STRATIFIED AND SYSTEMATIC SAMPLING DESIGNS

Introduction

These are other sampling techniques used to select elements from a population as a representative of the whole for analysis purposes.

Learning outcomes

By the end of this study unit, you should be able to;

- Distinguish between stratified and systematic sampling methods
- Determine sample sizes using stratified sampling methods
- Carry out computations

stratified sampling

With stratified sampling, the population is divided into homogeneous, mutually exclusive groups called strata, and then independent samples are selected from each stratum.

There are three main reasons for stratification.

- to make the sampling strategy more efficient than SRS or SYS.
- to ensure adequate sample sizes for specific domains of interest for which analysis is to be performed.
- to protect against drawing a 'bad' sample.

Units within a stratum should be similar with respect to the variable of interest) and the strata themselves must be as different as possible (with respect to the same variable of interest)

Advantages of stratified sampling

i. It can increase the precision of overall population estimates, resulting in a more efficient sampling strategy.
ii. It can guarantee that important subgroups, when defined as strata, are well represented in the sample, resulting in statistically efficient domain estimators.
iii. It can be operationally or administratively convenient.
i. can protect against selecting a 'bad' sample.
ii. It allows different sampling frames and procedures to be applied to different strata (e.g., SRS in one stratum, PPS in another).

Disadvantages of stratified sampling

- It requires that the sampling frame contain high quality auxiliary information for all units on the frame, not just those in the sample, that can be used for stratification.

- Frame creation is more costly and complex than for SRS or SYS, since the frame requires good auxiliary information.
- It can result in a sampling strategy that is less statistically efficient than SRS for survey variables that are not correlated to the stratification variables.
- Estimation is slightly more complex than for SRS or SYS.

estimators

Estimators can be computed for the stratified mean, population total, and standard errors.

- Population mean $\bar{X}_{st} = \sum_{h=1}^{H}(\frac{N_h}{N})\bar{x}_h$. Where ;
  H: number of strata, $\bar{x}_h$ is the sample mean for strata h, N_h the number of elements in the population for stratum h. where $N_h/N = W_h$ which is the stratum weight.
- Standard error of the population mean $S_{\bar{X}_{st}} = \sqrt{\frac{1}{N^2}\sum N_h(N_h - n_h)\frac{S_h^2}{n_h}}$
- Confidence interval of the population mean: $\bar{X}_{st} \pm ZS_{\bar{x}_{st}}$
- Estimator of the population total $\hat{\hat{X}} = NS_{\bar{x}_{st}}$

sample size determination

With stratified simple random sampling we can think of choosing a sample size as a two step process. First, a total sample size *n* must be chosen. Second, we must decide how to assign the sampled units to the various strata. Alternatively, we could first decide how large a sample to take in each stratum and then sum the stratum sample sizes to obtain the total sample size. This can be done using;

a) Neyman allocation. The sample size for each stratum is obtained using $n_h = n[\frac{N_h S_h}{\sum N_h S_h}]$

b) Proportional allocation. The sample size is established using $n_h = n[\frac{N_h}{N}]$

Systematic sampling (SYS)

Systematic sampling is often used as an alternative to simple random sampling. In some sampling situations, especially those with large populations, it can be time-consuming to select a simple random sample by first finding a random number and then counting or searching through the frame until the corresponding element is found.

Systematic sampling A method of choosing a sample by randomly selecting the first element and then selecting every *k*th element thereafter.

SYS has a number of **advantages** depending on the circumstances and objective of the survey:

- It is a proxy for SRS when there is no frame.
- It does not require auxiliary frame information, like SRS.
- It can result in a sample that is better dispersed than SRS (depending on the sampling interval and how the list is sorted).
- It has a well-established theory, just like SRS, and so estimates can be easily calculated.
- It is simpler than SRS since only one random number is required.

The **disadvantages** of SYS are:

- It can result in a 'bad' sample if the sampling interval matches some periodicity in the population.

- Like SRS, it does not use any auxiliary information that might be available on the frame, and thus it can result in an inefficient sampling strategy.
- The final sample size is not known in advance when a conceptual frame is used.
- It does not have an unbiased estimator of the sampling variance. In order to do variance estimation, the systematic sample is often treated as if it were a simple random sample. This is only appropriate when the list is sorted randomly.
- It can lead to a variable sample size if the population size, N, cannot be evenly divided by the desired sample size, n (but this can be avoided using circular SYS).

CLASSIFICATION OF SYSTEMATIC SAMPLING DESIGN

a) Circular systematic sampling design. Where elements are arranged in a circular format before selection of the required sample size.
b) Linear systematic sampling design. Elements are arranged in a linear manner before selection of the required sample size.