Comparing two samples (Chapter 11)

• Comparing two independent samples ← c.f. → paired samples (future lecture)

• Problem formulation and statistical modeling

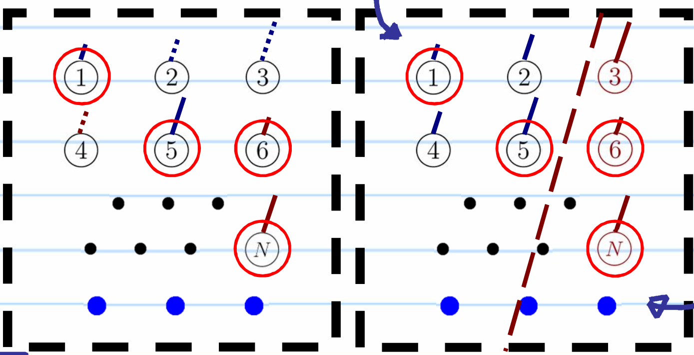
Data

U	V
1	X_1
⋮	⋮
1	X_n
2	Y_1
⋮	⋮
2	Y_m

subpopulation 1 cf. ⟷ stratum

population

population



Subpopulation 2

observed data from s.r.s (random variables)
$\{X_1, \ldots, X_n\}$
$\{Y_1, \ldots, Y_m\}$ ← Sample size

- X_i's, Y_j's are continuous quantities of same characteristic
- $X-Y$ is meaningful → the comparison of their means is meaningful

s.r.s., $N \to \infty$: without replacement $\approx$ with replacement ($\Rightarrow$ i.i.d.)

For example, in medical study,
- X_i's: treatment
- Y_j's: control

For example, in human population,
- X_i's: heights of males
- Y_j's: heights of females

Why?

- $X_1, \ldots, X_n \sim$ i.i.d. with a common continuous distribution F
- $Y_1, \ldots, Y_m \sim$ i.i.d. from a common continuous distribution G
- $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$ are independent

- Let random variables $Z_1, \ldots, Z_n$ and $Z_{n+1}, \ldots, Z_{n+m}$ represent the variablility of the $n+m$ members sampled from the population.

∵ s.r.s. $N \to \infty$

- Assume $Z_1, \ldots, Z_{n+m}$ are i.i.d. from a population distribution H.

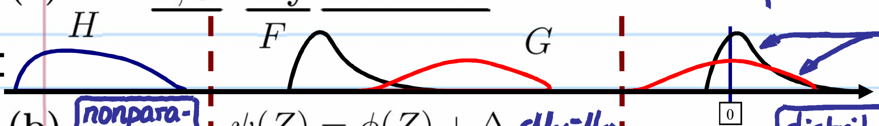
use e.g.'s in LN p.1 to understand

- Let F and G be the distributions of $X = \phi(Z)$ and $Y = \psi(Z)$, respectively. (cf.)
- The transformations ϕ and ψ might contain random components, e.g., $\phi(Z) = \phi^*(Z) + \delta$, where ϕ^*: a fixed function and δ: a random variable.
- Let μ_X and μ_Y be the means of F and G, respectively.

If $\underset{\uparrow Y}{\psi(Z)} = \underset{\uparrow X}{\phi(Z)} + \Delta$, F, G: cdf, $G(y) = P(Y \le y) = P(X + \Delta \le y) = P(X \le y - \Delta) = F(y - \Delta)$

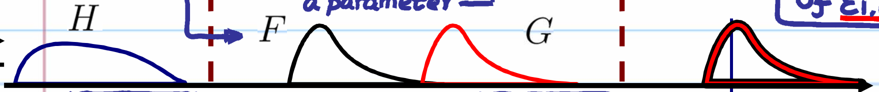
add restriction

(a) F, G: **any** continuous distribution ← nonparametric



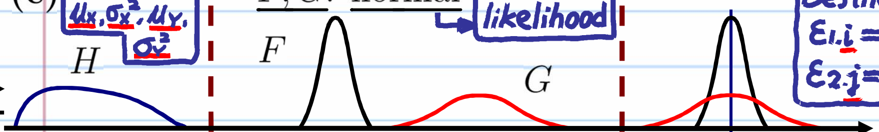
distributions of $\varepsilon_{1,i}$ & $\varepsilon_{2,j}$

(b) nonparametric $\psi(Z) = \phi(Z) + \underset{\text{a parameter}}{\Delta}$ ⟵ $\mu_Y - \mu_X$

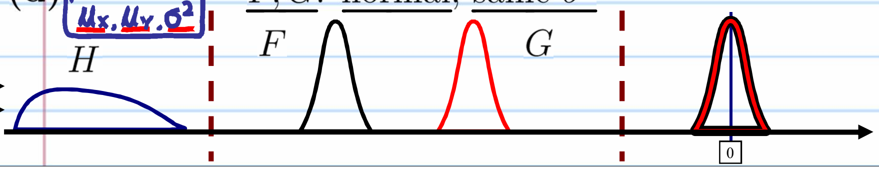


(c) parameters $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2$ ← F, G: normal → can use likelihood

Define $\varepsilon_{1,i} = X_i - \mu_X$, $\varepsilon_{2,j} = Y_j - \mu_Y$



(d) parameters μ_X, μ_Y, σ^2 ← F, G: normal, same σ^2



- Let $X_i = \phi(Z_i)$, $i = 1, \ldots, n$, $\Rightarrow X_i \sim F$ (r.v.) (cf.) (r.v.)
- Let $Y_j = \psi(Z_{n+j})$, $j = 1, \ldots, m$, $\Rightarrow Y_j \sim G$
- $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are independent

- $X_i = \mu_X + \epsilon_{1,i}$, $i = 1, \ldots, n$,
- $Y_j = \mu_Y + \epsilon_{2,j}$, $j = 1, \ldots, m$,
- $E(\epsilon) = 0$ and ϵ's are independent