STATISTIC COURCE SUMMERIZATION
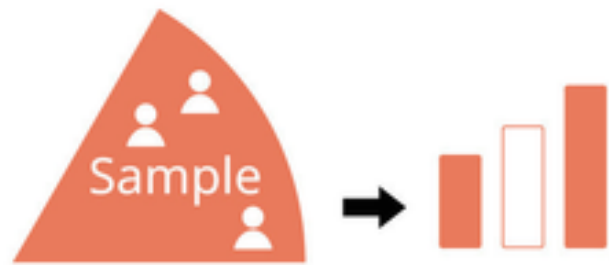
Made by

Rawan Hatem

WHAT ARE STATISTICS & WHAT ARE THEY DIVIDED INTO?
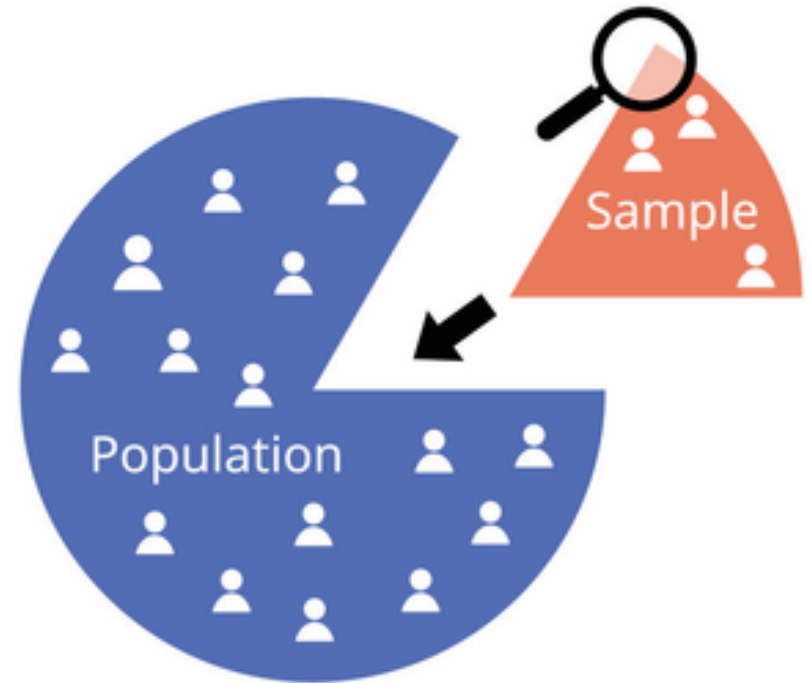
- **Statistics is** the science of collecting, analyzing, interpreting, and presenting data.

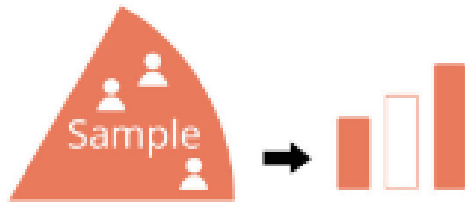- Statistics **Divided into** two categories: **descriptive statistics** and **inferential statistics**.

Sample

Sample

Population

VS

Descriptive
Statistics

Inferential
Statistics

# DISCRIPTIVE STATISTICS

Sample → Descriptive Statistics

- Descriptive statistics involves **organizing** and **summarizing** data using numbers and graphs.
- Data can also be organized using tables, such as frequency tables and frequency distribution tables.
- Descriptive statistics can be used to describe the shape of the data, such as whether it is symmetrical or skewed.
- Descriptive statistics can also be used to describe data using numbers, such as measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation).

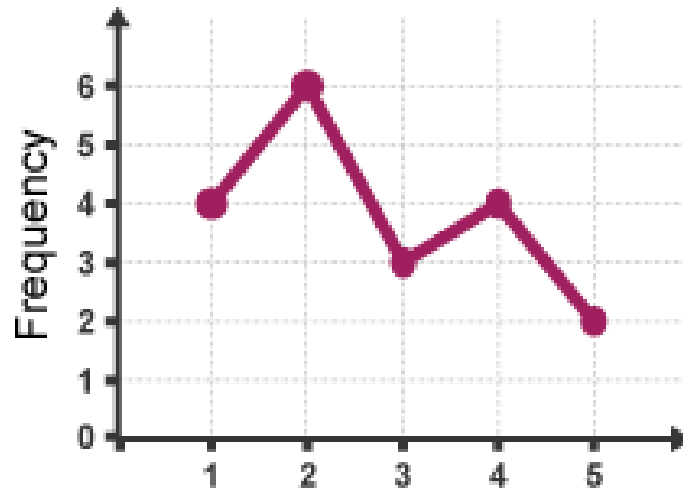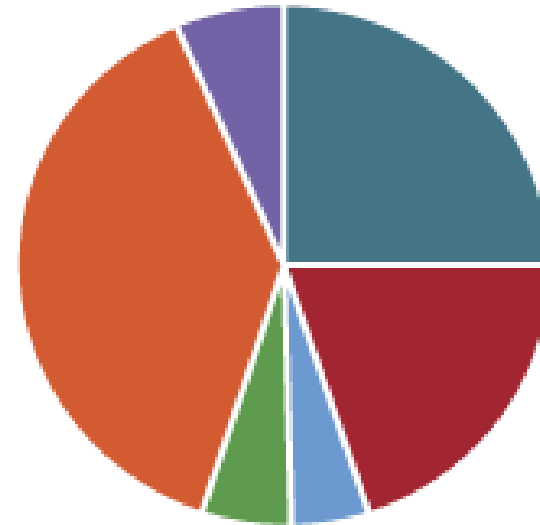# In Descriptive Statistics, Data can be Summarized Using
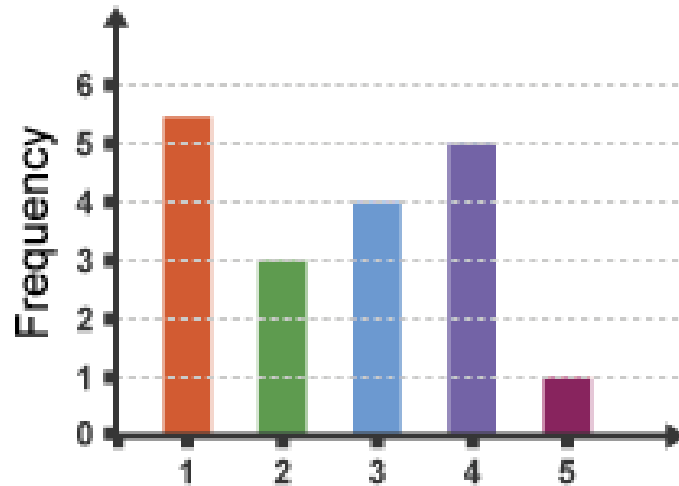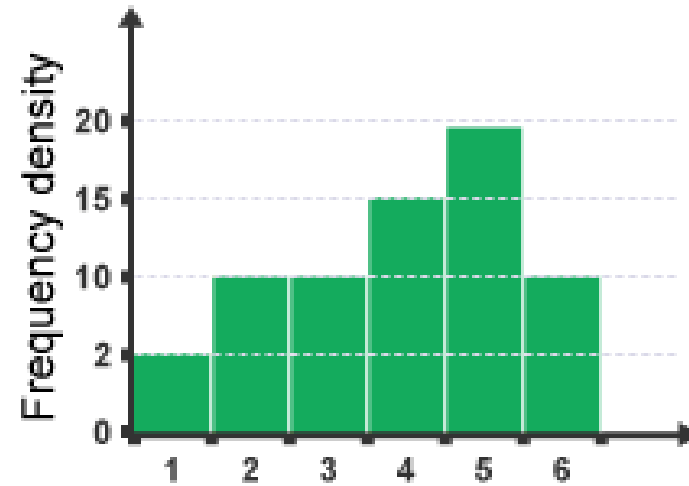
INFERENTIAL STATISTICS

- Inferential statistics involves using sample data to make inferences or draw conclusions about a population.
- Probability can be used to determine how confident we can be that the conclusions we make are correct.
- An example of inferential statistics is determining the percentage of people in a city who prefer blue cars, by taking a sample of the population and making an inference about the entire population.
- Confidence intervals can be used in inferential statistics to express the certainty of our conclusions.
- By increasing the sample size, we can make a better prediction of the population's preferences.

Inferential Statistics
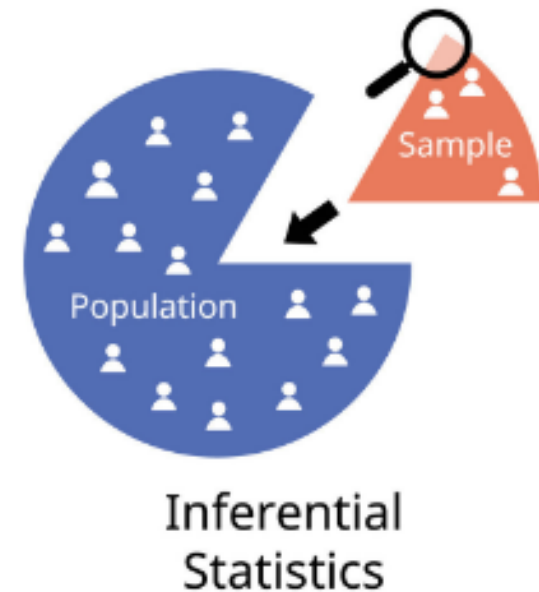
MEAN, MEDIAN, MODE & RANGE

- To calculate the mean, add up all the numbers and divide by the count of numbers.

- To find the median, arrange the numbers in ascending order and find the middle value.

- The mode is the number with the highest frequency.

- The range is the difference between the highest and lowest numbers.

- For a data set with an even number of elements, the median is the average of the two middle values.

Measures Of Tendency :
Mean, Median, Mode In Three Different Graphs
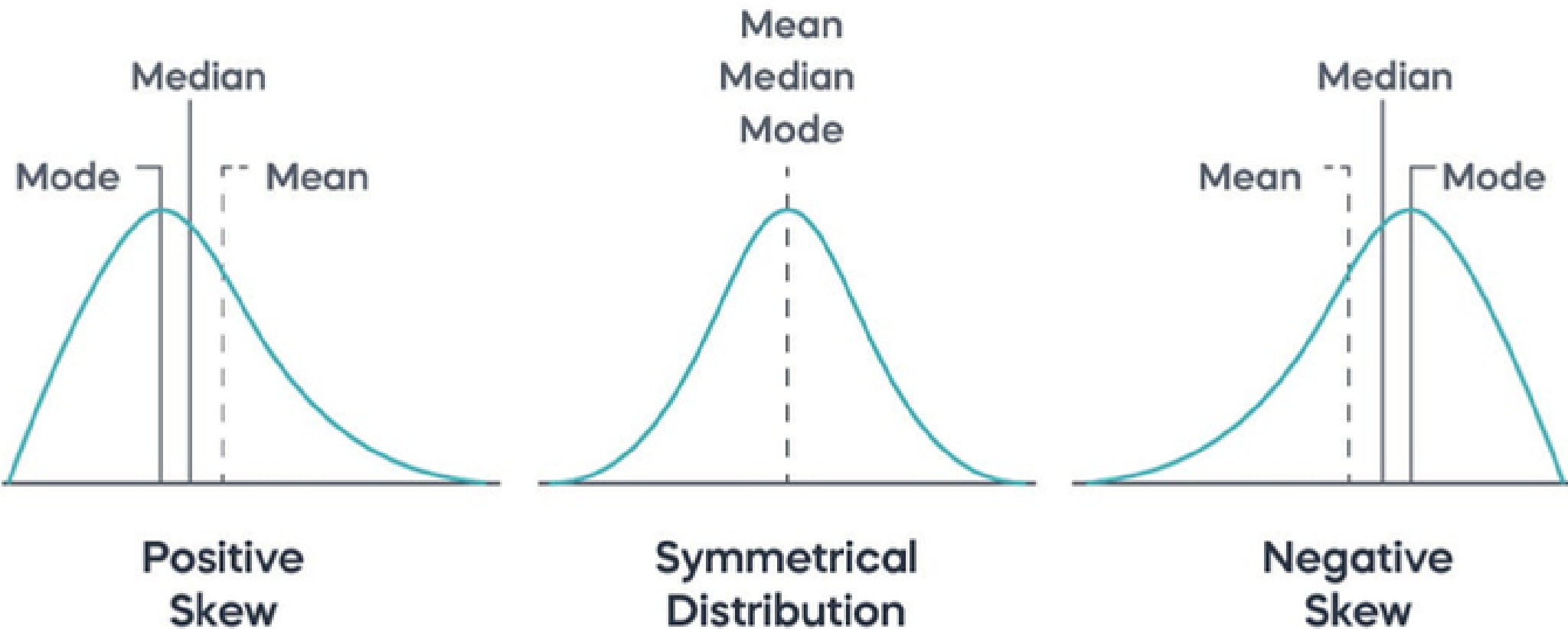
VARIANCE

- Variance is represented as s ^ 2 , where s is the sample standard deviation.
- To calculate variance, find the sum of the squared differences between each data point and the sample mean, then divide by n-1 (n is the size of the sample).
- $\sigma^2 = (\Sigma (x-\mu)^2) / N$
- Variance is a measure of data spread:
- A high variance indicates a larger spread in the data.
- A low variance indicates that the data points are close to the mean.

Population and Sample Variance Formula

Population Variance	Sample Variance
$$\sigma^2 = \frac{\displaystyle\sum_{i=1}^{N}\left(x_i - \mu\right)^2}{N}$$	$$s^2 = \frac{\displaystyle\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$
$\sigma^2 = $ population variance $x_i = $ value of ith element $\mu = $ population mean $N = $ population size	$s^2 = $ sample variance $x_i = $ value of ith element $\bar{x} = $ sample mean $n = $ sample size

STANDARD DEVIATION

- To calculate standard deviation, first find the mean (sum of numbers divided by the size of the sample).

- After finding the mean, use the formula for standard deviation: square the difference of each number and the mean, then divide the sum of those squares by n - 1.

- $\sigma = \sqrt{(\sum(x-\bar{x})\,(x-\bar{x})\,2\,/n)}$

SUMMARIZATION on VARIANCE & STANDARD DEVIATION

Calculation	Formula	Notes
Population Standard Deviation	$\sigma = \sqrt{\dfrac{\Sigma\,(X_i - \mu)^2}{N}}$	μ = population average X = individual values in population N = count of values in population
Sample Variance	$s^2 = \dfrac{\Sigma\,(x_i - \bar{x})^2}{(n - 1)}$	$\bar{X}$ = sample average x = individual values in sample n = count of individual values in sample
Sample Standard Deviation	$s = \sqrt{\dfrac{\Sigma\,(x_i - \bar{x})^2}{(n - 1)}}$	$\bar{X}$ = sample average x = individual values in sample n = count of individual values in sample

INTERQUARTILE RANGE (IQR)

- The minimum value is at the 0th percentile, the median (second quartile) is at the 50th percentile, and the maximum value is at the 100th percentile.

- Q1 is the median of the lower half of the data (25th percentile), and Q3 is the median of the upper half of the data (75th percentile).

- The interquartile range (IQR) represents the middle 50% of the data and is calculated as Q3 - Q1.

- To identify outliers, calculate the range: Q1 - 1.5 * IQR and Q3 + 1.5 * IQR. Any number outside this range is considered an outlier.

IQR REPRESENTATION IN BOX PLOT

Interquartile Range (IQR)

Outliers

Outliers

"Minimum" (Q1 - 1.5*IQR)

Median

Q1 (25th Percentile)

Q3 (75th Percentile)

"Maximum" (Q3 + 1.5*IQR)

-3   -2   -1   0   1   2   3