

# Types of Statistics; Variables and Types of Data

1.1 and 1.2

---

Jacob Ayers

MAT 110

Lesson #1

# Objectives

- Demonstrate knowledge of statistical terms
- Differentiate between descriptive and inferential statistics
- Identify types of data
- Identify levels of measurement

## Definition

*Statistics* is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

# Introduction to Statistics

## Definition

*Statistics* is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

You are probably familiar with basic concepts of probability through life experience, but high school math generally does not spend much time on this important branch of mathematics.

Statistics is used in just about every field:

## Definition

*Statistics* is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

You are probably familiar with basic concepts of probability through life experience, but high school math generally does not spend much time on this important branch of mathematics.

Statistics is used in just about every field:

- Sports: keeping records of players' performance

## Definition

*Statistics* is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

You are probably familiar with basic concepts of probability through life experience, but high school math generally does not spend much time on this important branch of mathematics.

Statistics is used in just about every field:

- Sports: keeping records of players' performance
- Medical: determine mortality rate of a virus

## Definition

*Statistics* is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

You are probably familiar with basic concepts of probability through life experience, but high school math generally does not spend much time on this important branch of mathematics.

Statistics is used in just about every field:

- Sports: keeping records of players' performance
- Medical: determine mortality rate of a virus
- Government: determine voting rate, polling

# Introduction to Statistics

Reasons to study statistics:



# Introduction to Statistics

Reasons to study statistics:

- Learn to read and understand studies performed in your desired field

# Introduction to Statistics

Reasons to study statistics:

- Learn to read and understand studies performed in your desired field
- Learn methods of conducting research that you can use in your field

# Introduction to Statistics

Reasons to study statistics:

- Learn to read and understand studies performed in your desired field
- Learn methods of conducting research that you can use in your field
- Become a better consumer and citizen

## Basic Terms

Statisticians study variables. In statistics, a *variable* is a characteristic or attribute that can assume different values.

## Basic Terms

Statisticians study variables. In statistics, a *variable* is a characteristic or attribute that can assume different values.

*Data* are the values that the variables can assume.

## Basic Terms

Statisticians study variables. In statistics, a *variable* is a characteristic or attribute that can assume different values.

*Data* are the values that the variables can assume.

We will be studying random variables in this course. A *random variable* is a variable whose value is determined by chance.

## Basic Terms

Statisticians study variables. In statistics, a *variable* is a characteristic or attribute that can assume different values.

*Data* are the values that the variables can assume.

We will be studying random variables in this course. A *random variable* is a variable whose value is determined by chance.

To study variables, statisticians collect data. This collection of data forms a *data set*. Each value in the set is called a *data value*.

## Basic Terms

We will often need to distinguish between a population and a sample.



## Basic Terms

We will often need to distinguish between a population and a sample.

A *population* consists of all subjects that are being studied.

## Basic Terms

We will often need to distinguish between a population and a sample.

A *population* consists of all subjects that are being studied.

A *sample* is a group of subjects that is selected from a population.

## Basic Terms

We will often need to distinguish between a population and a sample.

A *population* consists of all subjects that are being studied.

A *sample* is a group of subjects that is selected from a population.

One example of a study involving an entire population is the U.S. Census.

## Basic Terms

We will often need to distinguish between a population and a sample.

A *population* consists of all subjects that are being studied.

A *sample* is a group of subjects that is selected from a population.

One example of a study involving an entire population is the U.S. Census.

Most of the time, studies involve samples because it is very difficult to obtain data from an entire population.

## Basic Terms

When studying a sample, statisticians make efforts to ensure that the sample they select is representative of the population.

## Basic Terms

When studying a sample, statisticians make efforts to ensure that the sample they select is representative of the population.

Information is said to be *biased* if

## Basic Terms

When studying a sample, statisticians make efforts to ensure that the sample they select is representative of the population.

Information is said to be *biased* if

- the results from the sample are significantly different from the results of a census of the population, or
- the sample is not representative of the population

# Branches of Statistics

There are two main areas of statistics.



# Branches of Statistics

There are two main areas of statistics.

## **Descriptive Statistics**

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

# Branches of Statistics

There are two main areas of statistics.

## **Descriptive Statistics**

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

Key concept: In descriptive statistics, our goal is only to *describe* situations. We are not seeking to make predictions.

# Branches of Statistics

There are two main areas of statistics.

## **Descriptive Statistics**

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

Key concept: In descriptive statistics, our goal is only to *describe* situations. We are not seeking to make predictions.

Examples of descriptive statistics:

# Branches of Statistics

There are two main areas of statistics.

## Descriptive Statistics

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

Key concept: In descriptive statistics, our goal is only to *describe* situations. We are not seeking to make predictions.

Examples of descriptive statistics:

- Calculating median household income using census data
- Calculating a baseball player's batting average
- Rolling a die 100 times, then making a bar graph to illustrate how often each number comes up

# Branches of Statistics

## Inferential Statistics

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships between variables, and making predictions.

# Branches of Statistics

## Inferential Statistics

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships between variables, and making predictions.

Key concept: In inferential statistics, our goal is to use the data we have to draw an inference (prediction) of some sort.

# Branches of Statistics

## Inferential Statistics

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships between variables, and making predictions.

Key concept: In inferential statistics, our goal is to use the data we have to draw an inference (prediction) of some sort.

Examples of inferential statistics:

# Branches of Statistics

## Inferential Statistics

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships between variables, and making predictions.

Key concept: In inferential statistics, our goal is to use the data we have to draw an inference (prediction) of some sort.

Examples of inferential statistics:

- Determine whether median household income of a neighborhood is correlated with crime rates
- Using batting average (and other statistics) to estimate how much you should pay a player



## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

In a weight loss study using teenagers at Boston University, 52% of the group said that they lost weight and kept it off by counting calories.

Based on a sample of 2739 respondents, it is estimated that pet owners spend a total of 14 billion dollars on veterinary care for their pets.

## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

In a weight loss study using teenagers at Boston University, 52% of the group said that they lost weight and kept it off by counting calories.

No prediction or generalization is being made. Descriptive statistics were used.

Based on a sample of 2739 respondents, it is estimated that pet owners spend a total of 14 billion dollars on veterinary care for their pets.

## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

In a weight loss study using teenagers at Boston University, 52% of the group said that they lost weight and kept it off by counting calories.

No prediction or generalization is being made. Descriptive statistics were used.

Based on a sample of 2739 respondents, it is estimated that pet owners spend a total of 14 billion dollars on veterinary care for their pets.

An inference is being drawn based on the data from the sample. Inferential statistics were used.

## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

A study conducted by a research network found that people with fewer than 12 years of education had lower life expectancies than those with more years of education.

A survey of 1507 smartphone users showed that 38% of them purchased insurance at the same time as they purchased the phone.

## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

A study conducted by a research network found that people with fewer than 12 years of education had lower life expectancies than those with more years of education.

An inference is being drawn about the relationship between two variables. Inferential statistics were used.

A survey of 1507 smartphone users showed that 38% of them purchased insurance at the same time as they purchased the phone.

## Branches of Statistics

Determine whether descriptive or inferential statistics were used:

A study conducted by a research network found that people with fewer than 12 years of education had lower life expectancies than those with more years of education.

An inference is being drawn about the relationship between two variables. Inferential statistics were used.

A survey of 1507 smartphone users showed that 38% of them purchased insurance at the same time as they purchased the phone.

No prediction or generalization is being made. Descriptive statistics were used.

# Types of Variable

## Definition

*Qualitative variables* are variables that have distinct categories (not numeric)

## Definition

*Quantitative variables* are variables that can be counted or measured

# Types of Variable

## Definition

*Qualitative variables* are variables that have distinct categories (not numeric)

## Definition

*Quantitative variables* are variables that can be counted or measured

Examples of qualitative variables: gender, blood type, eye color



# Types of Variable

## Definition

*Qualitative variables* are variables that have distinct categories (not numeric)

## Definition

*Quantitative variables* are variables that can be counted or measured

Examples of qualitative variables: gender, blood type, eye color

Examples of quantitative variables: weight, heart rate, temperature

# Types of Variables

There are two types of quantitative variable:

# Types of Variables

There are two types of quantitative variable:

*Discrete variables* assume values that can be counted.

Examples: number of students in a class, number of text messages sent

# Types of Variables

There are two types of quantitative variable:

*Discrete variables* assume values that can be counted.

Examples: number of students in a class, number of text messages sent

*Continuous variables* can assume an infinite number of values between any two specific values; they are obtained by measuring.

Examples: time, temperature, length

# Types of Variables

Classify each variable as discrete or continuous:

**Votes received by a mayoral candidate in a city election**

**Systolic blood pressure readings**

**Temperatures at a seashore resort**

# Types of Variables

Classify each variable as discrete or continuous:

**Votes received by a mayoral candidate in a city election**

The number of votes can be counted; this is a discrete variable.

**Systolic blood pressure readings**

**Temperatures at a seashore resort**

# Types of Variables

Classify each variable as discrete or continuous:

## **Votes received by a mayoral candidate in a city election**

The number of votes can be counted; this is a discrete variable.

## **Systolic blood pressure readings**

While the accuracy of the measurement is limited by the accuracy of the device, systolic blood pressure can take on any positive real value. This is a continuous variable.

## **Temperatures at a seashore resort**

# Types of Variables

Classify each variable as discrete or continuous:

## **Votes received by a mayoral candidate in a city election**

The number of votes can be counted; this is a discrete variable.

## **Systolic blood pressure readings**

While the accuracy of the measurement is limited by the accuracy of the device, systolic blood pressure can take on any positive real value. This is a continuous variable.

## **Temperatures at a seashore resort**

Again, temperature can take on any real value. It is a continuous variable.



## Class Boundaries

Continuous data can take on infinitely many values between any two specific data points, but devices have limits in terms of accuracy.

Usually answers are rounded to the nearest given unit (e.g. heights to the nearest inch, weights to the nearest ounce)

## Class Boundaries

Continuous data can take on infinitely many values between any two specific data points, but devices have limits in terms of accuracy.

Usually answers are rounded to the nearest given unit (e.g. heights to the nearest inch, weights to the nearest ounce)

Example: If I report that my height is 72 inches, what is the range of possible values for my actual height?

## Class Boundaries

Continuous data can take on infinitely many values between any two specific data points, but devices have limits in terms of accuracy.

Usually answers are rounded to the nearest given unit (e.g. heights to the nearest inch, weights to the nearest ounce)

Example: If I report that my height is 72 inches, what is the range of possible values for my actual height?

Any height from 71.5 inches up to (but not including) 72.5 inches would round to 72 inches.

So the *boundary* of 72 inches is 71.5-72.5 inches

# Class Boundaries

The *boundary* of a number is the class that it would fall into before being rounded.

The boundaries of a continuous variable are given in one additional decimal place, and always end with a 5.

Find the boundary of each value:

24 ft

19.63 tons

200.6 joules

3.1415 in

# Class Boundaries

The *boundary* of a number is the class that it would fall into before being rounded.

The boundaries of a continuous variable are given in one additional decimal place, and always end with a 5.

Find the boundary of each value:

24 ft

23.5-24.5 ft

19.63 tons

200.6 joules

3.1415 in

# Class Boundaries

The *boundary* of a number is the class that it would fall into before being rounded.

The boundaries of a continuous variable are given in one additional decimal place, and always end with a 5.

Find the boundary of each value:

24 ft

23.5-24.5 ft

19.63 tons

19.625-19.635 tons

200.6 joules

3.1415 in

## Class Boundaries

The *boundary* of a number is the class that it would fall into before being rounded.

The boundaries of a continuous variable are given in one additional decimal place, and always end with a 5.

Find the boundary of each value:

24 ft

23.5-24.5 ft

19.63 tons

19.625-19.635 tons

200.6 joules

200.55-200.65 joules

3.1415 in

# Class Boundaries

The *boundary* of a number is the class that it would fall into before being rounded.

The boundaries of a continuous variable are given in one additional decimal place, and always end with a 5.

Find the boundary of each value:

24 ft	23.5-24.5 ft
19.63 tons	19.625-19.635 tons
200.6 joules	200.55-200.65 joules
3.1415 in	3.14145-3.14155 in



# Levels of Measurement

There are four *levels of measurement*, or *measurement scales* - two for qualitative data, and two for quantitative data.

# Levels of Measurement

There are four *levels of measurement*, or *measurement scales* - two for qualitative data, and two for quantitative data.

Qualitative data can be measured at the nominal level or at the ordinal level.

# Levels of Measurement

There are four *levels of measurement*, or *measurement scales* - two for qualitative data, and two for quantitative data.

Qualitative data can be measured at the nominal level or at the ordinal level.

Quantitative data can be measured at the interval level or at the ratio level.

# Levels of Measurement

## Nominal Level of Measurement

At the nominal level of measurement, data is classified into categories that cannot be ranked.

Given two data values, we can only say they are either the same or different from one another ( $=$ ,  $\neq$ ). We cannot say that one is better than another.

# Levels of Measurement

## Nominal Level of Measurement

At the nominal level of measurement, data is classified into categories that cannot be ranked.

Given two data values, we can only say they are either the same or different from one another ( $=$ ,  $\neq$ ). We cannot say that one is better than another.

Examples of nominal data:

- political party (Democrat, Independent, Libertarian, Republican, etc.)

# Levels of Measurement

## Nominal Level of Measurement

At the nominal level of measurement, data is classified into categories that cannot be ranked.

Given two data values, we can only say they are either the same or different from one another ( $=$ ,  $\neq$ ). We cannot say that one is better than another.

Examples of nominal data:

- political party (Democrat, Independent, Libertarian, Republican, etc.)
- eye color (blue, brown, green, red, etc.)

# Levels of Measurement

## Nominal Level of Measurement

At the nominal level of measurement, data is classified into categories that cannot be ranked.

Given two data values, we can only say they are either the same or different from one another ( $=$ ,  $\neq$ ). We cannot say that one is better than another.

Examples of nominal data:

- political party (Democrat, Independent, Libertarian, Republican, etc.)
- eye color (blue, brown, green, red, etc.)
- ethnicity (Asian, Black, Caucasian, Native American, etc.)

# Levels of Measurement

## Ordinal Level of Measurement

At the ordinal level of measurement, data is classified into categories that can be ranked.

Given two data values, we can tell whether they are the same or different from one another. In addition, we can compare the two using an inequality symbol ( $>$  or  $<$ )



# Levels of Measurement

## Ordinal Level of Measurement

At the ordinal level of measurement, data is classified into categories that can be ranked.

Given two data values, we can tell whether they are the same or different from one another. In addition, we can compare the two using an inequality symbol ( $>$  or  $<$ )

Examples of ordinal data:

- Letter grade earned in a course (A, B, C, D, F)

# Levels of Measurement

## Ordinal Level of Measurement

At the ordinal level of measurement, data is classified into categories that can be ranked.

Given two data values, we can tell whether they are the same or different from one another. In addition, we can compare the two using an inequality symbol ( $>$  or  $<$ )

Examples of ordinal data:

- Letter grade earned in a course (A, B, C, D, F)
- Size of a t-shirt (S, M, L, XL, XXL, etc.)

# Levels of Measurement

## Ordinal Level of Measurement

At the ordinal level of measurement, data is classified into categories that can be ranked.

Given two data values, we can tell whether they are the same or different from one another. In addition, we can compare the two using an inequality symbol ( $>$  or  $<$ )

Examples of ordinal data:

- Letter grade earned in a course (A, B, C, D, F)
- Size of a t-shirt (S, M, L, XL, XXL, etc.)
- Place earned in a race (1st, 2nd, 3rd, etc.)

# Levels of Measurement

## Interval Level of Measurement

At the interval level of measurement, there is a precise difference between any two data values.

We can determine whether two values are the same or different, we can compare using inequality symbols, and we can measure the distance between the two values. Interval data has no meaningful zero.

# Levels of Measurement

## Interval Level of Measurement

At the interval level of measurement, there is a precise difference between any two data values.

We can determine whether two values are the same or different, we can compare using inequality symbols, and we can measure the distance between the two values. Interval data has no meaningful zero.

Examples of interval data:

- Temperature ( $^{\circ}\text{C}$  or  $^{\circ}\text{F}$ )

# Levels of Measurement

## Interval Level of Measurement

At the interval level of measurement, there is a precise difference between any two data values.

We can determine whether two values are the same or different, we can compare using inequality symbols, and we can measure the distance between the two values. Interval data has no meaningful zero.

Examples of interval data:

- Temperature ( $^{\circ}\text{C}$  or  $^{\circ}\text{F}$ )
- IQ

# Levels of Measurement

## Interval Level of Measurement

At the interval level of measurement, there is a precise difference between any two data values.

We can determine whether two values are the same or different, we can compare using inequality symbols, and we can measure the distance between the two values. Interval data has no meaningful zero.

Examples of interval data:

- Temperature ( $^{\circ}\text{C}$  or  $^{\circ}\text{F}$ )
- IQ
- Date an apple was harvested

# Levels of Measurement

## Ratio Level of Measurement

At the ratio level of measurement, there is a precise difference between any two data values; furthermore, there is a true zero and a true ratio between any two data values.

We can tell whether two values are the same or different, we can compare using inequality symbols, we can measure the distance, and we can calculate the ratio.



# Levels of Measurement

## Ratio Level of Measurement

At the ratio level of measurement, there is a precise difference between any two data values; furthermore, there is a true zero and a true ratio between any two data values.

We can tell whether two values are the same or different, we can compare using inequality symbols, we can measure the distance, and we can calculate the ratio.

Examples of ratio level data:

- Revenue earned by companies

# Levels of Measurement

## Ratio Level of Measurement

At the ratio level of measurement, there is a precise difference between any two data values; furthermore, there is a true zero and a true ratio between any two data values.

We can tell whether two values are the same or different, we can compare using inequality symbols, we can measure the distance, and we can calculate the ratio.

Examples of ratio level data:

- Revenue earned by companies
- Age

# Levels of Measurement

## Ratio Level of Measurement

At the ratio level of measurement, there is a precise difference between any two data values; furthermore, there is a true zero and a true ratio between any two data values.

We can tell whether two values are the same or different, we can compare using inequality symbols, we can measure the distance, and we can calculate the ratio.

Examples of ratio level data:

- Revenue earned by companies
- Age
- Height

# Levels of Measurement

Here is a summary of the levels of measurement:

Measurement Level	Type	$= / \neq$	$> / <$	$+ / -$	$\times / \div$
Nominal	Qualitative	✓	✗	✗	✗
Ordinal	Qualitative	✓	✓	✗	✗
Interval	Quantitative	✓	✓	✓	✗
Ratio	Quantitative	✓	✓	✓	✓

# Levels of Measurement

Determine the level of measurement for each variable:

Amazon's quarterly profits

Colors of baseball hats sold in a store

Sizes of pizza (S, M, L)

Year of birth

# Levels of Measurement

Determine the level of measurement for each variable:

Amazon's quarterly profits

Ratio

Colors of baseball hats sold in a store

Sizes of pizza (S, M, L)

Year of birth

# Levels of Measurement

Determine the level of measurement for each variable:

Amazon's quarterly profits

Ratio

Colors of baseball hats sold in a store

Nominal

Sizes of pizza (S, M, L)

Year of birth

# Levels of Measurement

Determine the level of measurement for each variable:

Amazon's quarterly profits

Ratio

Colors of baseball hats sold in a store

Nominal

Sizes of pizza (S, M, L)

Ordinal

Year of birth



# Levels of Measurement

Determine the level of measurement for each variable:

Amazon's quarterly profits

Ratio

Colors of baseball hats sold in a store

Nominal

Sizes of pizza (S, M, L)

Ordinal

Year of birth

Interval

## Next Steps

- Read 1.3 and 1.4
- Watch Video Lesson #2
- Complete Assignment #1

Thanks for watching!