

I. Cluster Sample: A simple random sample in which each sampling unit is a collection (cluster) of listing units.

A. Cluster sampling is useful when:

1. a frame of all enumeration units is not available (feasibility).
2. cost per observation increases as the distance between numeration units increases (economy).

B. Single -Stage Cluster Sampling: All listing units in the chosen clusters are selected.

Example clusters: city blocks
listing unit: households (observe all)
elementary unit: person

C. Two-Stage Cluster Sampling: First selecting a simple random sample of clusters and then selecting a simple random sample of listing units from each sampled cluster.

Example at the second stage, select a SRS of households in the Previous example.

1. Simple 2-stage cluster sampling \Rightarrow the same fraction of listing units selected from each cluster at the second stage by SRS.
2. Probability Proportional to Size Sampling (PPS sampling): Clusters are not selected by SRS.

Multi - Stage Cluster Sampling: Several stages of sampling.

1. Example: Immunization Surveys –
 - 1st stage - sample counties,
 - 2nd stage - sample townships,
 - 3rd stage - sample school districts,
 - 4th stage - sample schools,
 - 5th stage - sample classrooms,
 - 6th stage - sample students.
2. Primary Sampling Unit (PSU): The clusters sampled at the first stage of sampling.
3. A sampling frame is required for each stage of sampling. They are compiled from only chosen clusters.

II. Advantages of Cluster Sampling:

A. Feasibility: The only frames available are lists of clusters.

B. Economy: Listing costs and traveling costs tend to be lower for cluster sampling (the field costs are lower).

EXAMPLE - Page 227.

III. Disadvantages of Cluster Sampling:

A. Sample estimates based on cluster samples tend to have higher standard error than those based on other sampling plans, for the same sample size.

1. Clusters tend to be homogeneous within, leading to redundant sampling within clusters, and high standard errors.
2. Cluster sampling is most efficient when clusters are heterogeneous within, and homogeneous without.