# 12. Amostragem Estratificada - AE

Existe uma outra forma de retirar uma amostra da micro-população apresentada na motivação (pág. 112), de modo a satisfazer a condição imposta.

Observe que a micro-população exibe uma estrutura particular.  Qual é essa estrutura ?

## Plano amostral

Em um plano AE, a população sob estudo é particionada em $H$ estratos de tamanhos $N_1 , N_2 , \dots , N_H$ , respectivamente:

$$U = \bigcup_{h=1}^{H} U_h ,$$

onde $U_h = \{ k \in U : k \in estrato\ h \}$ .

Um processo físico de aleatorização é empregado dentro de cada estrato $h$ , independentemente, para gerar uma amostra $s_h$ , de tamanho $n_h$ $(h = 1, 2, \dots, H)$. A amostra final (de tamanho $n$) é composta por todos os elementos selecionados, isto é:

$$s = \bigcup_{h=1}^{H} s_h ,$$

com $n = \sum_{h=1}^{H} n_h$.

Denote por $p_h$ o plano amostral implementado pela aleatorização imposta ao estrato $h$. Como as amostras $s_1, s_2, \ldots, s_H$ foram geradas independentemente, o plano AE atribui probabilidade de seleção da amostra $s$, dada por

$$p(s) = \prod_{h=1}^{H} p_h(s_h) .$$

## Esquema amostral

Um esquema amostral deve ser empregado para implementar cada plano escolhido $p_h$, dentro de cada estrato $h$. Embora não haja restrição imposta, via de regra, escolhe-se um único esquema aplicado a todos os estratos.

Por exemplo, um plano AE com uma amostra de Bernoulli em cada estrato.

O plano mais utilizado, dentro de estratos, é o de uma AAS. Por essa razão, chamaremos de uma amostra aleatória estratificada (AAE), uma amostra gerada pelo plano AE, com AAS em cada estrato.

## Justificativas freqüentes para a estratificação de uma população

- *Inferência para subpopulações*.

  Muitas vezes, antes da pesquisa ir a campo, é possível identificar subpopulações para as quais desejam-se estimativas com precisões pré-especificadas. Neste caso, cada subpopulação corresponderia a um estrato.

- *Razões administrativas/custo*.

  Por exemplo, imagine um levantamento de âmbito geográfico, em que o órgão responsável pela pesquisa tenha escritórios em várias regiões, cobrindo a área total a ser investigada. Neste caso, seria natural que cada região fosse um estrato. O levantamento de campo em cada estrato seria de responsabilidade do escritório regional correspondente.

- *Razões técnicas de estimação*.

  É possível ainda que, para algumas subpopulações específicas, o contexto (existência de informações auxiliares, por exemplo) indique um procedimento diferente de estimação.  Nestes casos, cada subpopulação específica seria um estrato.

## Estimação

Considere a estimação do total populacional de uma população finita sob um plano AE, com $H$ estratos.

Neste caso, o parâmetro de interesse pode ser escrito como:

$$t = \sum_{h=1}^{H} \sum_{k \in U_h} y_k = \sum_{h=1}^{H} t_h$$

**Resultado 12.1:** Sob uma AE, com $H$ estratos, o estimador de H-T assume a forma

$$\hat{t}_{\pi} = \sum_{h=1}^{H} \hat{t}_{h\pi}$$

e tem como variância

$$Var_{AE}(\hat{t}_{\pi}) = \sum_{h=1}^{H} Var_{p_h}(\hat{t}_{h\pi}).$$

Além disso,

$$V\hat{a}r_{AE}(\hat{t}_{\pi}) = \sum_{h=1}^{H} V\hat{a}r_{p_h}(\hat{t}_{h\pi})$$

será um estimador centrado para $Var_{AE}(\hat{t}_{\pi})$, desde que $V\hat{a}r_{p_h}(\hat{t}_{h\pi})$ seja um estimador centrado para $Var_{p_h}(\hat{t}_{h\pi})$, para $h = 1,2,\ldots,H$.

# Prova:

## EXERCÍCIO:

$$\pi_{hk} = \frac{n_h}{N_h}$$

Apresente a forma específica do estimador de H-T para um plano AAE, sua variância e um estimador centrado para a variância.

## 12.1 Alocação ótima, sob um plano AE

Para selecionar uma AE de tamanho $n$ de uma população particionada em $H$ estratos é preciso decidir quantas unidades amostrais selecionar em cada estrato.

Suponha que o custo total do levantamento amostral ($C$) pode ser aproximado por uma função linear do tipo:

$$C = c_0 + \sum_{h=1}^{H} c_h n_h \, . \qquad (11.1)$$

O problema de alocação ótima consiste em responder a uma ou outra das seguintes perguntas pertinentes:

a) Considerando que a verba disponível para execução do levantamento é fixa e dada por $C$, de que forma deve-se alocar o tamanho de amostra (permitido pelo orçamento) entre os estratos, de tal forma a minimizar a variância do estimador que planeja-se utilizar ?

b) Considerando que, dada a importância da pesquisa, o objetivo maior é o de conseguir estimativas com determinada precisão, de que forma deve-se alocar o tamanho de amostra (necessário para atingir o grau de precisão determinado) entre os estratos, de tal forma a minimizar o custo total do levantamento, dado por $C$ ?

**Resultado 12.2:** Considere um plano AE, tal que a variância do estimador de H-T assume a forma

$$Var_{AE}(\hat{t}_\pi) = \sum_{h=1}^{H} \frac{A_h}{n_h} + B = V,$$

onde $A_h$ e $B$ não dependem de $n_h$. Assumindo a função-custo linear (11.1) para $C$, a alocação ótima é dada pela escolha de $n_h$ proporcional ao fator

$$\left( \frac{A_h}{c_h} \right)^{1/2}.$$

**Prova:**

Como $B$ e $c_0$ não dependem de $n_h$, o problema de alocação ótima resume-se a encontrar valores de $n_h$ que minimizem a expressão dada por

$$\phi(n_h) = \left( \sum_{h=1}^{H} \frac{A_h}{n_h} \right) \left( \sum_{h=1}^{H} c_h n_h \right).$$

Defina $\quad a_h = (A_h/n_h)^{1/2}$

e $\quad b_h = (c_h n_h)^{1/2}$.

Assim, pela desigualdade de Cauchy-Schwarz,

$$\phi(n_h) = (\sum_{h=1}^{H} a_h^2)(\sum_{h=1}^{H} b_h^2) \geq \left( \sum_{h=1}^{H} a_h b_h \right)^2$$

$$\therefore \quad \phi(n_h) \geq \left( \sum_{h=1}^{H} (A_h c_h)^{1/2} \right)^2 .$$

$\phi(n_h)$ atinge o mínimo se e somente se

$$\frac{b_h}{a_h} = n_h \left( \frac{c_h}{A_h} \right)^{1/2} = cte\,,$$

o que equivale a $n_h \propto \left( \frac{A_h}{c_h} \right)^{1/2}$.

$\blacksquare$

EXERCÍCIO:

Apresente a forma específica da regra de alocação ótima para um plano AAE.

## 12.2 Alternativas de regras de alocação sob um plano AAE.

i. *Alocação ótima de Neyman (1934)*:

Considerando o custo dentro de cada estrato fixo, isto é, $c_h = c$, para $h = 1,\ldots,H$, a alocação ótima sob um plano AAE - chamada, neste caso, alocação ótima de Neyman - é dada por

$$n_h = n \frac{N_h S_{yU_h}}{\sum_{h=1}^{H} N_h S_{yU_h}}$$

ii. *Alocação ótima em relação a uma variável auxiliar* $x$ *:*

$$n_h = n \frac{N_h S_{xU_h}}{\sum_{h=1}^{H} N_h S_{xU_h}}$$

iii. *Alocação proporcional ao tamanho dos estratos:*

$$n_h = n \frac{N_h}{N}$$

## 12.3 Eficiência de um plano AAE com alocação proporcional ao tamanho do estrato

Sob alocação proporcional ao tamanho do estrato,

$$f_h = \frac{n_h}{N_h} = \frac{n}{N} = f$$

Daí, tem-se que

$$Var_{AAE,pr}(\hat{t}_\pi) = \sum_{h=1}^{H} (1 - f_h) N_h^2 \frac{S_{yU_h}^2}{n_h}$$

$$= (1 - f) \frac{N}{n} \sum_{h=1}^{H} N_h S_{yU_h}^2$$

$$= (1 - f) \frac{N}{n} \left( SQD + \sum_{h=1}^{H} S_{yU_h}^2 \right)$$

onde, $SQD$ é a soma de quadrados entre estratos de uma tabela de ANOVA.

| F.V. | G.L. | S.Q. |
|---|---|---|
| Entre estratos | $H\text{-}1$ | $SQE$ |
| Dentro dos estratos | $N\text{-}H$ | $SQD$ |
| Total | $N\text{-}1$ | $SQT$ |

$$SQE = \sum_{h=1}^{H} N_h (\bar{y}_{U_h} - \bar{y}_U)^2$$

$$SQD = \sum_{h=1}^{H} (N_h - 1) S^2_{yU_h}$$

$$SQD = (N - 1) S^2_{yU}$$

$$Var_{AAS}(\hat{t}_\pi) = \sum_{h=1}^{H} (1-f)N^2 \frac{S_{yU}^2}{n}$$

$$= (1-f)\frac{N^2}{n}\frac{SQT}{N-1}$$

$$= (1-f)\frac{N^2}{n(N-1)}(SQE + SQE)$$

$$= Var_{AAE,pr}(\hat{t}_\pi) +$$

$$+(1-f)\frac{N}{n(N-1)}\left[NSQE - \sum_{h=1}^{H} (N-N_h)S_{yU_h}^2\right]$$

Daí, tem-se que o plano AAE com alocação proporcional ao tamanho do estrato é mais eficiente que um plano AAS, a menos que

$$SQE < \sum_{h=1}^{H} (1 - \frac{N_h}{N})S_{yU_h}^2$$