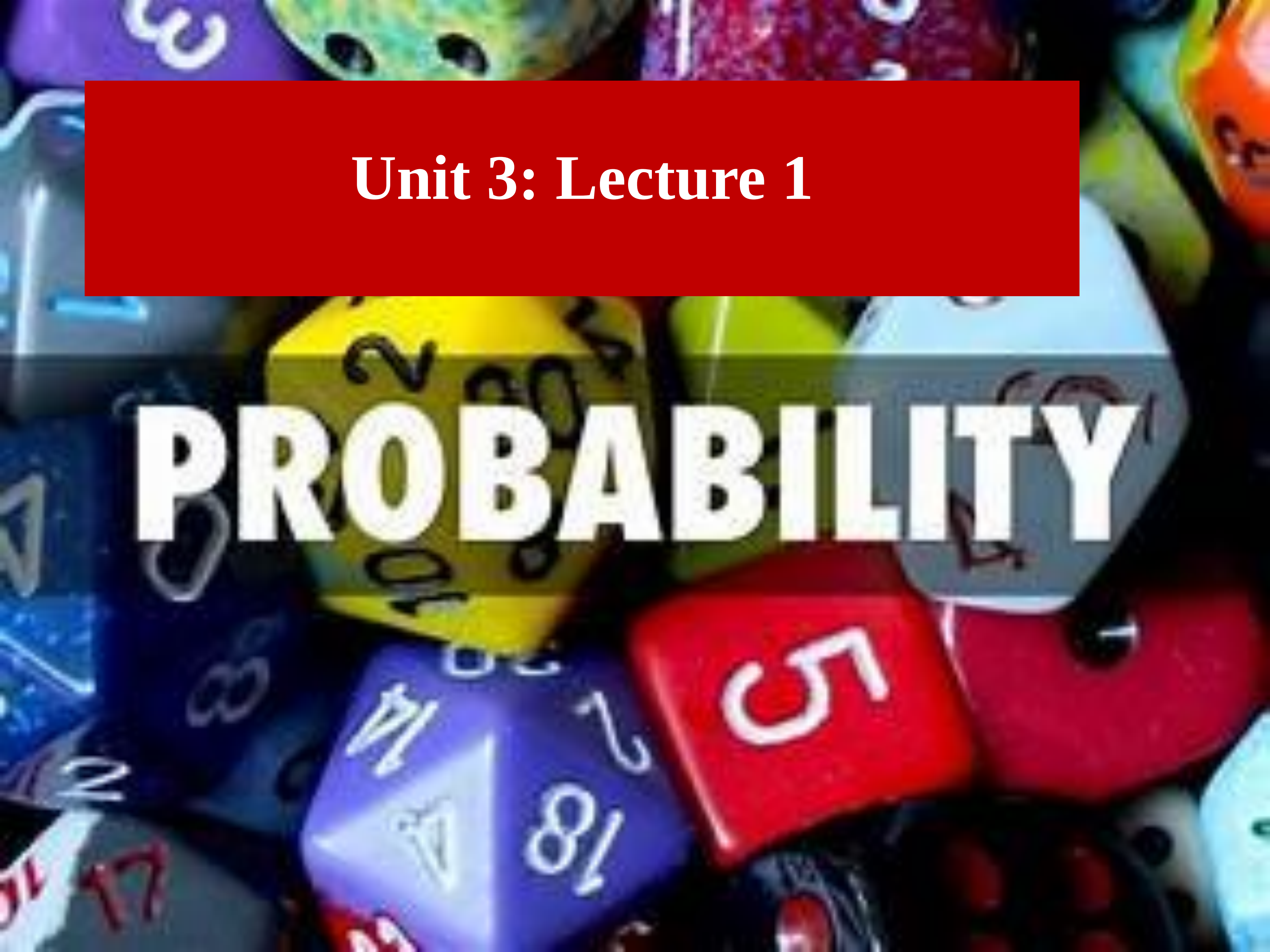
Unit 3: Lecture 1

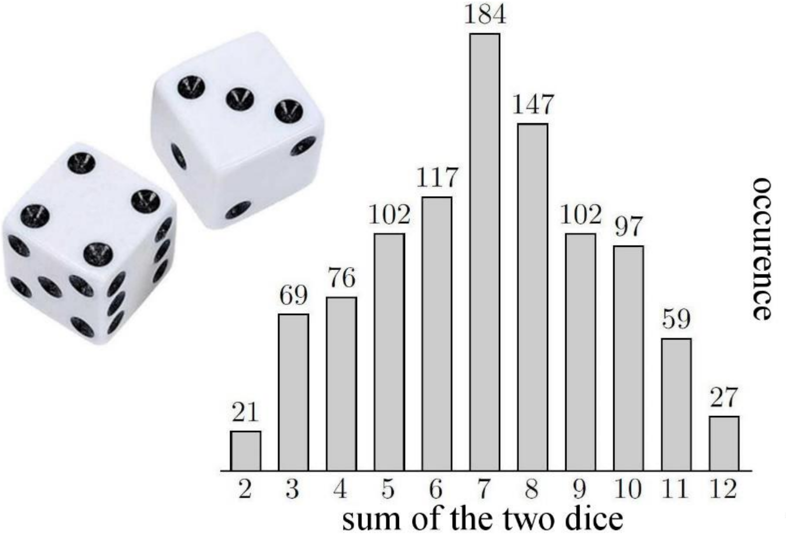


Probability

- ***Probability theory*** developed from the study of games of chance like dice and cards. Probability theory is the foundation for ***statistical inference***.
- A process like flipping a coin, rolling a die or drawing a card from a deck is called a ***probability experiment***.
- An ***outcome*** is a specific result of a single trial of a probability experiment.

Probability distributions

- A ***probability distribution*** is a way of presentation or a graph to show the values that a random variable may have.



A **random variable**

- Typically denoted as X, is a variable whose possible values are outcomes of a random process.

- There are two categories of random variables. These are:
  - ***discrete random variables***,

  And
  - ***continuous random variables***

Discrete Random Variables

Is a variable which can take on only a countable number of distinct values like 0, 1, 2, 3, 4, 5…100, 1 million, etc. Some examples of discrete random variables include:

- The number of times a coin lands on tails after being flipped 20 times.
- The number of times a dice lands on the number *4* after being rolled 100 times.

Discrete Probability Distributions

- Binomial distribution – the random variable can only assume 1 of 2 possible outcomes. There are a fixed number of trials, and the results of the trials are independent. i.e. flipping a coin and counting the number of heads in 10 trials.
- Poisson Distribution – Poisson probability distribution is used in situations where events occur randomly and independently a number of times on average during an interval of time or space. The random variable X associated with a Poisson process is discrete and therefore the Poisson distribution is discrete.
- Hospital emergencies receive on average 5 very serious cases every 24 hours.
- I receive on average 10 e-mails every 2 hours.
- Customers make on average 10 calls every hour to the customer help center

Discrete Random Variable

- A discrete random variable X has a finite number of possible values. The probability distribution of X lists the values and their probabilities.

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities must be 1.

- Find the probabilities of any event by adding the probabilities of the particular values that make up the event.

Example

- The instructor in a large class gives 15% each of A's and D's, 30% each of B's and C's and 10% F's. The student's grade on a 4-point scale is a random variable X (A=4).

Grade	F=0	D=1	C=2	B=3	A=4
Probability	0.10	.15	.30	.30	.15

- What is the probability that a student selected at random will have a B or better?
- ANSWER: P (grade of 3 or 4)=P(X=3) + P(X=4)

= 0.3 + 0.15 = 0.45

Example 2

What will be the probability of getting both heads or both tails when 2 coins are tossed?

- Out of 4 outcomes, viz., HH, HT, TH, TT, which are mutually exclusive, number of favorable cases of HH is ¼ (4 events) and that of TT is ¼.
- Therefore, probability of getting both heads or both tails will be
- ¼ + ¼ = ½

P**robability distribution** for a discrete random variable

- What is the probability that a dropped dice will be either a one or a six ? Given that all outcomes are equally likely, we can compute the probability of a one or a six using the formula:

Probability =Number of favorable outcomes/Number of possible equally-likely outcomes

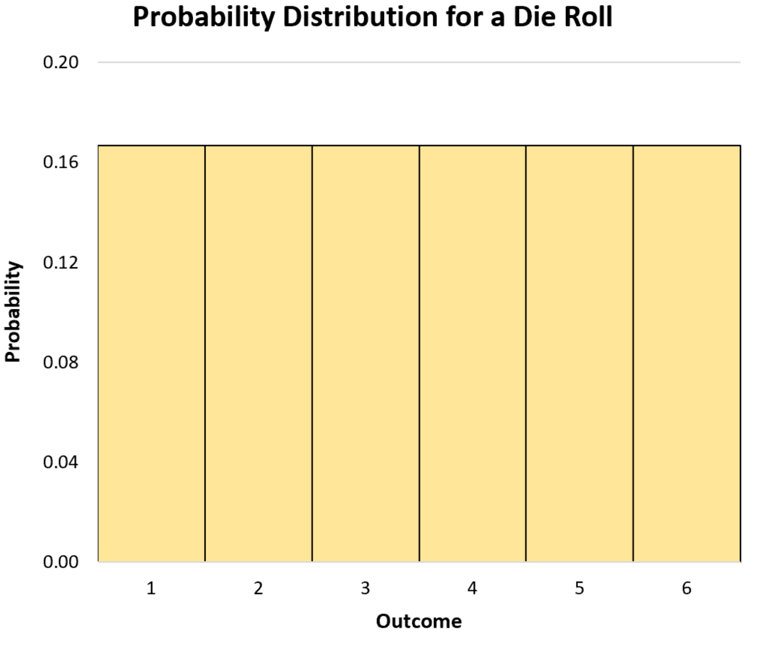
- This will tell us the probability that the random variable will take on certain values.

P**robability distribution** for a discrete random variable

- For example, suppose we roll a fair dice one time. If we let X denote the probability that the dice lands on a certain number, then the probability distribution can be written as:
- **P(X=1):** 1/6
- **P(X=2):** 1/6
- **P(X=3):** 1/6
- **P(X=4):** 1/6
- **P(X=5):** 1/6
- **P(X=6):** 1/6

For discrete probability distribution to be valid, it must satisfy the following two criteria:

- The probability for each outcome must be between 0 and 1.
- The sum of all of the probabilities must add up to 1.
- Notice that the probability distribution for the dice roll satisfies both of these criteria:



Continuous Random Variables

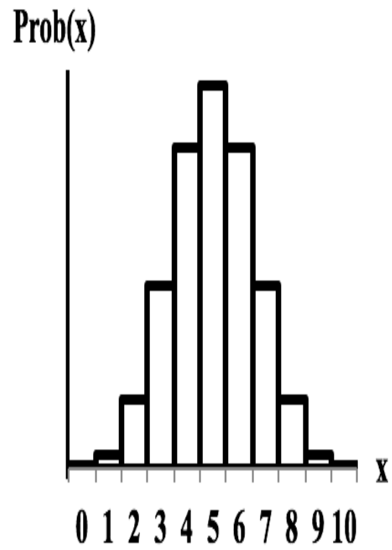
- A **continuous random variable** is one which can take on an infinite number of possible values. Some examples of continuous random variables include:
- Height of a person
- Weight of an animal
- For example, the height of a person could be 60.2 inches, 65.2344 inches, 70.431222 inches, etc. There are an infinite amount of possible values for height.

Rule of Thumb:

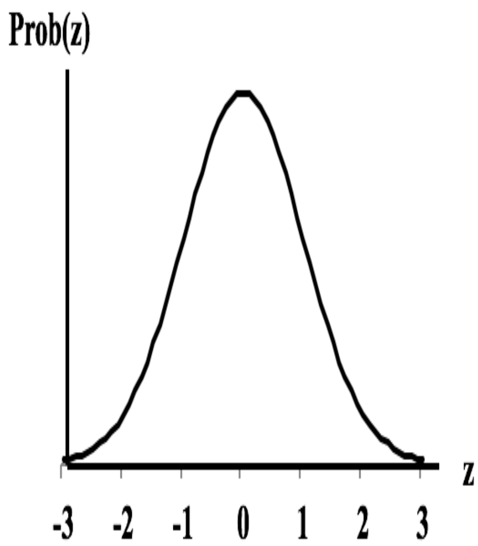
- If you can *count* the number of outcomes, then you are working with a discrete random variable (e.g. counting the number of times a coin lands on heads). But if you can *measure* the outcome, you are working with a continuous random variable (e.g. measuring, height, weight, time, etc.)

Probability Distributions

- A discrete variable follows a Binomial or a Poisson distribution. the variable is restricted to taking on integer values only.
- A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space.
- Between two values of a continuous random variable, we can always find a third.

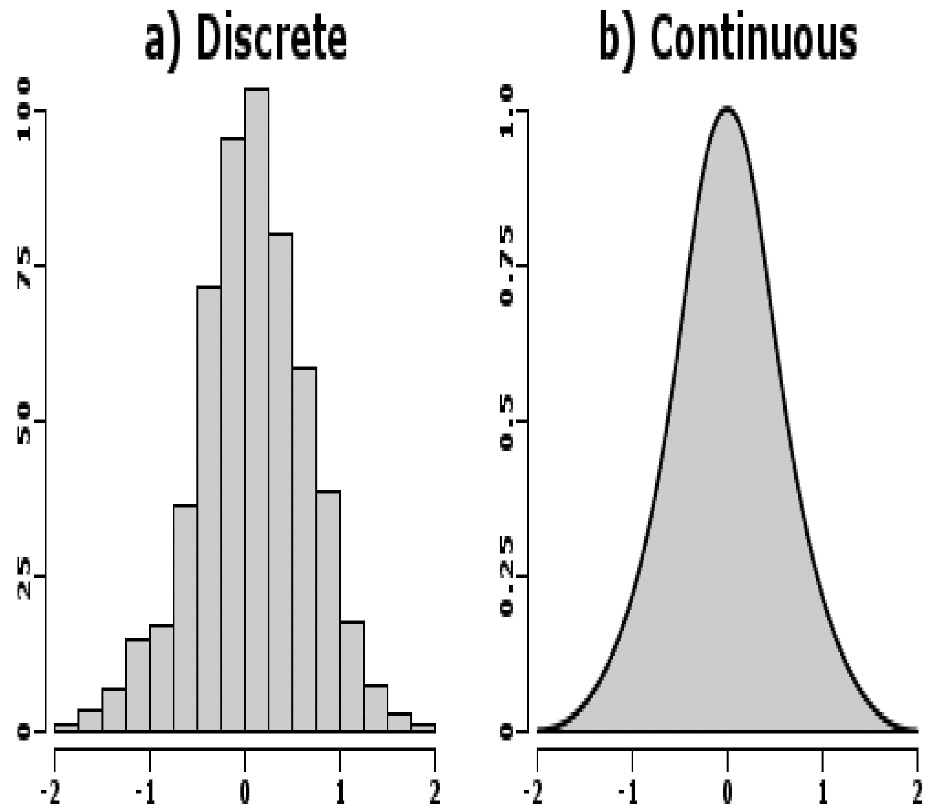


Binomial Distribution
Discrete Data & Discrete
Probability Curve



Standard Normal Distribution
Continuous Data and Continuous
Probability Curve

- A histogram is used to represent a discrete probability distribution.
- Smooth curve called the *probability density* is used to represent a continuous probability distribution.

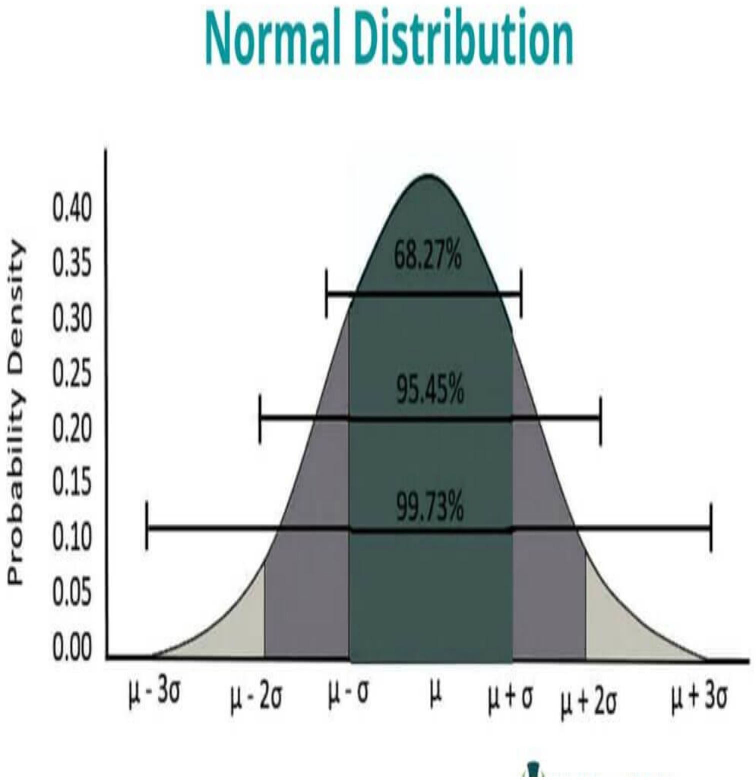


Continuous Variable

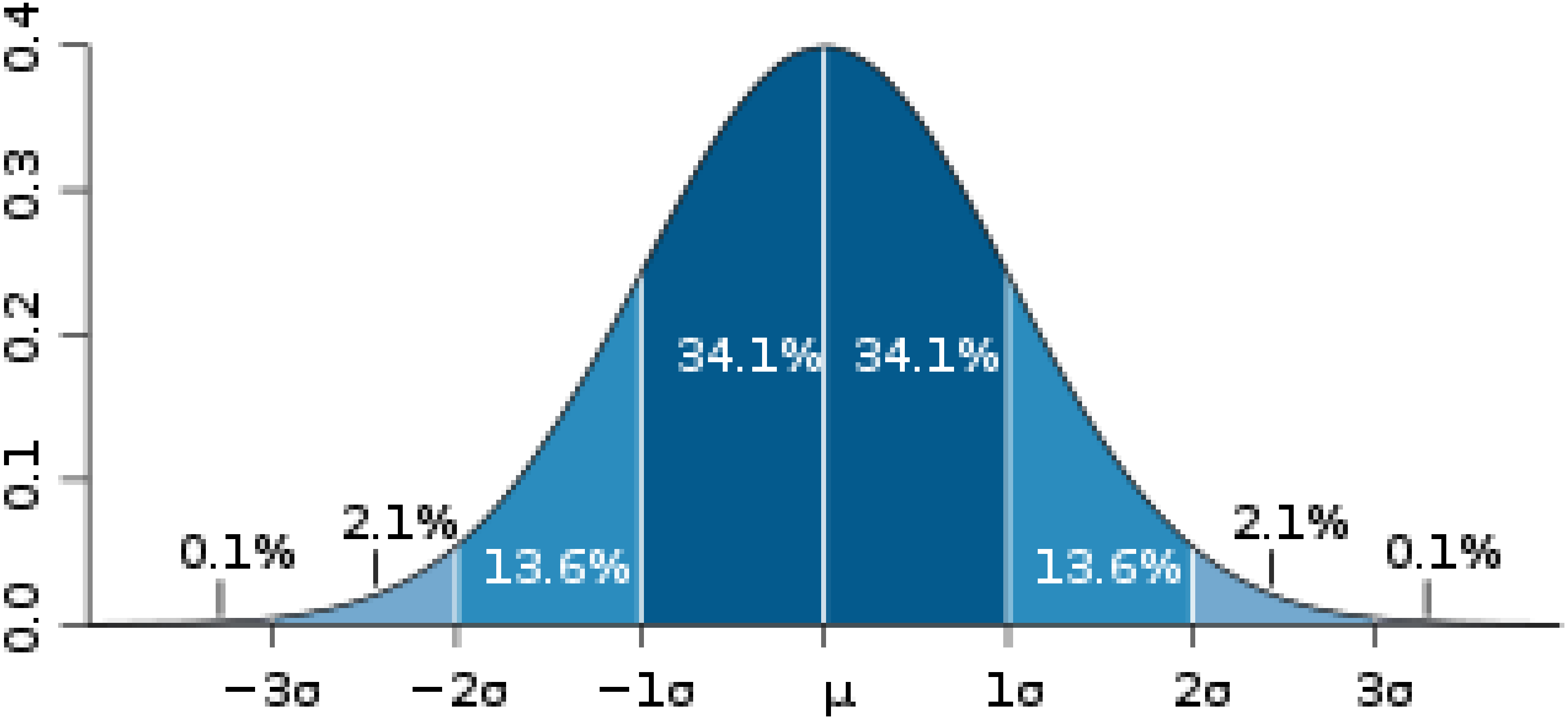
- A ***continuous probability distribution*** is a ***probability density function***.
- The area under the smooth curve is equal to 1
- The frequency of occurrence of values between any two points equals the total area under the curve between the two points and the x-axis.

Normal Distribution

- Also called bell shaped curve, normal curve, or Gaussian distribution.
- Given its name by the French mathematician Quetelet who, in the early 19th century noted that many human attributes, e.g. height, weight, intelligence appeared to be distributed normally.

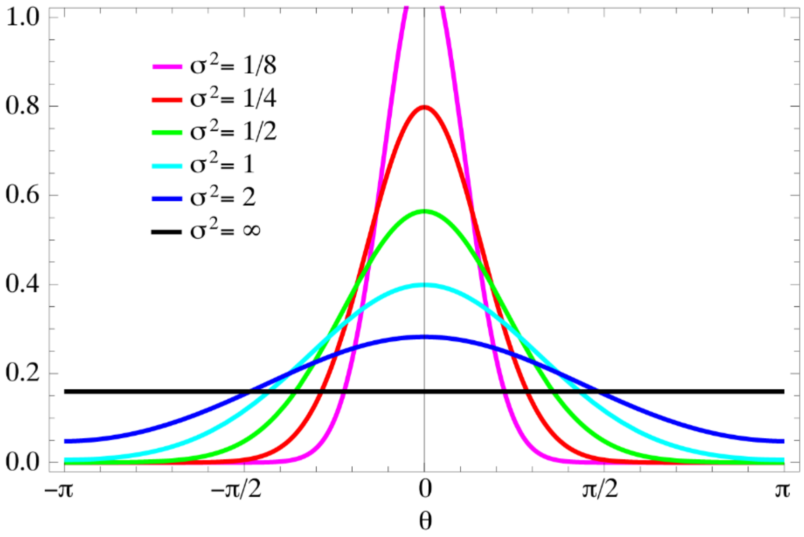


The normal distribution



Normal Distribution

- The normal curve is unimodal and symmetric about its mean (μ). It is not too peaked or flat.
- In this distribution the mean, median and mode are all identical.
- The standard deviation (σ) specifies the amount of dispersion around the mean.
- The two parameters μ and σ completely define a normal curve.



Properties of a Normal Distribution

- Also called a Probability density function.
- The probability is = to the area under the curve.
- It is symmetrical about m . The mean, median and mode are all equal.
- The total area under the whole curve = 1
- Therefore 50% is to the right of m and 50% is to the left of m.
- Perpendiculars of:
  - ±1 s contain about 68%;
  - ±2 s contain about 95%;
  - ±3 s contain about 99.7%

  of the area under the curve.

Importance of Normal Distribution to Statistics

- Most of the variables in natural and social sciences are approximately normally or approximately normally distributed. Height, birth weight, reading ability, job satisfaction, or SAT scores are just a few examples of such variables.
- Because normally distributed variables are so common, many statistical tests are designed for normally distributed populations.
- Many inferential statistics assume that the populations are distributed normally. Thus, understanding the properties of normal distributions means you can use inferential statistics to compare different groups and make estimates about populations using samples.

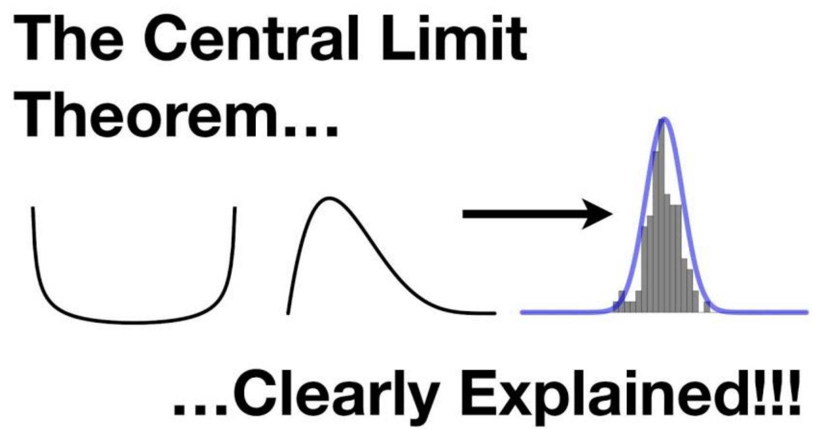
Normal Distribution

Q Is every variable normally distributed?

A Absolutely not

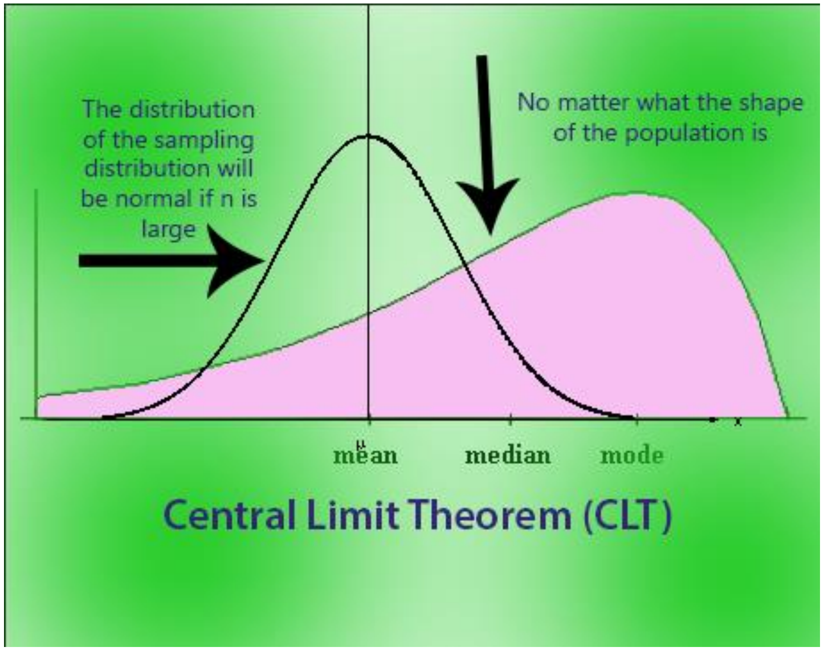
Q Then why do we spend so much time studying the normal distribution?

A Some variables are normally distributed; a bigger reason is the “Central Limit Theorem”!!!!!!!!!!!!!!!!
!!!!!!!!!!!???????????



Central Limit Theorem

- Describes the characteristics of the "**population of the means"** which has been created from the means of an infinite number of random population samples of size (N), all of them drawn from a given "**parent population**".



Central Limit Theorem (CLT)

[ˈsen-trəl ˈli-mət ˈthē-ə-rəm]

The principle that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.

Central Limit Theorem

- It predicts that regardless of the distribution of the parent population:
  - The **mean** of the samples means is always equal to the mean of the parent population from which the population samples were drawn.
  - The distribution of means will increasingly approximate a **normal distribution** as the size N of samples increases.

Why is the Central Limit Theorem important in statistics?

Because it allows to make inferences about population parameters using sample statistics. It underpins hypothesis testing by justifying the normality assumption of large sample statistics.

Can the Central Limit Theorem be applied to any distribution?

Yes, one of the remarkable aspects of the CLT is that it does not require the population to be normally distributed. It applies to any distribution with a finite mean and variance, making it a powerful tool in statistical analysis and data science.

What are the conditions necessary for the Central Limit Theorem to hold? The theorem requires that the samples are independent and identically distributed, the sample size is sufficiently large (usually $n \geq 30$), and the population from which samples are drawn has a finite variance. These conditions ensure the sampling distribution of the mean approximates normality.

Defining “Sufficiently Large”

- Central limit theorem states that the sampling distribution of a sample mean is approximately normal if the sample size is **“large enough”**, even if the population distribution is not normal.
- If the population distribution is symmetric, sometimes a sample size as small as 15 is sufficient.
- If the population distribution is skewed, generally a sample size of at least 30 is needed.
- If the population distribution is extremely skewed, then a sample size of 40 or higher may be necessary.

Example

A population has a mean of 12 and a standard deviation of 3. Find the mean and standard deviation if a sample of 36 is drawn from the distribution.

Given: $\mu = 12$, $\sigma = 3$, $n = 36$

As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}}$ μx = $\mu = 12$

$\mathbf{s^2 = \sigma^2 / n}$

9/36 =1/4 =

S= 0.5