DS & AI
CS & IT

Probability & Statistics

Lecture No. 09

By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic

→ BAYE'S THEOREM

→ Concept of with & w/o Replacement

Topics to be Covered

Topic

BASICS of STATISTICS

(Discrete Random Variable)

Statistics.

Random Variable → whenever we are not sure about the outcome of an Experiment, then such types of Experiments are called R.Exp & variable involve in R.Exp is called RANDOM VARIABLE.

Discrete R.V (X) → Counting Related variables are called D.R.V.

for eg; No. of students, No. of vehicles, No. of Deaths etc

Continuous Random Variable (x) → when R.V has infinite possibilities in a certain Range then it is called C.R.V

for eg: Height, weight, time etc

Random Variable

D.R.V (X)

↓

Discrete Prob Distribution

eg (BINOMIAL, POISSON, GEOMETRIC)

↓

Prob Mass func^n (p.m.f)

$p_i \geq 0, \ \sum p_i = 1$

C.R.V (x)

↓

Continuous Prob Distribution

eg (Exponential, Uniform, Normal)

↓

Prob. Density func^n (p.d.f)

$f(x) \geq 0 \ \& \ \int_{-\infty}^{\infty} f(x)\,dx = 1$

p.m.f

(i) Expected Value $E(X) = \sum p_i x_i$

(ii) Variance $(X) = E(X^2) - (E(X))^2$

(iii) $S.D(\sigma) = +\sqrt{Var(X)}$

p.d.f

(i) Expected Value $E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$

(ii) Variance $(x) = E(x^2) - (E(x))^2$

(iii) $S.D(\sigma) = +\sqrt{Var(x)}$

Note ① Random Variable: $X = \{$ which is Required should be assumed as X $\}$

② $E(X^2) = \sum p_i X_i^2$, $E(X^3) = \sum p_i X_i^3$, ...

(1) Measures of Central Tendency (Mean, Median, Mode)

MEAN (Central Value / Average / Expected Value)
(μ) ($\bar{x}$) $E(x)$

It is the Average of Random Variable (X)

$$\bar{X} = \frac{\sum X}{N}$$

$$E(X) = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i$$

MODE → The Data having highest Frequency is Called Mode.

OR

the data which is Repeating More as compare to others known as MODE.

MEDIAN → After Arranging the data either in Increasing order or in Decreasing order, the Middle Most Value is Called Median.

Case I: if N = odd then $M_d = \left(\frac{N+1}{2}\right)^{th}$ observation

Case II if N = even then $M_d = \dfrac{\left(\frac{N}{2}\right)^{th} + \left(\frac{N}{2}+1\right)^{th}}{2}$

MEASURES of DISPERSION

(Variance, S.D, Co-Variance)

Variance :→ (MSD) It measures the spread of Distribution about Central Value (μ) i.e for smaller Variance, individual Values lies Closer to Mean.

Defn: "Variance is the Average of Squares of Deviations from Central Value ($\bar{X}$)"

$$Var(X) = E(X-\bar{X})^2 = \boxed{\frac{\sum (X-\bar{X})^2}{N}} = \cdots\cdots = \boxed{E(X^2) - (E(X))^2}$$

S.D (σ) (RMSD) it has the same Physical Significance as that of Variance. & it is defined as $SD(\sigma) = +\sqrt{Var(X)}$

PODCAST: Consider 4 kids having weights 9 kg, 13 kg, 16 kg, 22 kg

(Analysis) Average weight $(\bar{x}) = \frac{9+13+16+22}{4} = 15\text{ kg}$

Average of Deviation from Central Value $= \frac{(9-15)+(13-15)+(16-15)+(22-15)}{4}$

$= \frac{(-6)+(-2)+(1)+(7)}{4} = \frac{0}{4} = 0\text{ kg}$

Average of Modulus of Deviations from Central Value $= \frac{\sum|x-\bar{x}|}{N}$

$= \frac{|-6|+|-2|+|1|+|7|}{4} = \frac{16}{4} = 4\text{ kg}$

✓ Average of Square of deviations from central Value $= \frac{\sum(x-\bar{x})^2}{N} = \frac{(-6)^2+(-2)^2+1^2+(7)^2}{4}$

i.e. Variance $= 22.5\text{ kg}^2$

(x) $S.D = +\sqrt{22.5\text{ kg}^2} = 4.75\text{ kg}$

Covariance → It measures the simultaneous variation of two R.V. X & Y & it is defined as,

$$\boxed{Cov(X,Y) = E(XY) - E(X).E(Y)}$$

Proof: $Cov(X,Y) = E\{(X-\bar{X}).(Y-\bar{Y})\} = \dfrac{\sum (X-\bar{X}).(Y-\bar{Y})}{N}$

$= \cdots\cdots\cdots = \boxed{E(XY) - E(X).E(Y)}$

Note: If X & Y are Independent R.V, then $\boxed{Cov(X,Y) = 0}$

eg After the age of 20 yrs, $Cov(Ht, Age) = 0$

while in case of wt & Age, $Cov(wt, Age) \neq 0$ throughout the life.

Some useful points –

(1) $\boxed{\text{Var or S.D} \geq 0}$ (T) (equality Holds in case of Constant Data set)

(∵ It Represents spread of data)

(2) Var or SD $\propto \dfrac{1}{\text{Consistency}}$ (T)

2022 (3) $\boxed{Cov(X, X) = Var(X)}$ (T)

Proof: $Cov(X, X) = E\left\{(X-\bar{X})(X-\bar{X})\right\} = E(X-\bar{X})^2 = \dfrac{\sum (X-\bar{X})^2}{N} = Var(X)$

ANALYSIS:

w.k.that $\bar{X} = \frac{\sum(X)}{N} = \sum p_i x_i = E(X)$

Similarly $Var(X) = \frac{\sum(X-\bar{X})^2}{N} = E(X-\bar{X})^2 = \cdots\cdots = \boxed{E(X^2) - E^2(X)}$

Again, $Cov(X,Y) = \frac{\sum(X-\bar{X})\cdot(Y-\bar{Y})}{N} = E\{(X-\bar{X})(Y-\bar{Y})\}$

$= - - - - - = \boxed{E(XY) - E(X)\cdot E(Y)}$

Some More Standard Results → Let X & Y are R.V & a, b, c are constants.

(i) $E(aX \pm bY \pm c) = aE(X) \pm bE(Y) \pm E(c)$

$= aE(X) \pm bE(Y) \pm c$

(ii) $Var(aX + b) = a^2 . Var(X) + Var(b)$

$= a^2 Var(X) + 0$

(iii) $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y) \pm 2ab\, Cov(X, Y)$ (w/o proof)

Q.8 : Find Mode & Median of following data

$2, 3, 3, \underbrace{4, 4, 4, 4, 4}, 5, 5, 5, 6, 7, 7, 8, 8, 8, 9, 10, 10, 11, 12, 13$

By observation, MODE = 4

Now $N = 23$ so $Md = \left(\frac{23+1}{2}\right)^{th} = 12^{th}$ observation $= 6$ Ans.

and $\text{Mean}(\bar{x}) = \frac{\sum x}{N} = \frac{2 + 2\times3 + 5\times4 + 3\times5 + 6 + 2\times7 + 3\times8 + 9 + 2\times10 + 11 + 12 + 13}{23}$

$= \frac{152}{23} = 6.608$

(PYQ) Marks obtained by 100 students in a test is shown in the following Table then find Me, Md & Mo of Marks obtained?

Marks (X)	No. of students (N)
25	20
30	20
35	40
40	20

$\Sigma X = 130$? $N = 100$

$$\underbrace{25, 25 \ldots, 25}_{20}, \underbrace{30, 30, \ldots 30}_{20}, \underbrace{35, 35 \ldots 35}_{40}, \underbrace{40, 40 \ldots 40}_{20}$$

Mode = 35 Marks

$$Md = \frac{\left(\frac{100}{2}\right)^{th} + \left(\frac{100}{2}+1\right)^{th}}{2} = \frac{50^{th} + 51^{st}}{2} = \cancel{50.5}$$

$$= \frac{35+35}{2} = 35$$

$$\text{Mean} = \frac{\Sigma X}{N} = \frac{20 \times 25 + 20 \times 30 + 40 \times 35 + 20 \times 40}{100} = \frac{3300}{100}$$

= 33 Marks

(M-I) $\cancel{\bar{X} = \frac{\Sigma X}{N} = \frac{130}{100}} = 1.3$, (M-II) $\bar{X} = \cancel{\frac{\Sigma X}{N} = \frac{130}{4}} = 32.5$

Aptitude
2014

Which of the following Batsman is most Consistent

Batsman	AV	S.D
K	65.2	5.79
L	43.7	4.75
M	54	6.21
N	58.3	5.11

useless Info.

$\because$ Consistency $\propto \frac{1}{SD}$

Ans = L.

Q. If the Difference b/w Expected Value of the square of Random Variable & square of the Expected Value is given as R then

(a) $R = 0$

(b) $R > 0$

(c) $R < 0$

(d) $R \geq 0$

Let X is the Random Variable. then

ATQ,

$$R = E(X^2) - (E(X))^2$$

$$= Var(X)$$

Q. If X & Y are two Ind Random Variables then which one is False?

(a) $Cov(X,Y) = 0$ (T)

(b) $E(XY) = E(X) \cdot E(Y)$ (T)

$$\left\{ \begin{aligned} \because Cov(X,Y) &= E(XY) - E(X) \cdot E(Y) \\ 0 &= E(XY) - E(X) \cdot E(Y) \end{aligned} \right\}$$

(c) $E(X^2Y^2) = E^2(X) \cdot E^2(Y)$ (F) while Correct version is

$$E(X^2Y^2) = E(X^2) \cdot E(Y^2)$$

(d) $Var(X-Y) = Var(X) + Var(Y)$ (T)

$$\because Var(aX - bY) = a^2 Var(X) + b^2 Var(Y) - 2ab\,Cov(X,Y)$$

$$Var(X-Y) = Var(X) + Var(Y) - 2(1)(1)\,Cov(X,Y)$$

11. Que if x & y are two Zero Mean Ind Random Variables having Variances $\frac{1}{4}$ & $\frac{1}{9}$ resp then find Mean & Variance of $(2x-3y)$?

(a) Mean = 0, Var = 4

~~(b)~~ Mean = 0, Var = 2

(c) Mean = 0, Var = $\sqrt{2}$

(d) Mean = −1, Var = 2

ATQ, $E(X) = E(Y) = 0$, $Var(X) = \frac{1}{4}$, $Var(Y) = \frac{1}{9}$, $Cov(X,Y) = 0$

Let $Z = 2X - 3Y$

$$E(Z) = E(2X-3Y) = 2E(X) - 3E(Y) = 2(0) - 3(0) = 0$$

Now, $Var(Z) = Var(2X-3Y)$

$$= 4\,Var(X) + 9\,Var(Y) + 2(2)(-3)\,Cov(X,Y)$$

$$= 4\left(\frac{1}{4}\right) + 9\left(\frac{1}{9}\right) + 0 = 2$$

Q. If mean & Variance of R.V x is given as μ & σ^2 resp then then mean & Variance of $\frac{x-\mu}{\sigma}$ are respectively ?

(a) $\{0,1\}$

(b) $\{0,\sigma\}$

(c) $\{\mu,\sigma\}$ ×

(d) $\{\mu,\sigma^2\}$ ×

ATQ, $E(x)=\mu$ & $Var(x)=\sigma^2$.

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

$$E(Z)=E\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sigma}E(x-\mu)=\frac{1}{\sigma}\left[E(x)-E(\mu)\right]=\frac{1}{\sigma}\left[\mu-\mu\right]=0$$

$$Var(Z)=Var\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sigma^2}Var(x-\mu)=\frac{1}{\sigma^2}\left[Var(x)+Var(-\mu)\right]$$

$$=\frac{1}{\sigma^2}\left[\sigma^2+0\right]=1.$$

If X and Y are random variable such that $E[2X + Y] = 0$ and $E[X + 2Y] = 33$, then

$$E[X] + E[Y] = \_\_\_\_\_.$$

M-II (1) + (2)

$3E(X) + 3E(Y) = 33$

so $E(X) + E(Y) = 11$ Ans

$2E(X) + E(Y) = 0$ — (1)

$E(X) + 2E(Y) = 33$ — (2)

$2E(X) + 4E(Y) = 66$ — (2)

$2E(X) + E(Y) = 0$

$\underline{2E(X) + 4E(Y) = 66}$

$-3E(Y) = -66$

$E(Y) = 22$ & $E(X) = 33 - 44 = -11$

Hence $E(X) + E(Y) = -11 + 22 = 11$

The following sequence of numbers is arranged in increasing order : 1, x, x, x, y, y, 9,16,18. Given that the mean and median are equal, and are also equal to twice the mode, the value of y is

(a) 5

(b) 6

(c) 7

(d) 8

By observation, Mode $= x$

$\because N = 9$ so

$Md = \left(\frac{9+1}{2}\right)^{th} = 5^{th}$ observation

ie $Md = y$

Let Mean $= \alpha$

ATQ, Mean $=$ Md $\Rightarrow \alpha = y$

& Mean $= 2$ Mode

$\alpha = 2x$

or $y = 2x$ — (1)

Now $\bar{X} = \frac{\Sigma X}{N} = \frac{1+x+x+x+y+y+9+16+18}{9}$

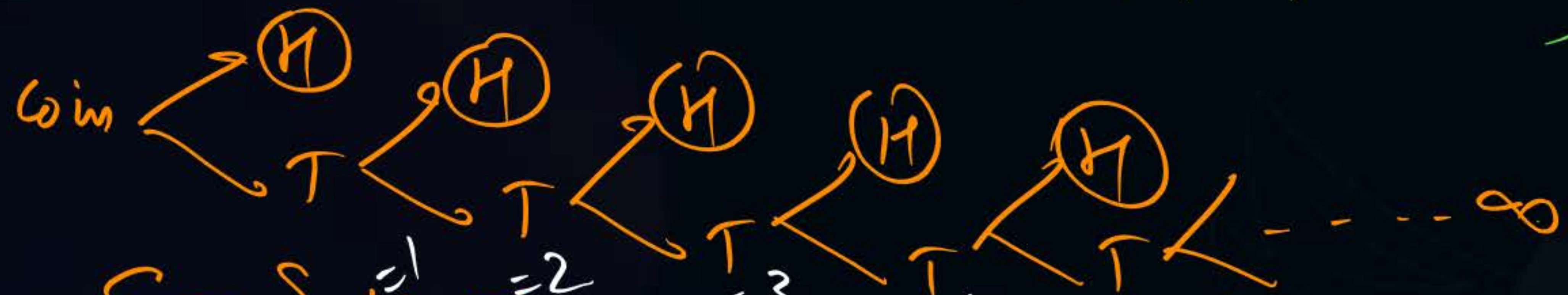
$y = \frac{3x+2y+44}{9} \Rightarrow 3x - 7y = -44$ — (2)

Solving x & y we get $y = 8$ and $x = 4$

Probability Dist — The Table representing Distribution of probabilities is called Prob Distribution.

eg A coin is tossed until H appears then find Prob Dist of No. of tosses.

Sol: $X = \{\text{Number of tosses}\} = \{1, 2, 3, 4, \ldots\ldots\}$



$S = \{\overset{=1}{H}, \overset{=2}{TH}, \overset{=3}{TTH}, \overset{=4}{TTTH}, \ldots\ldots\}$

Prob Dist:

X:	1	2	3	4	5	...... ∞
$P(X)$:	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	......

CROSS Check:-

$$\sum p_i = \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\dots\dots\infty$$

$$= \frac{1/2}{1-\frac{1}{2}} = \frac{1/2}{1/2} = 1$$

Hence verified.

Q: A coin is tossed until Head appears then find the Average No. of tosses Req.?

Sol: $X = \{\text{No. of tosses}\} = \{1, 2, 3, 4 \dots\}$

X:	1	2	3	4	-------
$P(X)$:	1/2	1/4	1/8	1/16	-------

$$E(X) = \sum p_i x_i = p_1x_1 + p_2x_2 + p_3x_3 + \dots\dots$$

$$= \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \dots\dots$$

$$= \frac{1}{2}\left[1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{2}\right)^2 + 4\left(\frac{1}{2}\right)^3 + \dots\dots\right]$$

$$\boxed{1 + 2x + 3x^2 + 4x^3 + \dots\dots = (1-x)^{-2}}$$

$$= \frac{1}{2}\left[\left(1-\frac{1}{2}\right)^{-2}\right] = \frac{1}{2}\left[\left(\frac{1}{2}\right)^{-2}\right] = \frac{2^2}{2} = 2$$

Average No. of tosses $= 2$

Note: Minimum Tosses Required $= 1$

Maximum " " $=$ No idea (∞)

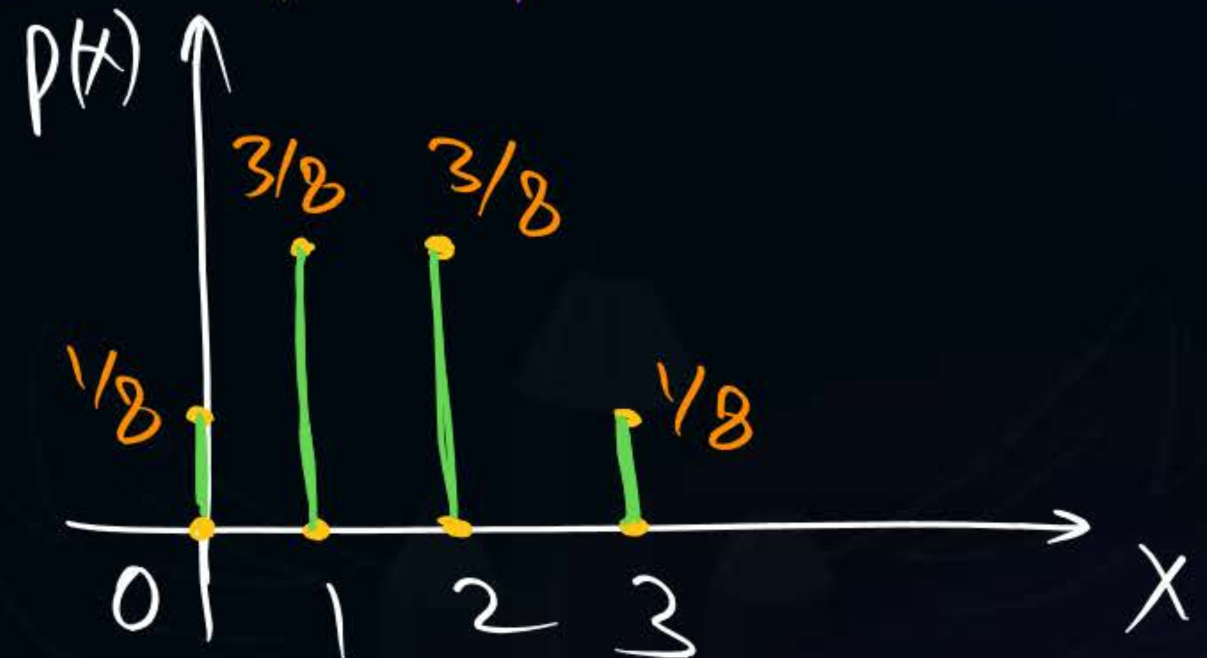
eg: A coin is tossed thrice then find the prob Dist of Number of Heads?

Sol: $X = \{\text{Number of Heads}\} = \{0, 1, 2, 3\}$

$$S = \{\underbrace{(HHH)}_{3H}, \underbrace{(HHT)}_{2H}, \underbrace{(HTH)}_{2H}, \underbrace{(HTT)}_{1H}, \underbrace{(THH)}_{2H}, \underbrace{(THT)}_{1H}, \underbrace{(TTH)}_{1H}, \underbrace{(TTT)}_{0H}\} = 8 \text{ Triplets}$$

Prob Dist:

X:	0	1	2	3
$P(X)$:	$1/8$	$3/8$	$3/8$	$1/8$



Combined Height of these Bars $= \sum p_i = 1$

(ii) Also find the Av. No. of tosses Req.

Q: A coin is tossed thrice then Find Mean, Variance & S.D of Numbers of Heads.

Sol: $X = \{\text{No. of tosses}\} = \{0, 1, 2, 3\}$

X:	0	1	2	3
$P(X)$:	$\frac{1}{8}$	$3/8$	$3/8$	$1/8$

$$E(X) = \sum p_i x_i$$

$$= p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4$$

$$= \frac{1}{8}(0) + \frac{3}{8}(1) + \frac{3}{8}(2) + \frac{1}{8}(3)$$

$$= \frac{12}{8} = \frac{3}{2} = 1.5$$

Average No. of Heads $= 1.5$

Now, $E(X^2) = \sum p_i x_i^2$

$$= p_1x_1^2 + p_2x_2^2 + p_3x_3^2 + p_4x_4^2$$

$$= \frac{1}{8}(0)^2 + \frac{3}{8}(1)^2 + \frac{3}{8}(2)^2 + \frac{1}{8}(3)^2$$

$$= 3$$

So $\text{Var}(X) = E(X^2) - (E(X))^2$

$$= 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} = 0.75$$

$$SD(\sigma) = +\sqrt{\text{Var}(X)} = +\frac{\sqrt{3}}{2} = 0.86$$

ANALYSIS: (1) Average w.r.to 1 = Prob & Av w.r.to 100 = %

A coin is tossed twice then Average No. of Heads = $\boxed{1.5}$, Variance = 0.75, SD = $\frac{\sqrt{3}}{2}$

" " once, then Average No. of Heads = 0.5, Var = 0.25, SD = $\frac{\sqrt{3}/2}{\sqrt{3}}$ = 0.5

" " 10000 times, Av No. of Heads = 5000, Var = 2500, SD = $0.5 \times \sqrt{10000}$ = 50

ie $\mu = 5000$, $\sigma = 50$

$\mu - 3\sigma = 5000 - 150 = 4850$

$\mu + 3\sigma = 5000 + 150 = 5150$

or we have very good chance (99.7% chance) of having No. of Heads lying in b/w (4850, 5150) ie

$$\boxed{4850 \leq \text{No of Heads} \leq 5150}$$

Q. A Coin is tossed until Head appears or Tail appears 4 times in succession, then find the Average Number of tosses Required

(a) 2.50

(b) 1.87

(c) 4

(d) 3.89

Syllabus

① Linear Algebra	✓
② Calculus	✓
③ Prob & Stats	✓

Don't Judge yourself before solving P.Y.Q.

Tel: drpuneetsirpw

Strategy:-

1. Live Class.
2. Revision
3. Short Notes (in Later Phase)
4. D.P.P
5. Chapterwise test (Sund)
6. P.Y.Q. → Judge
7. O.T.S → "

Book: → No Book is Needed only PYQ Book is required.

eg: L-Algebra: Class (150-200Q) | DPP (70-90) | WT (13Q) | OTS 100Q | PYQ (200Q) | ≈ 700-800 Q

Doubts: → Conceptual Doubts → you can ask anytime.
→ Generic Doubt → will be discussed after 9:30 AM

Parachute Landing → Conceptual Doubts are also not allowed.

PREREQUISITE of Engg Maths: → ✓ (25 Lectures)
(Foundation Series of Engg Maths)

* Engg Maths   CS / IT   DS / AI

$\frac{10M}{100}$ $\left(\frac{40-45}{100}\right)$

* **Maths:** is the Language of Engg.

* Maths → Language → understanding of Symbols.

Maths → Concept:
- → Information Based Concept (No Brain should be used)
- → Analysis Based Concept (Pure Maths) (useful to develop Neurons)
- → Application Based Concept. (Applied Maths) in LIVE Class.

* M.M. Imp Point → Try to have Patience as much as possible

Telegram: $\boxed{\text{drpuneet sir pw}}$

Thank

YOU