Statistics Assignment 1

Q1. What is Statistics?

Statistics is the science of Collecting, Organizing & Analyzing the Data & it's helps us to make a better decision.

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

There are two types of statistics
1. **Inferential Statistics**

   Inferential statistics take a random sample of data taken from population data to make inferences about the population. They used to take predictions, hypotheses & determine relationships.

   Examples are:
   - Z test
   - T test
   - Hypotheses Testing

2. **Descriptive Statistics**

   Descriptive Statistics Summarize & Describe the main features of Dataset.These Summarize may be graphical & Numerical.

   Examples are:

   - Measure of Central tendency(Mean,Median,Mode)
   - Measure of Dispersion(Variance, Standard Deviation)

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

- **Qualitative (Categorical) Data**
  - Qualitative data describe qualities or characteristics and are typically non-numeric. They can be further divided into two subtypes:
    - **Nominal Data:** These are categories with **no inherent** order. Examples include gender (male, female), blood type (A, B, AB, O), and marital status (single, married, divorced).
    - **Ordinal Data:** These are categories with a meaningful order or **ranking** but the intervals between the ranks are not necessarily equal. Examples

include education level (high school, bachelor's, master's, PhD), and customer satisfaction ratings (poor, fair, good, excellent).

- **Quantitative (Numerical) Data**
  - Quantitative data represent numerical values and can be measured. They are further divided into two subtypes:
    - **Discrete Data:** These are countable values that can take on a finite number of distinct values. Examples include the number of students in a class, number of cars in a parking lot, and number of pages in a book.
    - **Continuous Data:** These can take on any value within a range and are often measured. Examples include height, weight, temperature, and time.

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

Ans: Qualitative Data(Nominal)

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

Ans: Qualitative(Nominal)

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

Ans: Qualitative(Continuous Data)

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

Ans: Qualitative(Discrete Data)

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

In Data Collection Phase to measure that particular data through some scale we need to know the types of scale to measure the data

Scale	Description	Examples
Nominal	Categories with names or labels, no inherent order	Gender, marital status, blood type
Ordinal	Ordered categories, intervals between categories are not equal	Education level, customer satisfaction
Interval	Equal intervals between values, no true zero point	Temperature (Celsius, Fahrenheit), IQ scores, dates on a calendar
Ratio	Equal intervals between values, true zero point	Height, weight, age, income, temperature (Kelvin)

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

In Data Collection Phase to measure that particular data through some scale we need to know the types of scale to measure the data

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

Nominal are categories with **no inherent** order Examples include gender (male, female), blood type (A, B, AB, O), and marital status (single, married, divorced) but Ordinal Data are categories with a meaningful order or ranking but the intervals between the ranks are not necessarily equal. Examples include education level (high school, bachelor's, master's, PhD), and customer satisfaction ratings (poor, fair, good, excellent).

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

To display data in terms of range, you can use several types of plots depending on the context and the details you want to visualize. Here are a few common options:

Box Plot, Histogram

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

The main difference between inferential & descriptive statistics is inferential statistics used to take a decision & descriptive statistics is to describe & summarize the data.

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.

Measures of Central Tendency

These measures describe the center or typical value of a dataset.

1. Mean (Arithmetic Mean)
   ○ Definition: The sum of all data points divided by the number of data points.
   ○ Use: The mean is used to represent the average value of the dataset. It is sensitive to extreme values (outliers), which can skew the mean.
   ○ Example: For a dataset [2,4,6,8,10][2, 4, 6, 8, 10][2,4,6,8,10], the mean is (2+4+6+8+10)/5=6(2+4+6+8+10)/5 = 6(2+4+6+8+10)/5=6.
2. Median
   ○ Definition: The middle value of a dataset when it is ordered from least to greatest. If there is an even number of observations, the median is the average of the two middle numbers.
   ○ Use: The median is used to represent the central value of a dataset and is not affected by outliers. It is a better measure of central tendency for skewed distributions.
   ○ Example: For a dataset [2,4,6,8,10][2, 4, 6, 8, 10][2,4,6,8,10], the median is 666. For [2,4,6,8][2, 4, 6, 8][2,4,6,8], the median is (4+6)/2=5(4+6)/2 = 5(4+6)/2=5.
3. Mode
   ○ Definition: The value that appears most frequently in a dataset.
   ○ Use: The mode is useful for categorical data where we want to know the most common category. It can also be used for numerical data to identify the most frequent value.
   ○ Example: For a dataset [1,2,2,3,4][1, 2, 2, 3, 4][1,2,2,3,4], the mode is 222.

Measures of Variability

These measures describe the spread or dispersion of a dataset.

1. Range
   ○ Definition: The difference between the maximum and minimum values in a dataset.

- ○ Use: The range gives a basic indication of the spread of the data. It is very sensitive to outliers.
- ○ Example: For a dataset $[2,4,6,8,10][2, 4, 6, 8, 10][2,4,6,8,10]$, the range is $10-2=810 - 2 = 810-2=8$.

2. Variance
   - ○ Definition: The average of the squared differences between each data point and the mean.
   - ○ Use: Variance provides a measure of how much the data points vary from the mean. It is used in many statistical procedures, including hypothesis testing and regression analysis.
   - ○ Example: For a dataset $[2,4,6,8,10][2, 4, 6, 8, 10][2,4,6,8,10]$ with a mean of 666, the variance is $((2-6)2+(4-6)2+(6-6)2+(8-6)2+(10-6)2)/5=8((2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2)/5 = 8((2-6)2+(4-6)2+(6-6)2+(8-6)2+(10-6)2)/5=8$.

3. Standard Deviation
   - ○ Definition: The square root of the variance.
   - ○ Use: Standard deviation provides a measure of the average distance of each data point from the mean. It is easier to interpret than variance because it is in the same units as the data.
   - ○ Example: For the dataset $[2,4,6,8,10][2, 4, 6, 8, 10][2,4,6,8,10]$ with a variance of 888, the standard deviation is $8≈2.83\sqrt{8} \approx 2.838≈2.83$.

4. Interquartile Range (IQR)
   - ○ Definition: The difference between the first quartile (25th percentile) and the third quartile (75th percentile).
   - ○ Use: IQR measures the spread of the middle 50% of the data. It is not affected by outliers and is useful for skewed distributions.
   - ○ Example: For a dataset $[1,2,4,5,6,8,9,10,12][1, 2, 4, 5, 6, 8, 9, 10, 12][1,2,4,5,6,8,9,10,12]$, the first quartile (Q1) is 444 and the third quartile (Q3) is 999, so the IQR is $9-4=59 - 4 = 59-4=5$.