

Penarikan Contoh Acak Berlapis (*Stratified Random Sampling*)

Definisi

Contoh acak berlapis didapatkan dengan cara membagi populasi menjadi beberapa kelompok yang tidak saling tumpang tindih, dan kemudian mengambil secara acak dari setiap kelompok-kelompok itu. Kelompok tersebut dinamakan LAPISAN atau STRATA.

Karena tujuan dari penarikan contoh adalah mendapatkan contoh yang mewakili (*representative*) populasi, maka sifat lapisan adalah:

perbedaan objek antar lapisan tinggi, sedangkan dalam lapisan rendah.

Dengan kata lain, populasi disekat-sekat sehingga di setiap sekatan, objek memiliki karakteristik yang mirip.

Kelebihan PCAB dibandingkan PCAS

- 1 Menggunakan PCAB, akan memastikan bahwa contoh yang kita dapatkan terdiri atas berbagai kelompok. Jaminan ini tidak diberikan oleh PCAS. Misalnya saja, jika kita ingin mengambil contoh dari suatu kelas mahasiswa sebanyak 10 orang. Andaikan kita pisahkan terlebih dahulu berdasarkan jenis kelamin, kemudian secara acak kita ambil 5 orang dari kelompok perempuan dan 5 orang dari yang laki-laki, maka 10 orang yang diperoleh akan mewakili kedua jenis kelamin. Dengan PCAS, ada kemungkinan yang terambil perempuan saja, atau laki-laki saja. **Implikasinya, pendugaan menggunakan PCAB akan memberikan *bound of error* yang lebih kecil, dengan kata lain meningkatkan presisi hasil pendugaan. Hal tersebut benar hanya jika kita mampu membuat strata dengan tepat.**

Kelebihan PCAB dibandingkan PCAS

- 2 PCAB dapat mengurangi biaya survei, jika dibandingkan PCAS. Pada lokasi survei yang luas, hasil pengacakan menggunakan PCAS dapat saja terjadi kita harus pergi menyeberang pulau hanya untuk mendapatkan data dari satu orang responden. Menggunakan PCAB hal ini bisa diminimalkan.
- 3 Dengan PCAB dimungkinkan untuk melakukan pendugaan parameter di setiap sub-populasi secara terpisah, yaitu di setiap lapisan.

Membuat Lapisan (strata)

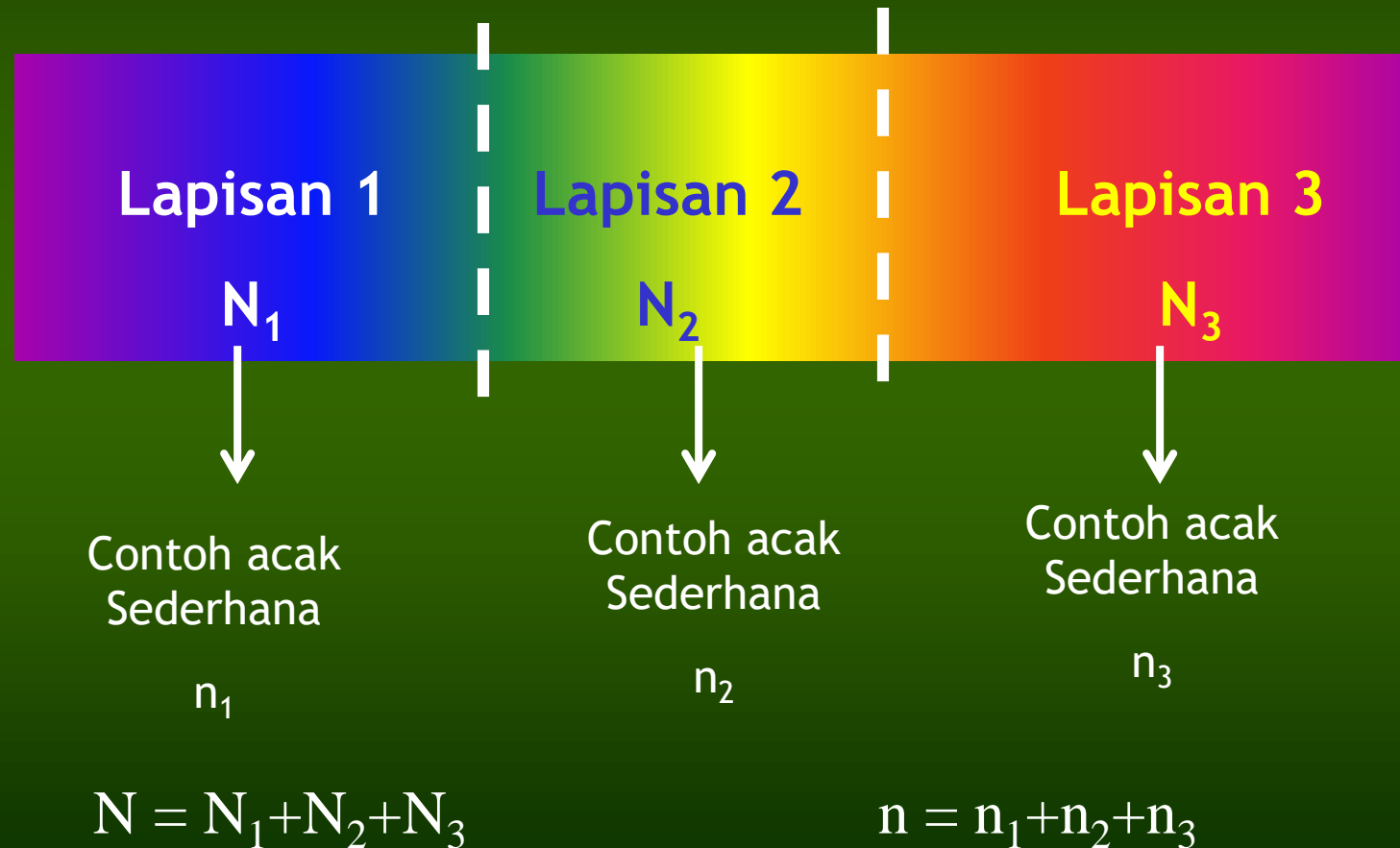
Lapisan adalah kumpulan objek yang memiliki karakteristik yang sama pada variabel yang ingin diambil datanya. Variabel yang menyekat populasi menjadi beberapa lapisan adalah variabel yang memiliki pengaruh atau berhubungan dengan variabel yang diambil datanya. Misalnya kita ingin menduga rata-rata waktu yang disediakan oleh mahasiswa untuk menonton televisi dalam sehari. Jika ada justifikasi yang kuat bahwa antara laki-laki dan perempuan terdapat perbedaan dalam hal tersebut, maka jenis kelamin dapat dipilih untuk dijadikan lapisan. Jika tidak demikian, harus dicari variabel lain. Penyusunan lapisan dapat dilakukan menggunakan lebih dari satu variabel.

Membuat Lapisan (strata) tugas

Tentukan variabel yang sesuai untuk menyekat populasi, jika kita ingin menduga rata-rata dari variabel berikut:

1. Tingkat konsumsi gas elpiji per rumahtangga per bulan di Kabupaten Bogor
2. Tingkat pemahaman masyarakat mengenai kewajiban membayar pajak
3. Waktu yang digunakan mahasiswa untuk belajar di luar jam kuliah
4. Kualitas Sekolah Dasar di Kabupaten/Kota Bogor

Cara menarik contoh acak berlapis



Pendugaan Rataan Populasi (μ)

$$\bar{y}_{st} = \frac{1}{N} [N_1 \bar{y}_1 + N_2 \bar{y}_2 + \dots + N_L \bar{y}_L] = \frac{1}{N} \sum_{i=1}^L N_i \bar{y}_i$$

$$\hat{V}(\bar{y}_{st}) = s_{\bar{y}_{st}}^2 = \frac{1}{N^2} \sum_{i=1}^L \left[N_i^2 \frac{N_i - n_i}{N_i} \frac{s_i^2}{n_i} \right]$$

Selang kepercayaan $(1-\alpha)100\%$ bagi μ :

$$\bar{y}_{st} \pm t_{\frac{\alpha}{2}} \sqrt{\hat{V}(\bar{y}_{st})}$$

Pendugaan Total Populasi (τ)

$$\hat{\tau}_{st} = N\bar{y}_{st} = [N_1\bar{y}_1 + N_2\bar{y}_2 + \cdots + N_L\bar{y}_L] = \sum_{i=1}^L N_i\bar{y}_i$$

$$\hat{V}(\hat{\tau}_{st}) = s_{\hat{\tau}_{st}}^2 = N^2 s_{\bar{y}_{st}}^2 = \sum_{i=1}^L \left[N_i^2 \frac{N_i - n_i}{N_i} \frac{s_i^2}{n_i} \right]$$

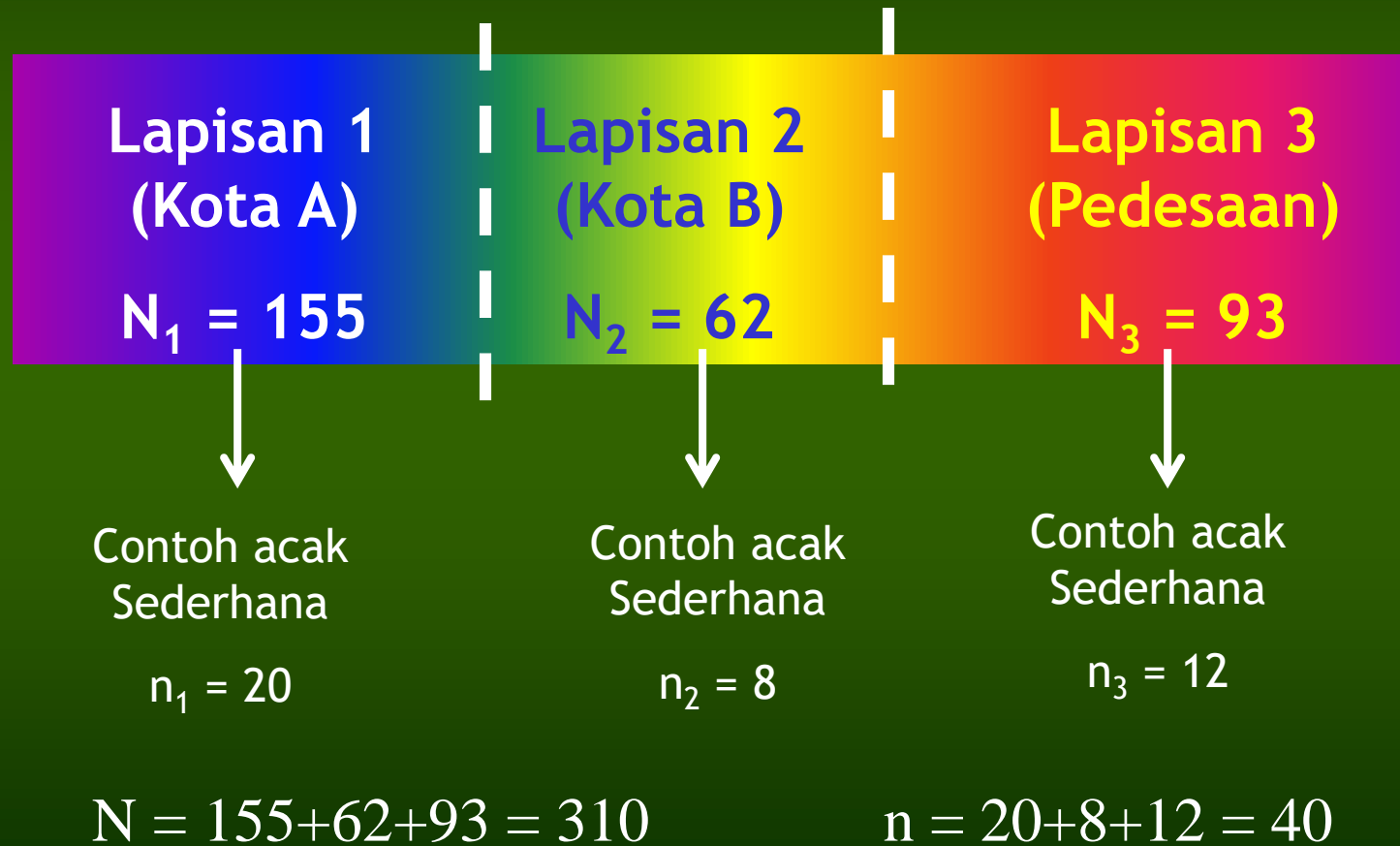
Selang kepercayaan $(1-\alpha)100\%$ bagi τ :

$$\hat{\tau}_{st} \pm t_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\tau}_{st})}$$

Ilustrasi

- Suatu perusahaan periklanan ingin menentukan berapa lama waktu yang akan dialokasikan untuk beriklan melalui televisi di suatu wilayah. Untuk tujuan tersebut perusahaan melakukan survei untuk menduga rata-rata lama waktu (jam) menonton televisi per rumahtangga per minggu di wilayah tersebut.
- Informasi yang tersedia mengenai wilayah tersebut sbb:
Wilayah terdiri dari dua kota, yaitu kota A dan kota B, serta daerah pedesaan yang mengitarinya. Kota A dibangun di sekitar pabrik, dan kebanyakan rumahtangga terdiri atas pekerja pabrik dengan anak-anak usia sekolah. Kota B adalah sub-urban eksklusif dari suatu kota di wilayah tetangga dan terdiri dari orang-orang yang lebih tua dengan sedikit anak yang tinggal di rumah.
- Buatlah rancangan penarikan contoh untuk masalah di atas.

Rancangan Survei



Data sampel: Waktu menonton TV (jam/minggu) di ketiga lapisan

Kota A

35	28	26	41
43	29	32	37
36	25	29	31
39	38	40	45
28	27	35	34

Kota B

27	4	49	10
15	41	25	30

Pedesaan

8	15	21	7
14	30	20	11
12	32	34	24

Ringkasan Perhitungan Setiap Lapisan

Lapisan	N_i	n_i	Rata-rata (\bar{y})	Ragam (s^2)
Kota A	155	20	33.900	35.358
Kota B	62	8	25.125	232.411
Pedesaan	93	12	19.000	87.636
Total	310	40		

Nilai dugaan rata-rata lama waktu menonton TV per keluarga (jam/minggu)

- Di seluruh wilayah

$$\begin{aligned}\bar{y}_{st} &= \frac{1}{N} [N_1 \bar{y}_1 + N_2 \bar{y}_2 + N_3 \bar{y}_3] \\ &= \frac{1}{310} [(155)(33.900) + (62)(25.125) + (93)(19.000)] \\ &= 27.675\end{aligned}$$

Nilai dugaan ragam rata-rata:

$$\begin{aligned}\hat{V}(\bar{y}_{st}) &= \frac{1}{N^2} \left[\sum_{i=1}^3 N_i (N_i - n_i) \frac{s_i^2}{n_i} \right] \\ &= \frac{1}{310^2} \left[(155)(155 - 20) \left(\frac{35.358}{20} \right) + (62)(62 - 8) \left(\frac{232.411}{8} \right) + (93)(93 - 12) \left(\frac{87.636}{12} \right) \right] \\ &= 1.97\end{aligned}$$

Nilai dugaan rata-rata lama waktu menonton per keluarga TV (jam/minggu)

- Selang kepercayaan kira-kira 95% bagi μ

$$\begin{aligned}\bar{y}_{st} \pm 2\sqrt{\hat{V}(\bar{y}_{st})} &\Leftrightarrow 27.675 \pm 2\sqrt{1.97} \\ &\Leftrightarrow 27.675 \pm 2.807 \\ &\Leftrightarrow 24.868 \text{ s / d } 30.482 \text{ jam / mg}\end{aligned}$$

- Selang kepercayaan kira-kira 95% bagi μ_2 (Rata-rata lama waktu menonton TV per keluarga di kota B:

$$\begin{aligned}\bar{y}_2 \pm 2\sqrt{\hat{V}(\bar{y}_2)} &\Leftrightarrow \bar{y}_2 \pm 2\sqrt{\left(\frac{N_2 - n_2}{N_2}\right) \frac{s_2^2}{n_2}} \\ &\Leftrightarrow 25.125 \pm 2\sqrt{\left(\frac{62 - 8}{62}\right) \frac{232.411}{8}} \\ &\Leftrightarrow 25.125 \pm 10.060 \Leftrightarrow 15.065 \text{ s / d } 35.185 \text{ jam / mg}\end{aligned}$$

Nilai dugaan total lama waktu menonton TV seluruh keluarga (jam)

- Penduga titik bagi τ

$$\hat{\tau} = N\bar{y}_{st} = 310(27.675) = 8579.250 \text{ jam}$$

- Nilai dugaam ragam total

$$\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}_{st}) = N^2\hat{V}(\bar{y}_{st}) = 310^2(1.97) = 189278.560 \text{ jam}$$

- Selang kepercayaan kira-kira 95% bagi τ (Total lama waktu menonton TV semua keluarga di wilayah tersebut):

$$\begin{aligned}\hat{\tau} \pm 2\sqrt{\hat{V}(\hat{\tau})} &\Leftrightarrow 8579.25 \pm 2\sqrt{189278.560} \\ &\Leftrightarrow 8579.25 \pm 870.12 \\ &\Leftrightarrow 7709.13 \text{ s / d } 9449.37 \text{ jam / mg}\end{aligned}$$

Penentuan Ukuran Contoh

$$n = \frac{\sum_{i=1}^L N_i^2 \sigma_i^2 / w_i}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$$

$$D = \frac{B^2}{4}, \text{ jika menduga } \mu$$

$$D = \frac{B^2}{4N^2}, \text{ jika menduga } \tau$$

B = *bound of error* pendugaan

$w_i = n_i/n$ = proporsi contoh masing-masing lapisan

Alokasi Contoh di Setiap Lapisan

1. Alokasi optimum



Meminimumkan $V(\bar{y})$ pada biaya yang tetap, atau meminimumkan biaya pada $V(\bar{y})$ yang tetap

$$n_i = n \left[\frac{N_i \sigma_i / \sqrt{c_i}}{\sum_{k=1}^L N_k \sigma_k / \sqrt{c_k}} \right]$$

c_i = biaya untuk memperoleh satu unit contoh dari lapisan ke- i

w_i

Alokasi Contoh di Setiap Lapisan

2. Alokasi Neyman



Jika biaya untuk memperoleh satu unit contoh dari setiap lapisan sama ($c_1 = c_2 = \dots = c_L$)

$$n_i = n \left[\frac{N_i \sigma_i}{\sum_{k=1}^L N_k \sigma_k} \right]$$

w_i 

Alokasi Contoh di Setiap Lapisan

3. Alokasi Proporsional

$$n_i = n \left[\frac{N_i}{\sum_{k=1}^L N_k} \right]$$

atau

$$n_i = n \left[\frac{N_i}{N} \right]$$

Formula di atas mengasumsikan biaya untuk memperoleh satu unit contoh dari setiap strata sama $\rightarrow c_1 = c_2 = \dots = c_L$, dan ragam antar strata sama $\rightarrow \sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2$



Pendugaan Proporsi Populasi

$$\hat{p}_{st} = \frac{1}{N} [N_1 \hat{p}_1 + N_2 \hat{p}_2 + \cdots + N_L \hat{p}_L] = \frac{1}{N} \sum_{i=1}^L N_i \hat{p}_i$$

$$\hat{V}(\hat{p}_{st}) = \frac{1}{N^2} \sum_{i=1}^L \left[N_i^2 \frac{N_i - n_i}{N_i} \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i - 1} \right]$$

Selang kepercayaan bagi p

$$\hat{p}_{st} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{p}_{st})}$$

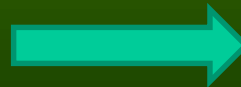
Penentuan Ukuran Contoh untuk Menduga Proporsi Populasi

$$n = \frac{\sum_{i=1}^L N_i^2 p_i (1 - p_i) / w_i}{N^2 \frac{B^2}{4} + \sum_{i=1}^L N_i p_i (1 - p_i)}$$

Alokasi Optimum



$$w_i = \frac{N_i \sqrt{\frac{p_i (1 - p_i)}{c_i}}}{\sum_{k=1}^L N_k \sqrt{\frac{p_k (1 - p_k)}{c_k}}}$$



$$n_i = n \left[\frac{N_i p_i (1 - p_i) / \sqrt{c_i}}{\sum_{k=1}^L N_k p_k (1 - p_k) / \sqrt{c_k}} \right]$$