# Task 3

## 1- What is the normal distribution?

**The normal distribution, also known as the Gaussian distribution or bell curve, is a probability distribution that describes how the values of a variable are distributed. In a normal distribution:**

### 1- Symmetry:

The distribution is perfectly symmetrical around its mean, which means the left side is a mirror image of the right side.

### 2- Mean, Median, and Mode:

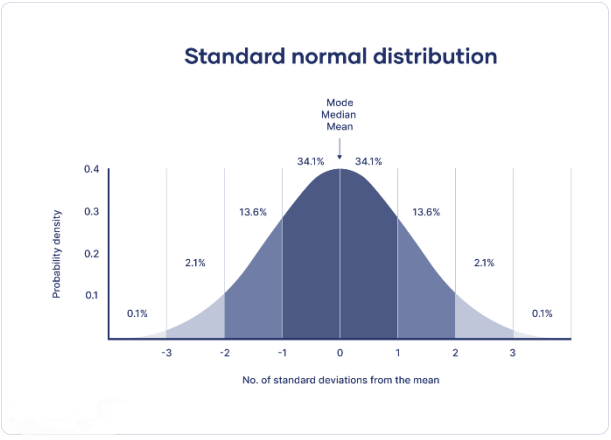
All three of these central measures are equal and located at the center of the distribution.

### 3- Bell-Shaped Curve:

The curve has a single peak, which occurs at the mean, and the tails of the curve approach, but never touch, the horizontal axis.

### 4- Empirical Rule:

About 68% of the data falls within one standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations.



## 2- The types of distribusion:

### 1- Discrete Distributions

**Bernoulli Distribution:** Describes a random variable that has only two possible outcomes (success or failure) with a single trial.

**Binomial Distribution:** Represents the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success.

**Poisson Distribution:** Models the number of times an event occurs within a fixed interval of time or space, assuming the events happen independently and at a constant average rate.

**Geometric Distribution:** Describes the number of trials needed to get the first success in a series of independent Bernoulli trials.

**Negative Binomial Distribution:** Counting the number of trials until a fixed number of successes occur.

**Hypergeometric Distribution:** Models the probability of a certain number of successes in a sample drawn without replacement from a finite population.

## 2- Continuous Distributions

**Normal Distribution:** Also known as the Gaussian distribution, it's the most common continuous distribution, characterized by its bell-shaped curve.

**Uniform Distribution:** Describes a situation where all outcomes in a range are equally likely. It has a constant probability density over its range.

**Exponential Distribution:** Models the time between events in a Poisson process. It's commonly used to describe waiting times.

**Gamma Distribution:** A generalization of the exponential distribution that models the time until an event occurs a certain number of times.

**Beta Distribution:** Used to model variables that are constrained within a range, typically [0, 1], like probabilities.

**Chi-Square Distribution:** Arises when a normal variable is squared. It's commonly used in hypothesis testing and constructing confidence intervals.

**Log-Normal Distribution:** Occurs when the logarithm of a variable is normally distributed. It’s used to model data that are positively skewed.

# 3- How to convert any distribution to the normal one?

## 1- Z-Score Normalization (Standardization)

**When to use:** When the data is already approximately normally distributed but has different means and variances.

**How it works:** This method converts your data to a standard normal distribution (mean = 0, standard deviation = 1).

**Formula:** $Z=(X-\mu)/\sigma$

- $X$ is the data point
- $\mu$ is the mean
- $\sigma$ is the standard deviation

## 2- Box-Cox Transformation

**When to use:** When the data is positive and you want to transform it to be more normally distributed.

**How it works:** This method finds an optimal power transformation (lambda) that stabilizes variance and makes the data more normal.

**Formula:**

$$y(\lambda)=\begin{cases}(y^{\lambda}-1)/\lambda & \text{if } \lambda \neq 0\\ \ln(y) & \text{if } \lambda=0\end{cases}$$

- $y$ is the original data and $\lambda$ is the transformation parameter

## 3- Log Transformation

**When to use:** When the data is positively skewed (long right tail).

**How it works:** This method compresses the right tail, making the distribution more symmetric.

**Formula:** $Y=\log(X)$

- $X$ is the original data

## 4- Square Root Transformation

**When to use:** When dealing with count data or when the data has a moderate right skew.

**How it works:** This method reduces the skewness of the distribution.

**Formula:** $Y=\sqrt{X}$

- $X$ is the original data

## 5- Reciprocal Transformation

**When to use:** When the data has a strong positive skew.

**How it works:** This method involves taking the reciprocal of each data point, which can significantly reduce skewness.

**Formula:** $Y=1 / X$

- $X$ is the original data

## 6- Rank-Based Transformation

**When to use:** When the data is not normally distributed and other methods fail.

**How it works:** This method replaces data points with their ranks and transforms them into a normal distribution using inverse normal distribution functions.

**Steps:** 1. Rank the data.

2. Map the ranks to the corresponding quantiles of a normal distribution.

### 7- Quantile Transformation

**When to use:** For any arbitrary distribution.

**How it works:** This method maps the data to a uniform distribution first and then applies an inverse cumulative distribution function (CDF) of a normal distribution.

**Steps:** 1. Compute the quantiles of the data.

2. Apply the inverse CDF of the normal distribution to these quantiles.

## 4- what is the difference between loc and iloc in pandas?

**loc:** Label-based indexing; uses row/column labels to access data.

Example: df.loc['row_label', 'col_label']

**iloc:** Integer-based indexing; uses row/column positions to access data.

Example: df.iloc[0, 1]

**Key Difference:**

loc selects by labels , iloc selects by positions.