

Tamanho Amostral e Amostragem Estratificada

Gilberto Pereira Sassi

Universidade Federal Fluminense
Instituto de Matemática e Estatística
Departamento de Estatística

07 de junho de 2016

Tamanho da Amostrai

Objetivo: Determinar o tamanho mínimo da amostra aleatória simples.

Erro Amostral diferença entre estimativa e parâmetro

Erro Amostral Tolerável erro amostral máximo, em proporção, aceitável pelo pesquisador.

Notações:

N número de elementos da população

n número mínimo de elementos da amostra

E_0 erro amostral tolerável (em proporção)

Tamanho Amostral:
$$n = \frac{N}{N \cdot E_0^2 + 1}$$

Exemplo

Planeja-se um levantamento para avaliar a porcentagem que moram em casa próprias em população em um bairro com $N = 200$ famílias. Qual deve ser o tamanho mínimo amostral para termos um erro amostral tolerável de 4%?

Desejamos achar n , conhecendo $N = 200$, $E_0 = \frac{4}{100} = 0,04$. O tamanho amostral mínimo é

$$n = \frac{200}{200 \cdot (0,04)^2 + 1} = 151,52.$$

Logo, o tamanho mínimo amostral é $n = 152$.

Note que $\left(\frac{152}{200}\right) \cdot 100 = 76\%$, ou seja, o tamanho mínimo amostral corresponde a 76% dos elementos da população. Este é um caso em que a amostragem pode não ser vantajoso.

Exemplo

Suponha agora que $N = 200000$, então

$$n = \frac{200000}{200000 \cdot (0,04)^2 + 1} = 623,053.$$

Logo, $n = 624$.

Note que $\frac{624}{200000} 100 = 0,312\%$, ou seja, o tamanho amostral corresponde a 0,312% dos elementos da população. Aqui, vemos claramente a vantagem de usar amostragem.

Observações:

- Com o mesmo erro amostral tolerável, usamos diferentes proporções da população.
- O erro amostral tolerável não é a proporção que devemos coletar da população.

Uso de Tabelas de Número Aleatórios

Para o próximo exemplo precisamos da tabela 1 de números aleatórios.

2	5	6	3	6	9	7	4	2	5	6	8	7	3	1	1
7	8	5	9	7	9	2	3	8	2	0	8	0	1	4	6
4	5	1	8	5	6	3	1	0	1	3	7	4	2	9	3
8	4	8	6	5	7	2	1	4	7	2	4	2	6	8	6
9	9	2	6	2	0	1	2	7	8	8	7	3	4	8	3
0	4	0	9	1	4	4	4	1	3	4	6	6	7	1	1

Tabela 1: Tabela de Números Aleatórios.

Exemplo

Considero a população composta de 200 crianças do sexo masculino (representados por $H1, \dots, H200$) e 100 crianças do sexo feminino (representados por $M1, \dots, M100$).

- a) Qual deve ser o tamanho da amostra para que o erro amostral tolerável seja 0,1?

$$n = \frac{N}{N \cdot E_0^2 + 1} = \frac{300}{300 \cdot (0,1)^2 + 1} = 75$$

- b) Retire uma amostra aleatória simples do tamanho $n = 5$.
Primeiro atribuímos um número a cada elemento da população:

H1	...	H200	M1	...	M100
1	...	200	1	...	100

Sorteamos 5 números usando a tabela de números aleatórios:
256, 178, 013, 147, 242.

A amostra aleatória simples é: $M56, H178, H13, H147, M242$.

Exemplo

Qual o erro amostral tolerável para uma amostra de tamanho $n = 10$.

$$\begin{aligned} E_0 &= \sqrt{\frac{N - n}{N \cdot n}} \\ &= \sqrt{\frac{300 - 10}{300 \cdot 10}} \\ &= \sqrt{\frac{290}{300 \cdot 10}} \\ &\approx 0,31 \end{aligned}$$

Amostra Estratificada

A técnica de amostragem estratificada consiste em dividir a população em estratos que devem ser mais homogêneos que a população em relação à variável ou variáveis sob estudo. Para usarmos amostragem estratificada, precisamos alguns pré-requisitos:

- Precisamos estabelecer os estratos ou a variável estratificadora;
- Precisamos ter um conhecimento prévio sobre a população;
- Precisamos ter uma listagem completa da população.

Exemplo

Deseja-se saber a porcentagem que cada pré-candidato receberia em uma possível eleição geral em 2016 entre os chefes de família no bairro Saco Grande II. O pesquisador sabe que o grau de escolaridade está fortemente relacionada com a intenção de voto de uma família: quanto mais instruído mais inclinado o chefe de família está em votar em um candidato de direita. Além disso, de um estudo anterior, temos as distribuições de frequência das tabelas 2, 3, 4 e 5 e os gráficos de barras da figuras 1, 2, 3 e 4. Notamos que o grau de instrução, que está fortemente associada a intenção de voto, não tem um distribuição homogênea nas regiões Monte Verde, Pq da Figueira e Encosta do Morro dentro do bairro Saco Grande II. Note que uma amostra aleatória poderia retirar todos ou a maioria dos elementos de uma única região e a estimativa pode não ser precisa. A ideia da amostra estratificada é escolher elementos da população mantendo a mesma proporção de moradores de cada região da população na amostra.

Exemplo

Tabela 2: Saco Grande II

Distribuição de frequência – total		
Grau de Instrução	Frequência	Proporção
Nenhum grau completo	38	0.32
Primeiro grau completo	37	0.31
Segundo grau completo	45	0.38
Total	120	1.00

Tabela 3: Monte Verde

Distribuição de frequência – Monte Verde		
Grau de Instrução	Frequência	Proporção
Nenhum grau completo	6	0.15
Primeiro grau completo	11	0.275
Segundo grau completo	23	0.575
Total	40	1

Tabela 4: Pq da Figueira

Distribuição de frequência – Pq da Figueira		
Grau de Instrução	Frequência	Proporção
Nenhum grau completo	14	0.33
Primeiro grau completo	14	0.33
Segundo grau completo	15	0.35
Total	43	1.00

Tabela 5: Encosta do Morro

Distribuição de frequência – Encosta do Morro		
Grau de Instrução	Frequência	Proporção
Nenhum grau completo	18	0.49
Primeiro grau completo	12	0.32
Segundo grau completo	7	0.19
Total	37	1.00

Exemplo

Figura 1: Gráfico de Barras – Total

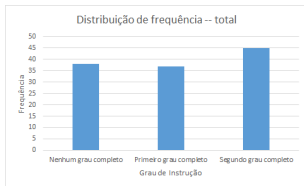


Figura 2: Gráfico de Barras – Monte Verde

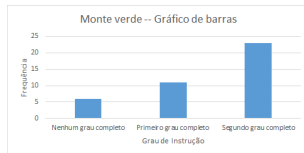


Figura 3: Gráfico de Barras – Pq da Figueira

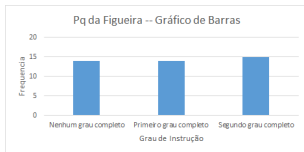
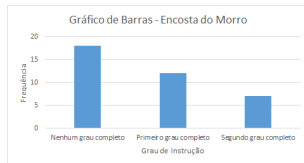
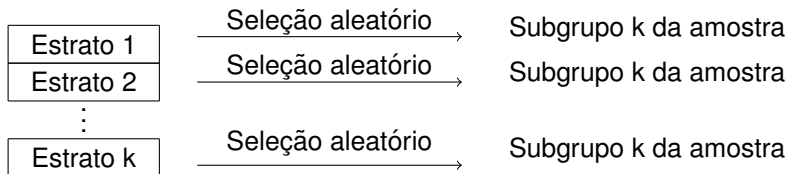


Figura 4: Gráfico de Barras – Encosta do Morro



Forma de Seleção

Sobre os diversos estratos da população são realizadas seleções aleatórias de forma independente. A amostra completa é obtida através da agregação das amostras em cada subgrupo.



Características da amostragem estratificada

- Em uma amostragem estratificada proporcional, selecionamos o número de elementos em cada subgrupo de tal forma que a proporção de elementos em cada estrato na amostra reflita a proporção de elementos em cada estrato na população.
- Se os estratos formam subgrupos mais homogêneos que a população em relação a uma variável, uma amostra estratificada tende a gerar resultados mais precisos.

Exemplo

Suponha que os chefes de família do bairro Saco Grande II tem a seguinte distribuição populacional

Localidade	População	Porcentagem
Monte	2000	0.1333
Pq da Figueira	3000	0.2
Encosta do Morro	10000	0.6667
Total	150000	1

Então, se desejamos uma amostra estratificada com 3100 chefes de famílias, precisamos selecionar o número de elementos da população em cada estrato por

Estrato	Tamanho do subgrupo	Tamanho do subgrupo ajustado
Monte	$0.1333 \cdot 3100 = 413.12$	414
Pq da Figueira	$3100 \cdot 0.2 = 620$	620
Encosta do Morro	2066.77	2067
Total	3100	3101

Exemplo

Com o objetivo de levantar o estilo de liderança preferido da comunidade escolar com 50 membros distribuídos de acordo com a tabela, retire uma amostra estratificada com 10 elementos.

População	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Servidores	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Alunos	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30

A distribuição de frequência da variável estratificadora é

Estrato	Proporção	Tamanho do Subgrupo
Professores	0.2	2
Servidores	0.2	2
Alunos	0.6	6
Total	1	10

Exemplo

professores Números sorteados: 2; 0. Professores na amostra: P2, P10;

servidores Números sorteados: 2; 0. Servidores na amostra: S2, S10;

Alunos Números sorteador: 20; 22; 27; 09; 13; 25; 24. Alunos na amostra: A20, A22, A27, A09, A13, A25, A24;

Amostra $\{P2, P10, S2, S10, A20, A22, A27, A09, A13, A25, A24\}.$