

Types of data in statistics

There are only 2 classes of data in statistics: quantitative data and qualitative data. This highest level of classification comes from the fact that data can either be measured or can be an observed feature of interest.

Qualitative data are also referred to as categorical data. They are an observed phenomenon and cannot be measured with numbers. Examples: a race, age group, gender, origin, and so on. Even if they contain a numerical value, they hold no meaning (1 for male and 0 for female).

Quantitative data, on the other hand, tells us about the quantities of things or the things we can measure. And, so they are expressed in terms of numbers. It is also known as numerical data and includes statistical data analysis. Examples: height, weight, distance, and so on.

We can further subdivide quantitative data and qualitative data into 4 subtypes as follows: nominal data, ordinal data, interval data, and ratio data.

Qualitative (Categorical) data type:

Qualitative data can be subdivided into nominal and ordinal data types. While both these types of data can be classified, ordinal data can be ordered as well.

Nominal Data: Nominal data is a type of data that represents discrete units which is why it cannot be ordered and measured. They are used to label variables without providing any quantitative value. Also, they have no meaningful zero.

Some examples of nominal data include:

- Gender (Male, Female)
- Hair color (Black, Brown, Gray, etc)
- Nationality (Indian, American, Chinese, etc)

Data scientists use hot encoding, to transform nominal data into a numeric feature.

The only logical operation that you can apply to them is equality or inequality which you can also use to group them. The descriptive statistics you can do with nominal data include frequencies, proportions, percentages, and central points. And, to visualize nominal data, you can use a pie chart or a bar chart.

Ordinal Data: Ordinal values represent discrete as well as ordered units. Unlike nominal, here the ordering matters. However, there is no consistency in the relative distance between the adjacent categories. And, similar to nominal data, ordinal data also don't have a meaningful zero.

Examples of ordinal data:

- Opinion (agree, mostly agree, neutral, mostly disagree, disagree)
- Socioeconomic status (low income, middle income, high income)
- Data scientists use label encoding to transform ordinal data into a numeric feature.

The descriptive statistics that you can do with ordinal data include frequencies, proportions, percentages, central points, percentiles, median, mode, and the interquartile range. Here the visualization methods that can be used are the same as nominal data.

Quantitative (Numerical) Data Types:

Two types of quantitative data are discrete data and continuous data. Discrete data have distinct and separate values. Therefore, they are data with fixed points and can't take any measures in between. So all counted data are discrete data. Some examples of discrete data include shoe sizes, number of students in class, number of languages an individual speaks, etc. Continuous data, on the other hand, represent an endless range of possible values within a specified range. It can be divided into finer parts to be measured but not counted. Continuous data examples include temperature range, height, weight, etc.

Continuous data can be visualized by histogram or box plot while bar graphs or stem plots can be used for discrete data.

Here are two types of continuous data:

Interval Data: It represents ordered data that is measured along a numerical scale with equal distances between the adjacent units. These equal distances are also referred to as intervals. So a variable contains interval data if it has ordered numeric values with the exact differences known between them.

Interval data can be continuous or discrete.

Examples of Interval data:

- IQ test's intelligence scale
- Time if measured using a 12-hour clock

You can compare the data with interval data and add/subtract the values but cannot multiply or divide as it doesn't have a meaningful zero. The descriptive statistics you can apply for interval data include central point, range, and spread.

Ratio Data: Like Interval data, ratio data are also ordered with the same difference between the individual units. However, they also have a meaningful zero so they cannot take negative values.

Examples of ratio data:

- Temperature on a Kelvin scale (0 degrees represent total absence of thermal energy)
- Height (zero is the starting point)

Now with real zero points, we can also multiply and divide the numbers. Besides, you can sort the values as well. The descriptive statistics you can do with ratio data are the same as interval data and include central point, range, and spread.

Overall, ratio data and interval data are the same with equal spacing between adjoining values but the former also has a meaningful zero. Besides addition and subtraction, you can also multiply and divide the data, which is impossible with interval data as it does not have an absolute zero. However, interval data can take negative values with no absolute zero while ratio data cannot.