BAB III
METODE *STRATIFIED RANDOM SAMPLING*

3.1 Pengertian *Stratified Random Sampling*

Dalam bukunya Elementary Sampling Theory, Taro Yamane menuliskan *"The process of breaking down the population into strata, selecting simple random samples from each stratum, and combining these into a single sampel to estimate population parameter is called stratified random sampling*". Berdasarkan kutipan di atas dapat dinyatakan bahwa *stratified random sampling* merupakan proses pengambilan sampel melalui proses pembagian populasi kedalam strata, memilih sampel acak sederhana dari setiap stratum, dan menggabungkannya ke dalam sebuah sampel untuk menaksir parameter populasinya.

Sampel yang representatif adalah sampel yang benar-benar dapat mewakili karakteristik seluruh populasi. Jika populasi bersifat homogen, maka sampel bisa diambil dari populasi yang mana saja, namun jika populasi bersifat heterogen, maka sampel harus mewakili dari setiap bagian yang heterogen dari populasi tersebut sehingga hasil penelitian dari sampel dapat terpenuhi terhadap setiap anggota populasi.

Proses pembagian populasi kedalam stratum bertujuan agar sampel yang diambil dari setiap stratum dapat merepresentasikan karakteristik populasi yang berukuran besar dan heterogen. Oleh karena itu, stratum harus dibentuk sehomogen mungkin dengan manganalisis karakteristik populasi dengan baik. Terdapat tiga tahapan yang harus dilakukan dalam mengambil sampel dengan menggunakan metode *stratified random sampling*, yaitu sebagai berikut:

1. Tahap Pertama

Populasi yang berukuran N dibagi menjadi sub-sub populasi yang masing-masing terdiri atas $N_1, N_2, N_3, \dots, N_L$ elemen. Diantara dua 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). Dalam pembentukan stratum harus diperhatikan variabel apa yang dijadikan sebagai dasar pembentukan stratum, yaitu variabel yang memiliki korelasi tinggi dengan variabel yang diteliti.

2. Tahap Kedua

Sampel diambil dari setiap stratum secara terpisah (independen) dengan ukuran sampel dari masing-masing stratum adalah $n_1, n_2, n_3, \dots, n_L$ dengan syarat $n_1 + n_2 + n_3 + \cdots + n_L = n$.

3. Tahap Ketiga

Setelah diperoleh sampel, selanjutnya dilakukan penaksiran terhadap parameter yang diperlukan dan selanjutnya dibuat kesimpulan untuk populasi berdasarkan hasil penaksiran sampel.

3.2 Total Populasi

3.2.1 Pengertian Total Populasi

Apabila N menyatakan banyak anggota populasi dan L menyatakan banyak stratum maka total populasi adalah jumlah dari total stratum dan didefinisikan sebagai berikut:

$$X = \sum_{h}^{L} X_h = \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}$$

Dimana X_h adalah total dari stratum h yang didefinisikan sebagai berikut:

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

X_{hi} adalah sampel ke-i pada stratum ke-h.

Rata-rata stratum didefiniskan sebagai berikut:

$$\bar{X}_h = \frac{X_h}{N_h}$$

Dan rata-rata populasi didefinisikan sebagai berikut:

$$\bar{X} = \frac{X}{N} = \frac{\sum_h^L N_h \bar{X}_h}{N}$$

3.2.2 Penaksir Total Populasi

Total populasi merupakan jumlah dari total stratum sehingga dalam menaksir total populasi dapat melalui penjumlahan dari taksiran total stratum. Taksiran total stratum dapat diperoleh dengan rumus sebagai berikut:

$$\hat{X}_h = N_h \bar{x}_h \tag{3.1}$$

Dimana $\bar{x}_h$ merupakan rata-rata sampel dari sebuah subsampel acak yang berukuran n_h dari stratum ke-h.

Taksiran total populasi $\boldsymbol{X}$ adalah jumlah dari taksiran total stratum seperti yang dijabarkan dalam persamaan berikut:

$$\hat{X}_{st} = N_1\bar{x}_1 + N_2\bar{x}_2 + \cdots + N_L\bar{x}_L = \sum_{h=1}^{L} N_h \bar{x}_h \tag{3.2}$$

dan taksiran rata-rata populasi menjadi

$$\widehat{\bar{X}_h} = \bar{x}_{st} = \frac{\hat{X}_{st}}{N} = \frac{N_1\bar{x}_1 + .. + N_L\bar{x}_L}{N_1 + \cdots + N_L} = \frac{\sum_{h=1}^{L} N_h \bar{x}_h}{N}$$

Karena rata-rata sampel stratum $\bar{x}_h$ yang diperoleh dengan sampling acak sederhana merupakan penaksir tak bias dari rata-rata stratum $\bar{X}_h$.

$$E(\bar{x}_{st}) = \bar{X}_h$$

Maka nilai ekspektasi $\bar{x}_{st}$menjadi

$$\mathrm{E}(\bar{x}_{st}) = \frac{E(N_1\bar{x}_1 + \ldots + N_L\bar{x}_L)}{N}$$

$$= \frac{N_1\bar{X}_1 + \cdots + N_L\bar{X}_L}{N} = \frac{X_1 + \cdots + X_L}{N} = \frac{X}{N} = \bar{X}$$

Jadi $\bar{x}_{st}$ merupakan penaksir tak bias untuk $\bar{X}$.

Informasi mengenai $N_1 dan\ N_2$ dibutuhkan untuk menentukan penaksir $\bar{x}_{st}$ karena E$(\bar{x}_{st}) = \frac{X}{N},\ sehingga\ E\ (N\bar{x}_{st}) = X$. Pada penjabaran di atas sudah diketahui taksiran dari rata-rata populasi sehingga akan diperoleh persamaan berikut:

$$N\bar{x}_{st} = \sum_{h}^{L} N_h \bar{x}_h = \hat{X}_{st}$$

Ini artinya, penaksir $\hat{X}_{st}$ juga merupakan penaksir tak bias untuk X karena dapat ditunjukkan bahwa $E(\hat{X}_{st}) = X$

3.2.3 Varians Penaksir Total Populasi dan Penaksirnya

Varians dari $\hat{X}_{st}$ diperoleh dengan menggunakan hasil dari varians $\bar{x}_{st}$. Sebelum membahas mengenai varians untuk penaksir total populasi, akan dijelaskan terlebih dahulu mengenai varians untuk rata-rata sampel.

Variansi dari $\bar{x}_{st}$didefinisikan oleh:

$$\sigma^2_{\bar{x}_{st}} = \frac{1}{M} \sum_{st=1}^{M} (\bar{x}_{st} - \bar{X})^2$$

M adalah banyaknya kemungkinan rata-rata sampel, dimana $M = \binom{N_1}{n_1}\binom{N_2}{n_2}$.

Selanjutnya akan dicari $V(\bar{x}_{st})$ dalam bentuk varians stratum $S^2{}_h$ yang dapat menunjukkan karakteristik dari $V(\bar{x}_{st})$. $S^2{}_h$ didefinisikan sebagai berikut:

$$S_h^2 = \frac{1}{N_h - 1} \sum_{h=1}^{L} (X_{hi} - \bar{X}_h)^2 \qquad (3.3)$$

Ketika populasi dan sampel cukup besar, maka kemungkinan rata-rata sampel M akan semakin besar sehingga dalam menghitung $V(\bar{x}_{st})$ menggunakan definisi $\sigma^2_{\bar{x}_{st}}$ akan sulit. Selanjutnya akan ditunjukkan bagaimana $V(\bar{x}_{st})$ dapat dijelaskan dalam bentuk S_h^2.

Diketahui,

$$\bar{x}_{st} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N} = w_1\bar{x}_1 + w_2\bar{x}_2$$

dimana $w_h = \frac{N_h}{N}$, dan disebut stratum *weight* (bobot). Selama n_h dipilih dengan sampling acak dan saling bebas antara satu dengan yang lainnya, maka diperoleh:

$$\mathrm{V}(\bar{x}_{st}) = w_1^2 V(\bar{x}_1) + w_2^2 \mathrm{V}(\bar{x}_2)$$

Berdasarkan definisi varians untuk rata-rata sampel yang dipilih secara acak dan tanpa pengembalian, yaitu $\sigma_{\bar{x}}^2 = \frac{N-n}{N}\frac{S^2}{n}$, maka persamaan di atas dapat diubah menjadi sebagai berikut:

$$\mathrm{V}(\bar{x}_{st}) \ = w_1^2 V(\bar{x}_1) + w_2^2 \mathrm{V}(\bar{x}_2)$$

$$= \mathrm{w}_1^2 \frac{\mathrm{N}_1 - \mathrm{n}_1}{\mathrm{N}_1}\frac{\mathrm{S}_1^2}{\mathrm{n}_1} + \mathrm{w}_2^2 \frac{\mathrm{N}_2 - \mathrm{n}_2}{\mathrm{N}_2}\frac{\mathrm{S}_2^2}{\mathrm{n}_2}$$

$$= \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h}$$

Persamaan di atas dapat dituliskan kembali dalam bentuk:

$$V(\bar{x}_{st}) = \frac{1}{N^2}\sum N_h^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h} \qquad (3.4)$$

S_h^2 menunjukkan varians dalam masing-masing stratum. Jadi dapat disimpulkan bahwa ketika varians dalam masing-masing stratum kecil, maka $\mathrm{V}(\bar{x}_{st})$ akan kecil dan ketelitian dari $\bar{x}_{st}$ akan tinggi.

Karena rumusan dari $\mathrm{V}(\bar{x}_{st})$ memuat S_h^2, maka tidak dapat digunakan pada masalah praktis dimana S_h^2 biasanya tidak diketahui. Oleh karena itu, dibutuhkan penkasir untuk S_h^2 dan diperoleh penkasir untuk $\mathrm{V}(\bar{x}_{st})$. Karena s_h^2 adalah sebuah penaksir tak bias dari S_h^2 maka selanjutnya akan disubtitusikan s_h^2 kedalam persamaan (3.4). Dimana, $\hat{V}(\bar{x}_{st})$ merupakan penaksir tak bias untuk $\mathrm{V}(\bar{x}_{st})$.

$$\hat{V}(\bar{x}_{st}) = \frac{1}{N^2}\sum N_h^2 \frac{N_h - n_h}{N_h}\frac{s_h^2}{n_h}$$

Setelah diperoleh varians $\bar{x}_{st}$ selanjutnya akan dibahas mengenai varians dari penaksir total populasi, yaitu sebagai berikut:

Varians dari $\hat{X}_{st}$ adalah:

$$\begin{aligned} V(\hat{X}_{st}) &= V(N\,\bar{x}_{st}) \\ &= N^2\,V(\bar{x}_{st}) \\ &= N^2\left(\frac{1}{N^2}\sum N_h^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h}\right) \\ &= \sum N_h^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h} \end{aligned}$$

Ini artinya $V(\hat{X}_{st})$ sama dengan $V(\bar{x}_{st})$ dengan syarat $\frac{1}{N^2}$ dihilangkan.

Estimator $V(\hat{X}_{st})$ diperoleh untuk mendapatkan penaksir $V(\bar{x}_{st})$, yaitu dengan cara mengganti S_h^2 menjadi s_h^2 . Oleh karena itu, estimatornya adalah:

$$\hat{V}(\hat{X}\mathrm{st}) = \sum_{h=1}^{L} N_h^2 \frac{N_h - n_h}{N_h}\frac{S_h^2}{n_h} \qquad (3.5)$$

Dimana,

$$s_h^2 = \frac{1}{n_h - 1}\sum^{n_h}(x_{hi} - \bar{x}_h)^2$$

Dengan menggunakan hasil $\hat{V}(\bar{x}_{st})$ yang merupakan penaksir tak bias dari $V(\bar{x}_{st})$, kita dengan mudah dapat menunjukkan $\hat{V}(\hat{X}_{st})$ merupakan penaksir tak bias dari $V(\hat{X}_{st})$. Diketahui bahwa:

$$E[\hat{V}(\bar{x}_{st})] = V(\bar{x}_{st})$$

Dimana $\bar{x}_{st} = \hat{X}_{st}/N$. Substitusikan $\bar{x}_{st}$ ke kedua ruas persamaan di atas, maka diperoleh:

$$E\left[\frac{1}{N^2}\hat{V}(\hat{X}_{st})\right] = \frac{1}{N^2}V(\hat{X}_{st})$$

Maka diperoleh:

$$E[\hat{V}(\hat{X}_{st})] = V(\hat{X}_{st})$$

3.3 Alokasi Sampel

Alokasi sampel merupakan suatu metode untuk menentukan ukuran sampel dari setiap stratum untuk didistribusikan kedalam sampel n. Ada dua masalah yang perlu dipertimbangkan oleh peneliti, yaitu menentukan ukuran sampel n dan mengalokasikan sampel ini diantara strata h untuk menentukan n_h. Terdapat beberapa metode yang dapat digunakan dalam mengalokasikan sampel dari setiap strata yaitu sebagai berikut:

3.3.1 Alokasi Sembarang

Alokasi sembarang merupakan suatu cara mengalokasikan sampel dimana ukuran sampel masing-masing strata ditentukan secara sembarang dengan syarat, minimal harus ada dua satuan pengamatan yang dipilih dari setiap stratanya. Dalam praktek, alokasi seperti ini jarang dan tidak disarankan untuk digunakan karena menyebabkan standar *error* membesar.

3.3.2 Alokasi Proporsional

Alokasi proporsional merupakan suatu metode untuk mengalokasikan sampel dimana ukuran sampel untuk setiap stratum sesuai dengan proporsi ukuran masing-masing stratum. Metode ini paling sering digunakan karena praktis dan jelas, tidak bergantung pada pertimbangan biaya dan peneliti hanya perlu mengetahui ukuran stratum. Metode alokasi proporsional bersifat sederhana dan lebih mudah bila dibandingkan dengan metode lainnya dengan tingkat ketepatan yang tidak berbeda jauh dengan metode lainnya.

Berikut rumus yang digunakan untuk mengalokasikan sampel secara proporsional:

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

Dimana n_h merupakan ukuran sampel dari setiap strata, dan rumus di atas menunjukkan bahwa n harus dialokasikan sesuai dengan N_h/N (secara proporsional). Pengambilan sampel dilakukan secara acak sederhana di setiap strata, sehingga peluang dari setiap unit sampling di strata h untuk terpilih sebagai subsampel n_h yaitu $\frac{n_h}{N_h} = f$. Setiap unit dalam populasi mempunyai peluang yang sama untuk terpilih sebagai sampel.

4. Alokasi Optimum

Dalam penarikan sampel stratifikasi, nilai dari ukuran sampel n_h dalam masing-masing strata dipilih oleh peneliti. Dalam melakukan penelitan, peneliti akan dihadapi oleh dua kemungkinan dalam mempertimbangkan biaya penelitian atau memperkecil error dalam tahap penarikan sampel, yaitu peneliti mungkin memilih untuk meminimumkan $V(\bar{x}_{st})$ dengan biaya tertentu untuk memperoleh sampel atau untuk meminimumkan biaya dengan sebuah nilai $V(\bar{x}_{st})$ tertentu. Metode untuk mengalokasikan sampel n diantara strata untuk meminimalkan $V(\bar{x}_{st})$ disebut alokasi optimum. Berikut ini adalah bentuk sederhana dari fungsi biaya, yaitu:

$$\text{Biaya} = c = c_0 + \sum^{L} c_h n_h$$

Biaya dalam setiap lapisan adalah proposional dengan ukuran sampel, tetapi biaya perunit c_h dapat bervariasi antara lapisan satu dengan lapisan lainnya.

Biaya overhead dinyatakan dengan c_0, fungsi biaya ini adalah *fixed cost* (ongkos tetap) bila sebagian besar biaya itemnya diperoleh dengan mengukur setiap unit dan tidak bergantung pada ukuran dari sampel survey. c_h merupakan *variable cost* dan menunjukkan ongkos tiap unit sampel pada stratum ke h.

Dalam penarikan sampel acak stratifikasi dengan sebuah fungi biaya seperti di atas, variansi perkiraan rata-rata x_{st} adalah minimum untuk biaya tertentu C, dan biaya adalah minimum untuk $V(\bar{x}_{st})$ tertentu.

Diketahui rumus variansi sebagai berikut:

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum^{L} N_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}$$

Masalah yang dihadapi adalah bagaimana memilih n_h agar meminimumkan $V(\bar{x}_{st})$ dengan biaya tertentu (dengan fungsi biaya linear). Perumusan untuk menentukan besarnya sampel dari setiap stratum agar meminimumkan $V(\bar{x}_{st})$ adalah sebagai berikut:

$$n_h = \frac{N_h S_h / \sqrt{c_h}}{\sum^{L} N_h S_h / \sqrt{c_h}} \bullet n$$

5. Alokasi Neyman

Alokasi Neyman digunakan apabila varians setiap strata berbeda-beda besarnya sedangkan ongkos per unit penarikan sampel dianggap relatif sama. Rumus ukuran sampel pada setiap strata untuk alokasi Neyman adalah sebagai berikut:

$$n_h = \frac{N_h S_h}{\sum_{h=I}^{L} N_h S_h} n$$

Perumusan di atas menunjukkan bahwa sampel berukuran n dialokasikan secara proporsi ke $N_h S_h$.

Alokasi Neyman dipergunakan juga ketika strata besar dan dari strata yang heterogen. Sebagai contoh jika sebuah kota dibagi dalam dua wilayah dan dalam wilayah satu terdapat sedikit perbedaan antara pendapatan keluarga, sedangkan di wilayah kedua terdapat variasi yang besar. Dengan rumus alokasi Neyman tersebut dapat diambil sampel wilayah kedua. Jelas bahwa distrik dengan varians yang besar dan sampel yang besar akan memeberikan sampel yang tidak representatif.