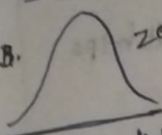
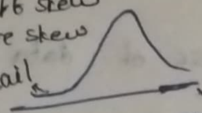
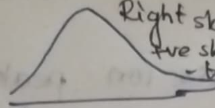
October-30:

B. 

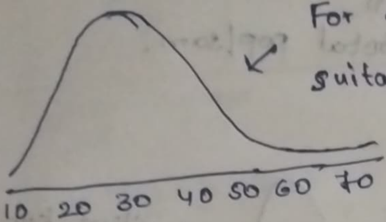
Normal distribution





These are not normally distributed as they are not symmetric (Skewness is based on tail growing side)

Skewness.

Ex: 

For diabetic record prediction this model will be not be suitable as the data is not collected properly.

the model becomes biased.

* Whenever the skewness is present,

1. Positive Skew / Right skew: Tail on the right side is longer & most of the data is in left.

2. Negative / Left skew: Tail on the left side is longer & most of the data is in right.

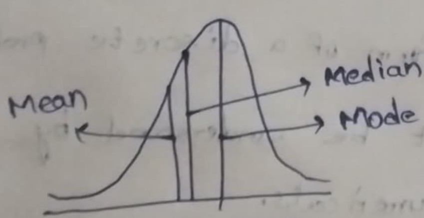
3. Zero skew (symmetric): The data is evenly distributed around the mean (like a normal distribution).

* We can identify the skewness based on the visual representation & the calculation.

$$\boxed{\text{Skewness} = \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}}$$

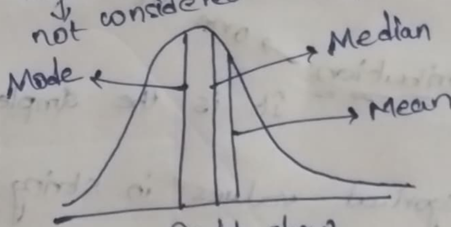
* In zero skew Mean = Median = Mode
↓
not considered mostly

(Mean presents always towards tail. Mode is the peak point.)



Left skew

Mean < Median



Right skew

Mean > Median.

* Skewness value ranges between -1 to +1.
  - If the value is near to -1, then it's -ve skew
  - If the value is near to +1 then its +ve skew
  - If the value is between "-0.5 to 0.5" it is zero skew.

kurtosis: Talks about outliers of the data.

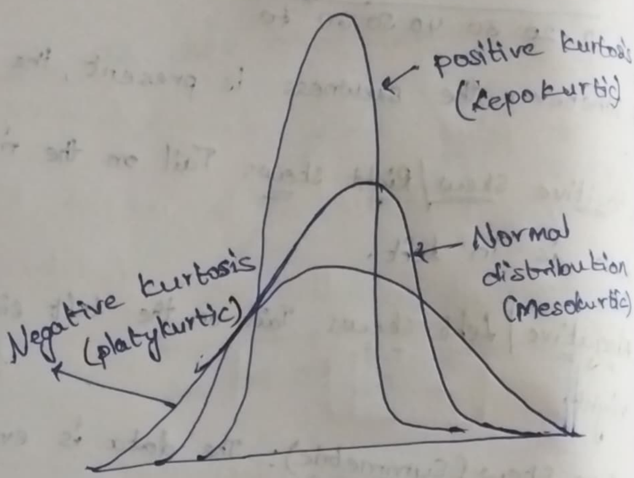
It measures the tailedness (or peakness of data distribution, it helps to find outliers.

$$\boxed{k = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)}$$

$\mu / \bar{x} \rightarrow$ mean.
$\sigma \rightarrow$ SD
$n \rightarrow$ total pop/sam.

Types of kurtosis:

+ Measure



1. Mesokurtic ($k=3$):
   - Normal distribution
   - No outliers
   - Moderate tail & peak

2. Lepokurtic ($k>3$):
   - Heavy tails
   - Sharp peak
   - More outliers.

3. Platykurtic ($k<3$):
   - Light tails & flat peak.
   - Less outliers

* Normal & Standard Normal Distributions, are continuous distributions.

Bernolli
3. Bernoulli Distribution: → one event. It is the simplest form of a discrete probabili

* The ~~disc~~ categorical values in string can't be understood by model, we change them to discrete numericals.

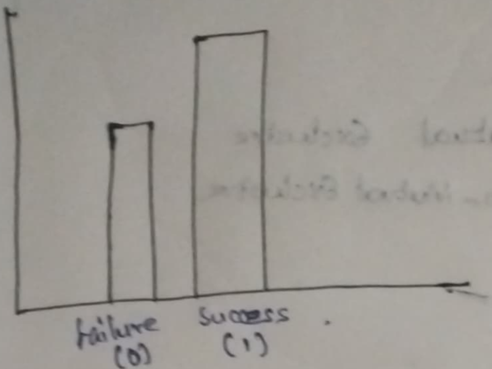
yes → 0
no → 1

Green - 1
Blue - 2
Red - 3.

* It is the simplest form of a discrete probability distribution & models a random experiment with exactly 2 outcomes.

Sucess denoted by = P
failure : 1 - P

(Total probability is 1. it ranges from 0 to 1.)



4. **Binomial Distribution:** It generalises the bernoulli distribution to multiple events or trails. It models the number of success in a fixed number of independent & identical bernoulli trials.

Multiple events & no fixed time.

$$\boxed{P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

k : no. of success
n : no. of trails
p : probability of success

5. **Poisson Distribution:** The poisson distribution is used to model the number of events that occur in a fixed time interval or space and occur independently, the parameter (λ) represents the average number of events in the interval.

Multiple events & fixed intervals.

$$\boxed{P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}}$$

λ : avg. no. of events in interval
k : no. of occurrences.

Inferential Statistics:

Probability: It is a measure of likelihood of an event.

Eg: Dice = {1, 2, 3, 4, 5, 6} Rolling a dice.

$P(4) = \frac{1}{6}$

$P(2, 5, 3) = \frac{3}{6} = \frac{1}{2} = 0.5$

$$\boxed{P(x) = \frac{\text{No. of favourable outcomes}}{\text{Total outcomes.}}}$$

- Tossing 2 coins {HH, HT, TH, TT}.

There are 2 rules in probability:

1. Addition Rule. → OR
2. Multiplication Rule → AND.