BAB III

STRATIFIED CLUSTER SAMPLING

3.1 Pengertian *Stratified Cluster Sampling*

Proses memprediksi hasil *quick count* sangat dipengaruhi oleh pemilihan sampel yang dilakukan dengan metode sampling tertentu. Sampel yang baik adalah sampel yang dapat mewakili karakteristik seluruh populasi. Ketika populasi bersifat heterogen dan sangat besar, akan sulit mengambil sampel secara acak dari populasi yang heterogen, hal tersebut disebabkan oleh sampel yang diambil secara acak belum tentu mewakili setiap bagian yang heterogen dari populasi tersebut. Sedangkan ketika populasi bersifat homogen, maka sampel yang diambil secara acak dari setiap anggota populasi dapat mewakili karakteristik populasi dengan baik. Selain itu, populasi yang besar akan menyulitkan dalam membuat daftar data populasi, sehingga membutuhkan waktu dan biaya yang cukup besar. Salah satu metode sampling yang dapat digunakan untuk menghasilkan sampel yang baik dari populasi yang besar tersebut adalah metode *stratified cluster sampling*.

Yamane (1967) menyatakan *"Stratified cluster sampling combines the characteristics of stratified sampling and cluster sampling. It breaks down the population into strata which are internally homogeneous, and therefore heterogeneous among one another, and clusters are selected from each stratum"*. Berdasarkan kutipan di atas, diketahui bahwa *stratified cluster sampling* merupakan proses pengambilan sampel yang menggabungkan karakteristik dari *stratified random sampling* dengan karakteristik *simple cluster sampling*. Pada *stratified cluster sampling*, populasi dikelompokkan ke dalam strata yang homogen didalamnya sehingga kelompok itu akan heterogen dengan kelompok lainnya dan proses selanjutnya yaitu pemilihan *cluster* dari tiap stratum. Proses pengelompokkan populasi ke dalam stratum bertujuan agar sampel yang diambil dari setiap stratum dapat merepresentasikan karakteristik populasi dengan baik.

Oleh karena itu, stratum harus dibentuk sehomogen mungkin dengan menganalisis karakteristik populasi dengan baik. Proses selanjutnya yaitu populasi pada masing-masing strata akan dikelompokkan ke dalam beberapa *cluster*. Proses ini bertujuan untuk mempermudah pengelompokkan populasi sehingga dapat mengefisiensikan waktu dan biaya yang ada. Ketika variasi yang besar terjadi pada antar stratum, pengambilan sampel di *stratified cluster sampling* menjadi lebih efisien. Oleh karena itu, keuntungan sampling dengan menggunakan metode *stratified cluster sampling* ini adalah sampling dengan metode ini akan memiliki variansi lebih kecil daripada *simple cluster sampling*.

Terdapat tahapan-tahapan yang harus dilakukan dalam pengambilan sampel dengan menggunakan metode *stratified cluster sampling,* yaitu sebagai berikut:

1. Tahap pertama yaitu populasi yang berukuran N dibagi ke dalam beberapa stratum (sub populasi), dimana setiap stratum bersifat homogen (memiliki kriteria yang sama) dan masing-masing strata terdiri atas $N_I, N_2, N_3, \dots, N_L$ elemen. Diantara dua stratum (sub populasi) tidak boleh ada yang saling tumpang tindih sehingga $N_1 + N_2 + N_3 + \cdots + N_L = N$. Setiap stratum dapat dipandang sebagai populasi tersendiri (sub populasi). Pada proses pembentukan stratum harus diperhatikan variabel apa yang akan dijadikan sebagai dasar pembentukan stratum, yaitu variabel yang memiliki korelasi tinggi dengan variabel yang diteliti.
2. Tahap kedua yaitu membagi populasi ke dalam M_h kelompok secara acak, hal ini berarti tidak ada kriteria tertentu yang mensyaratkan pembentukan suatu kelompok. M_h kelompok ini dinamakan *primary sampling units* (*psu*) atau unit sampling utama (usu).
3. Berdasarkan kelompok usu tersebut, tahapan ketiga yaitu memilih secara acak m_h kelompok yang akan dijadikan sampel. m_h kelompok sampel ini masing-masing berukuran N_{hi}. Selanjutnya m_h kelompok ini disebut *secondary sampling units* (*ssu*) atau unit sampling kedua (usk).

4. Selanjutnya tahap keempat adalah memilih secara acak n_{hi} buah dari masing-masing usk tersebut yang dinamakan kelompok *ultimate* (utama).
5. Pada tahap kelima, setelah memperoleh sampel, selanjutnya melakukan penaksiran terhadap parameter yang diperlukan dan membuat kesimpulan untuk populasi serta variansnya berdasarkan hasil penaksiran sampel.

3.2 Pengertian Total Populasi

Pada sebuah survei selain populasi, sampel menjadi sesuatu yang sangat penting. Oleh karena itu, hal yang dilakukan pada saat melakukan suatu survei adalah menentukan sifat-sifat, mengukur dan mencatat setiap unit dalam sampel. Sifat-sifat dari setiap unit dalam sampel ini dinamakan karakteristik populasi. Penarikan sampel mempunyai banyak tujuan, namun terdapat empat karakteristik populasi yang lebih sering digunakan (Yamane, 1967) yaitu:

1. Rata-rata populasi

   Rata-rata populasi adalah nilai rata-rata dari data populasi (Azhar, 2011). Rata-rata populasi dinotasikan dengan $\bar{Y}$, dan didefinisikan sebagai berikut:

$$\bar{Y} = \frac{Y_1+Y_2+\cdots+Y_N}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{Y}{N} \tag{3.1}$$

Sedangkan rata-rata sampel didefinisikan sebagai berikut:

$$\bar{y} = \frac{y_1+y_2+\cdots+y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{y}{n} \tag{3.2}$$

Penaksir dari rata-rata populasi dinotasikan dengan $\hat{\bar{Y}}$, dan penaksir tak bias dari rata-rata populasi adalah rata-rata sampel, dinyatakan sebagai berikut:

$$\hat{\bar{Y}} = \bar{y} \tag{3.3}$$

Pembuktian:

$$E(\bar{y}) = E\left[\frac{1}{n}(y_1 + y_2 + \cdots + y_n)\right]$$

$$= \frac{1}{n}[E(y_1) + E(y_2) + \cdots + E(y_n)]$$

$$= \frac{1}{n}(n\bar{Y}) = \bar{Y}$$

$$E(\bar{y}) = \bar{Y}$$

2. Jumlah populasi atau total populasi

   Menurut Hidayat (2013), total populasi adalah jumlah keseluruhan dari satuan-satuan atau individu-individu yang karakteristiknya hendak diteliti. Total populasi dinotasikan dengan Y, dan didefinisikan sebagai berikut:

$$Y = \sum_{i=1}^{N} Y_i = Y_1 + Y_2 + \cdots + Y_N \tag{3.4}$$

   atau berdasarkan persamaan (3.1) diperoleh:

$$Y = N\bar{Y} \tag{3.5}$$

   Sedangkan total sampel didefinisikan sebagai berikut:

$$y = \sum_{i=1}^{n} y_i = y_1 + y_2 + \cdots + y_n \tag{3.6}$$

   atau berdasarkan persamaan (3.2) diperoleh:

$$y = n\bar{y} \tag{3.7}$$

   Penaksir dari total populasi dinotasikan dengan $\hat{Y}$. Berdasarkan persamaan (3.3), diperoleh informasi bahwa penaksir tak bias untuk total populasi adalah total sampel, dinyatakan sebagai berikut:

$$\hat{Y} = N\bar{y} = N\frac{\sum_{i=1}^{n} y_i}{n} \tag{3.8}$$

   Pembuktian:

$$E(\hat{Y}) = E(N\bar{y})$$
$$= N[E(\bar{y})]$$
$$= N\bar{Y}$$
$$E(\hat{Y}) = Y$$

3. Rasio dari dua jumlah populasi atau dua rata-rata populasi

   Menurut Wibisaputro (2015), rasio adalah perbandingan antara pembilang (numerator) dan penyebut (denominator) yang saling terpisah dan tidak ada hubungannya. Rasio populasi dinotasikan dengan R dan didefinisikan sebagai berikut:

$$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}} \tag{3.9}$$

Penaksir dari rasio populasi dinotasikan dengan $\hat{R}$, dengan perumusan sebagai berikut:

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \quad (3.10)$$

Pembuktian:

$$E(\hat{R}) = E\left(\frac{\bar{y}}{\bar{x}}\right)$$

$$= \frac{E\ (\bar{y})}{E\ (\bar{x})}$$

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

$$E(\hat{R}) = R$$

4. Proporsi dari unit-unit sampel yang masuk dalam beberapa kelas tertentu

Menurut Wibisaputro (2015), proporsi adalah bentuk pecahan yang pembilangnya merupakan bagian dari penyebutnya. Proporsi digunakan untuk melihat komposisi suatu variabel dalam populasi. Bentuk proporsi ini sering dinyatakan dalam persen, yaitu dengan mengalikan pecahan proporsi dengan 100%. Proporsi tidak mempunyai satuan (dimensi), karena satuan dari pembilang dan penyebutnya sama, sehingga saling meniadakan. Perumusan proporsi adalah sebagai berikut:

$$Proporsi = \frac{X}{X+Y}.100\% \quad (3.11)$$

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

Perhatikan bahwa huruf-huruf besar biasanya menunjukkan karakteristik populasi, sedangkan karakteristik sampel biasanya diberi simbol huruf-huruf kecil.

Karakteristik populasi yang digunakan pada skripsi ini adalah total populasi. Alasan penggunaan karakterisitik total populasi, yaitu karena tujuan dari skripsi ini adalah untuk memperoleh total suara dari populasi. Selain itu, penggunaan total populasi ini diharapkan akan lebih mewakili fakta yang ada (Notoatmodjo, 2002).

Pada *stratified cluster sampling*, total populasi didefinisikan sebagai berikut:

$$X = \left(\sum_{h=1}^{L} \sum_{i=1}^{M_h} X_{hi}\right) \quad (3.12)$$

Sedangkan rata-rata populasi didefinisikan sebagai berikut:

$$\bar{X} = \frac{X}{LM_h} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{M_h} X_{hi}}{LM_h} \quad (3.13)$$

3.3 Penaksir Total Populasi *Stratified Cluster Sampling*

Sampel berkelompok tiga tahap (*three-stage cluster sampling*) adalah teknik pengambilan sampel yang dilakukan dalam 3 tahap. Tahap pertama adalah membagi populasi ke dalam beberapa kelompok *(cluster)* misalkan terdapat *L* (psu), kemudian dari *L* psu tersebut diasumsikan terpilih sebanyak *l* sampel acak dari psu. Tahap kedua, dari *i* (indeks sampel acak (psu)) masing-masing mempunyai M_i kelompok (ssu), kemudian asumsikan $\bar{m}$ dipilih dari setiap sampel acak (psu). Terakhir asumsikan terdapat N_{ij} (tsu) dalam *j* (indeks sampel kelompok (ssu)) dari *i* (indeks sampel acak (psu)) dan subsampel n_{ij} dipilih dari *j* sampel kelompok (ssu).

Sedangkan pada *stratified cluster sampling L* adalah strata, bukan kelompok *(cluster)* dengan *h* sebagai indeks dari strata *L*. Selain itu, pada *stratified cluster sampling* $L = l$ artinya bahwa pada *stratified cluster sampling* seluruh strata (*L*) yang berada dalam populasi akan dijadikan sampel penelitian. Oleh karena itu, penaksir dari total populasi untuk *stratified cluster sampling* diperoleh dari keadaan $L = l$ yang ditaksir dari total populasi *X* untuk *three-stage cluster sampling*. Penaksir total populasi untuk *three-stage cluster sampling* adalah sebagai berikut:

$$\begin{aligned} \hat{X} &= L\hat{\hat{X}} \\ &= \frac{L}{l} \sum_{i=1}^{l} \hat{X}_i \\ &= \frac{L}{l} \sum_{i=1}^{l} M_i \hat{\hat{\hat{X}}}_i \end{aligned}$$

$$= \frac{L}{l}\sum_{i=1}^{l}\frac{M_i}{\bar{m}}\sum_{j=1}^{\bar{m}}\hat{X}_{ij}$$

$$= \frac{L}{l}\sum_{i=1}^{l}\frac{M_i}{\bar{m}}\sum_{j=1}^{\bar{m}}N_{ij}\bar{x}_{ij}$$

$$\hat{X} = \frac{L}{l}\sum_{i=1}^{l}\frac{M_i}{\bar{m}}\sum_{j=1}^{\bar{m}}\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}}x_{ijk} \quad (3.14)$$

dimana *L* menyatakan *cluster*. Pada pembahasan sebelumnya telah dikemukakan, berbeda dengan *three-stage cluster sampling* bahwa pada *stratified cluster sampling L* menyatakan strata menggantikan *cluster* dan keadaan $L = l$ dipenuhi, maka dengan mengganti indeks *i* menjadi indeks *h* untuk mengindikasikan sebagai strata akan diperoleh penaksir tak bias dari total populasi untuk *stratified cluster sampling* yang diturunkan dari persamaan (3.14), diperoleh:

$$\hat{X} = \frac{L}{l}\sum_{\text{h}=1}^{l}\frac{M_h}{m_h}\sum_{\text{i}=1}^{m_h}\frac{N_{hi}}{n_{hi}}\sum_{\text{j}=1}^{n_{hi}}x_{hij}$$

$$= \frac{L}{L}\sum_{\text{h}=1}^{L}\frac{M_h}{m_h}\sum_{\text{i}=1}^{m_h}\frac{N_{hi}}{n_{hi}}\sum_{\text{j}=1}^{n_{hi}}x_{hij}$$

$$\hat{X} = \sum_{\text{h}=1}^{L}\frac{M_h}{m_h}\sum_{\text{i}=1}^{m_h}\frac{N_{hi}}{n_{hi}}\sum_{\text{j}=1}^{n_{hi}}x_{hij} \quad (3.15)$$

Persamaan $\left(\frac{N_{hi}}{n_{hi}}\right)\sum x_{hij}$ adalah penaksir dari total populasi untuk *cluster* ke-*i* di stratum ke-*h*. Oleh karena itu, $A = \sum\left(\frac{N_{hi}}{n_{hi}}\right)\sum x_{hij}$ adalah penaksir total populasi untuk sampel m_h *cluster* di stratum *h*. Persamaan $B = \left(\frac{M_h}{m_h}\right)A$ adalah penaksir total populasi dari stratum ke-*h*. Oleh karena itu $\sum B$ adalah penaksir total populasi untuk semua *L* strata.

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:

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

Dengan kata lain, penaksir total populasi ($\hat{X}$) merupakan penaksir yang tak bias untuk total populasi.

Pembuktian:

Ekspektasi dari $(\hat{X})$ harus dipandang dalam tiga tahapan yaitu ekspektasi yang berkaitan dengan tahapan pertama sampling, tahapan kedua sampling, dan ekspektasi bersyarat yang berkaitan dengan tahapan ketiga sampling, dengan menganggap tahapan pertama dan tahapan kedua konstan. E_j merupakan ekspektasi bersyarat sepanjang *j* dan menganggap tahapan pertama dan tahapan kedua konstan.

$$E(\hat{X}) = E_h E_i E_j\ (\hat{X})$$

$$= E_h E_i E_j\ \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{\mathrm{j}=1}^{n_{hi}} x_{hij}\right)$$

$$E(\hat{X}) = E_h E_i \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{\mathrm{j}=1}^{n_{hi}} E_j\ (x_{hij})\right) \qquad (3.16)$$

Pada metode *simple cluster sampling*, diberikan *i* sebagai indeks pada psu dan selanjutnya dari setiap psu tersebut dilakukan pemilihan sampel acak sebanyak n_i, sehingga diperoleh $E_j\left(x_{ij}\right) = \bar{\bar{X}}_{i}$. Hal yang sama juga terdapat pada metode *stratified cluster sampling*, karena diberikan *h* sebagai indeks pada strata, *i* sebagai indeks pada psu dan selanjutnya dari setiap psu tersebut dilakukan pemilihan sampel acak sebanyak n_{hi}, sehingga diperoleh $E_j\left(x_{hij}\right) = \bar{\bar{X}}_{hi}$.

$$E(\hat{X}) = E_h E_i \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{\mathrm{j}=1}^{n_{hi}} \bar{\bar{X}}_{hi}\right)$$

$$= E_h E_i \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} \frac{N_{hi}}{n_{hi}} (n_{hi}.\bar{\bar{X}}_{hi})\right)$$

$$= E_h E_i \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} N_{hi}.\bar{\bar{X}}_{hi}\right)$$

$$= E_h E_i \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} X_{hi}\right)$$

$$= E_h \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} E_i(X_{hi})\right)$$

$$= E_h \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h} \sum_{\mathrm{i}=1}^{m_h} \sum_{i=1}^{M_h} \frac{1}{M_h}\ X_{hi}\right)$$

$$= E_h \left(\sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h}\ m_h\ \frac{1}{M_h}\ \sum_{i=1}^{M_h} X_{hi}\right)$$

$$= E_h\left(\sum_{\mathrm{h}=1}^{L}\ \sum_{i=1}^{M_h} X_{hi}\ \right)$$

$$= \left(\sum_{\mathrm{h}=1}^{L}\ \sum_{i=1}^{M_h} E_h(X_{hi})\ \right)$$

$$= \left(\sum_{\mathrm{h}=1}^{L}\ \sum_{i=1}^{M_h}\ \sum_{h=1}^{L} \frac{1}{L} X_{hi}\ \right)$$

$$= \left(\sum_{\mathrm{h}=1}^{L} \frac{1}{L}\ \sum_{i=1}^{M_h} \sum_{h=1}^{L} X_{hi}\ \right)$$

$$= \left(L.\frac{1}{L}\sum_{i=1}^{M_h}\sum_{h=1}^{L} X_{hi}\ \right)$$

$$E(\hat{X}) = \left(\sum_{h=1}^{L}\sum_{i=1}^{M_h} X_{hi}\ \right) = X$$

Terbukti bahwa $E(\hat{X}) = X$, dengan kata lain $(\hat{X})$ merupakan penaksir yang tak bias untuk total populasi (X).

3.4 Variansi dari Penaksir Total Populasi dan Penaksirnya

3.4.1 Variansi dari Penaksir Total Populasi

Setelah memperoleh taksiran dari total populasi, langkah selanjutnya adalah menentukan variansi dari $\hat{X}$.

Varians dari penaksir tak bias $\hat{X}$ untuk *three-stage cluster sampling* diperoleh dengan menggabungkan dua varians *two-stage cluster sampling*.

Varians dari $\hat{X}$ untuk kasus *two-stage cluster sampling* adalah:

$$V(\hat{X}) = M^2\,\frac{M-m}{M}\frac{S_b^2}{m} + \frac{M}{m}\sum^{M} N_i^2\,\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i} \qquad (3.17)$$

dimana

$$S_b^2 = \frac{1}{M-1}\sum_i^M (X_i - \bar{X})^2$$

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

Dengan mensubstitusikan persamaan (3.17) ke psu dan ssu pada kasus *three-stage cluster sampling*, diperoleh:

$$L^2\,\frac{L-l}{L}\frac{S_b^2}{l} + \frac{L}{l}\sum_i^L M_i^2\,\frac{(M_i-\bar{m})}{M_i}\ \frac{S_i^2}{\bar{m}} \qquad (3.18)$$

dimana

$$S_b^2 = \frac{1}{L-1}\sum^{L}\,(X_i - \bar{X})^2$$

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

Selanjutnya, dengan mensubstitusikan persamaan (3.17) ke ssu dan tsu pada kasus *three-stage cluster sampling*, diperoleh:

$$M_i{}^2\,\frac{M_i-\bar{m}}{M_i}\frac{S_i^2}{\bar{m}}+\frac{M_i}{\bar{m}}\sum_j^{M_i}N_{ij}^2\,\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{S_{ij}^2}{n_{ij}} \qquad (3.19)$$

dimana

$$S_{ij}^2=\frac{1}{N_{ij}-1}\sum_k^{N_0}\left(X_{ijk}-\bar{\bar{X}}_{ij}\right)^2$$

Perumusan untuk $V(\hat{X})$ pada kasus *three-stage cluster sampling* diperoleh dengan menggabungkan persamaan (3.18) dan persamaan (3.19), diperoleh:

$$\begin{aligned}V\left(\hat{X}\right)&=L^2\,\frac{L-l}{L}\frac{S_b^2}{l}+\frac{L}{l}\sum^L\left(M_i^2\,\frac{M_i-\bar{m}}{M_i}\,\frac{S_i^2}{\bar{m}}+\frac{M_i}{\bar{m}}\sum_j^{M_i}N_{ij}^2\,\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{S_{ij}^2}{n_{ij}}\right)\\&=L^2\,\frac{L-l}{L}\frac{S_b^2}{l}+\frac{L}{l}\sum^L\,M_i^2\,\frac{M_i-\bar{m}}{M_i}\,\frac{S_i^2}{\bar{m}}\,+\frac{L}{l}\sum^L\,\frac{M_i}{\bar{m}}\sum_j^{M_i}N_{ij}^2\,\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{S_{ij}^2}{n_{ij}}\end{aligned}$$

Oleh karena itu, variansi dari $\hat{X}$ untuk *three-stage cluster sampling* adalah sebagai berikut:

$$V\left(\hat{X}\right)=L^2\,\frac{L-l}{L}\frac{S_b^2}{l}+\frac{L}{l}\sum_{\mathrm{i}=1}^{L}M_i^2\,\frac{M_i-\bar{m}}{M_i}\frac{S_i^2}{\bar{m}}+\frac{L}{l}\sum_{\mathrm{i}=1}^{L}\frac{M_i}{\bar{m}}\sum_{\mathrm{j}=1}^{M_i}N_{ij}^2\,\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{S_{ij}^2}{n_{ij}} \qquad (3.20)$$

Dengan memisalkan $L=l$ dan mengganti indeks *i* menjadi indeks *h* untuk mengindikasikan sebagai strata, maka akan diperoleh variansi dari $\hat{X}$ untuk *stratified cluster sampling* yaitu sebagai berikut :

$$V\left(\hat{X}\right)=L^2\,\frac{L-L}{L}\frac{S_b^2}{L}+\frac{L}{L}\sum_{\mathrm{h}=1}^{L}M_h^2\,\frac{M_h-\bar{m}}{M_h}\,\frac{S_h^2}{\bar{m}}+\frac{L}{L}\sum_{\mathrm{h}=1}^{L}\frac{M_h}{\bar{m}}\sum_{\mathrm{i}=1}^{M_h}N_{hi}^2\,\frac{N_{hi}-n_{hi}}{N_{hi}}\,\frac{S_{hi}^2}{n_{hi}} \qquad (3.21)$$

$$V\left(\hat{X}\right)=\sum_{\mathrm{h}=1}^{L}M_h^2\,\frac{M_h-\bar{m}}{M_h}\,\frac{S_h^2}{\bar{m}}+\sum_{\mathrm{h}=1}^{L}\frac{M_h}{\bar{m}}\sum_{\mathrm{i}=1}^{M_h}N_{hi}^2\,\frac{N_{hi}-n_{hi}}{N_{hi}}\,\frac{S_{hi}^2}{n_{hi}}$$

Seperti yang telah diperlihatkan, S_b^2, variansi antar *cluster* (dimana dalam kasus ini menjadi strata) dikeluarkan dari persamaan (3.21).

$$V\left(\hat{X}\right)=\sum_{\mathrm{h}=1}^{L}M_h^2\frac{M_h-m_h}{M_h}\frac{S_h^2}{m_h}+\sum_{\mathrm{h}=1}^{L}\frac{M_h}{m_h}\sum_{\mathrm{i}=1}^{M_h}N_{hi}^2\frac{N_{hi}-n_{hi}}{N_{hi}}\frac{S_{hi}^2}{n_{hi}} \qquad (3.22)$$

$$S_h^2=\frac{1}{M_h-1}\sum_{i=1}^{M_h}(X_{hi}-\bar{X}_h)^2 \qquad (3.23)$$

$$S_{hi}^2 = \frac{1}{N_{hi}-1}\sum_{i=1}^{N_{hi}}(X_{hij} - \bar{\bar{X}}_{hi})^2 \qquad (3.24)$$

Jika $m_h = \bar{m}$ dan $n_{hi} = \bar{n}$, persamaan (3.22) menunjukkan bahwa ketika diberikan $L\bar{m}\bar{n} = n$, $V(\hat{X})$ direduksi dengan menurunkan $\bar{n}$ dan menaikkan $\bar{m}$.

Besarnya $\bar{n}$ biasanya sekitar 5 – 15 (Yamane, 1967), sedangkan $\bar{m}$ mungkin sangat kecil atau sangat besar, bergantung pada permasalahnya (Yamane, 1967).

3.4.2 Penaksir Variansi dari Penaksir Total Populasi

Pada populasi berukuran besar, sulit untuk menentukan nilai dari V($\hat{X}$) secara langsung sehingga dapat menggunakan penaksirnya. Penaksir variansi dari $\hat{X}$ untuk *three-stage cluster sampling* adalah:

$$\hat{V}(\hat{X}) = L^2\frac{L-l}{L}\frac{s_b^2}{l} + \frac{L}{l}\sum_{i=1}^{l} M_i^2\frac{M_i-\bar{m}}{M_i}\frac{s_i^2}{\bar{m}} + \frac{L}{l}\sum_{i=1}^{l}\frac{M_i}{\bar{m}}\sum_{j=1}^{\bar{m}} N_{ij}^2\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{s_{ij}^2}{n_{ij}}$$

Dengan memisalkan $L = l$ dan mengganti indeks *i* menjadi indeks *h* untuk mengindikasikan sebagai strata, maka akan diperoleh penaksir variansi dari $\hat{X}$ untuk *stratified cluster sampling* yaitu sebagai berikut :

$$\hat{V}(\hat{X}) = L^2\frac{L-L}{L}\frac{s_b^2}{L} + \frac{L}{L}\sum_{h=1}^{L} M_h^2\frac{M_h-\bar{m}}{M_h}\frac{s_h^2}{\bar{m}} + \frac{L}{L}\sum_{h=1}^{L}\frac{M_h}{\bar{m}}\sum_{i=1}^{\bar{m}} N_{hi}^2\frac{N_{hi}-n_{hi}}{N_{hi}}\frac{s_{hi}^2}{n_{hi}}$$

$$\hat{V}(\hat{X}) = \sum_{h=1}^{L} M_h^2\frac{M_h-\bar{m}}{M_h}\frac{s_h^2}{\bar{m}} + \sum_{h=1}^{L}\frac{M_h}{\bar{m}}\sum_{i=1}^{\bar{m}} N_{hi}^2\frac{N_{hi}-n_{hi}}{N_{hi}}\frac{s_{hi}^2}{n_{hi}}$$

Jika $m_h = \bar{m}$ dan $n_{hi} = \bar{n}$ seperti yang telah dilakukan di atas, dapat dilihat bahwa $\hat{V}(\hat{X})$ dipengaruhi terutama oleh s_h^2.

$$\hat{V}(\hat{X}) = \sum_{h=1}^{L} M_h^2\frac{M_h-m_h}{M_h}\frac{s_h^2}{m_h} + \sum_{h=1}^{L}\frac{M_h}{m_h}\sum_{i=1}^{m_h} N_{hi}^2\frac{N_{hi}-n_{hi}}{N_{hi}}\frac{s_{hi}^2}{n_{hi}} \qquad (3.25)$$

$$s_h^2 = \frac{1}{m_h-1}\sum_{i=1}^{m_h}(\hat{X}_{hi} - \hat{\hat{X}}_h)^2$$

$$s_{hi}^2 = \frac{1}{n_{hi}-1}\sum_{j=1}^{n_{hi}}(x_{hij} - \bar{\bar{x}}_{hi})^2$$

$$\hat{X}_{hi} = \frac{N_{hi}}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}} x_{hij} = \frac{N_{hi}}{n_{hi}} x_{hi} = N_{hi}\bar{\bar{x}}_{hi}$$

$$\hat{\hat{X}}_h = \frac{1}{m_h}\sum_{\mathrm{i=1}}^{m_h}\hat{X}_{hi}$$

$$\bar{\bar{x}}_{hi} = \frac{1}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}} x_{hij}$$

x_{hij} adalah suara pemilu di TPS ke- j dari kelompok ke-i di stratum ke –h. Huruf x ditulis dengan huruf kecil, hal ini menandakan nilai (suara pemilu) berasal dari sampel.

$\hat{X}_{hi} = N_{hi}\bar{\bar{x}}_{hi}$ merupakan penaksir jumlah total dari kelompok ke-i di stratum ke-h, $\bar{\bar{x}}_{hi} = \frac{1}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}} x_{hij}$ merupakan rata-rata sampel dari subsampel n_{hi}, dan $\hat{\hat{X}}_h = \frac{1}{m_h}\sum_{\mathrm{i=1}}^{m_h}\hat{X}_{hi}$ merupakan rata-rata sampel dari $\hat{X}_{hi}, i = 1,2,..,m_h$.

Pembuktian:

- $E(\hat{X}_{hi}) = E\left(\frac{N_{hi}}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}} x_{hij}\right)$

$$= \frac{N_{hi}}{n_{hi}} E\left(\sum_{\mathrm{j=1}}^{n_{hi}} x_{hij}\right)$$

$$= \frac{N_{hi}}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}} E(x_{hij})$$

$$= \frac{N_{hi}}{n_{hi}}\sum_{\mathrm{j=1}}^{n_{hi}}\bar{\bar{X}}_{hi}$$

$$= \frac{N_{hi}}{n_{hi}}(n_{hi}.\bar{\bar{X}}_{hi})$$

$$= N_{hi}.\bar{\bar{X}}_{hi}$$

$$= X_{hi}$$

- $E(\bar{\bar{x}}_{hi}) = E\left(\frac{1}{n_{hi}}\sum_{\mathrm{j}=1}^{n_{hi}} x_{hij}\right)$

$$= \frac{1}{n_{hi}} E\left(\sum_{\mathrm{j}=1}^{n_{hi}} x_{hij}\right)$$

$$= \frac{1}{n_{hi}} \sum_{\mathrm{j}=1}^{n_{hi}} E(x_{hij})$$

$$= \frac{1}{n_{hi}} \sum_{\mathrm{j}=1}^{n_{hi}} \bar{\bar{X}}_{hi}$$

$$= \frac{1}{n_{hi}} (\, n_{hi}.\bar{\bar{X}}_{hi})$$

$$= \bar{\bar{X}}_{hi}$$

- $E\left(\hat{\bar{X}}_h\right) = E\left(\frac{1}{m_h}\sum_{\mathrm{i}=1}^{m_h} \hat{X}_{hi}\right)$

$$= \frac{1}{m_h} E\left(\sum_{\mathrm{i}=1}^{m_h} \hat{X}_{hi}\right)$$

$$= \frac{1}{m_h} \sum_{\mathrm{i}=1}^{m_h} E(\hat{X}_{hi})$$

$$= \frac{1}{m_h} \sum_{\mathrm{i}=1}^{m_h} X_{hi}$$

$$= \bar{X}_h$$

s_h^2 menunjukkan penaksir variansi di antara psu (kelompok) di dalam strata ke-*h*. Karena m_h adalah sampel acak dari M_h, $\hat{X}_{hi}$ merupakan penaksir jumlah total dari kelompok ke-*i* di stratum ke-*h*, dan $\hat{\bar{X}}_h$ merupakan rata-rata sampel dari $\hat{X}_{hi}$. Diketahui pula bahwa s_h^2 adalah penaksir tak bias dari S_h^2, sehingga

$E(s_h^2) = E\left(\frac{1}{m_h-1}\sum_{\mathrm{i}=1}^{m_h}(\hat{X}_{hi} - \hat{\bar{X}}_h)^2\right)$

$E(s_h^2) = \frac{1}{M_h - 1} \sum_{i=1}^{M_h} (X_{hi} - \bar{X}_h)^2$

$E(s_h^2) = S_h^2$

s_{hi}^2 menunjukkan penaksir variansi di dalam psu (kelompok) dari strata ke-h. Karena n_{hi} adalah sampel acak dari N_{hi}, dan $\bar{\bar{x}}_{hi}$ merupakan rata-rata sampel dari subsampel n_{hi}, diketahui pula bahwa s_{hi}^2 adalah penaksir tak bias dari S_{hi}^2, sehingga

$E(s_{hi}^2) = E\left(\frac{1}{n_{hi} - 1} \sum_{j=1}^{n_{hi}} (x_{hij} - \bar{\bar{x}}_{hi})^2\right)$

$E(s_{hi}^2) = \frac{1}{N_{hi} - 1} \sum_{\text{i}=1}^{N_{hi}} (X_{hij} - \bar{\bar{X}}_{hi})^2$

$E(s_{hi}^2) = S_{hi}^2$

Penaksir varians $\hat{V}(\hat{X})$ merupakan penaksir yang tak bias untuk varians, hal ini dapat dibuktikan dengan membuktikan $E\left(\hat{V}\left(\hat{X}\right)\right) = V(\hat{X})$ pada proses pembuktian berikut ini.

Pembuktian:

$$E\left(\hat{V}\left(\hat{X}\right)\right) = E\left(\sum_{\text{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h} + \sum_{\text{h}=1}^{L} \frac{M_h}{m_h} \sum_{\text{i}=1}^{m_h} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= E\left(\sum_{\text{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h}\right) + E\left(\sum_{\text{h}=1}^{L} \frac{M_h}{m_h} \sum_{\text{i}=1}^{m_h} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= E_h E_i \left(\sum_{\text{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h}\right) + E_h E_i E_j \left(\sum_{\text{h}=1}^{L} \frac{M_h}{m_h} \sum_{\text{i}=1}^{m_h} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= E_h \left(E_i \sum_{\text{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h}\right) + E_h E_i \left(\sum_{\text{h}=1}^{L} \frac{M_h}{m_h} E_j \sum_{\text{i}=1}^{m_h} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= E_h \left(L.\, M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h}\right) + E_h E_i \left(\sum_{\text{h}=1}^{L} \frac{M_h}{m_h}\, m_h\, N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= L E_h \left(M_h^2 \frac{M_h - m_h}{M_h} \frac{s_h^2}{m_h}\right) + E_h \left(E_i \sum_{\text{h}=1}^{L} M_h\; N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= L\frac{1}{L}\sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + E_h\left(\sum_{\mathrm{h}=1}^{L} E_i\left(M_h N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{s_{hi}^2}{n_{hi}}\right)\right)$$

$$= \sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + E_h\left(\sum_{\mathrm{h}=1}^{L} \frac{1}{M_h}\sum_{i=1}^{M_h} M_h N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= \sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + E_h\left(\sum_{\mathrm{h}=1}^{L} \frac{1}{M_h} M_h \sum_{i=1}^{M_h} N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= \sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + E_h\left(\sum_{\mathrm{h}=1}^{L} \sum_{i=1}^{M_h} N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{s_{hi}^2}{n_{hi}}\right)$$

$$= \sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + E_h\left(\sum_{\mathrm{h}=1}^{L} \sum_{i=1}^{M_h} 1.\, N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{s_{hi}^2}{n_{hi}}\right)$$

Karena ketika ukuran m_h kelompok mendekati ukuran M_h kelompok, maka $\frac{M_h}{m_h} \approx \frac{M_h}{M_h} \approx 1$

$$= \sum_{\mathrm{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + \sum_{\mathrm{h}=1}^{L} \frac{M_h}{m_h}\sum_{i=1}^{M_h} N_{hi}^2 \frac{N_{hi}-n_{hi}}{N_{hi}}\ \frac{S_{hi}^2}{n_{hi}}$$

$$= V(\hat{X})$$

Berdasarkan pembuktian di atas, ini menunjukkan bahwa $\hat{V}(\hat{X})$ adalah penaksir tak bias dari $V(\hat{X})$.

3.5 Alokasi Sampel

Permasalahan yang biasanya muncul pada pengalokasian sampel adalah berapa banyak kelas m_h dan berapa banyak n_{hi} dari kelas ke-*hi* yang harus dipilih. Apakah akan dipilih m_h kelas lebih sedikit dan lebih banyak n_{hi} atau sebaliknya? Prosedur untuk menyelidiki permasalahan ini adalah pertama-tama menentukan variansi dan fungsi biaya yang berfungsi sebagai kendala linear, dan kemudian menentukan m_h dan n_{hi} untuk meminimumkan variansi subjek fungsi biaya yang diberikan. Untuk menyederhanakan variansi, perhatikan subsampel-subsampel dari proporsi yang sama, seringkali mengambil dari *psu* itu, sehingga akan diasumsikan bahwa

$$\frac{N_{hi}}{n_{hi}} = f_{2h} \tag{3.26}$$

Misalkan, apabila $f_{2h} = 0{,}05$, berarti 5% dari N_{hi} diambil sebagai sampel acak. Sebagai ilustrasi, misalkan $M_h = 10$ kelas di strata ke-h, maka

$$\frac{n_{h1}}{N_{h1}} = \frac{n_{h2}}{N_{h2}} = \cdots = \frac{n_{h10}}{N_{h10}} = f_{2h} \tag{3.27}$$

Persamaan (3.27) dapat dinyatakan sebagai berikut:

$$\frac{\frac{1}{10}(n_{h1} + \cdots + n_{h10})}{\frac{1}{10}(N_{h1} + \cdots + N_{h10})} = f_{2h}$$

yang dapat dinyatakan sebagai

$$\frac{\bar{n}_h}{\bar{N}_h} = f_{2h} \tag{3.28}$$

dimana $\bar{N}_h$ adalah rata-rata jumlah populasi per kelas di strata ke-h dan juga dapat dianggap sebagai nilai ekspektasi dari N_{hi}. Hal ini dapat dinyatakan sebagai berikut:

$$\bar{N}_h = \frac{N_h}{M_h} = \frac{\sum_{\text{i}=1}^{M_h} N_{hi}}{M}$$

Demikian pula, $\bar{n}_h$ juga dapat dianggap sebagai nilai ekspektasi dari n_{hi}, dan dapat ditunjukkan sebagai

$$\bar{n}_h = f_{2h}\bar{N}_h$$

Perhatikan bahwa interpretasi ini berbeda dari

$$\bar{n}_h = \frac{1}{m_h}\sum_{\text{i}=1}^{m_h} n_{hi}$$

yang hanya rata-rata sampel.

Dengan menggunakan $\bar{n}_h$ dan $\bar{N}_h$ sebagaimana didefinisikan pada persamaan (3.28), perumusan variansi yang diberikan pada persamaan (3.22) menjadi:

$$V(\hat{X}) = \sum_{\text{h}=1}^{L} M_h^2 \frac{M_h - m_h}{M_h}\frac{S_h^2}{m_h} + \sum_{\text{h}=1}^{L}\frac{M_h}{m_h}\sum_{\text{i}=1}^{M_h}\bar{N}_h^2\frac{\bar{N}_h - \bar{n}_h}{\bar{N}_h}\frac{S_{hi}^2}{\bar{n}_h} \tag{3.29}$$

Ruas kanan persamaan (3.29) dapat disederhanakan lagi menjadi:

$$\sum_{h=1}^{L}\frac{M_h}{m_h}\sum_{i=1}^{M_h}\bar{N}_h^2\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}\frac{S_{hi}^2}{\bar{n}_h} = \sum_{h=1}^{L}\frac{M_h}{m_h}\sum_{i=1}^{M_h}\frac{N_h^2}{M_h^2}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}\frac{S_{hi}^2}{\bar{n}_h}$$

$$= \sum_{h=1}^{L}\frac{M_h}{m_h}\frac{N_h^2}{M_h^2}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}\frac{1}{\bar{n}_h}\sum_{i=1}^{M_h}S_{hi}^2$$

$$= \sum_{h=1}^{L}\frac{N_h^2}{m_h\bar{n}_h}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}\frac{1}{M_h}\sum_{i=1}^{M_h}S_{hi}^2$$

$$= \sum_{h=1}^{L}\frac{N_h^2}{m_h\bar{n}_h}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}S_{2h}^2$$

dengan

$$S_{2h}^2 = \frac{1}{M_h}\sum_{i=1}^{M_h}S_{hi}^2$$

$$S_{2h}^2 = \frac{1}{M_h}\sum_{i=1}^{M_h}S_{hi}^2 \cdot \left(\frac{\bar{N}_h}{\bar{N}_h}\right)$$

dimana ditetapkan $\bar{N}_h = N_{hi}$ dan $\bar{N}_h = N_h/M_h$, sehingga:

$$S_{2h}^2 = \frac{1}{M_h\bar{N}_h}\sum_{i}^{M_h}N_{hi}\,S_{hi}^2$$

$M_h\bar{N}_h$menunjukkan jumlah populasi dari strata ke-*h*, sedangkan $\sum_i^{M_h} N_{hi}\,S_{hi}^2$ dapat diinterpretasikan sebagai jumlah kuadrat variansi di dalam kelas di strata ke-*h* untuk semua M_h kelas. Oleh karena itu, S_{2h}^2 dapat dianggap mewakili dalam variansi kelas untuk strata ke-*h*. Dengan menggunakan S_{2h}^2, persamaan (3.29) menjadi:

$$V(\hat{X}) = \sum_{h=1}^{L}M_h^2\frac{M_h-m_h}{M_h}\frac{S_h^2}{m_h} + \sum_{h=1}^{L}\frac{N_h^2}{m_h\bar{n}_h}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}S_{2h}^2 \tag{3.30}$$

dan akhirnya diperoleh variansi sederhana yang akan digunakan untuk memudahkan analisis selanjutnya. Berdasarkan penaksir variansi pada persamaan (3.25), maka diperoleh penaksir variansi dengan alokasi sampel yaitu:

$$\hat{V}(\hat{X}) = \sum_{h=1}^{L}M_h^2\frac{M_h-m_h}{M_h}\frac{s_h^2}{m_h} + \sum_{h=1}^{L}\frac{N_h^2}{m_h\bar{n}_h}\frac{\bar{N}_h-\bar{n}_h}{\bar{N}_h}s_{2h}^2 \tag{3.31}$$

$$s_{2h}^2 = \frac{1}{M_h\bar{n}_h}\sum_{i=1}^{m_h}n_{hi}\,s_{hi}^2 \tag{3.32}$$

3.6 Perbandingan *Stratified Cluster Sampling* dan *Simple Cluster Sampling*

Pada bab sebelumnya telah dikemukakan bahwa s*tratified cluster sampling* memiliki varians lebih kecil daripada *simple random sampling*, *simple cluster sampling, dan stratified random sampling*. Oleh karena itu, *stratified cluster sampling* digunakan ketika ingin mengurangi variansi dari penaksir dan menurunkan biaya survei. Untuk mempertimbangkan pengurangan variansi, perlu dibandingkan taksiran variansi dari *stratified cluster sampling* dengan taksiran variansi dari metode lainnya. Berikut adalah perbandingan taksiran variansi dari *stratified cluster sampling* dengan taksiran variansi dari *simple cluster sampling*. Untuk perbandingan taksiran varians dengan metode yang lainnya dapat dilihat pada lampiran 4.

Untuk menyederhanakan variansi dari *stratified cluster sampling*, dapat dengan cara memisalkan:

$$M_h = \bar{M} = \frac{\sum^L M_h}{L} = \frac{M}{L} \tag{3.33}$$

$$m_h = \bar{m} = \frac{\sum^L m_h}{L} = \frac{m}{L} \tag{3.34}$$

$$N_{hi} = \bar{N} = \frac{1}{\bar{M}\,L}\sum^{L}\sum^{M} N_{hi} \tag{3.35}$$

Selanjutnya dengan mengasumsikan jumlah setiap subsampel sama dari setiap kelas (psu), maka variansi dari rata-rata untuk sampling stratifikasi proporsional (*proportional stratified random sampling)* adalah:

$$V\left(\bar{x}_{prop}\right) = \frac{N-n}{N}\sum^{L} \frac{N_h}{N}\frac{S_h^2}{n} \tag{3.36}$$

Kemudian dari persamaan (3.36), sampling unit utama *N* dan *n* keduanya digantikan oleh *M* (total populasi) dan *m* (total sampel) sehingga persamaan (3.36) menjadi

$$V(\bar{x}_{st}) = \frac{M-m}{M}\sum^{L} \frac{M_h}{M}\frac{S_h^2}{m} \tag{3.37}$$

$$= \frac{M-m}{M}\sum^{L}\frac{M/L}{M}\frac{S_h^2}{m}$$

$$= \frac{M-m}{M}\sum^{L}\frac{1}{L}\frac{S_h^2}{m}$$

$$S_h^2 = \frac{1}{\bar{M}-1}\sum_{i}^{\bar{M}}\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2 \qquad (3.38)$$

$$\bar{\bar{X}}_{hi} = \frac{X_{hi}}{N_{hi}} \qquad (3.39)$$

$$\bar{\bar{X}}_h = \frac{1}{\bar{M}}\sum_{i}^{\bar{M}}\bar{\bar{X}}_{hi}$$

Variansi dari rata-rata untuk metode sampling acak sederhana (*simple random sampling)* *m* cluster adalah:

$$\sigma_{ran}^2 = \frac{M-m}{M}\frac{(\frac{1}{M})\sum^{L}\sum^{\bar{M}}\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2}{m} \qquad (3.40)$$

$$\bar{\bar{X}} = \frac{1}{M}\sum^{L}\sum^{\bar{M}}\bar{\bar{X}}_{hi} = \frac{1}{\bar{M}L}\sum^{L}\sum^{\bar{M}}\bar{\bar{X}}_{hi}$$

Untuk mengevaluasi keuntungan stratifikasi, akan dibandingkan dua variansi, yaitu variansi pada persamaan (3.37) dan variansi pada persamaan (3.40):

$$\sigma_{st}^2 = \frac{M-m}{M}\sum^{L}\frac{1}{L}\frac{S_h^2}{m} \qquad (3.41)$$

$$= \frac{M-m}{Mm}\frac{1}{L}\frac{1}{\bar{M}-1}\sum^{L}\sum^{\bar{M}}\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2$$

$$\sigma_{ran}^2 = \frac{M-m}{Mm}\frac{1}{M}\sum^{L}\sum^{\bar{M}}\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2 \qquad (3.42)$$

$$= \frac{M-m}{Mm}\frac{1}{L}\frac{1}{\bar{M}}\sum^{L}\sum^{\bar{M}}\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2$$

Pada saat $1 \ll \bar{M}$, maka dengan memisalkan $\bar{M} = \bar{M} - 1$, persamaan (3.41) dan persamaan (3.42) dapat disederhanakan menjadi:

$$\sigma_{st}^2 = W \sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2 \tag{3.43}$$

$$\sigma_{ran}^2 = W \sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2 \tag{3.44}$$

dimana

$$W = \frac{M-m}{Mm} \frac{1}{L} \frac{1}{\bar{\bar{M}}}$$

Keuntungan absolut akibat stratifikasi ditemukan dengan:

$$\sigma_{ran}^2 - \sigma_{st}^2 = W \left[\sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2 - \sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2 \right] \tag{3.45}$$

Penyederhanaan tanda dalam kurung secara aljabar adalah sebagai berikut:

$$\sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2 - \sum^{L} \sum^{\bar{M}} \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2 = \sum^{L} \sum^{\bar{M}} \left[\left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}\right)^2 - \left(\bar{\bar{X}}_{hi} - \bar{\bar{X}}_h\right)^2 \right]$$

$$= \sum_{h}^{L} \sum_{i}^{\bar{M}} \left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)\left(2\bar{\bar{X}}_{hi} - \bar{\bar{X}} - \bar{\bar{X}}_h\right)$$

$$= \sum_{h}^{L} \left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)\left(2\bar{M}\bar{\bar{X}}_h - \bar{M}\bar{\bar{X}} - \bar{M}\bar{\bar{X}}_h\right)$$

$$= \sum_{h}^{L} \left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)^2 \bar{M}$$

Oleh karena itu persamaan (3.45) dapat dinyatakan dengan:

$$\sigma_{ran}^2 - \sigma_{st}^2 = W \sum_{h}^{L} \left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)^2 \bar{M} \tag{3.46}$$

$$= \frac{M-m}{Mm} \frac{1}{L} \sum_{h}^{L} \left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)^2$$

$$\sigma_{ran}^2 = \sigma_{st}^2 + \frac{M-m}{Mm}\frac{1}{L}\sum_{h}^{L}\left(\bar{\bar{X}}_h - \bar{\bar{X}}\right)^2 \qquad (3.47)$$

Hal ini menunjukkan bahwa *stratified cluster sampling* memiliki variansi lebih kecil daripada *simple cluster sampling*. Ketika ada perbedaan antar strata, maka dianjurkan untuk menggunakan *stratified cluster sampling*.