

# Métodos Estocásticos da Engenharia II

## Capítulo 1 - Princípios de Amostragem

Prof. Magno Silvério Campos

2019/1



# Bibliografia

Essas notas de aulas foram baseadas nas seguintes obras:

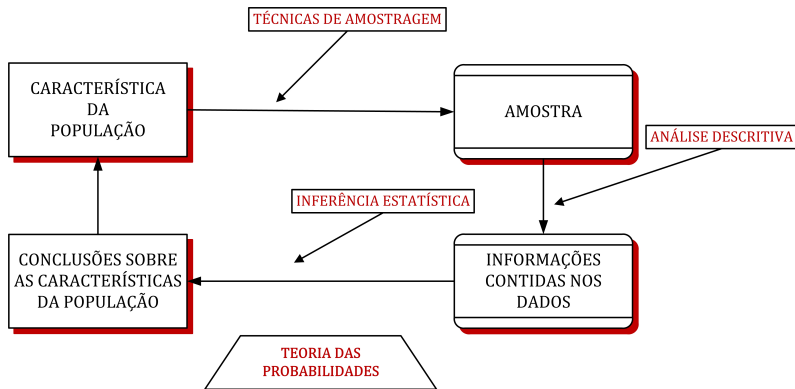
- 1 BRUNI, A. L. *Estatística Aplicada à Gestão Empresarial*. 3.ed. São Paulo: Atlas, 2011.
- 2 FARIAS, A. A.; SOARES, J. F.; CÉSAR, C. C. *Introdução à Estatística*. 2.ed. Rio de Janeiro: LTC, 2008.
- 3 LOESCH, C. *Probabilidade e Estatística*. Rio de Janeiro: LTC, 2012.

Aconselha-se pesquisá-las para se obter um **maior aprofundamento** e um **melhor aproveitamento** nos estudos.



# Introdução

## Etapas do Método de Análise Estatística



Fonte: adaptado de [Cancho(2010)]

# Conteúdo Programático

## 1 Seção 1 - Conceitos iniciais

- Censo *versus* amostragem;
- Amostras probabilísticas *versus* não-probabilísticas;
- Amostragem com reposição *versus* sem reposição;

## 2 Seção 2 - Amostragem probabilística

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

## 3 Seção 3 - Amostragem não-probabilística

- Amostragem acidental ou por conveniência;
- Amostragem por julgamento;
- Amostragem intencional ou proposital;
- Amostragem por quotas;

## 4 Seção 4 - Nível de confiança e tamanho amostral. ★



# Conceitos iniciais

## Censo *versus* amostragem

Quando o desejo é avaliar toda a população, elemento por elemento, deve-se realizar um ***censo***. Por exemplo, o IBGE realiza um censo demográfico da população brasileira a cada 10 anos, onde toda a população é "ouvida" e a partir disso, obtém-se informações da mesma.

Caso não haja recursos disponíveis para a execução do censo, como por exemplo, tempo hábil curto ou recursos financeiros restritos, deve-se optar por se realizar uma ***amostragem*** a partir da população de interesse.



## Amostras probabilísticas *versus* não-probabilísticas

Quando o desejo é fazer inferências a respeito dos dados amostrados, generalizando os resultados para toda a população, deve-se optar por uma *amostra probabilística*.

**Fato:** em uma amostragem probabilística, a amostra é representativa do universo populacional, isto é, a probabilidade de se encontrar elementos com certas características na amostra é igual à probabilidade de se encontrar elementos com as mesmas características na população.

Por exemplo, se em uma população de 700 itens produzidos existem 10% de defeituosos, uma amostra de 50 itens amostrados também deve apresentar esse mesmo percentual de defeituosos, ou seja, 5 itens.



Agora, caso não haja recursos suficientes para a realização de uma amostragem probabilística, opta-se por uma *amostra não-probabilística*, mais rápida e com custos menores, porém com o **viés** da impossibilidade de se realizar inferências.

## Fato

Se a característica de interesse na população for muito homogênea, pode-se utilizar qualquer estratégia de amostragem, seja ela probabilística ou não-probabilística.



## Amostragem com reposição *versus* sem reposição

Quando a população é **finita**, a amostragem pode ser feita de duas formas:

- **Com reposição** - o elemento selecionado volta a fazer parte da população após ser selecionado e analisado.

→ O espaço amostral permanece inalterado.

Exemplo: o lançar sucessivo de um dado não-viciado permite que uma mesma face seja sorteada mais de uma vez.

- **Sem reposição** - o elemento selecionado não volta a compor a população.

→ O espaço amostral é alterado.

Exemplo: retirar um item de uma linha de produção e testá-lo até a falha (*teste destrutivo*).



Quando a população é **infinita**, não há grandes impactos nos resultados ao se fazer amostragem com ou sem reposição.

Geralmente, costuma-se empregar o seguinte critério para saber se uma população pode ser considerada infinita em relação à amostra:

$$\text{se } \frac{n}{N} < 0,10 \quad \Rightarrow \quad \text{população infinita,}$$

sendo  $N$  e  $n$  os tamanhos populacional e amostral, respectivamente.



# Amostragem probabilística

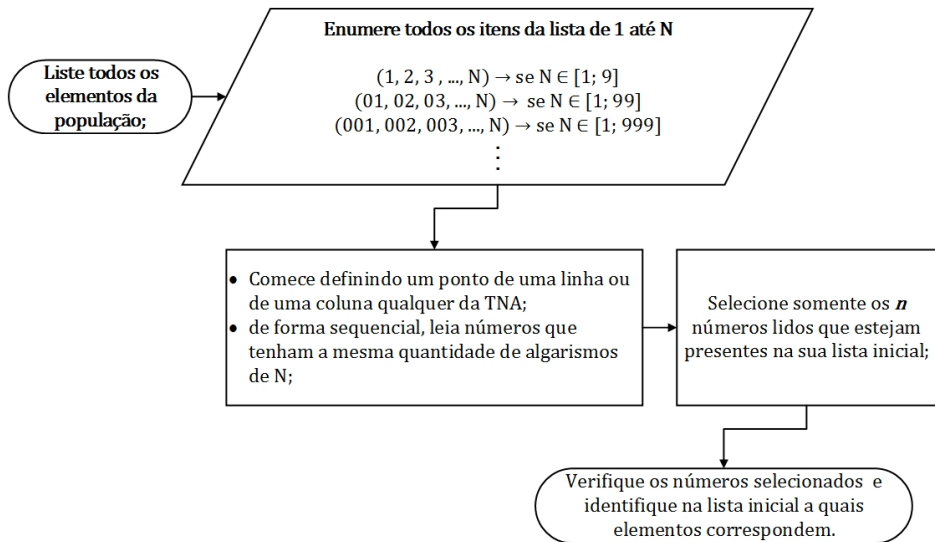
Os métodos mais usuais para se realizar uma amostragem probabilística são: *amostragem aleatória simples*, *amostragem sistemática*, *amostragem estratificada* e *amostragem por conglomerados*.

## [1] - Amostragem aleatória simples

- É ***estritamente necessário*** o conhecimento de todos os elementos que compõem a população
- A amostra é escolhida elemento a elemento;
- Cada elemento da população recebe um número de 1 a  $N$ ;
- Baseando-se em uma ***Tabela de Números Aleatórios (TNA)***, escolhem-se  $n$  números compreendidos entre 1 e  $N$ .



## Fluxograma - Método de amostragem aleatória simples



### Tabela de Números Aleatórios

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

- Conjunto de algarismos gerados por um programa computacional, através de um *gerador de números pseudo-aleatórios*;
- Cada um dos 10 algarismos (0, 1, ..., 9) aparece na tabela na mesma proporção, ou seja, 1/10;
- Cada uma das 100 combinações de 2 algarismos (00, 01, ..., 99) também aparece com igual probabilidade, 1/100, e assim por diante.



## Exemplo

Suponha que um estudo precise extrair um amostra aleatória simples formada por 8 departamentos acadêmicos da UFOP, escolhidos do universo de 48 departamentos existentes atualmente, conforme lista abaixo:

01 - DEACL	09 - DECBI	17 - DEDIR	25 - DEETE	33 - DEJOR	41 - DEMUL
02 - DEALI	10 - DECEA	18 - DEECO	26 - DEETE	34 - DELET	42 - DEMUS
03 - DEAMB	11 - DECGP	19 - DEEDU	27 - DEFAR	35 - DEMAT	43 - DENCS
04 - DEARQ	12 - DECIV	20 - DEEFD	28 - DEFIL	36 - DEMEC	44 - DEPRO
05 - DEART	13 - DECOM	21 - DEELT	29 - DEFIS	37 - DEMEC	45 - DEQUI
06 - DEBIO	14 - DECPA	22 - DEEMA	30 - DEGEO	38 - DEMET	46 - DESSO
07 - DECAD	15 - DECSI	23 - DEENP	31 - DEGEP	39 - DEMIN	47 - DETUR
08 - DECAT	16 - DECSO	24 - DEEST	32 - DEHIS	40 - DEMSC	48 - DEURB

Quais seriam os 8 departamentos amostrados, se a consulta da TNA se iniciasse na linha 3 - coluna 1?

**Observação** - E se a lista de elementos da população for infinita (discreta ou contínua)?

Nesse caso, indica-se trabalhar com faixas de valores, como exemplo, faixas  $0 \leq x \leq 5$ ;  $5 < x \leq 10$ ;  $10 < x \leq 15$ ; e assim sucessivamente.

Porém, é necessário que se garanta que as probabilidades de intervalos de valores que sejam incluídos na amostra sejam iguais às porcentagens da população. Por exemplo, se em uma população 40% dos elementos possuem massa corporal entre 50 e 70Kg, na amostra a porcentagem deve se manter: 40% dos elementos amostrados também devem ter massa corporal entre 50 e 70Kg.



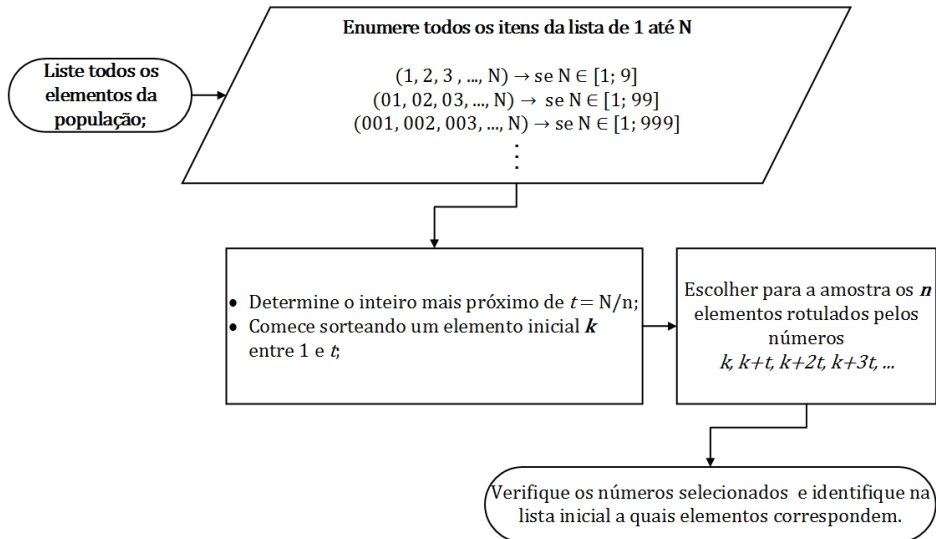
## [2] - Amostragem sistemática

A *amostragem sistemática* consiste em uma escolha dos elementos empregando o mesmo procedimento (sistemática).

- É *estritamente necessário* o conhecimento de todos os elementos que compõem a população;
- Cada elemento da população recebe um número de 1 a  $N$ ;
- Sorteia-se um elemento inicial  $k$  entre 1 e  $t$ , sendo  $t = \frac{N}{n}$ , isto é,  $t$  é o inteiro mais próximo da razão entre o tamanho da população e o tamanho da amostra;
- A amostra é escolhida elemento a elemento, a cada  $t$  elementos;
- Escolher para a amostra os  $n$  elementos rotulados pelos números  $k, k + t, k + 2t, k + 3t, \dots$



## Fluxograma - Método de amostragem sistemática





## Exemplo 1

Suponha que um estudo precise extrair uma amostra aleatória simples formada por 8 departamentos acadêmicos da UFOP, escolhidos do universo de 48 departamentos existentes atualmente, conforme lista abaixo:

01 - DEACL	09 - DECBI	17 - DEDIR	25 - DEETE	33 - DEJOR	41 - DEMUL
02 - DEALI	10 - DECEA	18 - DEECO	26 - DEETE	34 - DELET	42 - DEMUS
03 - DEAMB	11 - DECGP	19 - DEEDU	27 - DEFAR	35 - DEMAT	43 - DENCs
04 - DEARQ	12 - DECIV	20 - DEEFD	28 - DEFIL	36 - DEMEC	44 - DEPRO
05 - DEART	13 - DECOM	21 - DEELT	29 - DEFIS	37 - DEMEC	45 - DEQUI
06 - DEBIO	14 - DECPA	22 - DEEMA	30 - DEGEO	38 - DEMET	46 - DESSO
07 - DECAD	15 - DECSI	23 - DEENP	31 - DEGEP	39 - DEMIN	47 - DETUR
08 - DECAT	16 - DECSO	24 - DEEST	32 - DEHIS	40 - DEMSC	48 - DEURB

Quais seriam os 8 departamentos amostrados a partir de uma amostragem sistemática, adotando-se  $k = 04$ ?

### Observação 1

A amostragem sistemática é preferível à amostragem aleatória simples pois é mais fácil de ser executada (sem necessidade da TNA) e tem custo menor.

### Observação 2

Se o tamanho da população é desconhecido, não se pode determinar exatamente o valor de  $t$ . Logo, escolhe-se intuitivamente um valor razoável para  $t$ .



## Exemplo 2

Como proceder para extrair uma amostra de 100 usuários de um restaurante?



### [3] - Amostragem estratificada proporcional

A *amostragem estratificada proporcional* consiste em dividir a população em *subgrupos (ou estratos)* de elementos parecidos (homogêneos), aplicando em seguida, a amostragem aleatória simples ou a amostragem sistemática dentro de cada um dos subgrupos individuais. A partir disso, agrupa-se as amostras individuais.

Objetiva-se com esse procedimento, melhorar o critério de representatividade da amostra final.

São exemplos de estratificações: classe social, idade, gênero, escolaridade, profissão, etc.



O mais comum é utilizar a proporcionalidade, que consiste em selecionar os elementos da amostra em número proporcional ao tamanho de cada um dos *estratos*.

Isto é,

o número de elementos a serem sorteados de cada estrato é dado por:

$$n_1 = N_1 f \quad n_2 = N_2 f \quad n_3 = N_3 f \quad \dots \quad n_S = N_S f,$$

onde:

- $N_i$  é o número de elementos do estrato  $i, \forall i = 1, 2, \dots, S$ . Observe que  $N = N_1 + N_2 + N_3 + \dots + N_S$ ;
- $f = \frac{n}{N}$  é a proporção de amostragem da população.

Por fim, a amostra completa é dada por:

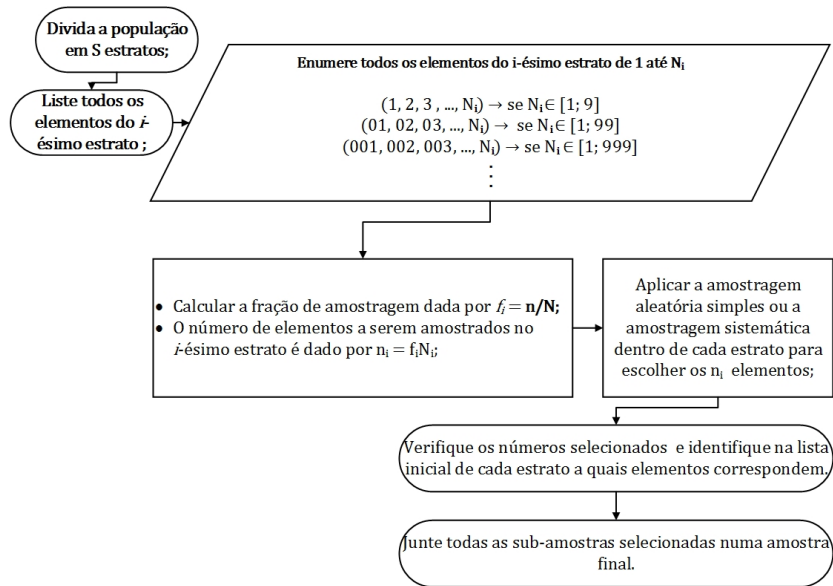
$$n = n_1 + n_2 + n_3 + \dots + n_S$$

## Observação

Quanto mais homogêneos forem os elementos dentro dos *estratos*, menor será o tamanho da amostra de cada subgrupo individual.



## Fluxograma - Método de amostragem estratificada



## Exemplo

Suponha que um estudo precise extrair uma amostra aleatória formada por 12 departamentos acadêmicos da UFOP, escolhidos do universo de 48 departamentos existentes atualmente, conforme lista abaixo:

Unidade Acadêmica	Departamentos											
CEAD	5	1 - DECSO	2 - DEEMA	3 - DEETE	4 - DEETE	5 - DEGEP						
CEDUFOP	1	1 - DEEFD										
EDTM	3	1 - DEDIR	2 - DEMUL	3 - DETUR								
EFAR	2	1 - DEACL	2 - DEFAR									
EMD	3	1 - DECGP	2 - DECPA	3 - DEMSC								
EM	11	01 - DEAMB	02 - DEARQ	03 - DECAT	04 - DECIV	05 - DEGEO	06 - DEMEC	07 - DEMEC	08 - DEMET	09 - DEMIN	10 - DEPRO	11 - DEURB
ENUT	2	1 - DEALI	2 - DENCs									
ICEA	4	1 - DECEA	2 - DECSI	3 - DEELT	4 - DEENP							
ICEB	7	1 - DEBIO	2 - DECBI	3 - DECOM	4 - DEEST	5 - DEFIS	6 - DEMAT	7 - DEQUI				
ICHS	3	1 - DEEDU	2 - DEHIS	3 - DELET								
ICSA	4	1 - DECAD	2 - DEECO	3 - DEJOR	4 - DESSO							
IFAC	3	1 - DEART	2 - DEFIL	3 - DEMUS								

Quais seriam os 12 departamentos amostrados a partir de uma amostragem estratificada por unidades acadêmicas?







## [4] - Amostragem por conglomerados

A *amostragem por conglomerados* propõe a divisão da população em conglomerados (grupos) de elementos **heterogêneos**, seguida de uma amostragem aleatória simples desses conglomerados.

Por exemplo, os bairros de uma cidade são conglomerados de pessoas com características variadas de idade, renda, cor, sexo, etc.

### Observação 1

- Na *amostragem estratificada*, seleciona-se uma amostra aleatória simples ou sistemática dentro de cada grupo (**estrato**);
- Já na *amostragem por conglomerados*, selecionam-se amostras aleatórias simples ou sistemáticas de grupos (**conglomerados**), e todos os elementos dentro desses conglomerados selecionados farão parte da amostra.

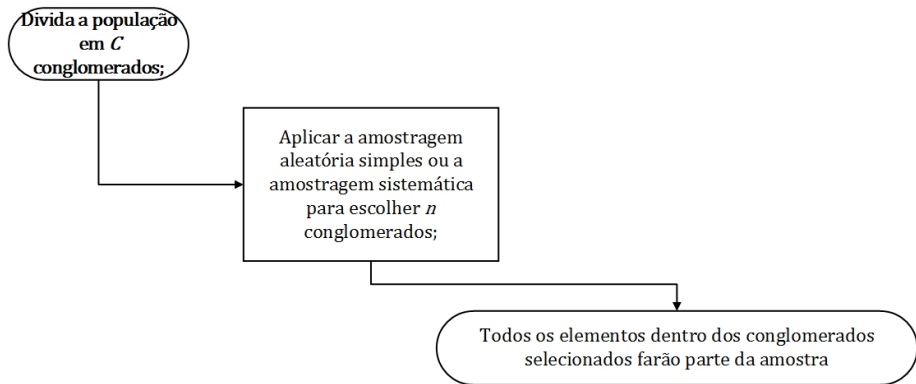
## Observação 2

A *amostragem por conglomerados* é indicada quando:

- 1 Não se tem a listagem de todos os elementos da população ou ela é cara;
- 2 Os elementos da população estão distantes fisicamente uns dos outros, o que faz os custos de obtenção de informação aumentarem com o aumento dessa distância.



## Fluxograma - Método de amostragem por conglomerados



## Exemplo

Qual a melhor forma de se escolher uma amostra para se estimar o rendimento médio familiar em uma metrópole?



## Amostragem não-probabilística

Os métodos mais usuais para se realizar uma amostragem não-probabilística são: *amostragem accidental ou por conveniência*, *amostragem por julgamento* e *amostragem por quotas*.

### [1] - Amostragem accidental ou por conveniência

Na *amostragem accidental* ou por conveniência, os elementos são escolhidos por serem os mais acessíveis ou fáceis de serem amostrados.

Por exemplo, para saber a preferência dos eleitores para prefeito de Ouro Preto, os pesquisadores poderiam se posicionar em locais de grande tráfego de eleitores, perguntando-os sobre suas intenções de voto.

Deve ser utilizada em um estudo exploratório inicial, abrindo campo para uma amostragem probabilística posterior.

## Vantagens

- Essa amostragem é rápida e de baixo custo;
- Não é necessário o conhecimento da lista de todos os elementos da população.

## Vieses

- Corre-se o risco de se formar amostras com grandes concentrações de determinadas características, pois a amostragem pode ser influenciada pelo local da coleta de informações (aeroportos, estações rodoviárias, shopping centers, camelódromo, etc.)
- Por ser não-probabilística, essa amostragem não permite a realização de inferências.





## [2] - Amostragem por quotas

A *amostragem por quotas* consiste em um refinamento da amostragem acidental ou por conveniência. Nela, os elementos a serem selecionados devem estar de acordo com as proporções de características da população.

Por exemplo, se em uma população, existem 20% de indivíduos da classe econômica A, 50% da classe B e 30% da classe C, a amostra acidental ou por conveniência deve respeitar essas proporções (amostragem por quotas), selecionando 20% de indivíduos da classe econômica A, 50% da classe B e 30% da classe C.

### Vantagens

- Essa amostragem é rápida e de baixo custo;
- Não é necessário o conhecimento de todos os elementos da população.

### Viés

Por ser não-probabilística, essa amostragem não permite a realização de inferências.

### [3] - Amostragem por julgamento

Na *amostragem por julgamento* o pesquisador seleciona aqueles elementos que ele julga representarem melhor a população.

Por exemplo, em uma pesquisa sobre desempenho acadêmico, um pesquisador resolve entrevistar somente aqueles alunos que tenham coeficiente de rendimento acadêmico acima de 7, julgando que estes dariam respostas mais condizentes com o assunto.

#### Vantagem

Essa amostragem é rápida e de menor custo.

#### Vieses

- A presença de subjetividade, preferências e preconceitos pode provocar tendências na representatividade da amostra;
- Por ser não-probabilística, essa amostragem não permite a realização de inferências.

# Nível de confiança e tamanho amostral

## Nota

Em momento oportuno do curso, retornaremos a este tópico.





Cancho, V., 2010. Notas de aulas sobre noções de estatística e probabilidade.

