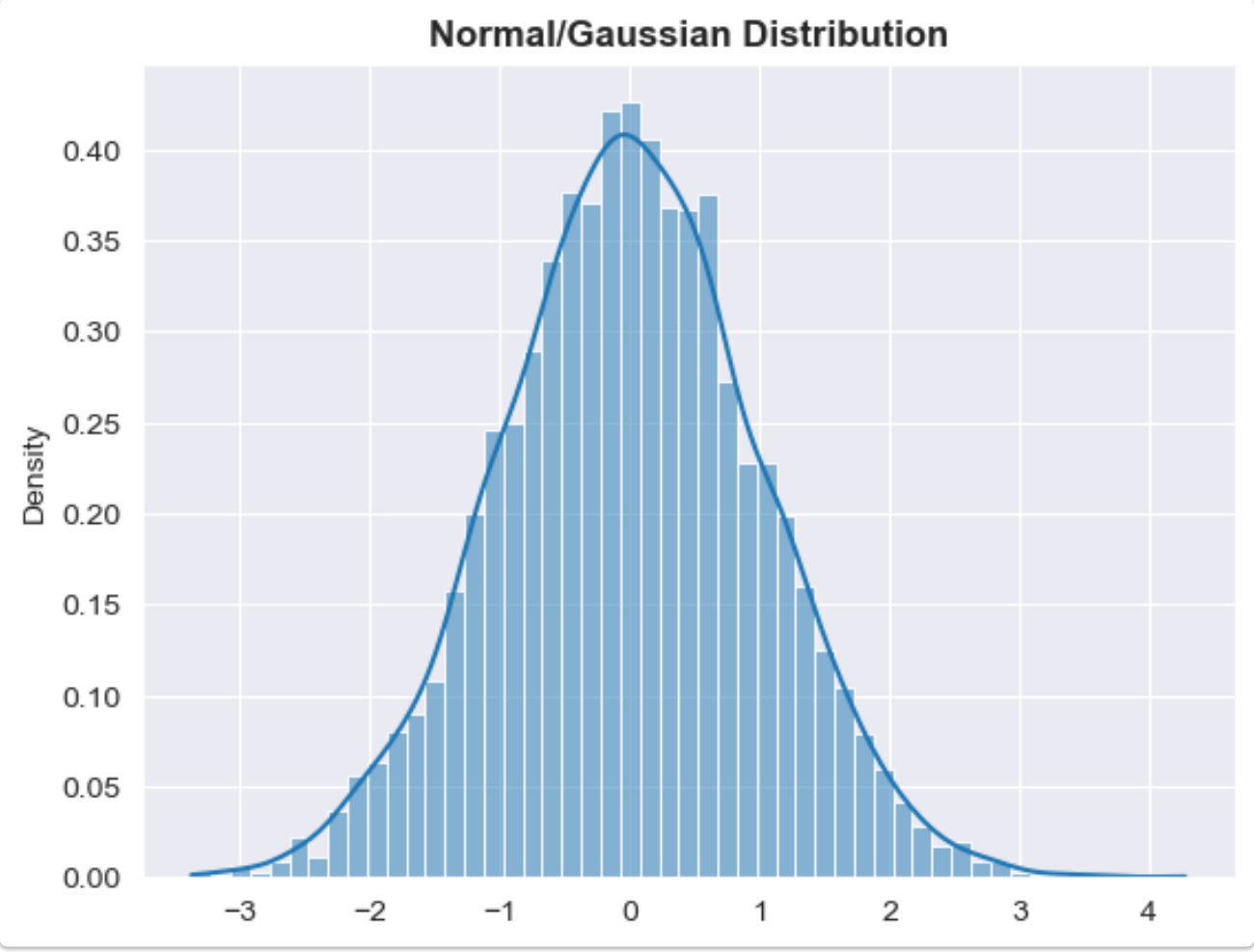
Ep 9 - Gaussian/Normal Distribution

Dec 4,2023

Gaussian/Normal Distribution

Normal distribution, also known as Gaussian distribution, is a probability distribution that is commonly used in statistical analysis. It is a *continuous probability distribution* that is **symmetrical around the mean**, with a bell-shaped curve.



A normal distribution has **2 tails** and is **asymptotic** in nature. It indicates that a *large number of values are present near the mean*, and *very few at the far ends* i.e. the tails.

$$\text{Gaussian Distribution PDF: } f(x|\mu,\sigma) = N(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The normal distribution is characterized by two parameters, namely, **mean(μ)** and **standard deviation(σ)**. The *mean represents the position/centre* of the distribution whereas the *standard deviation represents the spread* of the distribution.

Why is it of such Importance?

Commonality in Nature: Many natural phenomena follow a normal distribution, such as the heights of people, the weights of objects, the IQ scores of a population, and many more. Thus, the normal distribution provides a convenient way to model and analyse such data.

Standard Normal Variate(Z)

It is a standardized form of the normal distribution with $\mu = 0$ and $\sigma = 1$.
Standardizing a normal distribution allows us to compare different distributions with each other, and to calculate probabilities using standardized tables or software.

$$\text{Standard Normal Variate: } N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

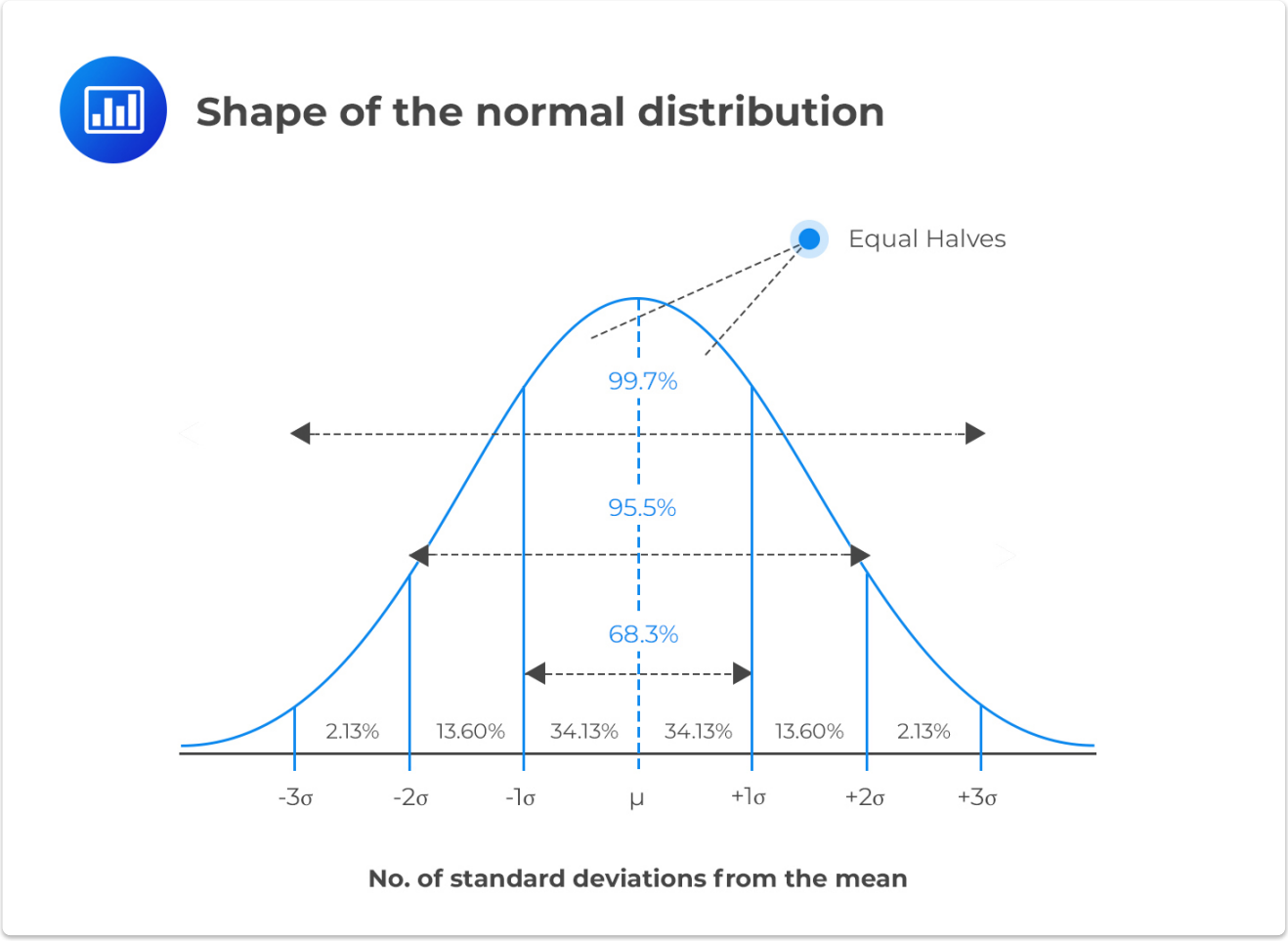
Z Table

A z-table tells you the area underneath a normal distribution curve to the left of the z-score.
Z Table

Properties of Normal Distribution

1. **Symmetricity**: The normal distribution is symmetric about its mean, which means that the *probability of observing a value above the mean is the same as the probability of observing a value below the mean*. The bell-shaped curve of the normal distribution reflects this symmetry.
2. **Measures of Central Tendencies are Equal**: $\text{mean} = \text{median} = \text{mode}$
3. **Area Under the Bell/Curve**: $\text{Area under the curve} = 1$
4. **Empirical Rule**: The $68 - 95 - 99.7$ rule.

Empirical Rule



Applications of Normal Distributions

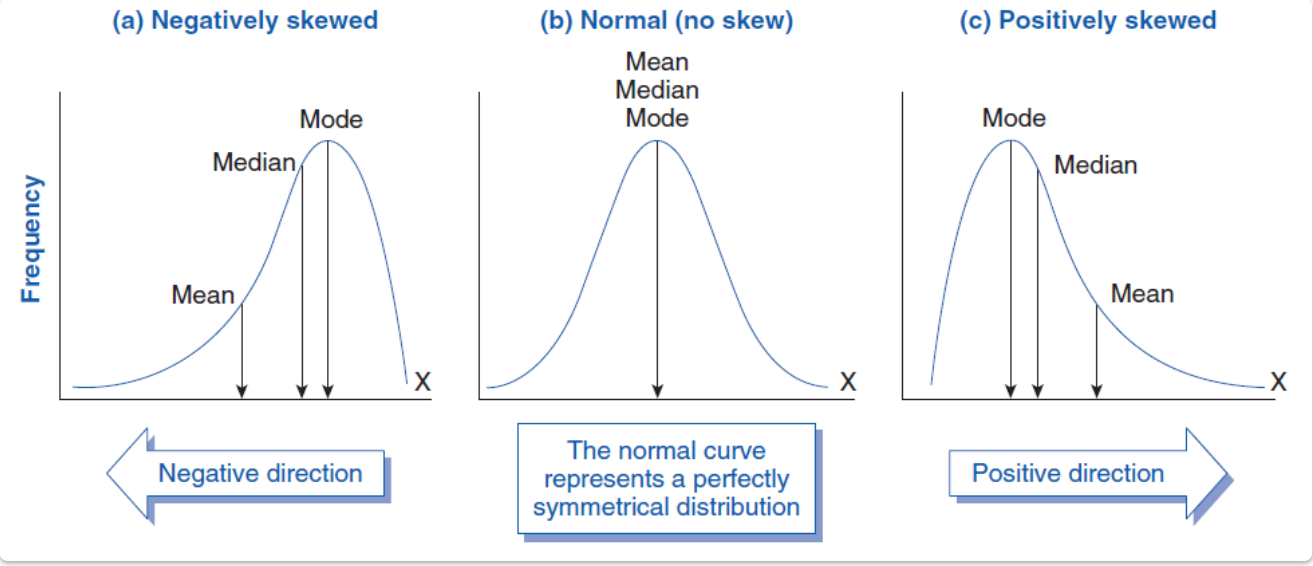
1. Outlier Detection
2. Central Limit Theorem
3. Hypothesis Testing

A Short Video on Normal Distribution
The Most Beautiful Video on Normal Distribution

Skewness

Skewness is a measure of the asymmetry of a probability distribution. It is a statistical measure that describes the degree to which a dataset deviates from the normal distribution.

In a skewed distribution, the mean, median, and mode are not equal, and the distribution tends to have a longer tail on one side than the other. The *greater the skew the greater the distance* between mean, median and mode.

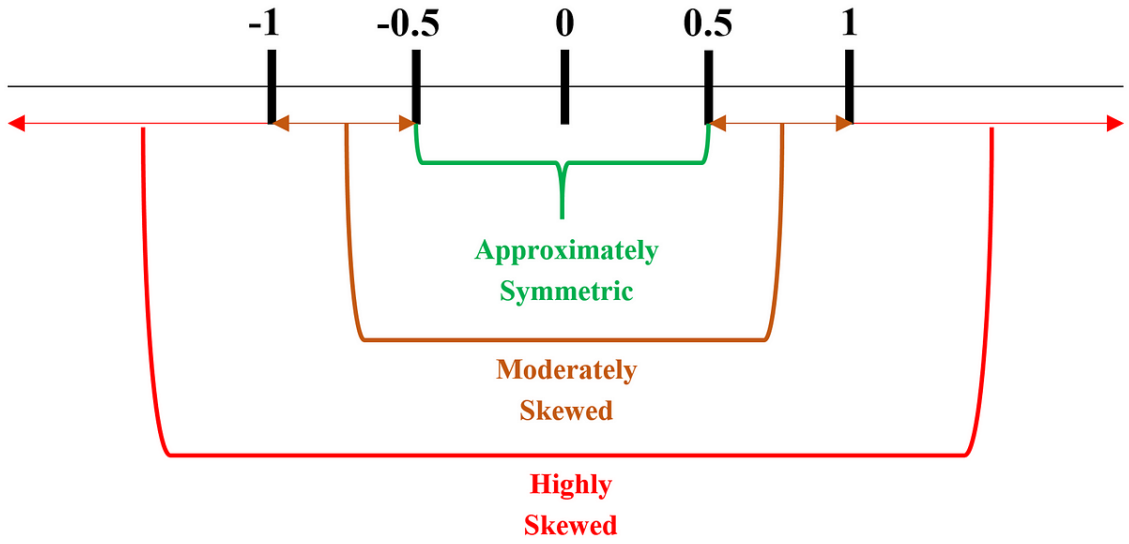


Skewness *can be positive, negative, or zero*. A positive skewness means that the tail of the distribution is longer on the right side, while a negative skewness means that the tail is longer on the left side. A zero skewness indicates a perfectly symmetrical distribution.

$$\text{Skewness: } S = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{3/2}}$$

Skewness *indicates how much a certain distribution resembles a normal distribution*.

Scale of Skewness:



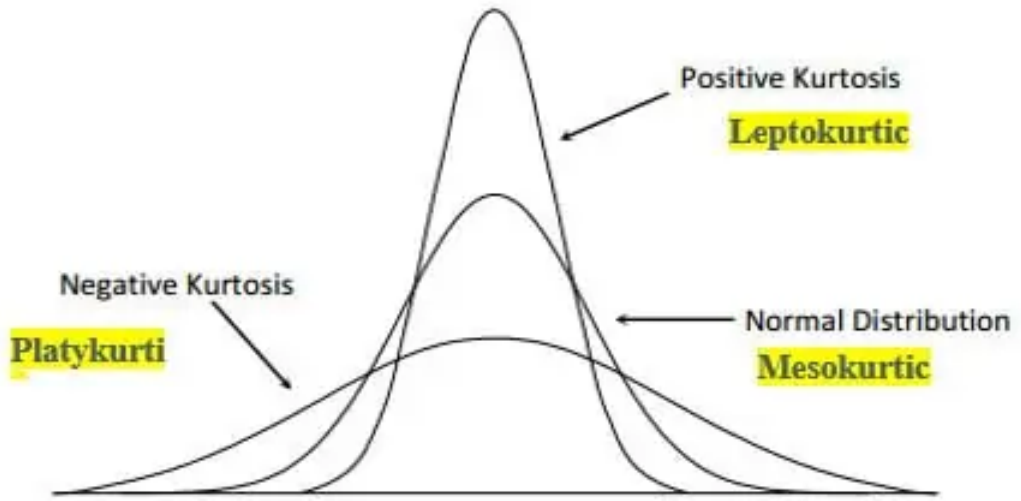
(-0.5, 0.5) = Low
(-1, -0.5) U (0.5, 1) = Moderate
(-1 & beyond) U (1 & beyond) = High

Kurtosis

In probability theory and statistics, kurtosis (meaning "curved, arching") is a **measure of the "tailedness"** of the probability distribution of a real-valued random variable. Like skewness, kurtosis describes a particular aspect of a probability distribution.
If a distribution has *high kurtosis, it means that there is a higher likelihood of extreme events occurring*, either positive or negative, compared to a normal distribution.

> **⚠ Warning**
>
> Kurtosis is the measure of tailedness and not the measure of peakedness as popularized.

> **Quote**
>
> As Westfall notes in 2014,"...its only unambiguous interpretation is in terms of tail extremity; i.e., either existing outliers (for the sample kurtosis) or propensity to produce outliers (for the kurtosis of a probability distribution)."



$$\text{Kurtosis: } K = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

ⓘ Info

Kurtosis is the 4th statistical moment. 3rd moment is the skew, 2nd is the variance and 1st moment being the mean.

Example

In finance, kurtosis risk refers to the risk associated with the possibility of extreme outcomes or "fat tails" in the distribution of returns of a particular asset or portfolio.

Kurtosis risk is important to consider because it indicates that there is a greater probability of large losses or gains occurring, which can have significant implications for investors. As a result, investors may want to adjust their investment strategies to account for kurtosis risk.

Excess Kurtosis is a measure of how much more peaked or flat a distribution is compared to a normal distribution, which is considered to have a kurtosis of 0.
It is calculated by subtracting 3 from the sample kurtosis coefficient.

1. **Leptokurtic** has fatter tails. This indicates that there are more extreme values or outliers in the distribution.
2. **Platykurtic** has thinner tails. This indicates that there are fewer extreme values or outliers in the distribution.
3. Distributions with zero excess kurtosis are called **Mesokurtic**.

Example

Assets with positive excess kurtosis are riskier and more volatile than those with a normal distribution, and they may experience sudden price movements that can result in significant gains or losses.

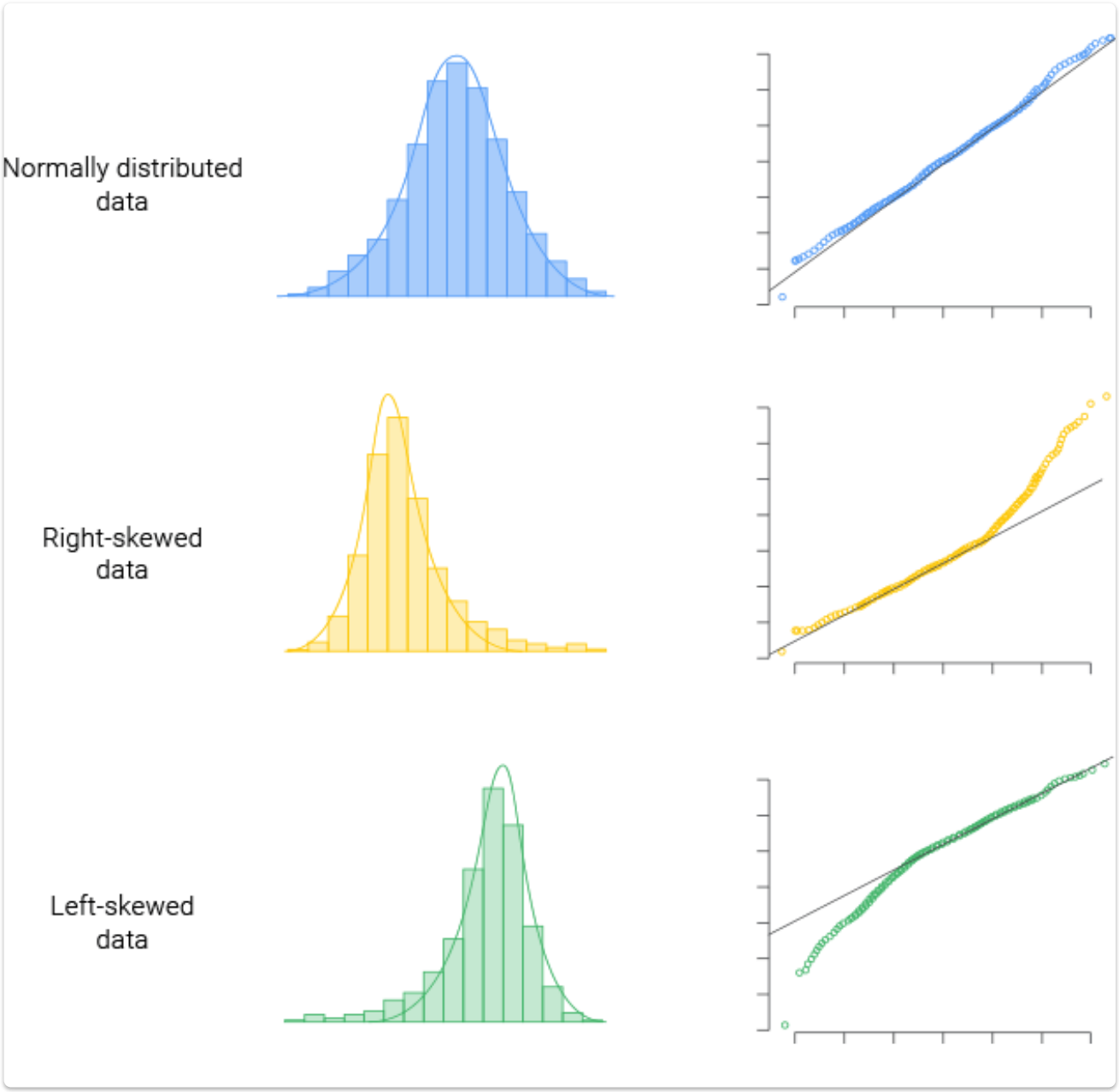
Assets with negative excess kurtosis are less risky and less volatile than those with a normal distribution, and they may experience more gradual price movements that are less likely to result in large gains or losses.

In finance, a mesokurtic distribution is considered to be the ideal distribution for assets or portfolios, as it represents a balance between risk and return.

Skew and Kurtosis

QQ Plot

A QQ plot (quantile-quantile plot) is a graphical tool used to assess the similarity of the distribution of two sets of data. It is particularly useful for determining whether a set of data follows a normal distribution.



⑦ How to check if a given distribution is normal or not?

We can check if a given distribution is normally distributed or not by the following methods:

1. Visual Inspection
2. Statistical Tests like the Shapiro-Wilk test, the Anderson-Darling test, and the Kolmogorov-Smirnov test.
3. QQ Plot

Does QQ Plot detect only normal distribution?

No. We can use QQ plot to detect any type of distribution.

QQ Plot Explained