# **TITLE:**

* Statistics for Machine Learning

# **OVERVIEW:**

# **OBJECTIVES:**

* Learn Statistics
* Learn Different type of statistics
* Descriptive vs Inferential statistics
* Measures of Central Tendency (Mean, Median, Mode)
* Measures of Variability (Range, Variance, Standard Deviation)
* Quantile/IQR
* Covariance – Co-relation
* Probability
* Data Distribution
* Transformation
* Central Limit Theorem
* Confidence Interval
* Hypothesis

# **CHAPTER 1:**

**Introduction to statistics:**

What is statistics, uses in real life, application in machine learning, example

# **CHAPTER 2:**

**Different type of statistics:**

Descriptive statistics vs Inferential statistics, topics of statistics

# **CHAPTER 3:**

**Descriptive Statistics:**

# **CHAPTER 4:**

**Inferential Statistics:**

# **CHAPTER 5:**

**Conclusion:**

# **TEXT BOOKS:**

# **REFERENCES:**

**CHAPTER 1:**

**Introduction to Statistics**

**Introduction:**

Statistics is the science of collecting, organizing, analyzing, interpreting, and presenting data. It encompasses a wide range of techniques for summarizing data, making inferences, and drawing conclusions.

Statistical methods help quantify uncertainty and variability in data, allowing researchers and analysts to make data-driven decisions with confidence.

**Applications of Statistics in Machine Learning:**

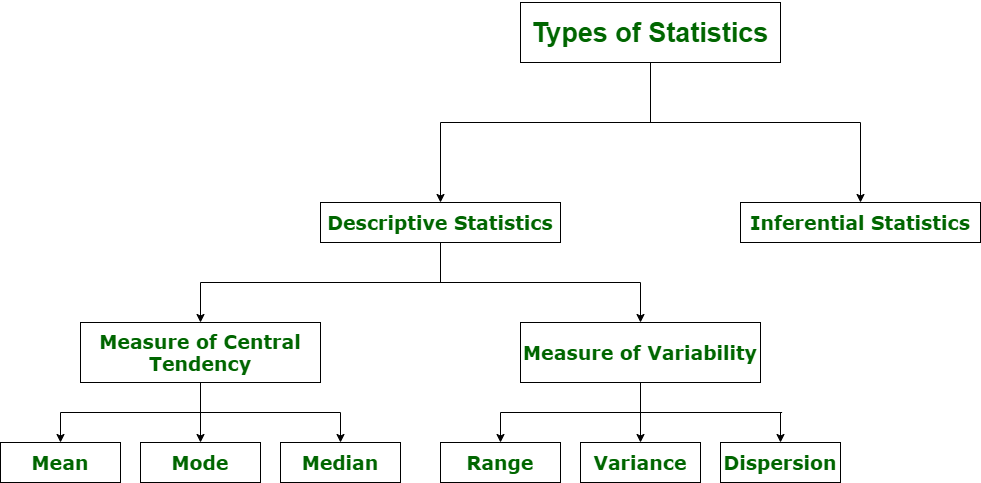
Statistics is a key component of machine learning, with broad applicability in various fields.

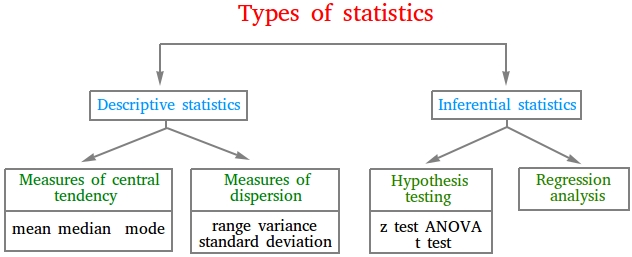
* Feature engineering relies heavily on statistics to convert geometric features into meaningful predictors for[machine learning algorithms.](https://www.geeksforgeeks.org/machine-learning-algorithms/)
* In image processing tasks like object recognition and segmentation, statistics accurately reflect the shape and structure of objects in images.
* Anomaly detection and quality control benefit from statistics by identifying deviations from norms, aiding in the detection of defects in manufacturing processes.
* Environmental observation and geospatial mapping leverage statistical analysis to monitor land cover patterns and ecological trends effectively.

Overall, statistics plays a crucial role in machine learning, driving insights and advancements across diverse industries and applications.

# **CHAPTER 2:**

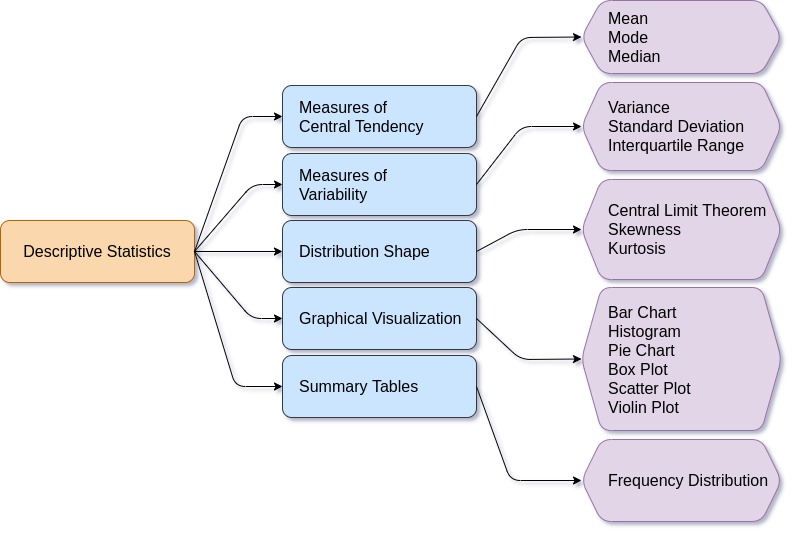
**Different type of statistics:**





# **CHAPTER 3:**

**Descriptive Statistics**

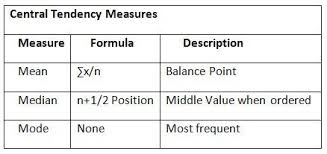


**Descriptive Statistics:**

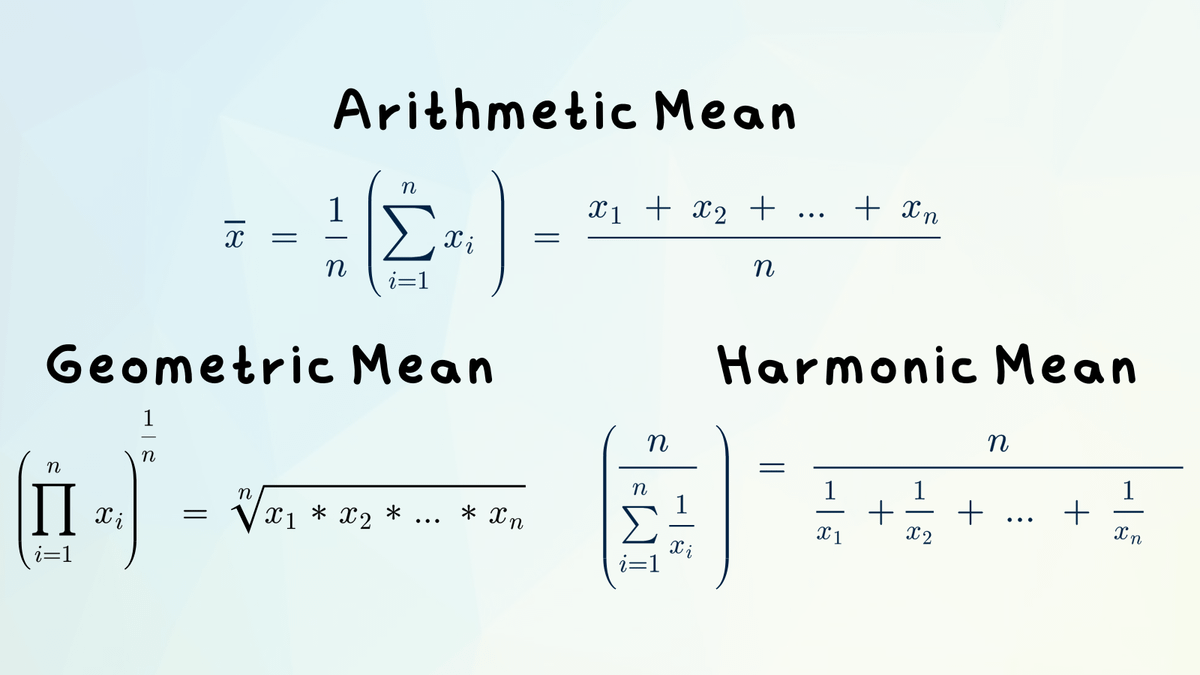
1. Measures of Central Tendency
   1. Mean, Median, Mode
2. Measures of Variability
   1. Range, Variance, Standard Deviation, IQR

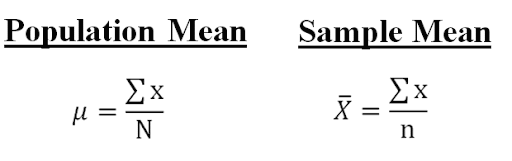
**Central Tendency:**

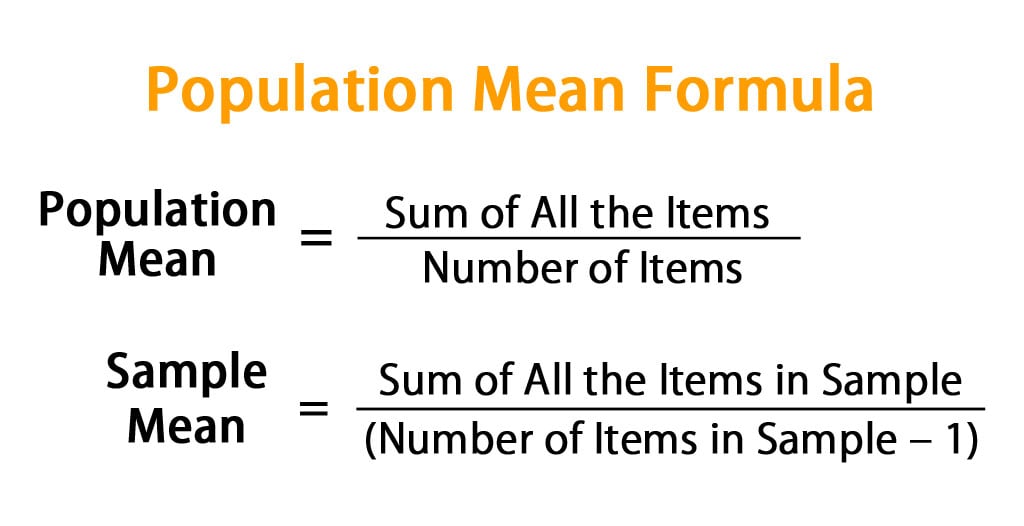
In [statistics](https://en.wikipedia.org/wiki/Statistics), a central tendency (or measure of central tendency) is a central or typical value for a [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution).



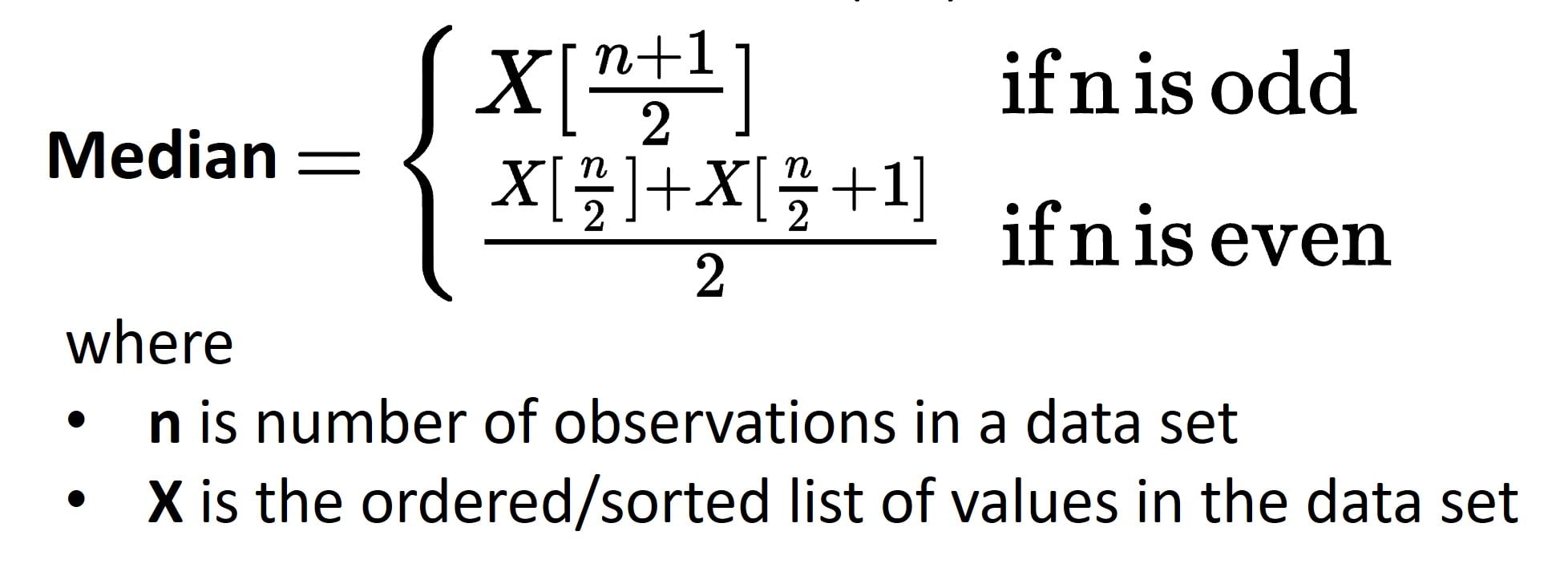
**Mean:**



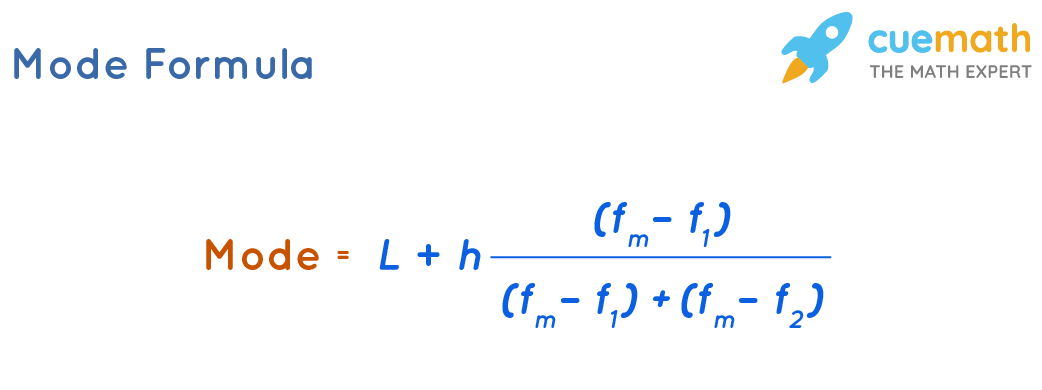




**Median:**

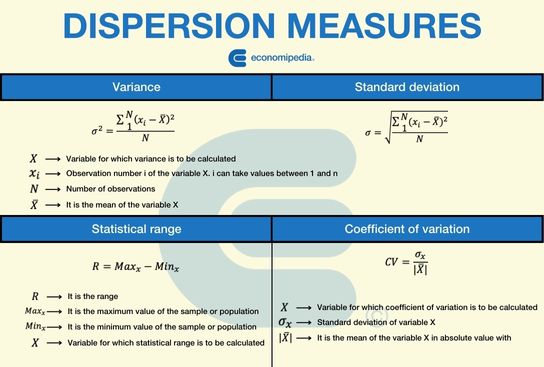


**Mode:**

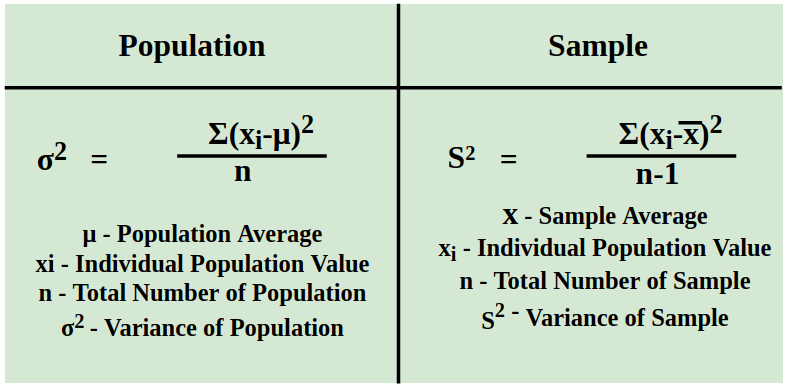


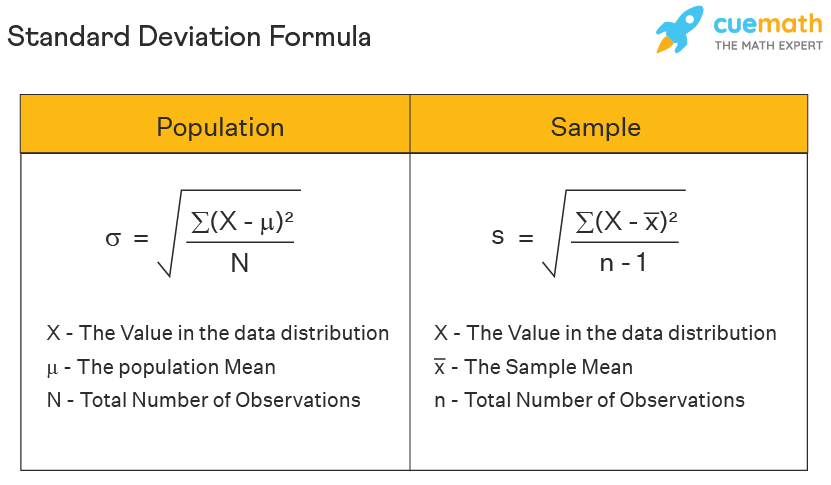
**Measures of Dispersions:** It measures the Spread of Data from Mean

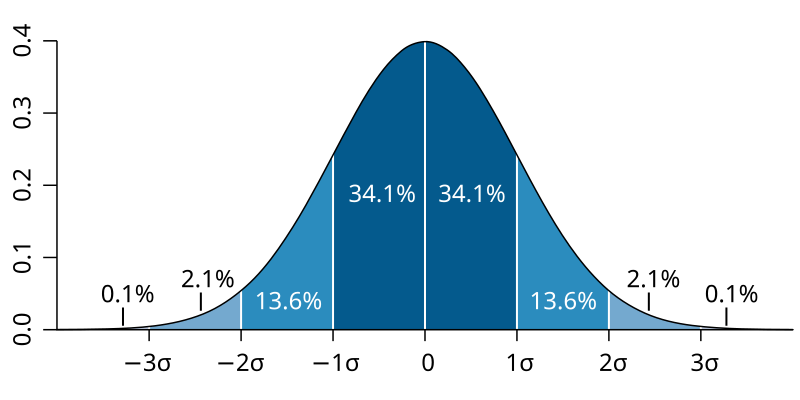
1. Range 🡪 Max-Min
2. Variance 🡪 Spread from mean
3. Standard Deviation 🡪 Square root of Variance



**Variance:**



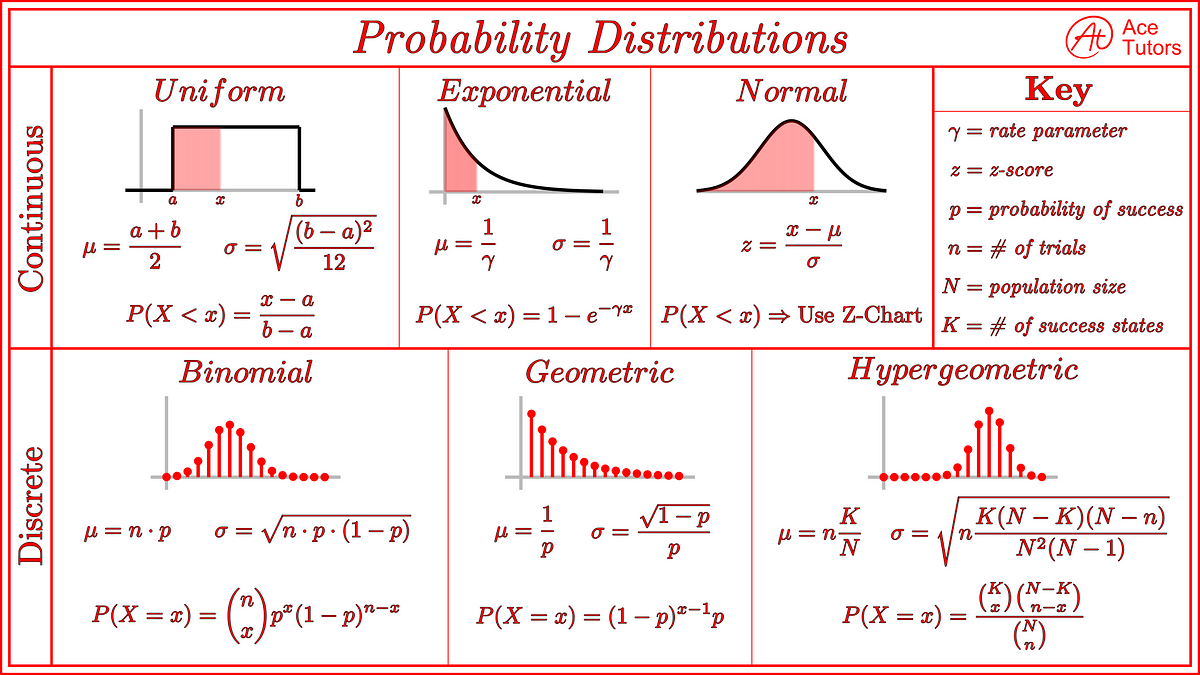
**Standard Deviation:**



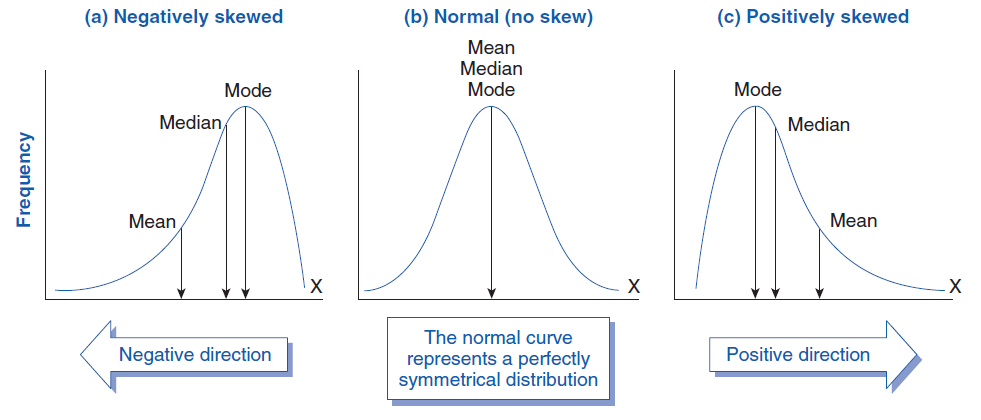
**68-95-99.7 Rule**

**Data Distribution:**

1. Normal Distribution
2. Binomial Distribution
3. Poisson Distribution
4. Exponential Distribution
5. Uniform Distribution
6. Student’s Distribution



**Skewed Distribution:**



# **CHAPTER 4:**

**Inferential Statistics**

# **CHAPTER 5:**

**Conclusion**