

Penarikan Contoh Gerombol (*Cluster Sampling*)

Departemen Statistika FMIPA IPB

Review: Beberapa Istilah dalam Penarikan Contoh

1

Unsur (*element*)

An object on which a measurement is taken

2 Populasi (*population*)

A collection of elements about which we wish to make an inference

3

Satuan penarikan contoh (*sampling units*)

Nonoverlapping collections of element from the population that cover entire population

4 Kerangka (*frame*)

A list of sampling units

5 Contoh (*sample*)

A collection of sampling units drawn from a frame(s)

A subset of the population



Setiap unit yang dapat diambil sebagai contoh

Unsur (elemen)

cluster (gerombol):
kumpulan elemen

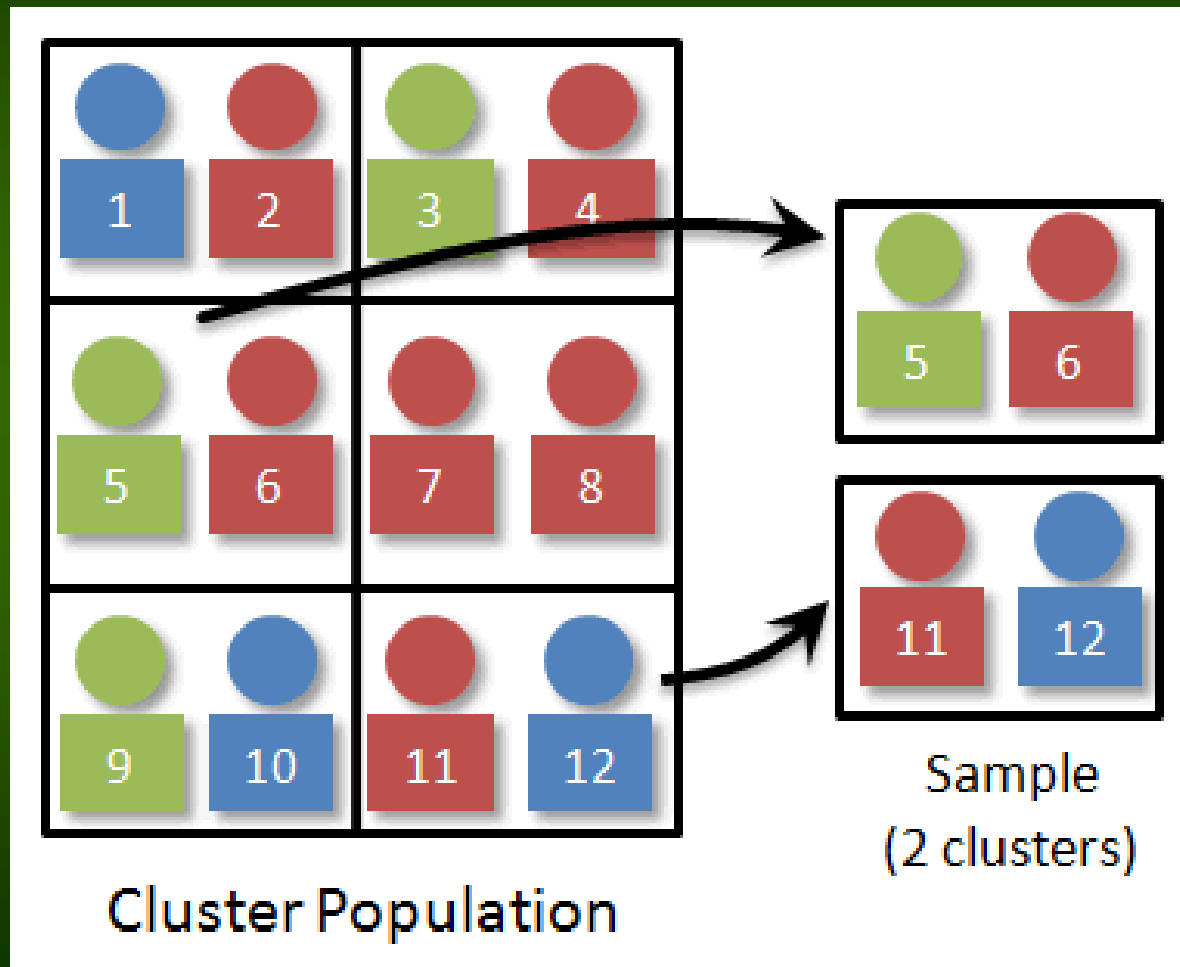
Definisi

Cluster Sample (Contoh Gerombol) adalah suatu contoh berpeluang yang satuan penarikan contohnya berupa gerombol (kumpulan elemen)

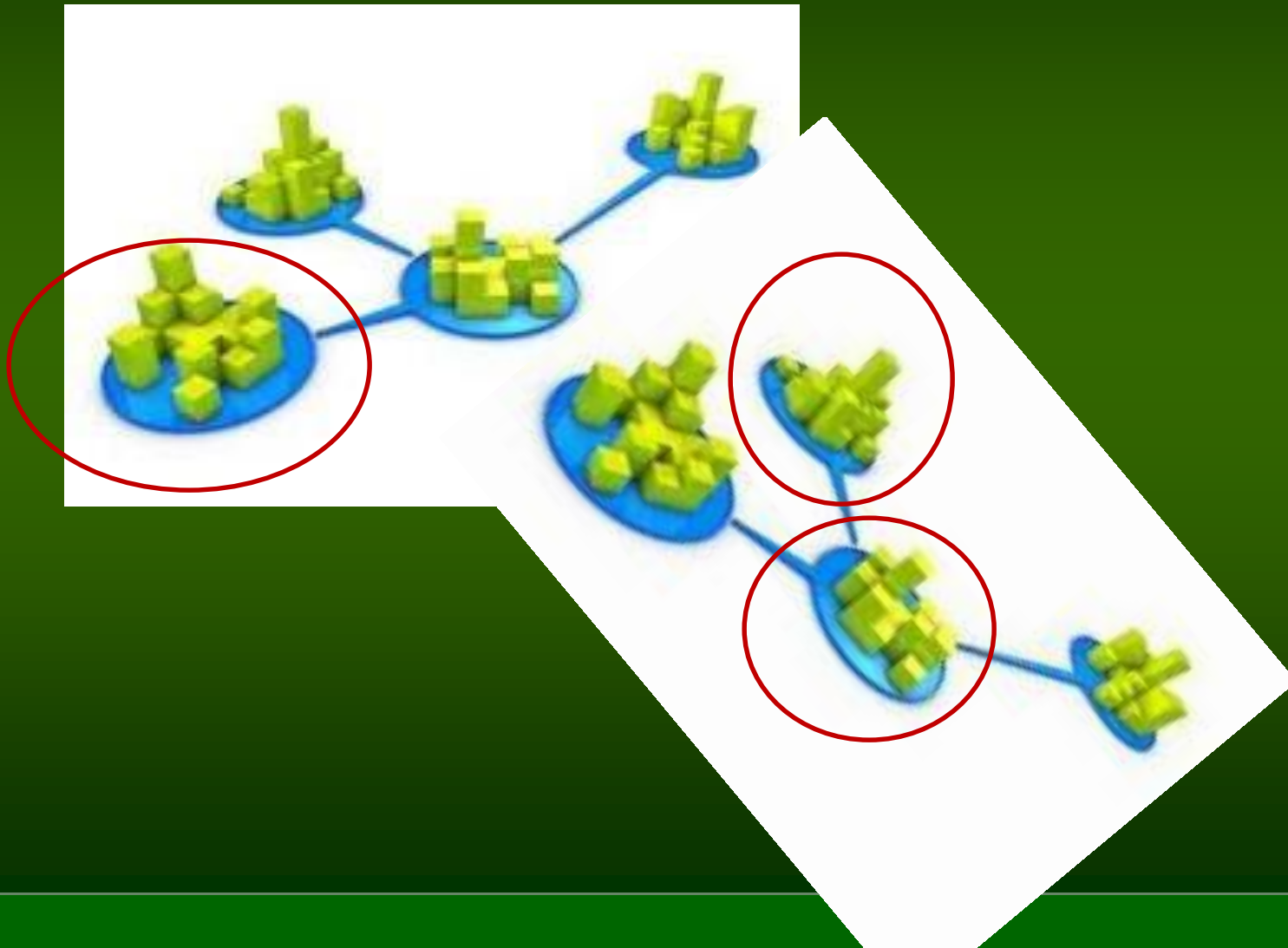
Cluster Sampling (Penarikan Contoh Gerombol = PCG) adalah penarikan contoh acak sederhana terhadap satuan contoh yang berupa gerombol

Dalam PCG semua elemen dalam gerombol yang terpilih sebagai contoh diamati semua

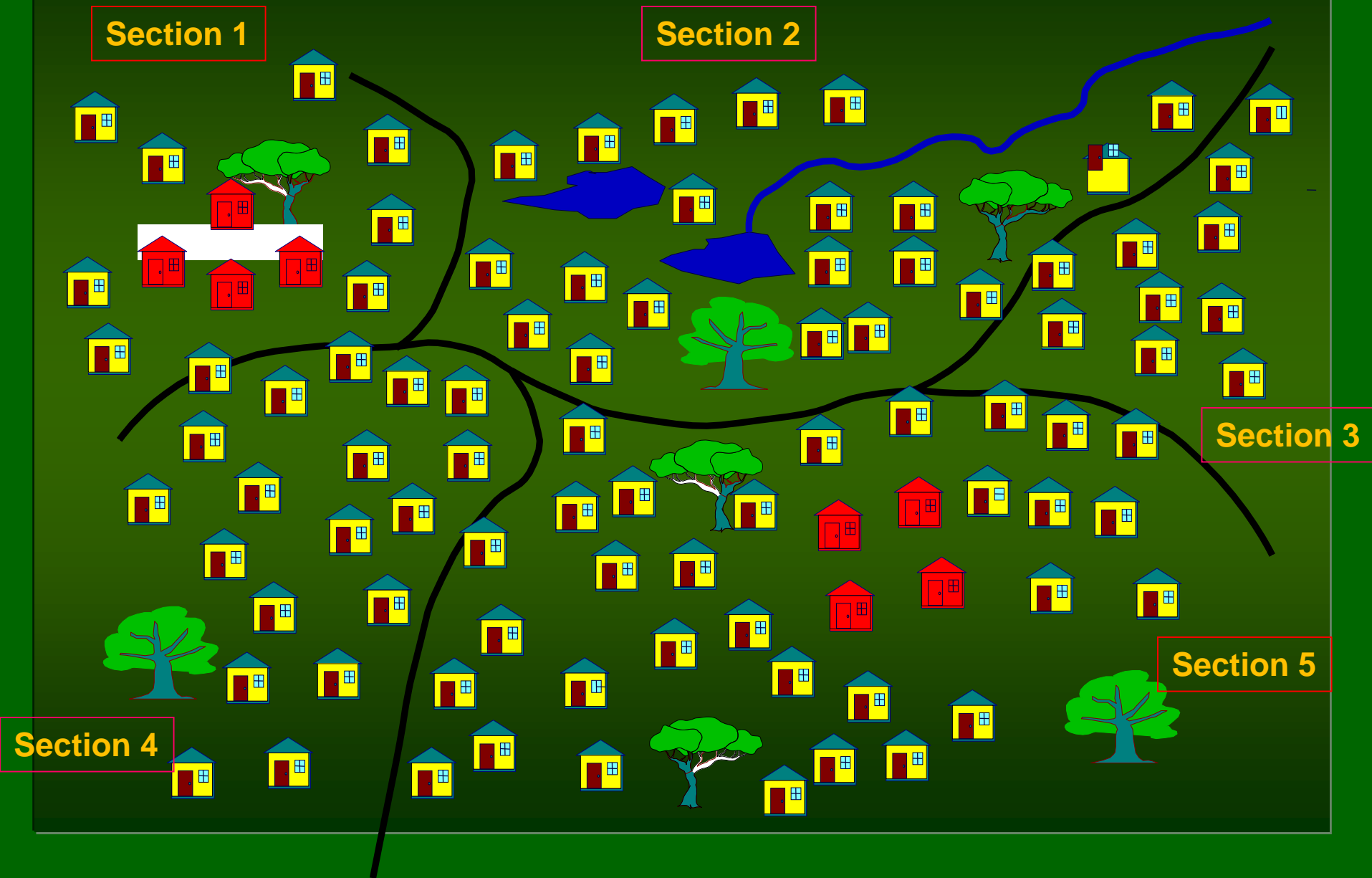
Example: Cluster Sampling



Example: Cluster Sampling



Example: Cluster Sampling



Ilustrasi

1

Populasi: seluruh penduduk yang terdaftar sebagai pemilih dalam suatu Pemilu

Elemen: setiap penduduk yang terdaftar sebagai pemilih

Gerombol: TPS (sebagai *sampling unit*)

2

Populasi: seluruh siswa aktif di sebuah SMA

Elemen: setiap siswa di SMA tsb

Gerombol: kelas (sebagai *sampling unit*)

Kapan Penarikan Contoh Gerombol Digunakan?

- Jika kerangka penarikan contoh elemen tidak tersedia, atau untuk mendapatkannya perlu biaya yang besar
- Jika biaya untuk memperoleh amatan meningkat dengan semakin jauhnya jarak antar elemen populasi

Contoh Kasus

Akan diduga rata-rata penghasilan rumahtangga di suatu kota kecil

Masalah: Tidak tersedia daftar semua rumahtangga di kota tsb



Gunakan desa/kelurahan/area sebagai gerombol, ambil beberapa desa/area secara acak, semua rumahtangga di desa/area yang terpilih sebagai contoh diamati

Cara Mengambil Contoh Gerombol

- Langkah pertama adalah mendefinisikan gerombol
- Pertimbangannya: (1) Kedekatan geografis antar elemen dalam gerombol, (2) Ukuran gerombol yang mudah ditangani
- Mana yang lebih baik? Mengambil banyak gerombol yang berukuran kecil atau sedikit gerombol yang berukuran besar?



Tergantung kemiripan karakteristik dalam gerombol

Stratified vs Cluster Sampling

- Kondisi dalam strata di *stratified sampling* adalah homogen dan antar strata heterogen
- Kondisi dalam gerombol (*cluster*) relatif heterogen dan antar *cluster* relatif mirip.

Notasi

N = banyaknya gerombol dalam populasi

n = banyaknya gerombol yang diambil sebagai contoh

m_i = banyaknya unsur dalam gerombol ke- i ,

$i = 1, 2, \dots, N$

$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$ = rata-rata ukuran gerombol untuk contoh

$M = \sum_{i=1}^N m_i$ = banyaknya unsur dalam populasi


$\bar{M} = \frac{M}{N}$ = rata-rata ukuran gerombol untuk populasi

\mathcal{Y}_i = total semua amatan dalam gerombol ke- i


Pendugaan Rataan Populasi (μ)

Penduga bagi μ adalah:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$$


$$\hat{V}(\bar{y}) = \left[\frac{N-n}{Nn\overline{M}^2} \right] \frac{\sum_{i=1}^n (y_i - \bar{y}m_i)^2}{n-1}$$

Penduga yang baik
bagi $V(y)$ jika $n \geq 20$



Selang Kepercayaan bagi μ :

$$\bar{y} \pm t_{\frac{\alpha}{2}(n-1)} \sqrt{\hat{V}(\bar{y})}$$

Bound of error = B



Ilustrasi 1

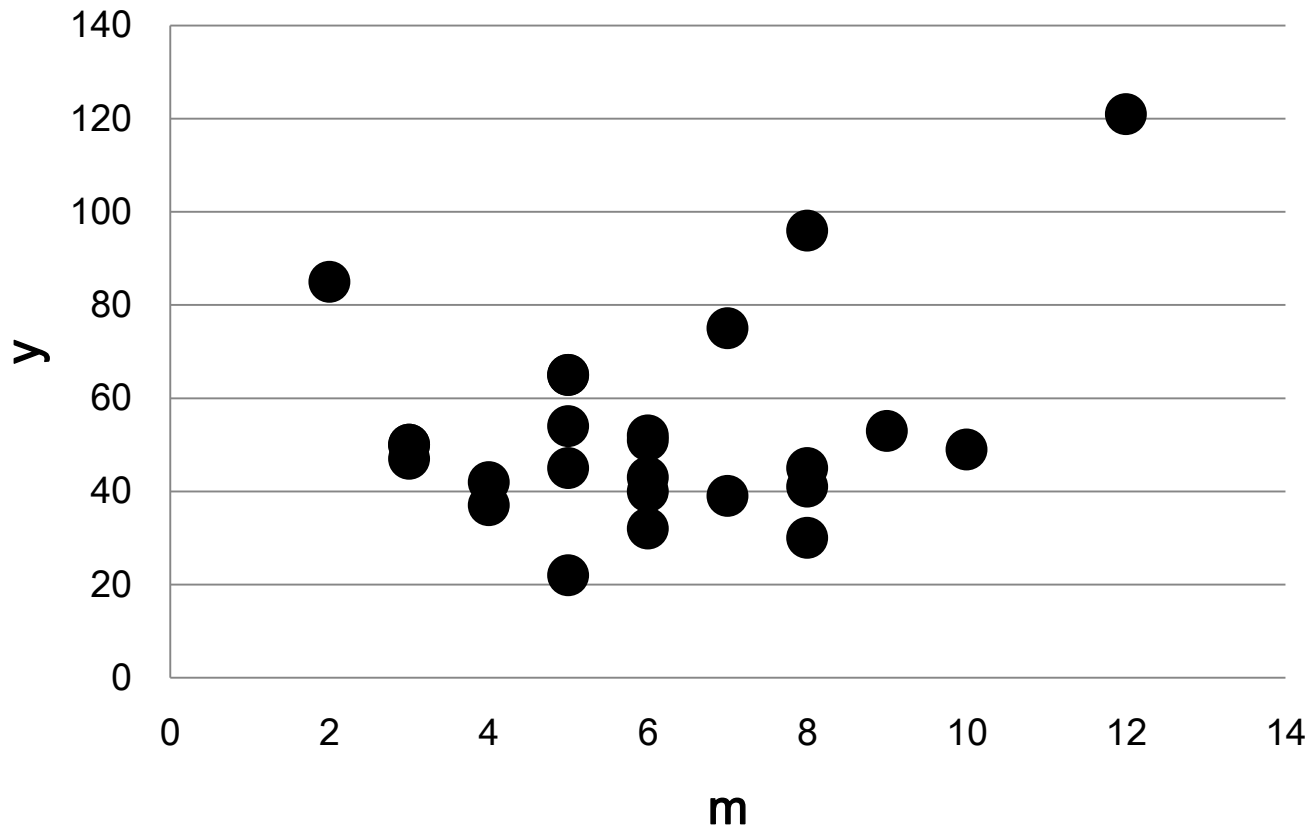
- Suatu survey dilakukan untuk menduga rata-rata pendapatan per rumahtangga di suatu kota kecil. Tidak ada daftar rumah tangga dewasa yang tersedia (tidak tersedia kerangka penarikan contoh unsur).
- Si peneliti memutuskan menggunakan penarikan contoh acak gerombol.
- Wawancara dilakukan terhadap rumah tangga di 25 gerombol terpilih.

Data yang diperoleh

Gerombol	Jml RT	y (ribu \$)	Gerombol	Jml RT	y (ribu \$)
1	8	96	14	10	49
2	12	121	15	9	53
3	4	42	16	3	50
4	5	65	17	6	32
5	6	52	18	5	22
6	6	40	19	5	45
7	7	75	20	4	37
8	5	65	21	6	51
9	8	45	22	8	30
10	3	50	23	7	39
11	2	85	24	3	47
12	6	43	25	8	41
13	5	54			

y = total pendapatan setiap gerombol

Scatter plot y vs m



Pertanyaan

- Dugalah rata-rata pendapatan per rumah tangga di kota tersebut, dan hitunglah *bound of error*
- Berikan selang kepercayaan 95% bagi rata-rata pendapatan per rumah tangga di kota tersebut.

Pendugaan Total Populasi (τ)

$$\tau = M\mu$$

$$\hat{\tau} = M\bar{y} = M \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$$

$$\hat{V}(\hat{\tau}) = \hat{V}(M\bar{y}) = N^2 \left(\frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}m_i)^2}{n-1}$$

Pendugaan Total Populasi (τ)

Jika M tidak diketahui:

$$\hat{\tau} = N\bar{y}_t = \frac{N}{n} \sum_{i=1}^n y_i$$

$$\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}_t) = N^2 \left(\frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n-1}$$

Pertanyaan

- Jika diketahui $M=2500$, dugalah total pendapatan semua rumah tangga di kota tersebut dan hitunglah *bound of error*
- Apabila dianggap M tidak diketahui, dugalah total pendapatan semua rumah tangga di kota tersebut dan hitunglah *bound of error*

Ukuran gerombol sama

$$m_1 = m_2 = \dots = m_N = m$$

- Perbandingan dengan *simple random sampling* (PCAS)

Penduga bagi μ adalah:

$$\bar{\bar{y}}_c = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i} = \frac{\sum_{i=1}^n y_i}{nm}$$

$$\hat{V}(\bar{\bar{y}}_c) = \left[1 - \frac{n}{N} \right] \frac{1}{nm} MSB$$

One-Way ANOVA

Kesimpulan.... Ada dua sumber keragaman (*source of variation*)

- Keragaman antar group, $SSB=JK(B)$, atau keragaman karena faktor
- Keragaman dalam group, $SSW =JKG$, atau keragaman yang tidak dapat diterangkan oleh faktor maka disebut keragaman galat.



TABEL SIDIK RAGAM
/ANOVA TABLE

Tabel Sidik Ragam

Sumber keragaman	df (db)	SS (JK)	MS (KT)
Antar gerombol	$n-1$	SSB	$MSB = SSB / (n-1)$
Dalam gerombol	$n(m-1)$	SSW	$MSW = SSW / (n(m-1))$
Total	$nm-1$	SST	

Note: $p = n$ dan $r = m$

Efisiensi Relatif PCG terhadap PCAS

$$RE(\bar{\bar{y}}_c / \bar{y}) = \frac{\hat{s}^2}{MSB}$$

$$\begin{aligned}\hat{s}^2 &= \frac{N(m-1)MSW + (N-1)MSB}{Nm-1} \\ &\approx \frac{1}{m} [(m-1)MSW + MSB]\end{aligned}$$

Ilustrasi 2

Ada 500 kotak telur, setiap kotak berisi 12 butir telur. Dari 500 kotak ini dipilih secara acak 5 kotak, kemudian setiap butir telur dari 5 kotak ini ditimbang. Jumlah bobot dari seluruh telur dari 5 kotak ini adalah 2900 gram. Berdasarkan data bobot setiap telur dari 5 kotak yang terpilih ini dapat diperoleh tabel analisis ragam sebagai berikut:

Sumber Keragaman	db	Jumlah Kuadrat (SS)
Antar Kotak	4	40.16
Dalam Kotak	55	743.05
Total	59	783.21

Ilustrasi 2 (Lanjutan)

Pertanyaan :

- Berikan nilai dugaan bagi rata-rata bobot sebutir telur dari 500 kotak telur itu.
- Dari tabel analisis ragam, berikan nilai dugaan bagi ragam dari rata-rata bobot sebutir telur pada bagian (a).
- Apakah penarikan contoh gerombol lebih efisien bila dibandingkan dengan penarikan contoh acak sederhana (dengan ukuran contoh yang sama) pada kasus ini? Berikan efisiensi relatif penarikan contoh gerombol terhadap penarikan contoh acak sederhana dengan menggunakan informasi yang tersedia.

Penentuan Ukuran Contoh

Tentukan dulu nilai *bound on the error estimation*, misalkan sebesar B

$$z_{\frac{\alpha}{2}} \sqrt{V(\bar{y})} = B \quad \longrightarrow \quad n = \frac{N \sigma_c^2}{N \frac{B^2 \bar{M}^2}{z^2} + \sigma_c^2}$$

$$s_c^2 = \frac{\sum_{i=1}^{n'} (y_i - \bar{y} m_i)^2}{n' - 1}$$



Nilai σ_c^2 dan M ditentukan berdasarkan informasi awal, atau melakukan survei pendahuluan terlebih dahulu dengan mengambil contoh awal berukuran n'

Pendugaan Proporsi Populasi

Jika a_i = banyaknya yang menjawab "Ya" dalam gerombol ke-i

m_i = banyaknya elemen dalam gerombol ke-i, maka:

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$$

dan

$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1}$$

Penduga yang baik bagi $V(\hat{p})$
jika $n \geq 20$

Ilustrasi 3

- Kembali ke ilustrasi (1). Pada saat bersamaan, ditanyakan apakah rumah yang bersangkutan merupakan rumah sewa atau rumah sendiri.
- Yang jadi minat peneliti adalah ingin menduga jumlah rumah tangga yang menyewa tempat tinggal, sekaligus dengan *bound of error* (batas kesalahan pendugaan)

Data yang diperoleh

Gerombol	Jml RT	# penyewa	Gerombol	Jml RT	# penyewa
1	8	4	14	10	5
2	12	7	15	9	4
3	4	1	16	3	1
4	5	3	17	6	4
5	6	3	18	5	2
6	6	4	19	5	3
7	7	4	20	4	1
8	5	2	21	6	3
9	8	3	22	8	3
10	3	2	23	7	4
11	2	1	24	3	0
12	6	3	25	8	3
13	5	2			

Ilustrasi 4

Sebuah perusahaan taksi ingin menduga proporsi banyaknya ban yang tidak aman dari 175 taksi yang dimiliki perusahaan tersebut. Untuk keperluan itu dipilih secara acak 25 buah taksi, kemudian diamati berapa banyaknya ban yang tidak aman dari setiap taksi. Berikut adalah hasil pengamatan banyaknya ban yang tidak aman dari 25 taksi yang diambil sebagai contoh:

2, 4, 0, 1, 2, 0, 4, 1, 3, 1, 2, 0, 1,
1, 2, 2, 4, 1, 0, 0, 3, 1, 2, 2, 1

Catatan: ban cadangan diabaikan (setiap taksi dianggap hanya memiliki 4 buah ban)

Ilustrasi 4 (Lanjutan)

- Dalam kasus di atas, apa yang menjadi gerombol (satuan penarikan contoh) nya? Apa pula yang menjadi unsurnya?
- Berikan nilai dugaan bagi proporsi banyaknya ban taksi yang tidak aman terhadap seluruh ban taksi yang dimiliki perusahaan tersebut.
- Berikan selang kepercayaan 95% bagi proporsi ban taksi yang tidak aman di perusahaan tersebut.
- Berikan nilai dugaan bagi banyaknya ban taksi yang tidak aman di perusahaan tersebut beserta batas galat pendugaannya.

TERIMA KASIH