

Introdução à estimadores razão e do tipo razão

Prof. Caio Azevedo

Motivação

- Além da construção/utilização de planos amostrais apropriados, uma outra forma de melhorar o processo de inferência é utilizar informação de variáveis auxiliares (**covariáveis**).
- O objetivo é produzir **estimadores mais precisos** (vício, variância, eqm).
- Uma outra situação de interesse consiste em não se dispor de certas informações necessárias para se utilizar os estimadores vistos até o momento.
- Nos focaremos nos planos **AAS_c** e **AAS_s**. Contudo, os desenvolvimentos aqui apresentados poderão ser estendidos para outros planos.

Motivação

- Suponha que seja de interesse estimar a quantidade de açúcar que pode ser extraída de um caminhão carregado de laranjas. As unidades populacionais são (as) laranjas.
- Seja, então, y_i a quantidade de açúcar extraída da laranja i , $i = 1, \dots, N$.
- Tem-se interesse em estimar $\tau_y = \sum_{i=1}^N y_i$.
- O estimador natural seria o **estimador expansão**, $\hat{\tau}_y = N\hat{\mu}_y$.

Motivação

- 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 forte (e positivamente) correlacionado com y_i , $i = 1, \dots, N$.
- Pode-se então definir a razão, quantidade média de açúcar por unidade de peso

$$r = \frac{\tau_y}{\tau_x} = \frac{\mu_y}{\mu_x}$$
$$\Rightarrow \tau_y = r\tau_x = \frac{\mu_y}{\mu_x}\tau_x$$

em que μ_y , μ_x e τ_x são, respectivamente, a quantidade média de açúcar, o peso médio e o peso total das laranjas do caminhão.

Cont.

- Além da média e do total, muitas vezes, a própria razão (r) é um parâmetro de interesse.
- Tais situações ocorrem, por exemplo, quando se tem interesse em se comparar determinadas quantidades em períodos sucessivos.
- Exemplo específico: razão de vendas de automóveis entre dois anos consecutivos.

Cont.

- Podemos, então, definir os seguintes estimadores:

$$\hat{r} = \frac{\hat{\mu}_y}{\hat{\mu}_x} = \frac{\bar{Y}}{\bar{X}}$$
$$\hat{\tau}_R = \hat{r}\tau_x \quad (1)$$

$$\hat{\mu}_R = \hat{r}\mu_x \quad (2)$$

Naturalmente, nos casos (1) e (2), teremos de conhecer τ_x e μ_x .

Exemplo hipotético

- 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
nº de trabalhadores	1	3	2	t_i

Cont.

- 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}_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

Cont.

- Da tabela anterior, (exercício) temos que $\mathcal{E}(\bar{F}) = 20$, $\mathcal{V}(\bar{F}) = 28$,
 $\mathcal{E}(\bar{F}_R) \approx 20,27$, $\mathcal{V}(\bar{F}_R) \approx 2,52$, $B(\bar{F}) = 0,27$,
 $EQM(\bar{F}_R) = \mathcal{V}(\bar{F}_R) + B(\bar{F}_R)^2 = 2,52 + 0,27^2 = 2,59$.
- Note que, apesar de \bar{F}_R ser viciado,
 $EQM(\bar{F}_R) = 2,59 < EQM(\bar{F}) = 28$.

Cont.

- Em geral, as distribuições exatas dos estimadores razão e do tipo razão $(\hat{r}, \hat{\tau}_R, \hat{\mu}_R)$ são difíceis 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 (n).
- Para amostras grandes, as distribuições aproximam-se da normal.

Propriedades dos estimadores razão

- Consideraremos, como mencionado, os planos AAS_c e AAS_s .
- Note que $\mathcal{E}_{A_i}(\hat{r}) = \mathcal{E}_{A_i}\left(\frac{\bar{Y}}{\bar{X}}\right)$, $i = 1, 2$. Como se espera que \bar{X} e \bar{Y} sejam dependentes, então $\mathcal{E}_{A_i}\left(\frac{\bar{Y}}{\bar{X}}\right) \neq \mathcal{E}_{A_i}(\bar{Y}) \mathcal{E}_{A_i}\left(\frac{1}{\bar{X}}\right)$.
- De qualquer forma, calcular $\mathcal{E}_{A_i}\left(\frac{1}{\bar{X}}\right)$ não seria uma tarefa fácil (sob qualquer plano amostral).
- Consideraremos, portanto, um cálculo aproximado baseado em expansões [séries de Taylor](#).

Propriedades dos estimadores razão

- Vamos obter, primeiramente, as esperanças de \hat{r} , $\hat{\mu}_R$ e $\hat{\tau}_R$.
- Note, num primeiro momento, que

$$\hat{r} - r = \frac{\bar{Y}}{\bar{X}} - r = \frac{\bar{Y} - r\bar{X}}{\bar{X}}. \quad (3)$$

- Agora, vamos considerar a expansão em séries de Taylor de $\frac{1}{\bar{X}}$ (em torno de μ_X):

$$\begin{aligned} \frac{1}{\bar{X}} &= \frac{1}{\mu_X + \bar{X} - \mu_X} = \frac{1}{\mu_X \left(1 + \frac{\bar{X} - \mu_X}{\mu_X}\right)} = \frac{1}{\mu_X} \left(1 + \frac{\bar{X} - \mu_X}{\mu_X}\right)^{-1} \\ &= \frac{1}{\mu_X} \left\{ 1 - \frac{\bar{X} - \mu_X}{\mu_X} + \left(\frac{\bar{X} - \mu_X}{\mu_X}\right)^2 - \dots \right\} \end{aligned} \quad (4)$$

Propriedades dos estimadores razão

- Portanto, de (4) em (3), vem que

$$\hat{r} - r = \frac{\bar{Y}}{\bar{X}} = \frac{\bar{Y} - r\bar{X}}{\mu_x} - \frac{(\bar{Y} - r\bar{X})(\bar{X} - \mu_x)}{\mu_x^2} + \dots \quad (5)$$

- Considerando-se a expansão de primeira ordem, temos que:

$$\mathcal{E}(\hat{r} - r) \approx \mathcal{E}\left[\frac{\bar{Y} - r\bar{X}}{\mu_x}\right] = \frac{\mu_y - r\mu_x}{\mu_x} = 0$$

- Note que este resultado vale para os planos AAS_c e AAS_s .
- Exercício: calcular as esperanças de $\mathcal{E}(\hat{\mu}_R)$ e $\mathcal{E}(\hat{\tau}_R)$.

Propriedades dos estimadores razão

- Assim, do resulado anterior, concluímos que o estimador razão é aproximadamente não viciado (para n suficientemente grande).
- Contudo, para amostras pequenas ou moderadas, ele pode apresentar um viés de magnitude razoável.
- O viés pode ser calculado considerando-se o truncamento na expansão em série de Taylor, a partir do termo de interessante.
- Quanto maior a ordem do truncamento, mais preciso (acurado) é o resultado.
- A área da Estatística que lida com cálculo do viés, correção de estimadores etc é a [Teoria Assintótica](#).
- Para calcular (estimar) o viés vamos considerar o segundo termo de (5), assim (próximo slide)

Cont.

$$\begin{aligned} -\mathcal{E} \left[\frac{(\bar{Y} - r\bar{X})(\bar{X} - \mu_x)}{\mu_x^2} \right] &= \frac{1}{\mu_x^2} \{ r\mathcal{E} [\bar{X}(\bar{X} - \mu_x)] - \mathcal{E} [\bar{Y}(\bar{X} - \mu_x)] \} \\ &= \frac{1}{\mu_x^2} \{ r\mathcal{V}(\bar{X}) - \text{Cov}[\bar{X}, \bar{Y}] \} \\ &= r \frac{\mathcal{V}(\bar{X})}{\mu_x^2} - \rho [\bar{X}, \bar{Y}] \frac{DP(\bar{X})}{\mu_x^2} DP(\bar{Y}) \\ &= rCV^2 [\bar{X}] - \rho [\bar{X}, \bar{Y}] \frac{CV(\bar{X})}{\mu_x} CV(\bar{Y})\mu_y \\ &= rCV^2 [\bar{X}] - \rho [\bar{X}, \bar{Y}] CV(\bar{X})CV(\bar{Y})r \\ &= rCV^2 [\bar{X}] \left\{ 1 - \rho [\bar{Y}, \bar{X}] \frac{CV [\bar{Y}]}{CV [\bar{X}]} \right\} \end{aligned}$$

Cont.

- Em que $CV[\bar{Y}] = DP[\bar{Y}]/\mu_y$, $CV[\bar{X}] = DP[\bar{X}]/\mu_x$ (coeficientes de variação).
- Além disso, $DP[\bar{Y}] = \sqrt{\sigma_y^2/n}$, $DP[\bar{X}] = \sqrt{\sigma_x^2/n}$ (AAS_c), e

$$\rho[\bar{X}, \bar{Y}] = \frac{Cov[\bar{X}, \bar{Y}]}{DP[\bar{X}]DP[\bar{Y}]}$$

$$Cov[\bar{X}, \bar{Y}] = \mathcal{E}(\bar{X}\bar{Y}) - \mathcal{E}(\bar{X})\mathcal{E}(\bar{Y}) = \frac{1}{nN} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

- Exercícios: provar as fórmulas acima e verificar como elas ficam no caso AAS_s .

Cont.

- Vamos explorar um pouco mais o resultado obtido.
- Sob os planos AAS, temos que

$$-\mathcal{E} \left[\frac{(\bar{Y} - r\bar{X})(\bar{X} - \mu_x)}{\mu_x^2} \right] = rCV^2(\bar{X}) \left\{ 1 - \rho[x, y] \frac{CV[y]}{CV[x]} \right\}$$

em que $CV[x] = \frac{\sigma_x}{\mu_x}$, $CV[y] = \frac{\sigma_y}{\mu_y}$,

$$\rho[x, y] = \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y} = \rho[\bar{X}, \bar{Y}]$$

- Pode-se, por outro lado, provar que,

$$\mathcal{E}(\hat{r} - r) \approx \frac{1}{n\mu_x^2} \{ r\sigma_x^2 - \rho[x, y]\sigma_x\sigma_y \} \quad (6)$$

- Exercício: obter expressões para os vícios de $\hat{\mu}_R$ e $\hat{\tau}_R$, usando (6).

Cont.

- Podemos provar que a expressão (6) pode ser escrita como:

$$\begin{aligned}\mathcal{E}(\hat{r} - r) &\approx \frac{1}{n} \left\{ \frac{r\sigma_x^2}{\mu_x^2} - \frac{\mu_y}{\mu_x} \rho[x, y] \frac{\sigma_x \sigma_y}{\mu_x \mu_y} \right\} \\ &= \frac{1}{n} \left\{ rCV[x]^2 - r\rho[x, y]CV[y]CV[x] \right\} \\ &= \frac{1}{n} \left\{ rCV[x]^2 \left\{ 1 - \rho[x, y] \frac{CV[y]}{CV[x]} \right\} \right\}\end{aligned}$$

Assim, para que o vício seja pequeno, devemos ter $\rho[x, y] \frac{CV[y]}{CV[x]} \approx 1$

Cálculo da variância dos estimadores

- Todos os desenvolvimentos serão feitos considerando-se “n” suficientemente grande.
- Primeiramente, vamos calcular a variância do estimador razão (usando-se a aproximação de primeira ordem (5)):

$$\mathcal{V}(\hat{r}) \approx \text{EQM}(\hat{r}) = \mathcal{E} [(\hat{r} - r)^2] = \frac{1}{\mu_x^2} \mathcal{E} [(\bar{Y} - r\bar{X})^2] = \frac{1}{\mu_x^2} \mathcal{E} [\bar{D}^2].$$

Cont.

- Em que $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$ é a média amostral das variáveis $d_i = y_i - rx_i, i = 1, 2, \dots, N$. Note que (exercício)

$$\bar{d} = 0; \sigma_d^2 = \frac{1}{N} \sum_{i=1}^N d_i^2$$

são, respectivamente, a média e a variância populacionais das variáveis d_i .

Cont.

- Assim, temos que $\mathcal{E}(\bar{D}) = 0$ (provar) e

$$\mathcal{V}(\hat{r}) = \frac{1}{\mu_x^2} \mathcal{E}(\bar{D}^2) = \frac{1}{\mu_x^2} \mathcal{V}(\bar{D})$$

- Sob os planos AAS_c e AAS_s temos, respectivamente, que

$$\mathcal{V}_{A_1}(\hat{r}) = \frac{1}{\mu_x^2} \mathcal{V}_{A_1}(\bar{D}) = \frac{1}{\mu_x^2} \frac{\sigma_d^2}{n} = \frac{1}{\mu_x^2} \frac{\sigma_R^2}{n} \quad (7)$$

$$\mathcal{V}_{A_1}(\hat{r}) = \frac{1}{\mu_x^2} \mathcal{V}_{A_2}(\bar{D}) = \frac{1}{\mu_x^2} (1-f) \frac{s_d^2}{n} = \frac{1}{\mu_x^2} (1-f) \frac{s_R^2}{n} \quad (8)$$

em que $f = n/N$, $\sigma_d^2 = \sigma_R^2 = \frac{1}{N} \sum_{i=1}^N d_i^2$ e $s_d^2 = s_R^2 = \frac{1}{N-1} \sum_{i=1}^N d_i^2$.

Cont.

- Das expressões (1), (2), (7) e (8), vem que

$$\mathcal{V}_{A_1}(\hat{\mu}_R) \approx \mu_x^2 \mathcal{V}(\hat{r}) = \mu_x^2 \frac{1}{\mu_x^2} \frac{\sigma_R^2}{n} = \frac{\sigma_R^2}{n} \quad (9)$$

$$\mathcal{V}_{A_2}(\hat{\mu}_R) \approx \mu_x^2 \mathcal{V}(\hat{r}) = \mu_x^2 \frac{1}{\mu_x^2} (1-f) \frac{s_R^2}{n} = (1-f) \frac{s_R^2}{n} \quad (10)$$

$$\mathcal{V}_{A_1}(\hat{\tau}_R) \approx \tau_x^2 \mathcal{V}(\hat{r}) = \tau_x^2 \frac{1}{\mu_x^2} \frac{\sigma_R^2}{n} = N^2 \frac{\sigma_R^2}{n} \quad (11)$$

$$\mathcal{V}_{A_2}(\hat{\tau}_R) \approx \tau_x^2 \mathcal{V}(\hat{r}) = \tau_x^2 \frac{1}{\mu_x^2} (1-f) \frac{s_R^2}{n} = N^2 (1-f) \frac{s_R^2}{n} \quad (12)$$

Os respectivos erros-padrão correspondem à raiz-quadrada positiva das expressões acima.

Cont.

- Estimação da variância populacional.
 - Das equações (9) - (12), notamos que, essencialmente, precisamos de um estimador apropriado para σ_R^2 , s_R^2 , para estimarmos as variâncias de \hat{r} , $\hat{\mu}_R$, $\hat{\tau}_R$.
 - Um estimador apropriado (aproximadamente não viciado), é

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

- Se N for desconhecido, não se pode calcular μ_x (em geral). Nesse caso substitui-se μ_x por \bar{X} . Sob tais circunstâncias, o vício, em geral, tende a aumentar.

Cont.

- Portanto, estimadores apropriados para as variâncias dos estimadores são dados, respectivamente, por:

$$\begin{aligned}\widehat{V}_{A_1}(\widehat{\mu}_R) &\approx \mu_x^2 \widehat{V}(\widehat{r}) = \mu_x^2 \frac{1}{\mu_x^2} \frac{\widehat{\sigma}_R^2}{n} = \frac{\widehat{\sigma}_R^2}{n} \\ \widehat{V}_{A_2}(\widehat{\mu}_R) &\approx \mu_x^2 \widehat{V}(\widehat{r}) = \mu_x^2 \frac{1}{\mu_x^2} (1-f) \frac{\widehat{S}_R^2}{n} = (1-f) \frac{\widehat{S}_R^2}{n} \\ \widehat{V}_{A_1}(\widehat{\tau}_R) &\approx \tau_x^2 \widehat{V}(\widehat{r}) = \tau_x^2 \frac{1}{\mu_x^2} \frac{\widehat{\sigma}_R^2}{n} = N^2 \frac{\widehat{\sigma}_R^2}{n} \\ \widehat{V}_{A_2}(\widehat{\tau}_R) &\approx \tau_x^2 \widehat{V}(\widehat{r}) = \tau_x^2 \frac{1}{\mu_x^2} (1-f) \frac{\widehat{S}_R^2}{n} = N^2 (1-f) \frac{\widehat{S}_R^2}{n} \quad (13)\end{aligned}$$

Estimadores para os erros-padrão correspondem à raiz-quadrada positiva das expressões acima.

Cont.

- Tais estimadores também são viciados mas, em geral, o vício diminui à medida em que n aumenta.
- Contudo, conforme discutido em [Cochran \(1977\)](#), os estimadores para as variâncias são consistentes, para N e n suficientemente grandes.
- Também é possível estimar as variâncias via [bootstrap](#).

Cont.

- Comparação entre os estimadores razão ($\hat{\tau}_R$) e expansão ($\hat{\tau}$) para o total populacional.
- Resultado: Se n for suficientemente grande, o planejamento adotado for AAS_c e

$$\rho[x, y] > \frac{\sigma_x/\mu_x}{2\sigma_y/\mu_y} = \frac{CV[x]}{2CV[y]}$$

então $\mathcal{V}[\hat{\tau}_R] < \mathcal{V}[\hat{\tau}]$

Cont.

- Temos que

$$\begin{aligned} \mathcal{V}_{A_1}(\widehat{\tau}_R) &\approx \frac{\tau_x^2}{n\mu_x^2 N} \sum_{i=1}^N (y_i - rx_i)^2 = \frac{N^2}{nN} \sum_{i=1}^N \{(y_i - \mu_y) - r(x_i - \mu_x)\}^2 \\ &= \frac{N^2}{nN} \left\{ \sum_{i=1}^N (y_i - \mu_y)^2 + r^2 \sum_{i=1}^N (x_i - \mu_x)^2 \right. \\ &\quad \left. - 2r \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \right\} \\ &= \frac{N^2}{n} \{ \sigma_y^2 + r^2 \sigma_x^2 - 2r\rho[x, y] \sigma_x \sigma_y \} \end{aligned} \quad (14)$$

Cont.

- Por outro lado, já vimos que

$$\mathcal{V}_{A_1}[\hat{\tau}] = \frac{N^2}{n} \sigma_y^2 \quad (15)$$

- Assim, de (14) e (15), temos que $\mathcal{V}[\hat{\tau}_R] < \mathcal{V}[\hat{\tau}]$, se e somente se

$$\sigma_y^2 + r^2 \sigma_x^2 - 2r\rho[x, y] \sigma_x \sigma_y < \sigma_y^2 \Leftrightarrow 2r\rho[x, y] \sigma_x \sigma_y > r^2 \sigma_x^2$$

desde que $r > 0$, donde o resultado segue.

- Note que $CV[x] = \sigma_x/\mu_x$ e $CV[y] = \sigma_y/\mu_y$ são os coeficientes de variação das variáveis x e y .

Cont.

- Para ganhos maiores de $\hat{\mu}_R$ com relação à $\hat{\mu}$, $CV[x]/CV[y]$ deve estar entre 0,5 e 1,3 e $\rho[x, y]$ deve ser maior do que 0,6 (Kish, 1995).
- Portanto, se a variável x for mais homogênea do que y , basta uma baixa correlação entre as variáveis para que se tenha vantagens em se usar o estimador do tipo razão.
- Exercício: pesquisar sobre como realizar inferência para a proporção populacional, usando o estimador razão.

Propriedades assintóticas dos estimadores

- 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 .
- Portanto, sob certas condições, em relação à estimação da média, temos que

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

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

Propriedades assintóticas dos estimadores

- Em que $EP_{PA}(\hat{\mu}_R)$ e $\widehat{EP}_{PA}(\hat{\mu}_R)$ são, respectivamente, o erro-padrão e um estimador do erro-padrão, associados à um determinado plano amostral (PA), relativos à $\hat{\mu}_R$.
- Note, entretanto, que a forma do estimador para μ pode mudar. Vide, por exemplo, o planejamento AE (amostragem estratificada).
- Resultados semelhantes podem ser obtidos para o total (τ) e para a razão (r) (Exercício).

Intervalos de confiança

- Analogamente aos casos anteriores ([link 1](#), [link 2](#)), temos que um intervalo assintótico, com coeficiente de confiança de aproximadamente γ , é dado por:

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

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

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 $Z_t = \frac{\hat{\mu}_R - \mu_0}{\widehat{EP}_{PA}(\hat{\mu}_R)}$.

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

Testes de Hipótese

- Defina $z_t = \frac{\tilde{\mu}_R - \mu_0}{\widehat{EP}_{PA}(\tilde{\mu}_R)}$ 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 29 a 31 dos slides: [link](#)

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: $z_\gamma \widehat{EP}_{PA}(\widehat{\mu}_R)$. Fixa-se um erro de estimativa de interesse.
- Probabilidade do módulo da diferença $P(|\widehat{\mu}_R - \mu| < \delta) > \gamma$, $\delta > 0$, $\gamma \in (0, 1)$.

Determinação do tamanho amostral: erro da estimativa

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

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

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

Em geral, o (um) valor de s_R^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 estiver estratificada, podemos considerar uma combinação dos resultados obtidos anteriormente, sob AE, com os resultados aqui obtidos para o estimador razão.
- A estrutura é a mesma daquela apresentada [aqui](#).
- 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:

$$\hat{\mu}_{Res} = \sum_{h=1}^H W_h \frac{\bar{Y}_h}{\bar{X}_h} \mu_{xh} = \sum_{h=1}^H W_h \hat{r}_h \mu_{xh} = \sum_{h=1}^H W_h \hat{\mu}_{Rh}$$

$$\hat{T}_{Res} = N \hat{\mu}_{Res} = \sum_{h=1}^H N_h \hat{\mu}_{Rh}$$

Sob amostragem estratificada

- Usando resultados anteriores (estimador razão e AE), temos que:

$$\mathcal{V}_{AE_1}(\hat{\mu}_{Res}) \approx \sum_{h=1}^H W_h^2 \frac{\sigma_{Rh}^2}{n_h}$$

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

em que $f_h = \frac{n_h}{N_h}$, $\sigma_{Rh}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{hi} - r_h x_{hi})^2$,

$s_{Rh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - r_h x_{hi})^2$ e $r_h = \frac{\mu_{yh}}{\mu_{xh}} = \frac{\tau_{yh}}{\tau_{xh}}$.

Sob amostragem estratificada

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

$$\hat{V}_{AE_1}(\hat{\mu}_{Res}) \approx \sum_{h=1}^H W_h^2 \frac{\hat{\sigma}_{Rh}^2}{n_h}$$

$$\hat{V}_{AE_2}(\hat{\mu}_{Res}) \approx \sum_{h=1}^H W_h^2 (1 - f_h) \frac{\hat{S}_{Rh}^2}{n_h}$$

em que $f_h = \frac{n_h}{N_h}$, $\hat{\sigma}_{Rh}^2 = \hat{S}_{Rh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (Y_{hi} - \hat{r}_h X_{hi})^2$, $\hat{r}_h = \frac{\bar{Y}_h}{\bar{X}_h}$.

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 ([link](#)).
- Intervalos de confiança e testes de hipótese podem ser desenvolvidos de modo semelhante ao exposto anteriormente ([link](#)).

Alocação ótima

- Sob alocação ótima e AAS_c dentro de cada extrato, como visto [aqui](#), temos que:

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

de acordo com [Bolfarine & Bussab \(2005\)](#), as vezes temos que $\sigma_{Rh} \propto \sqrt{\mu_{xh}}$ ou $\sigma_{Rh} \propto \mu_{xh}$. De qualquer forma, seria necessário ter-se uma “estimativa” para μ_{Rh} .

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

Comentários finais

- Em linhas gerais, quanto maior for a correlação entre y e x , melhor será o desempenho do(s) estimadores razão em relação aos usuais (visto até o presente momento).
- Conforme verificado em [Cochran \(1977\)](#) e [Rodrigues e Bolfarine \(1984\)](#), quando a relação entre y e x for explicada (de forma apropriada) pelo modelo $y_i = \beta x_i + \xi_i$ ($Y_i = \beta x_i + \xi_i$), ou seja, um modelo de regressão linear com **intercepto nulo**, então os estimadores razão serão os “melhores” (menor variância) estimadores na classe dos (aproximadamente) não viciados.

Exercício

- Considere o Exercício 5.2 do Livro Bolfarine e Bussab (2005), que consiste em estimar o número de árvores mortas de determinada espécie em uma reserva florestal. A reserva fora dividida em 200 áreas de 1,5 hectare. O número de árvores mortas fora avaliado por fotografia aérea (X - “foto aérea”) nas 200 áreas, apresentando uma contagem total de aproximadamente ($\tau_X =$)15.600 árvores mortas, da espécie. Em 10 das 200 áreas, o número de árvores mortas, além da avaliação por fotografia aérea, fora também avaliado por contagem terrestre (Y - “in loco”).
- Vamos comparar os estimadores **expansão** e razão do total (estes slides) (τ_Y), sob AAS_s

Dados: foto aérea e in loco são as “variáveis”

Área	foto aérea	in loco
1	12	18
2	30	42
3	24	24
4	24	36
5	18	24
6	30	36
7	12	14
8	6	10
9	36	48
10	42	54

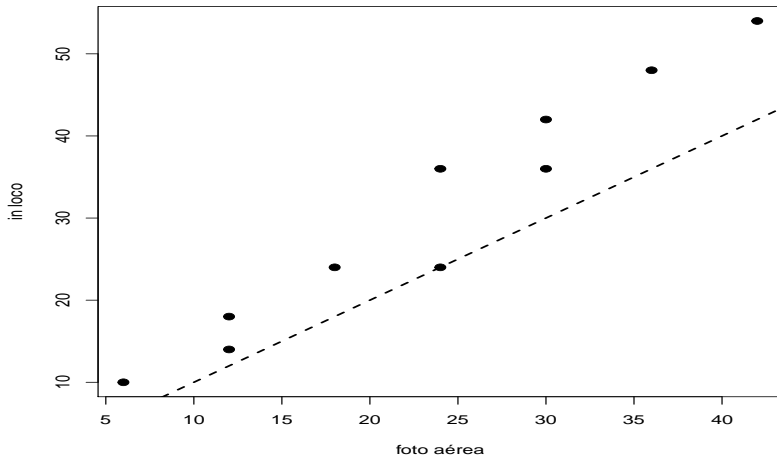
Fórmulas

- As respectivas fórmulas são dadas:
 - Estimador expansão: [link](#).
 - Estimador razão: nestes slides.
- Faremos análises descritivas e inferenciais.
- A correlação amostral de Pearson (estimativa (erro-padrão)) resultou em 0,972 (0,082). O teste para nulidade da correlação de Pearson resultou em $t = 11,92$ ($< 0,0001$). Estes resultados indicam que as variáveis são significativa (e positivamente) correlacionadas.

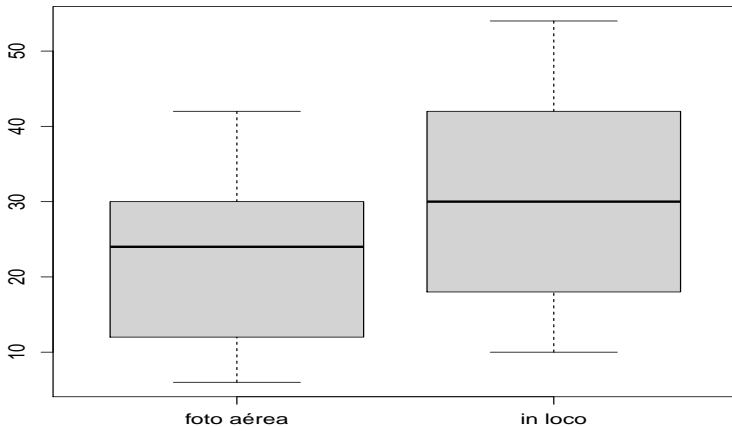
Medidas resumo

Medida resumo	foto aérea	in loco
média	23,40	30,60
dp	11,47	14,85
var.	131,60	220,49
cv(%)	49,02	48,53
mínimo	6,00	10,00
mediana	24,00	30,00
máximo	42,00	54,00
curtose	-1,41	-1,57
ca	0,04	0,12

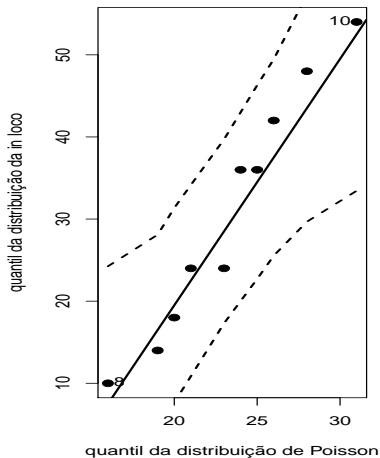
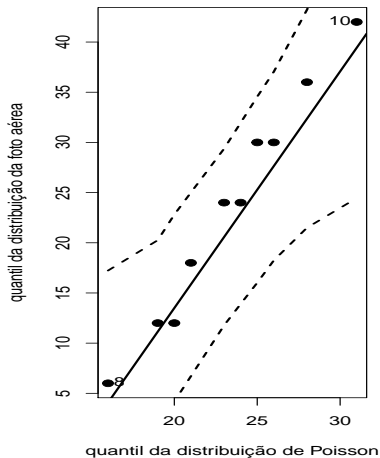
Gráfico de dispersão entre as variáveis



Boxplot entre as variáveis



QQ plot das variáveis



Comentários

- As duas variáveis são positiva e significamente correlacionadas.
- Se se fosse utilizar a [modelagem clássica](#) (distribuição dos dados) por se tratar de contagem, pensar-se-ia em considerar a [distribuição de Poisson](#). Contudo, há indícios de [superdispersão](#) (média < variância). Com efeito, apesar de todos os pontos estarem dentro das bandas de confiança dos QQplots, um teste para presença de superdispersão ([link](#), pag. 115), acusou sua presença (estatística (p-valor)) para ambas as variáveis: 64,85 (<0,0001) (foto aérea) e 64,85 (<0,0001) (in loco).
- Por outro lado, apesar de se basear em resultados assintóticos, a presente abordagem ([modelagem probabilística](#)), prescinde de suposições sobre a distribuição dos dados.

Estimativas

- Temos os seguintes resultados:

Estimador	Est.	EP	IC(%)
Expansão (EE)	6120,00	915,35	[4325,96 ; 7914,04]
Razão (ER)	20400,00	714,06	[19000,46 ; 21799,54]

- $EPA = \frac{\widehat{V}_{A_2}(\widehat{\tau}_R)}{\widehat{V}_{A_2}(\widehat{\tau})} = 0,610$.
- Temos claras diferenças entre os dois conjuntos de estimativas (pontual e intervalar). Com relação ao EPA, o ER apresenta grande vantagem em relação ao EE.
- Mesmo em se comparando o EQM, o vício do ER (pag. destes slides) é pequeno, de modo que $\frac{\widehat{EQM}_{A_2}(\widehat{\tau}_R)}{\widehat{EQM}_{A_2}(\widehat{\tau})} = 0,610$.

Comparação entre as estimativas

- Usando **modelos de regressão apropriados** (veja também a pag. 45 destes slides), temos que a relação entre as duas variáveis parecer ser “adequadamente” representável por uma reta passando pela origem (modelo de regressão normais lineares homocedásticos).
- Além disso, segundo critérios de superioridade do estimador razão em relação ao estimador usual (expansão) (pag. 27 destes slides) temos que $\frac{2\tilde{\rho}[x,y]\widetilde{CV}[y]}{\widetilde{CV}[x]} = 1,926 > 1$.