CHAPTER 4: SAMPLING DESIGN

- A complete enumeration of all items in the population is known as a census inquiry.
- This method is beyond the reach of ordinary researchers, and mostly done by government.
- In many cases, undertaking a census survey is not possible.
- Sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population, technically called samples.
- **SAMPLING** is the process of selecting a number of study units from a defined study population.
  - It aims at obtaining consistent and unbiased estimates of the population parameters.

The Need for Sampling

- There could be limitations in resource (time, finance, manpower, etc.) which would make it difficult to study the whole population.
- In some cases, tests may be destructive.
- Sampling provides much quicker results than does a census.
- Sampling is the only process possible if the population is **infinite**.
- There is also an argument that the quality of a study is often better with sampling than with a census.
  - The basis of the argument is that sampling possesses the possibility of better interviewing, more thorough investigation of missing, wrong, or suspicious information, better supervision, and better processing than is possible with complete coverage.

Some Fundamental Terms

- **Population:** is the theoretically specified aggregation of survey elements from which the survey sample is selected
- **Sampling Frame:** is the list of elements from which the sample is drawn
- **Sample:** a subset or some part of a larger population
- **Sample design:** is a definite plan for obtaining a sample frame
- **Sampling:** is the process of using a small number or part of a larger population to make conclusion about the whole population.
- **Element:** is a unit from which information is collected and provides the basis of analysis

Some Fundamental Terms…(cont'd)

- **Statistic:** is a characteristic of a sample.
- **Parameter:** is a characteristic of a population.

➢ When we work out certain measurement, like mean from a sample, it is called statistic.

  – But when such measure describe the characteristic of the population, they are called parameter (s); That is, population mean (μ) is a parameter Where as the sample mean (x) is a statistic.

Steps in Sampling Design

a) Identifying relevant population & sampling unit:

- Sampling Unit: A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a geographical one such as district, kebele, village, etc., or a social unit such as family, school, etc., or it may be an individual.

  Example: if the study concerns income, then the definition of the population elements as individuals or households can make a difference.

b) Determining the method of sampling:

- Whether a probability sampling procedure or a non-probability sampling procedure has to be used is also very important.

Steps in Sampling Design…(Cont'd)

c) Securing a sampling frame:

- A list of elements from which the sample is actually drawn is important and necessary. A source list should be comprehensive, correct, reliable and appropriate.

d) Identifying parameters of interest:

- what specific population characteristics (variables and attributes) may be of interest.

e) Determining the sample size

- The determination of the sample size depends on several factors:

i) **Degree of homogeneity**:

- The simple most important factor in determining the size of the sample need for estimating a parameter is the size of the population variance.

ii) Degree of confidence:

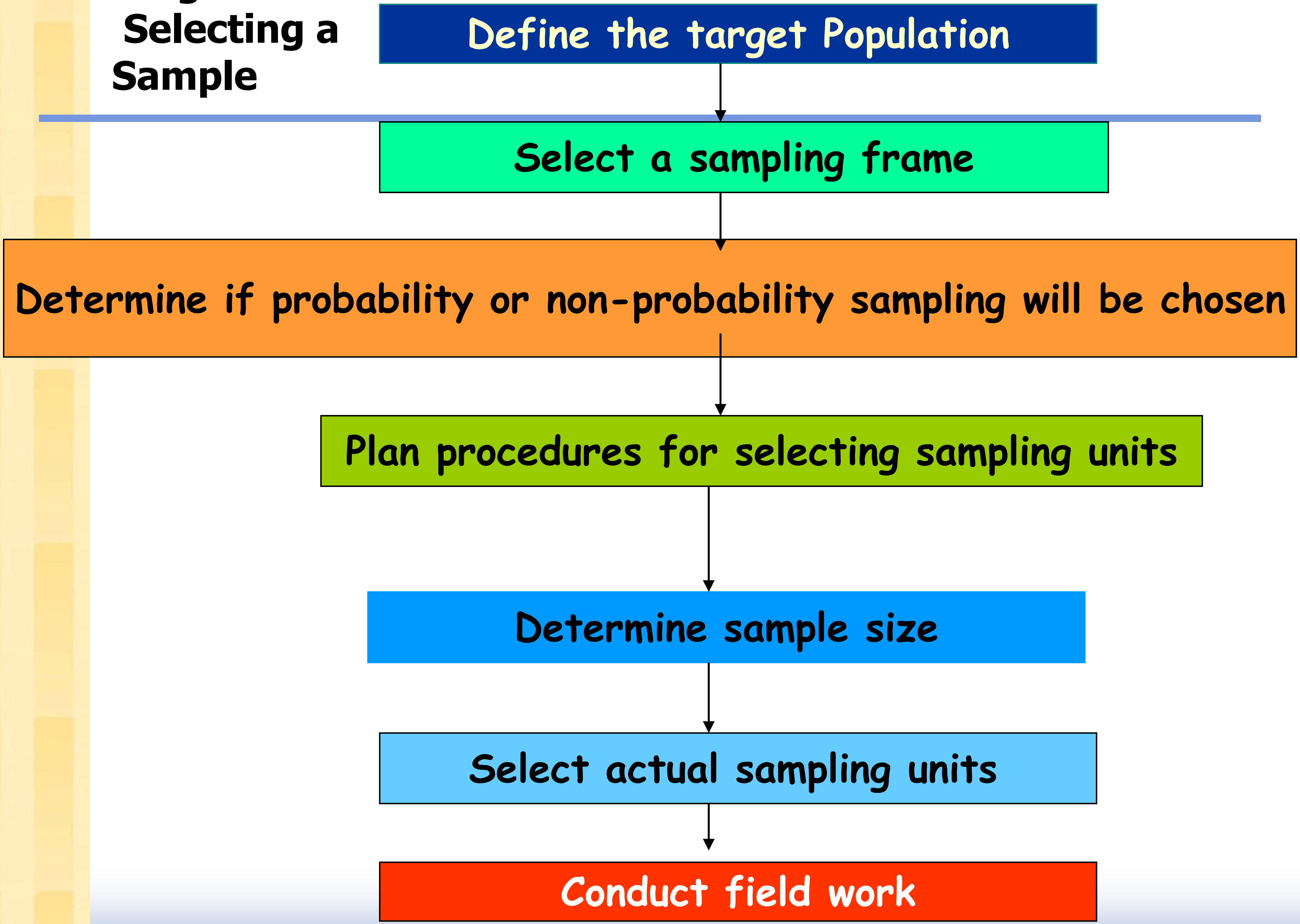
- since a sample can never reflect its population with certainty, the researcher must determine how much precision he/she needs.
- Precision is measured in terms of:
  - an interval range in which we would expect to find the parameter estimate.
  - the degree of confidence we wish to have in the estimate.

iii) Number of sub-groups to be studied:

- When the researcher is interested in making estimates concerning various sub-groups of the population then the sample must be large enough for each of these subgroups to meet the desired quality.

iv) Cost: all studies have some budgetary constraint & hence cost dictates the size of the sample.

Stages in Selecting a Sample



Sources of incorrect inference

- Three factors influence sample representativeness:
  - Sampling procedure
  - Sample size
  - Participation (response)
- There are two causes of incorrect inferences: namely, **systematic bias** and **sampling error**.

1) Systematic bias:

- Systematic bias results from errors in the sampling procedures, and it can not be reduced or eliminated by increasing the sample size.
  - Bias enters in when a sample fails to represent the population it was intended to represent.

Factors causing systematic bias

a) Inappropriate sampling frame

E.g., Using registrar's list vs. Class rosters.

b) Defective measuring device

E.g., Using questionnaire vs. interview; physical measuring device

c) Non-responses

d) Indeterminacy principle

– sometimes we find that individuals act differently when kept under observation than what they do when kept in non-observed situations.

e) Natural bias in the reporting of data

E.g., the downward bias in the income data

2) Sampling error

- Sampling errors are the random variations in the sample estimates around the true population parameters.
- It is the difference between the sample result and the result of a census conducted using identical procedures.
- It is a statistical fluctuation due to chance variations.
- Since they occur randomly and are equally likely to be in either direction, their nature happens to be of compensatory type.
- Sources of sampling error is inadequate sample size.
  – The smaller the sample, the more difficult it will be for that sample to truly capture the characteristics of a population; the larger the sample, the better. But, collecting large samples costs money and resources. In reality, there should be a balance b/n collecting extensive samples & spending a lot of money and resources.

Types of Sample Designs

- Sample designs are basically of two type:
  - Probability sampling and
  - non-probability sampling.

a) Probability Sampling Techniques

- It involves using random selection procedures to ensure that each unit of the sample is chosen on the basis of chance.
- A randomization process is used in order to reduce or eliminate sampling bias so that the sample is representative of the population from which it is drawn.
- A **basic principle of sampling** is that a sample will be representative of the population from which it is drawn if all members of the population have an equal chance of being included in the sample.
- Probability sampling requires a sampling frame (listing of all study units).

Types of probability sampling methods

- Generally speaking we could distinguish between the following types of sampling designs.
  - Simple Random Sampling Technique
  - Systematic sampling Technique
  - Stratified Sampling Technique
  - Cluster Sampling Technique.
  - Hybrid/multistage Sampling

1. Simple Random Sampling (SRS):

- The SRS is the simplest and easiest method of probability sampling.
- It is the sampling procedure in which each element of the population has an equal chance of being selected.
- It assumes that an **accurate sampling frame** exists.

Simple Random Sampling…cont'd

– SRS can only be applied in situation where the population **size is small** and **homogeneous**.

– Usually **two methods** are adopted to pick a sample.

(a) Random number table

(b) Lottery method

How to choose random sample using the random number table?

- To select the sample, each item is assigned a number from 1 to N. Then, n random numbers are selected from the table.
- To do this we select some random starting point and then a systematic pattern is used in proceeding through the table. We might start in the 7th row, 8th column and proceed down the column to the bottom of the table and then move to the top of the next column to the right.

Random Number Table

55	24	80	17	45	62	44	87	29	2	34	21	42	55	96	68	71	50	35	99
6	81	99	89	59	98	46	94	85	88	82	62	36	61	64	22	37	19	26	40
81	1	72	41	50	23	36	25	39	48	0	43	23	58	33	98	99	21	87	85
98	67	49	95	35	18	58	43	58	37	22	97	2	93	97	7	71	45	7	70
86	13	34	95	94	74	85	18	11	97	33	12	39	22	82	81	4	32	93	29
82	82	96	56	16	77	46	0	44	16	24	54	96	45	69	6	16	91	5	88
26	61	99	35	44	21	[illegible]	86	7	14	54	85	75	64	34	45	8	75	61	10
52	25	74	9	65	12	91	87	71	18	69	78	36	69	1	21	88	27	27	59
22	42	75	73	89	39	53	28	30	28	82	32	14	58	24	47	3	54	36	4
33	6	81	93	1	96	56	46	85	98	20	29	42	67	42	16	67	6	94	0
96	95	54	71	88	96	15	68	34	83	90	83	64	55	70	70	41	40	83	92
88	88	41	11	74	9	96	91	42	13	24	89	85	60	20	39	72	16	65	52
99	76	93	5	15	8	94	78	42	17	4	52	90	46	73	2	37	58	24	36
34	50	40	57	39	23	29	55	3	89	70	60	48	87	35	12	28	55	35	20
55	63	40	22	22	16	72	80	67	75	69	45	52	55	61	97	67	7	27	48
57	81	1	40	48	34	59	2	30	29	2	30	93	36	15	12	55	37	75	9
12	91	23	70	96	10	36	36	38	54	64	52	49	56	7	19	82	10	58	31
41	11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	33	97	45	94	33	[illegible]	[illegible]	[illegible]	[illegible]	54	42	40	33
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	76	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9

2. Systematic Sampling Technique

- This is also one of the most widely used probability sampling.
- In **Systematic Sampling** individuals are chosen at regular intervals (for example every fifth) from the sampling frame.
- **Steps**
  - Assign a sequence number to each member of the population.
  - Determine the skip interval by dividing the number of units in the population by the sample size. I=N/n where I is skip interval, N is population size, and n is sample size.
  - Select a starting point in a random digit table (it must be between 1 and I).
  - include that item in a sample and select every i^{th} item thereafter until total sample has been selected.

Example of Systematic Random Sample

N = 100

want n = 20

N/n = 5

select a random number from 1-5: chose 4

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Merits and Demerits of Systematic Sampling

- The **merits** of this sampling technique is:
  - the samples will spread evenly over the entire population.
  - It is also an easier
  - less costly method of sampling and can be conveniently used even in case of large populations.
- **Demerits**
  - if there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling.

3. Stratified Sampling

- The stratified sampling technique is particularly useful when we have heterogeneous populations
- Most populations can be segregated into a number of mutually exclusive sub populations or Strata.
- Sub-populations(strata) that are individually more homogeneous than the total population
- After a population is divided into the appropriate strata a simple random sample can be taken either using the SRS or the SS techniques from each stratum.

Steps in Stratified Random Sampling

1. Divide the population to be surveyed in to strata of similar study units or into areas with which similar social, environmental, or economic conditions exist.
2. Make a separate and complete list of the stratum and from each stratum draw a separate random sample of study units using these lists.
3. A similar survey is then done on the sample of study units in each of the strata i.e. the same questionnaire is used.

How to form strata?

- The strata are formed on the basis of common characteristic(s) of the items to be put in each stratum. i.e. strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata. Strata are purposively formed usually based on past experience and personal judgment of the researcher.
- **How should items be selected from each stratum?**
  - use simple random sampling, or
  - Systematic random sampling

How many items be selected from each stratum?

i) proportional to the sizes of the strata

- That is, if P_i represents the proportion of population included in stratum ***i***, and ***n*** represents the total sample size, the number of elements selected from stratum ***i*** (n_i) is $\mathbf{n*P_i}$. i.e. $n_i=\mathbf{n}(N_i/\mathbf{N})$
- For example, Suppose we want to take a sample of size n=30 to be drawn from a population of size N=800 which is divided into three strata of size $N_1=400$, $N_2=240$, and $N_3=160$. The sample size for stratum with $N_1=400$ is $n_1=30(400/800)=15$.
- The sample size for stratum with $N_2=240$ is $n_2=30(240/800)=9$.
- The sample size for stratum with $N_3=160$ is $n_3=30(160/800)=6$.

ii) Disproportionate sampling design

- When we account for both differences in stratum size and differences in stratum variability
- Formula: $$n_i = \frac{n.N_i.\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \ldots + N_k\sigma_k}$$
- For example, assume a population is divided into three strata so that N_1=5000, N_2=2000, and N_3=3000.Respective standard deviations are

$$\sigma_1 = 15, \sigma_2 = 18, \sigma_3 = 5$$

- How should a sample of size n=84 be allocated to the three strata , if we want optimum allocation using disproportionate sampling design?

$$n_1 = \frac{84(5000)(15)}{(5000)(15) + (2000)(18) + (3000)(5)} = 50$$

Pros and Cons of stratified sampling

- **Pros of stratified sampling:**

1) more reliable information is obtained for the same sample size if the population is stratified than they are for the population as a whole.
2) Comparisons between strata are easy.

- **Drawbacks to using stratified sampling:**

1) First, sampling frame of entire population has to be prepared separately for each stratum
2) Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata.
3) Finally, stratified sampling can potentially require a larger sample than would other methods.

4. Cluster sampling

- **This is an example of two stage sampling.**
  - **First stage a sample of areas is chosen; Second stage a sample of respondents within those areas is selected.**
  - **Divide population into a large number of groups, called clusters and then sample among clusters. Finally select all individuals within those clusters.**
- **Clusters are often geographic units (e.g., districts, villages) or organizational units (e.g., clinics, training groups, etc).**
- **Cluster sampling addresses two problems:**
  - **Researchers lack a good sampling frame for a dispersed population.**
  - **The cost to reach a sample element is very high and cluster sampling reduces cost by concentrating surveys in selected clusters.**
- **But certainly it is *less precise* than simple random sampling.**

Cluster Sampling.. Cont'd

Advantages :

- Cuts down on the cost of preparing a sampling frame.
- This can reduce travel and other administrative costs.

Disadvantages:

- sampling error is higher for a simple random sample of same size.

Differences between stratified sampling and cluster sampling

Stratified sampling	Cluster sampling
• We divide the population into a few subgroups, each with many elements in it. • The sub-groups are selected according to some criterion that is related to the variables under study.	• We divide the population into many subgroups, each with a few elements in it. • The sub-groups are selected according to some criterion of ease or availability in data collection.
• We try to secure homogeneity within sub-groups & heterogeneity b/n subgroups.	• We try to secure heterogeneity within subgroups and homogeneity b/n sub-groups, but we usually get the reverse.
• We randomly choose elements from within each sub-group.	• We randomly choose a number of sub-groups, which we want to study about.

Sampling with probability proportional to the cluster size

- In case the cluster sampling units do not have the same number of elements, it is considered appropriate to use a random selection process where the probability of each cluster being included in the sample is proportional to the size of the cluster.
- For this purpose, we have to list the number of elements in each cluster irrespective of the method of ordering the cluster.
- Then we must sample systematically the appropriate number of elements from the cumulative totals.

Multi-stage random sampling

- Complex form of cluster sampling in which two or more levels of units are embedded one in the other.
- Sampling plan that uses a combination of sampling methods in various stages
- First stage, random number of districts chosen in all states.
- Followed by random number of villages.
- Then third stage units will be houses.
- All ultimate units (houses, for instance) selected at last step are surveyed.

b) Non-probability Sampling Techniques

- **Non-probability** selection is non random i.e., each member does not have a known non-zero chance of being included.
- While useful for many studies, non-probability sampling procedures provide only a weak basis for generalization. In reality, the conclusions drawn from a study of a non-probability sample are limited to that sample and cannot be used for further generalization.
- The followig are the types non-probablity sampling methods

1) Convenience Sampling

- Sometimes known as **grab** or **opportunity sampling or accidental or haphazard sampling.**
- A type of non-probability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, readily available and convenient.
- This type of sampling is most useful for pilot testing.
- The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.
  - **Example**: the person on street and interview conducted regarding television programs are examples of convenient samples.

2) Purposive or Judgment sampling

- Purposive sampling occurs when one draws a non-probability sample that conforms to a certain criteria.
- The researcher chooses the sample based on who they think would be appropriate for the study.
  - When focusing on a limited number of informants, whom we select *strategically* so that their in-depth information will give optimal insight into an issue is known as *purposeful sampling.*
- It uses the judgment of the expert in selecting cases.
- Purposeful sampling should **not** be **haphazard**.
  - Care should be taken that for different categories of informants; **selection rules** are developed to prevent the researcher from sampling according to personal preference.

3) Quota Sampling

- Quotas are assigned to different strata group and interviewers are given quotas to be filled from different strata.
- But, the actual selection of the items is left to the interviewers discretion.
- In quota sampling a researcher first identifies categories of people (e.g., male, female) then decides how many to get from each category.
  - **Example**: the researcher interviews the first 5 male and the first 5 females that he or she encounters.
- The major limitation of this method is the absence of an element of randomization.
  - Consequently the extent of sampling error cannot be estimated.
- In spite of its limitations, quota sampling is used in opinion pollsters, marketing research and other similar research areas.

4) Snowball /Network/Referral/Sampling

- This is a method for identifying and sampling (or selecting) the cases in a network.
- Snowball sampling is based on an analogy to a snowball, which begins small but becomes larger as it is rolled on wet snow and pick up additional snow.
- Snowball sampling begins with one or a few people or cases and spread out on the basis of links to the initial case.
- You start with one or two information-rich key informants and ask them if they know persons who know a lot about your topic of interest.