

METODE SAMPLING

1. Penggunaan Metode Sampling

Data dapat diperoleh dari berbagai sumber seperti catatan administrasi, sensus, maupun survei sampel. Survei sampel adalah salah satu cara untuk mengumpulkan data melalui sebagian unit baik dengan *probability* (peluang) maupun *non probability* tergantung dari tujuan survei. Bila informasi dari survei sampel ingin disajikan data untuk menarik kesimpulan dari sebuah populasi, maka perlu digunakan sampling berpeluang.

Pada uraian selanjutnya akan lebih difokuskan pada metode sampling dengan menggunakan *probability sampling*, walaupun tetap akan dijelaskan juga mengenai *non probability sampling*. Suatu survei sampel dapat dilakukan juga dengan kombinasi antara *probability sampling* dan *non probability sampling* yang biasanya saling melengkapi satu sama lain. Sebagai contoh pada pengumpulan data konsumsi/pengeluaran rumah tangga, pada saat ini konsumsi makanan jadi cenderung makin meningkat. Makanan jadi terdiri dari berbagai komposisi mulai dari beras, sayur-sayuran, daging, sampai ke minuman, dan sebagainya. Yang menjadi pertanyaan, misalnya data total konsumsi beras rumah tangga, berapa sebenarnya konsumsi beras per kapita. Untuk keperluan ini perlu diadakan studi mendalam dengan meneliti dari makanan jadi kira-kira berapa komposisi berasnya. Survei sampel dapat dilakukan secara *purposive* dengan mengambil sub-sampel dengan kriteria tertentu serta dilakukan penelitian secara mendalam.

Sebagai dasar untuk mengkaji alternatif metode sampling yang akan digunakan, minimal ditentukan terlebih dahulu:

- a) Obyek dan tujuan survei
- b) Populasi dan target populasi
- c) Variabel yang akan dikumpulkan
- d) Tingkat ketelitian yang dikehendaki
- e) Kerangka sampel yang tersedia
- f) Inferensial berupa kajian dan analisis

2. Konsep Definisi

Sebelum mendalami metode sampling perlu memahami berbagai konsep definisi (istilah) yang digunakan terutama yang berkaitan dengan penyediaan kerangka sampel dan estimasi, antara lain:

a. Populasi

Populasi merupakan kumpulan dari seluruh elemen yang memiliki karakteristik tertentu dan bisa dibedakan satu sama lain. Dalam populasi perlu ditentukan hal-hal seperti isi, unit, cakupan, dan waktu. Sebagai contoh populasi adalah semua penduduk yang bertempat tinggal dalam rumah tangga biasa di Kecamatan Polobangkang Selatan, Kabupaten Takalar, pada bulan Agustus tahun 2007. Contoh lain populasi adalah semua sekolah SD Negeri yang ada di Kecamatan Polobangkang Selatan, Kabupaten Takalar.

b. Target populasi

Target populasi merupakan sub populasi dari elemen yang ada pada populasi yang akan dicari indikator-indikatornya atau dengan kata lain target populasi adalah poluasi yang terkait dengan tujuan penelitian. Sebagai contoh adalah: penduduk usia sekolah.

c. Populasi survei

Pupulasi survei adalah populasi yang terliput dalam survei. Jumlah muatan populasi yang terliput akan kurang dari atau sama dengan populasi target. Sebagai contoh adalah: penduduk usia usia sekolah yang berhasil di survei.

d. Elemen

Elemen adalah unit yang digunakan untuk mendapatkan informasi, misalnya rumah tangga. Elemen biasanya merupakan unit sampling terkecil (*ultimate sampling unit*).

e. Unit observasi

Unit observasi adalah unit dimana informasinya diperoleh baik secara langsung maupun melalui responden tertentu. Elemen sangat erat kaitannya dengan unit observasi. Elemen bisa sama dengan unit observasi, sebagai contoh rumah tangga adalah selain sebagai elemen juga dapat sebagai unit observasi, misal pengumpulan data keadaan tempat tinggal. Unit observasi bisa individu dari elemen, misalnya anggota rumah tangga.

f. Unit sampling

Unit sampling adalah unit yang dijadikan dasar penarikan sampel baik berupa elemen maupun kumpulan elemen (klaster). Sebagai contoh, rumah tangga dapat dijadikan sebagai unit sampling dan atau kumpulan rumah tangga pada wilayah tertentu dapat dijadikan unit sampling seperti blok sensus, RT/RW, bahkan desa. Di luar rumah tangga cukup banyak unit yang dapat dijadikan unit sampling sesuai dengan tujuan survei seperti sekolah, kelas, Puskesmas, perusahaan/usaha dan sebagainya.

g. Daftar unit (*listing unit*)

Daftar unit adalah daftar unit yang digunakan sebagai dasar penarikan sampel seperti daftar blok sensus, daftar rumah tangga, daftar perusahaan/usaha, daftar sekolah, daftar kelas, daftar murid, dan sebagainya.

h. Kerangka sampel (*sampling frame*)

Kerangka sampel merupakan seluruh unit dalam populasi yang akan dijadikan dasar penarikan sampel. Kerangka sampel perlu diteliti apakah memenuhi persyaratan kerangka sampel. Oleh karena kerangka sampel merupakan hal yang penting dan menjadi dasar untuk penerapan metode sampling.

i. Metode sampling

Metode sampling khususnya pada *probability sampling* digunakan sebagai alat untuk penentuan sampel agar dapat menyajikan data yang mewakili populasi ditinjau dari efisiensi desain dan biayanya. Informasi yang dihasilkan dapat disajikan dan dianalisis secara deskriptif dan statistik dan mempunyai tingkat ketelitian tertentu (presisi) yang dapat diukur ditinjau dari metode samplingnya (*sampling error*).

j. Peluang terpilihnya setiap unit

Unit sampel harus dipilih sedemikian rupa sehingga dapat mewakili populasi dan ditentukan metode sampling yang efisien baik ditinjau dari metode sampling maupun biaya. Dengan ketentuan harus mewakili populasi, berarti setiap unit dalam populasi 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.

k. Kebenaran yang diukur

Estimasi yang berasal dari sampel harus dapat diukur kebenarannya/ketelitiannya ditinjau dari metode sampling yang digunakan. Pengukuran ketelitiannya berupa nilai total, rata-rata, proporsi, dan sebagainya, dengan menyajikan besarnya kesalahan sampling serta selang kepercayaannya (*level of confidence*). Pengukuran kesalahan di luar metode sampling (*non sampling error*) tidak dapat diukur secara langsung dari hasil survei tetapi harus melalui cara lain, seperti melalui pasca evaluasi survei (PES). Dengan diadakannya PES, minimal dapat diperkirakan adanya kesalahan cakupan (*coverage error*) dan kesalahan isian (*content error*). PES merupakan pencacahan ulang dan dilakukan setelah survei tetapi dengan selang waktu yang tidak terlalu lama. Dengan melisting ulang dapat diperkirakan ketelitian pada saat listing dan dengan interview ulang dapat diperkirakan kesalahan isian sehingga kesalahan cakupan dapat diperkirakan. PES tidak untuk memperbaiki estimasi tetapi untuk masukan analisis dan penyempurnaan yang akan datang. Bila kesalahan

cakupan cukup signifikan, maka hasil tes dapat digunakan untuk acuan menyusun *adjustment factor* hasil sensus.

Pengukuran tingkat ketelitian disesuaikan dengan metode sampling yang digunakan dan dilakukan sebagai berikut:

- 1) Pada saat perencanaan berdasarkan data yang ada seperti hasil sensus atau survei-survei yang lalu digunakan untuk memperkirakan banyaknya sampel yang harus dipilih pada tingkat penyajian tertentu, dalam hal ini kecamatan, dan dengan tingkat ketelitian yang diinginkan/ditentukan. Tidak semua variabel diteliti tetapi dipilih beberapa variabel yang mewakili variabel dengan tingkat heterogenitas/homogenitas tinggi, sedang dan rendah, atau variabel dengan kejadian yang sering terjadi, sedang, atau langka (*rare cases*).
- 2) Besarnya sampel tergantung dari sifat variabel yaitu didasarkan tingkat heterogenitas/homogenitas karakteristik atau banyaknya kejadian. Makin heterogen karakteristik unit-unit dalam populasi makin besar sampel yang diperlukan. Untuk kajian penentuan besarnya sampel dengan penyajian tingkat kecamatan telah digunakan hasil Susenas 2001-2005.
- 3) Dengan telah ditentukan besarnya sampel tersebut selanjutnya dilakukan penarikan sampel dan penyusunan daftar sampel untuk pelaksanaan survei. Berdasarkan hasil pengolahan survei dihitung kembali perkiraan tingkat ketelitian sesuai metode sampling yang digunakan. Diharapkan hasil penghitungan *sampling error* antara perencanaan dan realisasi hasil survei tidak jauh berbeda.

1. Prasyarat

Desain sampling dengan peluang baru dapat diaplikasikan bila tersedia kerangka sampel sesuai metode sampling yang ditetapkan. Disamping itu metode sampling yang dipilih harus dapat diaplikasikan di lapangan ditinjau dari segi unit sampling dan biaya yang tersedia. Metode yang telah ditentukan harus benar-benar diikuti dan tidak boleh diubah.

m. Efisiensi

Efisiensi dalam hal ini adalah membandingkan suatu metode sampling dengan metode sampling lainnya dengan mengukur variansnya. Metode sampling dengan varians yang lebih kecil akan menunjukkan bahwa metode sampling yang digunakan lebih efisien dari metode sampling dengan varians yang lebih besar. Dalam hal ini perlu keseimbangan antara varians dan biaya. Hasil sensus atau survei yang lalu dapat digunakan untuk penghitungan efisiensi.

n. Efek dari desain (*design effect* atau *deff*)

Sebenarnya efek dari desain ini sama dengan efisiensi, hanya pada *design effect* adalah membandingkan suatu metode sampling dengan metode sampling dengan acak sederhana dengan elemen sampling. Dengan selalu membandingkan terhadap acak sederhana, maka dapat langsung dilihat metode sampling mana yang paling efisien yaitu dengan melihat *deff* yang paling kecil.

o. Statistik, Estimasi, dan Distribusi Sampling

1) Statistik

Statistik adalah nilai yang dihitung dari hasil survei sampel dari berbagai karakteristik dari variabel yang dikumpulkan melalui survei dengan tujuan untuk membuat estimasi populasi. Karakteristik dapat berasal dari variabel yang berhubungan dengan keadaan elemen dalam populasi seperti variabel SDGs (umur, pendidikan, kesehatan, keadaan tempat tinggal, dan sebagainya). Nilai yang dihitung dapat berupa angka mutlak, rasio, rata-rata, persentase, dan sebagainya atau kajian yang lebih mendalam seperti APM, kemiskinan, dan berbagai indikator SDGs lainnya.

2) Estimasi

Estimasi merupakan perkiraan nilai karakteristik variabel yang dikumpulkan pada survei dari sampel-sampel yang telah dipilih. Pada metode sampling telah disepakati adanya notasi dengan

“huruf besar” menyatakan data populasi dan “huruf kecil” menyatakan nilai sampel.

Beberapa notasi yang biasa digunakan antara lain:

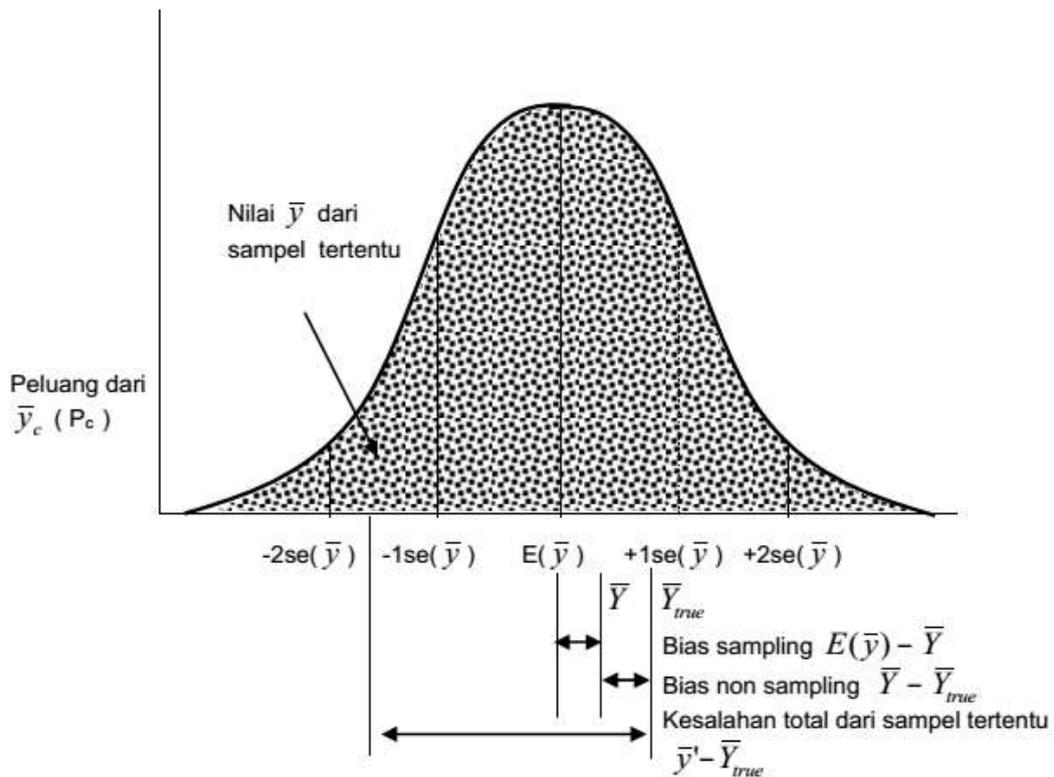
Rincian	Populasi	Sampel
1. Nilai karakteristik	Y	y
2. Rata-rata nilai karakteristik	\bar{Y}	$\bar{y} = \hat{Y}$
3. Banyaknya unit sampling	N	N
4. Varians	S^2	s^2

Notasi rinci akan disesuaikan dengan metode sampling yang digunakan.

3) Distribusi sampling

Penarikan sampel khususnya sampel berpeluang dilakukan dengan menggunakan angka random. Dengan demikian sampel yang terpilih tergantung dari angka randomnya. Secara keseluruhan akan terbentuk sejumlah kemungkinan gugus sampel (*all possible samples*), yang banyaknya tergantung dari metode sampling yang digunakan dan besarnya sampel yang ditentukan. Setiap *possible sample* merupakan satu gugus sampel dengan sebanyak n *sampling unit* dan masing-masing mempunyai nilai \bar{y} (rata-rata). Pada survei berskala besar seperti Susenas, distribusi sampling merupakan distribusi normal.

Sebaran dari nilai statistik, misalnya rata-rata (\bar{y}) dari setiap gugus sampel yang mungkin terbentuk disebut distribusi sampling dan dapat digambarkan sebagai berikut:



Nilai dari karakteristik pada distribusi sampling dinyatakan sebagai berikut:

- \bar{Y} : nilai rata-rata parameter dari populasi (rata-rata populasi)
- \bar{Y}_{true} : nilai rata-rata sesungguhnya dari populasi, yang tidak bisa diketahui
- \bar{y} : nilai rata-rata sampel
- $E(\bar{y})$: nilai rata-rata yang diperoleh dari rata-rata seluruh *possible sample*
- $E(\bar{y}) - \bar{Y}$: bias sampling, bila $E(\bar{y}) = \bar{Y}$ berarti rata-rata dari *all possible samples* sama dengan rata-rata populasi
- $\bar{Y} - \bar{Y}_{true}$: bias karena *non sampling*, bila berbedanya besar berarti kesalahan disebabkan *non sampling* seperti kesalahan isian, jawaban, pengolahan besar
- $\bar{y} - \bar{Y}_{true}$: kesalahan total, *sampling* dan *non sampling* dari suatu estimasi (gugus sampel tertentu)
- $se(\bar{y})$: *standard error* dari nilai rata-rata

Masalah pokok pada metode sampling adalah membuat sebaran/distribusi *possible samples* tidak terlalu melebar, sehingga menghasilkan *standard error* sesuai yang diharapkan.

Dalam bab ini diberikan berbagai metode sampling khusus *probability sampling* dengan tetap memberikan rumus-rumus dasar dengan memberikan:

- a) Contoh sederhana untuk memudahkan pengertian
- b) Contoh aplikasi sederhana dengan jumlah unit yang kecil termasuk persyaratan kerangka sampelnya

Untuk lebih jelasnya diberikan beberapa contoh penerapan *probability sampling* yang mungkin dilakukan.

Probability Sampling dan Non Probability Sampling

Dengan *probability sampling* diharapkan dapat menarik kesimpulan tentang sebuah populasi, sehingga data yang dihasilkan dapat mewakili keadaan populasi. Setiap unit dalam populasi harus mempunyai peluang untuk terpilih dalam sampel. Untuk keperluan penarikan sampel diperlukan kerangka sampel. Oleh karena setiap unit dalam populasi mempunyai peluang untuk terpilih dalam sampel dan besarnya juga telah diperhitungkan, maka dimungkinkan untuk menghasilkan estimasi parameter dari populasi seperti total, rata-rata, proporsi, dan sebagainya.

Non probability sampling tidak dapat menghasilkan estimasi parameter karena tidak semua unit dalam populasi mempunyai peluang untuk terpilih dalam sampel. Meskipun demikian *non probability sampling* sering digunakan untuk suatu keperluan tertentu seperti pemasaran, opini publik, kepuasan konsumen, atau dikombinasikan dengan *probability sampling* seperti dijelaskan sebelumnya, yaitu untuk melengkapi data yang dikumpulkan melalui *probability sampling*. Secara lebih rinci akan dijelaskan kemudian.

a. Metode Sampling

Metode sampling yang akan diuraikan adalah *probability sampling* sebagai alat penentuan sampel yang dapat digunakan untuk memperkirakan nilai populasi beserta cara penarikan sampel dan estimasinya.

Metode sampling yang akan diuraikan meliputi:

Sampling Elemen

Pada sampling elemen, penarikan sampel langsung ke elemen yang ada dalam populasi, metode sampling yang diterapkan adalah:

- Pengambilan sampel acak sederhana (*simple random sampling/SRS*)
- Sampling sistematis (*systematic sampling*)
- Sampling berstrata (*stratified sampling*)
- Sampling berpeluang sebanding dengan ukuran unit (*probability proportional to size/pps*)

Sampling Klaster

Pada sampling klaster, penarikan sampel tidak langsung ke elemen yang ada dalam populasi tetapi melalui kelompok elemen dan dibedakan sebagai berikut:

- Sampling klaster satu tahap, berarti dipilih langsung klaster dan kemudian semua unit dalam klaster dikumpulkan informasinya
- Sampling bertahap, pada tahap pertama dipilih terlebih dahulu klaster dan pada tahap berikutnya pada klaster terpilih dipilih elemennya

Cara penarikan sampel pada sampling klaster seperti halnya sampling elemen dapat dilakukan melalui 4 cara, yaitu:

- 1) secara acak sederhana
- 2) secara sistematis
- 3) secara berstrata, dan
- 4) secara berpeluang sebanding dengan ukuran

Prosedur Pemilihan Sampel

Ada dua metode yang digunakan dalam memilih sampel, yaitu:

- a. Metode Lotere;

Metode ini menyerupai acara arisan, dimana setiap elemen-elemen (dapat berupa nama, nomor, dsb.) yang dikumpulkan menjadi satu dalam sebuah wadah yang diberi lubang kecil pada permukaannya. Wadah tersebut kemudian dikocok hingga elemen dalam wadah itu keluar satu per satu. Elemen yang keluar itulah yang terpilih sebagai sampel.

b. Menggunakan Tabel Angka Random (TAR)

TAR merupakan angka-angka acak yang disusun dalam sebuah tabel. TAR yang biasa digunakan adalah yang terdiri dari 2 halaman, dimana setiap halaman terdiri dari 40 kolom dan 30 baris.

Halaman 1/2

Baris	Kolom							
	(1-5)	(6-10)	(11-15)	(16-20)	(21-25)	(26-30)	(31-35)	(36-40)
1	88347	17286	78607	56395	57187	49184	28747	93067
2	57140	14727	84858	96891	08337	06006	76040	43189
3	74686	19219	00336	86883	08091	96975	99600	41765
4	68013	47831	62237	74722	43311	60190	71402	49379
5	57477	01083	54076	77307	26245	59383	27506	11435
.
.
.
30	99559	91771	51621	2	70	67839		

Angka 08 terletak di baris 5 kolom 8 dan 9

- 1) **Penentuan halaman** yang ingin dipakai dapat dilakukan dengan cara memakai “**hari**” saat pengambilan sampel. Dengan ketentuan hari Senin, Rabu, Jum’at atau Minggu dianggap sebagai hari yang mewakili bilangan ganjil sehingga apabila pengambilan sampel dilakukan pada hari tersebut berarti halaman yang digunakan adalah halaman 1. Demikian juga sebaliknya untuk hari Selasa, Kamis atau Sabtu dianggap sebagai hari yang mewakili bilangan genap sehingga apabila pengambilan sampel dilakukan pada hari tersebut berarti halaman yang digunakan adalah halaman 2.
- 2) **Penentuan baris**, dapat dilakukan dengan cara memakai “**tanggal**” saat pengambilan sampel.
- 3) **Penentuan kolom**, dapat dilakukan dengan cara memakai “**bulan**” saat pengambilan sampel.
- 4) Apabila halaman, baris dan kolom sudah diketahui, pencarian dilakukan kearah bawah sampai baris terakhir. Dan apabila masih belum ditemukan, diteruskan kembali ke baris 1 pada kolom berikutnya. Begitu seterusnya hingga jumlah sampel yang diinginkan terpenuhi.

Catatan: pada dasarnya semua angka-angka yang ada pada tabel angka random sudah acak semua, jadi bagaimanapun tekniknya selalu menghasilkan angka yang acak sepanjang dilakukan secara obyektif.

Contoh: Misalkan kita ingin mengetahui angka berapa dari nomor urut 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; yang terpilih sebagai sampel bila dilakukan pada hari Rabu tanggal 5 Agustus 2008.

Jawab: Diketahui jumlah populasi sama dengan 10 ($N=10$). Karena 10 terdiri dari dua digit maka angka random yang diambil juga dua digit. Sedangkan hari Rabu tanggal 5 Agustus berarti menggunakan halaman satu, baris 5, kolom 8. Dari TAR terlihat angka yang bersesuaian dengan itu adalah angka 08, sehingga yang terpilih adalah nomor urut 8.

Sampling Elemen

1) Pengambilan Sampel Acak Sederhana (PSAS)

Pengambilan Sampel Acak sederhana pada sampling elemen digunakan bila pada populasi hanya tersedia daftar unit tanpa variabel pendukung dan unit tersedia berupa elemen. Sampling elemen pada umumnya hanya digunakan pada populasi yang unitnya tidak terlalu banyak dan areanya tidak terlalu luas. Pada level kecamatan tidak mungkin dilakukan sampling elemen, misalnya rumah tangga sebagai elemen dan ingin langsung dilakukan penarikan sampel rumah tangga. Untuk melakukan penarikan sampel rumah tangga secara langsung diperlukan daftar rumah tangga berikut nama kepala rumah tangga dan alamat lengkap. Penyediaan daftar nama kepala rumah tangga dan alamatnya yang mutakhir akan sulit karena memakan waktu, tenaga, dan biaya yang besar. Untuk mengatasinya digunakan sampling klaster yang terlebih dahulu mengelompokkan rumah tangga pada unit-unit yang sudah tersedia seperti blok sensus, RT/RW.

Meskipun demikian karena metode sampling acak sederhana ini merupakan dasar bagi metode selanjutnya, maka pengertian dari metode sampling serta cara estimasinya akan dibahas terlebih dahulu. Metode sampling lainnya dapat dipandang sebagai modifikasi atau pengembangan dari metode sampling acak sederhana.

Yang tidak boleh dilupakan dalam penerapan *probability sampling* adalah tersedianya kerangka sampel yang memenuhi persyaratan yaitu tersedia sampai satuan unit terkecil yang akan digunakan sebagai dasar penarikan sampel, batas jelas, tidak tumpang tindih/terlewat, ada korelasi dengan data yang diteliti dan mutakhir. Untuk masalah kerangka sampel akan dibahas tersendiri. Pada pembahasan metode sampling diasumsikan kerangka sampel tidak ada masalah.

Pada metode sampling acak sederhana ini, penarikan sampel biasanya dilakukan dengan menggunakan tabel angka random dan yang akan dibahas disini terbatas pada penarikan sampel acak sederhana tanpa pemulihan (*without replacement*). Tanpa pemulihan diartikan bila suatu unit yang sudah terpilih tidak ada kemungkinan untuk terpilih lagi, seperti misalnya kocokan pada arisan, lotere, dan sebagainya.

2) **Sampling Sistematis**

Pada penarikan sampel acak sederhana setiap unit dipilih dengan menggunakan tabel angka random. Dengan demikian kita harus menarik sampel sebanyak n kali, misal dari suatu kecamatan harus dipilih 300 rumah tangga berarti harus mengambil angka random sebanyak 300 kali. Untuk memperingan penarikan sampel ini maka diterapkan penarikan sampel secara sistematis, dengan hanya mengambil satu angka random saja dan lainnya akan mengikuti dengan menghitung intervalnya.

Salah satu yang sederhana adalah penggunaan sistematis linear dengan cara sebagai berikut:

- a. Hitung interval, yaitu $\frac{N}{n}$
- b. Tentukan satu angka random yang lebih kecil atau sama dengan intervalnya. Angka random ini selanjutnya disebut angka random pertama R_1 .

Angka random selanjutnya

$$R_2 = R_1 + I$$

$$R_3 = R_2 + I = R_1 + 2I$$

$$R_n = R_{n-1} + I = R_1 + (n-1)I$$

R_n digunakan sebagai kontrol apakah penarikan sampel sudah benar.

Misal banyaknya unit dalam populasi $N = 30$ dan banyaknya unit dalam sampel $n = 5$, maka $I = 6$ sehingga $R_1 \leq 6$ katakan 2, maka yang harus dipilih adalah nomor 2, 8, 14, 20, dan 26 yang harus dicek dengan $R_n = R_1 + (n-1)I = 2 + 4(6) = 26$.

Selain untuk mempermudah penarikan sampel, penarikan sampel sistematis juga dapat meningkatkan efisiensi, misal dengan mengadakan pengaturan unit-unit (*systematic arrangement*). Seperti pada contoh soal acak sederhana, elemen terletak dengan urutan sebagai berikut:

Elemen	1	2	3	4	5
Nilai karakteristik	8	6	12	10	4

urutan letak elemen diubah menjadi:

Elemen	1	2	3	4	5
Nilai karakteristik	4	6	8	10	12

Elemen diurutkan menurut besarnya nilai karakteristik, dalam hal ini dari nilai terkecil ke nilai terbesar, sehingga kalau dilakukan penarikan sampel secara sistematis, sampel akan menyebar dengan terwakili dari nilai kecil sampai dengan nilai besar.

Pada pemberian nomor urut wilayah desa/kota dan blok sensus misalnya dari ujung barat daya secara zig-zag, salah satu cara

pemberian nomor urut ini dimaksud apabila dilakukan penarikan sampel secara sistematis, sampel akan menyebar di seluruh wilayah.

Metode estimasi yang digunakan, karena biasanya sampling sistematis semata-mata hanya untuk mempermudah penarikan sampel, adalah sama dengan acak sederhana.

3) **Sampling Berstrata**

Dalam metode sampling dikenal dengan istilah strata, yaitu mengelompokkan unit-unit dalam populasi menjadi strata, dengan tujuan untuk efisiensi penggunaan metode sampling atau untuk keperluan lain seperti domain penyajian (daerah perkotaan dan daerah pedesaan, daerah miskin dan bukan daerah miskin, atau daerah sulit dan bukan daerah sulit). Penggunaan stratifikasi untuk efisiensi metode sampling adalah dengan mengusahakan pengelompokan elemen yang karakteristiknya lebih homogen. Pembentukan strata dapat tidak langsung mengelompokkan elemen, tetapi unit level di atasnya, sebagai contoh desa/kota dapat dijadikan dasar pembentukan strata sedangkan unit samplingnya tetap rumah tangga. Pada setiap strata dapat digunakan metode sampling yang berbeda karena estimasi digambarkan dari masing-masing strata. Estimasi total merupakan penjumlahan dari estimasi masing-masing strata. Dengan sampel berstrata diharapkan diperoleh estimasi dengan presisi yang lebih tinggi dengan banyaknya sampel yang sama.

Keuntungan Penggunaan Sampling Berstrata

1. Dapat diperoleh estimasi dengan presisi lebih tinggi
2. Pada setiap strata dapat digunakan metode sampling berbeda
3. Strata dapat dianggap populasi sendiri
4. Dalam beberapa hal membawa manfaat pada pengelolaan administrasi

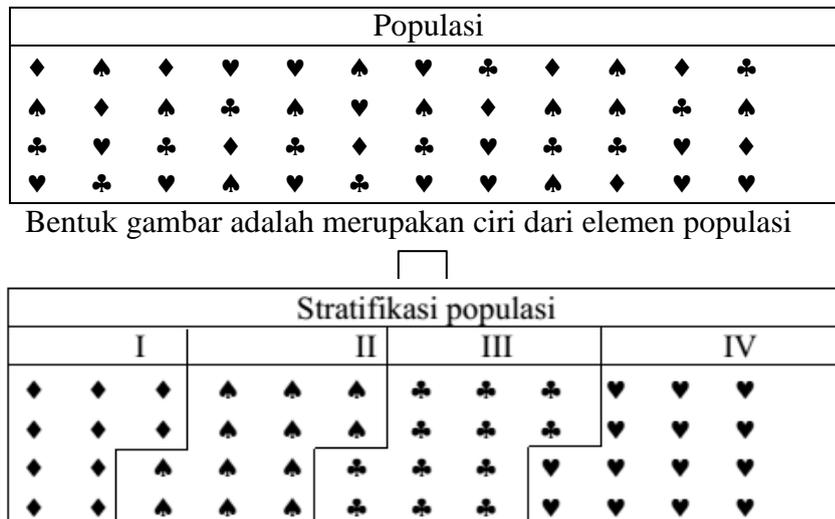
Pengelompokan unit sampling ke dalam strata yaitu membagi N unit sampling menjadi N_1, N_2, \dots, N_L yang masing-masing menunjukkan jumlah unit dalam strata, yaitu strata ke 1, ke 2, dan

seterusnya sampai dengan ke L . L menunjukkan banyak strata yang dibentuk pada populasi, sehingga $N_1 + N_2 + \dots + N_L = N$

Pembentukan Strata dengan Tujuan Meningkatkan Presisi

Untuk membentuk strata diperlukan variabel pendukung yang dapat digunakan untuk mengelompokkan unit sampling sehingga varians dari nilai variabel di dalam strata menjadi lebih homogen. Dan bila memungkinkan lebih baik lagi bila dapat diusahakan agar perbedaan rata-rata nilai karakteristik antar strata dibuat sebesar mungkin. Secara skematis pembentukan strata disajikan pada Gambar. 4.2 di bawah ini:

Gambar 4.2: Skema Pembentukan Strata



Berbagai cara dapat dilakukan untuk dasar pembentukan strata tergantung tujuan dari pembentukan strata dan sifat-sifat variabel yang akan dijadikan dasar pembentukan strata antara lain:

- a) Unit sampling itu sendiri, sebagai contoh blok sensus dikelompokkan menurut blok sensus dengan rumah tangga elit dan non-elit, blok sensus padat dan blok sensus tidak padat rumah tangga/penduduk. Jadi dalam kasus ini yang dikelompokkan adalah unit sampling itu sendiri dan karakteristiknya juga karakteristik dari blok sensus itu sendiri.

- b) Variabel wilayah administrasi misalnya desa perkotaan dan desa pedesaan.
- c) Variabel letak geografis, misalnya desa pantai dan desa bukan pantai.
- d) Variabel lainnya misalnya kepadatan penduduk, jenis lapangan usaha (daerah pertanian dan non pertanian).
- e) Perusahaan/usaha bisa dibedakan usaha skala besar, sedang, dan kecil, misalnya berdasarkan omzet atau jumlah tenaga kerja.
- f) Sekolah, bisa sekolah negeri dan sekolah swasta.

Untuk meningkatkan presisi maka perlu dipilih suatu variabel yang diperkirakan mempunyai korelasi dengan data yang akan dikumpulkan. Contoh variabel yang baik untuk dasar stratifikasi survei sosial ekonomi nasional antara lain pengelompokan wilayah elit dan non elit, atau daerah perkotaan dan daerah pedesaan. Dalam hal stratifikasi untuk meningkatkan presisi adalah membuat karakteristik di dalam strata homogen (varians diperkecil) dan antar strata heterogen (varians diperbesar). Dengan sampel yang sama diharapkan presisi lebih baik atau dengan tujuan yang lain yaitu dengan presisi yang sama dapat menurunkan banyaknya sampel yang harus terpilih, sehingga biaya dapat ditekan.

4) **Sampling Berpeluang Terhadap Ukuran (PPS)**

Pada acak sederhana penarikan sampel hanya didasarkan pada nomor urut unit dalam populasi. Penarikan acak sederhana ini menjadi kurang baik bila unit dalam populasi ukurannya bervariasi. Oleh karena itu digunakan informasi tambahan (*auxiliary information*) sebagai dasar pertimbangan di dalam penarikan sampel agar diperoleh *estimator* yang lebih efisien.

Informasi pendukung yang digunakan sebagai dasar penarikan sampel adalah informasi yang memiliki korelasi yang erat dengan variabel yang akan diteliti. Informasi pendukung yang dipertimbangkan sebagai dasar penarikan sampel selanjutnya disebut

ukuran. Prosedur penarikan sampel dimana peluang terpilihnya suatu unit sampel sebanding dengan ukuran disebut sebagai sampling berpeluang sebanding dengan ukuran unit atau *sampling with probability proportional to size* atau disingkat *pps sampling*.

Sampling Klaster

Metode sampling klaster sebenarnya sama dengan sampling elemen yang berupa acak sederhana, stratifikasi, sistematis dan sampling berpeluang. Perbedaan terletak pada unit sampling yang digunakan. Penarikan sampel pada klaster tidak langsung ke elemen, tetapi terlebih dahulu melalui kelompok elemen yang selanjutnya disebut unit sampling. Pada elemen sampling misalnya rumah tangga sebagai unit sampling, tetapi pada klaster, unit sampling adalah kelompok rumah tangga seperti desa, RT/RW, blok sensus, sekolah, dan sebagainya.

Ada dua cara dalam menentukan unit yang diteliti, yaitu:

- a) Semua elemen/unit yang ada dalam klaster terpilih dikumpulkan informasinya. Penarikan sampel ini disebut sampling satu tahap.
- b) Sebagian elemen/unit yang ada dalam klaster terpilih dikumpulkan informasinya. Penarikan sampel ini disebut sampling bertahap (*multistage cluster sampling*). Tahapan penarikan sampel dapat 2 tahap atau lebih dan ditinjau dari efisiensi desain sampel sebaiknya hanya 2 tahap karena makin banyak tahapannya makin kurang efisien.

Alasan penggunaan sampling klaster antara lain:

- a) Pengumpulan data pada unit yang berdekatan lebih mudah, murah, cepat, dan operasi lapangan lebih memungkinkan dibanding bila unit menyebar di seluruh populasi.
- b) Biaya transport antar individu unit mahal sehingga klaster akan lebih efisien.
- c) Kesulitan penyediaan kerangka sampel sampai ke elemen sebagai unit sampling.

Contoh: Klaster satu tahap

Tabel 1. Klaster, Unit listing, Elemen/unit Analisis, dan Aplikasi

Klaster	Unit listing/ daftar unit	Elemen/unit analisis	Aplikasi
(1)	(2)	(3)	(4)
1. Desa	Rumah tangga	Rumah tangga/orang	Estimasi jumlah rumah tangga/penduduk beserta karakteristiknya
2. Blok sensus	Rumah tangga	Rumah tangga/orang	Estimasi jumlah rumah tangga dan penduduk beserta karakteristiknya
3. Sekolah	Kelas	Murid	Estimasi banyaknya murid beserta karakteristiknya
4. Blok sensus	Rumah tangga pertanian	Rumah tangga pertanian	Estimasi banyaknya rumah tangga pertanian beserta karakteristiknya.
5. Bulan	Hari	Kendaraan	Estimasi kepadatan lalu lintas

Penghitungan pada sampling klaster untuk perkiraan nilai rata-rata karakteristik suatu variabel bukan merupakan rata-rata per klaster tetapi rata-rata per elemen. Dalam contoh di atas tidak diperkirakan rata-rata per desa, blok sensus, sekolah, dan sebagainya, tetapi ditujukan untuk penghitungan rata-rata per rumah tangga, orang, murid, dan sebagainya.

LATIHAN 4:

1. Jelaskan apa yang dimaksud dengan *probability sampling* dan *nonprobability sampling*!
2. Jelaskan pengertian dari populasi, target populasi, populasi survei, unit sampling, unit observasi dan kerangka sampel! Berikan contoh untuk menjelaskan konsep-konsep tersebut!
3. Apa perbedaan dari stratified sampling dan cluster sampling!

KESALAHAN DI LUAR DESAIN SAMPLING **(*NONSAMPLING ERROR*)**

1. Umum

Pengumpulan data perlu dirancang dengan baik dimulai dari penentuan obyek dan tujuan sampai dengan perencanaan lapangan, pengolahan, dan penyajian yang keseluruhannya merupakan desain pengumpulan data baik melalui sensus, survei maupun catatan administrasi. Di dalam desain survei termasuk di dalamnya desain sampel apabila pengumpulan data akan dilakukan secara sampel.

Desain sampel sendiri tergantung pada:

- a. Kerangka sampel yang tersedia yang harus memenuhi persyaratan kerangka sampel, yaitu tersedia sampai satuan unit terkecil sebagai dasar penarikan sampel, tidak tumpang tindih / terlewat, mempunyai batas jelas, ada korelasi dengan data yang dikumpulkan dan mutakhir.
- b. Penyediaan data pendukung untuk penetapan metode sampling yang sesuai.

Desain survei di luar desain sampel tergantung pada:

- a. Perencanaan survei secara menyeluruh, dimulai dari penetapan obyek dan tujuan sd perencanaan penyajian/analisis.
- b. Di dalam perencanaan sangat memegang peranan kegiatan antara lain:
 - 1) Penetapan topik dan variabel yang akan dikumpulkan
 - 2) Penetapan responden yang tepat
 - 3) Penetapan petugas dan pelatihan
 - 4) Desain dari kuesioner
 - 5) Rancang pelaksanaan lapangan dan organisasi

6) Rancangan pengolahan, penyajian, dan analisis

Pertanyaan: *Apakah hasil survei dijamin baik atau tingkat akurasi besar (total error/kesalahan kecil), bila perencanaan di atas telah dilakukan dengan baik?*

Jawaban: *Belum tentu.*

Mengapa: *Kesalahan kemungkinan besar juga terjadi di lapangan dan atau pada saat pengolahan.*

Jadi suatu survei yang didesain dengan baik belum tentu mempunyai tingkat akurasi yang baik, karena berbagai penyebab. Perencanaan memang perlu dirancang dengan baik agar mengurangi kesalahan dari hasil survei.

2. Desain Survei

Pada saat selesai mendesain suatu survei, maka sebelum dilaksanakan di lapangan perlu dikaji sekali lagi:

- i. Apakah perencanaan survei secara umum telah baik, dan telah ditetapkan organisasi serta penanggungjawabnya.
- ii. Survei desain dan struktur secara keseluruhan untuk penetapan unit sampling, unit observasi, metode pengumpulan data, lamanya interview, jadwal survei, efek terhadap non sampling error, dan sebagainya.
- iii. Kuesioner yang akan digunakan, seperti kelengkapan kuesioner listing, kuesioner kor, kuesioner modul, termasuk di dalamnya mengkaji ulang kemungkinan diterapkan.
- iv. Penetapan petugas termasuk kriteria, beban tugas, dan tatacara pelatihan.
- v. Penetapan responden serta kemungkinan kesulitan jawaban yang menyebabkan non response berkaitan variabel dan konsep definisi yang diterapkan.
- vi. Penetapan metode pengolahan yang berkaitan dengan kode, klasifikasi (menyiapkan pedoman editing coding), perekaman data, rancangan tabulasi, dan penyajian.

Apabila telah siap secara keseluruhan, maka segera diadakan persiapan dan pelaksanaan lapangan (pengumpulan data) sesuai jadwal ditetapkan. Perencana yang baik akan mempunyai efek positif terhadap hasil survei, tetapi perlu diimbangi dengan pelaksanaan lapangan dan pengolahan yang baik.

b. Pelaksanaan Lapangan dan Pengolahan

Sumber kesalahan dari pengumpulan data secara sampel di luar metode sampling sering dijumpai selama ini dan belum dilakukan pengkajian secara mendalam. Survei desain yang baik bila tidak diimbangi pelaksanaan lapangan dan pengolahan yang baik akan mengakibatkan hasil survei tidak akurat.

Kesalahan sampling dapat diukur dengan menggunakan formula sesuai dengan metode sampling yang digunakan, tetapi kalau data yang digunakan salah akan mengakibatkan hasil yang salah.

Sebagai contoh hasil listing rumahtangga dari suatu survei yang dilakukan tidak benar, tidak dilisting seluruh rumahtangga misalnya, akan mengakibatkan perkiraan (estimasi) menjadi lebih kecil dari kenyataan (under coverage).

Non sampling error yang terjadi di lapangan antara lain:

- i. Listing yang dilakukan tidak baik (tidak ada batas jelas, atau petugas kurang jeli).
- ii. Konsep-definisi yang salah interpretasi atau tidak diaplikasikan secara baik. Kuesioner yang terlalu tebal.
- iii. Kualifikasi petugas yang rendah dan pelatihan petugas lemah.
- iv. Petugas kurang bisa mendekati responden atau sebaliknya responden kurang partisipatif, bahkan cenderung menolak.
- v. Pengawasan, pemeriksaan, dan monitoring tidak dilakukan dengan baik.
- vi. Editing-coding, perekaman data tidak dilakukan dengan baik.
- vii. Dan sebagainya.

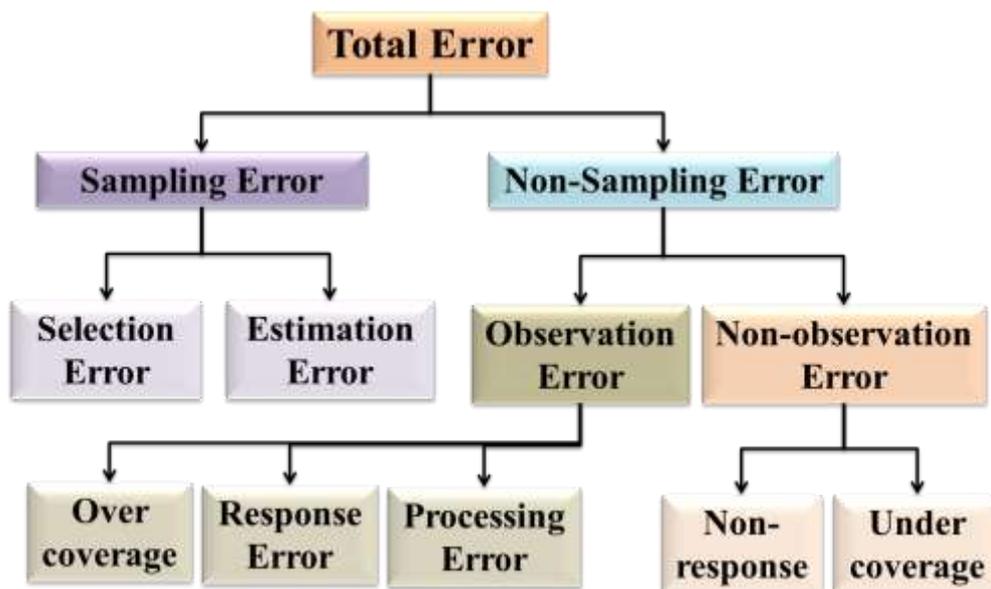
Kejadian tersebut boleh dikatakan merupakan kelemahan yang mengakibatkan adanya kesalahan di luar metode sampling (non sampling

error). Besarnya non sampling error sebenarnya dapat dikurangi dengan kesadaran mengikuti desain survei sesuai ketentuan oleh berbagai pihak. Kesalahan yang selama ini terjadi juga karena perencanaan kurang baik seperti kuesioner yang kurang baik susunannya atau pertanyaan terlalu banyak, sehingga responden maupun petugas menjadi jenuh. Petugas dapat saja mengarang jawaban atau tidak melakukan pencacahan. Sebaliknya responden asal-asalan dalam menjawab pertanyaan.

c. Jenis Kesalahan (Error)

Telah disinggung pada uraian sebelumnya berbagai contoh jenis kesalahan yang mungkin terjadi di luar desain sampling dan selanjutnya disebut *non sampling error*. Kesalahan yang terjadi dapat digambarkan sebagai berikut:

Gambar 1. Bagan Taxonomy of error



Bagan *taxonomy of error* di atas sebagai titik awal untuk mengantisipasi kemungkinan terjadi kesalahan, dimulai dari kejadian yang akan mengakibatkan *error* pada sampling dan *error* di luar metode sampling atau estimasi yang biasanya disebut *nonsampling error*. Para pelaksana pengumpulan data sudah dipastikan mengalami apa yang terjadi pada *non*

sampling error baik disadari maupun tanpa disadari. Disadari artinya petugas tidak dengan baik melakukan pengumpulan data atau responden sengaja menyembunyikan jawaban yang benar. Kesalahan ini perlu dihindari sejauh mungkin, agar data yang diperoleh dapat dipercaya oleh berbagai pihak. Organisasi suatu penyelenggara survei akan mempunyai kredibilitas baik apabila dapat menyajikan data yang dapat menggambarkan kondisi populasi. Walaupun tidak bisa terlepas dari kesalahan tetapi harus diusahakan kesalahan sekecil mungkin.

Sampling Error

Sampling error ini tidak akan dikaji secara mendalam terutama karena menyangkut teknik estimasi dalam metode sampling. *Sampling error* terdiri dari 2 jenis yaitu: *selection error* dan *estimation error*.

Selection error disebabkan karena kesalahan dalam menentukan sampel yang akan dipilih. Hal ini berkenaan dengan penggunaan kerangka sampel yang salah atau juga karena adanya pergantian sampel. Sedangkan *estimation error* dikarenakan penggunaan teknik sampling yang tidak sesuai dengan kondisi populasi ataupun penggunaan rumus yang tidak tepat serta penghitungan yang salah (termasuk penggunaan faktor pengali dan atau *weighting* yang salah). Tentu hal ini akan menyebabkan estimasi yang dihasilkan juga menjadi salah.

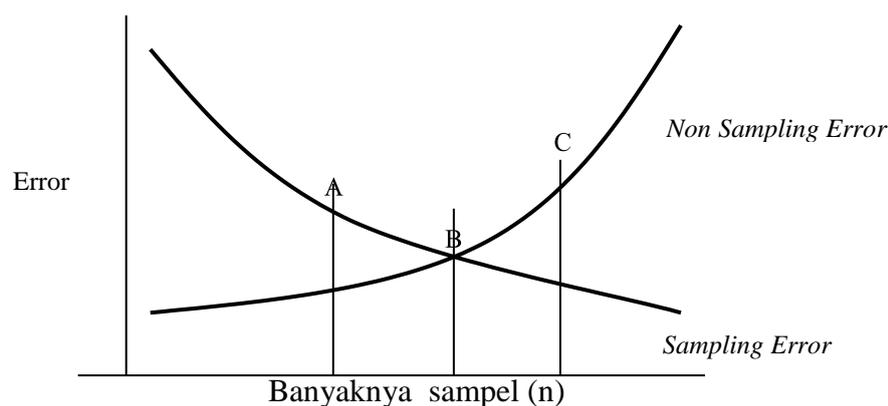
NonSampling Error

Penyebab terjadinya NonSampling error adalah *Observation error* dan *NonObservation error*. *Observation error* adalah kesalahan yang ditimbulkan akibat pengamatan, misalnya *over coverage*, *response error* dan *processing error*. Sedangkan pada *NonObservation error*, kesalahan yang ditimbulkan selain karena pengamatan, misalnya *nonresponse* dan *under coverage*.

Nonsampling error akan makin besar bila banyaknya sampel makin besar, karena secara otomatis pengumpulan data akan berhadapan dengan lebih banyak responden dan lebih banyak petugas. Dengan demikian

peluang untuk adanya kesalahan makin besar, apalagi bila dilakukan suatu sensus lengkap. Oleh karena itu dalam sensus lengkap, variabel yang dikumpulkan tidak boleh terlalu banyak. Pada kenyataan sering tidak dapat dihindari karena konsumen membutuhkan banyak jenis data untuk berbagai keperluan.

Gambar 2. Hubungan *Sampling Error* dan *NonSampling Error*



Dari gambar di atas dapat dilihat pengaruh banyaknya sampel (n) terhadap besarnya *error*, dengan penjelasan sebagai berikut:

- a) Ditinjau dari segi *sampling error*, makin besar sampel kesalahan makin kecil, dan akan menjadi nol untuk suatu sensus (pencacahan lengkap).
- b) Ditinjau dari segi *nonsampling error*, makin besar sampel kesalahan makin besar dan tidak terkontrol apabila pengumpulan data tidak dilakukan secara baik ditinjau dari semua aspek tahapan pengumpulan dan pengolahan data.

Ditinjau dari kedua sumber kesalahan tersebut, maka:

- a) Besarnya *sampling error* jauh lebih mudah dikontrol terutama bila tersedia kerangka sampel dan data pendukung yang baik. Kerangka sampel dan data pendukung yang baik artinya kerangka sampel dan data pendukung yang digunakan mempunyai non sampling error kecil atau bias kecil terhadap nilai parameter atau nilai populasi. Dalam metode sampling seperti dijelaskan terdahulu dapat ditetapkan metode yang

sesuai kondisi populasi yaitu penetapan sampel optimum dengan tingkat kesalahan ditetapkan atau kompromi antara biaya tersedia dan kesalahan yang dihasilkan.

- b) Besarnya *nonsampling error* jauh lebih sulit mengontrolnya dibanding *sampling error* karena tergantung dari berbagai aspek baik dari pengelola pengumpulan data, petugas maupun respondennya sendiri. Non sampling error dapat diperkecil dengan desain pengumpulan data atau survei, pelaksanaan lapangan dan pengolahan yang baik oleh berbagai pihak pengelola dan pelaksana survei. Pada suatu pelaksanaan pengumpulan data secara lengkap akan diperoleh *sampling error* sebesar nol. Kesalahan pada pengumpulan data secara lengkap hanya dipengaruhi non sampling error.

Dengan melihat sumber-sumber kesalahan di atas, maka dalam suatu pelaksanaan survei, total kesalahan adalah sebagai berikut:

$$\textit{Total Error} = \textit{Sampling Error} + \textit{NonSampling Error}$$

Usahakan total kesalahan atau total error sekecil mungkin. Ditinjau dari segi penggunaan metode sampling, khusus dalam pelaksanaan survei sampel kondisi pada kesalahan yang digambar pada B akan mempunyai total error paling kecil. Hal dapat diperkirakan pada saat perencanaan.

Pada kondisi A kesalahan karena sampling cukup besar, tetapi kesalahan karena non sampling error kecil, karena banyaknya sampel tidak memenuhi syarat. Sebaliknya pada kondisi C kesalahan karena sampling cukup kecil tetapi non sampling error besar. Pada desain sampel yang dicari kondisi B dengan persyaratan survei dilaksanakan dengan baik.

Pertanyaan: *Bagaimana mengurangi nonsampling error?*

Untuk menjawab pertanyaan ini maka perlu dipahami sumber-sumber dari *nonsampling error* serta cara mengatasinya.

Sumber-sumber NonSampling Error

Cukup banyak sumber-sumber kesalahan dari pelaksana survei yang menyebabkan *error* yang besar bila survei tidak dilaksanakan dengan baik. Juga perlu didukung oleh pelaksana (termasuk petugas) kesadaran pentingnya data dan akibatnya bila data tidak dikumpulkan secara benar dan serius. Data digunakan untuk berbagai perencanaan dan penetapan kebijakan sehingga bila kesalahan besar mengakibatkan kepercayaan terhadap data yang dikumpulkan tidak ada. Bahkan data yang dikumpulkan dengan biaya dan tenaga besar akan mubasir dan kepercayaan terhadap pengelola data menjadi berkurang atau data tidak dipakai.

Seperti telah disebutkan sebelumnya bahwa *nonsampling error* secara garis besar dikelompokkan menjadi 2 kelompok, yaitu:

- a) Kesalahan observasi (*observation error*) yaitu kesalahan akibat pengamatan yang kurang baik. Dalam kesalahan jenis ini terbagi menjadi (1). kesalahan karena berlebihnya unit sampling saat pembentukan kerangka sampel (*over coverage*), contoh: ada beberapa rumah tangga yang masuk ke dalam kerangka sampel padahal secara administrasi rumah tangga tersebut berada di luar wilayah pengamatan; (2). kesalahan yang berkaitan dengan pengumpulan yang menyangkut materi survei, petugas pengumpulan data, responden (*response error*), contohnya: petugas yang kurang memahami kuesioner ataupun responden yang menjawab semauanya; dan (3). kesalahan pada saat pengolahan (*processing error*), contohnya: salah memberi kode ataupun salah entri.
- b) Kesalahan di luar observasi (*NonObservation error*) yaitu kesalahan selain karena pengamatan. Dalam kesalahan jenis ini terbagi menjadi: (1). kesalahan akibat berkurangnya unit sampling dalam pembentukan kerangka sampel (*under coverage*), contohnya ada beberapa rumah tangga yang tidak masuk ke dalam kerangka sampel padahal secara administrasi rumah tangga tersebut berada di wilayah pengamatan; dan (2). kesalahan karena tidak ada respon dari unit sampling

(*nonresponse*), contohnya unit sampling yang tidak ditemukan ataupun menolak untuk di data.

Kesalahan Petugas Pengumpul Data

Kesalahan petugas pengumpul data sangat berpengaruh dalam menghasilkan data yang baik. Oleh karena itu penetapan petugas yang terus menerus untuk berbagai kegiatan survei dalam waktu yang berdekatan juga perlu menjadi pertimbangan untuk memperkecil terjadinya kesalahan. Adapun kesalahan-kesalahan yang mungkin terjadi yang diakibatkan oleh petugas pengumpul data antara lain:

- a) Petugas menganggap dirinya sudah memahami materi survei, sehingga tidak serius dalam pelatihan. Adanya penyempurnaan atau perubahan konsep misalnya menjadi tidak dihiraukan. Petugas telah tertanam dalam konsep lama dan tidak mengikuti perubahan. Akibatnya hasil survei akan kurang mengena sasaran.
- b) Petugas telah jenuh dengan pengumpulan data, sehingga tidak serius dalam pengumpulan data.
- c) Petugas telah hafal dengan pertanyaan pada kuesioner dan jawaban yang umum dari responden, sehingga ada kecenderungan untuk mengarahkan jawaban dari responden. Kejadian yang paling gawat, petugas memperkirakan sendiri jawaban responden. Responden lainnya diperkirakan jawaban dengan melihat kondisi dari luar.
- d) Contoh kasus yang pernah terjadi:
 - i. Pada suatu blok sensus terpilih terdapat beberapa penjual gado-gado, maka petugas hanya mencacah 2 atau 3 penjual dan lainnya tidak dilakukan wawancara hanya diwawancari nama saja. Langkah lanjut isian kuesioner lainnya dimodifikasi.
 - ii. Kasus ini terjadi pula pada misalnya usaha sablon, usaha angkot, dan sebagainya.
 - iii. Kasus lain bahkan lebih parah yaitu konsumsi/pengeluaran rumahtangga diperkirakan sendiri oleh petugas.

Kejadian semacam itu apabila terus menerus dilakukan akan menyebabkan data tidak akurat lagi dan kepercayaan terhadap pengelola survei akan hilang. Oleh karena itu secara teori petugas pengumpul data, sebaiknya maksimum 3 kali melakukan survei yang sama.

Pelatihan petugas salah satu tujuan adalah untuk mengurangi *nonsampling error* agar:

- a) Petugas menguasai materi survei, dimulai dari obyek dan tujuan survei, materi survei, konsep-definisi, teknik wawancara.
- b) Petugas pengawas harus mengikuti ketentuan pengawasan termasuk mengontrol kualitas jawaban.

Kurang Pemahaman Materi Survei

Pertanyaan sederhana pun dapat menjadi sumber kesalahan, apalagi pertanyaan yang sulit.

Sebagai contoh kesalahan yang terjadi yang berakibat non sampling error menjadi besar:

- a) Kesalahan penetapan anggota rumahtangga. Dimungkinkan petugas tidak memberi penjelasan pada responden, sehingga dijawab adalah anggota keluarga. Dalam hal ini termasuk anaknya yang tinggal di luar kota, sebaliknya pembantu rumahtangga yang tinggal serumah tidak dimasukkan. Padahal konsep yang ditetapkan adalah semua anggota yang tinggal bersama dalam satu atap dan makan dalam satu dapur. Apabila kejadian ini terjadi, maka data jumlah rumahtangga dan anggota rumahtangga (penduduk) akan menjadi lemah dan bercampur antara konsep rumahtangga dan atau KK (keluarga).
- b) Kesalahan dalam mengintepretasikan "mencari pekerjaan". Untuk mendapatkan data ini dalam survei angkatan kerja misalnya pertanyaan tidak sekedar "apakah mencari pekerjaan", tetapi ditanyakan secara bertahap sehingga memenuhi kriteria yang diinginkan seperti termasuk mencari pekerjaan bagi mereka yang masih ingin bekerja tetapi sudah bosan tidak cari kerja lagi, dan sebagainya.

- c) Dan sebagai misalnya kesalahan dalam penggolongan lapangan pekerjaan, jenis usaha, dan dalam pertanyaan konsumsi/pengeluaran rumahtangga (di dalamnya termasuk produksi sendiri, diberi, dan sebagainya).
- d) Non sampling error yang terjadi pada survei pertanian padi/palawija, yaitu pada data produksi per hektar (ubinan) dan data luas panen melalui perkiraan petugas pertanian (kasat mata, tanpa pengukuran). Dalam hal sampling sudah diperhitungkan perkiraan kesalahan karena sampling khususnya untuk produksi per hektar melalui ubinan. Pada perkiraan luas panen kesalahan karena sampling nol, karena tidak melalui survei sampel. Luas panen diperkirakan secara lengkap dari catatan administrasi petugas pertanian.

Kesalahan non sampling yang mungkin timbul antara lain:

- a) Pada perkiraan produksi per hektar kemungkinan adanya kesalahan pengukuran petak ubinan, kesalahan dalam penimbangan dan kemungkinan juga ada petugas yang tidak melakukan ubinan tetapi diperkirakan sendiri.
- b) Pada perkiraan luas panen, karena tidak melalui pengukuran areal panen, dipastikan terjadi kesalahan perkiraan yang sampai saat ini masih dipertanyakan. Metode pengumpulan data perlu disempurnakan.

Kemungkinan kesalahan ini perlu diantisipasi pada saat mendesain survei dan pada pelatihan petugas.

Partisipasi Responden

Partisipasi responden juga mempengaruhi akan timbulnya kesalahan non sampling, sehingga dalam desain survei perlu dikaji dan ditetapkan antara lain:

- a) Sifat-sifat dari responden
- b) Pendekatan pada responden termasuk penjelasan tentang maksud dan tujuan survei
- c) Situasi responden
- d) Diperlukan revisit atau tidak

Partisipasi responden juga dipengaruhi oleh sifat petugas dan diperlukan surat tugas, sehingga responden menjawab dengan benar. Untuk mengurangi non sampling error maka diperlukan kontrol yang ketat misalnya dengan mengadakan pengontrolan kualitas (quality control) baik dengan pengamatan langsung, pemeriksaan, atau dengan interview ulang secara sub-sampel. Juga perlu dibuat sistem untuk evaluasi pengawas/pemeriksa.

Kesalahan Pengolahan

Kesalahan pengolahan bisa timbul dimulai dari penerimaan dokumen, misalnya kurangnya kontrol kelengkapan dokumen atau salah pemberian identitas, tercampurnya dokumen antar wilayah. Dalam pengolahan pada satu bendel dokumen petugas pengolahan pada saat perekaman data bisa saja menset kode kabupaten/kota sehingga bendel tersebut seluruhnya dianggap kabupaten/kota sama. Perbedaan mungkin hanya terjadi pada nomor kode sampel. Dalam hal ini kurang diperhatikan perlunya identifikasi yang unik. Oleh karena itu kesalahan dalam pengolahan perlu dibuat sekecil mungkin, karena kesalahan pengolahan lebih dapat dideteksi dan diatasi dibandingkan kesalahan pengumpulan data.

Kontrol, deteksi, dan menghindari kesalahan pengolahan dilakukan antara lain:

- a) Melalui pengontrolan sistem batching dan pelaksanaannya
- b) Dicek sekali lagi identitas dari setiap dari setiap kuesioner dengan daftar sampel dan kelengkapannya
- c) Manual editing-coding perlu jelas dan diikuti serta tidak bertentangan dengan manual tatacara pengisian kuesioner.
- d) Perlu diikuti pedoman verifikasi atau kontrol isian melalui proses komputer
- e) Pemeriksaan konsistensi perlu diikuti dengan baik termasuk kewajaran isian

Petugas pengolahan yang kualifikasinya kurang baik akan menyebabkan non sampling error, sebagai contoh kurang teliti dalam

perekaman data. Kesalahan perekaman data akan sangat bahaya. Sebagai contoh angka 100 terketik 1000 atau terketik 10. Angka pertama akan menimbulkan estimasi terlampau besar (over estimation) dan angka kedua akan menimbulkan estimasi terlampau kecil (under estimation). Kesalahan pengetikan kode lapangan usaha misalnya akan menyebabkan penggolongan lapangan usaha berbeda, pertanian bisa berubah menjadi industri. Perlu perhatian dalam rancangan pengolahan dan pelaksanaan pengolahan, misal dilakukan editing pasca komputer (validasi).

Kesalahan Cakupan

Listing atau pendaftaran unit-unit dalam suatu wilayah dimulai dari pendaftaran bangunan baik bangunan tempat tinggal maupun bukan tempat tinggal, rumahtangga dan usaha atau kegiatan ekonomi sering dilakukan mendahului diadakan suatu pengumpulan data/survei, sebagai contoh:

- a) Dalam survei pendekatan rumahtangga seperti Survei Sosial Ekonomi Nasional (Susenas), Survei Angkatan Kerja (Sakernas), Survei Biaya Hidup (SBH), dan Survei Demografi dan Kesehatan (SDKI).
- b) Survei usaha kecil menengah (UKM), dengan mendaftar keberadaan unit-unit usaha yang ada di blok sensus, seperti usaha industri kecil/kerajinan rumahtangga, UKM perdagangan, dan UKM di sektor lainnya.
- c) Survei pertanian dengan pendekatan rumahtangga yang berusaha di bidang pertanian.

Pertanyaan: *Apakah listing selalu dilakukan dengan baik oleh petugas?*

Jawaban: *Tidak karena sering terjadi kesalahan listing, umumnya under coverage (lewat cakupan).*

Pertanyaan berikut: *Mengapa terjadi demikian dan apa disadari petugas akibatnya.*

Jawaban: *Petugas tidak sepenuhnya mempelajari batas-batas atau memang batas tidak jelas dari wilayah yang harus dilisting, misal batas blok sensus. Petugas saat itu tidak sepenuhnya menyadari.*

Kesalahan cakupan akan sangat mempengaruhi keakuratan dari estimasi, terutama yang mengkait faktor pengali. Akhir-akhir ini timbul pertanyaan dari berbagai konsumen data yang mempertanyakan keakuratan data, misalnya antara SP2000, P4B, BLT , yang bila dilihat sampai ke wilayah kecil makin besar perbedaannya dan tidak dapat dijelaskan penyebabnya (waktu pelaksanaan tidak terlalu jauh berbeda).

Kemungkinan kesalahan yang terjadi antara lain:

- a) Salah cakupan, batas blok sensus yang digunakan untuk listing tidak jelas sehingga petugas salah arah. Biasanya yang mengetahui batas rinci adalah petugas yang menggambar peta. Apakah petugas pembuat peta dapat selalu dihubungi?. Belum tentu dan Mantri Statistik tidak semuanya memahami batas blok sensus.
- b) Ada kesalahan dalam kualifikasi, blok sensus, atau salah pemberian kode, sehingga pada saat pengolahan tertukar. Sebagai akibat perbandingan data banyaknya rumahtangga antar unit tidak relevan (tidak dapat dibandingkan).
- c) Letak bangunan atau rumahtangga tidak beraturan sehingga saat listing petugas tidak melakukan secara teliti dan kemungkinan terlewat cakup banyak.
- d) Letak bangunan atau rumahtangga dalam areal luas dan berjauhan, sehingga adanya bangunan atau rumahtangga tidak dapat terlihat.

Kejadian tersebut selama ini sering terjadi dan biasanya cenderung under coverage (cakupan lebih kecil).

Sebenarnya keasalahan ini dapat diperkecil antara lain:

- a) Dengan membuat peta blok sensus yang lebih jelas dengan memberi tanda yang mudah dilihat. Pengalaman selama ini banyak peta blok sensus yang tidak jelas dan tidak dibuat menurut persyaratan yang ditentukan (lihat tatacara pemetaan wilayah).
- b) Petugas kurang memperhatikan tatacara listing termasuk mengevaluasi terlebih dahulu keberadaan peta.

Dua jenis peta perlu dikaji terlebih dahulu, yaitu:

- i. Peta desa yang terdiri dari blok sensus. Peta yang baik di dalamnya mencakup identitas RW/RT atau lingkungan setempat.
- ii. Peta blok sensus dengan identitasnya. Peta blok sensus yang baik terdapat identitas yang lengkap.

Kecenderungan tanpa dipelajari terlebih dahulu langsung terjun ke lapangan. Kejadian yang dilisting bukan blok sensus yang ada dalam daftar sampel.

- c) Petugas sering ingin segera selesai melakukan listing sehingga akibatnya tidak dicakup seluruh blok sensus.

Kesalahan tersebut menjadi kumulatif pada listing usaha, karena usaha terutama UKM tidak terlihat dari luar, seperti usaha makanan, menjahit, dan sebagainya, yang dilakukan di dalam rumah. Akibat dari kesalahan tersebut menyebabkan estimasi akan under estimate (estimasi lebih dari kondisi sebenarnya).

Sebagai contoh, pada blok sensus terpilih kondisi banyaknya rumahtangga pada saat kerangka sampel dan saat listing digambarkan pada Tabel 1 di bawah ini:

Tabel 1. Banyaknya Rumahtangga pada Saat Kerangka Sampel dan Listing pada 5 Blok Sensus Terpilih di Suatu Wilayah

Blok sensus	Saat Kerangka Sampel (M_i)	Saat survei (M_i^1)
(1)	(2)	(3)
1	121	110
2	140	90
3	81	75
4	70	71
5	110	95
Total	522	441

Dari pengalaman selama ini kecenderungan hasil listing under estimate. Apabila pada wilayah tersebut seluruh blok sensus adalah 20, dan bila penarikan sampel dilakukan dengan acak sederhana, maka estimasi banyaknya rumahtangga dari survei adalah:

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^n M_i^1 = \frac{20}{5}(441) = 1764$$

Bila digunakan data kerangka sampel, maka estimasi adalah:

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^n M_i = \frac{20}{5}(522) = 2088$$

Pertanyaan: *Apakah benar terjadi penurunan sekitar 15 persen?*

Jawaban: *Hal ini perlu dikaji apakah memang benar ada perubahan di atas atau kesalahan listing.*

Contoh di atas hanya contoh sederhana, dan akan menjadi beban yang lebih berat, apabila data salah digunakan juga untuk penetapan metode sampling seperti penetapan untuk stratifikasi atau peluang.

NonResponse

Nonresponse akan juga mengakibatkan bias dalam estimasi dan perlu dikaji bagaimana efek dari non response serta sumber-sumbernya. Non response dapat mengakibatkan sampel yang diperoleh tidak mewakili populasi.

Penyebab non response adalah responden, karena:

- a) Tidak berada di tempat, misal tidak berada di rumah atau tempat usaha
- b) Penolakan, yang sering terjadi pada perusahaan besar atau wilayah elite
- c) Tidak bisa menjawab pertanyaan, kemungkinan kondisi responden tidak mungkin diwawancarai
- d) Tidak bisa dijumpai
- e) Kuesioner hilang atau rusak.

Non response dapat terjadi pada seluruh kuesioner atau karena kesulitan menjawab ada pertanyaan yang tidak bisa diisi/dijawab responden.

Efek NonResponse

Nonresponse bisa sangat mempengaruhi hasil survei, terutama bila kondisi dari responden non response sangat berlainan dengan kondisi responden yang response (bisa mengarah bias yang positif atau bias negatif, under estimate atau over estimate). Apabila terjadi pada sensus dan jumlahnya cukup besar akan lebih rumit lagi, karena biasanya hasil suatu sensus lengkap harus disajikan sampai wilayah kecil, misal desa dan akan semakin terlihat kesalahannya.

Survei dengan pendekatan perusahaan/usaha biasanya non response cukup besar, seperti terjadi pada Survei Industri Besar/Sedang, Survei Perkebunan Besar, dan sebagainya, walaupun telah diusahakan berbagai cara misalnya dengan revisit atau pendekatan khusus.

Secara matematik, penulisan nilai populasi bila terbagi menjadi responden response (Y_1) dan responden non response (Y_2), maka nilai populasi menjadi:

$$Y = Y_1 + Y_2$$

$$\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$$

$$W_1 = \frac{N_1}{N} \text{ dan } W_2 = \frac{N_2}{N}$$

$$W_1 + W_2 = 1$$

dimana N: banyaknya unit/responden dalam populasi

N_1 : banyaknya unit/responden response

N_2 : banyaknya unit/responden non response

Apa yang terjadi bila \bar{Y}_1 berbeda jauh dengan \bar{Y}_2 ?. Perbedaan ini akan menyebabkan estimasi tidak mewakili populasi.

Sebagai contoh pada survei perusahaan misalnya perusahaan industri besar/ sedang atau perusahaan perkebunan, dalam pelaksanaan lapangan perusahaan yang besar non response tinggi, maka sampel hanya terwakili oleh perusahaan yang tidak terlalu besar/ sedang. Akibatnya rata-rata produksi misalnya akan cenderung under estimation: $\bar{Y}_1 < \bar{Y}$.

Akan terjadi bias, yang nilai relatifnya:

$$RB = \frac{\bar{Y}_1 - \bar{Y}}{\bar{Y}} = \frac{(\bar{Y}_1 - W_1\bar{Y}_1 - W_2\bar{Y}_2)}{\bar{Y}} = W_2 \frac{(\bar{Y}_1 - \bar{Y}_2)}{\bar{Y}}$$

Makin besar perbedaan antara \bar{Y}_1 dan \bar{Y}_2 maka bias akan semakin besar.

Demikian juga apabila W_2 makin besar bias juga akan makin besar.

Bias dapat positif atau negatif tergantung dari kondisi antara response dan non response.

Tabel 2. Persentase, Banyaknya Non response dari Survei Perusahaan Perkebunan Kopi an Rata-rata Banyaknya Pohon

Penerimaan dokumen	Persentase response	Jumlah pohon per perusahaan
(1)	(2)	(3)
Penerimaan 1	10	290
Penerimaan 2	17	382
Penerimaan 3	34	344
Non response	(39)	340
		650
Total	100	463

Contoh sederhana pada tabel di atas dimana pada survei perkebunan setelah berulang kali diadakan revisit sampai tiga kali hanya diterima 61 persen dokumen dengan rata-rata banyaknya pohon 344. Apabila data ini disajikan akan terjadi under estimate, rata-rata seharusnya 463 pohon. Dalam hal ini

$\bar{Y}_1 = 344$, $\bar{Y}_2 = 650$ dan $\bar{Y} = 463$ dengan $W_2 = 0,39$.

Signifikansi dari *nonresponse* tergantung pada:

- a) Bagaimana rates dari response
- b) Apakah karakteristik response dan non response berbeda
- c) Sumber-sumber dari non response dan cara mengatasi
- d) Cara mengatasi non response
- e) Adanya alokasi biaya untuk mengatasi non response (revisit, pendekatan, sosialisasi, dan sebagainya)

Sumber-sumber *nonresponse* yang perlu dikaji antara lain:

- a) Kuesioner yang memberatkan responden, misal terlalu panjang, adanya pertanyaan yang biasanya rahasia
- b) Kualitas petugas
- c) Sifat responden

d. Cara Mendeteksi *NonSampling Error*

Non sampling error tidak dapat diukur secara matematis seperti halnya sampling error. Non sampling error perlu dihindari dengan desain survei, pelaksanaan lapangan, dan pengolahan yang baik seperti dijelaskan sebelumnya. Pengawasan lapangan, monitoring dan pemeriksaan hasil survei dan hasil pengolahan merupakan hal mutlak yang harus dilaksanakan. Pasca evaluasi sensus/survei (PES) merupakan salah satu cara untuk mengkaji sampai sejauh mana kira-kira akurasi hasil suatu pengumpulan data. PES tidak untuk memperbaiki estimasi, tetapi mempunyai tujuan:

- i. Mengkaji sampai sejauh mana kira-kira adanya kesalahan, baik ditinjau dari cakupan (coverage) maupun dari isian kuesioner (content)
- ii. Sebagai pendukung analisis
- iii. Sebagai masukan untuk penyempurnaan metodologi dan pelaksanaan yang akan datang.

LATIHAN

1. Jelaskan jenis-jenis kesalahan yang mungkin terjadi dalam suatu kegiatan survei dan berikan contoh masing-masing!
2. Jelaskan hubungan antara *sampling error* dan *nonsampling error* terkait dengan ukuran sampe

