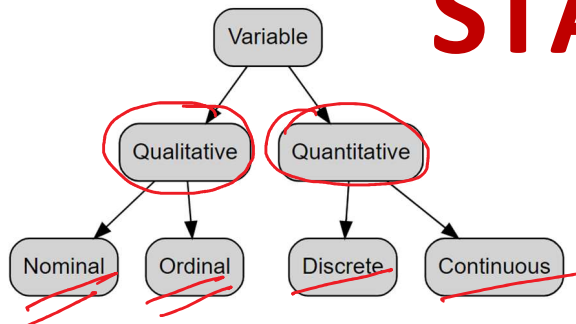- In statistics, a **variable** has two defining characteristics:
  - A variable is an attribute that describes a person, place, thing, or idea.
  - The value of the variable can "vary" from one entity to another.

START: 09:15

- A **quantitative** variable is a variable that reflects a notion of **magnitude**, that is, if the values it can take are **numbers**.

- A quantitative variable represents thus a measure and is numerical.

- Quantitative variables are divided into two types:
  - **discrete**
  - **continuous**

- **Quantitative discrete** variables are variables for which the values it can take are **countable** and have a **finite number of possibilities**.

- The values are often integers.
  - Number of children per family
  - Number of students in a class
  - Number of citizens of a country
  - Whatever the number of children in a family, it will never be 3.58 or 7.912 so the number of possibilities is a finite number and thus countable.

- **quantitative continuous** variables are variables for which the values are **not countable** and have an **infinite number of possibilities**.
- For example:
  - Age   Weight   Height
- we usually referred to years, kilograms and centimeters for age, weight and height respectively.
- However, a 28-year-old man could actually be 28 years, 7 months, 16 days, 3 hours, 4 minutes, 5 seconds, 31 milliseconds, 9 nanoseconds old.

- **Qualitative** variables are variables that are **not numerical** and which **values fits into categories**.
- In other words, a **qualitative** variable is a variable which takes as its values modalities, **categories** or even levels, in contrast to **quantitative** variables which measure a **quantity** on each individual.
- Qualitative variables are divided into two types:
  - **nominal**
  - **ordinal**.

- A **qualitative nominal** variable is a qualitative variable where **no ordering** is possible or implied in the levels.
- For example, the variable gender is nominal because there is no order in the levels female/male.
- Eye color is another example of a nominal variable because there is no order among blue, brown or green eyes.
- A nominal variable can have between two levels
  - (what is your gender? Female/Male)

- **Qualitative ordinal** variable is a qualitative variable with an **order implied in the levels**.

- good example is health, which can take values such as poor, reasonable, good, or excellent.

order
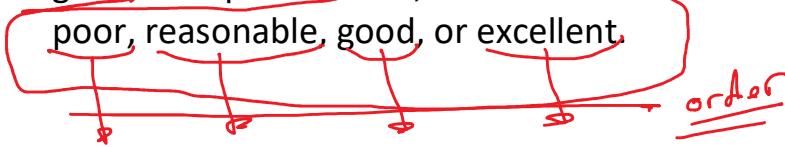
Example: Educational level might be categorized as
  1: Elementary school education
  2: High school graduate
  3: Some college
  4: College graduate
  5: Graduate degree

LEVELS OF MEASUREMENT

- When we talk about levels of measurement, we are talking about how we measure a variable.

- Variables have four different levels of measurement:

1. **Nominal**
2. **Ordinal**
3. **Interval**
4. **Ratio**

- **Nominal** variables are *categorical* variables where the categories are different only because they are named differently.
  - We cannot rank or order the categories.
  - Some examples include the following: race/ethnicity, gender, eye color, or neighborhood.

- **Ordinal** variables are *categorical* variables where the categories can be ordered or ranked.
  - Some examples include the following: class level (freshman, sophomore, junior, senior) and education level (less than HS, HS diploma, some college, college
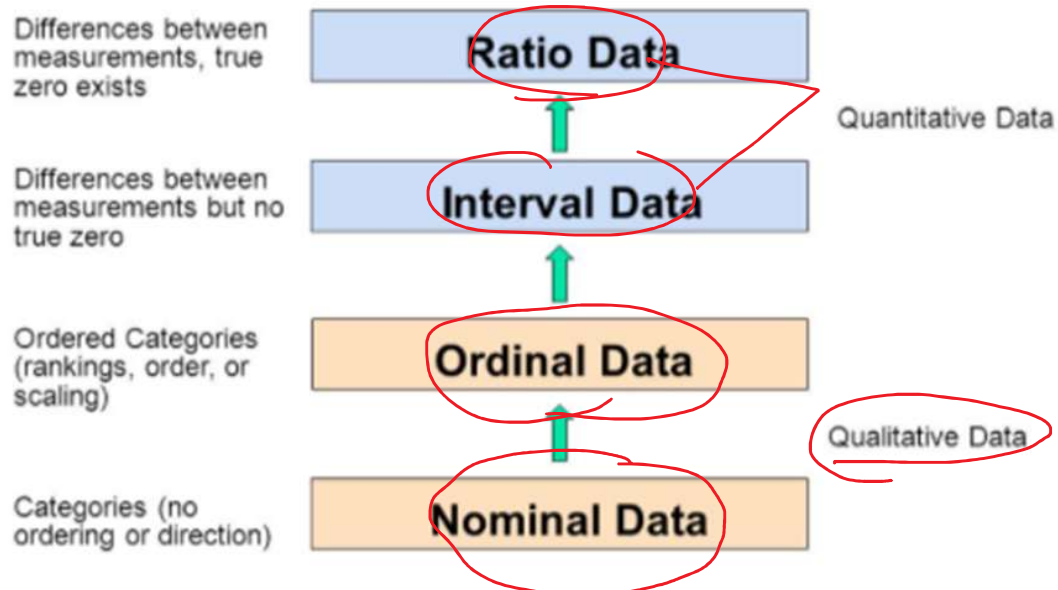
degree).

- **Interval** variables are *continuous/scale* variables with no meaningful/absolute zero.
  - A meaningful/absolute zero means that there is an absence of something.
  - In an interval variable, 0 is just another data point along the scale, it does NOT mean the absence of something.
  - For example, 0 degrees Fahrenheit is not the absence of heat or temperature, it is just another number along the temperature spectrum (it does mean it's pretty cold, though).

- **Ratio** variables are *continuous/scale* variables with a meaningful/absolute zero.
  - In a ratio variable, 0 means that there is nothing there.
  - For example, if I have 0 dollars, I have no money. If I have 0 hairs on my head, I am bald.

Types of Variables:

Variable types can be distinguished based on their scale. Typically, different statistical methods are appropriate for variables of different scales.

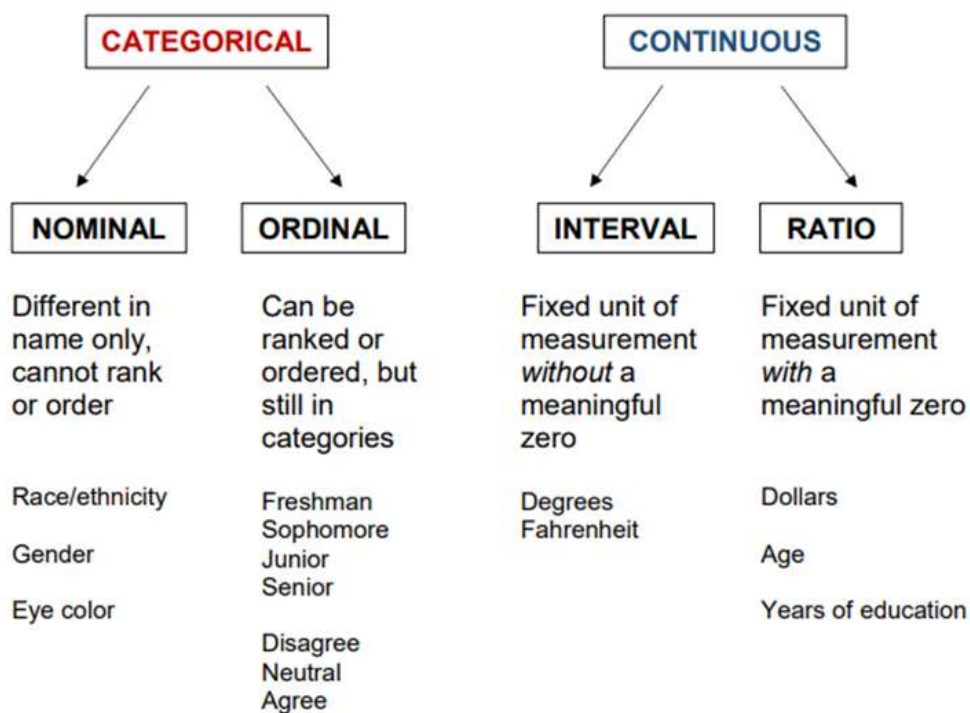| Scale	Characteristic Question	Examples
Nominal	Is A different than B?	Marital status Eye color Gender Religious affiliation Race
Ordinal	Is A bigger than B?	Stage of disease Severity of pain Level of satisfaction
Interval	By how many units do A and B differ?	Temperature SAT score
Ratio	How many times bigger than B is A?	Distance Length Time until death Weight

Differences between measurements, true zero exists — **Ratio Data**

Quantitative Data

Differences between measurements but no true zero — **Interval Data**

Ordered Categories (rankings, order, or scaling) — **Ordinal Data**

Qualitative Data

Categories (no ordering or direction) — **Nominal Data**

Data type	Mathematical operations	Measures of central tendency	Measures of variability
Nominal	• Equality (=, ≠)	• Mode	• None
Ordinal	• Equality (=, ≠) • Comparison (>, <)	• Mode • Median	• Range • Interquartile range
Interval	• Equality (=, ≠) • Comparison (>, <) • Addition, subtraction (+,−)	• Mode • Median • Arithmetic mean	• Range • Interquartile range • Standard deviation • Variance
Ratio	• Equality (=, ≠) • Comparison (>, <) • Addition, subtraction (+,−) • Multiplication, division (×, ÷)	• Mode • Median • Arithmetic mean • *Geometric mean	• Range • Interquartile range • Standard deviation • Variance • **Relative standard deviation

Central ten

1- Mean
2- mode
3- Median

LEVELS OF MEASUREMENT

CATEGORICAL

NOMINAL

Different in name only, cannot rank or order

Race/ethnicity

Gender

Eye color

ORDINAL

Can be ranked or ordered, but still in categories

Freshman
Sophomore
Junior
Senior

Disagree
Neutral
Agree

CONTINUOUS

INTERVAL

Fixed unit of measurement *without* a meaningful zero

Degrees Fahrenheit

RATIO

Fixed unit of measurement *with* a meaningful zero

Dollars

Age

Years of education

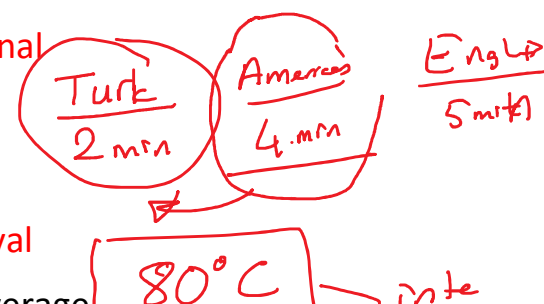## 2 Examples

1. Identify each of the following as qualitative or quantitative:

   (a) Gender Qualitative

   (b) High school GPA. Quantitative

   (c) The letter grade that you will receive for this course. Qualitative

   (d) Annual salary Quantitative

2. Identify each of the following as being at the nominal, ordinal, interval, or ratio level of measurement:

   (a) The years in which the LA Lakers won the NBA championship (for example, 2010) Interval

   (b) A restaurant's food and service rating (from 1–5 stars, with 1 being the lowest) Ordinal

   (c) A restaurant's food and service rating ("Horrible," "Poor", "Average," "Good," "Great") Ordinal

   (d) A collection of zip codes Nominal (a zip code is a number acting as a label)

   (e) Total touchdowns thrown by each quarterback in the NFL for the 2011 season Ratio

   (f) Annual salary Ratio

   (g) Today's temperature in degrees Fahrenheit Interval

   (h) Motion Picture Ratings (G, PG, PG-13, etc.) Ordinal

QUIZ I. Determine the level of measurement. (Nominal, Ordinal, Interval, Ratio)

- 1. Cars described as compact, midsize, and full-size.
- 2. Colors of M&M candies.
- 3. Weights of M&M candies
- 5. types of markers (washable, permanent, etc.)
- 6. time it takes to sing the National Anthem
- 7. total annual income for statistics students
- 8. body temperatures of bears in the north pole
- 9. teachers being rated as superior, above average, average, below average, or poor

- 1. Cars described as compact, midsize, and full-size. ordinal
- 2. Colors of M&M candies. nominal
- 3. Weights of M&M candies ratio
- 5. types of markers (washable, permanent, etc.) nominal
- 6. time it takes to sing the National Anthem ratio
- 7. total annual income for statistics students ratio
- 8. body temperatures of bears in the north pole interval
- 9. teachers being rated as superior, above average, average,

(handwritten annotations: National Anthem circled; Turk 2 min; America 4 mm; Engl Smith; 80°C; inte; 4s 2s 5s 10 toe)

- 8. body temperatures of bears in the north pole interval
- 9. teachers being rated as superior, above average, average, below average, or poor ordinal

80°C

40°C

inte
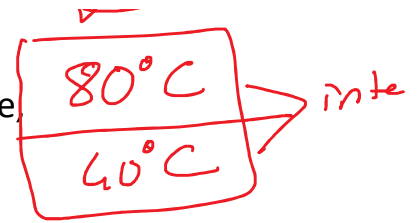
QUIZ II. Identify as Qualitative or Quantitative. If it is quantitative, tell if it is discrete or continuous

- 10. hair color of the math teachers at PHS
- 11. the number of people that prefer Pepsi over Coke
- 12. the weight of your sister's car (in pounds)
- 13. the number of criminal indictments against Michael Vick
- 14. the length of his jail sentence
- 16. how fast you were going when you were pulled over for speeding down Main Street (in MPH)

- 10. hair color of the math teachers at PHS qualitative
- 11. the number of people that prefer Pepsi over Coke quantitative discrete
- 12. the weight of your sister's car (in pounds) quantitative continuous
- 13. the number of criminal indictments against Michael Vick quantitative discrete
- 14. the length of his jail sentence quantitative continuous
- 16. how fast you were going when you were pulled over for speeding down Main Street (in MPH) quantitative continuous
-

Problem 1

Which of the following statements are true?

I. All variables can be classified as quantitative or categorical variables.
II. Categorical variables can be continuous variables.
III. Quantitative variables can be discrete variables.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) I and III

Solution

The correct answer is (E). All variables can be classified as quantitative or categorical variables. Discrete variables are indeed a category of quantitative variables. Categorical variables, however, are not numeric. Therefore, they cannot be classified as continuous variables.
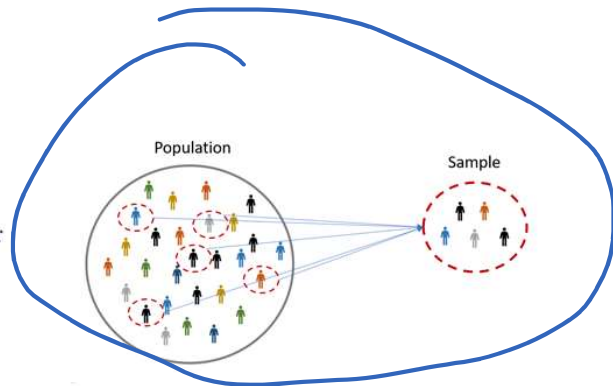
Population and Sample

- A **population** is the entire group that you want to draw conclusions about.

- A **sample** is the specific group that you will collect data from.

- The size of the sample is always less than the total size of the population.

- In research, a population doesn't always refer to people.

- It can mean a group containing elements of anything you want to study,

  - such as objects,

  - events,
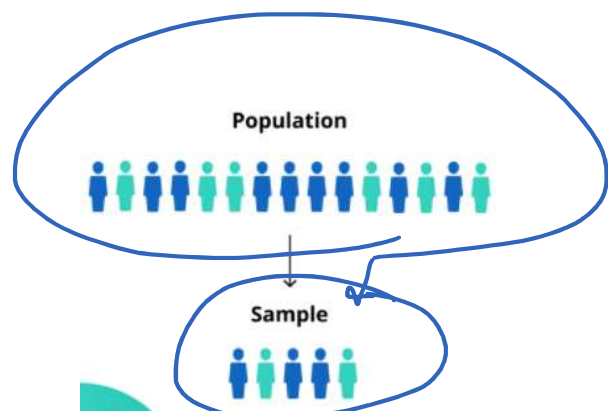
- organizations,
- countries,
- species,
- organisms, etc.

Use "population" when:
1. you know you have the entire population.
2. you have a sample of a larger population, but you are only interested in this sample (and you will not be generalizing your findings to the entire larger population).

Population   Sample

Use "sample" when:
1. you have a sample of a larger population, **and** you wish to generalize your findings from this sample to the entire larger population from which this sample was taken. The sample will be used as an estimate of the population.

Example: Student Housing

A survey is carried out at Penn State Altoona to estimate the proportion of all undergraduate students living at home during the current term.

total ⟶ *Survey*

Of the 3,838 undergraduate students enrolled at the campus, a random sample of 100 was surveyed.

- **Population**: All 3,838 undergraduate students at Penn State Altoona
- **Sample**: The 100 undergraduate students surveyed

We can use the data collected from the sample of 100 students to make inferences about the population of all 3,838 students.

Example: Polling Teachers
Educational policy researchers randomly selected 400 teachers at random from the National Science Teachers Association database of members and asked them

teachers at random from the National Science Teachers Association database of members and asked them whether or not they believed that evolution should be taught in public schools.
They received responses from 252 teachers. → Sample

- **Population**: All National Science Teachers Association members
- **Sample**: The 252 respondents

The researchers can use the data collected from the 252 teachers who responded to the survey to make inferences about the population of all National Science Teachers Association members.

Example: Flipping a Coin
A fair coin is flipped 500 times and the number of heads is recorded.

- **Population**: All flips of this coin
- **Sample**: The 500 flips recorded in this study

We can use data from these 500 flips to make inferences about the population of all flips of this coin.

Answer Please

A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School. What is the population?

A) the 100 selected students

B) all freshmen at Lincoln High School

C) all students at Lincoln High School

- all freshmen at Lincoln High School

A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School. What is the sample?

the 100 selected students	(A) sample
all freshmen at Lincoln High School	(B) → ?
all students at Lincoln High School	(C)

- the 100 selected students

Fifty bottles of water were randomly selected from a large collection of bottles in a company's warehouse. These fifty bottles are referred to as the

| parameter. |
| population. |
| **sample.** |

- sample

Fifty bottles of water were randomly selected from a large collection of bottles in a company's warehouse. The large collection of bottles is referred to as the

| parameter. |
| population. |
| sample. |

- population.

A mean is known as a statistic if it is computed from the **sample**

parameter.

population.

sample.

sample

START : 10 : 30