MA1351 - PROBABILITY AND STATISTICS

(II Information Technology , II Computer Science Engineering & II Artificial Intelligence and Data Science – III Semester)

UNIT I – PROBABILITY & RANDOM VARIABLES

PART – A

1. **Let A and B be two events such that $P(A)=\frac{1}{2},\ P(B)=\frac{1}{3}$ and $P(A\cap B)=\frac{1}{6}$. Compute $P\left(B/A\right)$ and $P(\overline{A}\cap B)$.** **(April/May 2021)**

Solution:

$$P\left(\overline{A}\cap B\right)=P(B)-P(A\cap B)=\frac{1}{3}-\frac{1}{6}=\frac{1}{6}$$

$$P\left(B/A\right)=\frac{P(B\cap A)}{P(A)}=\frac{1}{6}\times 2=\frac{1}{3}$$

2. **If *A* and *B* are mutually exclusive events $P(A) = 0.29$ and $P(B) = 0.43$ then find $P(A\cup B)$ and $P(\overline{A})$.** **(Nov/Dec 2019)**

Solution:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.29+0.43-0=0.72$$

$\because$ (A & B are mutually exclusive events, $P(A\cap B)=0$)

$$P\left(\overline{A}\right)=1-P(A)=1-0.29=0.71$$

3. **Let A and B be two events such that $P(A)=\frac{1}{3}, P(B)=\frac{3}{4},\ P(A\cap B)=\frac{1}{4}$. Compute $P\left(A/B\right)$ and $P\left(A\cap\overline{B}\right)$.** **(May/June 2019)**

Solution:

$$P\left(A/B\right)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{4}}{\frac{3}{4}}=\frac{1}{3}.$$

$$P\left(A\cap\overline{B}\right)=P(A)-P(A\cap B)=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$$

4. **State Baye's Theorem on Probability.**

Statement:

If $E_1, E_2 \ldots E_n$ are a set of exhaustive and mutually exclusive events associated with a random experiment and A is any other event associated with E_i.

Then $P(E_i / A)=\dfrac{P(E_i)P(A/E_i)}{\sum_{i=1}^{n} P(E_i)P(A/E_i)}$, i=1,2,..n

5. **Let X be a discrete RV with probability mass function** $P(X=x)=\begin{cases}\dfrac{x}{10}, & x=1,2,3,4\\ 0, & \text{otherwise}\end{cases}$

Compute $E\left(\frac{X}{2}\right)$.

Solution:

$X=x$	1	2	3	4
$P(X=x)$	1/10	2/10	3/10	4/10

$$E\left(\frac{X}{2}\right)=\frac{1}{2}E(X)=\frac{1}{2}\sum_{x=1}^{4}xP(X=x)=\frac{1}{2}\left\{1\times\frac{1}{10}+2\times\frac{2}{10}+3\times\frac{3}{10}+4\times\frac{4}{10}\right\}=\frac{1}{2}\left\{\frac{1+4+9+16}{10}\right\}=\frac{3}{2}$$

6. **Let X be the random variable which denotes the number of heads in three tosses of a fair coin. Determine the probability mass function of X.**

Solution:

If a coin is tossed three times, then the sample space is

$$S=\{TTT, HTT, THT, TTH, HTH, HHT, THH, HHH\}$$

Let X be the R.V denotes the number of heads.

The probability mass function is

$X=x$	0	1	2	3
$P(X=x)$	1/8	3/8	3/8	1/8

7. **Find the value of '*k*' for a continuous random variable X whose probability density function is given by $f(x)=kx^4;\ 0\le x\le 2$.** **(April/May2021)**

Solution:

Since X is a continuous random variable $f(x)\ge 0$ and $0\le x\le 2$ and $\int_0^2 f(x)\,dx=1$

$$\int_0^2 k\,x^4\,dx=1\Rightarrow k\left[\frac{x^5}{5}\right]_0^2=1\Rightarrow \frac{k}{5}[32-0]=1\Rightarrow k=\frac{5}{32}$$

8. **A random variable X has the p.d.f f(x) given by** $f(x)=\begin{cases} a(1+x^2), & 2<x<5\\ 0, & \text{otherwise}\end{cases}$ **. Find 'a' and P (X < 4).**

Solution:

Since X is a continuous random variable $f(x)\ge 0, \forall x$ and $\int_{-\infty}^{\infty} f(x)\,dx=1$

$$\int_2^5 a(1+x^2)dx=1\Rightarrow a\left[x+\frac{x^3}{3}\right]_2^5=1$$

$$\Rightarrow a\left[\left(5+\frac{5^3}{3}\right)-\left(2+\frac{2^3}{3}\right)\right]=1\Rightarrow 42a=1\Rightarrow a=\frac{1}{42}$$

$$P(X<4)=\int_2^4 f(x)dx=\int_2^4\frac{(1+x^2)}{42}dx=\frac{1}{42}\left[x+\frac{x^3}{3}\right]_2^4=\frac{1}{42}\left[\left(4+\frac{4^3}{3}\right)-\left(2+\frac{2^3}{3}\right)\right]=\frac{31}{63}$$

9.	**The CDF of a continuous random variable is given by** $F(x)=\begin{cases}0, & x<0\\ 1-e^{-\frac{x}{5}}, & x\geq 0\end{cases}$**. Find the PDF of X and mean of *X*.** **Solution:** The relation between pdf and cdf $f(x)=\frac{d}{dx}\left[F(x)\right]=\frac{d}{dx}\left[\begin{cases}0, & x<0\\ 1-e^{-\frac{x}{5}}, & x\geq 0\end{cases}\right]=\begin{cases}0, & x<0\\ \frac{1}{5}e^{-\frac{x}{5}}, & x\geq 0\end{cases}$ $E(X)=\int_{-\infty}^{\infty}xf(x)\,dx=\int_{0}^{\infty}\frac{1}{5}xe^{-\frac{x}{5}}\,dx=\frac{1}{5}\left[(x)\left(\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}}\right)-(1)\left(\frac{e^{-\frac{x}{5}}}{\frac{1}{25}}\right)\right]_0^{\infty}=\frac{25}{5}=5$
10.	**Let X be a random variable with E(X)=1, E [X (X – 1)] = 4. Find Var(X) &Var (2 – 3X).** **Solution:** $E(X)=1$ $E[X(X-1)]=4 \Rightarrow E[X^2-X]=4$ $\Rightarrow E[X^2]-E[X]=4 \Rightarrow E[X^2]-1=4 \Rightarrow E[X^2]=5$ $Var(X)=E\left[X^2\right]-\left(E\left[X\right]\right)^2=5-1=4$ $Var(2-3X)=(-3)^2Var(X)=9\times 4=36$ $\quad\because Var(aX+b)=(a)^2Var(X)$
11.	**Let** $M_X(t)=\frac{1}{1-t}$ **such that** $t\neq 1$**, be the mgf of R.V X. Find the MGF of Y = 2X +1.** **Solution:** $M_Y(t)=M_{2X+1}(t)=e^tM_X(2t)$ $\quad\left[\because M_{aX+b}(t)=e^{bt}M_X(at)\right]$ $=e^t\left[\frac{1}{1-t}\right]_{t\to 2t}=\frac{e^t}{1-2t}.$ $\quad\left[\because M_X(at)=\left[M_X(t)\right]_{t\to at}\right]$
12.	**If the random variable has the moment generating function** $M_x(t)=\frac{3}{3-t}$**, compute E[X**^{2}**].** **Solution:** $M_X(t)=\frac{3}{3-t}=\frac{3}{3\left(1-\frac{t}{3}\right)}=\left(1-\frac{t}{3}\right)^{-1}=1+\left(\frac{t}{3}\right)+\left(\frac{t}{3}\right)^2+\left(\frac{t}{3}\right)^3+\cdots=1+\frac{1}{3}\left(\frac{t}{1!}\right)+\frac{2}{9}\left(\frac{t^2}{2!}\right)+\frac{2}{9}\left(\frac{t^3}{3!}\right)+\cdots$ $E\left(X^r\right)=\mu_r'=\text{coefficient of}\left(\frac{t^r}{r!}\right)\ in\ M_X(t)$ $E\left(X^2\right)=\mu_2'=\text{coefficient of}\left(\frac{t^2}{2!}\right)\ in\ M_X(t)=\frac{2}{9}.$

13.	**A continuous RV X has the pdf $f(x) = \frac{x^2 e^{-x}}{2}$, x > 0 Find the rth moment of X about the origin.** **Solution:** $\mu_r' = E[X^r] = \int_{-\infty}^{\infty} x^r f(x)dx = \int_0^{\infty} x^r \frac{x^2 e^{-x}}{2} dx = \frac{1}{2}\int_0^{\infty} x^{r+2} e^{-x} dx = \frac{1}{2}\int_0^{\infty} x^{(r+3)-1} e^{-x} dx$ $= \frac{1}{2}\Gamma(r+3) \qquad \because \int_0^{\infty} x^{n-1} e^{-x} dx = \Gamma(n)$ $= \frac{1}{2}(r+2)! \qquad \because$ if n is positive integer $\Gamma(n) = (n-1)!$
14.	**For a Binomial distribution with mean 6 and standard deviation $\sqrt{2}$, find the first two terms of the distribution.** **Solution:** $np = 6 \ and \sqrt{npq} = \sqrt{2} \Rightarrow npq = 2 \Rightarrow 6q = 2 \Rightarrow \boxed{q = \frac{1}{3}} \therefore \boxed{p = \frac{2}{3}} \qquad \therefore n \times \frac{2}{3} = 6 \ \therefore \boxed{n = 9}$ $P(X = x) = nC_x p^x q^{n-x} = 9C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{9-x}, \ x = 0,1,2,3,\ldots 9$ $P(X = 0) = 9C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^9 = \left(\frac{1}{3}\right)^9$ $P(X = 1) = 9C_1 \left(\frac{2}{3}\right)^1 \left(\frac{1}{3}\right)^8 = 9 \times \frac{2}{3} \times \frac{1}{3^8} = \frac{2}{3^7}$
15.	**If X and Y are independent binomial variates following $B\left(5, \frac{1}{2}\right)$ and $B\left(7, \frac{1}{2}\right)$ respectively find $P[X + Y = 3]$.** **Solution:** Given $X \sim B\left(5, \frac{1}{2}\right), Y \sim B\left(7, \frac{1}{2}\right)$ By additive property, $X + Y$ is also a binomial variable with parameters $n_1 + n_2 = 12$ & $p = \frac{1}{2}$ $\therefore X + Y \sim B\left(12, \frac{1}{2}\right)$ $P(Z = z) = nC_z p^z q^{n-z}; z = 0,1,2,3.....n$ $\therefore P[X + Y = 3] = 12C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^9 = \frac{55}{2^{10}}$
16.	**One percent of jobs arriving at a computer system need to wait until weekends for scheduling, owing to core-size limitations. Find the probability that there are no jobs among a sample of 200 jobs.**

	Solution: Let X be the Poisson random variable denotes the number of jobs that have to wait *p = 1%= 0.01, n = 200, λ = np = (200)(0.01) =2,* By Poisson distribution, $P(X=x)=\dfrac{e^{-\lambda}\lambda^x}{x!},\quad x=0,1,2,.....$ $P(X=0)=\dfrac{e^{-2}2^0}{0!}=e^{-2}=0.1353$
17.	**If X is a Geometric variable taking values $1,2,3,...$, then find P(X is odd). (April/May 2017)** **Solution:** We know that for Geometric Distribution $P(X=x)=q^{x-1}p,\quad x=1,2,3,\cdots\infty$ $P(X \text{ is odd})=P(X=1,3,5,\cdots)=P(X=1)+P(X=3)+P(X=5)+P(X=7)+\cdots$ $=p+q^2p+q^4p+q^6p+\cdots=p(1+q^2+q^4+q^6+\cdots)=p(1-q^2)^{-1}=\dfrac{p}{(1-q^2)}=\dfrac{(1-q)}{(1-q)(1+q)}=\dfrac{1}{1+q}$
18.	**If X is a Uniformly distributed R.V with mean 1 and variance $\dfrac{4}{3}$ find P(X<0).** **Solution:** Mean = $\dfrac{a+b}{2}=1 \Rightarrow a+b=2$ --------------(1) variance = $\dfrac{(b-a)^2}{12}=\dfrac{4}{3} \Rightarrow b-a=4$ ---------------(2) (1) + (2) ⇒2b = 6 ⇒ b = 3 (1) – (2) ⇒2a = –2 ⇒ a = –1 Probability density function of Uniform distribution is $f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b\\ 0, & \text{otherwise}\end{cases} \Rightarrow f(x)=\begin{cases}\dfrac{1}{4}, & -1<x<3\\ 0, & \text{otherwise}\end{cases}$ $P(X<0)=\int_{-1}^{0}f(x)dx=\int_{-1}^{0}\dfrac{1}{4}dx=\dfrac{1}{4}[x]_{-1}^{0}=\dfrac{1}{4}.$
19.	**Suppose the length of life of an appliance has an exponential distribution with mean 10 years. What is the probability that the average life time of a random sample of the appliance is atleast 10.5 years?** **Solution:** Mean of the exponential distribution = E(X) = 1/λ ⇒10 = 1/λ $\lambda=\dfrac{1}{10},\quad f(x)=\lambda e^{-\lambda x},\ x>0 \Rightarrow f(x)=\dfrac{1}{10}e^{-\frac{x}{10}}, x>0$ $P(X>10.5)=\int_{10.5}^{\infty}f(x)dx=\int_{10.5}^{\infty}\dfrac{1}{10}e^{-\frac{x}{10}}dx=e^{-1.05}=0.3499$

20.	**If X is a normal variate with mean $=20$ and S.D = 10. Find $P[15 \le X \le 40]$.** **Solution:** X follows N(20, 10).$\therefore \mu = 20$ & $\sigma = 10$ Let $Z = \dfrac{X - \mu}{\sigma}$ be the standard normal variate $P[15 \le X \le 40] = P\left[\dfrac{15-20}{10} \le Z \le \dfrac{40-20}{10}\right] = P[-0.5 \le Z \le 2] = P[-0.5 \le Z \le 0] + P[0 \le Z \le 2]$ $= P[0 \le Z \le 0.5] + P[0 \le Z \le 2] = 0.1915 + 0.4772 = 0.6687$.

PART-B

1. **(i) A bag contains 3 black and 4 white balls. Two balls are drawn at random one at a time without replacement. (a) What is the probability that the second ball drawn is white?**
(b)What is the conditional probability that the first ball drawn is white if the second ball is known to be white? (May/June 2019)
Solution:

(a) The total number of balls in the bag=7 balls
The total number of possible outcomes=$7C_1$ X $6C_1$
Therefore the total number of favourable outcomes =$3C_1$ X $4C_1$ +$4C_1$ X $3C_1$

P(the second ball drawn is white) $= \dfrac{3C_1 X\ 4C_1}{7C_1 X\ 6C_1} + \dfrac{4C_1 X\ 3C_1}{7C_1 X\ 6C_1} = \dfrac{3(4)}{7(6)} + \dfrac{4(3)}{7(6)} = \dfrac{4}{7}$

(b)P(the second ball is white / first ball drawn is white) $= \dfrac{P(both\ \ are\ white)}{P(\ first\, draw\, is\, white)} = \dfrac{\left(\dfrac{4}{7}\times\dfrac{3}{6}\right)}{\dfrac{4}{7}} = \dfrac{\left(\dfrac{2}{7}\right)}{\left(\dfrac{4}{7}\right)} = \dfrac{1}{2}$

(ii) There are two boxes B_1 and B_2. B_1 contains two red balls and one green ball. B_2 contains one red ball and two green balls.
(a) A ball is drawn from one of the boxes randomly. It is found to be red, what is the probability that it is from B_1?
(b)Two balls are drawn randomly from one of the boxes without replacement. One is red and the other is green, what is the probability that they came from B_1?
(c)A ball is drawn from one of the boxes is green, what is the probability that it came from B_2?
(d)A ball is drawn from one of the boxes is white, what is the probability that it came from B_2?

Solution:
Let B_1 and B_2 be the events that the boxes B_1 and B_2 are selected respectively.

$P(B_1) = \dfrac{1}{2}$, $P(B_2) = \dfrac{1}{2}$

Let A be the event that a red ball is selected.

$P(A/B_1) = \dfrac{2}{3}$, $P(A/B_2) = \dfrac{1}{3}$

By Baye's theorem,

	(a) P(ball is from B_1, given it is red) $=P(B_1/A)=\dfrac{P(B_1)P(A/B_1)}{P(B_1)P(A/B_1)+P(B_2)P(A/B_2)}=\dfrac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{1}{3}\right)}=\dfrac{2}{3}$ (b) Let C be the event that a red ball and a green ball are selected. $P(C/B_1)=\dfrac{2_{C_1}\times 1_{C_1}}{3_{C2}}=\dfrac{2}{3},\ P(C/B_2)=\dfrac{1_{C_1}\times 2_{C_1}}{3_{C2}}=\dfrac{2}{3}$ P(B_1 was chosen given a red ball and a green ball were selected) $=P(B_1/C)=\dfrac{P(B_1)P(C/B_1)}{P(B_1)P(C/B_1)+P(B_2)P(C/B_2)}=\dfrac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{2}{3}\right)}=\dfrac{1}{2}$ (c) Let D be the event that a green ball is selected. $P(D/B_1)==\dfrac{1}{3},P(D/B_2)=\dfrac{2}{3}$ $=P(B_2/D)=\dfrac{P(B_2)P(D/B_2)}{P(B_1)P(D/B_1)+P(B_2)P(D/B_2)}=\dfrac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{1}{3}\right)+\left(\frac{1}{2}\times\frac{2}{3}\right)}=\dfrac{2}{3}$ (d) Let E be the event that a white ball is selected. The given two boxes does not contain a white ball, hence the probability is 0.
2.	**A random variable X has the following probability function:** **X : 0 1 2 3 4 5 6 7** **$P(X)$: 0 K 2K 2K 3K K^2 $2K^2$ $7K^2+K$** **(a) Find K (b) Evaluate $P(X<6)$, $P(X\geq 6)$ and $P(0<X<5)$ (c) Determine the distribution function of X (d) Find $P(1.5<X<4.5/X>2)$ (e) E(3X-4),Var(3X-4) (f) If $P[X\leq C]>\frac{1}{2}$, find the minimum value of *C*.** (a) We know that $\sum_i P(X=x_i)=1$ $\Rightarrow \sum_{x=0}^{7}P(X=x)=1, \Rightarrow K+2K+2K+3K+K^2+2K^2+7K^2+K=1 \Rightarrow 10K^2+9K-1=0$ $\Rightarrow$ $K=\dfrac{1}{10}$ *or* $K=-1$ (here $K=-1$ is impossible, since $P(X=x)\geq 0$) $\therefore K=\dfrac{1}{10}$ $\therefore$ The probability mass function is

$X=x$	0	1	2	3	4	5	6	7
$P(X=x)$	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{1}{100}$	$\frac{2}{100}$	$\frac{17}{100}$

(b) $P(X<6)=P(X=0)+P(X=1)+...+P(X=5)=\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}=\frac{81}{100}$

$P(X\geq 6)=1-P(X<6)=1-\frac{81}{100}=\frac{19}{100}$

$P(0<X<5)=P(X=1)+P(X=2)+P(X=3)+P(X=4)=K+2K+2K+3K$

$=8K=\frac{8}{10}=\frac{4}{5}$

(c) The distribution of X is given by $F_X(x)$ defined by $F_X(x)=P(X\leq x)$

$X=x$	$P(X=x)$	$F_X(x)=P(X\leq x)$
0	0	0 , x < 1
1	$\frac{1}{10}$	$0+\frac{1}{10}=\frac{1}{10}$, $1\leq x<2$
2	$\frac{2}{10}$	$0+\frac{1}{10}+\frac{2}{10}=\frac{3}{10}$, $2\leq x<3$
3	$\frac{2}{10}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}=\frac{5}{10}$, $3\leq x<4$
4	$\frac{3}{10}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}=\frac{8}{10}=\frac{4}{5}$, $4\leq x<5$
5	$\frac{1}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}=\frac{81}{100}$, $5\leq x<6$
6	$\frac{2}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+\frac{2}{100}=\frac{83}{100}$, $6\leq x<7$
7	$\frac{17}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+\frac{2}{100}+\frac{17}{100}=\frac{100}{100}=1$, $x\leq 7$

(d) $P(1.5<X<4.5/X>2)=\frac{P(X=2,3,4\cap X=3,4,5,6,7)}{P(X=3,4,5,6,7)}=\frac{P(X=3,4)}{P(X=3,4,5,6,7)}$

$=\frac{P(X=3)+P(X=4)}{P(X=3)+P(X=4)+P(X=5)+P(X=6)+P(X=7)}$

$=\frac{2K+3K}{2K+3K+K^2+2K^2+7K^2+K}=\frac{5K}{6K+10K^2}=\frac{5}{6+1}=\frac{5}{7}$

(e) To find $E(3X-4)$, Var$(3X-4)$

$E(3X-4)=3E(X)-E(4)=3E(X)-4$ - - - - - - (1)

$Var(3X-4)=3^2Var(X)-Var(4)=9Var(X)-0=9Var(X)$

$Var(3X-4)=9Var(X)$ - - - - - - - - - - (2)

$E(X)=\sum xP(X=x)$

$$=0\times P(X=0)+1\times P(X=1)+2\times P(X=2)+3\times P(X=3)+4\times P(X=4)$$
$$+5\times P(X=5)+6\times P(X=6)+7\times P(X=7)$$
$$=0+1\times K+2\times 2K+3\times 2K+4\times 3K+5\times K^2+6\times 2K^2+7\times(7K^2+K)$$
$$=K+4K+6K+12K+5K^2+12K^2+49K^2+7K \;=30K+66K^2=\frac{30}{10}+\frac{66}{100}=\frac{366}{100}$$
$$E(X^2)=\sum x^2P(X=x)$$
$$=0\times P(X=0)+1^2\times P(X=1)+2^2\times P(X=2)+3^2\times P(X=3)+4^2\times P(X=4)$$
$$+5^2\times P(X=5)+6^2\times P(X=6)+7^2\times P(X=7)$$
$$=0+1^2\times K+2^2\times 2K+3^2\times 2K+4^2\times 3K+5^2\times K^2+6^2\times 2K^2+7^2\times(7K^2+K)$$
$$=K+8K+18K+48K+25K^2+72K^2+343K^2+49K$$
$$=124K+440K^2=\frac{124}{10}+\frac{440}{100}=\frac{1240+440}{100}=\frac{1680}{100}=\frac{168}{10}=\frac{84}{5}$$
$$(1)\Rightarrow E(3X-4)=\frac{3\times 366}{100}-4=\frac{1098-400}{100}=\frac{698}{100}=69.8$$
$$(2)\Rightarrow Var(3X-4)=9Var(X)=9\left[E(X^2)-(E(X))^2\right]=9[16.8-13.3956]=30.6396$$

(f) <u>To find the minimum value of</u> C if $P[X\le C]>\frac{1}{2}$

$X=x$	$P(X=x)$	$P(X\le x)$
0	0	$0<\frac{1}{2}$
1	$\frac{1}{10}$	$0+\frac{1}{10}=\frac{1}{10}<\frac{1}{2}$
2	$\frac{2}{10}$	$0+\frac{1}{10}+\frac{2}{10}=\frac{3}{10}<\frac{1}{2}$
3	$\frac{2}{10}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}=\frac{5}{10}=\frac{1}{2}$
4	$\frac{3}{10}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}=\frac{8}{10}=\frac{4}{5}>\frac{1}{2}$
5	$\frac{1}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}=\frac{81}{100}>\frac{1}{2}$
6	$\frac{2}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+\frac{2}{100}=\frac{83}{100}>\frac{1}{2}$
7	$\frac{17}{100}$	$0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+\frac{2}{100}+\frac{17}{100}=\frac{100}{100}=1>\frac{1}{2}$

$\therefore$The minimum value of C is 4

3. **(i)**A random variable X has the following probability distribution.

$X=x$	**-2**	**-1**	**0**	**1**	**2**	**3**
$P(X=x)$	**0.1**	**k**	**0.2**	**2k**	**0.3**	**3k**

Find the value of k, mean of X and P(–2<X<2).

(April/May 2021)

Solution:

Since $\sum P(X) = 1$

$0.1 + K + 0.2 + 2K + 0.3 + 3K = 1$

$6K + 0.6 = 1$

$$6K = 0.4$$

$$K = \frac{0.4}{6} = \frac{1}{15}$$

Mean of X is defined by $E(X) = \sum xP(x)$

$$E(X) = \left(-2\times\frac{1}{10}\right) + \left(-1\times\frac{1}{15}\right) + \left(0\times\frac{1}{5}\right) + \left(1\times\frac{2}{15}\right) + \left(2\times\frac{3}{10}\right) + \left(3\times\frac{1}{5}\right)$$

$$= -\frac{1}{5} - \frac{1}{15} + \frac{2}{15} + \frac{3}{5} + \frac{3}{5} = \frac{16}{15}.$$

$P(-2 < X < 2) = P(X = -1, 0 \text{ or } 1)$

$$= P(X = -1) + P(X = 0) + P(X = 1) = \frac{1}{15} + \frac{1}{5} + \frac{2}{15} = \frac{1+3+2}{15} = \frac{6}{15} = \frac{2}{5}$$

(ii) A random variable X has the probability mass function f (x) = $\frac{1}{2^x}$, x= 1,2,3,...Find its M.G.F , mean and variance.

Solution:

M.G.F= $M_X(t) = E(e^{tX})$

$$= \sum_{x=1}^{\infty} e^{tx} f(x) = \sum_{x=1}^{\infty} e^{tx}\frac{1}{2^x} = \sum_{x=1}^{\infty}\left(\frac{e^t}{2}\right)^x$$

$$= \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \left(\frac{e^t}{2}\right)^4 + \dots$$

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

$$= \frac{e^t}{2}\left[1 - \frac{e^t}{2}\right]^{-1}$$

M.G.F= $M_X(t) = \frac{e^t}{2 - e^t}$ …………….. (1)

Mean = $E(X)$= $\left\{\frac{d}{dt}\left[M_X(t)\right]\right\}_{t=0} = \left\{\frac{d}{dt}\left[\frac{e^t}{2 - e^t}\right]\right\}_{t=0} = \left[\frac{(2 - e^t)e^t - e^t(-e^t)}{(2 - e^t)^2}\right]_{t=0} = 2$

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

Where $E(X^2) = \left\{\frac{d}{dt}\left[M_X'(t)\right]\right\}_{t=0} = \left\{\frac{d}{dt}\left[\frac{2e^t}{(2 - e^t)^2}\right]\right\}_{t=0} = \left[\frac{(2 - e^t)^2 e^t - e^t 2(2 - e^t)(-e^t)}{(2 - e^t)^4}\right]_{t=0} = 6$

Variance = $Var(X) = E(X^2) - (E(X))^2 = 6 - 4 = 2$

4.	**(i)If the density function of a continuous random variable X is given by** $\mathbf{f(x) = \begin{cases} ax, & 0 \le x \le 1 \\ a, & 1 \le x \le 2 \\ 3a - ax, & 2 \le x \le 3 \\ 0, & \text{elsewhere} \end{cases}}$ **(a)Find the value of 'a' (b) Find the c.d.f of X (c) Find** $\mathbf{P(X \le 1.5)}$ **(April/May 2017)** **Solution:** (a) Since $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_{-\infty}^{\infty} f(x)dx = 1$ $\Rightarrow \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$ $\int_0^1 ax\, dx + \int_1^2 a\, dx + \int_2^3 (3a - ax)dx = 1$ $a\left[\frac{x^2}{2}\right]_0^1 + a[x]_1^2 + a\left[3x - \frac{x^2}{2}\right]_2^3 = 1 \Rightarrow a = \frac{1}{2}$ (b) CDF If $0 \le x \le 1$ $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^x \frac{x}{2} dx = \left[\frac{x^2}{4}\right]_0^x = \frac{x^2}{4}$ *If* $1 \le x \le 2$ $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^1 \frac{x}{2}\, dx + \int_1^x \frac{1}{2} dx$ $= \left[\frac{x^2}{4}\right]_0^1 + \left[\frac{x}{2}\right]_1^x = \frac{x}{2} - \frac{1}{4}$ *If* $2 \le x \le 3$ $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^1 \frac{x}{2}\, dx + \int_1^2 \frac{1}{2} dx + \int_2^x \left(\frac{3}{2} - \frac{x}{2}\right) dx$ $= \left[\frac{x^2}{4}\right]_0^1 + \left[\frac{x}{2}\right]_1^2 + \left(\frac{3x}{2} - \frac{x^2}{4}\right)_2^x$ $= \frac{3x}{2} - \frac{x^2}{4} - \frac{5}{4}$

	$$F_X(x)=\begin{cases}0, & x<0\\ \dfrac{x^2}{2}, & 0\le x\le 1\\ \dfrac{x}{2}-\dfrac{1}{4}, & 1\le x\le 2\\ \dfrac{3x}{2}-\dfrac{x^2}{4}-\dfrac{5}{4}, & 2\le x\le 3\\ 1, & x>3\end{cases}$$ (c) $P(X\le 1.5)=F(1.5)$ $\quad(\because F(x)=P(X\le x))$ $$=\frac{1.5}{2}-\frac{1}{4}=0.5$$ **(ii) A continuous random variable X has the p.d.f $f(x)=kx^2e^{-x},\ x\ge 0.$ Find the r[th] moment of X about the origin. Hence find mean and variance of X.** **Solution:** Since $\int_0^\infty kx^2e^{-x}dx=1$ $$\Rightarrow k\left[x^2\left(\frac{e^{-x}}{-1}\right)-(2x)\left(\frac{e^{-x}}{1}\right)+(2)\left(\frac{e^{-x}}{-1}\right)\right]_0^\infty=1$$ $$k\left[-x^2e^{-x}-2x\,e^{-x}-2e^{-x}\right]_0^\infty=1\Rightarrow k\left[(0)-(-2)\right]=1\Rightarrow 2k=1\Rightarrow k=\frac{1}{2}.$$ $$E(X^r)=\mu_r'=\int_0^\infty x^r f(x)dx=\frac{1}{2}\int_0^\infty x^r x^2e^{-x}dx=\frac{1}{2}\int_0^\infty x^{r+2}e^{-x}dx \qquad \left(\because \Gamma n=\int_0^\infty e^{-x}x^{n-1}dx,\ n>0\right)$$ here $n=r+3$ $$=\frac{1}{2}\int_0^\infty e^{-x}x^{(r+3)-1}dx=\frac{1}{2}\Gamma(r+3)=\frac{(r+2)!}{2} \qquad \because \Gamma n=(n-1)!$$ Putting $r=1,\ E(X)=\mu_1'=\frac{3!}{2}=\frac{6}{2}=3$ $$r=2,\ E(X^2)=\mu_2'=\frac{4!}{2}=\frac{24}{2}=12$$ $\therefore$ Mean = $E(X)=\mu_1'=3$; Variance = $E(X^2)-\left[E(X)\right]^2=\mu_2'-\left(\mu_1'\right)^2$ $$\mu_2=12-(3)^2=12-9=3$$
5.	**A random variable** X **has the P.d.f** $f(x)=\begin{cases}2x, 0<x<1\\ 0\ \ , Otherwise\end{cases}$ **Find (a)** $P\left(X<\frac{1}{2}\right)$ **(b)** $P\left(\frac{1}{4}<x<\frac{1}{2}\right)$ **(c)** $P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)$

	Solution: $(a)\ P\left(X<\frac{1}{2}\right)=\int_0^{1/2} f(x)dx=\int_0^{1/2} 2x\,dx=2\left(\frac{x^2}{2}\right)_0^{1/2}=\frac{1}{4}$ $(b)\ P\left(\frac{1}{4}<X<\frac{1}{2}\right)=\int_{1/4}^{1/2} f(x)dx=\int_{1/4}^{1/2} 2xdx=2\left(\frac{x^2}{2}\right)_{1/4}^{1/2}=\frac{1}{4}-\frac{1}{16}=\frac{3}{16}$ $(c)\ P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)=\frac{P\left(X>\frac{3}{4}\cap X>\frac{1}{2}\right)}{P\left(X>\frac{1}{2}\right)}=\frac{P\left(X>\frac{3}{4}\right)}{P\left(X>\frac{1}{2}\right)}$ $P\left(X>\frac{3}{4}\right)=1-P\left(X<\frac{3}{4}\right)=1-\int_0^{3/4} 2xdx=1-2\left(\frac{x^2}{2}\right)_0^{3/4}=1-\left(\frac{9}{16}\right)=\frac{7}{16}$ $P\left(X>\frac{1}{2}\right)=1-P\left(X\le\frac{1}{2}\right)=1-\frac{1}{4}=\frac{3}{4}$ $P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)=\frac{\frac{7}{16}}{\frac{3}{4}}=\frac{7}{12}$
6.	**(i) Out of 800 families with 4 children each, how many families would be expected to have (a) 2 boys and 2 girls (b) atleast 1 boy (c) atmost 2 girls and (d) children of both genders. Assume equal probabilities for boys and girls. (April/May 2019)** **Solution:** Considering each child as a trial, $n=4$ Assuming that birth of a boy is a success, $p=\frac{1}{2}$ $\quad \therefore q=\frac{1}{2}$ Let X denote the number of successes (boys) (a) $P(\text{2 boys and 2 girls})=P(X=2)$ $=4c_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$ $=0.375$ $\therefore$ No. of families having 2 boys and 2 girls $=N.P(X=2)$ $=800(0.375)$ $=300$ (b) $P(\text{at least 1 boy})=P(X\ge 1)$ $=1-P(X<1)$ $=1-P(X=0)$

$$=1-4c_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4$$

$$=0.9375$$

$\therefore$ No. of families having at least 1 boy $=N.P(X\ge 1)$

$$=800(0.9375)$$

$$=750.$$

(c) $P(\text{atmost 2 girls})=P(\text{exactly 0 girl, 1girl, 2 girls})$

$$=P(X=4)+P(X=3)+P(X=2)$$

$$=4c_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0+4c_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1+4c_4\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$

$= 0.6875.$

$\therefore$No. of families having atmost 2 girls $=800(0.6875)=550$.

(d) $P(\text{Children of both genders})=1-P(\text{Children of the same gender})$

$$=1-\left\{P(\text{all are boys})+P(\text{all are girls})\right\}$$

$$=1-\left\{P(X=4)+P(X=0)\right\}$$

$$=1-\left[4c_4\left(\frac{1}{2}\right)^4+4c_0\left(\frac{1}{2}\right)^4\right]$$

$$=0.875.$$

$\therefore$No. of families having children of both genders $=800(0.875)$

$$=700.$$

ii) Derive the moment generating function, mean and variance of Poisson distribution.

Solution:

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!}; x=0,1,2,...;\lambda>0\\ 0, otherwise\end{cases}$$

M.G.F= $M_X(t)=E(e^{tX})$

$$\sum_{x=1}^{\infty}e^{tx}f(x)=\sum_{x=1}^{\infty}e^{tx}\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)=e^{-\lambda}\sum_{x=1}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!}$$

$$=e^{-\lambda}\left[1+\frac{\lambda e^t}{1!}+\frac{\left(\lambda e^t\right)^2}{2!}+\dots\right]$$

$$=e^{-\lambda}e^{\lambda e^t}$$

	$M_X(t) = e^{\lambda(e^t - 1)}$ Mean = $E(X) = \left\{\frac{d}{dt}\left[M_X(t)\right]\right\}_{t=0} = \left\{\frac{d}{dt}\left[e^{\lambda(e^t-1)}\right]\right\}_{t=0} = \left[\left(\lambda e^t\right)e^{\lambda(e^t-1)}\right]_{t=0} = \lambda$ Variance = $Var(X) = E(X^2) - (E(X))^2$ Where $E(X^2) = \left\{\frac{d}{dt}\left[M'_X(t)\right]\right\}_{t=0} = \left\{\frac{d}{dt}\left[\left(\lambda e^t\right)e^{\lambda(e^t-1)}\right]\right\}_{t=0} = \lambda\left[e^t.\left(\lambda e^t\right)e^{\lambda(e^t-1)} + e^{\lambda(e^t-1)}.e^t\right]_{t=0}$ $= \lambda(\lambda+1)$ Variance $= \mathrm{Var}(X) = E(X^2) - (E(X))^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$.
7.	**(i)Derive the moment generating function, mean and variance of Geometric distribution** **Solution:** **Moment Generating Function (MGF)** $M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} p(x)$ $= \sum_{x=1}^{\infty} e^{tx} q^{x-1} p$ $= p[e^t + e^{2t} q + e^{3t} q^2 + \cdots]$ $= pe^t[1 + qe^t + \left(qe^t\right)^2 + \cdots]$ $= pe^t[1 - qe^t]^{-1}$ $= \frac{pe^t}{1 - qe^t}$ **Mean and Variance** $\mu_1' = M_X'(0) = \left[\frac{d}{dt}\left(\frac{pe^t}{\left(1-qe^t\right)}\right)\right]_{t=0} = \left[\left(\frac{pe^t}{\left(1-qe^t\right)^2}\right)\right]_{t=0} = \frac{1}{p}$ $\mu_2' = M_X''(0) = \left[\frac{d^2}{dt^2}\left(\frac{pe^t}{\left(1-qe^t\right)}\right)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{pe^t}{\left(1-qe^t\right)^2}\right)\right]_{t=0} = \frac{1+q}{p^2}$ $Mean = \mu_1' = \frac{1}{p}$ $Variance = \mu_2' - \left(\mu_1'\right)^2 = \frac{1+q}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}$ **(ii) State and prove the memory less property of Geometric distribution.**

Solution:
Memoryless property of geometric distribution.

Statement:

If X is a random variable with geometric distribution, then X lacks memory, in the sense that $P\left[X > s+t / \mathrm{X} > s\right] = P\left[X > t\right] \quad \forall s,t > 0$.

Proof:

The probability mass function of the geometric random variable X is $P(X = x) = q^{x-1}p \quad , \quad x = 1,2,3,\ldots.$

$$P\left[X > s+t / X > s\right] = \frac{P\left[X > s+t \cap X > s\right]}{P\left[X > s\right]} = \frac{P\left[X > s+t\right]}{P\left[X > s\right]} \text{ - - - - - - (1)}$$

$$\therefore P\left[X > t\right] = \sum_{x=t+1}^{\infty} q^{x-1}p = q^t p + q^{t+1}p + q^{t+2}p + \ldots. = q^t p\left[1 + q + q^2 + q^3 + \ldots.\right]$$

$$= q^t p(1-q)^{-1} = q^t p(p)^{-1} = q^t$$

Hence $P\left[X > s+t\right] = q^{s+t}$ and $P\left[X > s\right] = q^s$

$$(1) \Rightarrow P\left[X > s+t / X > s\right] = \frac{q^{s+t}}{q^s} = q^t = P[X > t]$$

$$\Rightarrow P\left[X > s+t / \mathrm{X} > s\right] = P\left[X > t\right] .$$

8. **(i) Let X be a Uniformly distributed R.V over [-5,5]. Determine (a) P(X ≤ 2) (b) P(|X|≤ 2) (c) CDF of X (d) Var (X)**

Solution:
The R.V X ~ U[-5,5].
The p.d.f

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0 \quad , & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} \dfrac{1}{10} & for\ -5 \le x \le 5 \\ 0 & otherwise \end{cases}$$

$(1)\ P(X \le 2) = \int_{-5}^{2} f(x)dx = \int_{-5}^{2} \frac{1}{10}dx = \frac{1}{10}\int_{-5}^{2} dx = \frac{1}{10}[x]_{-5}^{2}$

$= \frac{1}{10}[2+5] = \frac{7}{10}$

(2) $P(|X| \le 2) = P(-2 \le X \le 2)$

$= \int_{-2}^{2} f(x)\,dx = \int_{-2}^{2} \frac{1}{10}\,dx = \frac{1}{10}[x]_{-2}^{2} = \frac{1}{10}[2+2] = \frac{4}{10} = \frac{2}{5}$

(3) *Cumulative distribution function of X*

If $x < -5$

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

If $-5 \le x < 5$

$F(x) = \int_{-5}^{x} f(x)dx = \int_{-5}^{x} \frac{1}{10}dx = \frac{1}{10}[x]_{-5}^{x} = \frac{x+5}{10}$

If $x \ge 5$

$F(x) = \int_{-5}^{5} f(x)dx + \int_{5}^{x} f(x)dx = \int_{-5}^{5} \frac{1}{10}dx + 0 = \frac{1}{10}[x]_{-5}^{5} = \frac{5+5}{10} = 1$

$$f(x) = \begin{cases} 0 & for\ x < -5 \\ \frac{x+5}{10} & for\ -5 \le x \le 5 \\ 1 & for\ x > 5 \end{cases}$$

$(4) Var(X) = \frac{(b-a)^2}{12} = \frac{(5-(-5))^2}{12} = \frac{100}{12} = \frac{25}{3}.$

(ii) Let $\mathbf{P(X = x) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x-1}, x = 1, 2, 3, ...}$ **be the probability mass function of the RV X. Compute**

(a) P(X > 4/ X >2) (b) Find mean and variance

Solution:

(a)To find P(X > 4/ X >2) (By Memoryless property)

	$P(X>4/X>2) = P(X>2)$ $= P(X=3)+P(X=4)+P(X=5)+\cdots$ $= \sum_{X=3}^{\infty} P(X=x)$ $= \sum_{X=3}^{\infty}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x-1}$ $= \left(\frac{3}{4}\right)\left[\left(\frac{1}{4}\right)^2+\left(\frac{1}{4}\right)^3+\left(\frac{1}{4}\right)^4+\cdots\right]$ $= \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left[1+\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)^2+\cdots\right]$ $= \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left[1-\frac{1}{4}\right]^{-1} = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^{-1} = \left(\frac{1}{4}\right)^2$ **(b)To find mean and variance** $E(X)=\frac{1}{p}=\frac{1}{\left(\frac{3}{4}\right)}=\frac{4}{3}$ $Var(X)=\frac{q}{p^2}=\frac{\left(\frac{1}{4}\right)}{\left(\frac{3}{4}\right)^2}=\frac{1}{4}\times\frac{16}{9}=\frac{4}{9}.$
9.	**(i) Trains arrive at a station at 15 minutes interval starting at 4 a.m. If a passenger arrives at a station at a time that is uniformly distributed between 9.00 a.m. and 9.30 a.m., find the probability that he has to wait for the train (i) less than 6 min (ii) more than 10 min.** **Solution:** Let X denotes number of minutes past 9.00 a.m. that the passenger arrives at the stop till 9.30a.m. X ~U[0,30] $\Rightarrow f(x)=\frac{1}{30},\ 0<x<30$ (i)**P(that he has to wait for the train for less than 6 minutes)** $=P\left[(9<x<15)\cup(24<x<30)\right]$ $=\int_9^{15} f(x)dx+\int_{24}^{30} f(x)dx$ $=\int_9^{15}\frac{1}{30}dx+\int_{24}^{30}\frac{1}{30}dx=\frac{1}{30}\left\{\left[x\right]_9^{15}+\left[x\right]_{24}^{30}\right\}=\frac{12}{30}=0.4$

(ii)**P** $\left(\textbf{that he has to wait for the train for } more \textbf{ than } 10 \textbf{ minutes}\right)$

$$= P\left[(0 < x < 5) \cup (15 < x < 20)\right]$$

$$= \int_0^5 f(x)dx + \int_{15}^{20} f(x)dx$$

$$= \int_0^5 \frac{1}{30}dx + \int_{15}^{20} \frac{1}{30}dx = \frac{1}{30}\left\{\left[x\right]_0^5 + \left[x\right]_{15}^{20}\right\} = \frac{10}{30} = 0.3333$$

(ii) A component has an exponential time to failure distribution with mean of 10,000 hours.

(a)The component has already been in operation for its mean life. What is the Probability that it will fail by 15,000 hours?

(b) At 15,000 hours the component is still in operation. What is the probability that it will operate for another 5000 hours?

Solution:

Let X be the random variable denoting the time to failure of the component following exponential distribution with Mean $= 10000$ hours.

$$\therefore \frac{1}{\lambda} = 10,000 \Rightarrow \lambda = \frac{1}{10,000}$$

The p.d.f. of X is $f(x) = \begin{cases} \frac{1}{10,000} e^{-\frac{x}{10,000}} & , x \geq 0 \\ 0 & , otherwise \end{cases}$

(a) Probability that the component will fail by 15,000 hours given that it has already been in operation for its mean life $= P\left[X < 15,000 / X > 10,000\right]$

$$= \frac{P\left[10,000 < X < 15,000\right]}{P\left[X > 10,000\right]} \text{ - - - - - - - - (1)}$$

$$P\left[10,000 < X < 15,000\right] = \int_{10,000}^{15,000} \frac{1}{10000} e^{-\frac{x}{10000}} dx$$

$$= \frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{10000}^{15000}$$

$$= -\left[e^{-\frac{x}{10000}}\right]_{10000}^{15000}$$

$$= -\left[e^{-\frac{15000}{10000}} - e^{-\frac{10000}{10000}}\right]$$

$$= -\left[e^{-\frac{3}{2}} - e^{-1}\right] = e^{-1} - e^{-1.5} \text{ - - - - (2)}$$

	$P\left[X>10,000\right]=\int_{10,000}^{\infty}\frac{1}{10000}e^{-\frac{x}{10000}}dx$ $=\frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{10000}^{\infty}$ $=-\left[e^{-\frac{x}{10000}}\right]_{10000}^{\infty}$ $=-\left[e^{-\infty}-e^{-1}\right]=e^{-1}$ - - - - - - - (3) Sub (2) & (3) in (1) (1) $\Rightarrow P\left[X<15,000/\text{X}>10,000\right]=\frac{e^{-1}-e^{-1.5}}{e^{-1}}=\frac{0.3679-0.2231}{0.3679}=0.3936$. (b) Probability that the component will operate for another 5000 hours given that it is in operation 15,000 hours $=P\left[X>20,000/X>15,000\right]$ $=P\left[X>5000\right]$ [By memoryless property] $=\int_{5000}^{\infty}f(x)\,dx$ $=\int_{5000}^{\infty}\frac{1}{10000}e^{-\frac{x}{10000}}dx$ $=\frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{5000}^{\infty}=-\left[e^{-\frac{x}{10000}}\right]_{5000}^{\infty}$ $=e^{-0.5}=0.6065$
10.	**X is a normal variate with mean 30 and $S.D$ 5 Find (a) $P\left[26\le X\le 40\right]$ (b) $P\left[X\ge 45\right]$ (c) $P\left[\lvert X-30\rvert>5\right]$** **Solution:** X follows N(30, 5) $\therefore \mu=30$ & $\sigma=5$ Let $Z=\frac{X-\mu}{\sigma}$ be the standard normal variate $(a)\ P\left[26\le X\le 40\right]=P\left[\frac{26-30}{5}\le Z\le\frac{40-30}{5}\right]=P\left[-0.8\le Z\le 2\right]$ $=P\left[-0.8\le Z\le 0\right]+P\left[0\le Z\le 2\right]$ $=P\left[0\le Z\le 0.8\right]+\left[0\le Z\le 2\right]$ $=0.2881+0.4772=0.7653$.

$(b)\quad P\left[X \geq 45\right] = P\left[Z \geq \frac{45-30}{5}\right] = P\left[Z \geq \frac{15}{5}\right] = P\left[Z \geq 3\right] = 0.5 - P\left[0 \leq Z \leq 3\right]$

$= 0.5 - 0.49865 \quad = 0.00135$

$(c)\quad P\left[|X-30| > 5\right] = 1 - P\left[|X-30| \leq 5\right] = 1 - P\left[25 \leq X \leq 35\right] = 1 - P\left[-1 \leq Z \leq 1\right]$

$= 1 - 2P\left[0 \leq Z \leq 1\right] = 1 - 2(0.3413) = 1 - 0.6826 = 0.3174$

UNIT – II TWO DIMENSIONAL RANDOM VARIABLES

1. **The following table gives the joint probability distribution of X and Y, find the marginal distribution function of X and Y.**

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

Solution:

Y \ X	1	2	3	p(y)
1	0.1	0.1	0.2	0.4
2	0.2	0.3	0.1	0.6
p(x)	0.3	0.4	0.3	1

The marginal distribution of X is

X	1	2	3
p(x)	0.3	0.4	0.3

The marginal distribution of Y is

Y	1	2
p(y)	0.4	0.6

2. **Determine the value of the constant c if the joint density function of two discrete random variables X and Y is given by** $p(x, y) = cxy,\ x = 1, 2, 3$ and $y = 1, 2, 3.$

Solution:

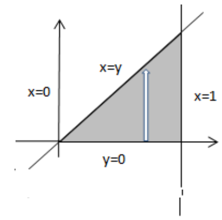
Y \ X	1	2	3	p(y)
1	c	2c	3c	6c
2	2c	4c	6c	12c
3	3c	6c	9c	18c
p(x)	6c	12c	18c	36c

Since p (x, y) is the joint pdf of X and Y, p (x, y) ≥ 0 , for all x ,y

	$\sum_x \sum_y p(x,y) = 1 \Rightarrow 36c = 1 \Rightarrow c = \frac{1}{36}$
3.	**Find the value of k, if the joint density function of (X , Y) is given by** $f(x,y) = \begin{cases} k(1-x)(1-y), 0 < x < 4, \ 1 < y < 5 \\ 0 \quad , otherwise \end{cases}$ **(April/May 2021)** **Solution:** Given the joint pdf of (X , Y) is f(x , y) = k (1 – x) (1 – y), 0 < x < 4, 1 < y < 5 $\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \Rightarrow \int_1^5\int_0^4 k(1-x)(1-y)dxdy = 1 \Rightarrow k\int_1^5\int_0^4 (1-x-y+xy)dxdy = 1$ $\Rightarrow k\int_1^5 \left[x - \frac{x^2}{2} - yx + y\frac{x^2}{2}\right]_0^4 dy = 1 \Rightarrow k\int_1^5 (-4+4y)dy = 1 \Rightarrow k\left[-4y + 4\frac{y^2}{2}\right]_1^5 = 1$ $\Rightarrow k(30+2) = 1 \Rightarrow 32k = 1 \Rightarrow k = \frac{1}{32}$
4.	**The joint p.d.f. of random variable (X,Y) is given as** $f(x,y) = \begin{cases} k, 0 < x, y \le 1 \\ 0, elsewhere \end{cases}$ **. Find k** **(April/May 2019)** **Solution:** $\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1 \Rightarrow \int_0^1\int_0^1 kdxdy = 1 \Rightarrow k\int_0^1 [x]_0^1\, dy = 1$ $\Rightarrow k\int_0^1 dy = 1 \Rightarrow k(y)_0^1 = 1$ $\Rightarrow k = 1$
5.	**The joint pdf of the random variable (X,Y) is given as** $f(x,y) = \frac{1}{2}xe^{-y}, 0 < x < 2, y > 0.$ **Calculate the marginal p.d.f of X.** **(April/May 2021)** **Solution**: The marginal p.d.f of X is given by $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{\infty}\frac{1}{2}xe^{-y}dy = \frac{1}{2}x\int_0^{\infty}e^{-y}dy = \frac{1}{2}x\left[\frac{e^{-y}}{-1}\right]_0^{\infty} = -\frac{1}{2}x[0-1] = \frac{x}{2}$ $f_X(x) = \frac{x}{2}, \quad 0 < x < 2$

6. **If** $f(x,y)=\begin{cases} 8xy, 0<x<1,\ 0<y<x \\ 0 \quad ,elsewhere\end{cases}$ **is the joint PDF of X & Y, find f(y/x).**

Solution:



$$f_X(x)=\int_y f(x,y)dy = \int_{y=0}^{x} 8xydy = \left[8x\frac{y^2}{2}\right]_0^x = 4x^3,\ 0<x<1$$

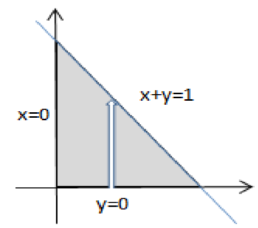
$$f(y/x)=\frac{f(x,y)}{f_X(x)}, = \frac{8xy}{4x^3} = \frac{2y}{x^2},\ 0<y<x, 0<x<1$$

7. **The joint probability density function of bivariate random variable (X , Y) is given by**

$f(x,y)=\begin{cases} 4xy, 0<x<1, \quad 0<y<1 \\ 0 \quad ,elsewhere\end{cases}$ **Find P (X + Y<1)**

Solution:

Given the joint pdf of (X , Y) is $f(x,y)=\begin{cases} 4xy, 0<x<1,\ 0<y<1 \\ 0 \quad ,elsewhere\end{cases}$



$$\therefore P(X+Y<1) = \int_0^1\int_0^{1-x} 4xydydx = 4\int_0^1 x\left[\frac{y^2}{2}\right]_0^{1-x} dx = 2\int_0^1 x(1-x)^2 dx$$

$$= 2\int_0^1 x(1-2x+x^2)dx = 2\int_0^1 (x-2x^2+x^3)dx$$

$$= 2\left[\frac{x^2}{2} - 2\frac{x^3}{3} + \frac{x^4}{4}\right]_0^1 = 2\left[\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right] = \frac{1}{6}$$

8. **If the joint cumulative distribution function of X and Y is given by**

$F(x,y)=(1-e^{-x})(1-e^{-y}), x>0, y>0$ **, find P (1 < X < 2 , 1 < Y < 2)**

Solution:

The joint pdf is $f(x,y)=\frac{\partial^2 F}{\partial x\partial y} = \frac{\partial^2\left[\left(1-e^{-x}\right)\left(1-e^{-y}\right)\right]}{\partial x \partial y} = \frac{\partial}{\partial x}\left(1-e^{-x}\right).e^{-y} = e^{-x}.e^{-y} = e^{-(x+y)}, x>0, y>0$

$$P(1<X<2, 1<Y<2) = \int_1^2\int_1^2 f(x,y)dxdy = \int_1^2\int_1^2 e^{-(x+y)}dxdy = \int_1^2\int_1^2 e^{-x}.e^{-y}dxdy$$

$$= \int_1^2 e^{-x}dx.\int_1^2 e^{-y}dy = \left[-e^{-x}\right]_1^2.\left[-e^{-y}\right]_1^2 = \left(e^{-1}-e^{-2}\right)^2 = \left(\frac{1}{e}-\frac{1}{e^2}\right)^2 = \left(\frac{e-1}{e^2}\right)^2 = 0.054$$

9. **Let X and Y be two random variables having joint density function**

$f(x,y)=\frac{3}{2}(x^2+y^2), 0\le x\le 1, 0\le y\le 1.$ **Determine** $P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right)$

Solution:

$$P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right) = \int_{x=-\infty}^{\frac{1}{2}} \int_{y=\frac{1}{2}}^{\infty} f(x,y)\,dy\,dx = \int_{x=0}^{\frac{1}{2}} \int_{y=\frac{1}{2}}^{1} \frac{3}{2}\left(x^2+y^2\right) dy\,dx$$

$$=\frac{3}{2}\int_0^{\frac{1}{2}}\left[x^2y+\frac{y^3}{3}\right]_{\frac{1}{2}}^{1} dx = \frac{3}{2}\int_0^{\frac{1}{2}}\left[x^2\left(1-\frac{1}{2}\right)+\frac{1}{3}\left(1-\frac{1}{8}\right)\right]dx = \frac{3}{2}\int_0^{\frac{1}{2}}\left[\frac{x^2}{2}+\frac{7}{24}\right]dx$$

$$=\frac{3}{2}\left[\frac{x^3}{6}+\frac{7x}{24}\right]_0^{\frac{1}{2}} = \frac{3}{2}\left[\frac{1}{6}.\frac{1}{8}+\frac{7}{24}.\frac{1}{2}\right] = \frac{3}{2}\left[\frac{8}{48}\right] = \frac{1}{4}$$

10. **If X and Y have joint p.d.f $f(x,y)=e^{-(x+y)}, x>0, y>0$. Check whether X and Y are independent.**
(April/ May 2021)

Solution:

The marginal pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{\infty} e^{-(x+y)}dy = e^{-x}\left[\frac{e^{-y}}{-1}\right]_0^{\infty} = -e^{-x}\left[0-1\right] = e^{-x}$$

Similarly the marginal pdf of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} e^{-(x+y)}dx = e^{-y}\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = -e^{-y}\left[0-1\right] = e^{-y}$$

$$\text{Now, } f_X(x).f_Y(y) = e^{-x}.e^{-y} = e^{-(x+y)} = f(x,y)$$

$$\therefore \text{X and Y are independent.}$$

11. **The joint p.d.f. of Random variables (X,Y) is given as $f(x,y)=\begin{cases} \frac{1}{x}, 0<y<x\le 1 \\ 0, elsewhere\end{cases}$. Find the marginal p.d.f. of Y.**

The marginal pdf of Y is $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_y^1 \frac{1}{x}dx = \left[\log x\right]_y^1 = \log 1 - \log y = -\log y,\ 0<y<1$

12. **Let X and Y be two independent random variables with Var(X) = 9 and Var(Y) = 3. Find Var (4X – 2Y + 6)**

Solution:

Var (4X – 2Y + 6) = 16 Var(X) + 4 Var(Y) = 16(9) + 4(3) = 156

13. **If X has mean 4 and variance 9, while Y has mean –2 and variance 5 and the two are independent find (a) E[XY] (b) E[XY^2]**

Solution:

Given E[X] = 4, E[Y] = –2, $\sigma_X^2 = 9$, $\sigma_Y^2 = 5$, X and Y are independent.

(a) E[XY] = E[X] E[Y] = 4(–2) = –8

(b) E[XY^2] = E[X] E[Y^2]

	$\therefore \sigma_Y^2 = E[Y^2] - [E[Y]]^2 \Rightarrow 5 = E[Y^2] - 4 \Rightarrow E[Y^2] = 9 \therefore E[XY^2] = 4(9) = 36$
14.	**If Y = –2X + 3, find Cov (X, Y).** **Solution:** Cov (X, Y) = E(XY) – E(X) E(Y) = E (X (–2X + 3)) – E(X) {E (–2X + 3)} = E ($-2X^2$ + 3X) – E(X) E (–2X + 3) = $-2E(X^2)$ + 3 E(X) + 2 $(E(X))^2$ – 3E(X) = $2(E(X))^2 - 2E(X^2)$ = –2 var(X)
15.	**Write down the formula for Covariance and Coefficient of correlation between two random variables X and Y.** **(April/May 2019)** **Solution:** Covariance between x and y $Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$ Coefficient of correlation between x and y $r_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$
16.	**The correlation coefficient of two random variables X and Y is $\frac{-1}{4}$ while their variances are 3 and 5. Find the covariance.** **Solution:** Given $r_{xy} = \frac{-1}{4}$, $\sigma_X^2 = 3, \sigma_Y^2 = 5$, $r_{xy} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}, \sigma_X \neq 0, \sigma_Y \neq 0$ $-\frac{1}{4} = \dfrac{Cov(X,Y)}{\sqrt{3}.\sqrt{5}} \Rightarrow Cov(X,Y) = -\frac{1}{4}.\sqrt{3}.\sqrt{5} = -0.968$
17.	**The lines of regression in a bivariate distribution are X + 9Y = 7 and Y + 4X = $\frac{49}{3}$. Find the coefficient of correlation.** **Solution:** x + 9y – 7 = 0 ----------- (1) y + 4x – $\frac{49}{3}$ = 0 ----------- (2) Let (1) be the regression line of Y on X and (2) be the regression line of X on Y. $\therefore y = -\frac{1}{9}x + \frac{7}{9} \Rightarrow b_1 = -\frac{1}{9}$ $x = -\frac{1}{4}y + \frac{49}{12} \Rightarrow b_2 = -\frac{1}{4}$ $\therefore r = \pm\sqrt{b_1 b_2} = \pm\sqrt{\left(-\frac{1}{9}\right)\left(-\frac{1}{4}\right)} = \pm\sqrt{\frac{1}{36}} = -\frac{1}{6} < 1$ Since both regression coefficients are negative, correlation coefficient is negative.
18.	**The regression equations are 3x + 2y = 26 and 6x + y = 31. Find the mean of X and Y.** **(Nov/Dec 2017)** **Solution:**

	Regression lines pass through the mean values of X and Y. Solving the two equations we get the mean values. Let 3x + 2y = 26 --------------(1) 6x + y = 31 --------------(2) Multiply equation (2) by 2 and subtract equation (2) $3x + 2y = 26$ $12x + 2y = 62$ $(-)\quad(-)\quad(-)$ - - - - - - - - - - $-9x = -36$ $x = 4$ Substitute in equation (1) 3 (4) – 2y = 26 ⇒ y = 7. ∴mean value of X = 4 and mean value of Y = 7
19.	**The two lines of regression are 4x – 5y + 33 = 0 and 20x – 9y = 107. Calculate the coefficient of correlation between X and Y.** **Solution:** 4x – 5y + 33 = 0 ----------- (1) 20x – 9y = 107 ----------- (2) Let (1) be the regression line of Y on X and let (2) be the regression line of X on Y. $\therefore y = \frac{4}{5}x + \frac{33}{5} \Rightarrow b_1 = \frac{4}{5}$ $x = \frac{9}{20}y + \frac{107}{20} \Rightarrow b_2 = \frac{9}{20}$ $\therefore r = \sqrt{b_1 b_2} = \sqrt{\frac{4}{5}.\frac{9}{20}} = \sqrt{\frac{9}{25}} = \frac{3}{5} = 0.6 < 1$
20.	**State central limit theorem** **Statement:** If $X_1, X_2....X_n$ is a sequence of independent random variable $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$, $i = 1, 2, ...n$ and if $S_n = X_1 + X_2 + ... + X_n$ then under certain general conditions S_n follows a Normal distribution with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$ as $n \to \infty$.
	PART-B
1.	**The joint probability mass function of (X,Y) is given by $P(x, y) = k(2x + 3y)$, x = 0,1,2; y = 1,2,3. Find the Marginal distributions and Conditional probability distributions. Also find the probability distribution of X +Y.** **Solution:** The joint probability mass function table:

Y / X	1	2	3	$P_X(x)$

0	$3k$	$6k$	$9k$	18k
1	$5k$	$8k$	$11k$	24k
2	$7k$	$10k$	$13k$	30k
$P_Y(y)$	15k	24k	33k	72k

We know that $\sum_{j=1}^{3}\sum_{i=0}^{2} P(x_i, y_j) = 1 \Rightarrow 72k = 1 \Rightarrow \boxed{k = \frac{1}{72}}$

MARGINAL DISTRIBUTION OF *X*:

$$P(X=0) = \sum_{j=1}^{3} P(X=0, Y=j) = 3k + 6k + 9k = 18k = \frac{18}{72}$$

$$P(X=1) = \sum_{j=1}^{3} P(X=1, Y=j) = 5k + 8k + 11k = 24k = \frac{24}{72}$$

$$P(X=2) = \sum_{j=1}^{3} P(X=2, Y=j) = 7k + 10k + 13k = 30k = \frac{30}{72}$$

MARGINAL DISTRIBUTION OF *Y*:

$$P(Y=1) = \sum_{i=0}^{2} P(X=i, Y=1) = 3k + 5k + 7k = 15k = \frac{15}{72}$$

$$P(Y=2) = \sum_{i=0}^{2} P(X=i, Y=2) = 6k + 8k + 10k = 24k = \frac{24}{72}$$

$$P(Y=3) = \sum_{i=0}^{2} P(X=i, Y=3) = 9k + 11k + 13k = 33k = \frac{33}{72}$$

CONDITIONAL DISTRIBUTION OF *X* GIVEN *Y*=1:

(i) If $X=0,\ P(X=0/Y=1) = \frac{P(X=0, Y=1)}{P(Y=1)} = \frac{3k}{15k} = \frac{1}{5}$

(ii) $X=1,\ P(X=1/Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{5k}{15k} = \frac{1}{3}$

(iii) $X=2,\ P(X=2/Y=1) = \frac{P(X=2, Y=1)}{P(Y=1)} = \frac{7k}{15k} = \frac{7}{15}$.

CONDITITNAL DISTRIBUTION OF *X* GIVEN *Y*=2:

(i) $X=0,\ P(X=0/Y=2) = \frac{P(X=0, Y=2)}{P(Y=2)} = \frac{6k}{24k} = \frac{1}{4}$

(ii) $X=1,\ P(X=1/Y=2) = \frac{P(X=1, Y=2)}{P(Y=2)} = \frac{8k}{24k} = \frac{1}{3}$

(iii) $X=2,\ P(X=2/Y=2) = \frac{P(X=2, Y=2)}{P(Y=2)} = \frac{10k}{24k} = \frac{5}{12}$

CONDITIONAL DISTRIBUTION OF *X* GIVEN *Y*=3:

(i) $X=0,\ P(X=0/Y=3)=\dfrac{P(X=0,Y=3)}{P(Y=3)}=\dfrac{9k}{33k}=\dfrac{3}{11}$

(ii) $X=1,\ P(X=1/Y=3)=\dfrac{P(X=1,Y=3)}{P(Y=3)}=\dfrac{11k}{33k}=\dfrac{1}{3}$

(iii) $X=2,\ P(X=2/Y=3)=\dfrac{P(X=2,Y=3)}{P(Y=3)}=\dfrac{13k}{33k}=\dfrac{13}{33}.$

CONDITIONAL DISTRIBUTION OF *Y* GIVEN *X*=0:

(i) $Y=1,\quad P(Y=1/X=0)=\dfrac{P(Y=1,X=0)}{P(X=0)}=\dfrac{3k}{18k}=\dfrac{1}{6}$

(i) $Y=2,\quad P(Y=2/X=0)=\dfrac{P(Y=2,X=0)}{P(X=0)}=\dfrac{6k}{18k}=\dfrac{1}{3}$

(ii) $Y=3,\quad P(Y=3/X=0)=\dfrac{P(Y=3,X=0)}{P(X=0)}=\dfrac{9k}{18k}=\dfrac{1}{2}.$

CONDITIONAL DISTRIBUTION OF *Y* GIVEN *X*=1:

(i) $Y=1,\quad P(Y=1/X=1)=\dfrac{P(Y=1,X=1)}{P(X=1)}=\dfrac{5k}{24k}=\dfrac{5}{24}$

(ii) $Y=2,\quad P(Y=2/X=1)=\dfrac{P(Y=2,X=1)}{P(X=1)}=\dfrac{8k}{24k}=\dfrac{1}{3}$

(iii) $Y=3,\quad P(Y=3/X=1)=\dfrac{P(Y=3,X=1)}{P(X=1)}=\dfrac{11k}{24k}=\dfrac{11}{24}.$

CONDITIONAL DISTRIBUTION OF *Y* GIVEN *X*=2:

(i) $Y=1,\quad P(Y=1/X=2)=\dfrac{P(Y=1,X=2)}{P(X=2)}=\dfrac{7k}{30k}=\dfrac{7}{30}$

(ii) $Y=2,\quad P(Y=2/X=2)=\dfrac{P(Y=2,X=2)}{P(X=2)}=\dfrac{10k}{30k}=\dfrac{1}{3}$

(iii) $Y=3,\quad P(Y=3/X=2)=\dfrac{P(Y=3,X=2)}{P(X=2)}=\dfrac{13k}{30k}=\dfrac{13}{30}.$

PROBABILITY DISTRIBUTION OF $(X+Y)$:

(i) $P(X+Y=1)=P(X=0,Y=1)=3k=\dfrac{3}{72}$

(ii) $P(X+Y=2)=P(X=0,Y=2)+P(X=1,Y=1)=6k+5k=\dfrac{11}{72}$

(iii) $P(X+Y=3)=P(X=0,Y=3)+P(X=1,Y=2)+P(X=2,Y=1)$

	$=9k+8k+7k=24k=\frac{24}{72}$ (iv) $P(X+Y=4)=P(X=1,Y=3)+P(X=2,Y=2)=11k+10k=21k=\frac{21}{72}$ (v) $P(X+Y=5)=P(X=2,Y=3)=13k=\frac{13}{72}$
2.	**The joint probability mass function of (X, Y) is given by**

X \ Y	0	1	2
0	0.1	0.04	0.06
1	0.2	0.08	0.12
2	0.2	0.08	0.12

(a) Compute the marginal probability mass function of X and Y (b) Check if X and Y are independent.

Solution:

a) Marginal Probability Mass function of X and y

X/Y	0	1	2	P(x)
0	0.1	0.04	0.06	0.2
1	0.2	0.08	0.12	0.4
2	0.2	0.08	0.12	0.4
P(y)	0.5	0.2	0.3	1

Marginal function of X:

X	0	1	2
P(x)	0.2	0.4	0.4

Marginal function of Y:

Y	0	1	2
P(y)	0.5	0.2	0.3

b) To check X & Y are independent:

Choose $X=1$, $Y=1$, then

$$P(X,Y)=P_X(x)\,P_Y(y)$$

$$P(1,1)=P_X(1)\,P_Y(1)$$

$$0.08\neq(0.4)(0.2)$$

Therefore X & Y are not independent

3. **The joint probability density function of a two dimensional random variable (X , Y) is given by** $f(x,y)=xy^2+\frac{x^2}{8},0\leq x\leq 2,0\leq y\leq 1$ **. Compute (a) P(X > 1) (b)** $P\left(Y<\frac{1}{2}\right)$ **(c)** $P\left(X<1/Y<\frac{1}{2}\right)$ **(d) P(X < Y) (e)** $P(X+Y\leq 1)$ **(f) Are X and Y independent?**

Solution:

The marginal pdf of X is

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^1\left(xy^2+\frac{x^2}{8}\right)dy=\left[x\frac{y^3}{3}+\frac{x^2}{8}y\right]_0^1=\frac{x}{3}+\frac{x^2}{8}=\frac{x}{24}(8+3x),0\le x\le 2$$

The marginal pdf of Y is

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_0^2\left(xy^2+\frac{x^2}{8}\right)dx=\left[\frac{x^2}{2}y^2+\frac{x^3}{24}\right]_0^2=2y^2+\frac{1}{3},0\le y\le 1$$

(a) P(X > 1) = $\int_1^{\infty} f_X(x)\,dx=\int_1^2\left(\frac{x}{3}+\frac{x^2}{8}\right)dx=\left[\frac{x^2}{6}+\frac{x^3}{24}\right]_1^2=\frac{1}{6}(4-1)+\frac{1}{24}(8-1)=\frac{3}{6}+\frac{7}{24}=\frac{19}{24}$

$$(b)\ P\left(Y<\frac{1}{2}\right)=\int_{-\infty}^{1/2} f_y(Y)\,dy=\int_0^{1/2}\left(2y^2+\frac{1}{3}\right)dy=\left(\frac{2y^3}{3}+\frac{y}{3}\right)_0^{1/2}=\frac{1}{4}$$

$$(c)\ P\left(X<1/Y<\frac{1}{2}\right)=\frac{P\left(X<1,Y<\frac{1}{2}\right)}{P\left(Y<\frac{1}{2}\right)}$$

$$P\left(X<1,Y<\frac{1}{2}\right)=\int_0^1\left(\int_0^{1/2} f(x,y)\,dy\,dx\right)=\int_0^1\left(\int_0^{1/2}\left(xy^2+\frac{x^2}{8}\right)dy\,dx\right)=\int_0^1\left(x\frac{y^3}{3}+\frac{x^2}{8}y\right)_0^{1/2}dx$$

$$=\int_0^1\left(\frac{x}{3}.\frac{1}{8}+\frac{x^2}{8}.\frac{1}{2}\right)dx=\frac{1}{8}\left(\frac{x^2}{6}+\frac{x^3}{6}\right)_0^1=\frac{1}{8}\left(\frac{2}{6}\right)=\frac{1}{24}$$

(d)To find P(X <Y)

$$P(X<Y)=\int_0^1\int_0^y\left(xy^2+\frac{x^2}{8}\right)dx\,dy=\int_0^1\left(\frac{x^2y^2}{2}+\frac{x^3}{24}\right)_0^y dy$$

$$=\int_0^1\left(\frac{y^4}{2}+\frac{y^3}{24}\right)dy=\left[\frac{y^5}{10}+\frac{y^4}{96}\right]_0^1=\frac{1}{10}+\frac{1}{96}=\frac{53}{480}$$

	(e) $P(X+Y\le 1)=\int_0^1\left(\int_0^{1-y} f(x,y)dxdy\right)=\int_0^1\left(\int_0^{1-y}(xy^2+\frac{x^2}{8})dxdy\right)=\int_0^1\left(\frac{x^2}{2}y^2+\frac{x^3}{24}\right)_0^{1-y}dy$ $=\int_0^1\left(\frac{(1-y)^2}{2}y^2+\frac{(1-y)^3}{24}\right)dy=\frac{1}{2}\int_0^1(y^2-2y^3+y^4)dy+\frac{1}{24}\int_0^1(1-y)^3dy$ $=\frac{1}{2}\left(\frac{y^3}{3}-2.\frac{y^4}{4}+\frac{y^5}{5}\right)_0^1+\frac{1}{24}\left(\frac{(1-y)^4}{-4}\right)_0^1=\frac{1}{2}\left(\frac{1}{3}-\frac{1}{2}+\frac{1}{5}\right)_0^1-\frac{1}{24\times 4}(0-1)_0^1$ $=\frac{1}{60}+\frac{1}{96}=\frac{13}{480}$ (f) Consider $f_X(x).\ f_Y(y)=\frac{x}{24}(8+3x).\ 2y^2+\frac{1}{3}\neq$ f(x , y) $\therefore$ X and Y are not independent.
4.	**Find the constant k such that $f(x,y)=\begin{cases}k(x+1)e^{-y}, & 0<x<1, y>0\\ 0 & ,otherwise\end{cases}$ is a joint probability density function of the continuous random variable (X,Y). Are X and Y Independent Random variables? Justify.** **Solution:** To find k : Given that f(x,y) is pdf of (X,Y) $\therefore f(x,y)\ge 0$, for all x ,y and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1\Rightarrow\int_0^{\infty}\int_0^1 k(x+1)e^{-y}dxdy=1$ $\Rightarrow k\int_0^1(x+1)dx.\int_0^{\infty}e^{-y}dy=1$ $\Rightarrow k\left[\frac{x^2}{2}+x\right]_0^1.\left[-e^{-y}\right]_0^{\infty}=1$ $\Rightarrow k\left(\frac{3}{2}\right)(1)=1\Rightarrow k=\frac{2}{3}$ $\therefore\ f(x,y)=\begin{cases}\frac{2}{3}(x+1)e^{-y}, & 0<x<1, y>0\\ 0 & ,otherwise\end{cases}$ The marginal PDF of X is $f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_0^{\infty}\frac{2}{3}(x+1)e^{-y}dy=\frac{2}{3}(x+1)\left[-e^{-y}\right]_0^{\infty}=\frac{2}{3}(x+1)\left[-e^{-\infty}+e^0\right]=\frac{2}{3}(x+1)(1)=\frac{2}{3}(x+1), 0<x<1$ The marginal PDF of Y is $f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx=\int_0^1\frac{2}{3}(x+1)e^{-y}dx=\frac{2}{3}e^{-y}\left[\frac{x^2}{2}+x\right]_0^1=\frac{2}{3}e^{-y}\left[\frac{1}{2}+1\right]=\frac{3}{2}.\frac{2}{3}e^{-y}=e^{-y},\ 0<y<\infty$

	Consider $f_X(x)\,.\,f_Y(y) = \frac{2}{3}(x+1)\,.\,e^{-y} = f(x,y)$ $\therefore$ X and Y are independent
5.	**(i)Let the joint p.d.f. of random variable (X,Y) be given as** $f(x,y) = \begin{cases} cxy^2, & 0 \le x \le y \le 1 \\ 0, & elsewhere \end{cases}$. **Determine** **(a) The value of c (b) the marginal p.d.f of X and Y (c) the conditional p.d.f. of X given Y = y.** **Solution:** $(a) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \Rightarrow \int_0^1\int_0^1 cxy^2dxdy = 1 \Rightarrow c\int_0^1 xdx\int_0^1 y^2dy = 1$ $\Rightarrow c\left[\left[\frac{x^2}{2}\right]_0^1\left[\frac{y^3}{3}\right]_0^1\right] = 1$ $\Rightarrow c\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = 1 \Rightarrow c = 6$ $(b)\ f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^1 6xy^2dy = 6x\left(\frac{y^3}{3}\right)_x^1 = 6x\left(\frac{1}{3} - \frac{x^3}{3}\right) = 2x\left(1 - x^3\right), 0 \le x \le 1$ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 6xy^2dx = 6y^2\left(\frac{x^2}{2}\right)_0^y = 6y^2\left(\frac{y^2}{2}\right) = 3y^4, 0 \le y \le 1$ (c) Conditional p.d.f of X on Y $f(x/y) = \frac{f(x,y)}{f_Y(y)} = \frac{6xy^2}{3y^4} = \frac{2x}{y^2}$ **(ii)Two random variables X and Y have the joint probability density function** $f(x,y) = kxye^{-(x^2+y^2)}, x > 0, y > 0$**. Find the value of k and also prove that X and Y are independent.** **Solution:** Given the joint pdf $f(x,y) = kxye^{-(x^2+y^2)}, x > 0, y > 0$ $\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ $\Rightarrow \int_0^{\infty}\int_0^{\infty} kxye^{-\left(x^2+y^2\right)}\,dx\,dy = 1$ $\Rightarrow k\int_0^{\infty} xe^{-x^2}dx.\int_0^{\infty} ye^{-y^2}dy = 1$

	Put $x^2 = u$, Put $y^2 = v$ $\therefore 2xdx = du$, $\therefore 2ydy = dv$ When $x = 0, u = 0$ When $y = 0, v = 0$ $x = \infty, u = \infty$ $y = \infty, v = \infty$ $\therefore \frac{k}{2}\int_0^\infty e^{-u}\,du . \frac{1}{2}\int_0^\infty e^{-v}\,dv = 1$ $\Rightarrow \frac{k}{4}\left[\frac{e^{-u}}{-1}\right]_0^\infty . \left[\frac{e^{-v}}{-1}\right]_0^\infty = 1$ $\Rightarrow \frac{k}{4}[0+1].[0+1] = 1$ $\Rightarrow k = 4$ $\therefore$ the joint pdf is $f(x,y) = 4xye^{-(x^2+y^2)}, x > 0, y > 0$ The marginal pdf of X is $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^\infty 4xye^{-(x^2+y^2)}dy$ $= 4xe^{-x^2}\int_0^\infty ye^{-y^2}dy$ $= 4xe^{-x^2}\left[\frac{1}{2}\right] = 2xe^{-x^2}, x > 0$ The marginal pdf of Y is $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^\infty 4xye^{-(x^2+y^2)}dx$ $= 4ye^{-y^2}\int_0^\infty xe^{-x^2}dx$ $= 4ye^{-y^2}\left[\frac{1}{2}\right] = 2ye^{-y^2}, y > 0$ Now, $f_X(x).f_Y(y) = \left(2xe^{-x^2}\right).\left(2ye^{-y^2}\right)$ $= 4xye^{-(x^2+y^2)}$ $f_X(x).f_Y(y) = f(x,y)$ $\Rightarrow$ X and Y are independent
6.	**The joint density function of two random variables X and Y is given by**

	$$f(x,y)=\begin{cases}\frac{6}{7}\left(x^2+\frac{xy}{2}\right), 0\le x\le 1,\ 0\le y\le 2\\ 0 \quad ,elsewhere\end{cases}.$$ **(a) Compute the marginal density function of X and Y? (b) Find E(X) and E(Y)** **(c)** $P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right)$ **Solution:** (a) The marginal pdf of X is $$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^2\frac{6}{7}\left(x^2+\frac{xy}{2}\right)dy=\frac{6}{7}\left[x^2y+\frac{x}{2}\frac{y^2}{2}\right]_0^2=\frac{6}{7}\left[2x^2+x\right], 0\le x\le 1$$ The marginal pdf of Y is $$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_0^1\frac{6}{7}\left(x^2+\frac{xy}{2}\right)dx=\frac{6}{7}\left[\frac{x^3}{3}+\frac{x^2}{2}\frac{y}{2}\right]_0^1=\frac{6}{7}\left[\frac{1}{3}+\frac{y}{4}\right], 0\le y\le 2$$ $$(b)\ E(X)=\int_{-\infty}^{\infty} x\, f_X(x)\,dx$$ $$=\frac{6}{7}\int_0^1 x\left(2x^2+x\right)dx=\frac{6}{7}\int_0^1\left(2x^3+x^2\right)dx=\frac{6}{7}\left(\frac{2x^4}{4}+\frac{x^3}{3}\right)_0^1=\frac{5}{7}$$ $$E(Y)=\int_{-\infty}^{\infty} y\, f_Y(y)\,dy$$ $$=\frac{6}{7}\int_0^2 y\left(\frac{1}{3}+\frac{y}{4}\right)dy=\frac{6}{7}\int_0^2\left(\frac{1}{3}y+\frac{y^2}{4}\right)dy=\frac{6}{7}\left(\frac{y^2}{6}+\frac{y^3}{12}\right)_0^2=\frac{8}{7}$$ **(c)** $$P(X<\frac{1}{2}, Y>\frac{1}{2})=\int_{-\infty}^{\frac{1}{2}}\int_{\frac{1}{2}}^{\infty} f(x,y)\,dy\,dx=\int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{2}\frac{6}{7}\left(x^2+\frac{xy}{2}\right)dy\,dx=\int_0^{\frac{1}{2}}\frac{6}{7}\left[x^2y+\frac{x}{2}\frac{y^2}{2}\right]_{\frac{1}{2}}^{2}dx$$ $$=\int_0^{\frac{1}{2}}\frac{6}{7}\left[2x^2+x-\frac{x^2}{2}-\frac{x}{16}\right]dx=\int_0^{\frac{1}{2}}\frac{6}{7}\left[\frac{3x^2}{2}+\frac{15x}{16}\right]dx=\frac{6}{7}\left[\frac{x^3}{2}+\frac{15x^2}{32}\right]_0^{\frac{1}{2}}$$ $$=\frac{6}{7}\left[\frac{1}{16}+\frac{15}{128}\right]=\frac{6}{7}\cdot\frac{23}{128}=\frac{69}{448}$$
7.	**If** X **and** Y **are two random variables having joint probability density function** $$f(x,y)=\begin{cases}\frac{1}{8}(6-x-y),\ 0<x<2,\ 2<y<4\\ 0 \quad \text{otherwise}\end{cases}$$ **. Find (a)** $P(X<1\cap Y<3)$ **(b)** $P(X+Y<3)$

(c) $P(X<1/Y<3)$. **(April/May 2021)**

Solution:

(a)To find $P(X<1\cap Y<3)$

$$P(X<1\cap Y<3)=\int_{y=-\infty}^{y=3}\int_{x=-\infty}^{x=1} f(x,y)\,dxdy$$

$$=\int_{y=2}^{y=3}\int_{x=0}^{x=1}\frac{1}{8}(6-x-y)\,dxdy$$

$$=\frac{1}{8}\int_{2}^{3}\int_{0}^{1}(6-x-y)\,dxdy$$

$$=\frac{1}{8}\int_{2}^{3}\left[6x-\frac{x^2}{2}-xy\right]_0^1 dy$$

$$=\frac{1}{8}\int_{2}^{3}\left[\frac{11}{2}-y\right]dy=\frac{1}{8}\left[\frac{11y}{2}-\frac{y^2}{2}\right]_2^3$$

$$=\frac{1}{8}\left[\left(\frac{33}{2}-\frac{9}{2}\right)-\left(\frac{22}{2}-\frac{4}{2}\right)\right]$$

$$P(X<1\cap Y<3)=\frac{3}{8}$$

(b). $P(X+Y<3)=\int_{0}^{1}\int_{2}^{3-x}\frac{1}{8}(6-x-y)\,dydx$

$$=\frac{1}{8}\int_{0}^{1}\left[6y-xy-\frac{y^2}{2}\right]_2^{3-x}dx$$

$$=\frac{1}{8}\int_{0}^{1}\left[6(3-x)-x(3-x)-\frac{(3-x)^2}{2}-[12-2x-2]\right]dx$$

$$=\frac{1}{8}\int_{0}^{1}\left[18-6x-3x+x^2-\frac{(9+x^2-6x)}{2}-(10-2x)\right]dx$$

$$=\frac{1}{8}\int_{0}^{1}\left[18-9x+x^2-\frac{9}{2}-\frac{x^2}{2}+\frac{6x}{2}-10+2x\right]dx$$

$$=\frac{1}{8}\int_{0}^{1}\left[\frac{7}{2}-4x+\frac{x^2}{2}\right]dx$$

$$=\frac{1}{8}\left[\frac{7x}{2}-\frac{4x^2}{2}+\frac{x^3}{6}\right]_0^1=\frac{1}{8}\left[\frac{7}{2}-2+\frac{1}{6}\right]$$

$$=\frac{1}{8}\left[\frac{21-12+1}{6}\right]=\frac{1}{8}\left(\frac{10}{6}\right)=\frac{5}{24}.$$

	(c).To find $P\left(X<1 \diagup Y<3\right)=\dfrac{P(x<1\cap y<3)}{P(y<3)}$ The Marginal density function of Y is $f_Y(y)=\int_0^2 f(x,y)dx$ $$=\int_0^2 \frac{1}{8}(6-x-y)dx$$ $$=\frac{1}{8}\left[6x-\frac{x^2}{2}-yx\right]_0^2$$ $$=\frac{1}{8}\left[12-2-2y\right]$$ $$=\frac{5-y}{4},\ 2<y<4.$$ $$P\left(X<1 \diagup Y<3\right)=\frac{\int_{x=0}^{x=1}\int_{y=2}^{y=3}\frac{1}{8}(6-x-y)\,dxdy}{\int_{y=2}^{y=3} f_Y(y)\,dy}$$ $$=\frac{\frac{3}{8}}{\int_2^3\left(\frac{5-y}{4}\right)dy}=\frac{\frac{3}{8}}{\frac{1}{4}\left[5y-\frac{y^2}{2}\right]_2^3}$$ $$=\frac{3}{8}\times\frac{8}{5}=\frac{3}{5}.$$
8.	**Two dimensional random variable (X, Y) have the joint probability density function** $f(x,y)=\begin{cases}8xy, 0<x<y<1\\ 0\ \ ,elsewhere\end{cases}$. **Find (a)** $P\left(X<\frac{1}{2}\cap Y<\frac{1}{4}\right)$ **(b) the marginal and conditional distributions. (c) Are X and Y independent?** **Solution:** (a) $P\left(X<\frac{1}{2}\cap Y<\frac{1}{4}\right)=\int_0^{\frac{1}{4}}\int_0^{y}8xy\,dx\,dy=8\int_0^{\frac{1}{4}}\left[\frac{x^2}{2}y\right]_0^y dy$ $$=4\int_0^{\frac{1}{4}}y^3dy=\left[y^4\right]_0^{\frac{1}{4}}=\frac{1}{256}$$ (b) The marginal pdf of X is $f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy=\int_x^1 8xy\,dy=8x\left[\frac{y^2}{2}\right]_x^1=8x\left[\frac{1}{2}-\frac{x^2}{2}\right]=4x(1-x^2),\ 0<x<1$ The marginal pdf of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 8xy\,dx = 8y\left[\frac{x^2}{2}\right]_0^y = 8y\left[\frac{y^2}{2}\right] = 4y^3,\ 0 < y < 1$$

The conditional probability density function of X given Y is

$$f(x/y) = \frac{f(x,y)}{f_Y(y)} = \frac{8xy}{4y^3} = \frac{2x}{y^2},\ 0 < x < y,\ 0 < y < 1$$

The conditional probability density function of Y given X is

$$f(y/x) = \frac{f(x,y)}{f_X(x)} = \frac{8xy}{4x(1-x^2)} = \frac{2y}{(1-x^2)},\ x < y < 1,\ 0 < x < 1$$

(c) To check whether X and Y are independent.

$$f_X(x).f_Y(y) = 4x(1-x^2).4y^3 = 16xy^3(1-x^2) \neq f(x,y)$$

$\therefore$ X and Y are not independent.

9. **(i) A joint probability mass function of the discrete random variable X and Y is given as**

P(X = x , Y = y) = $\begin{cases} \frac{x+y}{32}, & x = 1,2,\ y = 1,2,3,4 \\ 0 & , otherwise \end{cases}$ **. Compute the covariance of X and Y.**

Solution:

Y \ X	1	2	$P_Y(y)$
1	2/32	3/32	5/32
2	3/32	4/32	7/32
3	4/32	5/32	9/32
4	5/32	6/32	11/32
$P_X(x)$	14/32	18/32	1

The marginal distribution of X is

X	1	2
P(x)	14/32	18/32

The marginal distribution of Y is

Y	1	2	3	4
P(y)	5/32	7/32	9/32	11/32

$$E(X) = \sum_x xP(x) = 1\left(\frac{14}{32}\right) + 2\left(\frac{18}{32}\right) = \frac{25}{16}$$

$$E(Y) = \sum_y yP(y) = 1\left(\frac{5}{32}\right) + 2\left(\frac{7}{32}\right) + 3\left(\frac{9}{32}\right) + 4\left(\frac{11}{32}\right) = \frac{45}{16}$$

$$E(XY) = \sum_x\sum_y xyP(x,y) = 1\left(\frac{2}{32}\right) + 2\left(\frac{3}{32}\right) + 2\left(\frac{3}{32}\right) + 4\left(\frac{4}{32}\right) + 3\left(\frac{4}{32}\right) + 6\left(\frac{5}{32}\right) + 4\left(\frac{5}{32}\right) + 8\left(\frac{6}{32}\right)$$

$$= \frac{35}{8}$$

Cov(X , Y) = E(XY) – E(X)E(Y) = $\frac{35}{8} - \left(\frac{25}{16}\right)\left(\frac{45}{16}\right) = -0.0195$.

(ii) The joint pdf of the random variables X and Y is defined as $f(x,y) = \begin{cases} 25e^{-5y}, & 0 < x < 0.2, y > 0 \\ 0 & , elsewhere \end{cases}$

(a) Find the marginal Probability density function of X and Y (b) Covariance of (X,Y)

Solution:

(a) The marginal pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^{\infty} 25e^{-5y}\,dy$$

$$= 25\left(\frac{-e^{-5y}}{5}\right)_0^{\infty} = 5\left(-e^{\infty} + e^0\right) = 5, 0 < x < 0.2$$

The marginal pdf of Y is

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

$$= \int_0^{0.2} 25e^{-5y}\,dx$$

$$= 25e^{-5y}(x)_0^{0.2} = 25e^{-5y}(0.2) = 5e^{-5y}, 0 < y < \infty$$

(b) $E(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx$

$$= \int_0^{0.2} 5x\,dx = 5\left(\frac{x^2}{2}\right)_0^{0.2} = 0.1$$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)dy$$

$$= \int_0^{\infty} 5e^{-5y}\,dy = 5\left(\frac{-ye^{-5y}}{5} - \frac{e^{-5y}}{25}\right)_0^{\infty} = 0.2$$

$$E(XY) = \int_0^{\infty}\int_0^{0.2} xy(25e^{-5y})dx\,dy$$

$$= \int_0^{\infty} 25e^{-5y} \int_0^{0.2} x\,dx = 25\left(\frac{-ye^{-5y}}{5} - \frac{e^{-5y}}{25}\right)_0^{\infty}\left(\frac{x^2}{2}\right)_0^{0.2}$$

$$=(0.2)(0.02)$$

$$=0.004$$

Cov(x,y)=E(XY)-E(X)E(Y)=0.004-(0.1)(0.2)=-0.016

	$=25\left[-ye^{-y}-e^{-y}\right]_0^{\infty}\cdot\left[\dfrac{x^2}{2}\right]_0^{0.2} = 25\left[0+1\right]\cdot\left[\dfrac{0.04}{2}\right] = (25)(0.02) = 0.5$ Cov(x,y) = E(XY) – E(X)E(Y) = 0.5 – (0.5)(5) = –2
10.	**(i) Calculate the correlation coefficient for the following heights in inches of fathers (x) and their son (y):** (April/May 2021)

x:	**65**	**66**	**67**	**67**	**68**	**69**	**70**	**72**
y:	**67**	**68**	**65**	**68**	**72**	**72**	**69**	**71**

Solution:

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4359	4624
67	65	4355	4489	4285
67	68	4556	4355	4359
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
$\sum X = 544$	$\sum Y = 552$	$\sum XY = 37560$	$\sum X^2 = 37028$	$\sum Y^2 = 38132$

$$\overline{X} = \frac{\Sigma x}{n} = \frac{544}{8} = 68$$

$$\overline{Y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69$$

$$\overline{X}\,\overline{Y} = 68\times 69 = 4692$$

$$\sigma_X = \sqrt{\frac{1}{n}\Sigma x^2 - \overline{X}^2} = \sqrt{\frac{1}{8}(37028) - 68^2} = \sqrt{4628.5 - 4624} = 2.121$$

$$\sigma_Y = \sqrt{\frac{1}{n}\Sigma y^2 - y^2} = \sqrt{\frac{1}{8}(38132) - 69^2} = \sqrt{4766.5 - 4761} = 2.345$$

$$Cov(X,Y) = \frac{1}{n}\Sigma XY - \overline{X}\,\overline{Y} = \frac{1}{8}(37650) - 68\times 69$$

$$= 4695 - 4692 = 3$$

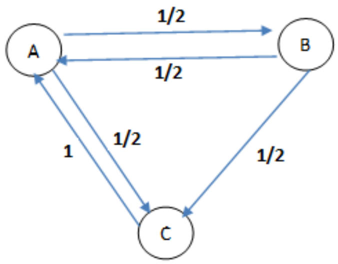
The correlation coefficient of X and Y is given by

	$r(X,Y)=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}=\dfrac{3}{(2.121)(2.345)}$ $=\dfrac{3}{4.973}=0.6032$ **(ii) The lifetime of a certain brand of an electric bulb may be considered as a random variable with mean 1200 hours and standard deviation 250 hours. Find the probability, using central limit theorem, that the average lifetime of 60 bulbs exceeds 1250 hours.** **Solution:** Let X_i (i=1,2,...60) denote the life time of the bulbs. Here μ=1200, $\sigma^2=250^2$ Let $\bar{X}$ denote the average life time of 60 bulbs. By Central limit theorem, $\bar{X}$ follows $N\left(\mu, \dfrac{\sigma^2}{n}\right)$. Let $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $Z=\dfrac{\bar{X}-1200}{\frac{250}{\sqrt{60}}}$ $P(\bar{X}>1250)=P(Z>1.55)=0.0606$
UNIT – III RANDOM PROCESSES	
PART-A	
1.	**Define (a) Continuous-time random process (b) Discrete state random process**. **(May 2017)** **Solution:** Consider a random process {X(t), t $\in$ T}, where T is the index set or parameter set. The values assumed by X(t) are called the states, and the set of all possible values of the states forms the state space E of the random process. (a) If the state space E and index set T are both continuous, then the random process is called continuous-time random process. (b) If the state space E is discrete and the index set T is continuous, then the random process is called discrete state random process
2.	**Define Wide sense stationary process** **(April/ May 2021)** **Solution:** A random process {X(t)} is called wide-sense stationary if the following conditions hold: (i) $E[X(t)]=a \ \ constant$ (ii) $R_{XX}(t_1,t_2)=E[X(t_1)X(t_2)]=R_{XX}(t_1-t_2)$ = function of time difference.
3.	**What is a random process? When do you say a random process is a random variable?** **Solution:** A random process is an infinite indexed collection of random variables X(t) : t ε T , defined over a common probability space. The index parameter t is typically time, but can also be a spatial dimension. Random processes are used to model random experiments that evolve in time such as Daily price of a stock.

	If the time(parameter) 't' is fixed, random process is called a random variable.
4.	**If the initial state probability distribution of a Markov chain is** $P^{(0)} = \begin{pmatrix} \frac{5}{6} & \frac{1}{6} \end{pmatrix}$ **and the transition probability matrix of the chain is** $\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$**, find the probability distribution of the chain after 2 steps.** **Solution:** Probability distribution after 2 steps = $P^{(2)} = P^{(1)} P = \{P^{(0)} P\} P$ Now $P^{(1)} = P^{(0)}P = \begin{pmatrix} \frac{5}{6} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1/12 & 11/12 \end{pmatrix}$ $\therefore P^{(2)} = P^{(1)}P = \begin{pmatrix} \frac{1}{12} & \frac{11}{12} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 11/24 & 13/24 \end{pmatrix}$
5.	**If the transition probability matrix of a Markov chain is** $\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ **find the steady-state distribution of the chain.** **(April/ May 2021)** **Solution:** Let $\pi = (\pi_1, \pi_2)$ be the limiting form of the state probability distribution on stationary state distribution of the Markov chain. By the property of π, $\pi P = \pi$ i.e., $(\pi_1, \pi_2) \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_1, \pi_2)$ $\frac{1}{2}\pi_2 = \pi_1$ ----------- (1) $\pi_1 + \frac{1}{2}\pi_2 = \pi_2$ ------ (2) We know that $\pi_1 + \pi_2 = 1$, since π is a probability distribution. Using (1) , $\frac{1}{2}\pi_2 + \pi_2 = 1 \Rightarrow \frac{3\pi_2}{2} = 1 \Rightarrow \pi_2 = \frac{2}{3}$ $\pi_1 = 1 - \pi_2 = 1 - \frac{2}{3} = \frac{1}{3} \therefore \pi_1 = \frac{1}{3}$ & $\pi_2 = \frac{2}{3}$
6.	**Examine whether the Poisson process {X(t)} is stationary or not.** **Solution:** A random process to be stationary in any sense, its mean must be a constant. We know that the mean of a Poisson process with rate λ is given by E{X(t)} = λt which depends on the time t. Thus the Poisson process is not a stationary process.
7.	**When is a Markov chain, called homogeneous?** **Solution:**

	If the one-step transition probability is independent of n, i.e., $p_{ij}(n, n+1) = p_{ij}(m, m+1)$, then the Markov chain is said to have stationary transition probabilities and the process is called as homogeneous Markov chain.
8.	**Is a Poisson process a continuous time Markov chain? Justify your answer** **Solution:** We know that Poisson process has the Markovian property. Therefore, it is a Markov chain as the states of Poisson process are discrete. Also, the time 't' in a Poisson process is continuous. Therefore, the Poisson process is a continuous time Markov chain.
9.	**Define transition probability matrix**. **Solution:** The transition probability matrix (TPM) of the process $\{X_n, n \ge 0\}$ is defined by $P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$ where the transition probabilities (elements of P) satisfy $p_{ij} \ge 0$ & $\sum_{j=1}^{\infty} p_{ij} = 1,\ i = 1, 2, 3, \cdots$
10.	**Define Markov process**. **Solution:** A random process or Stochastic process X(t) is said to be a Markov process if given the value of X(t), the value of X(v) for v > t does not depend on values of X(u) for u < t. In other words, the future behaviour of the process depends only on the present value and not on the past value.
11.	**Prove that first order stationary random process has a constant mean.** **Solution:** Let X(t) be a first-order stationary process. Then the first-order probability density function of X(t) satisfies $f_X(x_1 ; t_1) = f_X(x_1 ; t_1 + \varepsilon)$ …..(A) for all t_1 and ε. Consider any two time instants t_1 and t_2, and define the random variable $X_1 = X(t_1)$ and $X_2 = X(t_2)$. By definition, the mean values of X_1 and X_2 are given by $E(X_1) = E[X(t_1)] = \int_{x_1=-\infty}^{\infty} x_1 f_X(x_1; t_1) dx_1 \quad \cdots (1)$ $E(X_2) = E[X(t_2)] = \int_{x_2=-\infty}^{\infty} x_2 f_X(x_2; t_2) dx_2 \quad \cdots (2)$ $Let\ t_2 = t_1 + \varepsilon \quad (2) \Rightarrow E[X(t_1+\varepsilon)] = \int_{x_2=-\infty}^{\infty} x_2 f_X(x_2; t_1+\varepsilon) dx_2$ Using (A), $E[X(t_1+\varepsilon)] = \int_{x_2=-\infty}^{\infty} x_2 f_X(x_2; t_1+\varepsilon) dx_2 = \int_{x_1=-\infty}^{\infty} x_1 f_X(x_1; t_1) dx_1 = E[X(t_1)]$ which shows first order stationary random process has a constant mean.
12.	**Define Ergodic Random process.** **Solution:** A random process {X(t)} is called ergodic if all its ensemble averages are equal to appropriate time averages.

13.	**Define Markov Chain. When do you say that a Markov chain is irreducible?** **Solution:** **A** Markov process is called Markov chain if the states $\{X_i\}$ are discrete, no matter whether 't' is discrete or continuous. The Markov chain is irreducible if all states communicate with each other at some time.
14.	**State Chapman-Kolmogorov Equation.** **(May 2017)** **Solution:** The Chapman-Kolmogorov equation provides a method to compute the n-step transition probabilities. The equation can be represented as $P_{ij}^{n+m} = \sum_{k=0}^{\infty} p_{ik}^{n} p_{kj}^{m} \ \forall n, m \geq 0$.
15.	**Define a k^{th} order stationary process. When will it become a strict sense stationary process?** **Solution:** **(Apr 2017)** The process is said to be k^{th} order stationary if all its finite dimensional distributions are invariant under translation of time for all $t_1, t_2, \ldots, t_n$ and for n=1,2,3……k only and not for all n>k, then the process is called the k^{th} order stationary process. It becomes a strict sense stationary process when $k \to \infty$.
16.	**When do you say the Markov chain is regular? When do you say that state 'i' is periodic and aperiodic?** **Solution:** A regular Markov chain is defined as a chain having a transition matrix P such that for some positive power of P, it has only non-zero positive probability values. Let A be the set of all positive integers n such that $p_{ii}^{(n)} > 0$ and 'd' be the Greatest Common Divisor of the set A. We say state 'i' is periodic if d>1 and aperiodic if d = 1.
17.	**What are the properties of Poisson process?** **Solution:** (a) Sum of two independent Poisson processes is a Poisson process. (b) The Poisson process is a Markov process.
18.	**Consider the Markov chain consisting of the three states 0, 1, 2 and transition probability matrix** $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$ **. Is it irreducible? Justify.** **Solution:** $P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.25 & 0.25 \\ 0 & 0.33 & 0.67 \end{bmatrix}$ and $P^2 = \begin{bmatrix} 0.5 & 0.375 & 0.125 \\ 0.375 & 0.395 & 0.23 \\ 0.165 & 0.304 & 0.531 \end{bmatrix}$ Here $P_{ij}^{(n)} > 0, \forall i, j$ when n=1 or 2. So, P is irreducible.
19.	**Consider the Markov chain consisting of the three states A, B, C and transition probability matrix**

	$P=\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix}$. **Draw the state transition diagram.** **Solution:** 
20.	**A radioactive source emits particles at a rate of 5 per min in accordance with Poisson process. Each particle emitted has a probability 0.6 of being recorded. Find the probability that 10 particles are recorded in 4 min period.** **Solution:** The number of particles N(t) emitted is Poisson with parameter λ= np = 5(0.6) = 3 $P(N(t)=m)=\frac{e^{-3t}(3t)^m}{m!} \Rightarrow P(N(4)=10)=\frac{e^{-3(4)}(3(4))^{10}}{10!}=0.1048.$
	PART-B
1.	**(i)Show that the random process $X(t)=A\cos(w_0t+\theta)$ is wide-sense stationary, if A and w_0 are constants and θ is uniformly distributed random variable in (0,2π)** **(April 2017)** **Solution:** Since θ is uniformly distributed in $(0,2\pi)$ p.d.f. is $f(\theta)=\frac{1}{2\pi}$, $0\le\theta\le 2\pi$ $E\left[X(t)\right]=\int_0^{2\pi}\frac{1}{2\pi}A\cos(\omega_0t+\theta)d\theta=\frac{A}{2\pi}\left[\sin(\omega_0t+\theta)\right]_0^{2\pi}=\frac{A}{2\pi}\left[\sin(\omega_0t+2\pi)-\sin(\omega_0t)\right]$ $=\frac{A}{2\pi}\left[\sin\omega_0t-\sin\omega_0t\right]=0$ $E\left[X(t)\right]$ = constant $R(t_1,t_2)=E\left[X(t_1)X(t_2)\right]=E\left[A\cos(\omega_0t_1+\theta)A\cos(\omega_0t_2+\theta)\right]$ $=E\left[A^2\cos(\omega_0t_1+\theta)\cos(\omega_0t_2+\theta)\right]=\frac{A^2}{2}E\left[\cos\left[\omega_0(t_1+t_2)+2\theta\right]+\cos\omega_0(t_1-t_2)\right]$ $=\frac{A^2}{2}\int_0^{2\pi}\frac{1}{2\pi}\left[\cos(\omega_0(t_1+t_2)+2\theta)+\cos(\omega_0(t_1-t_2))\right]d\theta$

$$=\frac{A^2}{4\pi}\left\{\left[\frac{\sin(\omega_0(t_1+t_2)+2\theta)}{2}\right]_0^{2\pi}+\cos(\omega_0(t_1-t_2))[\theta]_0^{2\pi}\right\}$$

$$=\frac{A^2}{4\pi}\left\{\frac{\sin(\omega_0(t_1+t_2)+4\pi)}{2}-\frac{\sin(\omega_0(t_1+t_2))}{2}+\cos(\omega_0(t_1-t_2))(2\pi)\right\}$$

$$=\frac{A^2}{4\pi}\left[\frac{\sin\omega_0(t_1+t_2)}{2}-\frac{\sin\omega_0(t_1+t_2)}{2}+2\pi\cos\omega_0(t_1-t_2)\right]$$

$$=\frac{A^2}{4\pi}.2\pi\cos\omega_0(t_1-t_2)=\frac{A^2}{2}\cos\omega_0(t_1-t_2) \text{ = a function of } t_1-t_2$$

$\therefore$ The process $X(t)$ is W.S.S.

(ii) If $X(t)=Ycost+Zsint$ for all t & where Y & Z are independent binary random variables. Each of which assumes the values – 1 & 2 with probabilities $\frac{2}{3}$ & $\frac{1}{3}$ respectively, prove that $\{X(t)\}$ is WSS. **(April/May 2021)**

Solution:

Given

$$\begin{array}{lccc} Y=y & : & -1 & 2 \\ P(Y=y) & : & \frac{2}{3} & \frac{1}{3} \end{array}$$

$$E(Y)=E(Z)=-1\times\frac{2}{3}+2\times\frac{1}{3}=0$$

$$E(Y^2)=E(Z^2)=(-1)^2\times\frac{2}{3}+(2)^2\times\frac{1}{3}$$

$$E(Y^2)=E(Z^2)=\frac{2}{3}+\frac{4}{3}=\frac{6}{3}=2$$

Since Y & Z are independent

$E(YZ)=E(Y)E(Z)=0$ ---- (1)

Hence $E[X(t)]=E[ycost+zsint]$

$$=E[y]cost+E[z]sint$$

$E[X(t)]=0=$ is a constant. $[\because E(y)=E(z)=0]$

$$R_{XX}(t_1,t_2)=E[X(t_1)X(t_2)]$$

$$=E[(ycost_1+zsint_1)(ycost_2+zsint_2)]$$

$$=E[y^2cost_1cost_2+yzcost_1sint_2+zysint_1cost_2+z^2sint_1sint_2]$$

$$=E(y^2)cost_1cost_2+E[yz]cost_1sint_2+E[zy]sint_1cost_2+E[z^2]sint_1sint_2$$

$$=E(y^2)cost_1cost_2+E(z^2)sint_1sint_2$$

	$=2\left[cost_1cost_2+sint_1sint_2\right]\quad\left[\because E\left(y^2\right)=E\left(z^2\right)=2\right]$ $=2\cos\left(t_1-t_2\right)=$ is a function of time difference. $\therefore\left\{X(t)\right\}$ is WSS.
2.	**(i) The process {X(t)} whose probability distribution under certain condition is given by** $P\left[X(t)=n\right]=\begin{cases}\dfrac{(at)^{n-1}}{(1+at)^{n+1}}, & n=1,2,3,\cdots\\ \dfrac{at}{1+at}, & n=0.\end{cases}$ **. Show that {X(t)} is not stationary.** **Solution:** Given $P\left[X(t)=n\right]=\begin{cases}\dfrac{(at)^{n-1}}{(1+at)^{n+1}}, & n=1,2,3,\cdots\\ \dfrac{at}{1+at}, & n=0.\end{cases}$ $E\left[X(t)\right]=\sum_{n=0}^{\infty}nP\left\{X(t)=n\right\}=0+1P\left\{X(t)=1\right\}+2P\left\{X(t)=2\right\}+3P\left\{X(t)=3\right\}+\cdots$ $=1\left(\dfrac{1}{(1+at)^2}\right)+2\left(\dfrac{at}{(1+at)^3}\right)+3\left(\dfrac{(at)^2}{(1+at)^4}\right)+\cdots$ $=\dfrac{1}{(1+at)^2}\left[1+2\dfrac{at}{1+at}+3\left(\dfrac{at}{1+at}\right)^2+\cdots\right]=\dfrac{1}{(1+at)^2}\left[1-\dfrac{at}{1+at}\right]^{-2}$ $=\dfrac{1}{(1+at)^2}\left[\dfrac{1}{1+at}\right]^{-2}=1=a$ constant $E\left[X^2(t)\right]=\sum_{n=0}^{\infty}n^2P\left\{X(t)=n\right\}=\sum_{n=0}^{\infty}\left\{n(n+1)-n\right\}P\left\{X(t)=n\right\}$ $=\sum_{n=0}^{\infty}n(n+1)P\left\{X(t)=n\right\}-nP\left\{X(t)=n\right\}$ $=\sum_{n=0}^{\infty}n(n+1)P\left\{X(t)=n\right\}-1\qquad\left\{\because\sum_{n=0}^{\infty}nP\left\{X(t)=n\right\}=1\right\}$ $=\left[0+1.2P\left\{X(t)=1\right\}+2.3P\left\{X(t)=2\right\}+3.4P\left\{X(t)=3\right\}+\cdots\right]-1$ $=\left\{2\left(\dfrac{1}{(1+at)^2}\right)+6\left(\dfrac{at}{(1+at)^3}\right)+12\left(\dfrac{(at)^2}{(1+at)^4}\right)+\cdots\right\}-1$ $=\dfrac{2}{(1+at)^2}\left(1-\dfrac{at}{1+at}\right)^{-3}-1$

	$=\dfrac{2}{(1+at)^2}\left(\dfrac{1}{1+at}\right)^{-3}-1$ $=2(1+at)-1=2at+1$, not a constant So, X(t) is not a stationary process. **(ii) Show that the process $X(t)=A\cos\lambda t+B\sin\lambda t$ is wide sense stationary, if E(AB) =0; E(A)=E(B) = 0, & $E(A^2) = E(B^2)$, where A &B are random variables.** **Solution:** Given $X(t)=A\cos\lambda t+B\sin\lambda t$, E(A) = E(B) = E(AB) = 0, $E(A^2) = E(B^2)$ = k(say) To prove X(t) is wide sense stationary, we have to show (i) E{X(t)} = constant (ii) R_{XX}(t,t+τ) = function of time difference = function of τ Now, $E\left[X(t)\right]=E\left[A\cos\lambda t+B\sin\lambda t\right]=\cos\lambda t E(A)+\sin\lambda t E(B)=0 \therefore E\left[X(t)\right]=\text{constant}$ $R_{XX}(t,t+\tau)=E\{X(t)X(t+\tau)\}$ $=E\{\left[A\cos\lambda t+B\sin\lambda t\right]\left[A\cos(\lambda t+\lambda\tau)+B\sin(\lambda t+\lambda\tau)\right]\}$ $=E\left[A^2\cos\lambda t\cos(\lambda t+\lambda\tau)+AB\cos\lambda t\sin(\lambda t+\lambda\tau)+AB\sin\lambda t\cos(\lambda t+\lambda\tau)+B^2\sin\lambda t\sin(\lambda t+\lambda\tau)\right]$ $=\cos\lambda t\cos(\lambda t+\lambda\tau)E(A^2)+\cos\lambda t\sin(\lambda t+\lambda\tau)E(AB)+\sin\lambda t\cos(\lambda t+\lambda\tau)E(AB)+\sin\lambda t\sin(\lambda t+\lambda\tau)E(B^2)$ $=\cos\lambda t\cos(\lambda t+\lambda\tau)k+\sin\lambda t\sin(\lambda t+\lambda\tau)k$ $\left[\because E(AB)=0 \;\&\; E(A^2)=E(B^2)=k(say)\right]$ $=k\left[\cos\lambda t\cos(\lambda t+\lambda\tau)+\sin\lambda t\sin(\lambda t+\lambda\tau)\right]$ $=k\cos(\lambda t+\lambda\tau-\lambda t)=k\cos\lambda\tau=a$ function of time difference. $\therefore$ X(t) is wide sense stationary.
3.	**(i) Three boys A, B and C are throwing a ball to each other. A always throws the ball to B and B always throws the ball to C, but C is just as likely to throw the ball to B as to A. Show that the process is Markovian. Find the transition matrix and classify the states. (April/May 2021)** **Solution:** Since states of X_n depend only on states of X_{n-1} .{ X_n} is a Markov chain.

TPM= $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$

A always throws to B B always throws to C

$\therefore P_{01} = 1, P_{00} = 0, P_{02} = 0,$

$\therefore P_{10} = 1, P_{11} = 0, P_{12} = 1$

C throws A or B with equal probability

$\therefore P_{20} = \frac{1}{2}, P_{21} = \frac{1}{2}, P_{22} = 0$

Now $P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}$; $P^3 = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$

Therefore $P_{11}^{(3)} > 0$, $P_{13}^{(2)} > 0$, $P_{21}^{(2)} > 0$, $P_{22}^{(2)} > 0$, $P_{33}^{(2)} > 0$ and all other $P_{ij}(1)>0$.

Therefore, chain is irreducible.

$P^4 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$; $P^5 = \begin{bmatrix} 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.125 & 0.375 & 0.5 \end{bmatrix}$ and so on.

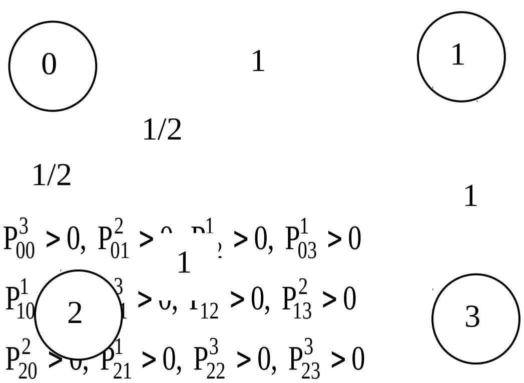
We note that $P_{ii}^{(2)}, P_{ii}^{(3)}, P_{ii}^{(5)}$ …. are greater than zero for i=2,3,.. and GCD of 2,3,5,6,….=1
Therefore, the state 1 is aperiodic. Since the chain is finite and irreducible, all its states are non-null Persistent. Þ The period of 2 and 3 is 1. The state with period 1 is aperiodic all states are ergodic

(ii)Let the Markov Chain consisting of the states 0, 1, 2, 3 have the transition probability matrix $P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$. **Determine which states are transient and which are recurrent by defining transient and recurrent states.**

Solution:

Transient state: A state 'a' is transient if $F_{aa} < 1$. Recurrent state: A state 'a' is recurrent if $F_{aa}=1$.

Here $F_{aa} = \sum_{n=0}^{\infty} f_{aa}^{(n)}$, where $f_{aa}^{(n)}$ = first time return probability of state 'a' after n steps.



Here $P^3_{00} > 0,\ P^2_{01} > 0,\ P^1_{02} > 0,\ P^1_{03} > 0$

$P^1_{10} > 0,\ P^3_{11} > 0,\ P^1_{12} > 0,\ P^2_{13} > 0$

$P^2_{20} > 0,\ P^1_{21} > 0,\ P^3_{22} > 0,\ P^3_{23} > 0$

$P^2_{30} > 0,\ P^1_{31} > 0,\ P^3_{32} > 0,\ P^3_{33} > 0$

Therefore, the Markov chain is irreducible. And also it is finite.
So, all the states are of same nature.
Consider the state '0'

$$f^1_{00} = 0;\ f^2_{00} = 0;\ f^3_{00} = \frac{1}{2} + \frac{1}{2} = 1;\ f^4_{00} = 0 \text{ and so on.}$$

Therefore, the state '0' is recurrent.
Since, the chain is irreducible, all the states are recurrent.

4. **The transition probability matrix of a Markov chain { X(t)}, n = 1,2,3,….} having three states 1, 2 and 3 is** $P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ **and the initial distribution is** $P^{(0)} = (0.7 \quad 0.2 \quad 0.1)$ **.Find**

$(i)\ P[X_2 = 3] \quad (ii)\ P[X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2]$ **.**

Solution:

We have $P^2 = P.P = \begin{pmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{pmatrix}$

(i) $P(X_2 = 3) = \sum_{i=1}^{3} P(X_2 = 3 / X_0 = i)\, P(X_0 = i)$

$= P(X_2 = 3 / X_0 = 1) P(X_0 = 1) + P(X_2 = 3 / X_0 = 2) P(X_0 = 2) + P(X_2 = 3 / X_0 = 3) P(X_0 = 3)$

$= P^2_{13} P(X_0 = 1) + P^2_{23} P(X_0 = 2) + P^2_{33} P(X_0 = 3) = 0.26 \times 0.7 + 0.34 \times 0.2 + 0.29 \times 0.1 = 0.279$

(ii) $P(X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2)$

$= P[X_0 = 2, X_1 = 3, X_2 = 3]\ P[X_3 = 2 / X_0 = 2, X_1 = 3, X_2 = 3]$

	$=P\left[X_0=2, X_1=3, X_2=3\right] P\left[X_3=2/X_2=3\right]$ $=P\left[X_0=2, X_1=3\right] P\left[X_2=3/X_0=2, X_1=3\right] P\left[X_3=2/X_2=3\right]$ $=P\left[X_0=2, X_1=3\right] P\left[X_2=3/X_1=3\right] P\left[X_3=2/X_2=3\right]$ $=P\left[X_0=2\right] P\left[X_1=3/X_0=2\right] P\left[X_2=3/X_1=3\right] P\left[X_3=2/X_2=3\right]$ $=(0.2)(0.2)(0.3)(0.4)=0.0048$
5.	**A gambler has Rs. 2. He bets Rs. 1 at a time and wins Rs. 1 with probability 0.5. He stops playing if he loses Rs. 2 or wins Rs. 4.** **(a) What is the tpm of the related Markov chain?** **(b) What is the probability that he has lost his money at the end of 5 plays?** **Solution:** Let X_n denote the amount with the player at the end of the n^{th} round of the play. The possible values of X_n = State space = {0, 1, 2, 3, 4, 5, 6} Initial probability distribution= $P^{(0)}=\begin{pmatrix}0 & 0 & 1 & 0 & 0 & 0 & 0\end{pmatrix}$ (a) The transition probability matrix is = $\begin{array}{c} \\ 0\\1\\2\\3\\4\\5\\6\end{array}\begin{array}{c}\begin{array}{ccccccc}0&1&2&3&4&5&6\end{array}\\ \begin{pmatrix}1&0&0&0&0&0&0\\ \frac{1}{2}&0&\frac{1}{2}&0&0&0&0\\ 0&\frac{1}{2}&0&\frac{1}{2}&0&0&0\\ 0&0&\frac{1}{2}&0&\frac{1}{2}&0&0\\ 0&0&0&\frac{1}{2}&0&\frac{1}{2}&0\\ 0&0&0&0&\frac{1}{2}&0&\frac{1}{2}\\ 0&0&0&0&0&0&1\end{pmatrix}\end{array}$ (b) Probability that he has lost his money at the end of 5 plays = $P\left[X_5=0\right]$ To find this we need $P^{(5)}$ $P^{(1)}=P^{(0)}P=\begin{pmatrix}0&0&1&0&0&0&0\end{pmatrix}\begin{bmatrix}1&0&0&0&0&0&0\\ \frac{1}{2}&0&\frac{1}{2}&0&0&0&0\\ 0&\frac{1}{2}&0&\frac{1}{2}&0&0&0\\ 0&0&\frac{1}{2}&0&\frac{1}{2}&0&0\\ 0&0&0&\frac{1}{2}&0&\frac{1}{2}&0\\ 0&0&0&0&\frac{1}{2}&0&\frac{1}{2}\\ 0&0&0&0&0&0&1\end{bmatrix}=\begin{pmatrix}0&\frac{1}{2}&0&\frac{1}{2}&0&0&0\end{pmatrix}$

$$P^{(2)}=P^{(1)}P=\begin{pmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0\end{pmatrix}\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}=\begin{pmatrix}\frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0\end{pmatrix}$$

$$P^{(3)}=P^{(2)}P=\begin{pmatrix}\frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0\end{pmatrix}\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}=\begin{pmatrix}\frac{1}{4} & \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{1}{8} & 0\end{pmatrix}$$

$$P^{(4)}=P^{(3)}P=\begin{pmatrix}\frac{1}{4} & \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{1}{8} & 0\end{pmatrix}\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}=\begin{pmatrix}\frac{3}{8} & 0 & \frac{5}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16}\end{pmatrix}$$

$$P^{(5)}=P^{(4)}P=\begin{pmatrix}\frac{3}{8} & 0 & \frac{5}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16}\end{pmatrix}\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}=\begin{pmatrix}\frac{3}{8} & \frac{5}{32} & 0 & \frac{9}{32} & 0 & \frac{1}{8} & \frac{1}{16}\end{pmatrix}$$

$$\therefore \boxed{P(X_5=0)=3/8}$$

6.	**A man either drives a car or catches a train to go to office each day. He never goes 2 days in a row by train but if he drives one day, then the next day he is just as likely to drive again as he is to travel by train. Now suppose that on the first day of the week, the man tossed a fair die and drove to work if and only if a 6 appeared. Find (a) the probability that he takes a train on the third day and (b) the probability that he drives to work in the long run.**

Solution:

State Space = (train, car)

The TPM of the chain is

$$P = \begin{array}{c} \\ T \\ C \end{array} \begin{array}{c} \begin{array}{cc} T & C \end{array} \\ \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

P (traveling by car) = P (getting 6 in the toss of the die)= $\frac{1}{6}$

& P (traveling by train) = $\frac{5}{6}$

$$P^{(1)} = \left(\frac{5}{6}, \frac{1}{6}\right)$$

$$P^{(2)} = P^{(1)}P = \left(\frac{5}{6}, \frac{1}{6}\right)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{1}{12}, \frac{11}{12}\right)$$

$$P^{(3)} = P^{(2)}P = \left(\frac{1}{12}, \frac{11}{12}\right)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{11}{24}, \frac{13}{24}\right)$$

P (the man travels by train on the third day) = $\frac{11}{24}$

Let $\pi = (\pi_1, \pi_2)$ be the limiting form of the state probability distribution or stationary state distribution of the Markov chain.

By the property of π , $\pi P = \pi$

$$(\pi_1 \ \pi_2)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_1 \ \pi_2)$$

$$\frac{1}{2}\pi_2 = \pi_1$$

$$\pi_1 + \frac{1}{2}\pi_2 = \pi_2$$

& $\pi_1 + \pi_2 = 1$

Solving $\pi_1 = \frac{1}{3}$ & $\pi_2 = \frac{2}{3}$

P {The man travels by car in the long run}= $\frac{2}{3}$.

7.	**A salesman territory consists of three cities A, B and C. He never sells in the same city on successive days. If he sells in city A, then the next day he sells in city B. However, if he sells in either city B or city C, the next day he is twice as likely to sell in city A as in the other city. In the long run how often does he sell in each of the cities?** **Solution:** States: A, B and C The transition probability matrix is given by $P=\begin{bmatrix}0&1&0\\2/3&0&1/3\\2/3&1/3&0\end{bmatrix}$ The long run probability is given by $\pi=\begin{bmatrix}a&b&c\end{bmatrix}$, where $\pi P=\pi$. Now $\pi P=\pi \Rightarrow \begin{bmatrix}a&b&c\end{bmatrix}\begin{bmatrix}0&1&0\\2/3&0&1/3\\2/3&1/3&0\end{bmatrix}=\begin{bmatrix}a&b&c\end{bmatrix}$ $\therefore\ 0.a+\frac{2}{3}b+\frac{2}{3}c=a \quad\Rightarrow\quad -a+\frac{2}{3}b+\frac{2}{3}c=0\cdots(1)$ $1.a+0.b+\frac{1}{3}c=b \quad\Rightarrow\quad a-b+\frac{1}{3}c=0\cdots(2)$ $0.a+\frac{1}{3}b+0.c=c \quad\Rightarrow\quad \frac{1}{3}b-c=0\cdots(3)$ Also, we know that $a+b+c=1\ \cdots(4)$ From (3), $c = b/3$ From (2), $a - b + b/9 = 0 \Rightarrow a = 8b/9$ From (4), $8b/9 + b + b/3 = 1 \Rightarrow 20b/9 = 1 \Rightarrow \boxed{b = 9/20}$ $\therefore \boxed{c = 3/20 \text{ and } a = 8/20}$ $\therefore$Long run probability = $\begin{bmatrix}8/20&9/20&3/20\end{bmatrix}$
8.	**(i)Prove that (a) difference of two independent Poisson processes is not a Poisson process and (b) Poisson process is a Markov process.** **Soltuion:** (a) Let $X(t)=X_1(t)-X_2(t)$ where $X_1(t)$ and $X_2(t)$ are poisson processes with λ_1 and λ_2 as the parameters $E[X(t)]=E[X_1(t)]-E[X_2(t)]=(\lambda_1-\lambda_2)t$ $E[X^2(t)]=E\{[X_1(t)-X_2(t)]^2\}=E[X_1^2(t)]+E[X_2^2(t)]-2E[X_1(t)X_2(t)]$ $=E[X_1^2(t)]+E[X_2^2(t)]-2E[X_1(t)]E[X_2(t)]$ $=(\lambda_1^2t^2+\lambda_1t)+(\lambda_2^2t^2+\lambda_2t)-2(\lambda_1t)(\lambda_2t)=(\lambda_1+\lambda_2)t+(\lambda_1^2+\lambda_2^2)t^2-2\lambda_1\lambda_2t^2$

$=(\lambda_1+\lambda_2)t+(\lambda_1-\lambda_2)^2t^2 \neq (\lambda_1-\lambda_2)t+(\lambda_1-\lambda_2)^2t^2$

$\therefore X_1(t)-X_2(t)$ is not a Poisson process.

(b) Consider $P\left[X(t_3)=n_3 / X(t_2)=n_2, X(t_1)=n_1)\right]=\dfrac{P\left[X(t_1)=n_1, X(t_2)=n_2, X(t_3)=n_3\right]}{P\left[X(t_1)=n_1, X(t_2)=n_2\right]}$

$$=\frac{\dfrac{e^{-\lambda t_3}\lambda^{n_3}t_1^{n_1}(t_2-t_1)^{n_2-n_1}(t_3-t_2)^{n_3-n_2}}{n_1!(n_2-n_1)!(n_3-n_2)!}}{\dfrac{e^{-\lambda t_2}\lambda^{n_2}t_1^{n_1}(t_2-t_1)^{n_2-n_1}}{n_1!(n_2-n_1)!}}=\frac{e^{-\lambda(t_3-t_2)}\lambda^{n_3-n_2}(t_3-t_2)^{n_3-n_2}}{(n_3-n_2)!}=P\left[X(t_3)=n_3 / X(t_2)=n_2\right]$$

$\therefore P\left[X(t_3=n_3 / X(t_2)=n_2, X(t_1)=n_1)\right]=P\left[X(t_3)=n_3 / X(t_2)=n_2\right]$

This means that the conditional probability distribution of $X(t_3)$ given all the past values $X(t_1)=n_1, X(t_2)=n_2$ depends only on the most recent values $X(t_2)=n_2$.

i.e., The Poisson process possesses Markov property.

(ii) Suppose that customers arrive at a bank according to Poisson process with mean rate of 3 per minute. Find the probability that during a time interval of two minutes a) exactly 4 customers arrive b) greater than 4 customers arrive c) fewer than 4 customers arrive.

Solution:

Mean of the Poisson process = λt.

Mean arrival rate = mean number of arrivals per minute (unit time) = λ

Given λ = 3. $P\{X(t)=k\}=\dfrac{e^{-\lambda t}(\lambda t)^k}{\underline{|k}}$

(a) Probability that exactly 4 customers arrive $P\{X(2)=4\}=\dfrac{e^{-6}6^4}{\underline{|4}}=0.133$

(b) Probability that greater than 4 customers arrive

$P\{X(2)>4\}=1-\left\{P[X(2)=0]+P[X(2)=1]+P[X(2)=2]+P[X(2)=3]+P[X(2)=4]\right\}$

$$=1-\left\{\frac{e^{-6}6^0}{\underline{|0}}+\frac{e^{-6}6^1}{\underline{|1}}+\frac{e^{-6}6^2}{\underline{|2}}+\frac{e^{-6}6^3}{\underline{|3}}+\frac{e^{-6}6^4}{\underline{|4}}\right\}=0.715$$

(c) Probability that fewer than 4 customers arrive

$P\{X(2)<4\}=P[X(2)=0]+P[X(2)=1]+P[X(2)=2]+P[X(2)=3]$

$$=\frac{e^{-6}6^0}{\underline{|0}}+\frac{e^{-6}6^1}{\underline{|1}}+\frac{e^{-6}6^2}{\underline{|2}}+\frac{e^{-6}6^3}{\underline{|3}}=0.151$$

9. On the average a submarine on patrol sights 6 enemy ships per hour. Assuming that the number of ships sighted in a given length of time is a Poisson variate. Find the probability of sighting 6 ships in the next half-an-hour, 4 ships in the next 2 hours and at least 1 ship in the next 15 minutes.

(April 2017)

Solution:

Mean arrival rate = mean number of arrivals per minute (unit time) = λ=6/hr

	Given $\lambda = 6$. $P\{X(t) = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{\lfloor k}$ (i) $P\left\{X\left(\frac{1}{2}\right) = 6\right\} = \frac{e^{-3}3^6}{\lfloor 6} = 0.0504$ (ii) $P\{X(2) = 4\} = \frac{e^{-12}(12)^4}{\lfloor 4} = 0.053$ (iii) $P\left\{X\left(\frac{1}{4}\right) \geq 1\right\} = 1 - P\left\{X\left(\frac{1}{4}\right) = 0\right\} 1 - \left\{\frac{e^{-\frac{3}{2}}\left(\frac{3}{2}\right)^0}{\lfloor 0}\right\} = 1 - e^{-\frac{3}{2}} = 0.776.$
10.	**A radioactive source emits particles at a rate of 5 per min in accordance with Poisson process. Each particle emitted has a probability 0.6 of being recorded. Find the probability that 10 particles are recorded in 4 min period.** **Solution:** The number of particles N(t) emitted is Poisson with parameter λ= np = 5(0.6) = 3 $P(N(t) = m) = \frac{e^{-3t}(3t)^m}{m!} \Rightarrow$ $P(N(4) = 10) = \frac{e^{-3(4)}(3(4))^{10}}{10!} = 0.1048.$
	UNIT – IV TESTING OF HYPOTHESIS
	PART-A
1.	**Define Population, Sample and Sample Size.** **Solution:** The group of individuals under study is called population. The population may be finite or infinite. A finite subset of statistical individuals in a population is called Sample. The number of individuals in a sample is called Sample Size (n).
2.	**Define Parameter and Statistic.** **Solution:** The Statistical constants in population namely mean μ and variance σ^2 which are usually referred to as parameters. Statistical measures computed from sample observations alone, i.e. mean $\bar{x}$ and variance s^2 which are usually referred to as statistic.
3.	**Define Sampling distribution.** **Solution:** The probability distribution of a sample statistic is called the sampling distribution
4.	**What is Standard Error?** **(April / May 2017)** **Solution:** The standard deviation of the sampling distribution of a statistic is known as its Standard error.

5.	**Find the standard error of sample mean from the following data** $n = 14$, $\mu = 18.5$, $\bar{x} = 17.85$, $S = 1.955$ **Solution:** $\text{Standard error} = \frac{S}{\sqrt{n-1}} = \frac{1.955}{\sqrt{14-1}} = 0.54$
6.	**Define Level of Significance.** **Solution:** The probability that the value of the statistic lies in the critical region is called the level of significance. It is denoted by α .
7.	**Define critical region and acceptance region?** **(Nov/Dec 2019)** **Solution:** A region corresponding to a statistic (t), in the sample space (s) which amounts to rejection of null hypothesis is called as critical region or region of rejection. The region complementary to the critical region is called acceptance region.
8.	**Define one - tailed and two - tailed test.** **Solution:** A test of any statistical hypothesis where the alternative hypothesis is one tailed (right or left tailed) is called a one tailed test. i.e. $H_o : \mu = \mu_o$ Vs $H_1 : \mu > \mu_o$ (right tailed) , (or) $H_1 : \mu < \mu_o$ (Left tailed) A test of statistical hypothesis whose alternative hypothesis is two tailed, such as $H_o : \mu = \mu_o$ Vs $H_1 : \mu \neq \mu_o$ is known as two tailed test.
9.	**Define Type-I and Type-II errors.** **(April / May 2018, 2019), (Nov/Dec2019)** **Solution:** Type I error: Reject Null hypothesis when it is true. The type I error is denoted by α, example of type I error is Reject a lot when it is good. Type II error: Accept Null hypothesis when it is false. The type II error is denoted by β, example of type II error is Accept a lot when it is bad.
10.	**Write 95% confidence interval of the population mean.** **Solution:** $\bar{x} - t_{0.05} \frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{0.05} \frac{S}{\sqrt{n-1}}$
11.	**Write the 95% confidence interval of population proportion.** **Solution:** 95% Confidence interval for P is $p - 1.96\sqrt{\frac{pq}{n}} \leq P \leq p + 1.96\sqrt{\frac{pq}{n}}$
12.	**List out the applications of t –distribution.** **Solution:** ❖ To test the significant difference between the means of two independent samples. ❖ To test the significant difference between the means of two dependent samples or paired observation. ❖ To test the significant difference between population mean and mean of a random sample. ❖ To test the significance of an observed correlation coefficient.
13.	**Mention the Properties of t – distribution.** **Solution:** ❖ The variable t distribution ranges from $-\infty$ to ∞. ❖ The t – distribution is symmetrical and has a mean zero.

	❖ The variance of the t – distribution is greater than one, but approaches one as the number of degrees of freedom and therefore the sample size become large.
14.	**Define student's t-test for single mean.** **(April/May 2021)** **Solution:** The t – distribution is used when sample size is 30 or less and the population standard deviation is unknown. The t – statistic is defined as $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ where $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n - 1}$. The t – distribution has been derived mathematically under the assumption of a normally distributed population.
15.	**What are the assumptions on which F-test is based?** **Solution:** ❖ Normality: The values in each group should be normally distributed. ❖ Independence of error: The variations of each value around its own group mean. i.e. error should be independent of each value. ❖ Homogeneity : The variances within each group should be equal for all groups.
16.	**What are the uses F – test?** **(April/May 2021)** **Solution:** ❖ To test equality of two population variances. ❖ To test the sample observation coming from normal population. ❖ To determine whether or not the two independent estimates of the population variances are homogeneous in nature.
17.	**State any two properties of χ^2 distribution.** **Solution:** (i)Chi – square curve is always positively skewed (ii)Chi – square values increase with the increase in degrees of freedom
18.	**Explain the various uses of Chi-square test.** **(April/May 2021)** **Solution:** ❖ Test of goodness of fit. ❖ Test of independence of attributes. ❖ Test of Homogeneity for a specified value of standard deviation.
19.	**State the assumptions of Chi-square test.** **Solution:** ❖ The sample observations should be independent. ❖ Constraints on the cell frequencies, if any must be linear. ❖ The total frequency should be atleast 50. ❖ No theoretical frequency is less than 5, then for the application of chi square test, it is pooled with the succeeding or preceding so that combined frequency is less than 5.

20. **Write down the values of χ^2 for a 2×2 contingency table with cell frequencies a, b, c, d.** **(April/May 2017)**

Solution:

	A	B	Total
A	a	b	a + b
B	c	d	c + d
Total	a + c	b + d	N=a+b+c+d

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

PART-B

1. **(i) The mean of two large sampling 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches?** **(April/May 2021)**

Solution:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Level of Significance: 5%

Test Statistic:

$$|Z| = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.16$$

$\therefore Z = 5.16$

Table value : $Z_\alpha = 1.96$

Conclusion : The calculated value is greater than the table value, hence we reject the null hypothesis.

(ii) In a sample of 900 members has the mean 3.4 cms and S.D 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms? If the population is normal and its mean is unknown, find the 95% confidence limits of true mean. **(April/May 2021)**

Solution:

Given $n = 900$, $\mu = 3.25$, $\bar{x} = 3.4cm$, $\sigma = 2.61$, $s = 2.61$

H_0: There is no significant difference between sample mean and population mean. (i.e) $\mu = 3.25$

H_1: There is a significant difference between sample mean and population mean. (i.e) $\mu \neq 3.25$

Test Statistic:

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = 1.724 \Rightarrow |z| = 1.724$$

Critical value: The critical value of z for two tailed test at 5% level of significance is 1.96

Conclusion:

	i.e., $z = 1.724 < 1.96 \Rightarrow$ calculated value < tabulated value Therefore We accept the null hypothesis H_0. i.e., The sample has been drawn from the population with mean $\mu = 3.25$ **To find confidence limit:** 95% confidence limits are $\bar{x} \mp 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \mp 1.96\left(\frac{2.61}{\sqrt{900}}\right) = 3.4 \mp 0.1705 = (3.57, 3.2295)$
2.	**Before an increase in excise duty on tea, 800 people out of a sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty.** **Solution:** **Hypothesis:** $H_0 : P_1 = P_2 \quad H_1 : P_1 > P_2$ **Level of Significance :** $\alpha = 0.05$ **Test Statistic :** $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ The Sample proportion, $p_1 = \frac{800}{1000} = 0.80, \quad p_2 = \frac{800}{1200} = 0.67, \quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.7273 \ \& \ Q = 1 - P = 0.2727$ $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 6.9905$ **Table value :** $Z_\alpha = 1.645$ **Conclusion :** The calculated value is greater than the table value, hence we reject the null hypothesis.
3.	**A random sample of 10 boys had the following I.Q's: 70,120,110,101,88,83,95,98,107,100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.** **Solution:** Hypothesis: $H_0 : \mu = 100$ $H_1 : \mu \neq 100$ Level of Significance : $\alpha = 0.05$ Test Statistic : $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$, *where* $s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ Analysis:

X	70	120	110	101	88	83	95	98	107	100	$972 = \sum X$
X^2	4900	14400	12100	10201	7744	6889	9025	9604	11449	10000	$96312 = \sum X^2$

	$s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{96312}{10} - \left(\frac{972}{10}\right)^2$ $\Rightarrow 9631.2 - 9447.84 = 183.36 \Rightarrow S = 13.54$ $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}, \text{ where } s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ $t = \frac{97.2 - 100}{13.54/\sqrt{10-1}} = \frac{2.8}{4.5133} = 0.6204$ Table value : $t_{\alpha,n-1} = t_{5\%,10-1} = t_{0.05,9} = 2.262$ Conclusion: The table value is greater than the calculated value; hence we accept the null hypothesis and conclude that the data are consistent with the assumption of mean I.Q of 100 in the population. Reasonable range in which most of the mean I.Q. values of samples of 10 boys lie: 95 % Confidence interval for the sample mean $\bar{x}$ $\lvert t\rvert \le t_{0.05,9} = 2.262$ $\Rightarrow \left\lvert \frac{\bar{x} - \mu}{S/\sqrt{n-1}} \right\rvert \le 2.262 \Rightarrow \left\lvert \frac{\bar{x} - 100}{13.54/\sqrt{10-1}} \right\rvert \le 2.262$ $\Rightarrow \left\lvert \frac{\bar{x} - 100}{4.5133} \right\rvert \le 2.262$ $\Rightarrow -2.262 \le \frac{\bar{x} - 100}{4.5133} \le 2.262$ $\Rightarrow -10.2091 \le \bar{x} - 100 \le 10.2091$ $\Rightarrow 89.7909 \le \bar{x} \le 110.2091$
4.	**(i)** The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	**24**	**27**	**26**	**21**	**25**	**-**
Sample B	**27**	**30**	**28**	**31**	**22**	**36**

Can it be said that both the samples have come from same normal population? (Apr/May 2021)

(i) F-test : (Equality of variance)

Let $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \ne \sigma_2^2$

x_1	$x_1 - \overline{x_1}$	$\left(x_1 - \overline{x_1}\right)^2$	x_2	$x_2 - \overline{x_2}$	$\left(x_2 - \overline{x_2}\right)^2$
24	-0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	-3.6	12.96	31	2	4
25	0.4	0.16	22	-7	49

			36	7	49
123		21.2	174		108

$\overline{x_1} = \frac{\sum x_1}{n_1} = \frac{123}{5} = 24.6, \overline{x_2} = \frac{\sum x_2}{n_2} = \frac{174}{6} = 29$

$s_1^2 = \frac{\sum\left(x - \overline{x}\right)^2}{n_1 - 1} = \frac{21.2}{4} = 5.3$, $s_2^2 = \frac{\sum\left(x_2 - \overline{x_2}\right)^2}{n_2 - 1} = \frac{108}{5} = 21.6$

$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{5(5.3)}{4} = 6.625 \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{6(21.6)}{5} = 25.92$

The test statistic is $F = \frac{S_2^2}{S_1^2}$ $(\sin ce\ S_1^2 < S_2^2)$

$$= \frac{25.92}{6.625} = 3.912$$

From the table, $F_{0.05}\left(n_2 - 1, n_1 - 1\right) = F_{0.05}(5,4) = 6.26$

Since $F < F_{0.05}$ $\therefore H_0$ is accepted

(ii) t-test:(Equality of means)

Null Hypothesis H$_0$: $\mu_1 = \mu_2$

Alternative Hypothesis H$_1$: $\mu_1 \neq \mu_2$

Under H_0 , the test statistic is $t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$,

$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{5(5.3) + 6(21.6)}{5 + 6 - 2}} = 4.164$

$t = \frac{24.6 - 29}{4.16\sqrt{\frac{1}{5} + \frac{1}{6}}} = -1.746 \quad \Rightarrow \ |t| = 1.746$

From the table ,with degrees of freedom $n_1 + n_2$-2=9, $t_{0.05}$=2.262

$\sin ce\ |t| < t_{0.05}$ $\therefore H_0$ is accepted ie. $\mu_1 = \mu_2$

Conclusion:

$\therefore$ The two samples could have been drawn from the same normal population.

(ii) The following data represent the biological values of protein from cow's milk and buffalo's milk. Examine whether the average values of protein in the two samples significantly differ at 5% level.

Cow's milk	**1.82**	**2.02**	**1.88**	**1.61**	**1.81**	**1.54**
Buffalo's milk	**2.00**	**1.83**	**1.86**	**2.03**	**2.19**	**1.88**

Solution:

n=6

Sample I				Sample II	
X_1	X_1^2			X_2	X_2^2
1.82	3.3124			2.00	4
2.02	4.0804			1.83	3.3489
1.88	3.5344			1.86	3.4596
1.61	2.5921			2.03	4.1209
1.81	3.2761			2.19	4.7961
1.54	2.3716			1.88	3.5344
10.68	19.167				23.2599

$$\overline{x_1} = \frac{1}{6}\times 10.68 = 1.78,\ s_1^2 = \frac{1}{6}\times 19.167 - (1.78)^2 = 0.0261$$

$$\overline{x_2} = \frac{1}{6}\times 11.79 = 1.965;\quad s_2^2 = \frac{1}{6}\times 23.2599 - (1.965)^2 = 0.0154$$

$H_0 : \overline{x_1} = \overline{x_2}$ $\quad and \quad$ $H_1 : \overline{x_1} \neq \overline{x_2}$ (two tailed test)

LOS is 5%

As the two samples are independent , the test statistic is given by t = $\dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^2 + s_2^2}{n-1}}}$

$$|t| = \frac{1.78 - 1.965}{\sqrt{\dfrac{0.0261 + 0.0154}{5}}} = -2.03 \quad and \ \ v = 10$$

From table $t_{5\%}(v = 10) = 2.23$

H_0 is accepted (i.e,) the difference between the mean protein values of the two varieties of milk is not significant at 5% level.

5. **The table below gives the number of aircraft accidents that occurred during the various days of the week. Test whether the accidents are uniformly distributed over the week.**

Days	**Mon**	**Tue**	**Wed**	**Thurs**	**Fri**	**Sat**
No. of accidents	**14**	**18**	**12**	**11**	**15**	**14**

Solution:

We want to test whether the accidents are uniformly distributed. So we apply χ^2 -test.

Null Hypothesis H_0: The accidents are uniformly distributed over the 6 days. (Monday to Saturday)

Alternative Hypothesis H_1: The accidents are not uniformly distributed.

Under H_o, the expected frequencies for each day = $\frac{84}{6}$ =14

The test statistic is $\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
14	14	0	0	0
18	14	4	16	1.143
12	14	-2	4	0.286
11	14	-3	9	0.643
15	14	1	1	0.071
14	14	0	0	0.000
84	84			χ^2 =2.143

Number of degrees of freedom v = n-1 = 6-1 = 5

For v=5 degrees of freedom , from the table of χ^2 at 5% level is $\chi^2_{0.05}$ = 11.07

$\therefore \chi^2 < \chi^2_{0.05}$

Conclusion: Since the calculated value of χ^2 < the table value of χ^2 , H_o is accepted at 5% level of significance.

i.e. ,the accidents are uniformly distributed over the 6 days .

6. (i) The following data are collected on two characters. Based on this can you say that there is no relation between smoking and literacy. (April / May 2017)

	Smokers	**Non Smokers**
Literates	**83**	**57**
Illiterates	**45**	**68**

Solution:

Null Hypthesis H_0: No difference between the two treatment.

Alternative Hypothesis H_1: difference between the two treatment

Level of significance: $\alpha = 5\%$ or 0.05

Degrees of freedom=(r-1)(s-1)=(2-1)(2-1)=1

The test statistic is $\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$

$$\chi^2 = \frac{[(83*68)-(45*57)]^2(253)}{(83+45)(57+68)(83+57)(45+68)} = 9.47$$

Conclusion: Since χ^2 =9.47 > 3.841, so we reject H_0 at 5% level of significance

(ii) The theory predicts that the proportion of beans in the four groups A, B, C, D should be 9:3:3:1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287, and 118. Does the experimental result support the theory?

Solution:

Ho : The experimental data support the theory

Based on Ho, the expected numbers of beans in the four groups are as follows

Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$

882	900	324	0.360
313	300	169	0.563
287	300	169	0.563
118	100	324	3.240
			4.726

$\therefore \chi^2 = \sum \frac{(O-E)^2}{E}$ =4.726

Calculated value of χ^2 =4.726

Tabulated value of χ^2 is 7.81 at 5% level of significance. Since calculated value < tabulated value. Therefore, we accept the null hypothesis. i.e. the experimental data support the theory.

7. Two independent samples of eight and seven items respectively had the following values of the variable:

Sample 1	**9**	**11**	**13**	**11**	**15**	**9**	**12**	**14**
Sample 2	**10**	**12**	**10**	**14**	**9**	**8**	**10**	

Do the two estimates of population variance differ significantly at 5% level of significance?

Let S_1^2, S_2^2 be the sample variances and let σ_1^2, σ_2^2 be the variances of the two populations we have to test the significance of the differences the variances of the two samples. So we apply F-test of equality of variances.

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ (Variances are equal)

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$ (Variances are not equal)

To find S_1^2 and S_2^2

Sample I	
X_1	X_1^2
9	81
11	121
13	169
11	121
15	225
9	81
12	144
14	196
94	1138

Sample II	
X_2	X_2^2
10	100
12	121
10	100
14	196
9	81
8	64
10	100
73	785

	$n_1 = 8, n_2 = 7, \sum x_1 = 94, \sum x_1{}^2 = 1138$ $\sum x_2 = 73, \sum x_2{}^2 = 785$ $\overline{x_1} = \frac{\sum x_1}{n_1} = \frac{94}{8} = 11.75 \quad , \overline{x_2} = \frac{\sum x_2}{n_2} = \frac{73}{7} = 10.43$ $s_1^2 = \frac{\sum x_1^2}{n_1} - \left(\overline{x_1}\right)^2 = \frac{1138}{8} - (11.75)^2 = 4.19,$ $s_2^2 = \frac{\sum x_2^2}{n_2} - \left(\overline{x_2}\right)^2 = \frac{785}{7} - (10.43)^2 = 3.39$ $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{8 \times 4.19}{7} = 4.79 \quad , \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{7 \times 3.39}{6} = 3.96$ $\text{Since } S_1^2 > S_2^2, \text{the test statistic is } F = \frac{S_1^2}{S_2^2} = \frac{4.79}{3.96} = 1.21$ Number of degrees of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1) = (7, 6)$ For $(v_1, v_2) = (7,6)$, the table value of F at 5% level is $F_{0.05} = 4.21$ $\therefore F < F_{0.05}$ **Conclusion:** Since the calculated value of F < the table value of F, H_0 is accepted at 5% level of significance. The two samples are drawn from populations with same variances.
8.	**In one sample of 10 observations, the sum of squares of deviations of sample values from the sample mean was 120 and in another sample of 12 observations it was 314.Test whether this difference is significant at 5% level of significance. (Nov/Dec 2019)** **Null Hypothesis:** There is no difference in population variances $H_0 : \sigma_1^2 = \sigma_2^2$ **Alternate Hypothesis:** There is a significant difference in population variances (i.e) $H_1 : \sigma_1^2 \neq \sigma_2^2$ **Level of significance:** $\alpha = 5\%$ **The test statistic:** $F = \frac{S_1^2}{S_2^2}$ Given that

	$n_1 = 10,\ n_2 = 12$ $\sum (x - \bar{x})^2 = 120$ $\sum (y - \bar{y})^2 = 314$ $S_1^2 = \frac{1}{n_1 - 1} \sum (x - \bar{x})^2 = \frac{120}{9} = 13.33$ $S_2^2 = \frac{1}{n_2 - 1} \sum (y - \bar{y})^2 = \frac{314}{11} = 28.55$ $F = \frac{S_1^2}{S_2^2} = \frac{28.55}{13.33} = 2.14$ Degrees of freedom= (9,11) The Tabulated value of $F_{0.05} = 3.11$ Conclusion : Since Calculated value of F< Table value of F We accept H_0 and conclude that there is no significant difference in the sample variances.
9.	**In 120 throws of a single die, the following distribution of faces was observed.**

Face	**1**	**2**	**3**	**4**	**5**	**6**
Frequency	**30**	**25**	**18**	**10**	**22**	**15**

Can you say that the die is biased? (Nov/Dec 2019)

Solution:

Given n=6

Null HypthesisH_0: The die is unbiased.

Alternative Hypothesis H_1: The die is biased.

Level of significance: $\alpha = 5\%$ or 0.05 Degrees of freedom=n-1=6-1=5

On the assumption H_0, the expected frequency for each face=120 /6=20

Face	**Observed Frequency(O)**	**Expected Frequency(E)**	**O-E**	**$(O-E)^2$**	$\frac{(O-E)^2}{E}$
1	**30**	**20**	**10**	**100**	**5**
2	**25**	**20**	**5**	**25**	**1.25**
3	**18**	**20**	**-2**	**4**	**0.2**
4	**10**	**20**	**-10**	**100**	**5**
5	**22**	**20**	**2**	**4**	**0.2**
6	**15**	**20**	**-5**	**25**	**1.25**
					12.9

The test statistic is $\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right]$ =12.9

For v=5 degrees of freedom , the table of χ^2 at 5% level is $\chi^2_{0.05}$ = 11.07$\therefore \chi^2 > \chi^2_{0.05}$

Conclusion: Since the calculated value of χ^2 < the table value of χ^2 ,H_o is rejected at 5% level of significance.i.e., The die can be regarded as biased.

10. Fit a Poisson's distribution to the following data and the goodness of fit. Test at 5% level of significance. **(April/ May 2019)**

x	**0**	**1**	**2**	**3**	**4**	**5**
f	**142**	**156**	**69**	**27**	**5**	**1**

Solution:

Null Hypothesis H_0: Poisson distribution fit the given data

Alternative Hypothesis H_1: Poisson distribution not fit the given data

$$Mean = \overline{X} = \frac{\sum f_i x_i}{\sum f_i} = 1 \Rightarrow \overline{X} = \lambda = 1$$

By Poisson distribution, $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Now, Expected frequency = $= f\,\frac{e^{-\lambda}\lambda^x}{x!}$

<table>
<tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td rowspan="2">O_i</td><td rowspan="2">142</td><td rowspan="2">156</td><td rowspan="2">69</td><td rowspan="2">27</td><td>5</td><td>1</td></tr>
<tr><td colspan="2">6</td></tr>
<tr><td rowspan="2">E_i</td><td rowspan="2">147</td><td rowspan="2">147</td><td rowspan="2">74</td><td rowspan="2">25</td><td>6</td><td>1</td></tr>
<tr><td colspan="2">7</td></tr>
</table>

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
142	147	-5	25	0.1700
156	147	9	81	0.5510
69	74	-5	25	0.3378
27	25	2	4	0.16
6	7	-1	1	0.1459
				1.3467

the test statistic is $\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right]$ =1.3467

For v= 5-2= 3 degrees of freedom , the table of χ^2 at 5% level is $\chi^2_{0.05}$ = 7.815,

$\therefore \chi^2 < \chi^2_{0.05}$

Conclusion: Since the calculated value of χ^2 < the table value of χ^2 , H_o is accepted at 5% level of significance.

UNIT V DESIGN OF EXPERIMENTS

	PART-A
1.	**What is the aim of the design of the experiments?** **Solution:** The main aim of the design of experiments is to control the extraneous variables and hence to minimize the experimental error so that the results of the experiments could be attributed only to the experimental variables.
2.	**What are the basic principles of design of experiments? (April/ May 2017), (Nov/Dec 2019)** **Solution:** (i) Randomization (ii) Replication (iii) Local Control
3.	**What are the three essential steps to plan design of experiment?** **Solution:** To plan an experiment the following three are essential. 1. A Statement of the objective: Statement should clearly mention the hypothesis to be tested. 2. A description of the experiment: Description should include the type of experimental material, size of the experiment and the number of replications. 3. The outline of the method of analysis : The outline of the method consists of analysis of variance.
4.	**Define Analysis of Variance.** **Solution:** Analysis of Variance is a technique that will enable us to test for the significance of the difference among more than two sample means.
5.	**What are the assumptions of analysis of variance? (April / May 2021)** **Solution:** (i) The sample observations are independent (ii) The Environmental effects are additive in nature (iii) Population from which samples are taken is normal
6.	**What is the purpose of ANOVA?** **Solution:** The purpose of ANOVA is to test the homogeneity of several means.
7.	**Define Replication.** **Solution:** The repetition of the treatments under investigation is known as replication.
8.	**Define Experimental Error.** **Solution:** The variation from plot to plot caused by uncontrolled factors, that is factors beyond the control of the experimenter, is known as experimental error.
9.	**Define one-way classification and two way classifications in ANOVA.** **Solution:** The entire experiment influences on only single factor is one way classification. The entire experiment influences on only two factors is two way classification.

10. **Write down the ANOVA table for one way classification.** **(April / May 2017)**

Solution:

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Samples	SSC	C-1	$MSC = \frac{SSC}{C-1}$	$F_C = \frac{MSC}{MSE}$ if $MSC > MSE$
Within Samples	SSE	N-C	$MSE = \frac{SSE}{N-C}$	

11. **Find the missing value of A, B, C, D from the ANOVA table.**

S.V	**D.F**	**S.S**	**M.S.S**	**F_{cal}**
Treatment	**2**	**A**	**3**	**1.66**
Error	**B**	**C**	**5**	**-**
Total	**9**	**D**	**-**	**-**

Solution:

A=2 x 3 =6, B = Total SS – Total D.F = 9-2 =7,

C = 7 x 5 =35, D = A+C = 6+35 =41

12. **What is Latin Square Design?** **(April / May 2019)**

Solution:

It is square array of the letters A, B, C, D etc., of Latin square alphabets in this square array each letter appears once and only once in each row and column. For Latin square design involving n treatments, it is necessary to include n^2 observations, 'n' for each treatment.

13. **Write down the ANOVA table for Randomized Block Design.**

Solution:

Source of Variation	Sum of Degrees	Degree of freedom	Mean Square	F- Ratio
Column Treatments	SSC	c-1	$MSC = \frac{SSC}{c-1}$	$F_C = \frac{MSC}{MSE}$ if MSC > MSE
Row Treatments	SSR	r-1	$MSC = \frac{SSR}{r-1}$	$F_R = \frac{MSR}{MSE}$ if MSR > MSE
Error (or) Residual	SSE	(r-1) (c-1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	

14. Find the missing values from the ANOVA table.

S.V	**D.F**	**S.S**	**M.S.S**	**F_{cal}**
Treatment	**4**	**A**	**40**	**5**
Block	**B**	**C**	**115**	**14.375**
Error	**D**	**96**	**8**	**-**
Total	**19**	**601**	**-**	**-**

Solution:

A = 4 x 40 =160

D = 96 / 8 = 12

B = 3

C = 115 x 3 = 345

15. Why a 2 x 2 Latin Square is not possible?

Solution:

Consider a n X n Latin square design, the degree of freedom for SSE is

$$=(n^2-1)-(n-1)-(n-1)-(n-1)=n^2-1-3n+3=n^2-3n+2=(n-1)(n-2)$$

For n = 2, degree of freedom of SSE = 0 and hence MSE is not defined.

Comparisons are not possible. Hence a 2 X 2 Latin Square Design is not possible.

16. What are the advantages of completely randomized block design?

Solution:

The advantages of completely randomized experimental design as follows:

(i) Easy to lay out. (ii) Allow flexibility (iii) Simple Statistical Analysis

(iv)The lots of information due to missing data is smaller than with any other design

17. Write down the ANOVA table for Latin Square Design.

Solution:

Source of Variation	Sum of Degrees	Degree of freedom	Mean Square	F- Ratio
Column Treatment	SSC	n-1	$MSC=\frac{SSC}{n-1}$	$F_C=\frac{MSC}{MSE}$ if MSC > MSE
Row Treatments	SSR	n-1	$MSR=\frac{SSR}{n-1}$	$F_R=\frac{MSR}{MSE}$ if MSR > MSE
Between Treatments	SSK	n-1	$MSK=\frac{SSK}{n-1}$	$F_K=\frac{MSK}{MSE}$ if MSK > MSE
Error (or) Residual	SSE	(n-1) (n-2)	$MSE=\frac{SSE}{(n-1)(n-2)}$	

18. Mention the advantages of Latin square design over other designs. (April/May 2021)

Solution:

The advantages of the latin square design over other designs are:

(i) With a two-way stratification or grouping, the latin square controls more of the variation than the CRD or the randomized completely block design. The two-way elimination of variation often results in small error mean square.

(ii) The analysis is simple.

(iii) Even with missing data the analysis remains relatively simple.

19. **Write down the ANOVA table for 2^2 factorial design.**
Solution:

Source of Variation	Sum of Squares	Degrees of freedom	Mean Square	F- Ratio
a	SSA	**1**	$MSA = \frac{SSA}{d.f}$	$if\ MSA > SSE\ then\ F_A = \frac{MSA}{SSE}$
B	SSB	**1**	$MSB = \frac{SSB}{d.f}$	$if\ MSB > SSE\ then\ F_B = \frac{MSB}{SSE}$
ab	$SSAB$	**1**	$MSAB = \frac{SSAB}{d.f}$	$if\ MSAB > SSE\ then\ F_{AB} = \frac{MSAB}{SSE}$
ERROR	SSE	**N-c-r+1**	$MSE = \frac{SSE}{d.f}$	

20. **Compare RBD, LSD, CRD.**
Solution:

CRD	RBD	LSD
To influence one factor	To influence two factor	To influence more than two factor
No restriction further treatments	No restriction on treatment and replications	The number of replication of each treatment is equal to the number of treatment
-	Use only rectangular or Square field	Use only Square filed

PART-B

1. **The following are the number of mistakes made in 5 successive days by 4 technicians working for a photographic laboratory. Test whether the difference among the four sample means can be attributed to chance. (Test at a level of significance $\alpha = 0.01$)**

Technicians			
I	**II**	**III**	**IV**
6	**14**	**10**	**9**
14	**9**	**12**	**12**
10	**12**	**7**	**8**
8	**10**	**15**	**10**
11	**14**	**11**	**11**

Solution:

H_0: There is no significant difference between the technicians

H_1 : Significant difference between the technicians

We shift the origin

Total	**X_1**	**X_2**	**X_3**	**X_4**	**TOTAL**	**X_1^2**	**X_2^2**	**X_3^2**	**X_4^2**
	-4	4	0	-1	**-1**	16	16	0	1
	4	-1	2	2	**7**	16	1	4	4
	0	2	-3	-2	**-3**	0	4	9	4
	-2	0	5	0	**3**	4	0	25	0

	1	4	1	1	**7**	1	16	1	1
	-1	**9**	**5**	**0**	**13**	**37**	**37**	**39**	**10**

N= Total No of Observations = 20 T=Grand Total = 13

Correction Factor = $\frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 8.45$

$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 37 + 37 + 39 + 10 - 8.45 = 114.55$

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{(-1)^2}{5} + \frac{(9)^2}{5} + \frac{(5)^2}{5} + \frac{(0)^2}{5} - 8.45 = 12.95$$

SSE = TSS – SSC = 114.55-12.95= 101.6

ANOVA Table

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**
Between Samples	SSC=12.95	C-1= 4-1=3	$MSC = \frac{SSC}{K-1} = 4.317$	$F_C = \frac{MSC}{MSE} = 1.471$
Within Samples	SSE=101.6	N-C=20-4=16	$MSE = \frac{SSE}{N-K} = 6.35$	

Cal F_C = 1.471 & Tab F_C (16,3)=5.29

Conclusion : Cal F_C < Tab F_C $\Rightarrow$ There is no significance difference between the technicians

2. **As part of the investigation of the collapse of the roof of a building, a testing laboratory is given all the available bolts that connected the steel structure at 3 different positions on the roof. The forces required to shear each of these bolts (coded values) are as follows:**

Position 1: 90 82 79 98 83 91

Position 2: 105 89 93 104 89 95 86

Position 3 : 83 89 80 94

Perform an analysis of variance to test at the 0.05 level of significance whether the differences among the sample means at the 3 positions are significant. (April / May 2019)

Solution:

H_0: There is no significant difference between the sample means at the three positions.

H_1: Significant difference between the sample means at the three positions.

We shift the origin

Total	X_1	X_2	X_3	TOTAL	X_1^2	X_2^2	X_3^2
	1	16	-6	11	1	256	36
	-7	0	0	-7	49	0	0
	-10	4	-9	-15	100	16	81
	9	15	5	29	81	225	25

	-6	0	-	-6	36	0	-
	2	6	-	8	4	36	-
	-	-3	-	-3	-	9	-
	-11	38	-10	17	271	542	142

N= Total No of Observations = 17 T=Grand Total = 17

Correction Factor = $\frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 17$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 - C.F = 271 + 542 + 142 - 17 = 938$$

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} - C.F = \frac{(-11)^2}{6} + \frac{(38)^2}{7} + \frac{(-10)^2}{4} - 17 = 234.44$$

SSE = TSS – SSC = 938-234.44= 703.56

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Samples	SSC=234.44	C-1= 3-1=2	$MSC = \frac{SSC}{C-1} = 117.22$	$F_C = \frac{MSC}{MSE} = 2.332$
Within Samples	SSE=703.56	N-C=17-3=14	$MSE = \frac{SSE}{N-C} = 50.25$	

Cal F_C = 2.332 & Tab F_C (14,2) =3.74

Conclusion : Cal $F_C <$ Tab $F_C \Rightarrow$ There is no significance difference between the given positions.

3. **The following table shows the lives in hours of 4 batches of electric bulbs.**

1	**1610**	**1610**	**1650**	**1680**	**1700**	**1720**	**1800**
2	**1580**	**1620**	**1620**	**1700**	**1750**	**-**	**-**
3	**1460**	**1550**	**1600**	**1620**	**1640**	**1740**	**1820**
4	**1510**	**1520**	**1530**	**1570**	**1600**	**1680**	**-**

Perform an analysis of variance of these data and show that a significance test does not reject their homogeneity.

Solution:

H_0: The means of the lives of the four brands are homogeneous.

H_1: The means of the lives of the four brands are not homogeneous.

We shift the origin (subtract 1640)

	X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
	30	60	180	130	**400**	900	3600	32400	16900
	30	20	90	120	**260**	900	400	8100	14400
	-10	20	40	110	**160**	100	400	1600	12100
	-40	-60	20	70	**-10**	1600	3600	400	4900
	-60	-110	0	40	**-130**	3600	12100	0	1600
	-80		-100	-40	**-220**	6400		10000	1600
	-160		-180		**-340**	25600		32400	
Total	**-290**	**-70**	**50**	**430**	**120**	**39100**	**20100**	**84900**	**51500**

N= Total No of Observations = 25 T=Grand Total = **120**

Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = \dfrac{120^2}{25} = 576$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 39100 + 20100 + 84900 + 51500 - 576 = 195024$$

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{(-290)^2}{7} + \frac{(-70)^2}{5} + \frac{(50)^2}{7} + \frac{(430)^2}{6} - 576$$

$$SSC = 43592.09$$

SSE = TSS – SSC = 195024 – 43592.09=151431.91

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Samples	SSC=43592.09	C-1= 4-1=3	$MSC = \dfrac{SSC}{C-1} = 14530.69$	$F_C = \dfrac{MSC}{MSE} = 2.02$
Within Samples	SSE=151431.91	N-C=25-4=21	$MSE = \dfrac{SSE}{N-C} = 7211.04$	

Cal F_C = 2.02 & Tab F_C (3,21)=3.07

Conclusion : Cal F_C< Tab F_C ⇒ There is no significance difference between the technicians

4. **Four varieties A, B,C, D of a fertilizer are tested in a randomized block design with 4 replication. The plot yields in pounds are as follows:**

Column / Row	**1**	**2**	**3**	**4**
1	**A(12)**	**D(20)**	**C(16)**	**B(10)**
2	**D(18)**	**A(14)**	**B(11)**	**C(14)**
3	**B(12)**	**C(15)**	**D(19)**	**A(13)**
4	**C(16)**	**B(11)**	**A(15)**	**D(20)**

Analyse the experimental yield. **(Nov/Dec 2019)**

Solution:

H_0: There is no significant difference between the fertilizers and replication

H_1 : Significant difference between the fertilizers and replication

Variety	**Block**				**Total varieties**				
	1 (X_1)	2 (X_2)	3 (X_3)	4 (X_4)		X_1^2	X_2^2	X_3^2	X_4^2
A	12	14	15	13	54	144	196	225	169
B	12	11	11	10	44	144	121	121	100
C	16	15	16	14	61	256	225	256	196
D	18	20	19	20	77	324	400	361	400
	58	**60**	**61**	**57**	236	868	942	963	865

N=16; T=Grand Total = 236

Correction Factor = $\dfrac{(Grand\ \ total)^2}{Total\ \ No\ \ of\ \ Observations} = \dfrac{(236)^2}{16} = 3481$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 868 + 942 + 963 + 865 - 3481 = 157$$

$$SSC = \frac{(\sum X_1)^2}{C_1} + \frac{(\sum X_2)^2}{C_2} + \frac{(\sum X_3)^2}{C_3} + \frac{(\sum X_4)^2}{C_4} - C.F = 841 + 900 + 930 + 812 - 3481 = 2$$

$$SSR = \frac{(\sum Y_1)^2}{R_1} + \frac{(\sum Y_2)^2}{R_2} + \frac{(\sum Y_3)^2}{R_3} + \frac{(\sum Y_3)^2}{R_4} - C.F = 729 + 484 + 930.25 + 1482.25 - 3481 = 144.5$$

SSE = TSS – SSC – SSR = 157-2-144.5=10.5

ANOVA Table

Source of Variation	**Sum of Squares**	**Degree of freedom**	**Mean Square**	**F- Ratio**	**F_{Tab}Ratio**
Between columns	SSC=2	c – 1=3	MSC = 0.67	F_C = 0.545	$F_{5\%}(3, 9)$ =3.86
Between	SSR=144	r - 1= 3	MSR=48.1	F_R = 39.27	$F_{5\%}(3,9)$ =

rows	.5		7		3.86
Residual	SSE = 10.5	(c – 1)(r – 1) = 9	MSE = 1.17		

Conclusion: Cal F_C<Tab F_C and Cal F_R> Tab F_R
There is no significant difference between the columns and there is a significant difference between the rows.

5. The following table gives the number of refrigerators sold by 4 salesmen in three months.

Months	Salesman			
	A	**B**	**C**	**D**
May	**50**	**40**	**48**	**39**
June	**46**	**48**	**50**	**45**
July	**39**	**44**	**40**	**39**

Is there a significant difference in the sale made by the four salesmen? Is there a significant difference in the sales made during different months? (April / May 2021)

Solution:

Null Hypothesis H_0**:** There is no significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

Alternate Hypothesis H_1**:** There is a significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

Test statistic:

To simplify calculations, we deduct 40 from each value

Seasons		**A** X_1	**B** X_2	**C** X_3	**D** X_4	**Seasons Total**	X_1^2	X_2^2	X_3^2	X_4^2
Y_1	Summer	10	0	8	-1	**17**	100	0	64	1
Y_2	Winter	6	8	10	5	**29**	36	64	100	25
Y_3	Monson	-1	4	0	-1	**2**	1	16	0	1
Total		15	12	18	3	**48**	137	80	164	27

Step1: N= Total No of Observations = 12

Step 2: T=Grand Total =48

Step 3: Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = \dfrac{T^2}{N} = \dfrac{48^2}{12} = 192$

Step 4: $\text{TSS} = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 137 + 80 + 164 + 27 - 192 = 216$

Step 5: $SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{15^2}{3} + \frac{12^2}{3} + \frac{18^2}{3} + \frac{3^2}{3} - 192 = 42$

Step 6: $SSR = \frac{\left(\sum Y_1\right)^2}{R_1} + \frac{\left(\sum Y_2\right)^2}{R_2} + \frac{\left(\sum Y_3\right)^2}{R_3} - C.F = \frac{17^2}{4} + \frac{29^2}{4} + \frac{2^2}{4} - 192 = 91.5$

Step 7: SSE=TSS-SSC-SSR $= 216 - 42 - 91.5$

SSE $= 82.5$

Step 8: ANOVA TABLE:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	varience	F – ratio
Between Columns (Salesmen)	SSC=42	c-1=4-1=3	$MSC = \frac{SSC}{c-1} = \frac{42}{3} = 14$	$MSC = \frac{MSC}{MSE} = \frac{14}{13.75} = 1.018$	$F_C(3,6) = 4.76$
Between rows (Seasons)	SSR $= 91.5$	r-1=3-1=2	$MSR = \frac{SSR}{r-1} = \frac{91.5}{2} = 45.75$	$MSR = \frac{MSR}{MSE} = \frac{45.75}{13.75} = 3.327$	$F_R(2,6) = 5.14$
Error	SSE=82.5	(c-1)(r-1)=6	$MSE = \frac{SSE}{(c-1)(r-1)} = \frac{82.5}{6} = 13.75$		
Total	216	11			

Conclusion:

1) Cal F_C **< Table** $F_{C,0.05}(3,6)$

Hence we accept the H_0 and we conclude that there is no significant difference between sales in the three seasons.

2) **Cal** F_R **< Table** $F_{R,0.05}(2,6)$.

Hence we accept the H_0 and we conclude that there is no significant difference between in the sales of 4 salesmen.

6. For doctors each test, 4 treatments for a certain diseases and observe the number of days each patient takes to recover. The results are as follows.

		TREATMENTS			
		I	**II**	**III**	**IV**
DOCTORS	**A**	**10**	**14**	**19**	**20**
	B	**11**	**15**	**17**	**21**
	C	**9**	**12**	**16**	**19**

	D	8	13	17	20

Discuss the difference between (i) doctors and (ii) treatments

Solution:

Null Hypothesis H_0 **:** There is no significant difference between doctors and treatments

Alternate Hypothesis H_1 **:** There is a significant difference between doctors and treatments

	A **X_1**	**B** **X_2**	**C** **X_3**	**D** **X_4**	**Total**	**X_1^2**	**X_2^2**	**X_3^2**	**X_4^2**
I	**10**	**14**	**19**	**20**	**63**	100	196	361	400
II	**11**	**15**	**17**	**21**	**64**	121	225	289	441
III	**9**	**12**	**16**	**19**	**56**	81	144	256	361
IV	8	**13**	**17**	**20**	**58**	64	169	289	400
Total	38	**54**	**69**	**80**	**241**	366	734	1195	1602

Step1: N= Total No of Observations = 16

Step 2: T=Grand Total =241

Step 3: Correction Factor = $\frac{(\text{Grand total})^2}{\text{Total No of Observations}} = \frac{T^2}{N} = \frac{241^2}{16} = 3630.06$

Step 4: $TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 366+734+1195+1602 - 3630.06 = 266.94$

Step 5:

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{38^2}{4} + \frac{54^2}{4} + \frac{69^2}{4} + \frac{80^2}{4} - 3630.06 = 250.19$$

Step 6: $SSR = \frac{\left(\sum Y_1\right)^2}{R_1} + \frac{\left(\sum Y_2\right)^2}{R_2} + \frac{\left(\sum Y_3\right)^2}{R_3} + \frac{\left(\sum Y_4\right)^2}{R_4} - C.F = \frac{63^2}{4} + \frac{64^2}{4} + \frac{56^2}{4} + \frac{58^2}{4} - 3630.06 = 11.19$

Step 7: $SSE = TSS - SSC - SSR = 266.94 - 250.19 - 11.19 = 5.56$

Step 8: ANOVA TABLE:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	variance	F – ratio
Between Columns (Salesmen)	SSC=250.19	c-1=4-1=3	$MSC = \frac{SSC}{c-1} = \frac{250.19}{3} = 83.40$	$MSC = \frac{MSC}{MSE} = \frac{83.40}{0.62} = 134.52$	$F_C(3,9) = 3.86$

Between rows (Seasons)	SSR =11.19	r-1=4-1=3	$MSR = \frac{SSR}{r-1} = \frac{11.19}{3} = 3.73$	$MSR = \frac{MSR}{MSE} = \frac{3.73}{0.62} = 6.02$	$F_R(3,9) = 3.86$
Error	SSE=5.56	(c-1)(r-1)=9	$MSE = \frac{SSE}{(c-1)(r-1)} = \frac{5.56}{9} = 0.62$		
Total	266.94	15			

Conclusion:

1) Cal F_C **< Table** $F_{C,0.05}(3,9)$

Hence we accept the H_0 and we conclude that there is no significant difference between sales in the three seasons.

2) **Cal** F_R **< Table** $F_{R,0.05}(3,9)$.

Hence we accept the H_0 and we conclude that there is no significant difference between in the sales of 4 salesmen.

7. **Analyse the variance in the following latin square of yields (in kgs) of paddy where A, B, C, D denote the different methods of cultivation.**

D 122	**A 121**	**C 123**	**B 122**
B 124	**C 123**	**A 122**	**D 125**
A 120	**B 119**	**D 120**	**C 121**
C 122	**D 123**	**B 121**	**A 122**

(April / May 2021)

Examine whether the different methods of cultivation have given significantly different yields.

Solution:

H_0 : There is no significant difference between the methods of cultivation.

H_1 : There is a significant difference between the methods of cultivation.

Let us take 120 as origin for simplifying the calculation

Variety	X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
Y_1	2	1	3	2	8	4	1	9	4
Y_2	4	3	2	5	14	16	9	4	25
Y_3	0	-1	0	1	0	0	1	0	1
Y_4	2	3	1	2	8	4	9	1	4
	8	6	6	10	30	24	20	14	34

N=Total No of Observations = 16 T=Grand Total = 30

Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}}$ = 56.25

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 24 + 20 + 14 + 34 - 56.25 = 35.75$$

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{(8)^2}{4} + \frac{(6)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 56.25 = 2.75$$

$$SSR = \frac{\left(\sum Y_1\right)^2}{R_1} + \frac{\left(\sum Y_2\right)^2}{R_2} + \frac{\left(\sum Y_3\right)^2}{R_3} + \frac{\left(\sum Y_4\right)^2}{R_4} - C.F = \frac{(8)^2}{4} + \frac{(14)^2}{4} + \frac{(0)^2}{4} + \frac{(8)^2}{4} - 56.25 = 24.75$$

To find SSK

Treatment	1	2	3	4	Total
A	1	1	3	7	12
B	-1	1	1	-1	0
C	5	3	-1	3	10
D	3	5	3	-1	10

SSE= TSS - SSC-SSR-SSK = 88-4-2-22=60

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Columns	SSC=4	n-1=3	$MSC = \frac{SSC}{n-1}$ =1.33	$F_C = \frac{MSC}{MSE}$ =7.52
Between Rows	SSR=2	n-1=3	$MSR = \frac{SSR}{n-1}$ =0.67	$F_R = \frac{MSR}{MSE}$ =14.9
Between Treatments	SST=22	n-1=3	$MSK = \frac{SSK}{n-1}$ = 7.33	$F_K = \frac{MSK}{MSE}$ =1.36
Error (or) Residual	SSE=60	(n-1) (n-2)=6	$MSE = \frac{SSE}{(n-1)(n-2)} = 10$	

Table value F(3,6) degrees of freedom 4.76

There is significant difference between treatments

We shift the origin $X_{ij} = x_{ij} - 100$; n = 4; N = 16

	I	II	III	IV	Total=T_{i*}	$[T_{i*}^2]/n$	ΣX_{*ij}^2
A	2	1	3	2	8	16	18
B	4	3	2	5	14	49	54
C	0	-1	0	1	0	0	2
D	2	3	1	2	8	16	18
Total=T_{*j}	8	6	6	10	30	81	92

$[T_{*j}^2]/n$	16	9	9	25	59
ΣX_{i*}^2	24	20	14	34	92

	Letters				Total=T_{i*}	$[T_{i*}^2]/n$
P	1	2	0	2	5	6.25
Q	2	4	-1	1	6	9
R	3	3	1	2	9	20.25
S	2	5	0	3	10	25
Total					30	60.5

T=Grand Total = 30 ;

Correction Factor = $\frac{(Grand\ total)^2}{Total\ \ No\ of\ \ Observations} = \frac{(30)^2}{16}$

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 92 - \frac{(30)^2}{16} = 35.75$$

$$SSR = \frac{\sum T_{i*}^2}{n} - C.F = 81 - \frac{(30)^2}{16} = 24.75$$

$$SSC = \frac{\sum T_{*j}^2}{n} - C.F = 59 - \frac{(30)^2}{16} = 2.75$$

$$SSL = \frac{\sum T_{i*}^2}{n} - C.F = 60.5 - \frac{(30)^2}{16} = 4.25$$

SSE = TSS – SSC – SSR-SSL = 35.75 – 24.75 – 2.75 – 4.25 = 4

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio	F_{Tab}Ratio (5% level)
Between Columns	SSC=2.75	n - 1= 3	MSC = 0.92	F_C = 1.37	Fc(3, 6)=4 .76
Between Rows	SSR=24.75	n - 1= 3	MSR=8.25	F_R= 12.31	F_R(3, 6)=4 .76
Between Treatments	SSL = 4.25	n - 1= 3	MSL = 1.42	F_T= 2.12	F_T(3, 6)=4 .76
Residual	SSE= 4	(n – 1)(n – 2) = 6	MSE = 0.67		
Total	35.75				

Conclusion :

Cal F_C< Tab F_C , Cal F_R> Tab F_R and Cal F_T< Tab F_T

⇒ There is no significant difference between the columns, There is a significant difference between the rows and There is no significant difference between the treatments.

8. **The following is a Latin square of a design when 4 varieties of seeds are being tested. Set up the analysis of variance table and state your conclusion. The following is a Latin square of a design when 4 varieties of seeds are being tested. Set up the analysis of variance table and state your conclusion. You may carry out suitable change of origin and scale.**

A 105	**B 95**	**C 125**	**D 115**
C 115	**D 125**	**A 105**	**B 105**
D 115	**C 95**	**B 105**	**A 115**
B 95	**A 135**	**D 95**	**C 115**

(April / May 2017)

Solution:

H_0 : Four varieties are similar

H_1 : Four varieties are not similar

Let us take 100 as origin and divide by 5 for simplifying the calculation

Variety	X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
Y_1	1	-1	5	3	8	1	1	25	9
Y_2	3	5	1	1	10	9	25	1	1
Y_3	3	-1	1	3	6	9	1	1	9
Y_4	-1	7	-1	3	8	1	49	1	9
	6	10	6	10	32	20	76	28	28

N=Total No of Observations = 16 T=Grand Total = 32

Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = 64$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 20 + 76 + 28 + 28 - 64 = 88$$

$$SSC = \frac{\left(\sum X_1\right)^2}{C_1} + \frac{\left(\sum X_2\right)^2}{C_2} + \frac{\left(\sum X_3\right)^2}{C_3} + \frac{\left(\sum X_4\right)^2}{C_4} - C.F = \frac{(6)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 64 = 4$$

$$SSR = \frac{\left(\sum Y_1\right)^2}{R_1} + \frac{\left(\sum Y_2\right)^2}{R_2} + \frac{\left(\sum Y_3\right)^2}{R_3} + \frac{\left(\sum Y_4\right)^2}{R_4} - C.F = \frac{(8)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(8)^2}{4} - 64 = 2$$

To find SSK

Treatment	1	2	3	4	Total
A	1	1	3	7	12
B	-1	1	1	-1	0
C	5	3	-1	3	10
D	3	5	3	-1	10

$$SSK = \frac{(\sum Y_1)^2}{K_1} + \frac{(\sum Y_2)^2}{K_2} + \frac{(\sum Y_3)^2}{K_3} + \frac{(\sum Y_4)^2}{K_4} - C.F = 22$$

SSE= TSS - SSC-SSR-SSK = 88-4-2-22=60

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Columns	SSC=4	n-1=3	$MSC = \frac{SSC}{n-1}$ =1.33	$F_C = \frac{MSC}{MSE}$ =7.5 2
Between Rows	SSR=2	n-1=3	$MSR = \frac{SSR}{n-1}$ =0.67	$F_R = \frac{MSR}{MSE}$ =14 .9
Between Treatment s	SST=22	n-1=3	$MSK = \frac{SSK}{n-1}$ = 7.33	$F_K = \frac{MSK}{MSE}$ =1.3 6
Error (or) Residual	SSE=60	(n-1) (n-2)=6	$MSE = \frac{SSE}{(n-1)(n-2)} = 10$	

Table value F(3,6) degrees of freedom 4.76

There is significant difference between treatments

9. **A variable trial was conducted on wheat with 4 varieties in a Latin Square Design. The plan of the experiment and the per plot yield are given below :**

C	**25**	**B**	**23**	**A**	**20**	**D**	**20**
A	**19**	**D**	**19**	**C**	**21**	**B**	**18**
B	**19**	**A**	**14**	**D**	**17**	**C**	**20**
D	**17**	**C**	**20**	**B**	**21**	**A**	**15**

Analyse data and interpret the result

Solution:

H_0 : Four varieties are similar

H_1 : Four varieties are not similar

Subtract 20 from all the items

Variety	X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
Y_1	5	3	0	0	8	25	9	0	0
Y_2	-1	-1	1	-2	-3	1	1	1	4

Y_3	-1	-6	-3	0	-10	1	36	9	0
Y_4	-3	0	1	-5	-7	9	0	1	25
	0	-4	-1	-7	-12	36	46	11	29

N=Total No of Observations = 16 T=Grand Total = -12

Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = 9$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 36 + 46 + 11 + 29 - 9 = 113$$

$$SSC = \frac{\left(\sum X_1^2\right)}{N_1} + \frac{\left(\sum X_2^2\right)}{N_1} + \frac{\left(\sum X_3^2\right)}{N_1} + \frac{\left(\sum X_4^2\right)}{N_1} - C.F = \frac{(0)^2}{4} + \frac{(-4)^2}{4} + \frac{(-1)^2}{4} + \frac{(-7)^2}{4} - 9 = 7.5$$

$$SSR = \frac{\left(\sum Y_1\right)^2}{N_1} + \frac{\left(\sum Y_2\right)^2}{N_2} + \frac{\left(\sum Y_3\right)^2}{N_2} + \frac{\left(\sum Y_4\right)^2}{N_2} - C.F = \frac{(8)^2}{4} + \frac{(-3)^2}{4} + \frac{(-10)^2}{4} + \frac{(-7)^2}{4} - 9 = 46.5$$

To find SSK

Treatment	1	2	3	4	Total
A	0	-1	-6	-5	-12
B	3	-2	-1	1	1
C	5	1	0	0	6
D	0	-1	-3	-3	-7

$$SSK = \frac{(-12)^2}{4} + \frac{(1)^2}{4} + \frac{(6)^2}{4} + \frac{(-7)^2}{4} - 9 = 48.5$$

SSE= TSS - SSC-SSR-SSK = 113-7.5-46.5-48.5=10.5

ANOVA Table

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
Between Columns	SSC=7.5	n-1=3	$MSC = \frac{SSC}{n-1}$ =2.5	$F_C = \frac{MSC}{MSE}$ =1.43
Between Rows	SSR=46.5	n-1=3	$MSR = \frac{SSR}{n-1}$ =15.5	$F_R = \frac{MSR}{MSE}$ =8.86
Between Treatments	SST=48.5	n-1=3	$MSK = \frac{SSK}{n-1}$ =16.17	$F_K = \frac{MSK}{MSE}$ =9.24
Error (or) Residual	SSE=10.5	(n-1) (n-2)=6	$MSE = \frac{SSE}{(n-1)(n-2)}$ $=1.75$	

	Table value F(3,6) degrees of freedom 4.76 There is significant difference between treatments
10.	**Analyse 2^2 factorial experiments for the following table.**

Treatment	**Replications**			
	I	**II**	**III**	**IV**
(1)	**12**	**12.3**	**11.8**	**11.6**
a	**12.8**	**12.6**	**13.7**	**14**
b	**11.5**	**11.9**	**12.6**	**11.8**
ab	**14.2**	**14.5**	**14.4**	**15**

SOLUTION:

Null hypothesis: All the mean effects are equal.

Let A and B be the two factors. Let n= number of replications=4

Subtract 12 from each

Treatment	Replications			
	I	II	III	IV
(1)	0	0.3	-0.2	-0.4
a	0.8	0.6	1.7	2
b	-0.5	-0.1	0.6	-0.2
ab	2.2	2.5	2.4	3

Let us find SS for the table

Treatment	Replications				Row Total R_i	R_i^2
	I	II	III	IV		
(1)	0	0.3	-0.2	-0.4	-0.3	0.09
a	0.8	0.6	1.7	2	5.1	26.01
b	-0.5	-0.1	0.6	-0.2	-0.2	0.04
ab	2.2	2.5	2.4	3	10.1	102.01
Column Total C_j	2.5	3.3	4.5	4.4	T=14.7	
C_j^2	6.25	10.89	20.25	19.36		

T=14.7; Correction factor= $\frac{T^2}{N}$ =13.5

TSS=21.19, SSC=0.688, SSR=18.54 , SSE=1.96

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio	F_{Tab}Ratio
b	S_B =1.63	1	MSB=1.63	F_B =7.409	10.56
a	S_A =15.41	1	MSA=15.41	F_A =70.04	10.56
ab	S_{AB} =1.50	1	MSAB=1.50	F_{AB} =6.81	10.56
Error	SSE=1.962	N-C-r+1=9	SSE=1.962		

Cal(F_A)=70.04 $\Rightarrow H_0$ is rejected at 1% level

Cal(F_B)=7.409 $\Rightarrow H_0$ is accepted at 1% level

Cal(F_{AB})=10.56 $\Rightarrow H_0$ is accepted at 1% level