(Aprox.)

+ Varianza del estimador:

$$V(\hat{R}) \simeq R^2 [C^2_{\bar{x}} + C^2_{\bar{y}} - 2C_{\bar{x}\bar{y}}] = \frac{1}{\bar{Y}^2}\left[V(\bar{x}) + R^2 V(\bar{y}) - 2R\,Cov(\bar{x},\bar{y})\right]$$

$$(SR) \rightarrow \frac{1-f}{n} \cdot \frac{1}{\bar{Y}^2}\left[S^2_X + R^2 S^2_Y - 2RS_{XY}\right]$$

$$(CR) \rightarrow \frac{1}{n} \cdot \frac{1}{\bar{Y}^2}\left[\sigma^2_X + R^2 \sigma^2_Y - 2R\sigma_{XY}\right]$$

+ Estimación de la varianza del estimador:

$$\hat{V}(\hat{R}) \simeq \hat{R}^2 [\hat{C}^2_{\bar{x}} + \hat{C}^2_{\bar{y}} - 2\hat{C}_{\bar{x}\bar{y}}]$$

$$(SR) \rightarrow \frac{1-f}{n} \cdot \frac{1}{\bar{y}^2}\left[\hat{S}^2_x + \hat{R}^2 \hat{S}^2_y - 2\hat{R}\hat{S}_{xy}\right]$$

$$(CR) \rightarrow \frac{1}{n} \cdot \frac{1}{\bar{y}^2}\left[\hat{S}^2_x + \hat{R}^2 \cdot \hat{S}^2_y - 2\hat{R}\hat{S}_{xy}\right]$$

5_ COMPARACIÓN de $\hat{X}_R$ con $\hat{X}$.

+ $V(\hat{X}_R) < V(\hat{X})$ si $\rho_{XY} > \frac{1}{2}\frac{C_Y}{C_X} > 0$.

MUEST_T4

1. ESTIMADORES NO LINEALES.
2. MT. GENERAL de LINEALIZACIÓN para estimación de VARIANZAS.
3. APLICACIÓN al COCIENTE de ESTIMADORES.
4. EL ESTIMADOR de RAZÓN: SESGO, VARIANZA y ESTIMACIONES.
5. Comparación del estimador de razón y del estimador de expansión

Aprox. SESGO

1_ ESTIMADORES NO LINEALES.

Existen métodos que aprovechan la información conocida relativa a una var. auxiliar Y correlacionada con la var. de estudio X, para conseguir estimaciones más precisas para X que las calculadas únicamente a partir de la variable que se estudia X.

La información disponible de la var. auxiliar puede ser probabilística o no probabilística. Las fuentes más típicas suelen ser:
- variables obtenidas en un censo anterior
- variables de la misma población en fechas anter.
- estimaciones de una pobl. ≠ pero correlacionada
- etc.

Los estimadores no lineales se utilizan además para estimar estructuras matemáticas más complejas de la población, distintas al total o la media.

Por ejemplo, si se quiere estimar un cociente de totales, un buen estimador no lineal es el cociente de los estimadores lineales de los totales.

$$R = \frac{X}{Y} \longrightarrow \hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Entre los métodos clásicos de estimación indirecta más utilizados, se encuentran:

- Mt. estimación por razón: basado en el cociente entre X e Y.
- Mt. estimación por regresión: basado en la regresión de X sobre Y
- Mt. estimación por diferencia: basado en la diferencia entre X e Y.

¿ampliar un poco con T5?

2_ MT. GENERAL de LINEALIZACIÓN para la ESTIMACIÓN de VARIANZAS

Sea θ el parámetro poblacional a estimar (lineal o no lineal) función de k parám. ~~k-dimensional~~.

$\theta = \varphi(\theta_1 \dots \theta_K)$, $\theta_i \equiv$ parámetro poblacional, $i = 1 \dots K$

A partir de una muestra aleatoria de tamaño n, se construye un estimador de θ, $\hat{\theta}$, a partir de los estimadores individuales de θ_i.

$$\left.\begin{aligned} \hat{\theta}_1 &= T_1(X_1 \dots X_n) \text{ estimador insesgado de } \theta_1 \\ &\dots \\ \hat{\theta}_K &= T_K(X_1 \dots X_n) \text{ estimador insesgado de } \theta_K \end{aligned}\right\} \hat{\theta} = \varphi(\hat{\theta}_1 \dots \hat{\theta}_K)$$

El desarrollo en serie de Taylor del estimador en un punto $\theta_0 = (\theta_{01} \dots \theta_{0K})$ es:

$$\hat{\theta} = \varphi(\hat{\theta}_1 \dots \hat{\theta}_K) = \varphi(\theta_1 \dots \theta_K) + \frac{1}{1!}\sum_{i=1}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} (\hat{\theta}_i - \theta_i) + r$$

Ignorando el último término, se puede escribir:

$$\varphi(\hat{\theta}_1 \dots \hat{\theta}_K) - \varphi(\theta_1 \dots \theta_K) \simeq \sum_{i=1}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} (\hat{\theta}_i - \theta_i)$$

que es equivalente a decir

$$\hat{\theta} - \theta \simeq \sum_{i=1}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} (\hat{\theta}_i - \theta_i)$$

$\hat{\theta}_i$ insesgado de θ_i

Por tanto, la varianza del estimador se puede aproximar:

$$V(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] \simeq E\left[\left(\sum_{i=1}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} (\hat{\theta}_i - \theta_i)\right)^2\right] = E\left[\sum_{i=1}^{K} \left(\left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0}\right)^2 (\hat{\theta}_i - \theta_i)^2 + \right.$$

$$\left. + \sum_{i=1}^{K}\sum_{j\neq i}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} \cdot \left.\frac{\partial \varphi}{\partial \hat{\theta}_j}\right|_{\theta_0} (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right] = \sum_{i=1}^{K} \left(\left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0}\right)^2 E(\hat{\theta}_i - \theta_i)^2 +$$

$$+ \sum_{i=1}^{K}\sum_{j\neq i}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} \cdot \left.\frac{\partial \varphi}{\partial \hat{\theta}_j}\right|_{\theta_0} E\left[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right] \Rightarrow$$

$$V(\hat{\theta}) \simeq \sum_{i=1}^{K} \left(\left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0}\right)^2 V(\hat{\theta}_i) + \sum_{i=1}^{K}\sum_{j\neq i}^{K} \left.\frac{\partial \varphi}{\partial \hat{\theta}_i}\right|_{\theta_0} \left.\frac{\partial \varphi}{\partial \hat{\theta}_j}\right|_{\theta_0} \operatorname{cov}(\hat{\theta}_i, \hat{\theta}_j)$$

3_APLICACIÓN al COCIENTE de ESTIMADORES

Particularizamos el resultado anterior al caso de dos parámetros poblacionales, y φ su cociente

$$R = \frac{\alpha}{\beta} = \varphi(\alpha, \beta)$$

Se toma una muestra aleatoria de tamaño n y se construyen los estimadores lineales insesgados para cada parámetro, y su cociente:

$$\left.\begin{array}{l}\hat{\alpha} = T_1(X_1 \dots X_n), \text{ estim. insesgado para } \alpha \\ \hat{\beta} = T_2(X_1 \dots X_n), \text{ estim. insesgado para } \beta\end{array}\right| \Rightarrow \hat{R} = \frac{\hat{\alpha}}{\hat{\beta}} = \varphi(\hat{\alpha}, \hat{\beta})$$

Para aproximar el valor de la varianza del estimador cociente de estimadores se puede aplicar el método de linealización anterior, basado en el desarrollo en serie de Taylor del cociente en el punto (α, β)

$$\underbrace{\varphi(\hat{\alpha}, \hat{\beta})}_{\hat{R}} = \underbrace{\varphi(\alpha, \beta)}_{R} + \left.\frac{\partial \varphi}{\partial \hat{\alpha}}\right|_{(\alpha,\beta)} (\hat{\alpha} - \alpha) + \left.\frac{\partial \varphi}{\partial \hat{\beta}}\right|_{(\alpha,\beta)} (\hat{\beta} - \beta) + r \Rightarrow$$

$$\hat{R} - R \simeq \left.\frac{\partial \varphi}{\partial \hat{\alpha}}\right|_{(\alpha,\beta)} (\hat{\alpha} - \alpha) + \left.\frac{\partial \varphi}{\partial \hat{\beta}}\right|_{(\alpha,\beta)} (\hat{\beta} - \beta)$$

cuya varianza es:

$$V(\hat{R}) = E[(\hat{R} - R)^2] \simeq E\left[\left(\left.\frac{\partial \varphi}{\partial \hat{\alpha}}\right|_{(\alpha,\beta)} (\hat{\alpha} - \alpha) + \left.\frac{\partial \beta}{\partial \hat{\beta}}\right|_{(\alpha,\beta)} (\hat{\beta} - \beta)\right)^2\right] =$$

$$= \dots \overset{\text{mt. general}}{=} \left(\left.\frac{\partial \varphi}{\partial \hat{\alpha}}\right|_{(\alpha,\beta)}\right)^2 V(\hat{\alpha}) + \left(\left.\frac{\partial \varphi}{\partial \hat{\beta}}\right|_{(\alpha,\beta)}\right)^2 V(\hat{\beta}) +$$

$$+ 2 \cdot \left.\frac{\partial \varphi}{\partial \hat{\alpha}}\right|_{(\alpha,\beta)} \left.\frac{\partial \varphi}{\partial \hat{\beta}}\right|_{(\alpha,\beta)} \operatorname{cov}(\hat{\alpha}, \hat{\beta}) =$$

$$\varphi(\hat{\alpha}, \hat{\beta}) = \frac{\hat{\alpha}}{\hat{\beta}} \left\{\begin{array}{l} \dfrac{\partial \varphi}{\partial \hat{\alpha}} = \left.\dfrac{1}{\hat{\beta}}\right|_{(\alpha,\beta)} = \dfrac{1}{\beta} \\ \dfrac{\partial \varphi}{\partial \hat{\beta}} = \left.-\dfrac{\hat{\alpha}}{\hat{\beta}^2}\right|_{(\alpha,\beta)} = -\dfrac{\alpha}{\beta^2} \end{array}\right.$$

$$V(\hat{R}) \simeq \frac{1}{\beta^2} V(\hat{\alpha}) + \left(-\frac{\alpha}{\beta^2}\right)^2 V(\hat{\beta}) - 2 \cdot \frac{1}{\beta} \cdot \frac{\alpha}{\beta^2} \operatorname{cov}(\hat{\alpha}, \hat{\beta})$$

$$V(\hat{R}) = \frac{V(\hat{\alpha}) + R^2 V(\hat{\beta}) - 2R \operatorname{cov}(\hat{\alpha}, \hat{\beta})}{\beta^2}$$

multiplicando y dividiendo lo necesario para que aparezca $R^2 = \frac{\alpha^2}{\beta^2}$, resulta:

$$V(\hat{R}) = \frac{\alpha^2}{\beta^2} \cdot \frac{1}{\alpha^2} V(\hat{\alpha}) + \frac{\alpha^2}{\beta^2} \cdot \frac{1}{\beta^2} V(\hat{\beta}) - 2 \cdot \frac{\alpha^2}{\beta^2} \cdot \frac{1}{\alpha\beta} \operatorname{cov}(\hat{\alpha}, \hat{\beta}) =$$

$$= R^2 \left[\frac{V(\hat{\alpha})}{\alpha^2} + \frac{V(\hat{\beta})}{\beta^2} - 2 \frac{\operatorname{cov}(\hat{\alpha}, \hat{\beta})}{\alpha\beta} \right]$$

donde: $\left.\begin{array}{l} \frac{V(\hat{\alpha})}{\alpha^2} = C^2_{\hat{\alpha}} \\ \frac{V(\hat{\beta})}{\beta^2} = C^2_{\hat{\beta}} \end{array}\right\} \rightarrow$ varianzas relativas de $\hat{\alpha}$ y $\hat{\beta}$, cuadrados de los coeficientes de variación (desv. típica/parámetro)
esperanza del estimador

$\frac{\operatorname{cov}(\hat{\alpha}, \hat{\beta})}{\alpha \cdot \beta} = C_{\hat{\alpha}\hat{\beta}} \rightarrow$ covarianza relativa de $\hat{\alpha}$ y $\hat{\beta}$.

luego la aproximación de la varianza del cociente de estimadores es:

$$V(\hat{R}) \simeq R^2 \left[C^2_{\hat{\alpha}} + C^2_{\hat{\beta}} - 2C_{\hat{\alpha}\hat{\beta}} \right].$$

4_ EL ESTIMADOR de RAZÓN: SESGO, VARIANZA y SUS ESTIMACIONES

¿sobra?

El método de estimación de la razón trata de mejorar la precisión de un estimador simple (estimador lineal obtenido por mas ó por ma estratificado, por ejemplo), utilizando la información proporcionada por una variable auxiliar Y, que se supone correlacionada con la var. de estudio X.

Al estimador clásico del total poblacional:

$$\hat{X} = N\sum_{i=1}^{n} \frac{X_i}{n} = N\bar{x} \longleftarrow \text{Estimador de expansión}$$

se ofrece como alternativa un estimador que utiliza información de Y, var. correlacionada con X:

$Y = \sum_{i=1}^{N} Y_i \rightarrow$ total poblacional (conocido)

$\hat{Y} = N \sum_{i=1}^{n} \frac{Y_i}{n} \rightarrow$ estimador del total.

El cociente $\frac{Y}{\hat{Y}}$ mide la representatividad de la muestra

$$\frac{Y}{\hat{Y}} \begin{cases} \simeq 1 & \rightarrow \text{muestra representativa} \\ \ll 1 & \rightarrow \text{muestra sobrevalora} \\ \gg 1 & \rightarrow \text{muestra infravalora} \end{cases}$$

Podemos utilizar este cociente para corregir el estimador insesgado de X con:

$\hat{X}_R = \hat{X} \cdot \frac{Y}{\hat{Y}} = \hat{R} \cdot Y \rightarrow$ estimador del total por el mt. de la razón

con $\hat{R} = \frac{\hat{X}}{\hat{Y}} \rightarrow$ estimador de la razón

Para la media: $\hat{\bar{X}}_R = \hat{R} \cdot \bar{Y}$

Para el total de clase y la proporción, sustituir por una variable cualitativa.

El estimador de la razón es ~~igual~~ el cociente entre los totales o las medias muestrales, y se utiliza como estimador de la razón poblacional, de totales o de medias ~~muestrales~~ poblacionales.

$$R = \frac{X}{Y} = \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i} = \frac{\sum_{i=1}^{N} X_i / N}{\sum_{i=1}^{N} Y_i / N} = \frac{\bar{X}}{\bar{Y}} \leftarrow \text{razón poblacional}$$

$$\hat{R} = \frac{\hat{X}}{\hat{Y}} = \frac{x}{y} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i} = \frac{\sum^{n} X_i / n}{\sum^{n} Y_i / n} = \frac{\bar{x}}{\bar{y}} \leftarrow \text{razón muestral}$$

SESGO del estimador

Mediante la estimación por razón se trata de estimar la razón poblacional R mediante la razón muestral $\hat{R}$. Aunque $\hat{X}$ e $\hat{Y}$ sean estimadores insesgados de X e Y, $\hat{R}$ no suele ser insesgado de R.

$$\left.\begin{array}{l} E[\hat{X}] = X \\ E[\hat{Y}] = Y \end{array}\right\} \text{ pero } E\left[\frac{\hat{X}}{\hat{Y}}\right] \neq \frac{E[\hat{X}]}{E[\hat{Y}]} \longrightarrow E[\hat{R}] \neq R.$$

En general, $E[\hat{R}] = R + B(\hat{R})$, donde $B(\hat{R}) \equiv$ sesgo del estimador

La expresión exacta del sesgo del estimador se debe a Hartley y Ross (1954), y se obtiene a partir de la covarianza entre el estimador de razón y la media muestral de la variable auxiliar, $\text{cov}(\hat{R}, \bar{y})$

$$\text{cov}(\hat{R},\bar{y}) = E[\hat{R}\bar{y}] - E[\hat{R}]E[\bar{y}] =$$
$$= E\left[\frac{\bar{x}}{\cancel{\bar{y}}}\cdot\cancel{\bar{y}}\right] - E[\hat{R}]\,\bar{Y} = \bar{X} - E[\hat{R}]\bar{Y} \Rightarrow$$
$$\Rightarrow E[\hat{R}] = \frac{\bar{X}}{\bar{Y}} - \frac{\text{cov}(\hat{R},\bar{y})}{\bar{Y}} = R - \frac{\text{cov}(\hat{R},\bar{y})}{\bar{Y}}$$

$$E[\hat{R}] = \cancel{R} + B(\hat{R}) = \cancel{R} - \frac{\text{cov}(\hat{R},\bar{y})}{\bar{Y}}$$, de donde

$$B(\hat{R}) = -\frac{\text{cov}(\hat{R},\bar{y})}{\bar{Y}} = \frac{-\rho_{\hat{R}\bar{y}}\cdot\sigma_{\hat{R}}\sigma_{\bar{y}}}{E[\bar{y}]} = -\rho_{\hat{R}\bar{y}}\cdot\sigma_{\hat{R}}\,C_{\bar{y}}$$ desviación típica relativa

De esta expresión se deduce que si $\hat{R}$ e $\bar{y}$ están incorrelacionadas, entonces el sesgo es nulo $\Rightarrow E[\hat{R}] = R \Rightarrow$ primera condición para la insesgadez del estimador de razón.

El sesgo se puede acotar en términos ~~del co~~ relativos:

$$B(\hat{R}) = -\frac{\text{cov}(\hat{R},\bar{y})}{\bar{Y}} = -\frac{\rho_{\hat{R}\bar{y}}\,\sigma_{\hat{R}}\,\sigma_{\bar{y}}}{E[\bar{y}]} = -\rho_{\hat{R}\bar{y}}\,\sigma_{\hat{R}}\cdot C_{\bar{y}}$$

Teniendo en cuenta que el valor absoluto del coef. de correlación está acotado por 1:

$$|B(\hat{R})| \le \sigma_{\hat{R}}\cdot C_{\bar{y}} \Rightarrow \left|\frac{B(\hat{R})}{\sigma_{\hat{R}}}\right| \le C_{\bar{y}} \quad \boxed{< 0'1}$$ despreciable si

Por lo que el sesgo relativo del estimador de razón siempre será menor que el coeficiente de variación de $\bar{y}$.

Para que el sesgo relativo sea despreciable bastará con que el coef. de variación de la media muestral de la var. auxiliar sea menor que 0'1, ~~que se consigue con tamaños muestrales~~ ~~$B(\hat{R}$~~ ~~mayores que $100\frac{\sigma_Y^2}{\bar{Y}^2}$ (CR)~~

$\longrightarrow$ mirar atrás

$C_{\bar{y}}$ = Coef. de variación de $\bar{y}$, $Cv(\bar{y}) = \dfrac{\sigma_{\bar{y}}}{E[\bar{y}]}$.

$$\left|\frac{B(\hat{R})}{\sigma_{\hat{R}}}\right| \text{ despreciable si } C_{\bar{y}} < 0'10 \sim n > 100\frac{\sigma_Y^2}{\bar{Y}^2} \quad (CR)$$

Una segunda condición para la insesgadez del estimador de razón es: "si la recta de regresión de la variable auxiliar Y sobre la variable en estudio X (o la recta X sobre Y) pasa por el origen de coordenadas, entonces el estimador de razón $\hat{R}$ es insesgado de R.

Lo vemos:

Si (X_i, Y_i) están en la recta Y/X que pasa por (0,0) se cumple que $Y_i = K X_i$, por lo que:

$$R = \frac{X}{Y} = \frac{\sum^N X_i}{\sum^N Y_i} = \frac{\sum X_i}{K\sum X_i} = \frac{1}{K} \quad (K \text{ para } X_i = KY_i).$$

$$\hat{R} = \frac{\bar{x}}{\bar{y}} = \frac{\sum^n X_i}{\sum^n Y_i} = \frac{\sum^n Y_i}{K\sum^n X_i} = \frac{1}{K} \quad (K \text{ para } X_i = KY_i).$$

Por lo que $E[\hat{R}] = E[\frac{1}{K}] = \frac{1}{K} = R \Rightarrow B(\hat{R})$ insesgado.

Un estimador de razón que siempre resulta insesgado es el estimador de Hartley y Ross:

En un muestreo con probabilidades iguales y sin reposición consideramos la media muestral de las razones poblacionales.

$$\hat{R}_1 = \frac{1}{n}\sum_{i=1}^{n} R_i \qquad \text{con } \hat{B}(\hat{R}_1) = -\frac{N-1}{N\bar{Y}}\cdot\frac{n}{n-1}(\bar{x} - \hat{R}_1\bar{y})$$

El estimador de Hartley y Ross es el estimador de $\hat{R}_1$ corregido del sesgo, que es insesgado de R.

$$\hat{R}_{HR} = \hat{R}_1 - \hat{B}(\hat{R}_1) = \hat{R}_1 + \frac{N-1}{N\bar{Y}}\cdot\frac{n}{n-1}(\bar{x} - \hat{R}_1\bar{y})$$

~~ESTIMACION~~ APROXIMACIÓN del SESGO:

La expresión exacta del sesgo del estimador no se puede obtener a partir de los datos de la muestra, ¿porqué?

Se puede hallar una expresión generalizada del sesgo considerando la diferencia entre la razón muestral y la razón poblacional:

$$\hat{R}-R = \frac{\bar{x}}{\bar{y}} - R = \frac{\bar{x}-R\bar{y}}{\bar{y}} \overset{\text{sp. } \bar{Y}\neq 0}{=} \frac{\bar{x}-R\bar{y}}{\bar{Y}} \cdot \frac{\bar{Y}}{\bar{y}} = \ldots$$

$$\frac{\bar{Y}}{\bar{y}} = \frac{\bar{Y}}{\bar{Y}+\bar{y}-\bar{Y}} = \underbrace{\frac{1}{1+\frac{\bar{y}-\bar{Y}}{\bar{Y}}}}_{\text{suma progr. geométrica de razón } -\frac{\bar{y}-\bar{Y}}{\bar{Y}}}$$

con la condición $\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right| < 1$

$$\hat{R}-R = \frac{\bar{x}-R\bar{y}}{\bar{Y}} \cdot \left[1 - \frac{\bar{y}-\bar{Y}}{\bar{Y}} + \left(\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)^2 - \ldots\right]$$

Despreciando los términos del desarrollo en serie de orden ≥ 2 y tomando esperanzas:

$$\underline{B(\hat{R})} = E[\hat{R}]-R = E[\hat{R}-R] \simeq E\left[\frac{\bar{x}-R\bar{y}}{\bar{Y}}\left(1-\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right)\right] =$$

$$= E\left[\frac{\bar{x}-R\bar{y}}{\bar{Y}} - \frac{(\bar{x}-R\bar{y})\cdot(\bar{y}-\bar{Y})}{\bar{Y}^2}\right] =$$

$$= \underbrace{\frac{E[\bar{x}-R\bar{y}]}{\bar{Y}}}_{=0} - \frac{E[(\bar{x}-R\bar{y})(\bar{y}-\bar{Y})]}{\bar{Y}^2} \overset{?}{=} \underline{\frac{R\cdot Var(\bar{y}) - cov(\bar{x},\bar{y})}{\bar{Y}^2}}$$

$$B(\hat{R}) = \frac{R\, Var(\bar{y}) - cov(\bar{x},\bar{y})}{\bar{Y}^2} = R\frac{Var(\bar{y})}{E[\bar{y}]^2} - \frac{\bar{X}}{\bar{Y}}\frac{cov(\bar{x},\bar{y})}{E[\bar{x}]E[\bar{y}]} =$$

$$\simeq \underline{R\left(C^2_{\bar{y}} - C_{\bar{x}\bar{y}}\right)}$$

Mirar atrás ⟹

ESTIMACIÓN de la APROX:
La expresión anterior puede estimarse mediante los valores muestrales:

$$\hat{B}(\hat{R}) \simeq \hat{R}\left[\hat{C}^2_{\bar{y}} - \hat{C}_{\bar{x}\bar{y}}\right]$$ ← estimador del sesgo del estimador de razón

donde: $\hat{C}^2_{\bar{y}} = \frac{\hat{V}(\bar{y})}{\bar{y}^2}$ ← estimador de la varianza relativa de $\bar{y}$

$\hat{C}_{\bar{x}\bar{y}} = \frac{\widehat{cov}(\bar{x},\bar{y})}{\bar{x}\,\bar{y}}$ ← estimador de la covarianza relativa.

que se puede particularizar para los ≠ tipos de muestreo:

– muestreo sin reposición:

$$\left.\begin{aligned}\hat{V}(\bar{y}) &= (1-f)\frac{\hat{S}^2_y}{n}\\ \widehat{cov}(\bar{x},\bar{y}) &= (1-f)\frac{\hat{S}_{xy}}{n}\end{aligned}\right\} \rightarrow \hat{B}(\hat{R}) \simeq \hat{R}\frac{(1-f)}{n}\left[\frac{\hat{S}^2_y}{\bar{y}^2} - \frac{\hat{S}_{xy}}{\bar{x}\bar{y}}\right]$$

El sesgo disminuye al aumentar el tamaño muestral n.
Para muestras suficientemente grandes, es despreciable.

– Muestreo con reposición:

$$\left.\begin{aligned}\hat{V}(\bar{y}) &= \frac{\hat{S}^2_y}{n}\\ \widehat{cov}(\bar{x},\bar{y}) &= \frac{\hat{S}_{xy}}{n}\end{aligned}\right\} \rightarrow \hat{B}(\hat{R}) \simeq \frac{\hat{R}}{n}\left[\frac{\hat{S}^2_y}{\bar{y}^2} - \frac{\hat{S}_{xy}}{\bar{x}\bar{y}}\right].$$

→ Igual.

<u>VARIANZA</u> del estimador de razón:

Aproximamos $\hat{R} - R$ por el primer término de su desarrollo en serie y le aplicamos el operador varianza:

$$\hat{R} - R \simeq \frac{\bar{x} - R\bar{y}}{\bar{Y}}$$

$$E[\hat{R} - R] \simeq E\left[\frac{\bar{x} - R\bar{y}}{\bar{Y}}\right] = \frac{1}{\bar{Y}}\left[\bar{X} - R\bar{Y}\right] = R - R = 0 \Rightarrow E[\hat{R}] = R.$$

$$V[\hat{R}] = V[\hat{R} - R] \simeq V\left[\frac{\bar{x} - R\bar{y}}{\bar{Y}}\right] = \frac{V(\bar{x}) + R^2 V(\bar{y}) - 2R\,COV(\bar{x},\bar{y})}{\bar{Y}^2}$$

que en términos de coeficientes de variación puede expresarse:

$$V[\hat{R}] \simeq \frac{\bar{X}^2}{\bar{Y}^2}\cdot\frac{V(\bar{x})}{\bar{X}^2} + R^2\cdot\frac{V(\bar{y})}{\bar{Y}^2} - 2R\cdot\frac{\bar{X}}{\bar{Y}}\cdot\frac{COV(\bar{x},\bar{y})}{\bar{X}\cdot\bar{Y}} \Rightarrow$$

$$V[\hat{R}] \simeq R^2\left[C^2_{\bar{x}} + C^2_{\bar{y}} - 2C_{\bar{x}\bar{y}}\right]$$

cuya expresión se puede detallar para los distintos tipos de muestreo:

$$SR \longrightarrow V[\hat{R}] \simeq \frac{1-f}{n\cdot\bar{Y}^2}\left[S^2_x + R^2S^2_y - 2RS_{xy}\right]$$

$$CR \longrightarrow V[\hat{R}] \simeq \frac{1}{n\,\bar{Y}^2}\left[\sigma^2_x + R^2\sigma^2_y - 2R\sigma_{xy}\right]$$

que se pueden estimar utilizando la información muestral: ¿insesgadas?

$$\hat{V}(\hat{R}) \simeq \hat{R}^2\left[\hat{C}^2_{\bar{x}} + \hat{C}^2_{\bar{y}} - 2\hat{C}_{\bar{x}\bar{y}}\right]$$

$$SR \longrightarrow \hat{V}(\hat{R}) \simeq \frac{1-f}{n\,\bar{y}^2}\left[\hat{S}^2_x + \hat{R}^2\hat{S}^2_y - 2\hat{R}\hat{S}_{xy}\right]$$

$$CR \rightarrow \hat{V}(\hat{R}) \simeq \frac{1}{n\bar{y}^2}\left[\hat{\sigma}^2_x + \hat{R}^2\hat{\sigma}^2_y - 2\hat{R}\hat{\sigma}_{xy}\right]$$

COMPARACIÓN del estimador de razón con el estimador de expansión

Sp. un muestreo aleatorio simple sin reposición:

Total poblacional $\longrightarrow X = \sum_{i=1}^{N} X_i$

$V(\hat{X}) = N^2(1-f)\frac{S_x^2}{n}$

Estimador de expansión $\longrightarrow \hat{X} = \frac{N}{n}\sum_{i=1}^{n} x_i = N\bar{x}$

Estimador de razón $\longrightarrow \hat{X}_R = \hat{X}\cdot\frac{Y}{\hat{Y}} = \hat{R}Y$

$$V(\hat{X}_R) = V(\hat{R}Y) \overset{Y \text{ conocido}}{=} Y^2 V(\hat{R}) \overset{SP}{\simeq} Y^2\frac{(1-f)}{\bar{Y}^2 n}\left[S_x^2 + R^2S_y^2 - 2RS_{xy}\right] =$$

$Y = N\bar{Y}$ $\longrightarrow$

$$= \frac{N^2\cancel{\bar{Y}^2}(1-f)}{\cancel{\bar{Y}^2}n}\left[S_x^2 + R^2S_y^2 - 2RS_{xy}\right] =$$

$$= V(\hat{X}) + N^2\frac{(1-f)}{n}R^2S_y^2 - 2N^2\frac{(1-f)}{n}RS_{xy} =$$

$$= V(\hat{X}) + N^2\frac{(1-f)}{n}RS_y\left[RS_y - \frac{2S_{xy}}{S_y}\right]$$

$\hat{X}_R$ más preciso que $\hat{X}$ si $V(\hat{X}_R) < V(\hat{X}) \Longleftrightarrow$

$$\underbrace{N^2\frac{(1-f)}{n}RS_y}_{>0}\left[RS_y - \frac{2S_{xy}}{S_y}\right] < 0 \Longleftrightarrow RS_y - \frac{2S_{xy}}{S_y} < 0$$

$\frac{S_{xy}}{S_xS_y} = \rho_{xy}$ $\longrightarrow$

$$\Longleftrightarrow RS_y - 2S_x\rho_{xy} < 0 \Longleftrightarrow$$

$$\underbrace{RS_y}_{>0} < 2S_x\rho_{xy} \Longleftrightarrow$$

$$\rho_{xy} > \frac{1}{2}R\frac{S_y}{S_x} = \frac{1}{2}\cdot\frac{S_y/Y}{S_x/X} = \frac{1}{2}\frac{C_y}{C_x}$$

el estimador de razón es más preciso que el estimador de expansión si la correlación entre X e Y es positiva y mayor que $\frac{1}{2}\frac{C_y}{C_x}$.