Amostragem - Profa. Lane

Estimadores usando H estratos

Parâmetros	Estimadores não viesados	Var_{AASc}	$\widehat{Var}_{AASc}$	$\widehat{Var}_{AASs}$
τ_Y	$T_{es}=\sum_{h=1}^{H} N_h\bar{y}_h$	$\sum_{h=1}^{H} N_h^2\frac{\sigma_h^2}{n_h}$	$\sum_{h=1}^{H} N_h^2\frac{s_h^2}{n_h}$	$\sum_{h=1}^{H} N_h^2(1-f_h)\frac{s_h^2}{n_h}$
μ_Y	$\bar{y}_{es}=\sum_{h=1}^{H} W_h\bar{y}_h$	$\sum_{h=1}^{H} W_h^2\frac{\sigma_h^2}{n_h}$	$\sum_{h=1}^{H} W_h^2\frac{s_h^2}{n_h}$	$\sum_{h=1}^{H} W_h^2(1-f_h)\frac{s_h^2}{n_h}$
P	$\hat{P}_{es}=\sum_{h=1}^{H} W_h\hat{P}_h$	$\sum_{h=1}^{H} W_h^2\frac{P_h(1-P_h)}{n_h}$	$\sum_{h=1}^{H} W_h^2\frac{\hat{P}_h(1-\hat{P}_h)}{n_h-1}$	$\sum_{h=1}^{H} W_h^2(1-f_h)\frac{\hat{P}_h(1-\hat{P}_h)}{n_h}$

Alocação proporcional: $n_h=nW_h$

Alocação ótima: $n_h=n\frac{W_h\sigma_h/\sqrt{c_h}}{\sum_{h=1}^{H} W_h\sigma_h/\sqrt{c_h}}$, $n=C'\frac{\sum_h N_h\sigma_h/\sqrt{c_h}}{\sum_h N_h\sigma_h\sqrt{c_h}}$ para o custo $C'=\sum_h n_h c_h$ e

$n=\frac{1}{Var(\bar{y}_{es})}\left(\sum_h W_h\sigma_h/\sqrt{c_h}\right)\left(\sum_h W_h\sigma_h\sqrt{c_h}\right)$ para $Var(\bar{y}_{es})$ fixada.

Estimadores Razão e seus derivados

Para obter as variâncias dos estimadores baseados na razão para AASs, basta trocar σ_R^2 por S_R^2. Usamos a t_{n-1}.

Parâmetros	Estimadores não viesados para n grande	Var_{AASc}	$\widehat{Var}_{AASc}$	$\widehat{Var}_{AASs}$
$R=\frac{\tau_Y}{\tau_X}=\frac{\mu_Y}{\mu_X}$	$\widehat{R}=r=\frac{\bar{y}}{\bar{x}}$	$\frac{\sigma_R^2}{n\mu_X^2}$	$\frac{s_R^2}{n\mu_X^2}$	$(1-f)\frac{s_R^2}{n\mu_X^2}$
τ_Y	$T_R=\widehat{\tau_Y}=r\tau_X$	$N^2\frac{\sigma_R^2}{n}$	$N^2\frac{s_R^2}{n}$	$N^2(1-f)\frac{s_R^2}{n}$
μ_Y	$\bar{y}_R=\widehat{\mu_Y}=r\mu_X$	$\frac{\sigma_R^2}{n}$	$\frac{s_R^2}{n}$	$(1-f)\frac{s_R^2}{n}$
$\sigma_R^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-RX_i)^2$	$AAS_c: s_R^2=\frac{1}{n-1}\sum_{i\in s}(Y_i-rX_i)^2$			
$S_R^2=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$	$AAS_s: s_R^2=\frac{1}{n-1}\sum_{i\in s}(Y_i-rX_i)^2$			

Estimador Regressão para AASs: $y_i=\beta_0+\beta_1x_i+e_i$

$$\begin{array}{lcl}
\bar{y}_{Reg} & = & \widehat{\beta_0}+\widehat{\beta_1}\mu_X=\bar{y}+\widehat{\beta_1}(\mu_X-\bar{x}) \text{ estimador de } \mu_Y\\
T_{Reg} & = & \widehat{\tau}_{Reg}=N\bar{y}_{Reg} \text{ para o total } \tau_Y\\
\hat{\beta}_0 & = & \bar{y}-\hat{\beta}_1\bar{x} \qquad \hat{\beta}_1=\dfrac{\sum_{i\in s}(Y_i-\bar{y})(X_i-\bar{x})}{\sum_{i\in s}(X_i-\bar{x})^2}\\
\widehat{Var}(\bar{y}_{Reg}) & = & QMErro(1-f)\left[\frac{1}{n}+\frac{(\mu_X-\bar{x})^2}{\sum_{i\in s}(x_i-\bar{x})^2}\right]
\end{array}$$

com $QMErro=\hat{Var}(e_i)=\frac{\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\hat{\beta}_1x_i\right)^2}{n-2}$ correspondendo ao quadrado médio do erros, que é o estimador não viesado da variância do erro no modelo de regressão linear. Usamos a t_{n-2}.

AASs de Conglomerados em único estágio: Estimadores para o total e para a média

Considere que há N conglomerados na população e cada conglomerado tem M_i unidades amostrais com $K=\sum_{i=1}^{N} M_i$. Temos AASs de n conglomerados.

O estimador NÃO VICIADO do total é $\widehat{\tau}_{cl}$ e o estimador razão $\widehat{\tau}_r$ é viesado.

Parâmetro	Estimador	Variância	Variância estimada
τ	$\widehat{\tau_{cl}}=\frac{N}{n}\sum_{i}^{n}\tau_i$	$Var(\widehat{\tau_{cl}})=\frac{N^2(1-f)}{n}S_t^2$	$\widehat{Var}(\widehat{\tau_{cl}})=N^2(1-f)\frac{s_t^2}{n}$
$\mu=\frac{\tau}{K}$	$\overline{y_{cl}}=\frac{\widehat{\tau}}{K}$	$Var(\overline{y_{cl}})=\frac{Var(\widehat{\tau_{cl}})}{K^2}$	$\widehat{Var}(\overline{y_{cl}})=\frac{\widehat{Var}(\widehat{\tau_{cl}})}{K^2}$
$S_t^2=\frac{\sum_{i=1}^{N}(\tau_i-\frac{\tau}{N})^2}{N-1}$	$s_t^2=\frac{1}{n-1}\sum_{i}^{n}\left(\hat{\tau}_i-\frac{\widehat{\tau}_{cl}}{N}\right)^2$		
τ	$\widehat{\tau}_r=rK$	$Var(\widehat{\tau}_r)=N^2(1-f)\frac{S_r^2}{n}$	$\widehat{Var}(\widehat{\tau}_r)=N^2(1-f)\frac{s_r^2}{n}$
$\mu=R=\frac{\tau}{K}$	$r=\frac{\sum_{i}^{n}\tau_i}{\sum_{i}^{n}M_i}$		$\widehat{Var}(\widehat{r})=(1-f)\frac{1}{\overline{M}^2}\frac{s_r^2}{n}$
S_r^2	$s_r^2=\frac{\sum_{i}^{n}(\tau_i-rM_i)^2}{n-1}$		

$\overline{M}=\frac{\sum_{i=1}^{N} M_i}{N}$ é o tamanho médio populacional dos conglomerados.

$\hat{\tau}_i=\tau_i=\sum_{j=1}^{M_i} y_{ij}$ é o total do conglomerado i e é conhecido para os conglomerados sorteados, pois observamos todos os elementos do conglomerado sorteado quando usamos um único estágio.

A proporção $P=\frac{\sum\tau_i}{\sum M_i}$ pode ser estimada pelo estimador razão $\widehat{p}_{cl}=\frac{\sum_{i=1}^{n}\tau_i}{\sum_{i=1}^{n}M_i}$ com variância $\frac{1}{n(n-1)\overline{M}^2}\sum(\tau_i-\widehat{p}_{cl}M_i)^2$.