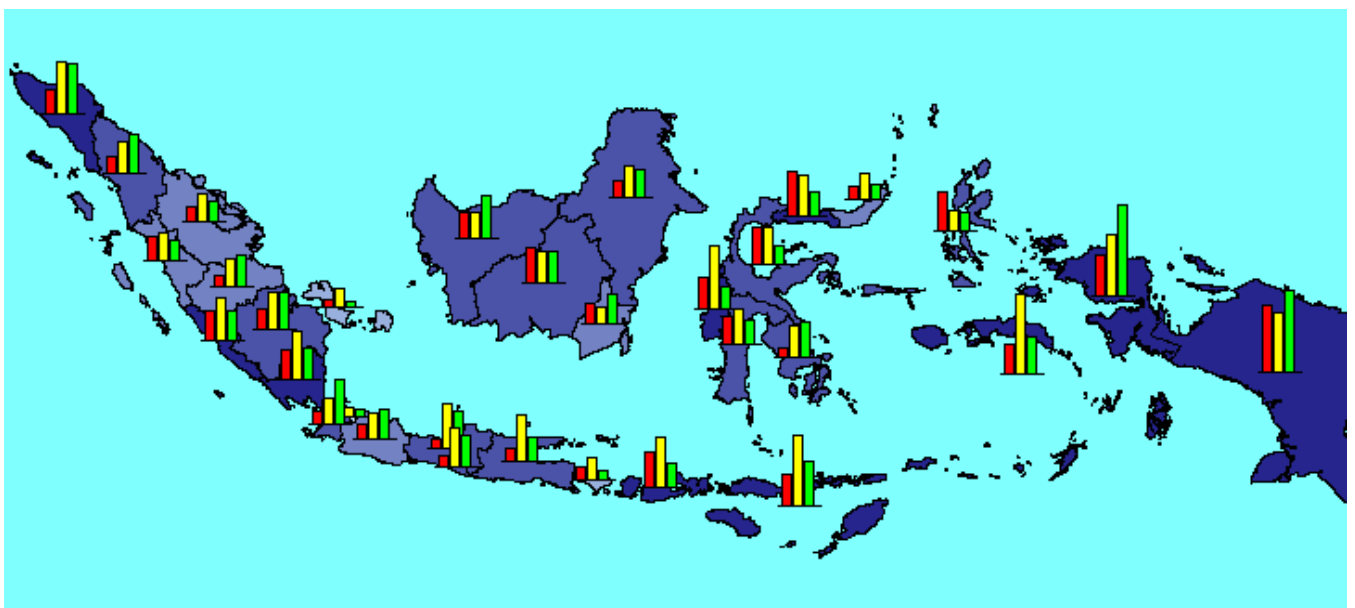


Metodologi Sampel Survey



Direktorat Metodologi Statistik

Diklat Teknis Kompetensi Kepala BPS Kabupaten/Kota
BPS Jakarta - Indonesia

BAB I. UMUM

1.1. Peranan Sampling

Data dapat diperoleh baik melalui sensus maupun survei sampel, dengan sampling berpeluang (*probability sampling*) atau sampling tidak berpeluang (*non probability sampling*). Pengumpulan data dengan mendapatkan informasi dari semua elemen dalam populasi disebut sensus. Apabila informasi hanya dikumpulkan dari sebagian elemen disebut survei sampel. Bila informasi dari survei sampel ingin disajikan untuk menarik kesimpulan tentang sebuah populasi maka perlu digunakan sampling berpeluang.

Yang menjadi pertanyaan di sini adalah:

“Bagaimana menarik sampel-sampel elemen/kelompok elemen agar dapat menarik kesimpulan tentang sebuah populasi?”

Metode sampling dengan berbagai teknik akan menjawab pertanyaan ini dan memegang peran penting untuk dapat menyajikan estimasi .

Dalam buku ini akan dibahas berbagai metode dari penarikan sampel untuk menarik kesimpulan tentang sebuah populasi bila informasi yang tersedia berasal dari sebagian elemen dalam populasi.

Dalam buku ini akan diberikan rumus-rumus tanpa pembuktian dengan memberikan:

- a. Contoh sederhana untuk memudahkan pengertian
- b. Contoh dari aplikasi yang pernah dilakukan, dengan menerapkan berbagai metode dengan kerangka sampel yang tersedia. Dengan demikian, diharapkan dapat membantu untuk melihat efisiensi antar berbagai metode sampling.

Pembahasan dari metode sampling akan dimulai dari elemen sampling, kluster sampling sampai dengan sampling bertahap.

Rumus-rumus yang dibahas meliputi:

- a. Rumus estimasi.

- b. Varians (*variance*) dan kesalahan sampling (*sampling error*).
- c. Selang kepercayaan (*confidence interval*).
- d. Penentuan ukuran sampel.

1.2. Sampling Berpeluang dan Tidak Berpeluang

Penggunaan metode sampling selain dengan sampling berpeluang adalah dengan sampling tidak berpeluang.

Dengan sampling berpeluang diharapkan dapat menarik kesimpulan tentang sebuah populasi, sehingga informasi/data yang dihasilkan dapat mewakili keadaan populasi. Setiap unit dalam populasi harus mempunyai peluang untuk terpilih dalam sampel. Untuk keperluan penarikan sampel perlu adanya kerangka sampel. Oleh karena dalam sampling berpeluang setiap unit harus mempunyai peluang untuk terpilih ke dalam sampel, maka dimungkinkan untuk menghasilkan estimasi parameter dari populasi seperti total, rata-rata, proporsi, dsb. Nilai-nilai ini diperoleh dari hasil sampel.

Sampling tidak berpeluang tidak dapat menghasilkan estimasi tersebut yang disebabkan adanya unit yang tidak mempunyai peluang untuk terpilih dalam sampel. Meskipun demikian, sampling tidak berpeluang sering digunakan untuk keperluan penelitian khusus seperti pemasaran, opini publik, dsb. Penelitian semacam ini tidak diperuntukkan estimasi populasi, tetapi sangat diperlukan untuk keperluan kajian tertentu. Sampling tidak berpeluang ini tergolong survei kuota (*quota survey*). Contoh lain dari sampling tidak berpeluang adalah sampel yang ditentukan berdasarkan pikiran/keahlian tertentu (*purposive* atau *judgement sampling*).

Survei semacam ini misalnya dilakukan pada survei harga, dengan menentukan pasar yang diperkirakan mewakili konsumen dalam populasi. Pada tulisan selanjutnya akan dibahas metode sampling yang berkaitan dengan sampling berpeluang karena biasanya suatu survei sampel harus

menghasilkan estimasi yang mewakili populasi. Statistik yang dihasilkan harus dapat menghasilkan perkiraan dan kesalahan samplingnya.

1.3. Keuntungan dan Kelemahan Penggunaan Metode Sampling

Keuntungan dari penggunaan metode sampling antara lain adalah:

a. Menghemat biaya

Sudah jelas menghemat biaya karena data hanya dikumpulkan dari sebagian unit dalam populasi. Karena sampel, maka petugas yang dibutuhkan lebih sedikit. Menghemat biaya pencetakan, pelatihan, pencacahan, dan pengolahan. Sebagai contoh Survei Sosial Ekonomi Nasional (Susenas) secara nasional hanya mencakup sampel sekitar 240.000 rumahtangga (lebih kurang 1/200 populasi rumahtangga). Informasi dari hasil survei diharapkan dapat menyajikan data sampai tingkat kabupaten/kota.

b. Mempercepat hasil survei

Pada umumnya data dibutuhkan segera, sehingga berbagai perencanaan segera dapat dilakukan dan mengena. Dengan melakukan survei sampel maka pelaksanaan lapangan dan pengolahan tentunya akan jauh lebih cepat diselesaikan.

c. Cakupan materi lebih besar

Data yang diperlukan biasanya beragam dan cukup banyak, sehingga tidak mungkin dikumpulkan melalui pencacahan lengkap. Data yang dikumpulkan melalui sensus lengkap biasanya sangat terbatas. Variabel yang dicakup sangat dibatasi pada variabel dasar saja. Untuk cakupan yang besar dibutuhkan kualifikasi petugas serta pemahaman yang baik, sehingga survei sampel sangat mendukung untuk keperluan ini.

d. Akurasi lebih tinggi

Pada sensus jumlah petugas dan responden yang besar akan mengakibatkan tingkat kesalahan yang juga besar, terutama kesalahan yang diakibatkan bukan oleh teknik sampling yang lazim

disebut sebagai *non sampling error*. Kesalahan yang timbul sebagai akibat tidak bisa terpenuhi kualifikasi petugas yang baik. Dalam sensus lengkap petugas yang diperlukan cukup besar. Sedangkan dalam survei sampel petugas yang diperlukan terbatas sehingga kualifikasi petugas yang sesuai dapat dicari. Demikian juga pelatihan dapat lebih intensif. *Non-sampling Error* juga dapat diakibatkan oleh: format dan alur pertanyaan (dalam daftar isian) yang kurang baik, konsep/definisi yang kurang tepat, jawaban responden salah, dan kesalahan dalam proses pengolahan.

Sebaliknya penggunaan metode sampling mempunyai kelemahan antara lain:

a. Penyajian wilayah kecil

Penyajian sampai wilayah kecil seperti kecamatan dan desa dengan sampel terbatas tidak dapat dipenuhi. Dalam metode sampling diperlukan jumlah sampel yang sesuai dengan tingkat ketelitian yang dikehendaki, khususnya yang menyangkut kesalahan sampling.

b. Penyajian variabel langka/proporsi kecil

Survei sampel sulit/ tidak dapat menyajikan variabel langka yaitu variabel yang kejadiannya kecil dalam populasi (proporsi kecil). Sebagai contoh apabila ingin diketahui lahir mati, tenaga S2 menurut profesi dan sebagainya.

c. Tren data

Apabila data diperlukan secara berkala untuk mengukur perubahan yang sangat kecil dari satu periode ke periode berikutnya, kemungkinan sampel diperlukan cukup besar. Perubahan kecil tadi dipengaruhi kesalahan sampling.

d. Tidak tersedianya kerangka sampel

Tidak tersedianya kerangka sampel sehingga persyaratan sampling berpeluang tidak bisa terpenuhi. Atau biaya untuk pembentukan

kerangka sampel cukup tinggi, sehingga berpengaruh besar terhadap total biaya.

Keuntungan dan kelemahan tersebut perlu dipertimbangkan dengan seksama sebelum memulai merencanakan sampel survei dengan mengkaji sifat variabel yang akan diteliti. Sifat variabel yang diteliti tentunya sangat erat kaitannya dengan obyek dan tujuan dari survei.

1.4. Prinsip Kegiatan Survei Sampel

Dalam melaksanakan suatu sampel survei perlu mengikuti ketentuan-ketentuan dimulai dari perencanaan, persiapan, pelaksanaan lapangan, pengolahan sampai dengan penyajian/analisa. Demikian juga perlu memperhatikan kesalahan baik yang disebabkan oleh metode sampling (sampling error) yang diukur dengan standard error maupun bukan metode sampling (non sampling error). Keduanya perlu mendapat perhatian yang sama agar total error kecil. Kesalahan bukan sampling dipengaruhi oleh banyak hal seperti daftar isian yang ambisi, konsep-definisi atau klasifikasi kurang jelas, susunan/struktur daftar isian yang kurang baik, pewawancara yang kualifikasinya kurang baik, jawaban responden yang kurang baik, pengolahan dsb.

Pada tulisan ini hanya dibatasi pada metode sampling yang diharapkan dapat menyajikan data dengan tingkat ketelitian/kesalahan sampling tertentu pada suatu level penyajian. Tingkat penyajian misalnya kabupaten/kota, propinsi, atau nasional serta dibedakan daerah perkotaan dan pedesaan. Tingkat penyajian sangat menentukan besarnya sampel yang harus dipilih.

Oleh karena itu dalam kegiatan survei perlu lebih dahulu menentukan minimal hal-hal sebagai berikut:

- a. Obyek dan tujuan dari survei
- b. Populasi survei/populasi yang akan disampel
- c. Data yang akan dikumpulkan, dalam hal ini variabelnya

- d. Tingkat ketelitian yang dikehendaki
- e. Kerangka sampel dalam populasi (*population frame*) dan penarikan sampelnya
- f. Target populasi yang akan disajikan (estimasi)
- g. Inferensial yang berupa berbagai kajian dan analisis.

Ketujuh hal di atas sebagai dasar untuk mengkaji alternatif metode sampling yang dapat digunakan. Populasi survei dan populasi dari kerangka sampel sangat erat kaitannya.

Dari obyek dan tujuan survei serta data yang akan dikumpulkan dapat diketahui sifat variabel. Sifat dari variabel mungkin menyebar rata di populasi, mengelompok pada wilayah tertentu, dengan frekwensi cukup banyak, atau hanya sedikit dan sebagainya. Kondisi ini harus dapat tergambar dalam kerangka sampel, apabila ingin menentukan metode sampling yang efisien. Oleh karena itu masalah kerangka sampel agar dapat menerapkan metode sampling yang baik perlu mendapat perhatian penuh. Tingkat ketelitian yang dikehendaki atau ditentukan didasarkan pada kesepakatan dan tidak dapat sama untuk semua variabel.

1.5. Populasi, Unit, dan Sampel

Suatu set sampel yang dipilih dengan peluang diharapkan dapat menjelaskan karakteristik dari populasi dengan menggunakan metode sampling dan prosedur estimasinya. Sebelum mendalami metode sampling perlu memperhatikan berbagai istilah yang akan digunakan terutama yang berkaitan dengan estimasi.

a. Elemen

Elemen adalah unit yang digunakan untuk mendapatkan informasi. Elemen merupakan individu yang informasinya akan digunakan untuk menghasilkan statistik (inferensial). Elemen berupa unit analisis dan ditentukan sesuai obyek dan tujuan dari survei.

b. Populasi

Populasi merupakan satu kesatuan dengan elemen. Populasi merupakan agregasi dari seluruh elemen yang perlu ditentukan. Populasi harus didefinisikan dengan menentukan isi, unit, cakupan dan waktu. Sebagai contoh semua penduduk yang tinggal dalam rumah tangga biasa di Indonesia pada Februari 2003. Pendefinisian ditentukan sesuai dengan obyek dan tujuan, juga dari populasi dapat ditentukan sub-populasi seperti dalam Survei Angkatan Kerja (Sakernas). Dalam Sakernas dapat didefinisikan subpopulasi yaitu penduduk yang berumur 15 tahun keatas yang termasuk dalam angkatan kerja, yang tinggal dalam rumah tangga biasa di Indonesia pada bulan Februari 2003.

c. Unit Observasi

Unit observasi adalah unit dimana informasi diperoleh baik secara langsung maupun melalui responden tertentu.

d. Unit Sampling (*Sampling Unit*)

Unit sampling adalah unit yang dijadikan dasar penarikan sampel baik berupa elemen atau kumpulan elemen (klaster). Sebagai contoh rumah tangga dapat dijadikan unit sampling dan atau kumpulan rumah tangga pada wilayah tertentu yaitu blok sensus. Contoh lain dari unit sampling adalah daftar individu perusahaan/usaha.

e. Daftar Unit (*Listing Unit*)

Daftar unit adalah daftar unit yang digunakan untuk dasar penarikan sampel seperti direktori perusahaan/ usaha atau daftar rumah tangga dari blok sensus.

f. Kerangka Sampel (*Sampling Frame*)

Kerangka sampel merupakan seluruh unit dalam populasi yang akan dijadikan dasar penarikan sampel. Kerangka sampel ini perlu diteliti apakah memenuhi persyaratan kerangka sampel induk (master sampling frame) sehingga dapat dijadikan dasar penarikan sampel sesuai dengan metode sampling yang digunakan. Oleh karena

kerangka sampel merupakan hal yang sangat penting maka akan dibahas secara khusus dalam bab berikutnya.

1.6. Konsep dan Definisi Sampling

- a. Metode sampling dalam penerapannya digunakan untuk dapat menyajikan data yang mewakili populasi ditinjau dari keefisienannya dan hematnya biaya. Informasi yang dihasilkan dapat disajikan dan dianalisis secara statistik dan harus mempunyai tingkat ketelitian tertentu (*precise*) yang dapat diukur.

- b. Peluang terpilihnya setiap unit

Sampel harus dipilih sedemikian rupa, sehingga mewakili dengan baik populasi yang diteliti. Ketentuan ini berarti setiap unit harus mempunyai peluang untuk terpilih dalam sampel (besarnya peluang tidak boleh sama dengan nol). Besarnya peluang dapat sama (*equal probability*) atau tidak sama (*unequal probability*) tergantung dari metode sampling yang digunakan.

- c. Kebenaran yang diukur

Estimasi yang berasal dari sampel harus dapat diukur tingkat kebenarannya/ ketelitiannya yang berupa perkiraan nilai total, rata-rata dan sebagainya dengan menyajikan besarnya kesalahan sampling serta selang kepercayaannya. Pengukuran tingkat ketelitian ini disesuaikan dengan metode sampling yang digunakan. Pengukuran tingkat ketelitian dilakukan sebagai berikut:

- i. Pada saat perencanaan berdasarkan data yang ada dari sensus atau survei guna memperkirakan banyaknya sampel yang harus dipilih dengan tingkat ketelitian tertentu. Dipilih beberapa variabel pokok yang berkaitan dengan data yang akan dikumpulkan dari berbagai sumber.
- ii. Dari hasil lapangan dihitung besarnya perkiraan tingkat ketelitian / kesalahan sampling berdasarkan data dari survei.

Diharapkan hasil penghitungan kesalahan sampling antara rencana dan realisasi tidak terlalu jauh berbeda.

d. Keserasian

Desain sampling baru dapat diaplikasikan bila tersedia kerangka sampel dan metode yang dipilih dapat diaplikasikan di lapangan baik ditinjau dari segi unit samplingnya maupun biaya yang tersedia. Metode yang telah dipilih harus benar-benar diikuti dan tidak boleh diubah.

e. Efisiensi

Dalam uraian sebelumnya telah disinggung mengenai tingkat ketelitian dan biaya, maka keduanya harus seimbang. Data yang ada perlu dimanfaatkan semaksimal mungkin sehingga dalam penghitungan besarnya sampel dapat dipilih metode sampling yang menghasilkan varians dan biaya sekecil mungkin. Harapan juga dikehendaknya informasi yang dikumpulkan sebanyak mungkin, dengan tetap memperhatikan kemampuan responden untuk menjawabnya.

1.7. Estimasi

a. Statistik

Statistik/inferensial adalah nilai yang dihitung dari hasil survei sampel mengenai karakteristik, biasanya untuk tujuan membuat estimasi populasi. Karakteristik dapat berupa variabel yang berhubungan dengan keadaan elemen dalam populasi seperti umur, pendidikan, status pekerjaan dan sebagainya. Nilai yang dihitung dapat berupa angka mutlak, rata-rata, ratio, persentase, dan sebagainya atau kajian yang lebih mendalam seperti kajian karakteristik sub-populasi, deviasi standar dan sebagainya.

b. Rumus Estimasi

Rumus estimasi adalah rumus yang digunakan untuk memperkirakan nilai karakteristik populasi berdasar nilai yang diperoleh dari sampel.

Notasi yang digunakan adalah

Y : nilai karakteristik populasi (parameter populasi)
 \bar{Y} : rata-rata nilai karakteristik populasi
 \hat{Y} : perkiraan nilai karakteristik populasi dari sampel
 N : banyaknya seluruh unit sampling dalam populasi
 n : banyaknya unit sampling terpilih
 $\hat{\bar{Y}} = \bar{y}$: perkiraan rata-rata nilai karakteristik populasi dari sampel sebanyak n .

Dalam metode sampling diadakan kesepakatan bahwa huruf besar menunjukkan nilai populasi dan huruf kecil nilai sampel (atau huruf besar menggunakan tanda ^).

Sebagai gambaran secara umum parameter dari populasi dan statistik dari suatu sampel dituliskan sebagai berikut:

Penghitungan	Parameter dari Populasi	Statistik dari Sampel
Total	$Y = \sum_{i=1}^N Y_i$	$y = \sum_{i=1}^n y_i$
Rata-rata	$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{y}{n}$
Proporsi	$P_y = \frac{Y}{N}$	$p_y = \frac{y}{n}$
Varians	$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$	$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$
Deviasi Standar	$\sigma_y = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2}$	$s_y = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2}$
Varians Proporsi	$\sigma_y^2 = P_y (1 - P_y)$	$s_y^2 = \frac{n}{(n-1)} p_y (1 - p_y)$

Catatan:

N : banyaknya elemen unit dalam populasi

Y_i : nilai karakteristik dari elemen unit ke-i

n : banyaknya elemen unit dalam sampel

y_i : nilai karakteristik dari elemen unit ke-i dari sampel

Rumus di atas adalah rumus sederhana untuk sekedar menggambarkan nilai populasi dan nilai sampel, yang selanjutnya akan dikembangkan sesuai metode samplingnya.

c. Distribusi Sampling

Penarikan sampel dilakukan dengan menggunakan angka random. Dengan demikian sampel yang terpilih tergantung dari angka randomnya. Secara keseluruhan akan terbentuk sejumlah kemungkinan set sampel (*all possible samples*), yang banyaknya tergantung dari metode penarikan sampel yang digunakan. Distribusi dari set sampel yang mungkin terbentuk disebut distribusi sampling.

Sebagai contoh dengan metode yang sederhana, yaitu dengan memilih n dari sebanyak N unit secara acak (sampel acak sederhana) dan unit tidak boleh terpilih lebih dari sekali tanpa pemulihan, akan diperoleh kemungkinan sampel terbentuk sebanyak

$${}^N C_n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Misalkan banyak unit dalam populasi $N=6$ dan $n=2$ maka sampel yang mungkin terbentuk adalah

$$\binom{6}{2} = \frac{6!}{2!4!} = \frac{6 \times 5}{2 \times 1} = 15$$

Kombinasi sampel yang dapat dibentuk masing-masing terdiri dari dua unit, yang meliputi unit ke:

1,2 1,5 2,4 3,4 4,5
1,3 1,6 2,5 3,5 4,6
1,4 2,3 2,6 3,6 5,6

Dalam hal ini tidak boleh ada unit yang terulang. Selanjutnya estimasi diperoleh dari unit sampel yang terpilih, yaitu salah satu dari 15 kombinasi di atas.

Secara rinci akan dijelaskan dalam metode sampling.

d. Metode Sampling

Pembahasan akan mencakup berbagai metode sampling termasuk cara penarikan sampel.

Penarikan sampel terdiri dari tiga cara, yaitu:

- Acak sederhana (*simple random sampling*)
- Dengan peluang sebanding ukuran/bobot unit (*probability proportional to size*)
- Sistematis.

Sedangkan metode sampling yang akan dibahas adalah sampling berpeluang yang meliputi:

1. Peluang sama untuk semua elemen	1. Peluang tidak sama/berbeda untuk masing-masing elemen
2. Sampling Elemen: satu tahap, sampling unit langsung elemen	2. Sampling Klaster: klaster terdiri dari beberapa elemen <ul style="list-style-type: none"> - Ukuran klaster sama - Ukuran klaster tidak sama - Sampling satu tahap - Sampling bertahap
3. Sampling tidak berstrata	3. Sampling berstrata

Dalam pembahasan, teknik-teknik sampling tersebut dikembangkan dengan sampling proporsi, rasio dan sebagainya.

BAB II. KERANGKA INDUK DAN KERANGKA SAMPEL INDUK

2.1. Populasi

Seperti telah dibahas sebelumnya populasi survei perlu didefinisikan terlebih dahulu sesuai dengan obyek dan tujuan survei dengan melihat isi, unit, cakupan dan waktu. Untuk mengarah pada penggunaan metode sampling berpeluang, tahap selanjutnya adalah menentukan populasi dari kerangka induk (frame population).

Dalam populasi survei perlu dipertimbangkan berbagai variabel yang akan dicakup baik sifat-sifat variabel maupun kemungkinan adanya non respons seperti menolak, tidak mampu menjawab, tidak ditemui dan sebagainya. Langkah lanjut adalah meneliti kemungkinan populasi kerangka induk yang sesuai dan kemungkinan penggunaannya. Penentuan populasi kerangka induk perlu mengkaitkan dengan kemungkinan terjadinya salah cakupan, tidak lengkapnya kerangka sampel dan sebagainya. Hal ini akan mempengaruhi target populasi yang perlu disepakati sebelum melangkah lanjut. Target populasi ini akan mempengaruhi metode estimasi yang digunakan.

2.2. Kerangka Induk

Kerangka Induk (*Master Frame*) merupakan kerangka dasar yang harus menjadi perhatian utama sebelum penentuan desain sampling dan pembentukan Kerangka Sampel Induk (*Master Sampling Frame*). Apa yang disebut Kerangka Induk dan Kerangka Sampel Induk perlu dipahami benar-benar yang meliputi pembentukan, penggunaan, dan pemeliharannya. Kerangka Sampel Induk yang tidak memenuhi persyaratan akan mengakibatkan target populasi tidak tercapai. Persyaratan yang tidak terpenuhi akan mengakibatkan bias yang cukup besar terutama bila dikehendaki estimasi total yang dapat mewakili populasi. Penyediaan Kerangka Induk tidak mudah disamping harus memenuhi persyaratan untuk populasi yang dicakup, juga perlu dapat dipergunakan untuk jangka waktu cukup lama. Penyediaan kerangka

induk cukup mahal, tetapi tanpa kerangka induk yang baik tidak mungkin dibentuk kerangka sampel. Sensus merupakan salah satu sumber untuk pembentukan kerangka induk. Pembentukan kerangka induk tanpa diintegrasikan dengan pelaksanaan sensus akan sangat mahal. Oleh karena itu adanya peta desa, peta blok sensus, serta karakteristik di dalamnya yang mencakup populasi dari hasil sensus dimungkinkan dibentuk kerangka sampel induk. Berhasilnya penerapan metode sampling sangat tergantung pada kualitas dari hasil sensus tersebut. Akurasi dari sketsa peta desa, sketsa peta blok sensus dan data di dalamnya sangat dituntut karena akan digunakan sebagai dasar estimasi. Apalagi selama ini survei-survei di BPS seperti Survei Sosial Ekonomi Nasional (Susenas), Survei Angkatan Kerja Nasional (Sakernas), Survei Pertanian, Survei Usaha Terintegrasi (Susi) dikehendaki estimasi yang mewakili populasi.

2.3. Persyaratan

Kerangka sampel induk yang dibentuk dari kerangka induk harus dapat untuk penerapan penarikan sampel berpeluang. Persyaratan kerangka sampel induk adalah sebagai berikut:

- a. Tersedia sampai satuan unit terkecil yang digunakan sebagai dasar penarikan sampel.
- b. Mempunyai batas jelas
- c. Tidak tumpang tindih atau terlewat
- d. Mempunyai korelasi dengan data yang akan diteliti
- e. Keadaan mutakhir (*up to date*)

Persyaratan tersebut sangat diperlukan agar tidak terjadi :

- a. Unit sampel yang tidak dijumpai
- b. Unit sampel yang duplikasi
- c. Unit sampel yang terpecah
- d. Unit sampel tergabung
- e. Unit baru

Dengan demikian dapat dihindari estimasi bias yang disebabkan kerangka sampel.

2.4. Efek Bias

Data hasil survei sampel akan dikalikan dengan suatu faktor pengali sesuai metode sampling yang digunakan. Faktor pengali ini merupakan kebalikan dari peluang pada saat penentuan sampel. Sampel sangat sederhana, misalnya penggunaan peluang sama maka penghitungan estimasi adalah

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^n y_i$$

N : banyaknya blok sensus dalam populasi

n : banyaknya blok sensus terpilih

y_i : nilai karakteristik dari blok sensus ke i .

Apabila N ternyata ada yang duplikasi atau sebaliknya, dipecah dan sebagainya akan mempengaruhi faktor pengali, yaitu

$$F = \frac{N}{n} \text{ menjadi } F' = \frac{N'}{n}.$$

Bila perbedaan N dan N' cukup besar akan sangat mempengaruhi estimasi.

Permasalahan akan lebih meluas bila digunakan perkiraan rasio, peluang tidak sama, sampling bertahap dsb.

Pengaruh tersebut dapat digambarkan sebagai berikut:

a. Perkiraan rasio

$$\hat{Y} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \sum_{i=1}^N X_i = \frac{y}{x} X = \hat{R} X$$

Dua kondisi yang akan mempengaruhi estimasi yaitu

$\sum_{i=1}^N X_i$: Merupakan total nilai karakteristik populasi dari kerangka sampel

- Bias pertama yang akan terjadi bila N tidak memenuhi syarat
- Bias kedua yang disebabkan oleh X_i , x_i dan y_i .

b. Sampling dengan peluang sebanding ukuran sampel

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{p_i} ; \text{ dengan } p_i = \frac{X_i}{X}$$

X_i dan X data dari kerangka sampel yang akan sangat menentukan estimasi.

c. Sampling dua tahap, keduanya dengan peluang sama

$$\hat{Y} = \frac{N}{n} \sum \frac{M_i}{m_i} \sum y_{ij}$$

N dan M_i adalah unit sampling dalam kerangka sampel yang akan sangat mempengaruhi estimasi. Misalnya M_i adalah banyaknya rumahtangga dalam blok sensus. Bila listing (pendaftaran rumahtangga untuk pembentukan unit sampling) tidak benar, maka estimasi juga tidak akan mewakili populasi. Oleh karena itu pelaksanaan listing perlu mendapat perhatian. Listing merupakan bagian juga dari kerangka sampel induk, yang dibentuk di lapangan bersamaan dengan pelaksanaan survei. Bila data listing tidak akurat akan menjadikan bias yang disebabkan kerangka sampel cukup besar. Oleh karena itu keberadaan kerangka sampel dan datanya perlu diadakan pengkajian antara lain:

- Ketersediaannya sebelum digunakan dengan meneliti kelengkapan persyaratan yang harus dipenuhi
- Kemungkinan perubahan di lapangan dan antisipasinya.

Setelah yakin kebenarannya baru dapat dijadikan sebagai kerangka sampel.

2.5. Penggunaan dan Pemeliharaan Kerangka Sampel

Penggunaan kerangka induk sebagai kerangka sampel induk perlu memperhatikan persyaratan-persyaratan yang harus dipenuhi. Kerangka sampel induk yang dibentuk diusahakan agar dapat dimanfaatkan untuk berbagai jenis survei, antara lain dengan cara:

- a. Menetapkan unit yang dapat dijadikan sampling unit untuk berbagai survei
- b. Mengisi unit tersebut dengan berbagai karakteristik dari berbagai sumber.
- c. Menghitung koefisien korelasi dari berbagai jenis variabel yang dapat dijadikan dasar untuk berbagai jenis survei.
- d. Menetapkan berbagai jenis variabel dasar yang dapat digunakan sebagai kajian untuk menentukan metode sampling.

Unit yang dijadikan unit sampling perlu dipelihara dan harus bersifat khas (*unique*) serta secara terus menerus dipantau untuk mengetahui apabila terjadi perubahan. Bila perubahan cukup signifikan maka perlu diadakan penyesuaian. Penyesuaian dapat melalui penyesuaian unit sampling atau metode sampling.

Salah satu contoh adalah penggunaan blok sensus sebagai unit sampling. Blok sensus digunakan untuk berbagai survei khususnya dengan pendekatan rumah tangga. Dengan demikian blok sensus memegang peranan penting sebagai kerangka sampel dan perlu mendapat perhatian akurasi. Data dari berbagai sumber seperti Sensus Penduduk, Sensus Pertanian dan sebagainya dapat digunakan untuk mengkaji bagaimana karakteristik dari blok sensus tersebut. Data ini dipergunakan sebagai dasar stratifikasi, penimbang, ratio dan sebagainya.

Persyaratan kerangka sampel bila dilaksanakan dengan baik akan sangat menunjang estimasi yang dihasilkan.

Dalam pemeliharaan kerangka sampel pertanyaan di bawah ini perlu menjadi pemikiran:

- a. Bagaimana menjadikan kerangka sampel konsisten dan unik serta mekanisme yang baik dalam pengelolaannya?
- b. Bagaimana mekanisme penyesuaian dan cara pembenahannya?

- c. Bagaimana mekanisme penggunaannya serta tanggung jawab pengelola dan pengguna?

2.6. Jenis Kerangka Sampel

Kerangka sampel yang dapat dibentuk dapat dikategorikan sebagai berikut:

- a. Dalam bentuk daftar individu, yang langsung dapat digunakan sebagai dasar penarikan sampel seperti daftar perusahaan industri besar/sedang atau direktori perusahaan sektor lainnya.
- b. Dalam bentuk daftar wilayah seperti halnya blok sensus. Kerangka sampel dalam bentuk wilayah harus dilengkapi dengan peta.

Dengan adanya persyaratan kerangka sampel, maka perlu diperhatikan hal-hal sebagai berikut:

- a. Kerangka Sampel yang berupa wilayah pembentukannya tentunya sudah memenuhi persyaratan batas jelas sehingga dapat dipertahankan dalam kurun waktu cukup lama. Kerangka sampel induk dapat dibentuk dengan melengkapi datanya.
- b. Kerangka sampel yang berupa individu seperti perusahaan/usaha lebih labil dan cepat berubah. Individu sebagai kerangka induk perlu ditentukan dengan jelas batasannya. Dengan batasan yang jelas baru dapat dibentuk kerangka sampel dan sekaligus melengkapi datanya. Untuk membentuk sampling unit individu tidak mudah, walaupun terbatas pada perusahaan/usaha besar. Perusahaan industri perubahan tidak terlalu cepat dan biasanya nyata di lapangan. Sedangkan sektor konstruksi misalnya perubahannya cepat dan sulit diikuti.

Kerangka sampel yang berupa wilayah, pada penerapannya akan terkait dengan sampling klaster, di mana didalamnya akan terdiri dari unit-unit yang lebih kecil, katakan elemen. Sebagai contoh blok sensus didalamnya bisa terdiri dari rumahtangga, penduduk, usaha atau luas lahan.

BAB III. PENARIKAN SAMPEL ACAK SEDERHANA

3.1. Pendahuluan

Penarikan sampel acak sederhana (PSAS) ialah suatu penarikan sampel yang dilakukan terhadap elemen-elemen di dalam populasi yang telah didefinisikan. Unit penarikan sampelnya berupa unit-unit yang terdaftar dalam kerangka sampel itu sendiri. Oleh karena itu PSAS adalah suatu prosedur penarikan sampel yang paling sederhana. Prosedur-prosedur penarikan sampel lainnya dapat dipandang sebagai suatu modifikasi atau bentuk pengembangan dari PSAS, yang dimaksudkan untuk menyesuaikan dengan kondisi dan situasi nyata yang dihadapi ketika akan melakukan penarikan sampel terhadap suatu populasi tertentu. Modifikasi dan pengembangan dapat dipertimbangkan karena adanya persyaratan yang tidak dapat dipenuhi dalam mengoperasikan PSAS atau didasarkan atas satu atau beberapa tujuan lain. Misalnya, yang berhubungan dengan alasan kepraktisan atau keterlaksanaan, biaya, tersedianya beberapa tambahan keterangan tentang unit dalam populasi, dan tentu saja untuk memperoleh dugaan yang lebih teliti dan tepat.

PSAS pada umumnya digunakan bila kita dihadapkan pada situasi-situasi, *pertama*, sangat terbatasnya pengetahuan terhadap unsur-unsur populasi. Keterangan sebelumnya yang lebih rinci dan diperlukan untuk menilai derajat keseragaman atau untuk menggolong-golongkan unsur-unsur populasi tidak diperoleh. *Kedua*, dari pengetahuan yang ada atau pengalaman selama ini, belum diperoleh suatu prosedur seleksi tandingan yang lebih efisien daripada PSAS.

Beberapa syarat yang perlu dipenuhi dalam mempergunakan PSAS ialah:

- a. Tersedianya suatu daftar kerangka sampel yang cermat dan lengkap yang mencakup seluruh elemen populasi. Jika kerangka sampel yang memuat elemen-elemen yang dimaksudkan belum tersedia, maka harus dibuat lebih dahulu sebelum penarikan sampel dapat dilakukan.

- b. Untuk variabel-variabel tertentu yang akan diamati, populasi data dapat dianggap bersifat cukup seragam. Penyimpangan yang besar terhadap anggapan ini akan memperbesar kemungkinan terjadinya bias akibat penarikan sampel.
- c. Dalam praktek penarikan sampel yang (langsung atau tak langsung) menyangkut soal geografis atau keruangan sebaran elemen-elemen populasi tidak terlalu terpencar-pencar dalam areal yang luas.

3.2. Prosedur Penarikan Sampel

Ada dua cara dalam seleksi unsur-unsur untuk suatu sampel PSAS, yaitu seleksi elemen dengan pemulihan (*without replacement-wor*) dan seleksi elemen tanpa pemulihan (*with replacement-wr*). Misalnya suatu populasi berukuran N dan ukuran sampel sebesar n , dengan penarikan sampel tanpa pemulihan akan diperoleh salah satu dari

$${}^N C_n = \frac{N!}{n!(N-n)!}$$

macam set sampel yang mungkin terbentuk, masing-masing terdiri atas n elemen yang berbeda.

Sedangkan dalam penarikan sampel dengan pemulihan akan diperoleh satu gugus sampel diantara N^n yang mungkin dapat dibentuk.

Langkah-langkah penarikan sampel yang harus dilakukan adalah sebagai berikut:

- a. Siapkan Tabel Angka Random (TAR) yang terdiri misalnya dua halaman. Setiap halaman terdiri atas 40 kolom dan 30 baris. Masing-masing halaman diberi nomor kolom 1, 2, 3,40 dan nomor baris 1, 2, 3,30.
- b. Tentukanlah kelipatan-kelipatan dari bilangan $c \times N$ (N adalah ukuran populasi) yang terdiri atas k digit (N terdiri atas k digit). Gunakan kelipatan-kelipatan bilangan $c \times N$ yang jumlah digitnya sama dengan k .
- c. Ambil sebuah pensil atau benda berujung tajam. Buka salah satu halaman dari dua halaman Daftar Bilangan Acak ini. Untuk keperluan ini dapat digunakan sembarang halaman. Picingkan mata atau alihkan pandangan ke tempat lain, dan letakkan ujung pensil di atas lembaran

Tabel Angka Random. Bilangan yang paling dekat dengan posisi ujung pensil adalah merupakan titik awal pembacaan angka random untuk menentukan halaman, baris, dan kolom yang akan digunakan untuk memilih sampel. Mulai dari titik ini bacalah beberapa angka ke kanan sesuai dengan keperluan.

Misalkan halaman yang digunakan untuk pembacaan ini adalah halaman pertama Tabel Angka Random ujung pensil jatuh pada bagian tertentu dari tabel seperti pada ilustrasi berikut :

Baris	Kolom				
	(11-15)		
...					
...					
12		. 93465			
...					

Lima angka di sebelah kanan tanda titik (.) adalah 93465

- Penentuan halaman pembacaan TAR

Karena ada 2 halaman TAR, angka random yang digunakan untuk menentukan halaman ini cukup satu angka saja. Untuk mudahnya, gunakan angka 0, 1, 2,,9 dengan perjanjian bahwa angka ganjil untuk menyatakan halaman pertama TAR, angka 0 (nol) dan genap untuk halaman kedua. Pada pembacaan di atas, yaitu 93465, digit pertama adalah 9 (bilangan ganjil). Oleh karena itu halaman yang terpilih adalah halaman pertama dari TAR.

- Penentuan baris

Karena pada setiap halaman ada 30 baris, maka untuk penentuan baris ini digunakan bilangan yang terdiri atas 2 digit. Untuk penghematan, ditentukan perjanjian bahwa bilangan 01, 31, dan 61 digunakan untuk menyatakan baris 1, bilangan 02, 32, dan 62 digunakan untuk menyatakan baris 2, dan seterusnya. Pada

pembacaan di atas, digit ke-2 dan ke-3 adalah 34, maka baris pembacaan jatuh pada baris ke-4, karena $34 - 30 = 4$

- Penentuan kolom

Karena pada setiap halaman ada 40 kolom, maka untuk penentuan kolom ini digunakan bilangan yang terdiri atas 2 digit. Untuk penghematan, ditentukan perjanjian bahwa bilangan 01, 41, dan 81 digunakan untuk menyatakan baris 1, bilangan 02, 42, dan 82, digunakan untuk menyatakan baris 2, dan seterusnya.

Pada pembacaan di atas, digit ke-4 dan ke-5 adalah 65, maka baris pembacaan jatuh pada kolom ke-25, karena $65 - 40 = 25$.

- Bukalah halaman pertama TAR. Bacalah k digit bilangan (ke kanan) yang dimulai dari pada baris ke-4 kolom ke-25 halaman tersebut. Bila k digit bilangan (misalnya R) yang tercantum pada TAR lebih besar dari N , maka angka random pemilihan sampel yang digunakan adalah $R - c.N$ dimana c adalah bilangan integer (1, 2, ...9). Nomor urut unit sampel yang sama dengan angka random yang terpilih adalah merupakan unit yang terpilih.

3.3. Estimasi parameter populasi

a. Notasi

N	ukuran populasi
n	ukuran sampel
Y	variabel yang diamati
y_i	nilai variabel Y pada pengamatan ke- i
$\frac{n}{N}$	fraksi sampling (<i>sampling fraction</i>)
$\mu = \bar{Y}$	rata-rata populasi
\bar{y}	rata-rata sampel
σ^2	varians populasi
s^2	varians sampel
P	proporsi populasi
p	proporsi sampel

b. Rata-rata (\bar{y})

Misalkan elemen-elemen populasi yang terseleksi untuk suatu sampel acak sederhana tanpa pemulihan ataupun dengan pemulihan yang berukuran n ialah u_1, u_2, \dots, u_n . Untuk suatu variabel, Y , sampel yang diperoleh itu misalnya mengandung nilai y_1, y_2, \dots, y_n . Statistik rata-rata dan varians yang didasarkan pada data sampel adalah merupakan nilai-nilai estimasi bagi rata-rata dan varians populasinya. Estimator tak bias bagi rata-rata dan varians dapat adalah sebagai berikut :

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

Dalam PSAS kita hanya mampu merealisasikan satu set sampel yang berukuran n . Oleh karena varians dari statistik yang dihasilkan harus diduga hanya berdasarkan satu set data sampel. Estimator varians dan standard error bagi rata-rata dibedakan menjadi dua sesuai dengan prosedur penarikan sampelnya, yaitu :

i. PCAS tanpa pemulihan

$$v(\bar{y}) = \frac{(N-n)}{N} \frac{s^2}{n}$$
$$se(\bar{y}) = \sqrt{\frac{(N-n)}{N} \frac{s^2}{n}}$$

ii. PCAS dengan pemulihan

$$v(\bar{y}) = \frac{s^2}{n}$$
$$se(\bar{y}) = \frac{s}{\sqrt{n}}$$

Dengan mengasumsikan, bahwa rata-rata dari seluruh kemungkinan sampel menyebar normal, maka selanjutnya dapat disusun nilai estimasi selang kepercayaan (*interval estimate*) bagi rata-rata.

Estimasi selang kepercayaan $(1 - \alpha).100\%$ bagi rata-rata yang sebenarnya adalah sebagai berikut :

$$\bar{y} - z_{\alpha/2} se(\bar{y}) < \mu < \bar{y} + z_{\alpha/2} se(\bar{y})$$

Estimasi selang kepercayaan adalah merupakan estimasi dalam bentuk selang antara dua nilai terendah dan tertinggi dengan tingkat kepercayaan sebesar $(1 - \alpha).100\%$. Batas terendah (*lower bound*) dari suatu estimasi selang adalah sebesar $L(\mu) = \bar{y} - z_{\alpha/2} se(\bar{y})$, sedangkan batas tertingginya (*upper bound*) adalah sebesar $U(\mu) = \bar{y} + z_{\alpha/2} se(\bar{y})$. Bila tingkat kepercayaan yang dikehendaki sebesar 95 %, maka besarnya $\alpha = 0,05$ dan $z_{\alpha/2} = 1,96$

c. Total

Dalam suatu penelitian, disamping ditujukan untuk memperkirakan rata-rata juga ditujukan untuk memperkirakan total suatu karakteristik tertentu. Besarnya nilai estimasi total adalah merupakan perkalian antara ukuran populasi (N) dan estimasi rata-rata (\bar{y}).

Mengingat penduga \bar{y} memiliki sifat yang tak bias, maka selanjutnya dapat digunakan untuk memperkirakan nilai total populasi.

Estimator tak bias bagi total adalah

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

dengan kesalahan sampling (*sampling error*)

$$se(\hat{Y}) = N se(\bar{y})$$

Dengan mengasumsikan, bahwa total dari seluruh kemungkinan sampel terbentuk menyebar normal, maka selanjutnya dapat disusun nilai perkiraan selang bagi total. Estimasi selang dengan tingkat kepercayaan $(1 - \alpha).100\%$ bagi nilai total yang sebenarnya adalah

$$\hat{Y} - z_{\alpha/2} N se(\bar{y}) < Y < \hat{Y} + z_{\alpha/2} N se(\bar{y})$$

d. Proporsi

Suatu variabel berskala ukur nominal atau ordinal adalah merupakan variabel kualitatif, tidak mempunyai satuan ukur, dan terdiri atas kategori-kategori yang kongkrit. Analisis yang layak dan lumrah untuk variabel semacam ini adalah didasarkan pada frekuensi-frekuensi atau proporsi-proporsi dari kategori-kategori.

Misalkan, Y_1, Y_2, \dots, Y_n ialah suatu populasi data untuk suatu variabel kualitatif. Dalam populasi berukuran N tersebut, masing-masing terdapat N_1, N_2, \dots, N_k elemen yang termasuk dalam kategori-1, kategori-2, ... dan kategori- k , sedangkan $N_1 + N_2 + \dots + N_k = N$. Frekuensi nisbi atau proporsi unsur dari kategori- g dalam populasi

data tersebut adalah $P_g = \frac{N_g}{N}$, untuk sembarang $g = 1, 2, \dots, k$. Jadi,

dalam menghitung proporsi kategori- g tersirat suatu pandangan memilah elemen-elemen populasi menjadi dua golongan, yaitu yang berciri kategori- g dan yang berciri bukan kategori- g . Yang bukan kategori- g dapat terdiri atas lebih dari satu kategori.

Elemen-elemen dengan ciri termasuk kategori- g (yang diperhatikan) masing-masing diberi nilai *dummy* sama dengan 1, sedangkan untuk kategori selainnya diberi nilai 0. Dengan pengertian ini, maka kita dapat memandang P_g sebagai suatu rata-rata populasi data variabel kategori- g . Jadi, P_g dapat dicatat sebagai $P_g = N_g/N = Y = Y/N$, sehingga $N_g = NP_g$. Dengan pandangan seperti ini, maka pernyataan tentang varians populasi untuk variabel *dichotomous* (atau yang dilandaskan pada suatu kategori tertentu yang lagi diperhatikan) dapat dinyatakan secara umum, seperti layaknya untuk suatu variabel kuantitatif. Sehingga varians populasi terhitung adalah

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - Y)^2$$

Untuk suatu variabel kategorik adalah sama dengan

$$S^2 = \frac{N}{N-1} P_g (1 - P_g)$$

Untuk simplifikasi, P_g cukup dicatat dengan P dan $1-P_g$ (yaitu frekuensi nisbi unsur-unsur dalam populasi yang bukan berciri kategori- g) dicatat sebagai Q . Sehingga persamaan dapat dicatat sebagai

$$S^2 = \frac{N}{N-1} P Q$$

Setara dengan pernyataan dalam persamaan-persamaan yang tercantum sebelumnya, maka varians dari suatu sampel PSAS dapat dicatat sebagai

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n}{n-1} p q$$

Penduga varians dan standard error bagi proporsi adalah sebagai berikut :

$$v(p) = \frac{(N-n)}{N} \frac{pq}{n}$$

$$se(p) = \sqrt{\frac{(N-n)}{N} \frac{pq}{n}}$$

Estimasi selang kepercayaan $(1 - \alpha).100\%$ bagi proporsi yang sebenarnya adalah

$$p - z_{\alpha/2} se(p) < P < p + z_{\alpha/2} se(p)$$

3.4. Penentuan ukuran sampel

Apabila dalam mengumpulkan data untuk menduga parameter populasi suatu variabel, dan pengaruh kesalahan bukan karena teknik sampling (*nonsampling error*) sepenuhnya dapat dihindarkan, maka simpangan atau beda antara nilai dugaan dan nilai parameter yang sebenarnya (*true value*) dapat dipandang sebagai suatu kesalahan sampling. Memperoleh suatu kesalahan sampling yang sekecil-kecilnya tetapi dalam batas-batas keterlaksanaan yang masih dapat diterima adalah menjadi salah satu tujuan dalam survei sampel. Karena parameter populasi tidak diketahui, maka lokasi hasil estimasi, yang diperoleh dari suatu sampel acak tertentu

terhadap nilai parameternya juga tidak diketahui dengan tepat. Lokasinya mungkin berada di kanan, di kiri, atau tepat pada lokasi parameter. Selisih antara nilai dugaan yang diperoleh dari suatu sampel acak terhadap nilai parameter selanjutnya disebut presisi (d), sedangkan batas-batas keterlaksanaan yang masih dapat diterima disebut sebagai tingkat keyakinan (z), dan standard error dari suatu statistik $se(\hat{\theta})$, maka secara statistik akan terdapat hubungan sebagai berikut :

$$d = z \cdot se(\hat{\theta})$$

$$presisi = tingkat\ keyakinan \times standard\ error$$

Untuk $z = 2$ berarti tingkat keyakinan yang dikehendaki sebesar 95 %.

Dari bentuk hubungan di atas selanjutnya dapat diturunkan formula untuk penentuan ukuran sampel yang dibutuhkan untuk menduga suatu parameter dengan presisi dan tingkat reliabilitas yang dikehendaki.

- a. Ukuran sampel yang dibutuhkan untuk menduga rata-rata dengan presisi d dan tingkat keyakinan $(1 - \alpha).100\%$ adalah

$$d = z_{\alpha/2} \sqrt{\frac{(N-n)S^2}{Nn}} \quad \text{sehingga} \quad n = \frac{N(z_{\alpha/2}S)^2}{Nd^2 + (z_{\alpha/2}S)^2}$$

- b. Ukuran sampel yang dibutuhkan untuk menduga total dengan presisi d' dan tingkat keyakinan $(1 - \alpha).100\%$ adalah

$$d' = z_{\alpha/2} N \sqrt{\frac{(N-n)S^2}{Nn}} \quad \text{sehingga} \quad n = \frac{N^2(z_{\alpha/2}S)^2}{d'^2 + N(z_{\alpha/2}S)^2}$$

- c. Ukuran sampel yang dibutuhkan untuk menduga proporsi dengan presisi d^* dan tingkat keyakinan $(1 - \alpha).100\%$ adalah

$$d^* = z_{\alpha/2} \sqrt{\frac{(N-n)PQ}{Nn}} \quad \text{sehingga} \quad n = \frac{Nz_{\alpha/2}^2 PQ}{Nd^{*2} + (z_{\alpha/2}PQ)^2}$$

Catatan: Bila nilai-nilai parameter S , P dan Q tidak tersedia, maka dapat digunakan nilai-nilai dugaannya, yaitu : s , p , dan q untuk variabel yang relevan dengan survei yang akan dilaksanakan yang diperoleh dari survei yang terdahulu.

BAB IV. Penarikan Sampel Acak Berstrata (*Stratified Random Sampling*)

4.1. Pendahuluan

Pada penarikan sampel acak sederhana, kita tidak memaksakan supaya suatu sampel mewakili populasinya. Kecenderungan untuk mewakili sudah dikandung oleh prosedur itu sendiri dan galat penarikan sampel dapat ditekan dengan menaikkan ukuran sampel. Akan tetapi, bila sebelumnya sudah diketahui suatu informasi mengenai populasinya, informasi tersebut mungkin dapat digunakan sebagai dasar pembentukan strata sehingga kesalahan sampling dapat ditekan.

Penarikan sampel berstrata adalah suatu metode dimana populasi, yang berukuran N , dibagi-bagi menjadi subpopulasi-subpopulasi yang masing-masing terdiri atas $N_1, N_2, N_3, N_4, \dots, N_L$ elemen. Diantara dua subpopulasi tidak boleh ada yang saling tumpang tindih, sehingga $N_1 + N_2 + N_3 + N_4 + \dots + N_L = N$. Selanjutnya setiap anak populasi disebut sebagai strata (*stratum*).

Dalam pembentukan strata harus diusahakan agar elemen-elemen yang hampir sama dimasukkan ke dalam satu strata sehingga varians di dalam masing-masing strata menjadi lebih homogen. Di samping itu, akan lebih baik lagi jika perbedaan rata-rata karakteristik antar strata dibuat sebesar mungkin perbedaannya. Secara skematis pembentukan strata disajikan pada Gambar 4.1.

Selanjutnya jika strata telah terbentuk kita bisa menarik sampel dari masing-masing strata secara terpisah (*independent*). Ukuran sampel yang dapat ditarik masing-masing strata adalah sebesar $n_1, n_2, n_3, n_4, \dots, n_L$. Jika penarikan sampel pada setiap strata dilakukan dengan metode penarikan sampel acak, maka prosedur ini disebut sebagai penarikan sampel acak berstrata.

Kenuntungan-keuntungan dari penerapan penarikan sampel berstrata adalah sebagai berikut :

- Dengan sampel berstrata dapat diperoleh nilai estimasi dengan presisi lebih tinggi, baik untuk setiap strata maupun untuk populasi secara keseluruhan;
- Pada setiap strata dapat dipergunakan rancangan penarikan sampel yang berbeda, tergantung keadaan setiap strata dan kebutuhannya,
- Setiap strata dapat dianggap sebagai populasi tersendiri, sehingga bisa saja menentukan presisi yang dikehendaki pada setiap strata, dan disajikan tersendiri
- Dalam beberapa hal membawa manfaat pada pengelolaan administrasi,

Populasi											
♦	♠	♦	♥	♥	♠	♥	♣	♦	♠	♦	♣
♠	♦	♠	♣	♠	♥	♠	♦	♠	♠	♣	♠
♣	♥	♣	♦	♣	♦	♣	♥	♣	♣	♥	♦
♥	♣	♥	♠	♥	♣	♥	♥	♠	♦	♥	♥
Bentuk gambar adalah merupakan ciri dari elemen populasi											



Populasi distratafikasi											
I			II			III			IV		
♦	♦	♦	♠	♠	♠	♣	♣	♣	♥	♥	♥
♦	♦	♦	♠	♠	♠	♣	♣	♣	♥	♥	♥
♦	♦		♠	♠	♠	♣	♣	♣	♥	♥	♥
♦	♦		♠	♠	♠	♣	♣	♣	♥	♥	♥

Gambar 4.1

Skema Pembentukan Strata

Yang perlu diperhatikan dalam penerapan rancangan penarikan sampel berstrata adalah variabel apa yang digunakan sebagai dasar pembentukan strata, alokasi sampel pada masing-masing strata, dan ukuran sampel yang diperlukan untuk menduga suatu statistik dengan presisi yang dikehendaki.

4.2. Variabel untuk pembentukan strata

Sebagai langkah awal penerapan rancangan penarikan sampel berstrata adalah menentukan variabel yang digunakan sebagai dasar pembentukan strata. Untuk survei sampel yang hanya mempelajari satu variabel saja, misalnya Y , maka variabel terbaik yang digunakan sebagai dasar stratifikasi adalah variabel Y itu sendiri. Pada kenyataannya hal tersebut jarang terjadi, karena survei sampel biasanya karakteristik-karakteristik yang dipelajari sangat rinci dan masing-masing variabel memiliki derajat kegunaan yang sama. Untuk mengatasi hal tersebut, cara terbaik sebelum menentukan variabel mana yang digunakan sebagai dasar pembentukan strata terlebih dahulu dipelajari besarnya korelasi antar variabel yang

dipelajari. Variabel yang digunakan sebagai dasar pembentukan strata adalah variabel yang memiliki korelasi yang erat dengan variabel-variabel yang diteliti.

Variabel yang digunakan sebagai dasar pembentukan strata dapat berupa variabel yang memiliki skala pengukuran selang atau katagorik. Misalnya survei industri, variabel yang berkorelasi erat dengan produksi adalah jumlah tenaga kerja; survei produktifitas padi, maka variabel kategorik yang berpotensi untuk digunakan sebagai dasar pembentukan strata adalah variabel letak geografis desa/kelurahan (pantai, dataran tinggi/pegunungan, daerah aliran sungai/DAS, dan dataran rendah).

4.3. Notasi

Misalkan suatu populasi yang berukuran N dibagi menjadi L strata dan penarikan sampel dilakukan pada setiap strata secara acak sederhana tanpa pemulihan. Selanjutnya perhatikan notasi-notasi berikut :

Y_{hi}	nilai karakteristik Y unit ke- i pada strata ke- h
N_h	jumlah unit pada strata ke- h
n_h	ukuran sampel pada strata ke- h
$f_h = \frac{n_h}{N_h}$	fraksi sampel pada strata ke- h
$W_h = \frac{N_h}{N}$	penimbang pada strata ke- h
$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$	rata-rata nilai karakteristik pada strata ke- h (berdasarkan N_h unit)
$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$	rata-rata nilai karakteristik sampel pada strata ke- h (berdasarkan n_h unit)
$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$	varians pada strata ke- h
$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$	varians sampel pada strata ke- h

4.4. Estimasi Rata-rata Populasi \bar{Y}

Suatu populasi dibagi dalam L strata, maka estimator untuk rata-ratanya adalah merupakan penjumlahan tertimbang dari rata-rata pada masing-masing strata.

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{hi} = \frac{1}{N} \sum_{h=1}^L N_h \bar{Y}_h = \sum_{h=1}^L W_h \bar{Y}_h$$

Bila pada setiap strata ditarik sampel secara bebas antar strata yang masing-masing berukuran n_h , maka rata-rata sampel pada strata ke- h adalah \bar{y}_h yang merupakan estimator tak bias bagi \bar{Y}_h . Dengan demikian, maka estimasi rata-rata \bar{Y} merupakan penjumlahan tertimbang dari rata-rata sampel masing-masing strata, yaitu :

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^L W_h \bar{y}_h$$

dengan varians penarikan sampel

$$V(\hat{\bar{Y}}_{st}) = \sum_{h=1}^L W_h^2 \frac{(N_h - n_h)}{N_h} \frac{S_h^2}{n_h} = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

Karena s_h^2 merupakan estimator tak bias bagi S_h^2 , maka estimasi bagi varians $\hat{\bar{Y}}_{st}$ yang dihitung berdasarkan data sampel adalah :

$$\hat{V}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^L W_h^2 \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h} = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{s_h^2}{n_h}$$

4.5. Estimator Proporsi

Misalkan suatu populasi dengan N elemen dibagi menjadi L strata sedemikian rupa sehingga $N_1 + N_2 + N_3 + N_4 + \dots + N_L = N$, dan y_{hi} adalah nilai variabel kualitatif Y dalam strata ke- h pada unit ke- i . Elemen-elemen dengan ciri termasuk dalam kategori \mathfrak{R} masing-masing diberi nilai 1, sedangkan untuk kategori lainnya diberi nilai 0. Proporsi elemen-elemen dalam strata ke- h yang termasuk kategori \mathfrak{R} adalah

$$P_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$$

dan variansnya adalah

$$S_h^2 = \frac{N_h}{N_h - 1} P_h (1 - P_h) = \frac{N_h}{N_h - 1} P_h Q_h$$

Proporsi elemen-elemen dalam populasi yang termasuk \mathfrak{K} kategori dapat dinyatakan sebagai

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

Pada masing-masing strata ditarik sampel berukuran n_h , maka estimator tak bias bagi P dan estimator varians penarikan sampel adalah sebagai berikut :

$$p_{st} = \sum_{h=1}^L \frac{N_h}{N} p_h \text{ dengan } p_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

dan

$$V(p_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h^2 \frac{N_h - n_h}{N_h} \frac{N_h}{N_h - 1} \frac{P_h Q_h}{n_h}$$

Dugaan tak bias bagi varians proporsi yang didasarkan pada data sampel adalah

$$v(p_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h^2 \frac{N_h - n_h}{N_h} \frac{N_h}{N_h - 1} \frac{p_h q_h}{n_h}$$

4.6. Alokasi Sampel

Penarikan sampel acak berstrata tidak mengharuskan ukuran sampel yang sama pada setiap strata. Selama setiap strata paling sedikit diwakili satu elemen ketentuan penarikan sampel acak berlapis sudah terpenuhi. Jika minimum ada dua elemen, maka kita dapat menduga rata-rata dan variansnya. Pada umumnya ukuran sampel pada setiap strata lebih dari dua elemen. Perlu ditentukan kriteria yang akan digunakan dalam menentukan ukuran sampel secara keseluruhan dan alokasi pada masing-masing strata.

Ada beberapa metode mengalokasikan sampel kedalam setiap strata, yaitu alokasi sembarang, alokasi sama (*equal allocation*), alokasi sebanding (*proportional allocation*), alokasi optimum (*optimum allocation*). Penerapan dari masing-masing metode tergantung pada ketersediaan informasi awal mengenai strata yang telah terbentuk.

a. Alokasi sembarang

Misalkan suatu populasi yang berukuran N dibagi-bagi kedalam L strata, sedemikian rupa sehingga $N_1 + N_2 + N_3 + N_4 + \dots + N_L = N$ dan total ukuran n dialokasikan kesetiap strata secara sembarang (berdasarkan pertimbangan subyektif peneliti) sedemikian rupa sehingga $n_1 + n_2 + n_3 + n_4 + \dots + n_L = N$.

b. Alokasi sama

Misalkan suatu populasi yang berukuran N dibagi-bagi kedalam L strata, sedemikian rupa sehingga $N_1 + N_2 + N_3 + N_4 + \dots + N_L = N$ dan total ukuran n dialokasikan kesetiap strata secara sama ke setiap strata, maka ukuran sampel pada setiap strata adalah :

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

c. Alokasi sebanding

Alokasi sebanding dapat dipergunakan bila rata-rata antara strata yang satu dengan strata lainnya berbeda sekali dan varians strata tidak tersedia.

Bila rata-rata antara satu strata dengan strata lainnya kira-kira sama, maka penurunan varians akan kecil sekali. Keuntungan dari alokasi sebanding adalah kepraktisan dalam pengolahan (tabulasi) hasil survei. Hal ini disebabkan karena dengan mengalokasikan sampel ke setiap strata sebanding dengan ukuran strata akan dihasilkan estimator-estimator yang tertimbang secara otomatis (*self-weighting*). Sebagai contoh total ukuran sampel sebesar n dialokasikan kedalam setiap strata sebanding terhadap ukuran setiap strata (N_h), maka

besarnya ukuran sampel pada setiap strata dapat ditentukan dengan cara sebagai berikut :

$$\frac{n_h}{n} = \frac{N_h}{N} \text{ atau } n_h = \frac{N_h}{N} \cdot n = W_h \cdot n$$

Sehingga estimator rata-rata bagi karakteristik Y dalam alokasi sebanding menjadi :

$$\hat{Y}_{st,prop} = \frac{1}{n} \sum_{h=1}^L \sum_{i=1}^{n_h} y_{hi}$$

dengan varians penarikan sampel bagi rata-rata adalah sebagai berikut :

$$V(\hat{Y}_{st})_{prop} = \frac{(1-f)}{n} \sum_{i=1}^L W_i^2 S_i^2$$

Karena s_h^2 merupakan estimator tak bias bagi S_h^2 , maka estimasi bagi varians \hat{Y}_{st} yang dihitung berdasarkan data sampel adalah :

$$v(\hat{Y}_{st})_{prop} = \frac{(1-f)}{n} \sum_{h=1}^L W_h^2 s_h^2$$

c. Alokasi optimum

Pada metode ini kita mengalokasikan sampel yang berukuran n ke dalam setiap strata sedemikian rupa sehingga diperoleh varians sekecil mungkin dengan biaya yang tersedia atau meminimumkan biaya dengan varians tertentu. Fungsi biaya yang sederhana dalam penarikan sampel acak berlapis adalah :

$$C = C_0 + \sum_{h=1}^L c_h n_h$$

dengan :

C total biaya (*total cost*)

C_0 biaya pendahuluan (*overhead cost*)

C_h biaya persatuan sampel pada strata ke- h

n_h ukuran sampel pada strata ke- h

Untuk total biaya tetap sebesar C , varians minimum dicapai bila ukuran sampel pada setiap strata sebanding terhadap $N_h S_h / \sqrt{c_h}$. Dengan demikian ukuran sampel pada setiap strata adalah sebagai berikut :

$$n_h = \frac{(C - C_0) N_h S_h / \sqrt{c_h}}{\sum_{h=1}^L N_h S_h / \sqrt{c_h}} n$$

Bila biaya perunit sampel antar strata sama, maka ukuran sampel pada masing-masing strata menjadi

$$n_i = \frac{(C - C_0) N_h S_h}{\sum_{h=1}^L N_h S_h} n$$

Alokasi ini dikenal sebagai *Neyman Allocation*.

Varians penarikan sampel bagi rata-rata di bawah Neyman Allocation adalah

$$V(\hat{Y}) = \frac{\left(\sum_{h=1}^L W_h S_h \right)^2}{n} - \frac{\sum_{h=1}^L W_h S_h^2}{N}$$

Karena s_h^2 merupakan estimator tak bias bagi S_h^2 , maka estimasi bagi varians \hat{Y}_{st} yang dihitung berdasarkan data sampel adalah :

$$v(\hat{Y})_{\min} = \frac{\left(\sum_{h=1}^L W_h s_h \right)^2}{n} - \frac{\sum_{h=1}^L W_h s_h^2}{N}$$

4.7. Penentuan Ukuran Sampel

Penentuan ukuran sampel sangat tergantung pada presisi (d), dan tingkat reliabilitas (z) yang dikehendaki. Dalam kaitannya dengan penentuan ukuran sampel terdapat hubungan antara presisi, tingkat reliabilitas dan varians bagi statistik yang akan diduga yang dirumuskan dalam bentuk $d^2 = z^2 V(\hat{\theta}_{st})$. Besarnya ukuran sampel, selain tergantung pada besaran-

besaran tersebut juga tergantung pada tipe alokasi sampel ke dalam setiap strata.

Misalkan untuk menduga rata-rata dengan presisi d_0 dan tingkat reliabilitas yang dikehendaki z_0 , maka besarnya ukuran sampel yang dibutuhkan untuk masing-masing tipe alokasi sampel ke dalam setiap strata adalah sebagai berikut : (*untuk penyederhanaan penulisan formula*

dimisalkan $D = \frac{d_0}{z_0}$)

a. Alokasi sama

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

b. Alokasi sebanding

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

c. Alokasi optimum

$$n = \frac{\left(\sum_{h=1}^L N_h S_h \sqrt{c_h} \right) \left(\sum_{h=1}^L N_h S_h / \sqrt{c_h} \right)}{N^2 D^2 + \sum_{h=1}^L N_h S_h^2}$$

d. Alokasi Neyman (biaya perunit sama antar strata)

$$n = \frac{\left(\sum_{h=1}^L N_h S_h \right)^2}{N^2 D^2 + \sum_{h=1}^L N_h S_h^2}$$

Dengan melihat rumus-rumus yang digunakan untuk penentuan ukuran sampel, maka dapat disimpulkan, bahwa pada suatu strata tertentu, ukuran sampel lebih besar bila :

- a. Ukuran strata lebih besar
- b. Strata lebih bervariasi karakteristiknya (*heterogen*)
- c. Strata yang biaya per unit sampelnya lebih murah

BAB V. SAMPLING SISTEMATIK

5.1. Pendahuluan

Pada penarikan sampel acak sederhana setiap unit dipilih dengan menggunakan tabel angka random. Penarikan sampel dengan cara ini agak berat dalam mengerjakannya. Dalam hal ini angka random yang diambil harus sebanyak n yang memenuhi syarat angka random. Penarikan sampel secara sistematis akan membantu mempermudah penarikan sampel, dengan hanya menggunakan satu angka random yang memenuhi persyaratan, sedangkan angka random berikutnya akan mengikuti dengan menghitung intervalnya.

$$\text{Interval sampel : } I = \frac{N}{n}$$

Metode estimasi sama dengan yang digunakan dalam acak sederhana. Sedangkan penghitungan kesalahan sampling salah satunya dapat didekati dengan cara seperti acak sederhana. Sampling sistematis tetap merupakan random sampling dan peluang setiap unit diketahui.

Keuntungan lain dari penggunaan sampling sistematis selain mempermudah penarikan sampel, adalah meningkatkan efisiensi dengan pertimbangan sebagai berikut.

- a. Apabila setiap unit/elemen dari populasi menyebar secara merata di seluruh populasi, dalam keadaan ini sering menghasilkan kesalahan sampling yang lebih kecil. Bila unit/elemen menyebar secara sembarang biasanya sampling sistematis sama efisien dengan acak sederhana.
- b. Dalam sampling sistematis memungkinkan diadakan pengaturan letak unit sampling menurut karakteristik, sehingga diharapkan sampel akan memiliki berbagai karakteristik.

5.2. Cara Penarikan Sampel

Penarikan sampel dengan sistematis dapat dilakukan dengan dua cara, yaitu :

- a. Sistematis linier (linear systematic)
- b. Sistematis sirkuler (circular systematic)

Cara pertama dan keduanya sebenarnya sama, yaitu dengan tetap menghitung interval sampel. Perbedaan adalah dalam penentuan angka random pertama (R_1)

Sistematik Linear

- Hitung interval, yaitu $I = \frac{N}{n}$
- Tentukan satu angka random yang lebih baik kecil atau sama dengan interval. Angka random ini selanjutnya disebut R_1 (angka random pertama). Angka random selanjutnya adalah :

$$R_2 = R_1 + I$$

$$R_3 = R_2 + I = R_1 + 2I, \text{ dst}$$

$$R_n = R_1 + (n - 1)I \quad (\text{Rumus ini sebagai kontrol apakah penarikan sampel sudah benar})$$

Contoh :

Misal banyaknya unit adalah :

$$N = 20$$

$$n = 4$$

$$I = \frac{20}{4} = 5$$

$$R_1 \leq 5 = 3$$

Maka nomor unit-unit terpilih adalah nomor 3,8,13, dan 18

$$R_n = R_4 = 3 + 3(5) = 18$$

Jadi dalam sampling sistematik unit dalam populasi yang akan dijadikan unit sampel diberi nomor urut dari 1 sampai dengan N .

Sistematik Sirkuler

- Hitung interval, yaitu $I = \frac{N}{n}$
- Tentukan satu angka random pertama (R_1) yang lebih kecil atau sama dengan N . Angka random selanjutnya adalah :

$$R_2 = R_1 + I$$

$$R_3 = R_1 + 2I$$

$$R_k = R_1 + (k - 1) I$$

$$R_n = R_1 + (n - 1) I$$

R_k = adalah penghitungan angka random terakhir sebelum nomor urut terakhir.

Apakah R_k melebihi N , maka R_k menjadi $R_k^1 = R_k - N$

Contoh :

Misal banyaknya unit dalam populasi : $N = 20$

$$n = 4$$

$$I = \frac{20}{4} = 5$$

$$R_1 \leq N = 12$$

Maka nomor urut terpilih adalah nomor 12, 17, 2 dan 7

$$R_1 = 12$$

$$R_2 = 12 + 5 = 17$$

$$R_3 = 17 + 5 - 20 = 2$$

$$R_4 = 2 + 5 = 7 \rightarrow R_n = R_1 + (n - 1) I \\ = 12 + 3 \times 5 = 27 - 20 = 7$$

Apabila I tidak berupa bilangan bulat, tetap dapat dilakukan penarikan sampel tersebut dengan melakukan pembulatan. Pembulatan ini menyebabkan peluang tidak tepat, tetapi untuk survei skala besar hal ini dapat diabaikan.

Penarikan sampel secara sistematis juga dapat diterapkan dalam sampling berpeluang tidak sama. Pemakaian sampel dapat didasarkan pada karakteristik yang digunakan sebagai ukuran unit. Sebagai contoh penarikan sampel blok sensus dapat dilakukan secara sistematis dengan banyaknya rumah tangga sebagai penimbang dalam penarikan sampel.

Tabel 5.1: Lembar Kerja Pemilihan Sampel Sistematis

No Urut Blok Sensus	Banyak Rumah tangga (X_1)	Kumulatif Rumah tangga	Angka Random
1	110	110	45
2	80	190	
3	45	235	
4	50	285	
5	87	372	318
6	105	477	
7	95	572	
8	80	652	591
9	68	720	
10	99	819	

Dari data di atas akan dipilih sebanyak 3 blok sensus dengan peluang sebanding dengan banyaknya rumah tangga.

$$I = \frac{819}{3} = 273$$

$$R_1 = \leq 273, \text{ misalnya } 45$$

$$R_2 = 45 + 273 = 318$$

$$R_3 = 318 + 273 = 591$$

Jadi nomor blok sensus terpilih adalah nomor 1,5, dan 8. cara ini adalah cara pertama dengan sistematis linear. Prosedur yang sama dapat diterapkan untuk sistematis sirkuler.

5.3. Kelemahan yang dapat terjadi dalam penggunaan sistematis

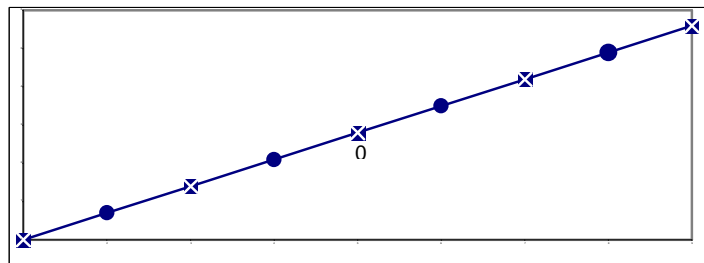
Kelemahan dalam sistematis akan terjadi apabila unit dalam populasi yang akan diambil sampel mengikuti suatu pola tertentu misalnya berfluktuasi secara periodik sehingga mengakibatkan :

- Unit sampel tertentu saja yang terambil sampel.
- Bila interval terlalu besar, maka suatu karakteristik dengan proporsi yang lebih kecil dari fraksi ($\frac{n}{N}$) tidak akan terpilih dalam sampel.

Misalnya bila semua rumah tangga beranggotakan 4 orang dengan susunan kepala keluarga, isteri, anak pertama dan anak kedua. Bila kita menarik sampel secara sistematis dan angka random jatuh pada nomor 1,

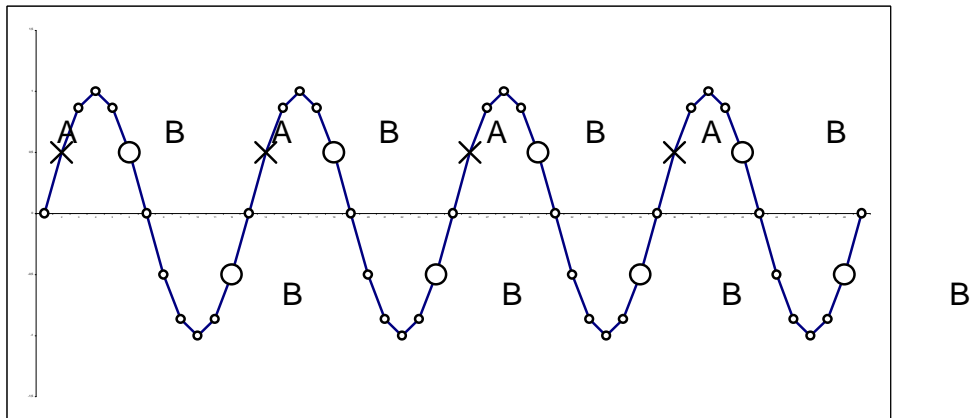
5, 9 dan seterusnya berarti akan selalu jatuh pada kepala rumah tangga. Apabila angka random jatuh pada nomor 2, 6, 10 dan seterusnya berarti akan selalu jatuh pada isteri, dan seterusnya. Kondisi ini biasanya jarang terjadi, tetapi apabila terjadi kondisi semacam ini mengakibatkan sampel tidak akan mewakili populasi. Sampel diharapkan dapat mewakili kepala rumahtangga, isteri maupun anak. Oleh karena itu sebelum menentukan penarikan sampel secara sistematis perlu dipelajari terlebih dahulu susunan kerangka induk dalam populasi dengan berbagai karakteristiknya yang ada kaitan dengan data yang akan diteliti.

Populasi yang unitnya tersusun mendekati gambaran berupa garis lurus; maka efisiensi dari sistematis sampel tergantung pada besarnya interval (I). Apabila karakteristik berupa nilai dari kecil ke besar, kemungkinan besar sampel akan diwakili karakteristik yang nilainya dari kecil ke besar. Sehingga dalam kondisi ini, sistematis sampling akan menguntungkan. Tetapi apabila urutan tadi suatu karakteristik misalnya anggota rumahtangga seperti contoh diatas, maka yang terpilih kemungkinan hanya kepala rumah tangga atau lainnya.



x: unit terpilih

Demikian pula kelemahan akan terjadi pada populasi yang unitnya mempunyai pola berupa variasi yang periodik (*periodic variation*), Efisiensi sampling sistematis tergantung dari besarnya interval.



Dari gambar di atas unit sampel A misalnya hanya mewakili karakteristik tertentu di atas garis, berarti estimasi tidak sesuai dengan yang diharapkan, sedangkan pada unit sampel B (interval lebih kecil dari A), sampel akan lebih mewakili nilai populasi.

5.4. Varians dalam sampling sistematis

Seperti telah dijelaskan sebelumnya estimasi dalam sampling sistematis didekati dengan acak sederhana.

$$\hat{Y} \approx \frac{N}{n} \sum_{i=1}^n y_i$$

Varians perkiraan total atau rata-rata sebenarnya merupakan varian yang diperoleh dari perbedaan sistematis sampel yang terbentuk yaitu sebesar k yang besarnya sama dengan I . Bila dari N unit dipilih n , maka akan terbentuk set sampel sebanyak $\frac{N}{n} = k = I$, dengan gambaran sebagai berikut

Tabel : Komposisi dari sebanyak k sampel

1	2	i	k
Y_1	y_2	y_i	y_k
y_{l+1}	y_{l+2}	y_{l+i}	y_{l+k}
$Y_{(n-1)l+1}$	$Y_{(n-1)l+2}$	$Y_{(n-1)l+i}$	$Y_{(n-1)l+k}$
\bar{y}_1	\bar{y}_2	\bar{y}_i	\bar{y}_k

k : banyaknya set sampel terbentuk, besarnya sama dengan l .

$$V(\bar{y}_{sy}) = \frac{1}{k} \sum_{i=1}^k (\bar{y}_i - \bar{Y})^2 = \frac{1}{n^2 k} \sum_{i=1}^k (n\bar{y}_i - n\bar{Y})^2$$

$$V(\hat{Y}_{sy}) = N^2 V(\bar{y}_{sy})$$

Cara lain dalam penghitungan varians adalah:

$$V(\bar{y}_{sy}) = \frac{N-1}{n} S^2 - \frac{k(n-1)}{N} S_{wsy}^2$$

$$S_{wsy}^2 = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

S_{wsy}^2 dalam sistematik diperhitungkan dari pada masing-masing set sampel dan dijumlahkan dari k set sampel

Varians di atas hanya dapat dihitung dalam perencanaan, misalnya kita akan menggunakan hasil sensus lengkap sebagai kerangka sampel. Dari hasil sensus dapat diperoleh gambaran tersebut di atas untuk memperkirakan besarnya n dengan varians tertentu pada survei yang akan dilaksanakan. Dalam praktek hanya akan diketahui nilai karakteristik dari sampel sebesar n sehingga penghitungan varians diadakan pendekatan sebagai berikut:

a. Acak Sederhana

$$v(\bar{y}_{sy}) \approx \frac{N-n}{N} \frac{s^2}{n}$$

$$v(\hat{Y}_{sy}) \approx N^2 v(\bar{y}_{sy})$$

- b. Model dengan penarikan sampel berpasangan (*paired selections model*)

$$v(\bar{y}_{sy}) \approx \frac{1-f}{n^2} \sum_{h=1}^{n/2} (y_{ha} - y_{hb})^2$$

Dalam hal ini diadakan modifikasi pada penarikan sampel, yaitu diambil dua angka random secara berpasangan sebagai angka random pertama dan dengan interval $I'=2I$. Misal dari N unit akan dipilih n unit, maka interval dihitung sebagai berikut :

$$I' = \frac{N}{\frac{1}{2}n} = 2I$$

$$N=30$$

$$n=6$$

$$I' = \frac{N}{\frac{1}{2}n} = \frac{30}{3} = 10$$

Bila penarikan sampel dengan sistematik linier

$$\left. \begin{matrix} R_1' \\ R_1'' \end{matrix} \right\} \leq 10 \rightarrow \begin{matrix} R_1' = 5 \\ R_1'' = 7 \end{matrix}$$

Maka unit sampel terpilih adalah

Unit ke	5	15	25
	y_{1a}	y_{2a}	y_{3a}
Unit ke	7	17	27
	y_{1b}	y_{2b}	y_{3b}

Sehingga varians dapat dihitung dengan rumus di atas.

Masalah dalam model ini bila n merupakan bilangan ganjil. Untuk mengatasi ini, maka ambil satu sampel secara acak, sehingga ada sampel yang dua kali diperhitungkan.

- c. Model dengan perbedaan yang berurutan (*Successive difference model*)

$$v(\bar{y}_{sy}) \approx \frac{1-f}{2n(n-1)} \sum_g^{n-1} (y_g - y_{g+1})^2$$

Contoh

$$\left. \begin{array}{l} N = 30 \\ n = 6 \end{array} \right\} \rightarrow I = \frac{30}{6} = 5; \quad R_1 = 5$$

Maka unit terpilih adalah :

Unit ke	5	10	15	20	25	30
	y_1	y_2	y_3	y_4	y_5	y_6

$y_g - y_{g+1}$ berarti $y_1 - y_2$ $y_2 - y_3$ $y_3 - y_4$ $y_4 - y_5$
 $y_5 - y_6$

5.5. Efisiensi Sampling Sistematis terhadap Acak Sederhana

$$V(\bar{y}_{SRS}) = \frac{N-n}{N} \frac{S^2}{n}$$

$$V(\bar{y}_{sy}) = \frac{N-n}{N} S^2 - \frac{k(n-1)}{N} S_{wsy}^2$$

Agar sampling sistematis lebih efisien maka

$$V(\bar{y}_{sy}) < V(\bar{y}_{SRS})$$

$$\frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S_{wsy}^2 < \frac{N-n}{N} \frac{S^2}{n}$$

$$k(n-1) S_{wsy}^2 > \left(N-1 - \frac{N-n}{N} \right) S^2 = k(n-1) S^2$$

$$S_{wsy}^2 > S^2$$

BAB VI. PENARIKAN SAMPEL DENGAN PELUANG SEBANDING

DENGAN UKURAN UNIT

(Sampling with Probability Proportional to Size – pps sampling)

6.1. Pendahuluan

Dalam penarikan sampel acak sederhana, informasi yang berkaitan dengan unit sampling yang digunakan hanya berupa identitas berupa nomor urut unit dalam populasi. Pada umumnya setiap unit sampling di dalam populasi ukurannya bervariasi, maka penerapan penarikan sampel acak sederhana bukanlah merupakan prosedur yang tepat

Bila setiap unit sampling di dalam populasi dilengkapi juga dengan informasi tambahan (*auxiliary information*), maka informasi tersebut dapat digunakan untuk dasar pertimbangan di dalam penarikan sampel sehingga dapat diperoleh estimator-estimator yang lebih efisien. Informasi tambahan yang berguna untuk dasar pertimbangan penarikan sampel adalah informasi yang memiliki korelasi yang erat dengan variabel-variabel yang akan diteliti. (informasi tambahan yang dipertimbangkan sebagai dasar penarikan sampel selanjutnya disebut ukuran). Prosedur penarikan sampel dimana peluang terpilihnya suatu unit sampel sebanding dengan ukuran disebut sebagai *Sampling with Probability proportional to Size* atau disingkat *pps sampling*

6.2. Prosedur Penarikan Sampel

a. Metode Kumulatif

Misalkan sampel berukuran n dipilih dari populasi yang berukuran N secara *pps-wr* dengan x_i sebagai ukuran unit ke- i . Tahapan penarikan sampelnya adalah sebagai berikut :

- i. Buatlah jumlah kumulatif dari ukuran yang digunakan untuk dasar penarikan sampel untuk seluruh unit di dalam populasi, yaitu : x_1 , $(x_1 + x_2)$, $(x_1 + x_2 + x_3)$, ..., $\sum x_i = X$
- ii. Bangkitkan angka random (R_k) yang besarnya antara 1 sampai dengan X
- iii. Unit ke- k terpilih sebagai sampel, bila terpenuhi kondisi $\sum_{i=1}^{k-1} x_i + 1 \leq R_k \leq \sum_{i=1}^k x_i$. Bila kondisi tersebut tidak terpenuhi, ulangi langkah ke-ii.
- iv. Ulangi langkah ke-ii hingga jumlah yang harus dipilih yaitu sebanyak n unit terpenuhi.

Ilustrasi 1 : Berikut ini adalah daftar nama 24 kepala rumahtangga berikut jumlah anggotanya untuk rumahtangga-rumahtangga yang tinggal di suatu desa pedesaan. Untuk mengetahui rata-rata pengeluaran per minggu, maka dipilih sampel sebanyak 4 rumahtangga secara *pps* dengan pemulihan dengan ukuran jumlah anggota rumahtangga. Untuk metode kumulatif maka perlu dibuat kumulatif ukuran unit.

Tabel 6.1: Daftar Kerangka Pemilihan Sampel Rumahtangga

No	Nama KRT	Jumlah art	Jumlah kumulati f art	No	Nama KRT	Jumlah art	Jumlah kumulati f art
01	Abiyoso	5	5	13	Mustakin	9	61
02	Banuarja	2	7	14	Normandino	4	65
03	Chaidir	9	16	15	Oposunggu	8	73
04	Darusmin	3	19	16	Pardosi	7	80
05	Efraim	7	26	17	Qurmanto	7	87
06	Faturrahman	4	30	18	Rusdi	5	92
07	Gusmianto	3	33	19	Samektopo	2	94
08	Hutajulu	1	34	20	Tarjono	2	96
09	Indrastoto	5	39	21	Umarsono	7	103
10	Jamaludin	5	44	22	Verianto	4	107

11	Kartosemoy o	5	49	23	Xiemenes	7	114
12	Luhantara	3	52	24	Yamantoro	6	120

- Misalnya Tabel Angka Random yang digunakan adalah Tabel Angka Random Halaman 1 kolom 1-3, angka random sebanyak 3 digit karena kumulatif terakhir terdiri dari 3 digit (120)
- Angka random dan rumahtangga terpilih adalah sebagai berikut.

Tabel 6.1: Lembar Kerja Pemilihan Sampel Rumahtangga

<i>Angka Random (R_i)</i>		<i>No. Rtmg yang berpadana n dengan R_i</i>	<i>Nama Kepala rumahtangga</i>	<i>Jumlah art</i>
3 - digit	<i>Sisa setelah dibagi 120</i>			
034	-	08	Hutajulu	1
977	17	04	Darusmin	3
167	47	11	Kartosemoyo	5
125	05	01	Abiyoso	5

b. *Metode Lahiri*

Metode ini tidak membutuhkan jumlah kumulatif dari ukuran unit sampling di dalam populasi. Misalkan sampel berukuran n dipilih dari populasi yang berukuran N secara *pps* dengan pemulihan dengan x_i sebagai ukuran unit ke- i . Tahapan penarikan sampelnya adalah sebagai berikut :

- Bangkitkan dua buah angka random secara serentak, yaitu :
 - $R_i^{(1)}$ yang besarnya antara 1 sampai dengan N . Angka random ini berkenaan dengan nomor urut unit sampling di dalam populasi.

- $R_i^{(2)}$ yang besarnya antara 1 sampai dengan x_{maks} . Angka random ini berkenaan dengan ukuran yang digunakan untuk penarikan sampel.
- ii. Bila $R_i^{(2)} \leq x_i$, maka pilih unit yang ke-i. Sebaliknya bila kondisi ini tidak dipenuhi, maka bangkitkan 2 angka random lagi secara serentak.
- iii. Ulangi langkah ke-ii hingga jumlah yang harus dipilih yaitu sebanyak n unit terpenuhi.

Ilustrasi 2 : Berikut ini adalah daftar nama 24 kepala rumahtangga berikut jumlah anggotanya untuk rumahtangga-rumahtangga yang tinggal di suatu desa pedesaan. Untuk mengetahui rata-rata pengeluaran per minggu, maka dipilih sampel sebanyak 4 rumahtangga secara *pps* dengan pemulihan dengan ukuran jumlah anggota rumahtangga. Metode yang digunakan adalah Metode Lahiri.

Tabel 6.3: Daftar Kerangka Pemilihan Sampel Rumahtangga

No	Nama KRT	Jumlah ahart	No	Nama KRT	Jumlah ahart
01	Abiyoso	5	13	Mustakin	9
02	Banuarja	2	14	Normandino	4
03	Chaidir	9	15	Oposunggu	8
04	Darusmin	3	16	Pardosi	7
05	Efraim	7	17	Qurmanto	7
06	Faturrahman	4	18	Rusdi	5
07	Gusmianto	3	19	Samektopo	2
08	Hutajulu	1	20	Tarjono	2
09	Indrastoto	5	21	Umarsono	7
10	Jamaludin	5	22	Verianto	4
11	Kartosemoyo	5	23	Xiemenes	7
12	Luhantara	3	24	Yamantoro	6

- Misalnya Tabel Angka Random yang digunakan adalah Tabel Angka Random Halaman 1 kolom 2-4 baris pertama. Kolom 2-3 untuk angka random $R_i^{(1)}$ dan kolom 4 untuk $R_i^{(2)}$
- Angka random yang diperoleh adalah sebagai berikut :

Tabel 6.4: Lembar Kerja Pemilihan Sampel Rumahtangga

Angka random untuk		x_i	Keputusan	Rumahtangga terpilih	
Rumahtangga ($R_i^{(1)}$)	ART ($R_i^{(2)}$)			No. urut	Nama Kepala Rumahtangga
34 (R 10)	7	5	Tolak	-	-
62 (R 14)	2	4	Terima	01	Normandino
44 (R 20)	2	2	Terima	02	Taryono
30 (R 06)	1	4	Terima	03	Faturahman
32 (R 08)	1	1	Terima	04	Hutajulu
<p><i>Catatan</i></p> <ul style="list-style-type: none"> • Angka setelah huruf <i>R</i> yang terdapat dalam tanda kurung menunjukkan sisa setelah dibagi dengan 24 • Tidak semua angka random yang dibangkitkan dan <u>ditolak</u> ditampilkan pada tabel di atas 					

Metode Lahiri digunakan sebenarnya hanya menghindari penghitungan kumulatif apabila ukuran unit cukup besar.

6.3. Pps sistematis

Bila ukuran sampel sebesar n dan X adalah total ukuran, maka interval penarikan sampelnya adalah :

$$I = \frac{X}{n} = \frac{1}{n} \sum_{i=1}^N x_i$$

- I bilangan bulat (*integer*), maka gunakan sistematik linear
Misalkan R_1 adalah merupakan angka random pertama (*random start*) yang lebih kecil atau sama dengan I , maka unit-unit yang berpadanan dengan $(R_1 + j \cdot I)$, $j = 0, 2, 3, \dots, (n-1)$, akan terpilih sebagai sampel. Secara umum, unit ke- i terpilih sebagai sampel bila dipenuhi kondisi :

$$\sum_{i=1}^{i-1} x_i < R_1 + j \cdot I \leq \sum_{i=1}^i x_i$$

- I bukan bilangan bulat, maka gunakan sistematik sirkuler.
Dalam sistematik sirkuler, angka random pertama R_1 besarnya antara 1 sampai dengan N (tidak harus lebih kecil sama dengan interval).

Ilustrasi 4 : Berikut ini adalah daftar nama 24 kepala rumahtangga berikut jumlah anggotanya untuk rumahtangga-rumahtangga yang tinggal di suatu desa pedesaan. Untuk mengetahui rata-rata pengeluaran per minggu, maka dipilih sampel sebanyak 4 rumahtangga secara *pps-sistematik linear* dengan ukuran jumlah anggota rumahtangga.

Tabel 6.5: Daftar Kerangka Pemilihan Sampel Rumahtangga

No	Nama KRT	Jumlah h art	Jumlah kumulatif if art	No	Nama KRT	Jumlah art	Jumlah kumulatif art
01	Abiyoso	5	5	13	Mustakin	9	61
02	Banuarja	2	7	14	Normandino	4	65
03	Chaidir	9	16	15	Oposunggu	8	73
04	Darusmin	3	19	16	Pardosi	7	80
05	Efraim	7	26	17	Qurmanto	7	87
06	Faturrahman	4	30	18	Rusdi	5	92
07	Gusmianto	3	33	19	Samektopo	2	94
08	Hutajulu	1	34	20	Tarjono	2	96
09	Indrastoto	5	39	21	Umarsono	7	103
10	Jamaludin	5	44	22	Verianto	4	107

11	Kartosemoyo	5	49	23	Xiemenes	7	114
12	Luhantara	3	52	24	Yamantoro	6	120

- $I = 120/4 = 30$
- Misalnya Tabel Angka Random yang digunakan adalah Tabel Angka Random Halaman 1 kolom 1-2 sesuai dengan banyaknya digit dari I . Dari Tabel Angka Random diperoleh $R_1 = 03$ (lebih kecil dari Interval sehingga langsung dapat digunakan).
- Angka random pertama s.d. keempat dan rumahtangga terpilih adalah sebagai berikut:

Tabel 6.6: Lembar Kerja Pemilihan Sampel Rumahtangga

R_i	Jumlah kumulatif yang berpadanan dengan R_i	No. Rtmg yang berpadanan dengan jumlah kumulatif	Rumahtangga terpilih		
			No.	Nama kepala rumahtangga	Jumlah art
03	5	01	01	Abiyoso	5
33	33	07	02	Gusmianto	3
63	65	14	03	Normandino	4
93	94	19	04	Samektopo	2

Catatan :

Kolom (1) dihitung dengan rumus

$R_1 = 03$ (sama dengan angka random)

$$R_2 = R_1 + I = 3 + 30 = 33$$

$$R_3 = R_2 + I = 33 + 30 = 63$$

$$R_4 = R_3 + I = 63 + 30 = 93$$

Ilustrasi 5 : Berikut ini adalah daftar nama 24 kepala rumahtangga berikut jumlah anggotanya untuk rumahtangga-rumahtangga yang tinggal di suatu desa pedesaan. Untuk mengetahui rata-rata pengeluaran per minggu, maka dipilih sampel sebanyak 7 rumahtangga secara *pps-sistematic sirkuler* dengan ukuran jumlah anggota rumahtangga.

- $I = 120/7 = 17$
- Misalnya Tabel Angka Random yang digunakan adalah Tabel Angka Random Halaman 1 kolom 3-4 baris 1. Dari Tabel Angka Random diperoleh $R_1 = 47$ (tidak harus lebih kecil dari Interval, tetapi antara 1 s.d 120)
- Angka random pertama s.d. ketujuh dan rumahtangga terpilih adalah sebagai berikut :

Tabel 6.7: Lembar Kerja Pemilihan Sampel Rumahtangga

Angka Random R_i		Jumlah kumulatif yang berpadanan dengan R_i	No. Rtmg yang berpadanan dengan jumlah kumulatif	Rumahtangga terpilih		
				No.	Nama kepala rumahtangga	Jumlah ART
(1)	(2)	(3)	(4)	(5)	(6)	(7)
47	-	49	11	01	Kartosemoyo	5
64	-	65	14	02	Normandino	4
81	-	87	17	03	Qurmanto	7
98	-	103	21	04	Umarsono	7
115	-	120	24	05	Yamantoro	6
132	(R 12)	16	02	06	Chaidir	9
149	(R 29)	30	06	07	Faturrahman	4

Catatan :

Kolom (1) : dihitung dengan rumus $R_1 = 47$ (sama dengan angka random pertama)

$$R_2 = R_1 + I. = 47 + 17 = 64$$

$$R_3 = 64 + 17 = 81$$

$$R_4 = 81 + 17 = 98$$

$$R_5 = 98 + 17 = 115$$

$$R_6 = 115 + 17 = 132 - 120 = 12$$

$$R_7 = 12 + 17 = 29$$

Kolom (2) : sisa setelah dikurangi dengan 120

6.4. Prosedur Estimasi

1. *PPS dengan pemulihan : Hunsen – Hurwitz Estimator*

Dalam suatu penarikan sampel, misalnya ukuran sampel ditentukan sebanyak n unit yang harus dipilih dari N unit di dalam populasi secara pps dengan pemulihan dengan ukuran x_i , maka besarnya peluang terpilihnya unit ke- i sebagai sampel adalah :

$$p_i = \frac{x_i}{\sum_{i=1}^N x_i} = \frac{x_i}{X}$$

Misalkan y_i adalah nilai variabel yang berpadanan dengan terpilihnya unit ke- i , maka estimasi tak bias bagi total adalah :

$$\hat{Y}_i = \frac{y_i}{p_i} = X \frac{y_i}{x_i}$$

\hat{Y}_i dengan ($i = 1, 2, 3, \dots, n$) adalah merupakan estimator-estimator yang *tak bias* bagi total Y dan saling bebas (*independent*), maka kombinasi dari estimator-estimator tersebut merupakan estimator yang tak bias bagi total Y , yaitu :

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i = \frac{X}{n} \sum_{i=1}^n \frac{y_i}{x_i}$$

dengan sampling varians

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^N p_i \left(\frac{y_i}{p_i} - Y \right)^2 = \frac{1}{N} \left(\sum_{i=1}^N \frac{y_i^2}{p_i} - Y^2 \right)$$

Estimasi tak bias bagi $V(\hat{Y})$ adalah

$$v(\hat{Y}) = \hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{y_i}{p_i} - \hat{Y} \right)^2$$

2. *Sembarang Rancangan : Horwitz – Thomson Estimator*

Untuk sembarang rancangan, *pps* dengan pemulihan ataupun tanpa pemulihan, misalnya diketahui π_i adalah probability unit ke-*i* masuk sebagai anggota sampel, maka estimator tak bias bagi total *Y* adalah sebagai berikut :

$$\hat{Y} = \sum_{i=1}^n \frac{y_i}{\pi_i}$$

dengan sampling varians :

$$V(\hat{Y}) = \sum_{i=1}^N \left(\frac{1 - \pi_i}{\pi_i} \right) y_i^2 + \sum_{i=1}^N \sum_{j \neq i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) y_i y_j$$

n : jumlah sampel terpilih

$\pi_i = 1 - (1 - p_i)^n$, adalah peluang unit ke-*i* masuk dalam anggota sampel

$\pi_{ij} = \pi_i + \pi_j - [1 - (1 - p_i - p_j)^n]$, adalah peluang unit ke-*i* dan ke-*j* keduanya termasuk dalam anggota sampel.

Dengan estimator tak bias bagi $V(\hat{Y})$ adalah sebagai berikut :

$$v(\hat{Y}) = \sum_{i=1}^n \left(\frac{1}{\pi_i^2} - \frac{1}{\pi_i} \right) y_i^2 + 2 \sum_{i=1}^n \sum_{j > i} \left(\frac{1}{\pi_i \pi_j} - \frac{1}{\pi_{ij}} \right) y_i y_j$$

$$= \sum_{i=1}^n \left(\frac{1 - \pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i=1}^n \sum_{j \neq i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) \frac{y_i y_j}{\pi_{ij}}$$

atau

$$= \sum_{i=1}^n \left(\frac{1 - \pi_i}{\pi_i^2} \right) y_i^2 + 2 \sum_{i=1}^n \sum_{j > i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) \frac{y_i y_j}{\pi_{ij}}$$

Oleh karena penghitungan probability cukup rumit dalam penghitungan di atas, maka untuk survei dengan sampel cukup besar seperti Susenas, Sakernas dan sebagainya digunakan pendekatan rumus *pps*

dengan pemulihan. Dalam hal ini jumlah unit cukup besar, sehingga kemungkinan terpilih lebih dari sekali sangat kecil.

BAB VII. ESTIMASI RASIO

7.1. Pendahuluan

Pada dasarnya metode sampling digunakan dengan maksud merencanakan sampel yang paling efisien (dalam arti meminimumkan kesalahan sampling) dengan menggunakan sebanyak mungkin informasi yang tepat sesuai dengan obyek dan tujuan survei. Berbagai informasi telah digunakan untuk melakukan stratifikasi baik dalam sampel dengan alokasi sebanding maupun alokasi optimum, bagaimana mengkaitkan dengan biaya per unit, dan bagaimana menentukan unit sampel yang sesuai. Pemanfaatan berbagai informasi yang tersedia selain dalam sampling berstrata juga diterapkan dalam sampling berpeluang sebanding dengan ukuran unit (PPS Sampling) seperti telah dibahas sebelumnya. Peningkatan efisiensi ini dapat juga dikembangkan pada estimasi rasio yang menggunakan informasi tadi sebagai variabel pembantu dalam estimasi. Bila akan menggunakan metode sampling tersebut, informasi harus tersedia pada setiap unit yang akan dijadikan dasar stratifikasi atau penghitungan peluang terpilihnya suatu unit atau untuk penerapan estimasi rasio. Persyaratan lain informasi tersebut harus mempunyai korelasi yang erat dengan data yang akan diteliti.

Pada uraian selanjutnya akan dibahas estimasi rasio yang akan diharapkan dapat menghasilkan varians atau kesalahan sampling yang lebih kecil. Estimasi rasio ini mendasarkan perubahan suatu karakteristik dari waktu ke waktu. Hal ini akan sangat bermanfaat terutama untuk survei-survei yang mencakup dua periode waktu yang berbeda dengan melihat perkembangan suatu karakteristik yang nantinya digunakan sebagai rasio dalam estimasi. Dengan demikian diharapkan akan memperoleh nilai estimasi yang lebih efisien.

Sebagai contoh kita dapat memperoleh informasi mengenai perkembangan banyaknya rumah tangga/penduduk, lahan pertanian, besarnya pembayaran upah dan gaji, dan rata-rata penghasilan per

rumah tangga dengan menggunakan rasio. Untuk data semacam ini estimasi rasio akan lebih efisien, karena yang diharapkan adalah gambaran perubahannya, dengan tidak menutup kemungkinan mendapatkan estimasi total.

7.2. Estimasi Rasio Pada Data Agregat

Di antara metode sampling yang ada, estimasi rasio paling banyak dipakai dan penerapannya paling mudah. Estimasi rasio cocok sekali dipakai bila unit-unit populasi mempunyai karakteristik yang berkorelasi positif. Makin besar korelasi tersebut makin besar pula efisiensi penggunaan metode tersebut.

Bentuk sederhana dari estimasi rasio adalah :

$$\hat{Y}_R = \frac{y}{x} X$$

y dan x harus diketahui yang berupa nilai agregat dua buah karakteristik yang diperoleh dari unit sampel terpilih.

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

$$x = \sum_{i=1}^n x_i$$

Sedangkan X berupa nilai populasi yang diketahui yaitu :

$$X = \sum_{i=1}^N X_i$$

Estimasi rasio dibedakan menjadi tiga kondisi, yaitu :

- a. Rasio terhadap karakteristik yang sama atau berhubungan dengan periode sebelumnya.

X adalah jenis karakteristik yang sama dengan Y , tetapi berasal dari periode sebelumnya yaitu pada saat sensus diadakan. X harus merupakan nilai karakteristik dari populasi. Misalnya dari hasil Sensus Penduduk 2000 (SP 2000) kita mendapatkan berbagai karakteristik yang merupakan nilai populasi seperti jumlah rumahtangga, penduduk menurut berbagai karakteristik sehingga data ini dapat digunakan sebagai nilai dari X . Selanjutnya dari unit sampel akan diperoleh nilai

sampel dari unit sampel terpilih (misal blok sensus) pada saat survei. Nilai sampel dari hasil survei adalah y . Dari unit sampel terpilih diketahui juga nilai x , yaitu keadaan pada SP 2000. Dengan demikian dapat diketahui nilai perubahan pada kedua periode tersebut, yaitu y/x . Apabila nilai perubahan ini dikalikan dengan X , maka akan merupakan perkiraan total nilai populasi.

- b. Rasio dari dua karakteristik yang berhubungan pada periode yang sama

X dan Y merupakan dua buah karakteristik yang berasal dari periode yang sama dan diketahui mempunyai korelasi yang positif.

Nilai sesungguhnya dari data agregat X diketahui baik dari sensus atau sumber lain yang merupakan nilai populasi.

Contoh :

Misal dalam survei pertanian dengan pendekatan rumah tangga tani ditanyakan luas lahan pertanian yang dikuasai (x_i) dan karakteristik lebih rinci yang diteliti adalah banyaknya pupuk yang digunakan (y_i), dari sumber lain misalnya pengumpulan data rutin, diketahui luas lahan pertanian yang dikuasai rumah tangga tani (X). Bila diketahui adanya korelasi yang besar antara penggunaan pupuk dan luas lahan, maka penggunaan estimasi rasio akan menurunkan kesalahan sampling cukup besar.

Dari data y dan x , dapat dihitung rata-rata pupuk per hektar. Apabila dikalikan dengan X akan merupakan pupuk yang digunakan pada lahan pertanian oleh rumah tangga tani.

- c. Rasio dari suatu sub-set total

Karakteristik Y merupakan bagian (sub-set) dari X yang diperkirakan perubahannya sebanding dengan X .

Contoh

- i). Kasus a di atas dikembangkan, yaitu karakteristik y tidak terbatas pada rumah tangga/penduduk tetapi diperluas menjadi karakteristik kependudukan lainnya seperti pendidikan, angkatan kerja dan sebagainya.

$\frac{X}{x}$ merupakan faktor pengali

- ii). Kasus b di atas, diperluas menjadi luas lahan yang ditanami tanaman tertentu (y tidak terbatas pada penggunaan pupuk)

Sama seperti contoh sebelumnya

$$\frac{X}{x} = \frac{\text{luas lahan dari populasi dan suatu sumber}}{\text{luas lahan dalam sampel}}$$

merupakan faktor pengali. X juga dapat merupakan angka proyeksi pada periode yang sama. Misalnya dalam Susenas digunakan proyeksi penduduk sebagai dasar rasio.

7.3. Varians dan Bias dari Estimasi Rasio

1. Varians

Estimasi rasio merupakan estimasi yang bias, tetapi dari segi aplikasi bila memenuhi persyaratan adanya korelasi yang besar dan positif antara x dan y , maka penggunaan estimasi rasio akan sangat efisien. Hal ini akan tergambar dalam perhitungan varians.

Apabila diperhatikan rumus estimasi rasio, yaitu $\hat{Y} = \frac{y}{x} X$, maka

kesalahan sampling hanya tergantung pada y/x karena X merupakan nilai populasi. X hanya mempunyai pengaruh pada faktor pengali yang konstan. Oleh karena itu penghitungan ragam akan sangat tergantung

pada $r = \frac{y}{x}$ sebagai perkiraan $R = \frac{Y}{X}$, penghitungan varians akan

lebih sulit, karena dipengaruhi hubungan 2 variabel yang masing-masing mempunyai kesalahan sampling (korelasi dan covariance).

Apabila diperhitungkan dari populasi, maka sampling varians dari estimasi rasio, adalah :

$$V(\hat{R}) = \frac{(1-f)}{n\bar{X}^2(N-1)} \sum_{i=1}^N (y_i - R x_i)^2$$

dengan $f = \frac{n}{N}$ adalah merupakan fraksi sampling

a. Sampling varians bagi rasio \hat{R}

$$V(\hat{R}) = \frac{(1-f)}{n} R^2 \left[\frac{S_y^2}{\bar{Y}^2} + \frac{S_x^2}{\bar{X}^2} - 2\rho \frac{S_x}{\bar{X}} \frac{S_y}{\bar{Y}} \right] = \frac{(1-f)}{n} R^2 (C_y^2 + C_x^2 - 2\rho C_x C_y)$$

b. Sampling varians bagi rata-rata rasio $\hat{\bar{Y}}_R$

$$V(\hat{\bar{Y}}_R) = \frac{(1-f)}{n} (S_y^2 + R^2 S_x^2 - 2\rho R S_x S_y) = \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + C_x^2 - 2\rho C_x C_y)$$

c. Sampling varians bagi total rasio \hat{Y}_R

$$V(\hat{Y}_R) = N^2 V(\hat{\bar{Y}}_R) = \frac{(1-f)}{n} \hat{Y}^2 (C_y^2 + C_x^2 - 2\rho C_x C_y)$$

$$\text{atau } V(\hat{Y}_R) = \frac{N^2(1-f)}{n} \sum_{i=1}^N \frac{(y_i - R x_i)^2}{N-1}$$

$$\text{Koefisien korelasi dihitung dengan rumus } \rho = \frac{\sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})}{(N-1) S_x S_y}$$

Dalam suatu survei, penghitungan varians tersebut didekati dari hasil surveinya sendiri, yaitu:

a. Estimasi varians bagi \hat{R}

$$v(\hat{R}) = \frac{(1-f)}{n\bar{X}^2} \sum_{i=1}^n \frac{(y_i - \hat{R} x_i)^2}{n-1} = \frac{(1-f)}{n\bar{X}^2} (s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_x s_y)$$

b. Estimasi varians bagi $\hat{\bar{Y}}_R$

$$v(\hat{\bar{Y}}_R) = \frac{(1-f)}{n} (s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_x s_y) = \frac{(1-f)}{n} \hat{\bar{Y}}^2 (c_y^2 + c_x^2 - 2\hat{R} \hat{\rho} c_x c_y)$$

c. Estimasi varians bagi \hat{Y}_R

$$v(\hat{Y}_R) = \frac{N^2(1-f)}{n} (s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} r s_x s_y)$$

Estimasi nilai koefisien korelasi dihitung dengan rumus

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{(n-1) s_x s_y}$$

Kesalahan sampling adalah

$$\sqrt{v(\hat{Y}_R)} \quad \text{atau} \quad \sqrt{v(\hat{\bar{Y}}_R)}$$

Sedangkan kesalahan relatifnya :

$$\frac{\sqrt{v(\hat{Y}_R)}}{\hat{Y}_R} = \frac{\sqrt{v(\hat{\bar{Y}}_R)}}{\hat{\bar{Y}}_R}$$

2. Perbandingan antara varians rasio dan acak sederhana

$$\text{Varians acak sederhana : } V(\hat{Y}_{SR}) = \frac{(1-f)}{n} S_y^2$$

$$\text{Varians acak sederhana : } V(\hat{Y}_R) = \frac{(1-f)}{n} (S_y^2 + R^2 S_x^2 - 2R \rho S_x S_y)$$

Agar estimasi rasio lebih efisien dari acak sederhana, maka :

$$S_y^2 + R^2 S_x^2 - 2R \rho S_x S_y < S_y^2$$

$$\rho > \frac{R S_x}{2 S_y} = \frac{1}{2} \frac{\left(\frac{S_x}{\bar{X}} \right)}{\left(\frac{S_y}{\bar{Y}} \right)} = \frac{1}{2} \frac{C_x}{C_y}$$

$C_x = \frac{S_x}{\bar{X}}$ adalah koefisien variasi dari karakteristik X

$C_y = \frac{S_y}{\bar{Y}}$ adalah koefisien variasi dari karakteristik Y.

Dengan demikian, bila :

$\rho > \frac{1}{2} \frac{C_x}{C_y}$, maka estimasi rasio lebih efisien.

$\rho = \frac{1}{2} \frac{C_x}{C_y}$, maka kedua estimasi mempunyai kesalahan sampling sama.

$\rho < \frac{1}{2} \frac{C_x}{C_y}$, maka estimasi rasio kurang efisien.

Jadi, dalam hal ini besarnya korelasi sangat mempengaruhi efisiensi dugaan. Korelasi ini dapat dipelajari dari berbagai sumber. Di Indonesia telah banyak dilakukan berbagai sensus dan survei. Dari data yang telah ada dapat dipelajari besaran dari korelasi untuk penyempurnaan penggunaan estimasi rasio.

a. Korelasi yang tinggi antara X dan Y

Korelasi yang tinggi akan sangat mempengaruhi penurunan ragam dari penggunaan estimasi rasio. Dalam hal nilai C_x^2 dan C_y^2 hampir sama dan bila korelasi antara X dan Y sebesar 0,90 atau 0,95; maka estimasi rasio akan menghasilkan akurasi yang berarti. Besarnya keuntungan yang diperoleh dapat dihitung sebagai berikut :

$$C_x^2 = C_y^2 \text{ dan } V(\hat{Y}_R) = \frac{(1-f)}{n} Y^2 (C_y^2 + C_x^2 - 2\rho C_x C_y)$$

Pengaruh efisiensi digambarkan oleh $(C_y^2 + C_x^2 - 2\rho C_x C_y)$. Bila $\rho = 0,20$, diperoleh $(2 C_y^2 - 2(0,90) C_y^2) = 0,20 C_y^2$. Dengan perkataan lain, penggunaan estimasi rasio menghasilkan penurunan besarnya ragam 0,80. Ragam dalam kasus ini menjadi 0,20 dari ragam semula.

Apabila $\rho = 0,95$, dengan penghitungan yang sama diperoleh ragam yang besarnya 0,10 dari ragam semula. Dengan kata lain estimasi rasio menghasilkan penurunan ragam sebesar 0,90 dari ragam semula.

Bila dilihat dari segi besarnya sampel, maka penduga rasio untuk menghasilkan ragam yang sama dengan sampel acak sederhana, hanya diperlukan sampel seperlimanya untuk $\rho = 0,90$ atau

sepersepuluhnya untuk $\rho = 0,90$. Hal ini akan sangat mempengaruhi pada pembiayaan dari survei serta faktor-faktor lain yang berkaitan seperti lamanya pelaksanaan dan pengolahan.

b. Korelasi yang rendah antara X dan Y

Untuk kasus dimana estimasi rasio digunakan untuk subset karakteristik atau karakteristik lain di luar yang digunakan sebagai dasar rasio dalam populasi maka kemungkinan korelasi rendah. Sebagai contoh dalam survei pertanian, penggunaan lahan pertanian digunakan sebagai dasar rasio, untuk penggunaan pupuk secara total. Mungkin dalam hal ini akan dikaji juga penggunaan pupuk per jenis tanaman. Atau dalam survei kependudukan yang digunakan jumlah rumah tangga/penduduk sebagai dasar rasio, juga digunakan untuk karakteristik lainnya, seperti angkatan kerja, termasuk didalamnya pengangguran. Dalam kasus ini kemungkinan adanya rasio yang kecil dari $\frac{Y}{X}$ atau $\frac{y}{x}$.

Dalam praktek, bila rasio itu lebih kecil dari 20 persen, kemungkinan estimasi rasio akan menaikkan kesalahan sampling, meskipun pada umumnya kenaikan tidak terlalu besar. Bila rasio sebesar 40 persen dan lebih, estimasi rasio akan lebih efisien sedangkan diantara 20 – 40 persen, perbedaan antar kedua jenis penduga akan kecil saja.

Dalam Survei Angkatan Kerja misalnya, penggunaan estimasi rasio mungkin menghasilkan kenaikan keefisienan yang besar dalam hal estimasi jumlah angkatan kerja yang bekerja (karena proporsinya tinggi terhadap penduduk), sebaliknya untuk jumlah pengangguran mungkin menyebabkan sedikit kenaikan pada kesalahan sampling. Dalam kondisi bagaimana estimasi rasio lebih efisien dapat dilihat pada rumus yang telah dibahas sebelumnya.

3. Bias pada Estimasi Rasio

Estimasi dengan menggunakan rasio adalah estimasi yang berbias (biased estimate). Pada setiap unit sampel dapat diperoleh nilai x_i dan y_i dan rasio dari y_i/x_i yang dihitung dari masing-masing unit sampel akan berbeda dengan y/x dari seluruh untuk sampel. Rata-rata rasio tidak sama persis dengan rata-rata yang sebenarnya. Walaupun demikian bias tersebut dapat diabaikan bila ukuran sampel cukup besar. Dalam praktek bias tersebut demikian kecilnya, sehingga dibandingkan dengan keuntungan yang diperoleh dengan penurunan sampling error, estimasi rasio disarankan untuk digunakan. Oleh karena itu survei-survei di BPS seperti Survei Sosial Ekonomi Nasional (Susenas), Survei Angkatan Kerja dan Survei Pertanian digunakan estimasi rasio. Karakteristik yang digunakan sebagai dasar rasio adalah karakteristik yang mempunyai korelasi dengan data yang diteliti. Bias dalam estimasi rasio di sini adalah bias dari penggunaan rumus estimasi rasio. Bias lainnya yang sebenarnya sangat perlu mendapat perhatian adalah bias di luar metode sampling seperti salah cakupan dalam pendaftaran rumahtangga, penerapan konsep definisi yang tidak baku antar petugas, struktur pertanyaan yang kurang baik, salah jawaban, pengolahan dan sebagainya. Pendaftaran rumahtangga atau listing sering menyebabkan bias yang besar, karena tidak semua rumahtangga dicakup.

Bias dalam estimasi rasio dapat dihitung dengan rumus :

$$B(\hat{R}) = \frac{(1-f)}{n\bar{X}^2} (RS_x^2 - \rho S_x S_y), \quad \text{dan} \quad \text{Bias} \quad \text{relatif}$$

$$\frac{B(\hat{R})}{\hat{R}} = \frac{(1-f)}{n\bar{X}\bar{Y}} (RS_x^2 - \rho S_x S_y)$$

Apabila nilai populasi tidak diketahui dapat diperkirakan dari sampel, yaitu :

$$\hat{B}(\hat{R}) = \frac{(1-f)}{n\bar{X}^2} (\hat{R} s_x^2 - \rho s_x s_y)$$

4. Estimasi rasio yang konsisten

Estimasi rasio walaupun bias, tetapi merupakan penduga yang konsisten. Apabila sampel cukup besar maka nilai estimasi yang diperoleh akan mendekati dengan nilai sebenarnya. Bila sampel besar maka bukan saja besarnya kesalahan sampling akan mengecil, tetapi juga biasanya. Survei-survei di BPS sampel cukup besar, sehingga digunakan penduga rasio. Memang, estimasi rasio akan mempunyai kelemahan apabila dikehendaki penggolongan karakteristik yang sangat rinci. Penggolongan misalnya orang yang bekerja menurut umur tahunan dan lapangan pekerjaan. Apabila rasio diperhitungkan untuk masing-masing dan dijumlahkan maka bias masing-masing bagian populasi akan menambah bias. Bias dalam kondisi ini akan menjadi sangat besar.

5. Estimasi Selang Kepercayaan (*Confidence Interval*)

Estimasi selang kepercayaan $(1-\alpha).100$ % bagi rasio, rata-rata rasio, dan total rasio adalah sebagai berikut:

a. Rata-rata

$$\bar{y}_R - z_{\alpha/2} se(\bar{y}_R) \leq \bar{Y}_R \leq \bar{y}_R + z_{\alpha/2} se(\bar{y}_R)$$

b. Total

$$\hat{Y}_R - z_{\alpha/2} se(\hat{Y}_R) \leq Y_R \leq \hat{Y}_R + z_{\alpha/2} se(\hat{Y}_R)$$

c. Rasio

$$\hat{R} - z_{\alpha/2} se(\hat{R}) \leq R \leq \hat{R} + z_{\alpha/2} se(\hat{R})$$

BAB VIII. SAMPLING KLASTER (CLUSTER SAMPLING)

8.1. Pendahuluan

Pada penarikan sampel acak sederhana, elemen-elemen/ unit-unit analisis seperti perusahaan/ usaha rumahtangga dan sebagainya telah tersusun dan tersedia dalam kerangka sampel. Dari daftar kerangka sampel ini langsung dipilih perusahaan/usaha atau rumahtangga dan informasinya dikumpulkan. Apabila penarikan sampel tidak langsung ke elemen/unit analisis, tetapi melalui kelompok dari unit analisis/elemen maka metode ini disebut sampling klaster. Persyaratan dari klaster adalah tidak boleh adanya unit yang tumpang tindih atau terlewat. Salah satu contoh klaster adalah blok sensus yang terdiri dari kelompok rumahtangga yang berdekatan pada suatu wilayah tertentu dengan batas jelas.

Ada dua cara untuk menentukan unit yang diteliti:

- a. Semua unit yang ada dalam klaster terpilih dimasukkan sebagai anggota sampel dan informasinya dikumpulkan. Cara penarikan sampel semacam ini disebut klaster. Dalam sampling, penarikan sampel ini termasuk metode sampling klaster satu tahap (single stage/ one stage cluster sampling).
- b. Sebagian unit yang ada dalam klaster terpilih dipilih dalam sampel dan hanya dari unit terpilih informasi dikumpulkan. Penarikan sampel ini termasuk metode sampling klaster bertahap (multi stage cluster sampling).

Alasan penting penggunaan sampling klaster adalah:

- a. Pengumpulan data pada unit yang berdekatan lebih mudah, murah, cepat, dan operasi lapangan lebih memungkinkan dibandingkan bila unit menyebar ke seluruh wilayah.
- b. Biaya transpor antar unit sampel mahal dan tidak sebanding dengan biaya penelitian per unit sampel.

- c. Bila dihadapkan pada kesulitan penyediaan kerangka sampel yang memuat seluruh unit sampel dalam populasi secara lengkap dan mutahir. Biaya untuk pembuatan kerangka sampel ini sangat mahal.

Pada umumnya di dalam praktek, suatu sampel terdiri dari sejumlah unit tertentu yang dipilih secara acak dari kerangka sampel mempunyai ragam yang lebih kecil dibandingkan dengan sejumlah sampel yang sama besarnya yang dipilih dari klaster terpilih. Dalam kondisi ini perlu ada keseimbangan antara besarnya sampel dan tingkat ketelitiannya serta biaya.

Banyaknya unit dalam klaster atau karakteristiknya dapat dijadikan dasar penarikan sampel. Klaster akan mempunyai peluang terpilih didasarkan pada kriteria yang ditentukan sebelumnya. Oleh karena itu penarikan sampel klaster tersebut memenuhi persyaratan penarikan sampel berpeluang. Berbagai metode sampling yang telah dibahas sebelumnya misalnya acak sederhana, strata dan sistematik dapat diterapkan dalam sampling klaster.

8.2. Pengertian Klaster

Pada penjelasan di atas telah dijelaskan klaster adalah kelompok unit yang dapat terdiri dari satu atau lebih unit listing (daftar unit) yang digabung. Klaster terdiri dari unit listing dan unit listing terdiri dari elemen/unit analisis. Untuk lebih jelas contoh dapat dilihat dalam tabel 1.

Tabel 1 : Klaster, Unit Listing, Elemen/Unit Analisis, dan Aplikasi

Klaster	Unit Listing/ Daftar Unit	Elemen/Unit Analisis	Aplikasi
(1)	(2)	(3)	(4)
1. Blok Sensus	Rumahtangga	Orang	Estimasi banyaknya rumahtangga/ penduduk beserta karakteristiknya
2. Desa	Sekolah	Guru/ Murid	Estimasi banyaknya guru/ murid beserta karakteristiknya
3. Sekolah	Kelas	Murid	Estimasi banyaknya murid beserta karakteristiknya

4. Halaman buku	Baris	Kata	Estimasi banyaknya kata dalam buku
5. Bulan	Hari	Hari	Estimasi rata-rata kepadatan lalu lintas.

Langkah-langkah:

- a. Bila kolom (1) merupakan unit sampling dan seluruh unit yang ada dalam klaster diteliti semua penarikan sampel tergolong klaster satu tahap. Apabila pada klaster terpilih dan didasarkan unit yang ada pada listing, dipilih unit dan selanjutnya diadakan penelitian pada unit tersebut, metode sampling tergolong pada klaster sampling dua tahap. Contoh lain misalnya klaster adalah sekolah, kemudian dari daftar kelas pada sekolah terpilih dipilih kelas dan selanjutnya dari kelas terpilih dipilih murid, maka cara penarikan ini disebut klaster sampling tiga tahap. Tahap di sini berarti tahapan dari penarikan sampel.
- b. Klaster dapat dipilih dengan berbagai metode sampling. Klaster dapat dipilih dengan acak sederhana, strata, peluang dengan ukuran besarnya klaster, sistematis dan sebagainya.
- c. Kerangka sampel yang harus disediakan tergantung pada metode sampling. Minimal untuk penarikan sampling tahap pertama (*primary sampling unit-psu*) harus tersedia dan selanjutnya kerangka sampel untuk dasar penarikan sampel elemen/unit analisis (*ultimate sampling unit – psu*) dapat dibentuk di lapangan. Sebagai contoh dilakukan penarikan sampel blok sensus, kemudian pada blok sensus terpilih dilakukan pendaftaran rumahtangga. Dengan dasar hasil pendaftaran rumahtangga pada blok sensus terpilih dipilih rumahtangga. Daftar rumahtangga dalam blok sensus terpilih disebut unit listing.

8.3. Klaster dengan Ukuran Sama (*Equal Cluster*)

- a. Notasi yang dipergunakan

Klaster dengan ukuran sama adalah klaster dimana banyaknya unit sampling dalam klaster sama antara satu klaster dengan klaster lainnya yaitu sebanyak M . Untuk dapat memahami sampling klaster, terlebih dahulu akan dibahas klaster dengan ukuran sama, terutama untuk memahami metode estimasi dan ragamnya.

Misalnya suatu populasi terdiri dari N klaster dan setiap klaster terdiri dari M elemen sebagai unit sampling selanjutnya disebut unit dan sebanyak n klaster dipilih secara acak sederhana. Seluruh elemen dalam klaster dikumpulkan informasinya.

Notasi dasar :

N : banyaknya klaster dalam populasi
 n : banyaknya klaster dalam sampel
 M : banyaknya elemen dalam klaster
 y_{ij} : nilai karakteristik dari elemen ke- j klaster ke- i , ($j = 1, 2, 3, \dots, M$)
 dan ($i = 1, 2, 3, \dots, N$)

Rata-rata per elemen dari klaster ke- i

$$\bar{y}_i = \frac{1}{M} \sum_{j=1}^M y_{ij}$$

Rata-rata per elemen dari rata-rata n sampel klaster

$$\bar{y}_n = \frac{1}{nM} \sum_{i=1}^n \sum_{j=1}^M y_{ij} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i.$$

Rata-rata per elemen dari seluruh elemen dalam populasi

$$\bar{Y} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M y_{ij} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i.$$

Varians (*deviasi standar kwadrat*) dari nilai karakteristik dalam klaster ke- i

$$S_i^2 = \frac{1}{M-1} \sum_{j=1}^M (y_{ij} - \bar{y}_i)^2$$

Rata-rata simpangan kuadrat di dalam klaster (*mean square within clusters*)

$$S_w^2 = \frac{1}{N} \sum_{i=1}^N S_i^2$$

Rata-rata simpangan kuadrat antar rata-rata klaster (*mean square between clusters*)

$$S_b^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{y}_i - \bar{Y})^2$$

Rata-rata simpangan kuadrat antar elemen di dalam populasi (*mean square between elemen*)

$$S^2 = \frac{1}{NM-1} \sum_{i=1}^N \sum_{j=1}^M (y_{ij} - \bar{Y})^2$$

Koefisien korelasi intraklas antar elemen di dalam klaster

$$\rho = \frac{\sum_{i=1}^N \sum_{j \neq k}^M (y_{ik} - \bar{Y})(y_{ij} - \bar{Y})}{(M-1)(NM-1)S^2}$$

Koefisien korelasi interklas ini dapat menunjukkan sampai sejauh mana hubungan karakteristik antara unit-unit dalam klaster. Makin besar nilai ρ berarti hubungan makin erat. Dalam praktek biasanya korelasi tersebut besar dan positif. Sebagai contoh misalnya karakteristik keadaan ekonomi rumahtangga, usaha rumahtangga dan sebagainya dalam suatu blok sensus cenderung kurang lebih sama, suatu blok sensus cenderung dihuni misalnya oleh rumahtangga golongan atas atau golongan bawah dan sebagainya. Apabila klaster terlalu homogen karakteristiknya maka penggunaan klaster satu tahap tidak efisien (akan diwakili karakteristik yang sama). Oleh karena itu

dalam penggunaan kluster lebih cenderung untuk diterapkan sampling bertahap.

Koefisien korelasi interklas bergerak antara $-1/(M-1) \leq \rho \leq 1$.

b. Estimasi total dan varians

Estimasi nilai karakteristik dari sampel sebesar n dari N kluster dengan ukuran sama, dapat dilakukan dengan metode acak sederhana, yaitu:

$$\hat{Y}_c = NM \bar{y}_n \text{ dengan } \bar{y}_n = \frac{1}{n} \sum_{i=1}^n \bar{y}_i.$$

Sedangkan perkiraan varians yang diperhitungkan dari sampel adalah:

$$v(\bar{y}_c) = \frac{1-f}{n} s_b^2 \text{ dengan } s_b^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y}_i - \bar{y}_n)^2$$

Dalam penarikan sampel di atas, maka unit yang diteliti akan berjumlah nM .

c. Efisiensi

Efisiensi sampling kluster dapat dibandingkan dengan acak sederhana, dengan membandingkan variansnya. Pendekatan efisiensi bila diperhitungkan dari sampel adalah:

$$E \approx \frac{v(\bar{y}_{SRS})}{v(\bar{y}_c)} \approx \frac{\frac{(1-f)}{nM} s^2}{\frac{(1-f)}{n} s_b^2} \approx \frac{s^2}{M s_b^2}$$

$$s^2 = \frac{\sum_{i=1}^n \sum_{j=1}^M (y_{ij} - \bar{y})^2}{nM - 1} \text{ dimana } \bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^M y_{ij}}{nM}$$

Bila $E > 1$, maka kluster lebih efisien

$E = 1$, maka sama efisien

$E < 1$, maka acak sederhana lebih efisien.

Dari rumus di atas jelas bahwa varians kluster sangat tergantung pada besaran dari n , M dan s_b^2 atau S_b^2 bila diperhitungkan dari seluruh kluster dalam populasi. Varians dari perkiraan rata-rata, bila diperhitungkan dari seluruh kluster dalam populasi:

$$V(\bar{y}_c) = \frac{1-f}{n} S_b^2 \text{ atau}$$

$$E = \frac{V(\bar{y}_{SRS})}{V(\bar{y}_c)} = \frac{\frac{(1-f)}{nM} S^2}{\frac{(1-f)}{n} S_b^2} = \frac{S^2}{MS_b^2}$$

Efisiensi suatu metode sampling dengan metode sampling lainnya adalah dengan membandingkan varians keduanya dan melihat nilai E seperti di atas.

Efek dari desain (*Design Effect*)

Efek dari desain adalah membandingkan suatu desain (metode sampling) dengan acak sederhana. Dalam hal ini penyebut selalu menunjukkan metode sampling acak sederhana.

Rumus dituliskan sebagai berikut:

$$Deff = 1 + (M - 1)\rho = \frac{V(\bar{y}_c)}{V(\bar{y}_{SRS})}$$

ρ biasanya diperkirakan dari survei yang pernah dilaksanakan dengan menghitung varians dari kluster dan dari acak sederhana.

Pendekatan deff dituliskan sebagai berikut:

$$Deff = 1 + (M - 1)roh = \frac{v(\bar{y}_c)}{v(\bar{y}_{SRS})}$$

Bila ukuran kluster diketahui dan deff dapat dihitung, maka korelasi intraklas dapat diperkirakan dengan rumus di atas dari suatu survei. Selanjutnya M dapat disesuaikan untuk mendapatkan deff yang sesuai, sehingga mengarah ke sampling bertahap. Dalam hal ini perlu penyesuaian n (kluster terpilih) dan m (unit terpilih dalam kluster).

ρ atau roh biasanya bernilai positif karena dalam kehidupan sehari-hari biasanya unit yang berdekatan mempunyai hubungannya erat satu sama lain.

d. Estimasi total dan variansnya

Selain perkiraan rata-rata dari hasil survei diharapkan perkiraan total seperti rumus sebelumnya. Dalam kluster estimasi total :

$$\hat{Y}_c = NM \bar{y}_n$$

$$v(\hat{Y}_c) = (1-f)N^2 M^2 \frac{s_b^2}{n}$$

NM merupakan banyaknya unit dalam populasi.

8.4. Klaster dengan Ukuran Tidak Sama (*Unequal Cluster*)

a. Estimasi

Klaster dengan ukuran tidak sama adalah klaster dimana banyaknya unit dalam klaster tidak sama antara satu klaster dengan klaster lainnya. Notasi yang dipergunakan sama dengan klaster berukuran sama, hanya mengganti M dengan M_i .

M_i : banyaknya unit dalam klaster ke- i

$$\bar{y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij}$$

Selanjutnya dalam penghitungan rata-rata unit pada populasi ada 3 cara estimasi, yaitu :

- 1). Rata-rata karakteristik per unit dari sebanyak n sampel klaster, yang diperhitungkan dari rata-rata klaster tanpa ditimbang dengan M_i

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n \bar{y}_i.$$

- 2). Rata-rata karakteristik per unit dari sebanyak n sampel klaster, yang diperhitungkan dari karakteristik seluruh unit dalam sampel

$$\bar{y}_n' = \frac{\sum_{i=1}^n M_i \bar{y}_i}{\sum_{i=1}^n M_i} = \frac{\sum_{i=1}^n \sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^n M_i}$$

- 3). Rata-rata karakteristik per unit dari sebanyak n sampel klaster, dengan memperhitungkan rata-rata unit unit per klaster dari populasi

$$\bar{y}_n^* = \frac{1}{n\bar{M}} \sum_{i=1}^n \sum_{j=1}^{M_i} y_{ij}, \text{ dengan } \bar{M} = \frac{1}{N} \sum_{i=1}^N M_i$$

Dari ketiga cara estimasi di atas yang paling sering digunakan adalah estimasi yang kedua dengan dasar

- cara pertama adalah rata-rata sederhana dan tidak mempertimbangkan ukuran klaster
- cara ketiga perlu diketahui ukuran klaster dalam populasi yang mungkin tidak diperoleh atau adanya perubahan ukuran klaster yang cepat pada saat kerangka sampel dan survei.

Antara cara kedua dan ketiga perlu dilihat sesuai dengan kondisi kerangka sampel yang tersedia.

Estimasi total dilakukan dengan mengalikan dengan banyaknya unit dalam populasi yaitu:

$$N\bar{M}' \text{ dengan } \bar{M}' = \sum_{i=1}^n M_i / n$$

atau

$$N\bar{M} \text{ dengan } \bar{M} = \sum_{i=1}^N M_i / N,$$

bila M_i diketahui untuk N klaster

sehingga estimasi totalnya menjadi

$$\hat{Y} = N\bar{M}' \bar{y}_n' \text{ atau } \hat{Y} = N\bar{M} \bar{y}_n^*$$

Varians

Seperti halnya perkiraan rata-rata atau total, maka perkiraan varians diperhitungkan juga dengan 3 cara, yaitu

$$1) v(\bar{y}_n) = \frac{1-f}{n} s_b^2 \text{ dengan } s_b^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y}_i - \bar{y}_n)^2$$

$$2) v(\bar{y}_n') = \frac{1-f}{n} s_b'^2 \text{ dengan } s_b'^2 = \frac{1}{\bar{M}'^2 (n-1)} \sum_{i=1}^n M_i^2 (\bar{y}_i - \bar{y}_n')^2; \bar{M}' = \frac{1}{n} \sum_{i=1}^n M_i.$$

$$3) v(\bar{y}_n^*) = \frac{1-f}{n} s_b^{*2} \text{ dengan } s_b^{*2} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{M_i \bar{y}_i}{\bar{M}} - \bar{y}_n^* \right)^2 \text{ dan } \bar{M} = \frac{1}{N} \sum_{i=1}^N M_i.$$

Penghitungan estimasi dari varians di atas dilakukan bila penarikan sampel dilakukan dengan acak sederhana dan sampling satu tahap. Metode ini dapat dikembangkan dengan:

- Sampling untuk proporsi
- Sampling dengan menggunakan peluang dan ukuran klaster (unit dalam klaster, atau salah satu karakteristik lainnya) sebagai dasar penarikan sampel. Metode ini disebut klaster dengan peluang sebanding dengan ukuran klaster (probability proportional to size sampling atau disingkat PPS sampling)
- Klaster dengan strata (*Stratified Cluster Sampling*)
- Estimasi dengan ratio dan sebagainya.

Sampling Klaster untuk Proporsi

Misalkan y_{ij} adalah merupakan suatu nilai variabel kualitatif, tidak mempunyai satuan ukur, dan terdiri atas kategori-kategori yang kongkrit. Elemen-elemen dengan ciri termasuk suatu kategori (yang diperhatikan) masing-masing diberi nilai dummy sama dengan 1, sedangkan selainnya diberi nilai 0. Dengan pengertian ini, P (proporsi) dapat dipandang sebagai rata-rata populasi data variabel dari suatu kategori tertentu.

Estimator tak bias bagi proporsi elemen-elemen populasi yang memiliki suatu kategori adalah

$$P = \frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N M_i}$$

Terdapat tiga pilihan estimator bagi proporsi elemen-elemen yang memiliki kategori tertentu, yaitu :

- 1) Proporsi tak tertimbang

$$\hat{P}_c = \bar{p}_n = \frac{1}{n} \sum_{i=1}^n p_i \quad \text{dengan} \quad p_i = \frac{a_i}{M_i}$$

dan estimasi varians bagi \bar{p}_c adalah

$$v(\hat{P}_c) = v(\bar{p}_n) = \frac{1-f}{n(n-1)} \sum_{i=1}^n (p_i - \bar{p}_n)^2$$

2) Tertimbang terhadap banyaknya elemen (M_i) dalam sampel klaster

$$\hat{P}_c' = \bar{p}' = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n M_i}$$

dan estimasi varians bagi \bar{p}' adalah

$$v(\hat{P}_c') = \frac{(1-f)}{n} \frac{\sum_{i=1}^n M_i^2 (p_i - \bar{p}')^2}{\bar{M}'^2 (n-1)}$$

3) Tertimbang terhadap rata-rata banyaknya elemen per klaster dalam populasi

$$(\bar{M} = \frac{1}{N} \sum_{i=1}^N M_i)$$

$$\hat{P}_c^* = \bar{p}^* = \frac{\sum_{i=1}^n a_i}{n\bar{M}}$$

dan estimasi varians bagi \bar{p}^* adalah

$$v(\hat{P}_c^*) = \frac{(1-f)}{n(n-1)} \sum_{i=1}^n \left(\frac{M_i p_i}{\bar{M}} - \bar{p}^* \right)^2$$

Sampling Klaster dengan menggunakan peluang tidak sama (PPS Cluster Sampling)

p_i = peluang terpilihnya klaster ke- i

$$\sum_{i=1}^N p_i = 1$$

Bila digunakan peluang dengan pemulihan:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i} \sum_{j=1}^{M_i} y_{ij}, \quad p_i = \frac{X_i}{X} \quad \text{dan} \quad \sum_{i=1}^N p_i = 1$$

atau

$$z_{ij} = \frac{M_i}{M_0} \frac{y_{ij}}{p_i}$$

$$\bar{z}_{i.} = \frac{M_i}{M_0} \frac{\bar{y}_{i.}}{p_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_0 p_i}$$

atau

$$z_{i.} = \frac{\sum_{j=1}^{M_i} y_{ij}}{p_i} \quad (\text{estimasi populasi dari kluster ke-}i).$$

$$\bar{z}_n = \frac{\sum_{i=1}^n \bar{z}_{i.}}{n}$$

$$v(\bar{z}_n) = \frac{\sum_{i=1}^n (\bar{z}_{i.} - \bar{z}_n)^2}{n(n-1)}$$

Bila yang digunakan sebagai peluang $p_i = \frac{M_i}{M_0}$ dan dengan ulangan, maka:

$$\bar{z}_n = \frac{\sum_{i=1}^n \bar{y}_{i.}}{n}$$

$$v(\bar{z}_n) = \frac{\sum_{i=1}^n (y_{i.} - \bar{y}_n)^2}{n(n-1)}$$

Sampling Kluster dan Strata (Stratified Cluster Sampling)

Sampling kluster dapat juga diaplikasikan dalam sampling berstrata. Sesuai dengan prinsip strata yaitu membagi populasi menjadi sub-populasi baik untuk efisiensi disesuaikan level penyajian, maupun disesuaikan dengan keadaan administrasi.

Dengan demikian estimasi dilakukan melalui masing-masing strata dan terakhir disajikan estimasi populasi. Estimasi dilakukan melalui estimasi

total pada masing-masing strata, dijumlahkan menjadi nilai total populasi. Atau bila telah dihitung nilai rata-rata per strata, perlu ditimbang dengan banyaknya unit pada masing-masing strata.

Secara umum rumus adalah sebagai berikut:

$$\begin{aligned}
 1) \hat{Y}_{st} &= \sum_{h=1}^L N_h \bar{M}'_h \bar{y}_{hn_h} & v(\hat{Y}_{st}) &= \sum_{h=1}^L N_h^2 \bar{M}'_h{}^2 v(\bar{y}_n) \\
 \bar{M}'_h &= \frac{\sum_{i=1}^n M_{hi}}{n_h} & &= \sum_{h=1}^L v(\hat{Y}_h) \\
 2) \hat{Y}'_{st} &= \sum_{h=1}^L N_h \bar{M}'_h \bar{y}'_{hn_h} & v(\hat{Y}'_{st}) &= \sum_{h=1}^L N_h^2 \bar{M}'_h{}^2 v(\bar{y}'_n) \\
 \bar{M}'_h &= \frac{\sum_{i=1}^n M_{hi}}{n_h} & &= \sum_{h=1}^L v(\hat{Y}'_h) \\
 3) \hat{Y}_{st}^* &= \sum_{h=1}^L N_h \bar{M}_h^* \bar{y}_{hn_h}^* & v(\hat{Y}_{st}^*) &= \sum_{h=1}^L N_h^2 \bar{M}_h^{*2} v(\bar{y}_h^*) \\
 \bar{M}_h^* &= \frac{\sum_{i=1}^{N_h} M_{hi}}{N_h} & &= \sum_{h=1}^L v(\hat{Y}_h^*)
 \end{aligned}$$

Nilai rata-rata atau total disesuaikan dengan metode penarikan sampelnya.

Kesalahan Sampling (Standard Error/ Sampling Error)

Kesalahan sampling seperti metode sampling lainnya adalah merupakan akar dari varians:

$$\sqrt{v(\hat{Y}_c)} \quad \text{atau} \quad \sqrt{v(\hat{\hat{Y}}_c)}$$

Sedangkan relatif kesalahannya adalah:

$$\frac{\sqrt{v(\hat{Y}_c)}}{\hat{Y}_c} \quad \text{atau} \quad \frac{\sqrt{v(\hat{\hat{Y}}_c)}}{\hat{\hat{Y}}_c}$$

b. Ukuran Klaster yang Optimal

Pada umumnya sampling klaster digunakan dengan pemilihan sampel bertahap. Klaster satu tahap kurang efisien disebabkan ukuran klaster yang biasanya cukup besar dan karakteristiknya homogen. Oleh karena itu diadakan kompromi yaitu dengan memperbesar klaster terpilih, tetapi dengan memperkecil ukuran klaster. Unit-unit dalam klaster hanya akan diteliti sebagian. Banyaknya unit yang dipilih dalam setiap klaster perlu ditentukan yang optimal sehingga dapat disesuaikan dengan biaya yang tersedia atau tingkat ketelitian yang dikehendaki.

Data dari sensus atau survei sebelumnya dengan karakteristik yang sesuai dengan survei yang akan dilaksanakan dapat digunakan untuk kajian ini. Secara rinci akan dibahas pada sampling bertahap.

BAB IX. PENARIKAN SAMPEL BERTAHAP

(*Multi-Stage Sampling*)

9.1. Pendahuluan

Penarikan sampel bertahap sebenarnya tidak jauh berbeda dengan penarikan sampel satu tahap (*single stage sampling*), hanya saja ada perluasan dalam penarikan sampelnya. Secara sederhana dalam penarikan sampel satu tahap, penarikan sampel langsung pada unit-unit yang terdaftar pada kerangka sampel.

Penerapan penarikan sampel bertahap didasarkan pada pertimbangan-pertimbangan yang rasional yang meliputi :

- a. Tidak tersedianya kerangka sampel yang memuat unit-unit sampel yang terkecil (*ultimate sampling unit*);
- b. Untuk membangun kerangka sampel yang memuat unit-unit sampel yang terkecil memerlukan biaya, tenaga, dan waktu yang besar;
- c. Dengan menerapkan penarikan sampel bertahap, maka pengawasan lapangan lebih dapat ditingkatkan sehingga non sampling error dapat ditekan;
- d. Ditinjau dari segi biaya, penarikan sampel bertahap jauh lebih efisien dibandingkan penarikan sampel acak sederhana.

Dalam bagian ini akan disajikan struktur penarikan sampel dua tahap, estimasi total berikut variansnya dalam berbagai kombinasi metode penarikan sampel.

9.2. Total dan rata-rata populasi

Misalkan banyaknya unit yang dapat dijadikan dasar untuk penarikan sampel tahap pertama – *pstp* (*first stages sampling unit – fsu*) adalah N , dan banyaknya unit yang dapat dijadikan dasar penarikan sampel tahap kedua – *pstd* (*secondary sampling unit – ssu*) pada setiap unit penarikan sampel tahap pertama yang ke- i adalah M_i .

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

a. Total nilai karakteristik Y pada pstp ke- i adalah:

$$Y_i = \sum_{j=1}^{M_i} y_{ij}$$

b. Rata-rata nilai karakteristik pstp ke- i adalah:

$$\bar{Y}_i = \frac{Y_i}{M_i}$$

c. Total nilai karakteristik dalam populasi adalah:

$$Y = \sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij} = \sum_{i=1}^N Y_i = \sum_{i=1}^N M_i \bar{Y}_i$$

d. Rata-rata nilai karakteristik per unit pstp dalam populasi:

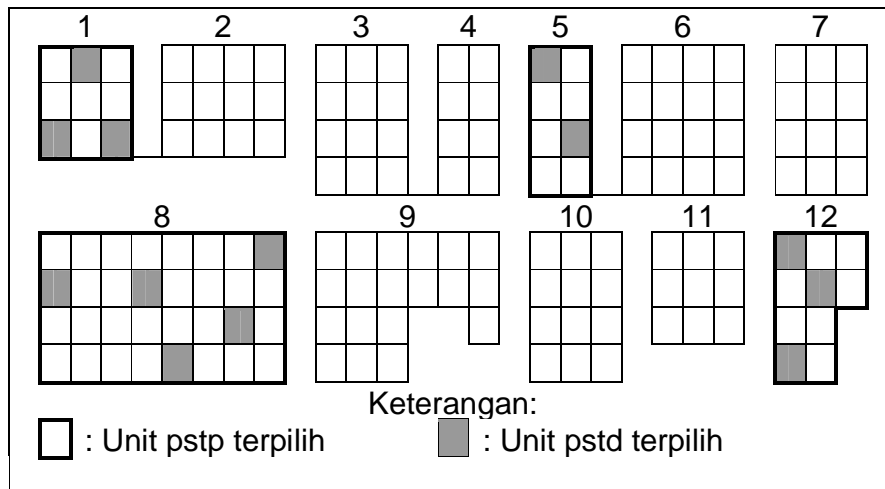
$$\bar{Y} = \frac{Y}{N} = \frac{1}{N} \sum_{i=1}^N M_i \bar{Y}_i$$

e. Rata-rata nilai karakteristik per unit pstd

$$\bar{Y} = \frac{Y}{M_o} = \frac{Y}{\sum_{i=1}^N M_i} = \frac{\sum_{i=1}^N M_i \bar{Y}_i}{\sum_{i=1}^N M_i}$$

9.3. Struktur penarikan sampel dua tahap, kedua tahap acak sederhana

Dari N unit pstp dipilih n unit, dan dari M_i unit pstp pada setiap ke- i dipilih sebanyak m_i unit. Penarikan sampel pada kedua tahap menerapkan metode penarikan sampel acak sederhana tanpa pemulihan. Banyaknya sampel pada pstd adalah $m_1+m_2+...+m_n$. Misalkan y_{ij} adalah nilai karakteristik Y pada unit pstd ke- j dan pstp ke- i yang terpilih ($j=1,2,...,m_i$) dan ($i=1,2,3,...,n$). Secara skematis penarikan sampel acak sederhana dua tahap dapat disajikan pada Gambar 9.1.



Gambar 9.1. Penyajian Secara Skematik
 Penarikan Sampel Acak Klaster Dua Tahap
 ($N=12$, $n=4$, $m_1=3$, $m_2=2$, $m_3=5$, $m_4=3$)

9.4. Pendugaan Parameter Populasi dari Sampel

a. Penarikan sampel dua tahap dengan metode PSAS dengan pemulihan

Rancangan penarikan sampel yang digunakan adalah rancangan penarikan sampel 2 tahap, dengan tahapan sebagai berikut :

- Tahap pertama, dari N unit sampling tahap pertama dipilih n unit dengan menerapkan metode PSAS dengan pemulihan,
- Tahap kedua, misalkan pada setiap unit pstp yang terpilih memuat M_i unit pstd, selanjutnya dipilih m_i unit dengan menerapkan metode PSAS dengan pemulihan.

Dari uraian rancangan penarikan sampel yang direncanakan dapat ditentukan peluang, dan fraksi sampling pada setiap tahap penarikan sampel seperti tercantum pada Tabel 9.1.

Tabel 9.1. Penarikan sampel dua tahap dengan metode PSAS dengan pemulihan

Tahap	Banyaknya unit di dalam		Metode penarikan sampel	Peluang pemilihan sampel	Fraksi sampling
	Populasi	Sampel			
Pertama	N	<i>n</i>	PSAS-DP	$\frac{1}{N}$	$\frac{n}{N}$
Kedua	M_i	m_i	PSAS-DP	$\frac{1}{M_i}$	$\frac{m_i}{M_i}$

Dengan demikian selanjutnya dapat ditentukan besarnya faktor pengali (*inflation factor*) pada masing-masing tahapan penarikan sampel yang merupakan kebalikan dari fraksi sampling dan factor pengali-pengalinya (*overall inflation factor*) adalah:

- Faktor pengali penarikan sampel tahap pertama, $F_1 = \frac{1}{f_1} = \frac{N}{n}$
- Faktor pengali penarikan sampel tahap kedua, $F_{2i} = \frac{1}{f_{2i}} = \frac{M_i}{m_i}$
- Faktor pengali keseluruhan, $F = F_1 \cdot F_{2i}$ yang berbeda antar pstp, kecuali bila $F_{2i} = F_2$ konstan, maka $F = F_1 \cdot F_2$ merupakan desain tertimbang sendiri (self-weighting design).

Misal y_{ij} menyatakan nilai karkteristik Y pada pengamatan ke- j dalam unit pstp ke- i , maka rumus umum estimasi yang tak bias bagi total adalah

$$\hat{Y} = \sum_{\text{sampel}} F_1 \cdot F_{2i} y_{ij} = \sum_{i,j} w_{ij} y_{ij}$$

dengan $w_{ij} = F_1 \cdot F_{2i} = \frac{N}{n} \frac{M_i}{m_i}$

Ambil w_{ij} dan diletakkan pada rumus umum, maka akan diperoleh estimasi bagi total karakteristik Y berdasarkan nilai-nilai sampel, yaitu :

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$$

$$= \frac{N}{n} \sum_{i=1}^n M_i \bar{y}_i$$

dan varians penarikan sampel bagi \hat{Y} adalah

$$V(\hat{Y}) = \frac{N^2 \bar{M}^2}{n} S_b^2 + \frac{N}{n} \sum_{i=1}^n \frac{M_i^2}{m_i} S_i^2$$

dengan $S_b^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{Y}_i - \bar{Y}_N)^2$ dan $S_i^2 = \frac{1}{M_i-1} \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_i)^2$

$$\bar{M} = \frac{M_0}{N} = \frac{1}{N} \sum_{i=1}^N M_i; \quad \frac{N-n}{N} \approx 1; \quad \frac{M_i - m_i}{M_i} \approx 1$$

S_b^2 dan S_i^2 masing-masing adalah varians antar unit penarikan sampel tahap pertama dan varians di dalam unit penarikan sampel tahap kedua pada unit penarikan sampel tahap pertama ke- i .

Dalam penarikan sample sampel dengan pemulihan, penduga tak bias bagi total karakteristik Y dapat didekati melalui estimasi yang diperoleh dari masing-masing unit penarikan sampel tahap pertama ke- i adalah

$$\hat{Y}_i = \frac{NM_i}{m_i} \sum_{j=1}^{m_i} y_{ij}, \text{ dalam hal ini } n=1.$$

Jadi ada sebanyak n estimasi dari setiap pstp. Dengan demikian penduga tak bias bagi varians \hat{Y} adalah

$$v(\hat{Y}) = \hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2$$

$$\bar{\hat{Y}} = \frac{N}{n} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i$$

Apabila penarikan sample tanpa pemulihan maka setiap pstp tidak merupakan estimasi yang bebas satu sama lain. Varians harus dihitung melalui pstp dan pstd.

$$v(\hat{Y}) = N^2 \bar{M}'^2 \frac{s_b^2}{n} + \frac{N}{n} \sum_{i=1}^n M_i'^2 \frac{s_i^2}{m_i}$$

$$s_b^2 = \frac{1}{\bar{M}'^2 (n-1)} \sum_{i=1}^n (\bar{y}_i - \bar{y}'_n)^2$$

$$\bar{y}'_n = \frac{\sum_{i=1}^n M_i \bar{y}_i}{\sum_{i=1}^n M_i}$$

$$s_i^2 = \frac{1}{(m_i - 1)} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

Rumus klaster yang kedua $\bar{M}' = \frac{1}{n} \sum_{i=1}^n M_i$

Dalam suatu survei dengan skala besar biasanya perkiraan varians didekati dengan penghitungan dengan pemulihan.

- b. Penarikan sampel dua tahap dengan metode penarikan sampel sebanding ukuran unit (*probability proportional to size – pps*) dan metode PSAS dengan pemulihan

Rancangan penarikan sampel yang direncanakan adalah penarikan sampel dua tahap, dengan tahapan sebagai berikut :

- Tahap pertama, dari N unit penarikan sampel tahap pertama dipilih n unit dengan menerapkan metode penarikan sampel sebanding terhadap ukuran unit x_i dengan pemulihan. Nilai-nilai x_i untuk seluruh unit untuk penarikan sampel tahap pertama harus tersedia sehingga dapat dihitung $X = \sum_{i=1}^N x_i$
- Tahap kedua, misalkan pada setiap unit pstp yang terpilih memuat M_i unit pstd, selanjutnya dipilih m_i unit dengan menerapkan metode PSAS dengan pemulihan.

Dari uraian rancangan penarikan sampel yang direncanakan dapat ditentukan peluang, dan fraksi sampling pada setiap tahap pemilihan sampel seperti tercantum pada Tabel 9.2.

Tabel 9.2. Rencana penarikan sampel 2 tahap dengan metode *pps* dan PSAS dengan pemulihan

Tahap	Banyaknya unit di dalam		Metode penarikan sampel	Peluang pemilihan sampel	Fraksi sampling
	Populasi	Sampel			
Pertama	N	n	<i>pps-dp</i>	$\frac{x_i}{X}$	$n \frac{x_i}{X}$
Kedua	M_i	m_i	<i>psas-dp</i>	$\frac{1}{M_i}$	$\frac{m_i}{M_i}$

Penduga tak bias bagi total karakteristik Y yang hanya didasarkan pada unit penarikan sampel tahap pertama ke-i adalah

$$\hat{Y}_i = \frac{1}{p_i} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij} = \frac{X}{x_i} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$$

Penduga tak bias bagi total populasi \hat{Y} dari seluruh unit penarikan sampel tahap pertama n adalah merupakan rata-rata sederhana dari \hat{Y}_i , yaitu

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i$$

Dengan demikian penduga tak bias bagi varians \hat{Y} adalah

$$\hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{Y}_i - \hat{Y})^2$$

9.5. Penentuan ukuran sampel

Secara umum tujuan daripada penarikan sampel adalah untuk menduga parameter-parameter populasi dengan presisi yang ditentukan pada biaya yang minimum. Presisi dalam hal ini biasanya diungkapkan dengan varians daripada statistik yang dihasilkan. Semakin rendah varians yang diharapkan, maka semakin tinggi biaya yang harus disiapkan. Fungsi biaya yang sederhana dalam kaitannya dengan rancangan penarikan sampel dua tahap dapat dinyatakan sebagai persamaan sebagai berikut :

$$C = c_1 n + c_2 \sum_{i=1}^n m_i \text{ atau } C = c_1 n + c_2 n \bar{m}$$

Biaya tetap (C_0) seperti biaya penentuan desain, analisis dan sebagainya yang tidak dipengaruhi besarnya sampel tidak dimasukkan dalam hitungan.

Ukuran sampel optimum dengan kendala varians minimum (diusahakan sekecil mungkin dengan biaya telah ditentukan) adalah sebagai berikut :

$$m_0 = \frac{\sqrt{V_2 c_1}}{\sqrt{V_1 c_2}} \text{ dan } n_0 = \frac{C \sqrt{V_1 / c_1}}{\sqrt{V_1 c_1} + \sqrt{V_2 c_2}} ; C: \text{ditentukan}$$

dengan :

$$V_1 = N^2 S_b^2, \text{ bila pstp menerapkan metode PSAS}$$

$$= \sum_{i=1}^N \frac{Y_i^2}{\pi_i} - Y^2, \text{ bila pstp menerapkan metode penarikan sampel } pps$$

$$V_2 = \sum_{i=1}^N M_i^2 S_i^2 = \frac{N}{n} \sum_{i=1}^N M_i^2 S_i^2 / m_i, \text{ bila pstd menerapkan metode PSAS}$$

$$= \sum_{i=1}^N \frac{1}{\pi_i} \left(\sum_{j=1}^{M_i} \frac{Y_{ij}^2}{\pi_{ij}} - Y_i^2 \right), \text{ bila pstd menerapkan metode penarikan sampel}$$

pps

$$\text{dan } V = \frac{V_1}{n} + \frac{V_2}{n m_0}$$

Bila varians (tingkat presisi), misalnya v telah ditetapkan, maka ukuran sampel tahap pertama yang optimum adalah

$$m_0 = \frac{\sqrt{V_2 c_1}}{\sqrt{V_1 c_2}} \text{ dan } n_0 = \frac{\sqrt{V_1 c_1} + \sqrt{V_2 c_2}}{V \sqrt{c_1 / V_1}}$$

Ukuran sampel tahap kedua yang optimum dapat juga dinyatakan dalam kaitannya dengan *intra-class correlation* ρ (rho), yaitu :

$$m_0 = \sqrt{\frac{c_1}{c_2} \left(\frac{1 - \rho}{\rho} \right)}$$

9.6 Penarikan Sampel Dua Tahap dalam Rancangan Penarikan Sampel Berstrata

a. Struktur penarikan sampel

Rancangan penarikan sampel bertahap dapat juga dilaksanakan dibawah rancangan penarikan sampel berstrata. Oleh sebab itu, sebelum penarikan sampel dilaksanakan, unit-unit penarikan sampel tahap pertama - *upstp* (*primary sampling unit – psu*) terlebih dahulu dikelompokkan berdasarkan kemiripan karakteristik yang dimilikinya. Variabel yang digunakan untuk dasar pembentukan strata dapat berupa variabel kuantitatif maupun variabel kategorik.

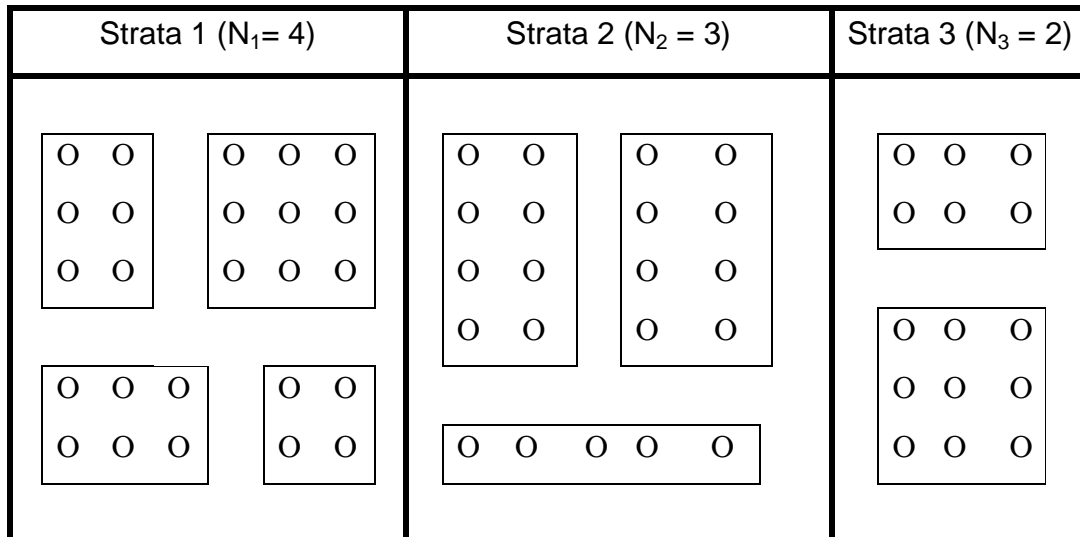
Misalkan suatu populasi dibagi habis menjadi L strata, N_h menyatakan banyaknya *upstp* dalam strata ke- h dan $N = \sum_{h=1}^L N_h$ adalah banyaknya

upstp pada seluruh strata, banyaknya *upstd* pada *upstp* ke- i dalam strata ke- h dinyatakan sebagai M_{hi} , dan $\sum_{h=1}^L \sum_i^{M_h} M_{hi}$ menyatakan

banyaknya unit penarikan sampel tahap kedua (*secondary sampling unit – upstd*) pada seluruh strata. Misalkan n_h adalah banyaknya sampel *upstp* yang ditarik dari strata ke- h , dan m_{hi} menyatakan banyaknya sampel *upstd* yang ditarik dari *upstp* ke- i strata ke- h .

Secara skematis struktur penarikan sampel dua tahap dalam rancangan penarikan sampel berstrata seperti tercantum pada Gambar 9.2.

Gambar 9.2 : Skema kerangka sampel penarikan sampel dua tahap dalam rancangan penarikan sampel berstrata



Keterangan : kotak bergaris tipis adalah *upstp*, lambang O adalah merupakan *upstd*

b. Estimasi total populasi

- i. Penarikan sampel tahap pertama dan kedua acak sederhana dengan pemulihan

Estimator tak bias bagi total karakteristik Y dalam strata dinyatakan sebagai Y_h , y_{hij} menyatakan nilai karakteristik Y pada *upstd* ke- j dalam *upstp* ke- i strata ke- h ($h = 1, 2, 3, \dots, L$; $i = 1, 2, 3, \dots, n_h$ dan $j = 1, 2, 3, \dots, m_{hi}$). Penarikan sampel tahap pertama pada setiap strata menerapkan metode penarikan sampel acak sederhana dengan pemulihan, yaitu dari N_h unit dipilih sebanyak n_h unit, demikian juga penarikan sampel pada tahap kedua dari M_{hi} *upstd* dipilih m_{hi} unit dengan menerapkan penarikan sampel acak sederhana dengan pemulihan. Jumlah sampel pada penarikan sampel tahap pertama dan

kedua masing-masing adalah $n = \sum_{h=1}^L n_h$ dan $m = \sum_{h=1}^L \sum_{i=1}^{n_h} m_{hi}$.

Tabel 9.3: Rencana Penarikan Sampel Acak Berstrata Dua Tahap.

Tahap	Banyaknya unit dalam strata h		Metode pemilihan sampel	Peluang pemilihan sampel	Fraksi sampel
	Universum	Sampel			
1	N_h	n_h	$psas-p$	$\frac{1}{N_h}$	$\frac{n_h}{N_h}$
2	M_{hi}	m_{hi}	$psas-p$	$\frac{1}{M_{hi}}$	$\frac{m_{hi}}{M_{hi}}$

Dugaan tak bias bagi total karakteristik Y strata ke- h adalah

$$\hat{Y}_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$

Untuk seluruh strata, estimasi total tak bias adalah merupakan penjumlahan dari seluruh estimasi total dari masing-masing strata, yakni

$$\hat{Y}_{st} = \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$

dan estimasi variansnya adalah

$$v(\hat{Y}) = \sum_{h=1}^L \frac{1}{n_h (n_h - 1)} \sum_{i=1}^{n_h} (\hat{Y}_{hi} - \hat{Y}_h)^2$$

dengan $\hat{Y}_{hi} = N_h \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$

Bila total ukuran sampel tahap pertama (n) dialokasikan sebanding terhadap ukuran strata (N_h), dan fraksi sampling pada penarikan sampel tahap kedua sama, yaitu sebesar k untuk seluruh unit penarikan sampel

tahap pertama, maka akan diperoleh estimator yang tertimbang secara otomatis (*self-weighting estimator*). Ukuran sampel pada penarikan sampel tahap kedua akan bervariasi sebanding terhadap ukuran populasi pada setiap unit penarikan sampel tahap pertama yang terpilih. Suatu rancangan penarikan sampel yang menghasilkan estimator yang tertimbang secara otomatis disebut sebagai rancangan penarikan sampel tertimbang otomatis (*self-weighting design*)

Ukuran sampel pada penarikan sampel tahap pertama untuk masing-masing strata adalah

$$n_h = \frac{N_h}{N} n$$

sedangkan ukuran sampel pada penarikan sampel tahap kedua untuk masing-masing unit penarikan sampel tahap pertama yang terpilih adalah

$$m_{hi} = k M_{hi} \text{ atau } k = \frac{m_{hi}}{M_{hi}}$$

Dengan demikian, maka estimator tak bias bagi total karakteristik Y adalah

$$\hat{Y} = \frac{N}{kn} \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} y_{hij}$$

dan estimasi variansnya

$$v(\hat{Y}) = \sum_{h=1}^L \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (\hat{Y}_{hi} - \hat{Y}_h)^2$$

dengan $\hat{Y}_{hi} = \frac{N_h}{k} \sum_{j=1}^{m_{hi}} y_{hij}$ dan $\hat{Y}_h = \frac{N}{k n} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} y_{hij}$

Penerapan rancangan penarikan sampel tertimbang otomatis memang menguntungkan bila ditinjau dari aspek pengolahan data (tabulasi), tetapi ditinjau dari aspek lapangan kurang praktis terlebih bila populasi penarikan sampel tahap kedua diperoleh langsung di lapangan.

- ii. Tahap pertama menerapkan metode penarikan sampel berpeluang sebanding terhadap ukuran dan tahap kedua acak sederhana

Estimator tak bias bagi total karakteristik Y dalam strata dinyatakan sebagai Y_h , dan y_{hij} menyatakan nilai karakteristik Y pada $upstd$ ke- j dalam $upstp$ ke- i strata ke- h ($h = 1, 2, 3, \dots, L$; $i = 1, 2, 3, \dots, n_h$ dan $j = 1, 2, 3, \dots, m_{hi}$). Penarikan sampel tahap pertama pada setiap strata menerapkan metode penarikan sampel berpeluang sebanding terhadap ukuran z_i , yaitu dari N_h unit dipilih sebanyak n_h unit, sedangkan penarikan sampel pada tahap kedua dari M_{hi} unit dipilih m_{hi} unit dengan menerapkan penarikan sampel acak sederhana dengan pemulihan.

Tabel 9.4 : Penarikan Sampel Acak Berlapis Dua Tahap
psps-p dan *psas-p*.

Tahap	Banyaknya unit dalam strata h		Metode pemilihan sampel	Peluang pemilihan sampel	Fraksi sampel
	Populasi	Sampel			
1	N_h	n_h	<i>psps-p</i>	$\frac{z_{hi}}{Z_h}$	$\frac{n_h z_{hi}}{Z_h}$
2	M_{hi}	m_{hi}	<i>psas-p</i>	$\frac{1}{M_{hi}}$	$\frac{m_{hi}}{M_{hi}}$

Dugaan total tak bias bagi karakteristik Y strata ke- h adalah

$$\hat{Y}_h = \frac{Z_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{z_{hi} m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$

Untuk seluruh strata, penduga total tak bias adalah merupakan penjumlahan dari seluruh penduga total dari masing-masing strata, yakni

$$\hat{Y}_{st} = \sum_{h=1}^L \frac{Z_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{z_{hi} m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$

dan estimasi variansnya adalah

$$v(\hat{Y}) = \sum_{h=1}^L \frac{1}{n_h (n_h - 1)} \sum_{i=1}^{n_h} (\hat{Y}_{hi} - \hat{Y}_h)^2$$

dengan $\hat{Y}_{hi} = \frac{Z_h}{n_h z_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$

c. Penduga rasio total dari dua variabel

Misalkan selain variabel Y , variabel X juga merupakan variabel yang diteliti, x_{hij} menyatakan nilai variabel X pada $upstd$ ke- j pada $upstp$ ke- i strata ke- h . Penduga tak bias bagi total yang diperoleh dari $upstp$ ke- i strata ke- h adalah

$$\hat{X}_{hi} = N_h \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{ij}$$

Penduga total bagi variabel X dari seluruh n $upstp$ adalah

$$\hat{X} = \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^n \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{ij}$$

Penduga varians bagi total X_h dan Y_h seperti yang tercantum pada anak bab sebelumnya. Penduga yang bias tetapi konsisten bagi rasio dari dua

nilai total pada strata ke- h $R_h = \frac{Y_h}{X_h}$ adalah merupakan rasio dari

\hat{Y}_h dan \hat{X}_h , yaitu :

$$\hat{R}_h = \frac{\hat{Y}_h}{\hat{X}_h}$$

dan penduga varians bagi \hat{R}_h adalah

$$v(\hat{R}_h) = \frac{1}{\hat{X}_h^2} [v(\hat{Y}_h) + \hat{R}_h^2 v(\hat{X}_h) - 2 \hat{R}_h \text{cov}(\hat{X}_h, \hat{Y}_h)]$$

dengan $\text{cov}(\hat{X}_h, \hat{Y}_h) = \frac{1}{n_h (n_h - 1)} \sum_{i=1}^{n_h} (\hat{X}_{hi} - \hat{X}_h)(\hat{Y}_{hi} - \hat{Y}_h)$ adalah merupakan

penduga tak bias bagi kovarians bagi \hat{Y}_h dan \hat{X}_h

Dengan demikian, maka penduga varians yang berbias bias tetapi konsisten bagi R adalah merupakan penjumlahan dari varians pada masing-masing strata, yaitu :

$$v(\hat{R}) = \sum_{h=1}^L v(\hat{R}_h) = \sum_{h=1}^L \frac{1}{\hat{X}_h^2} \left[v(\hat{Y}_h) + \hat{R}_h^2 v(\hat{X}_h) - 2 \hat{R}_h \text{cov}(\hat{X}_h, \hat{Y}_h) \right]$$

BAB X . TEKNIK BERWAWANCARA¹

10.1. Pendahuluan

Di dalam wawancara diperlukan kesediaan responden untuk memberikan keterangan. Kesediaan responden tersebut dapat dikondisikan dan biasanya sangat bergantung kepada sikap pewawancara pertama kali bertemu. Sikap duduk, kecerahan wajah, tutur kata, keramahan, kesabaran, dan keseluruhan penampilan pewawancara sangat mempengaruhi kelanjutan/kelancaran wawancara. Penampilan yang sopan dan ramah dengan sendirinya akan dapat mengurangi bahkan menghilangkan perasaan dan sikap penerimaan responden yang negatif, yang dapat merugikan penelitian, seperti: rasa curiga, rasa takut, rasa enggan, atau malu.

Beberapa hal yang penting untuk menciptakan hubungan baik dengan responden, antara lain:

- a. Dalam membuat janji wawancara dengan calon responden, sebaiknya memperhatikan waktu senggang dari responden tersebut, dan berusaha jangan sampai mengganggunya dalam kesibukan sehari-hari.
- b. Menggunakan bahasa yang dapat dimengerti oleh responden. Jika responden lebih mengerti bahasa daerah daripada bahasa Indonesia, maka gunakanlah bahasa daerah tersebut. Hal ini akan memperlancar jalannya wawancara.
- c. Sebelum memulai wawancara jangan lupa memperkenalkan diri, menunjukkan kartu pengenal jika perlu, serta menyebutkan lembaga atau badan yang menugaskannya. Kemudian menguraikan maksud wawancara serta tujuan penelitian yang dilakukan. Penting untuk disampaikan bahwa wawancara yang dilakukannya bukan suatu ujian atau test; tidak ada jawaban yang dibenarkan atau disalahkan dan informasikan bahwa semua pertanyaan yang diajukan akan mudah dijawab karena berhubungan

¹ Disalin dari Buku Pedoman Survei Pengetahuan Masyarakat Tentang Konstitusi dan MK

dengan pengalaman, kehidupan, pikiran dan perasaan responden sendiri. Sampaikanlah semuanya secara sederhana, tetapi cukup jelas.

- d. Dalam “obrolan” awal yang merupakan “intro” untuk membangun suasana yang kondusif ini jangan keluar dari konteks isi kuesioner. Arahkan perbincangan tersebut ke dalam isi kuesioner, namun demikian jangan menggunakan waktu terlalu lama.
- e. Berilah perhatian terhadap hal-hal yang sedang dibicarakan oleh responden selama berlangsungnya wawancara. Pewawancara dapat berperan sebagai seorang yang ingin tahu dan ingin belajar dari responden.
- f. Bila pewawancara kurang memahami sesuatu, maka dapat meminta responden tersebut untuk mengulangi jawaban terhadap pertanyaan yang diajukan. Sampaikan bahwa hal yang dikatakan responden sangat menarik, sehingga perlu dicatat atau mencoba mengulangi simpulan jawabannya guna meyakinkan bahwa yang dikatakan responden tidak salah ditafsirkan.
- g. Menjalankan tugasnya dengan penuh kepercayaan. Namun tidak dengan rasa percaya diri yang berlebihan, sehingga dirinya merasa lebih tinggi. Hal ini menimbulkan rasa antipati/rasa tidak suka dalam diri responden.
- h. Di dalam mengajukan pertanyaan yang bersifat sensitive, misalnya soal tindak kejahatan, dan soal-soal lain yang sejenis, usahakanlah agar pertanyaannya tidak menyinggung perasaan responden atau membuatnya merasa malu. Sebelum mengajukan pertanyaan tentang ini, dapat didahului dengan kata “maaf....”
- i. Gunakanlah waktu untuk wawancara dengan efektif, artinya dalam waktu yang sesingkat-singkatnya dapat diperoleh data yang sebanyak-banyaknya dan sejelas-jelasnya.
- j. Ucapkanlah terima kasih, bila wawancara tersebut telah selesai, karena responden telah meluangkan waktu serta memberikan keterangannya dalam wawancara tersebut.

10.2. Etika Berwawancara

Beberapa prinsip dasar yang dapat diikuti dalam wawancara, yaitu:

a. Berikan kesan pertama dengan baik

Pada saat pertama kali mengunjungi responden, usahakanlah agar responden merasa bebas/tidak tertekan. Pewawancara dapat memulai dengan ucapan selamat pagi/siang/sore, lalu dilanjutkan dengan memperkenalkan diri, misalnya: Nama saya Saya adalah petugas lapangan yang bertugas mengumpulkan keterangan mengenai guna

b. Usahakan selalu melakukan pendekatan yang positif

Jangan bersikap negatif dengan selalu meminta maaf atau jangan mengajukan pertanyaan sebagai berikut:

- Apakah Anda terlalu sibuk ?
- Apakah Anda berkebaratan meluangkan sedikit waktu untuk wawancara?
- Apakah Anda sudi menjawab berbagai pertanyaan yang akan saya ajukan ? Pertanyaan-pertanyaan seperti contoh di atas akan mengundang sikap tidak baik dari responden. Sebaiknya pewawancara bertanya seperti di bawah ini:

“Saya ingin mengajukan beberapa pertanyaan kepada Anda”. Kemudian mulailah mewawancarai responden.

c. Tekankan kerahasiaan jawaban

Sebelum pewawancara mengajukan pertanyaan, tekankan sekali lagi akan kerahasiaan jawaban yang diberikan. Tidak akan mencantumkan nama dalam melaporkan hasil penelitian ini, dan tidak menyebutkan nama responden lain atau pewawancara lain di hadapan responden.

d. Bila responden bertanya, jawablah semua pertanyaan-pertanyaan responden secara terus-terang

Sebelum responden bersedia untuk diwawancarai, mungkin responden akan mengajukan beberapa pertanyaan tentang penelitian dimaksud. Jawab

dengan tegas dan menyenangkan, tunjukkan kartu pengenal pewawancara. Tekankan sekali lagi kerahasiaan jawaban responden, apabila responden tampak ragu-ragu memberikan jawaban. Responden biasanya juga akan menanyakan tentang lamanya wawancara. Jawablah bahwa lama wawancara ini biasanya memakan waktu kurang lebih 30 menit.

e. Jangan memaksa responden yang tidak mau diajak wawancara

Bila seorang responden menolak untuk diwawancarai, janganlah terlalu memaksa. Jelaskanlah kegunaan dari keterangan-keterangan yang akan dikumpulkan, dan tawarkan kepadanya waktu lain jika ia tidak mau diwawancarai pada saat itu, dan lakukan kunjungan pada hari berikutnya. Apabila upaya ini tidak berhasil, ucapkan terima kasih, lalu minta kepada pihak Lapas/Rutan untuk berkenan mengganti responden.

f. Wawancarai responden sendirian

Apabila pada waktu wawancara hadir pihak ketiga, maka jelaskan pada responden bahwa pewawancara akan mengajukan pertanyaan-pertanyaan yang bersifat pribadi dan mintalah agar dapat mewawancarai di tempat lain, atau menjelaskan ke pihak ketiga bahwa keterangan yang diberikan responden tidak boleh diketahui oleh orang lain karena menyangkut kerahasiaannya.

g. Berusahalah bersikap netral selama wawancara

Biasanya bila seseorang menghadapi seorang tamu, maka akan bersifat sopan dan akan cenderung memberi jawaban yang menyenangkan. Jadi usahakan agar pewawancara bersikap senetral mungkin dalam mengajukan pertanyaan-pertanyaan dan menanggapi jawaban-jawaban responden. Jangan memberikan kesan melalui ekspresi muka atau suara bahwa jawaban yang diberikan oleh responden adalah “benar” atau “salah”. Tidak menyalahkan atau membenarkan secara langsung jawaban/pendapat yang dikemukakan oleh responden.

Ajukan pertanyaan sesuai dengan petunjuk yang telah diberikan dalam pedoman cara pengisian kuesioner. Bila ternyata responden memberikan jawaban yang meragukan cobalah mengajukan pertanyaan sebagai berikut:

- a. Dapatkah Anda menjelaskan hal tersebut sedikit lagi?
- b. Maaf saya tidak mendengar, dapatkah Anda menjelaskannya sekali lagi?

Kadang-kadang responden akan bertanya mengenai pendapat pewawancara tentang suatu pertanyaan yang terdapat dalam kuesioner. Dalam hal ini, jelaskan dengan sopan bahwa pendapat respondenlah yang penting dalam penelitian ini, bukan pendapat pewawancara.

- h. Jangan memberikan jawaban yang bersifat sugestif terhadap responden

Apabila jawaban responden tidak berhubungan dengan pertanyaan-pertanyaan yang diajukan, pewawancara jangan mengajukan pertanyaan-pertanyaan misalnya, "Saya rasa yang Anda maksud adalah bukankah demikian?" Hal tersebut tentunya akan mempengaruhi responden untuk menjawab seperti apa yang dikatakan oleh pewawancara, walaupun sebenarnya bukan jawaban tersebut yang dimaksudkan oleh responden. Hal ini tentunya dapat sangat merugikan penelitian. Karena itu, usahakanlah agar responden dapat memberikan penjelasan tentang jawaban seperti yang ia maksud.

- i. Jangan mengubah perkataan dan urutan pertanyaan kecuali ada petunjuk khusus

Perkataan dan urutan pertanyaan dalam kuesioner yang terstruktur tidak boleh diubah, kecuali ada petunjuk khusus. Apabila responden tidak mengerti atau salah mengerti akan pertanyaan-pertanyaan yang diajukan, maka ajukan sekali lagi pertanyaan tersebut secara perlahan-lahan dan jelas.

- j. Perlakukan responden yang ragu-ragu dengan bijaksana

Pada suatu saat dalam wawancara, mungkin terjadi bahwa responden akan menjawab: "saya tidak tahu", atau memberikan jawaban yang berlawanan dengan apa yang telah diutarakan responden sebelumnya atau menolak menjawab pertanyaan tertentu. Bila keadaan seperti ini terjadi, pewawancara harus mencari sebab mengapa responden bersikap demikian. Sebagai contoh, apabila pewawancara menduga bahwa responden merasa khawatir, coba usahakan untuk menghilangkan perasaan tersebut sebelum melanjutkan ke pertanyaan berikutnya.

Apabila responden memberikan jawaban yang tidak relevan dan mulai mengemukakan masalah-masalah pribadinya yang tidak berkaitan dengan pertanyaan, jangan menghentikan pembicaraannya secara mendadak, tapi arahkan pembicaraan tersebut kembali ke pertanyaan semula.

k. Jangan menunjukkan pengharapan tertentu terhadap responden

Pewawancara tidak boleh mengharapkan jawaban/sikap tertentu terhadap kemampuan dan pengetahuan dari responden. Jangan menganggap mereka yang berpendidikan rendah atau buta huruf tidak mengetahui hal-hal yang berkaitan dengan penelitian ini.

l. Jangan tergesa-gesa dalam melakukan wawancara

Ajukan pertanyaan-pertanyaan secara perlahan-lahan supaya responden mengerti maksudnya. Setelah pewawancara mengajukan pertanyaan, berilah waktu sesaat agar responden dapat memikirkan jawaban pertanyaan tersebut. Jika responden tidak diberi cukup waktu untuk memikirkannya, mungkin responden akan menjawab salah atau sekenanya. Berilah waktu yang cukup dan apabila terlihat bahwa responden memberi jawaban yang cepat untuk mempersingkat waktu wawancara, katakanlah "Anda tak perlu buru-buru, pikirkanlah baik-baik jawaban yang Anda berikan".

BAB XI. MAILING SISTEM ²

11.1. Pendahuluan

Salah satu keputusan awal yang harus diambil oleh peneliti survei, bahkan sebelum dia memilih sampel dan menyusun kuesioner, adalah apakah akan melaksanakan studi yang datanya diisi sendiri oleh responden (biasanya dengan menggunakan kuesioner yang dikirim melalui pos) atau studi yang datanya diperoleh dengan melakukan wawancara pada responden. Berikut adalah kelebihan dan kekurangan dari setiap studi tadi. Banyak bahan diskusi di bab 6 yang relevan dengan bahasan kita di bab ini.

11.2 Kelebihan dari Kuesioner yang Dikirim Melalui Pos

a. Sangat menghemat biaya

Belakangan ini biaya wawancara meningkat, begitu pula dengan upah pekerja. Biaya \$15 per wawancara (tidak termasuk biaya sampling atau analisa data) adalah biasa, dan biaya \$30 atau lebih per wawancara adalah tidak biasa. Banyak pimpinan studi wawancara melaksanakan kegiatan pelatihan petugas wawancara yang memakan waktu lama dan biaya yang banyak, mempekerjakan satu atau lebih pengawas (*supervisor*) kegiatan atau pengawas wawancara, dan bahkan membuka kantor di daerah sekitar yang menerima dan melatih petugas wawancara demi menjalin hubungan dengan publik. Meskipun kuesioner pada studi dengan surat biasanya lebih mahal daripada instrumen yang digunakan pada studi wawancara, mungkin dengan kertas yang berkualitas baik dan pencetakan dan sampul yang terperinci, namun tetap saja studi melalui pos masih jauh lebih murah daripada studi wawancara dengan jumlah sampelnya sama banyaknya. Ini adalah benar, bahkan bila digunakan pengiriman kelas pertama atau pos udara dan ada beberapa surat lanjutan, termasuk apabila lebih dari satu kuesioner dikirimkan kepada responden yang sama.

b. Menghemat waktu

Kuesioner yang dikirim melalui pos dapat dikirimkan kepada seluruh responden secara bersamaan, dan kebanyakan jawaban/balasan akan diterima dalam waktu satu minggu atau sedikit lebih lama (walaupun

pengembalian akhir memakan waktu beberapa minggu atau lebih), sedangkan wawancara biasanya dilakukan secara berurutan dan memakan waktu berbulan-bulan untuk bisa selesai.

² Diterjemahkan dari Bab 7 Buku METHODS OF SOCIAL RESEARCH
KARYA KENNETH D. BAILEY

c. Kuesioner bisa diisi sesuai dengan kenyamanan responden

Responden dapat meluangkan waktu yang lebih lama untuk mengisi kuesioner dibandingkan waktu yang diluangkannya pada studi wawancara, karena dia tidak dipaksa untuk melengkapi semua pertanyaan pada satu waktu. Dengan kuesioner yang dikirim melalui pos, responden bebas untuk menjawab satu atau lebih pertanyaan kapanpun ia memiliki waktu luang. Responden juga dapat menjawab pertanyaan yang mudah terlebih dahulu dan berpikir dahulu untuk menjawab pertanyaan yang lebih sulit.

d. Lebih menjamin anonimitas

Karena petugas wawancara yang dapat mengenali responden tidak ada, maka responden mungkin saja akan lebih bersedia memberikan jawaban yang tidak diinginkan secara sosial, atau jawaban yang melanggar norma-norma. Montero (1974) membandingkan tingkat respon (*response rate*) kuesioner yang dikirim melalui pos dengan studi wawancara untuk 23 pertanyaan yang tidak diinginkan secara sosial kepada sejumlah sampel orang Jepang-Amerika. Hasilnya diperoleh persentase yang lebih tinggi terhadap respon/jawaban yang tidak diinginkan secara sosial pada studi dengan kuesioner yang dikirim melalui pos untuk 18 dari 23 pertanyaan.

e. Kalimat yang dibuat standar

Perbandingan terhadap jawaban responden dimungkinkan oleh kenyataan bahwa setiap responden memberikan jawaban yang persis sama kalimatnya. Namun begitu, keuntungan ini bisa berkurang akibat perbedaan pengertian terhadap maksud pertanyaan diakibatkan perbedaan tingkat pendidikan antar responden.

f. Tidak ada bias petugas wawancara

Tidak ada kemungkinan bagi responden menjadi bias akibat petugas wawancara. Petugas wawancara dapat membuat jawaban menjadi bias dalam banyak cara, seperti gerak hati melalui perubahan nada suara, dengan mengasumsikan bahwa responden akan menjawab dengan suatu cara tertentu, atau mengatakan pendapat pribadinya kepada responden. Ada juga kemungkinan petugas akan salah membaca pertanyaan, sehingga responden akan jadi salah mengerti maksud dari petugas, atau petugas akan melakukan kesalahan dalam tulis menulis.

g. Keamanan informasi

Kuesioner yang dikirim melalui pos memungkinkan responden untuk mengonsultasikan catatan, membicarakan dengan koleganya, atau melakukan penyelidikan sebelum menjawab, sedangkan wawancara tidak memungkinkan hal ini.

h. Kemampuan mengakses

Responden yang tersebar luas secara geografis bisa dicapai seluruhnya hanya dengan seharga sebuah perangko, dibandingkan dengan biaya perjalanan yang mahal untuk petugas wawancara.

11.3 Kelemahan Kuesioner yang Dikirim Melalui Pos

a. Kurang Fleksibel

Tanpa kehadiran petugas wawancara, maka tidak akan ada variasi pada pertanyaan yang ditanyakan dan tidak ada penggalian yang lebih mendalam untuk mendapatkan jawaban yang lebih spesifik apabila jawaban pertama responden terlalu samar atau terlalu umum. Juga, apabila responden salah mengerti maksud pertanyaan, maka responden tidak dapat dikoreksi/diperbaiki. Terlebih lagi, tidak ada petugas wawancara yang "menyelamatkan" situasi dengan menenangkan responden yang marah yang tidak menyukai pertanyaan tertentu. Survei melalui surat, tidak hanya seringkali mendapatkan tingkat respon yang lebih rendah untuk pertanyaan-pertanyaan yang tidak diingingi secara sosial, tapi juga mendapatkan lebih banyak penghinaan dan caci makian dari responden yang marah yang merasa sangat terganggu dengan pertanyaan atau subyek yang sedang diteliti.

b. Tingkat respon rendah

Pada studi wawancara sebagian besar wawancaranya berhasil diselesaikan, dan alasan non respon biasanya dapat diketahui (misalnya responden meninggal, masalah bahasa). Namun, studi melalui surat kadang-kadang mendapatkan tingkat respon kurang dari 10%, dan 50 persen dianggap "mencukupi/memenuhi syarat" menurut Babbie (1973, hal 165). Selain itu, responden yang tidak menjawab umumnya bukan merupakan hasil pemilihan sampel secara acak tetapi memiliki

karakteristik yang membuat menjadi bias. Sebagai contoh, responden yang sudah tua lebih besar kecenderungannya untuk sakit dan tidak dapat menjawab. Responden yang sering berpindah-pindah lebih sering tidak memiliki alamat terbaru/terkini dan sulit menerima kuesioner. Responden yang tidak berpendidikan tidak dapat membaca kuesioner dan menuliskan jawaban. Bahkan banyak orang yang berpendidikan tinggi merasa bahwa mereka lebih dapat mengekspresikan dirinya dengan berbicara daripada menulis, atau terlalu malas untuk menulis paragraf yang panjang, atau merasa bahwa tata bahasa (*grammar*) atau ejaan mereka tidak sesuai dengan tingkat pendidikan mereka, dan merasa malu untuk mengajukan jawaban tertulis.

c. Hanya bisa untuk perilaku lisan

Tidak ada petugas wawancara yang hadir untuk mengamati perilaku non lisan atau melakukan pendekatan pribadi sesuai etnik responden, kelas sosial, dan karakteristik khas lainnya. Responden dari kalangan bawah akan mengganti identitas menjadi kalangan atas pada kuesioner yang dikirim melalui pos, tanpa teguran penolakan dari petugas wawancara.

d. Tidak ada kontrol atas lingkungan

Pada studi wawancara, petugas wawancara seringkali berjuang keras untuk memastikan setiap wawancara dilakukan pada lingkungan yang standar. Misalnya, wawancara akan dilakukan secara pribadi, tanpa suami/istri atau kehadiran anggota keluarga lainnya yang mendengar jawabannya, dan petugas akan mencoba memastikan bahwa ruangan yang digunakan cukup tenang dan responden tidak terburu-buru atau gugup. Pada kuesioner yang dikirim melalui pos tidak ada jaminan bahwa responden dapat menjawab pertanyaan secara pribadi. Suami/istri atau orang tua mungkin akan minta melihat kuesioner yang telah diisi dan mensensornya. Dan yang lebih buruk lagi adalah, ada orang lain yang mungkin mengisi kuesioner untuk menggantikan responden apabila responden merasa bahwa dirinya tidak memenuhi syarat atau sedang sibuk.

e. Tidak ada kontrol atas urutan pertanyaan

Karya agung urutan pertanyaan yang dirancang oleh peneliti untuk mengurangi bias respon, akan menjadi kacau oleh responden yang

membaca seluruh kuesioner sebelum menjawab, melewati/meloncati beberapa pertanyaan, atau tidak menjawab pertanyaan secara berurutan seperti tampilannya.

f. Banyak pertanyaan yang akan tetap tidak terjawab

Tanpa adanya pengawasan saat pengisian kuesioner, responden mungkin saja akan membiarkan beberapa pertanyaan tidak terjawab. Ketika 50 persen dari seluruh kuesioner kembali, peneliti akan mendapatkan hanya 10 persen responden yang menjawab pertanyaan sensitif.

g. Tidak dapat merekam jawaban spontan

Sulit mengumpulkan pendapat pertama yang spontan, karena responden memiliki kesempatan untuk menghapus jawaban yang diisi dengan buru-buru yang kemudian setelah dipertimbangkan kembali olehnya ternyata jawaban itu tidak diplomatis.

h. Sulit untuk memisahkan alamat yang salah dari non respon

(Lansing dan Morgan 1971, hal 117). Meskipun beberapa kuesioner yang gagal mencapai responden dikembalikan kepada peneliti, banyak diantaranya jatuh ke tangan penghuni baru yang kemudian membuang kuesioner tsb, sedangkan yang lain mengirimkannya ke alamat yang salah bukan mengembalikannya.

i. Tidak ada kontrol atas tanggal respon

Kurangnya kontrol atas waktu saat kuesioner diisi dapat sangat merusak suatu studi. Sebagai contoh, bila seseorang meneliti bencana alam dan badai atau gempa bumi terjadi ketika setengah dari responden telah selesai mengisi kuesioner mereka, maka jelas akan ada masalah dalam membandingkan jawaban sebelum dan setelah bencana terjadi. Tetapi ini juga merupakan masalah pada studi wawancara. Seorang petugas dapat memilih waktu baginya untuk tiba di suatu rumah, tetapi tidak dapat menjamin saat itu responden sedang ada di rumah atau setuju untuk diwawancara. Juga, wawancara sangat memakan waktu yang umumnya tidak dapat diselesaikan pada hari itu juga, padahal, paling tidak secara teori, kuesioner yang dikirim melalui pos bisa tiba pada hari yang sama.

j. Tidak dapat menggunakan format kuesioner yang rumit

Pertanyaan pada kuesioner yang dikirim melalui pos tidak hanya harus lebih mudah untuk dimengerti, tetapi juga format yang rumit dengan banyak kemungkinan pertanyaan akan membingungkan bagi sebagian besar responden. Pada banyak kasus kuesioner yang rumit dapat digunakan pada studi wawancara hanya karena petugas telah diberikan pelatihan yang intensif dalam memahami format kuesioner. Jelas bahwa responden pada studi melalui surat tidak dapat menerima pelatihan seperti itu dan akan menyerah/meninggalkan kuesioner yang penuh dengan panah dan "loncatan". Terlebih lagi, kalimat pertanyaan harus cukup sederhana bagi kebanyakan orang dengan tingkat pendidikan

rendah pada sampel untuk bisa mengerti. Hal ini akan berakhir pada penyederhanaan pertanyaan yang akan membuat responden dengan pendidikan tinggi merasa bahwa intelegensia/kecerdasan mereka telah dihina.

k. Kemungkinan sampel bias

Goode dan Hatt (1952, hal 174) menambahkan kekurangan lainnya: "Kuesioner," mereka mengatakan "bukan merupakan alat/sarana penelitian yang efektif bagi kebanyakan orang tetapi bisa untuk sekelompok responden yang dipilih secara khusus...karena diperoleh sampel yang bias." Mereka menyatakan kesimpulan ini berdasarkan fakta, seperti dibahas di atas, bahwa respon (dan juga non respon) bukanlah merupakan sampel random dari seluruh sampel tetapi umumnya bias dalam beberapa hal. Non respon cenderung merupakan orang-orang dengan pendidikan rendah dan yang sering pindah. Juga, pada masalah yang menyangkut emosi atau kontroversial, maka peneliti akan mendapatkan dua kemungkinan (*bimodal*) respon yaitu, mereka yang berkenan dan yang menentang tetapi menjawab, dan mereka yang netral tetapi tidak menjawab.

Namun begitu, beberapa studi telah menunjukkan bahwa kekurangan pada kuesioner yang dikirim melalui pos ini adalah tidak benar atau terlalu dilebih-lebihkan. McDonagh dan Rosenblum (1965) membandingkan hasil dari kuesioner yang dikirim melalui pos dan wawancara dengan mempelajari orang-orang yang menjawab kuesioner yang dikirim melalui pos dan orang-orang yang gagal menjawab. Mereka menyatakan:

Penemuan penting dalam studi ini yang berkaitan dengan metodologi adalah data menunjukkan bahwa tidak ada perbedaan yang signifikan secara statistik antara kuesioner yang dikirim melalui pos dan wawancara yang terstruktur yang keduanya menggunakan pertanyaan yang sama. Studi ini menyatakan bahwa kuesioner yang dikirim melalui pos memberikan respon yang mewakili (*representative*) walaupun tingkat pengembaliannya hanya setengah dari sampel populasi yang terpilih. Tidak ada perbedaan yang signifikan antara yang menjawab kuesioner yang dikirim melalui pos dengan yang tidak menjawab kuesioner wawancara. Orang yang bukan responden tampaknya tidak terlalu memilih pada beberapa variabel seperti yang diduga banyak peneliti. Hasil penelitian ini memberikan implikasi bahwa peneliti harus lebih percaya diri/yakin pada metode kuesioner sebagai sebuah alat utama penelitian (McDonagh dan Rosenblum 1965, pp. 135-6).

Ditambah lagi, seperti disebutkan di atas, beberapa studi (Knudsen et al. 1967; Montero 1974) telah menunjukkan bahwa survei melalui surat lebih baik daripada wawancara dalam mengumpulkan informasi yang sensitif atau hal tidak diinginkan secara sosial. Knudsen dan teman-temannya, yang mempelajari wanita yang mengalami kehamilan di luar nikah, mengatakan:

Pada situasi wawancara responden sepertinya mendukung publik dan batasan norma seksual, yang menurutnya dianut oleh petugas. Melalui perilaku yang taat ini si responden berharap dapat menghindari rasa malu karena telah melanggar norma. Kecenderungan mendukung norma seksual publik paling banyak dilakukan oleh masyarakat dengan status sosial-ekonomi rendah karena mereka mengacu pada norma yang ditampilkan oleh petugas dengan status yang lebih tinggi (Knudsen et al. 1967, p.296).

Pernyataan terakhir menyarankan dengan meningkatkan tingkat respon dari orang-orang dengan tingkat pendidikan rendah, studi wawancara menerima proporsi jawaban bias yang lebih besar. Makanya, responden kalangan bawah paling tidak senang mengisi kuesioner yang dikirim melalui pos dan merekalah yang apabila menjawab paling sering memberikan jawaban bias pada pertanyaan sensitif yang tidak diinginkan secara sosial.

11.4 Faktor-faktor yang Mempengaruhi Survei Melalui Surat

Selltiz et al. (1959) mendata tujuh variabel yang dapat mempengaruhi jumlah kuesioner yang kembali dan kecukupan data:

- a. Sponsor dukungan kuesioner
- b. Kecantikan format kuesioner
- c. Panjangnya kuesioner
- d. Kealamian surat penyerta yang meminta kerjasama
- e. Kemudahan mengisi kuesioner dan mengirimkannya kembali
- f. Bujukan untuk membalas
- g. Kepribadian orang yang dikirim kuesioner

Faktor lainnya mencakup:

- h. Tipe pengiriman (surat udara, balasan bisnis)
- i. Waktu minggu, bulan atau tahun ketika kuesioner dikirim
- j. Variasi tindak lanjut

Sponsor

Sponsor dapat mempengaruhi keinginan responden untuk mengembalikan kuesioner pos dengan menakut-nakuti responden atas legitimasi studi dan nilai, dengan membuat responden merasa takut untuk tidak menjawab, karena wibawa sponsor atau kekuasaan sponsor untuk memaksa responden menjawab.

Format dan Warna Kuesioner

Scott (1961, pp. 173-74) mendapati tingkat respon untuk kuesioner yang dicetak sebesar 95,2% dan untuk kuesioner hasil fotokopian sebesar 94,4%. Kuesioner yang dicetak dibalik sampul mendapatkan tingkat respon yang lebih tinggi (95,8%) daripada kuesioner pada lembar terpisah, hal ini mungkin mengindikasikan pilihan responden pada kuesioner yang lebih pendek.

Panjang Kuesioner

Scott (1961, pp. 173-74) membandingkan kuesioner versi satu halaman dan versi dua halaman. Versi dua halaman mendapatkan tingkat respon sebesar 94,8% dibandingkan 93,6% untuk versi satu halaman. Ini berarti kuesioner yang formatnya sedikit kacau lebih baik walaupun akhirnya menjadi kuesioner yang lebih panjang. Namun bagaimanapun tetap ada batasannya. Satu atau dua halaman hanya membuat perbedaan yang kecil, tapi setelah empat atau lima halaman tambahan maka akan menurunkan tingkat respon.

Surat Penyerta

Scott (1961) mencoba membuat dua surat yang sama kecuali yang satu berformat formal dan yang satunya informal. Surat yang formal berisi 22 kata ganti pribadi, sedangkan surat informal hanya 12. Tingkat respon jauh lebih tinggi pada versi yang lebih umum (91,4% vs. 89,6%), tapi perbedaan ini tidaklah signifikan secara statistik. Salah satu cara untuk membuat surat penyerta menjadi lebih personal adalah dengan menambahkan kata-kata tambahan pada akhir surat dengan tulisan tangan yang meminta responden untuk membalas/menjawab. Studi yang dilakukan Frazier dan Bird (1958) dan oleh Hope (1952) melaporkan peningkatan yang signifikan dalam tingkat respon (dari 25-31% dan dari 20-32% sebaliknya) dengan menggunakan permohonan tulisan tangan.

Scott membahas beberapa studi yang menemukan bahwa: (1) surat yang permisif mendapatkan tingkat respon yang lebih tinggi daripada surat yang tegas; (2) surat yang pendek dan mengena mendapatkan tingkat respon yang lebih tinggi daripada surat yang panjang, kata-kata tambahan lebih logis; (3) tidak ada perbedaan antara surat yang dialamatkan kepada "Dear Mr. Smith," "Dear Friend" dan "Dear Bulletin-User"; (4) tulisan alamat

dan tanda tangan dengan tulisan tangan tampaknya tidak membuat tingkat respon meningkat; dan (5) tidak ada perbedaan antara tandatangan asli dengan tandatangan faksimili. Scott (1961, p. 173) menyimpulkan bahwa "Pendek kata, isi surat jauh lebih penting daripada ornamen aksesorisnya"

Kemudahan dalam mengisi dan mengembalikan kuesioner

Adalah hal yang biasa bagi peneliti untuk menyediakan instruksi pengiriman surat secara tertulis dan menyediakan alamat yang telah tertulis dan amplop yang telah berperangko.

Bujukan untuk mengembalikan

Bujukan terbaik adalah dengan menyakinkan responden bahwa studi tsb sangat berarti dan kerjasama darinya sangatlah penting. Strategi lainnya adalah dengan membuat permintaan akan kebaikan responden, mengatakan padanya bahwa anda sangat memerlukan bantuannya.

Kepribadian responden

Pada sampel yang heterogen, ada konsensus bahwa orang yang paling tertarik pada studi tsb akan paling awal mengembalikan kuesioner.

Tipe Pengiriman

Telah diketahui bahwa kelas pengiriman kuesioner mempengaruhi tingkat pengembalian, dengan perangko yang lebih mahal umumnya akan merangsang pengembalian menjadi lebih banyak.

Hari, Minggu, atau Bulan Pengiriman

Peneliti tidak menginginkan kuesionernya tiba sehari sebelum liburan panjang, ketika sekolah dimulai kembali, atau ketika keluarga yang dituju sedang berlibur. Pada waktu-waktu seperti itu, responden punya hal lebih penting lainnya yang harus dikerjakan daripada menjawab kuesioner. Apabila kuesioner tiba saat responden sedang berlibur panjang, maka akan memperbesar peluang responden tidak akan menerima kuesioner atau akan menjadi masuk akal bahwa sudah “terlambat” untuk mengembalikannya.

Surat dan Telepon Lanjutan

Tampaknya telah menjadi konsensus bahwa tingkat respon akan meningkat tajam apabila diikuti dengan tindak lanjut, dan hampir semua studi dengan surat menggunakan tindak lanjut sebagai prosedur standar pelaksanaan. Tanpa tindak lanjut, maka hanya dapat diperoleh tingkat respon sebesar 50 atau 60%, sedangkan yang diikuti dengan tindak lanjut bisa diperoleh 70 atau 80% atau lebih.

