Bioestatística

População, amostra, variáveis, Técnicas de amostragem

Prof. Dr. Uirá do Amaral

Níveis de mensuração

- Escala nominal

  Os indivíduos são classificados em categorias segundo uma característica.

  Ex: sexo (masculino, feminino), hábito de fumar (fumante, não fumante), sobrepeso (sim, não), condição do domicílio (próprio já pago, próprio em pagamento, alugado, cedido, outra condição)

  Não existe ordem entre as categorias e suas representações, se numéricas, são destituídas de significado numérico.

  Ex: sexo masculino=1, sexo feminino = 2.

  Os valores 1 e 2 são apenas rótulos.

- Escala ordinal

  Os indivíduos são classificados em categorias que possuem algum tipo inerente de ordem. Neste caso, uma categoria pode ser "maior" ou "menor" do que outra.

  Ex: nível sócio-econômico (A, B, C e D; onde A representa maior poder aquisitivo);

  nível de retinol sérico (alto, aceitável, baixo, deficiente) onde alto: maior ou igual a 50,0 µg/dl; aceitável: 20,0 a 49,9 µg/dl; baixo: 10,0 a 19,9 µg/dl; deficiente: menor ou igual a 10,0 µg/dl. Estes critérios são do *Commitee on Nutrition for National Defense ICNND/USA*, 1963 (in Prado MS et al, 1995).

  Embora exista ordem entre as categorias, a diferença entre categorias adjacentes não têm o mesmo significado em toda a escala.

- Escala de razões discreta: o resultado numérico da mensuração é um valor inteiro.

  Ex: número de refeições em um dia (nenhuma, uma, duas, três, quatro, ...),

  frequência de consumo semanal de determinado alimento (1 vez, 2 vezes, 3 vezes, 4 vezes, 5 vezes, 6 vezes, 7 vezes) .

- Escala de razões contínua: o resultado numérico é um valor pertencente ao conjunto dos números reais R ={-∞; ...; 0; 0,2; 0,73; 1; 2,48;...; +∞}.

  Ex: idade (anos), peso (g), altura (cm),nível de retinol sérico (µg/dl), circunferência da cintura (cm).

De acordo com os níveis de mensuração, pode-se classificar a **natureza das variáveis** segundo a escala de mensuração em:

VARIÁVEL:
- qualitativa
  - nominal
  - ordinal
- quantitativa
  - discreta
  - contínua

O tipo da variável irá indicar a melhor forma para o dado ser apresentado em tabelas e gráficos, em medidas de resumo e, a análise estatística mais adequada.

Coleta de dados

- É a observação e registro da categoria ou medida de variáveis relacionadas ao objeto de estudo que ocorrem em unidades (indivíduos) de uma amostra ou população.

Tópicos iniciais de amostragem

- População: totalidade de elementos sob estudo. Apresentam uma ou mais características em comum.

  Supor o estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de Urutaí.

  População alvo – todas as crianças nesta faixa etária deste município.

  População de estudo – crianças matriculadas em escolas.

- Elementos: são unidades de análise; podem ser pessoas, domicílios, escolas, creches, células ou qualquer outra unidade.

- Amostra: é uma parte da população de estudo.

- Amostragem: processo para obtenção de uma amostra. Tem como objetivo estimar parâmetros populacionais.

- Parâmetro: Quantidade fixa de uma população.

  Ex: peso médio ao nascer de crianças que nascem no município de São Paulo ($\mu = 3100$ g);

  Proporção de crianças de 7 a 12 anos classificadas como obesas, no município de São Paulo ($\pi = 12\%$).

- Estimador: é uma fórmula matemática que permite calcular um valor (estimador por ponto) ou com um conjunto de valores (estimador por intervalo) para um parâmetro.

Ex: Média aritmética: $\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$,

onde $\sum_{i=1}^{N} X_i = X_1 + X_2 + \ldots + X_N$ e N = número de observações.

- Estimativa: Valor do estimador calculado em uma amostra. Estima o valor do parâmetro.

  Ex: Peso médio ao nascer, calculado em uma amostra de 10 crianças nascidas no Município de Urutaí-GO no ano de 2000:

  média amostral = $\bar{x} = 3000g$.

- Indicações para utilizar uma amostra
  - População muito grande
  - Processo destrutivo de investigação
  - Novas terapias

- Vantagens de realizar um estudo com amostragem:
  - Menor custo
  - Menor tempo para obtenção dos resultados
  - Possibilidade de objetivos mais amplos
  - Dados possivelmente mais fidedignos

- Desvantagens
  - Resultados sujeitos à variabilidade

Tipos de Amostragem

- **Probabilística:** cada unidade amostral tem probabilidade conhecida e diferente de zero de pertencer à amostra. É usada alguma forma de sorteio para a obtenção da amostra.

- **Não probabilística:** não se conhece a probabilidade de cada unidade amostral pertencer à amostra. Algumas unidades terão probabilidade zero de pertencer à amostra.

  Ex: amostragem intencional; por voluntários; acesso mais fácil; por quotas.

- Tipos de amostragem probabilística:
  - aleatória simples (com e sem reposição);
  - sistemática;
  - com partilha proporcional ao tamanho do estrato;
  - por conglomerado.

- <u>Amostragem aleatória simples (AAS)</u>

  É o processo de amostragem onde qualquer subconjunto de n elementos diferentes de uma população de N elementos tem mesma probabilidade de ser sorteado.

  - Suponha uma amostra representativa de 20% para uma pesquisa de estatura de 50 estudantes de uma turma.
  - 1° passo: Quanto é 20% de 50 = 10 estudantes
  - 2° passo: numeramos os estudantes de 1 a 50
  - 3° passo: escrevemos os número de 1 a 50 em pedaços de papel e depositamos dentro de uma urna. Misturamos bem e retiramos um a um, dez pedaços de papel que formarão a amostra.

Fórmula para cálculo do tamanho da amostra

- N = Tamanho da população
- E_0 = erro amostral tolerável
- n_0 = primeira aproximação do tamanho da amostra

$$n_0 = \frac{1}{E_0^2}$$

- n = tamanho da amostra

$$n = \frac{N.n_0}{N + n_0}$$

Observações:

6ª) Valores de $Z_{\alpha/2}$ para os níveis de confiança mais usados na prática:

Nível de confiança	α	α **/ 2**	$Z_{\alpha/2}$
90%	0,10	0,05	1,65
95%	0,05	0,025	1,96
99%	0,01	0,005	2,58

Distribuição normal reduzida P(0 < Z < z)

	Último dígito									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4658	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890

2. Determinação do tamanho da amostra

2.1 Para estimar a média populacional

- **Variância populacional conhecida**

$$n = \left(\frac{Z \cdot \sigma}{e} \right)^2$$

$$n = \frac{(Z)^2 . \sigma^2 . N}{e^2.(N - 1) + (Z)^2 . \sigma^2}$$

Exemplo1

- Que tamanho deve ter uma amostra com reposição para que possamos estimar a média de depósitos de todas as c/correntes da agência de um determinado banco, com 99% de confiança, sendo o erro de amostragem de 10% e o desvio padrão populacional deve estar torno de 20 u.m.

$$n = \left(2{,}58 \times 20)/0{,}10\right)^{2} = 266{,}256 \approx \textbf{267 clientes}$$

Exemplo2

- Que tamanho deve ter uma amostra sem reposição para que possamos estimar a média de depósitos de todas as c/correntes da agência (N=500) de um determinado banco, com 95% de confiança, sendo o erro de amostragem de 10% e o desvio padrão populacional deve estar torno de 20 u.m.

$$n = \frac{1{,}96^2 \times 20^2 \times 500}{0{,}10^2 \times (500 - 1) + 1{,}96^2 \times 20^2} = \mathbf{498\ clientes}$$

Exemplo cálculo do tamanho da amostra

N = 200 famílias

E_0 = erro amostral tolerável = 4% (E_0 = 0,04)

$n_0 = 1/(0,04)^2$ = 625 famílias

n (tamanho da amostra corrigido) =

n = 200x625/200+625 = 125000/825 = 152 famílias

E se a população fosse de 200.000 famílias?

n = (200.000)x625/(200.000 +625) = 623 famílias

Observe=se que se N é muito grande, não é necessário considerar o tamanho exato N da população. Nesse caso, o cálculo da primeira aproximação já é suficiente para o cálculo.

$$n = n_0 = \frac{1}{E_0^2}$$

Exercício Tamanho da amostra ...

5. Numa pesquisa para uma eleição presidencial, qual deve ser o tamanho de uma amostra aleatória simples, se se deseja garantir um erro amostral não superior a 2% ?

$$n = n_0 = 1/(0,02)^2 = 1/0,0004 = 2500 \text{ eleitores}$$

6. Numa empresa com 1000 funcionários, deseja-se estimar a percentagem dos favoráveis a certo treinamento. Qual deve ser o tamanho da amostra aleatória simples que garanta um erro amostral não superior a 5%?

$N = 1000$ empregados

E_0 = erro amostral tolerável = 5% ($E_0 = 0,05$)

$n_0 = 1/(0,05)^2 = 400$ empregados

$n = 1000x400/(1000+400) = 286$ empregados

Tamanho da amostra ...

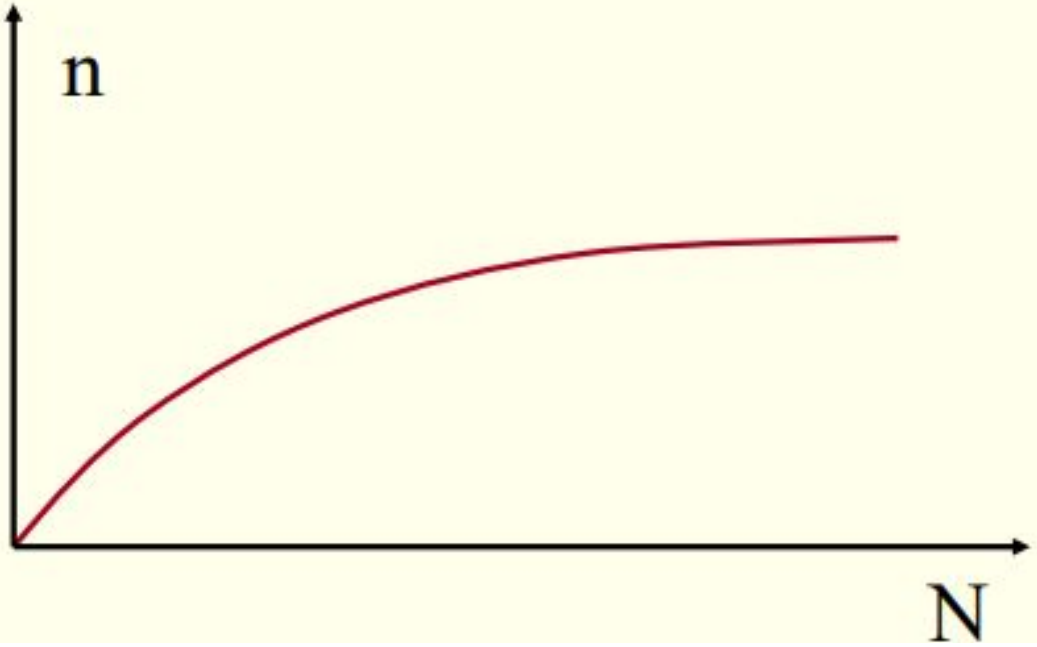
Observe que: N = 200 famílias, E_0 = 4%

n = 152 famílias ➔ 76% da população

Observe que: N = 200.000 famílias, E_0 = 4%

n = 623 famílias ➔ 0,3% da população

Logo, é errôneo pensar que o tamanho da amostra deve ser tomado como um percentual do tamanho da população para ser representativa



- <u>Amostragem sistemática</u>

  Utiliza-se a ordenação natural dos elementos da população (prontuários, casa, ordem de nascimento).

  - Intervalo de amostragem $k = \frac{N}{n}$, onde
    N= tamanho da população e n = tamanho da amostra.
  - Início casual i, sorteado entre 1 e k, inclusive.
  - Amostra sorteada é composta pelos elementos: i, i+k, i+2k, ..., i+(n-1)k

OBS: É necessário ter cuidado com a periodicidade dos dados, por exemplo se for feito sorteio de dia no mês, pode cair sempre em um domingo onde o padrão de ocorrência do evento pode ser diferente.

Exemplo:

N=80; n=10; $k = \frac{N}{n} = \frac{80}{10} = 8$; início casual: $1 \leq i \leq 8$

Começo casual **sorteado**: i=4

Amostra composta dos elementos:

i 4

i+k 12

i+2k 20

i+3k 28

i+4k 36

i+5k 44

i+6k 52

i+7k 60

i+8k 68

i+(n-1)k 76

Amostragem casual simples estratificada com partilha proporcional

A população possui estratos com tamanhos:

N_1; N_2; N_3, onde a soma dos estratos é o tamanho da população, ou seja $\sum N_i = N$

A amostra deve conter os elementos da população nas mesmas proporções dos estratos. Tem-se que os tamanhos dos estratos amostrais são n_1, n_2 e n_3 tal que $\sum n_i = n$

Aplicando-se a proporção:

$$\frac{n_i}{n} = \frac{N_i}{N} \Rightarrow n_i = n\frac{N_i}{N}$$

Exemplo:

N=500; N_1=50; N_2=150; N_3=300 e n=40

Estrato i	Tamanho do estrato na população N_i	Tamanho do estrato na amostra n_i	$\frac{n_i}{n} = \frac{N_i}{N}$
1	50	4	0,1
2	150	12	0,3
3	300	24	0,6
Total	500	40	

$$n_1 = 40\frac{50}{500} = 4\,;\; n_2 = 40\frac{150}{500} = 12\,;\; n_3 = 40\frac{300}{500} = 24$$

- <u>Amostragem por conglomerado</u>:

O conglomerado é um conjunto de elementos formando uma unidade amostral. Se a unidade amostral for indivíduo e forem sorteados domicílios, então a amostragem é por conglomerado.