

# Data Collection

## Key Definitions

- **Statistics:** Collecting, organizing, summarizing, and analyzing information to make a conclusion.
- **Data:** Information that is a fact or proposition that is used to draw a conclusion or to make a decision.
- **Population:** The entire group that is being studied. An example of a population is the student body of Regent University.
- **Sample:** A portion of the population that is being studied. An example of a sample is all English majors at Regent University.
- **Qualitative Variables:** Variables that are classified using attributes or characteristics. These would be classifications like gender, how good was your experience, etc.
- **Quantitative Variables:** Variables that are numerical measurements. These variables would come from questions such as, "How many children do you have?"
- **Discrete Variables:** A quantitative variable that is finite. This variable is countable. An example could be the number of people in a room.
- **Continuous Variables:** A quantitative variable that has an infinite number of outcomes. Continuous is not countable. An example could be the temperature as read by a thermometer.
- **Nominal Level of Measurement:** A qualitative, variable value that has no ranking or specific order. An example of a nominal level of measurement is gender since neither female nor male is ranked higher than the other.
- **Ordinal Level of Measurement:** A qualitative, variable value that has order or ranking. An example of an ordinal level of measurement is options such as "Very Good", "Good", "Impartial", "Bad", and "Very Bad," since "Very Bad" is worse than "Good".
- **Interval Level of Measurement:** A quantitative, variable value where zero does not mean zero. An example of this is temperature. Depending whether you use Celsius or Fahrenheit, zero has two different meaning, and neither of them mean that there is no temperature.
- **Ratio Level of Measurement:** A quantitative, variable value where zero means zero. An example is if you have no children and you are asked how many children you have, your answer would be zero since you have no children.
- **Explanatory Variable:** A variable in an experiment that is being tested as an explanation according to the hypothesis. This variable is also known as an independent variable.
- **Response Variable:** A variable in an experiment that is affected and responds to the explanatory variable. This is also known as the dependent variable.
- **Observational Study:** An observational study is done where the researcher only observes the behaviors and does not try to influence any of the variables.

- **Experiment:** An experiment is done where the researcher changes the value of the explanatory variable by introducing something new to the environment of the study.
- **Random Sampling:** In choosing your sample from a population, you leave it to chance as to who is part of the sample.
- **Simple Random Sample:** This sample is obtained by randomly selecting your sample from the population. The simple random sample is not to be confused with the convenience sample.
- **Stratified Sample:** This sample is obtained from separating the population into different, non-overlapping groups called strata, and then you get a simple random sample from each of the groups.
- **Systematic Sample:** This sample is obtained by separating the population into groups of a randomly chosen  $k$ th number, and then you select a random number between 1 and  $k$  called  $j$ . You, then, survey every  $j$ th person in the groups until you reach the sample size.
- **Cluster Sample:** This sample is obtained by selecting random samples of the population and surveying all individuals that are in the randomly selected groups.
- **Convenience Sample:** This is obtained by whichever individuals are easily obtained to survey. It is not based on randomness.
- **Bias:** A bias is when the results of a sample are not representative of a population. An example of this is if Regent University was surveying all the students enrolled in a MATH 201 course, but they only surveyed 1 of the classes as a representative of all the MATH 201 classes because it is possible that classes with a different professor or class location had a different experience.
- **Sampling Bias:** The technique used to obtain the sample favors one subgroup of the population over another.
- **Response Bias:** This bias happens when the answers on a survey do not reflect how the person being surveyed feels because they are aware that they are being studied.

## Applications

- **Explanatory Variable vs. Response Variable:**  
If you were to have a study about the association between eating habits and fatigue, where you see if what a person eats effects how fatigued they feel, you can distinguish the explanatory variable from the response variable. The explanatory variable being studied is the person's eating habits and the response variable being studied is how fatigued the person felt. This is because they are studying the effects of the explanatory variable on the response variable. The explanatory variable explains and the response variable responds to the explanation.
- **Observational Study vs. Experiment:**  
An observational study is a study where you do not change but only observe particular behaviors. An example of this is when you ask 200 people whether they prefer a burger from McDonald's or Burger King and report their responses. An experiment is where you

would give your sample two burgers, one from McDonald's and one from Burger King (without revealing which burger belongs to which chain), to 200 people and ask them which they prefer. This is an experiment because you are not just observing their opinion.

- **How to tell the difference between a Simple Random Sample and a Convenience Sample:**  
A convenience sample is more like being asked to take a survey after completing a phone call. Every person is given the option to take the survey and whoever they get is based off of the convenience of the person taking the survey. A simple random sample is more like having a poll and selecting a group of people that used your service to take the poll. A simple random sample is randomly selected by whoever is conducting the survey, and a convenience sample is offered to everyone to take.
- **Systematic Sample Example:**  
A company is conducting a survey on how the customers feel about the services they received. They decide to use a systematic sample to survey their customers. They want a sample size of 20 people. They decided they would pick one person every 14<sup>th</sup> person ( $k=14$ ). Between 1 and 14, they chose to survey the 6<sup>th</sup> person ( $j=6$ ). So they survey customers 6, 20 ( $k + j = 14+6$ ), 34 ( $20+14$ ), 48 ( $34+14$ ), ..., 272.
- **Sampling Bias vs. Response Bias:**  
A survey is distributed to each class you take here at Regent University. If the survey was distributed and collected by your professor, that would be an example of a response bias since you are more likely to give better reviews knowing it is being distributed and collected by your professor. If the survey was only given to classes that have an average grade of an A or A-, that would be a sampling bias since they are not representing the entirety of the class. This is why the surveys are distributed the way they are at Regent University to avoid bias.