II. LANDASAN TEORI

Definisi 2.1 Distribusi Sampling

Distribusi sampling adalah distribusi probibilitas dari suatu statistik. Distribusi tergantung dari ukuran populasi, ukuran sampel dan metode memilih sampel. Sampel yang diambil ialah sampel acak dan dari sampel tersebut nilai statistiknya dihitung. Nilai setiap statistik sampel akan bervariasi antar sampel.

Sampling adalah proses pengambilan atau memilih n buah elemen dari populasi yang berukuran N (Lohr, 1999). Dalam banyak hal, survai tidak mungkin melibatkan keseluruhan elemen populasi, karena akan memerlukan waktu, tenaga dan biaya yang cukup besar. Besarnya biaya dalam sensus terkadang tidak seimbang dengan manfaat dari informasi yang dikumpulkan. Tujuan teori sampling adalah membuat sampling menjadi lebih efisien, artinya dengan biaya yang lebih rendah diperoleh ketelitian yang sama tinggi atau dengan biaya yang sama diperoleh ketelitian yang lebih tinggi. Teori sampling mencoba untuk mengembangkan metode pemilihan sampel dan pembuatan perkiraan, sehingga diperoleh metode yang memungkinkan diperolehnya hasil penelitian dengan tingkat ketelitian tinggi sesuai dengan tujuan, tetapi dengan biaya yang lebih rendah.

Definisi 2.2 Teknik Sampling

Teknik Sampling merupakan teknik pengambilan sampel untuk mendapatkan sampel yang dapat mewakili karakteristik populasi. Terdapat berbagai teknik sampling untuk menentukan sampel yang akan digunakan dalam penelitian. Teknik sampling pada dasarnya dapat dikelompokkan menjadi dua yaitu probability sampling dan non probability sampling.

Yang termasuk dalam teknik *probability* sampling adalah :

a. Sampling Acak Sederhana

Definisi 2.3 Sampling Acak Sederhana

Teknik acak sederhana adalah teknik acak yang paling dasar dalam pengambilan sampel. Syarat utama agar suatu sampel mempunyai sifat acak, pemilihan harus melalui proses acak, yaitu suatu proses yang hasilnya tak dapat diketahui sebelumnya. Prinsip sampel acak sederhana, setiap anggota populasi mempunyai kesempatan yang sama untuk dipilih sebagai sampel.

Jika suatu sampel dengan n elemen dipilih dari suatu populasi dengan N elemen sedemikian rupa sehingga setiap kemungkinan sampel dengan n elemen mempunyai kesempatan yang sama untuk terpilih, prosedur sampling yang demikian disebut simple random sampling (Supranto,J. 1992).

Meskipun terlihat sangat sederhana, teknik sederhana ini mempunyai syarat yang khusus. Teknik acak sederhana bisa dipakai jika ada kerangka sampel yang baik dan lengkap yang memuat daftar nama anggota semua populasi. Oleh karena syarat yang ketat itu, teknik acak sederhana ini umumnya bisa dipakai dalam kondisi berikut :

1. Sampel acak sederhana efektif dipakai jika populasi tidak terlalu besar.
2. Sampel acak sederhana bisa dipakai jika populasi relatif homogen (Eriyanto, 2007).

Penduga rata-rata pada sampling acak sederhana adalah:

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

Penduga varian untuk $\hat{\mu}$ adalah:

$$V(\bar{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

Dimana:

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

b. Sampling Acak Berlapis atau *Stratified Random Sampling*

Definisi 2.4 Sampling Acak Berlapis

Sampling acak berlapis adalah bentuk sampling acak yang elemen populasinya dibagi ke dalam kelompok-kelompok homogen yang disebut strata. Ada beberapa alasan dalam pengguanaannya,antara lain:

1. Jika data diketahui ketelitian yang diinginkan untuk subkelompok tertentu dari populasi, ada baiknya memperlakukan setiap subkelompok sebagai suatu populasi tertentu.
2. Administrasi yang baik dapat memakai keguanaan strata.
3. Masalah penarikan sampel dapat berbeda dalam bagian populasi yang berbeda.
4. Pelapisan dapat menghasilkan suatu manfaat dalam ketelitian perkiraan dari karakteristik populasi. Hal ini memungkinkan untuk membagi sebuah populasi yang heterogen menjad subpopulasi-subpopulasi, dengan setiap subpopulasi menjadi homogen (Cochran,1991).

Dalam acak stratifikasi, sebelum sampel diambil dari populasi, dilakukan stratifikasi populasi terlebih dahulu berdasarkan karakteristik tertentu. Sampel yang diambil disesuaikan dengan proporsi dan populasi. Cara melakukan teknik stratifikasi adalah sebagai berikut: Setelah mendapatkan kerangka sampel, disusun terlebih dahulu stratifikasi. Anggota populasi dimasukkan ke dalam stratifikasi yang telah dibuat. Setelah itu baru ditarik sampel sesuai dengan strata masing-masing.

Apabila populasi heterogen, lebih baik menggunakan sampling acak berlapis daripada sampling acak sederhana oleh karena alasan berikut :

1. Setiap stratum homogen atau relatif homogen, sehingga sampel acak yang diambil dari setiap stratum akan memberikan perkiraan yang dapat mewakili stratum yang bersangkutan. Perkiraan gabungan yang diperoleh berdasarkan perkiraan dari setiap stratum akan memberikan perkiraan menyeluruh yang mewakili populasi.
2. Biaya untuk pelaksaan sampling acak berlapis lebih murah daripada sampling acak sederhana karena alasan administrasi.
3. Perkiraan bisa dibuat untuk setiap stratum yang dapat dianggap sebagai populasi yang berdiri sendiri dan mungkin bisa dilakukan oleh seorang peneliti saja (Supranto, J. 1992).

Penduga rata-rata pada sampling acak berlapis adalah:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{L} N_i \bar{y}_i$$

Penduga varian untuk $\hat{\mu}$ adalah:

$$V(\bar{y}) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \left(\frac{N_i - n_i}{N_i}\right)\left(\frac{S_i^2}{n_i}\right)$$

$$S_i^2 = \frac{\sum_{i=1}^{n} (Y_{ij} - \bar{Y})^2}{n-1}$$

Dengan

N = banyaknya elemen (sampling unit) dari populasi

N_i= banyaknya elemen dari stratum ke-i

n= banyaknya elemen sampel sebelum dikelompokan

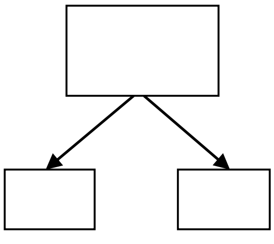
n_i= banyaknya elemen sampel dari stratum ke-i yang dipilih secara acak

c. Sampling Kelompok atau *Cluster Sampling*

Sampling kelompok adalah pengambilan sampel dari beberapa unit sampling yang merupakan kelompok dari elemen (Scheaffer, Mendenhall dan Ott., 1996). Langkah paling awal dalam penarikan sampel cluster yang harus dilakukan oleh peneliti adalah mengidentifikasi cluster atau satuan dimana individu menjadi anggota dalam cluster. Semua cluster yang ada dalam populasi harus bisa diidentifikasi. Setelah cluster diambil, disusun kerangka sampel berupa daftar nama individu yang menjadi anggota cluster terpilih. Setelah daftar itu bisa disusun, barulah dilakukan penarikan sampel seperti pada sampel acak sederhana, sistematis atau stratifikasi.

2.5 One-Stage Cluster Sampling

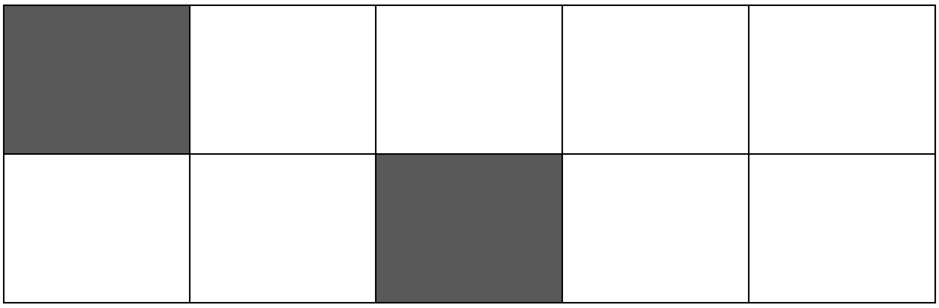
One-Stage Cluster Sampling dilakukan dengan didasarkan pada gugus (cluster). Asumsinya, individu adalah bagian dari gugus atau cluster tertentu, kerangka sampel berupa daftar nama individu memang tidak tersedia, tetapi daftar kelompok (gugus) itu pastilah tersedia. Teknik ini dapat dilakukan jika tidak tersedianya kerangka sampel berupa nama-nama individu anggota populasi dan kalaupun kerangka sampel itu tersedia masih diragukan akurasinya. Gambaran secara umum dari metode ini tersaji dalam Gambar 2.1.



Gambar 2.1. *One-stage cluster sampling*

Contoh kasus *one-stage cluster sampling*

Suatu populasi memiliki 10 cluster (N=10). Dari 10 cluster tersebut, dipilih secara acak 2 cluster untuk diamati. Secara lengkap tahapan samplingnya tersaji pada Gambar 2.2.



Gambar 2.2. Ilustrasi pengambilan sampel pada *one-stage cluster sampling*

Misal cluster yang terpilih adalah 1 dan 8, maka cluster terpilih dijadikan sebagai sampel penelitian.

Penduga bagi rata-rata populasi adalah :

$$\bar{y}_{one} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} \qquad (2.1)$$

Notasi : $y_i = \sum_{i=1}^{m_i} y_{ij}$

m_i = banyaknya elemen dalam kelompok i, dimana i = 1,2,3,...,N

Karena pemilihan cluster dilakukan dengan metode acak maka diperoleh penduga varians bagi *one-stage cluster* adalah :

$$\hat{V}(\bar{y}_{one}) = \left(\frac{N-n}{Nn\bar{M}^2}\right) S^2$$

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

Notasi N= jumlah cluster

n= jumlah cluster terpilih

$m_i = M_i$= jumlah elemen/unit sampel cluster terpilih ke-i

Selanjutnya akan dibuktikan varians penduga *one-stage cluster* dengan penduga rata-rata pada persamaan (2.1):

Dengan $\bar{y}_{one} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$ maka diperoleh

$$\hat{V}(\bar{y}_{one}) = \hat{V}\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}\right)$$

$$\hat{V}(\bar{y}_{one}) = = \frac{1}{(\sum_{i=1}^{n} m_i)^2} V(\sum_{i=1}^{n} y_i) \quad (2.2)$$

Dengan $\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n}$, maka diperoleh $\sum_{i=1}^{n} m_i = n\bar{m}$. Sehingga persamaan (2.2) menjadi

$$\hat{V}(\bar{y}_{one}) = \frac{1}{n^2\bar{m}^2} V(\sum_{i=1}^{n} y_i) \quad (2.3)$$

Dalam kasus one-stage cluster m_i=M_i, sehingga diperoleh $\sum_{i=1}^{n} m_i = n\bar{m}$=$\sum_{i=1}^{n} m_i = n\bar{M}$

Serta $\bar{y}^* = \frac{\sum_{i=1}^{n} y_i}{n}$, maka diperoleh $\sum_{i=1}^{n} y_i = n\bar{y}^*$. Sehingga persamaan (2.3) menjadi :

$$\hat{V}(\bar{y}_{one}) = \frac{1}{n^2\bar{M}^2} V(n\bar{y}^*)$$

$$= \frac{1}{n^2\bar{M}^2}\, n^2 V(\bar{y}^*)$$

$$= \frac{1}{\bar{M}^2}\, V(\bar{y}^*) \quad (2.4)$$

Diketahui bahwa $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$, sehingga $\frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} m_i$ atau $\bar{y}^* = \bar{y} m_i$.

Mengikuti konsep penduga varians pada kasus *simple random sampling*, maka diperoleh:

$$V(\bar{y}^*) = \left(\frac{N-n}{N}\right)\frac{S^2}{n} \quad (2.5)$$

Dengan mensubtitusikan persamaan (2.5) ke (2.4) maka diperoleh diperoleh:

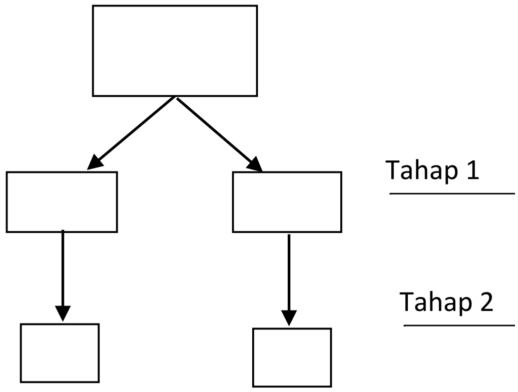
$$\hat{V}(\bar{y}_{one}) = \frac{1}{\bar{M}^2}\left(\frac{N-n}{N}\right)\frac{S^2}{n} \qquad (2.6)$$

$$\hat{V}(\bar{y}_{one}) = \frac{N-n}{Nn\bar{M}^2}\, S^2$$

Dengan $S^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}m_i)^2}{n-1}$ (terbukti)

2.6 *Two-stage Cluster sampling*

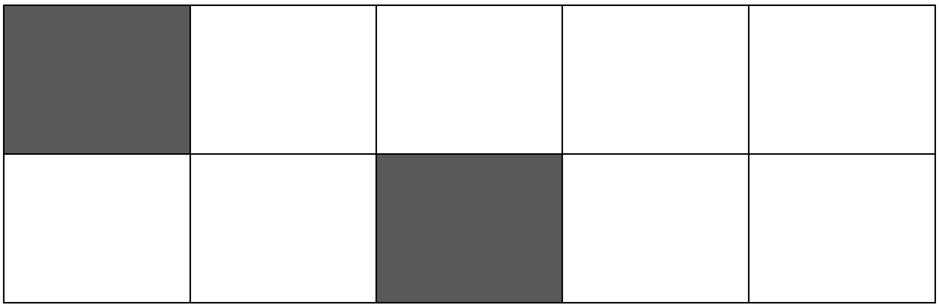
Metode *Two-Stage Cluster Sampling* merupakan pengembangan dari metode cluster sampling dimana pengambilan sampel dilakukan secara dua tahap, yaitu tahap pertama, memilih beberapa cluster dalam populasi secara acak sebagai sampel dan tahap kedua memilih elemen dari tiap cluster terpilih secara acak (Scheafer et.al., 1996).Gambaran secara umum dari metode ini tersaji dalam Gambar 2.3.



Gambar 2.3. Two-*stage cluster sampling*

Contoh kasus *two-stage cluster sampling*

Suatu populasi memiliki 10 cluster (N=10). Masing-masing cluster terdiri dari 6 sub cluster. Dari 10 cluster tersebut, dipilih secara acak 2 cluster. Kemudian dari masing-masing cluster terpilih tersebut, dipilih 2 sub cluster. Secara lengkap tahapan samplingnya tersaji pada Gambar 2.4.



Gambar 2.4 Ilustrasi pengambilan sampel pada *two-stage cluster sampling*

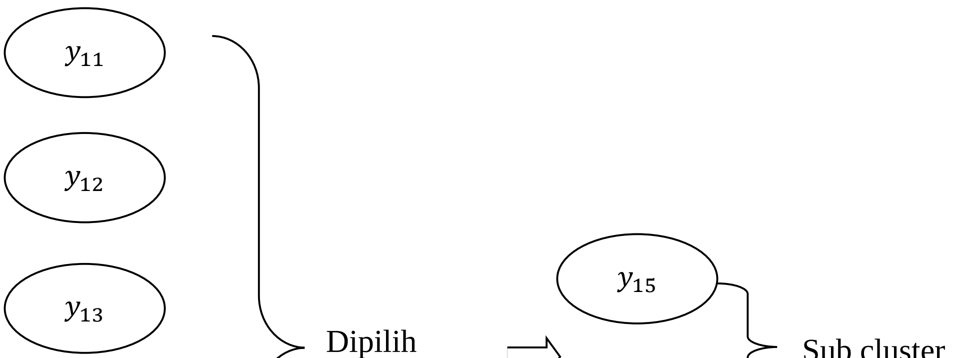
Misal pada tahap pertama cluster terpilih adalah 1 dan 8. Sub cluster dari masing-masing cluster terpilih disajikan dalam Tabel 2.1.

Tabel 2.1 Ilustrasi pengambilan sub kluster

Sub kluster 1	Sub kluster 4
y_{11}	y_{81}
y_{12}	y_{82}
y_{13}	y_{83}
y_{14}	y_{84}
y_{15}	y_{85}
y_{16}	y_{85}

Selanjutnya dari masing-masing cluster terpilih, dipilih 2 sub cluster dengan *simple random sampling*.

Ilustrasinya dapat dilihat pada Gambar 2.5.



Gambar 2.5 Ilustrasi pengambilan sub cluster terpilih

Persoalan yang dihadapi di dalam memilih sampel *Two-stage Cluster sampling* ialah memilih kelompok yang tepat. Dua syarat yang harus dipenuhi adalah :

1) Secara geografis elemen dalam kelompok harus saling berdekatan

2) Kelompok sedikit saja agar mudah mengadministrasikannya (Supranto, J. 1992).

Kelompok yang besar cenderung memeliki elemen yang heterogen dengan demikian diperlukan pemilihan banyak elemen dari setiap kelompok sehingga diperoleh hasil penelitian tingkat ketelitian yang tinggi. Keuntungan utama dari metode *two-stage cluster sampling* adalah bahwa metode ini lebih fleksibel daripada metode *one-stage cluster sampling sampling*. Bila subunit dalam unit mempunyai karakteristik yang sama yang sangat dekat, metode ini mengurangi penarikan sampel satu tahap berukuran besar sehingga peneliti mempunyai kesempatan mengambil beberapa nilai yang lebih kecil sehingga sampling menjadi lebih efisien.

Penduga bagi rata-rata populasi yaitu

$$\bar{y}_{two} = \frac{1}{\bar{\bar{M}}} \frac{\sum_{i=1}^{n} M_i \, \bar{y}_i}{n}$$

Notasi:

N = Jumlah kluster dalam populasi

n = Jumlah kluster terpilih

M_i = Jumlam elemen/unit sampling dari kluster ke-i

m_i = Jumlam elemen/unit sampling yang dipilih dari kluster terpilih ke-i

M = $\sum_{i=1}^{N} M_i$ = jumlah elemen/unit sampling dalam populasi

$\bar{M} = \frac{M}{N}$ = rata-rata jumlah elemen/unit sampling masing-masing kluster

Bukti :

Diketahui bahwa total populasi = $N\bar{y}$ dan M merupakan total jumlah elemen dalam populasi,sehingga akan dilakukan pendekatan dengan total pulasi. Maka diperoleh;

$$\bar{y}_{two} = \frac{N\,\bar{y}_{total}}{M} \quad (2.7)$$

Dimana $\bar{y}_{total} = \frac{\sum_{i=i}^{n} y_i}{n}$, sehingga menjadi

$$\bar{y}_{two} = \frac{N}{M}\,\frac{\sum_{i=i}^{n} y_i}{n}\,, \quad (2.8)$$

dimana y_i merupakan total pengamatan dari cluster terpilih ke-i :

$$y_i = \sum_{j=1}^{m_i} y_{ij} = M_i\bar{y}_i \quad (2.9)$$

Subtitusi persamaan (2.9) ke persamaan (2.8), sehingga diperoleh :

$$\bar{y}_{two} = \frac{N}{M}\,\frac{\sum_{i=i}^{n} y_i}{n}$$

$$\bar{y}_{two} = \frac{N}{N\bar{M}}\,\frac{\sum_{i=i}^{n} M_i\bar{y}_i}{n}$$

$$\bar{y}_{two} = \frac{1}{\bar{M}}\,\frac{\sum_{i=i}^{n} M_i\bar{y}_i}{n} \quad \text{(terbukti)}$$

Penduga varians bagi penduga rata-rata populasi adalah:

$$\hat{V}(\bar{y}_{two}) = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)S_b^2 + \frac{1}{nN}\sum_{i=1}^{n}\left(\frac{\bar{M}-m}{\bar{M}}\right)\frac{S_i^2}{m}$$

Dengan

$$S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i M_i - \bar{M}\mu)^2$$

$$S_i^2 = \frac{\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_i\right)^2}{m-1}$$

Bukti:

Menurut scheaffer, et.al (1996) penduga ragam bagi penduga rata-rata populasi untuk kasus *two-stage cluster sampling* dapat diuraikan sebagai berikut:

$$V(\bar{y}_{two}) = V_1[E_2(\bar{y}_{two})\ ] + E_1[V_2(\bar{y}_{two})]$$

Pertama menguraikan:

$$V_1[E_2(\bar{y}_{two})\ ] = V_1\left[\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i\right]$$

$= V_1[\bar{y}]$; varians *simple random sampling*

$$= \frac{N-n}{N}\frac{S_1^2}{n}$$

dengan $S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{y}_i - \bar{y})^2$

menurut scheaffer, et.al (1996) dengan menguraikan,

$\frac{1}{\bar{M}^2} S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \mu)^2$

maka diperoleh $\left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right) S_b^2$.

Kemudian yang kedua menguraikan:

$$E_1[V_2(\bar{y}_{two})] = E_1\left(V_2\left(\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i\right)\right)$$

$$= E_1 \left[\frac{1}{n^2} \left(\sum_{i=1}^{n} \bar{y}_i \right) \right]$$

$$= \ E_1 \left(\frac{1}{n^2} \ \left(V_2(\bar{y}_1) + V_2(\bar{y}_2) + \cdots + V_2(\bar{y}_n) \right) \right)$$

$$= E_1 \left(\frac{1}{n^2} \ n \, V(\bar{y}_i) \right)$$

$$= E_1 \left(\frac{1}{n} \ V(\bar{y}_i) \right)$$

$$= \frac{1}{n} E_1[V(\bar{y}_i)]$$

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

$$= \frac{1}{n} \frac{1}{N} \sum_{i=1}^{n} \frac{\bar{M} - m}{\bar{M}} \ \frac{S_i^2}{m}$$

dengan $S_i^2 \ = \frac{\sum_{i=1}^{n} \left(y_{ij} - \bar{y}_i \right)^2}{m-1}$

Dengan demikian diperoleh penduga varians bagi penduga rata-rata populasinya adalah:

$\hat{V}(\bar{y}_{two}) = \left(\frac{N-n}{N} \right) \left(\frac{1}{n\bar{M}^2} \right) S_b^2 + \ \frac{1}{nN} \sum_{i=1}^{n} \left(\frac{\bar{M}-m}{\bar{M}} \right) \frac{S_i^2}{m}$ (terbukti)

2.7 Sampling Error

Sampel berbeda dengan populasi. Dalam sampel kita hanya menyertakan sebagian anggota dari populasi untuk diamati. Karena hanya sebagian anggota populasi yang diamati, sehingga secara teoritis ada kesalahan hasil yang diperoleh dari suatu sampel. Kesalahan ini terjadi karena peneliti hanya mengamati sebagian anggota dan bukan keseluruhan anggota populasi.

Peneliti umumnya tidak mengetahui nilai populasi (parameter). Yang dihadapi oleh peneliti adalah hasil dari sampel. Peneliti tidak bisa membuat jawaban yang pasti. Yang bisa dilakukan adalah membuat interval kemungkinan nilai sebenarnya jika semua anggota populasi diamati. Sehingga dapat digunakan konsep sampling error. Sampling error menunjukan perbedaan antara statistik dan parameter. Dengan menggunakan sampling dari suatu penelitian, peneliti bisa memprediksi nilai sesungguhnya dalam populasi. Bila $\bar{y}$ digunakan untuk menduga μ, kita percaya (1-α)100% bahwa galatnya tidak akan melebihi $Z\alpha_{/2}\sqrt{\hat{V}(\hat{y})}$.

2.8 Selang Kepercayaan

Salah satu penduga titik bagi nilai tengah populasi μ adalah statistik $\bar{y}$. Sebaran penarikan sampel $\bar{y}$ berpusat di μ, dan dalam sebagian besar penerapannya ragamnya lebih kecil daripada ragam penduga-penduga lainnya. Jadi nilai tengah sampel $\bar{y}$ akan digunakan sebagai nilai dugaan titik bagi nilai tengah populasi μ. Selang kepercayan (1-α)100% memberikan ukuran sejauh mana ketelitian atau akurasi nilai dugaan titiknya. Bila μ memang pusat selang tersebut, maka $\bar{y}$ menduga μ tanpa galat. Tetapi, kecil sekali kemungkinannya , $\bar{y}$ tepat sama dengan μ, sehingga nilai dugaan tersebut mempunyai galat.

Bila $\bar{y}$ adalah nilai tengah sampel acak berukuan n yang diambil dari suatu populasi dengan varians σ^2 diketahui, maka selang kepercayaan (1-α)100% bagi μ adalah

$$\bar{y} \pm Z\alpha_{/2}\sqrt{\hat{V}(\hat{y})}$$

2.9 Relatif Bias

Salah satu alasan dasar untuk sampling adalah bahwa informasi yang terkandung dalam sampel berguna untuk mengestimasi parameter populasi. Penduga yang baik adalah penduga yang bersifat tak bias dan bervariansi minimum. $\hat{\theta}$ dikatakan penduga tak bias bagi parameter θ, jika $E(\hat{\theta}) = \theta$. Sebaliknya $\hat{\theta}$ dikatakan penduga bias bagi parameter θ, jika $E(\hat{\theta}) \neq \theta$. Akan tetapi tidak diharapkan suatu penduga akan menduga parameter tanpa kesalahan. Tidak beralasan mengharapkan $\hat{\theta}$ akan menaksir θ dengan tepat, tetapi tentunya diharapkan bahwa penduga yang dihasilkan tidak terlalu jauh menyimpang.

Kualitas suatu penduga dapat di evaluasi salah satunya dengan kriteria relatif bias. Jika θ suatu parameter dan $\hat{\theta}$ suatu penduga maka $\left|\hat{\theta} - \theta\right|$ merupakan kesalahan sampling. Relatif bias dapat dihitung dengan membagi kesalahan sampling dengan θ dikalikan dengan 100%. Semakin kecil nilai relatif bias maka penduga dapat dikatakan semakin baik.

$$B = \left(\frac{\left|\hat{\theta} - \theta\right|}{\theta}\right) 100\%$$

Notasi :

B = Relatif bias