

Amostragem I

(com o auxílio do software R)

Ricardo Alves de Olinda

Departamento de Estatística - UEPB
(Prof. Adjunto)

Amostragem I
Aula Motivacional: UEPB

10 de Agosto de 2020

Continuação...

Por que fazer amostragem ao invés de um censo?

- Vantagens da pesquisa por amostragem em relação ao censo:
 - 1 é mais barata;
 - 2 é mais rápida;
 - 3 é mais fácil de ser controlada por envolver operações menores.

Continuação...

Por que fazer amostragem ao invés de um censo?

- Vantagens da pesquisa por amostragem em relação ao censo:
 - ➊ é mais barata;
 - ➋ é mais rápida;
 - ➌ é mais fácil de ser controlada por envolver operações menores.
- Desvantagens da pesquisa por amostragem em relação ao censo:
 - ➊ o censo pode ser mais vantajoso quando a população é pequena e/ou as informações são de fácil obtenção;
 - ➋ os resultados da pesquisa por amostragem carregam erro;
 - ➌ se a população for muito heterogênea o erro pode ser muito grande (e a precisão muita baixa).

Neste caso pode ser necessária uma amostra muito grande.

Planos de Amostragem

- Para a definição do Plano Amostral devem-se ter bem definidos:

- ① *Unidade amostral*: indivíduos ou grupos de indivíduos (conglomerados);
- ② *Sistema de referência*: lista completa das unidades amostrais.
- ③ N = tamanho da população, é definido pelo número de indivíduos da população amostral;
- ④ n = tamanho da amostra, definido pelo número de indivíduos selecionados na amostra.

Os elementos da amostra devem ser selecionados da população amostral segundo alguma forma de sorteio.

Planos de Amostragem

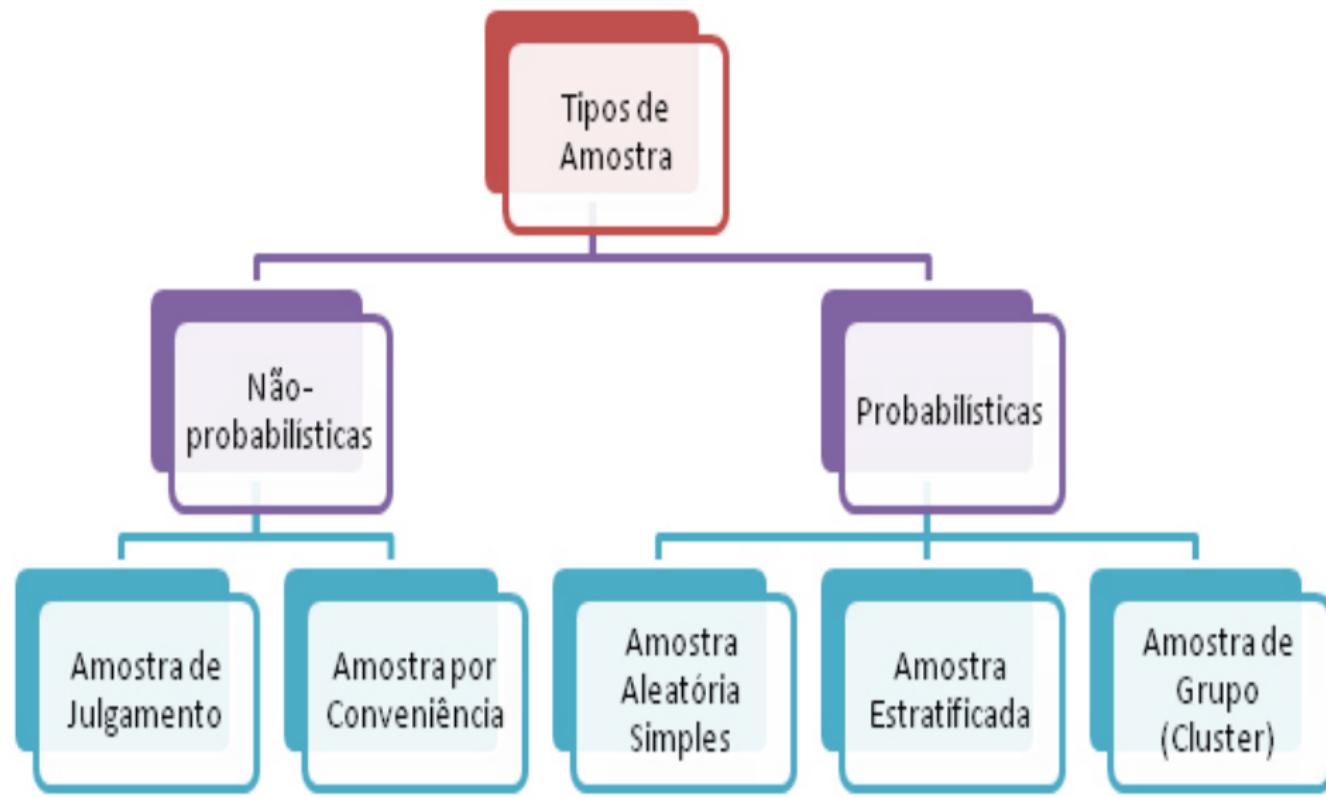
- Para a definição do Plano Amostral devem-se ter bem definidos:

- ① *Unidade amostral*: indivíduos ou grupos de indivíduos (conglomerados);
- ② *Sistema de referência*: lista completa das unidades amostrais.
- ③ N = tamanho da população, é definido pelo número de indivíduos da população amostral;
- ④ n = tamanho da amostra, definido pelo número de indivíduos selecionados na amostra.

- Fatores que interferem na escolha do Plano Amostral:

- ① Tamanho da população N ;
- ② Custo;
- ③ Heterogeneidade da população.

Os elementos da amostra devem ser selecionados da população amostral segundo alguma forma de sorteio.



① Amostragem Aleatória Simples (A.A.S.):

- Na A.A.S., a amostra de tamanho n é selecionada ao acaso dentre os N elementos da população amostral.

② Procedimento de sorteio:

- ① Um indivíduo é selecionado ao acaso dentre os N possíveis;
- ② O segundo indivíduo é selecionado ao acaso dentre os $(N - 1)$ restantes ...
- ③ ... e assim por diante, até que todos os n indivíduos sejam sorteados.

Esse procedimento tem a característica de ser “*sem reposição*”, o que significa que: *cada indivíduo aparece uma única vez na amostra*.

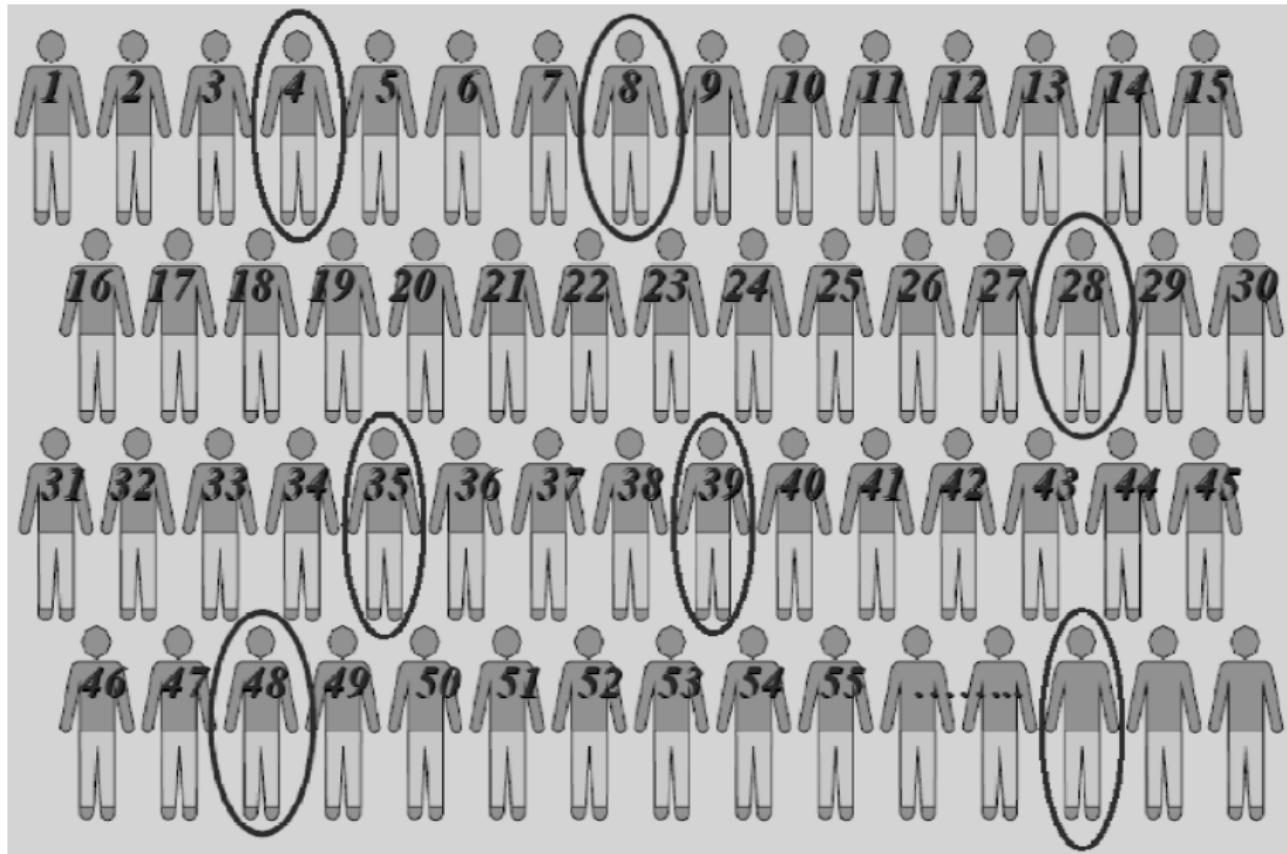
Procedimentos “*com reposição*”, quando o indivíduo pode aparecer mais de uma vez na amostra, não serão abordados por serem poucos comuns na prática.

- Na A.A.S. a probabilidade de qualquer indivíduo, ou elemento, da população fazer parte da amostra é igual a

$$\frac{n}{N}. \quad (1)$$

- *Como realizar o sorteio?*

- ① Geração números aleatórios, pelo computador;
- ② Tabela de números aleatórios;
- ③ Globos com bolinhas numeradas;
- ④ Qualquer outra forma aleatória de escolha que preserve a propriedade de que cada unidade amostral tenha a mesma chance de ser selecionada.



Vamos ao R!

Olinda, R. A.

ricardo.estat@yahoo.com.br

① Amostragem Aleatória Estratificada (A.A.E.):

- Quando a população é muito *heterogênea*, ou seja, quando as características observadas variam muito de um indivíduo para outro, é aconselhável *subdividir* a população em *estratos homogêneos*.
- A população é dividida em k estratos sendo que, uma A.A.S. é aplicada em cada um dos estratos.

② Definições:

- Tamanhos dos estratos: $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \dots, \mathbf{N}_k$.

$$\mathbf{N}_1 + \mathbf{N}_2 + \mathbf{N}_3 + \dots + \mathbf{N}_k = \mathbf{N}$$

- Tamanho das amostras por estratos: $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \dots, \mathbf{n}_k$.

$$\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3 + \dots + \mathbf{n}_k = \mathbf{n}$$

Obs: A A.A.E. produz *resultados mais precisos* do que a A.A.S. com o mesmo tamanho de amostra.

É mais cara, por segmentar a população.

Pergunta:

- ① Sabendo que o tamanho da amostra é **n**, como *alocar*, ou, determinar o número de indivíduos a serem selecionados em cada um dos estratos?

Pergunta:

- ① Sabendo que o tamanho da amostra é n , como *alocar*, ou, determinar o número de indivíduos a serem selecionados em cada um dos estratos?
- *Alocação por igual*: se desconfia de que os estratos são todos de tamanhos parecidos, ou seja,

$$\mathbf{N}_1 \approx \mathbf{N}_2 \approx \mathbf{N}_3 \dots \approx \mathbf{N}_k$$

Então pode-se fazer: $\mathbf{n}_1 = \mathbf{n}_2 = \mathbf{n}_3 = \dots = \frac{n}{k}$

Exemplo: Se o tamanho de uma amostra for $n = 56$ e, o número de estratos é $k = 4$, então, $\mathbf{n}_1 = \mathbf{n}_2 = \mathbf{n}_3 = \mathbf{n}_4 = 14$.

- **Alocação proporcional ao tamanho do estrato:** na alocação proporcional ao tamanho, os tamanhos das amostras devem seguir a mesma relação de proporcionalidade dos tamanhos dos estratos, ou seja,

$$\frac{n_1}{n} = \frac{N_1}{N}, \quad , \quad \frac{n_2}{n} = \frac{N_2}{N}, \quad \dots \quad , \quad \frac{n_k}{n} = \frac{N_k}{N}.$$

Desta forma, tem-se

$$n_1 = \frac{n \times N_1}{N}, \quad , \quad n_2 = \frac{n \times N_2}{N}, \quad \dots \quad , \quad n_k = \frac{n \times N_k}{N}.$$

Exemplo:

Considere uma amostra de tamanho $n = 48$ a ser selecionada de uma população dividida em 3 estratos, tais que $N_1 = 40$, $N_2 = 80$ e $N_3 = 120$, então

- $\mathbf{N} = 20 + 60 + 180 = 240$

$$n_1 = \frac{n \times N_1}{N} = \frac{48 \times 40}{240} \Rightarrow n_1 = 8.$$

$$n_2 = \frac{n \times N_2}{N} = \frac{48 \times 80}{240} \Rightarrow n_2 = 16.$$

$$n_3 = \frac{n \times N_3}{N} = \frac{48 \times 120}{240} \Rightarrow n_3 = 24.$$

Portanto, $n_1 = 8$, $n_2 = 16$ e $n_3 = 24$ é a alocação proporcional ao tamanho dos estratos.

Esse resultado significa que se deve selecionar 8 indivíduos do primeiro estrato, 16 do segundo estrato e 24 do terceiro.

- **Alocação ótima:** alocação que *otimiza* uma *relação conhecida (função)* e que normalmente envolve o tamanho dos estratos, as suas heterogeneidades e o custo da amostragem.

Por otimizar entende-se escolher os tamanhos de amostras em cada estratos que *maximizam*, ou *minimizam*, a função escolhida.



① Amostragem Sistemática:

- é aplicada de forma sistemática, tendo em mãos um sistema de referência de fácil acesso.

② Na **Amostragem Sistemática** além da facilidade de acesso ao sistema de referência, a informação a ser coletada também é de fácil acesso.

- Fichas de cadastro de assinantes (revistas, provedores de acesso à internet, serviço telefônico, etc); cadastro de funcionários; peças numa linha de produção; mudas num canteiro; etc.

③ **Procedimento:** com o sistema de referência em mãos

- Determina-se o intervalo de seleção, que é dado por $R = \frac{N}{n}$;
- Sorteia-se um indivíduo, ou item, dentre os **R** primeiros da relação;
- A partir daí, seleciona-se os indivíduos sistematicamente a cada intervalo de tamanho **R**.

① Exemplo:

- Se a população tem tamanho $N = 84$ e deve-se selecionar uma amostra de tamanho $n = 6$, então, tendo-se em mão uma relação com os 84 indivíduos da população:
- divide-se população em 6 seções de tamanho $\frac{84}{6} = 14$;
- seleciona-se aleatoriamente o primeiro indivíduo da amostra dentre os 14 primeiros (por exemplo, o de número 5);
- o segundo indivíduo a ser selecionado é o $5 + 14 = 19$, ou seja, o 19º da relação;
- o terceiro é o $19 + 14 = 33$, ou seja, o 33º da relação, e assim por diante.

ordem	Indivíduo selecionado
1	5º
2	19º
3	33º
4	47º
5	61º
6	75º

