

Modul 4: METODE CLUSTER SAMPLING

Capaian Pembelajaran:

1. *Mengenal sampling klaster sederhana*
2. *Menghitung taksiran jumlah populasi, variansi dan taksiran mean populasi*
3. *Mengenal karakteristik yang esensi dari sampling klaster*
4. *Menghitung alokasi optimum*

7.1 DEFINISI

Metode pengambilan sampel bloking (*cluster sampling*) adalah metode yang digunakan untuk memilih sampel yang berupa kelompok dari beberapa kelompok (*group/cluster*) dimana setiap kelompok terdiri atas beberapa unit yang lebih kecil (elements). Jumlah elemen dari masing-masing kelompok (*size of the clusters*) bisa sama maupun berbeda. Kelompok-kelompok (*groups*) tersebut dipilih baik dengan pengacakan pada kelompok pertamanya saja.

Contoh 1:

Sebuah Penelitian bertujuan untuk menentukan rata-rata tingkat pendapatan buruh bangunan di Kota Semarang dengan menggunakan *cluster sampling*.

Kota Semarang dibagi menjadi 16 Kecamatan, dari 16 Kecamatan dipilih 2 Kecamatan sebagai Populasi dari sampling I. Dari 2 Kecamatan masing-2 dipilih 2 Kelurahan sebagai Populasi dari sampel II. Dari 2 Kelurahan masing-2 dipilih 50 buruh bangunan sebagai sampel penelitian.

Contoh 2:

Metodologi Quick Count

1. **Populasi**

Seluruh pemilih di seluruh TPS (tempat pemungutan suara)

2. **Sampel**

Sebanyak 2000 TPS dipilih secara acak dari total populasi TPS (809.497) di seluruh Indonesia secara proporsional.

3. **Tingkat Kepercayaan**

Dengan sampel sebesar 2000 TPS, tingkat kepercayaan Quick Count sebesar 99% dengan toleransi kesalahan (*margin of error*) sebesar $+/- 1\%$. Artinya, perolehan suara kandidat dari hasil Quick Count ini bisa bergeser ke atas atau ke bawah hingga sebesar 1%.

4. **Metode Penarikan Sampel**

Sampel dipilih dengan metode *stratified-cluster random sampling* dari populasi tersebut.

5. **Metode Pengumpulan Data**

2000 relawan disebar di 2000 TPS terpilih, setiap relawan memonitor pemungutan suara sejak dibukanya TPS sampai penghitungan suara. Hasil penghitungan suara (CT) dikirim ke pusat data, kemudian data ditabulasi dan ditampilkan secara *real time*.

7.2 PENDUGAAN MEAN DAN VARIANSI

1. Rata – rata

$$\bar{y} = \frac{\sum y_{ij}}{\sum m_i}$$

Keterangan :

\bar{y} = Rata-rata populasi

y_{ij} = Total nilai pengamatan

$\sum m_i$ = Banyaknya elemen pada blok ke-i

2. Variansi dari rata – rata

$$V_{(\bar{y})} = \left(\frac{N-n}{nNM} \right) \left(\frac{\sum (y_i - y_{mi})^2}{n-1} \right)$$

Keterangan :

$V_{(\bar{y})}$ = varians dari rata-rata sampel

N = Banyaknya blok pada populasi

n = Banyaknya blok pada sampel acak sederhana

M = Rata-rata ukuran blok pada populasi

y_i = Total nilai pengamatan pada blok ke-i

3. Batas kesalahan total populasi

$$2\sqrt{V_{(\bar{y})}}$$

Variansi untuk populasi

$$\begin{aligned} V_{(\bar{y})} &= \frac{N-n}{n(n-1)N} \sum_{i=1}^n (p_i - \bar{p})^2 + \frac{1}{nN} \frac{(M-m)}{(m-1)M} \sum_{i=1}^n p_i(1-p_i) \\ &= \frac{N-n}{nN} \left(\frac{n \sum p_i^2 - (\sum p_i)^2}{n(n-1)} \right) + \frac{1}{nN} \frac{(M-m)}{(m-1)M} \sum_{i=1}^n p_i(1-p_i) \end{aligned}$$

Keterangan :

$$p_i = \frac{a_i}{m_i} \quad \text{= Proporsi dari elemen dengan atribut pada kelompok ke-}i.$$

7.3 KELEBIHAN DAN KEKURANGAN

Dalam cluster sampling terdapat beberapa kelebihan maupun kekurangan. Terdapat dua kelebihan utama dari metode sampel bloking ini.

1. Tidak perlunya menyusun kerangka sampling dari seluruh populasi yang ingin diteliti, seperti jika digunakan metode acak sederhana. Kerangka sampling cukup dibuat mengenai bloking-bloking (clusters) yang ada.
2. rendahnya biaya yang harus dikeluarkan ketimbang jika menggunakan metode stratified random sampling. Metode bloking ini tetap akan menjadi lebih murah karena sampel yang terambil pada akhirnya secara fisik akan terletak pada jarak/lokasi yang relative berdekatan. Dengan metode lainnya sampel yang terambil dapat sedemikian menyebarluas sehingga akan diperlukan waktu, tenaga dan biaya yang lebih lama atau banyak.

Kelemahan utama dari metode ini adalah kurang efisien dalam proses pemilihan anggota sampel dibandingkan metode lainnya, karena sub-sub dari setiap kelompok yang ditentukan dalam suatu stage diasumsikan memiliki karakteristik homogen (hampir homogen). Meskipun memiliki kelemahan, metode ini dapat lebih dipilih dari pada yang lain jika memang tidak dimiliki informasi atau data yang lengkap tentang kerangka sampelnya, ataupun jika pertimbangan biaya pendataan di lapangan yang rendah menjadi lebih penting daripada ketetapan atau presisi hasil yang diharapkan.

7.4 UKURAN SAMPEL

Ukuran sampel yang dimaksud dalam hal ini adalah banyaknya blok sampel (n). Nilai-nilai parameter pada rumus berikut diduga dengan nilai statistik yang diperoleh pada penelitian pendahuluan. Menarik sampel dari suatu populasi tidak selalu memuaskan, karena peneliti seringkali tidak yakin terhadap sampel yang sungguh-sungguh representatif terhadap populasi tersebut. Sering terjadi perbedaan antara sampel yang dipilih dengan populasi. Tetapi

bila sampel telah dipilih secara random dengan ukuran yang sesuai, maka perbedaan ini tampaknya tidak signifikan dan bersifat kebetulan.

Ukuran Sampel untuk menduga rata-rata

$$n = \frac{NZ_{\alpha}^2 \sigma_2}{Ne^2(\bar{M})^2 + Z_{\alpha}^2 \sigma^2}$$

Dalam hal ini σ^2 diduga dengan $S^2 = \frac{\sum(v_i - m_i \bar{v})^2}{n-1}$ yang diperoleh dari hasil penelitian pendahuluan Ukuran sampel untuk menduga proporsi

$$n = \frac{NZ_{\alpha}^2 \sigma_2}{Ne^2(\bar{M})^2 + Z_{\alpha}^2 \sigma^2}$$

Dalam hal ini σ^2 diduga dengan $S^2 = \frac{\sum(a_i - m_i p)^2}{n-1}$ dan $p = \frac{\sum a_i}{\sum m_i}$

Dimana a_i adalah banyaknya elemen yang masuk kriteria pada blok ke- i .

Contoh

Produsen gergaji melakukan penelitian untuk memperkirakan biaya perbaikan rata-rata per bulan yang telah dijual ke industri tertentu. Perusahaan tersebut tidak dapat memperoleh biaya perbaikan untuk setiap melihat saja, tapi dia bisa mendapatkan jumlah total yang dihabiskan untuk perbaikan gergaji dan jumlah gergaji yang dimiliki oleh masing-masing industri, sehingga ia memutuskan untuk menggunakan cluster sampling, dengan masing-masing industri sebagai cluster. Produsen memilih sampel acak sederhana $n = 20$ dari $N = 96$ industri layanan. Data pada total biaya perbaikan per industri dan jumlah gergaji per industri yang seperti yang diberikan pada tabel terlampir. Perkirakan biaya perbaikan rata-rata selama sebulan terakhir, dan tempatkan variabel terikat pada kesalahan estimasi.

Industri	Banyaknya Gergaji	Jumlah Biaya Perbaikan untuk sebulan (dolar)
1	3	50
2	7	110
3	11	230
4	9	140
5	2	60
6	12	280
7	14	240
8	3	45
9	5	60
10	9	230

11	8	140
12	6	130
13	3	70
14	2	50
15	1	10
16	4	60
17	12	280
18	6	150
19	5	110
20	8	120

Penyelesaian :

$$\text{Diketahui } n = 20, N = 96, \sum_{i=1}^{20} y_i = 2565, \sum_{i=1}^{20} m_i = 130, \bar{m} = \frac{m_i}{n} = \frac{130}{20} = 6.5.$$

Akan dicari estimasi rata-rata biaya perbaikan per gergaji selama sebulan terakhir.

Karena rata-rata jumlah anggota dalam populasi tidak diketahui, maka $\bar{m} = \bar{M} = 6.5$.

Sehingga estimasi rata-ratanya yaitu :

$$\bar{y} = \sum_{i=1}^{20} \frac{y_i}{m_i} = \frac{2565}{130} = 19,73\text{\$}$$

Kemudian diperoleh pula variansnya yaitu :

$$\begin{aligned} v_{(\bar{y})} &= \left(\frac{N-n}{nNM} \right) \left(\frac{\sum_{i=1}^{20} (y_i - \bar{y}m_i)^2}{n-1} \right) \\ &= \left(\frac{96-20}{(20)(96)(6.5)^2} \right) \left(\frac{16065,65}{20-1} \right) = 0.792 \end{aligned}$$

Nilai error dari estimasi rata-ratanya yaitu :

$$2\sqrt{V_{(\bar{y})}} = 1.78\text{\$}$$

Jadi, nilai estimasi rata-rata biaya pengeluaran perbaikan per gergaji selama sebulan terakhir adalah $9,73\text{\$} \pm 1.78\text{\$}$