Cap. 4 Resumo

População dividida em h estratos de Tamanhos N_h estratos

Tabela 4.1. Uma população estratificada

Estrato	Dados	Total	Média	Variância
1	$\boldsymbol{Y}_1$	τ_1	$\mu_1 = \overline{Y}_1$	σ_1^2 ou S_1^2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
h	$\boldsymbol{Y}_h$	τ_h	$\mu_h = \overline{Y}_h$	σ_h^2 ou S_h^2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
H	$\boldsymbol{Y}_H$	τ_H	$\mu_H = \overline{Y}_H$	σ_H^2 ou S_H^2

onde $\boldsymbol{Y}'_h = (Y_{h1}, \ldots, Y_{hN_h})$, é o vetor de dados no estrato h, $h = 1, \ldots, H$.

$N = \sum_{h=1}^{H} N_h$: tamanho do universo;

$W_h = N_h/N$: peso (proporção) do estrato h, com $\sum_{h=1}^{H} W_h = 1$;

$\tau = \sum_{h=1}^{H} \tau_h = \sum_{h=1}^{H} \sum_{i=1}^{N_h} Y_{hi} = \sum_{h=1}^{H} N_h \overline{Y}_h$: total populacional

$\mu = \overline{Y} = \frac{\tau}{N} = \frac{1}{N} \sum_{h=1}^{H} \sum_{i=1}^{N_h} Y_{hi} = \frac{1}{N} \sum_{h=1}^{H} N_h \overline{Y}_h = \sum_{h=1}^{H} W_h \overline{Y}_h$:

$$\sigma^2 = \sigma_d^2 + \sigma_e^2,$$

onde

$$\sigma_d^2 = \sum_{h=1}^{H} W_h \sigma_h^2$$

é a média das variâncias para cada estrato (variância dentro) e

$$\sigma_e^2 = \sum_{h=1}^{H} W_h (\overline{Y}_h - \overline{Y})^2$$

mede a variação das médias dos estratos (chamar-se-á de variância entre estratos).

$\hat{\mu}_h$ um estimador não viesado da média populacional μ_h do estrato h, ou seja, $E_A[\hat{\mu}_h] = \mu_h$, onde A é o plano usado no estrato h.

Então:

Teorema 4.1. *O estimador*

$$T_{es} = \sum_{h=1}^{H} N_h \overline{y}_h$$

é não viesado para o total populacional τ, com

$$Var_A[T_{es}] = \sum_{h=1}^{H} N_h^2 Var_A[\overline{y}_h].$$

Prova. Usando as relações (4.2.5) e (4.2.6) tem-se que para um plano amostral A que

$$E_A[T_{es}] = \sum_{h=1}^{H} N_h E_A[\overline{y}_h] = \sum_{h=1}^{H} N_h \mu_h = \sum_{h=1}^{H} \tau_h = \tau$$

e

$$Var_A[T_{es}] = \sum_{h=1}^{H} N_h^2 Var_A[\overline{y}_h].$$

Corolário 4.1. *O estimador*

$$\overline{y}_{es} = \frac{1}{N}\sum_{h=1}^{H} N_h \overline{y}_h = \sum_{h=1}^{H} W_h \overline{y}_h$$

é um estimador não viesado da média populacional μ *e*

$$Var_A[\overline{y}_{es}] = \sum_{h=1}^{H} W_h^2 Var_A[\overline{y}_h].$$

Corolário 4.2. *Considere agora que dentro de cada estrato a amostra foi sorteada por um processo AASc. Então, tem-se para as duas situações acima as seguintes fórmulas para as variâncias,*

$$Var[T_{es}] = \sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{n_h},$$

e

$$Var[\overline{y}_{es}] = \sum_{h=1}^{H} W_h^2 \frac{\sigma_h^2}{n_h},$$

com estimadores não viesados dados por

$$var[T_{es}] = \sum_{h=1}^{H} N_h^2 \frac{s_h^2}{n_h}$$

e

$$var[\overline{y}_{es}] = \sum_{h=1}^{H} W_h^2 \frac{s_h^2}{n_h}.$$

4.4.1. Alocação proporcional

Neste tipo de procedimento a amostra de tamanho n é distribuída proporcionalmente ao tamanho dos estratos, isto é,

$$n_h = nW_h = n\frac{N_h}{N}. \tag{4.4.1}$$

Este procedimento, é muitas vezes, também chamado de amostragem ”representativa”. Aqui será usada a notação ”Amostra Estratificada Proporcional” (AEpr).

Teorema 4.2. *Com relação a AEPr, o estimador $\overline{y}_{es}$ é igual a média amostral simples $\overline{y}$, com*

$$V_{pr} = Var[\overline{y}_{es}] = \sum_{h=1}^{H} W_h \frac{\sigma_h^2}{n} = \frac{\sigma_d^2}{n}$$

que é estimado por

$$var[\overline{y}_{es}] = \sum_{h=1}^{H} W_h \frac{s_h^2}{n}.$$

Na AAE com a função de custo linear, temos que V_{es} é mínimo para C fixado ou C é mínimo para V_{es} fixado se

$$(4.4.5) \qquad n_h = n\frac{W_h\sigma_h/\sqrt{c_h}}{\sum_{h=1}^{H} W_h\sigma_h/\sqrt{c_h}}, \quad h = 1, \ldots, H.$$

(i) Para C fixado, o tamanho ótimo da amostra é dado por

$$(4.4.11) \qquad n = (C - c_o)\frac{\sum_{h=1}^{H} N_h\sigma_h/\sqrt{c_h}}{\sum_{h=1}^{H} N_h\sigma_h\sqrt{c_h}}.$$

(ii) Para V_{es} fixado, o tamanho ótimo da amostra é dado por

$$n = \frac{(\sum_{h=1}^{H} W_h\sigma_h\sqrt{c_h})(\sum_{h=1}^{H} W_h\sigma_h/\sqrt{c_h})}{V_{es}},$$

onde $W_h = N_h/N$, como antes.

Corolário 4.5. *Para o caso de custo fixo, isto é,*

$$C = c_o + nc,$$

onde c *é o custo por unidade, a alocação ótima se reduz a*

$$n_h = n \frac{N_h \sigma_h}{\sum_{h=1}^{H} N_h \sigma_h}, \tag{4.4.13}$$

$h = 1, \ldots, H$. *Neste caso* V_{es} *reduz–se a*

$$V_{ot} = \frac{(\sum_{h=1}^{H} W_h \sigma_h)^2}{n} = \frac{\overline{\sigma}^2}{n}, \tag{4.4.14}$$

onde $\overline{\sigma} = \sum_{h=1}^{H} W_h \sigma_h$ *é um desvio padrão médio dentro de cada estrato.*

Com relação a AASc, tem-se que

$$V_{ot} \leq V_{pr} \leq V_c.$$

Prova. De acordo com (4.2.4), tem-se que

$$\begin{aligned} N\sigma^2 &= \sum_{h=1}^{H}\sum_{j=1}^{N_h}(Y_{hj} - \overline{Y})^2 \\ &= \sum_{h=1}^{H} N_h\sigma_h^2 + \sum_{h=1}^{H} N_h(\overline{Y}_h - \overline{Y})^2. \end{aligned} \tag{4.4.16}$$

Então, σ^2 em (4.4.16) acima pode ser escrita como

$$\sigma^2 = \sum_{h=1}^{H} W_h \sigma_h^2 + \sum_{h=1}^{H} W_h (\overline{Y}_h - \overline{Y})^2 = \sigma_d^2 + \sigma_e^2.$$

Conseqüentemente, escreve-se

$$\begin{aligned} V_c &= \frac{\sigma_d^2}{n} + \frac{\sigma_e^2}{n} \\ &= V_{pr} + \frac{\sigma_e^2}{n}. \end{aligned}$$

Já que σ_e^2/n é sempre não negativo,

$$V_c \geq V_{pr}.$$

Por construção, sabe-se que $V_{ot} \leq V_{pr}$. Por outro lado, (veja o Exercício 4.19)

$$V_{pr} - V_{ot} = \frac{1}{n}[\sum_{h=1}^{H} W_h\sigma_h^2 - (\sum_{h=1}^{H} W_h\sigma_h)^2]$$

$$= \frac{1}{n}\sum_{h=1}^{H} W_h(\sigma_h - \overline{\sigma})^2 = \frac{\sigma_{dp}^2}{n}, \tag{4.4.17}$$