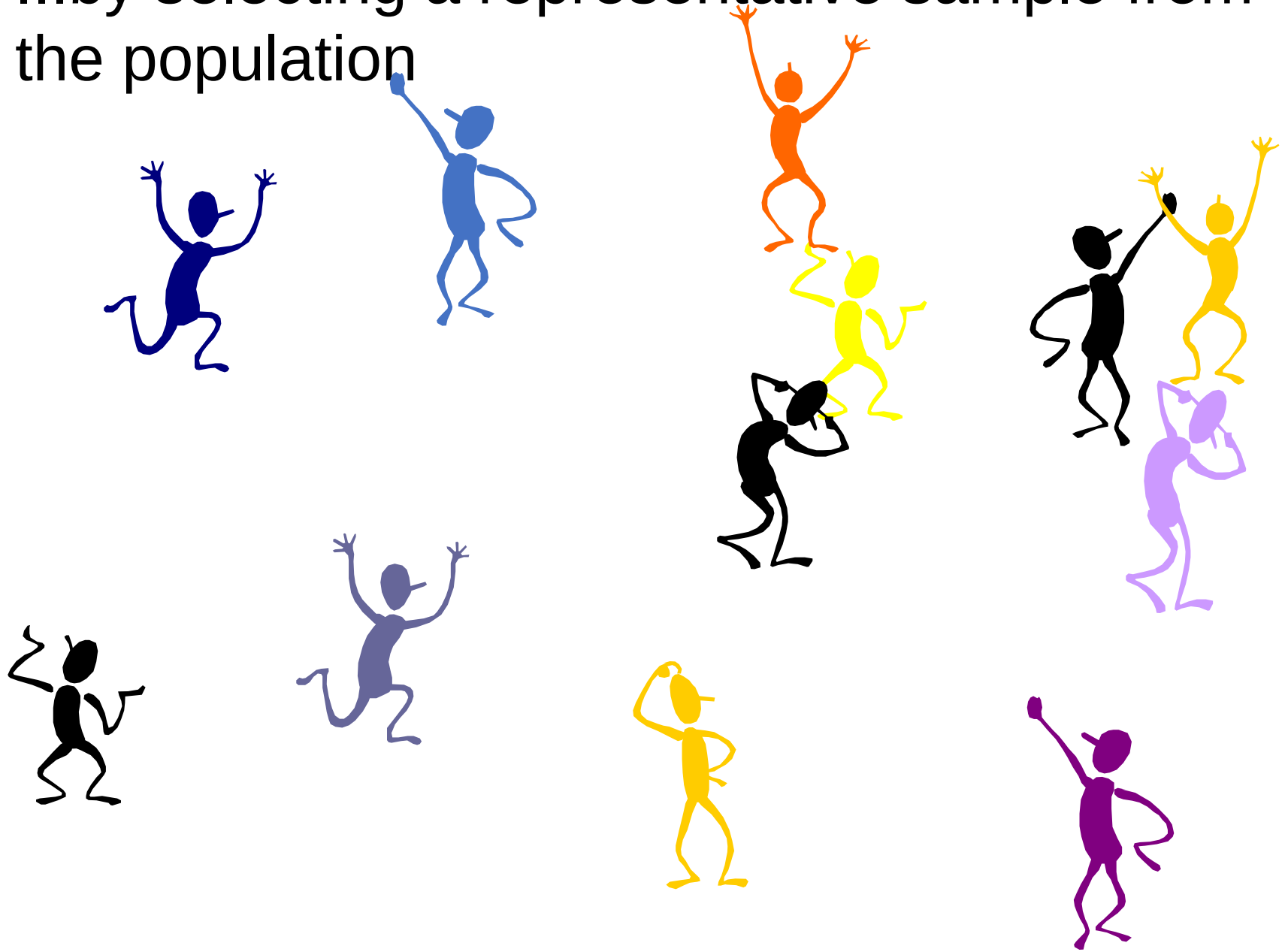

Sampling
Techniques

Aims of Sampling

- Reduces cost of research (e.g. political polls)
- Generalize about a larger population (e.g., benefits of sampling city r/t neighborhood)
- In some cases (e.g. industrial production) analysis may be destructive, so sampling is needed

Population inferences can be made...

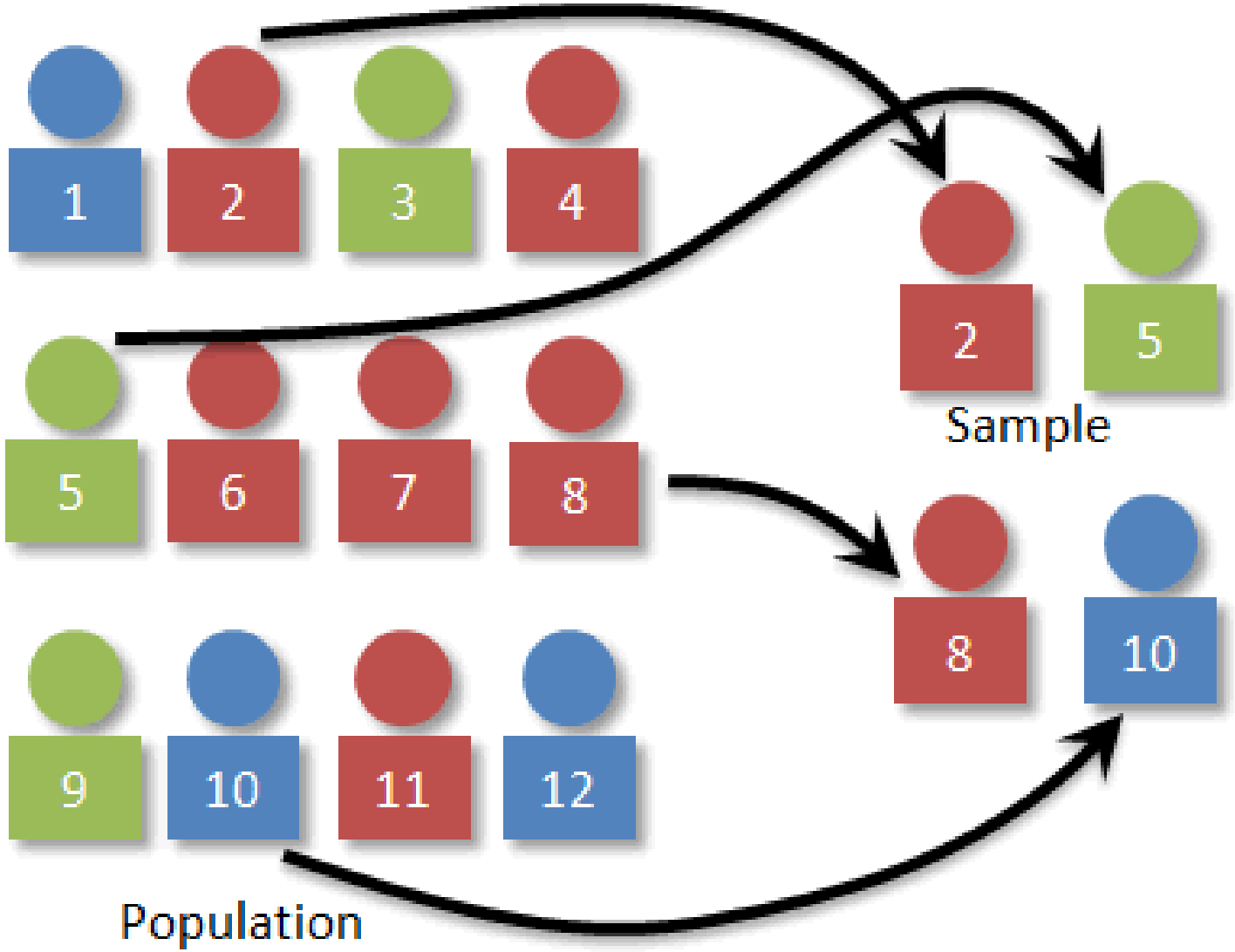
...by selecting a representative sample from the population



Types of Random Sampling

Random Sampling

- **Simple Random Sample** – A sample designed in such a way as to ensure that (1) every member of the population has an equal chance of being chosen and (2) every combination of N members has an equal chance of being chosen.
- This can be done using a computer, calculator, or a table of random numbers


1
2
3
4
5
6
7
8
9
10
11
12
2
5
Sample
8
10
Population

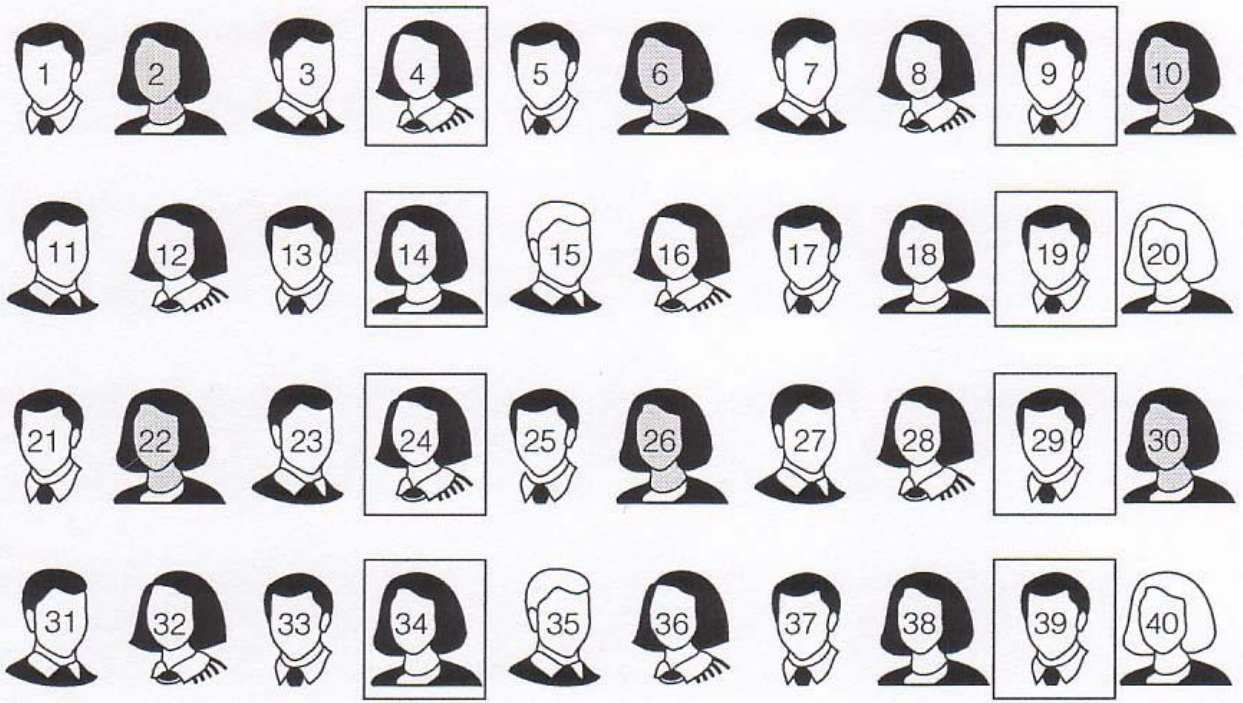
Random Sampling

- **Systematic random sampling** – A method of sampling in which every *K*th member (*K* is a ration obtained by dividing the population size by the desired sample size) in the total population is chosen for inclusion in the sample after the first member of the sample is selected at random from among the first *K* members of the population.

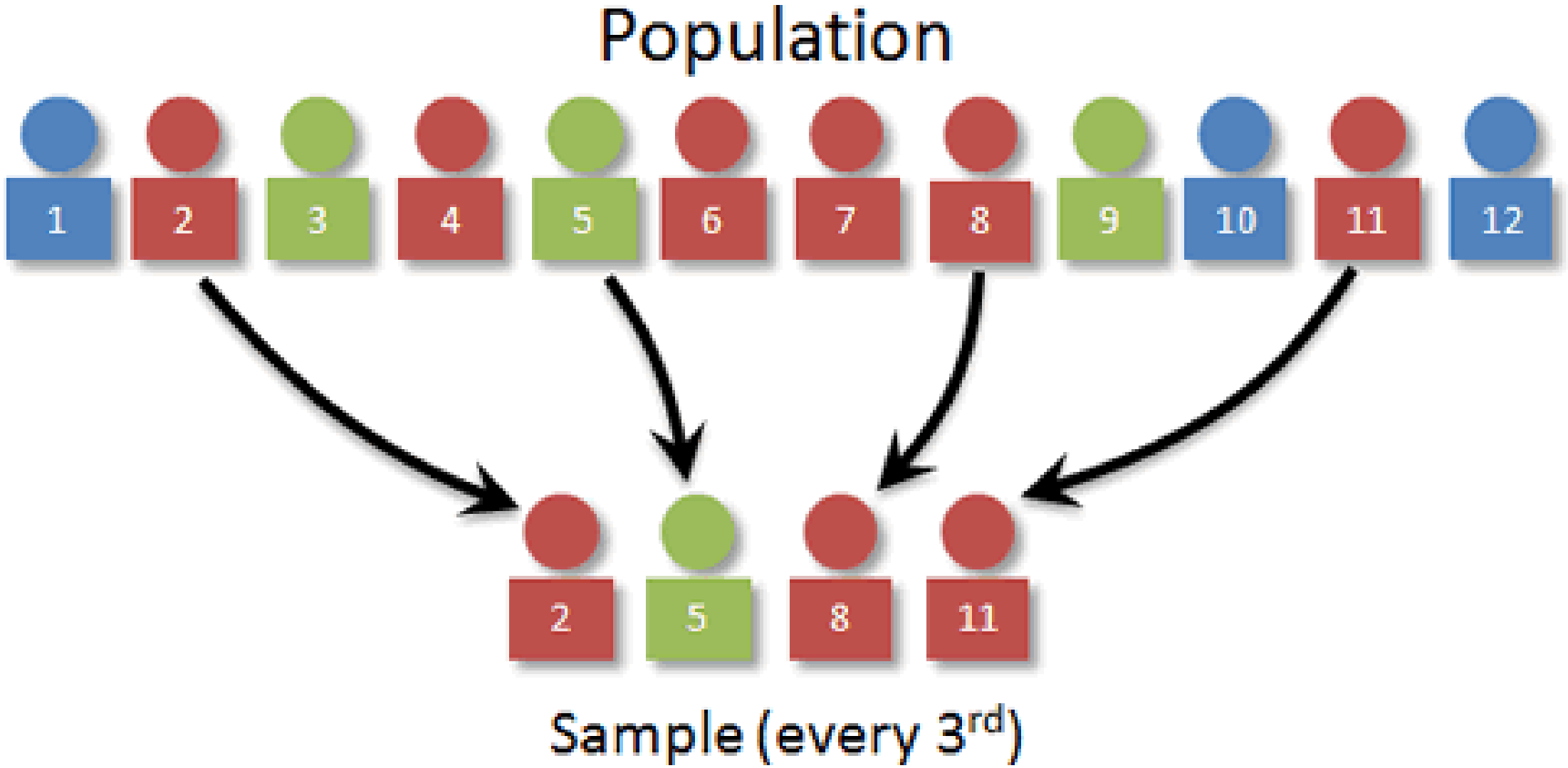
Systematic Random Sampling

Figure 11.2 **Systematic Random Sampling**

From a population of 40 students, let's select a systematic random sample of 8 students. Our skip interval will be 5 (40 ÷ 8 = 5). Using a random number table, we choose a number between 1 and 5. Let's say we choose 4. We then start with student 4 and pick every 5th student:



Our trip to the random number table could have just as easily given us a 1 or a 5, so all the students do have a chance to end up in our sample.


Population
1
2
3
4
5
6
7
8
9
10
11
12
2
5
8
11
Sample (every 3rd)

Stratified Random Sampling

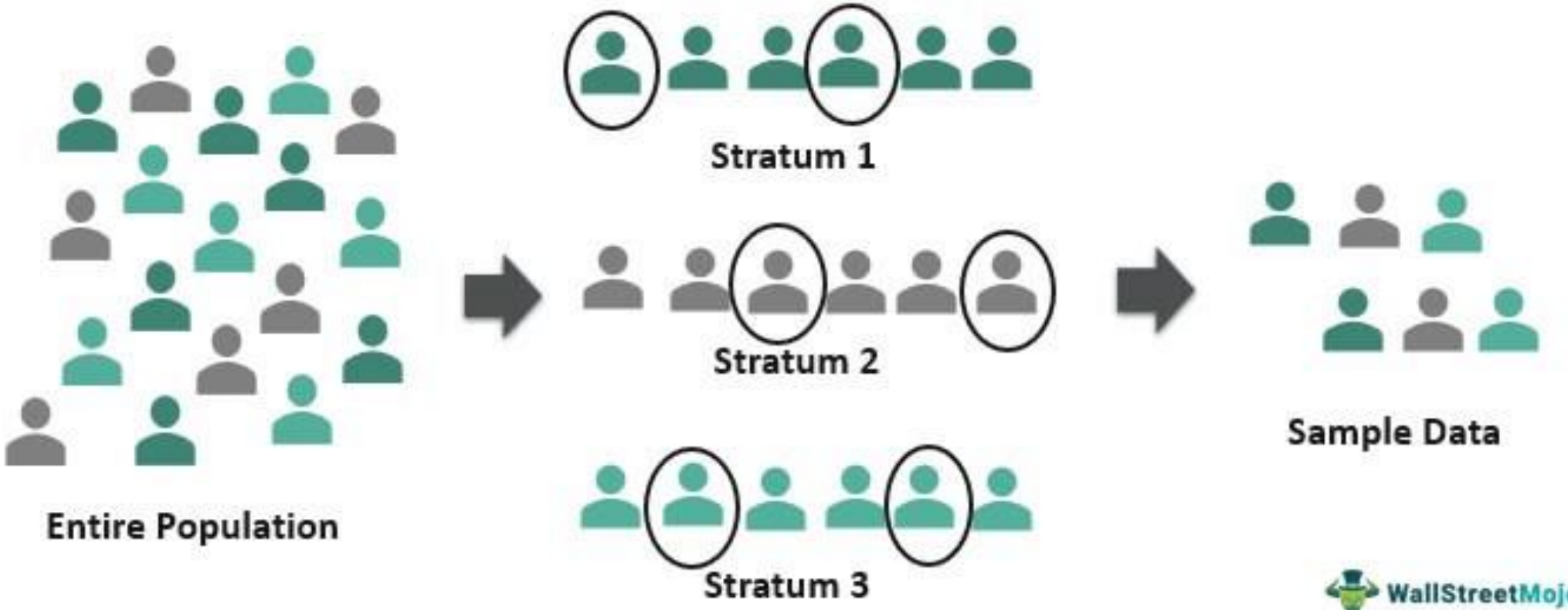
- **Stratified random sample** – A method of sampling obtained by

(1) dividing the population into subgroups based on one or more variables central to our analysis and

(2) then drawing a simple random sample from each of the subgroups

Stratified Sampling

Stratified sampling, refers to random sampling techniques that clubs items of whole population into different groups called strata, based on their similar characteristics. Then, samples from each stratum are taken, whether proportionately or disproportionately.

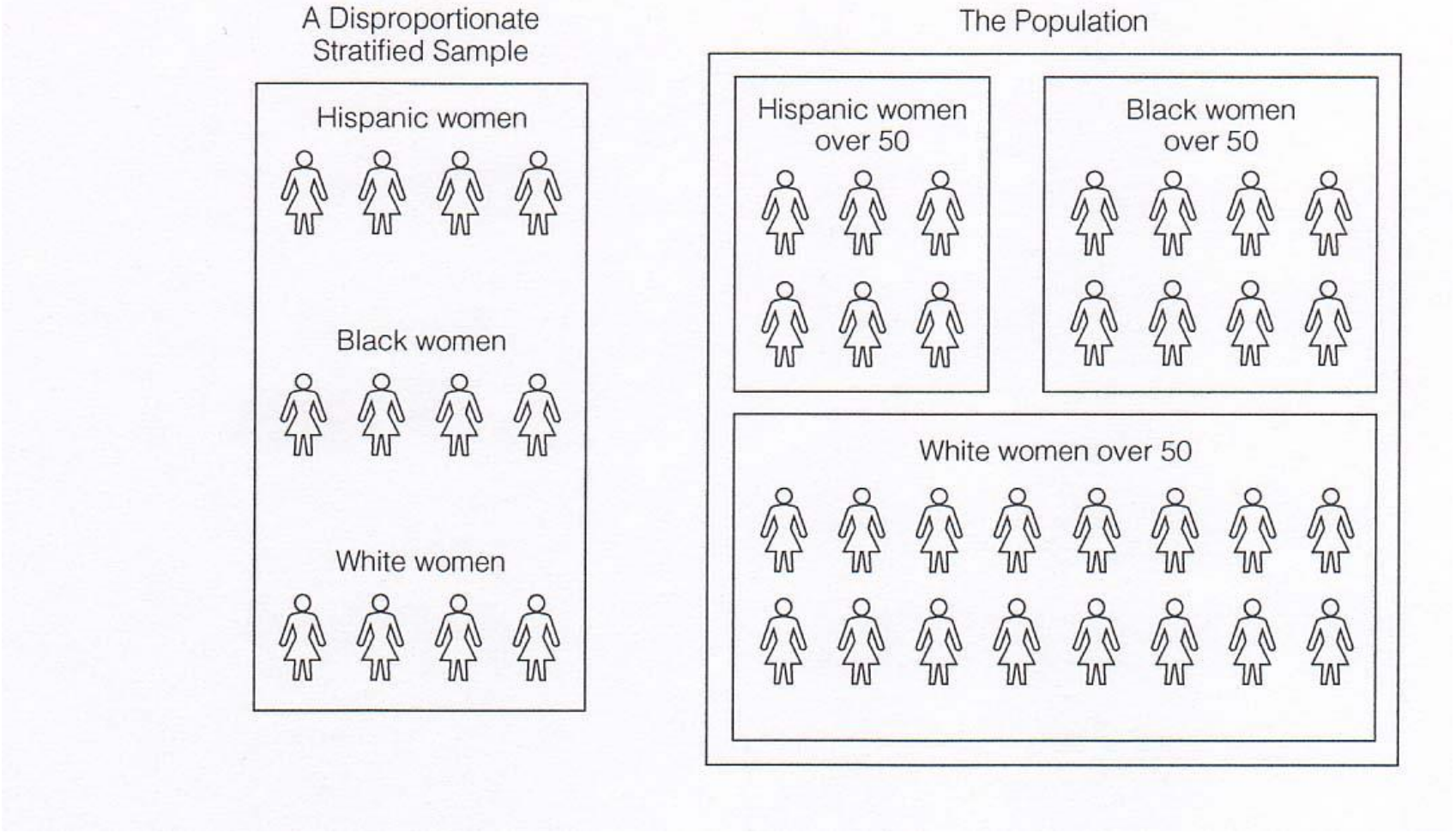


Stratified Random Sampling

- **Proportionate stratified sample** – The size of the sample selected from each subgroup is proportional to the size of that subgroup in the entire population. (Self weighting)
- **Disproportionate stratified sample** – The size of the sample selected from each subgroup is disproportional to the size of that subgroup in the population. (needs weights)

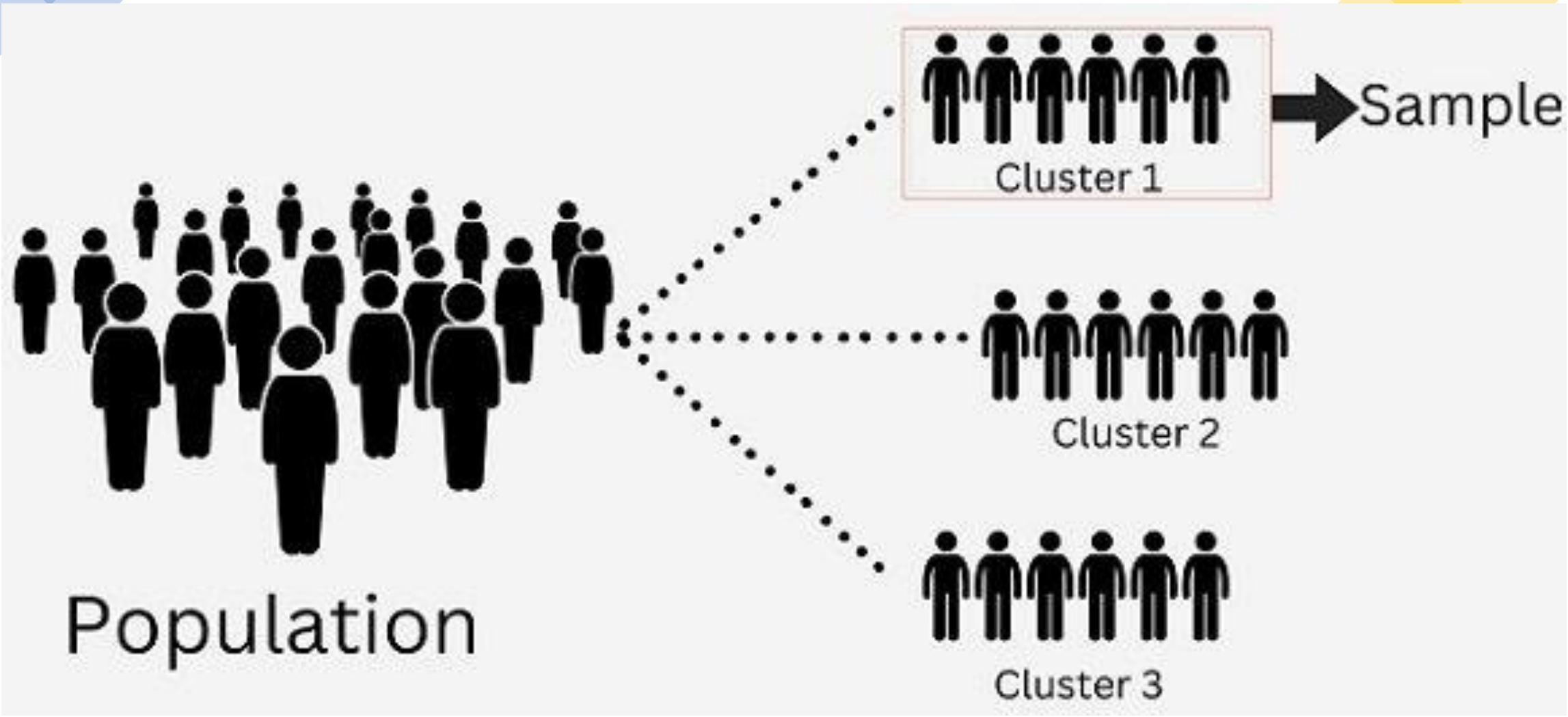
Disproportionate Stratified Sample

Figure 11.3 **A Random Sample Stratified by Race/Ethnicity**

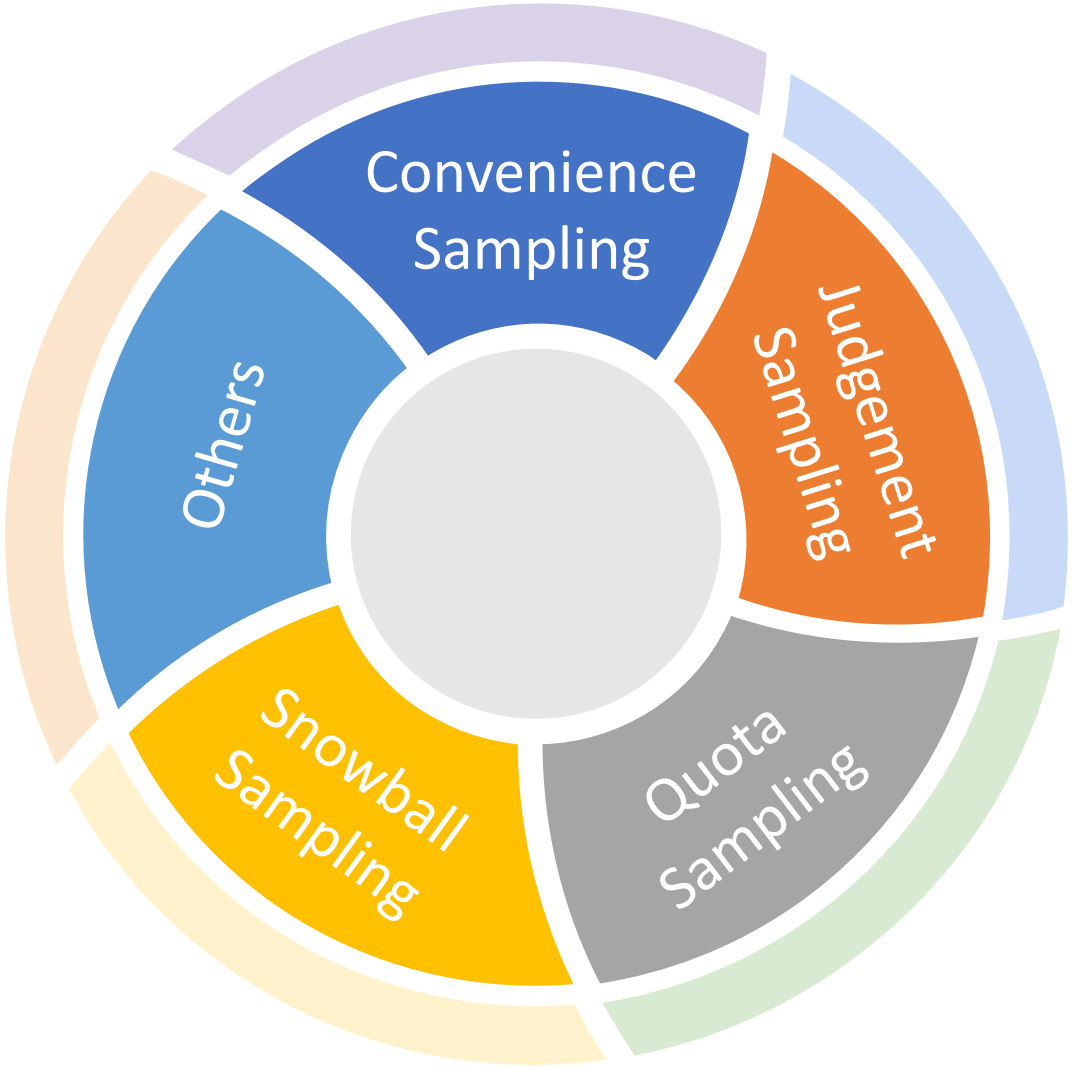


Cluster Random Sampling

- **Cluster random sample** – A method of sampling obtained by dividing population into groups called clusters. Then clusters are taken in random and all elements of cluster are taken as sample.
- Best results can be achieved if elements of clusters are as little alike as possible.
- Clusters represent small scale version of entire population.
- Can reduce the cost of sampling.


Cluster 1
Sample
Cluster 2
Population
Cluster 3

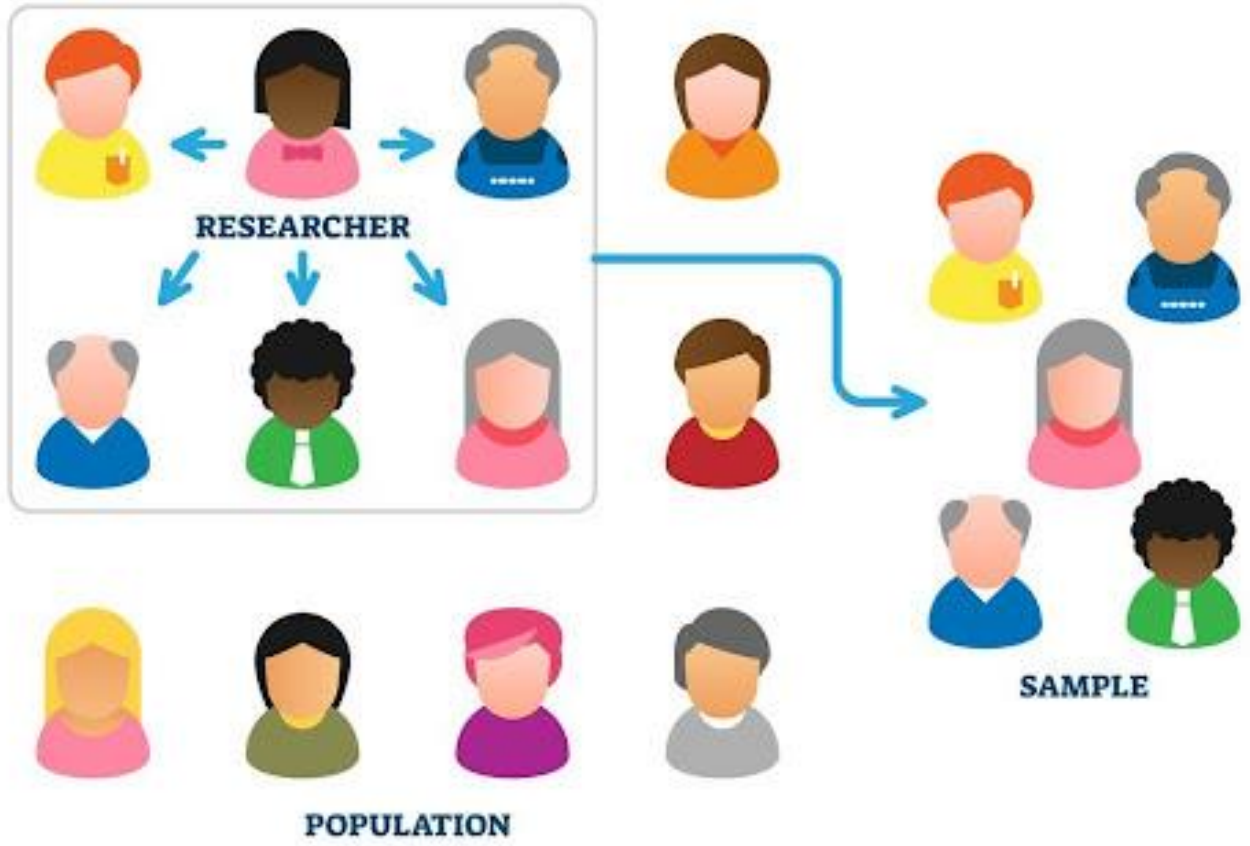
Types of Non-Random Sampling



Non-Random Sampling

- Convenience sampling
  - Not probability sampling.
  - Example: for medical studies, voluntaries are used.
  - It can not be guaranteed that voluntaries will represent whole population.
  - Easy to perform but results should be interpreted carefully.

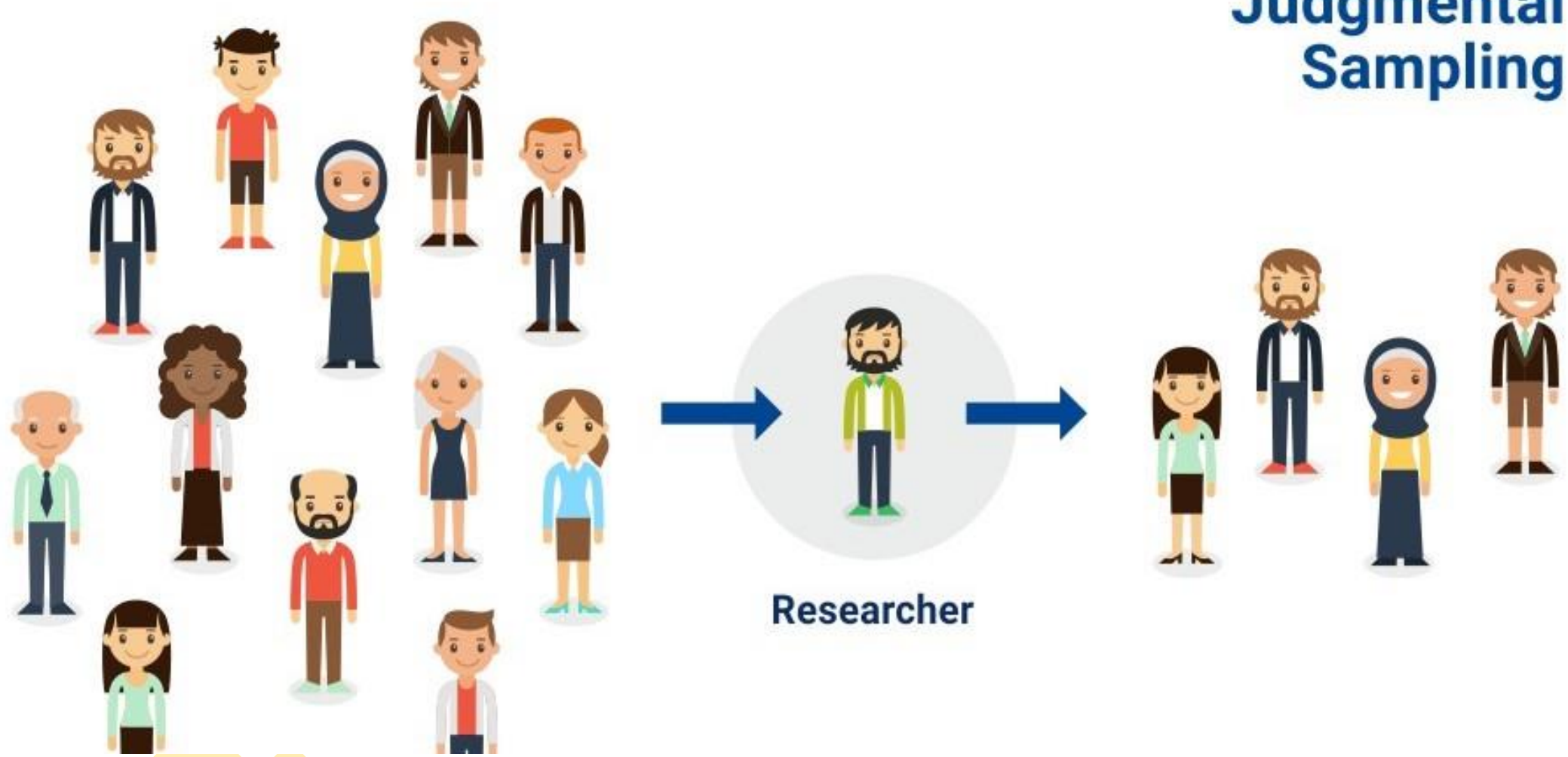
CONVENIENCE SAMPLING



Non-Random Sampling

- Judgment sampling
  - Persons who are most knowledgeable in the field select items for study.
  - Easy to perform but results again should be interpreted and used carefully.

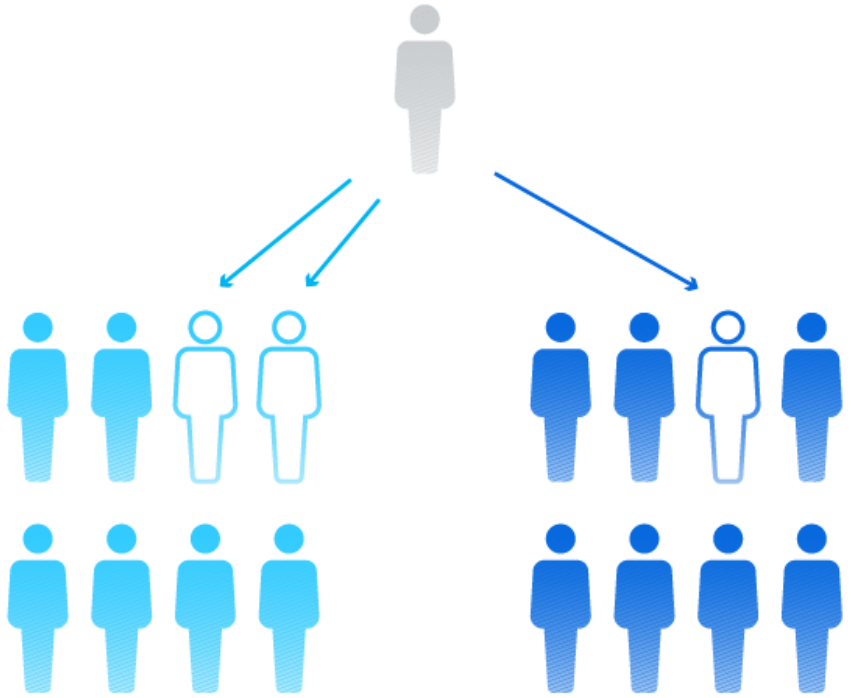
Judgmental Sampling



Non-Random Sampling

- Quota sampling
  - The makeup of the sample must reflect the makeup of the population on some selected characteristics.
  - Example: race, ethnic origin, gender, etc

Quota sampling

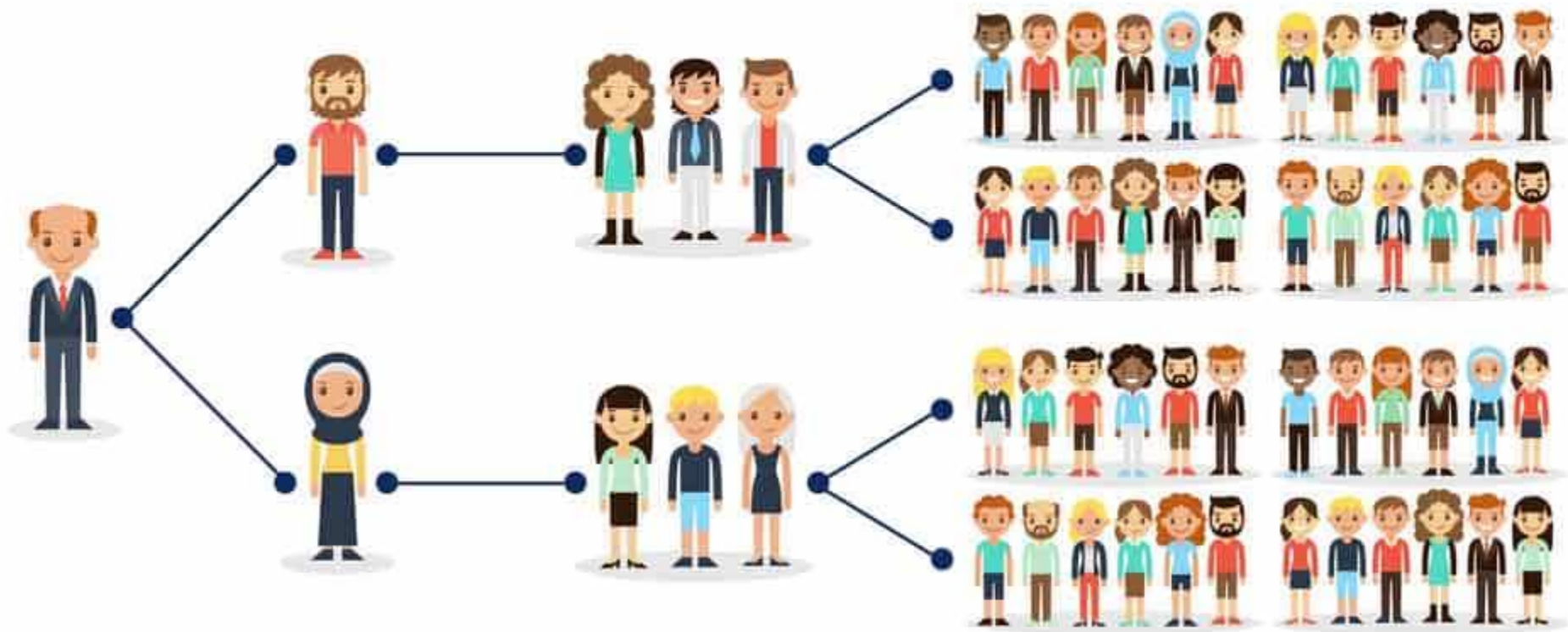


Non-Random Sampling

- Snowball sampling

  - Snowball sampling or chain-referral sampling is defined as a non-probability sampling technique in which the samples have rare traits.
  - This is a sampling technique, in which existing subjects provide referrals to recruit samples required for a research study.
  - This sampling method involves a primary data source nominating other potential data sources that will be able to participate in the research studies.

SNOWBALL SAMPLING



What size sample do I need?”

The answer to this question is influenced by a number of factors, including:

- the purpose of the study, population size, the risk of selecting a “bad” sample and the allowable sampling error.
- Data analysis plan e.g number of cells one will have in cross tabulation
- Most of all whether undertaking a **qualitative** or **quantitative** study

Sample size determination in qualitative study

- Probability sampling not appropriate as sample not intended to be statistically representative
- But, sample should have ability to represent salient characteristics in population.
- Sample size taken until point of theoretical saturation

……

- Sample size is usually small to allow in-depth exploration and understanding of phenomena under investigation
- Ultimately a matter of judgement and expertise in evaluating the quality of information against final use, research methodology , sampling strategy and results is necessary.
- In practice, qualitative sampling usually requires a flexible, pragmatic approach.

……

- The researcher actively selects the most productive sample to answer the research question.
- This can involve developing a framework of the

  variables that might influence an individual's contribution and will be based on the researcher's practical knowledge of the research area, the available literature and evidence from the study itself.
- This is a more intellectual strategy than the simple demographic stratification of epidemiological studies, though age, gender and social class might be important variables.

.....

- If the subjects are known to the researcher, they may be stratified according to known public attitudes or beliefs
- It may be advantageous to study a broad range of subjects :
  - (maximum variation sample)
  - outliers (deviant sample)
  - subjects who have specific experiences (critical case sample)
  - subjects with special expertise (key informant sample).

.....

- The iterative process of qualitative study design means that samples are usually theory driven (theoretical sampling) to a greater or lesser extent
- Theoretical sampling necessitates building interpretative theories from the emerging data and selecting a new sample to examine and elaborate on this theory.
- It is the principal strategy for the **grounded theoretical approach .**

Some suggestions of sample size in qualitative studies

- The smallest number of participants should be 15
- Should lie under 50
- 6-8 participants for FGDs AND at least 2 FGDs per population group

IMPORTANT

- Attainment of saturation
- Justification of choice of number

Strategies for Determining Sample Size

There are several approaches to determining the sample size.

- Using a census for small populations
- Imitating a sample size of similar studies
- Using published tables
- Applying formulas to calculate a sample size
- Use computer soft ware e.g EPI-info series

Using a Census for Small Populations

….

- One approach is to use the entire population as the sample.
- Although cost considerations make this impossible for large populations.
- Attractive for small populations (e.g., 200 or less).
- Eliminates sampling error and provides data on all the individuals in the population.
- Some costs such as questionnaire design and developing the sampling frame are “fixed,” that is, they will be the same for samples of 50 or 200.
- Finally, virtually the entire population would have to be sampled in small populations to achieve a desirable level of precision

Using a Sample Size of a Similar Study

- Use the same sample size as those of studies similar to the one you plan(Cite reference).
- Without reviewing the procedures employed in these studies you may run the risk of repeating errors that were made in determining the sample size for another study.
- However, a review of the literature in your discipline can provide guidance about “typical” sample sizes that are used.

Using Published Tables

- Published tables provide the sample size for a given set of criteria.
- Necessary for given combinations of precision, confidence levels and variability.
- The sample sizes presume that the attributes being measured are distributed normally or nearly so.
- Although tables can provide a useful guide for determining the sample size, you may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability.

Sample Size for ±5%, ±7% and ±10% Precision Levels where Confidence Level Is 95% and P=.5.

Size of Population	Sample Size (n) for Precision (e) of:		
	±5%	±7%	±10%
100	81	67	51
125	96	78	56
150	110	86	61
175	122	94	64
200	134	101	67
225	144	107	70
250	154	112	72
275	163	117	74
300	172	121	76
325	180	125	77
350	187	129	78
375	194	132	80
400	201	135	81
425	207	138	82
450	212	140	82

Using Formulas to Calculate a Sample Size

- Sample size can be determined by the application of one of several mathematical formulae.
- Formula mostly used for calculating a sample for proportions.

For example:

- For populations that are large, the Cochran (1963:75) equation yields a representative sample for proportions.
- Fisher equation, Mugenda etc

Cochran equation

$$n_0 = \frac{Z^2 pq}{e^2}$$

Where n_0 is the sample size,

Z^2 is the abscissa of the normal curve that cuts off an area α at the tails;

(1 – α) equals the desired confidence level, e.g., 95%);

e is the desired level of precision,

p is the estimated proportion of an attribute that is present in the population,and q is 1-p.

The value for Z is found in statistical tables which contain the area under the normal curve. e.g Z = 1.96 for 95 % level of confidence

Note

- The sample size formulas provide the number of responses that need to be obtained. Many researchers commonly add 10 % to the sample size to compensate for persons that the researcher is unable to contact.
- The sample size also is often increased by 30 % to compensate for non-response (e.g self administered questionnaires).

Further considerations

- The above approaches to determining sample size have assumed that a simple random sample is the sampling design.
- More complex designs, e.g. case control studies etc , one must take into account the variances of sub-populations, strata, or clusters before an estimate of the variability in the population as a whole can be made.