Statistics 480: Survey Sampling Techniques

Walter R. Mebane, Jr.

University of Michigan

GSIs: Adam Hall and Fabricio Vasselai
B844 East Hall, 14 February 2018

Outline

ratio estimation: motivation

- having a sample of measurements $y_1, y_2, \ldots, y_n$,
  - the ratio of the totals of two variables y and x

$$R = \frac{\tau_y}{\tau_x}$$

    may be of direct interest
  - using y and one or more subsidiary variables can produce better estimates for τ_y than $\hat{\tau}_y$, or better estimates for μ_y than $\bar{y}$
  - estimators for τ_y and μ_y can be improved if the subsidiary measures are strongly correlated with y
- terminology
  - instead of "subsidiary" variable the term "auxiliary" variable is often used

ratio estimation

- correlation, meaning the product moment correlation coefficient: $\rho = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

$$\hat{\rho} = \frac{s_{xy}}{s_x s_y}$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- see file `corr.R`

ratio estimation

- simple motivating example: total sugar content of a load of oranges
  - y_i: measured sugar context of orange i
  - x_i: measured weight of orange i
  - N: number of oranges in load (unknown)
  - τ_y, μ_y: total sugar content of load and mean sugar content
  - τ_x, μ_x: total weight of load and mean weight
- μ_y can be estimated by $\bar{y}$ from a sample of n oranges, so if N were known, the total could be estimated by $N\bar{y}$
- but N is unknown and expensive to count and to estimate
- τ_x, the total weight of the load is easy to measure, and μ_x is easy to estimate using $\bar{x}$

ratio estimation

- simple motivating example: total sugar content of a load of oranges
- but N is unknown and expensive but τ_x and μ_x are easy to determine
- an estimator can be based on the relationship

$$\frac{\mu_y}{\mu_x} = \frac{N\mu_y}{N\mu_x} = \frac{\tau_y}{\tau_x}$$

  solving for τ_y gives

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

- given a measure for τ_x (say done by weighing the truck), the ratio estimator for τ_y is the natural

$$\hat{\tau}_y = \tau_x \frac{\bar{y}}{\bar{x}}$$

ratio estimation

- simple motivating example: total sugar content of a load of oranges
- given a measure for τ_x (say done by weighing the truck), the ratio estimator for τ_y is the natural

$$\hat{\tau}_y = \tau_x \frac{\bar{y}}{\bar{x}}$$

- using the sample size n, equivalent expressions are given by

$$\hat{\tau}_y = \tau_x \frac{n\bar{y}}{n\bar{x}} = \tau_x \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$$

- even if N were known, the ratio estimator $\hat{\tau}_y = \tau_x \bar{y}/\bar{x}$ can be better than the estimator $\hat{\tau}_y = N\bar{y}$, if x and y are highly correlated
    - this is because the ratio estimator uses more information

ratio estimation

- to estimate the ratio of the totals of two variables y and x

$$R = \frac{\tau_y}{\tau_x}$$

- given a simple random sample of size n, an estimator for R:

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

- estimated variance of r:

$$\hat{V}(r) = \hat{V}\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}\right) = \left(\frac{N-n}{N}\right)\left(\frac{1}{\mu_x^2}\right)\frac{s_r^2}{n}$$

$$s_r^2 = \frac{\sum_{i=1}^{n}(y_i - r x_i)^2}{n-1}$$

$\bar{x}^2$ can be used for μ_x^2 in $\hat{V}(r)$ if μ_x^2 is unknown (when $\sqrt{V(\bar{x})}/\bar{x} < 0.10$).

ratio estimation

- Example 6.1: from a survey of housing, samples that measure the costs of home ownership in 13 MSAa in two years (1994 and 2002); estimate the ratio of mean costs for 2002 compared to costs in 1994, and calculate an appropriate margin of error
- see file `example6.1.R`

ratio estimation

- another way of writing the estimated variance of the ratio estimator for R: using $f = n/N$ (the sampling fraction)

$$\hat{V}(r) = \frac{1-f}{n}\left(\frac{1}{\mu_x^2}\right)\left(s_y^2 + r^2 s_x^2 - 2r\hat{\rho}s_x s_y\right)$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

ratio estimation

- yet another way of writing the estimated variance of the ratio estimator for R, based on

$$\hat{V}(r) = \frac{1-f}{n}\left(\frac{1}{\mu_x^2}\right)(s_y^2 + r^2 s_x^2 - 2r\hat{\rho}s_x s_y)$$

- replacing μ_x with $\bar{x}$,

$$\begin{aligned}\hat{V}(r) &= \frac{1-f}{n} r^2 \left(\frac{s_y^2}{\bar{y}^2} + \frac{s_x^2}{\bar{x}^2} - 2\hat{\rho}\frac{s_x s_y}{\bar{y}\bar{x}}\right) \\ &= \frac{1-f}{n} r^2 \left[(\mathrm{cv}(y))^2 + (\mathrm{cv}(x))^2 - 2\hat{\rho}\,\mathrm{cv}(y)\,\mathrm{cv}(x)\right]\end{aligned}$$

where $\mathrm{cv}(x)$ denotes the coefficient of variation of x:

$$\mathrm{cv}(x) = \frac{s_x}{\bar{x}}$$

ratio estimation

- given a simple random sample of size n, an estimator for τ_y:

$$\hat{\tau}_y = \left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \right) \tau_x = r\tau_x$$

- estimated variance of $\hat{\tau}_y$:

$$\begin{aligned} \hat{V}(\hat{\tau}_y) &= \tau_x^2 \left(\frac{N-n}{N} \right) \left(\frac{1}{\mu_x^2} \right) \frac{s_r^2}{n} \\ &= N^2 \left(\frac{N-n}{N} \right) \frac{s_r^2}{n} \end{aligned}$$

where

$$s_r^2 = \frac{\sum_{i=1}^n (y_i - rx_i)^2}{n-1}$$

ratio estimation

- given a simple random sample of size n, an estimator for τ_y:

$$\hat{\tau}_y = \left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \right) \tau_x = r\tau_x$$

- estimated variance of $\hat{\tau}_y$:

$$\hat{V}(\hat{\tau}_y) = N^2 \left(\frac{N-n}{N} \right) \frac{s_r^2}{n}, \quad s_r^2 = \frac{\sum_{i=1}^n (y_i - r x_i)^2}{n-1}$$

  - compare $\hat{V}(\hat{\tau}_y)$ to $\hat{V}(\hat{\tau})$, the variance of the SRS estimator of the total

$$\hat{V}(\hat{\tau}) = N^2 \left(\frac{N-n}{N} \right) \frac{s^2}{n}, \quad s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

ratio estimation

- Example 6.2: total sugar content of a load of oranges; given a random sample of size $n = 10$, with measures of sugar content (y) and weight (x), estimate the total sugar content of the load, given the total weight of the load $\tau_x = 1800$
- see file `example6.2.R`

ratio estimation

- given a simple random sample of size n, an estimator for μ_y:

$$\hat{\mu}_y = \left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \right) \mu_x$$

- estimated variance of $\hat{\mu}_y$:

$$\begin{aligned} \hat{V}(\hat{\mu}_y) &= \mu_x^2 \left(\frac{N-n}{N} \right) \left(\frac{1}{\mu_x^2} \right) \frac{s_r^2}{n} \\ &= \left(\frac{N-n}{N} \right) \frac{s_r^2}{n} \end{aligned}$$

where

$$s_r^2 = \frac{\sum_{i=1}^n (y_i - r x_i)^2}{n-1}$$

ratio estimation

- Example 6.3: given complete sugercane data from 1997 for $N = 32$ counties, and a sample of size $n = 6$ for 1999, estimate the total in 1999; average acres in 1997, 27,752 acres; mean production in 1997, 987,839 tons; estimate mean acreage in 1999 and calculate an appropriate margin of error
- see file `example6.3.R`

ratio estimation: sample size determination

- planning the sample size with ratio estimators
- for the estimator for R, the sample size computation is based on the familiar rule of solving for n in

$$2\sqrt{V(r)} = B$$

  where B is the desired margin of error
- $V(r)$ is not provided, but use $\hat{V}(r)$:

$$\hat{V}(r) = \left(\frac{N-n}{N}\right)\left(\frac{1}{\mu_x^2}\right)\frac{s_r^2}{n}$$

- replace s_r^2 with σ^2 (the population variance of y)

$$V(r) = \left(\frac{N-n}{N}\right)\left(\frac{1}{\mu_x^2}\right)\frac{\sigma^2}{n}$$

ratio estimation: sample size determination

- for the estimator for R, solve

$$2\sqrt{V(r)} = 2\sqrt{\left(\frac{N-n}{N}\right)\left(\frac{1}{\mu_x^2}\right)\frac{\sigma^2}{n}} = B$$

gives

$$n = \frac{N\sigma^2}{ND + \sigma^2}$$

where

$$D = \frac{B^2\mu_x^2}{4}$$

ratio estimation: sample size determination

- Example 6.4: estimate the ratio of last year's to this year's number of hours lost due to sickness; preliminary sample of $n' = 10$ employee records is available; last year $\tau_x = 16,300$ hours were lost for $N = 1000$ employees; determine a sample size to estimate R with bound $B = 0.01$
- see file `example6.4.R`

ratio estimation: sample size determination

- for the estimator for μ_y, the same procedure gives

$$n = \frac{N\sigma^2}{ND + \sigma^2}$$

where

$$D = \frac{B^2}{4}$$

ratio estimation: sample size determination

- Example 6.5: estimate average number of trees μ_y per acre on a plantation of $N = 1000$ acres using a sample of n 1-acre plots; the actual number of trees on each plot is y, which is estimated as x from an aerial survey; find the sample size needed to get a bound $B = 1.0$
- see file `example6.5.R`

ratio estimation: sample size determination

- for the estimator for τ_y, the same procedure gives

$$n = \frac{N\sigma^2}{ND + \sigma^2}$$

where

$$D = \frac{B^2}{4N^2}$$

ratio estimation: sample size determination

- Example 6.6: compare the actual dollar value of an inventory, τ_y, to the recorded inventory, τ_x; from $N = 2100$ item types and knowing $\tau_x = 45000$, determine the sample size n needed to estimate τ_y with bound $B = 500$
- see file `example6.6.R`

ratio estimation in stratified random sampling

- separate ratio estimator: estimate r in each stratum then compute a weighted average of the separate estimates

$$\hat{\mu}_{y\mathrm{RS}} = \sum_{i=1}^{L} \left(\frac{N_i}{N}\right) r_i \mu_{xi}, \quad r_i = \frac{\bar{y}_i}{\bar{x}_i}$$

$$\hat{V}(\hat{\mu}_{y\mathrm{RS}}) = \sum_{i=1}^{L} \left(\frac{N_i}{N}\right)^2 \left(\frac{N_i - n_i}{N_i}\right) \frac{s_{ri}^2}{n_i}, \quad s_{ri}^2 = \frac{\sum_{k=1}^{n_i} (y_k - r_i x_k)^2}{n_k - 1}$$

- combined ratio estimator: estimate μ_y by $\bar{y}_{\mathrm{st}}$, μ_x by $\bar{x}_{\mathrm{st}}$ and then estimate $R = \mu_y/\mu_x$ by $\bar{y}_{\mathrm{st}}/\bar{x}_{\mathrm{st}}$

$$\hat{\mu}_{y\mathrm{RC}} = \left(\frac{\bar{y}_{\mathrm{st}}}{\bar{x}_{\mathrm{st}}}\right) \mu_x, \quad r_C = \frac{\bar{y}_{\mathrm{st}}}{\bar{x}_{\mathrm{st}}}$$

$$\hat{V}(\hat{\mu}_{y\mathrm{RC}}) = \sum_{i=1}^{L} \left(\frac{N_i}{N}\right)^2 \left(\frac{N_i - n_i}{N_i}\right) \frac{s_{ri}^2}{n_i}, \quad s_{ri}^2 = \frac{\sum_{k=1}^{n_i} (y_k - r_C x_k)^2}{n_k - 1}$$

ratio estimation in stratified random sampling

- Example 6.7: estimate the ratio of last year's to this year's number of hours lost due to sickness; to the data from company A given in Example 6.4, a sample of $n_A = 10$ employee records, add a sample of size $n_B = 10$ from company B; last year $\tau_{xA} = 16,300$ for $N_A = 1000$ employees and $\tau_{xB} = 12,800$ for $N_B = 1500$ employees; find the separate ratio estimate of μ_x and its estimated variance
- see file `example6.7.R`
- Example 6.8: same situation as Example 6.7; find the combined ratio estimate of μ_x and its estimated variance
- see file `example6.8.R` (same as file `example6.7.R`)

ratio estimation in stratified random sampling

- as in Examples 6.7 and 6.8, the combined ratio estimator usually has a larger variance than the separate ratio estimator does, and so this motivates the latter's more frequent use
- but the separate ratio estimator may have a bigger bias, because L different estimates r_i are being used instead of a single estimate r_C
  - so if $n_i \geq 20$ for all strata, in which case the bias in each r_i is not too bad, use the separate ratio estimator, otherwise use the combined ratio estimator
  - also use the combined ratio estimator if the within-stratum ratios are all approximately equal, because in that case we may have $\hat{V}(\hat{\mu}_{y\mathrm{RS}}) > \hat{V}(\hat{\mu}_{y\mathrm{RC}})$

regression estimation

- regression estimation rather than ratio estimation may be appropriate when there is evidence that the linear relationship between y and x is best described as one that does not pass through the origin
  - the fact that the relationship does not go through the origin is additional information the use of which may improve the estimation of μ_y (or τ_y, etc.)
  - μ_x must be known before the regression estimator can be used, just as is the case with the ratio estimator
- a linear relationship between y and x may be represented as an equation for observations i:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where $E(u_i|x_i) = 0$; "goes through the origin" means $\beta_0 = 0$

regression estimation

- for observations indexed by i, define predictions

$$\hat{y}_i = a + bx_i$$

and residuals (prediction errors)

$$e_i = y_i - \hat{y}_i$$

such that a and b solve (least squares estimates)

$$\underset{a,b}{\operatorname{argmin}} \sum_i e_i^2$$

- using least squares in a sample of size n, the estimator for b is

$$b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

regression estimation

- starting with $\hat{y}_i = a + bx_i$ where a and b are least squares estimates, the intercept a satisfies

$$a = \bar{y} - b\bar{x}$$

- substituting this value for a into the prediction equation gives

$$\hat{y}_i = \bar{y} + b(x_i - \bar{x})$$

  this gives a prediction of $\hat{y}_i$ for any value of x_i

- the regression estimator $\hat{\mu}_{y\mathrm{L}}$ is given by substituting μ_x for x_i

regression estimation

- regression estimator for μ_y:

$$\hat{\mu}_{y\mathrm{L}} = \bar{y} + b(\mu_x - \bar{x})$$

where

$$b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- estimated variance of $\hat{\mu}_{y\mathrm{L}}$:

$$\hat{V}(\hat{\mu}_{y\mathrm{L}}) = \left(\frac{N-n}{Nn}\right)\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - b^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

where

$$\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - b^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

is the mean square error from the linear regression of y on x

regression estimation

- Example 6.9: $N = 486$ students have math achievement scores x_i, and a simple random sample of $n = 10$ have calculus grades y_i; given $\mu_x = 52$, estimate μ_y and put a bound on the error of estimation
- see file `example6.9.R`

difference estimation

- difference estimation is regression estimation with the coefficient b set equal to 1.0
- works when
  1. x is highly correlated with y
  2. x and y are measured on the same scale

difference estimation

- difference estimator for μ_y:

$$\hat{\mu}_{y\mathrm{D}} = \bar{y} + \mu_x - \bar{x} = \mu_x + \bar{d}$$

  where

$$\bar{d} = \bar{y} - \bar{x}$$

- estimated variance of $\hat{\mu}_{y\mathrm{D}}$:

$$\hat{V}(\hat{\mu}_{y\mathrm{D}}) = \left(\frac{N-n}{Nn}\right) \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}$$

  where

$$d_i = y_i - x_i$$

difference estimation

- Example 6.10: $N = 180$ items with a total book value of $\tau_x =$\$13320 are to be audited; there is a simple random sample of $n = 10$ audit values y_i; estimate μ_y and put a bound on the error of estimation
- see file `example6.10.R`

the design effect

- design effect: ratio of variance given some design to the variance given simple random sampling
- example: design effect of stratified random sampling (STSRS)

$$\mathrm{deff}(\bar{y}_{\mathrm{st}}) = \frac{V(\bar{y}_{\mathrm{st}})}{V(\bar{y})}$$

the two variance terms are

$$V(\bar{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$V(\bar{y}_{\mathrm{st}}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i-n_i}{N_i-1}\right) = \sum_{i=1}^{L}\frac{N_i^2}{N^2}\left(\frac{N_i-n_i}{N_i-1}\right)\frac{\sigma_i^2}{n_i}$$

the design effect

- example: design effect of STSRS

$$\mathrm{deff}(\bar{y}_{\mathrm{st}}) = \frac{V(\bar{y}_{\mathrm{st}})}{V(\bar{y})} = \frac{\displaystyle\sum_{i=1}^{L} \frac{N_i^2}{N^2}\left(\frac{N_i - n_i}{N_i - 1}\right)\frac{\sigma_i^2}{n_i}}{\displaystyle\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

$$= \sum_{i=1}^{L} \frac{N_i^2(N_i - n_i)(N-1)n}{N^2(N-n)(N_i-1)n_i}\frac{\sigma_i^2}{\sigma^2}$$

- for more details about $\mathrm{deff}(\bar{y}_{\mathrm{st}})$ see `SRSvSTSRS.pdf`
- Chapter 6 introduces an analogous concept called "relative efficiency"

relative efficiency

- let $E1$ and $E2$ be two unbiased estimators for parameter θ, based on the same sample size $n_1 = n_2 = n$
- the relative efficiency of $E1$ with respect to $E2$ is the ratio of their variances
$$\mathrm{RE}\left(\frac{E1}{E2}\right) = \frac{V(E2)}{V(E1)}$$
- in practice we use a similar ratio defined in terms of estimated variances
$$\widehat{\mathrm{RE}}\left(\frac{E1}{E2}\right) = \frac{\hat{V}(E2)}{\hat{V}(E1)}$$
- a similar concept often used is the "design effect":
$$\mathrm{deff}(\hat{\mu}_{\mathrm{E2}}) = V(\hat{\mu}_{\mathrm{E2}})/V(\bar{y})$$
where $\bar{y}$ is the simple random sampling estimator

relative efficiency

- the requirement that estimators be unbiased is, in practice, "nearly unbiased"
- the ratio estimator $\hat{\mu}_y$ is generally biased because $r = \bar{y}/\bar{x}$ is biased for $R = \mu_y/\mu_x$
  - if $y_i = \beta_1 x_i + u_i$, $E(u_i|x_i) = 0$, $\beta_1 \neq 0$, the bias is negligible
  - the relative bias is approximately

$$\frac{E(r) - R}{R} \approx \left(\frac{N-n}{Nn}\right)\left(\frac{s_x^2}{\bar{x}^2} - \hat{\rho}\frac{s_y s_x}{\bar{y}\bar{x}}\right) = \left(\frac{N-n}{Nn}\right)\left(\frac{s_x^2}{\bar{x}^2} - \frac{s_{xy}}{\bar{y}\bar{x}}\right)$$

- see files `example6.3bias.R`, `exercise6.5bias.R`
- judgment is still needed: see file `exercise6.7bias.R`

relative efficiency

- the regression estimator is biased in finite populations
  - bias is small if $y_i = \beta_0 + \beta_1 x_i + u_i$, $E(u_i|x_i) = 0$, $\beta_1 \neq 0$
  - bias is bigger if $y_i = \beta_0 + \beta_1 x_i + f(x) + u_i$, $E(u_i|x_i) = 0$, for some nonlinear function $f(x)$
- consider the simulation exercise in file `sim6.8.R`

relative efficiency

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_y}{\bar{y}}\right) = \frac{\hat{V}(\bar{y})}{\hat{V}(\hat{\mu}_y)} = \frac{s_y^2}{s_y^2 + r^2 s_x^2 - 2rs_{xy}} > 1\,,$$

$$\text{if } \hat{\rho} > \frac{1}{2}\frac{rs_x}{s_y} = \frac{1}{2}\frac{\mathrm{cv}(x)}{\mathrm{cv}(y)}$$

recall

$$\hat{\rho} = \frac{s_{xy}}{s_x s_y}$$

relative efficiency: simple random sampling, ratio and regression estimators

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\bar{y}}\right) = \frac{\hat{V}(\bar{y})}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2}{s_y^2(1-\hat{\rho}^2)} > 1\,, \text{if } \hat{\rho} \neq 0$$

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\hat{\mu}_y}\right) = \frac{\hat{V}(\hat{\mu}_y)}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2 + r(rs_x^2 - 2s_{xy})}{s_y^2(1-\hat{\rho}^2)} > 1\,, \text{if } (b-r)^2 > 0$$

relative efficiency: simple random sampling, ratio, regression and difference estimators

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{D}}}{\bar{y}}\right) = \frac{\hat{V}(\bar{y})}{\hat{V}(\hat{\mu}_{y\mathrm{D}})} = \frac{s_y^2}{s_y^2 + s_x^2 - 2s_{xy}} > 1\,, \text{if } \hat{\rho} > \frac{s_x}{2s_y}$$

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\hat{\mu}_{y\mathrm{D}}}\right) = \frac{\hat{V}(\hat{\mu}_{y\mathrm{D}})}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2 + s_x^2 - 2s_{xy}}{s_y^2(1-\hat{\rho}^2)} > 1\,, \text{if } \left(s_x - \frac{s_{xy}}{s_x}\right)^2 > 0$$

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{D}}}{\hat{\mu}_{y}}\right) = \frac{\hat{V}(\hat{\mu}_{y})}{\hat{V}(\hat{\mu}_{y\mathrm{D}})} = \frac{s_y^2 + r(rs_x^2 - 2s_{xy})}{s_y^2 + s_x^2 - 2s_{xy}} > 1\,,$$
$$\text{if } (1-r)\left[2s_{xy} - (1+r)s_x^2\right] > 0$$

relative efficiency: simple random sampling, ratio and regression estimators

- precise (for small n) values used for some solutions in the book

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\bar{y}}\right) = \frac{\hat{V}(\bar{y})}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2}{s_y^2(1-\hat{\rho}^2)\frac{n-1}{n-2}} > 1\,, \text{if } \hat{\rho} \neq 0$$

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\hat{\mu}_y}\right) = \frac{\hat{V}(\hat{\mu}_y)}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2 + r(rs_x^2 - 2s_{xy})}{s_y^2(1-\hat{\rho}^2)\frac{n-1}{n-2}} > 1\,, \text{if } (b-r)^2 > 0$$

$$\widehat{\mathrm{RE}}\left(\frac{\hat{\mu}_{y\mathrm{L}}}{\hat{\mu}_{y\mathrm{D}}}\right) = \frac{\hat{V}(\hat{\mu}_{y\mathrm{D}})}{\hat{V}(\hat{\mu}_{y\mathrm{L}})} = \frac{s_y^2 + s_x^2 - 2s_{xy}}{s_y^2(1-\hat{\rho}^2)\frac{n-1}{n-2}} > 1\,, \text{if } \left(s_x - \frac{s_{xy}}{s_x}\right)^2 > 0$$