

6th Mar Assignment

March 15, 2023

1 Assignment 30

Q1. What is Statistics?

Ans. Statistics is the science collecting, organizing and analyzing the data or simply you can say the numbers that you have collected in order to provide some information is called statistics.

There is Two types of statistics

1. Descriptive Statistics
2. Inferential Statistics

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 statistics: Descriptive stats consists of the organizing and analyzing the data. In descriptive we mainly talks about the measurement of central tendency (which is nothing just Mean, Mode and Median), measurement of dispersion (Variance and Standard Deviation), Histogram, Bar Chart and Pie Chart. Example: Student's grade point average
2. Inferential Statistics: Inferential stats is the tools that statisticians used to draw conclusions about the characteristics of the population, drawn from the characteristics of the sample and decide how certain can be the reliability of the conclusions. Example: You randomly select a sample of 11th graders in your state and collect data on their SAT scores and other characteristics. You can use inferential statistics to make estimates and test hypotheses about the whole population of 11th graders in the state based on your sample 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. There is two type of data and that further divided into two types

1. Quantitative - This data can be measured and is not observable. It can be numerically recorded and represented. Calculations and interpretations can be performed on this data type. E.g. Temperature, Prices, Income, Weights

Types of Quantitative -

- a) Discrete Data - The data values that can attain only certain specific values. It cannot attain a range of values. It can be represented by bar charts E.g. - No. of students, chips in bag, no. of stars in the sky
- b) Continuous Data - The data can contain values between a certain range that is within the highest and lowest values. Continuous data can be tabulated in what is called a frequency distribution. E.g. - Height and Weight of the Student, Daily temperature of the place, wind speed measurement

2. Qualitative - This is used to represent some characteristics or attribute of the data. The facts and figures depicted by the qualitative data cannot be computed. E.g. - Music and art, Intelligence, beauty of person

Types of Qualitative

- a) Nominal Data - It is a subcategory belonging to the sub category of the types of qualitative information. also know as the nominal scale. E.g. - Gender, Eye color, Marital Status
- b) Ordinal Data - Ordianl data is the specific type of data that follow a natural order. E.g. - Feedback Rank, Education Level

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 - Qualitative
- (ii) Colour of mangoes: yellow, green, orange, red - Qualitative
- (iii) Height data of a class: [178.9, 179, 179.5, 176, 177.2, 178.3, 175.8,...] - Quantitative
- (iv) Number of mangoes exported by a farm: [500, 600, 478, 672, ...] - Quantitative

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

Ans.Level of measurement also called scale of measurement, tell you how precisely variable is recorded. In scientific research a variable is anything that can take different values across your dataset. E.g. - heights and test scores

There are 4 level of measurement

1. Nominal - The data can be categorized - E.g. - City of birth, gender, car brands
2. Ordinal - The data can be categorized and rank E.g. - Olympic medal, Feedback rating
3. Interval - The data can be categorized and ranked and evenly spaced - E.g. - Test scores, Temperature
4. Ratio - The data can be categorized and ranked and evenly spaced and had a natural zero E.g. - Height, age, weight

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 essential when analyzing data because it determines the type of statistical analysis that can be performed, the appropriate measures of central tendency and dispersion, and the types of inferences that can be made from the data.

For example, suppose a company conducted a survey to determine the favorite type of ice cream among its customers. The responses could be categorized as follows:

Chocolate Vanilla Strawberry Mint chocolate chip Rocky road In this case, the data is nominal because the categories cannot be ranked. The company could calculate the percentage of customers who prefer each flavor, but they could not determine whether one flavor is more popular than another.

Now, suppose the same company conducted a survey to determine the satisfaction level of its customers with its service. The responses could be categorized as follows:

Very satisfied Satisfied Neutral Dissatisfied Very dissatisfied In this case, the data is ordinal because the categories can be ranked, but the distance between categories is unknown. The company could calculate the percentage of customers who are very satisfied, satisfied, neutral, dissatisfied, and very dissatisfied, but they could not determine the exact difference in satisfaction level between customers who are very satisfied and those who are satisfied.

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

Ans.Nominal data is categorical data that can only be classified into categories or groups, but the categories cannot be ranked. For example, gender, eye color, or race are all examples of nominal data. In nominal data, there is no inherent order, and there is no numerical value assigned to the categories. Nominal data can be analyzed using frequency distribution and percentages.

Ordinal data is also categorical data, but the categories can be ranked or ordered based on some criteria. For example, the ratings of a movie (excellent, good, fair, poor) or the educational level (high school, college, graduate school) are examples of ordinal data.

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

Ans. A range plot is a graph that shows the minimum and maximum values of a variable within a given range. The range chart is useful in monitoring variability in data over time or across different groups. The range chart typically displays a center line that represents the mean value and upper and lower control limits that represent the acceptable range of variation.

A range plot can also be used to display the range of a variable for different categories or groups. For example, a range plot can be used to compare the range of salaries for different job titles or the range of test scores for different schools. In this case, the range chart displays the range of the variable for each category or group along with the mean and upper and lower control limits.

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 are used to summarize and describe the main features of a dataset. Descriptive statistics provide information about the central tendency (e.g., mean, median, mode), variability (e.g., range, variance, standard deviation), and shape of the distribution of the data. Descriptive statistics can be used to provide a summary of the data and to identify patterns or relationships among variables. An example of descriptive statistics is calculating the average score of a group of students on a test, the range of the scores, and the distribution of the scores in terms of frequency.

Inferential statistics, on the other hand, are used to make inferences about a population based on a sample of data. Inferential statistics involve making generalizations about a population based on the characteristics of a sample. Inferential statistics can be used to estimate population parameters, test hypotheses, and assess the reliability of the findings. An example of inferential statistics is conducting a hypothesis test to determine whether there is a significant difference between the average test scores of two groups of students, where one group received a new teaching method and the other did not.

To summarize, descriptive statistics are used to describe and summarize the main features of a dataset, while inferential statistics are used to make inferences about a population based on a sample of data. Descriptive statistics are often used in exploratory data analysis, while inferential statistics are used in hypothesis testing and making decisions based on 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.

Ans.Measures of Central Tendency:

- Mean - The mean is the average value of the dataset and is calculated by adding up all the values and dividing by the number of values. The mean is a useful measure of central tendency when the data is normally distributed and has no outliers. However, the mean is sensitive to outliers, and if there are extreme values in the dataset, the mean may not be a good representation of the typical value.
- Median - The median is the middle value of the dataset when the values are ordered from lowest to highest. The median is a robust measure of central tendency and is less sensitive to outliers than the mean. The median is a useful measure when the data is skewed or has extreme values.
- Mode - The mode is the value that occurs most frequently in the dataset. The mode is useful for identifying the most common value in the dataset, but it is not always a useful measure of central tendency, especially when the dataset has multiple modes.

Measures of variability

- Range - The range is the difference between the maximum and minimum values in the dataset. The range is a useful measure of variability, but it is sensitive to outliers and may not provide a good representation of the spread of the data.
- Variance - The variance is a measure of how spread out the dataset is around the mean. It is calculated by taking the average of the squared differences between each value and the mean. A higher variance indicates that the data is more spread out, while a lower variance indicates that the data is more tightly clustered around the mean.
- Standard Deviation - The standard deviation is the square root of the variance and is a commonly used measure of variability. The standard deviation provides a measure of how much the values deviate from the mean. A higher standard deviation indicates that the values are more spread out, while a lower standard deviation indicates