Statistics Basic-1

January 11, 2024

1 Q1. What is Statistics?

1.0.1 ANS:

- Statistics is a branch of mathematics that deals with the collection, organization, analysis, interpretation, and presentation of data . It is used in many different fields such as business, medicine, biology, psychology and social sciences . The main purpose of using statistics is to plan the collected data in terms of experimental designs and statistical surveys .

[ ]:

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

2.0.1 ANS: There are two main types of statistics: Descriptive Statistics and Inferential Statistics.

- Descriptive Statistics is used to describe the characteristics of a data set, such as its center, spread, and shape. It provides a summary of the data in a form that is easy to understand . For example, if we want to know the average height of students in a class, we can use descriptive statistics to calculate the mean height of the students .

- Inferential Statistics is used to make predictions or inferences about a population based on a sample of data . It involves estimating the characteristics of a population based on a sample of data . For example, if we want to know the average height of all students in a school, we can use inferential statistics to estimate the mean height of all students in the school based on a sample of students .

[ ]:

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.

3.0.1 ANS: There are four types of data in statistics: Nominal, Ordinal, Interval, and Ratio.

- Nominal data is a type of categorical data that is used to label variables without providing any quantitative value . For example, the colors of cars in a parking lot can be classified as

nominal data .

- Ordinal data is a type of categorical data that is used to label variables with a specific order or ranking . For example, the grades of students in a class can be classified as ordinal data .

- Interval data is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful . For example, temperature measured in Celsius or Fahrenheit can be classified as interval data .

- Ratio data is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful and there is a true zero point . For example, weight, height, and age can be classified as ratio data .

[ ]:

4 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, ...]

4.0.1 ANS: Here are the different types of data for each dataset:

- (i) Grading in exam: A+, A, B+, B, C+, C, D, E ##### This dataset is an example of ordinal data, which is a type of categorical data that is used to label variables with a specific order or ranking.

- (ii) Colour of mangoes: yellow, green, orange, red ##### This dataset is an example of nominal data, which is a type of categorical data that is used to label variables without providing any quantitative value .

- (iii) Height data of a class: [178.9, 179, 179.5, 176, 177.2, 178.3, 175.8,...] ##### This dataset is an example of interval data, which is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful .

- (iv) Number of mangoes exported by a farm: [500, 600, 478, 672, ...] ##### This dataset is an example of ratio data, which is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful and there is a true zero point .

[ ]:

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

5.0.1 ANS: In statistics, levels of measurement refer to the different ways in which data can be categorized and analyzed . There are four levels of measurement: Nominal, Ordinal, Interval, and Ratio .

- Nominal data is a type of categorical data that is used to label variables without providing any quantitative value . For example, the colors of cars in a parking lot can be classified as nominal data .

- Ordinal data is a type of categorical data that is used to label variables with a specific order or ranking . For example, the grades of students in a class can be classified as ordinal data .

- Interval data is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful . For example, temperature measured in Celsius or Fahrenheit can be classified as interval data .

- Ratio data is a type of numerical data that is used to measure variables on a scale where the difference between two values is meaningful and there is a true zero point . For example, weight, height, and age can be classified as ratio data .

Here are some examples of variables for each level of measurement:

- Nominal data: Gender (Male/Female), Marital status (Married/Single/Divorced), Blood type (A/B/O/AB), Eye color (Blue/Green/Brown), Hair color (Blonde/Brown/Black/Red).

- Ordinal data: Grades (A+, A, B+, B, C+, C, D, E), Education level (High school diploma/Bachelor's degree/Master's degree/Doctorate), Military rank (Private/Corporal/Sergeant/Lieutenant/Captain/Major/Colonel/General).

- Interval data: Temperature (measured in Celsius or Fahrenheit), IQ scores, Dates (measured in days, months, years), Time (measured in hours, minutes, seconds).

- Ratio data: Height, Weight, Age, Income, Distance.

[ ]:

6 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 is crucial when analyzing data because it determines the type of statistical analysis that can be performed on the data . The level of measurement of a variable determines the type of statistical tests that can be used to analyze the data . For example, nominal data can only be categorized, while ordinal data can be categorized and ranked . Interval data can be categorized, ranked, and evenly spaced, while ratio data can be categorized, ranked, evenly spaced, and has a natural zero .

For instance, if we have nominal data, we can only use descriptive statistics such as frequency tables, mode, and percentages to analyze the data . However, if we have ratio data, we can use more advanced statistical techniques such as regression analysis, ANOVA, and t-tests to analyze the data .

- Let's consider an example to illustrate the importance of understanding the level of measurement. Suppose we want to analyze the relationship between the height of students and their grades in a class . If we use nominal data to categorize the height of students, we can only use descriptive statistics such as frequency tables and mode to analyze the data . However, if we use interval data to measure the height of students, we can use more advanced statistical techniques such as correlation analysis to analyze the data .

[ ]:

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

7.0.1 ANS: Nominal data and ordinal data are two types of categorical data. The main difference between nominal data and ordinal data is that nominal data is used to label variables without providing any quantitative value, while ordinal data is used to label variables with a specific order or ranking .

- For example, the colors of cars in a parking lot can be classified as nominal data because they are labels without any inherent order or ranking . On the other hand, the grades of students in a class can be classified as ordinal data because they are labels with a specific order or ranking .

7.0.2 Another difference between nominal data and ordinal data is that nominal data can only be categorized, while ordinal data can be categorized and ranked .

- For example, we can categorize the colors of cars in a parking lot as red, blue, green, etc., but we cannot rank them . However, we can categorize the grades of students in a class as A, B, C, D, and E, and we can also rank them from highest to lowest .

[ ]:

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

ANS:

- A box plot is a type of plot that can be used to display data in terms of range . It is also known as a box-and-whisker plot . A box plot displays the distribution of a dataset by showing the median, quartiles, and outliers . The box represents the interquartile range (IQR), which is the range between the first and third quartiles . The whiskers extend from the box to the minimum and maximum values, excluding outliers . Outliers are plotted as individual points beyond the whiskers .

- Box plots are useful for comparing the distribution of different datasets or for comparing the distribution of a single dataset at different times or under different conditions . They are

particularly useful for identifying outliers and for visualizing the spread of the data .

[ ]:

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

9.0.1 ANS:

- Descriptive statistics is a branch of statistics that deals with the collection, organization, analysis, interpretation, and presentation of data . It is used to describe the characteristics of a data set, such as its center, spread, and shape . For example, if we want to know the average height of students in a class, we can use descriptive statistics to calculate the mean height of the students . Descriptive statistics is useful for summarizing and organizing data in a form that is easy to understand .

- Inferential statistics is a branch of statistics that deals with making predictions or inferences about a population based on a sample of data . It involves estimating the characteristics of a population based on a sample of data . For example, if we want to know the average height of all students in a school, we can use inferential statistics to estimate the mean height of all students in the school based on a sample of students . Inferential statistics is useful for making predictions and inferences about the population based on the sample data .

Here are some examples of descriptive and inferential statistics:

- Descriptive statistics: Suppose we have a set of raw data that shows the test scores of 1,000 students at a particular school. We might be interested in the average score and create a graph that helps us visualize the distribution of scores. This allows us to understand the test scores of the students much more easily compared to just staring at the raw data .

- Inferential statistics: Suppose we want to know the average height of all students in a school. We can use inferential statistics to estimate the mean height of all students in the school based on a sample of students . This allows us to make predictions and inferences about the population based on the sample data .

[ ]:

9.0.2 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.

9.0.3 ANS:

- In statistics, measures of central tendency are used to describe the center of a dataset, while measures of variability are used to describe the spread of a dataset.

- Measures of central tendency include the mean, median, and mode . The mean is the sum of all values divided by the total number of values . The median is the middle value in an ordered dataset . The mode is the most frequent value in a dataset . These measures can be

used to describe the center of a dataset and provide a summary of the data in a form that is easy to understand .

- Measures of variability include the range, variance, and standard deviation . The range is the difference between the highest and lowest values in a dataset . The variance is the average of the squared differences from the mean . The standard deviation is the square root of the variance . These measures can be used to describe the spread of a dataset and provide information about how much the data values deviate from the mean .

- For example, suppose we have a dataset that contains the test scores of 100 students. We can use the mean to describe the average score of the students, the median to describe the middle score of the students, and the mode to describe the most frequent score of the students . We can also use the range to describe the difference between the highest and lowest scores, the variance to describe how much the scores deviate from the mean, and the standard deviation to describe the spread of the scores .

[ ]: