Stratified Random Sampling

When we know that parts of the population differ with respect to the quantity that we are estimating, we can obtain better estimates by using stratification. A **stratified random sample** is obtained by separating the population elements into nonoverlapping groups, called **strata**, and then selecting a simple random sample from each stratum. Stratification is particularly useful because: i) it can yield smaller error bounds than SRS, especially when the measurement is homogeneous within strata, ii) The cost per observation can be lowered by appropriate choice of strata, and iii) it will yield stratum-specific estimates of population parameters.

Drawing a Stratified Random Sample, notation: Choose strata, then take a SRS from each. L = number of strata, N_i = population size from stratum i, $N = N_1 + N_2 + ... + N_L$ = total population size, n_i = sample size from stratum i, $n = n_1 + n_2 + ... + n_L$ = total sample size.

Estimation of a population mean and total from stratified random sampling: To estimate the mean μ from stratified random sampling, we use the sample stratified mean:

$$\widehat{\mu} = \bar{y}_{st} = \frac{1}{N}\left[N_1\bar{y}_1 + N_2\bar{y}_2 + ... + N_L\bar{y}_L\right] = \frac{1}{N}\sum_{i=1}^{L} N_i\bar{y}_i \ .$$

Since the strata are independent, we obtain the estimated variance by summing the estimated variances from each stratum:

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

To estimate the total τ from stratified random sampling, we just estimate the total τ_i from each stratum and sum. Also, since the strata are independent, the estimated variance of the total is just the sum of the estimated variances from each stratum:

$$\widehat{\tau} = \sum_{i=1}^{L} N_i\bar{y}_i \text{ and}$$

$$\hat{V}(\widehat{\tau}) = \sum_{i=1}^{L} N_i^2\left(\frac{N_i - n_i}{N_i}\right)\frac{s_i^2}{n_i}$$