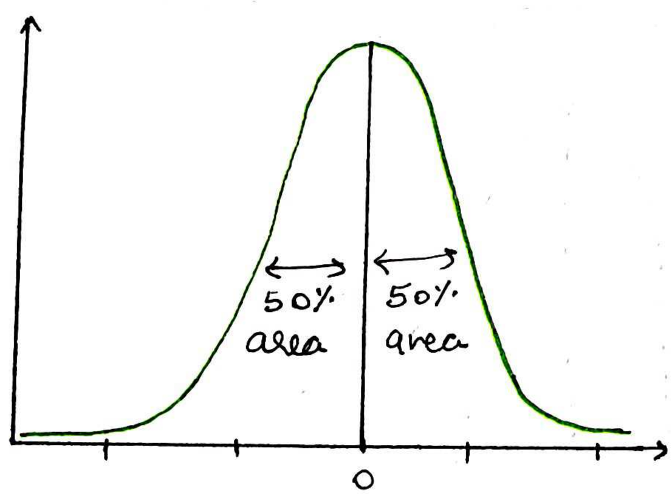
Properties

1. Symmetricity

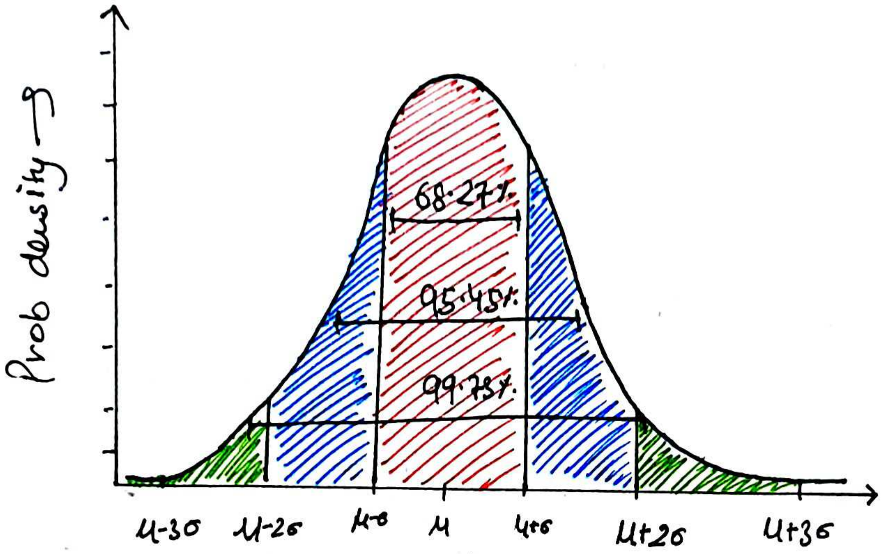
The normal distb^n is symmetric about its mean which means that the probability of observing a value above the mean is the same as the probability of observing a value below the mean. The bell-shaped curve of the normal distribution reflects this symmetry.



2. Measure of Central Tendencies are equal
   Mean, Median, Mode

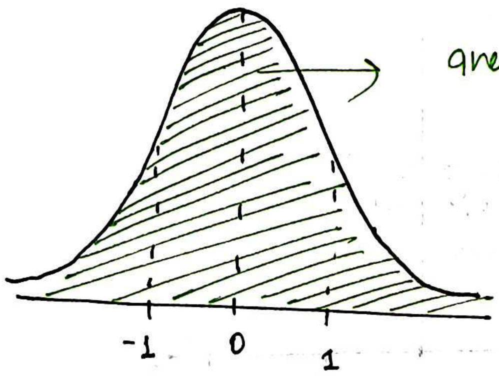
3. Empirical Rule

The normal distb^n has well-known empirical rule also called 68-95-99.7 rule, which states that approximately 68% of the falls within one standard deviation of the mean, about 95% of the data falls within two standard deviation of the mean, and about 99.7% of the data falls with three std of the mean.



68 - 95 - 99.7 Rule

4. The Area under the curve

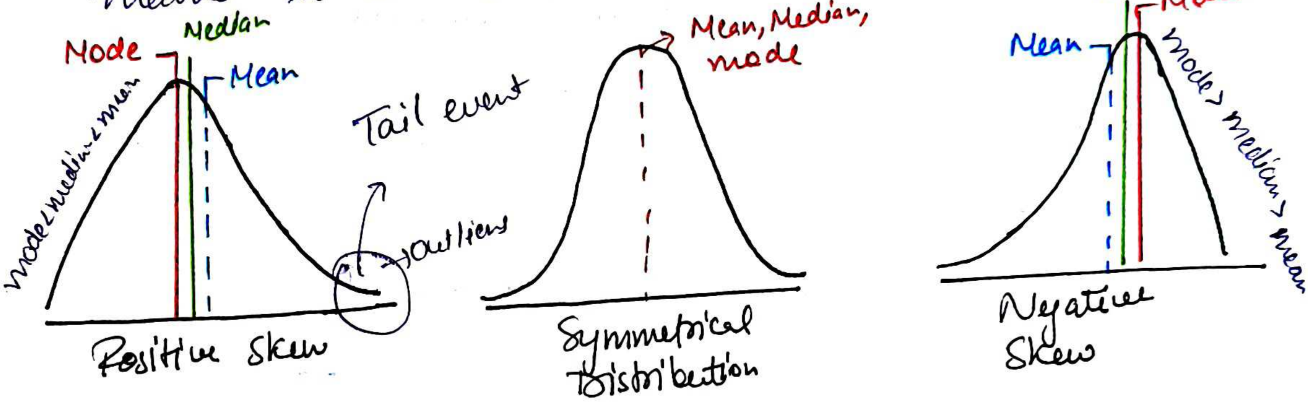


area under curve is 1 for (PDF) every PDF

Skewness

* What is Skewness?

A normal distribution is a bell-shaped, symmetrical distribution with a specific mathematical formula that describe how the data is spread out. Skewness indicates that the data is not symmetrical, which means it is not normally distributed.



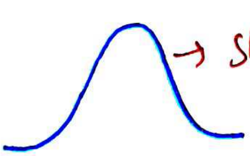
The greater the skew the greater distn betn mode, median, mean.

* How skewness is calculated?

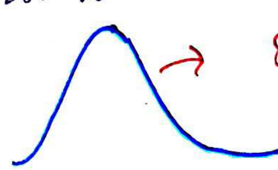
$$\boxed{\frac{n}{(n-1)(n-2)}\sum\left(\frac{x-\bar{x}}{s}\right)^3} \rightarrow \text{3rd Moment}$$

1st mont → mean

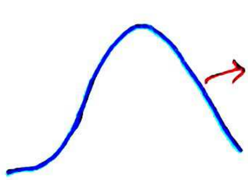
2st mont → variance



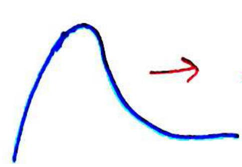
→ Skew = 0



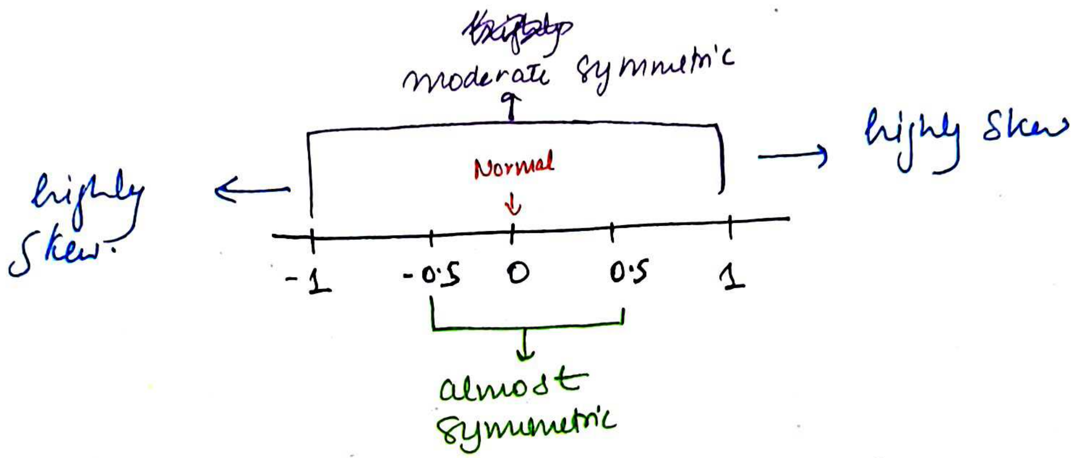
→ Skew = 1



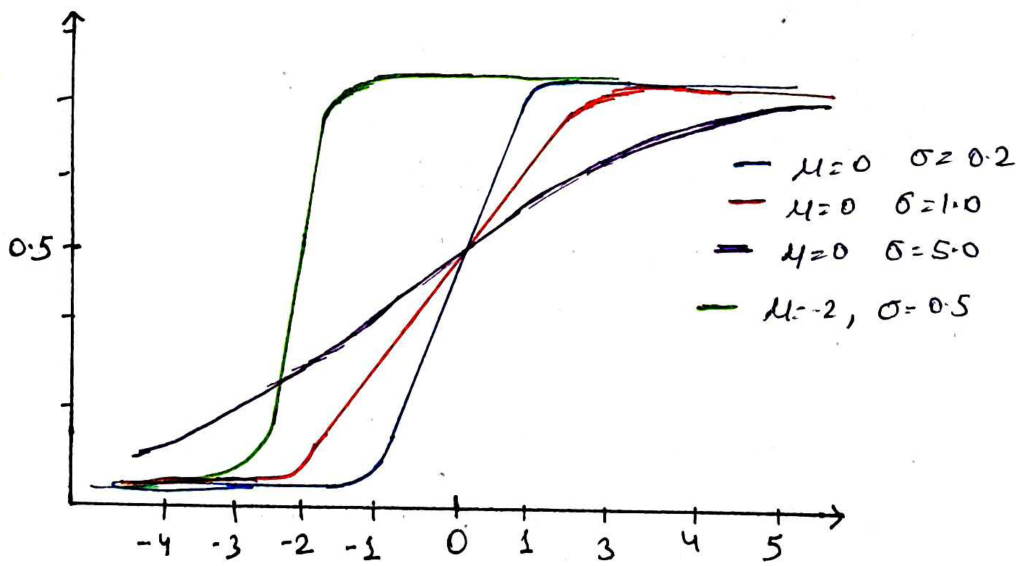
→ Skew = 0.5



→ Skew = 1.5

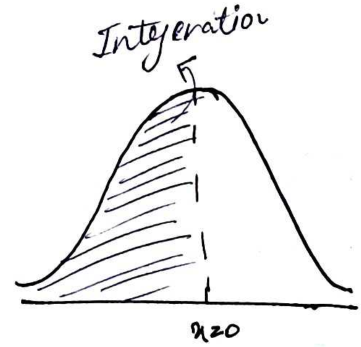


CDF of Normal Distribution



$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$



Use in data science

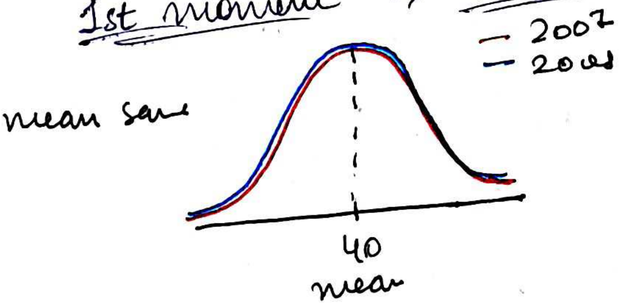
- Outlier detection
- Assumption on data for ML algo ⇒ LR and GMM
- Hypothesis testing → Central limit theorm.

Kurtosis

* What is kurtosis?

Kurtosis is the 4th statistical moment. 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.

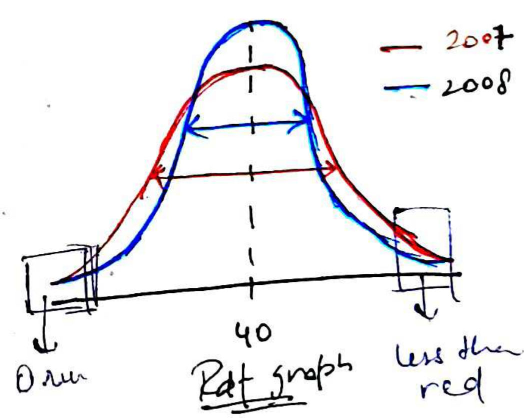
1st moment → mean



Batsman → Sachin

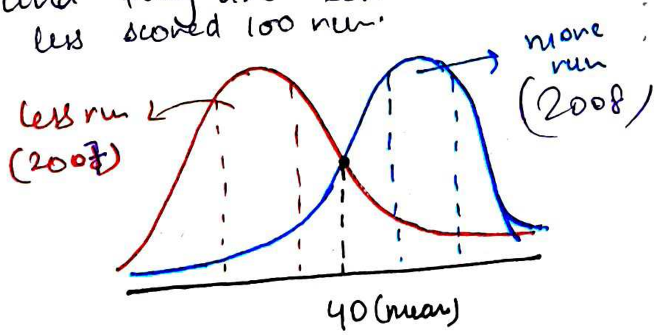
2007 → 100 matches → 40 avg

2008 → 100 matches → 40 avg



2nd moment → Std

Sachin more consistent in 2008 bcz most of the near the mean. and they are less out in 0 run. less scored 100 run.

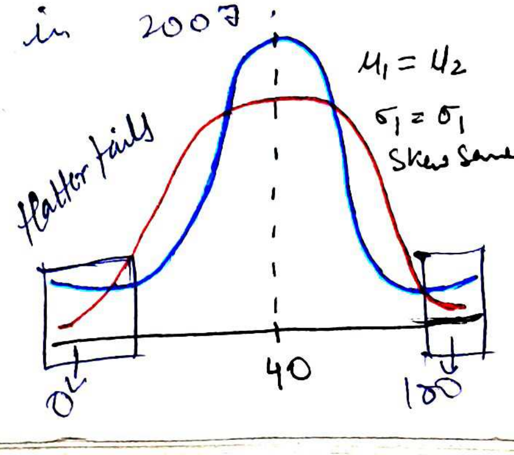


3rd moment → Skew

Sachin scored maximum run in 2008 and most of the less run in 2007.

4th moment → Kurtosis

In 2008, Sachin out maximum time at 0 and also scored 100 maximum time. As compare to 2007.



Formula

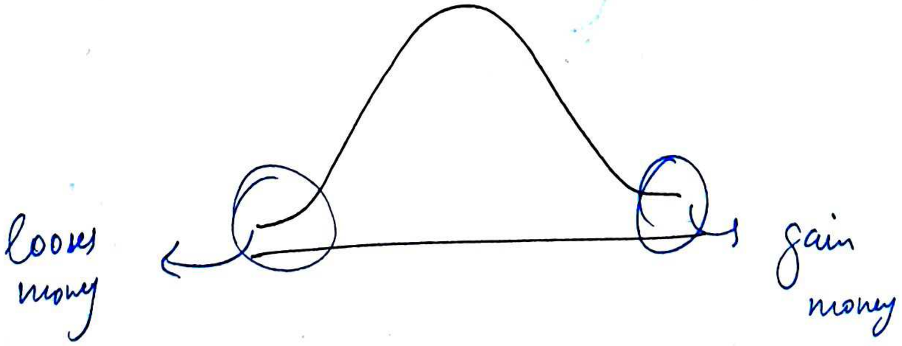
$$\left\{\frac{n*(n+1)}{(n-1)*(n-2)*(n-3)} * \sum_{j}^{n}\left(\frac{x_i-\bar{x}}{s}\right)\right\} - \frac{3*(n-1)^2}{(n-2)*(n-3)}$$

Practical Use-case

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

If a distribution has higher kurtosis, it means that there is a higher likelihood of extreme events occuring, either positive or negative, compared to a normal distribution.

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



Excess Kurtosis & Types

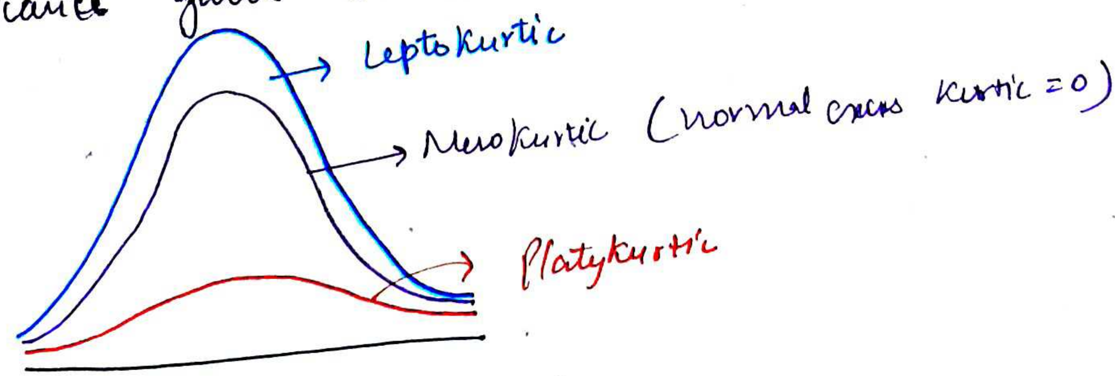
Excess kurtosis measure of how much tailed, flattered or thinner tail of our graph or distribution as compared to given normal distribution. Normal distribution's kurtosis is 0. Excess kurtosis calculated by subtracting 3 from the sample kurtosis coefficient.

Leptokurtic

A distribution with positive excess kurtos is called leptokurtic. "Lepto" mean "slender". In term of shape, a leptokurtic distribution has fatter tails. This indicates that there are more extreme values or outlier in the distribution.

$$\boxed{\text{Kurtosis} - 3 > 0}$$

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.



Platykurtic

A distribution with negative excess kurtosis is called platykurtic. "Platy" means "~~bra~~ broad". In term of shape, a platykurtic distribution has thinner tails. This indicates that there are fewer - extreme values or outliers in the distribution.

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

$$\boxed{\text{Kurtosis} - 3 < 0}$$

Mesokurtic

Distribution with zero excess kurtosis are called mesokurtic. The most prominent example of a mesokurtic distribution is the normal distribution family, regardless of the value of its paramete.

Mesokurtic is a term used to describe a distribution with a excess kurtosis of 0, indicating that it has the same degree of "peakedness" or "flatness" as a normal distribution.

Example: In finance, a mesokurtic distribution is considered to be the ~~total~~ ideal distribution for assets or portfolio, as it represent a balance both risk and return.

Q-Q Plot

* How to find if a given distribution is normal or not?

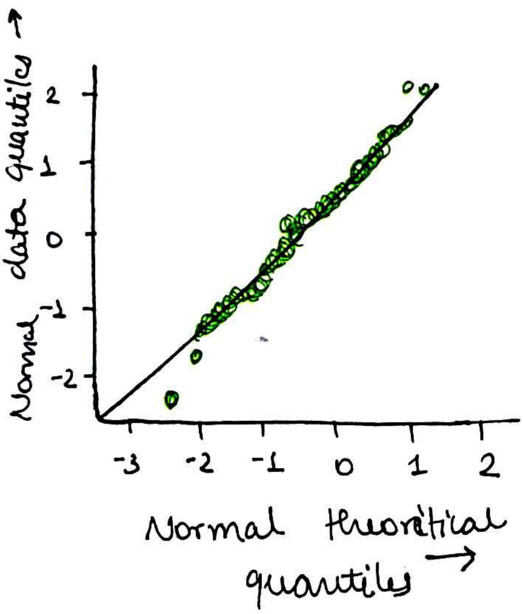
Visual inspection : One of the easiest ways to check for normality is to visually inspect a histogram or a density plot of the data. A normal distribution has a bell shaped curve, which means that the majority of the data falls in the middle, and the tails taper off symmetrically. If the distribution looks approximately bell shaped, it is likely to be normal.

QQ Plot : Another way to check for normality is to create a ~~normality~~ normal ~~is to create a~~ ~~normal probability plot plots the observed~~ prob plot ~~plots the observed~~ (also known as a QQ plot) of the data. A normal prob plot plots the observed data against the expected values of a normal distn. If the data points fall along a straight line, the distribution is likely to be normal.

Statistical tests : There are several statistical tests that can be used to test for normally, such as the shapiro-wilk test, the Anderson-darling test and the koimogorov-smirnov test. These tests compare the observed data to the expected value of a normal distn and provide a p-value that indicates whether the data is likely to be normal or not. A p-value less than the significance level (usually 0.05) suggest that the data is not normal.

* What is a QQ plot and how is it plotted?

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



Step 1 → theoritical data which is normal distribution

Step 2 → Our data → sort

Step 2 → Out data → quantiles

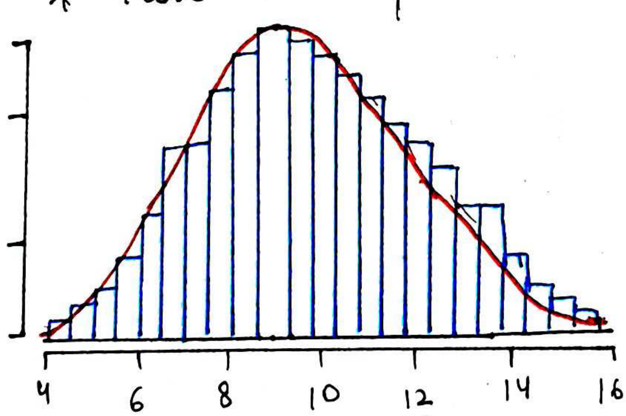
1 percentile, 2 percentile ... 100 percentile

Step 3 → ~~plot~~ theoritical data → sort

↳ quantile

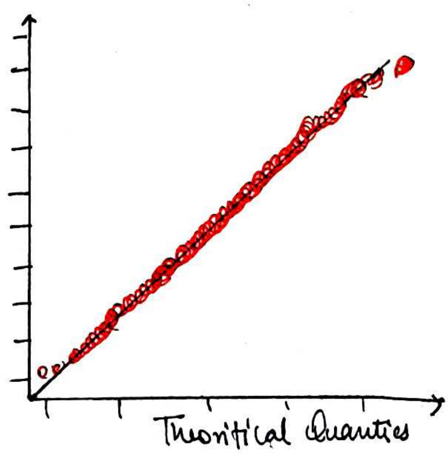
Step 4 :- plot our data quantiles and thoritical data.

* How to Interpret QQ plots



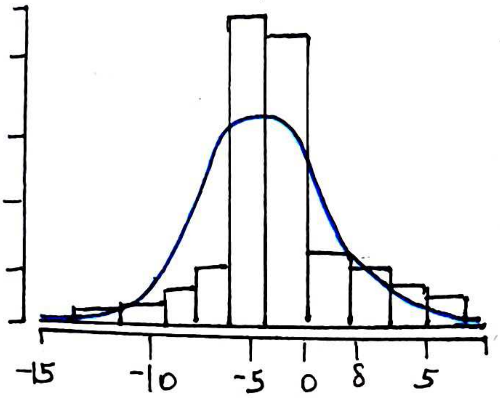
Histogram of Data

Normally distributed data

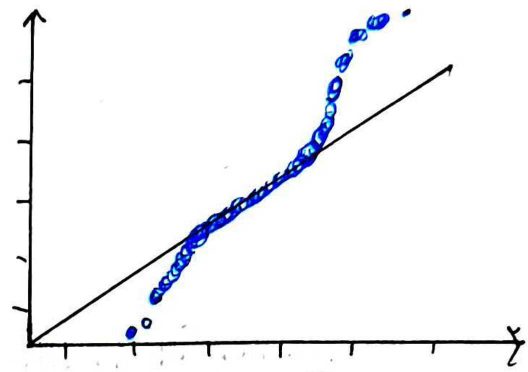


Normal Q-Q Plot

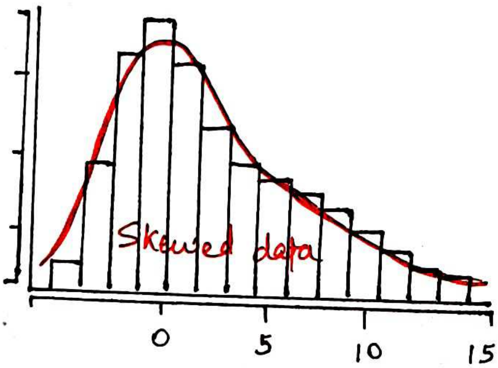
Data too peaked in middle



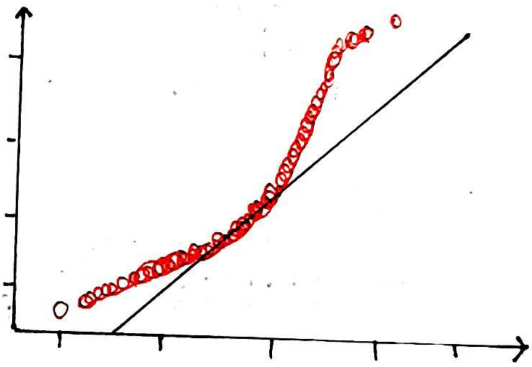
Histogram of data



Theoritical Quantile

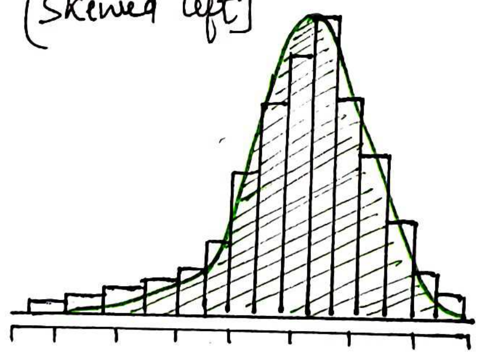


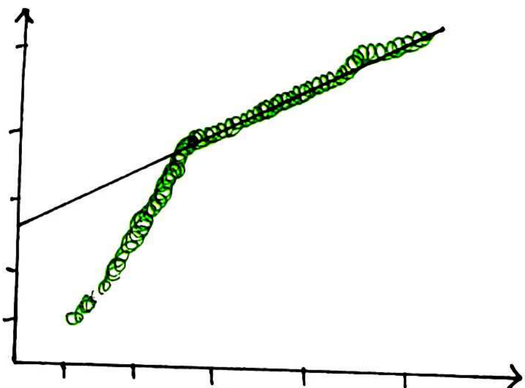
Histogram of data



Theoritical Quantile

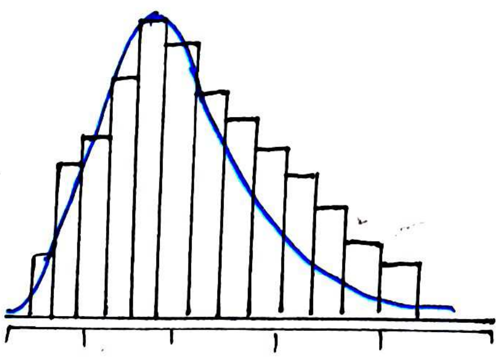
[Skewed left]

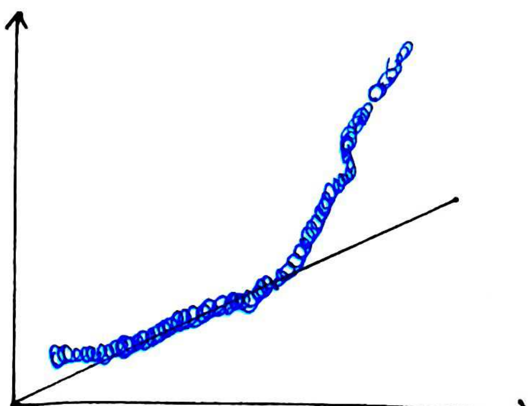




Theoritical Quantile

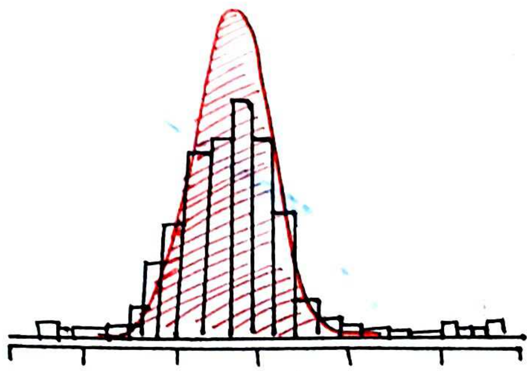
[Skewed Right]

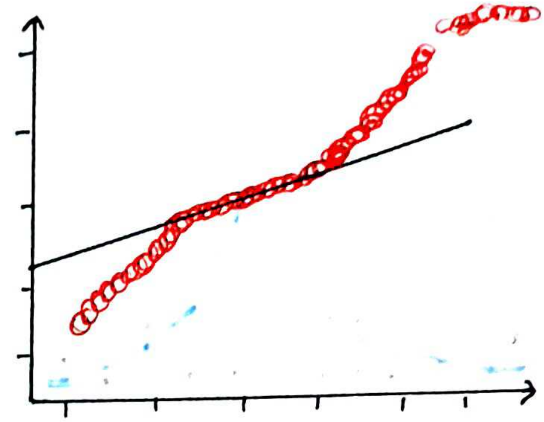




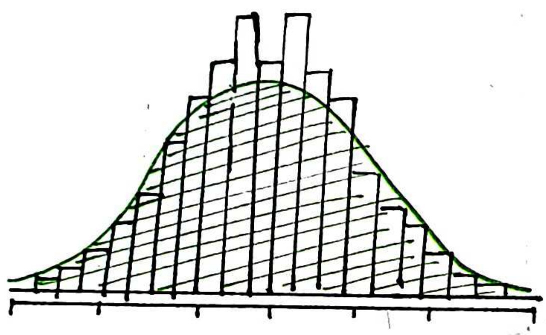
Theoritical Quantiles

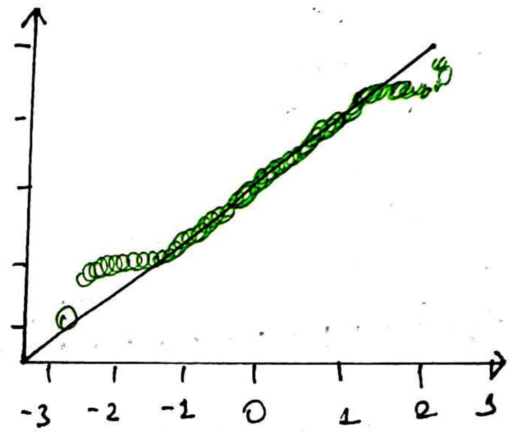
{Fat Tails}





{Thin Tails}





* Does QQ plot only detect normal distribution?