

BAB III

METODE *MULTISTAGE CLUSTER SAMPLING*

A. Pendahuluan

Metode *multistage cluster sampling* adalah proses pengambilan sampel yang dilakukan melalui dua tahap pengambilan sampel atau lebih (Cochran, 1977:314). Penarikan sampel dengan metode ini sebenarnya tidak jauh berbeda dengan penarikan sampel dengan metode *cluster sampling*. Penarikan sampel dengan metode *cluster sampling* dilakukan secara langsung ke elemen atau ke *cluster* yang langsung mencakup seluruh elemen dalam *cluster* untuk dikumpulkan datanya. Sedangkan pada metode *multistage cluster sampling* ada perluasan dalam penarikan sampelnya, yaitu tidak langsung dilakukan penarikan sampel pada elemen, tetapi melalui *cluster* terlebih dahulu seperti dijelaskan di atas. Pada metode *multistage cluster sampling* unit sampling yang dipilih pada tahap pertama disebut unit sampling primer (PSU). Sedangkan unit sampling yang dipilih pada penarikan sampel tahap kedua disebut unit sampling sekunder dan seterusnya (Hansen, 1953:315). Penarikan sampel dengan metode *multistage cluster sampling* didasarkan pada:

1. Tidak tersedianya kerangka sampel yang memuat unit-unit sampel yang terkecil (*ultimate sampling unit*)
2. Untuk membuat kerangka sampel yang memuat unit-unit sampel terkecil memerlukan biaya, tenaga dan waktu yang besar

3. Dengan menggunakan metode *multistage cluster sampling*, maka pengawasan lapangan dapat lebih ditingkatkan sehingga *nonsampling error* dapat ditekan
4. Ditinjau dari segi biaya, metode *multistage cluster sampling* lebih efisien dibandingkan metode sampel acak sederhana (SRS)

Bentuk estimasi pada metode *multistage cluster sampling* tergantung pada metode sampling yang dipakai pada setiap tahap.

B. Proses Penarikan Sampel

Proses penarikan sampel merupakan cara pemilihan sampel yang akan menjadi dasar perkiraan. Prosedur penarikan sampel terdiri dari beberapa tahapan diantaranya adalah sebagai berikut:

- Tahap pertama yaitu memilih populasi dan membagi populasi menjadi beberapa fraksi sebagai dasar untuk penarikan sampel pada tahap pertama atau *primary sampling unit* (PSU) kemudian diambil sampelnya.
- Tahap kedua yaitu sampel fraksi yang dihasilkan dibagi lagi menjadi fraksi-fraksi yang lebih kecil kemudian diambil sampelnya. Pembagian fraksi terus dilakukan sampai pada unit sampel yang diinginkan.

Contoh: Akan diadakan penelitian pola pemanfaatan sarana pelayanan kesehatan di kabupaten Bogor, maka:

- Penduduk Bogor : populasi studi
- Kecamatan : unit sampling primer (PSU)
- Rumahtangga : unit sampel

Dengan cara sebagai berikut:

- Diambil satu kecamatan secara acak sebagai kecamatan sampel (tahap pertama)
- Satu tempat dalam kecamatan terpilih dijadikan sampel tahap kedua
- Seluruh RT dalam daerah tersebut dijadikan unit sampling dan dipilih secara acak sesuai dengan jumlah sampel yang diinginkan

Cara ini dipergunakan bila:

- Populasinya cukup homogen
- Jumlah populasinya sangat besar
- Populasi menempati daerah yang sangat luas
- Biaya penelitian kecil

C. Perkiraan Besarnya Sampel

Dalam penentuan besarnya sampel, maka diadakan kajian dari survei yang pernah dilakukan sebelumnya guna penentuan tingkat heterogenitas karakteristik di antara *cluster*, tingkat heterogenitas karakteristik di dalam *cluster*, prevalensi suatu kejadian yang disesuaikan terhadap target populasi dan *design effect*. Dengan melihat besaran variabel tersebut dikaji efisiensi desain dan besarnya sampel yang diperlukan. Data yang digunakan biasanya berupa data hasil survei. Untuk memperoleh sampel yang cukup besar, maka dapat menggunakan cara *rotating sample*. Menurut Kish (1979:2) *Rotating samples* adalah hasil pendataan survei beberapa tahun dijadikan dasar untuk estimasi wilayah yang lebih kecil. Hasilnya akan merupakan nilai rata-rata. Hal ini disarankan bila dana tersedia

tidak memungkinkan untuk melakukan survei sampel dengan level penyajian lebih kecil. Misal data level penyajian kecamatan diperkirakan berdasarkan data beberapa tahun dari survei dengan level penyajian kabupaten/kota.

Dari data survei yang ada ditentukan target populasi, kejadian terhadap target populasi dan kejadian target populasi terhadap populasi dari variabel tertentu yang ada pada data survei. Angka ini dapat digunakan untuk memperkirakan besarnya sampel (m) dengan rumus yang digunakan berupa *probability proportional to size* (PPS).

$$m = \frac{Z_{\alpha/2}^2 (p)(1-p)(deff)(1.05)}{(e.p)^2 (k)(\bar{x})} \quad (3.1)$$

dengan

m : banyaknya sampel yang diperlukan

$Z_{\alpha/2}$: 1,96 untuk tingkat kepercayaan 95%

P : perkiraan proporsi kejadian dari indikator yang akan diperkirakan
(proporsi kejadian terhadap target populasi)

1.05 : besaran untuk antisipasi non respons sebesar 5%

$deff$: *design effect*, rasio varians suatu metode sampling dibagi dengan varians dari metode sampel acak sederhana (SRS)

e : persentase *margin of error* (*relative standard error*) terhadap p (bisa ditentukan berbeda untuk setiap indikator)

k : proporsi kejadian dari target populasi terhadap populasi

\bar{x} : rata-rata banyaknya anggota rumahtangga

Dalam penentuan besarnya sampel dipengaruhi oleh banyaknya kejadian atau proporsi kejadian baik terhadap target populasi dan target populasi terhadap populasi secara keseluruhan serta rata-rata anggota rumahtangga. Disamping itu perlu diperhitungkan juga *level of confidence* (selang kepercayaan), *margin of error* (batasan kesalahan yang diharapkan) dan tambahan sampel untuk mengatasi apabila terjadi *non response*.

Non response perlu diperhitungkan karena pada pelaksanaan lapangan sering terjadi adanya *non response* seperti responden menolak menjawab, responden tidak dijumpai, responden telah pindah, daftar isian hilang atau daftar isian tidak dapat diolah. Masalah ini perlu diantisipasi dan diamati sejak awal sehingga jumlahnya tidak terlalu banyak karena akan menyebabkan bias dalam estimasi. Menurut Kish (1979) *non response* untuk survei sampel diperhitungkan sebesar 5 persen dengan nilai 1,05.

D. Estimasi karakteristik sampel

1. Total dan rata-rata populasi

Misalkan banyaknya unit yang dapat dijadikan dasar untuk penarikan sampel pada tahap pertama (PSU) adalah N dan banyaknya unit yang dapat dijadikan dasar penarikan sampel pada tahap kedua atau *secondary sampling unit* (SSU) pada setiap unit penarikan sampel pada tahap pertama yang ke- i adalah M_i .

Bila Y_{ij} menyatakan nilai karakteristik Y pada unit SSU ke- j dalam unit PSU ke- i , maka nilai total dan rata-rata dapat dinyatakan sebagai berikut

a. Total nilai karakteristik Y pada setiap PSU ke- i adalah

$$Y_i = \sum_{j=1}^{M_i} y_{ij} \quad (3.2)$$

dengan

y_{ij} : nilai karakteristik dari unit ke-j dari *cluster* ke-i

M_i : banyaknya unit sampel pada *cluster* terpilih ke-i

Y_i : nilai total karakteristik Y pada setiap PSU ke-i

b. Rata-rata nilai karakteristik PSU ke-i adalah

$$\bar{Y}_i = \frac{\sum_{j=1}^{M_i} y_{ij}}{m} \quad (3.3)$$

dengan

\bar{Y}_i : nilai rata-rata karakteristik PSU ke-i

m : banyaknya unit sampel terpilih pada setiap *cluster* (konstan pada setiap *cluster*)

c. Total nilai karakteristik dalam populasi adalah

$$Y = \sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij} \quad (3.4)$$

$$= \sum_{i=1}^N Y_i \quad (3.5)$$

$$= \sum_{i=1}^N M_i \bar{Y}_i \quad (3.6)$$

dengan

Y : nilai total karakteristik dalam populasi

N : jumlah *cluster* dalam populasi

- d. Rata-rata nilai karakteristik per unit PSU dalam populasi

$$\bar{Y} = \frac{Y}{N} \quad (3.7)$$

$$= \frac{1}{N} \sum_{i=1}^N M_i \bar{Y}_i \quad (3.8)$$

dengan

\bar{Y} : nilai rata-rata karakteristik per unit PSU dalam populasi

- e. Rata-rata nilai karakteristik per unit SSU dalam populasi

$$\bar{Y} = \frac{Y}{M} \quad (3.9)$$

$$= \frac{Y}{\sum_{i=1}^N M_i} \quad (3.10)$$

$$= \frac{\sum_{i=1}^N M_i \bar{Y}_i}{\sum_{i=1}^N M_i} \quad (3.11)$$

dengan

\bar{Y} : nilai rata-rata karakteristik per unit SSU dalam populasi

2. Estimasi total dan varians estimasi total

- a. Rata-rata nilai karakteristik per unit sampel dari *cluster* ke-*i*

$$\bar{y}_i = \frac{\sum_{j=1}^m y_{ij}}{m} \quad (3.12)$$

dengan

\bar{y}_i : nilai rata-rata karakteristik per unit sampel dari *cluster* ke- i

b. Rata-rata nilai karakteristik unit sampel pada masing-masing *cluster*

$$y = \frac{\sum_{i=1}^n \sum_{j=1}^m y_{ij}}{nm} \quad (3.13)$$

$$= \frac{\sum_{i=1}^n \bar{y}_i}{n} \quad (3.14)$$

dengan

n : banyaknya *cluster* terpilih

y : nilai rata-rata karakteristik unit sampel pada masing-masing *cluster*

c. Varians dari nilai karakteristik unit sampel antar *cluster*

$$s_b^2 = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n-1} \quad (3.15)$$

dengan

s_b^2 : varians dari nilai karakteristik unit sampel antar (*between*) *cluster*

d. Varians dari estimasi rata-rata nilai karakteristik sampel

$$v(\bar{y}) = \frac{N-n}{N} \left(\frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n-1} \right) \quad (3.16)$$

$$v(\bar{y}) = \frac{N-n}{N} \frac{s_b^2}{n} \quad (3.17)$$

Apabila nilai $\frac{N-n}{N} = 1$ sehingga diperoleh

$$v(\bar{y}) = \frac{s_b^2}{n} \quad (3.18)$$

dengan

$v(\bar{y})$: varians dari estimasi rata-rata nilai karakteristik sampel

e. *Standard error* dari estimasi rata-rata nilai karakteristik sampel

$$se(\bar{y}) = \sqrt{v(\bar{y})} \quad (3.19)$$

dengan

$se(\bar{y})$: perkiraan kesalahan yang disebabkan oleh metode sampling
(*standard error*)

f. *Relative standard error* dari estimasi rata-rata nilai karakteristik sampel

$$rse(\bar{y}) = \frac{se(\bar{y})}{\bar{y}} \quad (3.20)$$

dengan

$rse(\bar{y})$: relative standard error, merupakan rasio *standard error* \bar{y} terhadap nilai rata-rata karakteristik unit sampel pada masing-masing *cluster*

g. Varians dari karakteristik unit sampel dalam *cluster* ke- i

$$s_{wi}^2 = \frac{\sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{m-1} \quad (3.21)$$

dengan

s_{wi}^2 : varians dari karakteristik unit sampel dalam (*within*) *cluster* ke- i

h. Rata-rata varians per *cluster* dari karakteristik sampel

$$s_w^2 = \frac{\sum_{i=1}^n \left(\sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 / (m-1) \right)}{n} \quad (3.22)$$

$$s_w^2 = \frac{\sum_{i=1}^n s_{wi}^2}{n} \quad (3.23)$$

dengan

s_w^2 : rata-rata varians per *cluster* dari karakteristik unit sampel yang
diperhitungkan dari masing-masing *cluster*

Dengan menghitung varians antar *cluster* dan varians dalam *cluster* maka dapat dihitung *intercluster (intra class) correlation coefficient* atau sering disebut *rate of homogeneity*. Dalam analisis ini digunakan data survei, sehingga penghitungan *intercluster correlation coefficient* tersebut merupakan suatu pendekatan.

$$\hat{\rho} = \frac{(n-1)m \left(\sum_{i=1}^n (\bar{y}_i - \bar{y})^2 / (n-1) \right) - n \left(\sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 / (m-1) \right) / n}{(n-1)m \left(\sum_{i=1}^n (\bar{y}_i - \bar{y})^2 / (n-1) \right) + n(m-1) \left(\sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 / (m-1) \right) / n} \quad (3.24)$$

$$\hat{\rho} = \frac{(n-1)m \left(\sum_{i=1}^n (\bar{y}_i - \bar{y})^2 / (n-1) \right) - n \left(\frac{s_{wi}^2}{n} \right)}{(n-1)m \left(\sum_{i=1}^n (\bar{y}_i - \bar{y})^2 / (n-1) \right) + n(m-1) \left(\frac{s_{wi}^2}{n} \right)} \quad (3.25)$$

$$\hat{\rho} = \frac{(n-1)ms_b^2 - ns_w^2}{(n-1)ms_b^2 + n(m-1)s_w^2} \quad (3.26)$$

$\hat{\rho}$: *intercluster correlation coefficient*, menunjukkan tingkat keeratan

hubungan suatu karakteristik antar unit di dalam *cluster* atau disebut *rate*

of homogeneity (roh), nilai roh di antara $-\frac{1}{M-1}$ dan 1

$$deff = \frac{\text{varians metode sampling}}{\text{varians metode sampel acak sederhana}}$$

$$= 1 + (m-1)\hat{\rho}$$

deff : *design effect*, yaitu efek dari penggunaan metode sampling yang merupakan rasio antara varians suatu metode sampling dan metode sampel acak sederhana (SRS)

Rasio = 1 berarti metode tersebut sama efisien dengan metode sampel acak sederhana (SRS)

Rasio < 1 berarti metode sampel acak sederhana (SRS) kurang efisien

Rasio > 1 berarti metode sampel acak sederhana (SRS) lebih efisien

Untuk tingkat kepercayaan sebesar 95% maka $Z_{\alpha/2} = 1,96$. Tingkat kepercayaan ini menunjukkan bahwa estimasi yang diperoleh nantinya diharapkan akan terletak pada selang $\pm Z_{\alpha/2} \times (\text{standard error})$, dengan kriteria tersebut maka selang dari estimasi p menjadi $p \pm Z_{\alpha/2}(e p)$, dimana:

e : *persentase margin of error* atau *relative standard error*

$e p$: *standard error*