Collecting Data

Department of Mathematics

Salt Lake Community College

(Slides by Adam Wilson)

Definition

The **population** of a study is the group the collected data is intended to describe.

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

- All first-year community college students in the state of Washington.

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the state of Washington.

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the entire United States.

Definition

The **population** of a study is the group the collected data is intended to describe.

Example 1

If we wish to study the amount of money spent on textbooks by first-year students, our intended population may be any of:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the entire United States.

Note

Sometimes the intended population is referred to as the **target population**.

Definition

A **parameter** is a value (average, percentage, etc) calculated using all the data from a population.

Definition

A **parameter** is a value (average, percentage, etc) calculated using all the data from a population.

Note

Parameters are seldom used, since it's hard to survey an entire population.

Definition

A **parameter** is a value (average, percentage, etc) calculated using all the data from a population.

Note

Parameters are seldom used, since it's hard to survey an entire population.

Definition

A survey of an entire population is called a **census**.

Definition

A **parameter** is a value (average, percentage, etc) calculated using all the data from a population.

Note

Parameters are seldom used, since it's hard to survey an entire population.

Definition

A survey of an entire population is called a **census**.

Example 2

Every ten years the United States federal government attempts to collect demographic information from every resident.

Definition

A **parameter** is a value (average, percentage, etc) calculated using all the data from a population.

Note

Parameters are seldom used, since it's hard to survey an entire population.

Definition

A survey of an entire population is called a **census**.

Example 2

Every ten years the United States federal government attempts to collect demographic information from every resident.

Example 3

A store manager wants to know the average number of hours her employees worked in the last month.

Definition

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

Definition

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

Example 4

To determine the average length of trout in a lake, researchers catch 20 fish and measure them.

What is the sample?

Definition

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

Example 4

To determine the average length of trout in a lake, researchers catch 20 fish and measure them.

What is the sample?

The trout the researchers caught.

Definition

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

Example 4

To determine the average length of trout in a lake, researchers catch 20 fish and measure them.

What is the sample?

The trout the researchers caught.

What is the population?

Definition

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

Example 4

To determine the average length of trout in a lake, researchers catch 20 fish and measure them.

What is the sample?

The trout the researchers caught.

What is the population?

All the trout in the lake.

Definition

A **statistic** is a value (average, percentage, etc) calculated using the data from a sample.

Definition

A **statistic** is a value (average, percentage, etc) calculated using the data from a sample.

Example 5

A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. Of those asked, 60% indicate they are supportive of the initiative.

What is the sample and population?

Definition

A **statistic** is a value (average, percentage, etc) calculated using the data from a sample.

Example 5

A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. Of those asked, 60% indicate they are supportive of the initiative.

What is the sample and population?

> *The sample is the 500 shoppers questioned. While the intended population of this survey was all Tacoma citizens, the effective population was mall shoppers. There is no reason to assume that the mall shoppers would be representative of all Tacoma citizens.*

Definition

A **statistic** is a value (average, percentage, etc) calculated using the data from a sample.

Example 5

A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. Of those asked, 60% indicate they are supportive of the initiative.

What is the sample and population?

> *The sample is the 500 shoppers questioned. While the intended population of this survey was all Tacoma citizens, the effective population was mall shoppers. There is no reason to assume that the mall shoppers would be representative of all Tacoma citizens.*

Is the 60% value a parameter or a statistic?

Definition

A **statistic** is a value (average, percentage, etc) calculated using the data from a sample.

Example 5

A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. Of those asked, 60% indicate they are supportive of the initiative.

What is the sample and population?

> *The sample is the 500 shoppers questioned. While the intended population of this survey was all Tacoma citizens, the effective population was mall shoppers. There is no reason to assume that the mall shoppers would be representative of all Tacoma citizens.*

Is the 60% value a parameter or a statistic?

> *The 60% value was based on a sample, so it is a statistic*

Categorizing Data

Categorical (qualitative) data are pieces of information that allow us to classify the objects under investigation into various categories.
Quantitative data are responses that are numerical in nature and which allow us to perform meaningful arithmetic calculations.

Categorizing Data

Categorical (qualitative) data are pieces of information that allow us to classify the objects under investigation into various categories.
Quantitative data are responses that are numerical in nature and which allow us to perform meaningful arithmetic calculations.

Example 6

We could survey the class to find out what each persons favorite movie.

Categorizing Data

Categorical (qualitative) data are pieces of information that allow us to classify the objects under investigation into various categories.
Quantitative data are responses that are numerical in nature and which allow us to perform meaningful arithmetic calculations.

Example 6

We could survey the class to find out what each persons favorite movie.

When we conduct such a survey, the responses would look like:

- Finding Nemo
- Plan 9 from Outer Space
- The Avengers

There are no numerical calculations we can perform on "Finding Nemo". We would have qualitative data.

Categorizing Data

Categorical (qualitative) data are pieces of information that allow us to classify the objects under investigation into various categories.
Quantitative data are responses that are numerical in nature and which allow us to perform meaningful arithmetic calculations.

Example 6

We could survey the class to find out what each persons favorite movie.

When we conduct such a survey, the responses would look like:

- Finding Nemo
- Plan 9 from Outer Space
- The Avengers

There are no numerical calculations we can perform on "Finding Nemo". We would have qualitative data.

If instead we asked how many movies each person in class saw in 2016, this would be quantitative data.

Definition

A sampling method is **biased** if every member of the population doesn't have equal likelihood of being in the sample.

Definition

A sampling method is **biased** if every member of the population doesn't have equal likelihood of being in the sample.

Definition

A **random sample** is one in which each member of the population has an equal probability of being chosen. A **simple random sample** is one in which every member of the population and any group of members has an equal probability of being chosen.

Definition

A sampling method is **biased** if every member of the population doesn't have equal likelihood of being in the sample.

Definition

A **random sample** is one in which each member of the population has an equal probability of being chosen. A **simple random sample** is one in which every member of the population and any group of members has an equal probability of being chosen.

Definition

The natural variation of samples is called **sampling variability**.

Definition

A sampling method is **biased** if every member of the population doesn't have equal likelihood of being in the sample.

Definition

A **random sample** is one in which each member of the population has an equal probability of being chosen. A **simple random sample** is one in which every member of the population and any group of members has an equal probability of being chosen.

Definition

The natural variation of samples is called **sampling variability**.

Note

Variability is unavoidable and expected in random sampling, and in most cases is not an issue.

Definition

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

Definition

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

Example 7

Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents.

Definition

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

Example 7

Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents.

In a sample of 1000 people, they would then expect to get about 390 Democrats, 370 Republicans and 240 independents.

Definition

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

Example 7

Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents.

In a sample of 1000 people, they would then expect to get about 390 Democrats, 370 Republicans and 240 independents.

To accomplish this, they could randomly select 390 people from among those voters known to be Democrats, 370 from those known to be Republicans, and 240 from those with no party affiliation.

Definition

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

Definition

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

Example 8

Suppose the pollsters call people at random, but once they have met their quota of 390 Democrats, they only gather people who do not identify themselves as a Democrat.

Definition

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

Example 8

Suppose the pollsters call people at random, but once they have met their quota of 390 Democrats, they only gather people who do not identify themselves as a Democrat.

Definition

In **cluster sampling**, the population is divided into subgroups (clusters), and a set of subgroups is selected to be in the sample.

Definition

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

Example 8

Suppose the pollsters call people at random, but once they have met their quota of 390 Democrats, they only gather people who do not identify themselves as a Democrat.

Definition

In **cluster sampling**, the population is divided into subgroups (clusters), and a set of subgroups is selected to be in the sample.

Example 9

If the college wanted to survey students, since students are already divided into classes, they could randomly select 10 classes and give the survey to all the students in those classes.

Definition

In **systematic sampling**, every n^{th} member of the population is selected to be in the sample.

Definition

In **systematic sampling**, every n^{th} member of the population is selected to be in the sample.

Example 10

To select a sample using systematic sampling, a pollster calls every 100th name in the phone book.

Definition

In **systematic sampling**, every n^{th} member of the population is selected to be in the sample.

Example 10

To select a sample using systematic sampling, a pollster calls every 100th name in the phone book.

Note

Systematic sampling is not as random as a simple random sample but it can yield acceptable samples. (If your name is Alexis Aardvark and your sister Alice Aardvark is right after you in the phone book, there is no way you could both end up in the sample.)

Bad Sampling Methods

- **Convenience sampling** is samples chosen by selecting whoever is convenient.
- **Voluntary response sampling** is allowing the sample to volunteer.

Bad Sampling Methods

- **Convenience sampling** is samples chosen by selecting whoever is convenient.
- **Voluntary response sampling** is allowing the sample to volunteer.

Example 11

A pollster stands on a street corner and interviews the first 100 people who agree to speak to him. This is a convenience sample.

Bad Sampling Methods

- **Convenience sampling** is samples chosen by selecting whoever is convenient.
- **Voluntary response sampling** is allowing the sample to volunteer.

Example 11

A pollster stands on a street corner and interviews the first 100 people who agree to speak to him. This is a convenience sample.

Example 12

A website has a survey asking readers to give their opinion on a tax proposal. This is a self-selected sample, or voluntary response sample, in which respondents volunteer to participate.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.
   Voluntary response

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.
   Voluntary response
4. A website randomly selects 50 of their customers to send a survey to.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.
   Voluntary response
4. A website randomly selects 50 of their customers to send a survey to.
   Simple Random

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.
   Voluntary response
4. A website randomly selects 50 of their customers to send a survey to.
   Simple Random
5. To survey voters in a town, a polling company randomly selects 10 city blocks, and interviews everyone who lives on those blocks.

Example 13

Which sampling method was used in each case?

1. Every 4th person in the class was selected.
   Systematic
2. A sample was selected to contain 25 men and 35 women.
   Stratified or Quota
3. Viewers of a new show are asked to vote on the show's website.
   Voluntary response
4. A website randomly selects 50 of their customers to send a survey to.
   Simple Random
5. To survey voters in a town, a polling company randomly selects 10 city blocks, and interviews everyone who lives on those blocks.
   Cluster

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Self-interest study: Bias that can occur when the researchers have an interest in the outcome.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Self-interest study: Bias that can occur when the researchers have an interest in the outcome.

Response bias: When the responder gives inaccurate responses for any reason.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Self-interest study: Bias that can occur when the researchers have an interest in the outcome.

Response bias: When the responder gives inaccurate responses for any reason.

Perceived lack of anonymity: When the responder fears giving an honest answer for any reason.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Self-interest study: Bias that can occur when the researchers have an interest in the outcome.

Response bias: When the responder gives inaccurate responses for any reason.

Perceived lack of anonymity: When the responder fears giving an honest answer for any reason.

Loaded (or leading) questions: When the question wording influences the responses.

Sources of Bias

Sampling bias: When the sample is not representative of the population.

Voluntary response bias: the sampling bias the ofter occurs when the sample is made of volunteers.

Self-interest study: Bias that can occur when the researchers have an interest in the outcome.

Response bias: When the responder gives inaccurate responses for any reason.

Perceived lack of anonymity: When the responder fears giving an honest answer for any reason.

Loaded (or leading) questions: When the question wording influences the responses.

Non-response bias: When people refusing to participate in the study can influence the validity of the outcome.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.
   Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.
   Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.
   Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll.
   Voluntary response bias. The sample is self-selected.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.
   Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll.
   Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks. Lack of anonymity.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks. Lack of anonymity.
5. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks. Lack of anonymity.
5. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk. Self-interest study.

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year. Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll. Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks. Lack of anonymity.
5. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk. Self-interest study.
6. A poll asks "Do you support a new transportation tax, or would you prefer to see our public transportation system fall apart?"

Example 14

In each case, what potential source of bias exists?

1. A survey asks how many sexual partners a person has had this year.
   Response bias. Historically, men are likely to over-report, and women are likely to under-report their sexual partners.
2. A radio station asks readers to phone in their choice in a daily poll.
   Voluntary response bias. The sample is self-selected.
3. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. Sampling bias.
4. High school students are asked if they have consumed alcohol in the last two weeks. Lack of anonymity.
5. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk. Self-interest study.
6. A poll asks "Do you support a new transportation tax, or would you prefer to see our public transportation system fall apart?"
   Loaded question.

Definition

An **observational study** is a study based on observations or measurements.

Definition

An **observational study** is a study based on observations or measurements.

Note

Most of the studies we have discussed are **observational**.

Definition

An **observational study** is a study based on observations or measurements.

Note

Most of the studies we have discussed are **observational**.

Definition

An **experiment** is a study in which the effects of a **treatment** are measured.

Definition

An **observational study** is a study based on observations or measurements.

Note

Most of the studies we have discussed are **observational**.

Definition

An **experiment** is a study in which the effects of a **treatment** are measured.

Example 15

A pharmaceutical company tests a new medicine for treating Alzheimer?s disease by administering the drug to 50 elderly patients with recent diagnoses. The treatment here is the new drug.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.
   Observational study.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.
   Observational study.
2. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.
   Observational study.
2. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.
   Experiment, the treatment is the jumping jacks.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.
   Observational study.
2. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.
   Experiment, the treatment is the jumping jacks.
3. Twenty coffee drinkers and twenty tea drinkers are given a concentration test.

Example 16

Is each scenario describing an observational study or an experiment?

1. The weights of 30 randomly selected people are measured.
   Observational study.
2. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.
   Experiment, the treatment is the jumping jacks.
3. Twenty coffee drinkers and twenty tea drinkers are given a concentration test.
   Experiment, the treatments are coffee and tea.

Definition

Confounding occurs when there are two potential variables that could have caused the outcome and it is not possible to determine which actually caused the result.

Definition

Confounding occurs when there are two potential variables that could have caused the outcome and it is not possible to determine which actually caused the result.

Example 17

A drug company study about a weight loss pill reports that people lost an average of 8 pounds while using their new drug.

Definition

Confounding occurs when there are two potential variables that could have caused the outcome and it is not possible to determine which actually caused the result.

Example 17

A drug company study about a weight loss pill reports that people lost an average of 8 pounds while using their new drug.

However, in the fine print you find a statement saying that participants were encouraged to also diet and exercise.

Definition

Confounding occurs when there are two potential variables that could have caused the outcome and it is not possible to determine which actually caused the result.

Example 17

A drug company study about a weight loss pill reports that people lost an average of 8 pounds while using their new drug.

However, in the fine print you find a statement saying that participants were encouraged to also diet and exercise.

It is not clear in this case whether the weight loss is due to the pill, to diet and exercise, or a combination of both. In this case confounding has occurred.

Definition

When using a **control group**, the participants are divided into two or more groups, typically a **control group** and a **treatment group**. The treatment group receives the treatment being tested and the control group does not receive the treatment.

Definition

When using a **control group**, the participants are divided into two or more groups, typically a **control group** and a **treatment group**. The treatment group receives the treatment being tested and the control group does not receive the treatment.

Example 18

To determine if a two day prep course would help high school students improve their scores on the SAT test, a group of students was randomly divided into two subgroups. The first group, the treatment group, was given the prep course. The second group, the control group, was not. Afterwards, both groups were given the SAT.

Definition

When using a **control group**, the participants are divided into two or more groups, typically a **control group** and a **treatment group**. The treatment group receives the treatment being tested and the control group does not receive the treatment.

Example 18

To determine if a two day prep course would help high school students improve their scores on the SAT test, a group of students was randomly divided into two subgroups. The first group, the treatment group, was given the prep course. The second group, the control group, was not. Afterwards, both groups were given the SAT.

Example 19

A company testing a new plant food grows two crops of the same plants in adjacent fields, the treatment group receiving the new plant food and the control group not. The crop yield would then be compared. By growing them at the same time in adjacent fields, they avoid confounding factors.

Definition

The **placebo effect** is when the effectiveness of a treatment is influenced by the patient's perception of how effective they think the treatment will be, so a result might be seen even if the treatment is ineffectual.

[1] Levine JD, Gordon NC, Smith R, Fields HL. (1981) Analgesic responses to morphine and placebo in individuals with postoperative pain. Pain. 10:379-89.

Definition

The **placebo effect** is when the effectiveness of a treatment is influenced by the patient's perception of how effective they think the treatment will be, so a result might be seen even if the treatment is ineffectual.

Example 20

A study[1] found that when doing painful dental tooth extractions, patients told they were receiving a strong painkiller while actually receiving a saltwater injection found as much pain relief as patients receiving a dose of morphine.

[1] Levine JD, Gordon NC, Smith R, Fields HL. (1981) Analgesic responses to morphine and placebo in individuals with postoperative pain. Pain. 10:379-89.

Definition

A **placebo** is a fake treatment given to control for the placebo effect. An experiment that gives the control group a placebo is called a **placebo controlled experiment**.

Definition

A **placebo** is a fake treatment given to control for the placebo effect. An experiment that gives the control group a placebo is called a **placebo controlled experiment**.

Note

In some cases it is more appropriate to compare to a conventional treatment than a placebo.

For example, in a cancer study, it would not be ethical to deny any treatment to the control group or to give a placebo treatment. In this case, the currently acceptable medical treatment would be given to the second group, called the **comparison group**.

Example 21

- In a study for a new medicine that is dispensed in a pill form, a sugar pill could be used as a placebo.

Example 21

- In a study for a new medicine that is dispensed in a pill form, a sugar pill could be used as a placebo.
- In a study on the effect of alcohol on memory, a non-alcoholic beer might be given to the control group as a placebo.

Example 21

- In a study for a new medicine that is dispensed in a pill form, a sugar pill could be used as a placebo.
- In a study on the effect of alcohol on memory, a non-alcoholic beer might be given to the control group as a placebo.
- In a study of a frozen meal diet plan, the treatment group would receive the diet food, and the control could be given standard frozen meals stripped of their original packaging.

Definition

A **blind study** is one in which the participant does not know whether or not they are receiving the treatment or a placebo.

Definition

A **blind study** is one in which the participant does not know whether or not they are receiving the treatment or a placebo.

Example 22

A classic example is the Pepsi Challenge. A tester, often a marketing person, prepares two sets of cups of cola labeled "A" and "B". One set of cups is filled with Pepsi, while the other is filled with Coca-Cola. The tester knows which soda is in which cup but is not supposed to reveal that information to the subjects. Volunteer subjects are encouraged to try the two cups of soda and polled for which ones they prefer.

Definition

A **blind study** is one in which the participant does not know whether or not they are receiving the treatment or a placebo.

Example 22

A classic example is the Pepsi Challenge. A tester, often a marketing person, prepares two sets of cups of cola labeled "A" and "B". One set of cups is filled with Pepsi, while the other is filled with Coca-Cola. The tester knows which soda is in which cup but is not supposed to reveal that information to the subjects. Volunteer subjects are encouraged to try the two cups of soda and polled for which ones they prefer.

Note

One problem with a single-blind test like this, is that the tester can unintentionally give subconscious cues which influence the subjects.

Definition

A **double-blind study** is one in which those interacting with the participants don't know who is in the treatment group and who is in the control group.

Definition

A **double-blind study** is one in which those interacting with the participants don't know who is in the treatment group and who is in the control group.

Example 23

In a study about anti-depression medicine, you would not want the psychological evaluator to know whether the patient is in the treatment or control group either, as it might influence their evaluation, so the experiment should be conducted as a double-blind study.