Descriptive Statistics UNIT 2

Dr. Praveen Barapatre

Unveiling the Story in Your Data: Descriptive Statistics

- Descriptive statistics are the tools we use to summarize, describe, and organize data. Through this process, we can uncover hidden patterns, trends, and gain valuable insights from our data.

The Two Branches of Statistics

- Statistics can be broadly categorized into two main branches: Descriptive Statistics and Inferential Statistics.
- Descriptive Statistics focus on summarizing and describing data sets.
- Inferential Statistics allow us to draw conclusions about a larger population based on a sample.

The Two Branches of Statistics

- Think of statistics like a tree with two main branches. Descriptive statistics lay the groundwork by summarizing and describing the data we have.
- Inferential statistics take us a step further, allowing us to use this information to make informed guesses about a larger population, even if we haven't observed every single member.

Taming the Data: Why Use Descriptive Statistics?

- Raw data can be overwhelming and difficult to interpret.
- Descriptive statistics help us to organize and condense data into a manageable format
- They enable us to identify central tendencies, how spread out the data is, and even the shape of the data distribution.

Taming the Data: Why Use Descriptive Statistics?

- Imagine a messy desk overflowing with papers. Just looking at it can be overwhelming.
- Descriptive statistics are like the filing cabinet for our data. They help us organize and categorize the information, making it easier to understand the big picture. By using descriptive statistics, we can identify key characteristics of our data, such as where most of the values fall (central tendency), how spread out the data is (dispersion) and even if the data is skewed in a particular direction (shape).

Central Tendency: Finding the Middle Ground

- Measures of central tendency tell us about the typical or average value in a data set.
- Common measures of central tendency include:
  - Mean (average)
  - Median (middle value when data is ordered)
  - Mode (most frequent value)

Central Tendency: Finding the Middle Ground

- Central tendency is like finding the middle ground in our data set. It tells us where most of the values tend to cluster.
- There are three main ways to measure central tendency: the mean, median, and mode. The mean is the most common average, calculated by adding all the values and dividing by the number of values.
- The median is the middle value when the data is ordered from least to greatest. The mode is the most frequent value in the data set.

Dispersion: How Spread Out is the Data?

- Measures of dispersion tell us how spread out the data points are in a data set.
- Common measures of dispersion include:
  - Range (difference between the highest and lowest values)
  - Variance (average of the squared deviations from the mean)
  - Standard Deviation (square root of the variance)

Dispersion: How Spread Out is the Data?

- Dispersion is like measuring how spread out the data points are in our data set. Imagine a group of people standing on a field. If everyone is clustered close together, the dispersion is low.
- If they are spread out all over the field, the dispersion is high. There are three main ways to measure dispersion: range, variance, and standard deviation.
- The range is simply the difference between the highest and lowest values. The variance is the average of the squared deviations of each data point from the mean.