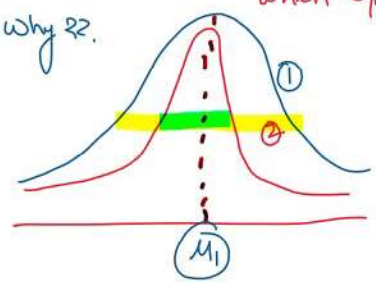
Variance:

Why ??



which Gaussian curve is having more variance ① Blue one

What??

Variance measures the spread of the data / variability of data from the avg. value ($\bar{x}$) or mean of the dataset.

univariate → one variable.

Population variance: $Var(x) = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$

sample variance: $Var(x) = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

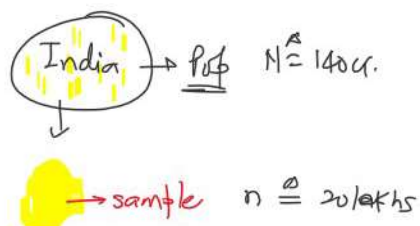
where: N: no. of observations in the population

n: no of observations in the sample

x_i: ith observation of data

μ: Population Mean

$\bar{x}$: sample mean.



conclusion. variance can only have positive number.

Higher the variance, higher the variability in the dataset

Std. deviation $\sigma = \sqrt{var(x)}$

x — height in cms

100

188

$\boxed{X - cm}$

var(X) — cm^2.

$\boxed{\sigma - cm}$

Normal distribution

$\mu \pm \sigma$:

$\mu \pm 2\sigma$:

Covariance

It measures how the two variables are varying together and the degree to which the deviation of one variable (X) from its mean is related to the deviation of another variable (Y) from its mean.

Population variance: $Var(X) = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ $\quad (x_i - \mu) * (x_i - \mu)$

Population covariance: $cov(X, Y) = \dfrac{\sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)}{N}$

Sample Covariance: $cov(X, Y) = \dfrac{\sum_{i=1}^{n} \overset{+ve}{(x_i - \bar{x})} \overset{-ve}{(y_i - \bar{y})}}{n-1}$

Conclusion: Unlike variance, covariance could be +ve, −ve and even '0'

Positive covariance: It indicates the two variables (X, Y), on an average, move in the same direction.

$X \uparrow\uparrow \rightarrow Y \uparrow\uparrow$ (and vice-versa)

Disclaimer: (No trading tips given)

stocks in Nifty 50 ↑↑↑ → Nifty 50 index ↑↑↑

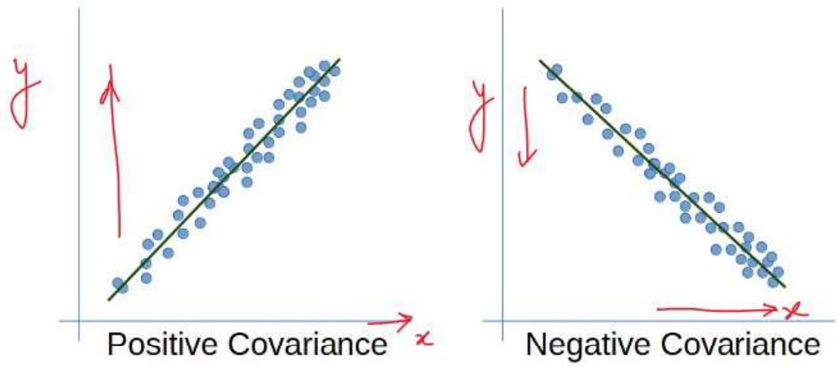
Negative covariance: indicates the two variables (X & Y) on an average move in the opposite direction.

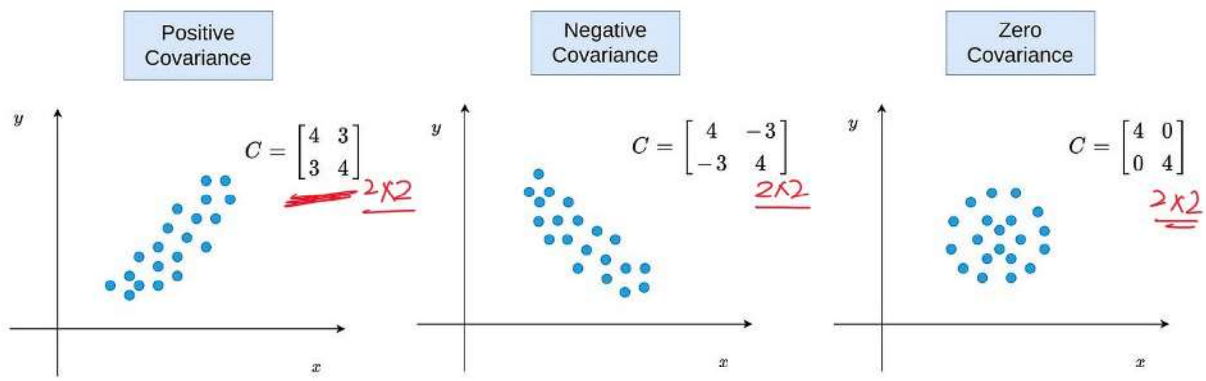
Price increases ↑↑↑ → Sale decreases ↓↓↓.

supply decreases ↓↓↓ → Price increases ↑↑↑↑

(Tomatoes)

zero covariance: there is no relationship between the two variables (X and Y)





Correlation:

While covariance measures how the two variables are varying together, correlation (or correlation coefficient) indicates how strongly the two variables are related to each other and measures both direction and strength of the relationship.

$$\text{Pearson's Correlation } \rho_{xy} = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad \text{: Population}$$

$$-1 \leq \rho \leq 1$$

$\rho = 0$ No correlation $\qquad 0 < |\rho| \leq 0.3$: weak correlation

$\rho = 0$ No correlation

$\rho > 0$ +ve correlation

$\rho < 0$ −ve correlation

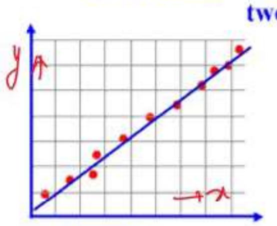
$0 < |\rho| \leq 0.3$: weak correlation

$0.3 < |\rho| \leq 0.7$: Moderate correlation

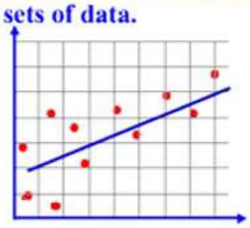
$|\rho| > 0.7$: strong correlation

SCATTERPLOTS & CORRELATION

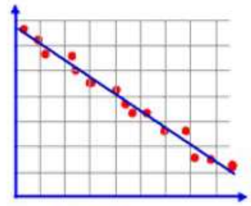
Correlation - indicates a relationship (connection) between two sets of data.



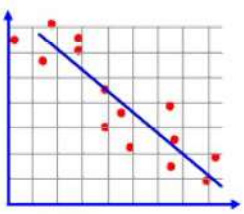
Strong positive correlation



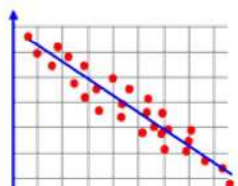
Weak positive correlation



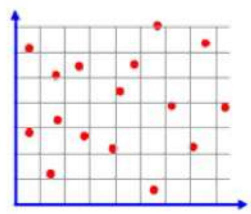
Strong negative correlation



Weak negative correlation



Moderate negative correlation



No correlation