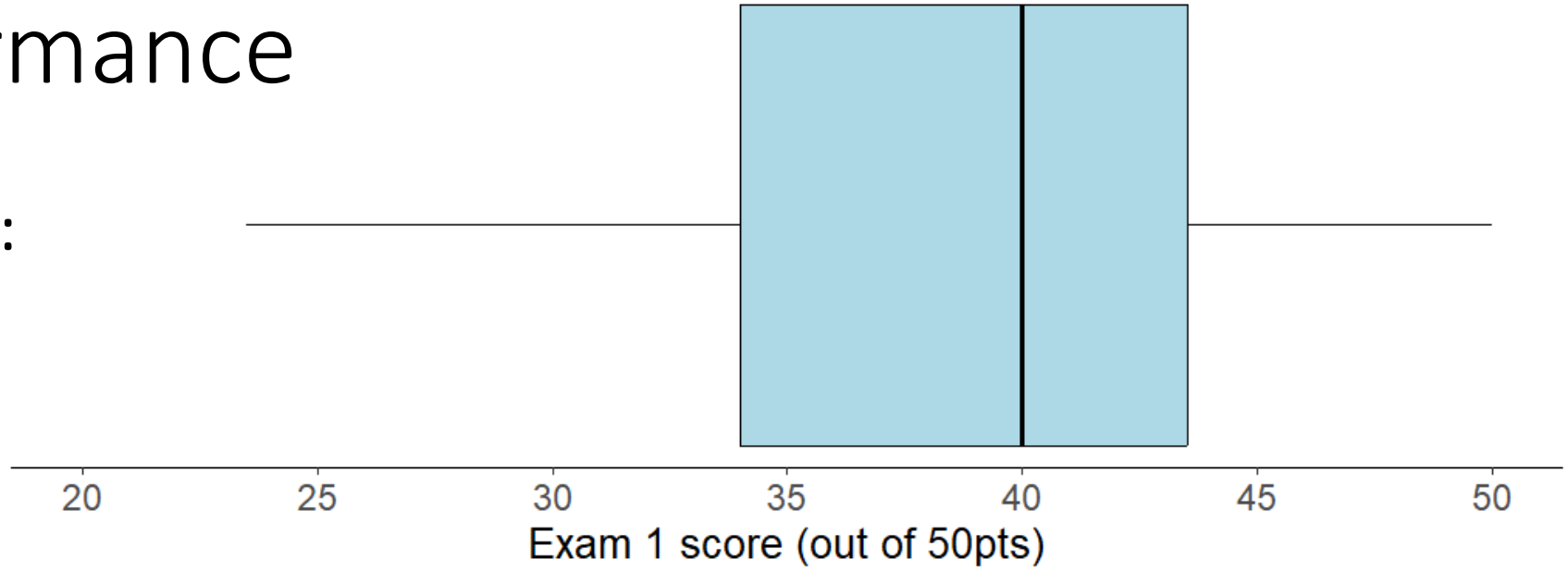
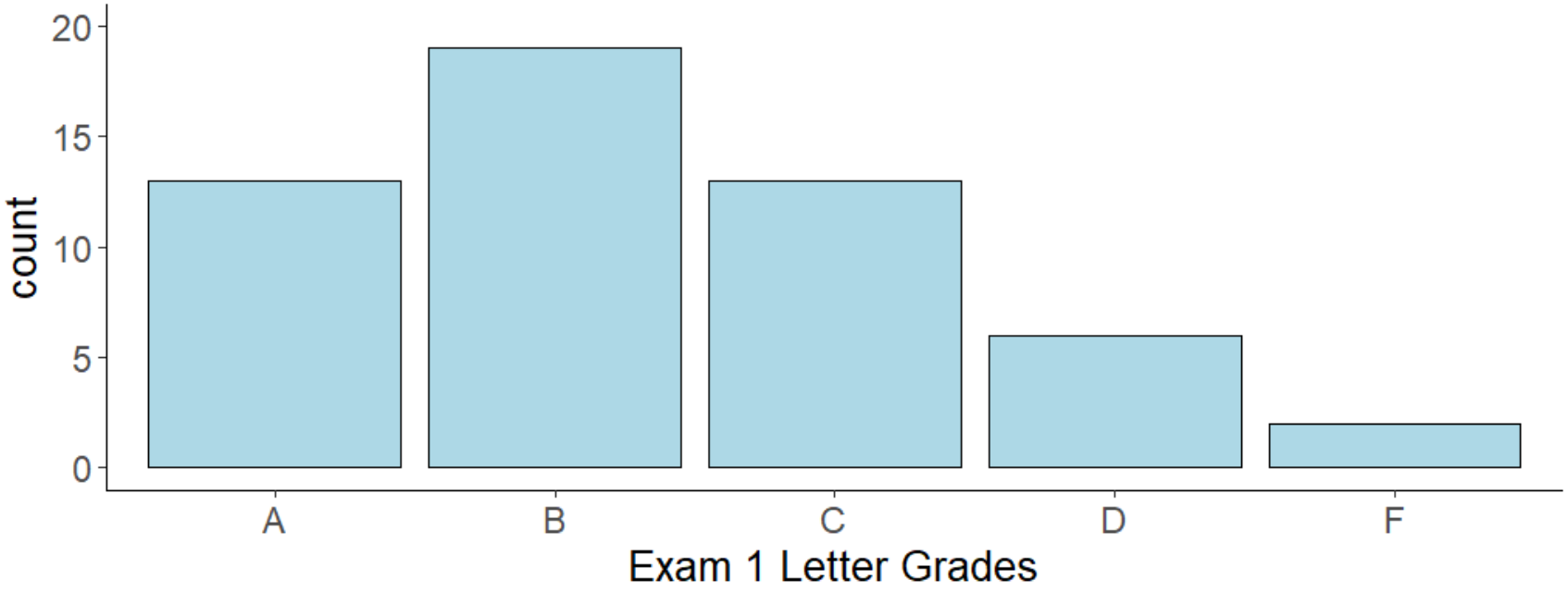
Lecture 9
Types of Studies-Continued

Exam 1 Performance

- Five number summary:

23 34 40 34 50





Review From Wednesday 2/7

- Sampling distributions
  - What is a sampling distribution?
- Margin of Error
  - What is the margin of error?
- What is a statistically significant result?
- Observation vs Experimental Study
- Causation vs Association

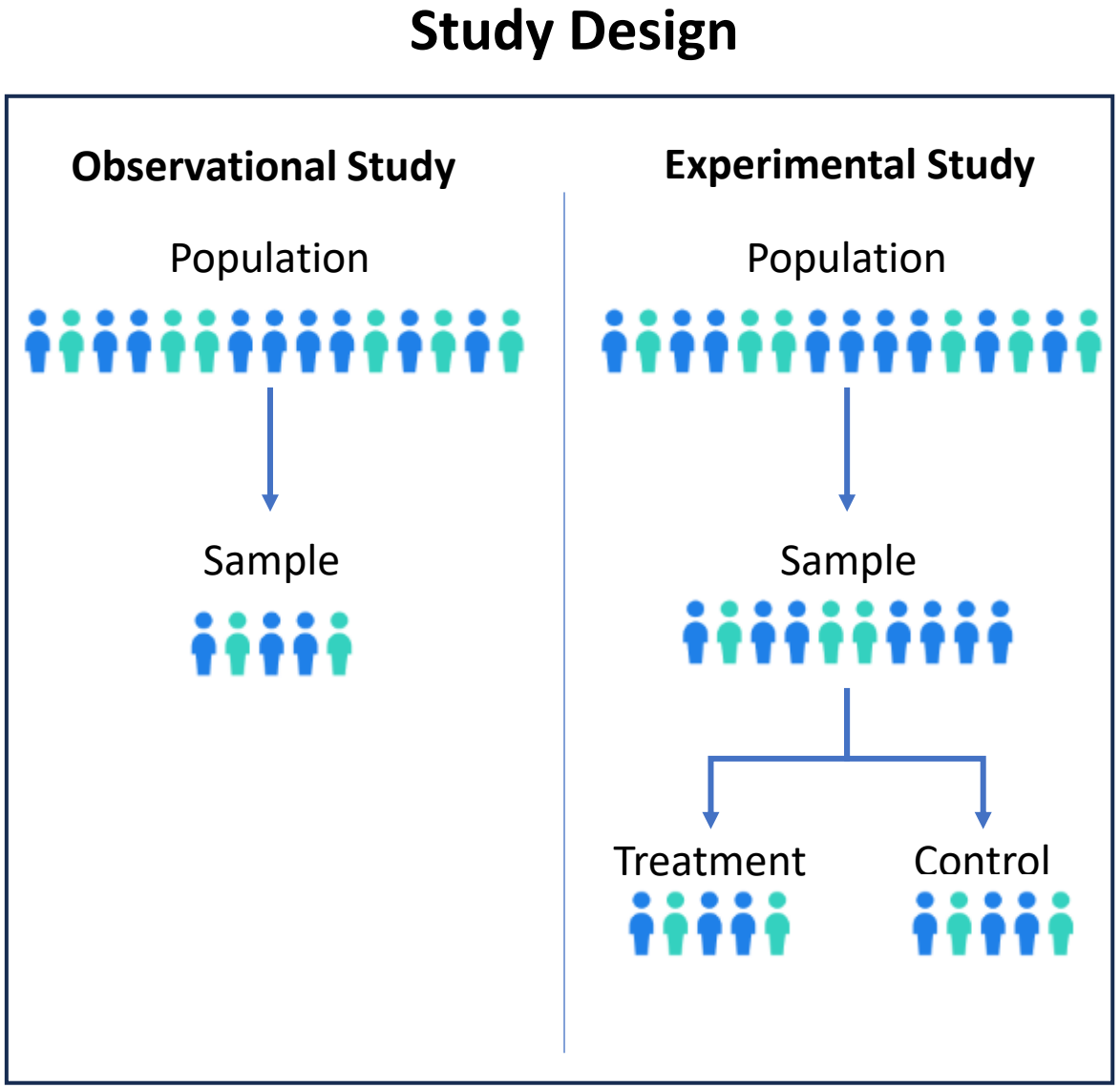
Advantages of Experimental Studies

- Observational studies cannot definitively establish causation
- Observational studies are prone to **lurking variables** – a variable unknown to the researchers that is not included in the study and has an association with both the response and explanatory variables

- Lurking variables can induce false associations between response and explanatory variables.

- In experimental studies subjects (observations) are randomly assigned to treatment groups.

  this randomization balances the effect of lurking variables between the treatment groups and removes their influence on the association between the response and explanatory variables



Hallmarks of a good experiment

- Control group – a group of subjects in the experiment who do not receive the treatment
  - reduces bias in the experiment because by design the only difference between the two groups is the treatment

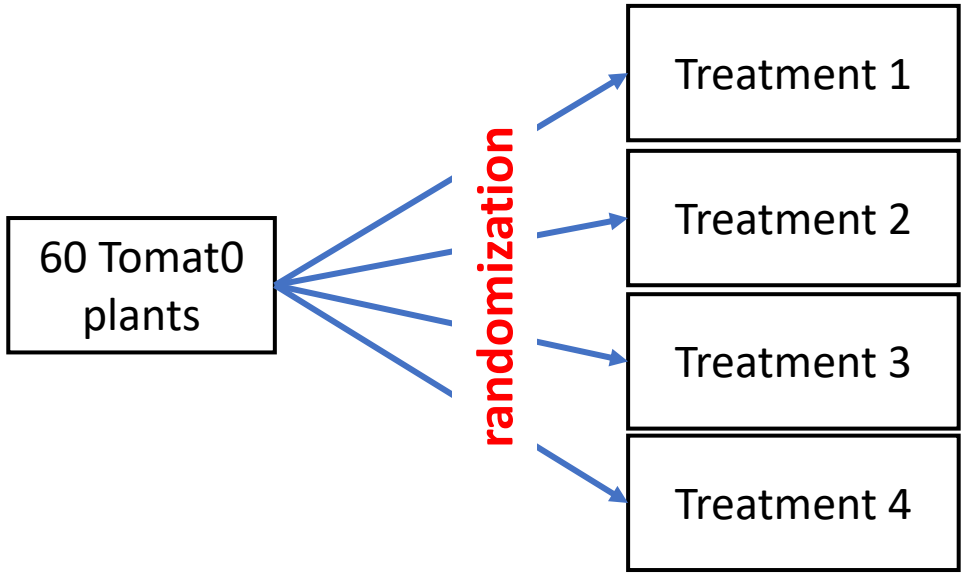
- Blinding – designing the experiment to ensure the subjects are unaware if they are in the treatment or control group.

- Double blinding – When subjects as well as the researchers are unaware of who is assigned to the treatment group and who is assigned to the control group.

Some Experimental Designs

- **Completely Randomize Design** – Subjects are randomly assigned to treatment groups.
  - compares response to a single factor
  - each unit has the same chance of being in the treatment or control groups

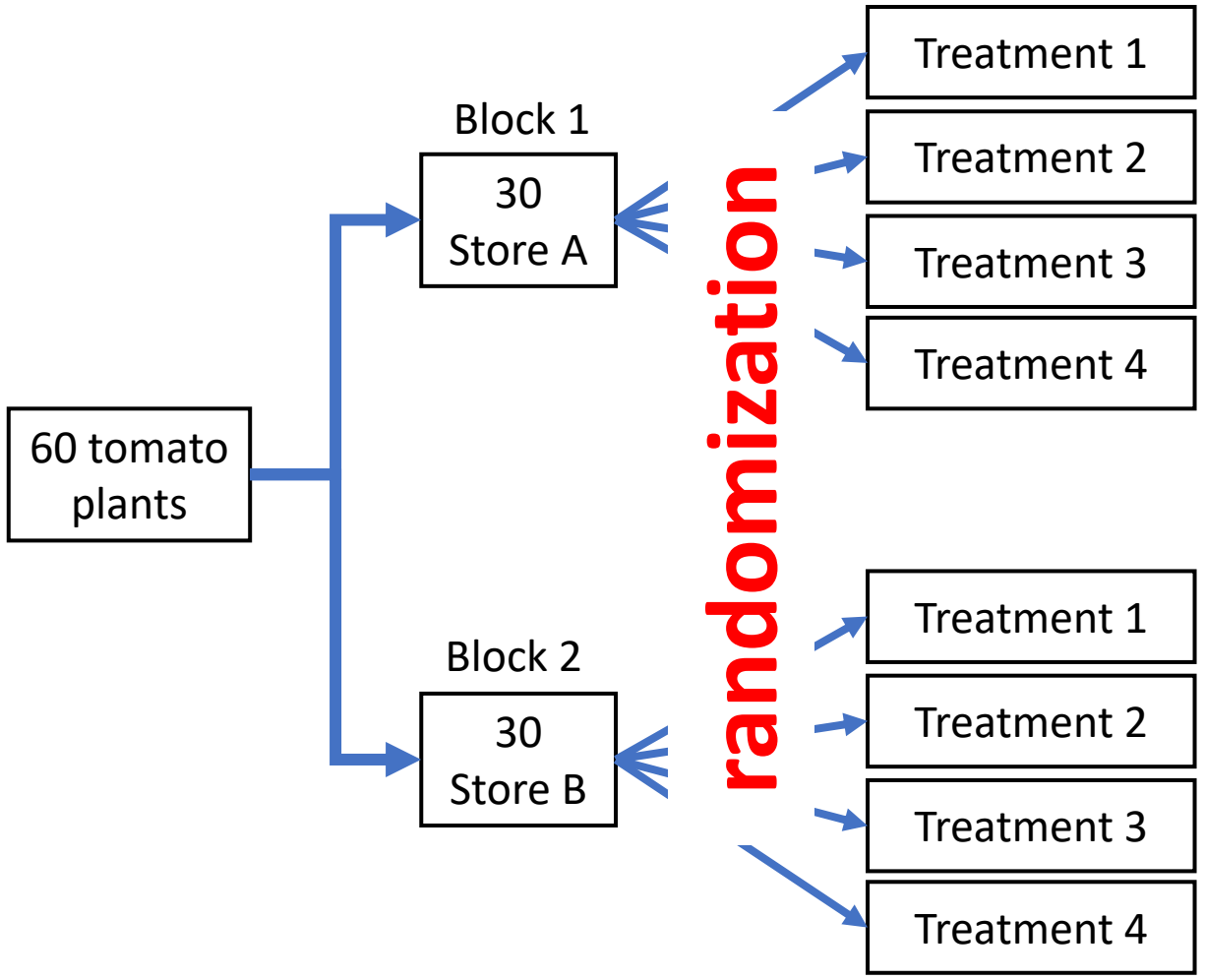
- **Multifactor experiments** – An experiment which compares multiple factors simultaneously
  - cheaper than conducting an experiment for each factor separately
  - we can learn more from a multifactor experiment



		Fertilizer A: YES	Fertilizer A: NO
Fertilizer B	YES	**Treatment 1** Fertilizer A + Fertilizer B	**Treatment 2** Fertilizer B + Placebo
	NO	**Treatment 3** Fertilizer A + Placebo	**Treatment 4** Placebo only

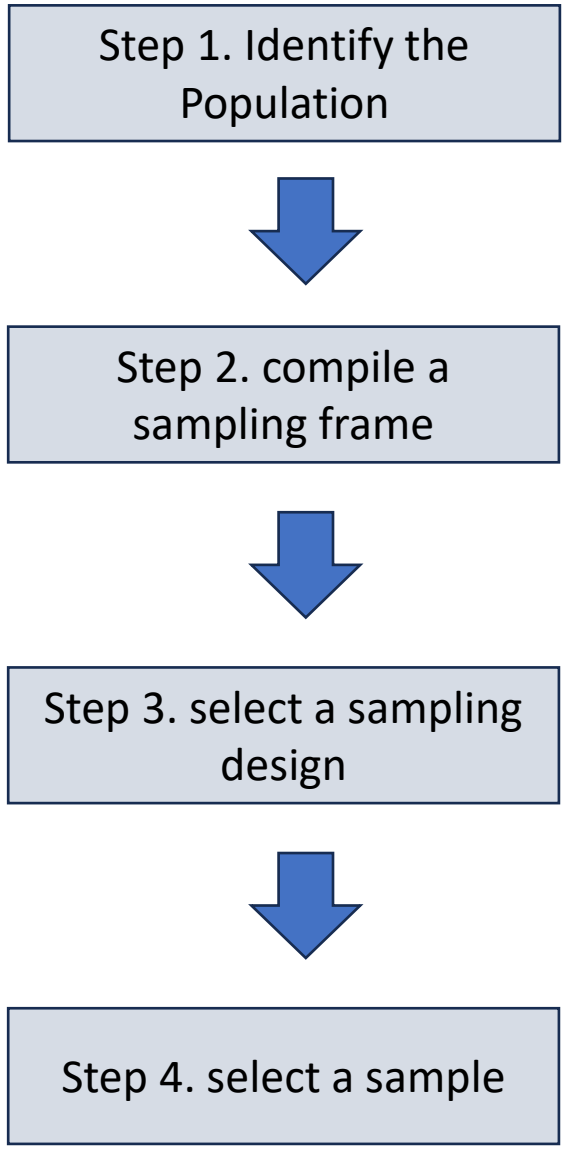
- **Randomized Complete Block Design** – When subjects are not similar enough, detecting differences among the treatment groups can be difficult. We instead create groups called **blocks.** Blocks are organized so that units inside a block are more similar. Each block sees all treatments in random order

- **Matched Pair Designs** – a design which takes measurements on each subject, usually once before the treatment and once after the treatment producing a set of paired measurements



Surveys

- A **sample survey** selects a sample of subjects from a population and collects data
  - In statistics, a survey is any information gathered from a sample of subjects.
  - It is a type of non-experimental study
- A **census** attempts to gather data for all (or nearly all) subjects in a population
- A **sampling frame** is a list of subjects in the population from which the sample will be collected
- A **sampling design** is the method that will be used to select subjects from the sampling frame
- We seek a sampling design that will lead to a sample that is representative of the entire population we are trying to estimate



Example of a Survey

- Suppose I want to see what proportion of people in Moscow Idaho liked the Star Wars sequel trilogy. So, I acquire a phone book for Latah county, ID and call the first 100 people with an address in Moscow ID and ask them to rate the sequels on a scale of 1-10.
- What is the sampling frame?

The phone book for Latah County

- Do you think such a sample will be representative of the population we are trying to estimate?

NO, the sampling frame does not cover all possible people in the population of interest

Simple Sampling Designs:

- **Convenience sampling** is a type of non-probability sampling that involves the sample being drawn from that part of the population that is close to hand**.**

  Volunteer sample – a type of sampling where participants self elect to be part of the study because they volunteer when asked or respond to advertisement

  - the most common type of convenience sampling
  - often required when we don't have a sampling frame for the population
  - this is the type of sampling used for most medical experiments

- A sample is more likely to be representative of the population if we let *chance,* rather than *convenience*, determine which subjects are sampled.

- In **simple random sampling** (also just called random sampling) each subject in the sampling frame to has an equal probability of being selected for the sample.

Sampling Designs: Simple Random Sampling

- If we **sample with replacement,** then each time we sample a subject from the population we put the subject back so that it can be sampled again

  In general, for a population of size N each subject will have a $\frac{1}{N}$ chance of being included in the sample.

- If we **sample without replacement,** then each time we sample a subject from the population we remove that subject from the sampling frame so that we cannot select them again.

  - This means that first subject will have a $\frac{1}{N}$ chance of being selected, the second a $\frac{1}{N-1}$, the third a $\frac{1}{N-2}$ … and so on
  - Sampling without replacement is common in most surveys because the sample size is usually small in comparison to the population size (i.e $n \ll N$) it is approximately the same as sampling with replacement

Sources of Bias In Surveys

- **Bias** – when a sample is not representative of the population of interest.

- **Undercoverage** – Bias introduced by having a sampling frame that lacks representation from parts of the population
  - non-random sampling designs are prone to undercoverage

- **Nonresponse Bias** – When some of the sampled subjects cannot be reached or refuse to participate
  - most surveys suffer from this kind of bias
  - Current population survey of the U.S Census Bureau has a nonresponse rate of about 7%

- **Response Bias** – When survey question is asked in a leading way or a subjects emotions affect how they respond

- A large sample size does NOT guarantee an unbiased sample!

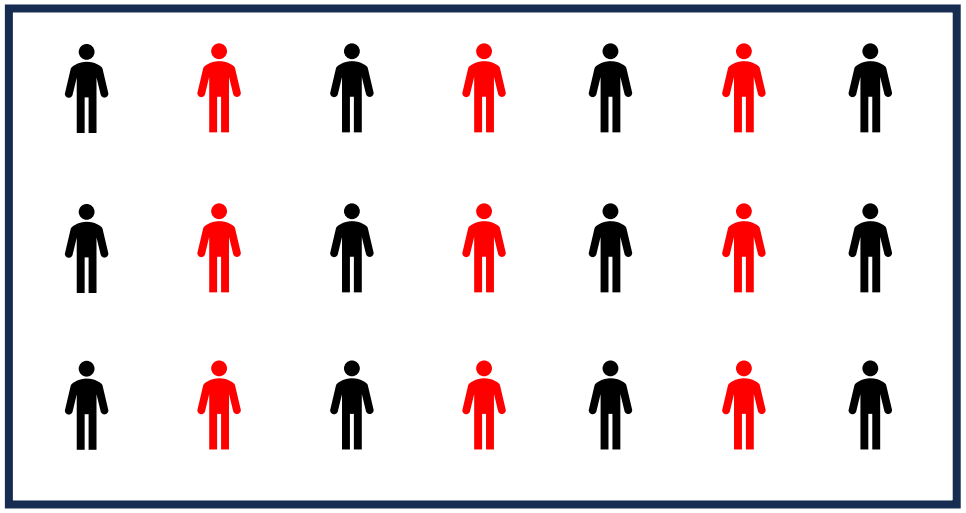
More complex methods of sampling

- **Systematic Sampling** – A sampling method in which the researcher selects every k^{th} subject from an ordered sampling frame

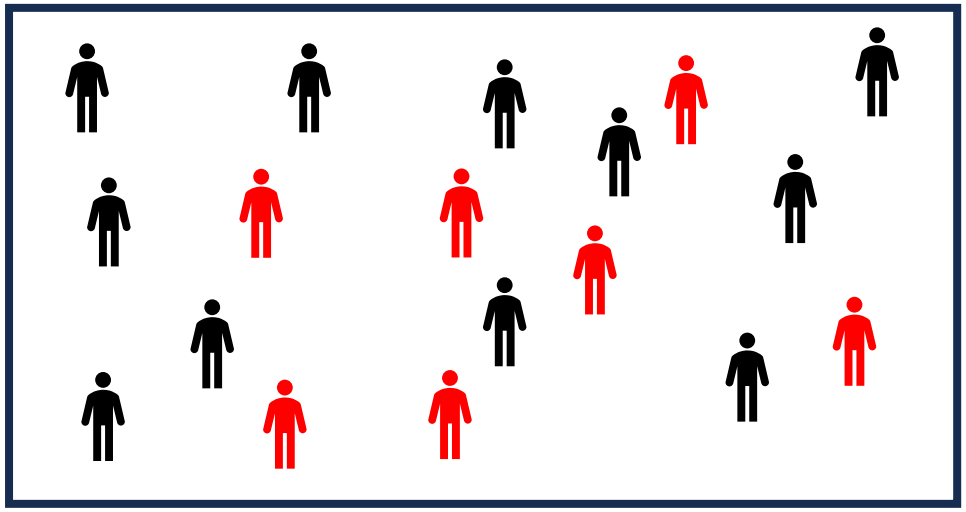
- **Cluster sampling** – A type of sampling method in which the population is divided in a set of clusters and the researcher selects a simple random sample of the clusters. The sample then comprises all subjects in the selected clusters.

- **Stratified Random Sampling** – A type of sampling method in which the population is separated into groups, call **strata**, based on some characteristic about the subjects. A simple random sample is then taken from each stratum.

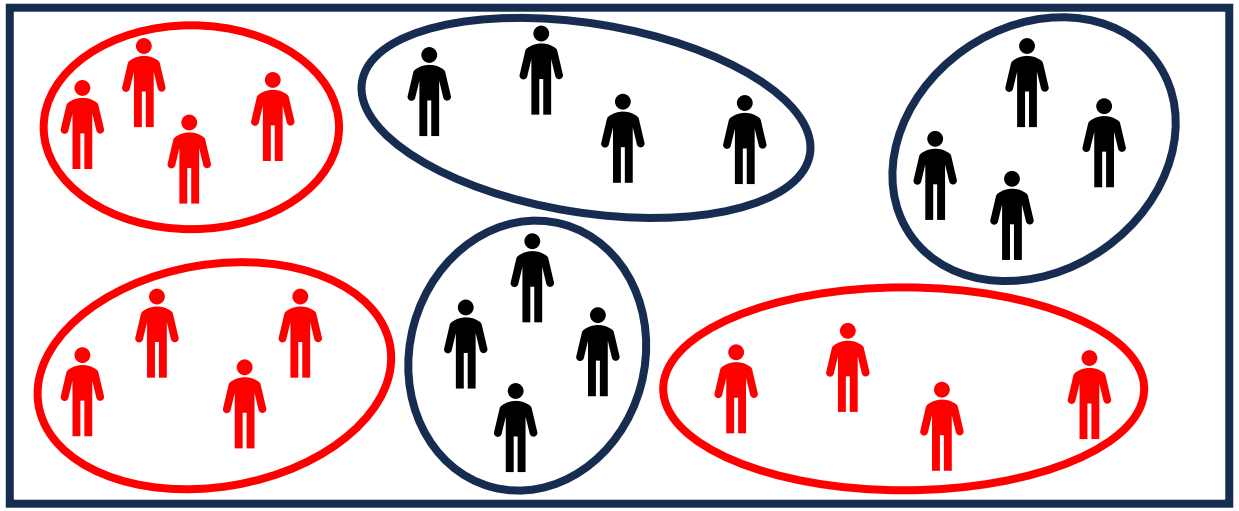
Systematic Sampling



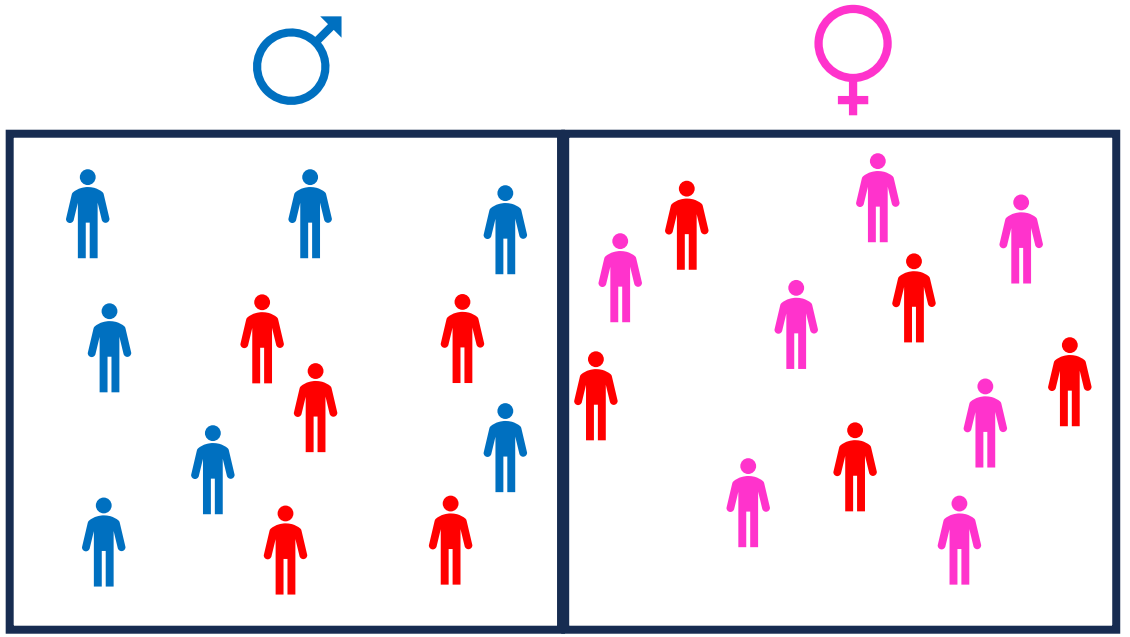
Simple Random Sampling



Cluster Sampling



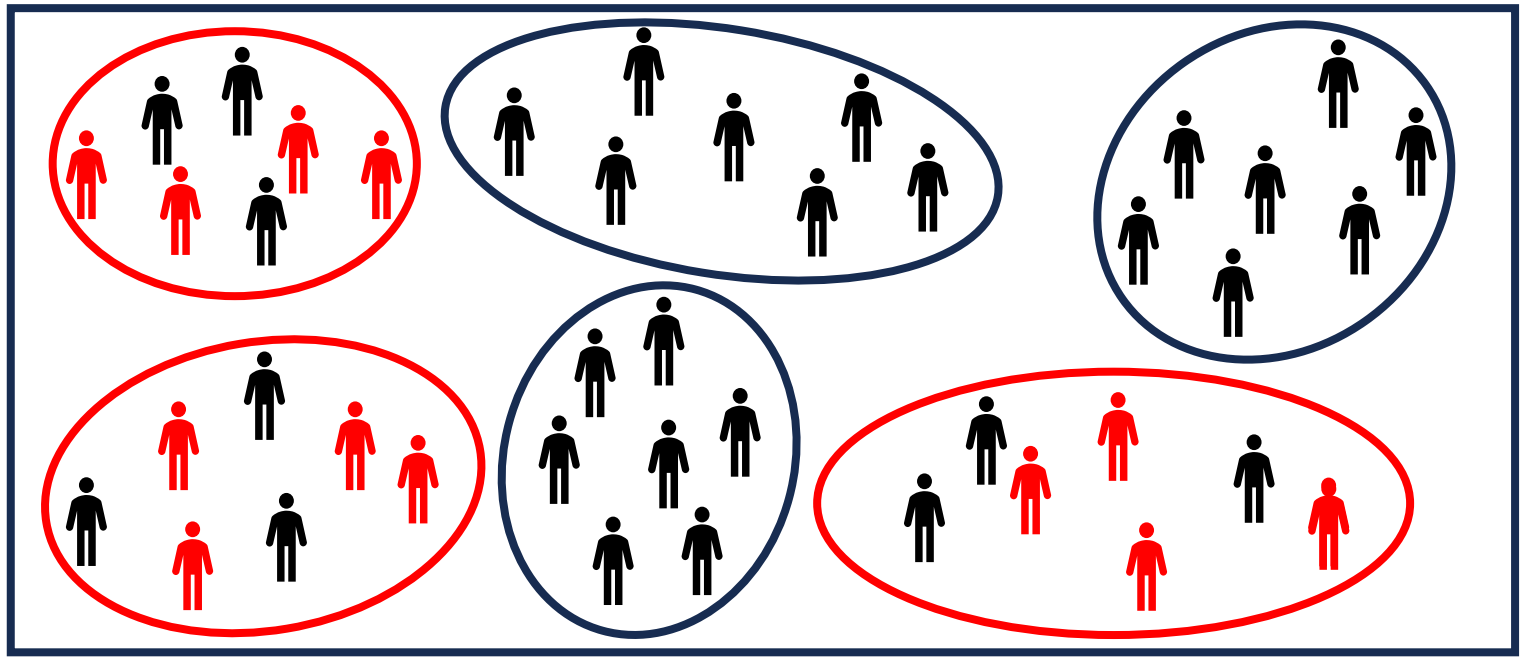
Stratified Random Sampling



More complex methods of sampling

- **Two – Stage cluster Sampling** - A type of sampling method in which the population is divided into a set of clusters and the researcher selects a simple random sample of the clusters. A simple random sample is then applied to each cluster

Two – Stage Cluster Sampling



Advantages and Disadvantages of Sampling Designs

Simple Random Sampling

- Mathematically simple to compute estimates such as $\bar{x}$ and s^2
- Samples tend to be a good representation of the population

Systematic Sampling:

- Sometimes useful when there is no sampling frame available.
- Lower margin of error than simple random sampling and some cluster sampling designs.

Advantages and Disadvantages of Sampling Designs

Stratified Random Sampling:

- **Administrative convenience** - It may be easier to conduct several smaller simple random sampling designs than coordinate one larger simple random sampling design.

- **Interest in individual strata** - The design ensures samples from all strata. A simple random sampling design might sample few or no elements from a stratum of interest.

- **Smaller margin of error** - By assuring samples from each strata, the combined sample tends to be more representative of the population, resulting in a smaller margin of error

Advantages and Disadvantages of Sampling Designs

Cluster Sampling:

- The advantages of cluster sampling are that (a) **it can be less expensive than simple or stratified random sampling** and (b) it **can be used when a sampling frame is unavailable**
- A disadvantage of cluster sampling is that **the margin of error is often larger** than what it would be for simple random sampling or stratified random sampling

Two-Stage Cluster Sampling:

- Same advantages as above
- Usually has a smaller margin of error, because we can control two sample sizes: the number of clusters to sample, and the number of elements to sample from each sampled cluster

Practice: Identify the Design

- Suppose I want to estimate the average height of my students in STAT 251 section 01. To do so, I use the registrars list to get the names of the students registered for my section. I number the students from 1 to N and select students 4, 8, 12, 16… to be my sample.
- What is the sampling frame?

The list of students in the class

- What is the sampling design?

Systematic sampling

Practice: Identify the Design

- Suppose I want to see what proportion of people in Moscow Idaho liked the Star Wars sequel trilogy. So, I acquire a cadastral map (a map that shows the boundaries and ownership of land parcels) for Moscow, Idaho and group the the houses into city blocks. I take a random sample of city blocks and for each city block I selected and put a questionnaire in the mailboxes of all houses on that block.

- What is the sampling frame?

The cadastral map of Moscow

- What is the sampling design?

Cluster Sampling