Estimadores do tipo razâo

População de $\mathcal{U}$ de N unidades

Unidade i: $\rightarrow (X_i, Y_i)$

Y_i: $\rightarrow$ Quantidade de interesse

X_i: $\rightarrow$ Variável auxiliar

Temos para uma amostra AASc de tamanho n

$\bar{x}$: média amostra de X
$\bar{y}$: média amostral de Y

Temos para a variável auxiliar X:

τ_x e $\mu_x = \bar{X}$: total e média populacional de X (v. auxiliar)

Estimadores:
Razão:

$$\bar{R} = \frac{\bar{y}}{\bar{x}}.$$

Média Y

$\mu_y = \bar{Y}$:

$$\hat{\mu}_{Ry} = \hat{R}.\bar{X} = r\bar{X} = \frac{\bar{y}}{\bar{x}}\bar{X}.$$

Total Y

$\tau_Y = N.\mu_y$:

$$\hat{\tau}_{Ry} = N.\hat{\mu}_{Ry} = \hat{R}.\tau_x = r\tau_x = \frac{\bar{y}}{\bar{x}}\tau_x.$$

Propriedades:

Estimadores $\hat{R}$, τ_{Ry}, $\bar{\mu}_{Ry}$ aproximadamente não viesados.

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R \approx \frac{\bar{y} - \bar{x}R}{\mu_x}.$$

Portanto, substituindo $\bar{x}$ por μ_x no denomidador temos

$$E[\hat{R} - R] = E[\frac{\bar{y}}{\bar{x}} - R] \approx \frac{E[\bar{y} - \bar{x}R]}{\mu_x} = 0.$$

Variância:

$$Var[\hat{R}] \approx \frac{1}{\mu_x^2} Var[\bar{y} - R\bar{x}] = \frac{1}{\mu_x^2} Var[\bar{d}],$$

onde $\bar{d} = \sum_{i \in s} D_i$, $D_i = Y_i - RX_i$, de modo que

$$Var[\bar{d}] = \frac{\sigma_R^2}{n},$$

e daí,

$$Var[\hat{R}] \approx \frac{1}{\mu_x^2} Var[\bar{d}] = \frac{\sigma_R^2}{n\mu_x^2},$$

de modo que

$$Var[\hat{\mu}_{Ry}] = \frac{\sigma_R^2}{n},$$

$$Var[\hat{\tau}_{Ry}] = N^2 \frac{\sigma_R^2}{n},$$

Estimadores para as variâncias são obtidos substituindo-se

$$\sigma_R^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - RX_i)^2,$$

por

$$s_R^2 = \hat{\sigma}_R^2 = \frac{1}{n-1}\sum_{i\in s}(Y_i - \hat{R}X_i)^2.$$

Temos os ICs:

$$\hat{R} \mp z_\alpha\sqrt{\frac{s_R^2}{\mu_x^2 n}}.$$

$$\hat{\mu}_y \mp z_\alpha\sqrt{\frac{s_R^2}{n}}.$$

$$\hat{\tau}_y \mp z_\alpha\sqrt{N^2\frac{s_R^2}{n}}.$$

Exemplo: Quantidade de suco em um carregamento de laranjas.
Amostra AASc com n=10 unidades.$\tau_X = 1800kg$.
X: 0.4, 0.48, 0.43,0.42,0.5,0.46,0.39,0.41,0.42,0.44 (Kg.)
Y: 0.021, 0.03, 0.025, 0.022, 0.033, 0.027, 0.019, 0.021, 0.023, 0.025 (Kg.)
de modo que

$$\bar{x} = \sum_{i \in s} X_i = 4.35/10 = 0.435, \qquad \bar{y} = \sum_{i \in s} Y_i = .246/10 = 0.246.$$

$$r = \hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{0.246}{4.35} = 0.05655$$

$$s_R^2 = \frac{1}{9}[(0.021 - .05655(.40))^2 + \cdots + (.025 - .05655(.44))^2] = (.00241)^2$$

Portanto,

$$\hat{\tau}_y = \frac{.246}{4.35} 1800 = (1800)(.05655) = 102.6 \quad kg$$

$$\widehat{Var[\hat{\tau}_y]} = 1800^2 . \frac{s_R^2}{n\mu_x^2} = (1800)^2 \frac{(.00241)^2}{(10).(.435)^2} = 9.949.$$

Portanto,

$$\hat{\tau}_y \mp 1.96\sqrt{9.949} = (94.66, 108.92).$$

O estimador razão estratificado

População esteja dividida em H estratos

$\overline{y}_h$, $\overline{x}_h$ e τ_{Xh} as médias amostrais correspondentes as variáveis Y e X e o total da variável X no estrato h.

No caso da média populacional μ_y, pode-se considerar o seguinte estimador

$$\overline{y}_{Res} = \sum_{h=1}^{H} W_h \frac{\overline{y}_h}{\overline{x}_h} \mu_{xh} = \sum_{h=1}^{H} W_h \overline{y}_{Rh}.$$

Como um estimador do total τ_y:

$$T_{Res} = \hat{\tau}_{Res} = \sum_{h=1}^{H} N_h \overline{y}_{Rh}.$$

$$\bar{y}_{Rh} = \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h.$$

Variância:

$$Var[\overline{y}_{Res}] \simeq \sum_{h=1}^{H} \frac{W_h^2}{n_h}(\sigma_{Yh}^2 + R_h^2\sigma_{Xh}^2 - 2R_h\rho_{XYh}\sigma_{Yh}\sigma_{Xh})$$
$$= \sum_{h=1}^{H} W_h^2 \frac{\sigma_{Rh}^2}{n_h},$$

com

$$\sigma_{Rh}^2 = \sigma_{Yh}^2 + R_h^2\sigma_{Xh}^2 - 2R_h\rho_{XYh}\sigma_{Yh}\sigma_{Xh})$$

$$Var[\overline{y}_{Rh}] \simeq \frac{1}{n_h N_h} \sum_{j=1}^{N_h} (Y_{hj} - R_h X_{hj})^2$$

$$= \frac{1}{n_h}(\sigma^2_{Yh} + R^2_h \sigma^2_{Xh} - 2R_h \rho_{XYh} \sigma_{Yh} \sigma_{Xh}),$$

$$Var[\overline{y}_{Res}] \simeq \sum_{h=1}^{H} W_h^2 \frac{\sigma^2_{Rh}}{n_h},$$

onde

$$\sigma^2_{Rh} = \sigma^2_{yh} - 2R_h \rho_h(X, X) \sigma_{Yh} \sigma_{Xh} + R^2_h \sigma^2_{Xh},$$

$$n_h = n \frac{N_h \sigma_{Rh}/\sqrt{c_h}}{\sum_{h=1}^{H} N_h \sigma_{Rh}/\sqrt{c_h}}, \tag{5.4.2}$$

$h = 1, \ldots, H.$

Tem-se então que dispor de uma estimativa piloto de σ^2_{Rh} em cada estrato para que (5.4.2) seja operacional.

Exemplo 5.3. Considere uma companhia que dispõe de duas indústrias em locais diferentes. O objetivo principal da pesquisa é avaliar se houve variação no tempo médio gasto por um empregado no último ano em relação ao anterior.

A população base é formada pelos empregados das duas indústrias. Essa população será dividida em dois estratos, sendo que o estrato 1 de tamanho $N_1 = 1000$, engloba os empregados da indústria 1 e o estrato 2, de tamanho $N_2 = 1500$, contém os empregados da indústria 2. De cada um dos estratos, uma amostra de tamanho 10 é observada usando a AASs. Em cada uma das amostras observa-se

Y_{ij}: número de horas gastas com visitas ao médico pelo empregado j da indústria i, no corrente ano e

X_{ij}: número de horas gastas com visitas ao médico pelo empregado j da indústria i, no ano anterior.

Obteve-se para o estrato 1,

$n_1 = 10, \overline{y}_1 = 18,7, \overline{x}_1 = 17,8, \tau_{X1} = 16300, s^2_{R1} = 3,47$

e para o estrato 2,

$n_2 = 10, \overline{y}_2 = 4,6, \overline{x}_2 = 7,8, \tau_{X2} = 12800, s^2_{R2} = 9,72.$

Como $N = 2500$, calcula-se

$$\overline{y}_{Res} = \frac{N_1}{N}\overline{y}_{R1} + \frac{N_2}{N}\overline{y}_{R2} = 19,75$$

e também uma estimativa de $V_{Res} = Var[\overline{y}_{Res}]$,

$$\hat{V}_{Res} = (\frac{N_1}{N})^2\frac{s^2_{R1}}{n_1} + (\frac{N_2}{N})^2\frac{s^2_{R2}}{n_2} = 7,22.$$