Ratio Estimation – an estimation procedure that can

1. estimate $\bar{X}/\bar{Y}$ with $\bar{x}/\bar{y}$;
2. provide an estimate of X (population total) that is more accurate than $x' = \left(\frac{N}{n}\right)x$.

I. Ratio Estimation for SRS of size n from a population of size N.

A. Population Ratio

$$R = \frac{X}{Y} = \frac{X/N}{Y/N} = \bar{X}/\bar{Y}$$

B. Ratio Estimate

$$r = \frac{x'}{y'} = \frac{x}{y} = \bar{x}/\bar{y}$$

1. Both numerator and denominator are subject to sampling variation.
2. r in general is not unbiased.
3. The bias is usually small – r is widely used.

C. Standard Error of r

1. Exact expressions for SE(r) can not be derived due to sampling variability of each variable.

2. Approximate SE(r) (Hansen, Hurwitz, Madow)
If $V(\bar{y}) \le 0.05$,

where $V(\bar{y}) = \frac{SE(\bar{y})}{E(\bar{Y})} = \frac{\frac{\sigma_y}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)^{\frac{1}{2}}}{\bar{Y}}$,

then

$$SE(r) \approx \left(\frac{R}{\sqrt{n}}\right)\left(V_x^2 + V_y^2 - 2\rho_{xy}V_xV_y\right)^{\frac{1}{2}}\sqrt{\left(\frac{N-n}{N-1}\right)} \qquad (7.1)$$

where ρ_{xy} is the correlation coefficient between x and y,

$$\rho_{xy} = \frac{\sum_{i=1}^{N}\frac{(X_i - \bar{X})(Y_i - \bar{Y})}{N}}{\sigma_x\sigma_y} \qquad (7.2)$$

3. Estimated SE(r)

If $\hat{V}(\bar{y}) = \left(\frac{\frac{s_y}{\sqrt{n}}}{\bar{y}}\right)\left(\frac{N-n}{N}\right)^{\frac{1}{2}} \le .05$, then

$$\hat{SE}(r) = \left(\frac{r}{\sqrt{n}}\right)\left(\hat{V}_x^2 + \hat{V}_y^2 - 2\hat{\rho}_{xy}\hat{V}_x\hat{V}_y\right)^{\frac{1}{2}}\sqrt{\left(\frac{N-n}{N-1}\right)} \qquad (7.3)$$

where

$$\hat{V}_x^2 = \left(\frac{N-1}{N}\right)\frac{s_x^2}{\bar{x}^2}, \quad \hat{V}_y^2 = \left(\frac{N-1}{N}\right)\frac{s_y^2}{\bar{y}^2}, \quad \hat{\rho}_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

4. <u>100(1-α)% C.I. for R:</u>
   $$r \pm Z_{1-\frac{\alpha}{2}}\hat{S}E(r)$$

II. <u>Ratio Estimation of Totals for SRS</u>

A. <u>Ratio Estimate of X</u>

$$R = \frac{X}{Y} \Rightarrow X = RY \Rightarrow \hat{X} = \hat{R}Y = rY$$

$$\boxed{x'' = rY} \qquad (7.5)$$

B. <u>Recall</u>: Inflation estimates of totals: $x' = \left(\frac{N}{n}\right)x$, $y' = \left(\frac{N}{n}\right)y$

Therefore,

$$x'' = rY = \left(\frac{x}{y}\right)Y \Rightarrow x'' = \left(\frac{x'}{y'}\right)Y \Rightarrow \boxed{x'' = \left(\frac{Y}{y'}\right)x'}$$

1. $y' < Y \Rightarrow x'' > x'$
2. $y' > Y \Rightarrow x'' < x'$

i.e. since X and Y are correlated, if y' underestimates Y, then x' underestimates X. However, x'' is adjusted up to compensate.

C. x' is unbiased and x'' is biased, but in general

$$MSE(x'') < MSE(x')$$

$$E(x'' - X)^2 < E(x' - X)^2$$

D. <u>Approximate Standard Error of x''</u>

$$VAR(x'') = VAR(rY) = Y^2 VAR(r)$$

$$SE(x'') = Y(SE(r)) \approx \left(\frac{YR}{\sqrt{n}}\right)\left(V_x^2 + V_y^2 - 2\rho_{xy}V_xV_y\right)^{\frac{1}{2}}\left(\frac{N-n}{N-1}\right)^{\frac{1}{2}} \qquad (7.6)$$

if $V(\bar{y}) < .05$.

E. <u>Estimated Standard Error of x''</u>

$$\hat{SE}(x'') = \left(\frac{Yr}{\sqrt{n}}\right)\left(\hat{V}_x^2 + \hat{V}_y^2 - 2\hat{\rho}_{xy}\hat{V}_x\hat{V}_y\right)^{\frac{1}{2}}\sqrt{\left(\frac{N-n}{N-1}\right)} \qquad (7.7)$$

if $\hat{V}(\bar{y}) < .05$.

F. <u>$100(1-\alpha)\%$ C.I. for X:</u>

$$x'' \pm Z_{1-\frac{\varepsilon}{2}}\hat{SE}(x'')$$

G. A comparison of the ratio estimate (x'') and the inflation estimate (x') of X

<u>Recall</u> $MSE(x'') = VAR(x'') + B^2(x'')$, and

$MSE(x') = VAR(x')$.

Ignoring $B^2(x'')$, which is negligible for large n,

$$\frac{MSE(x'')}{MSE(x')} \approx \frac{VAR(x'')}{VAR(x')}$$

$$= \frac{\left(\frac{Y^2R^2}{n}\right)\left(V_x^2 + V_y^2 - 2\rho_{xy}V_xV_y\right)\left(\frac{N-n}{N-1}\right)}{\left(\frac{X^2}{n}\right)V_x^2\left(\frac{N-n}{N-1}\right)} = \frac{V_x^2 + V_y^2 - 2\rho_{xy}V_xV_y}{V_x^2}$$

$$\Rightarrow \boxed{VAR(x'') < VAR(x') \; when \; \frac{V_y}{V_x} < 2\rho_{xy}} \qquad (7.8)$$

Thus, for fixed population C.V., greater correlation between X and Y $\Rightarrow$ x'' has a smaller variance than x'.