Recap :- PDF < PMF, PDF

Random variable < Discret → PMF, continuous → PDF

PMF, PDF → CDF

$P(X \leq x)$ = CDF (Algebraic)

$X = 4$: PMF

$X \leq 4$: CDF

PDF < Param, non-param — KDE

Param < Assumption, Normal

non-param: There is no assumption

PMF: Dis → PMF → Dis

{H, T}, {1, 2, 3, 4, 5, 6}

PDF: cont

marks (CGPA) = {1-10}

EMI —

$\mu = 50$, $\sigma = 5$

$\mu = 60$, $\sigma = 2$

iris

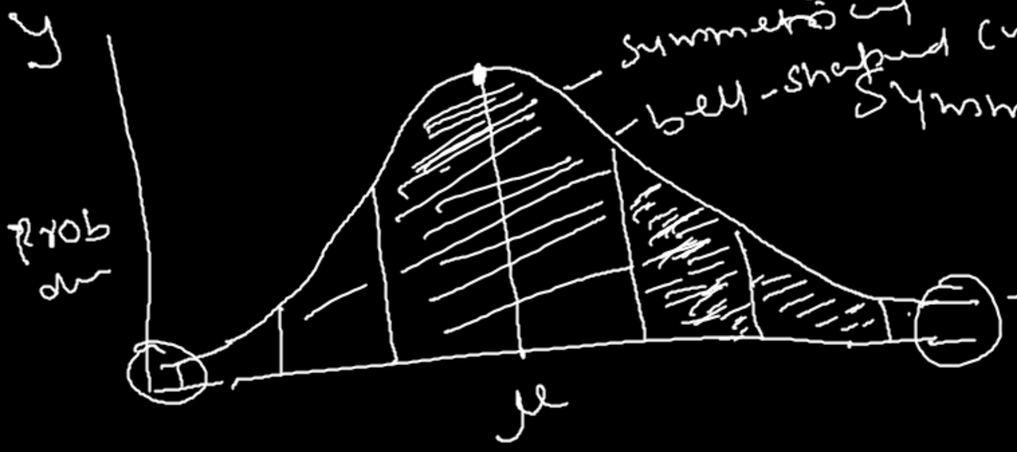
Discrete Dist

1. Uniform dist
2. Binomial dist / Bernoulli
3. Poisson dist

Continuous Distributions

1. Normal / Gaussian dist
2. Standard Normal dist

* **Normal Distribution** → Gaussian dist



- symmetrical
- bell-shaped curve
- Symmetrical around the mean / bell curve
- tail ∞

$X \sim N(\mu, \sigma)$

$\mu + \sigma, \mu + 2\sigma, \mu + 3\sigma$

$\mu + 5\sigma =$

μ = mean

σ = std

$$Y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

x = value given
μ = mean
σ = Std

Parameter of Normal dist

$$Y = e^x \qquad y = e^{-x^2} = e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$e = 2.718$

→ The normal dist is the most important Prob dist in Stats, it fits in many natural phenomena

eg - height, blood press, pulse, IQs,

Normal Dist

→ Always a symmetrical μ, σ

→ Long Tail

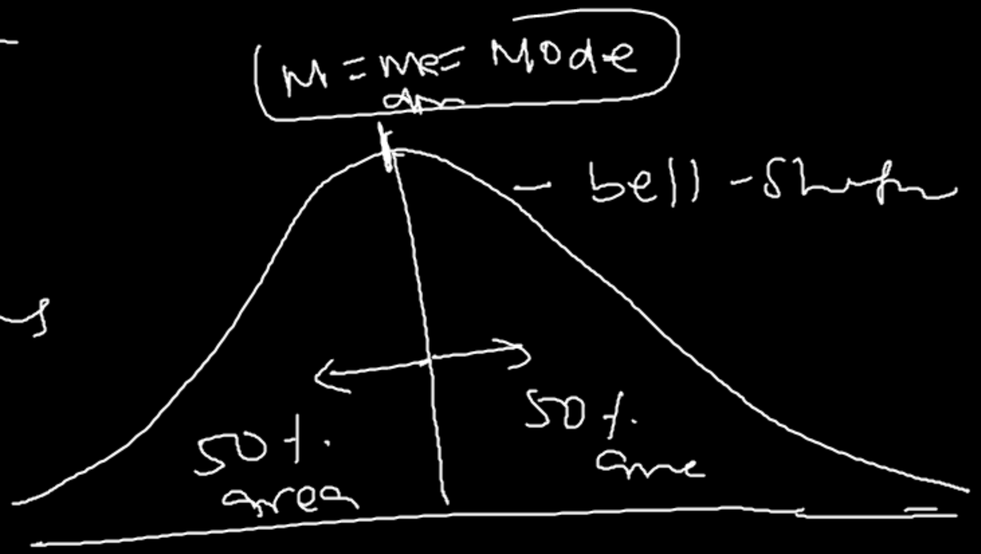
→ Bell shaped

→ Mean / mode & median all are same

→ Two factors define the shape of the curve

→ mean & Std

Properties of Normal distribution



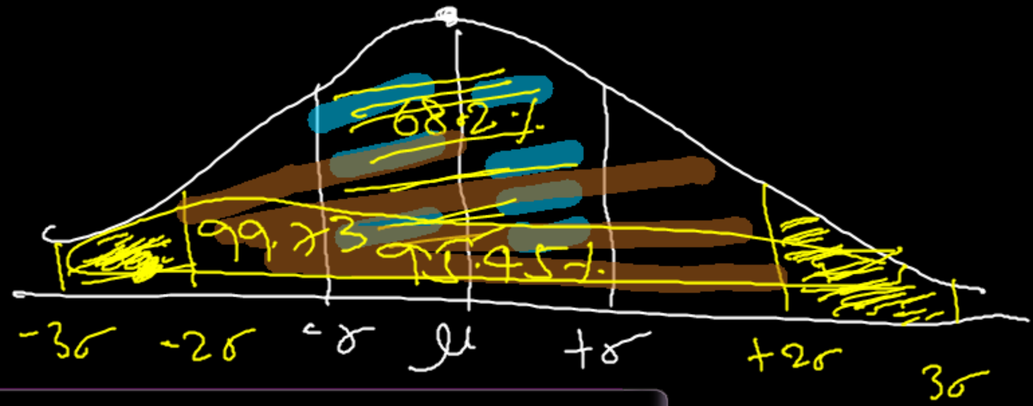
(1) Symmetricity

(2) Measures of central Tendencies are equal

mean = median = mode

(3) Empirical Rule (68-95-99.7 - Rule) — Z-Table

$\mu \pm \sigma$ or $\bar{x} \pm s$ (46-48)



Empirical or 68 – 95 – 99.7 Rule

(1) What % of data falls within a certain number of Std from the mean.

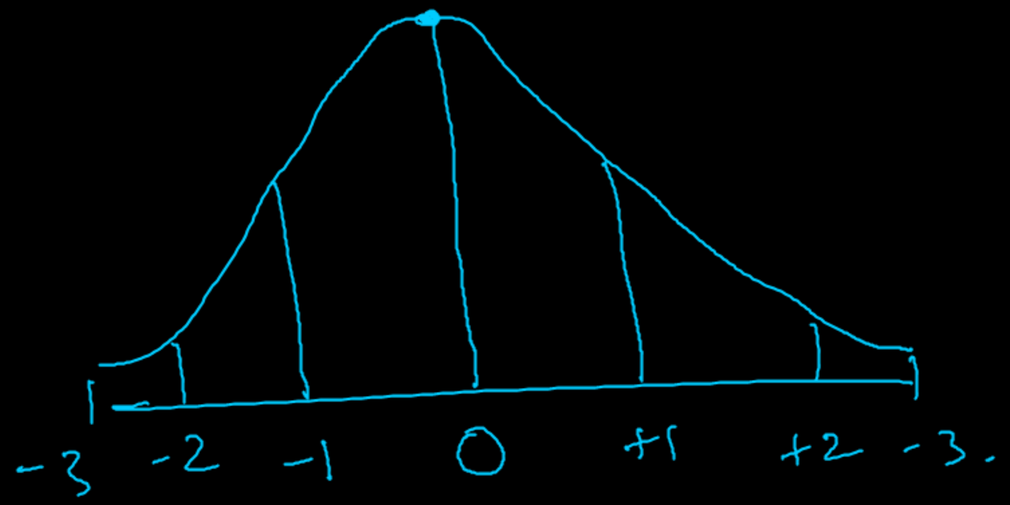
(5) 68.27 % of data falls within one Std of the mean.

(6) 95.45 % " " " " two " " " "

(9) 99.73 % " " " " three " " " "

* Standard Normal Variate

A Standard Normal variate (z) is a standardized form of the normal dist with mean = 0 & Std = 1



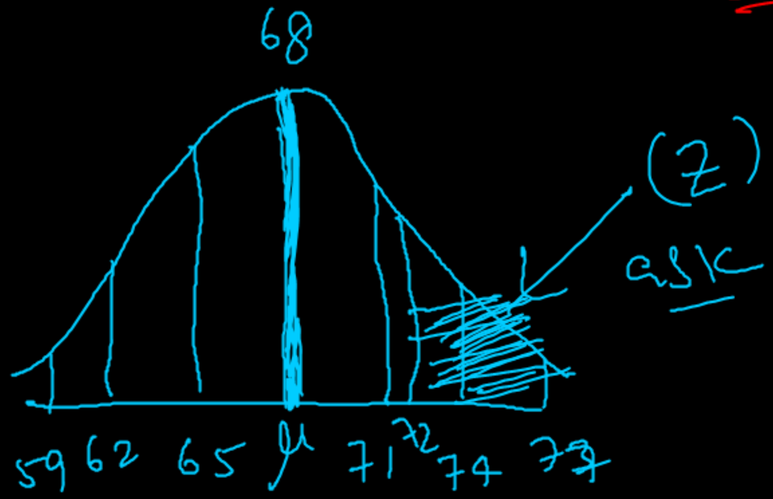
$X \sim N(68, 3)$

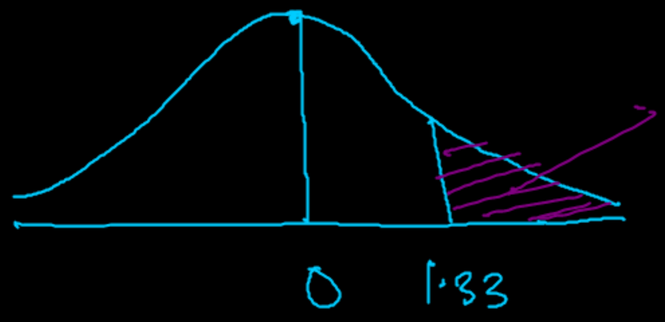
72 inches

$\bar{x}$

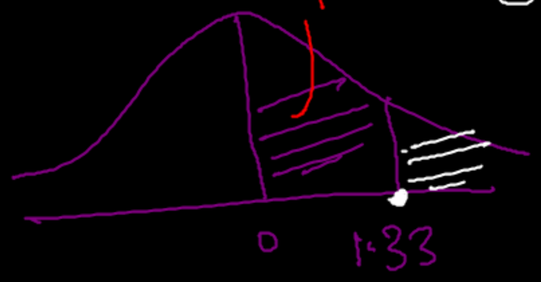
$$Z = \frac{x - \mu}{\sigma}$$ Formula

$$Z = \frac{72 - 68}{3} = \frac{4}{3} = 1.33$$





90.8% Z-Table



100 − 90.8 = 9.2%

100 − 90.8 = 9.2%

1.33 = 0.03

1.3 + 0.03 = 1.33

0.908

90.8

$$Z = \frac{x - \mu}{\sigma} = \frac{60 - 68}{3} = \frac{-8}{3} = -2.66$$

Q:(i) Mr. X 192 out 200

(ii) mean = 150 & σ = 25

(iii) Normal Distn, what would be the chances to be hired Mr. X ?

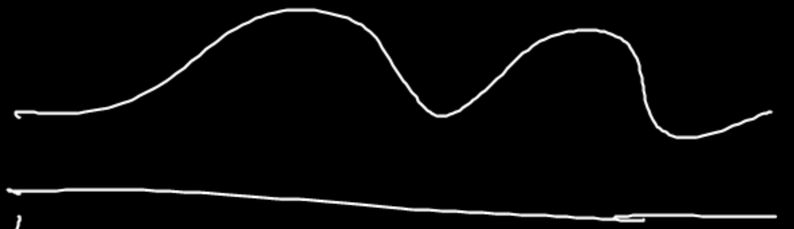
Q:- Virat Kohli Score - 138, India will win the match?

Virat, $\mu = 100, 20$

Skewness

↓

asymmetry of prob dist



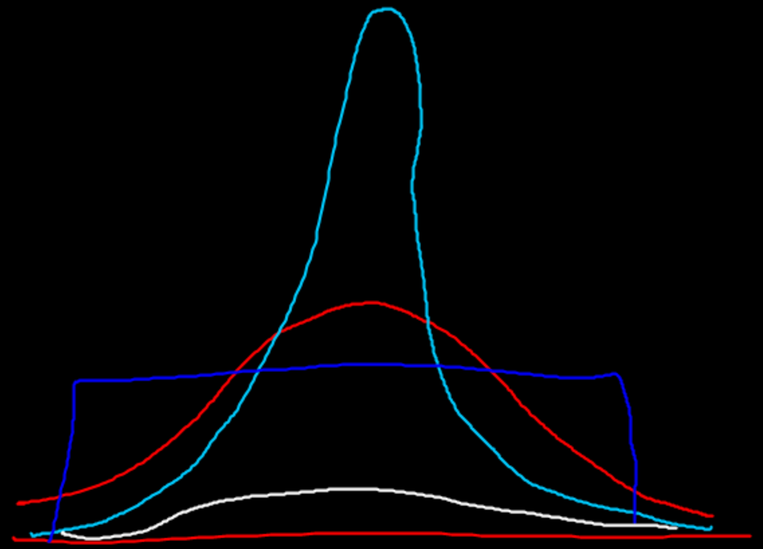
→ In a symmetrical dist, the mean, median & mode are equal.
In contrast, in a skewed dist (Asymmetrical dist), the mean, median & mode are not equal.

→ Skewness can be positive, negative or zero.

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x-\bar{x}}{s}\right)^3$$ — Skewness

moment

1 mu — mean/mu
2 moment — variance
3 moment — skew (x^3)
4th " — kurtosis



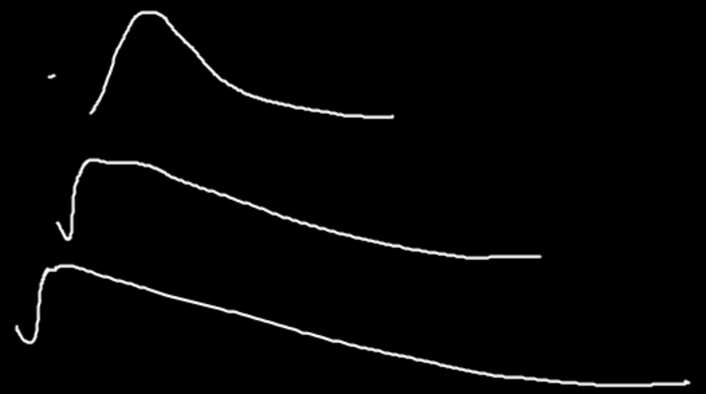
Skew = 0 — Symmetrical

Skew = 0.5

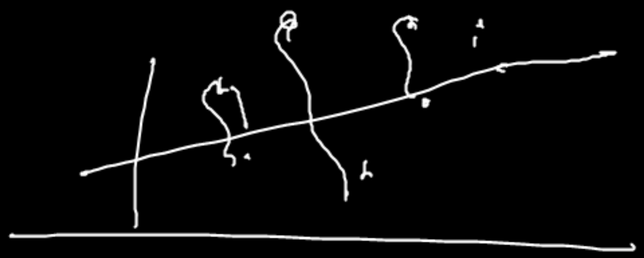
Skew = 1

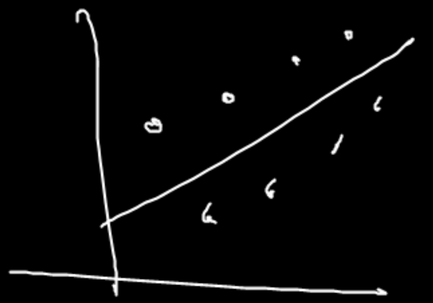
Skew = 2

Skew = -1, -2, ..



Data science





1. Outlier Detection
2. Assumption on data for ML algorithms
   - ↳ Linear Regression & GMM Q-Q Plot
3. Hypothesis testing
4. central limit theorem