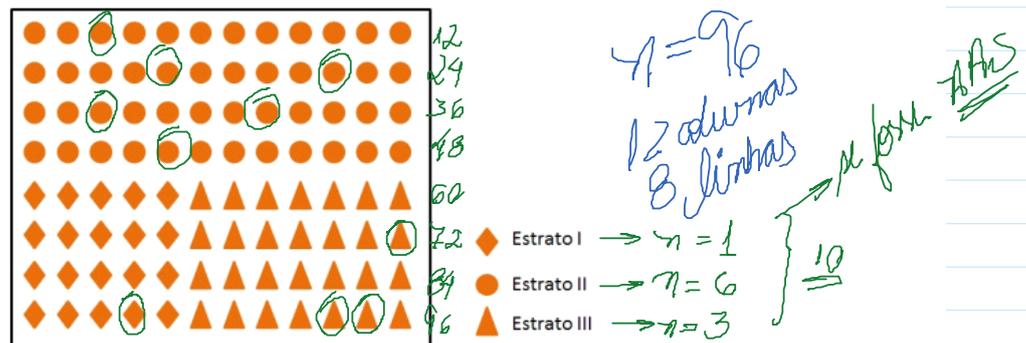
`sample(seq(1,70,1),10)`

\rightarrow nº inicial
 \rightarrow nº final
 \rightarrow incrementos
 \rightarrow tamanho de amostra

Amostragem Estratificada

Exemplo: Seja uma área florestal, onde, através de técnicas de interpretação de imagens fotográficas, foram detectados três estratos: Estrato I (floresta alta sem babaçu), Estrato II (floresta alta com babaçu), Estrato III (floresta baixa cipoálica).

- Deseja-se estimar o volume de madeira dessa área (m^3) observando-se apenas 10 unidades amostrais.
- Como você realizaria essa amostragem?
- Você recomendaria o uso da amostragem aleatória simples?



Se fosse por AAS:

```
> sample(seq(1,96,1),10)
[1] 72 27 95 88 17 3 32 41 22 94
```

formulas:

- ◆ Estrato I - 20 $\rightarrow \frac{20}{96} = 0,20 \times 10 = 2$
- Estrato II - 48 $\rightarrow \frac{48}{96} = 0,50 \times 10 = 5$
- ▲ Estrato III - 28 $\rightarrow \frac{28}{96} = 0,29 \times 10 = 3$

10

Na **amostragem estratificada** a população de tamanho N é dividida em L subpopulações disjuntas, denominadas estratos de tamanho N_1, N_2, \dots, N_L . Em seguida é realizada uma amostragem aleatória simples dentro de cada estrato.

Amostragem estratificada proporcional

é uma amostragem estratificada onde o número n_i de elementos amostrados em cada estrato é proporcional ao tamanho N_i e cada um dos estratos ($i = 1, 2, \dots, L$).

Procedimento para a obtenção de uma amostra estratificada de tamanho n :

Considerando:

- N o número de elementos da população
 - L o número de estratos
 - N_i o número de elementos do i -ésimo estrato tal que $N = N_1 + N_2 + \dots + N_L$
 - n o tamanho da amostra a ser selecionada,
- 1 Calcular a fração da amostragem dada por $f = \frac{n}{N}$
 - 2 Calcular o número de elementos a serem sorteados em cada estrato:

$$n_1 \approx N_1 f \quad n_2 \approx N_2 f \quad \dots \quad n_L \approx N_L f,$$

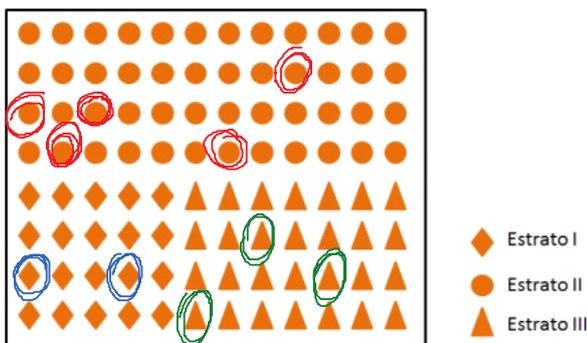
com $n = n_1 + n_2 + \dots + n_L$.

Cálculo da média

$$\bar{x}_{AE} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2 + \dots + N_L \bar{x}_L}{N},$$

onde \bar{x}_i corresponde a média aritmética simples dos n_i elementos amostrados do estrato i .

Exercício: Considerando-se o esquema apresentado anteriormente, obter uma amostra estratificada de tamanho $n = 10$ e indicar as plantas selecionadas.



forma da

◆ Estrato I - 20 $\rightarrow \frac{20}{96} = 0,20 \times 10 = 2$
● Estrato II - 48 $\rightarrow \frac{48}{96} = 0,50 \times 10 = 5$
▲ Estrato III - 32 $\rightarrow \frac{32}{96} = 0,33 \times 10 = 3$
10



$$28 \Rightarrow 28/9 = 3 \times 9 + 1 = 27 + 1 = 28$$

Códigos em R:

```
sample(seq(1,20,1),2)
```

```
sample(seq(1,48,1),5)
```

```
sample(seq(1,28,1),3)
```

```
> sample(seq(1,20,1),2)
```

```
[1] 14 11 ←
```

```
> sample(seq(1,48,1),5)
```

```
[1] 25 21 38 43 27 ←
```

```
> sample(seq(1,28,1),3)
```

```
[1] 22 10 19 ←
```

Amostragem por Conglomerado

Amostragem por conglomerado

é uma amostra aleatória simples na qual cada unidade de amostragem é um grupo de elementos, chamado conglomerado.

O número de elementos de um conglomerado deve ser pequeno e o número de conglomerados razoavelmente grande.

Procedimento para a obtenção de uma amostra por conglomerado

- 1 Definir conglomerado
- 2 Selecionar amostras aleatórias simples de grupos (conglomerados)
- 3 Todos os elementos dentro do grupo selecionado constituirão a amostra.

Amostragem Sistemática

Exemplo: Um restaurante deseja conhecer melhor o perfil de seus clientes e tem condições para entrevistar aproximadamente 15% dos mesmos. Os clientes que visitaram o restaurante num certo dia, por ordem de chegada, foram:

JJ	MR	SSZ	RA	RMA
CRG	CGB	DB	JEC	AAS
TZC	CTD	SMP	EM	DF
LB	SS	ROS	GS	SMC
MCN	JB	AFI	VFC	RSM
CF	MHT	PHA	MAP	AO
...				

Como você faria essa amostragem?

Procedimento para a obtenção de uma amostra sistemática (1 em cada k elementos)

1. Definir a porcentagem $P\%$ de elementos da população que farão parte da amostra

2. Obter o valor de k , inteiro, dado por:

$$k \approx \frac{1}{P} \times 100$$

15%
 $k = \frac{1}{15} \cdot 100 \approx 7$

3. Sortear um número r inteiro entre 1 e k

Sorteio: 3

4. A amostra será composta pelos elementos de ordem

$$r, r + k, r + 2k, r + 3k, \dots \quad 3, 10, 17, 24, \dots$$

Exemplo: Um restaurante deseja conhecer melhor o perfil de seus clientes e tem condições para entrevistar aproximadamente 15% dos mesmos. Os clientes que visitaram o restaurante num certo dia, por ordem de chegada, foram:

JJ	MR	SSZ	RA	RMA
CRG	CGB	DB	JEC	AAS
TZC	CTD	SMP	EM	DF
LB	SS	ROS	GS	SMC
MCN	JB	AFI	VFC	RSM
CF	MHT	PHA	MAP	AO
...				

Indicar os indivíduos selecionados numa amostragem sistemática supondo $P = 30\%$ e $P = 10\%$.

Códigos em R:

```
(k<- round((1/15)*100))  
(r<- sample(seq(1,k,1),1))  
seq(1,30,1)[c(r,r+k,r+2*k, r+3*k, r+4*k)]
```

Exercício:

A figura abaixo apresenta os diâmetros, em cm, de uma população de 150 árvores. Os valores em **negrito** representam as árvores com 3 anos de idade, os valores sublinhados em itálico, árvores com 5 anos de idade e os valores restantes, árvores com 4 anos de idade. (Obs.: x representam mortas)

			11,8	20,0	31,5	39,6	32,1	34,7				
	15,9	19,5	10,1	14,6	30,8	38,9	31,2	38,1	35,4			
	17,8	12,5	12,6	19,4	39,1	x	36,6	32,4	34,1			
10,3	11,0	19,0	13,1	15,5	35,5	32,1	32,8	33,9	34,8	<u>50,2</u>		
16,3	17,9	16,6	14,0	10,2	37,6	36,6	34,3	30,3	32,1	<u>57,0</u>		
13,7	13,4	x	18,0	14,6	39,9	34,3	34,6	30,0	33,9	<u>59,4</u>	<u>56,2</u>	
15,3	10,3	18,3	11,1	17,2	34,8	33,2	39,1	34,9	35,3	<u>57,6</u>	<u>52,1</u>	
15,7	10,5	18,9	11,9	10,1	36,8	36,6	35,4	31,9	39,6	<u>59,4</u>	<u>50,3</u>	
13,7	12,0	14,6	16,4	14,3	39,9	32,7	39,2	33,2	34,2	<u>50,5</u>	<u>53,9</u>	
16,6	18,5	16,6	19,5	37,3	37,8	34,3	31,4	38,9	30,3	x	<u>53,1</u>	
	14,2	15,1	19,5	35,1	31,2	34,8	39,5	31,7	30,9	<u>60,0</u>	<u>58,6</u>	
	16,3	19,1	32,2	36,7	38,1	39,5	33,6	34,3	34,4	<u>53,8</u>	<u>55,2</u>	
		19,3	36,7	37,9	38,5	x	36,5	33,7	31,4	<u>53,5</u>	<u>52,8</u>	
		17,3	35,2	39,8	39,2	36,5	33,7	39,0	<u>56,0</u>	<u>53,5</u>		
			34,6	39,0	32,9	31,1	31,8	32,8	<u>56,9</u>			

- a) Realizar duas amostragens aleatórias estratificadas de tamanho 9 com 3 unidades por estrato;
- b) Realizar duas amostragens aleatórias estratificadas **proporcionais** de tamanho 9;
- c) Para cada uma das amostras obtidas nos itens (a) e (b), calcular:
- Diâmetro médio dentro de cada estrato;
 - Diâmetro médio da amostra.