9/27

- the population ratio $r_{xy} = \frac{\mu_y}{\mu_x} = \frac{\tau_y}{\tau_x}$: a natural estimator of r_{xy} is

If U, V are independent r.v.'s, $E(\frac{V}{U}) = E(V \cdot \frac{1}{U}) = E(V)E(\frac{1}{U}) \stackrel{?}{=} \frac{E(V)}{E(U)}$; $E(\frac{1}{U}) \stackrel{?}{=} \frac{1}{E(U)}$

a plug-in estimator → $R \equiv \frac{\overline{Y}}{\overline{X}} = \frac{T_y}{T_x}$. ← total estimators of τ_y & τ_x

Note 10 (Some notes about the mean and variance of R)

① bias = ? unbiased ? ② accuracy of estimator ③ estimated standard error ④ MSE = Var + bias² ⑤ asymptotic normality $N(\mu_R, \sigma_R^2)$

- To study the estimation properties of R, we wish to derive expressions for $E(R)$ and $Var(R)$.
- Since R is a *nonlinear* function of $\overline{X}$ and $\overline{Y}$, we cannot always do this in closed form. (μ_R, σ_R^2)

(**cf.** $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$;

$s_x^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n-1}\left[\left(\sum_{k=1}^{n} X_k^2\right) - n\,\overline{X}^2\right]$;

$s_{xy} = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y}) = \frac{1}{n-1}\left[\left(\sum_{k=1}^{n} X_kY_k\right) - n\,\overline{X}\,\overline{Y}\right]$) ← linear form

- **Q**: How can we derive approximate expressions for $E(R)$ and $Var(R)$?
- Consider the problem. we have a rough idea about the shape of F_U
  - $Z = g(U)$, where U is a random variable and g is a known function.
  - Suppose we know only the mean μ_U and variance σ_U^2 of U, but **not** the exact distribution F_U of U (i.e., do not know the cdf or pdf/pmf of U).
  - **Q**: Can we derive the exact distribution of Z? ← No in general (But, consider the indicator function case in Ex.5, LNp.21)

Then, cannot derive $E(Z)$, $Var(Z)$

- If not, can we "roughly" characterize the mean and variance of Z?

(**Note.** $E(Z) = E\,[\,g(U)\,] \neq g\,[\,E(U)\,]$ in general.) → E and g cannot exchange in general.

Theorem 14 (δ-method, propagation of error)

- univariate case $Z = g(U)$: (g known; $E(U-\mu_U)=0$, $Var(U-\mu_U)=Var(U)=\sigma_U^2$)

$$Z = g(U) \approx g(\mu_U) + (U-\mu_U)g'(\mu_U) \;(+\cdots)\quad \text{(by Taylor expansion)}$$

$$\Rightarrow E(Z) \approx g(\mu_U) \quad \text{← only need to know } \mu_U$$

$$Var(Z) \approx Var(U)[g'(\mu_U)]^2 = \sigma_U^2\,[g'(\mu_U)]^2$$

or

$$Z = g(U) \approx g(\mu_U) + (U-\mu_U)g'(\mu_U) + (1/2)(U-\mu_U)^2\,g''(\mu_U) \;(+\cdots)$$

$$\Rightarrow E(Z) \approx g(\mu_U) + (1/2)\sigma_U^2\,g''(\mu_U) \quad \text{← only need to know } \mu_U, \sigma_U^2$$

If $U = \overline{X}$ with repl. $\mu_{\overline{X}} = \mu$, $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, $n \to \infty$, → 0 → very small neighborhood

Note. How good these approximations are depends on whether g can be reasonably well approximated by the 1st- or 2nd-order polynomials in a neighborhood of μ_U and on the size of σ_U. ← Recall Chebychev's inequality

- case of 2 random variables U, V and $Z = g(U, V)$: Let $\boldsymbol{\mu} = (\mu_U, \mu_V)$.

$$Z = g(U,V) \approx g(\boldsymbol{\mu}) + (U-\mu_U)\frac{\partial g(\boldsymbol{\mu})}{\partial u} + (V-\mu_V)\frac{\partial g(\boldsymbol{\mu})}{\partial v} \;(+\cdots)$$

$$\Rightarrow E(Z) \approx g(\boldsymbol{\mu}) \quad \text{← only need to know } \mu_U, \mu_V$$

$$Var(Z) \approx \sigma_U^2\left[\frac{\partial g(\boldsymbol{\mu})}{\partial u}\right]^2 + \sigma_V^2\left[\frac{\partial g(\boldsymbol{\mu})}{\partial v}\right]^2 + 2\,\sigma_{UV}\left[\frac{\partial g(\boldsymbol{\mu})}{\partial u}\right]\left[\frac{\partial g(\boldsymbol{\mu})}{\partial v}\right]$$

only need to know $\mu_U, \mu_V, \sigma_U, \sigma_V, \sigma_{UV}$

or $\quad Z = g(U,V) \approx g(\boldsymbol{\mu}) + (U-\mu_U)\dfrac{\partial g(\boldsymbol{\mu})}{\partial u} + (V-\mu_V)\dfrac{\partial g(\boldsymbol{\mu})}{\partial v}$

$$+\frac{1}{2}(U-\mu_U)^2\frac{\partial^2 g(\boldsymbol{\mu})}{\partial u^2} + \frac{1}{2}(V-\mu_V)^2\frac{\partial^2 g(\boldsymbol{\mu})}{\partial v^2}$$

$$+(U-\mu_U)(V-\mu_V)\frac{\partial^2 g(\boldsymbol{\mu})}{\partial u\partial v} + \cdots$$

Q: What characteristics do we need to know about the distribution of (U, V) if g is approximated by a 3rd-order polynomial?

only need to know $\mu_U, \mu_V, \sigma_U, \sigma_V, \sigma_{UV}$

$$\Rightarrow \quad E(Z) \approx g(\boldsymbol{\mu}) + \frac{1}{2}\sigma_U^2\left[\frac{\partial^2 g(\boldsymbol{\mu})}{\partial u^2}\right] + \frac{1}{2}\sigma_V^2\left[\frac{\partial^2 g(\boldsymbol{\mu})}{\partial v^2}\right] + \sigma_{UV}\left[\frac{\partial^2 g(\boldsymbol{\mu})}{\partial u\partial v}\right]$$

- **Note.** A function g of k random variables can be worked out similarly.

Example 15 (Application of δ-method on the mean and variance of $g(U, V)=V/U$)

$U=\overline{X}$, $V=\overline{Y}$, $\sigma_{\overline{X}\,\overline{Y}} = Cov(\overline{X},\overline{Y}) = ?$

- Let $Z = g(U,V) = V/U$. Then, for $g(u,v) = v/u$,

$$\frac{\partial g}{\partial u} = \frac{-v}{u^2},\quad \frac{\partial g}{\partial v} = \frac{1}{u},\quad \frac{\partial^2 g}{\partial u^2} = \frac{2v}{u^3},\quad \frac{\partial^2 g}{\partial v^2} = 0,\quad \frac{\partial^2 g}{\partial u\partial v} = \frac{-1}{u^2}.$$

- By δ-method, after substituting (μ_U, μ_V) for (u, v), we have

$$E(Z) \approx \frac{\mu_V}{\mu_U} + \frac{1}{2}\sigma_U^2\frac{2\mu_V}{\mu_U^3} + \frac{1}{2}\sigma_V^2\,0 + \sigma_{UV}\frac{-1}{\mu_U^2} = \frac{\mu_V}{\mu_U} + \frac{1}{\mu_U^2}\left(\sigma_U^2\frac{\mu_V}{\mu_U} - \sigma_{UV}\right).$$

- Similarly, by δ-method,

$$Var(Z) \approx \sigma_V^2\frac{\mu_V^2}{\mu_U^4} + \sigma_V^2\frac{1}{\mu_U^2} + 2\sigma_{UV}\frac{-\mu_V}{\mu_U^2}\frac{1}{\mu_U} = \frac{1}{\mu_U^2}\left(\sigma_U^2\frac{\mu_V^2}{\mu_U^2} + \sigma_V^2 - 2\sigma_{UV}\frac{\mu_V}{\mu_U}\right).$$

Theorem 15 (Covariance of the two sample mean)

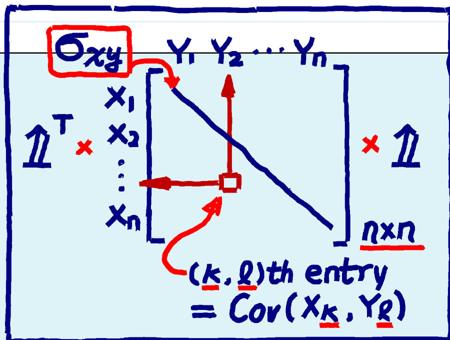
$\rho_{\overline{X}\,\overline{Y}} = Cor(\overline{X},\overline{Y}) = \sigma_{\overline{X}\,\overline{Y}}/(\sigma_{\overline{X}}\sigma_{\overline{Y}}) = \sigma_{xy}/(\sigma_x\sigma_y) = \rho_{xy} = Cor(X_k, Y_k)$

- Under s.r.s. with replacement (from F_0)

$$\sigma_{\overline{X}\,\overline{Y}} = Cov(\overline{X},\overline{Y}) = \frac{\sigma_{xy}}{n}.$$

- Under s.r.s. without replacement (cf.)

$$\sigma_{\overline{X}\,\overline{Y}} = Cov(\overline{X},\overline{Y}) = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right).$$



Proof: First, under s.r.s., no matter with or without replacement, we have

$$Cov(\overline{X},\overline{Y}) = E\left[(\overline{X}-\mu_x)(\overline{Y}-\mu_y)\right] = E\left[\left(\sum_{k=1}^{n}\frac{X_k-\mu_x}{n}\right)\left(\sum_{l=1}^{n}\frac{Y_l-\mu_y}{n}\right)\right]$$

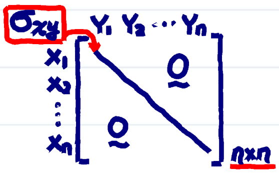
$(X_k, Y_k) \sim F_0$, $Cov(X_k, Y_k) = \sigma_{xy}$

$$= \frac{1}{n^2}\left(\sum_{k=1}^{n}\sum_{l=1}^{n} E\left[(X_k-\mu_x)(Y_l-\mu_y)\right]\right) \quad \leftarrow Cov(X_k, Y_l)$$

$$= \frac{1}{n^2}\sum_{k=1}^{n} E\left[(X_k-\mu_x)(Y_k-\mu_y)\right] + \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n} E\left[(X_k-\mu_x)(Y_l-\mu_y)\right]$$

($\leftarrow \sigma_{xy}$; $k=l$ (diagonal); $k \neq l$ (off-diagonal))

$$= \frac{\sigma_{xy}}{n} + \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n} Cov(X_k, Y_l) \qquad (*)$$



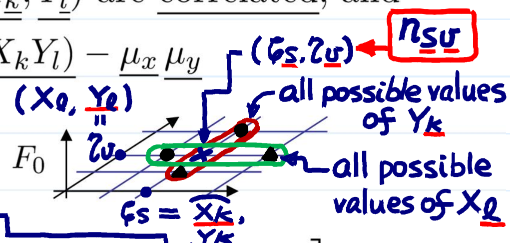
- Under s.r.s. with replacement, when $k \neq l$, X_k and Y_l are independent. Thus, for $k \neq l$, $Cov(X_k, Y_l) = 0$, and $(*)$ equals σ_{xy}/n.

- Under s.r.s. without replacement, for $k \neq l$, (X_k, Y_l) are correlated, and

$$Cov(X_k, Y_l) = E(X_kY_l) - E(X_k)\,E(Y_l) = E(X_kY_l) - \mu_x\,\mu_y$$

where

$$E(X_kY_l) = \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\,\eta_v\, P(X_k = \zeta_s, Y_l = \eta_v)$$



$$= \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left[\sum_{t=1}^{m_x}\sum_{u=1}^{m_y} P\left((X_k, Y_k) = (\zeta_s, \eta_u), (X_l, Y_l) = (\zeta_t, \eta_v)\right)\right]$$

$$= \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v\, P\Big((X_k, Y_k) = (\zeta_s, \eta_v),\ (X_l, Y_l) = (\zeta_s, \eta_v)\Big)$$

(same)

$$+ \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left[\sum_{\substack{t=1\\(\zeta_s,\eta_u)\neq(\zeta_t,\eta_v)}}^{m_x}\sum_{u=1}^{m_y} P\Big((X_k, Y_k) = (\zeta_s, \eta_u),\ (X_l, Y_l) = (\zeta_t, \eta_v)\Big)\right]$$

(different) — check LNp.41

$$= \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left[\frac{n_{sv}(n_{sv}-1)}{N(N-1)}\right] + \sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left[\sum_{\substack{t=1\\(\zeta_s,\eta_u)\neq(\zeta_t,\eta_v)}}^{m_x}\sum_{u=1}^{m_y} \left(\frac{n_{su}n_{tv}}{N(N-1)}\right)\right]$$

(pmf of F_0; if $s=t$, $u=v$ it is n_{sv}^2)

$$= \frac{N}{N-1}\sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left(\sum_{t=1}^{m_x}\sum_{u=1}^{m_y} \frac{n_{su}n_{tv}}{N\cdot N}\right) - \frac{1}{N-1}\sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\eta_v \left(\frac{n_{sv}}{N}\right)$$

$$= \frac{N}{N-1}\sum_{s=1}^{m_x}\sum_{v=1}^{m_y} \zeta_s\,\eta_v \left(\frac{n_{s\cdot}}{N}\right)\left(\frac{n_{\cdot v}}{N}\right) - \frac{1}{N-1}E(X_kY_k)$$

$$= \frac{N\,[E(X_k)\,E(Y_l)]}{N-1} - \frac{\sigma_{xy} + \mu_x\mu_y}{N-1} = -\frac{1}{N-1}\sigma_{xy} + \mu_x\mu_y.$$

$P(Y_l = \eta_v) = n_{\cdot v}/N = \left(\sum_{t=1}^{m_x}\frac{n_{tv}}{N}\right)\times\left(\sum_{u=1}^{m_y}\frac{n_{su}}{N}\right)$, $n_{s\cdot}/N = P(X_k = \zeta_s)$. $E(X_k) = \left(\sum_{s=1}^{m_x}\zeta_s\frac{n_{s\cdot}}{N}\right)\times\left(\sum_{v=1}^{m_y}\eta_v\frac{n_{\cdot v}}{N}\right) = E(Y_l)$ → $\mu_x\mu_y$. $\sigma_{xy} = E(X_kY_k) - \mu_x\mu_y$. marginal pmfs of X_k & Y_l (LNp.39).

Thus, $Cov(X_k, Y_l) = -\dfrac{\sigma_{xy}}{N-1}$ if $k \neq l$, and

$$(\ast) = \frac{\sigma_{xy}}{n} + \frac{1}{n^2}[n(n-1)]\left(-\frac{\sigma_{xy}}{N-1}\right) = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right).$$

Why negative? • $\sigma_{xy} > 0$: $X_k\uparrow$, $Y_k\uparrow$, $Y_l\downarrow$ • What if $\sigma_{xy} < 0$? LNp.46. (σ_{xy}, $Y_1, Y_2 \cdots Y_n$, $X_1, X_2, \ldots, X_n$, $n\times n$)

Theorem 16 (approximate mean of R)

- Under s.r.s. with replacement,

$$\mu_R \equiv E(R) \approx \frac{\mu_{\bar{Y}}}{\mu_{\bar{X}}} + \frac{1}{\mu_{\bar{X}}^2}\left(\sigma_{\bar{X}}^2\frac{\mu_{\bar{Y}}}{\mu_{\bar{X}}} - \sigma_{\bar{X}\bar{Y}}\right) = r_{xy} + \frac{1}{n}\times\frac{1}{\mu_x^2}(r_{xy}\,\sigma_x^2 - \sigma_{xy})$$

(cf. Ex.15; $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$)

- Under s.r.s. without replacement,

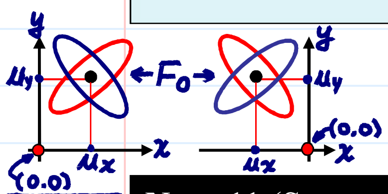
$$\mu_R \equiv E(R) \approx r_{xy} + \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\times\frac{1}{\mu_x^2}(r_{xy}\,\sigma_x^2 - \sigma_{xy})$$

$$= r_{xy} + \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\times\frac{1}{\mu_x^2}(r_{xy}\sigma_x^2 - \rho_{xy}\sigma_x\sigma_y).$$

(bias; $\to 0$ as $n\to\infty$)

$R = \bar{Y}/\bar{X} \xrightarrow{e} r_{xy} = \mu_y/\mu_x$

- with replacement,
  $\mu_{\bar{X}} = \mu_x$, $\mu_{\bar{Y}} = \mu_y$,
  $\sigma_{\bar{X}}^2 = \frac{\sigma_x^2}{n}$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_y^2}{n}$,
  $\sigma_{\bar{X}\bar{Y}} = \frac{\sigma_{xy}}{n}$.



Proof: From δ-method, Ex.15 (LNp.45), and Theorem 15 (LNp.46), the results follows.

- without replacement,
  $\mu_{\bar{X}} = \mu_x$, $\mu_{\bar{Y}} = \mu_y$,
  $\sigma_{\bar{X}}^2 = \frac{\sigma_x^2}{n}\left(1 - \frac{n-1}{N-1}\right)$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_y^2}{n}\left(1 - \frac{n-1}{N-1}\right)$,
  $\sigma_{\bar{X}\bar{Y}} = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right)$.

Note 11 (Some notes about the approximate mean of R)

F_0: red |bias|↓ F_0: blue |bias|↑

- strong correlation ρ_{xy} of the same sign as $r_{xy} = \mu_y/\mu_x$ decreases the bias
- the bias is large if $|\mu_x|$ is small ← $R = \bar{Y}/\bar{X}$, if $\bar{X}\approx 0$ usually, R is more variable
- the bias is of the order $1/n$, denoted by "bias $\sim O(n^{-1})$," and its contribution to the MSE is of the order $1/n^2$, i.e., bias$^2 \sim O(n^{-2})$

$MSE(R) = Var(R) + Bias^2(R)$

Theorem 17 (approximate variance of R)

- Under s.r.s. with replacement,

$$\sigma_R^2 \equiv Var(R) \approx \frac{1}{\mu_{\overline{X}}^2}\left(\sigma_{\overline{X}}^2 \frac{\mu_{\overline{Y}}^2}{\mu_{\overline{X}}^2} + \sigma_{\overline{Y}}^2 - 2\,\sigma_{\overline{X}\,\overline{Y}}\frac{\mu_{\overline{Y}}}{\mu_{\overline{X}}}\right) = \frac{1}{n}\times\frac{1}{\mu_x^2}\left(r_{xy}^2\sigma_x^2 + \sigma_y^2 - 2\,r_{xy}\,\sigma_{xy}\right).$$

- Under s.r.s. without replacement,

$$\sigma_R^2 \equiv Var(R) \approx \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\times\frac{1}{\mu_x^2}\left(r_{xy}^2\sigma_x^2 + \sigma_y^2 - 2r_{xy}\,\sigma_{xy}\right)$$

$$= \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\times\frac{1}{\mu_x^2}\left(r_{xy}^2\sigma_x^2 + \sigma_y^2 - 2\,r_{xy}\,\rho_{xy}\sigma_x\sigma_y\right).$$

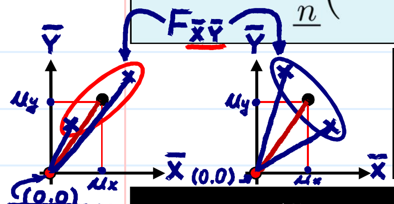
- with replacement, $\mu_{\overline{X}} = \mu_x$, $\mu_{\overline{Y}} = \mu_y$, $\sigma_{\overline{X}}^2 = \frac{\sigma_x^2}{n}$, $\sigma_{\overline{Y}}^2 = \frac{\sigma_y^2}{n}$, $\sigma_{\overline{X}\,\overline{Y}} = \frac{\sigma_{xy}}{n}$.

Ex.15 · $\rho_{xy}\sigma_x\sigma_y$ · cf.

Proof: From δ-method, Ex.15 (LNp.45), and Theorem 15 (LNp.46), the results follows.

- without replacement, $\mu_{\overline{X}} = \mu_x$, $\mu_{\overline{Y}} = \mu_y$, $\sigma_{\overline{X}}^2 = \frac{\sigma_x^2}{n}\left(1 - \frac{n-1}{N-1}\right)$, $\sigma_{\overline{Y}}^2 = \frac{\sigma_y^2}{n}\left(1 - \frac{n-1}{N-1}\right)$, $\sigma_{\overline{X}\,\overline{Y}} = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right)$.

R = Y̅/X̅ ↓ le r_{xy} = μ_y/μ_x



Note 12 (Some notes about the approximate variance of R)

- strong correlation ρ_{xy} of the same sign as $r_{xy} = \mu_y/\mu_x$ decreases the variance
- the variance is large if $|\mu_x|$ is small (**Note**. small values of $\overline{X}$ in the ratio $R = \overline{Y}/\overline{X}$ cause R to fluctuate wildly)
- the variance is of the order $1/n$, i.e., "Var $\sim O(n^{-1})$"
- the contributions of the Var and the bias^2 to the MSE (=Var+bias^2) are of the order $1/n$ and $1/n^2$, respectively $\Rightarrow$ for samples of large n, the bias is negligible compared to the standard error of the estimator

$F_{\overline{X}\overline{Y}}$: red, var ↓; $F_{\overline{X}\overline{Y}}$: blue, var ↑. very common property. Var rules. When $n\to\infty$, $1/n\to 0$ much faster than $1/\sqrt{n}\to 0$

Definition 14 (an intuitive estimators of the standard error of R) → to evaluate the accuracy of R

- Under s.r.s. with replacement, an estimator of the $\sigma_R^2 = Var(R)$ is

$$s_R^2 = \frac{1}{n}\times\frac{1}{\overline{X}^2}\left(R^2\,s_x^2 + s_y^2 - 2R\,s_{xy}\right).$$

$$\sigma_R^2 \approx \frac{1}{n}\times\frac{1}{\mu_x^2}\left(r_{xy}^2\,\sigma_x^2 + \sigma_y^2 - 2\,r_{xy}\,\sigma_{xy}\right)$$

LNp 42

The quantity s_R $\left(= \sqrt{s_R^2}\right)$ is an estimated standard error of R.

- Under s.r.s. without replacement, an estimator of the $\sigma_R^2 = Var(R)$ is

$$s_R^2 = \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\left(\frac{N-1}{N}\right)\times\frac{1}{\overline{X}^2}\left(R^2\,s_x^2 + s_y^2 - 2R\,s_{xy}\right)$$

$$= \frac{1}{n}\left(1 - \frac{n}{N}\right)\times\frac{1}{\overline{X}^2}\left(R^2 s_x^2 + s_y^2 - 2Rs_{xy}\right)$$

The quantity s_R $\left(= \sqrt{s_R^2}\right)$ is an estimated standard error of R.

S_R, S_R^2: same notation, different formula. $\times\left(1-\frac{n-1}{N-1}\right)$. cf. check estimators of σ^2 & σ_{xy} in Thm9 (LNp.24) & Def.13 (LNp.42). S_x^2 in Thm9 (LNp.24)

Theorem 18 (asymptotic sampling distribution of R) → ① probability of error ② C.I.

For samples of large size n, ($= \overline{Y}_n/\overline{X}_n$)

- truncating the Taylor series (in Thm 14, LNp.44) to the 1st order provides a good approximation, since the deviations $\overline{X}_n - \mu_x$ and $\overline{Y}_n - \mu_y$ are likely to be small (by LLN) ⇒ $\overline{X}_n \xrightarrow{P} \mu_x$, $\overline{Y}_n \xrightarrow{P} \mu_y$
- to this order of approximation, $R \approx \frac{\mu_y}{\mu_x} - \frac{\mu_y}{\mu_x^2}(\overline{X}_n - \mu_x) + \frac{1}{\mu_x}(\overline{Y}_n - \mu_y)$ (from Ex. 15, LNp.45), where $\overline{X}_n \stackrel{D}{\approx} N\left(\mu_x, \sigma_{\overline{X}_n}^2\right)$ and $\overline{Y}_n \stackrel{D}{\approx} N\left(\mu_y, \sigma_{\overline{Y}_n}^2\right)$ (by CLT).

higher-order terms: $(\overline{X}-\mu_x)^k \times (\overline{Y}-\mu_y)^\ell \approx 0$

- an argument based on the CLT can be used to show that R is approximately normally distributed, i.e., $R \overset{D}{\approx} N\left(\mu_R, \sigma_R^2\right)$, when sample size n is large.
- **Applications**
  - probability of estimation error $\in [a, b]$, e.g., $P\left(\left|\frac{R-r_{xy}}{s_R}\right| > \delta\right) \approx 2[1 - \Phi(\delta)]$
  - approximate $100(1-\alpha)\%$ confidence interval of r_{xy}: $R \pm z(\alpha/2)\, s_R$

Example 16 (estimate population ratio r_{xy})

- Suppose that 100 people who recently bought houses are surveyed, and

  y: mortgage payment $\qquad x$: gross income

  are observed. The $r_{xy} = \tau_y/\tau_x$ is the percentage of the total mortgage amount to the total gross income of all people who recently bought houses.
- Suppose that the population size N is missing, but it is known that $100 \ll N$.
- Suppose that $\overline{X} = 3100$, $s_x = 1200$, $\overline{Y} = 868$, $s_y = 250$, $\hat{\rho}_{xy} = 0.85$.

  We have $\quad R = 868/3100 = 0.28.$
- Neglecting the finite population correction, the estimated standard error of R is

$$s_R = \frac{1}{10} \times \frac{1}{3100}\sqrt{0.28^2(1200^2) + 250^2 - 2(0.28)(0.85)(250)(1200)} = 0.006.$$

Note that s_R is small because x and y are highly positively correlated, $r_{xy} > 0$, and $\overline{X}$ is large.