**1)What is Statistics?**

Ans- Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data.

**2)Define the different types of statistics and give an example of when each type might be used.**

Ans-The two types of statistics are: Descriptive and inferential.

Descriptive Statistics

In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or [standard deviation](https://byjus.com/maths/standard-deviation/).

Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures. For example, the collection of people in a city using the internet or using Television.

Descriptive statistics are also categorised into four different categories:

* Measure of frequency
* Measure of dispersion
* Measure of central tendency
* Measure of position

### Inferential Statistics

This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.

Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population. It grants us permission to give statements that goes beyond the available data or information. For example, deriving estimates from hypothetical research.

**3) What are the different types of data and how do they differ from each other? Provide an example of**

**each type of data.**

#### Ans- **The data is classified into four categories:**

* Nominal data.
* Ordinal data.
* Discrete data.
* Continuous data.

Nominal Data

Nominal Data is used to label variables without any order or quantitative value. The color of hair can be considered nominal data, as one color can’t be compared with another color.

The name “nominal” comes from the Latin name “nomen,” which means “name.” With the help of nominal data, we can’t do any numerical tasks or can’t give any order to sort the data. These data don’t have any meaningful order; their values are distributed into distinct categories.

Examples of Nominal Data :

* Colour of hair (Blonde, red, Brown, Black, etc.)
* Marital status (Single, Widowed, Married)
* Nationality (Indian, German, American)
* Gender (Male, Female, Others)
* Eye Color (Black, Brown, etc.)

### **Ordinal Data**

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness, etc., but we can’t do any arithmetical tasks on them.

Ordinal data is qualitative data for which their values have some kind of relative position. These kinds of data can be considered “in-between” qualitative and quantitative data. The ordinal data only shows the sequences and cannot use for statistical analysis. Compared to nominal data, ordinal data have some kind of order that is not present in nominal data.

#### **Examples of Ordinal Data :**

* When companies ask for feedback, experience, or satisfaction on a scale of 1 to 10
* Letter grades in the exam (A, B, C, D, etc.)
* Ranking of people in a competition (First, Second, Third, etc.)
* Economic Status (High, Medium, and Low)
* Education Level (Higher, Secondary, Primary)

### **Discrete Data**

The term discrete means distinct or separate. The discrete data contain the values that fall under integers or whole numbers. The total number of students in a class is an example of discrete data. These data can’t be broken into decimal or fraction values.

The discrete data are countable and have finite values; their subdivision is not possible. These data are represented mainly by a bar graph, number line, or frequency table.

#### **Examples of Discrete Data :**

* Total numbers of students present in a class
* Cost of a cell phone
* Numbers of employees in a company
* The total number of players who participated in a competition
* Days in a week

### **Continuous Data**

Continuous data are in the form of fractional numbers. It can be the version of an android phone, the height of a person, the length of an object, etc. Continuous data represents information that can be divided into smaller levels. The continuous variable can take any value within a range.

The key difference between discrete and continuous data is that discrete data contains the integer or whole number. Still, continuous data stores the fractional numbers to record different types of data such as temperature, height, width, time, speed, etc.

#### **Examples of Continuous Data :**

* Height of a person
* Speed of a vehicle
* “Time-taken” to finish the work
* Wi-Fi Frequency
* Market share price

**4) Categorise the following datasets with respect to quantitative and qualitative data types:**

**(i) Grading in exam: A+, A, B+, B, C+, C, D, E**

**(ii) Colour of mangoes: yellow, green, orange, red**

**(iii) Height data of a class: [178.9, 179, 179.5, 176, 177.2, 178.3, 175.8,...]**

**(iv) Number of mangoes exported by a farm: [500, 600, 478, 672, ...]**

Ans- (i) **Ordinal Data**

(ii) [Nominal](https://www.scribbr.com/statistics/nominal-data/) data

(iii) **Continuous Data**

**(iv) Discrete Data**

**5) Explain the concept of levels of measurement and give an example of a variable for each level.**

Ans- Levels of measurement, also called scales of measurement, tell you how precisely [variables](https://www.scribbr.com/methodology/independent-and-dependent-variables/) are recorded. In scientific research, a variable is anything that can take on different values across your data set (e.g., height or test scores).

There are 4 levels of measurement:

* [Nominal](https://www.scribbr.com/statistics/nominal-data/): the data can only be categorized
* [Ordinal](https://www.scribbr.com/statistics/ordinal-data/): the data can be categorized and ranked
* [Interval](https://www.scribbr.com/statistics/interval-data/): the data can be categorized, ranked, and evenly spaced
* [Ratio](https://www.scribbr.com/statistics/ratio-data/): the data can be categorized, ranked, evenly spaced, and has a natural zero.

Nominal level :- You can categorize your data by labelling them in mutually exclusive groups, but there is no order between the categories.

* City of birth
* Gender
* Ethnicity
* Car brands
* Marital status

Ordinal level:- You can categorize and rank your data in an order, but you cannot say anything about the intervals between the rankings.

Although you can rank the top 5 Olympic medallists, this scale does not tell you how close or far apart they are in number of wins.

* Top 5 Olympic medallists
* Language ability (e.g., beginner, intermediate, fluent)
* [Likert-type questions](https://www.scribbr.com/methodology/likert-scale/)  (e.g., very dissatisfied to very satisfied)

Interval level:- You can categorize, rank, and infer equal intervals between neighboring data points, but there is no true zero point.

The difference between any two adjacent temperatures is the same: one degree. But zero degrees is defined differently depending on the scale – it doesn’t mean an absolute absence of temperature.

The same is true for test scores and personality inventories. A zero on a test is arbitrary; it does not mean that the test-taker has an absolute lack of the trait being measured.

* Test scores (e.g., IQ or exams)
* Personality inventories
* Temperature in Fahrenheit or Celsius

Ratio level:- You can categorize, rank, and infer equal intervals between neighboring data points, and there is a true zero point.

A true zero means there is an absence of the variable of interest. In ratio scales, zero does mean an absolute lack of the variable.

For example, in the Kelvin temperature scale, there are no negative degrees of temperature – zero means an absolute lack of thermal energy.

Height

Age

Weight

Temperature in Kelvin

**6) Why is it important to understand the level of measurement when analyzing data? Provide an example to illustrate your answer.**

Ans- The level at which you measure a variable determines how you can analyze your data.

The different levels limit which descriptive statistics you can use to get an overall [summary](https://www.scribbr.com/working-with-sources/how-to-summarize/) of your data, and which type of [inferential statistics](https://www.scribbr.com/statistics/inferential-statistics/) you can perform on your data to support or refute your [hypothesis](https://www.scribbr.com/statistics/hypothesis-testing/).

In many cases, your variables can be measured at different levels, so you have to choose the level of measurement you will use before [data collection](https://www.scribbr.com/methodology/data-collection/) begins.

You can measure the variable of income at an ordinal or ratio level.

* Ordinal level: You create brackets of income ranges: $0–$19,999, $20,000–$39,999, and $40,000–$59,999. You ask participants to select the bracket that represents their annual income. The brackets are coded with numbers from 1–3.
* Ratio level: You collect data on the exact annual incomes of your participants.

| Participant | Income (ordinal level) | Income (ratio level) |
| --- | --- | --- |
| A | Bracket 1 | $12,550 |
| B | Bracket 2 | $39,700 |
| C | Bracket 3 | $40,300 |

At a ratio level, you can see that the difference between A and B’s incomes is far greater [than](https://www.scribbr.com/commonly-confused-words/then-vs-than/) the difference between B and C’s incomes.

At an ordinal level, however, you only know the income bracket for each participant, not their exact income. Since you cannot say exactly how much each income differs from the others in your data set, you can only order the income levels and group the participants.

**7) How nominal data type is different from ordinal data type.**

Ans- Nominal scale is a naming scale, where variables are simply “named” or labeled, with no specific order. Ordinal scale has all its variables in a specific order, beyond just naming them. Interval scale offers labels, order, as well as, a specific interval between each of its variable options.

**8) Which type of plot can be used to display data in terms of range?**

Ans- A histogram is a graphical representation of data points organized into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins.

**9) Describe the difference between descriptive and inferential statistics. Give an example of each type of statistics and explain how they are used.**

Ans- Descriptive statistics:

Describe the features of populations and/or samples.

Organize and present data in a purely factual way.

Present final results visually, using tables, charts, or graphs.

Draw conclusions based on known data.

Use measures like central tendency, distribution, and variance.

Inferential statistics:

Use samples to make generalizations about larger populations.

Help us to make estimates and predict future outcomes.

Present final results in the form of probabilities.

Draw conclusions that go beyond the available data.

Use techniques like hypothesis testing, confidence intervals, and regression and correlation analysis.

Descriptive and inferential statistics need to be used hand in hand so as to analyze the data in the best possible way. Some examples of descriptive and inferential statistics are given below:

Suppose the scores of 100 students belonging to a specific country are available. The performance of these students needs to be examined. This data by itself will not yield any valuable results. However, by using descriptive statistics, the spread of the marks can be obtained thus, giving a clear idea regarding the performance of each student.

Now suppose the scores of the students of an entire country need to be examined. Using a sample of, say 100 students, inferential statistics is used to make generalizations about the population.

**10) What are some common measures of central tendency and variability used in statistics? Explain how each measure can be used to describe a dataset.**

## Ans- Mean

The mean is the arithmetic average, and it is probably the measure of central tendency that you are most familiar. Calculating the mean is very simple. You just add up all of the values and divide by the number of observations in your dataset.

## Median

The median is the middle value. It is the value that splits the dataset in half, making it a natural measure of central tendency.

## Mode

The mode is the value that occurs the most frequently in your data set, making it a different type of measure of central tendency than the mean or median.