- In the following discussion of this topic, we consider the without-replacement case, but neglect the finite population correction. Actually, this is equivalent to the with-replacement case.

$n_l \ll N_l$

assume n_l's $\in \mathbb{R}$

$l \to \sigma_l$

The r.v. Z in LNp.59

cf.

Z_σ (r.v.) $\in \{\sigma_1, \ldots, \sigma_L\}$ with probabilities $W_1, \ldots, W_L$

$E(Z_\sigma) = \bar{\sigma}$.

$Z_\sigma^2 = Var(X|Z)$ in LNp.61

Theorem 24 (optimal allocation of the sample size n in a stratified random sampling)

Neglecting the finite population correction, the subsample sizes $n_1, \ldots, n_L$ that minimize $Var(\overline{X}_{\mathbb{S}})$ subject to the constraint $n_1 + n_2 + \cdots + n_L = n$ are

$$\underbrace{W_l\sigma_l \propto}\ n_l = \underbrace{n}_{\text{total sample size}} \times \frac{W_l\sigma_l}{\sum_{l'=1}^{L} W_{l'}\sigma_{l'}} = n \times \frac{W_l\sigma_l}{\bar{\sigma}}, \quad l = 1, 2, \ldots, L,$$

sum over l $\Rightarrow$ equal 1

where $\bar{\sigma} = \sum_{l'=1}^{L} W_{l'}\sigma_{l'}$ is a weighted average of $\sigma_1, \ldots, \sigma_L$.

minimize $Var(\overline{X}_{\mathbb{S}})$ subject to $(n_1 + \cdots + n_L) - n = 0$

Proof. Introduce a Lagrange multiplier λ, and minimize

$$\underbrace{L(n_1, \cdots, n_L, \lambda)}_{\text{target function}} = Var(\overline{X}_{\mathbb{S}}) + \lambda\left(\sum_{l'=1}^{L} n_{l'} - n\right) = \underbrace{\sum_{l'=1}^{L} \frac{W_{l'}^2\sigma_{l'}^2}{n_{l'}}}_{\text{Thm22, LNp.65}} + \lambda\left(\sum_{l'=1}^{L} n_{l'} - n\right).$$

Setting the partial derivatives equal to zero

$$0 = \frac{\partial L}{\partial n_l} = -\frac{W_l^2\sigma_l^2}{n_l^2} + \lambda, \quad l = 1, \cdots, L, \quad \text{and} \quad \frac{\partial L}{\partial \lambda} = \sum_{l'=1}^{L} n_{l'} - n = 0$$

we have

$$n_l = \frac{W_l\sigma_l}{\sqrt{\lambda}}, \quad l = 1, \cdots, L \quad \Rightarrow \quad n = \sum_{l'=1}^{L} n_{l'} = \frac{1}{\sqrt{\lambda}}\sum_{l'=1}^{L} W_{l'}\sigma_{l'}.$$

$n_l \propto W_l\sigma_l$

Thus,

$$\frac{1}{\sqrt{\lambda}} = \frac{n}{\sum_{l'=1}^{L} W_{l'}\sigma_{l'}} \quad \Rightarrow \quad n_l = n \times \frac{W_l\sigma_l}{\sum_{l'=1}^{L} W_{l'}\sigma_{l'}}.$$

proportion of the sample size n allocated to $\mathbb{S}_l$

Note 18 (Some notes about the optimal allocation scheme)

- This theorem shows that those strata with large $W_l\,\sigma_l$ should be sampled heavily. This makes sense intuitively because
  - W_l is large $\Rightarrow$ $\mathbb{S}_l$ contains a large fraction of Ω $\Rightarrow$ sample more
  - σ_l is large $\Rightarrow$ $x_{i,l}$'s in $\mathbb{S}_l$ are quite variable $\Rightarrow$ a relatively large n_l is required to obtain a good determination of μ_l ← $Var(\overline{X}_l) \propto \sigma_l^2/n_l$

subpopulation parameter

- This optimal allocation scheme depends on the within-stratum variances $\sigma_1^2, \ldots, \sigma_L^2$, which generally is unknown before sampling.
- If a survey measures several attributes, it is usually impossible to find an allocation optimal for all attributes. ← x_i's, y_i's, z_i's, $\cdots \Rightarrow \sigma_{l,x}$'s $\neq \sigma_{l,y}$'s $\neq \sigma_{l,z}$'s $\neq$

cannot find optimal allocation

Definition 23 (optimal stratified estimator)

- This optimal allocation scheme is called **Neyman allocation**.
- Denote the stratified estimator under this optimal allocation scheme by

Data from scheme 1 → $\sum_l W_l\overline{X}_l$ → $\overline{X}_{\mathbb{S},o}$.

$n_l = n \times \left(\frac{W_l\sigma_l}{\bar{\sigma}}\right)$

$n_l \ll N_l$

Theorem 25 (variance of the optimal stratified estimator)

Neglecting the finite population correction, and substituting the optimal values of n_l's in Thm. 24 (LNp.69) for the variance of the stratified estimator $\overline{X}_{\mathbb{S}}$ of μ presented in Thm. 22 (LNp.65) gives us

$[E(Z_\sigma)]^2$ (Z_σ: check LNp.69)

$$Var(\overline{X}_{\mathbb{S},o}) \approx \sum_{l=1}^{L} W_l^2\left(\frac{\sigma_l^2}{n_l}\right) = \sum_{l=1}^{L} \frac{W_l^{\cancel{2}}\sigma_l^{\cancel{2}}}{n(\cancel{W_l\sigma_l})/\bar{\sigma}} = \frac{\bar{\sigma}^2}{n} = \frac{1}{n}\left(\sum_{l=1}^{L} W_l\sigma_l\right)^2$$

Definition 24 (Proportional allocation and its stratified estimator)

- **Proportional allocation**. A simple and popular alternative method of allocation is to use the same sampling fraction in each stratum, i.e.,

optimal allocation $n_\ell \propto W_\ell \sigma_\ell$ cf. $n_\ell \propto W_\ell$

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_L}{N_L} = p \left(= \frac{n}{N} \right),$$

$n_\ell = pN_\ell,\ \ell = 1, \cdots, L$

$n = \sum_\ell n_\ell = p \sum_\ell N_\ell = pN$

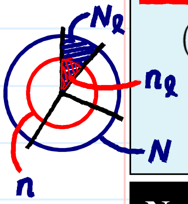
$\Rightarrow p = n/N$

$W_\ell = \frac{n_\ell}{n} = N_\ell/N$

which holds iff $n_l = n(N_l/N) = n\,W_l$ for $l = 1, \cdots, L$.

- Denote the stratified estimator under the proportional allocation scheme by

Data from scheme 2 $\rightarrow \sum_\ell W_\ell \bar{X}_\ell \rightarrow$ $\overline{X}_{\mathbb{S},p}$.



(**Note**. $\overline{X}_{\mathbb{S},o}$ and $\overline{X}_{\mathbb{S},p}$ are the estimator $\overline{X}_{\mathbb{S}}$ under two different allocation schemes (different n_l's, different possible samples (LNp.62), different joint distribution of data (LNp.64)). They are not different estimators.) Note 13, LNp.55

based on same data. e.g.

Note 19 (Some notes about the proportional allocation)

- Compared to the optimal allocation schemes, the proportional allocation neglects the difference in within-stratum variances σ_l's. same n_ℓ's
- If $\sigma_1 = \sigma_2 = \cdots = \sigma_L = \overline{\sigma}$, proportional allocation = optimal allocation, and $\overline{X}_{\mathbb{S},p}$ and $\overline{X}_{\mathbb{S},o}$ have same variance (accuracy). check LNp.65 $\bar{X}$ in s.r.s. cf.
- Under the proportional allocation,

$$\overline{X}_{\mathbb{S},p} = \sum_{l=1}^{L} W_l \overline{X}_l = \sum_{l=1}^{L} \left(\frac{n_l}{n}\right)\left(\frac{1}{n_l}\sum_{k=1}^{n_l} X_{k,l}\right) = \frac{1}{n}\sum_{l=1}^{L}\sum_{k=1}^{n_l} X_{k,l}$$

which is the unweighted mean of the sampled data. equal weights

Example 21 (optimal and proportional allocations, cont. Ex.20 in LNp.68)

Consider the population of 393 hospitals. The sampling fractions of the 4 strata are

Stratum	A	B	C	D	
optimal allocation	0.106	0.210	0.250	0.434	$\Leftarrow W_l\sigma_l/(\sum_l W_l\sigma_l)$
	$\wedge$			$\vee$	
proportional allocation	0.249	0.249	0.249	0.252	$\Leftarrow W_l$

Theorem 26 (variance of the stratified estimator under proportional allocation)

Ignoring the finite population correction, and substituting the proportional allocation, $n_l = nW_l$, for the variance of the stratified estimator $\overline{X}_{\mathbb{S}}$ of μ presented in Thm. 22 (LNp.65) gives us

$E(Z_\sigma^2)$

$$Var(\overline{X}_{\mathbb{S},p}) \approx \sum_{l=1}^{L} W_l^2 \left(\frac{\sigma_l^2}{n_l}\right) = \sum_{l=1}^{L} \frac{W_l^{\cancel{2}}\sigma_l^2}{n\cancel{W_l}} = \frac{1}{n}\left(\sum_{l=1}^{L} W_l\,\sigma_l^2\right).$$

$n_\ell \ll N_\ell$

Theorem 27 (variance difference between the optimal and proportional allocations)

Ignoring the finite population correction, from Thm.25 (LNp.70) and Thm.26,

$$Var(\overline{X}_{\mathbb{S},p}) - Var(\overline{X}_{\mathbb{S},o}) \approx \frac{1}{n}\left(\sum_{l=1}^{L} W_l\sigma_l^2\right) - \frac{1}{n}\,\overline{\sigma}^2 = \frac{1}{n}\sum_{l=1}^{L} W_l(\sigma_l - \overline{\sigma})^2 \geq 0.$$

more accurate $E(Z_\sigma^2)$ − $[E(Z_\sigma)]^2$ $Var(Z_\sigma)$

Note that

- If $\sigma_1 = \sigma_2 = \cdots = \sigma_L$, then $Var(\overline{X}_{\mathbb{S},p}) = Var(\overline{X}_{\mathbb{S},o})$. the impact of total sample size
- The more variable these σ_l's are, the better it is to use (if feasible) the optimal allocation.

σ_ℓ's are more distinct, not they are large

$W_1, \cdots, W_L$
Z_σ (r.v.) $\in \{\sigma_1, \cdots, \sigma_L\}$
$E(Z_\sigma) = \bar{\sigma}$, $E(Z_\sigma^2) = ?$, $Var(Z_\sigma) = ?$

Example 22 (comparison of optimal and proportional allocations, cont. Ex.20, LNp.67)

- Consider the population of 393 hospitals and the 4 strata.
- Under the sampling fractions given in Ex.21 (LNp.72),

$$\frac{Var(\overline{X}_{\mathbb{S},p})}{Var(\overline{X}_{\mathbb{S},o})} = 1 + \frac{\overbrace{\sum_l W_l(\sigma_l - \bar{\sigma})^2}^{Var(Z_\sigma)}}{\underbrace{\bar{\sigma}^2}_{[E(Z_\sigma)]^2}} = 1 + 0.218.$$

$Var(\overline{X}_{S,o}) + [Var(\overline{X}_{S,p}) - Var(\overline{X}_{S,o})]$

CV^2 of Z_σ (check Def.16, LNp.56)

- Under proportional allocation, the variance of the stratified estimator $\overline{X}_{\mathbb{S}}$ of μ is about 20% larger than it is under the optimal allocation.

- **Q**: **When** can a stratified random sample based on proportional allocation perform better than a simple random sample? Note that
  - under proportional allocation, $\overline{X}_{\mathbb{S},p}$ and $\overline{X}$, where $\overline{X}$ is the sample mean of the data from a s.r.s., have the same functional form, ← Note19 (LNp.71)
  - a s.r.s. has more possible samples than a stratified random sample (check Thm. 21, LNp.62).

But, not every stratified random sampling can successfully exclude biased samples

s.r.s., proportional allocation → same estimator $\overline{X}$ $(= \overline{X}_{S,p})$ → μ

- Recall.
  - Under a s.r.s. without replacement, neglecting the finite population correction, we have $Var(\overline{X}) \approx \sigma^2/n$ (check Thm.3, LNp.18). $n << N$
  - From Thm.20 (LNp.61),

$$\sigma^2 = \underbrace{\sum_{l=1}^{L} W_l\sigma_l^2}_{E[Var(X|Z)]} + \underbrace{\sum_{l=1}^{L} W_l(\mu_l - \mu)^2}_{Var[E(X|Z)]}.$$

$n \times Var(\overline{X}_{S,p})$, Z_σ^2, Z_μ

Theorem 28 (variance difference between s.r.s. and proportional allocation)

Ignoring the finite population correction, from Thm.26 (LNp.71), we have

$$Var(\overline{X}) - Var(\overline{X}_{\mathbb{S},p}) \approx \frac{\sigma^2}{n} - \frac{1}{n}\underbrace{\left(\sum_{l=1}^{L} W_l\sigma_l^2\right)}_{E[Var(X|Z)]} = \frac{1}{n}\underbrace{\left[\sum_{l=1}^{L} W_l(\mu_l - \mu)^2\right]}_{Var[E(X|Z)]} \geq 0.$$

$n_l << N_l$, $n << N$

> 0 if μ_l's not equal

Note 20 (Some notes about s.r.s., proportional allocation, and optimal allocation)

- Stratified random sampling with proportional allocation is better than s.r.s., which is a result of excluding some unwanted simple random samples.
- Comparing the equations for the variances under s.r.s., proportional allocation, and optimal allocation, we see that

$$\begin{aligned} Var(\overline{X}) \\ \geq Var(\overline{X}_{\mathbb{S},p}) \\ \geq Var(\overline{X}_{\mathbb{S},o}) \end{aligned}$$

every terms ≥ 0

$$\frac{\sigma^2}{n} \overset{(Thm26)}{=} \frac{1}{n}\sum_{l=1}^{L} W_l\sigma_l^2 + \frac{1}{n}\sum_{l=1}^{L} W_l(\mu_l - \mu)^2$$

$$= \frac{\bar{\sigma}^2}{n} + \frac{1}{n}\sum_{l=1}^{L} W_l(\sigma_l - \bar{\sigma})^2 + \frac{1}{n}\sum_{l=1}^{L} W_l(\mu_l - \mu)^2$$

$\boxed{\overline{X}}$ $\boxed{\overline{X}_{\mathbb{S},p}}$ $\boxed{\overline{X}_{\mathbb{S},o}}$ $\boxed{\overline{X}_{\mathbb{S},p} \to \overline{X}_{\mathbb{S},o}}$ $\boxed{\overline{X} \to \overline{X}_{\mathbb{S},p}}$

$E(Z_\sigma^2)$, $[E(Z_\sigma)]^2$ (Thm25), $Var(Z_\sigma)$ (Thm27), $E(Z_\mu)$, $Var(Z_\mu)$ (Thm28)

$\varrho(Z) \to \sigma_\varrho(Z_\sigma)$; Z_σ (r.v.) $\in \{\sigma_1, \cdots, \sigma_L\}$ with $W_1, \cdots, W_L$

$\varrho(Z) \to \mu_\varrho(Z_\mu)$; Z_μ (r.v.) $\in \{\mu_1, \cdots, \mu_L\}$ with $W_1, \cdots, W_L$

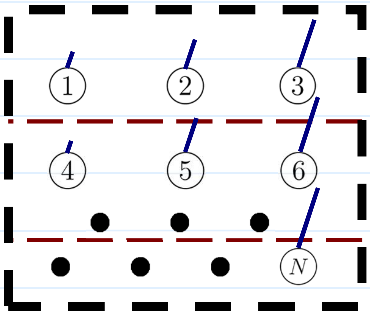
depends on the construction of strata

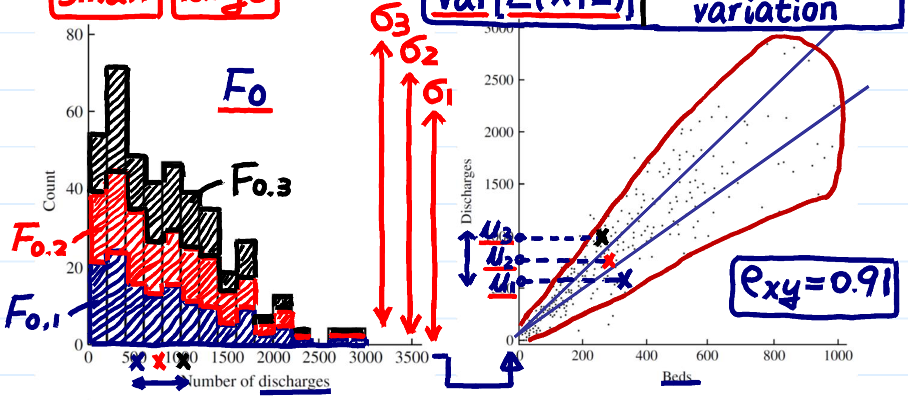
  - $\boxed{\overline{X} \to \overline{X}_{\mathbb{S},p}}$: $\overline{X}_{\mathbb{S},p}$ much better than $\overline{X}$ if μ_l's are quite variable
  - $\boxed{\overline{X}_{\mathbb{S},p} \to \overline{X}_{\mathbb{S},o}}$: $\overline{X}_{\mathbb{S},o}$ much better $\overline{X}_{\mathbb{S},p}$ if σ_l's are quite variable
- The gain from $\overline{X} \to \overline{X}_{\mathbb{S},p}$ is often greater than the gain from $\overline{X}_{\mathbb{S},p} \to \overline{X}_{\mathbb{S},o}$.

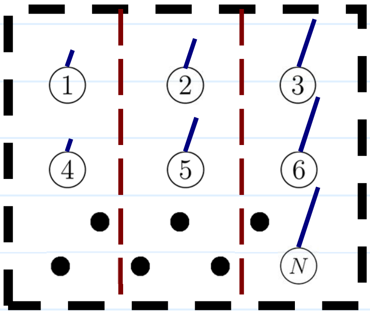
- **Q**: Which one is a better way to form strata (i.e., to partition the population)?

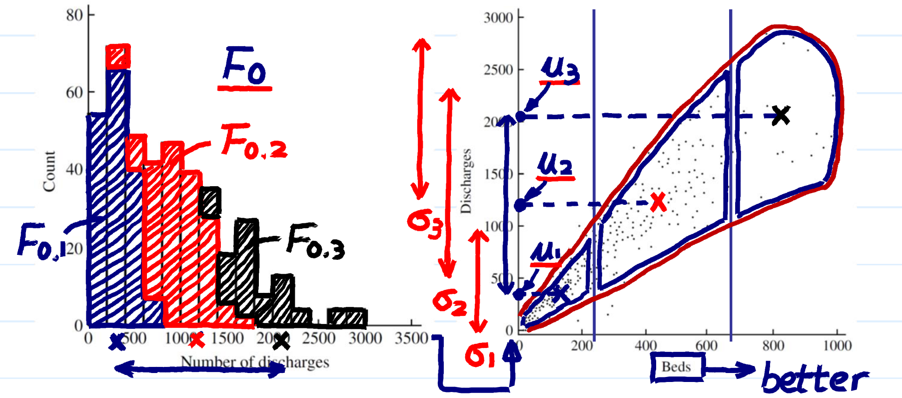
$$\sigma^2 = \underbrace{\sum_{l=1}^{L} W_l \sigma_l^2}_{\substack{\text{within-stratum variation} \\ E[Var(X|Z)] \\ \text{small}}} + \underbrace{\sum_{l=1}^{L} W_l (\mu_l - \mu)^2}_{\substack{\text{between-strata variation} \\ Var[E(X|Z)] \\ \text{large}}}$$

$\boxed{\overline{X} \to \overline{X}_{\mathbb{S},p}}$









❖ **Reading**: textbook, 7.5

❖ **Further reading**: textbook, 7.6 (**systematic sampling**, **cluster sampling**, **practical difficult**)