**Definition 15** (estimated standard error of ratio estimator; C.I. based on ratio estimator)

as $n \to \infty$, $R \overset{D}{\approx}$ normal $\Rightarrow \overline{Y}_R \overset{D}{\approx}$ normal$(\mu_y, \sigma^2_{\overline{Y}_R})$

- By Thm. 19 (LNp.55) and Def. 13 (LNp.42), the variance of $\overline{Y}_R$ can be estimated by

$$s^2_{\overline{Y}_R} = \frac{1}{n}\left(1 - \frac{n}{N}\right)(R^2 s_x^2 + s_y^2 - 2 R s_{xy}), \xleftrightarrow{\text{cf.}} S_R^2 \text{ (Def.14, LNp.50)}$$

The quantity $s_{\overline{Y}_R}\ \left(= \sqrt{s^2_{\overline{Y}_R}}\right)$ is an estimated standard error of $\overline{Y}_R$.

a known constant — C.I. of $r_{xy}$

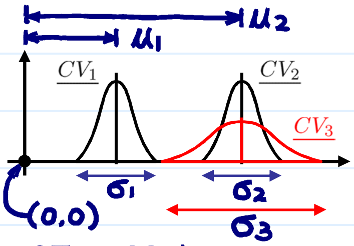
- Since $\overline{Y}_R = \mu_x R$, by Thm. 18 (LNp.51), an approximate $100(1-\alpha)\%$ confidence interval for $\mu_y$ $(= \mu_x r_{xy})$ is $\overline{Y}_R \pm z(\alpha/2)\, s_{\overline{Y}_R}$

(**cf.** the C.I. of $\mu_y$ based on $\overline{Y}$: $\overline{Y} \pm z(\alpha/2)\, s_{\overline{Y}}$ in App. 2, LNp.33.

When can $\sigma^2_{\overline{Y}_R} < \sigma^2_{\overline{Y}}$? (check Note13, LNp55) → **Q**: which C.I. of $\mu_y$ has a shorter width? under what condition?)

**Definition 16** (coefficient of variation)

For a distribution $F$ with mean $\mu \neq 0$ and variance $\sigma^2$, its **coefficient of variation** is defined as $CV = \sigma/\mu$, which gives $\sigma$ as a proportion of $\mu$. parameter when $F$ unknown




- $CV_1 > CV_2$, but $\sigma_1 = \sigma_2$
- $CV_3 > CV_2$, but $CV_3 \approx CV_1$

**Note 14** (Some notes about coefficient of variation)

- In some cases, $CV$ is more meaningful in explaining variation than $\sigma$, e.g., communication systems. → $\mu$: (average) sent signal, $\varepsilon$: random noise ($E(\varepsilon)=0$, $Var(\varepsilon)=\sigma^2$)
- $CV = \sigma/\mu$ is sometimes called **noise-to-signal ratio**.
- The value of $CV$ is free of unit. $\mu + \varepsilon$: received signal (mean $= \mu$, var $= \sigma^2$)



**Example 18** (Comparison of sample mean and ratio estimator, cont. Ex.17 in LNp.53)

- In the population of 393 hospitals, $x$: # of beds, $y$: # of discharges

known → $\mu_x = 274.8$, $\sigma_x = 213.2$, $\mu_y = 814.6$, $\sigma_y = 589.7$, $r_{xy} = 2.96$, $\rho_{xy} = 0.91$.

- For a sample of size $n = 64$, ← parameters (unknown) cf.

$\times$: unknown in sampling survey

– the standard error of the ratio estimator $\overline{Y}_R$ is (by Thm.19, LNp.55)

$$\sigma_{\overline{Y}_R} \approx \sqrt{\frac{1}{64}\left(1 - \frac{63}{392}\right)} \times \sqrt{(2.96^2)(213.2^2) + 589.7^2 - 2(2.96)(0.91)(213.2)(589.7)} = 30.0$$

cf.

– the standard error of the ordinary estimator $\overline{Y}$ is (by Thm.3, LNp.18)

When the parameters are unknown, can compare $S_{\overline{Y}_R}$ & $S_{\overline{Y}}$

$$\sigma_{\overline{Y}} = \sqrt{\frac{1}{64}\left(1 - \frac{63}{392}\right)} \times 589.7 = 67.5$$

The comparison of $\sigma_{\overline{Y}}$ to $\sigma_{\overline{Y}_R}$ is consistent with the substantial reduction in variability shown in the graph of Ex.17 (LNp.53).

$Var \sim O(n^{-1})$, $Bias^2 \sim O(n^{-2})$

cf. – the bias of the ratio estimator $\overline{Y}_R$ is (by Thm.19, LNp.54)

$\overline{Y}$ is unbiased

$$E(\overline{Y}_R) - \mu_y \approx \frac{1}{64}\left(1 - \frac{63}{392}\right) \times \frac{1}{274.8}[(2.96)(213.2^2) - (0.91)(213.2)(589.7)] = 1.0, \text{ cf.}$$

MSE($\overline{Y}_R$) < MSE($\overline{Y}$)

which is a slight and negligible bias compared to the variation reduction.

- An alternative interpretation of $\sigma^2_{\overline{Y}_R}/\sigma^2_{\overline{Y}}$. Neglecting finite population correction, an ordinary estimator $\overline{Y}_{n_1}$ from a sample of size $n_1$ will have about the same variance as a ratio estimator $\overline{Y}_{R,n_2}$ from a sample of size $n_2$ if

assume $n_1, n_2 \ll N$

Recall. For an unbiased estimator $\hat{\theta}$, $MSE(\hat{\theta}) = Var(\hat{\theta}) >$ C-R bound (TBp 300)

$(1/n_1) \times 589.7^2 \approx (1/n_2) \times [(2.96^2)(213.2^2) + 589.7^2 - 2(2.96)(0.91)(213.2)(589.7)]$

Thus, $n_2/n_1 \approx (30.0/67.5)^2 = \sigma^2_{\overline{Y}_{R,n}} / \sigma^2_{\overline{Y}_n} = 0.198$, i.e., we can obtain *same* precision from $\overline{Y}_R$ using *a sample about* 80% *smaller* than the sample of $\overline{Y}$. Note that this comparison neglects the bias of $\overline{Y}_R$, justifiable in this case.

- This is a case in which a biased estimator performs substantially better than an unbiased estimator.
- In this case, the biased estimator is better because the bias is quite small and the reduction in variance is quite large.

| | MSE = | Var + | Bias² |
|---|---|---|---|
| $\overline{Y}$ | large | large | =0 |
| $\overline{Y}_R$ | small | small | ≠0 |

**Definition 17** (ratio estimator of population total $\tau_y$)

Since $\tau_y = N\,\mu_y = N\,\mu_x r_{xy} = \tau_x\, r_{xy}$, an intuitive ratio estimator of $\tau_y$ is

known

$T_R = \tau_x (\overline{Y}/\overline{X}) = N\,\overline{Y}\,(\mu_x/\overline{X}) = N\,\overline{Y}_R.$ (cf.)

**Note 15** (Some notes about the ratio estimator of population total)

- Since $E(T_R) = N\,E(\overline{Y}_R)$ and $Var(T_R) = N^2\,Var(\overline{Y}_R)$, the approximate bias and variance of $T_R$ can be derived from Thm. 19 (LNp.54).
- The condition for $Var(T_R)$ to be smaller than $Var(T)$, where $T = N\,\overline{Y}$, is same as that given in Note 13 (LNp.55).
- An estimated standard error of $T_R$ is $s_{T_R} = N\,s_{\overline{Y}_R}$, and an approximate $100(1-\alpha)\%$ C.I. of $\tau_y$ is $T_R \pm z(\alpha/2)\,s_{T_R}$ (following from Def.15, LNp.56)

❖ **Reading**: textbook, 7.4



## • Stratified random sampling

information about $\mu_x \gtrless \overline{X}$

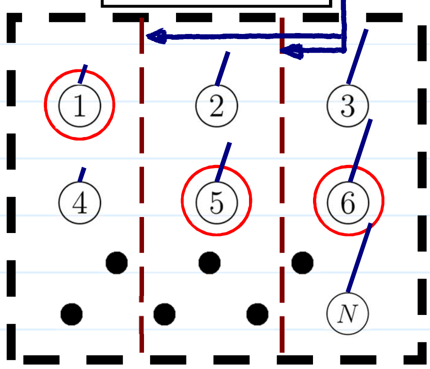
s.r.s. without repl.: $\binom{N}{n}$ samples

Samples with large estimation errors

- **Recall**. In the discussion of ratio estimator, extra information is used to adjust the sample mean of a **biased sample** and increase accuracy.
- **Q**: Is it possible to **exclude** some biased samples in s.r.s.? → extra information

not representitive

population




a stratum: a sub-population (subset) of the $N$ members in the population

the population is ***partitioned*** into $L$ strata

**Some notations for stratified random sampling**

- Let $\Omega = \{1, 2, \ldots, N\}$ be the population.
- Let $\mathbb{S}_l$, $l = 1, \ldots, L$, be a subset of $\Omega$, and $\mathbb{S}_1, \ldots, \mathbb{S}_L$ form a partition of $\Omega$, i.e.,

  $\mathbb{S}_l$'s are disjoint and $\mathbb{S}_1 \cup \cdots \cup \mathbb{S}_L = \Omega$.
- Each $\mathbb{S}_l$ is called a **stratum** of $\Omega$, and the number of strata is $L$.
- Denote the number of members in $\mathbb{S}_l$ by $N_l$ (subpopulation size), $l = 1, \ldots, L$. Then, (assume known)

  $N = N_1 + N_2 + \cdots + N_L.$
- Let $W_l = N_l/N$, $l = 1, \ldots, L$, be the fraction of the population in the $l$th stratum. (stratum fraction)
- Let $x_{i,l}$, $i = 1, \ldots, N_l$, $l = 1, \ldots, L$, denote the value associated with the $i$th member in the stratum $\mathbb{S}_l$.

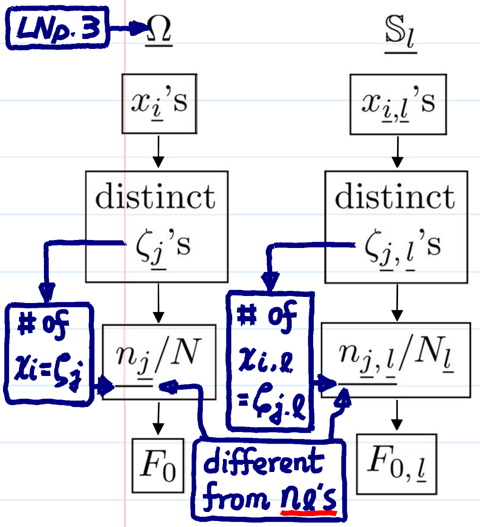
Z (r.v.)

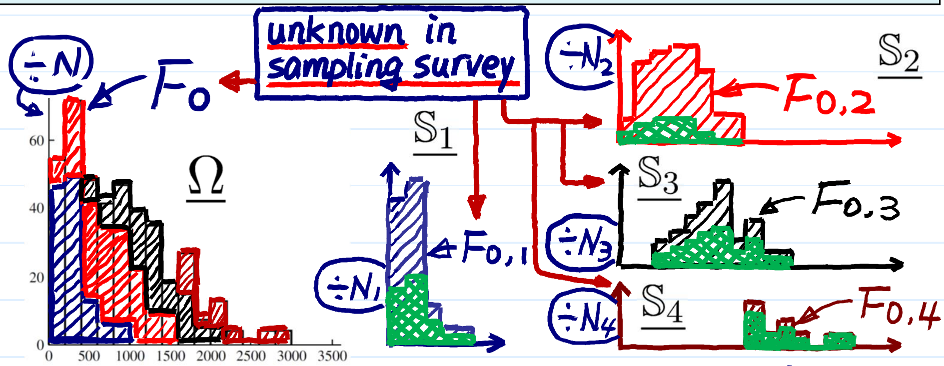
$W_1$ $W_2$ … $W_L$

1 2 … L (strata index)

a distribution

- Let $F_0$ denote the population distribution of the whole population $\Omega$.
- unknown — Let $F_{0,l}$, $l = 1, \ldots, L$, denote the population distribution of the subpopulation $\mathbb{S}_l$. Q₁: What are stratum fractions $W_\ell = N_\ell / N$? Q₂: What are sampling fractions $n_\ell / N_\ell$?




in producing representitive samples

Q: Is it better to have $F_{0,\ell} \approx F_0$, or to have $F_{0,\ell}$'s very different from $F_0$? (check LNp.63)

**Definition 18** (Some population parameters that are often of interest for $F_0$ and $F_{0,l}$)

- For $\Omega$, population mean, total, variance, and standard deviation of $F_0$, denoted by $\mu$, $\tau$, $\sigma^2$, $\sigma$, respectively, are defined as in Def. 3 (LNp.5). cf.
- For each $\mathbb{S}_l$, subpopulation mean, total, variance, and standard deviation of $F_{0,l}$, denoted by $\mu_l$, $\tau_l$, $\sigma_l^2$, $\sigma_l$, respectively, are similarly defined as above.



**Theorem 20** (some relations between the parameters of population and subpopulation)

- The two sampling schemes are equivalent: ← identical sampling probability
  - $\frac{1}{N}$ → [a]. Perform an s.r.s. from $\Omega$ to get one observation $X$ ← same distribution
  - $\frac{N_\ell}{N} \times \frac{1}{N_\ell} = \frac{1}{N}$ → [b]. (1) Randomly select a stratum $Z$, say $Z = l$, with probability proportional to the stratum size $N_l$; (2) Perform an s.r.s from $\mathbb{S}_l$ to get an $X$
- Thus, we have the distributions of $X$, $Z$, and $X|Z$ as follows:
  - $X \sim F_0$ (from [a]) ← marginal dist. of $X$ — $F_0 \xleftrightarrow{\text{cf.}} F_{0,\ell}$
  - joint dist. of (X, Z):
  - $Z \in \{1, \ldots, L\}$ and $P(Z = l) = W_l = N_l/N$, $l = 1, \ldots, L$ (from [b].(1)) ← marginal dist. of $Z$
  - $X|Z = l \sim F_{0,l}$ (from [b].(2)) ← conditional dist. of $X|Z$
- mean: $\mu = \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{N_l} x_{i,l} = \frac{1}{N} \sum_{l=1}^{L} N_l \mu_l = \sum_{l=1}^{L} W_l \mu_l$ (weighted average)

  Law of total expectation (TBp.149) → (**Recall.** $E(X) = E_Z\left[E_{X|Z}(X|Z)\right]$) ← $= \mu_\ell$ with probability $W_\ell$ (cf. stratum fraction in LNp.59)

  conditional mean $E(X|Z=\ell)$
- total: $\tau = N\mu = \sum_{l=1}^{L} \sum_{i=1}^{N_l} x_{i,l} = \sum_{l=1}^{L} N_l \mu_l = \sum_{l=1}^{L} \tau_l$
- variance: $\sigma^2 = \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{N_l} (x_{i,l} - \mu)^2$ ← $-\mu_\ell + \mu_\ell$ → $(x_{i,\ell} - \mu_\ell)^2 + (\mu_\ell - \mu)^2$ ~~+ cross term~~ $= 0$

  $= \frac{1}{N} \sum_{l=1}^{L} N_l \sigma_l^2 + \frac{1}{N} \sum_{l=1}^{L} N_l (\mu_l - \mu)^2 = \sum_{l=1}^{L} W_l \sigma_l^2 + \sum_{l=1}^{L} W_l (\mu_l - \mu)^2$ ← weights; conditional variance

  variance decomposition (TBp.151) → (**Recall.** $Var(X) = E_Z\left[Var_{X|Z}(X|Z)\right] + Var_Z\left[E_{X|Z}(X|Z)\right]$)

**Definition 19** (Stratified random sampling)

the s.r.s. discussed before can be regarded as a stratified random sampling with only one stratum.

In a stratified random sampling, to obtain a sample of size $n$, a simple random sampling (either with replacement or without replacement, but consistent in all strata) is taken independently within each stratum $\mathbb{S}_l$ to draw a subsample of size $n_l$, $l = 1, \ldots, L$, where $n = n_1 + n_2 + \cdots + n_L$. Results from the strata are combined to estimate the population parameters.

determined by investigators

cf. $n_i$, $n_{i,l}$ in LNp.60

all strata with or all strata without

Note. r.v.'s (data) appears after sampling

**Theorem 21** (Q: how many different possible samples? how many s.r.s are excluded?)

- Under with replacement, the number of all possible stratified random samples of size $n$ is

$$N_1^{n_1} \times N_2^{n_2} \times \cdots \times N_L^{n_L} < N^n,$$

how to prove "<"?

$\begin{cases} N_1 + \cdots + N_L = N \\ n_1 + \cdots + n_L = n \end{cases}$

where $N^n$ = number of all possible s.r.s. of size $n$ with replacement.

- Under without replacement, the number of all possible stratified random samples of size $n$ is

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times \cdots \times \binom{N_L}{n_L} < \binom{N}{n},$$

where $\binom{N}{n}$ = number of all possible s.r.s. of size $n$ without replacement.

sampling fraction in $\mathbb{S}_l$: $n_l/N_l$ (check graph in LNp.60)

Q: Are most excluded samples biased samples? Ans. Not necessary. Excluding itself is not enough to guarantee that will happen.

- Q: What is a good way of partitioning the population $\Omega$ into strata?
- Q: How to choose a good sampling schemne? $\Rightarrow$ How to allocate the sample size $n$ to each stratum, i.e., how to determine $n_1, \ldots, n_L$?

$F_0$, $F_{0,l}$'s are unknown



**Example 19** (Applications of stratified random sampling)

- In auditing financial transactions, the trnasactions may be grouped into strata on the basis of their nominal values, e.g., high-value, medium-value, and low-value strata.
- In human populations, geographical area often form natural strata.

**Note 16** (Advantages of stratified random sampling)

- It provides information about each subpopulation $\mathbb{S}_l$ in addition to the population $\Omega$ as a whole, e.g., in an industrial application,
  - population = all items produced by a manufacturing process;
  - subpopulations = items produced from different shifts or lots.
- It guarantees a prescribed number $n_l$ of observations from each $\mathbb{S}_l$.

not a feature in s.r.s. from $\Omega$

- Stratified sample mean can be considerably more precise ← i.e., smaller variance than the mean of a simple random sample (shown in later slides), expecially if the partition of the population into strata
  - is homogeneous within each stratum, and
  - has large variation between strata.

defined in LNp.65 $\bar{X}_S$

small $\sigma_l^2$'s

check the graph in LNp.60

Thm 20 (LNp.61)

**Recall.** $\sigma^2 = \sum_{l=1}^{L} W_l\, \sigma_l^2 + \sum_{l=1}^{L} W_l\, (\mu_l - \mu)^2$

weights (distribution)

$\sum_{l=1}^{L} W_l \mu_l$

population variation

within-stratum variation — small

between-strata variation — large