BAB III
SAMPLING BERKELOMPOK DAN SAMPLING BERKELOMPOK DENGAN *PROBABILITY PROPORTIONAL TO SIZE* (PPS)

3.1 Sampling Berkelompok

Populasi memiliki kondisi yang berbeda–beda jika dilihat berdasarkan ukurannya. Pada pembahasan subbab ini, ketika suatu populasi tersebar sangat luas, dalam arti sulit untuk dibuat kerangka sampelnya, maka akan sulit melakukan pengambilan sampel dengan metode–metode yang diharuskan memiliki kerangka sampel, salah satunya metode sampel acak sederhana. Hal ini mempunyai makna, ketika suatu populasi memiliki jumlah atau batasan kuantitatif yang jelas, maka tidak akan sulit untuk membuat *list* (daftar) elemen–elemen pada populasi. Sementara itu pada kasus pengambilan sampel untuk suatu populasi cukup besar, akan menghadapi beberapa permasalahan, diantaranya pengambilan sampel tersebut akan membutuhkan waktu dan biaya yang tidak sedikit, selain itu akan ada kesulitan dalam membuat daftar elemen populasi walaupun terkadang daftar populasi tersebut dapat dibuat. Oleh karena itu, untuk mengatasi permasalahan-permasalahan tersebut, pengambilan sampel dapat dilakukan dengan menggunakan metode sampling berkelompok, sebagai salah satu alternatif untuk mengatasi permasalahan pada pengambilan sampel untuk populasi yang cukup besar. Selain populasi yang berukuran cukup besar, metode sampling berkelompok juga dapat digunakan ketika populasi bersifat heterogen yaitu populasi yang unsur-unsurnya memiliki sifat atau keadaaan yang bervariasi. Hal ini dikarenakan pengelompokan pada metode sampling berkelompok tidak mesyaratkan ketentuan apapun. Oleh karena itu, maka tidak menjadi suatu permasalahan apabila populasinya bersifat heterogen.

Sampling berkelompok (*cluster sampling*) merupakan sampling probabilitas dimana masing-masing unit sampel (*sampling unit*) merupakan kumpulan (klaster) dari elemen (Scheaffer et al.1990). Secara garis besar, penarikan sampel dengan metode ini tidak langsung kepada elemen, melainkan melalui kelompok elemen terlebih dahulu yang disebut dengan unit sampling.

Alasan penggunaan metode sampling berkelompok pada populasi yang berjumlah banyak, antara lain:

1. Dengan menggunakan sampling berkelompok, maka pengelompokan populasi akan lebih mudah.
2. Biaya ketika akan melakukan penelitian akan lebih murah, karena kelompok yang dibuat akan lebih efisien.
3. Akan lebih mudah melakukan rencana pengambilan sampel, karena unit sampling tidak tersebar dengan luas.

Metode sampling berkelompok dilakukan dalam beberapa tahapan, yaitu sebagai berikut.

a. Tahap pertama yaitu membagi populasi kedalam M kelompok (*cluster*) secara acak, hal ini berarti tidak ada kriteria tertentu yang mensyaratkan pembentukan suatu kelompok. M kelompok (*cluster*) selanjutnya disebut sebagai unit sampling utama (usu) atau yang dikenal dengan *primary sampling units* (psu).
b. Tahap kedua, setelah populasi terbagi kedalam M kelompok, tahapan selanjutnya yaitu memilih secara acak m kelompok yang akan dijadikan sampel. m kelompok ini selanjutnya disebut dengan *secondary sampling units* (ssu) atau unit sampling kedua (usk). Masing-masing m kelompok ini berukuran N.
c. Tahap ketiga, setelah mendapatkan m kelompok, tahapan selanjutnya yaitu memilih n buah anggota sampel dari masing-masing usk yang disebut dengan kelompok utama (*ultimate cluster*).

Untuk menentukan total populasi beserta dengan variansinya tersebut, dilakukan dengan cara menentukan penaksirnya dengan menggunakan sampel yang diperoleh pada tahapan tersebut di atas. Dengan kata lain sampel yang diperoleh pada tahapan di atas tersebut akan digunakan untuk menaksir ukuran-ukuran populasi.

3.2 Pengertian Total Populasi

Sebelum membahas mengenai pengertian total populasi dalam sampling berkelompok, menurut Taro Yamane (1967) populasi memiliki empat karakteristik yang lebih sering diperhatikan, yaitu:

1. Rata–rata populasi

   Rata-rata merupakan jumlah keseluruhan data dibagi dengan banyaknya data tersebut. Jumlah total keseluruhan data populasi dinotasikan dengan X dan banyaknya data populasi dinotasikan dengan N sehingga perumusan untuk rata-rata populasi yang dinotasikan dengan $\bar{X}$ dinyatakan dalam perumusan berikut:

   $$\bar{X} = \frac{X}{N} \qquad (3.1)$$

2. Jumlah total populasi

   Secara umum pada sampling acak sederhana, jumlah total populasi merupakan hasil kali antara banyaknya data populasi (N) dengan rata–rata populasi $(\bar{X})$, dan dinyatakan dalam perumusan berikut :

   $$X = N\bar{X} \qquad (3.2)$$

3. Rasio populasi

   Rasio populasi merupakan perbandingan antara pembilang (*numerator*) dan penyebut (*denumenator*) yang saling terpisah dan tidak ada hubungannya. Pembilang dan penyebut dalam pembahasan ini dapat sebagai dua jumlah total populasi atau dua rata-rata populasi. Rasio populasi dinyatakan dalam perumusan berikut :

   $$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}} \qquad (3.3)$$

4. Proporsi populasi

   Proporsi populasi merupakan bentuk pecahan yang pembilangnya merupakan bagian dari penyebutnya. Proporsi dipergunakan untuk melihat komposisi suatu variabel dalam populasi. Bentuknya sering dinyatakan dalam persen, yaitu dengan mengalikan pecahan tersebut dengan 100%. Proporsi tidak mempunyai satuan (dimensi), karena satuan dari pembilang dan penyebutnya sama, sehingga saling meniadakan. Nilai proporsi berada pada interval tertutup antara 0 dan 1. Secara umum perumusan proporsi adalah sebagai berikut :

$$\text{Proporsi} = \frac{X}{X+Y} \times 100\% \tag{3.4}$$

dimana X merupakan bagian dari jumlah populasi dan Y merupakan jumlah populasi yang telah dikurangi oleh X.

Berdasarkan informasi mengenai empat karakteristik populasi tersebut, karena yang ingin diperoleh pada penelitian adalah total suara Pemilu dari seluruh populasi, maka itu, karakteristik populasi yang akan dibahas dalam skripsi ini adalah karakteristik total populasi.

Penaksir total populasi dan penaksir rata-rata berdasarkan pengertian umum ada dalam metode sampling acak sederhana. Misalkan $X_1, X_2, \dots, X_N$ adalah populasi yang berukuran N dan $x_1, x_2, \dots, x_n$ adalah sampel yang berukuran n. Rata-rata populasi ($\bar{X}$) dan rata-rata sampel ($\bar{x}$) didefinisikan sebagai berikut:

$$\bar{X} = \frac{1}{N}(X_1 + X_2 + \cdots + X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{3.5}$$

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.6}$$

Rata–rata sampel merupakan penaksir tak bias dari rata–rata populasi, dan dinyatakan sebagai berikut:

$$\hat{\bar{X}} = \bar{x} \tag{3.7}$$

Pembuktian :

$$\begin{aligned} E(\bar{x}) &= E\left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right] \\ &= \frac{1}{n}[E(x_1) + E(x_2) + \cdots + E(x_n)] \\ &= \frac{1}{n}(n\bar{X}) = \bar{X} \\ E(\bar{x}) &= \bar{X} \end{aligned}$$

Berdasarkan persamaan (3.7), dapat diperoleh informasi bahwa total populasi merupakan penaksir tak bias untuk total populasi, dinyatakan sebagai berikut:

$$\hat{X} = X \tag{3.8}$$

Pembuktian :

$$E(\widehat{X}) = E(N\bar{x})$$
$$= N[E(\bar{x})]$$
$$= N\bar{X}$$
$$E(\widehat{X}) = X$$

Pada sampling berkelompok, total populasi didefinisikan sebagai berikut:

$$X = \sum_{i=1}^{M} X_i \tag{3.9}$$

Sedangkan rata–rata populasi didefinisikan sebagai berikut:

$$\bar{X} = \frac{X}{M} = \frac{\sum_{i=1}^{M} X_i}{M} \tag{3.10}$$

3.3 Penaksir Total Populasi Sampling Berkelompok

Ciri dari sampling berkelompok yaitu proses pemilihan unit-unit sampling dilakukan dalam dua tahap. Tahap pertama adalah pemilihan sejumlah m kelompok yaitu unit sampling utama dari M, dan tahap selanjutnya adalah pemilihan n_i (dimana i = 1, 2, 3, …, m) dari N_i unit sampling kedua (usk).

Dengan kata lain, proses penaksiran total populasi pada sampling berkelompok dilakukan dalam dua tahap juga. Tahap pertama adalah menaksir total kelompok m$(\hat{x})$, dan tahap selanjutnya yaitu menggunakan penaksir yang diperoleh pada tahap pertama untuk menaksir total dari kelompok M$(\hat{X})$.

Penaksir dari total populasi dinotasikan dengan $\hat{X}$ dan didefinisikan sebagai berikut:

$$\hat{X} = \frac{M}{m} \sum_{i}^{m} \hat{X}_i = \frac{M}{m} \sum_{i}^{m} \frac{N_i}{n_i} \sum_{j}^{n_i} x_{ij} \tag{3.11}$$

dimana $\hat{X}_i$ merupakan notasi untuk menyatakan penaksir total populasi dari masing-masing kelompok, dan x_{ij} merupakan notasi untuk menyatakan elemen-elemen di kelompok utama.

Seperti yang telah dikemukakan pada subbab sebelumnya, bahwa rata–rata sampel merupakan penaksir yang tak bias bagi rata–rata populasi, sehingga untuk penaksir total populasi diperoleh:

$$E(\widehat{X}) = X$$

Untuk memudahkan perhitungan, persamaan (3.11) dapat diuraikan menjadi:

$$\hat{X} = M\left[\frac{1}{m}\sum_{i=1}^{m} N_i\left(\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}\right)\right]$$

$$= M\left[\frac{1}{m}\sum_{i=1}^{m} N_i(\bar{x}_i)\right]$$

$$= M\left[\frac{1}{m}\sum_{i=1}^{m} \hat{X}_i\right]$$

$$\hat{X} = M\left(\hat{\bar{X}}\right) \tag{3.12}$$

Berdasarkan persamaan (3.12), langkah awal yang dilakukan yaitu dengan menaksir rata-rata kelompok utama ke-i ($\bar{x}_i$). Selanjutnya mengalikan $\bar{x}_i$ dengan N_i, sehingga akan memperoleh penaksir total populasi X_i. Selanjutnya, hal yang dilakukan yaitu menentukan penaksir dari rata-rata kelompok utama ($\hat{\bar{X}}$). Setelah itu, mengalikan $\hat{\bar{X}}$ dengan M, sehingga akhirnya diperoleh penaksir dari total populasi.

3.4 Variansi dari Penaksir Total Populasi dan Penaksirnya

3.4.1 Variansi dari Penaksir Total Populasi Sampling Berkelompok

Variansi dari $\widehat{X}$ diperlukan untuk menilai presisinya. Terdapat dua tahapan proses yang perlu dilakukan dalam penentuan $V(\widehat{X})$. Selain itu, perlu diketahui bahwa pada varians terdapat dua komponen varians, yaitu komponen pertama merupakan varians yang disebabkan oleh pengambilan psu yang disebut dengan varians diantara psu dan komponen kedua merupakan varians yang disebabkan oleh sampel acak yang dipilih dari psu dan disebut varians di dalam psu. Variansi dari $\widehat{X}$ dinyatakan sebagai berikut :

V($\widehat{X}$) = (varians diantara psu) + (varians di dalam psu)

atau secara matematis dapat dinyatakan sebagai berikut:

$$V(\hat{X}) = M^2\frac{M-m}{M}\frac{S_b^2}{m} + \frac{M}{m}\sum_{i=1}^{M} N_i^2\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \tag{3.13}$$

Pembuktian :

Berdasarkan definisi, varians dari $\hat{X} = \frac{M}{m}\sum^{m}\hat{X}_i$ adalah

$$V(\hat{X}) = E(\hat{X} - X)^2 \tag{3.14}$$

Perhatikan bahwa $(\hat{X}-X)^2$ dapat dirubah secara aljabar sebagai berikut:

$$(\hat{X}-X)^2 = \left[\frac{M}{m}\sum^m \hat{X}_i - X\right]^2$$

$$= \left[\left(\frac{M}{m}\sum^m \hat{X}_i - \frac{M}{m}\sum^m X_i\right) + \left(\frac{M}{m}\sum^m X_i - X\right)\right]^2$$

$$= \left(\frac{M}{m}\sum^m X_i - X\right)^2 + 2\left(\frac{M}{m}\sum^m X_i - X\right)\left(\frac{M}{m}\sum^m \hat{X}_i - \frac{M}{m}\sum^m X_i\right) + \left(\frac{M}{m}\sum^m \hat{X}_i - \frac{M}{m}\sum^m X_i\right)^2$$

$$= \left(\frac{M}{m}\sum^m X_i - X\right)^2 + 2\left(\frac{M}{m}\right)\left(\frac{M}{m}\sum^m X_i - X\right)\sum(\hat{X}_i - X_i) + \left(\frac{M}{m}\right)^2\left[\sum^m(\hat{X}_i - X_i)\right]^2$$

$$= \left(\frac{M}{m}\sum^m X_i - X\right)^2 + 2\left(\frac{M}{m}\right)\left(\frac{M}{m}\sum^m X_i - X\right)\sum(\hat{X}_i - X_i) + \left(\frac{M}{m}\right)^2\sum^m(\hat{X}_i - X_i)^2 + \left(\frac{M}{m}\right)^2\sum^m_{i\neq i'}(\hat{X}_i - X_i)(\hat{X}_{i'} - X_i)$$

= A + B + C + D

Kemudian menentukan nilai ekspektasi dari $(\hat{X}-X)^2$ dengan i (psu) konstan. Karena itulah mengapa pada penentuan nilai ekspektasi melibatkan ssu yang dinotasikan dengan j.

$$E_j(\hat{X}-X)^2 = E_j(A) + E_j(B) + E_j(C) + E_j(D) \qquad (3.15)$$

- Penentuan $E_j(A)$

$$E_j\left(\frac{M}{m}\sum^m X_i - X\right)^2 = \left(\frac{M}{m}\sum^m X_i - X\right)^2$$

- Penentuan $E_j(B)$

$$E_j(B) = E_j\left[2\left(\frac{M}{m}\right)\left(\frac{M}{m}\sum^m X_i - X\right)\sum(\hat{X}_i - X_i)\right]$$

$$= 2\left(\frac{M}{m}\right)\left(\frac{M}{m}\sum^m X_i - X\right)E_j\sum^m(\hat{X}_i - X_i)$$

karena $\hat{X}_i$ adalah statistik yang diperoleh dari sampling acak pada ssu. Bagaimanapun, telah diketahui bahwa $E(\hat{X}_i) = X_i$ dan diketahui dari teori statistika bahwa:

$$E\sum X = \sum EX$$

Karena itu,

$$E_j \sum^{m} (\hat{X}_i - X_i) = \sum^{m} E_j (\hat{X}_i - X_i) = 0$$

Sehingga diperoleh $E_j(B) = 0$.

- Penentuan $E_j(D)$

  Untuk, $E_j(D)$ analog dengan penentuan $E_j(B)$ diperoleh

$$E_j \left(\frac{M}{m}\right)^2 \sum_{i \neq i\prime}^{m} (\hat{X}_i - X_i)(\hat{X}_{i\prime} - X_i) = 0$$

- Penentuan $E_j(C)$

$$E_j \left[\left(\frac{M}{m}\right)^2 \sum^{m} (\hat{X}_i - X_i)^2\right] = \left(\frac{M}{m}\right)^2 E_j \sum^{m} (\hat{X}_i - X_i)^2$$

$$= \left(\frac{M}{m}\right)^2 \sum^{m} E_j (\hat{X}_i - X_i)^2$$

Dengan menggunakan sampling acak sederhana, diperoleh

$$E_j (\hat{X}_i - X_i)^2 = N_i^2 \frac{N_i - n_i}{N_i} \frac{S_i^2}{n_i}$$

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j}^{N_i} (X_{ij} - \bar{X}_i)^2$$

S_i^2 yang mana merupakan varians untuk X_{ij} ketika sampling acak sederhana digunakan.

Selanjutnya mensubstitusikan hasil dari penentuan $E_j(A)$ dengan penentuan $E_j(C)$ diperoleh:

$$E_j(\hat{X}-X)^2 = \left(\frac{M}{m}\sum_i^m X_i - X\right)^2 + \left(\frac{M}{m}\right)^2 \sum^m N_i^2 \frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \tag{3.16}$$

i pada $E_i(\hat{X}-X)^2$ tidak dianggap konstan dan $E_j(\hat{X}-X)^2 = y_i$ menjadi variabel acak. Permasalahan selanjutnya terletak pada ekspektasi dari variabel acak tersebut yang diperlihatkan sebagai berikut:

$$E_i\left[E_j(\hat{X}-X)^2\right] = E_i\left[\frac{M}{m}\sum_i^m X_i - X\right]^2 + E_i\left(\frac{M}{m}\right)^2 \sum^m N_i^2 \frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \tag{3.17}$$

Misalkan bagian II pada ruas kanan dari persamaan (3.17) ditulis sebagai berikut:

$$E_i\left(\frac{M}{m}\right)^2 \sum^m U_i$$

dimana

$$U_i = N_i^2 \frac{N_i - n_i}{N_i}\frac{S_i^2}{n_i}$$

Sebagaimana yang ditunjukkan di atas, U_i $(i = 1,2,3,\dots,M)$ adalah variabel acak dengan M nilai yang mungkin dimana masing-masing memiliki probabilitas $\frac{1}{M}$, karena masing-masing psu dipilih menggunakan sampling acak sederhana, maka:

$$\begin{aligned}
E_i\left(\frac{M}{m}\right)^2 \sum^m U_i &= \left(\frac{M}{m}\right)^2 \sum_i^m E_i U_i \\
&= \left(\frac{M}{m}\right)^2 \sum_i^m \left[\sum_i^M \frac{1}{M} U_i\right] \\
&= \left(\frac{M}{m}\right)^2 m \frac{1}{M} \sum^M N_i^2 \frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \qquad (3.18)
\end{aligned}$$

Seperti yang telah ditunjukkan di atas, persamaan (3.18) adalah variansi karena S_i^2 merupakan variansi dalam psu ke i. Selanjutnya untuk bagian I ruas kanan dari persamaan (3.17), diperoleh:

$$E_i\left[\frac{M}{m}\sum^{m} X_i - X\right]^2$$

dimana

$$\frac{M}{m}\sum^{m} X_i$$

dapat dipertimbangkan sebagai estimasi dari X berdasarkan pada sampling acak dari m psu. Kemudian, dengan menggunakan perumusan pada metode sampling acak sederhana, diperoleh:

$$E_i\left[\frac{M}{m}\sum^{m} X_i - X\right]^2 = M^2 E_i\left[\frac{1}{m}\sum^{m} X_i - \bar{X}\right]^2$$

$$= M^2\,\frac{M-m}{M}\,\frac{1}{m}\,\frac{\sum(X_i-\bar{X})^2}{M-1} \qquad (3.19)$$

Berdasarkan persamaan (3.17), persamaan (3.18), dan persamaan (3.19), diperoleh varians dari $\hat{X}$ adalah :

$$V(\hat{X}) = \frac{M^2}{m}\frac{M-m}{M}S_b^2 + \frac{M}{m}\sum^{M} N_i^2\,\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i}$$

dengan

$$S_b^2 = \frac{1}{M-1}\sum_{i=1}^{M}(X_i-\bar{X})^2 \qquad (3.20)$$

$$S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(X_{ij}-\bar{\bar{X}}_i)^2 \qquad (3.21)$$

$$\bar{X} = \frac{X}{M} \qquad (3.22)$$

$$\bar{\bar{X}}_i = \frac{X_i}{N_i} \qquad (3.23)$$

dimana S_b^2 merupakan variansi populasi diantara total kelompok yang menunjukkan sebaran X_i di sekitar $\bar{X}$, dan S_i^2 merupakan variansi populasi di dalam kelompok yang menunjukkan sebaran X_{ij} di sekitar $\bar{\bar{X}}_i$.

Pada sampling acak sederhana, $\frac{S_i^2}{n_i}$ menunjukkan variansi sampel dari $\bar{\bar{x}}_i$ di sekitar $\overline{\overline{X}}_\iota$, dimana $\bar{\bar{x}}_i$ merupakan rata-rata sampel dari sebuah sampel berukuran n_i yang diambil dari N_i. Diketahui ${N_i}^2\left(\frac{S_i^2}{n_i}\right)$ merupakan variansi sampling dari $\hat{X}_i = N_i\bar{x}_i$ disekitar $X_i = N_i\bar{X}_i$. Ketika $N_i = n_i$, $\bar{x}_i = \bar{X}_i$, maka $\hat{X}_i = X_i$ dan variansi sampling dari $\hat{X}_i$ disekitar X_i menjadi sama dengan nol. Sehingga apabila semua unit sampling dalam psu dipilih ($N_i = n_i$), maka

$$V(\hat{X}) = M^2\,\frac{M-m}{M}\frac{S_b^2}{m} + 0 \qquad (3.24)$$

Persamaan (3.24) merupakan varians dari penaksir total pada keadaan semua unit sampling dalam psu dipilih. Hal tersebut akan mengakibatkan variansi dalam psu menjadi sama dengan nol dan karenanya $V(\hat{X})$ hanya dipengaruhi oleh S_b^2. Begitupun sebaliknya, pada keadaan apabila semua *secondary sampling units* (ssu) diambil dari semua M, dalam arti M = m, maka:

$$V(\hat{X}) = 0 + \frac{M}{m}\sum_{i=1}^{M} N_i^2\,\frac{N_i - n_i}{N_i}\frac{S_i^2}{n_i}$$

$$V(\hat{X}) = \frac{M}{m}\sum_{i=1}^{M} N_i^2\,\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \qquad (3.25)$$

Hal tersebut akan mengakibatkan variansi diantara kelompok menjadi sama dengan nol dan karenanya $V(\hat{X})$ hanya diperngaruhi oleh S_i^2.

3.4.2 Penaksir Variansi dari Penaksir Total Populasi Sampling Berkelompok

Pada populasi yang berukuran cukup besar, sulit untuk menentukan V($\hat{X}$) secara langsung, sehingga untuk penentuannya dapat dilakukan dengan menggunakan penaksirnya. Penaksir dari V($\hat{X}$) dinotasikan dengan $\hat{\mathrm{V}}$($\hat{X}$). Seperti telah dikemukakan sebelumnya bahwa V($\hat{X}$) dibentuk dari dua komponen varians, yaitu varians diantara psu (S_b^2) dan varians dalam psu (S_i^2). Oleh karena itu, penaksir $V(\hat{X})$ dapat diperoleh dengan menggunakan penaksir-penaksir dari S_b^2 dan S_i^2. Berdasarkan penjelasan di atas, maka penaksir $V(\hat{X})$ dirumuskan sebagai berikut:

$$\hat{V}(\hat{X}) = M^2 \frac{M-m}{M}\frac{s_b^2}{m} + \frac{M}{m}\sum_{i=1}^{m} N_i^2 \frac{(N_i-n_i)}{N_i}\frac{s_i^2}{n_i} \qquad (3.26)$$

dimana

$$s_b^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{X}_\iota - \hat{\hat{X}})^2 \qquad (3.27)$$

$$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij} - \bar{\bar{x}}_i)^2 \qquad (3.28)$$

$\hat{X}_\iota = N_i\bar{\bar{x}}_i$ merupakan penaksir total dari kelompok ke-i, $\bar{\bar{x}}_i = \frac{x_i}{n_i}$ merupakan rata-rata sampel dari subsampel n_i, dan $\hat{\hat{X}} = \frac{1}{m}\sum_{i=1}^{m}\hat{X}_i$ merupakan rata-rata sampel dari $\hat{X}_\iota, i = 1,2,..,m$.

Seperti telah diketahui bahwa s_i^2 merupakan varians dari x_{ij} dalam kelompok utama dari psu ke-i. Karena n_i adalah sampel acak dari N_i dan $\bar{\bar{x}}_i$ adalah rata-rata sampel dari n_i, maka dapat diketahui bahwa s_i^2 adalah penaksir tak bias dari S_i^2 dan dinyatakan sebagai berikut:

$$E(s_i^2) = S_i^2 \qquad (3.29)$$

Varians antar psu (kelompok) dinotasikan dengan s_b^2, namun untuk s_b^2 ternyata bukan merupakan penaksir tak bias dari S_b^2. Hal ini dapat dilihat pada pembahasan dibawah ini.

$$E(s_b^2) = S_b^2 + \frac{1}{M}\sum_{i=1}^{M} N_i^2 \frac{(N_i-n_i)}{N_i}\frac{S_i^2}{n_i} \qquad (3.30)$$

Pembuktian :

Sebagaimana yang telah dikemukakan sebelumnya bahwa $E_j(s_i^2) = S_i^2$, namun $E_i(s_b^2) \neq S_b^2$, sehingga hal tersebut mengakibatkan S_b^2 tidak dapat ditaksir berdasarkan sampelnya yaitu s_b^2. s_b^2 didefinisikan sebagai berikut:

$$s_b^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{X}_\iota - \hat{\hat{X}}\right)^2$$

$$= \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{X}_\iota - \frac{\sum_{i=1}^{m}\hat{X}_\iota}{m}\right)^2$$

$$= \frac{1}{m-1}\frac{m}{m}\sum_{i=1}^{m}\left(\hat{X}_\iota - \frac{1}{M}\frac{M}{m}\sum_{i=1}^{m}\hat{X}_\iota\right)^2$$

$$s_b^2 = \frac{m}{m-1}\frac{1}{m}\sum_{i=1}^{m}\left(\hat{X}_\iota - \frac{\hat{X}}{M}\right)^2 \qquad (3.31)$$

Persamaan (3.31) dapat dinyatakan dalam bentuk persamaan berikut:

$$\frac{m-1}{m}s_b^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{X}_i - \frac{\hat{X}}{M}\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[\hat{X}_i^2 - 2\hat{X}_i\left(\frac{\hat{X}}{M}\right) + \left(\frac{\hat{X}}{M}\right)^2\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{m}\hat{X}_i^2 - 2m\left(\frac{\hat{X}}{M}\right)^2 + m\left(\frac{\hat{X}}{M}\right)^2\right]$$

$$\frac{m-1}{m}s_b^2 = \frac{1}{m}\sum_{i=1}^{m}\hat{X}_i^2 - \left(\frac{\hat{X}}{M}\right)^2 \qquad (3.32)$$

Selanjutnya menentukan nilai ekspektasi dari persamaan (3.32), diperoleh:

$$E\left(\frac{m-1}{m}s_b^2\right) = E_i\left(E_j\left(\frac{1}{m}\sum_{i=1}^{m}\hat{X}_i^2\right)\right) - E\left(\frac{\hat{X}}{M}\right)^2$$

$$= E_i\left(\frac{1}{m}\sum_{i=1}^{m}E_j\left(\hat{X}_i^2\right)\right) - \frac{E\left(\hat{X}^2\right)}{M^2}$$

$$= \frac{1}{m}.m.\sum_{i=1}^{M}\frac{1}{M}E_j\left(\hat{X}_i^2\right) - \frac{E\left(\hat{X}^2\right)}{M^2}$$

$$E\left(\frac{m-1}{m}s_b^2\right) = \frac{1}{M}\sum_{i=1}^{M}E_j\left(\hat{X}_i^2\right) - \frac{E\left(\hat{X}^2\right)}{M^2}$$

Selanjutnya adalah menentukan $E_j\left(\hat{X}_i^2\right)$ dan $E\left(\hat{X}^2\right)$. Penentuan $E_j\left(\hat{X}_i^2\right)$ dan $E\left(\hat{X}^2\right)$ dapat dilakukan dengan menggunakan perumusan umum dari varians, yaitu:

$$E(\bar{x} - \bar{X})^2 = E(\bar{x}^2) - \bar{X}^2$$

$$E(\bar{x}^2) = E(\bar{x} - \bar{X})^2 + \bar{X}^2$$

Dengan menggunakan perumusan umum tersebut di atas, $E_j\left(\hat{X}_i^2\right)$ dapat dinyatakan dalam bentuk berikut ini.

$$\hat{X}_i = \hat{X}_i - X_i + X_i$$

$$\hat{X}_i^2 = \left(\hat{X}_i - X_i\right)^2 + X_i^2 + 2\left(\hat{X}_i - X_i\right)X_i$$

Pada ekspektasi bersyarat E_j sepanjang j dengan i dianggap konstan, akan diperoleh

$$E_j\left(\hat{X}_i^2\right) = E_j\left(\hat{X}_i - X_i\right)^2 + E_j(X_i^2) + 0$$

$$E_j\left(\hat{X}_i^2\right) = V_j\left(\hat{X}_i\right) + X_i^2$$

Namun $V_j(\hat{X}_i)$ tersebut merupakan varians untuk $\hat{X}_i$ pada kondisi sampling acak sederhana ketika *i* diasumsikan telah ditetapkan dan X_i konstan ketika diasumsikan nilai *i* telah ditetapkan. Oleh karena itu,

$$E_j(\hat{X}_i^2) = N_i^2 \frac{N_i - n_i}{N_i}\frac{S_i^2}{n_i} + X_i^2$$

Selanjutnya menentukan $E(\hat{X}^2)$ analog dengan $E_j(\hat{X}_i^2)$, akan diperoleh:

$$\hat{X} \ = \hat{X} - X + X$$

$$\hat{X}^2 = (\hat{X} - X)^2 + X^2 + 2(\hat{X} - X)X \qquad (3.33)$$

Selanjutnya menentukan ekspektasi pada kedua ruas persamaan (3.33) diperoleh

$$E\hat{X}^2 = E(\hat{X} - X)^2 + EX^2 + 0$$

$$= V(\hat{X}) + X^2$$

$$= M(M - m)\frac{S_b^2}{m} + \frac{M}{m}\sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i} + X^2$$

Sehingga, pada akhirnya diperoleh bahwa

$$E\left(\frac{m-1}{m}s_b^2\right) = \frac{1}{M}\sum_{i=1}^{M}\left(N_i^2\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} + X_i^2\right)$$

$$-\left(\frac{1}{M}\right)^2\left[M(M - m)\frac{S_b^2}{m} + \frac{M}{m}\sum_{i=1}^{M}(N_i - n_i)\frac{S_i^2}{n_i} + X^2\right]$$

$$= \frac{1}{M}A + \frac{1}{M}\sum_{i=1}^{M} X_i^2 - \left(\frac{1}{M}\right)^2\left[M(M - m)\frac{S_b^2}{m} + \frac{M}{m}A + X^2\right]$$

$$= \frac{1}{M}\left(1 - \frac{1}{m}\right)A + \left(\frac{1}{M}\sum_{i=1}^{M} X_i^2 - \frac{X^2}{M^2}\right) - \frac{M-m}{M}\frac{S_b^2}{m}$$

$$= \frac{1}{M}\left(1 - \frac{1}{m}\right)A + \left(\frac{M-1}{M}S_b^2\right) - \frac{M-m}{M}\frac{S_b^2}{m}$$

$$E\left(\frac{m-1}{m}s_b^2\right) \ = \frac{1}{M}\left(\frac{m-1}{m}\right)A + \frac{m-1}{m}S_b^2 \qquad (3.34)$$

dengan $A = \sum_{i=1}^{M} N_i^2 \frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i}$.

Selanjutnya dengan mengeluarkan $\left(\frac{m-1}{m}\right)$ dari kedua ruas pada persamaan (3.34), akan diperoleh:

$$E(s_b^2) \ = \frac{1}{M}A + S_b^2$$

$$E(s_b^2) \ = S_b^2 + \frac{1}{M}\sum_{i=1}^{M} N_i^2 \frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \qquad (3.35)$$

Pembahasan selanjutnya yaitu mengenai pembuktian yang akan menunjukkan bahwa $\hat{V}(\hat{X})$ merupakan penaksir yang tak bias dari $V(\hat{X})$. Seperti telah diketahui, bahwa $\hat{V}(\hat{X})$ harus merupakan penaksir tak bias dari $V(\hat{X})$, dengan kata lain harus memenuhi ketentuan berikut:

$$E[\hat{V}(\hat{X})] = V(\hat{X}) \tag{3.36}$$

Pada pembahasan sebelumnya telah diketahui bahwa perumusan penaksir dari $V(\hat{X})$ adalah seperti yang tersajikan pada persamaan (3.26). Pada perumusan (3.26), penjumlahan bentuk kedua pada ruas kanan dijumlahkan sepanjang m bukan sepanjang M. Sebagaimana telah diketahui bahwa s_i^2 merupakan penaksir tak bias dari S_i^2, namun s_b^2 bukan merupakan penaksir tak bias dari S_b^2. Oleh karena itu, penaksir dari $V(\hat{X})$ tidak dapat diperoleh secara langsung dengan cara mengganti notasi-notasi dari S_b^2 dan S_i^2 dengan notasi-notasi s_b^2 dan s_i^2. Walaupun demikian, sebagaimana telah dikemukakan bahwa $\hat{V}(\hat{X})$ merupakan penaksir tak bias untuk $V(\hat{X})$. Oleh karena itu, harus dibuktikan bahwa $E\left(\hat{V}(\hat{X})\right) = V(\hat{X})$.

<u>Pembuktian :</u>

Ekspektasi dari $V(\hat{X})$ harus dipandang dalam dua tahapan yaitu ekspektasi yang berkaitan dengan tahapan pertama sampling dan ekspektasi bersyarat yang berkaitan dengan tahapan kedua sampling, dengan menganggap tahapan pertama psu konstan.

$$E\left(\hat{V}(\hat{X})\right) = M(M-m)\frac{E(s_b^2)}{m} + \frac{M}{m}E_i\left[E_j\left(\sum_{i=1}^{m} N_i(N_i - n_i)\frac{s_i^2}{n_i}\right)\right] \tag{3.37}$$

dengan E_j merupakan ekspektasi bersyarat sepanjang j dan menganggap psu ke-i konstan.

Untuk mempermudah pembuktian, ruas kanan persamaan (3.37) dibagi menjadi dua bagian, yaitu bagian I dan bagian II. Pertama-tama substitusikan persamaan (3.30) pada bagian I ruas kanan persamaan (3.37), sehingga bagian I ruas kanan persamaan (3.37) menjadi:

$$M(M-m)\frac{1}{m}E(s_b^2) = M(M-m)\frac{1}{m}s_b^2 + (M-m)\frac{1}{m}\sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i} \tag{3.38}$$

Selanjutnya menentukan ekspektasi bersyarat dari sepanjang j dengan i dianggap konstan, dan kemudian ambil ekspektasi sepanjang i dari bagian II ruas kanan persamaan (3.37). Karena pemilihan ssu berasal dari sampel acak sederhana, dengan i dianggap konstan, maka $E_j(s_i^2)$ ekuivalen dengan penentuan ekspektasi untuk kasus sampel acak sederhana. Oleh karena itu, diketahui bahwa $E_j(s_i^2) = S_i^2$, sehingga bagian II ruas kanan dari persamaan (3.37) menjadi :

$$\frac{M}{m} E_i\left[E_j\left(\sum_{i=1}^{m} N_i(N_i - n_i)\frac{s_i^2}{n_i}\right)\right] = \frac{M}{m} m \, \frac{1}{M}\sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i}$$

$$\frac{M}{m} E_i\left[E_j\left(\sum_{i=1}^{m} N_i(N_i - n_i)\frac{s_i^2}{n_i}\right)\right] = \sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i} \quad (3.39)$$

Selanjutnya, dengan mensubstitusikan persamaan (3.38) dan persamaan (3.39) pada persamaan (3.37), akan diperoleh:

$$E\left[\hat{V}\left(\hat{X}\right)\right] = M(M-m)\frac{S_b^2}{m} + (M-m)\frac{1}{m}\sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i} + \sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i}$$

$$= M(M-m)\frac{S_b^2}{m} + \frac{M}{m}\sum_{i=1}^{M} N_i(N_i - n_i)\frac{S_i^2}{n_i}$$

$$E\left[\hat{V}\left(\hat{X}\right)\right] = V(\hat{X})$$

Hal ini menunjukan bahwa $\hat{V}\left(\hat{X}\right)$ adalah penaksir tak bias dari $V\left(\hat{X}\right)$.

3.4.3 Hubungan $V\left(\hat{\bar{\bar{X}}}\right)$ dengan $V(\hat{X})$

Pada subbab ini akan membahas mengenai hubungan antara varians dari $\hat{\bar{\bar{X}}}$ dengan varians dari $\hat{X}$**.** Karena pada studi kasus untuk sampling berkelompok pada skripsi ini yang akan ditentukan adalah varians dari $\hat{X}$, karena itu, perlu untuk melihat hubungan antara varians dari $\hat{X}$ dengan varians dari $\hat{\bar{\bar{X}}}$.

Varians dari $\hat{\bar{\bar{X}}}$ dinyatakan sebagai berikut:

$$V\left(\hat{\bar{\bar{X}}}\right) = V\left(\frac{\hat{X}}{N}\right)$$

$$= \frac{1}{N^2}V\left(\hat{X}\right)$$

$$V\left(\hat{\bar{\bar{X}}}\right) = \frac{V(\hat{X})}{N^2} \quad (3.40)$$

Selanjutnya dengan mensubstitusikan persamaan (3.13) pada persamaan (3.40), akan diperoleh:

$$V\left(\hat{\bar{\bar{X}}}\right) = \frac{1}{N^2}\left(M^2\frac{M-m}{M}\frac{S_b^2}{m} + \frac{M}{m}\sum^{M} N_i^2\frac{N_i - n_i}{N_i}\frac{S_i^2}{n_i}\right) \quad (3.41)$$

Berdasarkan persamaan (3.40), diketahui bahwa hubungan antara $V\left(\hat{\bar{\bar{X}}}\right)$ dengan $V(\hat{X})$ berbanding lurus, hal ini mempunyai makna bahwa semakin besar nilai $V(\hat{X})$, maka semakin besar pula nilai $V\left(\hat{\bar{\bar{X}}}\right)$ begitu pun sebaliknya.

3.5 Sampling Berkelompok dengan *Probability Proportional to Size* (PPS)

Pada sampling berkelompok, probabilitas pemilihan sebuah psu disamaratakan. Akan tetapi, ketika beberapa *cluster* berukuran besar dan yang lainnya berukuran kecil, maka sebaiknya *cluster* yang berukuran besar tersebut harus mendapat probabilitas yang lebih besar pada pemilihan sampel. Hal tersebut tidak berlaku pada sampling berkelompok. Prosedur pada pemilihan psu untuk menjadi sampel, merupakan point pembeda antara metode sampling berkelompok dan metode sampling berkelompok dengan *Probability Proportional To Size* (PPS).

Pada sampling berkelompok dengan PPS akan lebih disoroti mengenai bagaimana memilih sampel yang representatif bagi populasi yaitu dengan cara memberikan kesempatan yang berbeda pada setiap psu berdasarkan ukurannya untuk terpilih menjadi sampel. Alasan untuk mendesain prosedur pengambilan sampel dengan probabilitas psu yang berbeda adalah untuk membentuk sebuah metode pemilihan yang akan memberikan penaksir-penaksir yang tak bias dari rata-rata populasi dan juga akan membuat presisi yang lebih besar daripada metode sampling berkelompok. Besar peluang pemilihan *cluster* ke-i dinotasikan dengan p_i dan didefiniskan sebagai berikut:

$$p_i = \frac{N_i}{\sum N_i} \tag{3.42}$$

3.5.1 Penaksir Rata-Rata dan Total Populasi Sampling Berkelompok dengan PPS

Penaksir rata–rata pada sampling berkelompok dengan PPS dinotasikan dengan $\bar{\bar{x}}_{pps}$ dan dirumuskan sebagai berikut:

$$\hat{\bar{\bar{X}}} = \bar{\bar{x}}_{pps} = \frac{1}{m\bar{n}} \sum_{i=1}^{m} \sum_{j=1}^{\bar{n}} x_{ij} \tag{3.43}$$

dimana $\bar{\bar{x}}_{pps}$ merupakan penaksir tak bias dari $\bar{\bar{X}}$. Berdasarkan informasi tersebut, penaksir total populasi dinyatakan sebagai berikut:

$$\hat{X} = \frac{N}{m\bar{n}} \sum_{i=1}^{m} \sum_{j=1}^{\bar{n}} x_{ij} \tag{3.44}$$

Pembuktian :

Akan membuktikan bahwa $\bar{\bar{x}}_{pps}$ merupakan penaksir tak bias dari $\bar{\bar{X}}$.

Sebelumnya, pada sampling berkelompok rata-rata populasi didefinisikan sebagai berikut:

$$\bar{\bar{X}} = \frac{X}{N}$$

Sementara itu pada pembahasan sebelumnya telah dikemukakan bahwa, penaksir dari total populasi X adalah $\hat{X}$ dan dirumusakan sebagai berikut:

$$\hat{X} = \frac{M}{m} \sum^{m} \frac{N_i}{n_i} \sum^{n_i} x_{ij} \tag{3.45}$$

Berdasarkan perumusan di atas (3.45) maka proses untuk mencari penaksir dari $\bar{\bar{X}}$ sangatlah sederhana hanya dengan membagi penaksir dari total populasi oleh N seperti berikut.

$$\hat{\bar{\bar{X}}} = \frac{\hat{X}}{N} = \frac{1}{N}\frac{M}{m} \sum^{m} \frac{N_i}{n_i} \sum^{n_i} x_{ij} \tag{3.46}$$

Perhatikan bahwa $\hat{\bar{\bar{X}}}$ bukan merupakan rata-rata sampel, melainkan penaksir dari rata-rata populasi. Telah ditunjukkan bahwa $\hat{X}$ merupakan penaksir tak bias dari X. Oleh karena itu itu, nilai ekspektasi dari $\hat{\bar{\bar{X}}}$ adalah

$$E\left(\hat{\bar{\bar{X}}}\right) = \frac{E(\hat{X})}{N} = \frac{X}{N} = \bar{\bar{X}}$$

Terbukti bahwa $\hat{\bar{\bar{X}}}$ merupakan penaksir tak bias dari $\bar{\bar{X}}$.

Selanjutnya berdasarkan (3.45) dengan mengasumsikan bahwa $N_i = \bar{N} = \frac{N}{M}$ dan $n_i = \bar{n} = \frac{n}{m}$, maka perumusan $\hat{\bar{\bar{X}}}$ akan menjadi :

$$\hat{\bar{\bar{X}}} = \frac{1}{m\bar{n}} \sum_{i=1}^{m} \sum_{i=1}^{\bar{n}} x_{ij} \tag{3.47}$$

Karakteristik dari persamaan (3.47) adalah bahwa $\hat{\bar{\bar{X}}}$ merupakan rata-rata sampel dalam bentuk sederhana berukuran $m\bar{n}$ dan tidak ada pembobotan apapun yang digunakan untuk memperoleh $\hat{\bar{\bar{X}}}$. Penaksir jenis ini dikatakan sebagai penaksir dengan pembobotan diri.

Pada sampel pembobotan diri, probabilitas untuk memasukkan sebuah ssu dari populasi ke dalam sampel berukuran $m\bar{n} = n$ adalah sama untuk semua anggota populasi. Hal ini berarti, probabilitas ssu yang masuk ke dalam sampel berukuran $m\bar{n} = n$ adalah $\frac{n}{N}$. Sebanyak m psu dipilih dari M psu secara sampling acak sederhana, oleh karena itu probabilitas semua psu yang termasuk dalam sampel adalah $\frac{m}{M}$.

Sementara itu, untuk ssu yang termasuk dalam subsampel ketika $n_i = n$ akan memiliki probabilitas sebesar $\frac{\bar{n}}{\bar{N}}$, sehingga probabilitas ssu yang termasuk dalam sampel total atau $m\bar{n} = n$ adalah

$$\frac{m}{M} \times \frac{\bar{n}}{\bar{N}} = \frac{m}{M} \times \frac{\frac{n}{m}}{\frac{N}{M}} = \frac{n}{N} \tag{3.48}$$

Berdasarkan perhitungan di atas (3.48), maka probabilitas ssu yang masuk dalam sampel total menjadi $\frac{n}{N}$ dan juga penaksir rata-rata populasi adalah rata-rata sampel yang tidak membutuhkan pembobotan, sehingga dapat mempermudah perhitungan.

Pada sampling berkelompok dengan PPS, rata-rata sampel yang tidak diboboti menjadi penaksir tak bias dari rata-rata populasi. Hal ini diharapkan akan memberikan hasil yang lebih baik dan berguna, karena meskipun ukuran *cluster* bervariasi, namun hanya diperlukan rata-rata sampel untuk menaksir rata-rata populasi. Pada sampling berkelompok dengan PPS, ukuran *cluster* hanya dipergunakan sebagai kriteria dalam pemilihan psu.

Pada pemilihan psu untuk sampling berkelompok dengan PPS, probabilitas psu untuk menjadi sampel adalah sebesar $\frac{N_i}{N}$. Penaksir total populasi untuk masalah pemilihan psu dengan probabilitas berbeda dinyatakan sebagai berikut:

$$\hat{X} = \frac{x_i}{\hat{p}_i} \tag{3.49}$$

dan karena $\hat{p}_i = \frac{X_i}{X}$, maka $\hat{X}$ sama dengan X.

Namun, notasi $\hat{p}_i$ tidak dikenal dalam permasalahan sampling. Sebagai gantinya, dapat menggunakan probabilitas dari X_i. Probabilitas ini dinyatakan oleh p_i. Dengan asumsi bahwa p_i adalah perkiraan yang baik terhadap $\hat{p}_i$, sehingga penaksir total populasi dapat dinyatakan sebagai :

$$\hat{X} = \frac{X_i}{p_i} \tag{3.50}$$

Semakin p_i berbeda dari $\hat{p}_i = \frac{X_i}{X}$, semakin besar pula ketidaksesuaian antara $\hat{X}$ dan X.

Dengan menggunakan probabilitas p_i sebagai pengganti $\hat{p}_i$, maka penaksir $\hat{X}$ menjadi penaksir yang tidak bias dari X. Ini dapat dilihat sebagai berikut:

$$\hat{X} = \frac{X_i}{p_i}$$

$$E(\hat{X}) = \sum_{i=1}^{M} p_i \frac{X_i}{p_i} = \sum_{i=1}^{M} X_i = X \tag{3.51}$$

dengan demikian $\hat{X}$ merupakan penaksir yang tidak bias dari X.

Sebuah penaksir total populasi dapat diperoleh sebagai rata-rata dari penaksir-penaksir tersebut, yaitu:

$$\hat{X} = \frac{1}{m}\sum_{i=1}^{m} \frac{X_i}{p_i} \tag{3.52}$$

Persamaan (3.52) dapat dianggap sebagai rumus umum untuk menaksir X dan juga merupakan penaksir yang tidak bias dari X. Berikut ini merupakan pembuktian bahwa $\hat{X}$ merupakan penaksir yang tidak bias dari X.

$$E(\hat{X}) = \frac{1}{m}\sum_{i=1}^{m} E\frac{X_i}{p_i}$$

$$E(\hat{X}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{i=1}^{M} p_i \frac{X_i}{p_i} = \sum_{i=1}^{M} X_i = X$$

Hubungan antara probabilitas p_i dan proporsi $\hat{p}_i$ adalah jika p_i semakin mendekati $\hat{p}_i = \frac{X_i}{p_i}$, maka semakin tinggi presisi penaksir tersebut. Ketika $p_i = \hat{p}_i$, maka $\hat{X} = X$ dan variansinya adalah nol.

Permasalahan selanjutnya adalah bagaimana cara untuk menentukan p_i yang akan mendekati $\hat{p}_i = \frac{X_i}{p_i}$. Karena X adalah parameter yang tidak diketahui dan yang akan ditaksir, maka $\hat{p}_i$ juga tidak diketahui. Cara menentukan p_i adalah memilih psu menggunakan sampling berkelompok dengan PPS dengan harapan bahwa probabilitasnya akan mendekati $\hat{p}_i$.

Pada prosedur sampling berkelompok, pemilihan psu dilakukan secara sampling acak sederhana, yaitu setiap psu diberikan probabilitas sama yaitu sebesar $\frac{1}{M}$. Sekarang, yang diinginkan adalah menetapkan probabilitas p_i yang akan mendekati proporsi $\hat{p}_i$ untuk mengurangi ketidaksesuaian antara $\hat{X}$ dan X. Dengan kata lain, ingin mengurangi varians $\hat{X}$. Secara umum dari sampling acak sederhana, diketahui bahwa

$$\hat{X}_i = \frac{N_i}{n_i} x_i \tag{3.53}$$

dan diketahui pula dari sampling acak sederhana, bahwa

$$E(\hat{X}_i) = N_i \bar{\bar{X}}_i = X_i \tag{3.54}$$

dengan mengkombinasikan hasil dari proses psu tahap pertama dan proses ssu tahap kedua serta dengan mensubstitusikan persamaan (3.53) kedalam persamaan (3.52), akan diperoleh

$$\hat{X} = \frac{1}{m}\sum^{m} \frac{\hat{X}_i}{p_i}$$

$$= \frac{1}{m}\sum^{m} \frac{1}{p_i}\left(\frac{N_i}{n_i} x_i\right)$$

$$\hat{X} = \frac{1}{m}\sum^{m} \frac{1}{p_i}\left(\frac{N_i}{n_i}\sum^{n_i} x_{ij}\right) \tag{3.55}$$

yang merupakan hasil umum yang ingin ditentukan dimana p_i adalah probabilitas pemilihan X_i, dan persamaan (3.55) merupakan penaksir yang tidak bias dari X. Hal ini dapat diperlihatkan pada penjabaran berikut ini.

$$E(\hat{X}) = E_i\left(E_j \frac{1}{m}\sum^{m} \frac{1}{p_i}\hat{X}_i\right)$$

$$= E_i\left(\frac{1}{m}\sum^{m}\frac{1}{p_i}X_i\right)$$

$$= \frac{1}{m}\sum^{m}\sum^{M}p_i\left(\frac{1}{p_i}X_i\right)$$

$$E(\hat{X}) = \sum^{M}X_i = X$$

Dengan menggunakan persamaan (3.55), maka diperoleh penaksir-penaksir dari sampling berkelompok dan sampling berkelompok dengan PPS sebagai kasus khusus. Pada kasus sebelumnya pada sampling berkelompok, $p_i = \frac{1}{M}$, kemudian substitusikan $p_i = \frac{1}{M}$ ini ke persamaan (3.55), sehingga diperoleh:

$$\hat{X} = \frac{M}{m}\sum^{m}\frac{N_i}{n_i}x_i$$

$$\hat{X} = \frac{M}{m}\sum^{m}\frac{N_i}{n_i}\sum^{n_i}x_{ij} \qquad (3.56)$$

Untuk sampling berkelompok dengan PPS, $p_i = \frac{N_i}{N}$. Kemudian substitusikan $p_i = \frac{N_i}{N}$ ini ke persamaan (3.55), sehingga diperoleh:

$$\hat{X} = \frac{1}{m}\sum^{m}\frac{N}{N_i}\frac{N_i}{n_i}x_i$$

$$= \frac{N}{m}\sum^{m}\frac{1}{n_i}x_i$$

$$\hat{X} = \frac{N}{m}\sum^{m}\frac{1}{n_i}\sum^{n_i}x_{ij} \qquad (3.57)$$

yang merupakan rumus umum sampling berkelompok dengan PPS. Apabila $n_i = \bar{n} = \frac{n}{m}$, maka persamaan (3.57) dapat disederhanakan menjadi:

$$\hat{X} = \frac{N}{m\bar{n}}\sum^{m}\sum^{\bar{n}}x_{ij} \qquad (3.58)$$

Penaksir rata-rata populasi diperoleh hanya dengan membagi $\hat{X}$ oleh N, sehingga diperoleh :

$$\hat{\bar{X}} = \frac{1}{m\bar{n}}\sum^{m}\sum^{\bar{n}}x_{ij} \qquad (3.59)$$

Seperti yang diuraikan di atas, persamaan (3.59) merupakan rata-rata sampel dari sampel $n = m\bar{n}$ dan merupakan pembobotan diri, sehingga merupakan penaksir yang tidak bias dari $\bar{X}$. Oleh Karena $\hat{\bar{\bar{X}}}$ adalah rata-rata sampel $n = m\bar{n}$, dinotasikan dengan:

$$\hat{\bar{\bar{X}}} = \bar{\bar{x}}_{pps} = \frac{1}{m\bar{n}} \sum^{m} \sum^{\bar{n}} x_{ij} \tag{3.60}$$

3.5.2 Variansi dari $\bar{\bar{x}}_{pps}$

Pada subbab ini akan membahas mengenai variansi dari $\bar{\bar{x}}_{pps}$ .Variansi dari $\bar{\bar{x}}_{pps}$ didefinisikan sebagai berikut:

$$V\left(\bar{\bar{x}}_{pps}\right) = E(\bar{\bar{x}}_{pps} - \bar{\bar{X}})^2 \tag{3.61}$$

Perlu diingat bahwa E($\bar{\bar{x}}_{pps}$) = $\bar{\bar{X}}$. Untuk mengevaluasi persamaan (3.61), alangkah baiknya apabila terlebih dahulu menjabarkan $E(\bar{\bar{x}}_{pps} - \bar{\bar{X}})^2$.

Misalkan:

$$\bar{\bar{x}}_{pps} - \bar{\bar{X}} = \left(\bar{\bar{x}}_{pps} - \frac{1}{m}\sum^{m} \bar{\bar{X}}_i\right) + \left(\frac{1}{m}\sum^{m} \bar{\bar{X}}_i - \bar{\bar{X}}\right) = A + B$$

diperoleh,

$$\left(\bar{\bar{x}}_{pps} - \bar{\bar{X}}\right)^{\mathbf{2}} = A^2 + B^2 + 2AB$$

sehingga varians dari $\bar{\bar{x}}_{pps}$ dapat dinyatakan sebagai berikut:

$$V(\bar{\bar{x}}_{pps}) = E\left(\bar{\bar{x}}_{pps} - \bar{\bar{X}}\right)^2 = E(A^2) + E(B^2) + E(2AB) \tag{3.62}$$

Untuk mempermudah dalam proses penurunan rumusanya, maka ruas kanan persamaan (3.62) akan dijabarkan bagian per bagian.

- Penjabaran bagian 2AB

$$2AB = 2\left(\frac{1}{m}\sum^{m} \bar{\bar{x}}_i - \frac{1}{m}\sum^{m} \bar{\bar{X}}_i\right)\left(\frac{1}{m}\sum^{m} \bar{\bar{X}}_i - \bar{\bar{X}}\right)$$

$$= 2\frac{1}{m}\sum^{m} (\bar{\bar{x}}_i - \bar{\bar{X}}_i)\left(\frac{1}{m}\sum^{m} \bar{\bar{X}}_i - \bar{\bar{X}}\right)$$

Misal E_j merupakan ekspektasi yang diambil alih j dengan psu ke-i diketahui, sehingga:

$$E_i[E_j(2AB)] = E_i\left[E_j\left(2\frac{1}{m}\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_\iota)\left(\frac{1}{m}\sum^{m}\bar{\bar{X}}_i - \bar{\bar{X}}\right)\right)\right]$$

Karena A dan B *independent* (karena *cluster* terakhir *independent*) dan $E(\bar{\bar{x}}_i) = \bar{\bar{X}}_i$, maka:

$$E_j(2AB) = 0 \tag{3.63}$$

- Penjabaran bagian A^2

$$A^2 = \left(\bar{\bar{x}}_{pps} - \frac{1}{m}\sum^{m}\bar{\bar{X}}_i\right)^2$$

$$= \left(\frac{1}{m}\sum^{m}\bar{\bar{x}}_i - \frac{1}{m}\sum^{m}\bar{\bar{X}}_i\right)^2$$

$$= \frac{1}{m^2}\left[\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)\right]^2$$

$$= \frac{1}{m^2}\left[\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)^2 + 2\sum_{i}^{m}\sum_{i\neq i'}^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)(\bar{\bar{x}}_{i'} - \bar{\bar{X}}_i)\right]$$

Sehingga, ekspektasi dari A^2 adalah:

$$E_i(E_j(A^2)) = E_i\left[E_j\frac{1}{m^2}\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)^2 + E_j 2\sum_{i}^{m}\sum_{i\neq i'}^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)(\bar{\bar{x}}_{i'} - \bar{\bar{X}}_i)\right] \tag{3.64}$$

Karena *cluster* yang terakhir bersifat *independent*, maka suku kedua dari ruas kanan persamaan (3.64) bernilai 0. Sedangkan untuk suku pertama di ruas kanan persamaan (3.64), diperoleh:

$$E_i\left(E_j\left(\frac{1}{m^2}\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)^2\right)\right) = E_i\left(\frac{1}{m^2}\sum \frac{N_i-\bar{n}}{N_i}\frac{{S_i}^2}{\bar{n}}\right) \tag{3.65}$$

dimana ${S_i}^2 = \frac{1}{N_i-1}\sum^{N_i}(X_{ij} - \bar{\bar{X}}_\iota)^2$ yang merupakan *standard error* dari rata-rata sampel $\bar{\bar{x}}_i$, untuk sampling acak sederhana.

$$E_i\left(E_j\left(\frac{1}{m^2}\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)^2\right)\right) = E_i\left(\frac{1}{m^2}\sum \frac{N_i-\bar{n}}{N_i}\frac{{S_i}^2}{\bar{n}}\right)$$

$$= \frac{1}{m^2}m\sum^{m}\frac{N_i}{N}\frac{N_i-\bar{n}}{N_i}\frac{{S_i}^2}{\bar{n}}$$

$$E_i\left(E_j\left(\frac{1}{m^2}\sum^{m}(\bar{\bar{x}}_i - \bar{\bar{X}}_i)^2\right)\right) = \frac{1}{mN}\sum^{m}(N_i - \bar{n})\frac{{S_i}^2}{\bar{n}}$$

$$E_i(A^2) = \frac{1}{mN}\sum^m (N_i - \bar{n})\frac{{S_i}^2}{\bar{n}} \tag{3.66}$$

- Penjabaran bagian B^2

$$B^2 = \left(\frac{1}{m}\sum^m \bar{\bar{X}}_i - \bar{\bar{X}}\right)^2$$

$$B^2 = \frac{1}{m^2}\left[\sum^m (\bar{\bar{X}}_i - \bar{\bar{X}})^2 + 2\sum_i^m \sum_{i\neq i'}^m (\bar{\bar{X}}_i - \bar{\bar{X}})(\bar{\bar{X}}_{i'} - \bar{\bar{X}})\right] \tag{3.67}$$

Karena psu adalah sampel dengan penggantian dan $E_i(\bar{\bar{X}}_i) = \sum^m \frac{N_i}{N}\bar{\bar{X}}_i = \bar{\bar{X}}$, maka suku kedua ruas kanan persamaan (3.67) bernilai sama dengan 0. Sedangkan untuk suku pertama di ruas kanan persamaan (3.67), diperoleh:

$$\frac{1}{m^2}E_i\left[\sum^m (\bar{\bar{X}}_i - \bar{\bar{X}})^2\right] = \frac{1}{m^2}m\sum^m \frac{N_i}{N}(\bar{\bar{X}}_i - \bar{\bar{X}})^2$$

$$\frac{1}{m^2}E_i\left[\sum^m (\bar{\bar{X}}_i - \bar{\bar{X}})^2\right] = \frac{1}{mN}\sum^m N_i(\bar{\bar{X}}_i - \bar{\bar{X}})^2$$

Berdasarkan itu, maka ekspektasi dari B^2 dapat ditentukan, yaitu:

$$E_i(B^2) = \frac{1}{mN}\sum^m N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 \tag{3.68}$$

Selanjutnya mensubstitusikan persamaan (3.63), persamaan (3.66), dan persamaan (3.68), ke persamaan (3.62), akan diperoleh perumusan varians dari $\bar{\bar{x}}_{pps}$ adalah sebagai berikut:

$$V(\bar{\bar{x}}_{pps}) = \frac{1}{mN}\sum^m N_i(\bar{\bar{X}}_i - \bar{\bar{X}})^2 + \frac{1}{mN}\sum^m (N_i - \bar{n})\frac{{S_i}^2}{\bar{n}} \tag{3.69}$$

dimana

$${S_i}^2 = \frac{1}{N_i - 1}\sum^{N_i} (X_{ij} - \bar{\bar{X}}_i)^2$$

3.5.3 Penaksir Tak Bias dari $V(\bar{\bar{x}}_{pps})$

Salah satu karakteristik yang menarik dari sampling berkelompok dengan PPS adalah penaksir tak bias dari $V(\bar{\bar{x}}_{pps})$ sangat sederhana dan mudah untuk ditentukan. Penaksir tak bias dari $V(\bar{\bar{x}}_{pps})$ adalah:

$$\hat{V}(\bar{\bar{x}}_{pps}) = \frac{1}{m(m-1)}\sum^m \left(\bar{\bar{x}}_i - \bar{\bar{x}}_{pps}\right)^2 \tag{3.70}$$

dengan, $\bar{\bar{x}}_i = \frac{x_i}{\bar{n}}$, dan $\bar{\bar{x}}_{pps} = \frac{1}{m\bar{n}} \sum^m \sum^{\bar{n}} x_{ij} = \frac{1}{m} \sum^m \bar{\bar{x}}_i$

Pembuktian bahwa $\hat{V}(\bar{\bar{x}}_{pps})$ merupakan penaksir yang tak bias dari $V(\bar{\bar{x}}_{pps})$ adalah sebagai berikut:

<u>Pembuktian :</u>

Pada proses pembuktian ini yang akan dilakukan yaitu akan menunjukkan bahwa persamaan (3.70) adalah penaksir tak bias dari

$$V(\bar{\bar{x}}_{pps}) = \frac{1}{mN} \sum^m N_i (\bar{\bar{X}}_i - \bar{\bar{X}})^2 + \frac{1}{mN} \sum^m (N_i - \bar{n}) \frac{{S_i}^2}{\bar{n}}$$

Langkah pertama yaitu melakukan perubahan secara aljabar pada $\left(\bar{\bar{x}}_i - \bar{\bar{x}}_{pps}\right)$.

$$\bar{\bar{x}}_i - \bar{\bar{x}}_{pps} = \bar{\bar{x}}_i - \frac{1}{m} \sum \bar{\bar{x}}_i$$

$$= \bar{\bar{x}}_i - \frac{1}{m} \bar{\bar{x}}_i - \frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'}$$

$$\bar{\bar{x}}_i - \bar{\bar{x}}_{pps} = \frac{m-1}{m} \bar{\bar{x}}_i - \frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \tag{3.71}$$

Selanjutnya mensubstitusikan persamaan (3.71) pada persamaan (3.70), diperoleh:

$$\hat{V}(\bar{\bar{x}}_{pps}) = \frac{1}{m} \frac{1}{(m-1)} \sum^m \left(\frac{m-1}{m} \bar{\bar{x}}_i - \frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \right)^2$$

$$= \frac{1}{m} \left[\frac{1}{m-1} \sum^m \left\{ \left(\frac{m-1}{m} \bar{\bar{x}}_i \right)^2 - 2 \left(\frac{m-1}{m} \bar{\bar{x}}_i \right) \left(\frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \right) + \left(\frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \right)^2 \right\} \right]$$

$$\hat{V}(\bar{\bar{x}}_{pps}) = \frac{1}{m} (A - C + B) \tag{3.72}$$

dengan
$$A = \frac{1}{m-1} \sum^m \left(\frac{m-1}{m} \bar{\bar{x}}_i \right)^2$$

$$B = \frac{1}{m-1} \sum_{i}^{m} \left(\frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \right)^2$$

$$C = \frac{1}{m-1} \sum_i^m 2 \left(\frac{m-1}{m} \bar{\bar{x}}_i \right) \left(\frac{1}{m} \sum_{i' \neq i}^{m} \bar{\bar{x}}_{i'} \right)$$

Langkah kedua adalah mereduksi bagian A, B, dan C secara aljabar untuk menyederhanakan bentuk sehingga dapat menerapkan proses untuk memperoleh hasil ekspektasi.

- Pereduksian bagian A :

$$A = \frac{1}{m-1}\sum_i^m \left(\frac{m-1}{m}\bar{\bar{x}}_i\right)^2$$

$$= \frac{1}{m-1}\frac{(m-1)^2}{m^2}(\sum_i^m \bar{\bar{x}}_i)^2$$

$$A = \frac{m-1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\,2} \qquad (3.73)$$

- Pereduksian bagian B :

$$B \;=\; \frac{1}{m-1}\sum_i^m \left(\frac{1}{m}\sum_{i'\neq i}^m \bar{\bar{x}}_{i'}\right)^2$$

$$= \frac{1}{m-1}\frac{1}{m^2}\sum_i^m \left(\sum_{i'\neq i}^m \bar{\bar{x}}_{i'}\right)^2$$

$$= \frac{1}{m^2(m-1)}\sum_i^m (\sum_i^m \bar{\bar{x}}_i - \bar{\bar{x}}_i)^2$$

$$= \frac{1}{m^2(m-1)}\sum^m\left[(\sum_i^m \bar{\bar{x}}_i)^2 - 2\bar{\bar{x}}_i\sum^m \bar{\bar{x}}_i + \bar{\bar{x}}_i^{\,2}\right]$$

$$= \frac{1}{m^2(m-1)}\left[m(\sum_i^m \bar{\bar{x}}_i)^2 - 2(\sum_i^m \bar{\bar{x}}_i)^2 + \sum_i^m \bar{\bar{x}}_i^{\,2}\right]$$

$$= \frac{1}{m^2(m-1)}\left[(m-2)(\sum_i^m \bar{\bar{x}}_i)^2 + \sum_i^m \bar{\bar{x}}_i^{\,2}\right]$$

$$= \frac{1}{m^2(m-1)}\left[(m-2)\left(\sum_i^m \bar{\bar{x}}_i^{\,2} + \sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right) + \sum_i^m \bar{\bar{x}}_i^{\,2}\right]$$

$$= \frac{1}{m^2(m-1)}\left[(m-1)\sum_i^m \bar{\bar{x}}_i^{\,2} + (m-2)\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right]$$

$$B = \frac{1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\,2} + \frac{m-2}{m^2(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'} \qquad (3.74)$$

- Pereduksian bagian C :

$$C \;= \frac{1}{m-1}\sum^m\left[2\left(\frac{m-1}{m}\bar{\bar{x}}_i\right)\left(\frac{1}{m}\sum_{i'\neq i}^m \bar{\bar{x}}_{i'}\right)\right]$$

$$= \frac{1}{m-1}\frac{m-1}{m}\left[\frac{2}{m}\sum^m \bar{\bar{x}}_i\sum^m\left(\sum_{i'\neq i}^m \bar{\bar{x}}_{i'}\right)\right]$$

$$= \frac{2}{m^2}\sum_i^m \bar{\bar{x}}_i\sum_{i'\neq i}^m \bar{\bar{x}}_{i'}$$

$$C = \frac{2}{m^2}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'} \qquad (3.75)$$

Selanjutnya mensubstitusikan persamaan (3.73), persamaan (3.74), dan persamaan (3.75) ke dalam persamaan (3.72), akan diperoleh:

$$\hat{V}\left(\bar{\bar{x}}_{pps}\right) = \frac{1}{m}\left[\frac{m-1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\ 2} - \frac{2}{m^2}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'} + \frac{1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\ 2} + \frac{m-2}{m^2(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right]$$

$$= \frac{1}{m}\left[\frac{m-1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\ 2} + \frac{1}{m^2}\sum_i^m \bar{\bar{x}}_i^{\ 2} + \frac{m-2}{m^2(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'} - \frac{2}{m^2}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right]$$

$$\hat{V}\left(\bar{\bar{x}}_{pps}\right) = \frac{1}{m}\left[\frac{1}{m}\sum_i^m \bar{\bar{x}}_i^{\ 2} - \frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right] \qquad (3.76)$$

$$\hat{V}\left(\bar{\bar{x}}_{pps}\right) = \frac{1}{m}D \qquad (3.77)$$

dengan $D = \frac{1}{m}\sum_i^m \bar{\bar{x}}_i^{\ 2} - \frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}$

Langkah ketiga adalah menentukan ekspektasi dari $\hat{V}\left(\bar{\bar{x}}_{pps}\right)$. Penentuan ekspektasi dari $\hat{V}\left(\bar{\bar{x}}_{pps}\right)$ ini dapat diselesaikan dalam dua tahap. Tahap pertama untuk kasus dimana psu ke-*i* diberikan dan tahap kedua untuk kasus dimana *i* berubah-ubah dari seluruh kemungkinan M psu.

$$E\left[\hat{V}\left(\bar{\bar{x}}_{pps}\right)\right] = E_i\left(E_j\left(\frac{1}{m}D\right)\right)$$

$$E\left[\hat{V}\left(\bar{\bar{x}}_{pps}\right)\right] = \frac{1}{m}E_i\left(E_j(D)\right) \qquad (3.78)$$

Selanjutnya yang dilakukan yaitu menguraikan $E_j(D)$ menjadi:

$$E_j(D) = E_j\left[\frac{1}{m}\sum_i^m \bar{\bar{x}}_i^{\ 2} - \frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}\right]$$

$$E_j(D) = E_j[F - G] \qquad (3.79)$$

dimana $F = \frac{1}{m}\sum_i^m \bar{\bar{x}}_i^{\ 2}$

$$G = \frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{x}}_i\,\bar{\bar{x}}_{i'}.$$

Perhatikan ruas kanan dari persamaan (3.79). Untuk mempermudah dalam menentukan ekspektasi ruas kanan persamaan (3.79), maka penentuan ekspektasinya dilakukan satu persatu. Pertama-tama menentukan $E_j(F)$.

$$E_jF = E_j\left(\frac{1}{m}\sum_i^m \bar{\bar{x}}_i^{\ 2}\right) = \frac{1}{m}\sum_i^m E_j\bar{\bar{x}}_i^{\ 2} \qquad (3.80)$$

Seperti telah diketahui dari metode sampling acak sederhana bahwa:

$$E(\bar{x} - \bar{X})^2 = E(\bar{x}^2) - \bar{X}^2$$

$$E(\bar{x}^2) = \bar{X}^2 + E(\bar{x} - \bar{X})^2$$

$$E(\bar{x}^2) = \bar{X}^2 + \frac{N-n}{N}\frac{S^2}{n} \qquad (3.81)$$

Dengan menggunakan hubungan pada persamaan (3.81), maka persamaan (3.80) dapat dinyatakan sebagai berikut:

$$E_j(F) = \frac{1}{m}\sum_i^m \left(\bar{\bar{X}}_i^2 + \frac{N_i - \bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right) \quad (3.82)$$

dimana $S_i^2 = \frac{1}{N_i - 1}\sum^{N_i}\left(X_{ij} - \bar{\bar{X}}\right)^2$

Selanjutnya menentukan $E_j(G)$.

$$E_j(G) = \frac{1}{m(m-1)} E_j \sum_{i \neq i'}^m \bar{\bar{x}}_i \, \bar{\bar{x}}_{i'}$$

$$E_j(G) = \frac{1}{m(m-1)} \sum_{i \neq i'}^m \bar{\bar{X}}_i \, \bar{\bar{X}}_{i'} \quad (3.83)$$

Karena pemilihan ssu dilakukan dengan menggunakan metode sampling acak sederhana, maka diasumsikan bahwa psu ke-i yang sama akan terpilih secara berulang kali, ketika sampel yang terambil adalah sampel yang sama dengan sebelumnya, maka sampel tersebut dikembalikan lalu diambil lagi sampel yang lain sehingga terpilih sampel lain yang berbeda, maka dari itu $\bar{\bar{x}}_i$ dan $\bar{\bar{x}}_{i'}$ independen.

Selanjutnya mensubtitusikan persamaan (3.82) dan persamaan (3.83) ke dalam persamaan (3.79), akan diperoleh:

$$E_i\left(E_j(D)\right) = E_i\left[E_j(F) - E_j(G)\right]$$

$$= E_i\left[\frac{1}{m}\sum_i^m\left(\bar{\bar{X}}_i^2 + \frac{N_i - \bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right) - \frac{1}{m(m-1)}\sum_{i \neq i'}^m \bar{\bar{X}}_i \, \bar{\bar{X}}_{i'}\right]$$

$$= E_i\left[\frac{1}{m}\sum_i^m \bar{\bar{X}}_i^2 + \frac{1}{m}\sum_i^m \frac{N_i - \bar{n}}{N_i}\frac{S_i^2}{\bar{n}} - \frac{1}{m(m-1)}\sum_{i \neq i'}^m \bar{\bar{X}}_i \, \bar{\bar{X}}_{i'}\right]$$

$$E_i\left(E_j(D)\right) = E_i\left[\frac{1}{m}\sum_i^m \bar{\bar{X}}_i^2\right] + E_i\left[\frac{1}{m}\sum_i^m \frac{N_i - \bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right] - E_i\left[\frac{1}{m(m-1)}\sum_{i \neq i'}^m \bar{\bar{X}}_i \, \bar{\bar{X}}_{i'}\right] \quad (3.84)$$

- Untuk $E_i\left[\frac{1}{m}\sum_i^m \bar{\bar{X}}_i^2\right]$

$$E_i\left[\frac{1}{m}\sum_i^m \bar{\bar{X}}_i^2\right] = \frac{1}{m} m\left(E_i\left(\bar{\bar{X}}_i^2\right)\right) = E_i\left(\bar{\bar{X}}_i^2\right)$$

$$E_i\left[\frac{1}{m}\sum_i^m \bar{\bar{X}}_i^2\right] = \bar{\bar{X}}^2 + \frac{1}{N}\sum^M N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 \quad (3.85)$$

- Untuk $E_i\left[\frac{1}{m}\sum_i^m \frac{N_i-\bar{n}}{N_i}\frac{S^2}{\bar{n}}\right]$

$$E_i\left[\frac{1}{m}\sum_i^m \frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right] = \frac{1}{m}m\left(E_i\left(\frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right)\right)$$

$$= E_i\left(\frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right)$$

$$E_i\left[\frac{1}{m}\sum_i^m \frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}}\right] = \sum^M \frac{N_i}{N}\frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}} \tag{3.86}$$

- Untuk $E_i\left[\frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{X}}_i\,\bar{\bar{X}}_{i'}\right]$

$$E_i\left[\frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{X}}_i\,\bar{\bar{X}}_{i'}\right] = \frac{1}{m(m-1)}m(m-1)\bar{\bar{X}}^2$$

$$E_i\left[\frac{1}{m(m-1)}\sum_{i\neq i'}^m \bar{\bar{X}}_i\,\bar{\bar{X}}_{i'}\right] = \bar{\bar{X}}^2 \tag{3.87}$$

Berdasarkan uraian tersebut di atas, serta dengan mensubstitusikan persamaan (3.85), persamaan (3.86), dan persamaan (3.87) ke dalam persamaan (3.84), akan diperoleh

$$E_i\left(E_j(D)\right) = \bar{\bar{X}}^2 + \frac{1}{N}\sum^M N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 + \sum^M \frac{N_i}{N}\frac{N_i-\bar{n}}{N_i}\frac{S_i^2}{\bar{n}} - \bar{\bar{X}}^2$$

$$E_i\left(E_j(D)\right) = \frac{1}{N}\sum^M N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 + \frac{1}{N}\sum^M (N_i-\bar{n})\frac{S_i^2}{\bar{n}} \tag{3.88}$$

Langkah terakhir yaitu mensubstitusikan persamaan (3.88) ke dalam persamaan (3.78), maka diperoleh :

$$E\left[\hat{V}\left(\bar{\bar{x}}_{pps}\right)\right] = \frac{1}{m}E_i\left(E_jD\right)$$

$$= \frac{1}{mN}\sum^M N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 + \frac{1}{mN}\sum^M (N_i-\bar{n})\frac{S_i^2}{\bar{n}}$$

$$E\left[\hat{V}\left(\bar{\bar{x}}_{pps}\right)\right] = V(\bar{\bar{x}}_{pps})$$

Terbukti bahwa $\hat{V}\left(\bar{\bar{x}}_{pps}\right)$ merupakan penaksir tak bias dari $V(\bar{\bar{x}}_{pps})$.

3.6 Perbandingan $V\left(\bar{\bar{x}}_{pps}\right)$ dengan $V\left(\hat{\hat{\bar{\bar{X}}}}_{cl}\right)$

Alasan utama untuk memilih *primary sampling units* (psu) dengan metode PPS pada sampling berkelompok adalah untuk menghasilkan sampel yang lebih representatif dari populasi. Pada metode ini, psu atau *cluster* dengan ukuran yang berbeda mempunyai probabilitas yang berbeda pula disesuaikan dengan ukurannya, berbeda dengan metode sampling berkelompok yang tidak memperhatikan ukuran *cluster* oleh karena itu, presisi dari penaksir akan bertambah jika dibandingkan dengan sampling berkelompok. Karena metode PPS memiliki presisi yang lebih tinggi, maka metode ini pun harus memiliki varians yang lebih kecil daripada metode sampling berkelompok.

Seperti telah dikemukakan sebelumnya bahwa perbedaan utama antara sampling berkelompok dan sampling berkelompok dengan PPS yaitu terletak pada saat pemilihan psu. Sampling berkelompok memilih psu dengan menggunakan konsep sampling acak sederhana, sedangkan sampling berkelompok dengan PPS memilih psu dengan menggunakan konsep PPS. Untuk pemilihan *secondary sampling units* (ssu), kedua metode sampling berkelompok tersebut menggunakan konsep yang sama yaitu konsep sampling acak sederhana.

Bedasarkan uraian di atas, perbedaan prosedur dalam proses pemilihan psu tersebut, akan memungkinkan indikasi hasil presisi yang berbeda. Oleh karena itu, dirasa perlu untuk membandingkan $V\left(\hat{\hat{\bar{\bar{X}}}}_{cl}\right)$ dan $V\left(\bar{\bar{x}}_{pps}\right)$ dengan hanya mempertimbangkan psu.

Berdasarkan uraian pada subbab-subbab sebelumnya telah diketahui bahwa varians dari $\bar{\bar{x}}_{pps}$ $\left(V\left(\bar{\bar{x}}_{pps}\right)\right)$, dinyatakan sebagai berikut:

$$V\left(\bar{\bar{x}}_{pps}\right) = \frac{1}{mN}\sum^{M} N_i\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 + \frac{1}{mN}\sum^{M}\left(N_i - \bar{n}\right)\frac{S_i^2}{\bar{n}}$$

sedangkan varians dari $\hat{\hat{\bar{\bar{X}}}}_{cl}$ $\left(V\left(\hat{\hat{\bar{\bar{X}}}}_{cl}\right)\right)$, dinyatakan sebagai berikut:

$$V\left(\hat{\hat{\bar{\bar{X}}}}_{cl}\right) = \frac{1}{N^2}\left(M^2\frac{M-m}{M}\frac{S_b^2}{m} + \frac{M}{m}\sum^{M} N_i^2\frac{N_i - n_i}{N_i}\frac{S_i^2}{n_i}\right)$$

Selanjutnya dengan memisalkan $N_i = n_i$ dan mengeliminasi pengaruh ssu pada varians, maka perumusan $\left(V(\bar{\bar{x}}_{pps})\right)$, akan menjadi:

$$V(\bar{\bar{x}}_{pps}) = \frac{1}{mN}\sum^{M} N_i \left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 \quad (3.89)$$

atau dapat dituliskan dalam bentuk persamaan berikut ini :

$$V(\bar{\bar{x}}_{pps}) = \frac{1}{m}\sum^{M} \frac{N_i}{N}\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2 \quad (3.90)$$

Sedangkan perumusan $\left(V\left(\hat{\bar{\bar{X}}}_{cl}\right)\right)$, akan menjadi:

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) = \frac{M^2}{N^2}\frac{M-m}{M}\frac{S_b^2}{m}$$

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) = \frac{1}{\bar{N}^2}\frac{M-m}{M}\frac{1}{m}\frac{1}{M-1}\sum^{M}(X_i - \bar{X})^2 \quad (3.91)$$

dengan $S_b^2 = \frac{1}{M-1}\sum^{M}(X_i - \bar{X})^2$ dan $\bar{N}^2 = \frac{N^2}{M^2}$. Selanjutnya misalkan $\frac{(M-m)}{M} = 1$ dan $(M-1) = M$, maka persamaan (3.90) menjadi:

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) = \frac{1}{m}\frac{1}{M}\sum^{M}\left(\frac{X_i}{\bar{N}} - \bar{\bar{X}}\right)^2$$

$$= \frac{1}{mM}\sum^{M}\left(\frac{X_i - \bar{N}\bar{\bar{X}}}{\bar{N}}\right)^2$$

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) = \frac{1}{mM}\sum^{M}\frac{\left(N_i\bar{\bar{X}}_i - \bar{N}\bar{\bar{X}}\right)^2}{\bar{N}^2} \quad (3.92)$$

Selanjutnya adalah membandingkan $V(\bar{\bar{x}}_{pps})$ dengan $V\left(\hat{\bar{\bar{X}}}_{cl}\right)$ dengan menggunakan persamaan (3.90) dan persamaan (3.92). Proses perbandingan $V(\bar{\bar{x}}_{pps})$ dan $V\left(\hat{\bar{\bar{X}}}_{cl}\right)$ dilakukan dengan cara menggunakan operasi pengurangan antara dua varians tersebut, akan diperoleh:

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V(\bar{\bar{x}}_{pps}) = \frac{1}{mM}\sum^{M}\frac{\left(N_i\bar{\bar{X}}_i - \bar{N}\bar{\bar{X}}\right)^2}{\bar{N}^2} - \frac{1}{m}\sum^{M}\frac{N_i}{N}\left(\bar{\bar{X}}_i - \bar{\bar{X}}\right)^2$$

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V(\bar{\bar{x}}_{pps}) = \frac{1}{m}\frac{\sum^{M}(N_i - \bar{N})^2}{N\bar{N}}\bar{\bar{X}}_i^2 + \frac{1}{m}\frac{1}{N}\sum^{M}(N_i - \bar{N})\left(\bar{\bar{X}}_i^2 - \bar{\bar{X}}^2\right) \quad (3.93)$$

Hal pertama yang harus diperhatikan dari persamaan (3.93) yaitu kondisi pada saat $N_i = \bar{N} = \frac{N}{M}$, maka diperoleh:

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V(\bar{\bar{x}}_{pps}) = 0 + 0$$

$$V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V(\bar{\bar{x}}_{pps}) = 0$$

Karena $V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V\left(\bar{\bar{x}}_{pps}\right) = 0$, hal ini berarti kedua prosedur tersebut mempunyai presisi yang sama. Hal ini akan dengan mudah dipahami dengan memperhatikan bahwa

$$\frac{N_i}{N} = \frac{\bar{N}}{N} = \frac{\frac{N}{M}}{N} = \frac{1}{M}$$

Berdasarkan pernyataan di atas, ini memperlihatkan bahwa probabilitas dari pemilihan psu pada sampling berkelompok dengan PPS adalah $\frac{1}{M}$ sama hal nya dengan memilih psu pada sampling berkelompok. Oleh karena itu, diperoleh perkiraan bahwa $V\left(\hat{\bar{\bar{X}}}_{cl}\right) = V\left(\bar{\bar{x}}_{pps}\right)$.

Pada saat N_i bervariasi dan pada saat $V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V\left(\bar{\bar{x}}_{pps}\right) > 0$, presisi dari sampling berkelompok dengan PPS lebih baik daripada sampling berkelompok. Untuk membuktikannya, perhatikan kembali persamaan (3.93), diketahui bahwa ruas kanan dari persamaan (3.93) dapat dibagi menjadi dua komponen yaitu, komponen I ruas kanan dan komponen II ruas kanan. selanjutnya perhatikan bahwa komponen I ruas kanan pada persamaan (3.93) selalu bernilai positif, dan dinyatakan sebagai berikut:

$$\frac{1}{m}\frac{\sum^{M}(N_i - \bar{N})^2}{N\bar{N}}\bar{\bar{X}}_i^2 > 0$$

Sementara itu untuk komponen II ruas kanan pada persamaan (3.93), yaitu:

$$\frac{1}{m}\frac{1}{N}\sum^{M}(N_i - \bar{N})\left(\bar{\bar{X}}_i^2 - \bar{\bar{X}}^2\right)$$

terdapat beberapa poin yang harus diperhatikan, yaitu:

- Pada dasarnya metode sampling berkelompok, biasanya mengharapkan agar varians antar rerata kelompok (*cluster*) tetap kecil. Agar hal tersebut terpenuhi, maka harus mereduksi $V\left(\hat{\bar{\bar{X}}}_{cl}\right)$ atau $V\left(\bar{\bar{x}}_{pps}\right)$. Sedangkan metode sampling berkelompok dengan pps, mengharapkan bahwa nilai absolut dari $\left(\bar{\bar{X}}_i^2 - \bar{\bar{X}}^2\right)$ akan relatif kecil dibandingkan dengan $\bar{\bar{X}}_i^2$ atau $\bar{\bar{X}}^2$.

- $N_i - \overline{N} < 0$ hanya ketika ukuran kelompok (*cluster*) N_i kecil.

Dengan demikian, diperoleh kesimpulan bahwa $V\left(\hat{\bar{\bar{X}}}_{cl}\right) - V\left(\bar{\bar{x}}_{pps}\right) > 0$ pada kondisi umum ketika sampling berkelompok dengan PPS dipergunakan. Hal itulah yang menjadi alasan mengapa presisi dari sampling berkelompok dengan PPS akan lebih baik daripada sampling berkelompok.