

Statistics Basic-1

Assignment

Q1. What is Statistics?

Statistics is a science of collecting, organizing & analysing data. its primary purpose is to extract meaningful information from data, make informed decision and draw conclusion about various phenomena or population statistics is widely applied in various fields.

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

The main type of statistics are –

1.Descriptive statistics:-

Purpose: To summarize and describe the main features of a dataset.

E.g Mean: calculating the average score of students in a class to understand their performance.

Median: Finding the middle value of a set of data to represent the central tendency.

Range: Determining the difference between the highest and lowest temp. in a week to understand temp. Variability.

2. Inferential Statistics:

Purpose: To make predictions, inferences, or generalizations, about a population based on a sample.

e.g Hypothesis testing: Determining of a new drug is effective by comparing the treatment group to a control group.

Confidence intervals: Estimating the average income of a population with a certain level of confidence based on a sample of 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.

Data can be categorised into various types based on its nature and characteristics. The main types of data include:

1.Qualitative Data (Categorical data):

Qualitative data consists of non-numeric information& is typically describe in nature.

Nominal data: Nominal data represents categories or labels with no inherent order or ranking. Examples: colors (red, blue, green), types of fruits (apple, banana, orange)

Ordinal data: Ordinal data represents categories with a meaningful order but no fixed interval between them. examples: education levels (e.g high school, college, graduate) customer satisfaction rating (e.g: very satisfied, satisfied, neutral, dissatisfied, and very dissatisfied)

2. Quantitative Data(Numerical Data): Quantitative data consists of numeric values and can be further categorical into discrete and continuous data.

Discrete data: Discrete data consist of distinct, separate values and usually represents counts or whole numbers.

Statistics Basic-1

Assignment

Example: numbers of cars in a parking area, numbers of students in a class.

Continuous Data: Continuous data represents values that can take any real numbers within a given range. It often involves measurements.

Examples: heights of individuals, temp. of degree of Celsius, weight of products

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 - **Qualitative Data**

(ii) Colour of mangoes: yellow, green, orange, red – **Qualitative Data**

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

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

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

Levels of measurement refer to the difference way in which data can be categorized & analysed based on the characteristics of the variables being measured. There are four level of measurement

Nominal level of measurement: nominal data are categorical and represent distinct categories or labels with no inherent order or ranking.

Exam: eye color (blue, brown, black), gender (male, female ,transgender), types of fruit(apple, banana, orange)

Ordinal level of measurement: ordinal data present categories with a meaningful order or ranking but the intervals between the categories are not uniform or meaningful .

e.g education level(high school, bachelors, master, PhD) customers satisfaction rating (very dissatisfied, dissatisfied, neutral, satisfied, very satisfied) socioeconomic status (income, middle income, high income)

Interval level of measurement: Intervals data have a meaningful order, and the intervals between values are equal & meaningful, but there is no true zero pt.

Statistics Basic-1

Assignment

Ex: temp. in Celsius or Fahrenheit (20c, 30c,40c), iq scores(100,110,120), calendar years(2000, 2010, 2020)

Ratio level of measurement: Ratio data have all the characteristic of interval data, but they also have a true zero pt. which signifies the absence of the measurement attributes.

Ex: age (in years, where 0 represents birth), height (in centimetres or inches, where 0 represents no height), income (in dollars, where 0 represents no income)

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

Understanding the level of measurement is crucial when analysing data because it determines the types of statistical analysis and operations that can be applied to the data, as well as the meaningfulness of those operations.

e.g- suppose you are analysing survey data about people's favourite colors, you can count the no. of respondents who prefer each color, but you cannot calculate an average or perform other mathematical operations on the categories.

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

Nominal data & ordinal data consists of categories or labels with no inherent order or ranking it is used to classify data into distinct gps. Or classes.

Ex: gender (male, female, trans-gender), types of vehicles (car, bicycle, bus) there is not meaningful way to ranks or order these categories.

Ordinal data on the other hand, represents categories with a meaningful order or ranking. While the intervals between categories are not necessarily uniform or well-defined there is a clear sense of greater than or less than between the categories.

Ex- Educational attainment (high school diploma, bachelor's degree, master's degree) or customer satisfaction rating (very dissatisfied, dissatisfied, neutral, satisfied, very satisfied).

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

A type of plot that can be used to display data in terms of range is a box plot or box and whisker plot. A box plot provides a visual representation of the distribution of a dataset, including its minimum, maximum, median, quartiles and potential outliers. The "box" in plot represents the interquartile range (the middle 50% of the data), and the "whiskers" extend from the box to the minimum & maximum values in the data.

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 are used to summarize & describe the main features of a dataset. They provide a concise & meaningful representation of data through measures of central

Statistics Basic-1

Assignment

tendency. e.g (mean, median, mode) and measures of variability (e.g range, variance, s.d) . Descriptive statistics help in understanding the characteristics of a good dataset without making inferences Q9. Describe the difference between descriptive and inferential statistics. Give an example of each type of statistics and explain how they are used. About a larger population.

Ex: calculating the average income of a gp of individuals to understand their financial status.

Inferential statistics: inferential statistics used to make inference or predictions about a population based on a sample of data it involves hypothesis testing confidence intervals, and regression analysis, among other techniques. Inferential statistics allow us to draw conclusion about a population beyond the immediate dataset under the study.

Ex: conducting a hypothesis test to determine if a new drug is effective in treating a specific medical condition based on a sample of patients, and then making an inference about its effectiveness for the border 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.

Mean: the mean is the arithmetic average of a dataset and is calculated by adding up all the values and dividing by the number of data pt. it represents the “typical” value in the dataset.

Median: the median is the middle values when the data is ordered from smallest to largest. It is less affected by outliers than the mean & represents the central value.

Mode: the mode is the most frequently occurring value in the dataset. A dataset can have one mode (unimodal) or multiple modes (multimodal).

Measure of variability:

Range: the range is the difference between the max. & min. values in the datasets, providing a simple measure of the spread of the data.

Variance: var measures the avg. squared difference between each data pt. & the mean. It quantifies the overall dispersion of the data.

Standard deviation: the s.d is the square root of the variance & provides a measure of how much individual data pts. Typically deviate from the mean.

Statistics Basic-1

Assignment