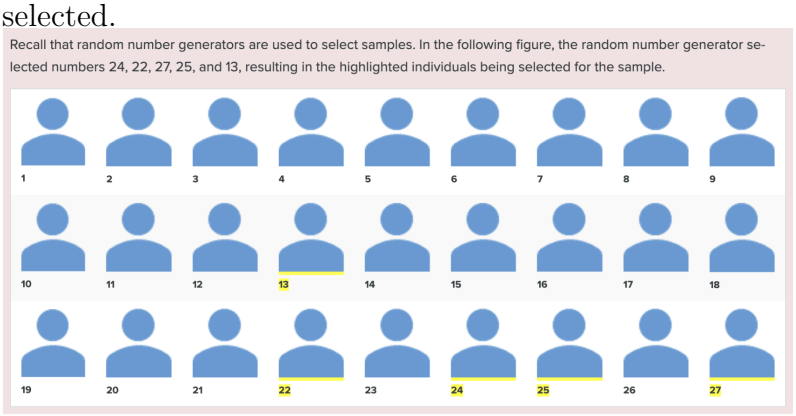
# Chapter 2: Sampling Techniques, Observational Studies, and Experimental Design

## 1 Samples and Population

A quick recap: Think of a population as containing **everything** you are studying whereas the sample is a small part of that large population. You can take as many samples as you'd like from one population. A parameter describes the whole population and statistic describes the whole sample.

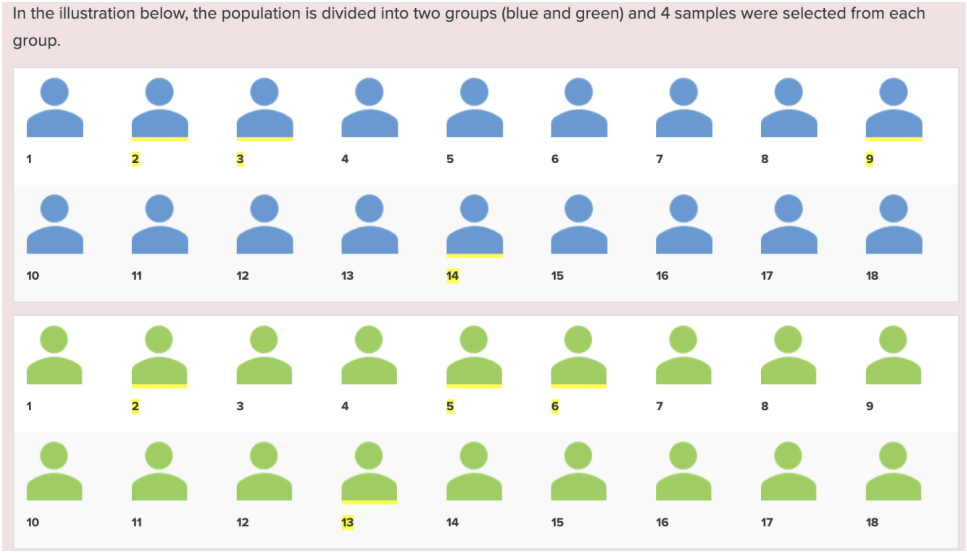
## 2 Common sampling methods

- Simple Random Sampling: Every sample of a given size has the same chance of being selected.




- Systematic Sampling: Every individual in the population is given a number and individuals are chosen at regular intervals, with a random starting point.

- Stratified Sampling: Population is divided into two or more groups (strata) according to some criterion and a sample is selected from each strata using simple random sampling or systematic sampling.




- Convenience Sampling: Sample of individuals who are most accessible to the researcher.

Example: A student at a local high school in Florida is interested in doing a research project about student diversity in the three public high schools in her county. She wants to investigate the proportion of public high school students in her county who self-identify as American Indian/Alaskan native, Asian, Black, Hispanic, White, or Multi-ethnic. What is the population of interest? What should she consider if she wants to take a representative sample from this population?

Population of interest: All the public high school students in her county.

Sampling: To produce a representative sample, she should decide on a sample size and then use a random mechanism to choose a sample from that population so that every sample of that particular size is equally likely to be chosen.

# 3 Bias

The most common sources of biases are as follows:

- Undercoverage: When some groups of the population are left out of the sampling process and the individuals in these groups do not have an equal chance of being selected for the sample. Example: a sample survey of households in a country may miss people who are homeless, prison inmates, or students living in dorms.
- Non-Response Bias: When an individual chosen for a sample cannot be contacted or decides to not participate in the study or research. This type of bias occurs after the sample has been selected and can create potential bias in the data collected.
- Response Bias: This type of bias can occur when a person does not understand a question or feels influenced to respond to a question in a certain way. Response bias can also occur as a result of the wording of questions that are of a sensitive nature.

- Voluntary Response Bias: The sample is not random or representative of the population. The people who volunteer for a study or survey may be more inclined to respond to questions or report certain behaviors. Examples: online reviews and post-call surveys.

Example: Let us revisit the previous example: A student at a local high school in Florida is interested in doing a research project about student diversity in the three public high schools in her county. She wants to investigate the proportion of public high school students in her county who self-identify as American Indian/Alaskan native, Asian, Black, Hispanic, White, or Multi-ethnic.

- If the student takes a simple random sample of students from her high school and asks those students about their race/ethnicity. What are, if any, the potential sources of bias in her sampling method?
  Undercoverage. This sampling method has the potential for undercoverage because only students from one of the three public high schools in the county were asked.

- Assume the student creates a questionnaire that asks about race/ethnicity and then asks for volunteers at all three high schools to take the questionnaire. What are, if any, the potential sources of bias in her sampling method?
  This sampling method has the potential for voluntary response bias and response bias. Since she is using volunteers, the sample might not be representative of the population because students of a certain race/ethnicity might be more willing to volunteer, which would lead to overrepresentation of a certain race/ethnicity (voluntary response bias). Also, depending on how the questionnaire is written, it could be confusing and lead to inaccurate answers(response bias).

- Since there are three public high schools in the student's county, which sampling method could she use to collect the sample?
  She could use a stratified sample consisting of three strata, one for each high school in her county. She could randomly select students by using student IDs. She could ask for permission to access the demographic information to reduce the chance of under coverage and eliminate non-response, response, and voluntary response bias.

# 4 Random Sampling

The purpose of taking a simple random sample from a population are:

- To produce a sample that is representative of the population.
- To eliminate bias in selecting the sample.
- To allow us to generalize our results from the sample to the population.

**Example:** Elyse is surveying the students at her campus to determine how much they value eco-friendly options when they shop. Elyse would like to collect 200 samples. Which of the following sampling strategies is an example of simple random sampling?

- Creating an email list of every student enrolled in campus and randomly selecting 200 emails to send a digital survey to
- Asking every student on campus to complete the survey

- Using the student ID numbers for actively enrolled students and randomly selecting 200 ID numbers and then locating and surveying those students
- Randomly selecting 40 students from each of the 5 departments at her school

Creating an email list of every student enrolled in campus and randomly selecting 200 emails to send a digital survey to is the correct answer as it gives each student in campus equal probability of being selected.

# 5 Observational studies and experimental design

**Key terms:**

**Factor of interest/Explanatory variable/Independent variable:** The variable that the researcher purposely changes or manipulates to see if it impacts a specific outcome is called an explanatory variable.

**Response variable/Dependent Variable:** An objective measure of the research question that is measured at the end of an experiment and compared across the different levels of the factor of interest (or explanatory variable). In simpler terms, a response variable measures an outcome or result of a study.

**Subjects**: The individuals studied in an experiment are often called subjects.

**Treatment**: A treatment is any specific experimental condition applied to the subjects. If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables.

**Placebo and placebo effect:** A placebo is a dummy treatment with no active ingredients. Many patients respond favorably to any treatment, even a placebo. This response to a dummy treatment is the placebo effect.

**Lurking variable:** A variable that is neither the explanatory variable nor the response variable but has a relationship with the response and the explanatory variable. It is not considered in the study but could influence the relationship between the variables in the study.

**Confounding variable:** Two variables are confounded when their effects on a response variable cannot be distinguished from each other. The confounded variables may be either explanatory variables or lurking variables.

Example: An optimistic account of learning online reports a study at Nova Southeastern University, Fort Lauderdale, Florida. The authors of the study claim that students taking undergraduate courses online were "equal in learning" to students taking the same courses in class.

- Subjects - College students are the subjects in this study
- Explanatory variable - Setting for learning (in class or online)
- Response variable - Student's score on a test at the end of the course

Example: Suppose we want to study the relationship between diet and heart disease risk. A lurking variable in this case might be genetics, as people with a family history of heart disease may be more likely to have a higher risk of heart disease. Genetics is a lurking variable because it is not directly measured or included in the analysis, but it may influence the relationship between diet and heart disease risk.

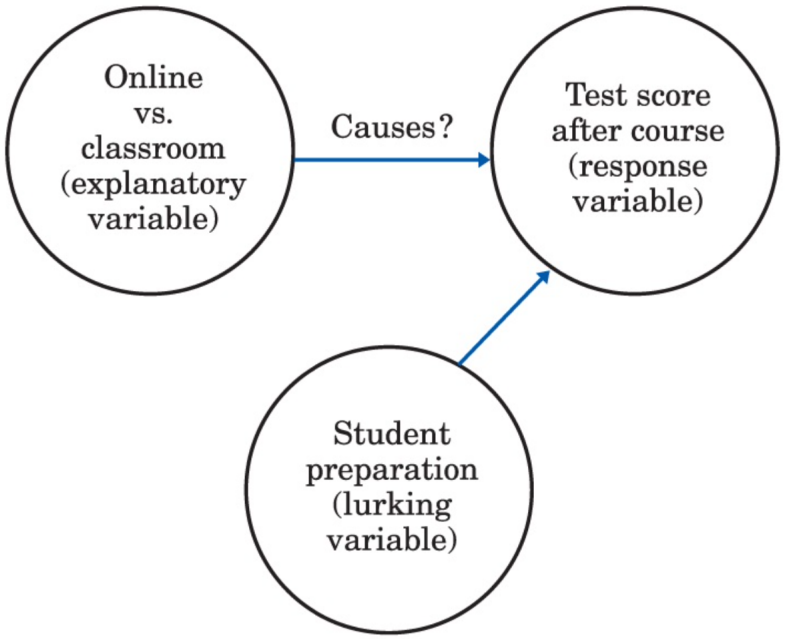
Observational studies: Researchers observe the population and record data without disrupting the population.

- This means that researches will take a "passive role" and simply record what they observe without having any control over the population and without making the population aware of their methods.
- Researchers can learn about characteristics of a population or compare two populations.
- An example would be observing behavioral tendencies in children. A researcher may collect a random sample of children and have them all play and interact with one another without having any influence on the children or telling them what to do.

Experiment:

- The first goal in designing an experiment is to ensure that it will show us the effect of the explanatory variables on the response variables. Confounding often prevents one-track experiments from doing this.

  Example: This figure below is an example of one track experiment. In this experiment, students chose for themselves whether to enroll in a classroom or an online version of a course. The study simply measured their learning.



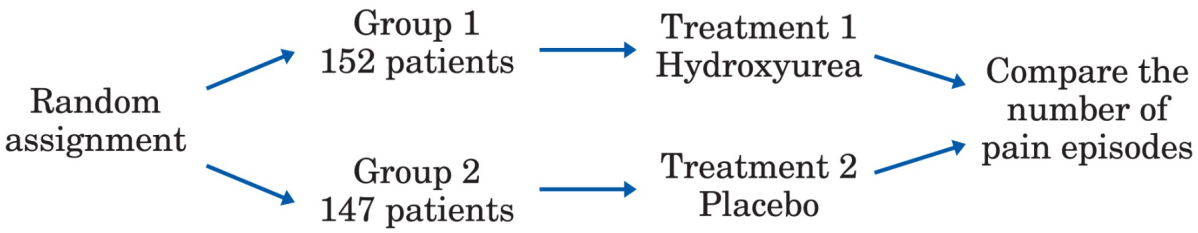

Moore/Notz, *Statistics: Concepts and Controversies*, 10e,
© 2020 W. H. Freeman and Company

- A good experiment compares two or more treatments. Subjects are divided into treatment group and control group. Control group are given a placebo and treatment group

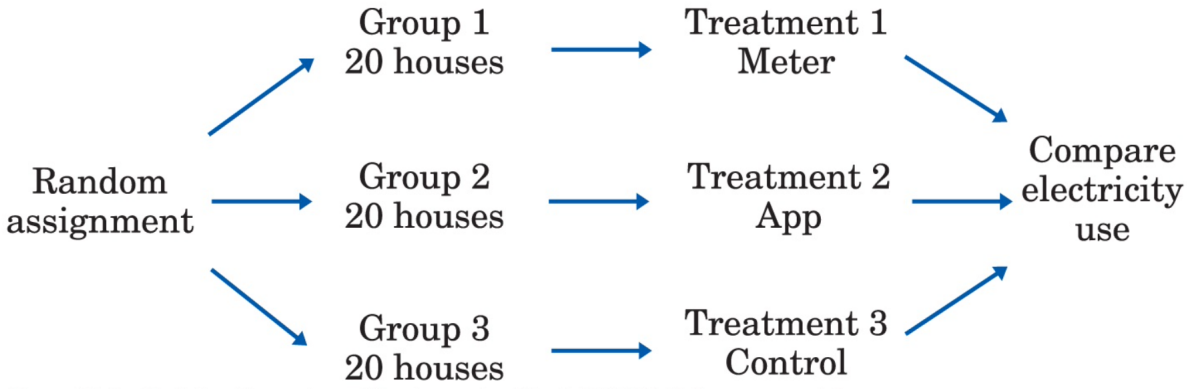
are given a treatment. Then, their responses are recorded and analyzed.

Example: The National Institutes of Health carried out a clinical trial of the drug hydroxyurea for treatment of sickle- cell anemia. The subjects were 299 adult patients who had had at least three episodes of pain from sickle-cell anemia in the previous year. Attached below outlines the desgion of a randomized comparative experiment to compare hydroxyurea with a placebo for treating sickle-cell anemia.




Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Example: An electricity company decides to design an experiment on 60 houses to monitor energy using two methods: placing a meter outside customers' house or providing the customers with an app to monitor their energy use. The experiment below compares these two approaches (meter, app) and also a control. The control group of customers receives information about energy conservation but no help in monitoring electricity use.




Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

## Experiment or observational study?

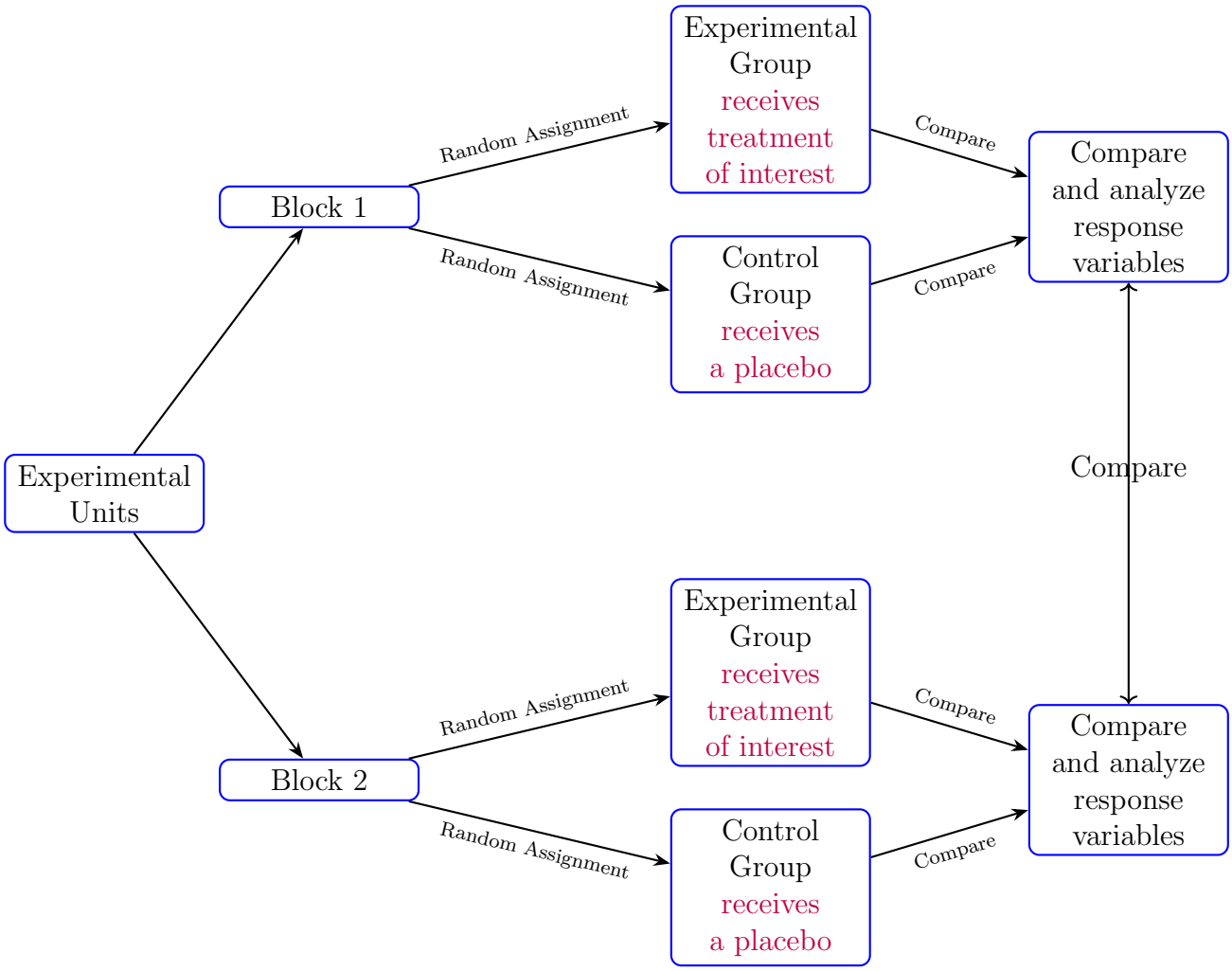
For the following two examples identify whether they are experiments or observational studies.

- A study took random sample of adults and asked them about their bedtime habits. The data showed that people who drank a cup of tea before bedtime were more likely to go to sleep earlier than those who didn't drink tea. Observational Study
- Another study took a group of adults and randomly divided them into two groups. One group was told to drink tea every night for a week, while the other group was told not to drink tea that week. Researchers then compared when each group fell asleep. Experiment
- A study took a random sample of people and examined their social media habits. Each person was classified as either a light, moderate, or heavy social media user. The researchers looked at which groups tended to be happier. Observational Study

Randomized block design:

Another way a researcher might conduct an experiment is by using a completely randomized block design. This design is used when the experimental units are divided into groups called blocks. Within each block, the experimental units are randomly assigned to treatments. The only difference to the outline of the experiments we studied above to randomized block design is creating several of the blocks. Randomized block design accounts for sampling variability between the experiments. The following diagram illustrates this design.

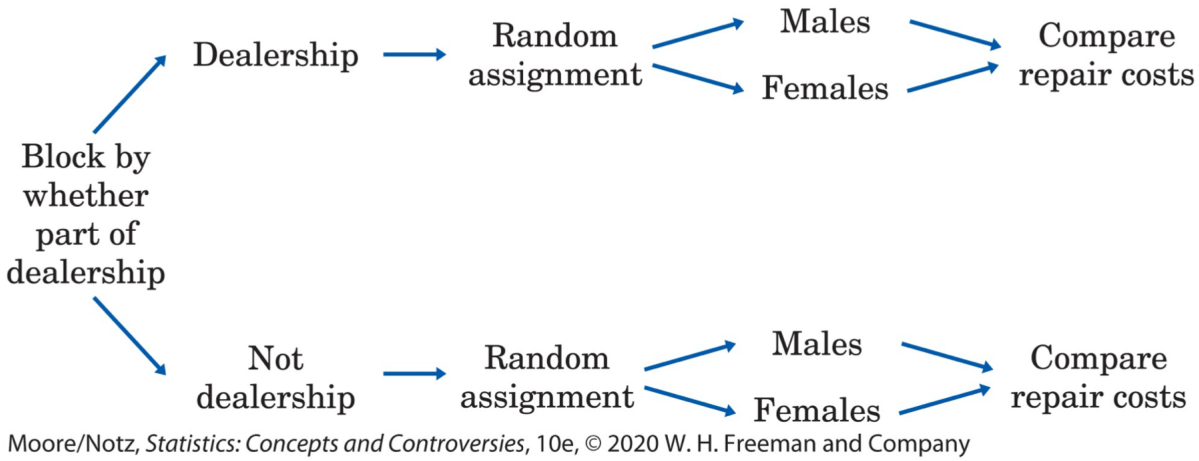



Randomized block designs are particularly useful in agricultural, medical, and psychological research, where they help to ensure that differences in response are due to the treatments themselves rather than other confounding factors. This approach improves the validity and reliability of the conclusions drawn from the experiment.

The most appropriate use of randomized complete block designs is when there is a known or suspected source of variation in one direction. The randomized complete block design is one of the most widely used designs in real world.

**Example: Car repairs.** Does your sex affect the price you are quoted for car repairs? A car with a specific problem is to be taken to repair shops in a large city for a quote on how much it will cost to have the problem fixed. For some repair shops, the person bringing in the car is a woman and for others a man. The researchers thought that there might be a difference between repair shops that are part of a dealership and those that are not. So the researchers decided to block on whether or not the repair shop was part of a dealership. Use a diagram to outline a block design for this experiment.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

**Try it yourself: Student Tutoring Effectiveness.** Does the type of tutoring method affect student performance in mathematics? A group of students with similar math proficiency levels is to be tutored using different methods, and their performance will be assessed afterward. The students are divided into three groups based on their learning preferences: visual learners, auditory learners, and kinesthetic learners. The researchers aim to determine whether the tutoring method has a different effect on these three groups of learners. Therefore, they decide to block based on learning preference (visual vs. auditory vs. kinesthetic). Use a diagram to outline a block design for this experiment.