

Introdução à estimadores regressão

Prof. Caio Azevedo

Motivação

- Vimos que, quando há informações (em nível populacional) sobre uma variável auxiliar (x_i) e que a relação desta com a variável de interesse (y_i) puder ser expressa como uma reta passando pela origem (além de outros fatores: magnitude dos coeficientes de variação e da correlação entre as duas variáveis), estimadores razão podem ser mais apropriados do que os estimadores usuais.
- Tal abordagem também pode ser útil quando não conhecemos, por exemplo, o tamanho da população (N).

Cenário

- Exploraremos, agora, estimadores que são mais apropriados quando a relação entre tais variáveis puder ser expressa como uma reta que não passa pela origem, ou seja: $y_i = \alpha + \beta x_i + e_i$ (nível populacional) $Y_i = \alpha + \beta X_i + E_i$ (nível amostral).
- Consideraremos as mesmas notações e suposições (acerca dos parâmetros populacionais da variável x) feitas para o estimador razão.

Cenário

- O estimador regressão para a média é definido como:

$$\hat{\mu}_{Reg} = \hat{\mu} + b(\mu_x - \bar{X})$$

em que b é uma estimativa apropriada de β (consideraremos, portanto, em princípio, que b não é aleatório) e $\hat{\mu} = \bar{Y}$.

- Note que se $b > 0$ e \bar{x} é pequeno, com relação à μ_x então, devido à linearidade entre y e x , a diferença entre $\hat{\mu}_{Reg}$ e $\hat{\mu}$ também é pequena.

Propriedades dos estimadores

- Note ainda que o estimador $\hat{\mu}_{Reg}$ faz uma “correção” em $\hat{\mu}$, isto é, adiciona a $\hat{\mu}$ uma quantidade proporcional a $\mu_x - \bar{X}$, ou seja, $b(\mu_x - \bar{X})$.
- O estimador regressão para o total populacional é dado por
$$\hat{\tau}_{Reg} = N\hat{\mu}_{Reg}.$$
- Consideraremos, inicialmente, que $b = b_0$ (é um valor conhecido).
- Note que, nesse caso, temos duas variáveis aleatórias no estimador $\hat{\mu}_{Reg}$, nomeadamente $\hat{\mu}$ e \bar{X} .

Propriedades dos estimadores

- Temos que

$$\begin{aligned}\mathcal{E}(\hat{\mu}_{Reg}) &= \mathcal{E}[\hat{\mu} + b_0(\mu_x - \bar{X})] = \mathcal{E}(\hat{\mu}) + b_0[\mu_x - \mathcal{E}(\bar{X})] \\ &= \mu_y + b_0(\mu_x - \mu_x) = \mu_y\end{aligned}$$

- Para o cálculo da variância, defina

$D_i = Y_i - b_0(X_i - \mu_x), i = 1, 2, \dots, n$ e note que

$$\begin{aligned}\hat{\mu}_{Reg} &= \frac{1}{n} \sum_{i=1}^n Y_i - b_0 \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu_x \right) = \frac{1}{n} \sum_{i=1}^n [Y_i - b_0(X_i - \mu_x)] \\ &= \frac{1}{n} \sum_{i=1}^n D_i = \bar{D}\end{aligned}$$

Propriedades dos estimadores

- Defina agora $d_i = y_i - b_0(x_i - \mu_x)$. Note que

$$\begin{aligned}\mu_d &= \frac{1}{N} \sum_{i=1}^N d_i = \frac{1}{N} \sum_{i=1}^N [y_i - b_0(x_i - \mu_x)] \\ &= \frac{1}{N} \sum_{i=1}^N y_i - b_0 \left(\frac{1}{N} \sum_{i=1}^N x_i - \mu_x \right) = \mu_y - b_0(\mu_x - \mu_x) = \mu_y\end{aligned}$$

■ E

$$\begin{aligned}
 \sigma_d^2 &= \frac{1}{N} \sum_{i=1}^N [d_i - \mu_d]^2 = \frac{1}{N} \sum_{i=1}^N [y_i - b_0(x_i - \mu_x) - \mu_y]^2 \\
 &= \frac{1}{N} \sum_{i=1}^N [(y_i - \mu_y) - b_0(x_i - \mu_x)]^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \left[(y_i - \mu_y)^2 - 2b_0(x_i - \mu_x)(y_i - \mu_y) + b_0^2(x_i - \mu_x)^2 \right]^2 \\
 &= \frac{1}{N} \sum_{i=1}^N (y_i - \mu_y)^2 - 2b_0 \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \\
 &\quad + b_0^2 \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2 \\
 &= \sigma_y^2 - 2b_0\sigma_{xy} + b_0^2\sigma_x^2
 \end{aligned}$$

Propriedades dos estimadores

- Em que $\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu_y)^2$ (analogamente para s_x^2),
 $\sigma_{xy} = \text{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$.
- Portanto, temos que

$$\mathcal{V}_{A_1}(\hat{\mu}_{Reg}) = \mathcal{V}_{A_1}(\bar{D}) = \frac{\sigma_d^2}{n}$$

$$\mathcal{V}_{A_2}(\hat{\mu}_{Reg}) = \mathcal{V}_{A_2}(\bar{D}) = (1 - f) \frac{s_d^2}{n}$$

em que $f = \frac{n}{N}$, $s_d^2 = s_y^2 - 2b_0 s_{xy} + b_0^2 s_x^2$, $s_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2$
(analogamente para s_x^2) e $s_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$.

- Estimadores não viesados para as variâncias anteriores são dados por:

$$\begin{aligned}\hat{V}_{A_1}(\hat{\mu}_{Reg}) &= \frac{\hat{\sigma}_d^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n [(Y_i - \bar{Y}) - b_0(X_i - \bar{X})]^2 \\ &= \frac{1}{n} (\hat{\sigma}_y^2 - 2b_0\hat{\sigma}_{xy} + b_0^2\hat{\sigma}_x^2)\end{aligned}$$

$$\begin{aligned}\hat{V}_{A_2}(\hat{\mu}_{Reg}) &= (1-f)\frac{\hat{s}_d^2}{n} = \frac{1-f}{n(n-1)} \sum_{i=1}^n [(Y_i - \bar{Y}) - b_0(X_i - \bar{X})]^2 \\ &= \frac{1}{n} (\hat{s}_y^2 - 2b_0\hat{s}_{xy} + b_0^2\hat{s}_x^2)\end{aligned}$$

- Em que $\hat{\sigma}_d^2 = \hat{s}_d^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \hat{\mu}_{Reg})^2$,
 $\hat{\sigma}_y^2 = \hat{s}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$, $\hat{\sigma}_x^2 = \hat{s}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ e
 $\hat{\sigma}_{xy} = \hat{s}_{xy} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})$.
- Prova: Basta utilizar algumas propriedades do valor esperado e lembrar que $\mathcal{E}_{A_1}(\hat{\sigma}_y^2) = \sigma_y^2$, $\mathcal{E}_{A_1}(\hat{\sigma}_x^2) = \sigma_x^2$, $\mathcal{E}_{A_2}(\hat{s}_y^2) = \sigma_y^2$,
 $\mathcal{E}_{A_2}(\hat{s}_x^2) = \sigma_x^2$, $\mathcal{E}_{A_1}(\hat{\sigma}_{xy}) = \sigma_{xy}$, $\mathcal{E}_{A_1}(\hat{s}_{xy}) = s_{xy}$.

Otimidade do estimador regressão

- O valor de b_0 que minimiza $\mathcal{V}_{AE_1}(\hat{\mu}_{Reg})$ é dado por

$$B_0 = \frac{\sum_{i=1}^N (y_i - \mu_y)(x_i - \mu_x)}{\sum_{i=1}^N (x_i - \mu_x)^2} = \frac{\sigma_{xy}}{\sigma_x^2} \quad (1)$$

Além disso, para o b_0 acima, $\mathcal{V}_{A_1(min)}(\hat{\mu}_{Reg}) = \frac{\sigma_y^2}{n} (1 - \rho^2[x, y])$,
 $\rho[x, y] = \sigma_{xy}/(\sigma_x\sigma_y)$.

- Prova: Seja $b_0 = B_0 + c$, $c \in \Re$. Tem-se, então, para este b_0 que

$$\begin{aligned}\mathcal{V}_{A_1}(\widehat{\mu}_{Reg}) &= \frac{1}{n} [\sigma_y^2 - 2(B_0 + c)\sigma_{xy} + (B_0 + c)^2\sigma_x^2] \\ &= \frac{1}{n} \left(\sigma_y^2 - 2\frac{\sigma_{xy}^2}{\sigma_x^2} - 2c\sigma_{xy} + \frac{\sigma_{xy}^2}{\sigma_x^2} + 2c\sigma_{xy} + c^2\sigma_x^2 \right) \\ &= \frac{1}{n} \left\{ \left(\sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} \right) + c^2\sigma_x^2 \right\}\end{aligned}$$

que é mínimo quanto $c = 0$, lembrando que $\rho[x, y] = \sigma_{xy}/\sigma_x\sigma_y$.

- Como na prática não é possível utilizar B_0 (parâmetro), devemos considerar um estimador apropriado, ou seja:

$$\hat{B}_0 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- Como estimador para $\mathcal{V}_{A_1}(\hat{\mu}_{Reg})$ consideramos

$$\hat{\mathcal{V}}_{A_1}(\hat{\mu}_{Reg}) = \frac{1}{n} \left(\hat{\sigma}_y^2 - 2\hat{B}_0\hat{\sigma}_{xy} + \hat{B}_0^2\hat{\sigma}_x^2 \right)$$

- Note que, neste caso, tal estimador não é, necessariamente, não viciado.
- Exercício: Repetir os desenvolvimentos acima para $\mathcal{V}_{A_2}(\hat{\mu}_{Reg})$.

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.
- Considere que o objetivo é estimar \bar{f} a média da renda bruta familiar.

Cont.

Variável	Valores			Notação
Unidade	1	2	3	i
nome do chefe	Ada	Beto	Ema	a_i
sexo (0: F, 1:M)	0	1	0	x_i
idade	20	30	40	y_i
fumante (0: N, 1:S)	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 $\bar{F}_R = \mu_t \left(\frac{\bar{F}}{\bar{T}} \right)$
- Dessa forma, temos que:

s	11	12	13	21	22	23	31	32	33
\bar{F}	12	21	15	21	30	24	15	24	18
\bar{T}	1	2	1,5	2	3	2,5	1,5	2,5	2
\bar{F}_{Reg}	21,0	21,0	19,5	21	21,0	19,5	19,5	19,5	18,0
$P(\mathbf{s})$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

- De onde obtemos que $\mathcal{E}(\bar{F}_{Reg}) = 20$, $\mathcal{V}(\bar{F}_{Reg}) = 1$, $\mathcal{E}(\bar{F}_R) \approx 20,27$, $\mathcal{V}(\bar{F}_R) \approx 2,52$.

- Neste caso temos que $EQM(\bar{F}_{Reg}) = 1 < EQM(\bar{F}_R) = 2,59$ (este último obtido em http://www.ime.unicamp.br/~cnaber/aula_Erazaao%20Amost%20S%202018.pdf, página 9). Em geral, tal resultado é válido.
- Podemos ainda obter a variância do estimador regressão utilizando a fórmula obtida anteriormente. Note que $\mu_t = 2$, $\mu_f = 20$, $\sum_{i=1}^3 f_i t_i = 138$, $\sum_{i=1}^3 f_i^2 = 1368$, $\sum_{i=1}^3 t_i^2 = 14$. Assim,

$$B_0 = \frac{\sum_{i=1}^N f_i t_i - N\bar{f}\bar{t}}{\sum_{i=1}^N t_i^2 - N\bar{t}} = \frac{18}{2} = 9$$

- Também, com $b_0 = B_0 = 9$, $\sigma_F^2 = 56$, $\sigma_t^2 = 2/3$ e $\sigma_{ft} = 6$, tem-se, para $n = 2$, que

$$\mathcal{V}(\bar{F}_{Reg}) = \frac{1}{n} (\sigma_f^2 - 2b_0\sigma_{ft} + b_0^2\sigma_t^2) = 1.$$

Comparação entre os estimadores razão e regressão

- Considerando B_0 dado em 1 e sob AAS_c , temos que

1 $\mathcal{V}(\hat{\mu}_{Reg}) \leq \mathcal{V}(\hat{\mu})$.

2 $\mathcal{V}(\hat{\mu}_{Reg}) \leq \mathcal{V}(\hat{\mu}_R)$.

- Prova: Temos que $\mathcal{V}(\hat{\mu}) = \frac{\sigma_y^2}{n}$,
 $\mathcal{V}(\hat{\mu}_{Reg}) = \frac{\sigma_y^2}{n} (1 - \rho[x, y]^2) = \mathcal{V}(\hat{\mu}) (1 - \rho[x, y]^2)$, de onde segue o item 1), pois $0 \leq \rho[x, y]^2 \leq 1$.

Comparação entre os estimadores razão e regressão

- Por outro lado, temos que

$$\mathcal{V}(\hat{\mu}_R) \approx \frac{1}{n} (\sigma_y^2 - 2\rho[x, y]\sigma_x\sigma_y + r^2\sigma_x^2)$$

- Portanto, vem que

$$\begin{aligned}\mathcal{V}(\hat{\mu}_R) - \mathcal{V}(\hat{\mu}_{Reg}) &\approx \frac{\sigma_y^2}{n} - \frac{2}{n}\rho[x, y]\sigma_x\sigma_y + \frac{1}{n}r^2\sigma_x^2 - \frac{\sigma_y^2}{n} + \frac{\sigma_y^2}{n}\rho[x, y]^2 \\ &= \frac{1}{n} (\rho^2[x, y]\sigma_y^2 - 2r\rho[x, y]\sigma_x\sigma_y + r^2\sigma_x^2) \\ &= (\rho[x, y]\sigma_y - r\sigma_x)^2 \geq 0\end{aligned}$$

De onde se conclui o item 2).

Comparação entre os estimadores razão e regressão

- Para o total, temos que

$$\mathcal{E}(\widehat{\tau}_{Reg}) = \mathcal{E}(N\widehat{\mu}_{Reg}) = N\mathcal{E}(\widehat{\mu}_{Reg}) = N\mu_y = \tau_y.$$

- Variâncias do estimador: $\mathcal{V}_{A_i}(\widehat{\tau}_{Reg}) = N^2\mathcal{V}_{A_i}(\widehat{\mu}_{Reg})$,

$$\mathcal{V}_{A_1}(\widehat{\tau}_{Reg}) = N^2\frac{\sigma_d^2}{n} \text{ e } \mathcal{V}_{A_2}(\widehat{\tau}_{Reg}) = (1-f)N^2\frac{s_d^2}{n}.$$

- Estimativas da variância do estimador: $\widehat{\mathcal{V}}_{A_1}(\widehat{\tau}_{Reg}) = N^2\frac{\widehat{\sigma}_d^2}{n}$ e

$$\widehat{\mathcal{V}}_{A_2}(\widehat{\tau}_{Reg}) = (1-f)N^2\frac{\widehat{s}_d^2}{n}.$$

- Intervalos de confiança, testes de hipótese, e tamanho de amostra podem ser obtidos de forma semelhante aos estimadores usuais (para a média e total) sob AAS_c e AAS_s .

Comentários

- À rigor, portanto, o estimador regressão (para a média) é dado por:

$$\hat{\mu}_{Reg}^* = \hat{\mu} + \hat{B}_0 (\mu_x - \bar{X})$$

- Note que as maioria das expressões (esperança e variância dos estimadores regressão) foram obtidas considerando-se $b_0(B_0)$ fixo (não aleatório). Como $\hat{B}_0 \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{P} B_0$, então os resultados são válidos, para n ($N-n$) suficientemente grandes.

- Ou seja, $|\hat{\mu}_{Reg}^* - \hat{\mu}_{Reg}| \xrightarrow[n \rightarrow \infty, N-n \rightarrow \infty]{P} 0$.

- Assim, $\mathcal{E}_{PA}(\hat{\mu}_{Reg}^*) \approx \mathcal{E}_{PA}(\hat{\mu}_{Reg})$ e $\mathcal{V}_{PA}(\hat{\mu}_{Reg}^*) \approx \mathcal{V}_{PA}(\hat{\mu}_{Reg})$

Teoria assintótica

- Para n e N suficientemente grandes, a convergência em distribuição para a normal padrão é válida, de modo semelhante ao caso dos estimadores usuais sob AAS_c e AAS_s (PA: indica um plano amostral apropriado).
- Portanto, sob certas condições, em relação à estimação da média, temos que

$$\frac{\hat{\mu}_{Reg} - \mu}{EP_{PA}(\hat{\mu}_{Reg})} \xrightarrow[n \rightarrow \infty, N - n \rightarrow \infty]{D} N(0, 1)$$

$$\frac{\hat{\mu}_{Reg} - \mu}{\widehat{EP}_{PA}(\hat{\mu}_{Reg})} \xrightarrow[n \rightarrow \infty, N - n \rightarrow \infty]{D} N(0, 1)$$

Intervalos de confiança

- Analogamente aos casos anteriores, temos que um intervalo assintótico, com coeficiente de confiança de aproximadamente γ , é dado por:

$$IC(\mu, \gamma) \approx \left[\hat{\mu}_{Reg} - z_\gamma \widehat{EP}_{PA}(\hat{\mu}_{Reg}); \hat{\mu}_{Reg} + z_\gamma \widehat{EP}_{PA}(\hat{\mu}_{Reg}) \right]$$

- Erro da estimativa: $z_\gamma \widehat{EP}_{PA}(\hat{\mu}_{Reg})$.

Testes de Hipótese

- Hipóteses usuais (μ_0 conhecido)

- 1 $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$.

- 2 $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$.

- 3 $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

- Estatística do teste $Z_t = \frac{\hat{\mu}_{Reg} - \mu_0}{\widehat{EP}_{PA}(\hat{\mu}_{Reg})}$.

- Sob H_0 , vimos que $Z_t \approx N(0, 1)$, para n e $N-n$ suficientemente grandes.

Testes de Hipótese

- Defina $z_t = \frac{\tilde{\mu}_{Reg} - \mu_0}{\widehat{EP}_{PA}(\hat{\mu}_{Reg})}$ o valor calculado da estatística do teste e z_c o(s) valor(es) crítico(s).
- Defina ainda $Z \sim N(0, 1)$. Os mecanismos de tomada de decisão são basicamente aqueles descritos nas páginas 27 a 29 dos slides http://www.ime.unicamp.br/~cnaber/aula_AAS%20sem%20reposicao%20parte%201%20Amost%202S%202018.pdf

Determinação do tamanho amostral

- Estabelece-se algum critério de interesse acerca da acurácia/precisão na estimativa da média populacional.
- Sob o estimador proposto, calcula-se o tamanho da amostra, com base em sua distribuição assintótica obtida e critério estabelecido.
- Erro de estimativa: $z_\gamma \widehat{EP}_{PA}(\widehat{\mu}_{Reg})$. Fixa-se um erro de estimativa de interesse.
- Probabilidade do módulo da diferença $P(|\widehat{\mu}_{Reg} - \mu| < \delta) > \gamma$,
 $\delta > 0$, $\gamma \in (0, 1)$.

Determinação do tamanho amostral: erro da estimativa

$$\delta = z_\gamma \sqrt{\frac{\sigma_d^2}{n}} \rightarrow n = \frac{z_\gamma^2 \sigma_d^2}{\delta^2}$$

Em geral, o (um) valor de σ_R^2 é obtido através de pesquisas anteriores ou de uma amostra piloto, de tamanho apropriado.

Determinação do tamanho amostral: erro da estimativa

$$\begin{aligned}\delta &= z_\gamma \sqrt{\frac{(1-f)s_d^2}{n}} \rightarrow \left(\frac{1}{n} - \frac{1}{N}\right) = \frac{\delta^2}{z_\gamma^2 s_d^2} \rightarrow \frac{1}{n} = \frac{\delta^2}{z_\gamma^2 s_d^2} + \frac{1}{N} \\ \rightarrow \frac{1}{n} &= \frac{\delta^2 N + z_\gamma^2 s_d^2}{N z_\gamma^2 s_d^2} \rightarrow n = \frac{N z_\gamma^2 s_d^2}{\delta^2 N + z_\gamma^2 s_d^2} = \frac{1}{\frac{\delta^2}{s_d^2 z_\gamma^2} + \frac{1}{N}}\end{aligned}$$

Em geral, o (um) valor de s_d^2 é obtido através de pesquisas anteriores ou de uma amostra piloto, de tamanho apropriado.

Exercício: Construir IC's e testes de hipótese para o total populacional.

Sob amostragem estratificada

- Se a população está estratificada, podemos considerar uma combinação dos resultados obtidos, anteriormente, sob AE, com os resultados obtidos para o estimador razão.
- A estrutura é a mesma daquela apresentada em http://www.ime.unicamp.br/~cnaber/aula_AE%20Amost%20S%202018.pdf.
- Defina: \bar{Y}_h , \bar{X}_h , μ_{xh} , as médias amostrais das variáveis y e x e a média populacional da variável x , respectivamente, no estrato h .

Sob amostragem estratificada

- Como estimadores para a média e o total populacionais, podemos considerar, respectivamente:

$$\hat{\mu}_{Reges} = \sum_{h=1}^H W_h [\hat{\mu}_h + b_{0h} (\mu_{xh} - \bar{X}_h)] = \sum_{h=1}^H W_h \hat{\mu}_{Regh}$$

$$\hat{\tau}_{Reges} = N \hat{\mu}_{Reges} = \sum_{h=1}^H N_h \hat{\mu}_{Regh}$$

Sob amostragem estratificada

- Usando resultados anteriores, temos que:

$$\mathcal{V}_{AE_1}(\widehat{\mu}_{Res}) = \sum_{h=1}^h W_h^2 \frac{\sigma_{dh}^2}{n_h}$$

$$\mathcal{V}_{AE_2}(\widehat{\mu}_{Res}) = \sum_{h=1}^h W_h^2 (1 - f_h) \frac{s_{dh}^2}{n_h}$$

em que $f_h = \frac{n_h}{N_h}$, $\sigma_{dh}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (\sigma_{yh}^2 - 2b_0\sigma_{xyh} + b_0^2\sigma_x^2)$ e
 $s_{dh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (s_{yh}^2 - 2b_0s_{xyh} + b_0^2\sigma_{xh}^2)$.

Sob amostragem estratificada

- Estimadores para as variâncias são obtidos substituindo-se as quantidades desconhecidas por estimadores apropriados, ou seja:

$$\widehat{V}_{AE_1}(\widehat{\mu}_{Res}) = \sum_{h=1}^h W_h^2 \frac{\widehat{\sigma}_{dh}^2}{n_h}$$

$$\widehat{V}_{AE_2}(\widehat{\mu}_{Res}) = \sum_{h=1}^h W_h^2 (1 - f_h) \frac{\widehat{S}_{dh}^2}{n_h}$$

em que $\widehat{\sigma}_{dh}^2 = \widehat{S}_{dh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(\widehat{S}_{yh}^2 - 2b_0 \widehat{S}_{xyh} + b_0^2 \widehat{\sigma}_{xh}^2 \right)$.

Sob amostragem estratificada

- Exercício: desenvolver as fórmulas das variâncias e dos estimadores das variâncias para o estimador do total populacional \hat{T}_{Res} .
- O comportamento assintótico dos estimadores são semelhantes àquele relativos aos estimadores usuais sob AE.
- Intervalos de confiança e testes de hipótese podem ser desenvolvidos de modo semelhante ao exposto anteriormente.

Alocação ótima

- Sob alocação ótima e AAS_c dentro de cada extrato, como visto em http://www.ime.unicamp.br/~cnaber/aula_AE%20Amost%20S%202018.pdf, temos que:

$$n_h = n \frac{N_h \sigma_{dh} / \sqrt{c_h}}{\sum_{h=1}^H N_h \sigma_{dh} / \sqrt{c_h}}$$

- Raciocínio análogo pode ser considerado sob AAS_s dentro de cada extrato.

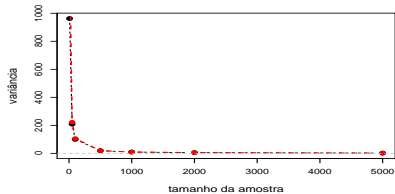
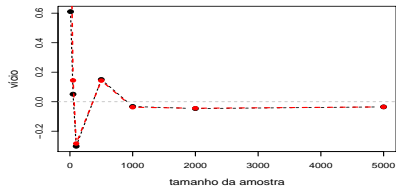
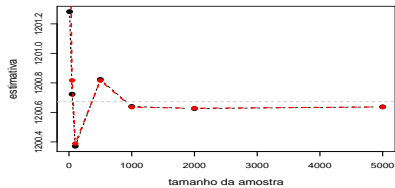
Comparaç3o entre os estimadores raz3o e regress3o

- Duas vari3aveis (y-resposta, x-auxiliar) foram simuladas considerando $N=100000$ em dois cen3arios: $Y_i = 1,2X_i + E_i$ (cen3ario 1),
 $Y_i = 100 + 1,2X_i + E_i$ (cen3ario 2), $X_i \overset{i.i.d.}{\sim} N(1000, 10000)$ e
 $E_i \overset{i.i.d.}{\sim} N(0, 10000)$.
- Objetivo: estimar μ_y .

Comparaçãõ entre os estimadores razão e regressãõ

- R= 1000 amostras foram selecionadas, sob diversos tamanhos $n = (10, 50, 100, 500, 1000, 2000, 5000)'$.
- Foram calculados a média, vício, variância e raiz quadrática do erro quadrático médio dos estimadores razão e regressão (respectivamente em preto e vermelho nos dois gráficos a seguir).
As linhas em cinza representam o valor verdadeiro de μ_y (primeiro gráfico) ou o valor zero (gráficos restantes).

Cenário 1



Cenário 2

