♠

Introduction: What is Statistics

quick notes @ranton95

Contents:

Key Definition:

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

- Statistics is both the science of uncertainty and the technology of extracting information from data.

- Statistics are used to help us make decisions.

What are Individuals and Variables?

Individuals: People or objects included in a study.

- For example, 5 individuals could be 5 people, 5 records, or 5 reports

Variable: is the characteristic of the individual to be measured or observed.

- For example, the age of an individual person

- The time an individual record was entered

- The diagnosis listed on an individual report

https://www.youtube.com/watch?v=EqeVXI4WNHM

What is a population and what is a sample?

The first step in every statistical analysis is to determine if the data we are provided is a population or a sample.

Population (**N**): Collection of all items of interest to our study. The numbers we obtain when using a population are called *parameters*.

Sample (**n**): A subset of the population. The numbers we obtain when working with a sample are called *statistics*.

A sample has two key characteristics:

A sample must be **random** and **representative** for an insight to be precise.

A **random sample** is collected when each member is chosen strictly by chance. We must ensure each member is equally likely to be chosen.

A **representative sample** is a sample that accurately reflects the members of the entire population.

Summary: Populations are harder to gather. They are hard to observe and contact. Sample on the other hand is less time-consuming and cheaper. It is easy to observe and contact.

What is the difference between 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.

What are the different types of data?
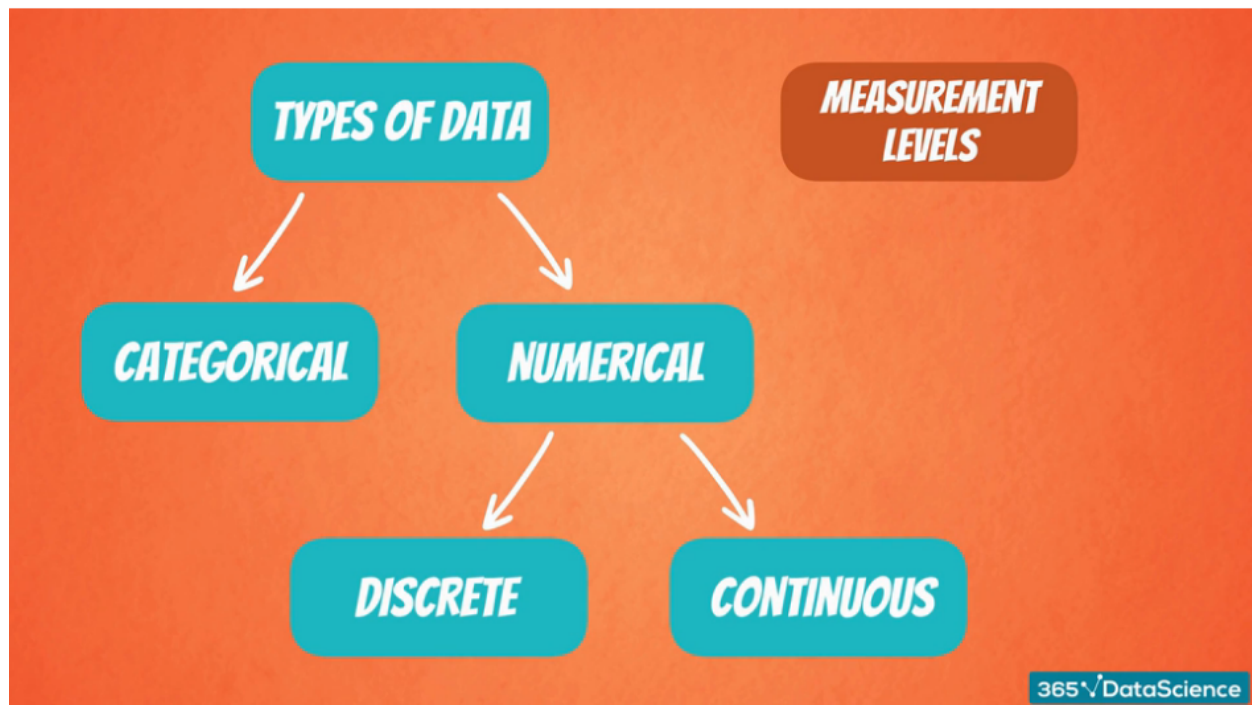
Categorical: categories of groups

Example: Car brands, Yes and No questions

Numerical: describing numbers; further divided into <u>discrete</u> and <u>continuous</u> data.

Continuous data is infinite and impossible to count. For example, our body weight changes although the scale can only show the approximations and not the discrete values. Height, area, distance, time are other examples of continuous.

The number of children, for example, is a great example of discrete values. Other examples include grades, number of objects, physical money (cannot pay $1.2860254802 in dollars)

What are the four levels of classifying data for measurement?
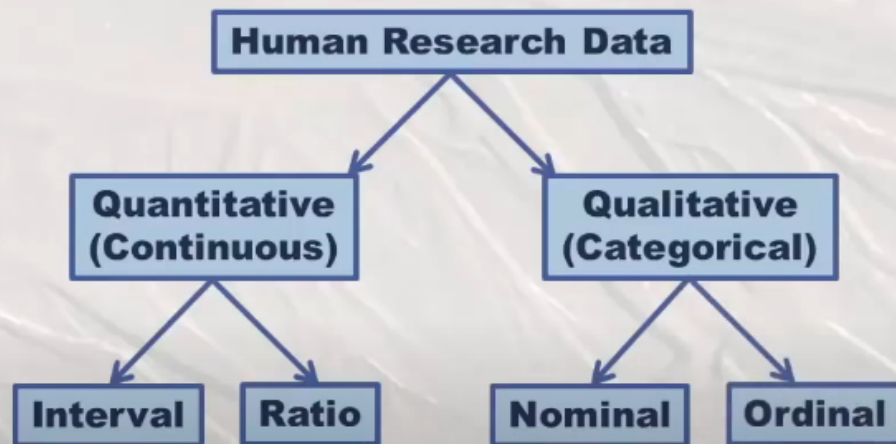
Quantitative (continuous) is a numerical measurement of something.

Example: Time of audit, Year of diagnosis, Platelet count

Interval vs. Ratio data

Interval: Differences between data values are meaningful, however, there is **no** true zero

Ratio: The difference between data values is meaningful, however, there is a true zero.

Qualitative (categorical) refers to the quality or categorical characteristics of something.

 Example: Type of health insurance, country of origin, stages of something

Nominal vs Ordinal data

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

Ordinal: applies to data, that **can be** arranged in the order in categories. However, here the difference between data values cannot be determined or is meaningless.