# Introdução à estimadores razão e do tipo razão

Prof. Caio Azevedo

28 de setembro de 2011

- Suponha que seja de interesse estimar a quantidade de açúcar que pode ser extraída de um caminhão carregado de laranjas. As unidades populacionais são laranjas. Seja então $y_i$ a quantidade de açúcar extraída da laranja i, i = 1,...,N. Tem-se interesse em estimar $\tau_y = \sum_{i=1}^{N} y_i$. O estimador natural seria o estimador expansão, $\widehat{\tau}_y = N\widehat{\mu}_y$. Mas tal estimador não pode ser utilizado, pois não se conhece o número de laranjas no caminhão. Por outro lado, sabe-se que o peso da laranja i, $x_i$, é fortemente (e positivamente) correlacionado com $y_i$, i =1,...,N. Pode-se então definir a razão, quantidade média de açúcar por unidade de peso

$$\begin{aligned} r &= \frac{\tau_y}{\tau_x} = \frac{\mu_y}{\mu_x} \\ \Rightarrow \tau_y &= r\tau_x = \frac{\mu_y}{\mu_x}\tau_x \end{aligned}$$

- Além da média e do total, muitas vezes, a própria razão $r$ é um parâmetro de interesse.
- Podemos, então, definir os seguintes estimadores:

$$\begin{aligned}
\widehat{r} &= \frac{\widehat{\mu}_y}{\widehat{\mu}_x} = \frac{\overline{Y}}{\overline{X}} \\
\widehat{\tau}_R &= \widehat{r}\tau_x \\
\widehat{\mu}_R &= \widehat{r}\mu_x
\end{aligned}$$

Naturalmente, nesses casos, teremos de conhecer $\tau_x$ e $\mu_x$.

# Exemplo

- Considere uma população formada por três domicílios, $\mathcal{U} = \{1, 2, 3\}$ e que se observam as seguintes variáveis: nome (do chefe), sexo, idade, fumante ou não, renda bruta (mensal em salários mínimos) familiar e número de trabalhadores.

# Cont.

Tabela: População de três domicílios

| Variável | Valores | | | Notação |
|---|---|---|---|---|
| Unidade | 1 | 2 | 3 | $i$ |
| nome do chefe | Ada | Beto | Ema | $a_i$ |
| sexo | 0 | 1 | 0 | $x_i$ |
| idade | 20 | 30 | 40 | $y_i$ |
| fumante | 0 | 1 | 1 | $g_i$ |
| renda bruta familiar | 12 | 30 | 18 | $f_i$ |
| n$^o$ de trabalhadores | 1 | 3 | 2 | $t_i$ |

- Considere um plano AASc e n = 2. Além disso, defina $\overline{F}_R = \mu_t \left( \frac{\overline{F}}{\overline{T}} \right)$
- Dessa forma, temos

| **s** | 11 | 12 | 13 | 21 | 22 | 23 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|
| $\overline{F}$ | 12 | 21 | 15 | 21 | 30 | 24 | 15 | 24 | 18 |
| $\overline{T}$ | 1 | 2 | 1,5 | 2 | 3 | 2,5 | 1,5 | 2,5 | 2 |
| $\overline{F}_R$ | 24 | 21 | 20 | 21 | 20 | 19,2 | 20 | 19,2 | 18 |
| $P(\mathbf{s})$ : | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

- De onde obtemos que $\mathcal{E}(\overline{F}) = 20$, $\mathcal{V}(\overline{F}) = 28$, $\mathcal{E}(\overline{F}_R) \approx 20,27$, $\mathcal{V}(\overline{F}_R) \approx 2,52$
- Note que, apesar de $\overline{F}_R$ ser viciado, $EQM(\overline{F}_R) = 2,59 < EQM(\overline{F}) = 28$.

- Em geral, as distribuições exatas dos estimadores razão $(\widehat{r}, \widehat{\tau}_R, \widehat{\mu}_R)$ são difícies de serem obtidas e, além disso, assimétricas.
- Em geral, tais estimadores são viciados, embora tais vícios tendam a 0 com o aumento do tamanho da amostra.
- Para amostras grandes, as distribuições aproximam-se da normal.