DS & AI
CS & IT

Probability and Statistics

Lecture No. 15

By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic

NORMAL DISTRIBUTION

(Part 1)

Topics to be Covered

Topic

1. NORMAL DISTRIBUTION (Part 2)
2. Curve fitting
3. Correlation – Regression

Q: In a school of 1000 students, Average height of student is 68.22 inches & Variance 10.8 (inches)^2 then find the Number of students over 6 feet tall If we have a confidence of 37.46% with in the limits $z = 0$ to 1.15

(HW)

(a) 875

(b) 125

(c) 100.

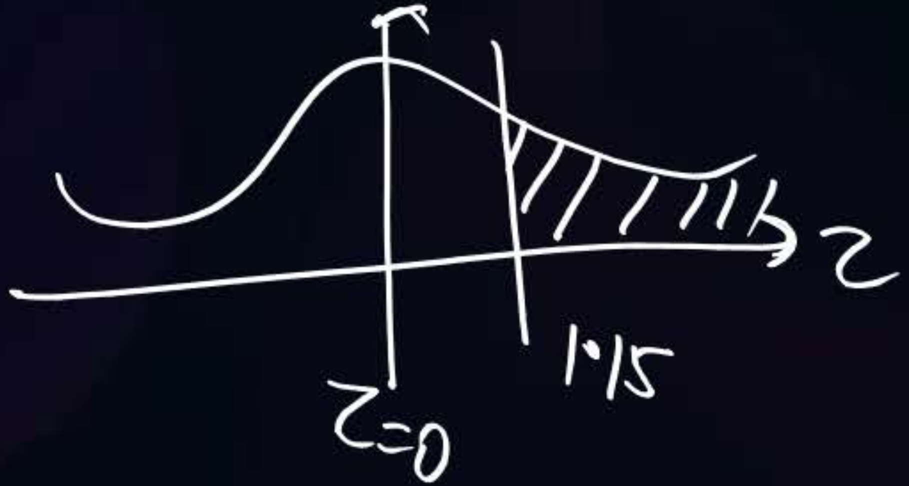
(d) 72

Sol: $N = 1000$, for single student: $x = \{\text{Height of this single student}\}$

$$\mu = 68.22 \text{ inches}, \quad \sigma = \sqrt{10.8} \text{ inches}, \quad z = \frac{x-\mu}{\sigma} = \frac{72 - 68.22}{\sqrt{10.8}} = 1.15$$

$$P(x > 72) = P(z > 1.15) = (\text{Right side area}) - P(0 \le z \le 1.15)$$

$$= 0.5 - 0.3746 = 0.1254 = \frac{125.4}{1000} \approx \frac{125}{1000}$$



HW8 if x is Zero Mean, unit Variance $\boxed{\text{Gaussian Variable}}$ then find $E(|x|) = ?$

(a) 0

(b) 0.5

(c) 0.8

(d) 0.4

ATQ, $x \sim N\{\mu, \sigma^2\}$

$\Rightarrow x \sim N\{0, 1\}$

No Need to Convert x into z

$\because$ it is already S.N.V.

& it's p.d.f is

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} = \text{Even func}^n$$

$$E\{|x|\} = \int_{-\infty}^{\infty} \underbrace{|x| f(x)}_{\text{Even.}} dx = 2\int_{0}^{\infty} |x| f(x) dx$$

$$= 2\int_{x=0}^{\infty} (+x) \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot dx$$

$$= \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} x \cdot e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} e^{-y} dy$$

Put $\frac{x^2}{2} = y$

$x\,dx = dy$

$$= \sqrt{\frac{2}{\pi}} \left[\frac{e^{-y}}{-1}\right]_{y=0}^{\infty} = \sqrt{\frac{2}{\pi}}(0-1)$$

$$= \sqrt{\frac{2}{\pi}} = 0.8 \text{ Ans}$$

Q. Evaluate $\int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{8}\right\} dx = ?$

(M-I) Using Gamma func^n → Do yourself.

(M-II) Using N Dist : → w.k. that p.d.f of N.D is given as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$

& Area under $f(x) = 1$

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

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = 1$$

If we take $\mu = 0$, $\sigma = 2$ then

$$\int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-0}{2}\right)^2\right\} dx = 1$$

$$\int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{x^2}{8}\right\} dx = 1$$

$$2\int_{0}^{\infty} \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{x^2}{8}\right\} dx = 1$$

$$\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{8}\right\} dx = 1$$

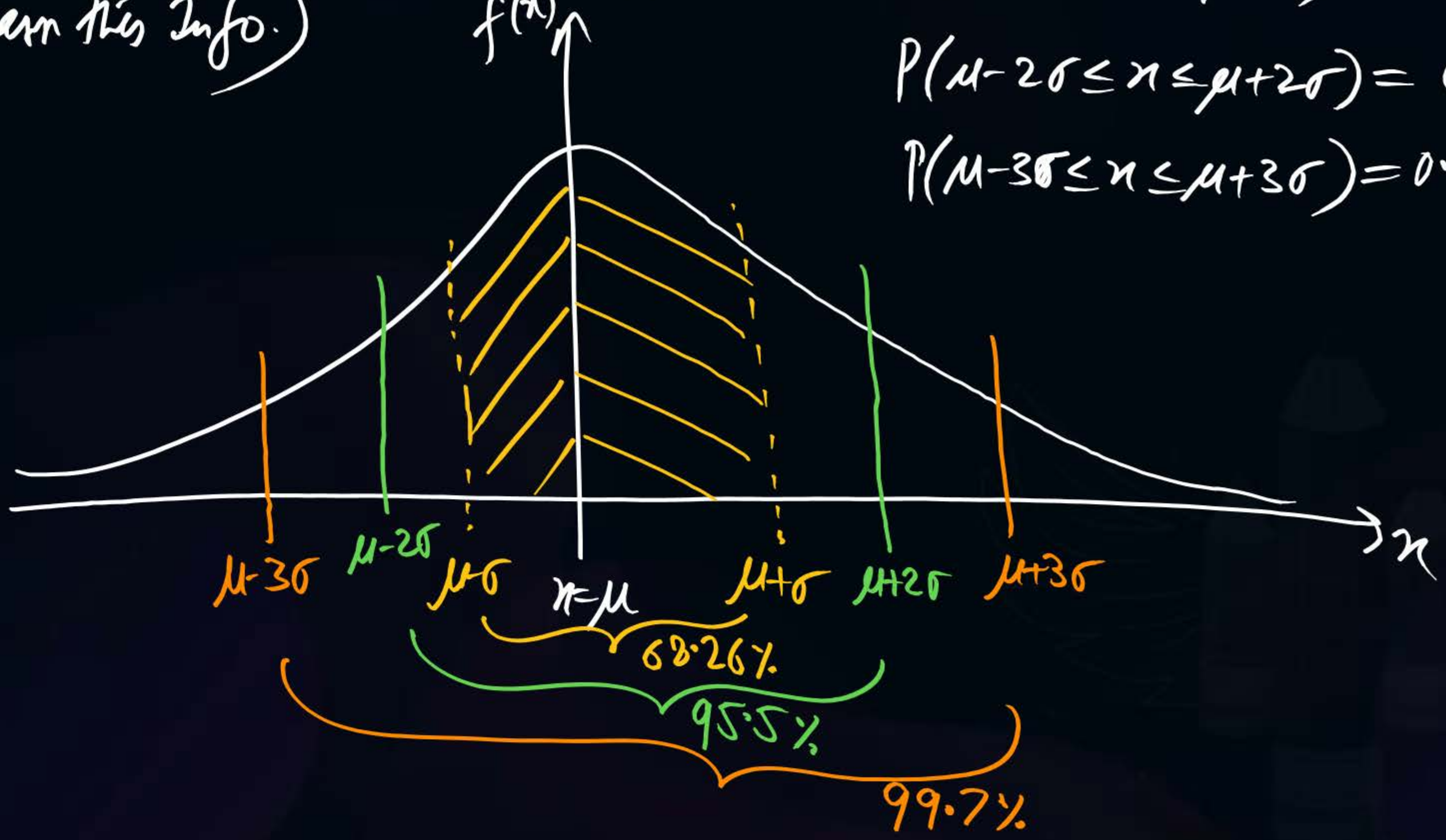
Some More theoretical Concepts of N. Dist →

(Learn this Info.)

$P(\mu-\sigma \le x \le \mu+\sigma) = 0.6826$

$P(\mu-2\sigma \le x \le \mu+2\sigma) = 0.955$

$P(\mu-3\sigma \le x \le \mu+3\sigma) = 0.997$



Note

(1)

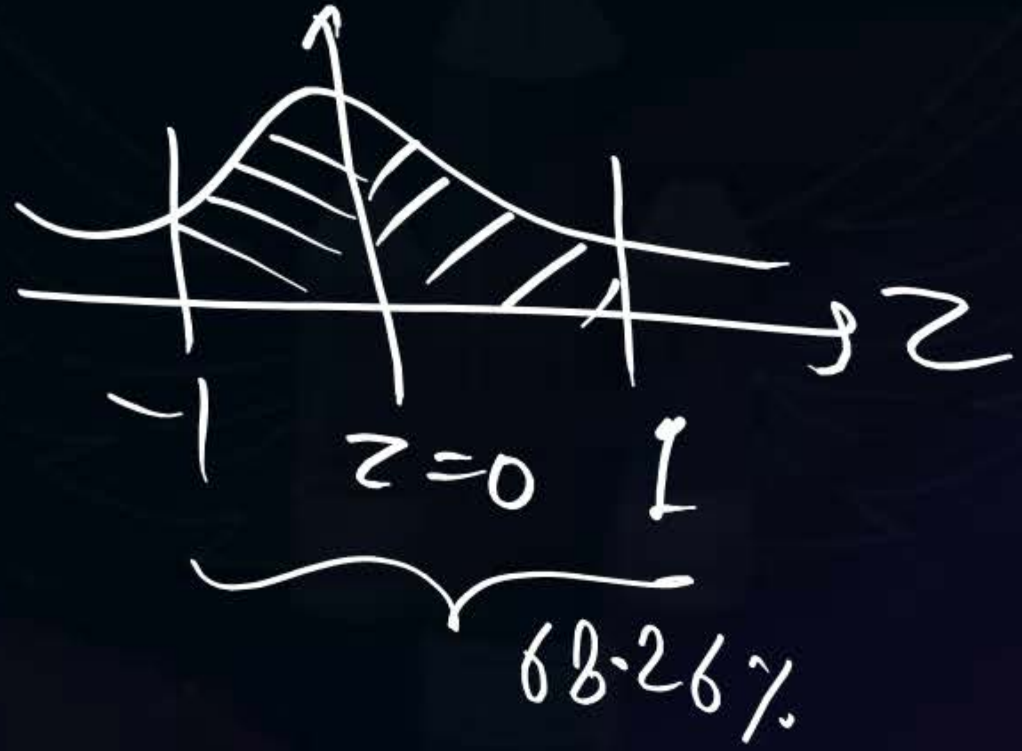
$$\left.\begin{aligned} P(\mu-\sigma \le x \le \mu+\sigma) &= 0.6826 \\ P(\mu-2\sigma \le x \le \mu+2\sigma) &= 0.955 \\ P(\mu-3\sigma \le x \le \mu+3\sigma) &= 0.997 \end{aligned}\right\} \xrightarrow[x \approx Z]{\mu=0,\ \sigma=1} \begin{aligned} P(-1 \le Z \le 1) &= 0.6826 \\ P(-2 \le Z \le 2) &= 0.955 \\ P(-3 \le Z \le 3) &= 0.997 \end{aligned}$$

(2) **Gate**: Evaluate $P(0 \le Z \le 1) = ?$ where Notations have their usual meaning.

Sol: w.k.that $P(-1 \le Z \le 1) = 0.6826$

$$2P(0 \le Z \le 1) = 0.6826$$

$$P(0 \le Z \le 1) = 0.3413 \quad \text{Ans}$$



Q: if $x \sim N\{102, 27^2\}$ then evaluate $P(90 \le x \le 102) = ?$

(a) × 68%

(b) × 34%

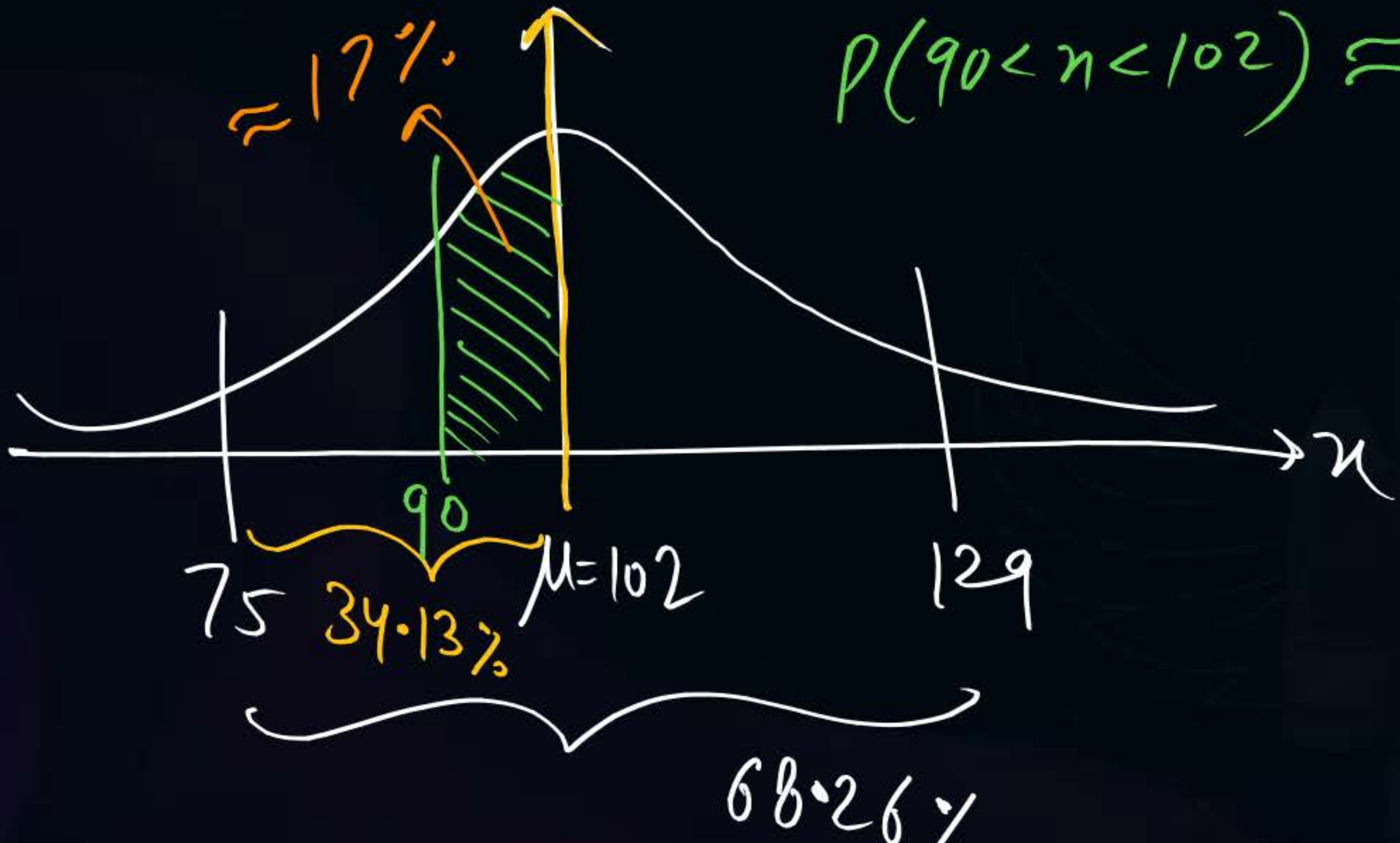
(c) × 50%

(d) 16.7%

ATQ, $\mu = 102$, $\sigma = 27$

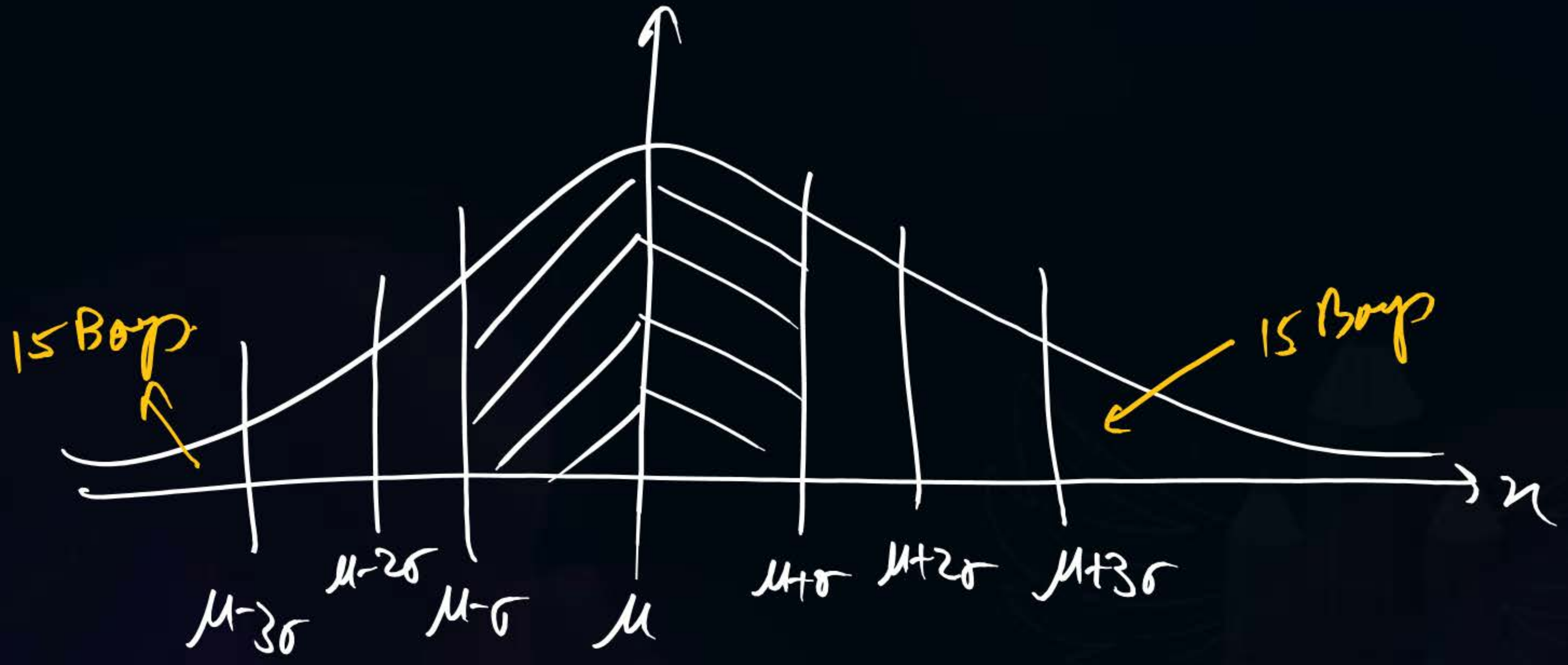
$\mu - \sigma = 102 - 27 = 75$

$\mu + \sigma = 102 + 29 = 129$



$P(90 < x < 102) \approx 17\%$

PODCAST: $N = 10000$ persons, $x = \{$Habits of Boys$\}$

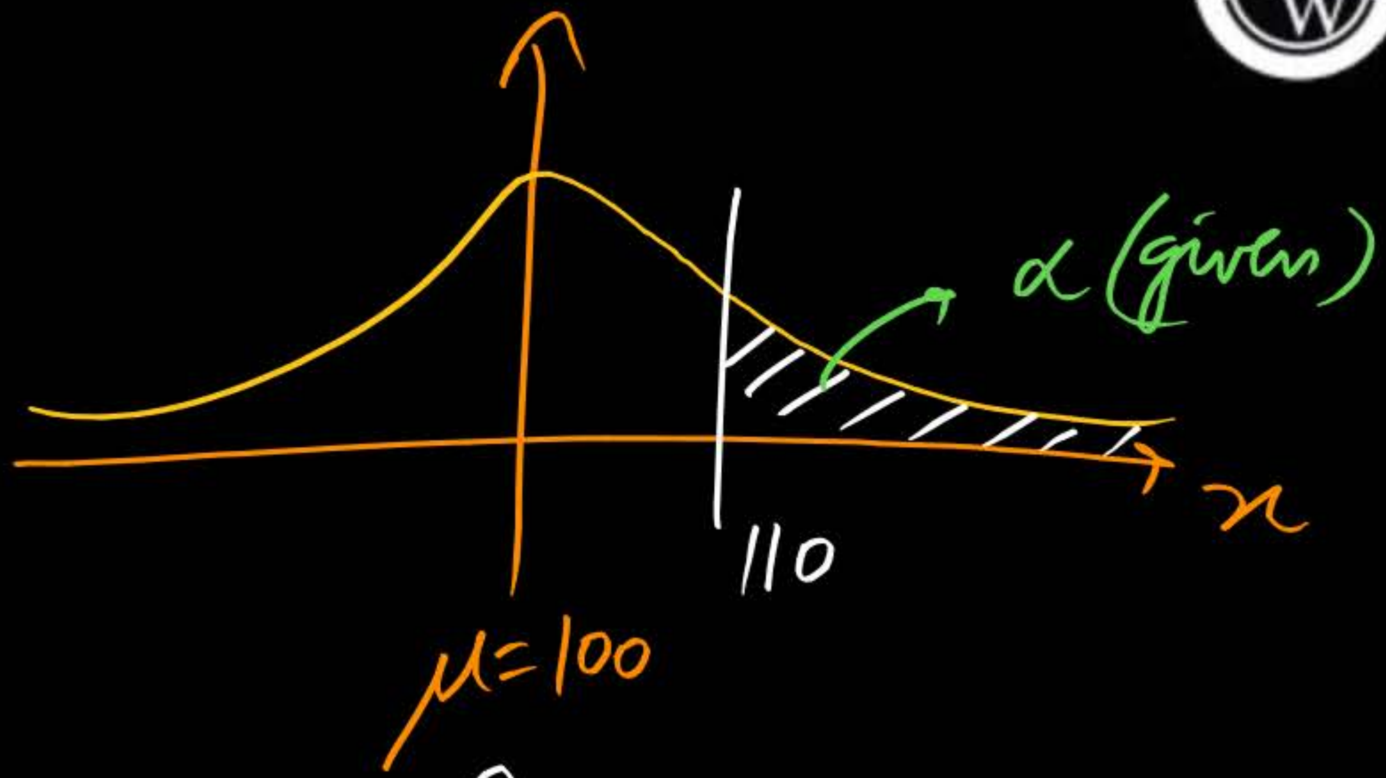


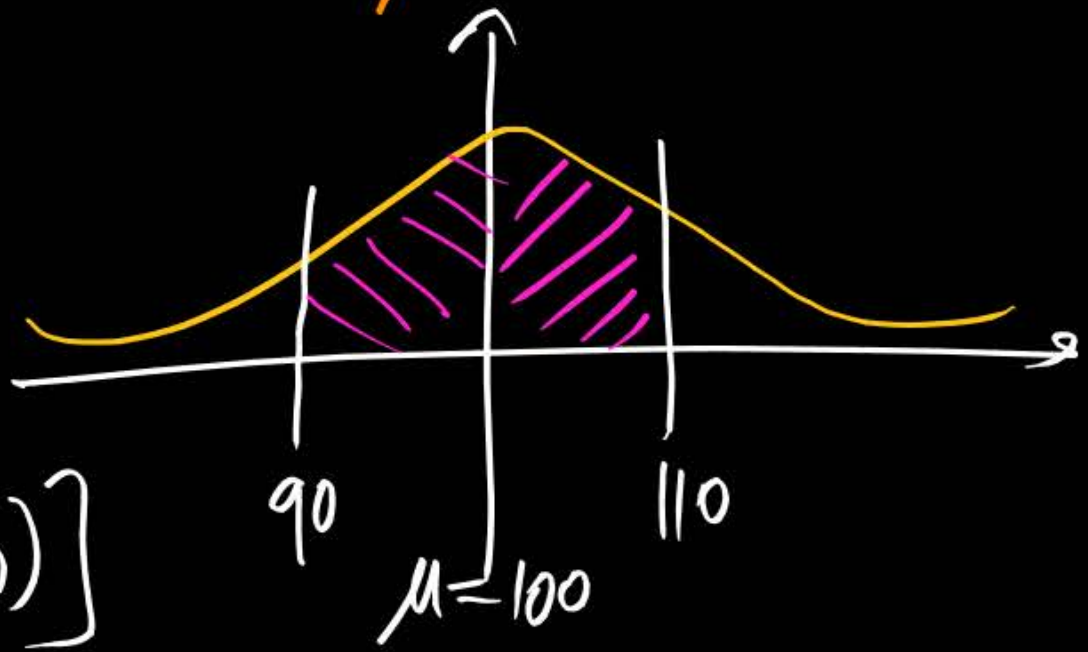
$99.7\% = 9970$ Boys.

For a random variable $x(-\infty < x < \infty)$ following normal distribution, the mean is $\mu = 100$. If the probability is $\boxed{P = \alpha \text{ for } x \geq 110}$. Then the probability of x lying between 90 and 110 i.e. $P(90 \leq x \leq 110)$ and equal to

(a) $1-2\alpha$ (b) $1-\alpha$

(c) $1-\alpha/2$ (d) 2α





$P(90 \le x \le 110) = 2P(100 \le x \le 110)$

$= 2[\text{Right side area} - P(x > 110)]$

$= 2[0.5 - \alpha] = 1 - 2\alpha$

PYQ

The area (in percentage) under standard normal distribution curve of random variable Z within limits from -3 to $+3$ is ____. 99.7%

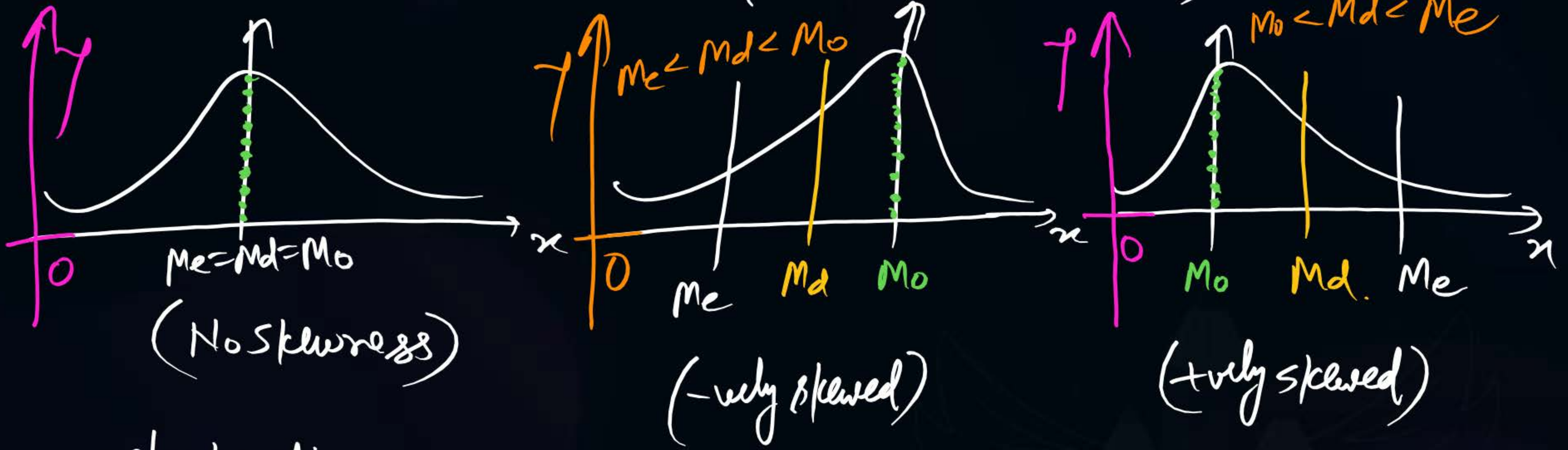
(a) 0.997

(b) 99.7

(c) 6

(d) 3

Some More Info for N Curve → (Skewness in N Curve) →



-vely skewed :- when Tail is occurring more in the LHS then it is called -vely skewed.

+vely Skewed :- " " " more " RHS " " +vely "

Which one of the following statements is not true?

(a) The measure of skewness is dependent upon the amount of dispersion (T)

(b) In a symmetric distribution, the values of mean, mode and median are the same (T)

(c) In a positively skewed distribution, mean > median > mode $M_o < M_d < M_e$ (T)

(d) In a negatively skewed distribution, mode > mean > median $M_d < M_e < M_o$

??

NORMAL SUM Theorem → If x_1 & x_2 are Independent, $\boxed{\text{N.R.V}}$ with mean μ_1, μ_2 & Var σ_1^2, σ_2^2 resp. then the variable defined as $c_1 x_1 \pm c_2 x_2$ (c_1, c_2 are const) is also N.R.V with mean $(c_1\mu_1 \pm c_2\mu_2)$ & Var. $(c_1^2\sigma_1^2 + c_2^2\sigma_2^2)$

OR.

ie $x_1 \sim N\{\mu_1, \sigma_1^2\}$
$x_2 \sim N\{\mu_2, \sigma_2^2\}$
then $\boxed{x = c_1 x_1 \pm c_2 x_2}$ is also $\boxed{\text{N.R.V}}$ defined as.

Here x_1 & x_2 are Ind

$$x \sim N\left\{(c_1\mu_1 \pm c_2\mu_2), (c_1^2\sigma_1^2 + c_2^2\sigma_2^2)\right\}$$

Note → If x_1, x_2 are $\boxed{\text{Not N.R.V}}$ still x will be N.R.V. (By C.L.Th.)

Eg: if $x_1 \sim N\{\overset{\mu}{2}, \overset{\sigma^2}{9}\}$, $x_2 \sim N\{\overset{\mu}{3}, \overset{\sigma^2}{16}\}$ s.t. x_1 & x_2 are Ind then find Mean & Variance of $\boxed{x = 2x_1 - 5x_2}$

M-I By N.S. Th. i.e. $\boxed{x = 2x_1 - 5x_2}$

$\mu_1 = 2, \mu_2 = 3$

So $\mu_x = 2\mu_1 - 5\mu_2 = 2(2) - 5(3) = -11$

Now $\sigma_1^2 = 9, \sigma_2^2 = 16$ then

$$\sigma_x^2 = 4\sigma_1^2 + 25\sigma_2^2$$

$$= 4(9) + 25(16) = 436$$

M-II Using Properties of Mean & Variance

$E(x_1) = 2, E(x_2) = 3, Cov(x_1, x_2) = 0$

So $E(x) = E(2x_1 - 5x_2) = 2E(x_1) - 5E(x_2)$

$\text{Mean}(x) = 2(2) - 5(3) = -11$

$\because Var(x_1) = 9, Var(x_2) = 16.$

So $Var(x) = Var(2x_1 - 5x_2)$

$$= 4Var(x_1) + 25Var(x_2) + 2(2)(-5)Cov(x_1, x_2)$$

$$= 4(9) + 25(16) - 20(0) = 436$$

NWQ If u & v are two Independent, Zero Mean Gaussian Variables having variances $\frac{1}{4}$ & $\frac{1}{9}$ resp then $P(3v \geq 2u) = ?$

(a) 0

(b) 0.25

(c) 0.5

(d) 0.75

$$P[3v \geq 2u] = P[(3v-2u) \geq 0] = P[x \geq 0] = 0.5 \text{ Ans}$$

let $\boxed{x = 3v - 2u}$

$$\mu_x = 3\mu_v - 2\mu_u = 3(0) - 2(0) = 0$$

$$\sigma_x^2 = 9\sigma_v^2 + 4\sigma_u^2 = 9\left(\frac{1}{9}\right) + 4\left(\frac{1}{4}\right) = 2$$

i.e. $x \sim N\{\mu_x, \sigma_x^2\}$

i.e. $x \sim N\{0, 2\}$

$x = 0$

Let X_1, X_2 and X_3 be independent and identically distributed random variables with the uniform distribution on [0, 1]. The probability $P\{X_1 + X_2 \le X_3\}$ is ______

(a) 0.1587 (b) 0.21

(c) 0.5 (d) 0.3413

$$P(x_1+x_2 \le x_3) = P(x_1+x_2-x_3 \le 0) = P(x \le 0) = ?$$

Let $\boxed{x = x_1 + x_2 - x_3}$

$\because x_1, x_2, x_3$ are U.R.V in $[0,1]$

So $E(x_1) = E(x_2) = E(x_3) = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}$

$E(x) = E(x_1+x_2-x_3) = \frac{1}{2}+\frac{1}{2}-\frac{1}{2} = \frac{1}{2}$

$\because x_1, x_2, x_3$ are U.R.V in $[0,1]$

So $Var(x_1) = Var(x_2) = Var(x_3) = \frac{(b-a)^2}{12}$

$= \frac{(1-0)^2}{12} = \frac{1}{12}$

So $Var(x) = Var(x_1 + x_2 - x_3)$

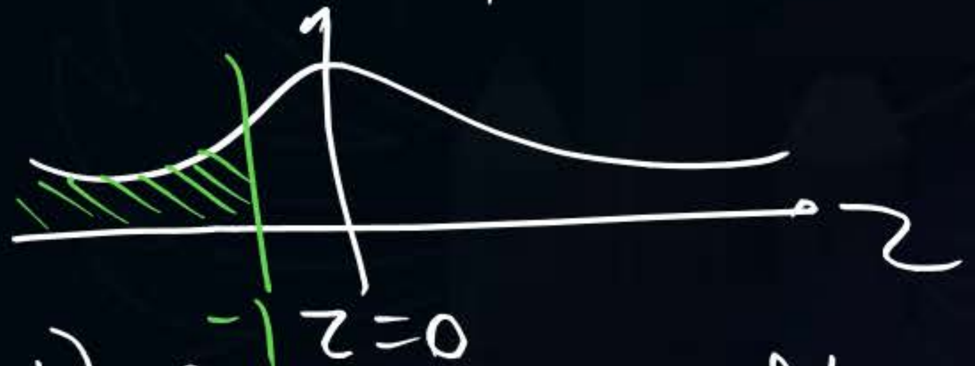
$= Var(x_1) + Var(x_2) + Var(x_3)$ $\{\because x_1, x_2, x_3$ are Ind$\}$

$= \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{4}$

ie $\mu_x = \frac{1}{2}$, $\sigma_x^2 = \frac{1}{4}$ & $\boxed{x \sim N\{\mu, \sigma_x^2\}}$ $\{\because x_1, x_2, x_3$ are iid$\}$

or $x \sim N\{\frac{1}{2}, \frac{1}{4}\}$

$$Z = \frac{x - \mu_x}{\sigma_x} = \frac{0 - 1/2}{1/2} = -1.$$



So $P(x_1 + x_2 \le x_3) = P(x \le 0) = P(Z \le -1) = P(Z \ge 1) = 0.5 - P(0 \le Z \le 1)$

$= 0.5 - 0.3413 = 0.1587$

Consider a binomial random variable X. If X_1, X_2, ... X_n are independent and identically distributed samples from the distribution of X with sum $\boxed{Y = \sum_{i=1}^{n} X_i}$, then the distribution of Y as $n \to \infty$ can be approximated as

(a) Exponential (b) Bernoulli

(c) Binomial (d) Normal

$\boxed{\text{i.i.d.}} \Rightarrow$ Normal

$y = (x_1 + x_2 + x_3 + \dots + x_n)$

when $n \to \infty$

Using Either by Normal Sum Theorem.

OR by C.L.Th.

CURVE FITTING

Case I: Line of Best Fit: Let us consider a set of values given as,

x_1	x_2	x_3	......	x_{n-1}	x_n
y_1	y_2	y_3	......	y_{n-1}	y_n

i.e. n points are given.

Let line of best fit is $\boxed{y = a + bx}$ — (1)

where a & b can be obtained by solving following eqn's;

$$\sum y = na + b\sum x$$

& $$\sum xy = a\sum x + b\sum x^2$$

} these two eqn's are called Normal eqn's for the line of best fit.

Method of least squares:-

Let $y = a + bx$ is the line of Best fit & $y = f(x)$ is the exact equ^n's.

At $x = x_i$, $\boxed{\text{Error} = \text{Approx Value} - \text{Exact Value}}$

$$e_i = y_i - f(x_i)$$

Let $u = e_i^2 = (y_i - f(x_i))^2$ —— (1)

for Minimum e_i, we must have,

$$\frac{\partial u}{\partial a} = 0 \ \& \ \frac{\partial u}{\partial b} = 0$$

To Develop Neurons.

- - - - - - - - - - - -

$\boxed{\Sigma y = na + b\Sigma x}$ & $\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$

Parabola of Best fit:- Consider the set of Data given as

x_1	x_2	x_3	...	x_n
y_1	y_2	y_3	...	y_n

Let the parabola that fits above data is as follows

$$\boxed{y=a+bx+cx^2} \longleftarrow (1)$$

Normal equ^s are;

$$\left.\begin{aligned}\sum y &= na+b\sum x+c\sum x^2\\ \sum xy &= a\sum x+b\sum x^2+c\sum x^3\\ \sum x^2y &= a\sum x^2+b\sum x^3+c\sum x^4\end{aligned}\right\}$$

after solving these 3 equ^s, we can find a, b, c

& Hence Parabola of Best fit can be obtained.

Q. Find the line of Best fit for the following Data ?

$x=$	1	2	3	4	5
$y=$	3	5	7	9	11

(M-I) (By observation); Line of Best fit is $\boxed{y = 1+2x}$ Ans

(M-II) Let the line of Best fit is $y = a+bx$ & Normal eqn's are;

x	y	xy	x^2
1	3	3	1
2	5	10	4
3	7	21	9
4	9	36	16
5	11	55	25
$\sum x = 15$	$\sum y = 35$	$\sum xy = 125$	$\sum x^2 = 55$

$n=5$

$$\left.\begin{aligned} \sum y &= na + b\sum x \\ \sum xy &= a\sum x + b\sum x^2 \end{aligned}\right\} \Rightarrow \begin{aligned} 35 &= 5a + 15x \\ 125 &= 15a + 55b \end{aligned}$$

$a=1,\ b=2$

So line of Best fit is $y = 1+2x$ Ans

(HW) Find the line of Best fit for

$x=$	1	2	3	4	5	6	7
$y=$	9	8	10	12	11	13	14

(M-I) By observation → Not possible

(M-II)

Ans: $y = \underset{a}{7.28} + \underset{b}{0.928}\,x$

(HW)

Three points in the x-y plane are (–1, 0.8), (0, 2.2) and (1, 2.8). The value of the slope of the best fit straight line in the least square sense is ________ (Round off to 2 decimal places).

(a) 2 (b) –1

?? (c) 1 (d) 1.93

x	y	xy	x^2
–1	0.8	–0.8	1
0	2.2	0	0
1	2.8	2.8	1
$\Sigma =$ 0	5.8	2	2

$\boxed{y = a + bx}$ — Line of Best fit.

N.L Equ^n are; $\left.\begin{array}{l}\Sigma y = na + b\Sigma x \\ \Sigma xy = a\Sigma x + b\Sigma x^2\end{array}\right\} \Rightarrow$

$5.8 = 3a + 0$

$2 = 0 + b(2)$

$b = 1$, $a = \frac{5.8}{3}$

Hence slope $= b = 1$ Ans

& line of Best fit is $y = x$

CORRELATION

Covariance: → it measures the simultaneous variation of two Random Variables x & y

Correlation: → it " " " " " " " " "

$$\boxed{Cov(x, y) = E(xy) - E(x).E(y)}$$

& Correlation coeff $\boxed{r = \dfrac{Cov(x, y)}{\sigma_x . \sigma_y}}$

i.e $-\infty < Cov(x, y) < \infty$

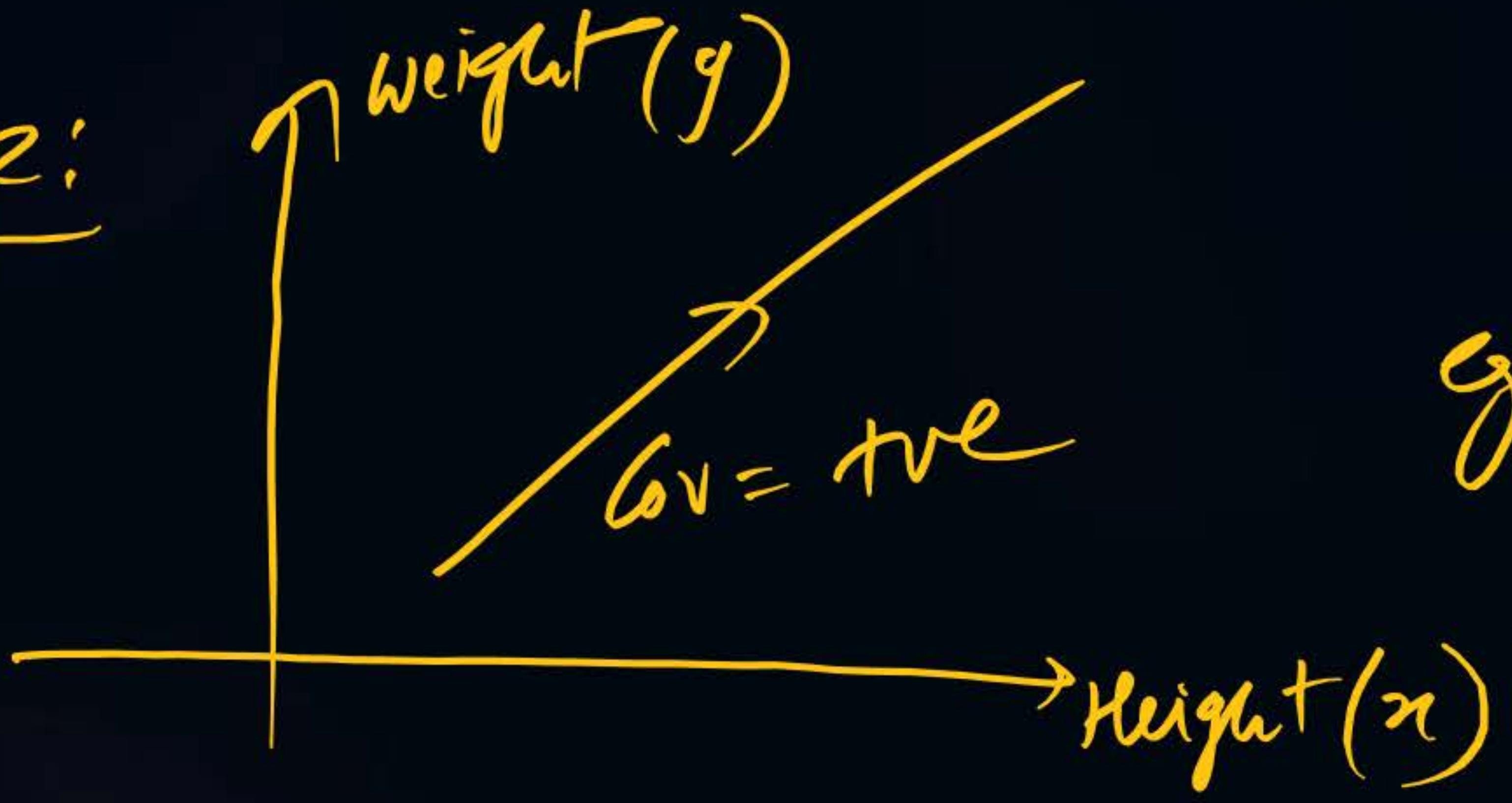
i.e $-1 \le r \le 1$

Note:
1. if $r = 1$ then x & y are perfectly correlated in +ve sense.
2. if $r = -1$ " " " " " -ve sense
3. if $r = 0$ then x & y are Not Related / Independent

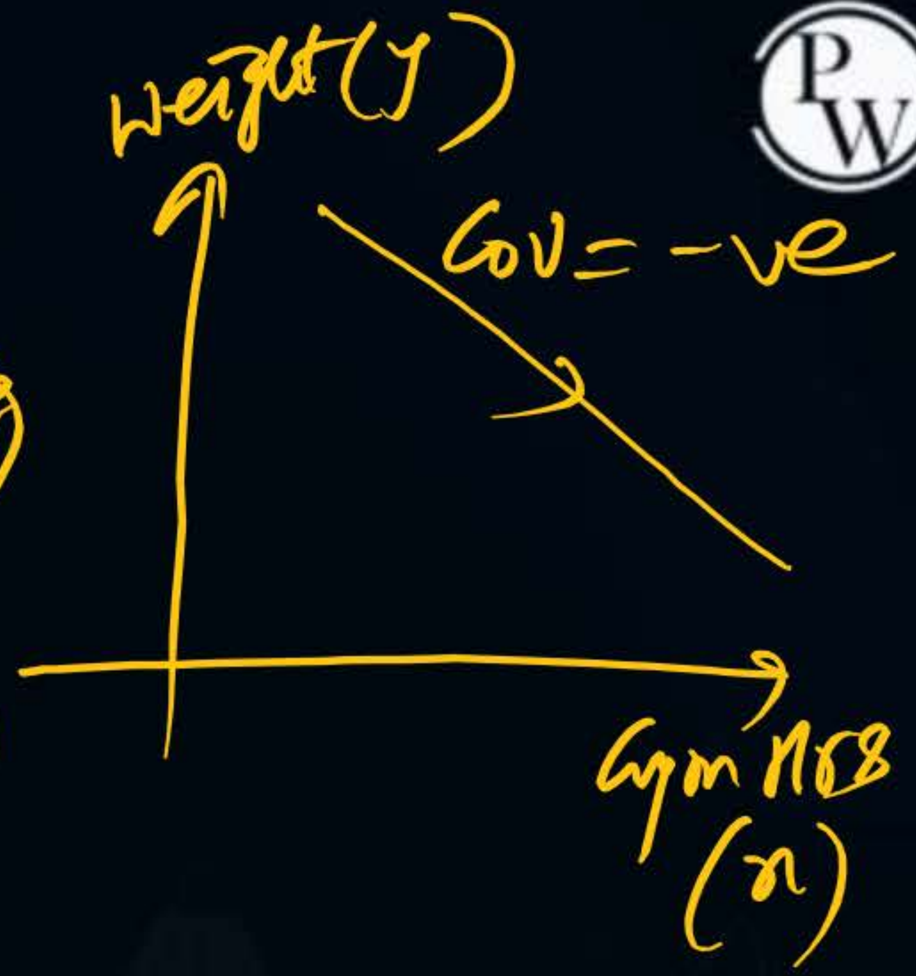
Covariance	Correlation
(1) It measures the direction of Relation	(1) It measures the dir.^{n} as well as the strength of Relationship.
(2) $-\infty < Cov(x,y) < \infty$	(2) $-1 \le r \le 1$
(3) $Cov(x,y) = E(xy) - E(x).E(y)$	(3) $r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$
(4) if x & y are Ind then $Cov(x,y) = 0$	(4) if x & y are Ind, $r = 0$
(5) In different units, Covariance is also different	(5) In different units, Correlation Coeff is same

Ef of Covariance:

eg



eg

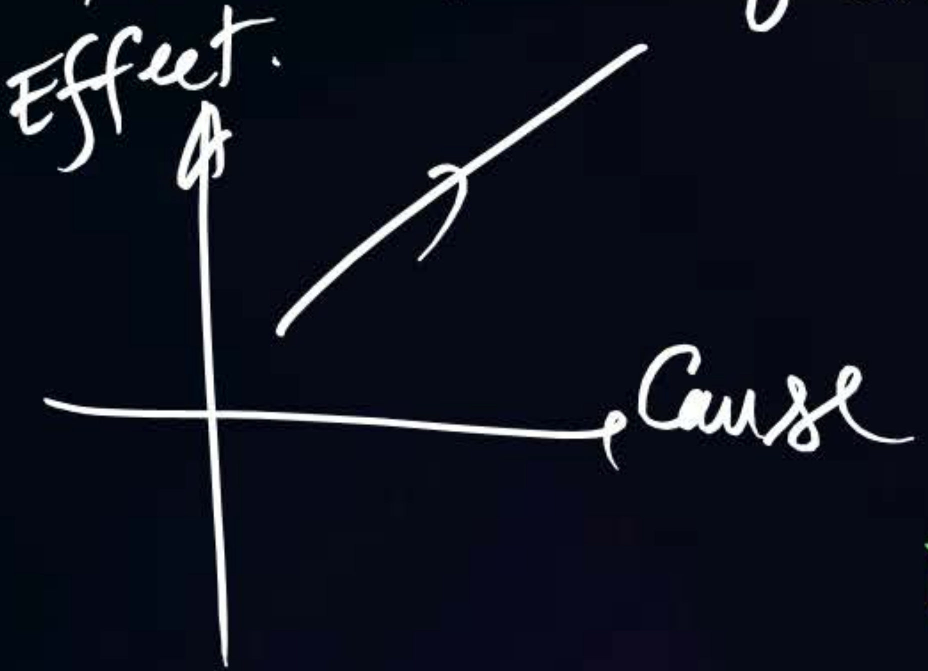


REGRESSION

eg If supply is constant then

- $\overset{(y)}{\text{Price}} \propto \overset{(x)}{\text{Demand}}$ (from shopkeeper point of view)
- $\underset{(x)}{\text{Demand}} \propto \dfrac{1}{\text{Price}(y)}$ (from customer point of view)

this is the Reason of existence of Two Regression lines.



$$\boxed{y = f(x)}$$

y → Dep. Variable, x → Ind. Variable

$$\boxed{\text{Effect} = f(\text{Cause})}$$

$$\boxed{x = f(y)}$$

x → Dep. V, y → Ind. V.

$$\boxed{\text{Effect} = f(\text{Cause})}$$

Correlation

1. It measures the direction as well as strength of relationship

2. It has nothing to do with cause & effect relationship.

3. $r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

Regression

1. It measures the value of dependent variable w.r.to independent variable.

2. It measures the cause & effect relationship i.e

$$\boxed{\text{Effect} = f(\text{Cause})}$$

$y = f(x)$

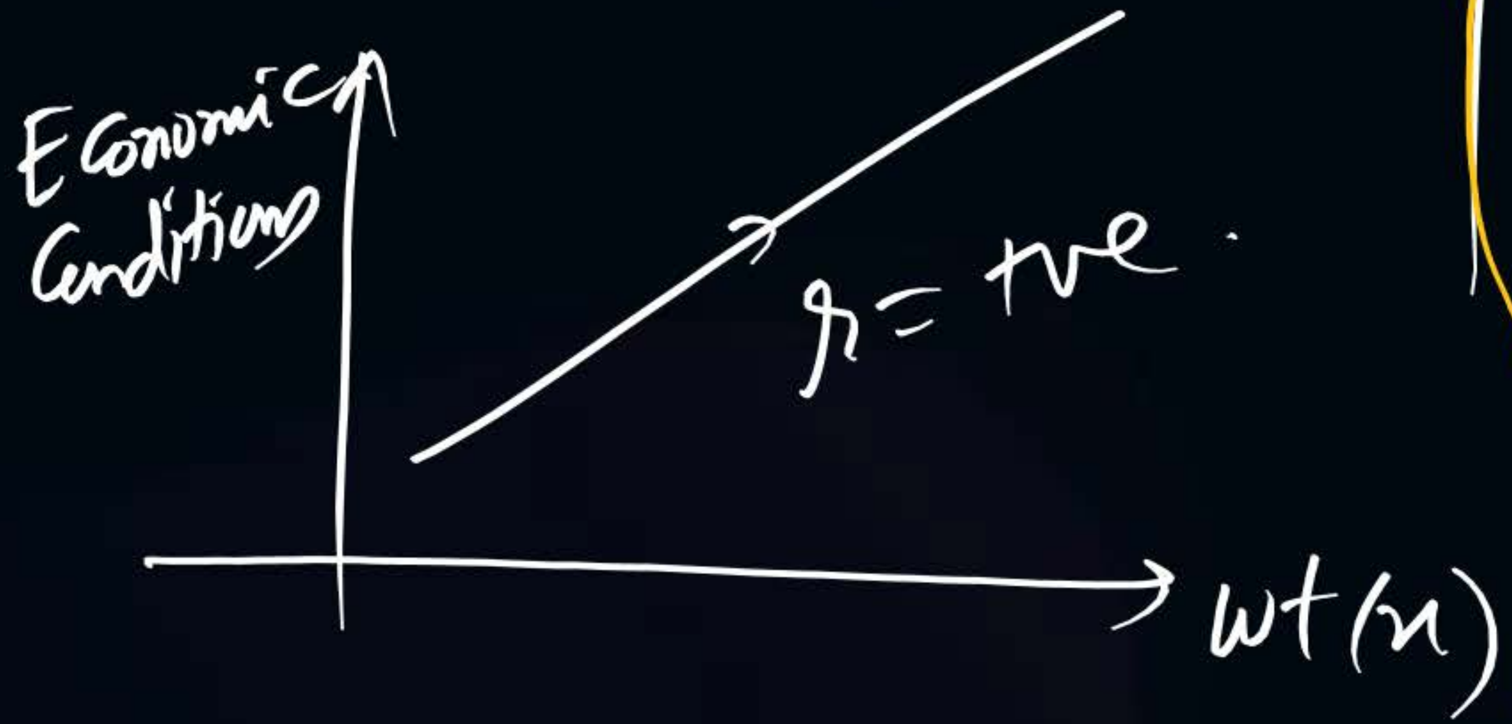
$x = f(y)$

3. R-line y on x :- $y - \bar{y} = b_{yx}(x - \bar{x})$

R-line x on y :- $x - \bar{x} = b_{xy}(y - \bar{y})$

where $b_{yx} = r\frac{\sigma_y}{\sigma_x}$, $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ & these are R-coeff.

(4) Sometimes there may be Non-Sense Correlation



(4) Here we do not have any Non sense Regression lines.

Analysis :- ① Intersecting Points of Two Regression lines is nothing but $(\bar{x}, \bar{y})$ (ESE 2022)

② Correlation coeff is the G.M of R. Coeff. i.e $\boxed{r = \pm\sqrt{b_{yx} \cdot b_{xy}}}$

Proof: $b_{yx} \cdot b_{xy} = \left(r \cdot \frac{\sigma_y}{\sigma_x}\right) \cdot \left(r \cdot \frac{\sigma_x}{\sigma_y}\right) = r^2 \Rightarrow \pm\sqrt{b_{yx} \cdot b_{xy}} = r$

③ r, b_{yx} & b_{xy} are of Same Sign (T)

i.e either all are +ve or all are -ve.

④ if $b_{yx} > 1$ then $b_{xy} < 1$ but Converse need not be True.

$\Longrightarrow$ / $\not\Longleftarrow$

i.e Both may be < 1 also. (GAZAB KI BAAT) But Both Can't greater than 1.

(5) Angle b/w two Regression lines :

$$\boxed{\tan\theta = \left(\frac{1-r^2}{r^2}\right)\cdot\frac{\sigma_x\cdot\sigma_y}{\sigma_x^2+\sigma_y^2}} \Rightarrow \theta = ?$$

Q If $\underbrace{\text{two R. lines are } \perp^r}_{\theta = 90^\circ}$ then correlation coeff will be ?

$$\tan 90^\circ = \left(\frac{1-r^2}{r^2}\right)\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

$$\infty = \left(\frac{1-r^2}{r^2}\right)\left(b\right)$$

$$\Rightarrow r = 0$$

Geometric Mean: (GM) for two +ve Nos a & b; $GM = +\sqrt{ab}$

for two -ve nos a & b; $GM = -\sqrt{ab}$

for one +ve & one -ve Number; $GM = \boxed{N.D}$

eg G.M of 4 & 9 is $= \sqrt{(4)(9)} = \sqrt{36} = +6$

eg GM of -9 & -4 is $= -\sqrt{(-9)(-4)} = -\sqrt{36} = -6$

eg GM of -9 & 4 is = N.D.

Q. Find Correlation Coeff, Regression Coeff, R. Line y on x & R. Line x on y for

1	2	3	4	5
3	5	7	9	11

$N = 5$

x	y	xy	x^2	y^2
1	3	3	1	9
2	5	10	4	25
3	7	21	9	49
4	9	36	16	81
5	11	55	25	121
Σ 15	35	125	55	285

$\bar{X} = E(X) = \frac{\Sigma X}{N} = \frac{15}{5} = 3$

$\bar{Y} = E(Y) = \frac{\Sigma Y}{N} = \frac{35}{5} = 7$

$\overline{X^2} = E(X^2) = \frac{\Sigma X^2}{N} = \frac{55}{5} = 11$

$\overline{Y^2} = E(Y^2) = \frac{\Sigma Y^2}{N} = \frac{285}{5} = 57$

$\overline{XY} = E(XY) = \frac{\Sigma XY}{N} = \frac{125}{5} = 25$

$Cov(X, Y) = E(XY) - E(X) \cdot E(Y) = 25 - 3 \times 7 = 4$

$Var(X) = E(X^2) - E^2(X) = 11 - (3)^2 = 2, \quad \sigma_x = \sqrt{2}$

$Var(Y) = E(Y^2) - E^2(Y) = 57 - (7)^2 = 8, \quad \sigma_y = 2\sqrt{2}$

(i) $r = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{4}{(\sqrt{2})(2\sqrt{2})} = 1$

(ii) $b_{yx} = r\frac{\sigma_y}{\sigma_x} = 1\cdot\left(\frac{2\sqrt{2}}{\sqrt{2}}\right) = 2$

(iii) $b_{xy} = r\cdot\frac{\sigma_y}{\sigma_x} = (1)\left(\frac{\sqrt{2}}{2\sqrt{2}}\right) = \frac{1}{2}$

(iv) R·line y on x :-

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 7 = 2(x-3)$$

$$y = 2x+1$$

(v) R·line x on y :- $x - \bar{x} = b_{xy}(y - \bar{y})$

$$x - 3 = \frac{1}{2}(y-7)$$

$$x = \frac{1}{2}y - \frac{1}{2}$$

Deep Analysis :→

① $\left.\begin{array}{l} y = 2x+1 \\ x = \frac{1}{2}y - \frac{1}{2} \end{array}\right\} \Rightarrow (x, y) = (3, 7) = (\bar{x}, \bar{y})$

② $b_{yx} = 2$, $b_{xy} = \frac{1}{2}$ i.e when one R. Coeff is > 1 then other is < 1

③ $b_{yx} = +2$, $b_{xy} = +\frac{1}{2}$, $r = +1$ i.e all three have same sign.

④ GM of b_{yx} & $b_{xy} = ? = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{2 \times \frac{1}{2}} = 1 = r$

i.e GM of R. Coeff is Correlation Coeff.

⑤ $\tan\theta = \left(\frac{1-r^2}{r^2}\right) \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} = \left(\frac{1-1^2}{1^2}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right) = 0 \Rightarrow \theta = 0^\circ$ Ans

Q. if $\bar{x}=10$, $\bar{y}=90$, $\sigma_x=3$, $\sigma_y=12$ & $r=0.8$

then find R. Line y on x and x on y ?

Sol: $b_{yx}=r\frac{\sigma_y}{\sigma_x}=0.8\left(\frac{12}{3}\right)=3.2$

$b_{xy}=r\frac{\sigma_x}{\sigma_y}=0.8\left(\frac{3}{12}\right)=0.2$

R Line y on x :

$$y-\bar{y}=b_{yx}(x-\bar{x})$$

$$y-90=3.2(x-10)$$

$$y=3.2x+58$$

R line x on y :

$$x-\bar{x}=b_{xy}(y-\bar{y})$$

$$x-10=0.2(y-90)$$

$$x=0.2y-\frac{90}{5}+10$$

$$x=0.2y-8$$

Analysis :-

(1) $\left.\begin{aligned} y &= 3.2x + 58 \\ x &= 0.2y - 8 \end{aligned}\right\} \Rightarrow (x, y) = (10, 90) = (\bar{x}, \bar{y})$

(2) $b_{yx} = 3.2$ i.e > 1 & $b_{xy} = 0.2$ ie < 1 (i.e justified)

(3) $r = 0.8$ (i.e +ve), $b_{yx} = 3.2$ (i.e +ve) & $b_{xy} = 0.2$ (ie +ve)

$\Rightarrow$ all three have same sign.

(4) $r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{3.2 \times 0.2} = \sqrt{0.64} = 0.8$

(5) $\tan\theta = \left(\dfrac{1 - r^2}{r^2}\right) \dfrac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} = \dfrac{1 - (0.8)^2}{(0.8)^2} \left[\dfrac{3 \times 12}{9 + 144}\right] \Rightarrow \theta = \tan^{-1}(?) = 13.24^\circ$

Find r, b_{yx}, b_{xy}, R. line y on x & R. line x on y. for the data;

$x =$	1	2	3	4	5	6	7
$y =$	9	8	10	12	11	13	14

HW

$$\because \cancel{b}_{yx} = \cancel{r}\frac{\sigma_y}{\sigma_x} \Rightarrow \sigma_x = \sigma_y$$

Ans: $b_{yx} = b_{xy} = r = 0.929$

& R line y on x is $y = 0.929x + 7.28$

R line x on y is $x = 0.929y - 6.219$

Analysis: ① $\left.\begin{array}{l} y = 0.929x + 7.28 \\ x = 0.929y - 6.219 \end{array}\right\} \Rightarrow (x, y) = (\_\_, \_\_) = (\bar{x}, \bar{y})$

② $b_{yx} = 0.929$ & $b_{xy} = 0.929$ i.e. Both are less than 1

$(b_{yx} < 1)$ $\quad (b_{xy} < 1)$

③ $b_{yx} = b_{xy} = r = +0.929$ i.e. all three have same sign.

④ $r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{(0.929)(0.929)} = 0.929$ i.e. r is G.M of b_{yx} & b_{xy}

⑤ $$\tan\theta = \left(\frac{1-r^2}{r^2}\right)\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} = \frac{1-(0.929)^2}{(0.929)^2} \cdot \left[\frac{\sigma_x^2}{\sigma_x^2 + \sigma_x^2}\right] = \frac{1}{2}\left[\frac{1-(0.929)^2}{(0.929)^2}\right] = ?$$

For the regression equations

$$y = 0.516x + 33.73 \quad \text{—(1)}$$

and $$x = 0.512y + 32.52 \quad \text{—(2)}$$

the means of x and y are nearly

(a) 67.6 and 68.6 (b) 68.6 and 68.6

(c) 67.6 and 58.6 (d) 68.6 and 58.6

Mean $= (\bar{x}, \bar{y}) = (\_\_, \_\_)$

= Intersecting Point of (1) & (2)

5 Consider the following regression equations obtained from a correlation table :

$$y = 0.516x + 33.73$$

$$x = 0.512y + 32.52$$

The value of the correlation coefficient will be

(a) 0.514 (b) 0.586

(c) 0.616 (d) 0.684

M-I $r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$ = Tough

M-II $r = +\sqrt{b_{yx} b_{xy}} = +\sqrt{0.516 \times 0.512} = +0.514$

Thank
YOU