DS & AI
CS & IT

Statistics -1
(Discrete Random Variable)
Lecture - 01

Recap of previous lecture

Topic

→ BAYE'S THEOREM

→ Concept of with & w/o Replacement

→ Questions Based on Tree Diagram.

Topics to be Covered

Topic

BASICS OF STATISTICS

(Discrete Random Variable)

Thumb Rule of this Chapter → Try to avoid making Question by using following words;

"If, what if, AGAR, YADI, TOH, ..."

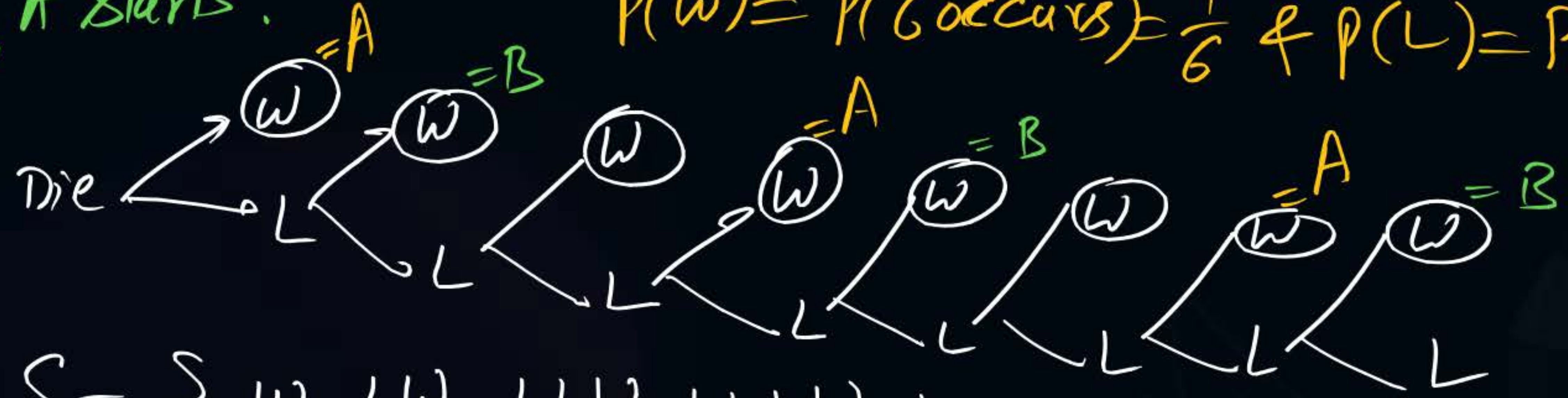
OR

Don't Try to develop Question (by your $\boxed{\text{little mind}}$ until you have a complete understanding of the chapter) & try to solve the Quest.

HW Q. Three persons A, B, C play a Game of Dice, $\boxed{\text{One after another}}$ (in succession) in which anyone can win if 6 appears 1st time then find their Respective chances of winning? if A starts.

$P(W) = P(6 \text{ occurs}) = \frac{1}{6}$ & $P(L) = P(\bar{6}) = \frac{5}{6}$

Sol:



$S = \{\underset{A}{W}, LW, LLW, \underset{A}{LLLW}, LLLLW, LLLLLW, \underset{A}{LLLLLLW}, \ldots\}$

fav cases for A = {W, LLLW, LLLLLLW, ...}

fav cases for B = {LW, LLLLW, LLLLLLLW, ...}

$P(A \text{ win}) = P(W \text{ or } LLLW \text{ or } LLLLLLW \text{ or } \ldots)$

$= \frac{1}{6} + \left(\frac{5}{6}\right)^3 \frac{1}{6} + \left(\frac{5}{6}\right)^6 \frac{1}{6} + \ldots$

$= \frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^6 + \ldots\right] = \frac{1}{6}\left[\frac{1}{1-\left(\frac{5}{6}\right)^3}\right] = \frac{36}{91}$

$P(B \text{ win}) = P(LW \text{ or } LLLLW \text{ or } LLLLLLLW \text{ or } \ldots)$

$= \frac{5}{6}\cdot\frac{1}{6} + \left(\frac{5}{6}\right)^4 \frac{1}{6} + \left(\frac{5}{6}\right)^7 \frac{1}{6} + \ldots$

$= \frac{5}{6}\cdot\frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^6 + \ldots\right] = \frac{5}{36}\left[\frac{1}{1-\left(\frac{5}{6}\right)^3}\right] = \frac{30}{91}$

$P(C \text{ win}) = 1 - \left(\frac{36}{91} + \frac{30}{91}\right) = \frac{25}{91}$

Q. Aishwarya studies either CS or Maths on each day.
If she studies CS on a day then the prob of studying M on next day is 0.6 &
If " " M " " " " " CS " " is 0.4.
Given that, $\boxed{\text{Aish studies CS on Monday}}$, then find the prob that she will also study CS on wednesday.

condition

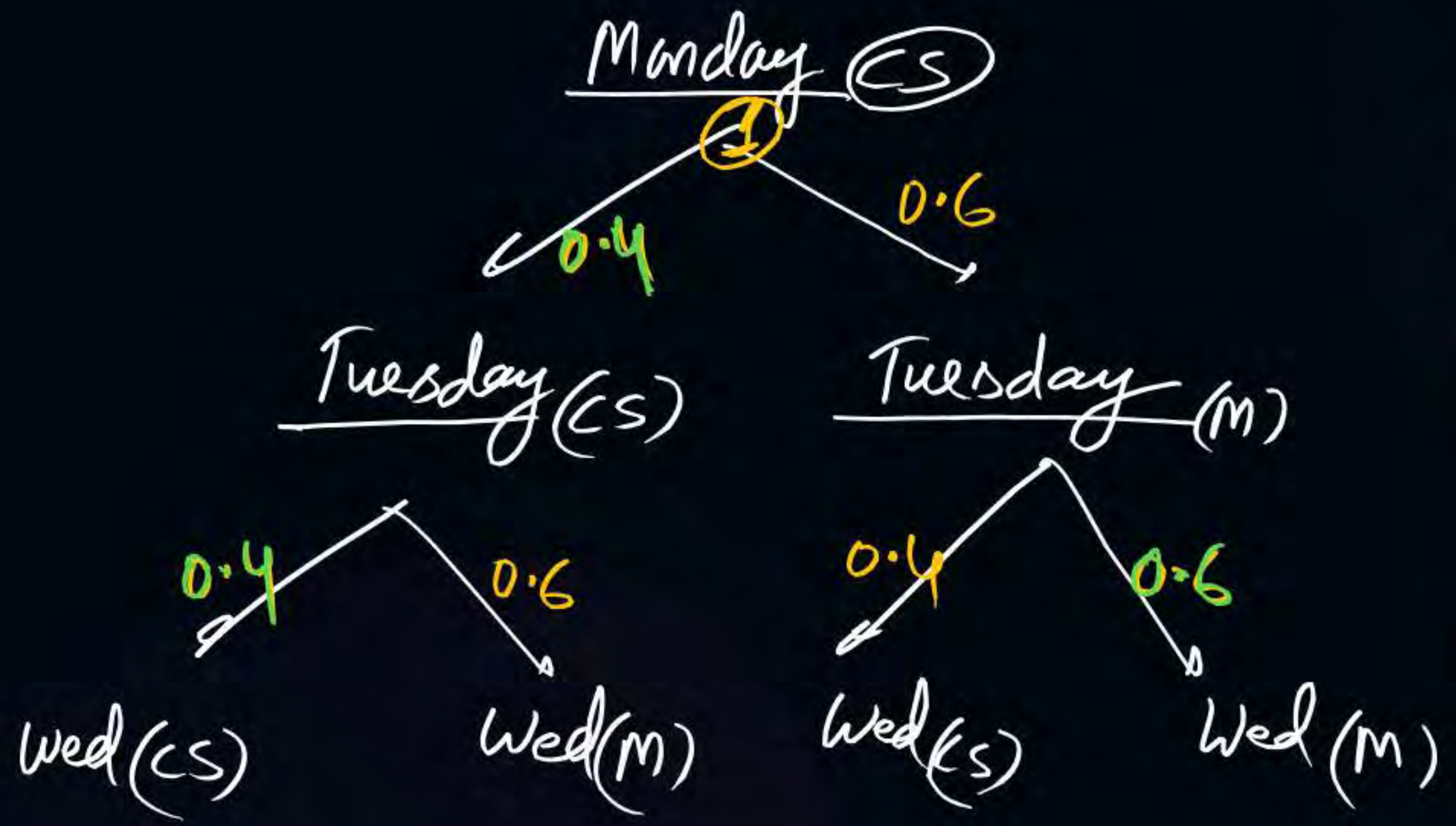
(Ans = 0.4 or 0.6)
(?) (?)

INFO TREE

$P(M \text{ after } CS) = 0.6$, $P(CS \text{ after } M) = 0.4$ $P(CS \text{ on Mon}) = 1$

$$S = \left\{ \begin{matrix} (CS, CS, CS), (CS, CS, M), (CS, M, CS), (CS, M, M) \\ (M, CS, CS), (M\ CS\ M), (M, M, CS), (M, M, M) \end{matrix} \right\} = 8 \text{ Triplets}$$

$$RSSp = \{CS \text{ on Mond}\} = \{(CS, CS, CS), (CS, CS, M)\ (CS, M, CS), (CS, M, M)\} = 4 \text{ Triplets.}$$



Fav Cases = {(CS CS CS), (CS M CS)} = 2 Triplets

So Req Prob = ~~$\frac{\text{Fav Triplets}}{\text{R. Triplets}} = \frac{2}{4} = \frac{1}{2} = 0.5$~~

∵ Triplets are not equally likely so we can't use $\frac{\text{Fav}}{\text{Total}}$ Approach

App III

Fav. Cases = {(CS CS CS), (CS M CS)}

Req Prob = P[(CS CS CS) ∩ (CS M CS)]

$= 1 \times 0.4 \times 0.4 + 1 \times 0.6 \times 0.4$

$= 0.16 + 0.24$

$= 0.40 = 40\%$ chance that she will also study CS on Wed.

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 (Geometric, Binomial, Poisson)

↓

Prob Mass func^n (p.m.f) $= p_i$

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

C.R.V (x)

↓

Continuous Prob Distribution

eg (Exponential, Uniform, Normal)

↓

Prob. Density func^n (p.d.f) $= f(x)$

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

p.m.f

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

(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)

It is the Average of Random Variable

$\rightarrow \bar{X} = \dfrac{\Sigma x}{N}$ (Childhood Method)

$\rightarrow E(X) = \dfrac{\Sigma p_i x_i}{\Sigma p_i} = \Sigma p_i x_i$ (Mature Method)

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 $M_d = \left(\dfrac{N+1}{2}\right)^{th}$ observation

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

P.Q : Find Mode & Median of following data

2, 3, 3, 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 $Md = \left(\frac{N+1}{2}\right)^{th} = \left(\frac{23+1}{2}\right)^{th} = 12^{th}$ observation

$= 6$ Ans

(iii) Also find Mean = ? $= \frac{\text{Sum of observation}}{N} = \frac{152}{23} = 6.61$ Ans

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.

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

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

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

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

$$\text{Average of Deviation from Central Value} = \frac{\sum(x-\bar{x})}{N} = \frac{(9-15)+(13-15)+(16-15)+(22-15)}{4} = 0\,kg$$

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

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

$$\text{Average of } \underbrace{\text{Square of deviations}} \text{ from central Value} = \frac{\sum(x-\bar{x})^2}{N}$$

$$= \frac{(-6)^2+(-2)^2+(1)^2+(7)^2}{4} = 22.5\,kg^2 = \text{Variance}$$

So $S.D = \sqrt{22.5\,kg^2} = 4.75\,kg$

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: Covariance is the Average of simultaneous deviations of X & Y from their Central Value $\bar{X}$ & $\bar{Y}$ resp. i.e

$$Cov(X,Y) = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N} = \cdots\cdots = E(XY) - E(X).E(Y)$$

Note if X & Y are Independent R.V, then $Cov(X,Y) = 0$

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 Var or S.D ≥ 0 True (equality holds in case of constant Data set)

($\because$ It represents spread of data)

② Var or SD $\propto \dfrac{1}{\text{consistency}}$ (True)

2022 ③ $Cov(X, X) = Var(X)$ True

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

with Breakup $\propto \dfrac{1}{\text{W/O Breakup}}$

Analysis: w.k. that $E(x) \begin{cases} = \dfrac{\sum x}{N} \\ = \sum p_i x_i \end{cases}$

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

& $Cov(x, y) = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{N} = E\{(x-\bar{x})(y-\bar{y})\} = \cdots\cdots = \boxed{E(xy) - E(x).E(y)}$

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

(i) $E(aX \pm bY \pm c) = a\,E(X) \pm b\,E(Y) \pm E(c)$

$= a\,E(X) \pm b\,E(Y) \pm c$

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

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

eg $Var(-X+3) = (-1)^2\,Var(X) + Var(3)$

$= Var(X) + 0$ //

(iii) $\boxed{Var(aX \pm bY) = a^2\,Var(X) + b^2\,Var(Y) \pm 2ab\,Cov(X, Y)}$

Learn only w/o Proof.

eg Consider $X = \underbrace{62, 62, 62, 62, 62, 62, 62}_{\text{Constant Data}}$ then $\bar{X} = 62$ kg & $Var = 0$, $SD = 0$

(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

$\sum(X) = ? = 3300 \quad N = 100$

① Mode = 35 Marks

② $\text{Median} = \dfrac{\left(\frac{100}{2}\right)^{th} + \left(\frac{100}{2}+1\right)^{th}}{2} = \dfrac{50^{th} + 51^{st}}{2} = \dfrac{35+35}{2} = 35$

③ $\text{Mean}(\bar{X}) = \dfrac{\sum X}{N} = \dfrac{25(20)+30(20)+35(40)+40(20)}{100}$

$= \dfrac{3300}{100} = 33$ Marks.

Marks: $\underbrace{25, 25, 25, \ldots 25}_{20\text{ students}}, \underbrace{30, 30, 30, \ldots 30}_{20\text{ students}}, \underbrace{35, 35, 35, \ldots 35}_{40\text{ students}}, \underbrace{40, 40, 40, \ldots 40}_{20\text{ students}}$

PODCAST:

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

$$\text{Mean}(\bar{X}) = \frac{\Sigma x}{N}$$

$$= \frac{25\,(20\text{ times}) + 30\,(20\text{ times}) + 35\,(40\text{ times}) + 40\,(20\text{ times})}{100}$$

$$= \frac{25 \times 20 + 30 \times 20 + 35 \times 40 + 40 \times 20}{100} = \frac{3300}{100} = 33\ M$$

Note: why the standard Result $\text{Mode} = 3M_d - 2M_e$ is not applicable here in this Q.??

bcoz this Result is Valid only for Moderately skewed Data set.

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

$\because$ Consistency $\propto \dfrac{1}{S.D}$

So L is More Consistent

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,

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

$$= Var(X)$$

$\Rightarrow Cov(X,Y)=0$

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

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

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

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

~~(c)~~ $E(X^2Y^2)=E^2(X)\cdot E^2(Y)$ False

$\therefore$ Correct Version is $\because X$ & Y are Ind $\Rightarrow X^2$ & Y^2 are also Ind

So $Cov(X^2,Y^2)=0$

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

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

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

So $Var(X-Y)=1^2Var(X)+(-1)^2Var(Y)+2(1)(-1)\underbrace{Cov(X,Y)}_{0}$

$=Var(X)+Var(Y)-0$

Q. 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 = −10, Var = 2

Given, $E(x)=E(y)=0$, $Var(x)=\frac{1}{4}$, $Var(y)=\frac{1}{9}$

x & y are Ind $\Longrightarrow Cov(x,y)=0$

Let $\boxed{Z=2x-3y}$

So $E(Z)=E(2x-3y)=2E(x)-3E(y)=2(0)-3(0)=$
$=0-0=0$

Now $Var(Z)=Var(2x-3y)$

$=(2)^2 Var(x)+(-3)^2 Var(y)+2(2)(-3)\cancel{Cov}(x,y)$

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

4. Qs 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) $\{\mu, \sigma^2\}$

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

$E[X] + E[Y] =$ __11__.

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

$$E(X+2Y)=33 \Rightarrow E(X)+2E(Y)=33$$

$$3\left[E(X)+E(Y)\right]=33$$

So $E(X)+E(Y)=11$

NIRMAN
2025

Dr Puneet Sir pw

@DRPUNEETSIRPW

Thank

YOU