Populations and Sampling

We analyze data to answer a question about a population

Population the who or what we are interested in learning about → consists of all the individuals we might collect data about

Unfortunately we cannot know every individuals data

If we try and collect data about the entire population, this is known as a census.

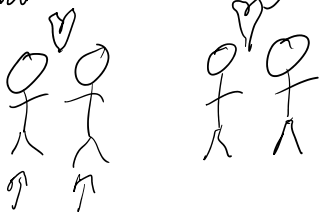
Sample A subset of the population of interest that we gather data about. Select some individuals (Ideally at random) from the population to gather data about.

There are a number of different ways to take a sample.

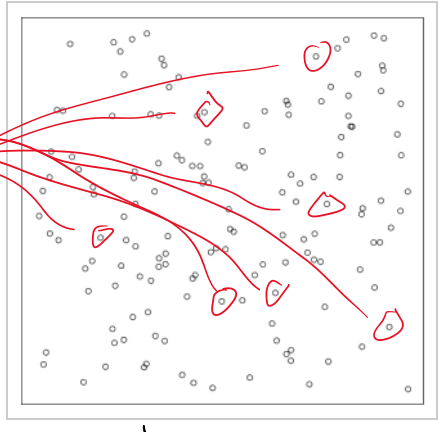
Best **Simple Random Sample**

Every individual has an equal chance of being selected

Every possible sample has an equal chance of being selected.



Population



Benefits

Without Bias

Because it is randomized

Downsides

Difficult → we must know entire population

Expensive to contact individuals

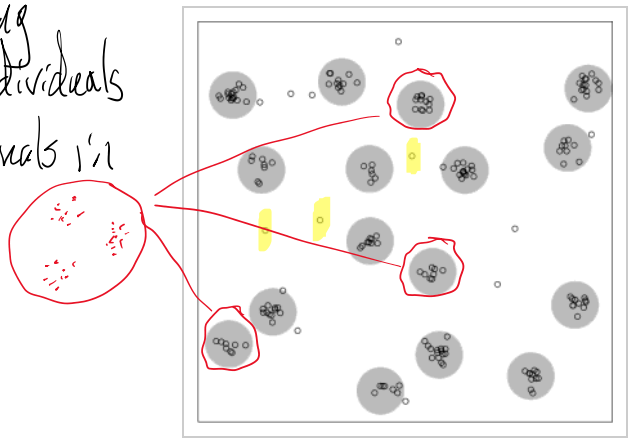
Cluster Sampling

Population

We our sample by selecting . .

We our sample by selecting pre existing clusters of individuals and sampling all individuals in the cluster.

Population



Doctors

Clusters are Hospitals, clinics etc.

Students

Classrooms

Benefits

It is easy to contact individuals through clusters.

Downsides

Individuals outside clusters are not able to be selected

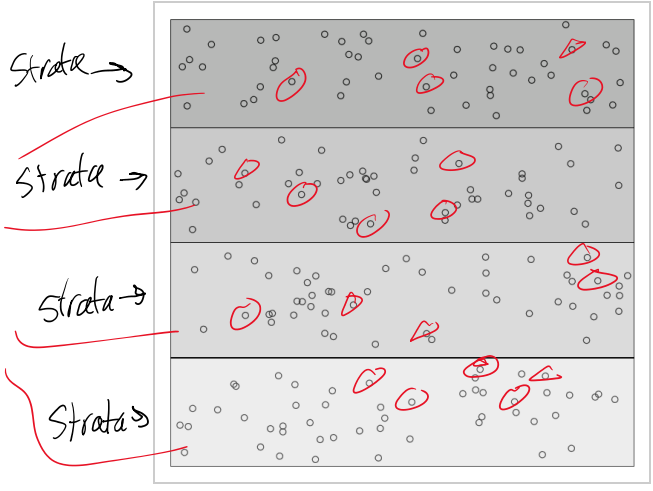
Individuals within clusters are correlated (Similar, affected by each other, affected by cluster)

Stratified Sampling

We sample (SRS) within strata. This ensures representation from each strata.

We are typically interested in comparing individuals within strata.

Population



Doctors

Type of care

→ GP, Surgery, Orthopedics.

Students

→ Majors

→ Grade levels

Benefits

We can specifically target very small strata to ensure their population is represented in the sample

Downsides

It hard to define full population of each Strata.

Voluntary Response - Every individual in the population is able to self select to be a member of the sample.

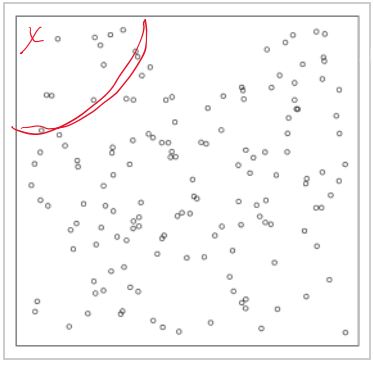
Ex: Product review online

Benefit
Extremely easy to conduct

Downsides
People are more likely to self select if they have a strong opinion.

Convenience Sample
We select our sample by pure convenience.

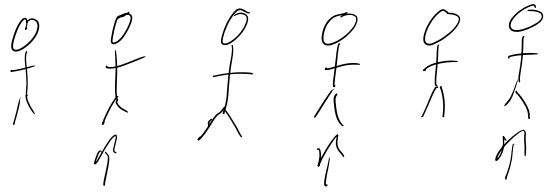
Ex. I ask you to respond to a survey.



Downsides
Highly Biased

Benefits
Easy

Systematic Sample
Make a list of population & take every k^{th} individual



Benefits
structured

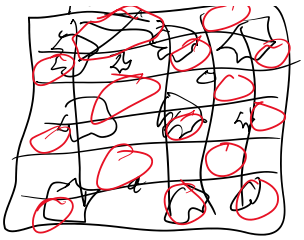
Downside
We need a list of entire population

Example
When should we record temp?

↓

Systematic sampling is common when we define samples over

Forest



define samples over space or time