LECTURE 8: STRATIFIED RANDOM SAMPLING 5.1 – 5.4

5.1 Stratified Random Sample

A stratified random sampling is a sampling method in which the sampling frame is partitioned into groups (called strata, singular stratum) and then a random sample is drawn independently within each group (stratum).

The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded.

The strata should be relatively homogeneous groups with respect to some variables (called stratification variables).

In order for a stratified sample to be a"miniature version" of the population, the size of the sample in each stratum is taken proportional to the size of the stratum. This approach is called proportional allocation.

Example Suppose that in a engineering company the employees are categorized as follows:

full-time, male: 90
full-time, female: 18
part-time, male: 9
part-time, female: 63
Total: 180

Suppose we want to take a sample of 40 employees, stratified according to the four given strata. We are using a stratified simple random sample with proportional allocation.

Step 1. We calculate the proportion of elements in each stratum.

full-time, male: 90/180=0.5
full-time, female: 18/180=0.1
part-time, male: 9/180=0.05
part-time, female: 63/180=0.35

Step 2. We know that in the sample of 40 employees, the proportions in each stratum should be the same as in the population. Thus, we should sample:

full-time, male: (0.5)(40)=20
full-time, female: (0.1)(40)=4
part-time, male: (0.05)(40)=2

part-time, female: (0.35)(40)=14.

5.4 Population Parameters for Strata

Suppose N population elements are grouped into L strata, and there are $N_1, N_2, \ldots, N_L$ sampling units within each stratum, so that

$$N = \sum_{h=1}^{L} N_h \,.$$

Denote by $X_{h,i}$ the measurement on the ith population element within stratum h, $i = 1, \ldots, N_h$, $h = 1, \ldots, L$.

The total within a stratum h is

$$X_{h+} = \sum_{i=1}^{N_h} X_{h,i} \,.$$

The population total is

$$X = \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{h,i} = \sum_{h=1}^{L} X_{h+} \,.$$

The mean for a stratum h is

$$\bar{X}_h = \frac{X_{h+}}{N_h} = \frac{\sum_{i=1}^{N_h} X_{h,i}}{N_h} \,.$$

The population mean is

$$\bar{X} = \frac{X}{N} = \frac{\sum_{h=1}^{L} X_{h+}}{N} = \frac{\sum_{h=1}^{L} N_h \bar{X}_h}{N} = \sum_{h=1}^{L} W_h \bar{X}_h$$

where $W_h = N_h/N$. Hence, the population mean $\bar{X}$ is a weighted average of the means $\bar{X}_h$ of individual strata, with the weights proportional to the size of each stratum.

The variance within a stratum h is

$$\sigma_{hx}^2 = \frac{\sum_{i=1}^{N_h} (X_{h,i} - \bar{X}_h)^2}{N_h} \,.$$