Definitions :

- **Descriptive Statistics -** procedures used to organize and present data in a convenient, usable and communicable form
- **Mean -** Average value of a sample or population

Population Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	N = number of items in the population

Sample Mean	$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$	n = number of items in the sample

- **Weighted mean -** Sum of a set of observations multiplied by their respective weights, divided by the sum of the weights

$$\bar{x}_w = \frac{\sum_{i=1}^{n}(w_i x_i)}{\sum_{i=1}^{n}(w_i)}$$

where as

$\bar{x}_w$ is the weighted mean variable
w_i is the allocated weighted vlue
x_i is the obsrved values

- **Median -** Value at the centre
- **Mode -** Value that occurs most
- **Variance -** The average of square differences between observations and their mean

$$\sigma^2 = \sum (X_i - \bar{X})^2 / N$$

σ^2 = variance
X_i = the value of the ith element
$\bar{X}$ = the mean of X
N = the number of elements

- **Standard Deviation -** Square root of the variance

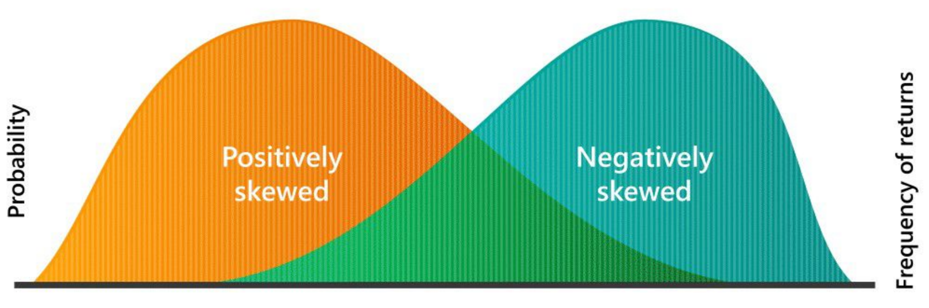
$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

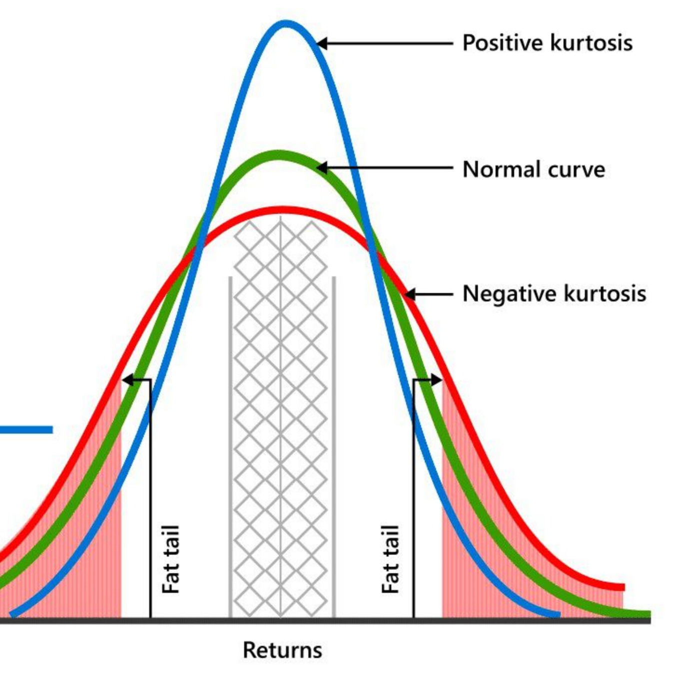
Interpreting σ :

- **Chebyshev's rule -** for any population at least 75% of the observations lie within 2σ of μ, at least 89% of the observations lie within 3σ of μ, at least 100(1 − 1/m2)% of the observations lie within m × σ of the mean μ.
- **IQR (Interquartile Range) -** The distance between the (n + 1)/4th and 3 × (n + 1)/4th observations in an ordered data set. These two values are called the first and third quartiles.

The measure of symmentry :

Skewness is the asymmetry of a distribution. A positively skewed distribution has a "tail" pulled in the positive direction. A negatively skewed distribution has a "tail" pulled in the negative direction. Most stock market returns are negatively skewed.

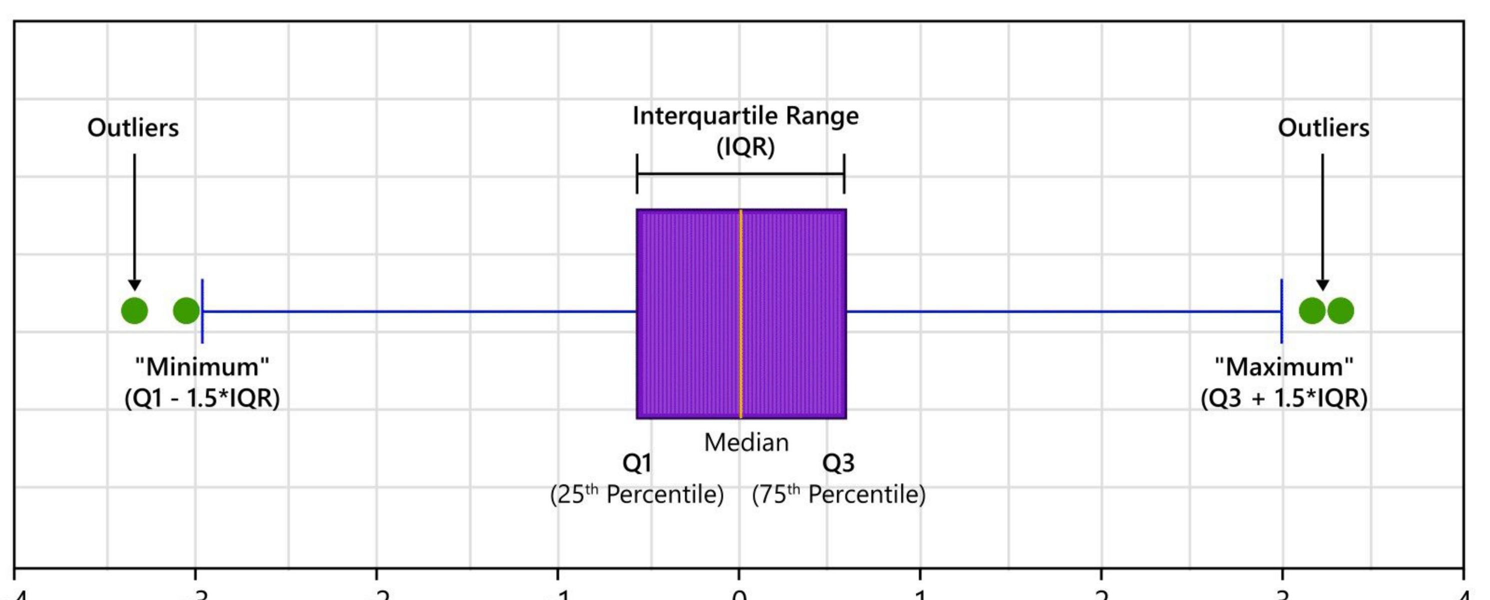




Normal not always the norm

Kurtosis refers to how peaked the curve is:steeper means positive kurtosis and fatter means negative kurtosis. Fat tails occur when there are more outsize returns on the downside or upside,or both, than the normal curve suggests.

Box-and-whisker plot : A graphic that summarizes the data using the median and quartiles, and displays outliers. Good for comparing several groups of data.



Correlation :

When there is some relationship between two things

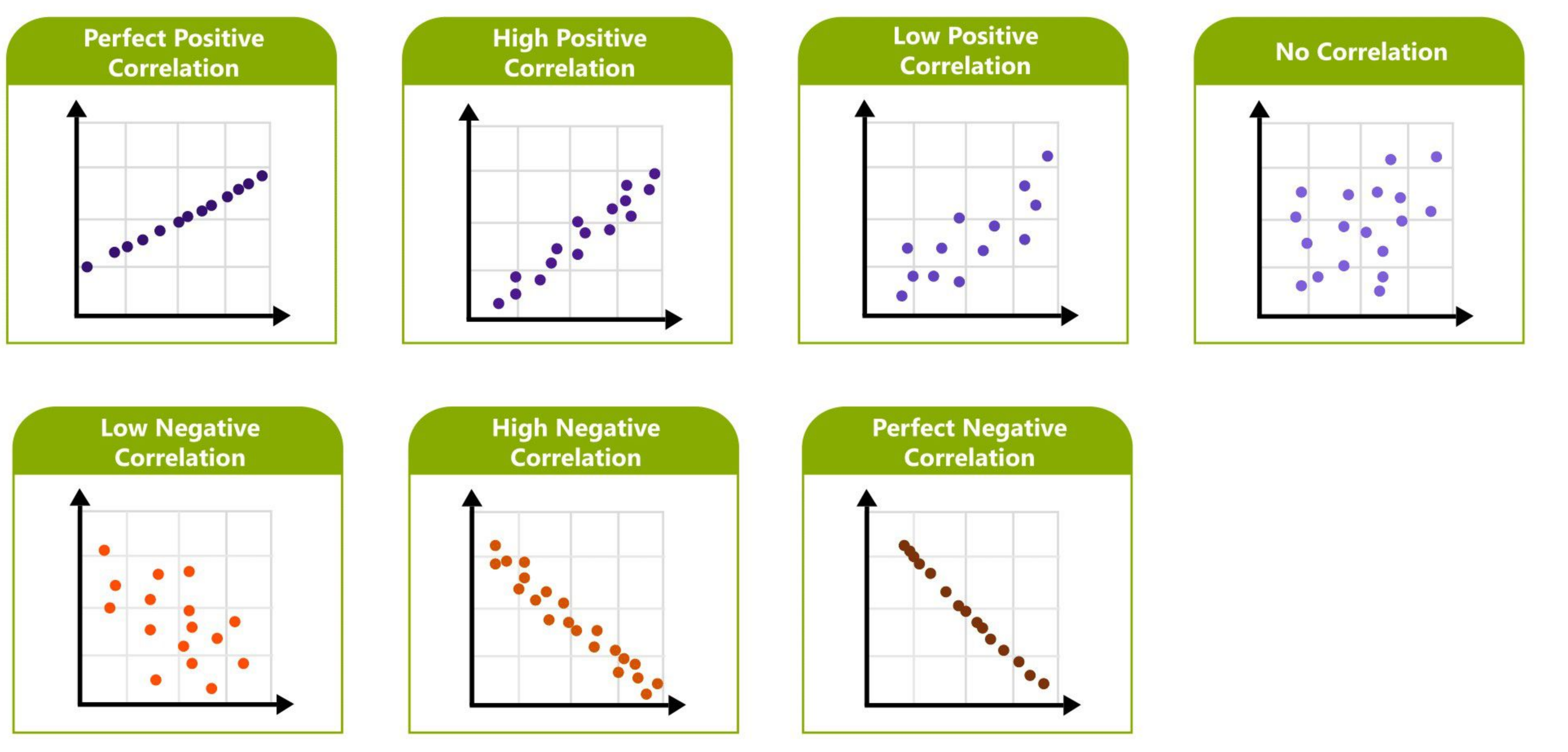
- Correlation always take values between-1 and 1

  -1 is a perfect negative correlation, which means as one thing gets bigger the other thing gets smaller

  0 is no correlation at all, basically is no relationship between these things

  1 is a perfect positive correlation, which means that when one thing gets bigger so does the other

- The closer the correlation value is -1 to 1, the tighter (more linear) the relationship will be on a scatter plot (see below on Pearson's coefficient)



Here's how we calculate correlation (Pearson's way) :

In this example we have two things to compare, X and Y

1. First calculate Mean (average) of **X**
2. Calculate Mean (average) of **Y**
3. Subtract Mean of **X** from each of **X** values (we'll call thease **A**), and subtract Mean of **Y** from each of **Y** values (we'll call thease **B**)
4. Square **A**'s (we'll call thease **C^2's**)
5. Square **B**'s (we'll call thease **D^2's**)
6. Multiply all **A**'s by **B**'s (we'll call thease **AB's**)
7. Add up all **AB's**
8. Add up all **C2's**
9. Add up all **D2's**
10. Now perform calculation below...

$$\text{Correlation} = \frac{\text{Sum of all AB's}}{\sqrt{(\text{Sum of C2's}) \times (\text{Sum of D2's})}}$$

Probabilities... (Chance) :

How likely something (an event) is to happen

Kind of Probabilities :

- **Conditional Probabilities -** Probability of an event happening based on whether or not something else happened
- **Joint Probabilities -** Probability of two events happening at the same time
- **Unconditional Probabilities -** Are just the summation of all probabilities

$$\text{Probabbility} = \frac{\textit{How many times event happened}}{\textit{Total Outcomes}}$$

Kind of Events :

- **Mutually Exclusive -** Events that can't happen at same time
- **Non-Mutually Exclusive -** Events that can happen at the same time
- **Independent -** When an event's probability isn't affected by anything else happening or not happening(e.g. a coin toss isn't affected by previous coin toss)
- **Dependent -** Events whose probabilities change based on each other happening or not happening

Cumulative Distribution Function

$$F_X(x) = \mathbb{P}(X \le x)$$

Cumulative Distribution Function

$$F_X(x) = \int_{-\infty}^{\infty} f_X(t)\,dt$$

$$\int_{-\infty}^{\infty} f_X(t)\,dt = 1$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Probability Distributions :

- **Poisson Distribution :**

notation	$Poisson(\lambda)$
cdf	$e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$
pmf	$\frac{\lambda^k}{k!} \cdot e^{-\lambda}$ for $k \in \mathbb{N}$
expectation	λ
variance	λ
mgf	$\exp(\lambda(e^t - 1))$
ind. sum	$\sum_{i=1}^{n} X_i \sim Poisson\left(\sum_{i=1}^{n} \lambda_i\right)$

Story - the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event

- **Binomial Distribution**

notation	$N(0,1)$
cdf	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt$
pdf	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
expectation	$\frac{1}{\lambda}$
variance	$\frac{1}{\lambda^2}$
mgf	$\exp\left(\frac{t^2}{2}\right)$

story: normal distribution with $\mu = 0$ and $\sigma = 1$.

Story - the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p

- **Normal Distribution :**

notation	$N(\mu, \sigma^2)$
pdf	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$
expectation	μ
variance	σ^2
mgf	$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
ind. sum	$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

Story - describes data that cluster around the mean

- **Standard Normal Distribtion :**

notation	$N(0,1)$
cdf	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt$
pdf	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
expectation	$\frac{1}{\lambda}$
variance	$\frac{1}{\lambda^2}$
mgf	$\exp\left(\frac{t^2}{2}\right)$

story: normal distribution with $\mu = 0$ and $\sigma = 1$.

Story - normal distribution with μ = 0 and σ = 1