**Q1. What is Statistics?**

**ANS**

* Statsistics is brunch of science where we collect data, analyse the data and infer a concusion

**Q2. Define the different types of statistics and give an example of when each type might be used.**

**Ans**

* There are 2 types of Statistics

A) Descriptive. B) Inferential

**A) Descriptive : Oraganisisng and summarising data**

* i) Measure of Centeral Tendency

mean, median, mode

* ii) Measure of Disperson

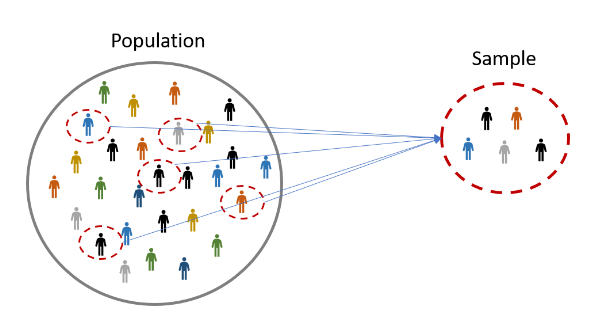
standard daviation and variance

* iii) Distribution of Data

PDF, PMF, Histogram, Box plot

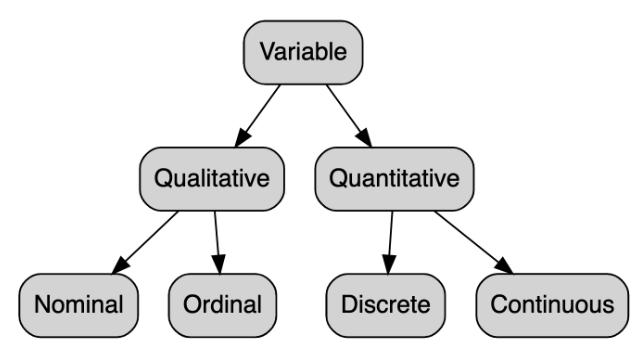
**B). Inferential Stats : Using data which we have measured to get an conclusion form other related data**

* Population : The main data
* Sample : The subset of the data



**Q3. What are the different types of data and how do they differ from each other? Provide an example of each type of data.**

**ANS:**



**Data can be divied in to 2 main categories:**

**A) Quantitative : Numbers**

**i) Descrete: The whole number**

Age of the students, Number of computer at home

**ii) Continious: Decimal Numbers**

Weight, Height, Speed

**B) Qualitative: Categorical**

**i) Nominal: Order does not matters, Cna't be comapired data with each other**

Favorit food, Gender, Blood Group

**ii) Ordinal: Oreders matters**

Designation, Students marks, Roll number

**Q4. 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) Grading in exam: A+, A, B+, B, C+, C, D, E**

* Ordinal Data Type (Ordinal Scale of Measurement)

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

* Nominal data (Scale of measurement : Nominal)

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

* Ordinal (Scale of Measurement Ratio Scale as it can be measured, 0 true starting points, we can order the values, diffrence can be measured)

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

* Ordinal (Scale of Measurement Ratio Scale as it can be measured, 0 true starting points, we can order the values, diffrence can be measured)

**Q5. Explain the concept of levels of measurement and give an example of a variable for each level.**

**ANS**

* Levels of measurement, also known as scales of measurement, refer to the ways in which variables can be classified based on the nature of their values. There are four levels of measurement: nominal, ordinal, interval, and ratio.

**Nominal level of measurement:**

* This is the lowest level of measurement where variables are classified into categories with no intrinsic order. The numbers assigned to the categories are arbitrary and do not reflect any quantitative value.

Examples of nominal variables include gender, race, religion, and marital status. For instance, in a survey, respondents may be asked to choose their gender by selecting 1 for male or 2 for female.

**Ordinal level of measurement:**

* This level involves variables that can be ranked or ordered, but the differences between the values cannot be measured.

Examples of ordinal variables include educational level, socioeconomic status, and customer satisfaction. For instance, a customer satisfaction survey may ask respondents to rate their level of satisfaction with a product or service as excellent, good, fair, or poor.

**Interval level of measurement:**

* This level involves variables where the differences between the values are meaningful and can be measured, but there is no true zero point.

Examples of interval variables include temperature measured in degrees Celsius or Fahrenheit and IQ scores. For instance, if the temperature outside is 20°C, it does not mean that there is no heat present.

**Ratio level of measurement:**

* This is the highest level of measurement where variables have a true zero point, and the differences between the values are meaningful and can be measured. Examples of ratio variables include weight, height, distance, and income. For instance, if someone weighs 50 kilograms, it means that they have no weight at all.

An example of a ratio variable is the income earned by a person. If a person earns $0, it means that they have no income at all. Also students marks

**Q6. Why is it important to understand the level of measurement when analyzing data? Provide an example to illustrate your answer.**

**ANS**

* It is important to understand the level of measurement when analyzing data because the level of measurement determines the type of statistical analysis that can be applied to the data. Using the wrong statistical analysis based on the level of measurement can result in inaccurate conclusions and misleading results.

**For example:**

* If we have a variable that is measured at the nominal level, such as political party affiliation (Republican, Democrat, Independent, etc.), we cannot compute a mean or standard deviation for this variable. Instead, we would use frequency counts or percentages to describe the distribution of the variable. If we were to apply a statistical analysis that assumes a continuous measurement scale, such as a t-test, the results would be invalid and misleading.
* On the other hand, if we have a variable that is measured at the ratio level, such as income, we can compute meaningful summary statistics such as mean, median, and standard deviation. However, if we were to apply a statistical analysis that assumes an ordinal or nominal level of measurement, such as a chi-square test, we would not be able to take full advantage of the information provided by the variable.
* Therefore, it is important to understand the level of measurement of the variables being analyzed to select appropriate statistical analyses and draw valid conclusions.

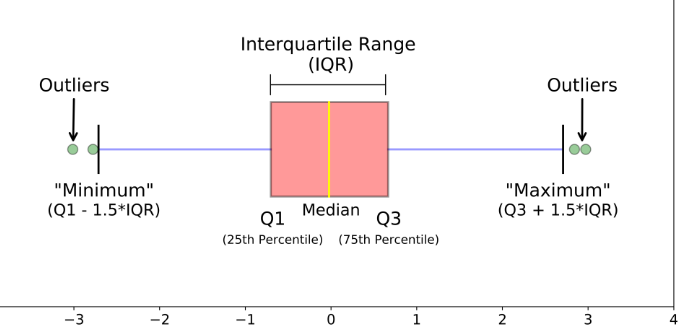
**Q7. How nominal data type is different from ordinal data type.**

**ANS**

* Nominal and ordinal data types are two levels of measurement used in statistics that are different from each other in several ways.
* Nominal data is a categorical data type where variables are classified into categories with no intrinsic order. The numbers assigned to the categories are arbitrary and do not reflect any quantitative value. Examples of nominal variables include gender, race, religion, and marital status. Nominal data can be analyzed using frequency counts or percentages, but we cannot compute a meaningful order or rank among the categories. In other words, we cannot say that one category is "higher" or "lower" than another.
* On the other hand, ordinal data is also a categorical data type, but the categories can be ranked or ordered. Ordinal data has an inherent order, but the differences between the values cannot be measured.
* Examples of ordinal variables include educational level, socioeconomic status, and customer satisfaction. We can say that one category is "higher" or "lower" than another based on the rank or order, but we cannot say how much higher or lower.
* To illustrate the difference between nominal and ordinal data types, let's consider the example of a survey that asks respondents to rate their level of agreement with a statement on a scale of 1 to 5, where 1 represents strongly disagree and 5 represents strongly agree. The responses can be analyzed as either nominal or ordinal data.
* If we treat the responses as nominal data, we would simply count the number of respondents in each category, such as the number of respondents who selected 1, 2, 3, 4, or 5. In this case, the responses are simply different categories with no inherent order.
* If we treat the responses as ordinal data, we would assign a rank or order to each category, such as 1 for strongly disagree, 2 for disagree, 3 for neutral, 4 for agree, and 5 for strongly agree. In this case, we can say that the responses are ordered, but we cannot measure the difference between each rank.
* In summary, nominal and ordinal data types are both categorical data types, but they differ in whether the categories can be ranked or ordered and whether the differences between the values can be measured.

**Q8. Which type of plot can be used to display data in terms of range?**

**ANS**



A box plot, also known as a box and whisker plot, can be used to display data in terms of range. A box plot is a graphical representation of a dataset that displays the distribution of the data based on the five-number summary: the minimum value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum value.

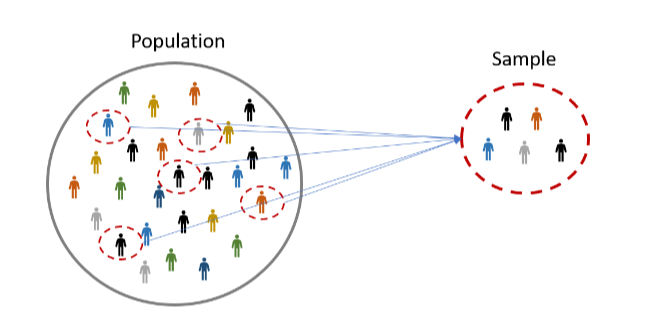
The box in the box plot represents the interquartile range (IQR), which is the range of values between the first quartile and the third quartile. The whiskers extend from the box to the minimum and maximum values of the dataset, and any outliers are represented as individual points beyond the whiskers.

By displaying the range of the data using the five-number summary, a box plot provides a visual summary of the central tendency, variability, and range of the data. Box plots are useful for comparing distributions of different datasets or for comparing the same dataset under different conditions.

For example, a box plot could be used to display the range of salaries for employees in different departments of a company. The box plot would show the median salary, the range of salaries between the first and third quartiles, and any outliers in each department. This would allow us to compare the salary ranges between departments and identify any departments with unusually high or low salaries.

**Q9. 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 and inferential statistics are two branches of statistics used to analyze data in different ways.**

**Descriptive statistics :**

* is concerned with summarizing and describing the characteristics of a dataset. It involves calculating summary statistics such as mean, median, mode, standard deviation, and range to describe the central tendency, variability, and range of the data. Descriptive statistics is useful for organizing and presenting data in a meaningful way, but it does not involve any generalization beyond the specific data being analyzed.
* An example of descriptive statistics would be calculating the mean and standard deviation of test scores for a group of students. These summary statistics would provide an overview of the average performance and the variability of the test scores in the group. Descriptive statistics can be used to identify patterns, trends, and outliers in the data, but it does not involve making any inferences beyond the specific data being analyzed.

**Inferential statistics:**

* is on the other hand, is concerned with making generalizations and drawing conclusions about a population based on a sample of data. It involves using probability theory and statistical models to test hypotheses, estimate parameters, and make predictions about the population. Inferential statistics is used to make inferences beyond the specific data being analyzed and to draw conclusions about the larger population.
* An example of inferential statistics would be conducting a hypothesis test to compare the mean test scores of two different groups of students. By collecting a sample of test scores from each group and using inferential statistics, we can make a statement about whether the difference in mean test scores is statistically significant or due to chance. Inferential statistics can be used to test hypotheses, make predictions, and estimate parameters with a certain level of confidence.
* In summary, descriptive statistics is used to describe and summarize the characteristics of a dataset, while inferential statistics is used to make inferences and draw conclusions about a population based on a sample of data. Descriptive statistics is useful for organizing and presenting data, while inferential statistics is used for hypothesis testing, parameter estimation, and prediction making.

**Q10. 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**

* Measures of central tendency and variability are used in statistics to describe the distribution of a dataset. They provide a summary of the location and spread of the data, respectively. Some common measures of central tendency and variability are:

**Measures of central tendency:**

**Mean:**

The mean is the average value of the dataset. It is calculated by adding up all the values in the dataset and dividing by the number of values. The mean is used to describe the typical value or central location of the data. However, it can be sensitive to outliers in the data.

**Median:**

The median is the middle value of the dataset when it is ordered from smallest to largest. It is used to describe the central location of the data and is less sensitive to outliers than the mean.

**Mode:**

The mode is the most common value in the dataset. It is used to describe the peak or most frequent value in the data and can be used for categorical data. Measures of variability:

**Range:**

The range is the difference between the maximum and minimum values in the dataset. It is used to describe the spread of the data, but it can be sensitive to outliers. Standard deviation: The standard deviation is a measure of how much the values in the dataset vary from the mean. It is used to describe the spread of the data and is more robust to outliers than the range. It provides a measure of how much the data is dispersed around the mean.

**Interquartile range:**

The interquartile range (IQR) is the range of the middle 50% of the data when it is ordered from smallest to largest. It is calculated by subtracting the first quartile from the third quartile. It is used to describe the spread of the data and is less sensitive to outliers than the range.

In [ ]: