

Section 1.1: Intro to Statistics

- **Population:** All people or things that you are studying
- **Parameter:** A numerical description measuring the variable of interest in the POPULATION
- **Sample:** A Representative subset of the population
- **Statistic:** A numerical description measuring the variable of interest in the SAMPLE
- **Variable:** Any characteristic, Number, or Quantity that can be measured or quantified.
- **Data:** Counts, Measurements or Observations Recorded about a specific variable.
- **Census:** When data is studied from every member of the population
- **Descriptive Statistics:** Data that is collected, organized, summarized and displayed.
- **Inferential Statistics:** Using data that is gathered to make **predictions (Inferential)** about the population parameter.

Section 1.2: Classifying Data

- **Qualitative Variable:** These variables represent categories. Numbers can be used here but we wouldn't find value in analyzing the numbers.
 - **Nominal:** data consisting of labels or names
 - **Ordinal:** data can be arranged in a meaningful order, **but calculations don't make sense.**
- **Quantitative Variable:** These variables represent measurable quantities and are expressed numerically. Typically records the amount of degree of something.
 - **Continuous:** Float data that can take on any value in an interval and are usually measurements.
 - **Interval:** Ratio: data can be arranged in a meaning order and differences between data entries are meaningful. **Zero has no meaning.**
 - **Ratio:** data can be arranged in a meaning order and differences between data entries are meaningful. **Zero indicates the absence of something.**
 - **Discrete:** Integers data that can take on only particular values and are Integer
 - **Interval:** Ratio: data can be arranged in a meaning order and differences between data entries are meaningful. **Zero has no meaning.**
 - **Ratio:** data can be arranged in a meaning order and differences between data entries are meaningful. **Zero indicates the absence of something**

Section 1.3.1: Introduction to Statistics Studies

- **Statistical Study:** In the real world we often want to answer a question or gather knowledge about a particular subject. To do this in a way in which our results can be trusted and verified, we must follow the process of a statistical study.
1. **Determine the design of the study:**
 - a. **What question are we trying to answer? (Why?)**
 - b. **What is the population? (Who?)**
 - c. **What variables will we record? (What?)**
 - d. What is the sampling method? (How?)
 - i. **Observation Study**
 - **Retroactive:** deals with actions or effects that influence the **past**
 - **Proactive:** involves taking measures to prepare for or prevent future events or issues
 - ii. **Experiments**
 2. Collect the data
 3. Organize the data Analyze the data to answer the question

Section 1.3.2: **Observation Study**

- **Observation Study:** collect data about our population. However, it is too costly or difficult to perform a census, so we instead find a sample that has the same **RELEVANT** characteristic as the population. **Representative sample** has the same relevant characteristic as the population and does not favor one group from the population over another.
- **Techniques use to collect Sample Data:**
 - **Random Sample:** broad term that covers any sampling method where the selection is random but may not be equal for all.
 - **Simple Random Sample (SRS):** ensures equal probability for every individual being selected.
 - **Stratified Sample:** the population is divided into groups or strata based on certain characteristics and then a random sample is taken from each group.
 - **Cluster Sample:** the population is divided into clusters, and entire clusters are randomly selected.
 - **Systematic Sample:** sampling method where you select every k^{th} individual from a list or population.
 - **Convenience Sample:** sample is a non-probability sampling method where the sample is chosen based on availability.
- **Types of Observational Studies:**
 - **Cross-Sectional Study:** Data are collected at a single point in time, providing a snapshot of a population or phenomenon at that moment.
 - **Longitudinal Study:** Data are gathered by following the same group or subjects over a period, allowing for the study of changes and developments.
 - **Meta-Analysis:** A study that compiles and statistically analyzes data from multiple previous studies to derive a more comprehensive conclusion.
 - **Case Study:** An in-depth examination of a single subject, event, or entity, looking at multiple variables that may influence the outcome or situation.

Section 1.3.3: **Experiments**

- **Experiments** are the only way to show a cause-and-effect relationship.
 - **Treatment:** A condition applied to a group of subjects in an experiment.
 - **Subjects:** Individuals (or things) being studied in an experiment; also referred to as participants.
 - **Response Variable:** The variable that is measured to assess the effects of the treatment; it responds to the changes made in the experiment.
 - **Explanatory Variable:** The variable that the researcher manipulates or changes to examine its effect on the response variable; also called the independent variable.
 - **Control group:** A group of subjects that do not receive the experimental treatment, serving as a baseline for comparison; they might receive a placebo or no treatment at all.
 - **Placebo effect:** A phenomenon where subjects experience a perceived improvement in their condition simply because they believe they are receiving treatment, even if the treatment is inactive.
 - **Treatment group:** A group of subjects that receive experimental treatment.

Section 1.3.3: Experiments (Continue)

- **Single Blind Experiment:** An experimental setup where the subjects are unaware of whether they are in the control group or the treatment group, but the researchers or individuals administering the experiment know the group assignments. This helps to reduce bias in subjects' responses.
- **Double Blind Experiment:** Neither the subjects nor the researchers administering the treatments know which subjects belong to the control or treatment group. This design further reduces bias by preventing both subjects and experimenters from influencing the outcomes based on their knowledge of group assignment.
- **Confounding Variables:** Variables that are not part of the treatment but may still affect the outcome of the experiment. These variables make it difficult to determine whether the observed effect is due to the treatment or some other factor.
- **Institutional Review Board (IRB):** A committee of experts that reviews the design and methodology of a study to ensure ethical standards are met and that no unnecessary harm will occur to participants. This is especially important for studies involving human subjects.
- **Informed Consent:** A process in which participants are fully informed about the purpose, procedures, risks, and benefits of a study before agreeing to take part. Participants' consent must be obtained voluntarily, ensuring they understand what the experiment involves.
- **Experiments: How to design an experiment**
 1. Randomize the control and treatment groups
 2. Control for outside effects on the response variable
 3. Replicate the experiment a significant number of times to see meaningful patterns.

Section 1.4: Critiquing a Published Study

1. Consider the Source

- Who paid for the study?
 - It's essential to know who funded the study, as the funding source could introduce bias. For example, if a pharmaceutical company funds a study on a new drug, there's a potential for conflict of interest.
- Where was the data collected?
 - The location of data collection matters, as geographical or cultural factors might influence the study's outcomes. Results may not be generalizable if the data were collected in a specific area that doesn't represent the larger population.
- When was the information collected?
 - Timing can impact study results. Data collected during a pandemic, for instance, might differ significantly from data collected under normal circumstances.
- Who published the study?
 - The reputation and rigor of the journal or organization that published the study are important. Was it peer-reviewed? Is the publisher reputable? Publishing in a respected journal usually indicates a higher level of scrutiny.

2. Consider the Variable

- Are the explanatory and response variables clearly defined?
- Are confounding variables identified and controlled?
- Are the variables relevant and appropriate for the study?

3. Consider the Setup

- **Bias:** Are there signs of favoring certain outcomes? Was the study designed or interpreted in a way that favors the sponsor or researchers' goals?
- **Sampling Bias:** Was the sample representative of the population? If certain groups were over- or underrepresented, the results might not be generalizable.
- **Dropouts:** Are participants who begin a study but fail to complete it.
- **Processing errors:** Are errors that occur from the data being processed, such as typos.
- **Nonadherent:** Participant who remains in the study but strays from the directions.
- **Researcher** Did the researchers' expectations or actions influence the results?
- **Response bias:** Occurs when a researchers behavior causes a participant to alter their response or when a participant gives an inaccurate response.
- **Participation bias:** Occurs when there is a problem with the participation or lack thereof -of those chosen for the study.
- **Nonresponse bias:** Occurs when there is a lack of participation in a self-selected sample from certain segments of a population, when a person refuses to participate in a survey, or when a respondent omits questions when answering a survey.

4. Consider the Conclusion

- Do the data support the conclusion?
- Do the results present the whole picture or just a part?
- Could there be another conclusion drawn?
- Could there be other reasons for the same conclusion drawn?
- Does the study have any practical applications?