

Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"



Estatística Geral

Professor

Fábio Prata



Conceitos iniciais de Amostragem

População:

Conjunto de elementos sobre os quais se deseja informação.

Recenseamento:

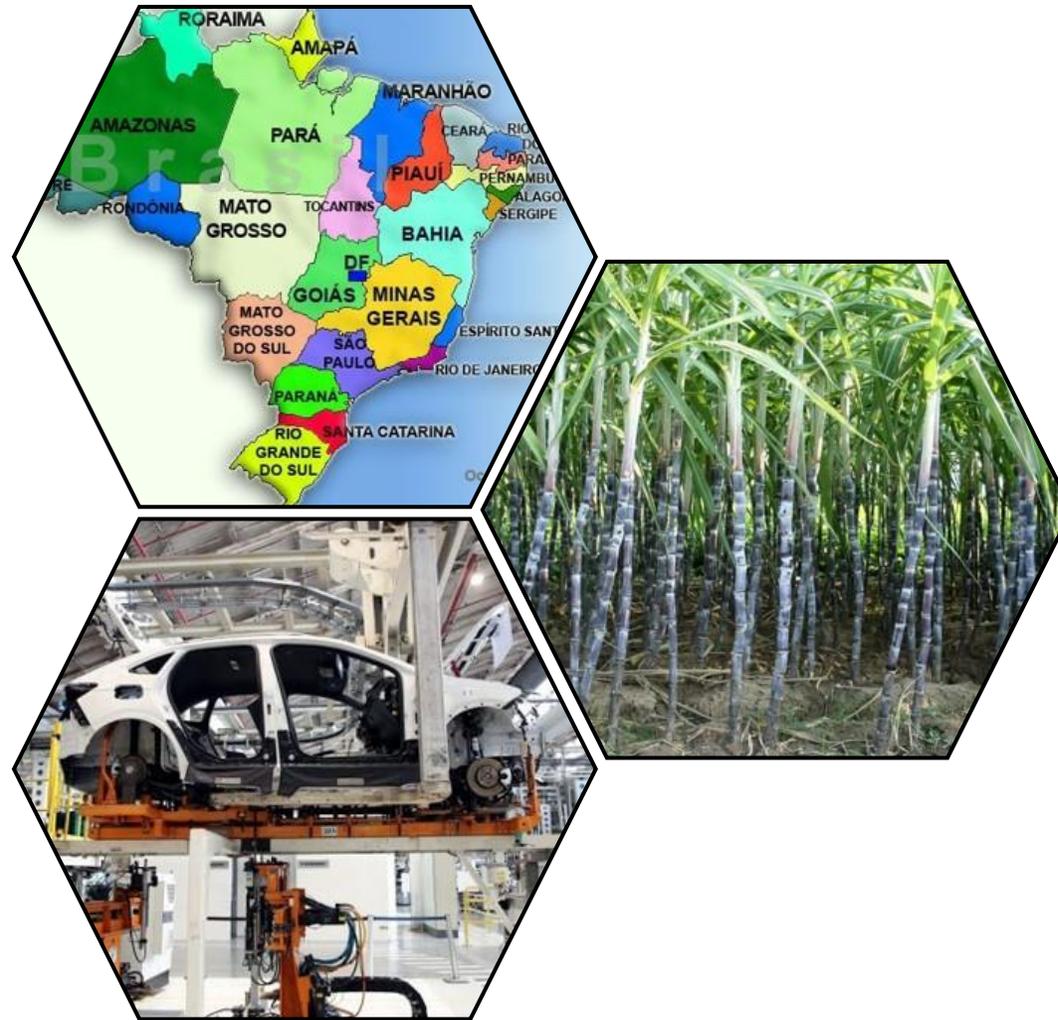
Processo de se coletar dados de uma população.

Censo:

É o conjunto de dados obtidos por meio do recenseamento.

Fatores limitantes do recenseamento

- Populações infinitas;
- Custo;
- Tempo;
- Processo destrutivo;
- Valor científico.



Conceitos iniciais de Amostragem

Amostra:

Subconjunto da população.

Amostragem:

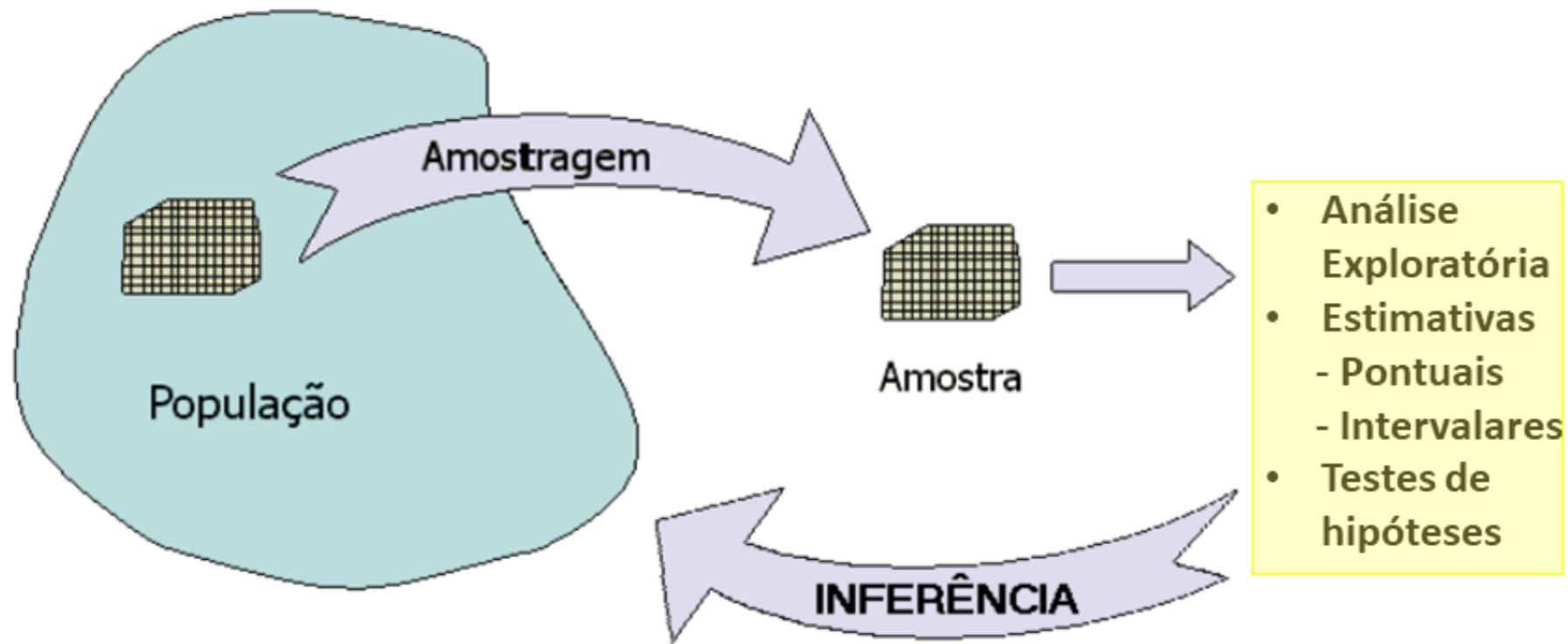
Conjunto de técnicas para se obterem amostras da população.

Parâmetro:

São quantidades da população, em geral desconhecidas, sobre as quais tem-se interesse.

Estimativa:

São valores numéricos associados aos parâmetros, obtidos por meio da amostra.



Exemplos

O objetivo de um estudo e determinar a taxa media de ocorrência de determinada praga numa fazenda produtora de

População:
Unidade Amostral:
Variável:
Amostra:
Parâmetro:
Estatística:



Exemplos

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:

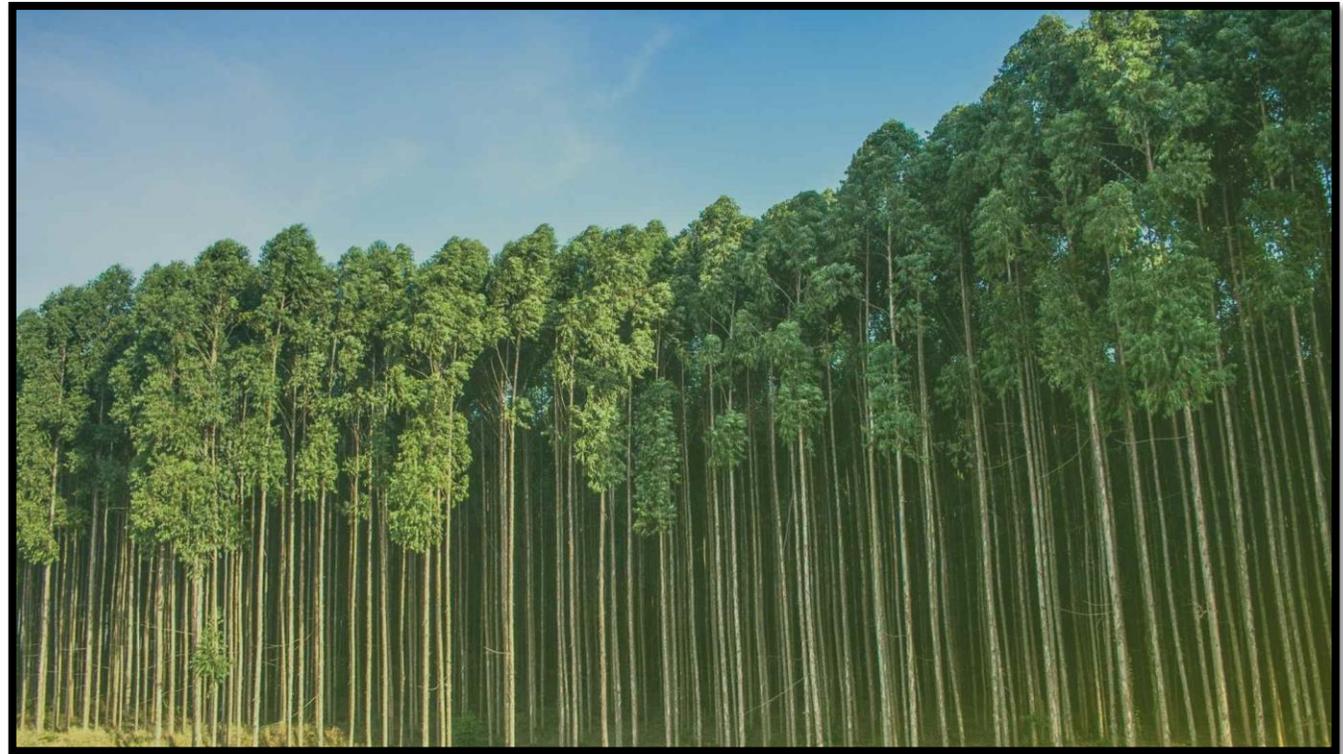
Unidade Amostral:

Variável:

Amostra:

Parâmetro:

Estatística:



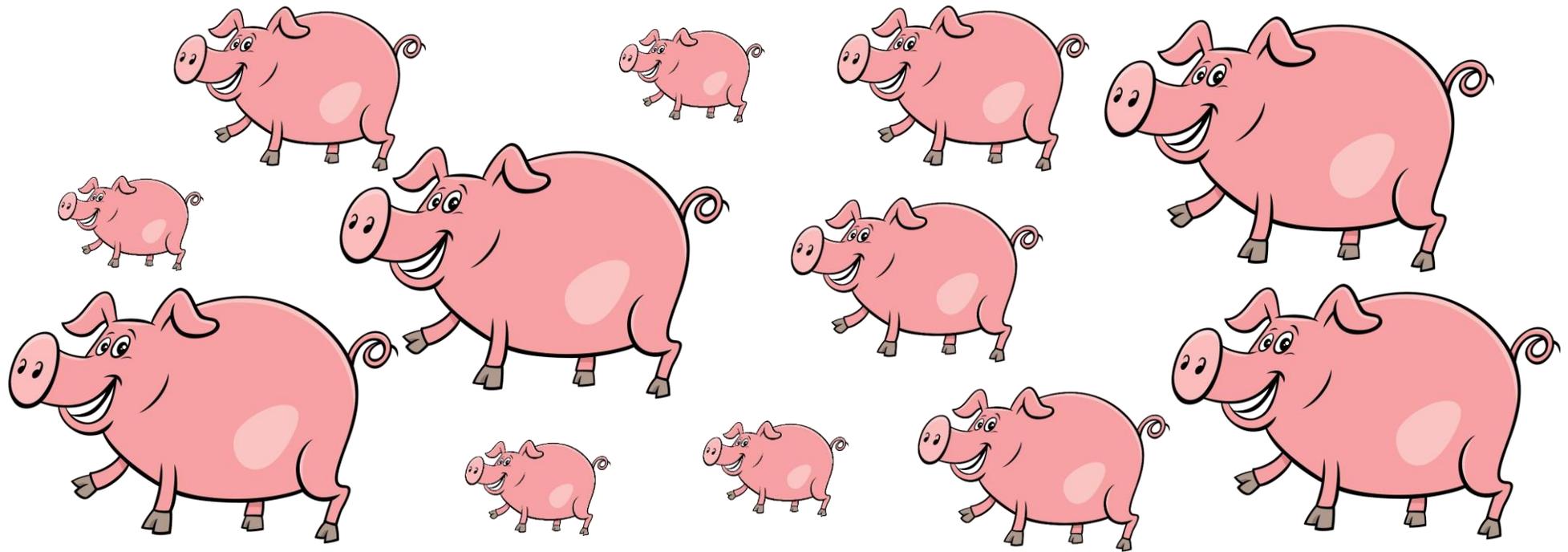
Exemplos

Pesquisa de opinião.



Exemplos

Experimentação Animal.



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.

Precisão

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



**Sem Viés
Sem Precisão**



**Com Viés
Com Precisão**



**Sem Viés
Com Precisão**

Tipos de Amostras

- **Amostra probabilística:**
 - Os itens são selecionados com base em probabilidades conhecidas;
 - Importante para a estatística pois permitem determinar estimativas para os valores da população.

- **Amostra não-probabilística:**
 - Os itens são escolhidos sem se conhecer suas respectivas probabilidades de seleção;
 - Devem ser evitados pois não permitem determinar estimativas para a população.

Problemas com amostras:

Aleatoriedade é importante na obtenção de amostras

Amostras tendenciosas: não representativas da população de que foram obtidas.

Problemáticas:

- **amostra de resposta voluntária (auto-selecionada):** os respondentes decidem, eles mesmos, se serão ou não incluídos.

Ex:

- pesquisas feitas pela internet, nas quais as pessoas podem decidir se respondem ou não.
- Pesquisas por telefone, nas quais um anúncio pede para você ligar para dar opinião

Conclusões a partir dessas amostras são válidas para o grupo específico que escolheu participar. Não deveria ser usada para fazer afirmações sobre uma população maior.

Amostras pequenas: Amostras pequenas podem não ser representativas da população.

Ex: Em uma região, entre os alunos de escola secundária suspensos, 67% tinham sido suspensos pelo menos três vezes.

Um detalhe: o resultado apresentado se baseou em uma amostra de 3 estudantes.

Note que: mesmo amostras grandes podem ser ruins se não forem coletadas adequadamente (não aleatória)

Em pesquisas com questionários. Atenção para:

Questões direcionadas: cuidado com questões de pesquisa direcionadas ou intencionalmente fraseadas para provocar uma resposta desejada.

Ex: Resultados de uma pesquisa com a mesma pergunta mas de maneira diferente:

Está sendo gasto muito pouco dinheiro com assistência social?
19% concordam

Está sendo gasto muito pouco dinheiro com assistência aos pobres?
63% concordaram

Ordem das questões:

Exemplo de pesquisa com mesma questão mas com ordem diferente:

-Você diria que o trafego contribui mais ou menos do que a industria para a poluição do ar? 45% tráfego, 27% industria

-Você diria que a industria contribui mais ou menos do que o tráfego para a poluição do ar? 24% tráfego, 57% industria

Não resposta: algumas pessoas se recusam a responder questionário. Talvez elas tenham comportamento diferente das que não recusam. Nesse caso se perde informação importante.

* Taxa de recusa tem crescido devido a pesquisas de opinião que na verdade são vendas

Dados ausentes:

Exemplo: Pesquisa realizada nos EUA em 1936: Franklin. D. Roosevelt e Alfred Landon.

Pesquisa antes da eleição pela revista Literary Digest para presidente indicava derrota de F.D. Rosevelt. (com amostra muito grande, 2.3 milhões de respostas com 10 milhões de questionários enviados)

Resultado da eleição: **Rosevelt ganhou!!!**

Má seleção de amostra: inquérito por telefone, donos de carro e assinantes da revista!!! por que?

George Gallup's American Institute of Public Opinion acertou com amostras menores (50.000), mas selecionada corretamente.

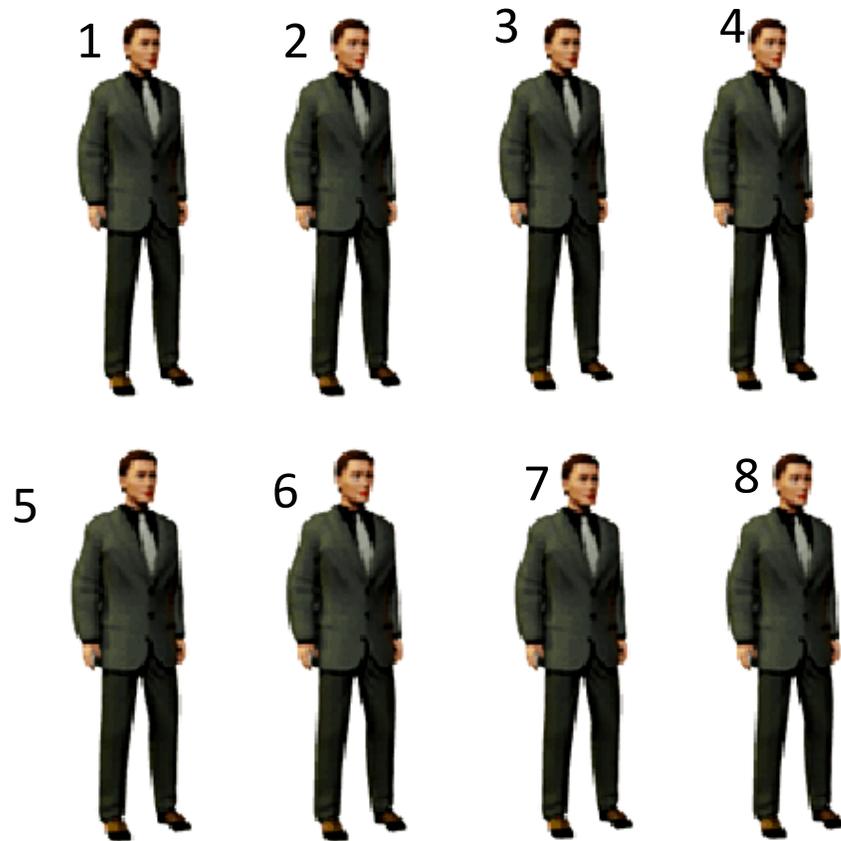
Amostragem aleatória simples (AAS)

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.

Amostragem Aleatória Simples: Seleção tal que cada elemento tem uma chance igual de ser selecionado.



Por exemplo: sorteio

Ou use um tabela de números aleatórios.

Por exemplo: temos uma lista de 21 nomes da qual queremos sortear 6 aleatoriamente.

Procedimento: Elaborar a relação dos dados (nomes) da população, atribuindo aos nomes uma numeração aleatória.

Como temos 21 elementos, iniciariamos pelo 00 até o 19, usando 2 dígitos.

Armando 00	Paulo 01	Joao 02	Ana 03	henrique 04
Marcia 05	Maria 06	Carolina 07	Marcelo 08	Juliana 09
Sandra 10	Fabio 11	Isaias 12	Júlio 13	Mariana 14
Amália 15	Gustavo 16	Artur 17	Rodrigo 18	Renato 19 Celia 20

Vamos a tabela de número aleatórios e pegamos pares de números numa linha ou coluna até completar a amostra.....

É só ir procurando numa linha ou coluna até obter o número de elementos da amostra

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

56 nt, 24 nt, 10 sandra, 04 henrique, 30 nt,
20 célia, 46 nt, 29 nt, 90 nt, 53 nt,
11 fabio, 05 marcia, 84 nt, 41 nt, 21 nt, 64, nt, 79 nt, 19 renato

Já obtivemos os 6 nomes.

Então a amostra será: {sandra, henrique, celia, fabio, marcia, renato}

nt significa não tem
Ignora-se números
que aparecem uma
segunda vez

Amostragem Estratificada : Subdivide a população em pelo menos dois diferentes subgrupos (estratos) que tenham as mesmas características, então extraia uma amostra aleatória simples de cada subgrupo.

mulheres

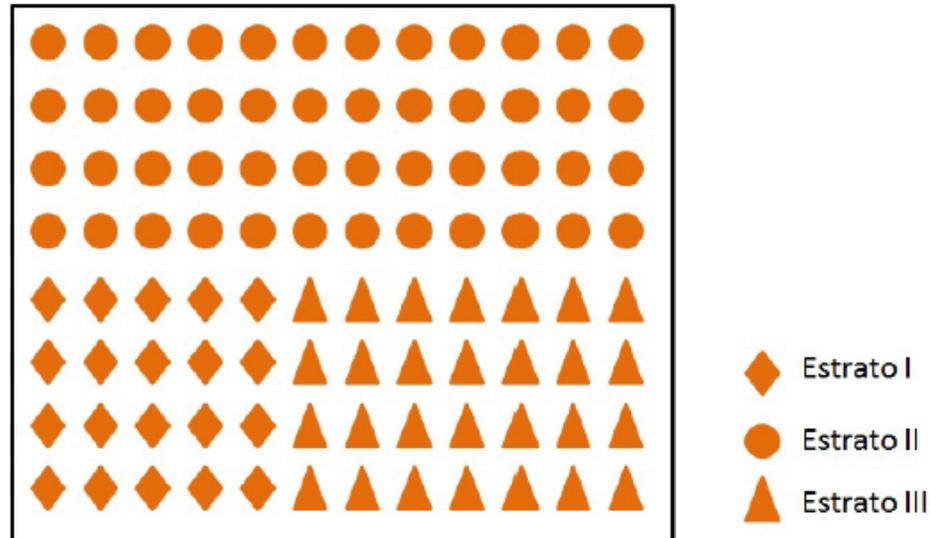


homens



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?



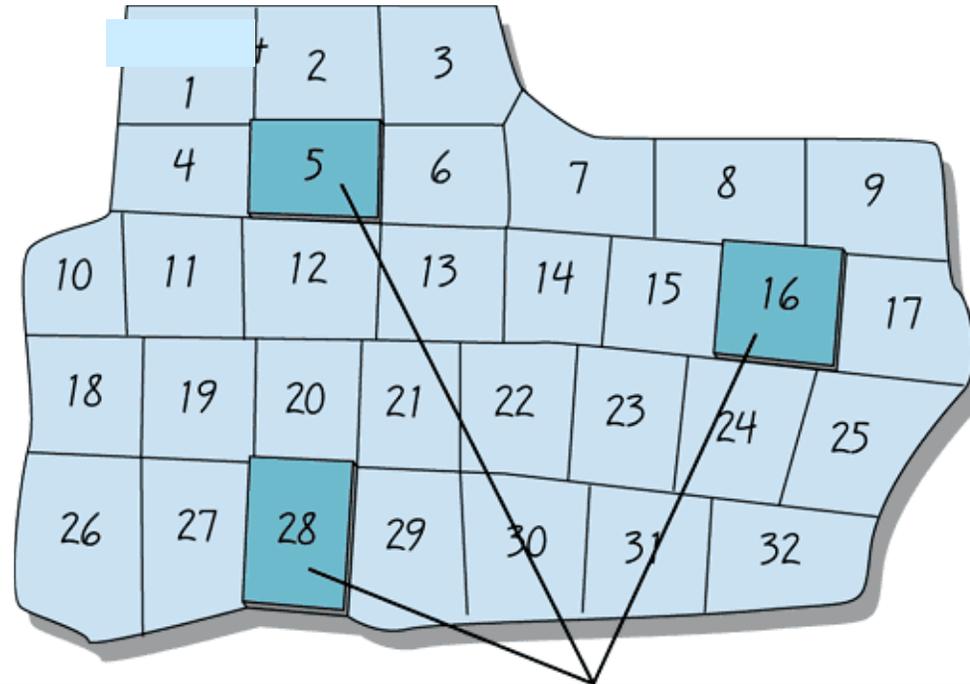
Amostragem Sistemática: Escolha algum ponto inicial e a seguir seleccione cada k -ésimo (por exemplo, o 3º elemento) elemento da população. Não é aleatória



É conveniente operacionalmente quando dispomos de uma lista, por exemplo, Lista telefônica, lista de chamada, ...

Pode ser não representativa pois pode apresentar periodicidade (repetição de característica)

Amostragem por Conglomerado: Subdivide a população em seções (conglomerados); seleciona aleatoriamente alguns destes conglomerados; escolhe todos os membros destes conglomerados selecionados



Exemplo: Entreviste todos os eleitores das zonas sombreadadas

Amostragem por conveniência: Usa resultados que são fáceis de serem obtidos



Exemplo: Os dados as vezes surpreendem

Estudo sobre o uso de “stents” na prevenção de AVC:

- Chimowitz MI, Lynn MJ, Derdeyn CP, et al. 2011. Stenting versus Aggressive Medical Therapy for Intracranial Arterial Stenosis. *New England Journal of Medicine* 365:993-1003. <http://www.nejm.org/doi/full/10.1056/NEJMoa1105335>. NY Times article reporting on the study: <http://www.nytimes.com/2011/09/08/health/research/08stent.html>.

O uso de stents (usados em problemas cardíacos) reduz o risco de AVC?

Dados de 451 pacientes de um grupo de risco (voluntários). Cada paciente foi sorteado para um dos grupos:

Grupo de tratamento (224 pessoas): receberam stents e acompanhamento médico (remédios, controle de fatores de risco, estilo de vida)

Grupo de controle (227 pessoas): somente acompanhamento médico.

Analísados dois períodos: 30 e 365 dias após a intervenção

Alguns elementos da tabela de dados:

Patient	group	0-30 days	0-365 days
1	treatment	no event	no event
2	treatment	stroke	stroke
3	treatment	no event	no event
⋮	⋮	⋮	
450	control	no event	no event
451	control	no event	no event

Stroke: AVC

Como analisar os dados: podemos usar uma tabela com uma estatística descritiva, no caso o número de pacientes numa dada

Estatística descritiva do uso dos stents

	0-30 days		0-365 days	
	stroke	no event	stroke	no event
treatment	33	191	45	179
control	13	214	28	199
Total	46	405	73	378

Estatística resumo que mostra o comportamento dos dados:

Após 1 ano:

Proporção com AVC no grupo de tratamento: $45/224 = 0.20 = 20\%$

Proporção com AVC no grupo de controle: $28/227 = 0.12 = 12\%$

SURPRESA: contrário ao esperado pelos médicos a proporção de AVC no grupo de tratamento foi 8% maior.

Mais difícil é a questão estatística: os dados apresentam realmente uma diferença entre os grupos? Ou a diferença foi devido ao acaso? Temos que estudar estatística para responder...

Cuidados: não generalizar o resultado para todos os pacientes e stents. O pacientes foram voluntários e podem ter características específicas e não representar todos os pacientes com AVC. O estudo usou um só tipo de stents.

Mas os dados podem revelar surpresas e estimular novas pesquisas.

Mais difícil é a questão estatística: os dados apresentam realmente uma diferença entre os grupos? Ou a diferença foi devido ao acaso? Temos que estudar estatística para responder...

Cuidados: não generalizar o resultado para todos os pacientes e stents. O pacientes foram voluntários e podem ter características específicas e não representar todos os pacientes com AVC. O estudo usou um só tipo de stents.

Mas os dados podem revelar surpresas e estimular novas pesquisas.

- Delineamento Inteiramente Casualizado (DIC).
 - Leva em conta apenas os princípios da repetição e casualização.
 - Os tratamentos são distribuídos de forma aleatória nas unidades experimentais (U.E.).
 - Exige que as condições de estudo sejam homogêneas, U.E.

T1	T5	T2	T3	T1
T4	T3	T4	T5	T4
T2	T4	T1	T4	T2
T5	T3	T2	T5	T3
T1	T2	T3	T1	T5

Os dados apresentados foram coletados de um experimento instalado no DBC, cujo objetivo é comparar nove porta-enxertos para laranjeira Valência. Cada parcela era constituída por duas plantas e as produções de laranja (número médio de frutos por planta) tomadas dois anos após a instalação do experimento são:



Porta Enxertos	Blocos			Totais	Médias
	I	II	III		
1. Tangerina sunki	145	155	166	466	155,33
2. Limão rugoso nacional	200	190	190	580	193,33
3. Limão rugoso da Flórida	183	186	208	577	192,33
4. Tangerina Cleópatra	190	175	186	551	183,16
5. Citranger-troyer	180	160	156	496	165,33
6. Trifoliata	130	160	130	420	140,00
7. Tangerina cravo	206	165	170	541	180,33
8. Laranja caipira	250	271	230	751	250,33
9. Limão cravo	164	190	193	547	182,33
Totais	1648	1652	1629	4929	