

In [1]:
from IPython import display
display.Image("1.png")

Out[1]:
Q1. What is Statistics?

Ans. Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. It involves the use of quantitative methods to extract meaning and insights from data sets. Statistics can be applied to a wide range of fields, including business, economics, social sciences, natural sciences, medicine, engineering, and many others. In essence, statistics is all about making sense of the information we have by using mathematical techniques and tools to find patterns, relationships, and trends within the data. This helps us make informed decisions, draw accurate conclusions, and solve real-world problems. For example : Imagine a company wants to know how many of their customers are satisfied with their product. They could collect data from a survey and use statistical methods to analyze the results, such as calculating the percentage of satisfied customers, identifying common themes in feedback, and comparing results across different demographics. This information can help the company make improvements to their product and increase customer satisfaction.

In [2]:
from IPython import display
display.Image("2.png")

Out[2]:
Q2. Define the different types of statistics and give an example of when each type might be used.

Ans. There are two types of statistics: 1: Descriptive , 2:Inferential

1. Descriptive statistics: Descriptive statistics involves summarizing and describing data through measures such as mean, median, mode, range, standard deviation, and variance. These statistics help us understand the characteristics of the data, such as its central tendency, variability, and distribution.

Example : A researcher wants to study the distribution of ages in a population. They collect data on the ages of a sample of people and use descriptive statistics to calculate the mean, median, and standard deviation of the ages. This information can help the researcher describe the age distribution of the population and identify any outliers or patterns in the data.

1. Inferential statistics: Inferential statistics involves making predictions or drawing conclusions about a population based on a sample of data. This type of statistics uses probability theory to estimate the likelihood that a particular conclusion is true.

Example: A pharmaceutical company wants to test the effectiveness of a new drug. They conduct a randomized controlled trial with a sample of patients and use inferential statistics to determine whether the drug has a significant effect compared to a placebo. This information can help the company make decisions about whether to pursue further research or seek approval from regulatory agencies.

In [3]:
from IPython import display
display.Image("3.png")

Out[3]:
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 are of two types: 1- Qualitative data -Qualitative data consists of descriptive or categorical information. 2- Quantitative data - Quantitative data consists of numerical measurements.

Qualitative and Quantiative are further classified into below categories :

. Qualitative Data :

1.Nominal Data: Descriptive or categorical information that cannot be ranked or ordered. Example- gender, car manufacturer, Cuisine.

2.Ordinal Data: Descriptive or categorical information that can be ranked or ordered, but the intervals between categories are not necessarily equal. Example-Education level, socioeconomic status,Clothing Sizes(S,M,L,XL).

. Quantiative Data :

1.Discrete Data: Numerical values that can only take on specific values, typically integers.Example-Number of children in a family, number of cars in a household.

2.Continuous Data: Numerical values that can take on any value within a range.Example-Length, Temperature, Humidity, Time.

In [1]:
from IPython import display
display.Image("4.png")

Out[1]:
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. (1). Qualitative data type.

(2). Qualitative data type.

(3). Quantitative data type.

(4). Quantitative data type.

In [5]:
from IPython import display
display.Image("5.png")

Out[5]:
Q5. Explain the concept of levels of measurement and give an example of a variable for each level.

Ans. There are four commonly recognized levels of measurement: nominal, ordinal, interval, and ratio.

1. Nominal: Categorical variables with no inherent order or numerical value. Example-Country of origin, favorite color,political affiliation.
- 2.Ordinal: Categorical variables with a specific order or rank, but unequal or non-meaningful differences between categories. Example- Likert scale responses (strongly disagree, disagree, neutral, agree, strongly agree), stage of cancer (stage I, stage II, stage III, stage IV).
- 3.Interval: Numerical variables with a specific order and equal intervals between values, but no true zero point. Example- Temperature (Celsius or Fahrenheit), Standardized test scores (SAT, ACT), credit scores, calendar years.
4. Ratio: Numerical variables with a specific order, equal intervals between values, and a true zero point. Example- income,distance traveled, age

In [2]:
from IPython import display
display.Image("6.png")

Out[2]:
Q6. Why is it important to understand the level of measurement when analyzing data? Provide an example to illustrate your answer.

Ans. Understanding the level of measurement of a variable is important because it determines the types of statistical analyses that can be performed and the appropriate interpretation of the results. In other words, the level of measurement helps researchers to identify the appropriate statistical procedures and tests to use when analyzing data. Using the wrong statistical procedure for a given level of measurement can result in misleading or inaccurate conclusions.

In summary, understanding the level of measurement of variables is crucial for selecting appropriate statistical methods and interpreting results accurately in data analysis.

For example, let's consider a study that examines the relationship between income and job satisfaction among employees in a company. The income variable is a ratio-level variable because it has a true zero point (no income) and equal intervals between values. In contrast, job satisfaction is an ordinal-level variable because it has a specific order or rank, but unequal or non-meaningful differences between categories. If the researcher were to use a correlation analysis to examine the relationship between these two variables, the result would be incorrect because correlation is only appropriate for ratio-level variables. Instead, a more appropriate statistical test for analyzing the relationship between a ratio-level and ordinal-level variable would be a chi-squared test or a t-test.

In [7]:
from IPython import display
display.Image("7.png")

Out[7]:
Q7. How nominal data type is different from ordinal data type.

Ans. Nominal data type: Nominal data is the lowest level of measurement and refers to data that is categorized without any numerical or quantitative value. In other words, nominal data consists of categories that are mutually exclusive and do not have any inherent order or ranking. Examples of nominal data include gender (male/female), hair color (blonde/brunette/red), and marital status (single/married/divorced). Ordinal data type: Ordinal data, on the other hand, is the second level of measurement and refers to data that has an inherent order or ranking, but the differences between categories may not be equal or meaningful. Examples of ordinal data include educational level (elementary school, high school, college), level of agreement (strongly disagree, disagree, neutral, agree, strongly agree), and rating scales (poor, fair, good, excellent).

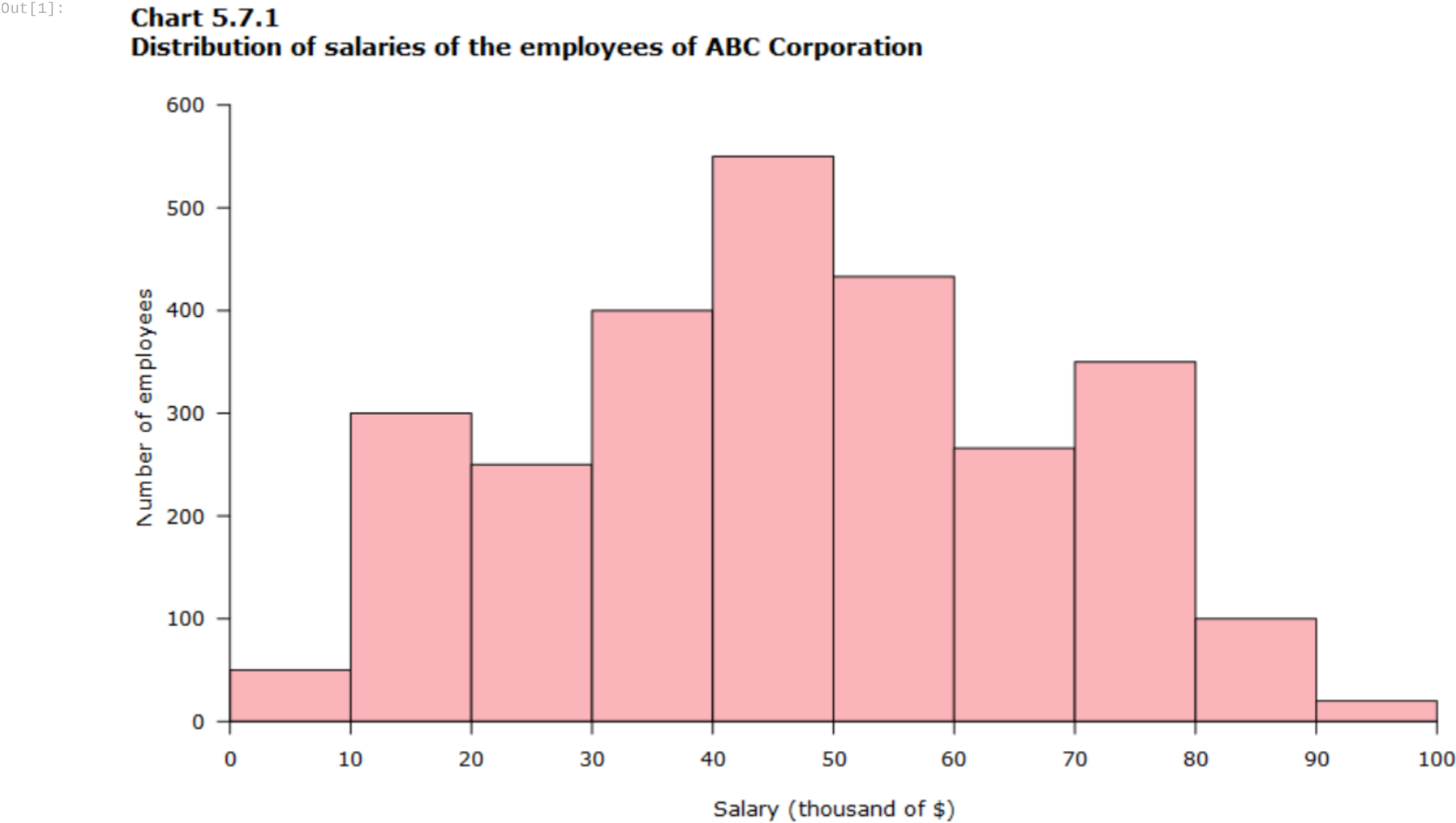
In [8]:
from IPython import display
display.Image("8.png")

Out[8]:
Q8. Which type of plot can be used to display data in terms of range?

Ans. Histogram:

A histogram is particularly useful for displaying continuous data, as it allows you to see the shape of the distribution, including any peaks or clusters, as well as the range and spread of the data. It can also help identify outliers or unusual patterns in the data.

In [1]:
from IPython import display
display.Image("@.png")



In [9]:
from IPython import display
display.Image("9.png")

Out[9]:
Q9. Describe the difference between descriptive and inferential statistics. Give an example of each type of statistics and explain how they are used.

Ans. 1.Descriptive statistics:

Descriptive statistics are used to summarize and describe the basic features of a dataset. They are used to analyze the characteristics of the data, such as central tendency, variability, and distribution. Descriptive statistics do not allow for generalization of the data beyond the sample being analyzed.

An example of descriptive statistics is calculating the mean, median, and mode of a dataset. These measures provide information about the central tendency of the data and are used to summarize the dataset. Another example of descriptive statistics is creating a histogram or bar chart to visually represent the distribution of the data.

2.Inferential statistics:

Inferential statistics, on the other hand, are used to make inferences or draw conclusions about a larger population based on a smaller sample of data. They allow researchers to make generalizations about the population and estimate the likelihood of certain outcomes based on the sample data.

An example of inferential statistics is conducting a hypothesis test to determine if there is a significant difference between two groups of data. This involves calculating a test statistic and comparing it to a critical value to determine if the difference is statistically significant. Another example is calculating a confidence interval to estimate the range of values that a population parameter (e.g. mean or proportion) is likely to fall within.

In [10]:
from IPython import display
display.Image("10.png")

Out[10]:
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. Central tendency

1. Mean- Arithmetic average of a set of values.
2. Median- Middle value in a dataset when it is ordered from smallest to largest.
3. Mode- Most common value in a dataset.

Variability

1. Range- Difference between the largest and smallest values in a dataset. 2.Variance- Average of the squared differences from the mean.
2. Standard Deviation- Square root of the variance.

In []: