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.

Terminologies:

1. **Individuals** : People or Objects in a study
2. **Variable**: A characteristic of the individual to be measured

Parameters vs Statistics

What is a Population?

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 sample?

A sample is a small portion of the population. It could be a representative sample or a biased one.

Population vs. Sample Data

- (N) Entire population = census , data available for every individual
- (n) Sample Data , data available for some individuals

Parameter	Statistics
A parameter is a measure that describes the entire population.	A statistic is a measure that describe only a sample of a population.

Types of statistics

1. **Descriptive statistics** involve methods of organizing, picturing and summarizing data from sample and population.
2. **Inferential statistics** involves methods of using information from samples to draw conclusions regarding population.

Types of data (variable measurement scales)

Qualitative data refers to information about qualities, or information that cannot be measured. It's usually descriptive and textual. Examples include someone's eye color or the type of car they drive. In surveys, it's often used to categorize 'yes' or 'no' answers.
Qualitative data is numerical. It's used to define information that can be counted. Some examples of quantitative data include distance, speed, height, length and weight. It's easy to remember the difference between qualitative and quantitative data, as one refers to qualities, and the other refers to quantities.

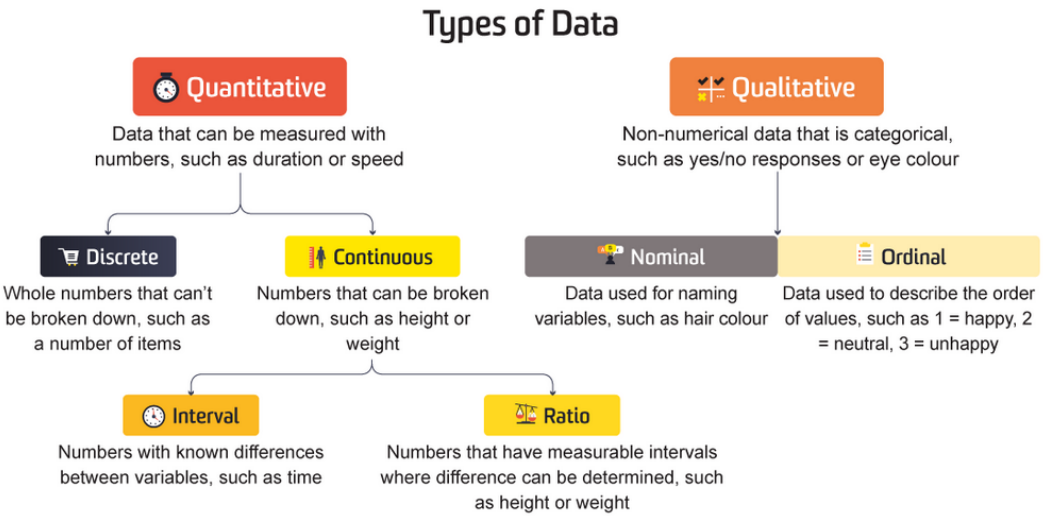

image @ https://studyonline.unsw.edu.au/sites/default/files/field/image/types of data.png

1. **Nominal Data**

Nominal data are used to label variables where there is no quantitative value and has no order. So, if you change the order of the value then the meaning will remain the same.
Thus, nominal data are observed but not measured, are unordered but non-equidistant, and have no meaningful zero.
The only numerical activities you can perform on nominal data is to state that perception is (or isn't) equivalent to another (equity or inequity), and you can use this data to amass them. You can't organize nominal data, so you can't sort them.
Neither would you be able to do any numerical tasks as they are saved for numerical data. With nominal data, you can calculate frequencies, proportions, percentages, and central points.

example : Native languages: English, Marathi, French, Hindi, Punjabi

2. **Ordinal Data**

Ordinal data is almost the same as nominal data but not in the case of order as their categories can be ordered like 1st, 2nd, etc. However, there is no continuity in the relative distances between adjacent categories.
Ordinal Data is observed but not measured, is ordered but non-equidistant, and has no

meaningful zero. Ordinal scales are always used for measuring happiness, satisfaction, etc. With ordinal data, likewise, with nominal data, you can amass the information by evaluating whether they are equivalent or extraordinary.
As ordinal data are ordered, they can be arranged by making basic comparisons between the categories, for example, greater or less than, higher or lower, and so on.
You can't do any numerical activities with ordinal data.
With ordinal data, you can calculate the same things as nominal data like frequencies, proportions,
percentage, central point but there is one more point added in ordinal data that is summary statistics and similarly Bayesian statistics.

example: Opinions: Mostly agree, Agree, Neutral , Disagree, etc

3. **Interval Data**

Interval Data are measured and ordered with the nearest items but have no meaningful zero.
The central point of an Interval scale is that the word 'Interval' signifies 'space in between', which is
the significant thing to recall, interval scales not only educate us about the order but additionally about the value between every item.
Interval data can be negative, though ratio data can't. Even though interval data can show
up fundamentally the same as ratio data, the thing that matters is in their characterized zero-points. If the zero-point of the scale has been picked subjectively, at that point the data can't be ratio data and should be interval data.
Hence, with interval data you can easily correlate the degrees of the data and also you can add or subtract the values.

example: Temperature (°C or F, but not Kelvin)

4. **Ratio Data**

Ratio Data are measured and ordered with equidistant items and a meaningful zero and never be negative like interval data.
Ratio data enlightens us regarding the order for variables, the contrasts among them, and they have absolute zero. It permits a wide range of estimations to be performed.
Ratio data is fundamentally the same as interval data, aside from zero means none.

example: Age (from 0 years to 100+)