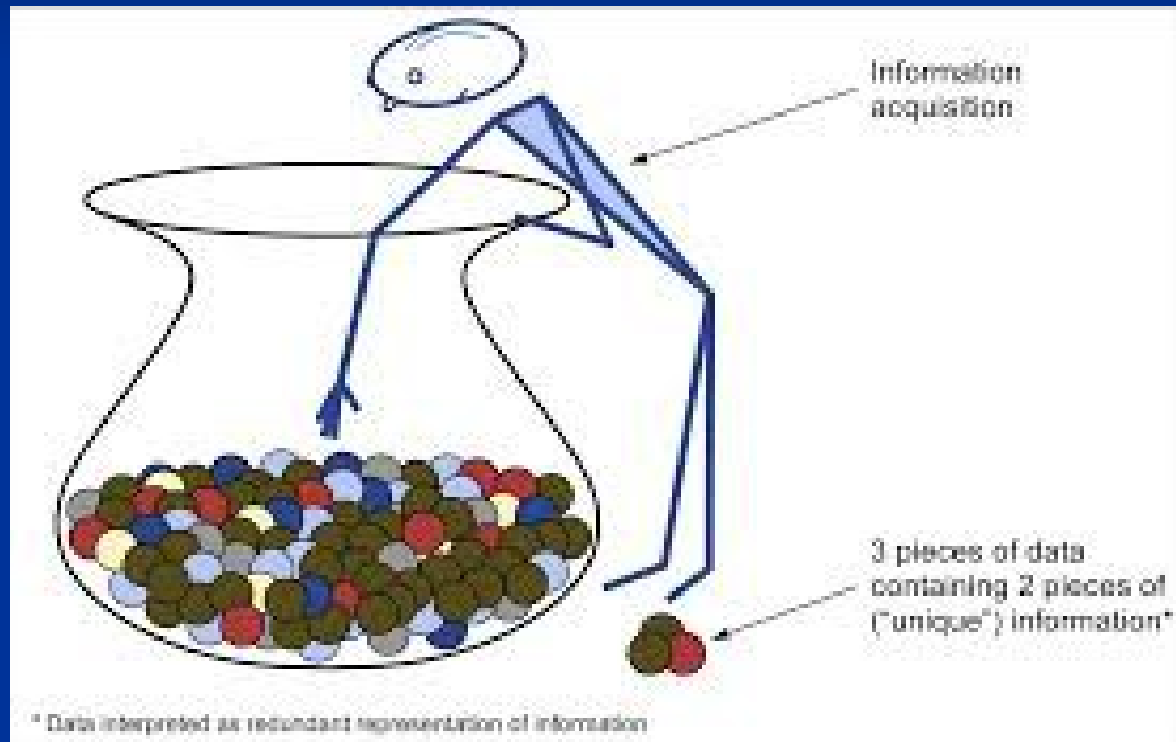


# Sampling



# Sampling concepts

## Population:

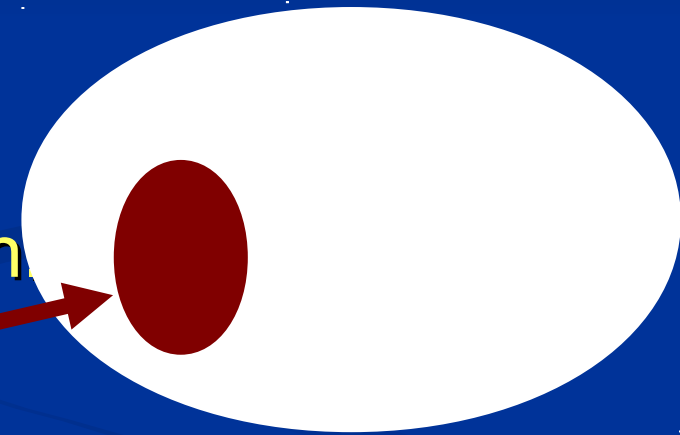
The collection of all objects in study.

## Element:

Single member of the population

## Sample:

A subset of the population



# Sampling concepts

Sampling frame: All elements of population with proper identification available for

.selection

. Sampling Unit: Single member of sample

# ?Why sampling

Get information about large populations

- ↓ Less costs
- ↓ Less field time
- ↑ More accuracy i.e. Can Do A Better Job of Data Collection
- ↘ When it's impossible to study the whole population

## Target Population:

The population to be studied/ to which the investigator wants to generalize his results

## Sampling scheme

Method of selecting sampling units from sampling frame

Sampling: Process of selection of adequate sample from population.

:Sampling in real life

Sample vs census: census is used when there is a lot of heterogeneity and .population size is small

# Types of sampling

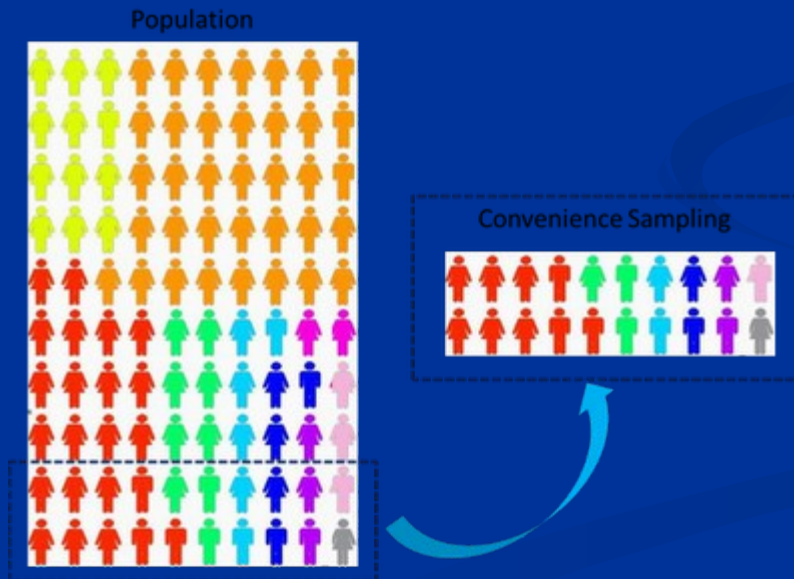
---

- Non-probability samples
- Probability samples

# Non probability samples

---

- Convenience samples (ease of access)  
sample is selected from elements of a population that are easily accessible





# Non probability samples

---

Probability of being chosen is unknown

Cheaper- but unable to generalise  
potential for bias

- Snowball sampling (friend of friend....etc.)
- Purposive sampling (judgemental)
- Quota sample

# Probability samples

---

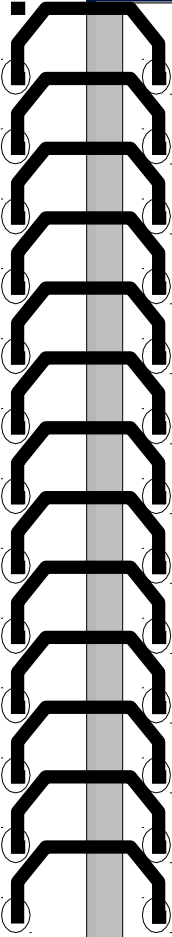
- Random sampling
  - Each subject has a known probability of being selected
- Allows application of statistical sampling theory to results to:
  - Generalise
  - Test hypotheses

# Methods used in probability samples

---

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Multi-stage sampling
- Cluster sampling

# Simple random sampling



The diagram shows a vertical spinner with 48 numbered slots, each with a pointer. The spinner is divided into two halves by a vertical line. The numbers 1 through 48 are arranged in two columns, with 1-24 on the left and 25-48 on the right. The spinner is shown in a state where the pointer is pointing to the number 20 on the left side. The numbers 20, 27, 29, 32, 40, and 48 are circled in red, indicating they are the selected sample.

1	Albert D.	25	Monique Q.
2	Richard D.	26	Réçine D.
3	Belle H.	27	Lucille L.
4	Raymond L.	28	Jérémy W.
5	Stéphane B.	29	Gilles D.
6	Albert T.	30	Renaud S.
7	Jean William V.	31	Pierre K.
8	André D.	32	Etienne M.
9	Jeremy W.	33	Marie M.
10	Anthony Q.	34	Gaétan Z.
11	James B.	35	Fidèle D.
12	Denis G.	36	Maria P.
13	Amanda L.	37	Anne Marie G.
14	Jennifer L.	38	Michel K.
15	Philippe K.	39	Gaston C.
16	Eve F	40	Alain M.
17	Priscilla O.	41	Olivier P.
18	Robert D	42	Geneviève M.
19	Brian F.	43	Berthe D.
20	Hellène H.	44	Jean Pierre P.
21	Isabelle R.	45	Jacques B.
22	Jean T.	46	François P.
23	Samanta D.	47	Dominique M.
24	Berthe L.	48	Antoine C.

# Systematic sampling

## Sampling fraction

Ratio between sample size and population size

# Systematic sampling

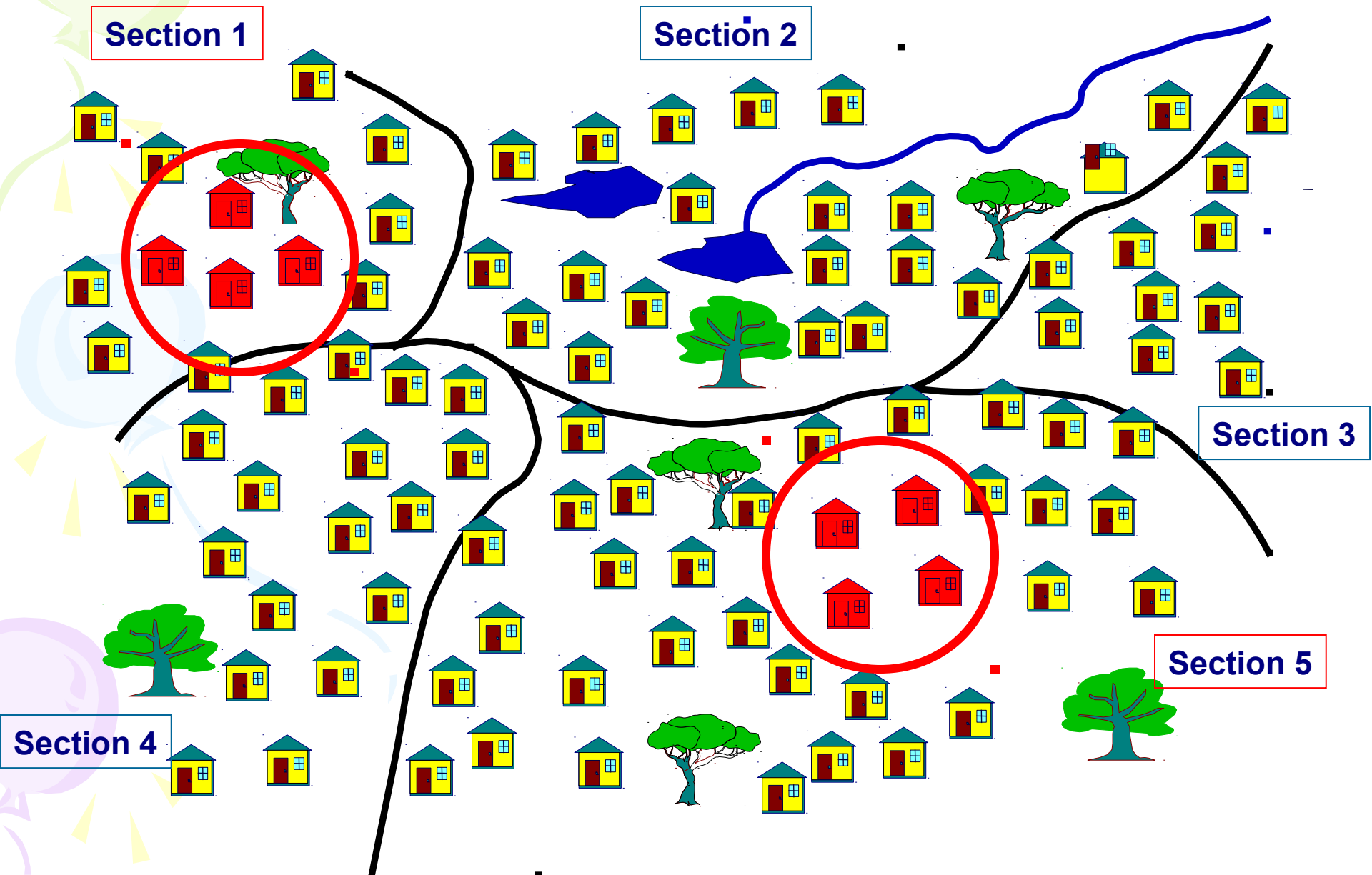
1	Albert D.	25	Monique Q.
2	Richard D.	26	Régine D.
3	Belle H.	27	Lucille L.
4	Raymond L.	28	Jérémy W.
5	Stéphane B.	29	Gilles D.
6	Albert T.	30	Renaud S.
7	Jean William V.	31	Pierre K.
8	André D.	32	Etienne M.
9	Jeremy W.	33	Marie M.
10	Anthony Q.	34	Gaétan Z.
11	James B.	35	Fidèle D.
12	Denis G.	36	Maria P.
13	Amanda L.	37	Anne-Marie G.
14	Jennifer L.	38	Michel K.
15	Philippe K.	39	Gaston C.
16	Eve F.	40	Alain M.
17	Priscilla O.	41	Olivier P.
18	Robert D.	42	Geneviève M.
19	Brian F.	43	Berthe D.
20	Hellène H.	44	Jean Pierre P.
21	Isabelle R.	45	Jacques B.
22	Jean T.	46	François P.
23	Samanta D.	47	Dominique M.
24	Berthe L.	48	Antoine C.

# Cluster sampling

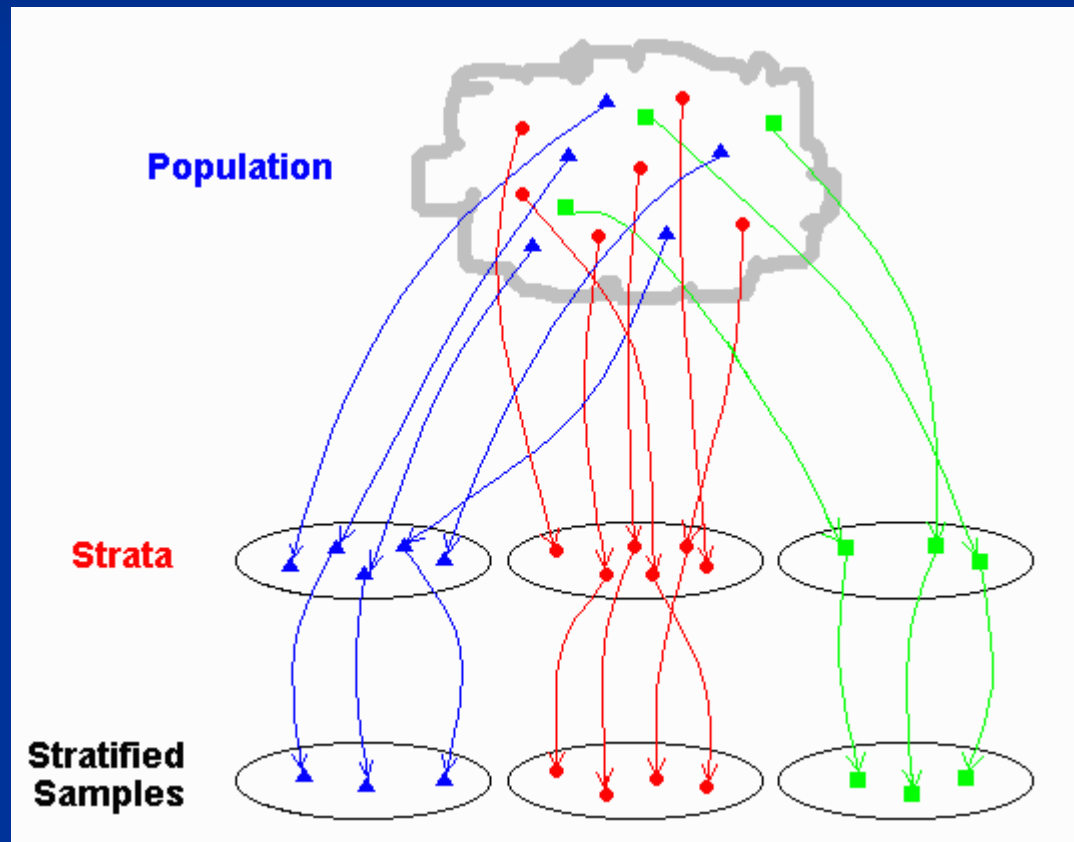
Cluster: a group of sampling units close to each other i.e. crowding together in the same area or neighborhood



# Cluster sampling



- Stratified sampling
- Multi-stage sampling



# Errors in sample

- Systematic error (or bias)
  - Inaccurate response (information bias)
  - Selection bias
- Sampling error (random error)
  - Sample is not representing the population

# Type 1 error

---

- Type I error, also known as a “false positive”: the error of rejecting a null hypothesis when it is actually true. In other words, this is the error of accepting an alternative hypothesis
- Known as the  $\alpha$  (or “type 1 error”)
- Usually set at 5% (or 0.05)

# Type 2 error

---

- Type II error, also known as a "false negative": the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature.
- Known as the  $\beta$  (or "type 2 error")

# Sample size

$$n = \frac{Z^2 \sigma^2}{e^2}$$

## Problem 1

A study is to be performed to determine a certain parameter in a community. From a previous study a sd of 46 was obtained.

If a sample error of up to 4 is to be accepted. How many subjects should be included in this study at 99% level of confidence?

# Answer

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$$n = \frac{2.58^2 \times 46^2}{4^2} = 880.3 \sim 881$$