

Statistics Basic -1

July 3, 2023

Q1. What is Statistics?

Statistics is a science of collecting organizing and analyzing the data is called as statistics

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) Descriptive Statistics # 2) Inferential Statistics

- 1) Descriptive statistics —> It consists of organizing and summarizing the data measure of central tendency which in measure of central tendency topics are like mean median mode measure of dispersion variance and standard deviation different type of distribution of data eg histogram, pdf, pmf
- 2) inferential statistics —> it consists of using data we have measured to form conclusion in inferential statistics topics are like Z-test T-test CHI square Hypothesis testing P value Ho value H value significance value

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

there are various types of data each with its own characteristics and suitable analysis methods here are some common types of data Nominal Data: Nominal data represents categories or groups that have no intrinsic order or ranking. It involves qualitative distinctions, and the data points are typically labels or names. Examples include gender (male, female), eye color (blue, brown, green), or country of origin (USA, UK, Canada).

Ordinal Data: Ordinal data also represents categories, but they have a specific order or ranking associated with them. The intervals between the categories may not be uniform. Examples include survey ratings like 1-5 scale satisfaction levels (1 = very dissatisfied, 5 = very satisfied) or education levels (e.g., elementary, middle, high school, college).

Interval Data: Interval data represents numeric or continuous values with equal intervals between them, but there is no true zero point. Arithmetic operations such as addition and subtraction can be performed on interval data. Examples include temperature measured in degrees Celsius or Fahrenheit, or years (e.g., 1990, 2000, 2010).

Ratio Data: Ratio data is similar to interval data, but it has a true zero point, representing the absence of the measured quantity. This type of data allows for all arithmetic operations, including multiplication and division. Examples include weight, height, or time taken to complete a task.

Discrete Data: Discrete data consists of whole numbers or countable values that have distinct separations. These values typically result from counting or enumerating. Examples include the number of students in a class, the number of cars in a parking lot, or the number of goals scored in a soccer match.

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, ...]
- i) Nominal Data(Qualitative)
 - ii) Ordinary Data(Qualitative)
 - iii) Continous Data(Quantitative)
 - iv) District Data(Quantitative)

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

—> Understanding the level of measurement is crucial when analyzing data because it determines the appropriate statistical techniques and operations that can be applied to the data. It helps in selecting the right analysis methods, interpreting the results accurately, and drawing meaningful conclusions. Here's an example to illustrate the importance of understanding the level of measurement:

Let's consider a dataset that contains information about the favorite colors of a group of individuals. The dataset includes the following data:

John: Blue Mary: Green David: Red Sarah: Blue Michael: Green If we treat this dataset as quantitative data without understanding the level of measurement, we might mistakenly perform mathematical operations such as calculating the mean or standard deviation. However, since the data represents qualitative categories (colors), these calculations would not provide meaningful insights

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

Nominal and ordinal data types are both types of qualitative data, but they differ in terms of the nature of the categories and the level of measurement. Here's how nominal data type is different from ordinal data type:

Nominal Data:

Nominal data represents categories or labels that have no inherent order or ranking. The categories in nominal data are mutually exclusive and do not have a numerical value associated with them. Nominal data can be thought of as "naming" or classifying data into distinct groups or categories. Examples of nominal data include gender (male, female), eye color (blue, brown, green), or types of fruit (apple, orange, banana)

Ordinal Data:

Ordinal data represents categories or labels with an inherent order or ranking. The categories in ordinal data have a meaningful sequence or hierarchy. The order between the categories is meaningful but the intervals between them may not be uniform or quantifiable. Ordinal data captures the notion of relative rankings or preferences. Examples of ordinal data include survey ratings (1

= strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree), education levels (elementary, middle, high school, college), or military ranks (private, sergeant, captain, general).

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

—> To display data in terms of range, a common plot that can be used is a box plot (also known as a box-and-whisker plot). A box plot provides a visual representation of the range, as well as other statistical measures such as the median, quartiles, and outliers.

A box plot typically consists of a box that represents the interquartile range (IQR), with a line inside indicating the median. The whiskers extend from the box to show the minimum and maximum values within a specified range or to the extent of the data. Outliers, if any, are shown as individual points beyond the whiskers.

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

—> Descriptive Statistics: Descriptive statistics involves summarizing and describing the main characteristics of a dataset. It focuses on organizing, presenting, and analyzing data to provide a clear and meaningful description of the data. Descriptive statistics aims to uncover patterns, trends, and insights within the data without making inferences or generalizations beyond the observed data.

Example: Suppose you have a dataset containing the heights of students in a classroom. Descriptive statistics would involve calculating measures such as the mean height, median height, and standard deviation to summarize the central tendency and variability of the heights within the observed data.

inferential Statistics: Inferential statistics involves drawing conclusions, making predictions, or generalizing findings from a sample to a larger population. It goes beyond the observed data to make inferences about the population based on statistical analysis. Inferential statistics uses probability theory and sampling techniques to estimate population parameters and test hypotheses.

Example: Consider a scenario where you want to determine if there is a significant difference in the average heights of male and female students in the entire school population. You could collect a sample of heights from both genders, conduct statistical tests (such as a t-test) on the sample data, and use inferential statistics to make conclusions about the entire population.

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.

—> In statistics, measures of central tendency and variability are used to describe the distribution of a dataset. Here are some common measures for each category and their purposes:

Measures of Central Tendency:

Mean: The mean is the average of all the values in a dataset. It is calculated by summing up all the values and dividing by the total number of observations. The mean represents the “typical” value or the center of the distribution. It is sensitive to extreme values and can be influenced by outliers.

Median: The median is the middle value in a dataset when it is arranged in ascending or descending order. If there is an even number of observations, the median is the average of the two middle values. The median represents the value that divides the dataset into two equal parts. It is less affected by extreme values and is a useful measure when there are outliers.

Mode: The mode is the value or values that occur most frequently in a dataset. It represents the most common value(s) in the distribution. The mode is useful for categorical data and discrete variables. Unlike the mean and median, the mode does not provide information about the overall distribution of the data

[]: