Viés do Estimador Razão

Teorema 5.1. Para um plano amostral com $E(\bar{x}) = \mu_X$ e $E(\bar{y}) = \mu_Y$, tem-se que para n suficientemente grande

$E(\hat{R}) = \frac{\bar{y}}{\bar{x}} \approx R = \frac{\mu_Y}{\mu_X}$

Demonstração (Rocha, F.M.M.)

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}} = f(\bar{x})(\bar{y} - R\bar{x}), \tag{1}$$

com $f(\bar{x}) = \frac{1}{\bar{x}}$.

O cálculo da esperança fica mais difícil por termos uma variável aleatória $\bar{x}$ no denominador.

Expansão em séries de Taylor de $f(x)$ em torno de a:

$$f(x) = f(a)(x-a)^0 + \frac{f'(a)(x-a)^1}{1!} + \frac{f''(a)(x-a)^2}{2!} + \frac{f'''(a)(x-a)^3}{3!} + \dots \tag{2}$$

Então, expandindo $f(\bar{x}) = \frac{1}{\bar{x}}$ em torno de $a = \mu_X$, temos:

$$f(\bar{x}) = \frac{1}{\bar{x}} \Longrightarrow f(\mu_X) = \frac{1}{\mu_X}$$

$$f'(\bar{x}) = -\frac{1}{\bar{x}^2} \Longrightarrow f'(\mu_X) = -\frac{1}{\mu_X^2}$$

$$f''(\bar{x}) = +2\frac{1}{\bar{x}^3} \Longrightarrow f''(\mu_X) = \frac{2}{\mu_X^3}$$

$$f'''(\bar{x}) = -6\frac{1}{\bar{x}^4} \Longrightarrow f'''(\mu_X) = -\frac{6}{\mu_X^4}$$

Expansão de $f(\bar{x}) = \frac{1}{\bar{x}}$ em torno de $a = \mu_X$:

$$\begin{aligned} f(\bar{x}) &= f(\mu_X) + f'(\mu_X)(\bar{x} - \mu_X) + \frac{f''(\mu_X)(\bar{x} - \mu_X)^2}{2} + \frac{f'''(\mu_X)(\bar{x} - \mu_X)^3}{3!} + \dots = \\ &= \frac{1}{\mu_X} - \frac{1}{\mu_X^2}(\bar{x} - \mu_X) + \frac{1}{\mu_X^3}(\bar{x} - \mu_X)^2 - \frac{1}{\mu_X^4}(\bar{x} - \mu_X)^3 + \dots \end{aligned}$$

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}} = f(\bar{x})(\bar{y} - R\bar{x}) = \frac{\bar{y} - R\bar{x}}{\mu_X} - \frac{(\bar{y} - R\bar{x})(\bar{x} - \mu_X)}{\mu_X^2} + \dots$$

Calcula a esperança usando a aproximação de primeira ordem,

$$E(\hat{R} - R) \approx E\left(\frac{\bar{y} - R\bar{x}}{\mu_X}\right) = \frac{\mu_Y - R\mu_X}{\mu_X} = 0,$$

Viés

Para plano AAS (AASc ou AASs), temos:

$$\begin{aligned}
E(\hat{R}-R) &\approx E\left(\frac{\bar{y}-R\bar{x}}{\mu_X}-\frac{(\bar{y}-R\bar{x})(\bar{x}-\mu_X)}{\mu_X^2}\right)\\
&= 0+\frac{1}{\mu_X^2}\{R\,E[\bar{x}(\bar{x}-\mu_X)]-E[\bar{y}(\bar{x}-\mu_X)]\}\\
&= \frac{1}{\mu_X^2}\{R\,E[(\bar{x}-\mu_X)(\bar{x}-\mu_X)]-E[(\bar{y}-\mu_Y)(\bar{x}-\mu_X)]\}\\
&= \frac{1}{\mu_X^2}\{R\,Var(\bar{x})-Cov(\bar{x},\bar{y})\}\\
&= R\frac{Var(\bar{x})}{\mu_X^2}-\rho[\bar{x},\bar{y}]\frac{DP(\bar{x})DP(\bar{y})}{\mu_X^2}\\
&= R\,CV^2(\bar{x})-\rho[\bar{x},\bar{y}]\frac{CV(\bar{x})CV(\bar{y})}{\mu_X}\mu_Y\\
&= R\,CV^2(\bar{x})-\rho[\bar{x},\bar{y}]CV(\bar{x})CV(\bar{y})R\\
&= R\,CV^2(\bar{x})\left[1-\rho[\bar{x},\bar{y}]\frac{CV(\bar{y})}{CV(\bar{x})}\right]
\end{aligned}$$

Temos que:

$$CV(\bar{x}) = \frac{DP(\bar{x})}{\mu_X}=\frac{\sigma_X}{\mu_X\sqrt{n}}=\frac{CV(X)}{\sqrt{n}} \quad CV(\bar{y})=\frac{DP(\bar{y})}{\mu_Y}=\frac{\sigma_Y}{\mu_Y\sqrt{n}}$$

Lembrando para AASc: $f_i \sim Bin(n,1/N)$, $Var(f_i)=\frac{n}{N}\left(1-\frac{1}{N}\right)$ e $Cov(f_i,f_j)=-\frac{n}{N^2}$.

$$\begin{aligned}
Cov(X,Y) &= \frac{\sum_{i=1}^N (X_i-\mu_X)(Y_i-\mu_Y)}{N}\\
Cov(\bar{x},\bar{y}) &= \frac{1}{n^2}Cov(\sum_{i=1}^N f_iX_i,\sum_{j=1}^N f_jY_j)=\frac{1}{n^2}\sum_{i=1}^N\sum_{j=1}^N X_iY_iCov(f_i,f_j)=\\
&= \frac{1}{n^2}\sum_{i=1}^N X_iY_i\frac{n}{N}(1-\frac{1}{N})-\frac{1}{n^2}\sum_{i=1}^N\sum_{j\neq i}^N X_iY_i\frac{n}{N^2}=\\
&= \frac{1}{nN^2}(N-1)(\sum_{i=1}^N X_iY_i)-\frac{1}{nN^2}\sum_{i=1}^N\sum_{j\neq i}^N X_iY_i=\\
&= \frac{1}{nN^2}N(\sum_{i=1}^N X_iY_i)-\frac{1}{nN^2}\sum_{i=1}^N\sum_{j=1}^N X_iY_i\\
&= \frac{1}{nN}(\sum_{i=1}^N X_iY_i)-\frac{1}{n}\mu_X\mu_Y fracNN=\\
&= \frac{1}{nN}\sum_{i=1}^N(X_i-\mu_X)(Y_i-\mu_Y)=\frac{1}{n}Cov(X,Y)
\end{aligned}$$

$$\rho(\bar{x},\bar{y})=\frac{Cov(\bar{x},\bar{y})}{DP(\bar{x})DP(\bar{y}))}=\frac{\frac{1}{n}Cov(X,Y)}{\frac{\sigma_X}{\sqrt{n}}\frac{\sigma_Y}{\sqrt{n}}}=\rho(X,Y)$$

Voltando ao viés

$$\begin{aligned} E(\hat{R}-R) &\approx E\left(\frac{\bar{y}-R\bar{x}}{\mu_X} - \frac{(\bar{y}-R\bar{x})(\bar{x}-\mu_X)}{\mu_X^2}\right) \\ &= R\,CV^2(\bar{x})\left[1-\rho[\bar{x},\bar{y}]\frac{CV(\bar{y})}{CV(\bar{x})}\right] \\ &= R\,\frac{CV^2(X)}{n}\left[1-\rho[X,Y]\frac{CV(Y)}{CV(X)}\right] \end{aligned}$$

Agora podemos ver que quanto maior o tamanho da amostra n, menor o viés.

Lembrando, temos a aproximação:

$$\hat{R}-R = \frac{\bar{y}}{\bar{x}} - R = \frac{\bar{y}-R\bar{x}}{\bar{x}} = f(\bar{x})(\bar{y}-R\bar{x}) = \frac{\bar{y}-R\bar{x}}{\mu_X} - \frac{(\bar{y}-R\bar{x})(\bar{x}-\mu_X)}{\mu_X^2} + \ldots$$

Aproximando a variância usando o Erro Quadrático Médio e usando só o primeiro termo da nossa aproximação, temos:

$$Var(\hat{R}) = E(\hat{R}-E(\hat{R}))^2 = E(\hat{R}-R+R-E(\hat{R}))^2 \approx E(\hat{R}-R)^2 = E\left(\frac{\bar{y}-R\bar{x}}{\mu_X}\right)^2 = \frac{E\left(\bar{d}\right)^2}{\mu_X^2}$$

em que $d_i = Y_i - RX_i$ e $\sum_{i=1}^N d_i = 0$.

Estimaremos a média de Y (μ_Y), usando $\hat{\mu}_R = \hat{R}\mu_X$ e estimaremos o total de T (τ_Y), usando $\hat{\tau}_R = \hat{R}\tau_X$.

Teorema 5.2. Se n é grande, tem-se para o plano AASc que

$$\begin{aligned} Var(\hat{R}) &\approx \frac{1}{n\mu_X^2}\sum_{i=1}^N \frac{(Y_i-RX_i)^2}{N} = \frac{1}{\mu_X^2}\frac{\sigma_R^2}{n} \\ Var(\hat{\mu}_R) &\approx \frac{1}{n}\sum_{i=1}^N \frac{(Y_i-RX_i)^2}{N} = \frac{\sigma_R^2}{n} \\ Var(\hat{\tau}_R) &\approx \frac{\tau_X^2}{n\mu_X^2}\sum_{i=1}^N \frac{(Y_i-RX_i)^2}{N} = N^2\frac{\sigma_R^2}{n} \\ \sigma_R^2 &= \sum_{i=1}^N \frac{(Y_i-RX_i)^2}{N} \end{aligned}$$

Um estimador de σ_R^2 é dado por

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

Cochran (1977) comenta que os estimadores das variâncias são consistentes.

As expressões para AASc e AASs estão no formulário.

Para n grande, faremos intervalos de confiança utilizando a distribuição normal, sendo os intervalos para R, μ_y e τ_y respectivamente:

$$IC = [\hat{R} \mp 1.96\sqrt{\widehat{Var}(\hat{R})}],\ IC = [\hat{\mu_R} \mp 1.96\sqrt{\widehat{Var}(\hat{\mu_R})}],\ IC = [\hat{\tau_R} \mp 1.96\sqrt{\widehat{Var}(\hat{\tau_R})}]$$

Observação: Só podemos usar o estimador razão para grandezas proporcionais, ou seja, ao fazermos o gráfico de dispersão entre y e x, teremos que ter reta a+bx ajustada passando pela origem (com a nulo).