SAMPLES AND SAMPLING

Definition of SAMPLE

-- a **representative part** or a **single item** from a larger whole or group especially when presented for inspection or shown as evidence of quality.

-- a **finite part** of a statistical population whose properties are studied to gain information about the whole.

Definition of SAMPLING

-- the act, process, or technique of **selecting a representative part** of a population for the purpose of determining parameters or characteristics of the whole population.

The sampling process comprises several stages:

1. **Defining** the population of concern
2. **Specifying a sampling frame**, a set of items or events possible to measure
3. **Specifying a sampling method** for selecting items or events from the frame
4. Determining the **sample size**
5. **Implementing** the sampling plan
6. **Sampling** and **data collecting**

Population

-- Successful statistical practice is based on **focused problem definition.**

In sampling, this includes **defining the population** from which our sample is drawn.

-- **A population can be defined as including all people or items with the characteristic one wishes to understand.**

-- The goal becomes finding a **representative sample** (or subset) of that population.

Sampling frame

-- In the most straightforward case, it would be most desirable to identify and **measure every single item** in the population and to include **any one of them in our sample.**

-- However, in the more general case this is **not usually possible or practical.**

-- As a remedy, we seek a **sampling frame** which has the property that **we can identify every single element** and **include any in our sample**.

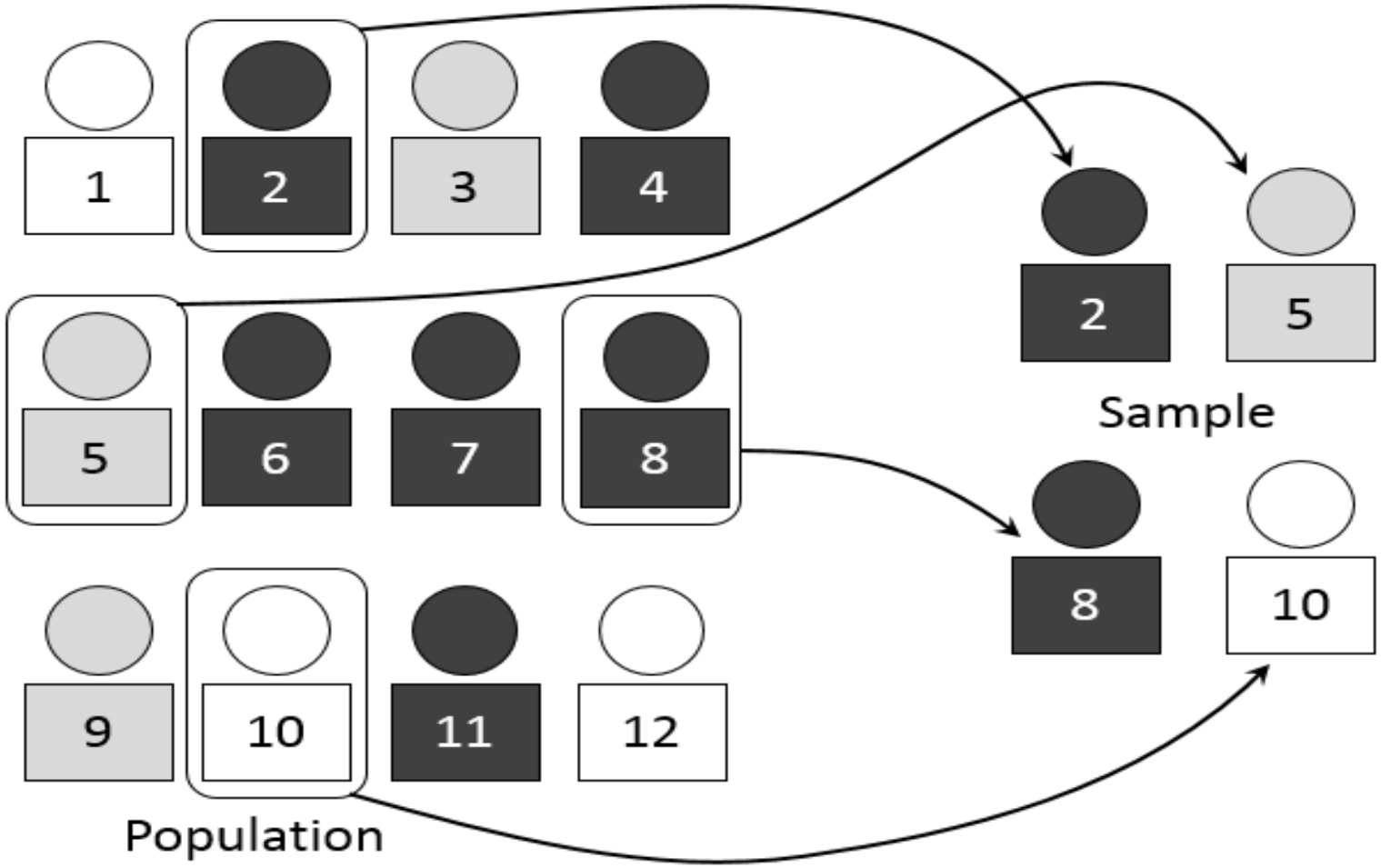
Sampling methods

Simple random sampling

-- In a **simple random sample (SRS)** of a given size, all subsets are given an **equal probability**.

-- Furthermore, any given pair of elements has the **same chance of selection** as any other such pair.

-- This **minimizes bias** and **simplifies analysis of results**.



simple random sampling

Systematic sampling

-- **Systematic sampling** (also known as **interval sampling**) relies on arranging the study population according to some **ordering scheme** and then selecting elements at **regular intervals** through that ordered list.

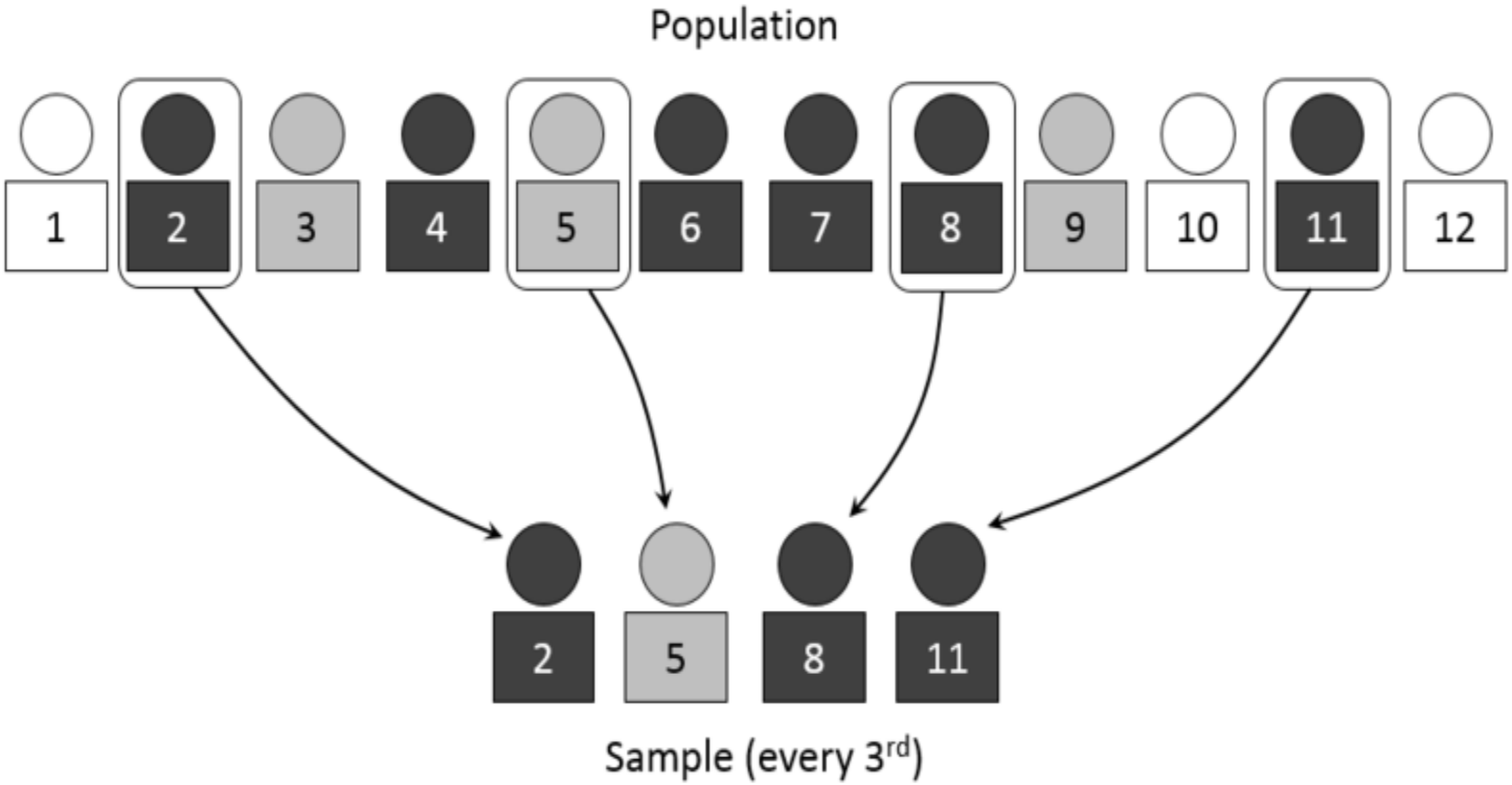
-- Systematic sampling involves a **random start** and then proceeds with the selection of **every *k*th element** from then onwards.

In this case, k= (population size/sample size).

-- It is important that the **starting point is not automatically the first in the list,** but is instead **randomly chosen** from within the first to the *k*th element in the list.

A simple example would be to start from say 40th , 55th or 70th name from a telephone directory and then select every next 10th name.

-- As long as the starting point is **randomized**, systematic sampling is a type of **probability sampling**.

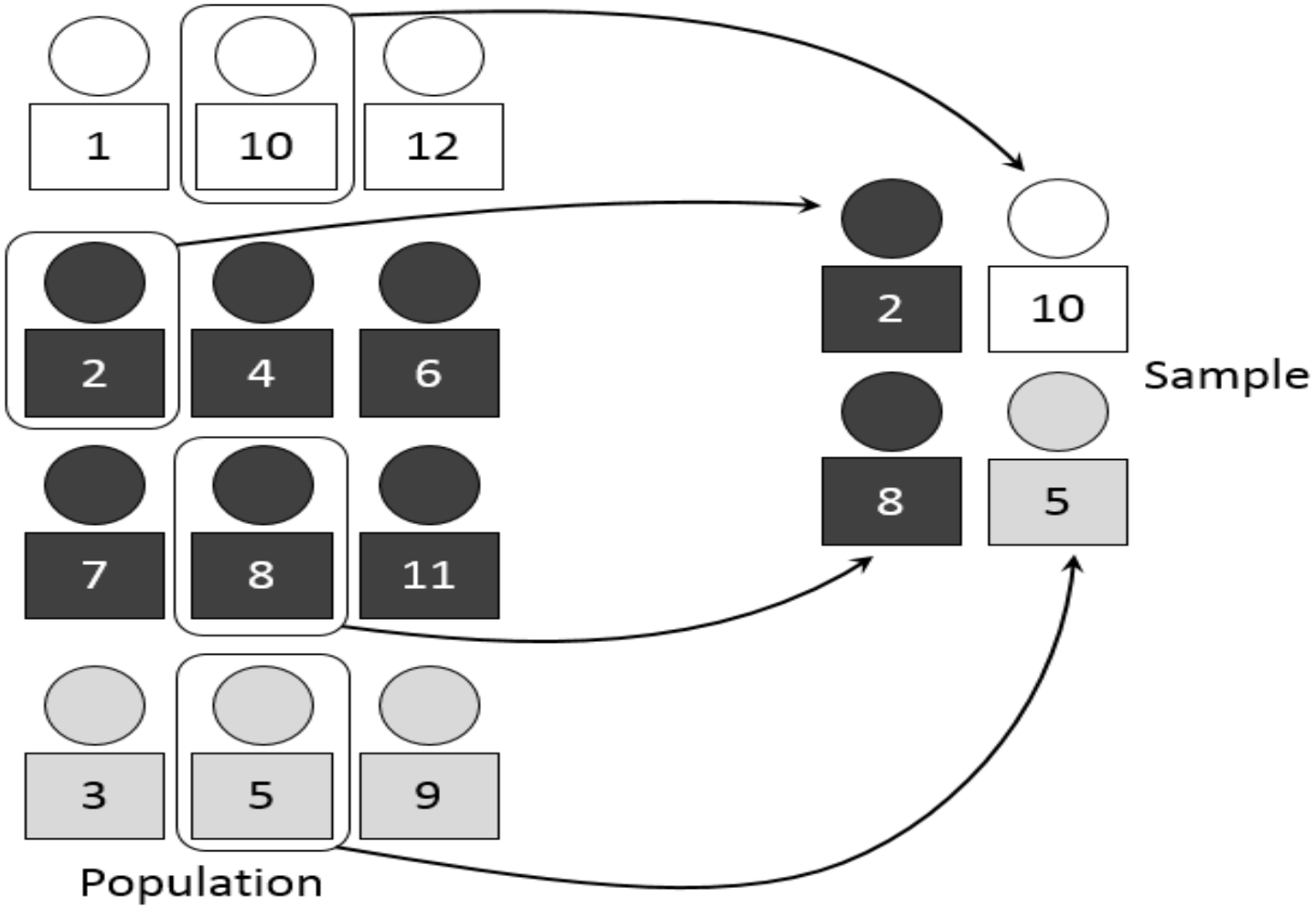


Selecting a random sample using systemic sampling technique

Stratified sampling

-- When the population has a number of distinct categories, the frame can be organized by these categories into **separate "strata."**

-- Each stratum is then sampled as **an independent sub-population,** out of which individual elements can be randomly selected.



Random sampling using stratified random sampling technique

Advantages over other sampling methods:

1. Focuses on **important subpopulations** and **ignores irrelevant ones.**

2. Allows use of **different** sampling techniques for different subpopulations.

3. Improves the **accuracy/efficiency** of estimation.

4. Permits greater balancing of statistical power of tests of differences between strata by sampling **equal numbers** from strata varying widely in size.

Disadvantages

-- Is not useful when there are no **homogeneous subgroups**.

-- Can be **expensive** to implement.

Cluster sampling

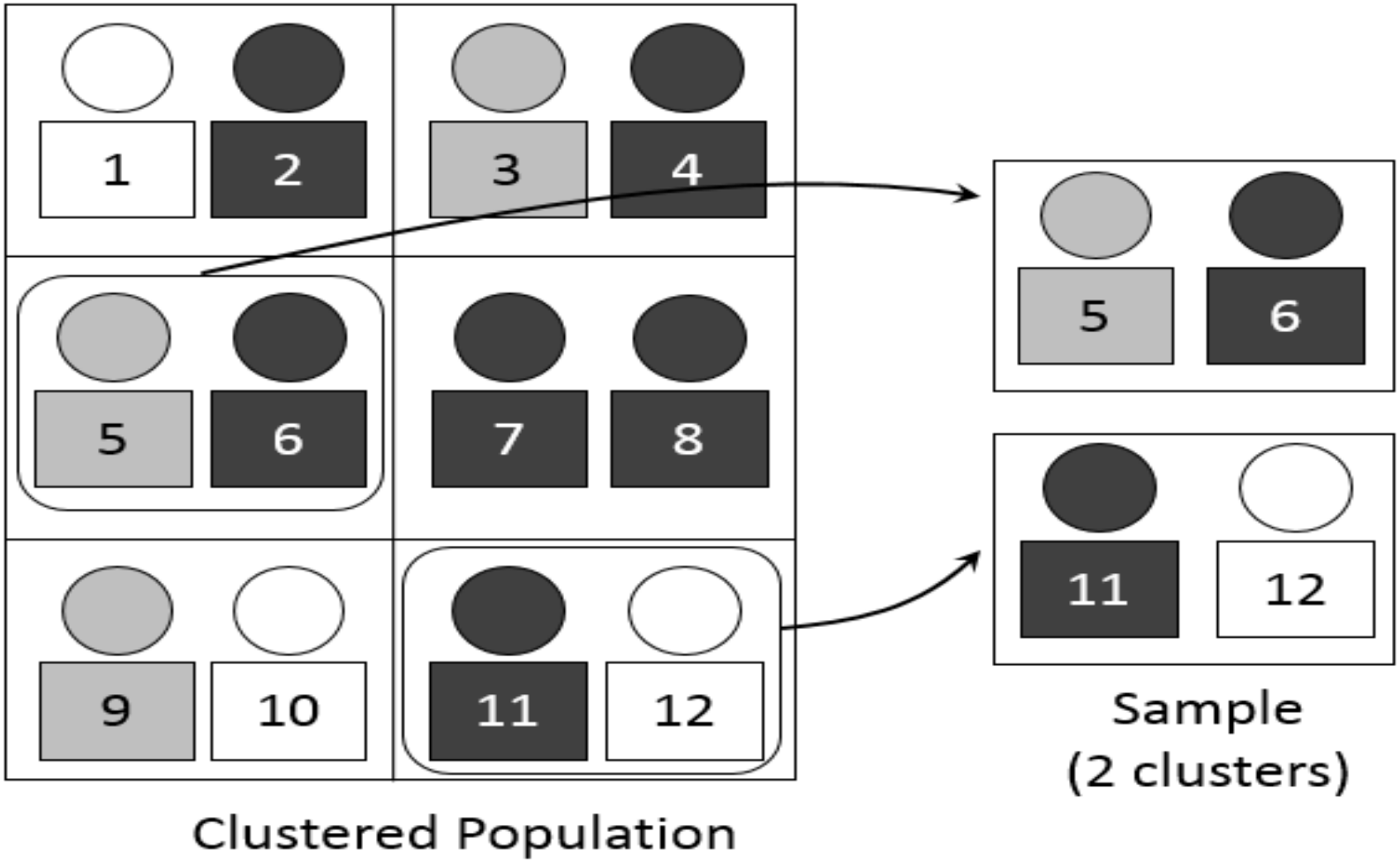
-- Sometimes it is more **cost-effective** to select respondents in groups ('clusters').

-- Sampling is often clustered by geography, or by time periods.

For instance, if surveying households within a city, we might choose to select 100 city blocks and then interview every household within the selected blocks.

-- Clustering can **reduce travel** and **administrative costs**.

In the example above, an interviewer can make a single trip to visit several households in one block, rather than having to drive to a different block for each household.



Random sampling by cluster sampling technique

Accidental sampling

-- Accidental sampling (sometimes known as **convenience** or **opportunity sampling**) is a type of **nonprobability sampling.**

-- A population is selected because **it is readily available and convenient.**

-- It may be through meeting the person or including a person in the sample when one meets them or chosen by finding them through technological means such as the internet or through phone.

-- The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because **it would not be representative enough**.

-- This type of sampling is most useful for **pilot testing**.

Panel sampling

-- **Panel sampling** is the method of first selecting a group of participants through a **random sampling method** and then asking that group for (potentially the same) information **several times over a period of time**.

-- Therefore, each participant is **interviewed at two or more time points**; each period of data collection is called a **"wave".**

-- This **longitudinal** sampling method allows estimates of changes in the population, for example with regard to chronic illness to job stress to weekly food expenditures.

-- Panel sampling can also be used to inform researchers about **within-person health changes** due to age or to help explain changes in continuous dependent variables such as spousal interaction.

Snowball sampling

-- Snowball sampling involves finding a **small group of initial respondents** and **using them to recruit more respondents.**

-- It is particularly useful in cases where the **population is hidden** or **difficult to enumerate**.

Errors in sample surveys

-- Survey results are typically subject to some **error.**

-- Total errors can be classified into **sampling errors** and **non-sampling errors**.

-- The term **"error"** here includes **systematic biases** as well as **random errors**.

Sampling errors and biases

Sampling errors and biases are induced by the **sample design**.

They include:

Selection bias: When the true selection probabilities differ from those assumed in calculating the results.

Random sampling errors: Random variation in the results due to the elements in the sample being selected at random.

Non-sampling error

Non-sampling errors are other errors which can impact the final survey estimates, caused by problems in data collection, processing, or sample design. They include:

1. **Over-coverage**: Inclusion of data from outside of the population.
2. **Under-coverage**: Sampling frame does not include elements in the population.
3. **Measurement error**: e.g. when respondents misunderstand a question, or find it difficult to answer.
4. **Processing error**: Mistakes in data coding.
5. **Non-response**: Failure to obtain complete data from all selected individuals.

After sampling, a review should be held of the **exact process** followed in sampling, **rather than that intended**.

A particular problem is that of **non-response**.

Two major types of non-response exist:

1. **Unit nonresponse** (referring to lack of completion of any part of the survey)

2. **Item non-response** (submission or participation in survey but failing to complete one or more components/questions of the survey).


Thank You!