

STRATIFIED RANDOM SAMPLING (STRA)

PENGERTIAN

- Populasi sebanyak N unit dikelompokkan menjadi L subpopulasi, masing-masing subpopulasi terdiri dari N_1, N_2, \dots, N_L unit.
- Subpopulasi yang terbentuk tidak boleh saling tumpang tindih (overlapping)
- Jumlah unit dari semua subpopulasi sama dengan jumlah populasi, sehingga:

$$N_1 + N_2 + \cdots + N_L = N$$

- Subpopulasi ini disebut strata.
- Penarikan sampel dilakukan untuk setiap strata dan bersifat independent antara strata satu dengan strata lainnya.

$$n_1 + n_2 + \cdots + n_L = n$$

- Jika penarikan sampel di setiap strata dilakukan secara SRS, maka prosedur itu disebut stratified random sampling.

KEUNTUNGAN

- Meningkatkan efisiensi desain/presisi estimasi karakteristik populasi.

Prinsip:

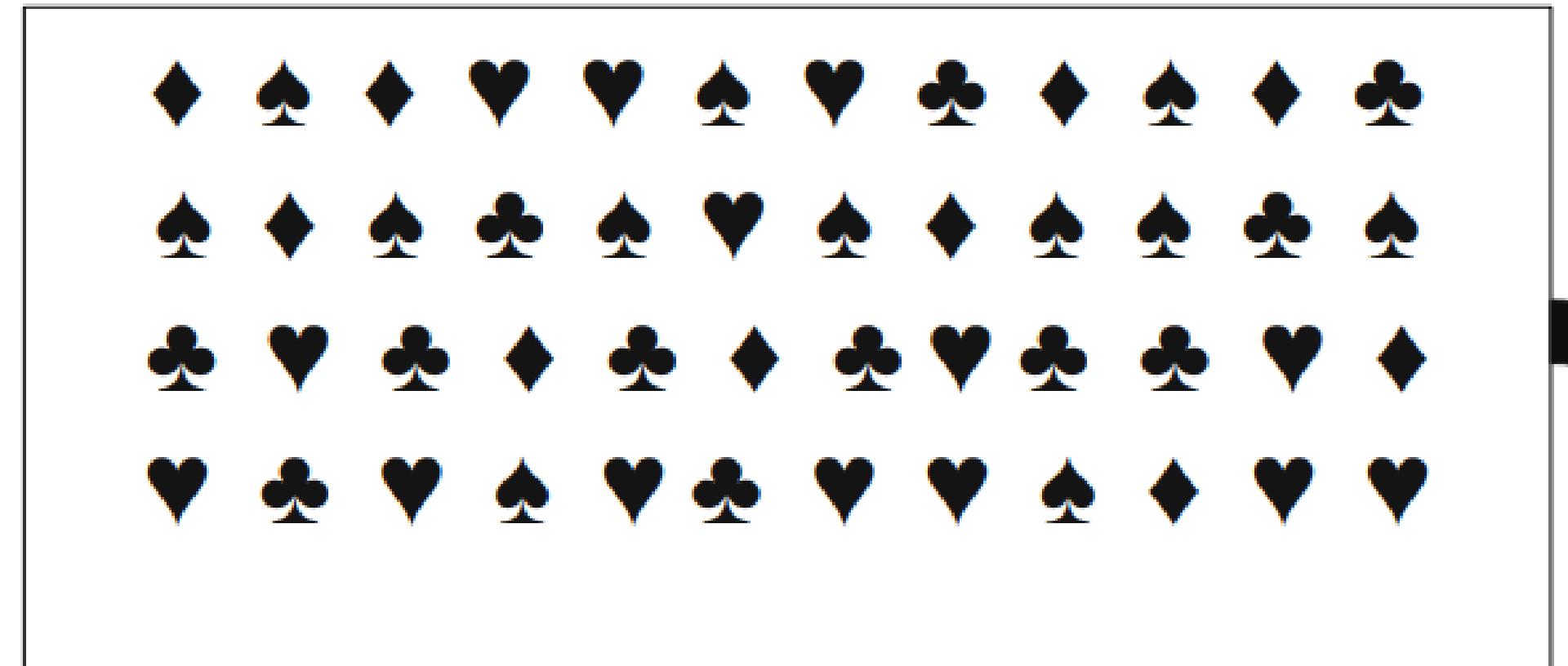
1. Unit/element yang karakteristiknya hampir sama dikelompokkan dalam satu strata
Unit-unit dalam strata (within stratum) --> homogen

2. Perbedaan rata-rata karakteristik antar strata dibuat sebesar mungkin
Unit-unit antar strata (between stratum) --> heterogen

- Masing-masing strata bisa dianggap sebagai populasi tersendiri sehingga bisa diterapkan desain sampling yang berbeda.
- Estimasi bisa dilakukan untuk penyajian sampai level strata
- Untuk kemudahan administratif

SKEMA PEMBENTUKAN STRATA

POPULASI



STRATIFIKASI
POPULASI



ESTIMASI RATA-RATA

Estimasi	Domain	
	Strata ke-h	Populasi
Rata-rata	$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$	$\bar{y}_{st} = \sum_{h=1}^L \frac{N_h}{N} \cdot \bar{y}_h = \sum_{h=1}^L W_h \cdot \bar{y}_h$
Varians rata-rata	$v(\bar{y}_h) = (1 - f_h) \frac{s_h^2}{n_h}$	$v(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \cdot v(\bar{y}_h)$
Standar Error	$se(\bar{y}_h) = \sqrt{v(\bar{y}_h)}$	$se(\bar{y}_{st}) = \sqrt{v(\bar{y}_{st})}$
RSE	$RSE(\bar{y}_h) = \frac{se(\bar{y}_h)}{\bar{y}_h} \times 100\%$	$RSE(\bar{y}_{st}) = \frac{se(\bar{y}_{st})}{\bar{y}_{st}} \times 100\%$
(1 - α)% Confidence interval	$\left[\bar{y}_h - Z_{\frac{\alpha}{2}} \cdot se(\bar{y}_h); \bar{y}_h + Z_{\frac{\alpha}{2}} \cdot se(\bar{y}_h) \right]$	$\left[\bar{y}_{st} - Z_{\frac{\alpha}{2}} \cdot se(\bar{y}_{st}); \bar{y}_{st} + Z_{\frac{\alpha}{2}} \cdot se(\bar{y}_{st}) \right]$

ESTIMASI TOTAL

Estimasi	Domain	
	Strata ke-h	Populasi
Total	$\hat{Y}_h = N_h \cdot \bar{y}_h$	$\hat{Y}_{st} = N \cdot \bar{y}_{st} = \sum_{h=1}^L \hat{Y}_h$
Varians Total	$v(\hat{Y}_h) = N_h^2 \cdot v(\bar{y}_h)$	$v(\hat{Y}_{st}) = N^2 \cdot v(\bar{y}_{st}) = \sum_{h=1}^L v(\hat{Y}_h)$
Standar Error	$se(\hat{Y}_h) = \sqrt{v(\hat{Y}_h)}$	$se(\hat{Y}_{st}) = \sqrt{v(\hat{Y}_{st})}$
RSE	$RSE(\hat{Y}_h) = \frac{se(\hat{Y}_h)}{\hat{Y}_h} \times 100\%$	$RSE(\hat{Y}_{st}) = \frac{se(\hat{Y}_{st})}{\hat{Y}_{st}} \times 100\%$
(1 - α)% Confidence interval	$\left[\hat{Y}_h - Z_{\frac{\alpha}{2}} \cdot se(\hat{Y}_h) ; \hat{Y}_h + Z_{\frac{\alpha}{2}} \cdot se(\hat{Y}_h) \right]$	$\left[\hat{Y}_{st} - Z_{\frac{\alpha}{2}} \cdot se(\hat{Y}_{st}) ; \hat{Y}_{st} + Z_{\frac{\alpha}{2}} \cdot se(\hat{Y}_{st}) \right]$

ESTIMASI PROPORSI

Estimasi	Domain	
	Strata ke-h	Populasi
Proporsi	$p_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} = \frac{a_h}{n_h}$	$p_{st} = \sum_{h=1}^L \frac{N_h}{N} \cdot p_h = \sum_{h=1}^L W_h \cdot p_h$
Varians proporsi	$v(p_h) = (1 - f_h) \frac{p_h q_h}{n_h - 1}$	$v(p_{st}) = \sum_{h=1}^L W_h^2 \cdot v(p_h)$
Standar Error	$se(p_h) = \sqrt{v(p_h)}$	$se(p_{st}) = \sqrt{v(p_{st})}$
RSE	$RSE(p_h) = \frac{se(p_h)}{p_h} \times 100\%$	$RSE(p_{st}) = \frac{se(p_{st})}{p_{st}} \times 100\%$
($1 - \alpha$)% <i>Confidence interval</i>	$\left[p_h - Z_{\frac{\alpha}{2}} \cdot se(p_h) ; p_h + Z_{\frac{\alpha}{2}} \cdot se(p_h) \right]$	$\left[p_{st} - Z_{\frac{\alpha}{2}} \cdot se(p_{st}) ; p_{st} + Z_{\frac{\alpha}{2}} \cdot se(p_{st}) \right]$

ALOKASI SAMPEL

- ❖ Alokasi sembarang --> jarang digunakan
- ❖ Alokasi sama (*equal*)
- ❖ Alokasi sebanding (*proportional*)
- ❖ Alokasi Neyman
- ❖ Alokasi optimum

ALOKASI SAMA (EQUAL)

- ❖ Alokasi ini sering digunakan jika varians strata (S_h^2) hampir sama.
- ❖ Jumlah sampel untuk setiap strata sama.
- ❖ Ukuran sampel untuk strata ke-h

$$n_h = \frac{n}{L}$$

n : jumlah sampel

n_h : jumlah sampel di strata ke-h

L : jumlah strata

- ❖ Ukuran sampel keseluruhan:

$$n = \frac{L \sum_{h=1}^L N_h^2 \cdot S_h^2}{N^2 D^2 + \sum_{h=1}^L N_h \cdot S_h^2}$$

Keterangan:

$$D = \frac{d}{Z_{\alpha/2}}$$

y_{hi} : nilai karakteristik unit ke-i strata ke-h

N : jumlah populasi

N_h : jumlah populasi di strata ke-h

$$\sum_{h=1}^L N_h = N$$

W_h : penimbang strata ke-h (*stratum weight*)

$$W_h = \frac{N_h}{N}$$

n : jumlah sampel

n_h : jumlah sampel di strata ke-h

$$\sum_{h=1}^L n_h = n$$

f_h : fraksi sampling strata ke-h

$$f_h = \frac{n_h}{N_h}$$

CONTOH

- ❖ Suatu survei dilakukan untuk mengetahui karakteristik wanita usia subur (WUS) di suatu kecamatan dengan menggunakan desa sebagai strata. Dari survei terdahulu, diperoleh rata-rata WUS beserta standar deviasinya.:

No	Strata	Populasi rumah tangga (N_h)	Rata-rata jumlah WUS per rumah tangga (\bar{y}_h)	Standar deviasi (s_h)
1	Desa A	1200	1.5	0.5
2	Desa B	800	1.25	0.4
3	Desa C	600	1.2	0.32
4	Desa D	400	0.8	0.18

- ❖ Jika alokasi sampel untuk survei di atas dilakukan secara ***equal allocation***, berapakah ukuran sampel (n) dan ukuran sampel tiap desa (n_h) ? Diketahui tingkat kepercayaan 95% dan persentase *margin of error* 5% dari nilai rata-ratanya.