# Section 1.1

## *What is Statistics?*

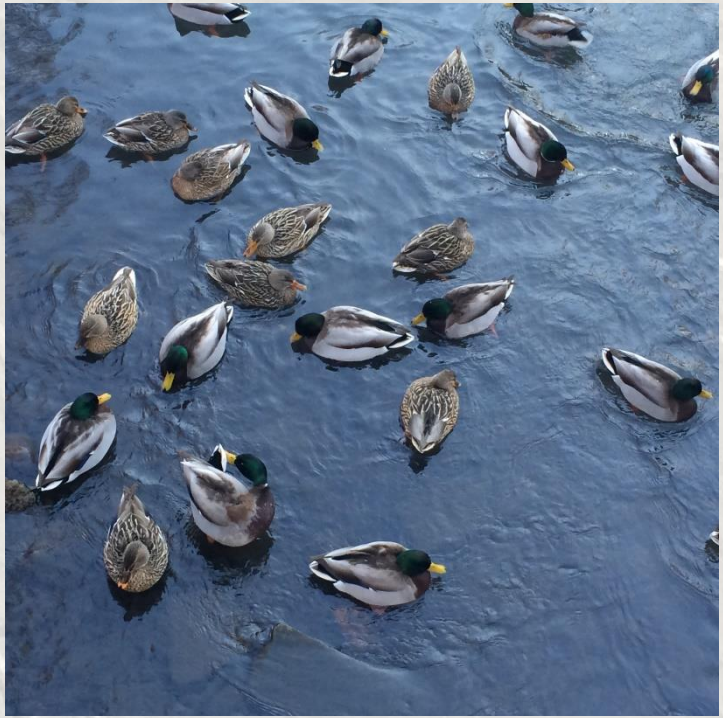
# Learning Objectives

At the end of this lecture, the student should be able to:

- State at least one definition of statistics.
- Give one example of a population parameter, and one example of a sample statistic.
- Be able to classify a variable into quantitative or qualitative, and as nominal, ordinal, interval or ratio.

# Outline

- Definition of Statistics
- Population Parameter & Sample Statistic
- Classifying Levels of Measurement



# Definition of Statistics

Statistics Helps Decision-Making

# Concepts in Statistics

- What is statistics?
- Individuals and variables
- Examples of statistics, individuals and variables in healthcare

# What is Statistics?

- Statistics is the study of how to collect, organize, analyze, and interpret numerical information and data.

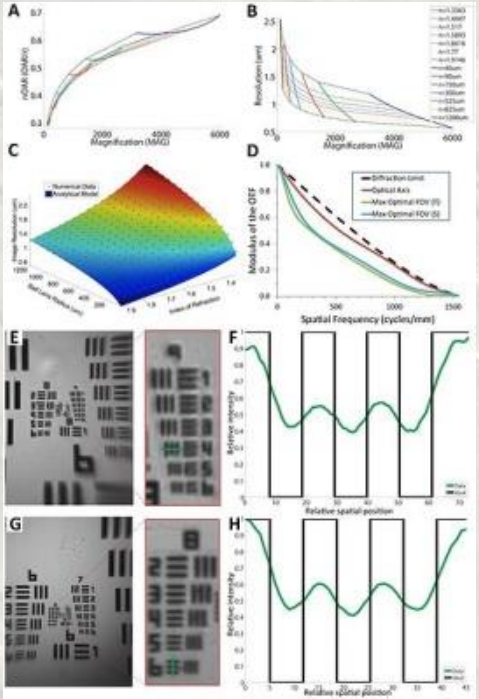



*Image from Cybulski J, Clements J, Prakash M (2014). "Foldscope: Origami-Based Paper Microscope". PLOS ONE. DOI: 10.1371/journal.pone.0098781*

# What is Statistics?

- Statistics is the study of how to collect, organize, analyze, and interpret numerical information and data.
- Statistics is both the science of uncertainty and the technology of extracting information from data.

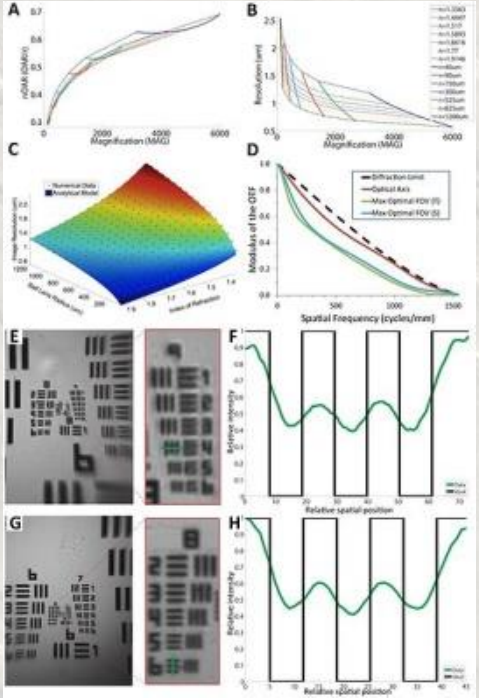



*Image from Cybulski J, Clements J, Prakash M (2014). "Foldscope: Origami-Based Paper Microscope". PLOS ONE. DOI: 10.1371/journal.pone.0098781*

# What is Statistics?

- Statistics is the study of how to collect, organize, analyze, and interpret numerical information and data.
- Statistics is both the science of uncertainty and the technology of extracting information from data.
- Statistics is used to help us make decisions. This is especially important in health care and public health.

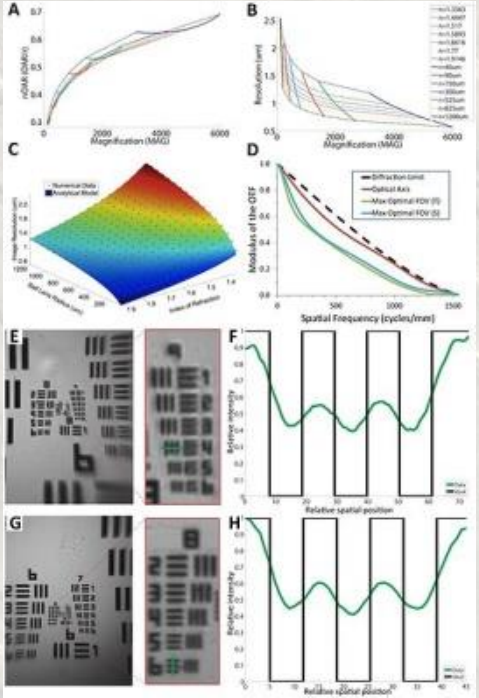



*Image from Cybulski J, Clements J, Prakash M (2014). "Foldscope: Origami-Based Paper Microscope". PLOS ONE. DOI: 10.1371/journal.pone.0098781*

# Example: CDC & the Flu Vaccine

- During the year, the United States Centers for Disease Control and Prevention (CDC) *collects, organizes, analyzes, and interprets numerical information and data*

*Photo by Marcela.*

# Example: CDC & the Flu Vaccine

- During the year, the United States Centers for Disease Control and Prevention (CDC) *collects, organizes, analyzes, and interprets numerical information and data*
- They *extract information from the data* and *make decisions* about what to include in next year's flu vaccine.

*Photo by Marcela.*

# Individuals & Variables

**Meaning Outside Statistics**

- Individuals are people.
  - We expect 50 individuals at the graduation.
- A variable is a factor that can vary, possibly causing a problem.
  - The time the shop takes with my car is an unknown variable, and I can’t predict it.

# Individuals & Variables

## Meaning Outside Statistics

- Individuals are people.
  - We expect 50 individuals at the graduation.
- A variable is a factor that can vary, possibly causing a problem.
  - The time the shop takes with my car is an unknown variable, and I can’t predict it.

## Meaning in Statistics

- Individuals are people or objects included in a study.
  - 5 individuals could be 5 people, 5 records, or 5 reports.
- A variable is a characteristic of the individual to be measured or observed.
  - The age of an individual person
  - The time an individual record was entered
  - The diagnosis listed on an individual report

# A Few Examples

| Individual | Variable |
|---|---|
| Kidney dialysis patient | Number of blood transfusions |
| Baby born to a mother who smokes cigarettes | Birthweight |
| Post-menopausal woman | Compliance with health screening recommendations |
| State with casinos | Rate of pathological gambling |
| County with hospitals | Medication error rate |
| Urban city | Rate of overdose deaths |

# Concepts in Statistics


- Statistics is used in healthcare and other disciplines to help aid in decision-making.
- Understanding statistics is necessary to understand certain processes in healthcare.

*Photo by Jorge Franganillo.*

# Population Parameter & Sample Statistic

Important Differences

# Parameters vs. Statistics

- What is a population and what is a sample?
- Difference between population and sample data
- Population parameters and sample statistics
- Examples of parameters and statistics

# What is a Population?

**Definition**

- A population is a group of people or objects with a common theme.
- When every member of that group is considered, it is a population.

# What is a Population?

**Definition**

- A population is a group of people or objects with a common theme.
- When every member of that group is considered, it is a population.

**Example**

- *Theme:* Nurses who work at Massachusetts General Hospital (MGH)
- *Population:* List from Human Resources of every currently employed nurse at MGH

# What is a Sample?

**Definition**

- A sample is a *small portion of the population.*
- It can be a representative sample.
- But it can also be a biased sample.

# What is a Sample?

**Definition**

- A sample is a *small portion of the population.*
- It can be a representative sample.
- But it can also be a biased sample.

**Example**

- Only survey ICU nurses at MGH
  - Not a representative sample
- At least one nurse from each department
  - More representative sample

# Population vs. Sample Data

**Population Data**

- In population data, data from every individual in the population is available.
- Entire population = census

# Population vs. Sample Data

**Population Data**

- In population data, data from every individual in the population is available.
- Entire population = census

**Sample Data**

- In sample data, data is only available from some of the individuals in the population
- Very commonly used in research studies of patients

# Examples of Population Data

- Medicare Claims Dataset – has all the insurance claims filed by Medicare population

# Examples of Population Data

- Medicare Claims Dataset – has all the insurance claims filed by Medicare population
- United States Census (conducted every 10 years)

# Examples of Sample Data

*Photo courtesy of KayEss*

- The Medicare Beneficiary Survey (MBS) is a survey of a sample of individuals on Medicare
- American Community Survey (ACS) conducted yearly by the United States Census Bureau

# Statistical Notation

**Total Population**

$N$

# Statistical Notation

**Total Population**

**N**

**Sample of Population**

**n**

# Parameter vs. Statistic

A parameter is a measure that describes the entire population

*Photo courtesy of Che/Wikimedia Commons*

# Parameter vs. Statistic

A parameter is a measure that describes the entire population

*Photo courtesy of Che/Wikimedia Commons*

A statistic is a measure that describes only a sample of a population

*Photo courtesy of Sandstein/Wikimedia Commons*

# Examples of Parameters & Statistics

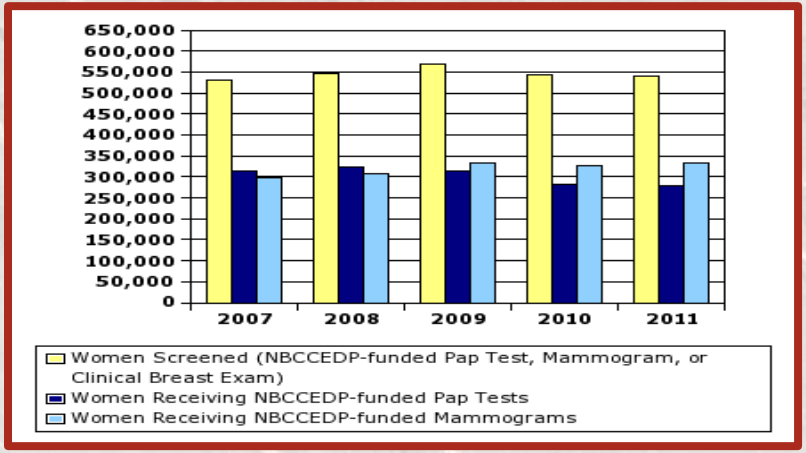
| Parameter | Statistics |
| --- | --- |
| Mean age of every American on Medicare | Mean age of Americans on Medicare estimated using the MBS |
| The proportion of Americans addicted to cigarettes | The proportion of Americans in the Behavioral Risk Factor Surveillance Survey (BRFSS) who admit they are addicted to cigarettes |
| Actual voter turnout | The proportion of people in opinion polls who say they plan to vote |

# Don’t Get Confused!

- When you hear a number on the television or radio – do they mention if it is a population parameter or a sample statistic?
- Clues that the number is a population parameter:
  - A dataset was used that encompasses the entire population (like Medicare)
  - Analyses were done by the government or on behalf of the government
- Clues that number is about a sample statistic:
  - It was from a study recruiting volunteers
  - The report mentions only surveying or measuring a sample of individuals

# Describing vs. Inferring

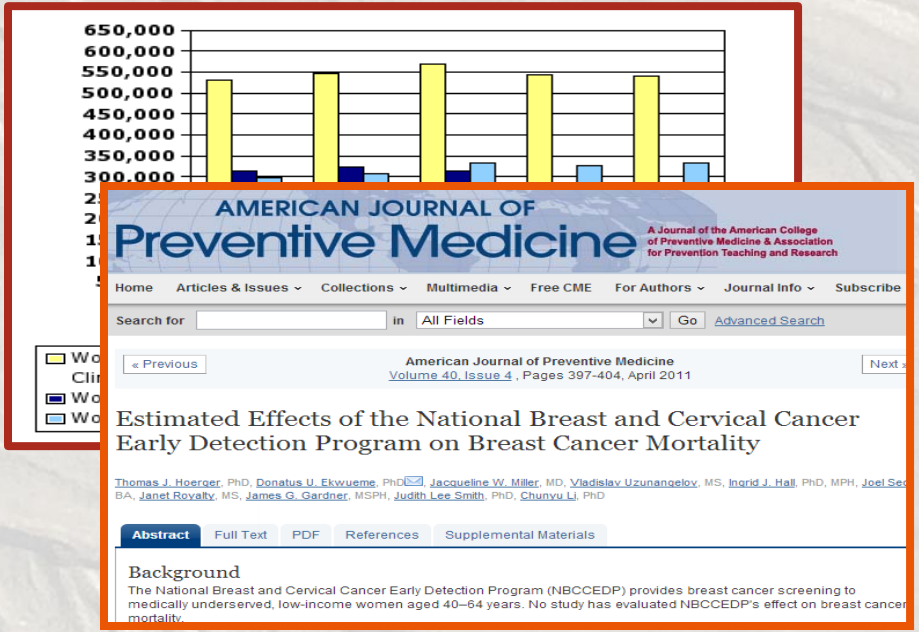
- Descriptive statistics involve methods of organizing , picturing, and summarizing information *from samples and populations.*




*Images courtesy of the Center for Disease Control and Prevention and the American Journal of Preventive Medicine.*

# Describing vs. Inferring

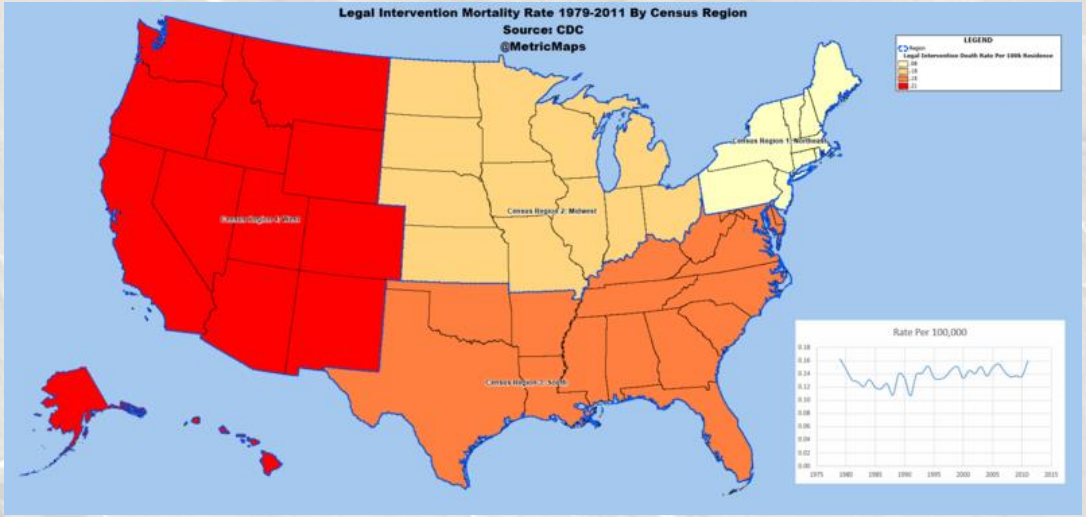
- Descriptive statistics involve methods of organizing , picturing, and summarizing information *from samples and populations.*
- Inferential statistics involves methods of using information *from a sample to draw conclusions regarding the population.*




*Images courtesy of the Center for Disease Control and Prevention and the American Journal of Preventive Medicine.*

# Parameters vs. Statistics

- In statistics, it is important to properly identify measures as population parameters or sample statistics.
- Different types of data are used for parameters and statistics.




*Map by Metric Maps.*

# Classifying Levels of Measurement

Four-level system

# Classifying Variables

- Quantitative vs. qualitative data
  - Interval vs. ratio data
  - Nominal vs. ordinal data
- Examples of how to classify healthcare data

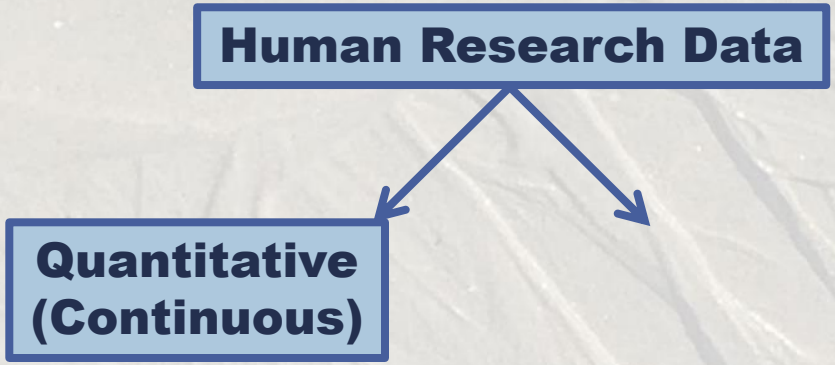

*Photo courtesy of Toho-u*

# Four-level Data Classification

**Human Research Data**

# Four-level Data Classification

# Four-level Data Classification



# Quantitative vs. Qualitative

Quantitative is a numerical measurement of something

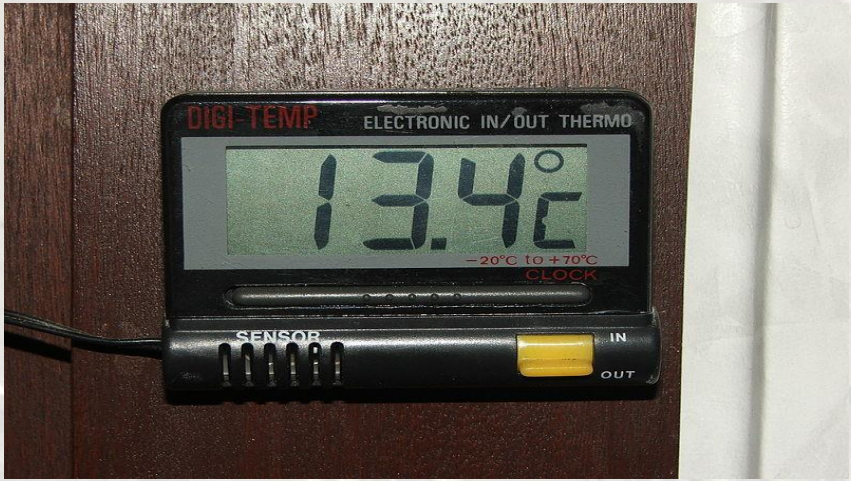


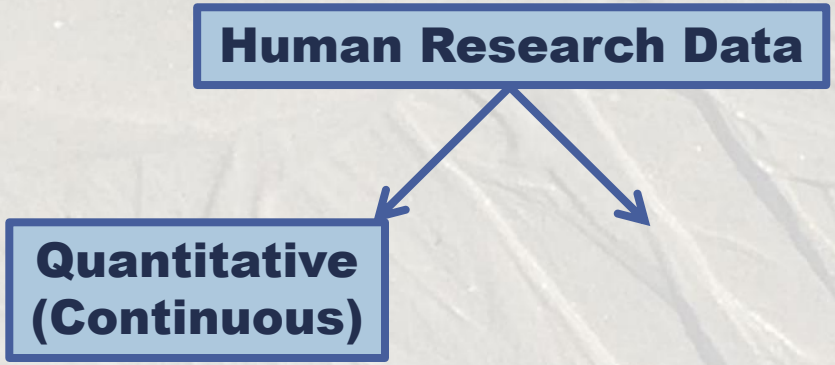

*Photo courtesy of Xell/Wikimedia Commons*

# Examples of Quantitative Variables in Healthcare

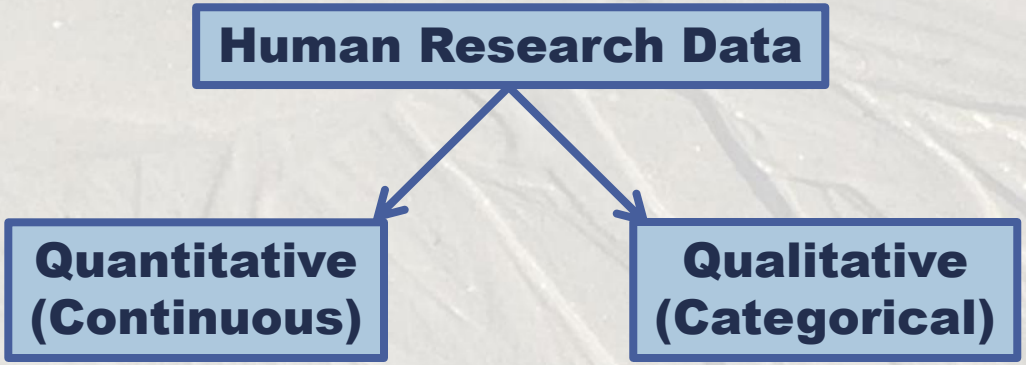
- Time of admit
- Year of diagnosis
- Systolic blood pressure (SBP)
- Platelet count


*Photo courtesy of US Navy*

# Four-level Data Classification



# Four-level Data Classification



# Quantitative vs. Qualitative

Quantitative is a numerical measurement of something

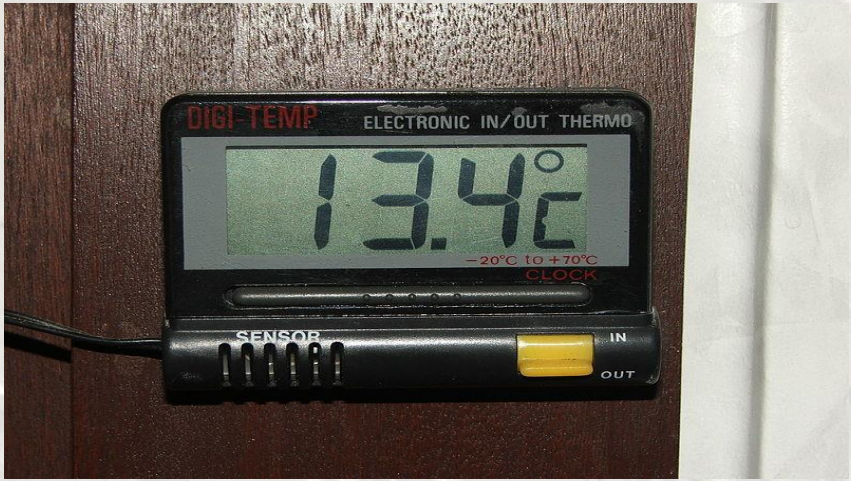



*Photo courtesy of Xell/Wikimedia Commons*

# Quantitative vs. Qualitative

Quantitative is a numerical measurement of something

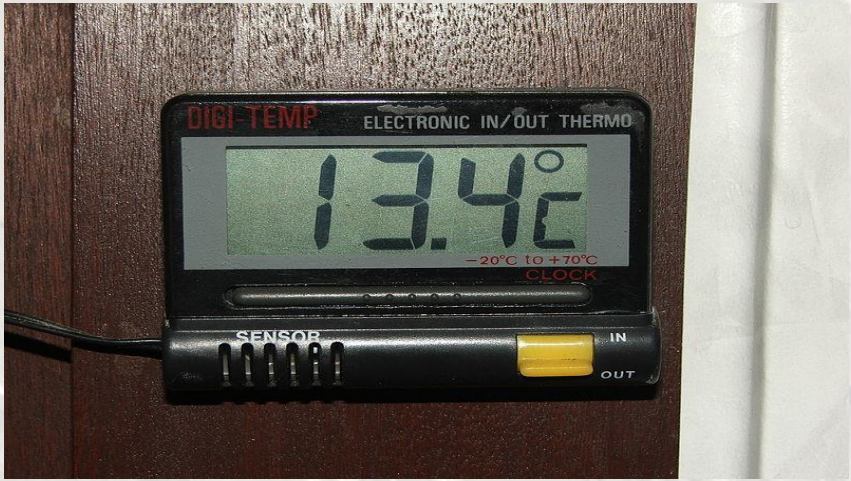



*Photo courtesy of Xell/Wikimedia Commons*

Qualitative refers to a "quality" or categorical characteristic of something

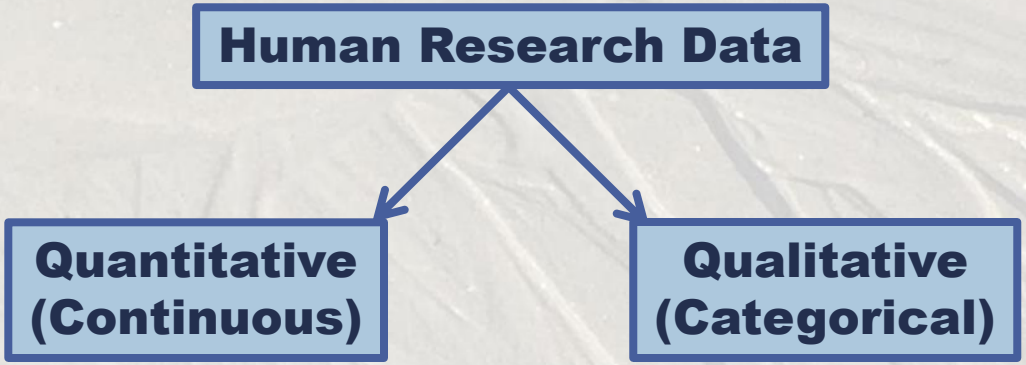
*Photo courtesy of Ceedz/Wikimedia Commons*

# Examples of Qualitative Variables in Healthcare

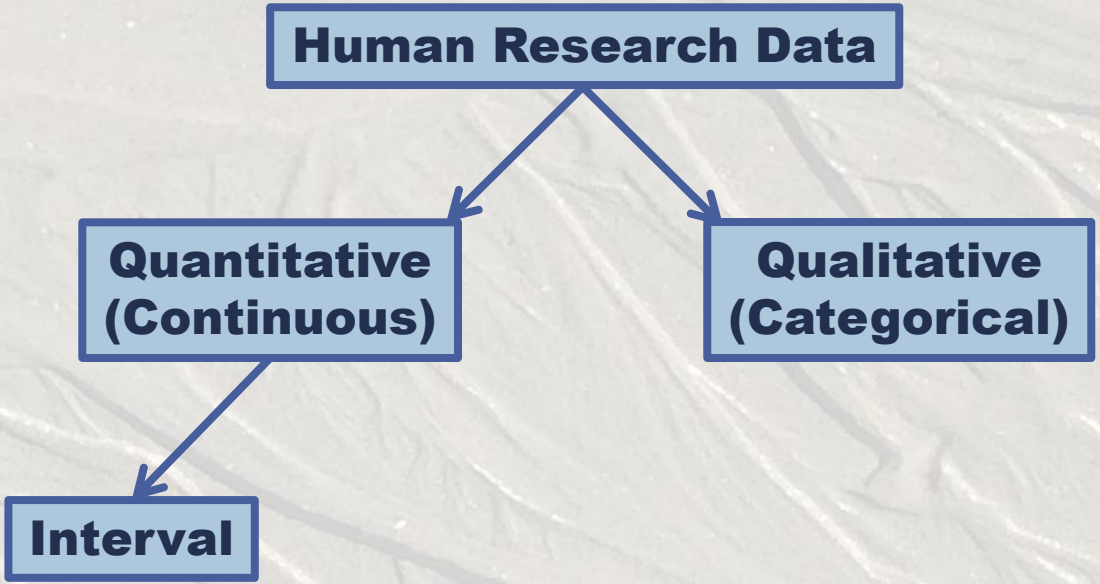

- Type of health insurance
- Country of origin
- Stage of  cancer
- Trauma center level

*Photo courtesy of US Navy*

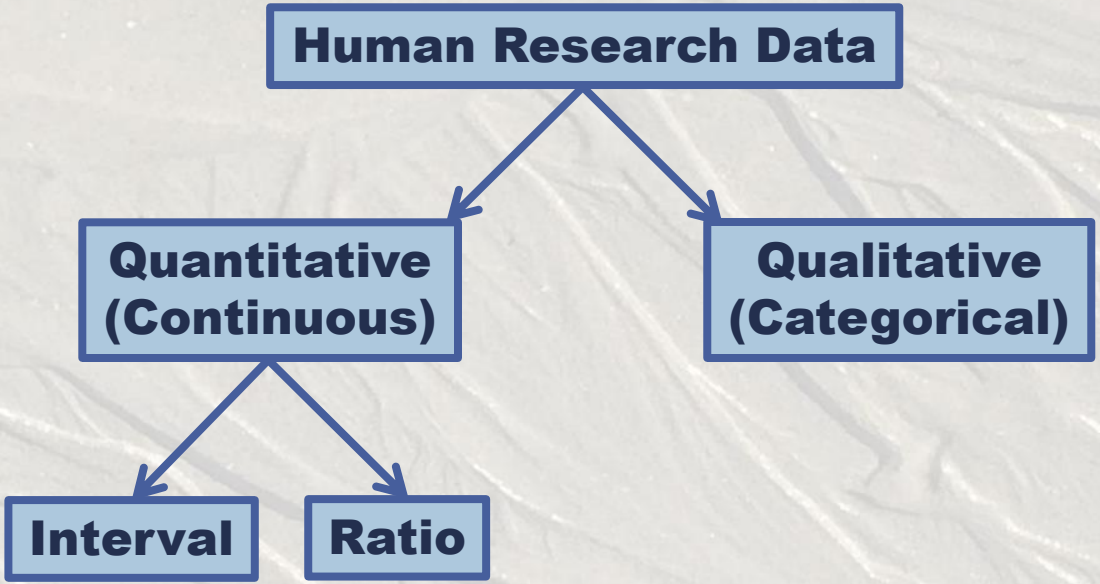
# Four-level Data Classification



# Four-level Data Classification



# Four-level Data Classification



# Interval vs. Ratio

**Interval**

- Quantitative (continuous) data
  - Differences between data values are meaningful

**Ratio**

- Quantitative (continuous) data
- Differences between data values are meaningful

# Interval vs. Ratio

**Interval**

- Quantitative (continuous) data
  - Differences between data values are meaningful
- There *is no* true zero

**Ratio**

- Quantitative (continuous) data
- Differences between data values are meaningful
- There *is a* true zero

# Examples of Quantitative Variables in Healthcare

- Time of admit
- Year of diagnosis
- Systolic blood pressure (SBP)
- Platelet count


*Photo courtesy of US Navy*

# Examples of Quantitative Variables in Healthcare

- Time of admit
- Year of diagnosis
- Systolic blood pressure (SBP)
- Platelet count

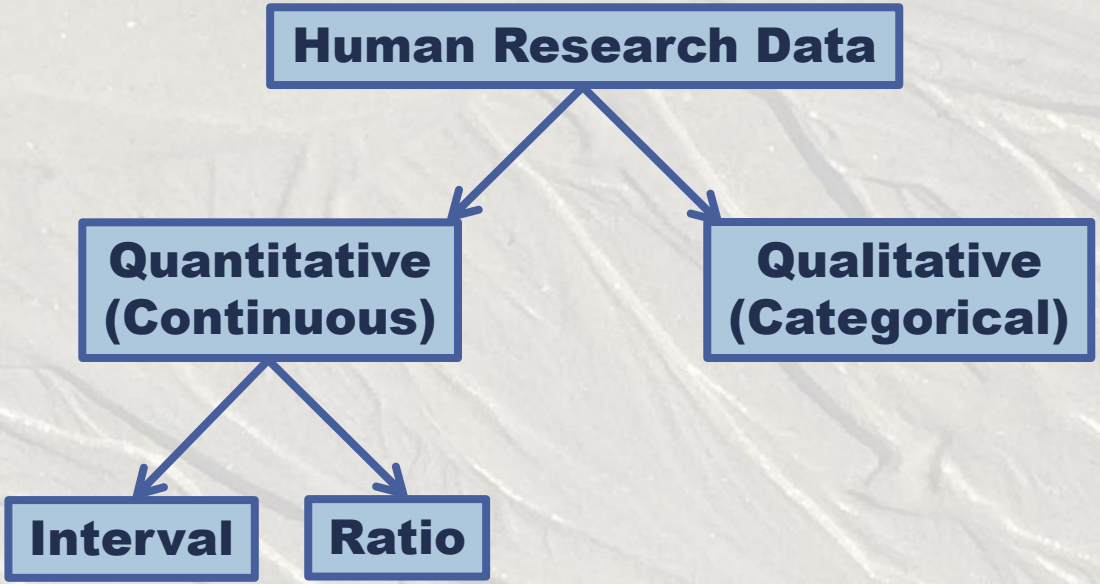
- Ratio = there is a true zero
  - If you are dead, you have a SBP = 0
  - Same is true with platelet count

# Examples of Quantitative Variables in Healthcare

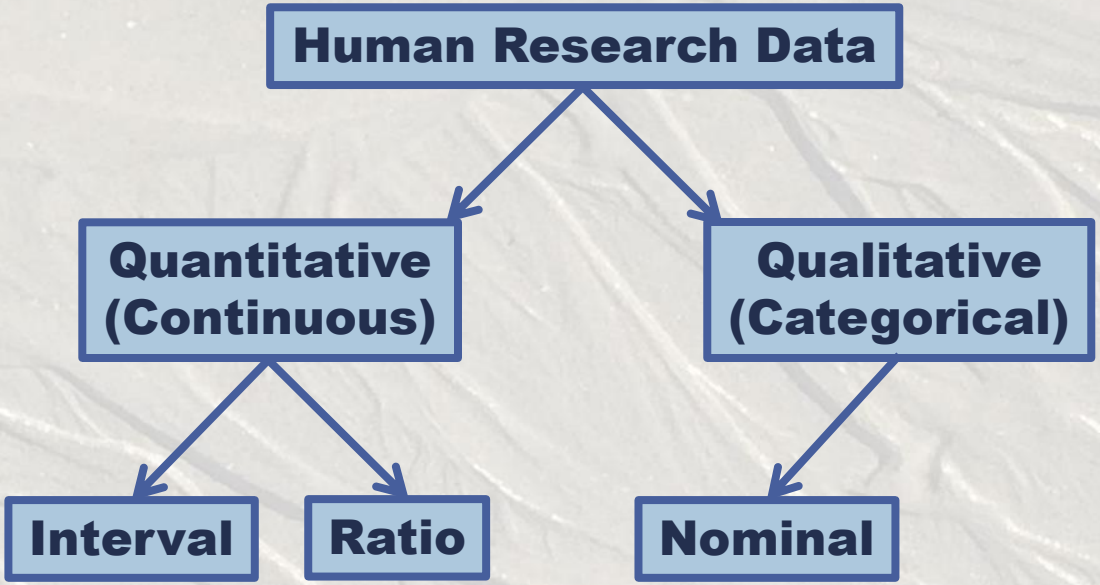
- Time of admit
- Year of diagnosis
- Systolic blood pressure (SBP)
- Platelet count

- Interval = no true zero
- "Time" cannot have a true zero
  - Time of admit = 8:09 am
  - Year of diagnosis = 1999

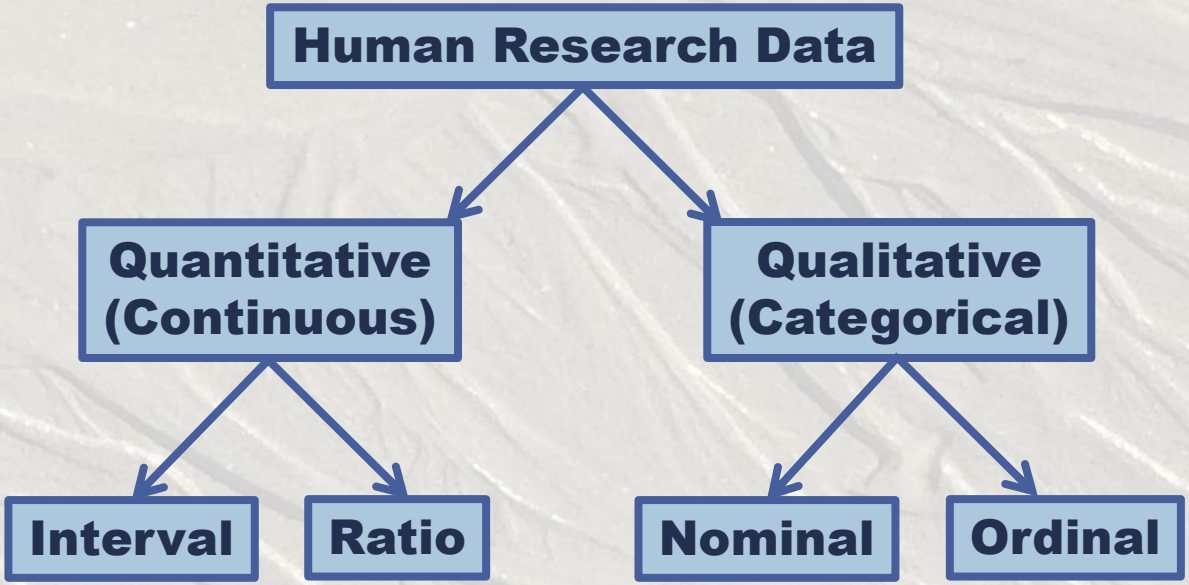
# Four-level Data Classification



# Four-level Data Classification



# Four-level Data Classification



# Nominal vs. Ordinal

**Nominal**

- Nominal applies to categories, labels or names, and *cannot* be ordered from smallest to largest

# Nominal vs. Ordinal

**Nominal**

- Nominal applies to categories, labels or names, and *cannot* be ordered from smallest to largest

**Ordinal**

- Ordinal applies to data that can be arranged in order in categories,
  - but the difference between data values cannot be determined, or is meaningless.

# Examples of Qualitative Variables in Healthcare


- Type of health insurance
- Country of origin
- Stage of  cancer
- Trauma center level

*Photo courtesy of US Navy*

# Examples of Qualitative Variables in Healthcare

- Nominal = Cannot be ordered

- Type of health insurance
- Country of origin
- Stage of  cancer
- Trauma center level

# Examples of Qualitative Variables in Healthcare

- Nominal = Cannot be ordered
- Ordinal = natural order
  - Differences between levels is meaningless

- Type of health insurance
- Country of origin
- Stage of cancer
- Trauma center level

# Classifying Variables

- All data can be classified as quantitative or qualitative.
- Data can be further classified as interval, ratio, nominal or ordinal.
- It is important to know how to classify data in healthcare.

*Photograph by Producer*

# Conclusion

- Definition of Statistics
- Population Parameter & Sample Statistic
- Classifying Levels of Measurement