

Lesson 1: Fundamental Concepts, Frequency Distributions and the Graphical Presentation of Data

Statistics refers to data and also to methods of analyzing data.

- **Descriptive statistics** refer to how large volumes of data are converted into useful, readily understood information by summarizing their important characteristics.
 - **Inferential statistics** (or statistical inference) refer to methods used to make forecasts, estimates, or draw conclusions about a larger set of data based on a smaller representative set.
-
- A **population** includes *all* the members of a particular group.
 - A descriptive measure of a population characteristic is known as a **parameter**.
-
- A **sample** is a relatively small subset of the population.
 - A descriptive measure of a sample is known as a **sample statistic**.

Nominal Scales

- Represent the weakest level of measurement.
- Categorize/count data but do not rank them.

Ordinal Scales

- Represent a stronger level of measurement than nominal scales.
- Sort data in categories that are ranked according to a certain characteristic.

Interval Scales

- Rank observations in such a manner that the differences between scale values are equal.
- Values can be added and subtracted meaningfully.

Ratio Scales

- Represent the strongest level of measurement.
- Have all the characteristics of interval scales and have a true zero point as the origin.
- Meaningful ratios can also be computed with ratio scales.

A **frequency distribution** is a tabular illustration of data categorized into a relatively small number of intervals or classes.

- The intervals must be such that each observation must fall into only one interval (mutually exclusive).
- The set of intervals must cover the entire range of values represented in the data (all inclusive).
- The data presented in a frequency distribution may be measured using any type of measurement scale.

Example

The scores of 20 students on a 100-point exam are 80, 15, 26, 51, 90, 95, 45, 70, 65, 58, 76, 54, 10, 27, 33, 76, 81, 54, 56, and 61. Construct a frequency distribution for this data.

Relative Frequency

- The proportion or fraction of total observations that lies in that particular interval.
- Calculated by dividing its absolute frequency by the total number of observations.

Cumulative Absolute Frequency/ Cumulative Frequency

- The number of observations that are less than the upper bound of the interval.
- Sum of the frequencies of all intervals less than and including the said interval.

Cumulative Relative Frequency

- The proportion of total observations that is less than the upper bound of the interval.
- Calculated by adding the relative frequencies of all intervals lower than, and including the said interval.
- Can also be calculated by dividing the cumulative absolute frequency of the interval by the total number of observations.

Interval	Frequency	Absolute Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
$0 < x \leq 25$	2				
$25 < x \leq 50$	4				
$50 < x \leq 75$	8				
$75 < x \leq 100$	6				
Total	20				

Histograms

- Graphically represent the data contained in frequency distribution.
- Intervals are listed on the horizontal axis.
- Frequencies are scaled on the vertical axis.

Frequency Polygons

- Graphically illustrates the data in a frequency distribution.
- Each coordinate or point on the frequency polygon is the frequency of each interval plotted against the midpoint of the interval.

Cumulative Frequency Distribution

- Plots the **cumulative** frequency of each interval against the upper bound of the interval.

