Descriptive Statistics

CCDATSCL | Data Science

What is Descriptive Statistics?

Statistics is the science, or a branch of mathematics that involves collecting, classifying, analyzing, interpreting, and presenting numerical facts and data.

Statistics further breaks down into two types: descriptive statistics and inferential statistics

Descriptive statistics refers to a branch of statistics that involves summarizing, organizing, and presenting data meaningfully and concisely.

It focuses on describing and analyzing a dataset's main features and characteristics without making any generalizations or inferences to a larger population.

The primary goal of descriptive statistics is to provide a clear and concise summary of the data, enabling researchers or analysts to gain insights and understand patterns, trends, and distributions within the dataset.

Descriptive and Inferential Statistics

Descriptive statistics give you a clear picture of what your current data shows.

Inferential statistics makes projections based on that data.

Inferential statistics takes a random sample of data from a portion of the population and describes and makes inferences about the entire population.

Inference – a conclusion reached on the basis of evidence.

What is Descriptive Statistics?

This summary typically includes measures such as:

1. Central tendency (e.g., mean, median, mode)
2. Dispersion (e.g., range, variance, standard deviation)
3. Shape of the distribution (e.g., skewness, kurtosis)

Types of Descriptive Statistics

1. Frequency Distribution
2. Central Tendency
3. Dispersion

Frequency Distribution

Shows the number of times a value appears in the dataset.

Example: 1, 1, 2, 2, 3

Value	Frequency
1	2
2	2
3	1

Central Tendency

- Measures of central tendency estimate a dataset's average or center.
- Central tendency is defined as "the statistical measure that identifies a single value as representative of an entire distribution."
- It aims to provide an accurate description of the entire data.
- Measures of central tendency estimate a dataset's average or center.

Measure	Definition
MEAN	Add up all the scores and divide by the number of scores.
MEDIAN	Arrange the scores in ascending order and find the middle value
MODE	Identify the score(s) that appear(s) most frequently

Dispersion

- Measures the spread of the data.

Measure	Definition
RANGE	Calculate the difference between the highest and lowest scores
VARIANCE	Calculate the average of the squared differences from the mean
STANDARD DEVIATION	Take the square root of the variance