MA8391 - PROBABILITY AND STATISTICS
(II Information Technology & II Bio Technology – IV SEMESTER)
UNIT I – PROBABILITY AND RANDOM VARIABLES

PART – A

1. **If A and B are two events such that** $P(A\cup B)=\frac{3}{4},\ P(A\cap B)=\frac{1}{4},\ P(\bar{A})=\frac{2}{3},\ find\ P\left(\bar{A}/B\right)$.

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$\frac{3}{4}=\frac{1}{3}+P(B)-\frac{1}{4}\ \Rightarrow P(B)=\frac{2}{3}$$

$$P\left(\frac{\bar{A}}{B}\right)=\frac{P\left(\bar{A}\cap B\right)}{P(B)}=\frac{P(B)-P(A\cap B)}{P(B)}=\frac{\frac{2}{3}-\frac{1}{4}}{\frac{2}{3}}=\frac{5}{8}$$

2. **Write down the axioms of probability.**

i) $P(E)\geq 0$ **ii)** $P(S)=1$ **iii)** $P(\bigcup_{i=1}^{n}E_i)=\sum_{i=1}^{n}P(E_i)$ if E_i's are mutually exclusive events

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

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

4. **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.** **(Nov/Dec 2015)**

A coin is tossed three times then the sample space is $S=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Let *X* be an event getting head.

The probability mass function is $P(X=x)=\begin{cases}\frac{1}{8} & for\ x=0\\ \frac{3}{8} & for\ x=1\\ \frac{3}{8} & for\ x=2\\ \frac{1}{8} & for\ x=3\end{cases}$

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

5. **Test whether** $f(x)=\begin{cases}|x|, & -1\leq x\leq 1\\ 0, & otherwise\end{cases}$ **can be the probability density function of a continuous random variable?** **(April / May 2015)**

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

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{-1}^{1}|x|\,dx=2\int_{0}^{1}x\,dx=2\left[\frac{x^2}{2}\right]_0^1=1-0=1$$

$\Rightarrow$ Yes, *f(x)* can be the probability density function of a continuous random variable.

6. **Let X be a random variable with E(X)=1, E[X(X-1)]= 4. Find Var(X), Var(3-2X).** **(April/May 2008)**

	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 $Var(X)=E\left[X^2\right]-\left(E\left[X\right]\right)^2=5-1=4$ $Var(2-3X)=(-3)^2\,Var(X)=9\times 4=36 \qquad \because Var(aX+b)=(a)^2\,Var(X)$
7.	**Find the value of '*k*' for a Continuous random variable X whose probability density function is given by $f(x)=kx^2e^{-x};\ x\geq 0$. (April/May 2017)** Since *X* is a continuous random variable $f(x)\geq 0$ and $x\geq 0$ and $\int_0^\infty f(x)\,dx=1$ $\int_0^\infty k\,x^2\,e^{-x}\,dx=1$ $\Rightarrow k\left[(x^2)(-e^{-x})-(2x)(e^{-x})+(2)(-e^{-x})\right]_0^\infty=1$ $\Rightarrow k\left[0-(-2)\right]=1$ $\Rightarrow 2k=1\Rightarrow \boxed{k=\frac{1}{2}}$
8.	**If a random variable X has the p.d.f $f(x)=\begin{cases}\frac{1}{2}(x+1); -1<x<1\\ 0 \quad ;otherwise\end{cases}$. Find the mean and variance of X.** $\text{Mean}=\int_{-\infty}^{\infty} xf(x)\,dx=\frac{1}{2}\int_{-1}^{1}x(x+1)\,dx=\frac{1}{2}\int_{-1}^{1}(x^2+x)\,dx=\frac{1}{2}\left(\frac{x^3}{3}+\frac{x^2}{2}\right)_{-1}^{1}=\frac{1}{3}$ $\mu_2'=\int_{-\infty}^{\infty}x^2f(x)dx=\frac{1}{2}\int_{-1}^{1}(x^3+x^2)\,dx=\frac{1}{2}\left[\frac{x^4}{4}+\frac{x^3}{3}\right]_{-1}^{1}=\frac{1}{2}\left[\frac{1}{4}+\frac{1}{3}-\frac{1}{4}+\frac{1}{3}\right]=\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{3}$ $Variance=\mu_2'-\left(\mu_1'\right)^2=\frac{1}{3}-\frac{1}{9}=\frac{3-1}{9}=\frac{2}{9}$.
9.	**Let $M_x(t)=\frac{1}{1-t}$ such that t≠ 1, be the mgf of r.v X. Find the mgf of Y = 2X +1.** $M_Y(t)=M_{2X+1}(t)=e^tM_X(2t) \qquad \because M_{aX+b}(t)=e^{bt}M_X(at)$ $=e^t\left[\frac{1}{1-t}\right]_{t\to 2t}=\frac{e^t}{1-2t}$
10.	**If the random variable has the moment generating function $M_X(t)=\frac{3}{3-t}$, compute E[X^2].** **(May/June 2016)** $M_X(t)=\frac{3}{3-t}=\frac{3}{3\left(1-\frac{t}{3}\right)}=\left(1-\frac{t}{3}\right)^{-1}$

	$M_X(t)=\frac{3}{3-t}=\frac{3}{3\left(1-\frac{t}{3}\right)}=\left(1-\frac{t}{3}\right)^{-1}$ $=1+\left(\frac{t}{3}\right)+\left(\frac{t}{3}\right)^2+\left(\frac{t}{3}\right)^3+\cdots=1+\frac{1}{3}\left(\frac{t}{1!}\right)+\frac{2}{9}\left(\frac{t^2}{2!}\right)+\frac{2}{9}\left(\frac{t^3}{3!}\right)+\cdots$ $E\left(X^r\right)=\mu_r'=\text{coefficient of}\left(\frac{t^r}{r!}\right)\text{ in } M_X(t)$ $E\left(X^2\right)=\mu_2'=\text{coefficient of}\left(\frac{t^2}{2!}\right)\text{ in} M_X(t)=\frac{2}{9}.$
11.	**A random variable X has density function given by $f(x)=\begin{cases}2e^{-2x}; x\ge 0\\ 0 \quad ; x<0\end{cases}$. Find m.g.f of X** $M_X(t)=E\left(e^{tx}\right)=\int_{-\infty}^{\infty}e^{tx}f(x)dx=\int_0^{\infty}e^{tx}2e^{-2x}dx=2\int_0^{\infty}e^{-(2-t)x}dx$ $=2\left[\frac{e^{-(2-t)x}}{-(2-t)}\right]_0^{\infty}=\frac{-2}{2-t}[e^{-\infty}-e^{-0}]=\frac{-2}{2-t}[0-1]=\frac{2}{2-t}, t<2.$
12.	**A continuous RV X has the pdf $f(x)=\frac{x^2e^{-x}}{2},\ x\ge 0$. Find the r^{th} moment of X about the origin.** $\mu_r'=E[X^r]=\int_{-\infty}^{\infty}x^r f(x)dx=\int_0^{\infty}x^r\frac{x^2e^{-x}}{2}dx=\frac{1}{2}\int_0^{\infty}x^{r+2}e^{-x}dx=\frac{1}{2}\int_0^{\infty}x^{(r+3)-1}e^{-x}dx$ $=\frac{1}{2}\Gamma(r+3) \qquad \because \int_0^{\infty}x^{n-1}e^{-x}dx=\Gamma(n)$ $=\frac{1}{2}(r+2)! \qquad \because \text{if } n \text{ is positive integer } \Gamma(n)=(n-1)!$
13.	**For a Binomial distribution with mean 2 and standard deviation $\sqrt{2}$, find the first two terms of the distribution. (May/June 2014)** $np=6 \text{ and} \sqrt{npq}=\sqrt{2}\Rightarrow npq=2\Rightarrow 6q=2\Rightarrow \boxed{q=\frac{1}{3}}\ \therefore \boxed{p=\frac{2}{3}}\quad n\times\frac{2}{3}=6\ \therefore \boxed{n=9}$ $P(X=x)=nC_xp^xq^{n-x}=9C_x\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{9-x}$ $P(X=0)=9C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^9=\left(\frac{1}{3}\right)^9$ $P(X=1)=9C_1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^8=9\times\frac{2}{3}\times\frac{1}{3^8}=\frac{2}{3^7}$
14.	**Define Binomial Distribution .What are its mean and variance? (April/May 2017)** The Probability of 'x' successes in 'n' trials is given by $P(X=x)=nC_x\, p^x\, q^{n-x}$, $x=0,1,2,\ldots$ Mean = np and variance = npq

15.	**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 among a sample of 200 jobs there are no jobs that have to wait until weekends.** X be the Random variable denoting the no. of jobs that have to wait *p = 1%= 0.01, n = 200,* $\lambda = np = (200)(0.01) = 2$, By Poisson distribution, $P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!},\quad x=0,1,2,.....$ Probability that there are no jobs that have to wait until weekends = $P(X=0)=\frac{e^{-2}2^0}{0!}=e^{-2}=0.1353$
16.	**A quality control inspector rejects 40% of a certain product. Find the probability that the first acceptable product is the third one inspected.** Probability of rejection = q = 0.4 Probability of acceptance = p = 0.6 $P(X=x)=q^{x-1}p,\ \ x=1,2,3,....$ P(X=3) = $q^{3-1}p = (.4)^2(.6) = 0.096$
17.	**If the probability that a target is destroyed on any one shot is 0.5, find the probability that it would be destroyed on 6th attempt. (Nov / Dec 2013)** Given that, the probability that a target is destroyed on any one shot is 0.5 $p=0.5 \Rightarrow q=1-p=1-0.5=0.5$ By Geometric Distribution, $P(X=x)=q^{x-1}p,\ \ x=1,2,3,....$; $P(X=6)=(0.5)^{6-1}(0.5)=(0.5)^6=0.0156$
18.	**If X is a Uniformly distributed R.V with mean 1 and variance $\frac{4}{3}$, find P(X<0).** Mean = $\frac{a+b}{2}=1 \Rightarrow\ a+b=2$ --------------(1) variance = $\frac{(b-a)^2}{12}=\frac{4}{3}\ \Rightarrow\ b-a=4$ ---------------(2) (1) + (2) $\Rightarrow$2b = 6 $\Rightarrow$ b = 3 (1) − (2) $\Rightarrow$2a = -2 $\Rightarrow$ a = -1 $f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{otherwise}\end{cases} \Rightarrow f(x)=\begin{cases}\frac{1}{4} & , -1<x<3\\ 0 & , \text{otherwise}\end{cases}$ $P(X<0)=\int_{-1}^{0} f(x)dx = \int_{-1}^{0}\frac{1}{4}dx = \frac{1}{4}[x]_{-1}^{0} = \frac{1}{4}.$
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 appliances is atleast 10.5?** Mean of the exponential distribution = E(X) = $1/\lambda \Rightarrow 10 = 1/\lambda$ $\lambda=\frac{1}{10},\ \ f(x)=\lambda e^{-\lambda x},\ x>0 \Rightarrow f(x)=\frac{1}{10}e^{-\frac{x}{10}},x>0$ $P(X>10.5)=\int_{10.5}^{\infty} f(x)dx=\int_{10.5}^{\infty}\frac{1}{10}e^{-\frac{x}{10}}dx\ = e^{-1.05}\ = 0.3499$

20. **If X is a normal variate with mean $=30$ and S.D = 5. Find $P\left[26\le X\le 40\right]$**

X follows N(30, 5) $\therefore \mu=30$ & $\sigma=5$

Let $Z=\dfrac{X-\mu}{\sigma}$ be the standard normal variate

$$P\left[26\le X\le 40\right]=P\left[\frac{26-30}{5}\le Z\le\frac{40-30}{5}\right]=P\left[-0.8\le Z\le 2\right]=P\left[-0.8\le Z\le 0\right]+P\left[0\le Z\le 2\right]$$

$$=P\left[0\le Z\le 0.8\right]+\left[0\le Z\le 2\right]=0.2881+0.4772=0.7653.$$

PART – B

1. i) 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	$2K^2$	$7K^2+K$

Find (i) K (ii) Evaluate $P(X<6), P(X\ge 6)$ and $P(0<X<5)$ (iii) Determine the distribution function of X (iv) $P(1.5<X<4.5/X>2)$ (v) $E(3X-4)$, $\mathrm{Var}(3X-4)$ (vi) If $P\left[X\le C\right]>\frac{1}{2}$, find the minimum value of C. **(April/May 2015)**

Solution:

(i) We know that $\sum_i P(X=x_i)=1$

$$\Rightarrow \sum_{x=0}^{7}P(X=x)=1,\ \Rightarrow K+2K+2K+3K+K^2+2K^2+7K^2+K=1$$

$$\Rightarrow 10\mathrm{K}^2+9\mathrm{K}-1=0\ \Rightarrow K=\frac{1}{10}\ \ or\ \ K=-1$$ (here $K=-1$ is impossible, since $P(X=x)\ge 0$)

$$\therefore K=\frac{1}{10}$$

$\therefore$ The probability mass function is

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

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

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

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

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

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

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

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

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

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

(v) To find $E(3X-4)$, $\text{Var}(3X-4)$

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

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

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

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

$$=0\times P(X=0)+1\times P(X=1)+2\times P(X=2)+3\times P(X=3)+4\times P(X=4)$$
$$+5\times P(X=5)+6\times P(X=6)+7\times P(X=7)$$

$$=0+1\times K+2\times 2K+3\times 2K+4\times 3K+5\times K^2+6\times 2K^2+7\times(7K^2+K)$$

$$=K+4K+6K+12K+5K^2+12K^2+49K^2+7K \quad =30K+66K^2=\frac{30}{10}+\frac{66}{100} \quad =\frac{366}{100}$$

$$E(X^2)=\sum x^2P(X=x)$$

$$=0\times P(X=0)+1^2\times P(X=1)+2^2\times P(X=2)+3^2\times P(X=3)+4^2\times P(X=4)$$
$$+5^2\times P(X=5)+6^2\times P(X=6)+7^2\times P(X=7)$$

$$=0+1^2\times K+2^2\times 2K+3^2\times 2K+4^2\times 3K+5^2\times K^2+6^2\times 2K^2+7^2\times(7K^2+K)$$

$$=K+8K+18K+48K+25K^2+72K^2+343K^2+49K$$

$$=124K+440K^2=\frac{124}{10}+\frac{440}{100}=\frac{1240+440}{100}=\frac{1680}{100}=\frac{168}{10}=\frac{84}{5}$$

$(1) \Rightarrow E(3X-4) = \frac{3\times 366}{100} - 4 = \frac{1098-400}{100} = \frac{698}{100} = 69.8$

$(2) \Rightarrow Var(3X-4) = 9Var(X) = 9\left[E(X^2) - (E(X))^2\right] = 9[16.8 - 13.3956] = 30.6396$

(vi) To find the minimum value of 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

ii) **Trains arrive at a station at 15 minutes interval starting at 4 a.m. If a passenger arrive 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 for (i) less than 6 minutes (ii) more than 10 minutes. (May/June 2014)**

Solution:

Let X denotes number of minutes past 9.00 a.m. that the passenger arrives at the stop till 9.30a.m.

$X \sim U[0,30] \Rightarrow f(x) = \frac{1}{30},\ 0 < x < 30$

(i)**P**(**that he has to wait for the train for less than 6 minutes**)

$= P\left[(9 < x < 15) \cup (24 < x < 30)\right]$

$$= \int_9^{15} f(x)dx + \int_{24}^{30} f(x)dx = \int_9^{15} \frac{1}{30}dx + \int_{24}^{30} \frac{1}{30}dx = \frac{1}{30}\left\{[x]_9^{15} + [x]_{24}^{30}\right\} = \frac{12}{30} = 0.4$$

(ii)**P**(**that he has to wait for the train for** *more* **than** 10 **minutes**)

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

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

	iii)	**Find the moment generating function of a poisson distribution. Hence find mean and variance. (APR / MAY' 19)**

Solution:

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

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

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

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

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

$$M_X(t)=e^{\lambda\left(e^{t}-1\right)}$$

Mean = $E(X)$= $\left\{\frac{d}{dt}\left[M_X(t)\right]\right\}_{t=0}=\left\{\frac{d}{dt}\left[e^{\lambda\left(e^{t}-1\right)}\right]\right\}_{t=0}=\left[\left(\lambda e^{t}\right)e^{\lambda\left(e^{t}-1\right)}\right]_{t=0}=\lambda$

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

Where $E(X^2)=\left\{\frac{d}{dt}\left[M'_X(t)\right]\right\}_{t=0}=\left\{\frac{d}{dt}\left[\left(\lambda e^{t}\right)e^{\lambda\left(e^{t}-1\right)}\right]\right\}_{t=0}=\lambda\left[e^{t}.\left(\lambda e^{t}\right)e^{\lambda\left(e^{t}-1\right)}+e^{\lambda\left(e^{t}-1\right)}.e^{t}\right]_{t=0}$

$$=\lambda(\lambda+1)$$

Variance $=Var(X)=E(X^2)-\left(E(X)\right)^2=\lambda(\lambda+1)-\lambda^2=\lambda$.

2.	i)	**A continuous random variable X has the p.d.f $f(x)=kx^3e^{-x},\ x\geq 0$. Find the r[th] order moment of X about the origin. Hence find m.g.f, mean and variance of X. (Nov/Dec 2015)**

Solution:

Since $\int_0^{\infty}kx^3e^{-x}dx=1\Rightarrow k\left[x^3\left(\frac{e^{-x}}{-1}\right)-(3x^2)\left(\frac{e^{-x}}{1}\right)+(6x)\left(\frac{e^{-x}}{-1}\right)-(6)\left(\frac{e^{-x}}{1}\right)\right]_0^{\infty}=1$

$$k\left[-x^3e^{-x}-3x^2e^{-x}-6xe^{-x}-6\right]_0^{\infty}=1\Rightarrow k\left[(0)-(-6)\right]=1\Rightarrow 6k=1\Rightarrow k=\frac{1}{6}.$$

$$E(X^r)=\mu_r'=\int_0^{\infty}x^r f(x)dx=\frac{1}{6}\int_0^{\infty}x^r x^3e^{-x}dx=\frac{1}{6}\int_0^{\infty}x^{r+3}e^{-x}dx \qquad \because \Gamma n=\int_0^{\infty}e^{-x}x^{n-1}dx,\ n>0$$

here $n=r+4$ $\qquad =\frac{1}{6}\int_0^{\infty}e^{-x}x^{(r+3+1)-1}dx=\frac{1}{6}\Gamma(r+4)=\frac{(r+3)!}{6} \qquad \because \Gamma n=(n-1)!$

Putting $r=1$, $E(X)=\mu_1'=\frac{4!}{6}=\frac{24}{6}=4$

$$r=2, E(X^2)=\mu_2'=\frac{5!}{6}=\frac{120}{6}=20$$

$\therefore$ Mean $= E(X)=\mu_1'=4$; Variance $=E(X^2)-[E(X)]^2=\mu_2'-(\mu_1')^2$

$$\mu_2=20-(4)^2=20-16=4$$

To find M.G.F

$$M_X(t)=E(e^{tX})=\int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$M_X(t)=\int_{-\infty}^{\infty} e^{tx}\frac{1}{6}x^3e^{-x}dx$$

$$=\frac{1}{6}\int_0^{\infty} x^3e^{tx-x}dx=\frac{1}{6}\int_0^{\infty} x^3e^{-(1-t)x}dx$$

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

$$=\frac{1}{6}\left[-x^3\frac{e^{-(1-t)x}}{(1-t)}-3x^2\frac{e^{-(1-t)x}}{(1-t)^2}-6x\frac{e^{-(1-t)x}}{(1-t)^3}-6\frac{e^{-(1-t)x}}{(1-t)^4}\right]_0^{\infty}$$

$$=\frac{1}{6}\left[(0)-\left(\frac{-6}{(1-t)^4}\right)\right]$$

$$\therefore M_X(t)=\frac{1}{(1-t)^4}$$

ii) **a) A machine manufacturing bolts is known to produce 5% defective. In a random sample of 15 bolts, what is the probability that there are (1). Exactly 3 defective bolts, and (2). Not more than 3 defective bolts?** **(NOV/DEC 2018)**

Solution:

Given $n=15,\ p=0.05,\ q=0.95$

By Binomial Distribution, $P(X=x)=nC_x\, p^x\, q^{n-x}$

$$=15C_x\,(0.05)^x\,(0.95)^{15-x}$$

(1) $P(\text{Exactly 3 defective bolts})=P(X=3)=15C_3\,(0.05)^3\,(0.95)^{15-3}=0.0307$

(2) $P(\text{Not more than 3 defective bolts})=P(X\le 3)=P(X=0)+P(X=1)+P(X=2)+P(X=3)$

$=15C_0\,(0.05)^0\,(0.95)^{15-0}+15C_1\,(0.05)^1\,(0.95)^{15-1}+15C_2\,(0.05)^2\,(0.95)^{15-2}+15C_3\,(0.05)^3\,(0.95)^{15-3}$

$=0.994$

b) Six coins are tossed 6400 times. Using the Poisson distribution, what is the approximate probability of getting six heads 10 times?

Solution:

Probability of getting six heads in one toss of six coins is $p=\left(\frac{1}{2}\right)^6$ $\lambda=np=6400\times\left(\frac{1}{2}\right)^6=100$

		Let X be the number of times getting 6 heads $P(X=10)=\dfrac{e^{-100}(100)^{10}}{10!}=1.025\times10^{-30}$
3.	**i)**	**A random variable *X* has the probability mass function $f(x)=\dfrac{1}{2^x}$, x= 1,2,3,… Find its (i) M.G.F (ii) Mean (iii) Variance.**
		Solution: M.G.F= $M_X(t)=E(e^{tX})$ $=\sum_{x=1}^{\infty}e^{tx}f(x)=\sum_{x=1}^{\infty}e^{tx}\dfrac{1}{2^x}=\sum_{x=1}^{\infty}\left(\dfrac{e^t}{2}\right)^x$ $=\dfrac{e^t}{2}+\left(\dfrac{e^t}{2}\right)^2+\left(\dfrac{e^t}{2}\right)^3+\left(\dfrac{e^t}{2}\right)^4+\ldots$ $=\dfrac{e^t}{2}\left[1+\left(\dfrac{e^t}{2}\right)+\left(\dfrac{e^t}{2}\right)^2+\left(\dfrac{e^t}{2}\right)^3+\ldots\right]$ $=\dfrac{e^t}{2}\left[1-\dfrac{e^t}{2}\right]^{-1}$ M.G.F= $M_X(t)=\dfrac{e^t}{2-e^t}$ ……………… (1) Mean = *E(X)*= $\left\{\dfrac{d}{dt}[M_X(t)]\right\}_{t=0}=\left\{\dfrac{d}{dt}\left[\dfrac{e^t}{2-e^t}\right]\right\}_{t=0}=\left[\dfrac{(2-e^t)e^t-e^t(-e^t)}{(2-e^t)^2}\right]_{t=0}=2$ Variance = $Var(X)=E(X^2)-(E(X))^2$ Where $E(X^2)=\left\{\dfrac{d}{dt}\left[M'_X(t)\right]\right\}_{t=0}=\left\{\dfrac{d}{dt}\left[\dfrac{2e^t}{(2-e^t)^2}\right]\right\}_{t=0}=\left[\dfrac{(2-e^t)^2e^t-e^t2(2-e^t)(-e^t)}{(2-e^t)^4}\right]_{t=0}=6$ Variance = $Var(X)=E(X^2)-(E(X))^2$ = 6 - 4 =2
	ii)	**A component has an exponential time to failure distribution with mean of 10,000 hours. (i) The component has already been in operation for its mean life. What is the Probability that it will fail by 15,000 hours? (ii) At 15,000 hours the component is still in operation. What is the probability that it will operate for another 5000 hours? (Nov/Dec 2015)**
		Solution: Let X be the random variable denoting the time to failure of the component following exponential distribution with Mean $=10000$ hours. $\therefore \dfrac{1}{\lambda}=10,000\Rightarrow\lambda=\dfrac{1}{10,000}$ The p.d.f. of X is $f(x)=\begin{cases}\dfrac{1}{10,000}e^{-\frac{x}{10,000}} & ,x\ge0\\ 0 & ,otherwise\end{cases}$ (i) Probability that the component will fail by 15,000 hours given that it has already been in operation for its mean life $=P[X<15,000/X>10,000]$

		$$=\frac{P[10,000<X<15,000]}{P[X>10,000]} \text{ ---------(1)}$$ $$P[10,000<X<15,000] = \int_{10,000}^{15,000}\frac{1}{10000}e^{-\frac{x}{10000}}dx$$ $$=\frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{10000}^{15000} = -\left[e^{-\frac{x}{10000}}\right]_{10000}^{15000}$$ $$=-\left[e^{-\frac{15000}{10000}}-e^{-\frac{10000}{10000}}\right]=-\left[e^{-\frac{3}{2}}-e^{-1}\right]=e^{-1}-e^{-1.5} \text{ ----(2)}$$ $$P[X>10,000]=\int_{10,000}^{\infty}\frac{1}{10000}e^{-\frac{x}{10000}}dx=\frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{10000}^{\infty}=-\left[e^{-\frac{x}{10000}}\right]_{10000}^{\infty}$$ $$=-\left[e^{-\infty}-e^{-1}\right]=e^{-1} \text{ -------(3)}$$ Sub (2) & (3) in (1) $$(1)\Rightarrow\ P[X<15,000/X>10,000]=\frac{e^{-1}-e^{-1.5}}{e^{-1}}=\frac{0.3679-0.2231}{0.3679}=0.3936.$$ (ii) Probability that the component will operate for another 5000 hours given that it is in operation 15,000 hours $=P[X>20,000/X>15,000]$ $$=P[X>5000] \quad \text{[By memoryless property]}$$ $$=\int_{5000}^{\infty}f(x)dx=\int_{5000}^{\infty}\frac{1}{10000}e^{-\frac{x}{10000}}dx$$ $$=\frac{1}{10000}\left[\frac{e^{-\frac{x}{10000}}}{\frac{-1}{10000}}\right]_{5000}^{\infty}=-\left[e^{-\frac{x}{10000}}\right]_{5000}^{\infty}=e^{-0.5}=0.6065$$
	iii)	**An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours. (April/May 2018)**
		Solution: Given $X \simeq N(\mu,\sigma)$ *where* $\mu=800,\ \sigma=40,\ z=\frac{x-\mu}{\sigma}=\frac{x-800}{40}$ The standard normal z values corresponding to $x_1=778,\ x_2=834$ are $$z_1=\frac{778-800}{40}=-0.55 \text{ and } z_2=\frac{834-800}{40}=0.85$$ $$P(778<X<834)=P(-0.55<Z<0.85)$$ $$=P(0<Z<-0.55)+P(0<Z<0.85)$$

$$= P(0 < Z < 0.55) + P(0 < Z < 0.85)$$

$$= 0.2088 + 0.3023 = 0.5111$$

Hence the probability that a bulb burns between 778 and 834 hours is 0.5111

4. **i)**

If the density function of a continuous random variable X is given by $f(x) = \begin{cases} ax, & 0 \le x \le 1 \\ a, & 1 \le x \le 2 \\ 3a - ax, & 2 \le x \le 3 \\ 0, & elsewhere \end{cases}$

(i) Find the value of a (ii) Find the c.d.f of X (iii) Find ***P(X*** ≤ 1.5***).***

Solution:

(i) Since $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_{-\infty}^{\infty} f(x)dx = 1$

$$\Rightarrow \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$$

$$\int_{0}^{1} ax\ dx + \int_{1}^{2} a\ dx + \int_{2}^{3} (3a - ax)dx = 1$$

$$a\left[\frac{x^2}{2}\right]_0^1 + a\left[x\right]_1^2 + a\left[3x - \frac{x^2}{2}\right]_2^3 = 1 \Rightarrow a = \frac{1}{2}$$

(ii) 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 $2 \le x \le 3$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{1} \frac{x}{2}\,dx + \int_{1}^{2} \frac{1}{2}dx + \int_{2}^{x} \left(\frac{3}{2} - \frac{x}{2}\right)dx$$

$$= \left[\frac{x^2}{4}\right]_0^1 + \left[\frac{x}{2}\right]_1^2 + \left(\frac{3x}{2} - \frac{x^2}{4}\right)_2^x = \frac{3x}{2} - \frac{x^2}{4} - \frac{5}{4}$$

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{2}, & 0 \le x \le 1 \\ \frac{x}{2} - \frac{1}{4}, & 1 \le x \le 2 \\ \frac{3x}{2} - \frac{x^2}{4} - \frac{5}{4}, & 2 \le x \le 3 \\ 1, & x > 3 \end{cases}$$

		$(iii) P(X \le 1.5) = F(1.5) \qquad [F(x) = P(X \le x)]$ $= \frac{1.5}{2} - \frac{1}{4} = 0.5$
	ii)	**A given lot of product contains 2% defective products. Each product is tested before delivery. The probability that the product is good given that is actually good is 0.95 and the probability that the product is defective given that it is actually defective is 0.94. If a tested product is defective, what is the probability that it is actually defective? (NOV/DEC 2018)**
		Solution: Let A be an event that defective products and let B be an event that product is good. Let D be an event that defective product is tested after delivery. $P(A) = 0.02,\ P(B) = 0.98,\ P(\bar{D}/B) = 0.95,\ P(D/A) = 0.94,\ P(D/B) = 0.05$ $P(A/D) = \frac{P(D/\mathrm{A})\ P(\mathrm{A})}{P(D/\mathrm{A})\ P(\mathrm{A}) + P(D/\mathrm{B})\ P(\mathrm{B})}$ $= \frac{(0.94)(0.02)}{(0.94)(0.02)+(0.05)(0.98)} = \frac{0.0188}{0.0678} = 0.27729$
5.	**i)**	**The average percentage of marks of candidates in an examination is 42 with a standard deviation of 10. If the minimum mark for pass is 50% and 1000 candidates appear for the examination, how many candidates can be expected to get the pass mark? If it is required, that double the number of the candidates should pass, what should be the minimum mark for pass? (Nov/Dec 2015)**
		Solution: Let X denote the marks of the candidates, then $X \sim N(42, 10^2)$ Let $z = \frac{X-42}{10}$, $P[X \ge 50] = P[z \ge 0.8] = 0.5 - P[0 < z < 0.8] = 0.5 - 0.2881 = 0.2119$ If 1000 students write the test, $1000 P[X \ge 50] \cong 212$ students would pass the examination. If double that number should pass, then the no of passes should be 424. We have to find z_1, such that $P[z \ge z_1] = 0.424$ $\therefore P[0 < z < z_1] = 0.5 - 0.424 = 0.076$ From tables, z_1= 0.19 , $\therefore z_1 = \frac{50 - x_1}{10} \Rightarrow x_1 = 50 - 10 z_1 = 50 - 1.9 = 48.1$ The pass mark should be 48 nearly.
	ii)	**Derive the MGF, mean and variance of Geometric distribution and also state and prove the special property of it. (May/June 16)**
		Solution: $P(X = x) = pq^{x-1}, x = 1, 2, 3, \cdots$ **Moment Generating Function** $M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} p(x) = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = p[e^t + e^{2t} q + e^{3t} q^2 + \cdots]$ $= pe^t[1 + qe^t + (qe^t)^2 + \cdots] = pe^t[1 - qe^t]^{-1} = \frac{pe^t}{1 - qe^t}$

Mean and Variance

$$\mu_1' = M_X'(0) = \left[\frac{d}{dt}\left(\frac{pe^t}{(1-qe^t)}\right)\right]_{t=0} = \left[\left(\frac{pe^t}{(1-qe^t)^2}\right)\right]_{t=0} = \frac{1}{p}$$

$$\mu_2' = M_X''(0) = \left[\frac{d^2}{dt^2}\left(\frac{pe^t}{(1-qe^t)}\right)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{pe^t}{(1-qe^t)^2}\right)\right]_{t=0} = \frac{1+q}{p^2}$$

$$Mean = \mu_1' = \frac{1}{p}$$

$$Variance = \mu_2' - (\mu_1')^2 = \frac{1+q}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}$$

Memoryless property of geometric distribution.

Statement:

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

Proof:

The p.m.f of the geometric random variable X is $P(X = x) = q^{x-1}p$, $x = 1,2,3,....$

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

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

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

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

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

iii) Find mean, variance and MGF of exponential distribution. Also prove the lack of memory property of the Exponential distribution. (APR / MAY' 19)

Solution:

We know that $f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, otherwise \end{cases}$

$$M_X(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx$$

$$= \lambda \int_0^{\infty} e^{-x(\lambda - t)} dx$$

$$= \lambda \left[\frac{e^{-x(\lambda-t)}}{-(\lambda - t)}\right]_0^{\infty} = \frac{\lambda}{\lambda - t}$$

$$\text{Mean} = \mu_1' = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = \left[\frac{\lambda}{(\lambda-t)^2}\right]_{t=0} = \frac{1}{\lambda}$$

$$\mu_2' = \left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0} = \left[\frac{\lambda(2)}{(\lambda-t)^3}\right]_{t=0} = \frac{2}{\lambda^2}$$

$$\text{Variance} = \mu_2' - \left(\mu_1'\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

MEMORYLESS PROPERTY

Statement: If X is exponentially distributed with parameters λ, then for any two positive integers 's' and 't', $P[X > s+t / X > s] = P[X > t] \quad \forall s,t > 0$

Proof:

$$\text{The p.d.f of X is } f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0 \quad , Otherwise \end{cases}$$

$$\therefore P[X > k] = \int_k^\infty \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_k^\infty = e^{-\lambda k}$$

$$\therefore P[X > s+t / x > s] = \frac{P[x > s+t \cap x > s]}{P[x > s]}$$

$$= \frac{P[X > s+t]}{P[X > s]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P[X > t]$$

6.	i)	**Let X be a Uniformly distributed R.V over [-5, 5]. Determine (May/June2016)** $(1)\, P(X \leq 2)$ $(2)\, P(\lvert X\rvert > 2)$ (3) *Cumulative distribution function of* X $(4) Var(X)$.

Solution:

The R.V X ~ U[-5,5].

The p.d.f

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

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

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

$$(2)\, P(\lvert X\rvert > 2) = 1 - P(\lvert X\rvert \leq 2) = 1 - P(-2 \leq X \leq 2)$$

$$\therefore P(-2 \leq X \leq 2) = \int_{-2}^{2} f(x)dx = \int_{-2}^{2} \frac{1}{10}dx = \frac{1}{10}\int_{-2}^{2} dx = \frac{1}{10}[x]_{-2}^{2} = \frac{1}{10}[2+2] = \frac{4}{10}$$

		(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** $\mathbf{P(X = x)} = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x-1}$, $x = 1, 2, 3, \cdots$ **be the probability mass function of the R.V. X. Compute** $(1) P(X > 4)$ $(2) P(X > 4 / X > 2)$ $(3)\ E(X)$ $(4) Var(X)$. **(May/June2016)**
		Solution: $(1) P(X > 4) = P(X = 5) + P(X = 6) + P(X = 7) + \cdots$ $= \sum_{X=5}^{\infty} P(X = x) = \sum_{X=5}^{\infty} \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x-1}$ $= \left(\frac{3}{4}\right)\left[\left(\frac{1}{4}\right)^4 + \left(\frac{1}{4}\right)^5 + \left(\frac{1}{4}\right)^6 + \cdots\right]$ $= \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4\left[1 + \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)^2 + \cdots\right]$ $= \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4\left[1 - \frac{1}{4}\right]^{-1} = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^{-1} = \left(\frac{1}{4}\right)^4$

$$(2)P(X>4/X>2)=P(X>2)$$
$$=P(X=3)+P(X=4)+P(X=5)+\cdots$$
$$=\sum_{X=3}^{\infty}P(X=x)=\sum_{X=3}^{\infty}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x-1}$$
$$=\left(\frac{3}{4}\right)\left[\left(\frac{1}{4}\right)^2+\left(\frac{1}{4}\right)^3+\left(\frac{1}{4}\right)^4+\cdots\right]$$
$$=\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left[1+\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)^2+\cdots\right]$$
$$=\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left[1-\frac{1}{4}\right]^{-1}=\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^{-1}=\left(\frac{1}{4}\right)^2$$

$$(3)\ E(X)=\frac{1}{p}=\frac{1}{\left(\frac{3}{4}\right)}=\frac{4}{3}$$

$$(4)\ Var(X)=\frac{q}{p^2}=\frac{\left(\frac{1}{4}\right)}{\left(\frac{3}{4}\right)^2}=\frac{1}{4}\times\frac{16}{9}=\frac{4}{9}.$$

iii) **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.**

(1) 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?

(2) 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?

(3) A ball is drawn from one of the boxes is green. What is the prob. that it came from B_2?

(4) A ball is drawn from one of the boxes is white what is the prob. that it came from B_2?

Solution:

Let B_1 & B_2 be the events that the boxes B_1 & B_2 respectively are selected. $P(B_1)=\frac{1}{2};\quad P(B_2)=\frac{1}{2}$

1) Let A be the event that a red ball is selected. $P(A/B_1)=\frac{2}{3};\quad P(A/B_2)=\frac{1}{3}$

P(ball is from B_1, given it is red) $=P(B_1/A)=\dfrac{P(B_1)P(A/B_1)}{P(B_1)P(A/B_1)+P(B_2)P(A/B_2)}$

$$=\frac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{1}{3}\right)}=\frac{2}{3}$$

2) Let C be the event that a red ball and a green ball are selected.

$$P(C/B_1)=\frac{2_{C_1}\times 1_{C_1}}{3_{C2}}=\frac{2}{3};\qquad P(C/B_2)=\frac{1_{C_1}\times 2_{C_1}}{3_{C2}}=\frac{2}{3}$$

P(B_1 was chosen given a red ball and a green ball were selected) $= \boldsymbol{P(B_1/C)}$

$$= \frac{P(B_1)P(C/B_1)}{P(B_1)P(C/B_1)+P(B_2)P(C/B_2)}$$

$$= \frac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{2}{3}\right)} = \frac{1}{3}$$

3) Let D be the event that a green ball is selected. $P(D/B_1) = \frac{1}{3}; \quad P(D/B_2) = \frac{2}{3}$

P(ball is from B_1, given it is green) $= \boldsymbol{P(B_2/D)} = \frac{P(B_2)P(D/B_2)}{P(B_1)P(D/B_1)+P(B_2)P(D/B_2)}$

$$= \frac{\frac{1}{2}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{1}{3}\right)+\left(\frac{1}{2}\times\frac{2}{3}\right)} = \frac{2}{3}$$

4) Let E be the event that a white ball is selected.
The given two boxes does not contained a white ball, hence the probability is 0

iv) **Four boxes A, B, C, D contain fuses. The boxes contain 5000,3000,2000 and 1000 fuses respectively. The percentages of fuses in boxes which are defective are 3%, 2%, 1% and 0.5% respectively. One fuse is selected at random arbitrarily from one of the boxes. It is found to be defective fuse. Find the probability that it has come from box D. (APR / MAY' 19)**

Solution:

Let E_1, E_2, E_3, E_4 be the event that boxes A,B,C,D respectively are selected.

$P(E_1) = P(E_2) = P(E_3) = P(E_4) = \frac{1}{4}$

Let D be the event that defective fuse is selected.

Given $P(D/E_1) = 0.03, P(D/E_2) = 0.02, P(D/E_3) = 0.01, P(D/E_4) = 0.005,$

By Bayes theorem,

$$P(E_4/D) = \frac{P(E_4)P(D/E_4)}{P(E_1)P(D/E_1)+P(E_2)P(D/E_2)+P(E_3)P(D/E_3)+P(E_4)P(D/E_4)}$$

$$= \frac{\left(\frac{1}{4}\right)(.005)}{\left(\frac{1}{4}\right)(.03)+\left(\frac{1}{4}\right)(.02)+\left(\frac{1}{4}\right)(.01)+\left(\frac{1}{4}\right)(.005)} = \frac{0.005}{0.065} = 0.07692$$

UNIT – II TWO DIMENSIONAL RANDOM VARIABLES

PART – A

1. **Given the joint probability density function of X and Y as** $f(x,y) = \begin{cases} \frac{1}{6}, 0<x<2, 0<y<3 \\ 0, otherwise \end{cases}$ **, determine the marginal density functions.**

The marginal function of X is $f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy \;=\int_0^3 \frac{1}{6}dy=\left[\frac{y}{6}\right]_0^3=\frac{1}{2},\; 0<x<2$

The marginal function of Y is $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx \;=\int_0^2 \frac{1}{6}dx=\left[\frac{x}{6}\right]_0^2=\frac{1}{3},\; 0<y<3$

2\. **Find the value of k, if the joint density function of (X , Y) is given by**

$$f(x,y)=\begin{cases} k(1-x)(1-y), 0<x<4, 1<y<5 \\ 0 \qquad , otherwise \end{cases}$$

Given the joint pdf of (X , Y) is f(x , y) = k (1 – x) (1 – y), 0 < x < 4, 1 < y < 5

$$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1 \Rightarrow \int_1^5\int_0^4 k(1-x)(1-y)dxdy=1$$

$$\Rightarrow k\int_1^5\left[x-\frac{x^2}{2}-yx+y\frac{x^2}{2}\right]_0^4 dy=1$$

$$\Rightarrow k\int_1^5(-4+4y)dy=1 \Rightarrow k\left[-4y+4\frac{y^2}{2}\right]_1^5=1 \Rightarrow k(30+2)=1 \Rightarrow 32k=1 \Rightarrow k=\frac{1}{32}$$

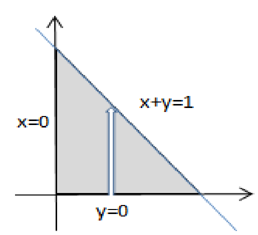
3\. **The joint probability density function of bivariate random variable (X , Y) is given by** $f(x,y)=\begin{cases} 4xy, 0<x<1, 0<y<1 \\ 0 \quad , elsewhere \end{cases}$. **Find P (X + Y<1)**

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

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

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

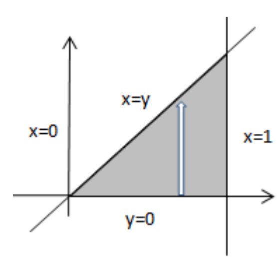
$$=2\int_0^1(x-2x^2+x^3)dx=2\left[\frac{x^2}{2}-2\frac{x^3}{3}+\frac{x^4}{4}\right]_0^1=2\left[\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right]=\frac{1}{6}$$



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

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

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



5. **The regression equations of X on Y and Yon X are respectively 5x – y = 22 and 64x – 45y = 24. Find the mean values of X and Y.**

Regression lines pass through the mean values of X and Y. Solving the two equations we get the mean values.

Let $5x - y = 22$ --------------(1)

$64x - 45y = 24$ --------------(2)

Multiply equation (1) by 45 and subtract equation (2)

$$\begin{array}{rl} 225x - 45y &= 990 \\ -\ 64x + 45y &= -\ 24 \\ \hline 161x &= 966 \Rightarrow x = 6 \end{array}$$

Substitute in equation 1

$5(6) - y = 22 \Rightarrow y = 8$. $\therefore$ mean value of X = 6 and mean value of Y = 8

6. **The joint p.d.f. of R.V. (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 $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_y^1 \frac{1}{x}dx = [\log x]_y^1 = \log 1 - \log y = -\log y,\ 0 < y < 1$

7. **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

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

8. **Let X and Y be two independent R.Vs with Var(X) = 9 and Var(Y) = 3. Find Var(4X – 2Y + 6)**

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

9. **The joint pdf of a two dimensional random variable (X,Y) is given by** $f(x,y) = kxe^{-y}, 0 \le x \le 2, y > 0$. **Find the value of k.**

Given that f(x,y) is pdf of (X,Y)

$\therefore f(x,y) \ge 0$, for all x ,y

	$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \Rightarrow \int_0^{\infty}\int_0^2 kxe^{-y}dxdy = 1 \Rightarrow k\int_0^2 xdx.\int_0^{\infty} e^{-y}dy = 1 \Rightarrow k\left[\frac{x^2}{2}\right]_0^2.\left[-e^{-y}\right]_0^{\infty} = 1$ $\Rightarrow k(2)(1) = 1 \Rightarrow k = \frac{1}{2}$
10.	**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)** The joint pdf is $f(x,y) = \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2}{\partial x \partial y}\left(1-e^{-x}\right)\left(1-e^{-y}\right) = \frac{\partial}{\partial x}\left(1-e^{-x}\right).e^{-y} = e^{-x}.e^{-y} = e^{-(x+y)}, x>0, y>0$ $P(1<X<2, 1<Y<2) = \int_1^2\int_1^2 f(x,y)dxdy = \int_1^2\int_1^2 e^{-(x+y)}dxdy = \int_1^2\int_1^2 e^{-x}.e^{-y}dxdy$ $= \int_1^2 e^{-x}dx.\int_1^2 e^{-y}dy = \left[-e^{-x}\right]_1^2.\left[-e^{-y}\right]_1^2 = \left(e^{-1}-e^{-2}\right)^2 = \left(\frac{1}{e}-\frac{1}{e^2}\right)^2 = \left(\frac{e-1}{e^2}\right)^2 = 0.054$
11.	**The lines of regression in a bivariate distribution are X + 9Y = 7 and Y + 4X =** $\boldsymbol{\frac{49}{3}}$ **. Find the coefficient of correlation.** x + 9y – 7 = 0 ----------- (1) y + 4x – $\boldsymbol{\frac{49}{3}}$ **= 0** ----------- (2) Let (1) be the regression line of Y on X and let (2) be the regression line of X on Y. $\therefore y = -\frac{1}{9}x + \frac{7}{9} \Rightarrow b_1 = -\frac{1}{9}$ $x = -\frac{1}{4}y + \frac{49}{12} \Rightarrow b_2 = -\frac{1}{4}$ $\therefore r = \pm\sqrt{b_1 b_2} = \sqrt{-\frac{1}{9}.-\frac{1}{4}} = \sqrt{\frac{1}{36}} = \frac{1}{6} < 1$ Since both regression coefficients are negative, correlation coefficient is negative.
12.	**If Y = -2X + 3, find Cov (X , Y).** Cov(X,Y) = E(XY) – E(X) E(Y) = E(X(-2X + 3)) – E(X){E(-2X + 3)} = [E(-2X^2 + 3X) - E(X)]{-2E(X) + 3} = -2E(X^2) + 3 E(X) + 2 (E(X))^2 - 3E(X) = 2(E(X))^2 – 2 E(X^2) = -2 var(X)
13.	**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** $P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right)$ $P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right) = \int_{x=-\infty}^{\frac{1}{2}}\int_{y=\frac{1}{2}}^{\infty} f(x,y)dydx = \int_{x=0}^{\frac{1}{2}}\int_{y=\frac{1}{2}}^{1} \frac{3}{2}\left(x^2+y^2\right)dydx = \frac{3}{2}\int_0^{\frac{1}{2}}\left[x^2y+\frac{y^3}{3}\right]_{\frac{1}{2}}^1 dx$ $= \frac{3}{2}\int_0^{\frac{1}{2}}\left[x^2\left(1-\frac{1}{2}\right)+\frac{1}{3}\left(1-\frac{1}{8}\right)\right]dx = \frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{x^2}{2}+\frac{7}{24}\right)dx = \frac{3}{2}\left[\frac{x^3}{6}+\frac{7x}{24}\right]_0^{\frac{1}{2}}$ $= \frac{3}{2}\left[\frac{1}{6}.\frac{1}{8}+\frac{7}{24}.\frac{1}{2}\right] = \frac{3}{2}\left[\frac{8}{48}\right] = \frac{1}{4}$

14. **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.**

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

$p(x,y) \geq 0$, for all x ,y

$$\sum_m \sum_n p(x,y) = 1 \Rightarrow 36c = 1 \Rightarrow c = \frac{1}{36}$$

15. **The joint probability mass function of X and Y is**

X\Y	0	1	2
0	0.1	0.04	0.02
1	0.08	0.2	0.06
2	0.06	0.14	0.3

Check if X and Y are independent.

X\Y	0	1	2	$p_X(x)$
0	0.1	0.04	0.02	0.16
1	0.08	0.2	0.06	0.34
2	0.06	0.14	0.3	0.5
$p_Y(y)$	0.24	0.38	0.38	1

$p_X(0) \cdot p_Y(0) = (0.16)(0.24) \neq 0.1 = p(0,0)$ $\therefore$ X and Y are not independent.

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.**

Given $r_{XY} = -\frac{1}{4}$, $\sigma_X^2 = 3, \sigma_Y^2 = 5$ $r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$, $\sigma_X \neq 0, \sigma_Y \neq 0$

$Cov(X, Y) = r_{XY}\, \sigma_X \sigma_Y = -\frac{1}{4}\sqrt{3} \cdot \sqrt{5} = -0.968$

17. **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[XY²]**

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[XY2] = E[X] E[Y2]

$\because \sigma_Y^2 = E[Y^2] - [E[Y]]^2 \Rightarrow 5 = E[Y^2] - 4 \Rightarrow E[Y^2] = 9 \therefore E[XY^2] = 4(9) = 36$

18. **Given the joint density function of X and Y as f(x , y) = $\begin{cases} xe^{-y}, 0 < x < 2, y > 0 \\ 0 \quad , elsewhere \end{cases}$. Find the range space for the transformation X + Y.**

Let the auxillary random variable be V = Y. The transformation functions are u = x + y, v = y, $y > 0 \Rightarrow v > 0$ and $0 < x < 2 \Rightarrow 0 < u - v < 2 \Rightarrow v < u < v + 2$

19. **The joint probability mass function of the discrete random variable (X , Y) is given by the table**

Y \ X	2	4
1	1/10	1.5/10
3	2/10	3/10
5	1/10	1.5/10

Find the conditional probability P (X = 2 / Y = 3)

Y \ X	2	4	$P_Y(y)$
1	1/10	1.5/10	2.5/10
3	2/10	3/10	5/10
5	1/10	1.5/10	2.5/10
$P_X(x)$	4/10	6/10	1

P (X = 2 / Y = 3) = $\dfrac{P(X=2,Y=3)}{P_Y(3)} = \dfrac{2/10}{5/10} = \dfrac{2}{5}$

20. **The two lines of regression are 4x – 5y + 33 = 0 and 20x – 9y = 107. Calculate the coefficient of correlation between X and Y.**

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.

$\therefore y = \frac{4}{5}x + \frac{33}{5} \Rightarrow b_1 = \frac{4}{5}$

$x = \frac{9}{20}y + \frac{107}{20} \Rightarrow b_2 = \frac{9}{20} \quad \therefore r = \sqrt{b_1 b_2} = \sqrt{\frac{4}{5}\cdot\frac{9}{20}} = \sqrt{\frac{9}{25}} = \frac{3}{5} = 0.6 < 1$

PART – B

1. **(i) 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 PDFs of X and Y (b) cov (X,Y)**

Solution:

The marginal PDF of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{\infty} 25e^{-y}dy = 25\left[-e^{-y}\right]_0^{\infty} = 25\left[-e^{-\infty} + e^0\right] = 25(1) = 25,\ 0 < x < 0.2$$

The marginal PDF of Y is $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{0.2} 25e^{-y}dx = 25e^{-y}\left[x\right]_0^{0.2} = 25e^{-y}\left[0.2\right] = 5e^{-y},\ 0 < y < \infty$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{0.2} x(25)dx = 25\left[\frac{x^2}{2}\right]_0^{0.2} = 25\left[\frac{0.04}{2}\right] = 0.5$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^{\infty} y(5e^{-y})dx = 5\left[-ye^{-y} - e^{-y}\right]_0^{\infty} = 5\left[0+1\right] = 5$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_0^{\infty}\int_0^{0.2} xy(25e^{-y})dxdy = \int_0^{\infty} 25ye^{-y}dy.\int_0^{0.2} xdx$$

$$= 25\left[-ye^{-y} - e^{-y}\right]_0^{\infty}.\left[\frac{x^2}{2}\right]_0^{0.2} = 25\left[0+1\right].\left[\frac{0.04}{2}\right] = (25)(0.02) = 0.5$$

Cov(x,y) = E(XY) – E(X)E(Y) = 0.5 – (0.5)(5) = –2

(ii). Find the constant k such that $f(x,y) = \begin{cases} k(x+1)e^{-y}, 0<x<1, y>0 \\ 0 \quad ,otherwise \end{cases}$ **is a joint p.d.f. of the continuous random variable (X,Y). Are X and Y independent R.Vs? Explain.**

Solution:

To find k : Given that *f*(x,y) is pdf of (X,Y)

$$\therefore f(x,y) \geq 0 \text{ , for all x ,y and } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \Rightarrow \int_0^{\infty}\int_0^{1} k(x+1)e^{-y}dxdy = 1$$

$$\Rightarrow k\int_0^1 (x+1)dx.\int_0^{\infty} e^{-y}dy = 1$$

$$\Rightarrow k\left[\frac{x^2}{2}+x\right]_0^1.\left[-e^{-y}\right]_0^{\infty} = 1$$

$$\Rightarrow k\left(\frac{3}{2}\right)(1) = 1 \Rightarrow k = \frac{2}{3}$$

$$\therefore f(x,y) = \begin{cases} \frac{2}{3}(x+1)e^{-y}, 0<x<1, y>0 \\ 0 \quad ,otherwise \end{cases}$$

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

2. (i). Let the joint p.d.f. of R.V. (X,Y) be given as $f(x,y) = \begin{cases} 4xy, 0\leq x\leq 1, 0\leq y\leq 1 \\ 0 \quad ,elsewhere \end{cases}$ **, find the marginal densities of X and Y and the conditional densities of X given Y = y. (April/May 2018)**

Solution:

The marginal density function of X is

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

The marginal density function of Y is

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

The conditional densities function of X given Y = y is

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

(ii). The joint density function of two random variable X and Y is given by

$$f(x,y) = \begin{cases} \frac{6}{7}\left(x^2+\frac{xy}{2}\right), 0\le x\le 1,\ 0\le y\le 2 \\ 0 \qquad , elsewhere \end{cases}.$$

(a) Compute the marginal p.d.f of X and Y ? (b) Find E(X) & E(Y) (c) $P\left(X<\frac{1}{2}, Y>\frac{1}{2}\right)$

Solution:

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

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 \frac{6}{7}\left(x^2+\frac{xy}{2}\right)dx = \frac{6}{7}\left[\frac{x^3}{3}+\frac{x^2}{2}\frac{y}{2}\right]_0^1 = \frac{6}{7}\left[\frac{1}{3}+\frac{y}{4}\right], 0\le y\le 2$$

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^1 x\frac{6}{7}\left[2x^2+x\right]dx = \frac{6}{7}\left[2x^3+x^2\right]_0^1 = \frac{6}{7}\left[\frac{2x^4}{4}+\frac{x^2}{2}\right]_0^1 = \frac{6}{7}\left[\frac{1}{2}+\frac{1}{2}\right] = \frac{6}{7}$$

$$E(Y) = \int_{-\infty}^{\infty} yf_Y(y)dy = \int_0^2 y\frac{6}{7}\left[\frac{1}{3}+\frac{y}{4}\right]dy = \frac{6}{7}\left[\frac{y}{3}+\frac{y^2}{4}\right]_0^2 = \frac{6}{7}\left[\frac{y^2}{6}+\frac{y^3}{12}\right]_0^2 = \frac{6}{7}\left[\frac{2}{3}+\frac{2}{3}\right] = \frac{8}{7}$$

$$P(X<\frac{1}{2}, Y>\frac{1}{2}) = \int_{-\infty}^{\frac{1}{2}}\int_{\frac{1}{2}}^{\infty} f(x,y)\,dy\,dx = \int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{2}\frac{6}{7}\left(x^2+\frac{xy}{2}\right)dy\,dx = \int_0^{\frac{1}{2}}\frac{6}{7}\left[x^2y+\frac{x}{2}\frac{y^2}{2}\right]_{\frac{1}{2}}^{2}dx$$

$$= \int_0^{\frac{1}{2}}\frac{6}{7}\left[2x^2+x-\frac{x^2}{2}-\frac{x}{16}\right]dx = \int_0^{\frac{1}{2}}\frac{6}{7}\left[\frac{3x^2}{2}+\frac{15x}{16}\right]_0^{\frac{1}{2}}dx = \frac{6}{7}\left[\frac{x^3}{2}+\frac{15x^2}{32}\right]_0^{\frac{1}{2}}$$

$$= \frac{6}{7}\left[\frac{1}{16}+\frac{15}{128}\right] = \frac{6}{7}.\frac{23}{128} = \frac{69}{448}$$

3\. **(i). If X,Y and Z are uncorrelated random variables with zero means and standard deviations 5 , 12 and 9 respectively and if U = X + Y, V = Y + Z, find the correlation coefficient between U and V.**

	Solution: Given E(X) = E(Y) = E(Z) = 0 $Var(X) = E(X^2) - (E(X))^2 = 25 \quad \therefore E(X^2) = 25$ $Var(Y) = E(Y^2) - (E(Y))^2 = 144 \quad \therefore E(Y^2) = 144$ $Var(Z) = E(Z^2) - (E(Z))^2 = 81 \quad \therefore E(Z^2) = 81$ Also given that X, Y and Z are uncorrelated $\therefore r_{XY} = 0$, i.e., $E(XY) - E(X)\,.\,E(Y) = 0 \quad \Rightarrow E(XY) = 0$ $\therefore r_{YZ} = 0$, i.e., $E(YZ) - E(Y)\,.\,E(Z) = 0 \quad \Rightarrow E(YZ) = 0$ $\therefore r_{XZ} = 0$, i.e., $E(XZ) - E(X)\,.\,E(Z) = 0 \quad \Rightarrow E(XZ) = 0$ Now, $E(U) = E(X+Y) = E(X) + E(Y) = 0$ and $E(V) = E(Y+Z) = E(Y) + E(Z) = 0$ $E(U^2) = E((X+Y)^2) = E(X^2 + Y^2 + 2XY) = E(X^2) + E(Y^2) + 2E(XY) = 25 + 144 + 0 = 169$ $E(V^2) = E((Y+Z)^2) = E(Y^2 + Z^2 + 2YZ) = E(Y^2) + E(Z^2) + 2E(YZ) = 144 + 81 + 0 = 225$ $\sigma_U^2 = E(U^2) - (E(U))^2 = 169$ and $\sigma_V^2 = E(V^2) - (E(V))^2 = 225$ $E(UV) = E\{(X+Y)(Y+Z)\} = E(XY) + E(XZ) + E(Y^2) + E(YZ) = 0 + 0 + 144 + 0 = 144$ $r_{UV} = \frac{E(UV) - E(U)E(V)}{\sigma_U . \sigma_V} = \frac{144}{(13)(15)} = \frac{48}{65}$
	(ii). The joint pdf of the continuous R.V (X,Y) is given as $f(x,y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0 \\ 0 & ,elsewhere \end{cases}$ **. Find the pdf of the random variable** $U = \frac{X}{Y}$**.**
	Solution: The transformation functions are $u = \frac{x}{y}$ *and* $v = y$ Solving for x , we get $u = \frac{x}{v} \Rightarrow x = uv$ The Jacobian of the transformation is J = $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$ The joint density of U and V is $f_{UV}(u,v) = \lvert J \rvert\, f_{XY}(x,y) = \lvert v \rvert e^{-(x+y)} = v\,e^{-v(u+1)}$ The range space of (U , V) is obtained from the range space of (X , Y) and the transformations x = uv, y = v $\therefore x > 0$ **and** $y > 0$ **we have** $uv > 0$ **and** $v > 0$ $\Rightarrow u > 0$ **and** $v > 0$ $f_{UV}(u,v) = \begin{cases} v\,e^{-v(u+1)}, & u > 0, v > 0 \\ 0\ , & elsewhere, \end{cases}$ The pdf of U is the marginal density function of U,

	$f_U(u) = \int_{-\infty}^{\infty} f(u,v)dv = \int_0^{\infty} v\,e^{-v(u+1)}\,dv = \left[v.\frac{e^{-v(u+1)}}{-(u+1)} - 1.\frac{e^{-v(u+1)}}{(u+1)^2} \right]_0^{\infty} = 0 + \frac{1}{(u+1)^2} = \frac{1}{(u+1)^2}\ , u > 0$
	(iii). The life time of a certain brand of an electric bulb may be considered as a random variable with mean 1200h and standard deviation 250h. Find the probability, using central limit theorem, that the average lifetime of 60 bulbs exceeds 1250hours. (NOV/DEC 2018)
	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 $N\left(\mu, \frac{\sigma^2}{n}\right)$. Let $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $P(\bar{X} > 1250) = P(Z > 1.55) = 0.0606$
4.	**(i). Obtain the equation of the regression line Y on X from the following data.**

X	3	5	6	8	9	11
Y	**2**	**3**	**4**	**6**	**5**	**8**

Solution:

X	Y	U = X – 6	V = Y – 6	U^2	V^2	UV
3	2	–3	–4	9	16	12
5	3	–1	–3	1	9	3
6	4	0	–2	0	4	0
8	6	2	0	4	0	0
9	5	3	–1	9	1	–3
11	8	5	2	25	4	10
		6	–8	48	34	22

$n = 6, \sum U = 6, \sum V = -8, \sum U^2 = 48, \sum V^2 = 34, \sum UV = 22\ \bar{U} = \frac{\sum U}{n} = \frac{6}{6} = 1,$

$\bar{V} = \frac{\sum V}{n} = \frac{-8}{6} = -1.33,\ \sigma_U^2 = \frac{\sum U^2}{n} - (\bar{U})^2 = \frac{48}{6} - 1 = 7,\ \sigma_U = 2.646,$

$\sigma_V^2 = \frac{\sum V^2}{n} - (\bar{V})^2 = \frac{34}{6} - (-1.33)^2 = 3.898,\ \sigma_V = 1.974$

Cov(U , V) = $\frac{\sum UV}{n} - \bar{U}\,\bar{V} = \frac{22}{6} - (1)(-1.33) = 4.997$

$\therefore r_{UV} = \frac{Cov(U,V)}{\sigma_U.\sigma_v} = \frac{4.997}{(2.646)(3.898)} = 0.484$

$\therefore r_{XY} = 0.484$

$\overline{X} = \overline{U} + 6 \Rightarrow \overline{X} = 1 + 6 = 7$

$\overline{Y} = \overline{V} + 6 \Rightarrow \overline{Y} = -1.33 + 6 = 4.67$

$\sigma_X = \sigma_U \Rightarrow \sigma_X = 2.646$

$\sigma_Y = \sigma_V \Rightarrow \sigma_Y = 3.898$

The regression line of Y on X is

$$Y - \overline{Y} = \frac{r\,\sigma_Y}{\sigma_X}\left(X - \overline{X}\right)$$

$$\Rightarrow Y - 4.67 = \frac{(0.484)(3.898)}{(2.646)}\left(X - 7\right)$$

$\Rightarrow Y = 0.713X - 0.321$

(ii) X and Y are two random variables having the joint probability mass function $f(x, y) = k(3x + 5y);\ x = 1,2,3 : y = 0,1,2$. Find the marginal distribution and conditional distribution of X, P(X = x_i /Y = 2), P(X ≤2 / Y ≤ 1). **(April/May 2019)**

Solution:

Y \ X	1	2	3	$P_Y(y)$
0	3k	6k	9k	18k
1	8k	11k	14k	33k
2	13k	16k	19k	48k
$P_X(x)$	24k	33k	42k	99k

To find k:

We know that Total probability = 1

ΣΣP(x,y) = 1 ⇒ 99k = 1

⇒ k = 1/99

The marginal distribution of X is

X	1	2	3
p(x)	24/99	33/99	42/99

The marginal distribution of Y is

Y	0	1	2
p(y)	18/99	33/99	48/99

Conditional distribution of X given Y = 2

P(X = x_i /Y = 2)

X	1	2	3
P(X = x_i / Y = 2) = [P(X = x_i ,Y = 2)]/ P(Y = 2)	13/48	16/48	19/48

P(X ≤ 2 / Y ≤ 1) = P(X=1, Y= 0) + P(X=1, Y= 1) + P(X=2, Y= 0)+ P(X=2, Y= 1)

= 3k + 8k + 6k + 11k

= 28k = 28/99

5. (i) In a partially destroyed laboratory record only the lines of regressions and variance of X are available. The regression equations are 8x – 10y + 66 = 0 and 40x – 18y = 214 and variance of X = 9. Find (a) the correlation coefficient between X and Y (b) Mean values of X and Y (c) variance of Y.

Solution:

Given 8x – 10y = –66 ……(1)

40x – 18y = 214 ……(2)

Let (1) be the regression line of y on x and (2) be the regression line of x on y.

$\therefore 10y = 8x + 66 \Rightarrow y = \frac{8x}{10} + \frac{66}{10}$ $\therefore$ the regression coefficient of y on x is $b_1 = \frac{8}{10} = \frac{4}{5}$

$\therefore 40x = 18y + 214 \Rightarrow x = \frac{18y}{40} + \frac{214}{40}$ $\therefore$ the regression coefficient of x on y is $b_2 = \frac{18}{40} = \frac{9}{20}$

$\therefore b_1 b_2 = \left(\frac{4}{5}\right)\left(\frac{9}{20}\right) = \frac{9}{25} < 1$

Let r be the correlation between x and y.

$\therefore r = \sqrt{b_1 b_2} = \sqrt{\frac{9}{25}} = \frac{3}{5} = 0.6$ [Since both regression coefficients are positive, r is positive]

Let $(\bar{x}, \bar{y})$ be the point of intersection of the two regression lines.

Solving (1) and (2) we get $\bar{x}, \bar{y}$

5 x (1) $\Rightarrow$ 40x – 50y = – 330

40x – 18y = 214

Subtracting – 32y = – 544

$\therefore$ y = 17

Now, 8x – 10y = – 66 $\Rightarrow$ 8x – 10(17) = – 66 $\Rightarrow$ 8x = 170 – 66 $\Rightarrow$ 8x = 104 $\Rightarrow$ x = 13

$\therefore (\bar{x}, \bar{y})$ = (13 , 17) is the mean of X and Y.

We know, $\frac{\sigma_Y^2}{\sigma_X^2} = \frac{b_1}{b_2} \Rightarrow \sigma_Y^2 = \frac{b_1}{b_2}\sigma_X^2 \Rightarrow \sigma_Y^2 = \frac{\frac{4}{5}}{\frac{9}{20}}.(9) \Rightarrow \sigma_Y^2 = 16 \therefore$ Variance of Y is 16

(ii) The joint probability density function of a two dimensional random variable (X,Y) is given by $f(x,y) = xy^2 + \frac{x^2}{8}, 0 \le x \le 2, 0 \le y \le 1$. Compute (i) P(X > 1), (ii) P(Y < ½), (iii) P(X < Y) (iv) 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

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

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

	$P\left(Y<\frac{1}{2}\right)=\int_{-\infty}^{1/2} f_Y(y)\,dy=\int_0^{1/2}\left(2y^2+\frac{1}{3}\right)dy=\left[\frac{2y^3}{3}+\frac{1}{3}\right]_0^{1/2}$ $=\frac{2}{3}\cdot\frac{1}{8}+\frac{1}{3}\cdot\frac{1}{2}=\frac{3}{12}=\frac{1}{4}$ $P(X<Y)=\int_0^1\int_0^y\left(xy^2+\frac{x^2}{8}\right)dx\,dy=\int_0^1\left(\frac{x^2y^2}{2}+\frac{x^3}{24}\right)_0^y dy$ $=\int_0^1\left(\frac{y^4}{2}+\frac{y^3}{24}\right)dy=\left[\frac{y^5}{10}+\frac{y^4}{96}\right]_0^1=\frac{1}{10}+\frac{1}{96}=\frac{53}{480}$ 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.
	UNIT III - TESTING OF HYPOTHESIS
	PART – A
1.	**Define Population, Sample and Sample Size.** 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.** **(APRIL / MAY '15)** 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.	**List out the applications of t –distribution.** **(NOV / DEC '13)** ❖ 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 significance of the mean of a random sample. ❖ To test the significance of an observed correlation coefficient.
4.	**Mention the Properties of t – distribution.** ❖ 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.
5.	**What is Standard Error?** **(APRIL / MAY '11) (APRIL / MAY '17)** The standard deviation of the sampling distribution of a statistic is known as its Standard error.
6.	**State any two properties of χ^2 distribution.** (i)Chi – square curve is always positively skewed (ii)Chi – square values increase with the increase in degrees of freedom
7.	**Explain the various uses of Chi-square test.** **(APRIL / MAY '14) (NOV / DEC '14)** Test of goodness of fit, Test of independence of attributes, Test of Homogeneity for a specified value of standard deviation
8.	**What are the assumptions on which F-test is based?**

	The F test is based on the following assumptions: a. Normality : The values in each group should be normally distributed. b. Independence of error : The variations of each value around its own group mean. i.e. error should be independent of each value. c. Homogeneity : The variances within each group should be equal for all groups.
9.	**Define Sampling of distribution.** The probability distribution of a sample statistic is called the sampling distribution.
10.	**Define Level of Significance. (APRIL / MAY '12) (NOV / DEC '13)** The probability that the value of the statistic lies in the critical region is called the level of significance. (i.e., the level of significance is denoted by $\alpha = 1\%$ or 5%.
11.	**Define one - tailed and two - tailed test.** 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. $\mathbf{H_o : \mu = \mu_o \text{ Vs } H_1 : \mu > \mu_o \text{ (or) } H_1 : \mu < \mu_o}$ **(right tailed) (Left tailed)** A test of statistical hypothesis whose alternative hypothesis is two tailed, such as $\mathbf{H_o : \mu = \mu_o \text{ Vs } H_1 : \mu \neq \mu_o}$ is known as two tailed test.
12.	**What do you mean by critical region and acceptance region?(APRIL / MAY '12) (NOV / DEC '12)** 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 of the sample space s which amounts to the acceptance of null hypothesis is called acceptance region.
13.	**Define Type-I and Type-II errors. (APRIL / MAY '14) (MAY / JUNE '16) (APRIL / MAY '18) (APRIL / MAY '19)** 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: Reject 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.
14.	**What are the uses t – test?** ❖ To test whether significance of the difference of the mean of random sample and the population mean. ❖ To test whether significance difference between two sample means. ❖ To test the significance of an observed sample correlation coefficient.
15.	**What are the uses F – test?** ❖ 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.
16.	**State the assumptions of Chi-square test.** ❖ 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.

17.	**Write 95% confidence interval of the population mean.** $$\bar{x}-t_{0.05}\frac{S}{\sqrt{n-1}}\le\mu\le\bar{x}+t_{0.05}\frac{S}{\sqrt{n-1}}$$
18.	**Write the 95% confidence interval of population proportion.** 95% Confidence interval for P is $p-1.96\sqrt{\frac{pq}{n}}\le P\le p+1.96\sqrt{\frac{pq}{n}}$
19.	**Write down the values of χ^2 for a 2×2 contingency table with cell frequencies a,b,c,d.** **(APRIL / MAY '15) (APRIL / MAY '17)** Expected frequency table: <table><tr><td>$\frac{(a+b)(a+c)}{N}$</td><td>$\frac{(a+b)(b+d)}{N}$</td></tr><tr><td>$\frac{(a+c)(c+d)}{N}$</td><td>$\frac{(c+d)(b+d)}{N}$</td></tr></table>
20.	**Find the standard error of sample mean from the following data.** $n=4,\ \mu=18.5,\ \bar{x}=17.85,\ S=1.955$. Standard error $\frac{S}{\sqrt{n-1}}=\frac{1.955}{\sqrt{14-1}}=0.54$
	PART B
1(i)	**In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400 , the mean is 15. Could the samples have been drawn from the same population with S.D. 4 ? (APR/ MAY '17)**
	Solution: $H_0:\bar{x}_1=\bar{x}_2$ $H_1:\bar{x}_1\ne\bar{x}_2$ **Level of Significance:** 1% **Test Statistic:** $$Z=\frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{20-15}{4\sqrt{\frac{1}{500}+\frac{1}{400}}}=18.6$$ **Table value :** $Z_\alpha=2.58$ **Conclusion :** The calculated value is greater than the table value, hence we reject the null hypothesis.
(ii)	**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. (MAY / JUNE '16)**
	Solution: **Hypothesis:** $H_0:P_1=P_2$ $H_1:P_1>P_2$ **Level of Significance :** $\alpha=0.05$

	Test Statistic : $Z=\dfrac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ The Sample proportion, $p_1=\dfrac{800}{1000}=0.80,\quad p_2=\dfrac{800}{1200}=0.67,\quad P=\dfrac{n_1p_1+n_2p_2}{n_1+n_2}=0.7273\ \ \&\ Q=1-P=0.2727$ $Z=\dfrac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=6.9905$ **Table value :** $Z_\alpha=1.645$ **Conclusion :** The calculated value is greater than the table value, hence we reject the null hypothesis.
2(i)	**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. (APRIL / MAY '15)**

Solution:

Hypothesis:

$H_0:\mu=100$

$H_1:\mu\neq 100$

Level of Significance : $\alpha=0.05$

Test Statistic : $t=\dfrac{\bar{x}-\mu}{S/\sqrt{n-1}}$, where $s^2=\dfrac{\sum x^2}{n}-\left(\dfrac{\sum x}{n}\right)^2$

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	$96312=\sum X^2$

$s^2=\dfrac{\sum x^2}{n}-\left(\dfrac{\sum x}{n}\right)^2=\dfrac{96312}{10}-\left(\dfrac{972}{10}\right)^2$

$\Rightarrow 9631.2-9447.84=183.36\Rightarrow S=13.54$

$t=\dfrac{\bar{x}-\mu}{S/\sqrt{n-1}}$, where $s^2=\dfrac{\sum x^2}{n}-\left(\dfrac{\sum x}{n}\right)^2$

$t=\dfrac{97.2-100}{13.54/\sqrt{10-1}}=\dfrac{2.8}{4.5133}=0.6204$

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 $\bar{x}$ $\quad |t| \le t_{0.05,9} = 2.262$ $\Rightarrow \left\| \frac{\bar{x}-\mu}{S/\sqrt{n-1}} \right\| \le 2.262 \Rightarrow \left\| \frac{\bar{x}-100}{13.54/\sqrt{10-1}} \right\| \le 2.262$ $\Rightarrow \left\| \frac{\bar{x}-100}{4.5133} \right\| \le 2.262$ $\Rightarrow -2.262 \le \frac{\bar{x}-100}{4.5133} \le 2.262$ $\Rightarrow -10.2091 \le \bar{x}-100 \le 10.2091$ $\Rightarrow 89.7909 \le \bar{x} \le 110.2091$
(ii)	**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 .

(iii) **Two random samples are drawn from normal populations are given below :**

Sample 1	**17**	**27**	**18**	**25**	**27**	**29**	**13**	**17**

Sample 2	16	16	20	27	26	25	21	

Can we conclude that the two samples are drawn from the same population? Test at 5% level of significance. (APRIL / MAY '19)

Solution:

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.

F test

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 1		Sample 2	
X	X^2	Y	Y^2
17	289	16	256
27	729	16	256
18	324	20	400
25	625	27	729
27	729	26	676
29	841	25	625
13	169	21	441
17	289		
$\Sigma X = 173$	$\Sigma X^2 = 3995$	$\Sigma Y = 151$	$\Sigma Y^2 = 3383$

$$n_1 = 8, \sum X = 173, \bar{X} = \frac{\sum X}{n_1} = \frac{173}{8} = 21.6$$

$$n_2 = 7, \sum Y = 151, \bar{Y} = \frac{\sum Y}{n_2} = \frac{151}{7} = 21.57$$

$$s_1^2 = \frac{\sum (X)^2}{n_1 - 1} - \left\{\frac{\sum (X)}{n_1 - 1}\right\}^2 = 36.27$$

$$s_2^2 = \frac{\sum (Y)^2}{n_2 - 1} - \left\{\frac{\sum (Y)}{n_2 - 1}\right\}^2 = 20.95$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{(8)(36.27)}{7} = 41.45$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{(7)(20.95)}{6} = 24.44$$

$$\therefore \quad F = \frac{S_1^2}{S_2^2} = \frac{41.45}{24.44} = 1.696$$

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 $F_{0.05} = 4.21$

$\therefore F < F_{0.05}$

Since the calculated value of F < the table value of F, H_0 is accepted at 5% level of significance.

t test

Null Hypothesis H₀: $\mu_1 = \mu_2$

Alternative Hypothesis H₁: $\mu_1 \neq \mu_2$

The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$where \ s = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(36.27) + 7(20.95)}{8+7-2} = 33.6$$

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{21.6 - 21.57}{33.6\left(\sqrt{\frac{1}{8} + \frac{1}{7}}\right)} = \frac{0.03}{(33.6)(0.518)} = 0.001724$$

$v = n_1 + n_2 - 2 = 13$ degrees of freedom at 5% level of significance

$t_{.05}(13) = 2.16$

$t_{.05}(13) > |t|$, **H₀** is accepted

Conclusion: The two samples are drawn from populations with same variances.

3(i) **The following data represent the biological values of protein from cow's milk and buffalo's milk:**

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

Examine whether the average values of protein in the two samples significantly differ at 5% level.

Solution:

n=6

$$\overline{x_1} = \frac{1}{6} \times 10.68 = 1.78, \ s_1^2 = \frac{1}{6} \times 19.16 - (1.78)^2 = 0.0261$$

$$\overline{x_2} = \frac{1}{6} \times 11.79 = 1.965; \quad s_2^2 = \frac{1}{6} \times 23.25 - (1.965)^2 = 0.0154$$

$$H_0 : \overline{x_1} = \overline{x_2} \quad and \quad H_1 : \overline{x_1} \neq \overline{x_2}$$

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

	$t=\dfrac{1.78-1.965}{\sqrt{\dfrac{0.0261+0.0154}{5}}}=-2.03 \quad and \quad v=10$ Two tailed test is to be used. LOS is 5% 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.
(ii)	**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 '17)** <table><tr><th></th><th>Smokers</th><th>Non Smokers</th></tr><tr><td>Literates</td><td>83</td><td>57</td></tr><tr><td>Illiterates</td><td>45</td><td>68</td></tr></table>
	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 $\chi^2=\dfrac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$ $\chi^2=\dfrac{[(83*68)-(45*57)]^2(253)}{(83+45)(57+68)(83+57)(45+68)}=9.47$ Conclusion: Since χ^2 =9.47 > 3.841, so we reject H_0 at 5% level of significance
4(i)	**A manufacturer of electric bulbs, by some process, find the standard deviation of the lamps to be 100hrs.He wants to change the process if the new process results in even smaller variation in the life of lamps. In adopting the new process a sample of 150 bulbs gave the standard deviation of 95 hrs. Is the manufacturer justified in changing the process.**
	Solution: Given $\alpha=0.05,\ \sigma=100,\ S=95,\ n=150$ The Parameter of interest is σ. Null Hypothesis H_0: σ =S Alternative Hypothesis H_1: $\sigma \neq$ S Level of significance: α =5% or 0.05 $z=\dfrac{s-\sigma}{\dfrac{\sigma}{\sqrt{2n}}}=\dfrac{95-100}{\dfrac{100}{\sqrt{300}}}=-0.866$ $i.e., \quad \lvert z\rvert=0.866<1.96$ Conclusion: The calculated value is less than the table value; hence we accept the null hypothesis. So the manufacture finds no justification in changing the process on this evidence alone.
(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 ? (MAY/JUNE '14) (APR/MAY '15)**

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 $\chi^2 = 4.726$

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.

(iii) **Two independent samples of eight and seven items respectively had the following values of the variable:** **(MAY / JUNE '16)**

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.

5(i) **Test significance of the difference between the means of the samples, drawn from two normal populations with the same SD using the following data: (APRIL / MAY '15) (NOV / DEC '13)**

	Size	Mean	Standard Deviation
Sample I	100	61	4
Sample II	200	63	6

Solution:

$H_o : \overline{x_1} = \overline{x_2} \text{ or } \mu_1 = \mu_2$

$H_1 : \overline{x_1} \neq \overline{x_2} \text{ or } \mu_1 \neq \mu_2$

Two tailed test is to be used.

The test statistic is $z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1{}^2}{n_2} + \frac{s_2{}^2}{n_1}}} = \frac{61 - 63}{\sqrt{\frac{4^2}{200} + \frac{6^2}{100}}} = -3.02$

Tabulated value is 1.96 at 5% level of significance.

$|z| > z_\alpha$ The difference between $\overline{x_1}$ and $\overline{x_2}$ is significant at 5% level of significance. i.e. H_o is rejected and H_1 is accepted. Therefore, the two normal populations, from which the samples are drawn, may not have the same mean though they may have the same S.D.

(ii) **A sample of size 13 gave and estimated population variance 3.0 while another sample of size 15 gave an estimate of 2.5. Could both samples be from population with the same variance? (APRIL/MAY '17)**

	Solution: **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) $n_1 = 13, \sigma_1{}^2 = 3.0\, and\ v_1 = 12$ $n_2 = 15, \sigma_2{}^2 = 2.5\, and\ v_2 = 14$ $F = \frac{\sigma_1{}^2}{\sigma_2{}^2} = \frac{3.0}{2.5} = 1.2$ $v_1 = 12\ and\ v_2 = 14\ \ F_{5\%} = 2.53$ from the F-table. Since $F < F_{5\%}$ $\therefore$ **H$_0$ is accepted** **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.
(iii)	**A die was thrown 498 times. Denoting x to be the number appearing on the top face of it, the observed frequency of x is given below:** x: 1, 2, 3, 4, 5, 6 f: 69, 78, 85, 82, 86, 98 **What opinion you would form for the accuracy of the die? (MAY / JUNE '16)**

x	1	2	3	4	5	6
f	69	78	85	82	86	98

Solntion:
Given n=6
Null Hypothesis H$_0$: There is no significant difference
Alternative Hypothesis H$_1$: There is a significant difference
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=498 /6=83**

Face	Observed Frequency(O)	Expected Frequency(E)	O-E	(O-E)^2	$\frac{(O-E)^2}{E}$
1	69	83	-14	196	2.36
2	78	83	-5	25	0.30
3	85	83	2	4	0.05
4	82	83	-1	1	0.01
5	86	83	3	9	0.11
6	98	83	15	225	2.71
				460	5.54

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

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 accepted at 5% level of significance. i.e., The die is accurate.

(iv) **Fit a Poisson's distribution to the following data and the goodness of fit. Test at 5% level of significance. (APRIL / MAY '19)**

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, $P(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

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

<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 IV – DESIGN OF EXPERIMENTS

PART-A

1. **Define Analysis of Variance.** **(APRIL / MAY '11) (MAY / JUNE '16)**

Analysis of Variance is a technique that will enable us to test for the Significance of the difference among more than two sample means.

2.	**What are the assumptions of analysis of variance?** **(i)** The sample observations are independent (ii) The Environmental effects are additive in nature (iii) Sample observation are coming from normal
3.	**What is the purpose of ANOVA?** We want to test whether more than two population mean we use analysis of Variance technique
4.	**Define one-way classification and two way classifications in ANOVA.** The entire experiment influences on only single factor is one way classification. The entire experiment influences on only two factors is two way Classification.
5.	**What are the basic principle of design of experiments? (APRIL / MAY '15) (APRIL / MAY '17)** (i) Randomization (ii) Replication (iii) Local Control
6.	**What is the aim of the design of the experiments? (NOV / DEC '13)** 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.
7.	**What is Latin Square Design? (APRIL / MAY '19)** It is square array of the letters A, B, C, D etc., of Latin square alphabets in this square array each letters 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.
8.	**Compare RBD, LSD, CRD**

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**

9. **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	**199**	**601**	**-**	**-**

A = 4 x 40 =160, D = 96 / 8 = 12 B = 3,
C = 115 x 3 = 435

10. **What ate the advantages of completely randomized block design?**

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

11. **Find the missing values 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	-	-

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 are the three essential steps to plan Design of experiment?**

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

13. **Define Replication.**

Replication is a repeating the experimental units

14. **Write down the ANOVA table for One way classification (APRIL / MAY '17)**

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

15. **Write down the ANOVA table for Randomized Block Design (APRIL / MAY '11)**

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

16. **Write down the ANOVA table for Latin Square Design.**

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

17. **Define Experimental Error.**

The estimation of the amount of variations due to each of the independent factors separately and then comparing these estimates due to assignable factors with the estimate due to chance factor is known as experimental error.

18. **Why a 2 x 2 Latin Square is not possible? (APRIL / MAY '15) (APRIL / MAY '14) (MAY / JUNE '16)**

Consider a n X n Latin square design, the 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, d.f of SSE = 0 and hence MSE is not defined.
Comparisons are not possible. Hence a 2 X 2 Latin Square Design is not possible.

19. **Mention the advantages of latin square design over RBD. (NOV / DEC '14)**

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.

20 **Write down the ANOVA table for 2^2 factorial designs.**

Source of Variation	Sum of Squares	Degree of freedom	Mean Square	F- Ratio
a	S_A	**1**	$MS_A = \frac{SS_A}{d.f}$	$if\ MS_A > SS_E\ then\ F_A = \frac{MS_A}{SS_E}$
b	S_B	**1**	$MS_B = \frac{SS_B}{d.f}$	$if\ MS_B > SS_E\ then\ F_B = \frac{MS_B}{SS_E}$
ab	S_{AB}	**1**	$MS_{AB} = \frac{SS_{AB}}{d.f}$	$if\ MS_{AB} > SS_E\ then\ F_{AB} = \frac{MS_{AB}}{SS_E}$
ERROR	S_E	**4(r-1)**	$MS_E = \frac{SS_E}{d.f}$	

PART B

1. **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**

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

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 ;

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

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

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

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

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

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 Rows	SSR=24.75	n - 1= 3	MSR=8.25	F_R= 12.31	$F_R(3, 6)$=4.76
Between Columns	SSC=2.75	n - 1= 3	MSC = 0.92	F_C = 1.37	$F_C(3, 6)$=4 .76
Between Letters	SSL = 4.25	n - 1= 3	MSL = 1.42	F_L = 2.12	$F_L(3, 6)$=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_L< Tab F_L and Cal F_R> Tab F_R $\Rightarrow$ There is significant difference between the rows, no significant difference between the letters and no significant difference between the columns. $\Rightarrow$ There is no significant difference between the different methods of cultivation.
2.	**Analyze 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

3. **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. **(MAY / JUNE '16)**

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

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

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 868 + 942 + 963 + 865 - 3481 = 157$$

$$SSC = \frac{\sum T_{*j}^{\ 2}}{h} - C.F = 841 + 900 + 930 + 812 - 3481 = 2$$

$$SSR = \frac{\sum T_{i*}^{\ 2}}{k} - C.F = 729 + 484 + 930.25 + 1482.25 - 3481 = 144.5$$

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 varieties	SSR=144.5	h - 1= 3	MSR= 48.17	F_R = 39.27	$F_{5\%}(3,9)$ = 3.86
Between blocks	SSC=2	k – 1=3	MSC = 0.67	F_C = 0.545	$F_{5\%}(3, 9)$ = 3.86
Residual	SSE = 10.5	(h – 1)(k – 1) = 9	MSE = 1.17		

Conclusion: Cal F_C<Tab F_C and Cal F_R> Tab F_R $\Rightarrow$Therefore null hypothesis is rejected. Hence four varieties are not similar. But the varieties are similar along block wise

4 **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. (APRIL / MAY '17)

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**

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

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

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

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(6)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 64 = 4$$

$$SSR = \frac{(\sum Y_1)^2}{N_1} + \frac{(\sum Y_2)^2}{N_2} + \frac{(\sum Y_3)^2}{N_2} + \frac{(\sum Y_4)^2}{N_2} - C.F = \frac{(8)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(8)^2}{4} - 64 = 2$$

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
Column Treatment	SSC=4	n-1=3	$MSC = \frac{SSC}{n-1}$ =1.33	$F_C = \frac{MSC}{MSE}$=7.52
Row Treatments	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 8.94

There is significant difference between treatments

5. **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:** **(APR / MAY '19)**

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.

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

	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	-
Total	-11	38	-10	17	271	542	142

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

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

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

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(-11)^2}{6} + \frac{(38)^2}{7} + \frac{(-10)^2}{4} - 17 = 234.44$$

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	$MSC = \frac{SSC}{C-1}$ =117.2 2	$F_C = \frac{MSC}{MSE}$ =2.332
Within Samples	SSE=703.56	N-C=17-3=14	$MSE = \frac{SSE}{N-C}$ =50.2 5	

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

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

UNIT– V: Statistical Quality Control

PART-A

1. Write down the objectives of statistical quality control.

To achieve better utilization of raw materials, to control waste and scrap and to optimize the quality of the product without any defects

2. Define control chart. **(MAY / JUNE '16) (APRIL / MAY '17)**

It is a useful graphical method to find whether a process is in statistical quality control.

3. What are the uses of Quality control chart?

It helps in determining whether the goal set is being achieved by finding out whether the Process is in control or not.

4. What are the different types of control chart? **(APRIL / MAY '17)**

Control chart for variable – Range and mean chart, Control chart for attributes- p-chart, C-chart, np-chart.

5. Write down the control limits for mean chart.

Central limit = $\bar{\bar{x}}$, upper control limit = $\bar{\bar{x}}$ +$A_2\bar{R}$, lower control limit = $\bar{\bar{x}}$ -$A_2\bar{R}$ where $\bar{x}$ is the mean of the sample and R is the range.

6. Write down the control limits for range chart.

CL= $\bar{R}$, UCL = $D_4\bar{R}$, LCL=$D_3\bar{R}$.

7. Define p-chart.

Control chart for fraction defectives is called p-chart.

8. Define C-chart.

Control chart for number of defects is called c-chart.

9. Write down the control limits for c-chart.

CL = $\bar{c}$ UCL = $\bar{c}$ +3 $\sqrt{\bar{c}}$ LCL = $\bar{c}$ -3 $\sqrt{\bar{c}}$

10. The total number of defects in 20 pieces is 220 .what is the UCL and LCL? **(MAY / JUNE '16)**

UCL = UCL = $\bar{c}$ +3 $\sqrt{\bar{c}}$ = 20.95 and LCL = $\bar{c}$ -3 $\sqrt{\bar{c}}$ = 1.05.

11.	**Write down the control limits for p-chart.** UCL= $\overline{np}+3\sqrt{\overline{npq}}$, LCL= $\overline{np}-3\sqrt{\overline{npq}}$ CL = $\overline{np}$
12.	**Define np –chart.** Control chart for number of defectives is called np chart.
13.	**What is two sided tolerance limits?** Two sided tolerance limits are values determined from a sample of size n so that one can claim with $(1-\alpha)$% confidence that atleast δ proportion of the population is included between these values.
14.	**What is the tolerance limit?** **(APR / MAY '19)** Tolerance Limits of a quality characteristic are defined as those values between which nearly all the manufactured items will lie. If the measurable quality characteristics X is assumed to be normally distributed with mean μ and S.D. σ, then the tolerance limits are usually taken as $\mu \pm 3\sigma$, since only 0.27% of all the items produced can be expected to fall outside these limits.
15.	**When the process is under control and if n=5, $\overline{\overline{X}}$ =1.1126 and $\overline{R}$ = 0.0054, find the tolerance limits?** The tolerance limits its are 1.056,1.196
16.	**Find the lower and upper control limits for $\overline{X}$ - chart and R- chart, when each sample is of size 4 and $\overline{\overline{X}}$ =10.80 and $\overline{R}$ =0.46.?** For $\overline{X}$ -chart, LCL=10.46, UCL=11.14: For R-chart, LCL=0, UCL=1.05.
17.	**Find the lower and upper control limits for $\overline{X}$ - chart and s-chart, if n=5 $\overline{\overline{X}}$ =15 and $\overline{s}$ =2.5?** For $\overline{X}$ -chart, LCL=11.43, UCL=18.57; For s-chart, LCL=0, UCL=5.22.
18.	**Find the lower and upper control limits for p- chart and np – chart, when n=100 and $\overline{P}$ = 0.085?** For p-chart, LCL =0.0013, UCL=0.1687; For np-chart, LCL=0.134, UCL=16.867
19.	**A garment was sampled on 10 consecutive hours of production. The number of defects found per garment is given below:** **(APR / MAY '19)** **Defects: 5,1,7,0,2,3,4,0,3,2. Compute upper and lower control limits for monitoring number of defects.** $\overline{c}$ =2.7 UCL = $\overline{c}$ +3 $\sqrt{\overline{c}}$ =7.6295 LCL = $\overline{c}$ -3 $\sqrt{\overline{c}}$ = -2.295
20.	**What are control limits for an R-chart in terms of σ values?** UCL= $D_2\sigma$; LCL = $D_1\sigma$
	PART B
1.	**The following data gives the average life in hours and range in hours of 12 samples each of 5 lamps. Construct $\overline{X}$ - hart and R- chart, comment on state of control.** **(APR / MAY '19)**

Sample No.	1	2	3	4	5	6	7	8	9	10	11	12
Mean $\overline{X}_i$	120	127	152	157	160	134	137	123	140	144	120	127
Range R_i	30	44	60	34	38	35	45	62	39	50	35	41

Solution:

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X}_i$$
$$= \frac{1}{12}\left[120+127+152+157+160+134+137+123+140+144+120+127\right] = 136.75$$

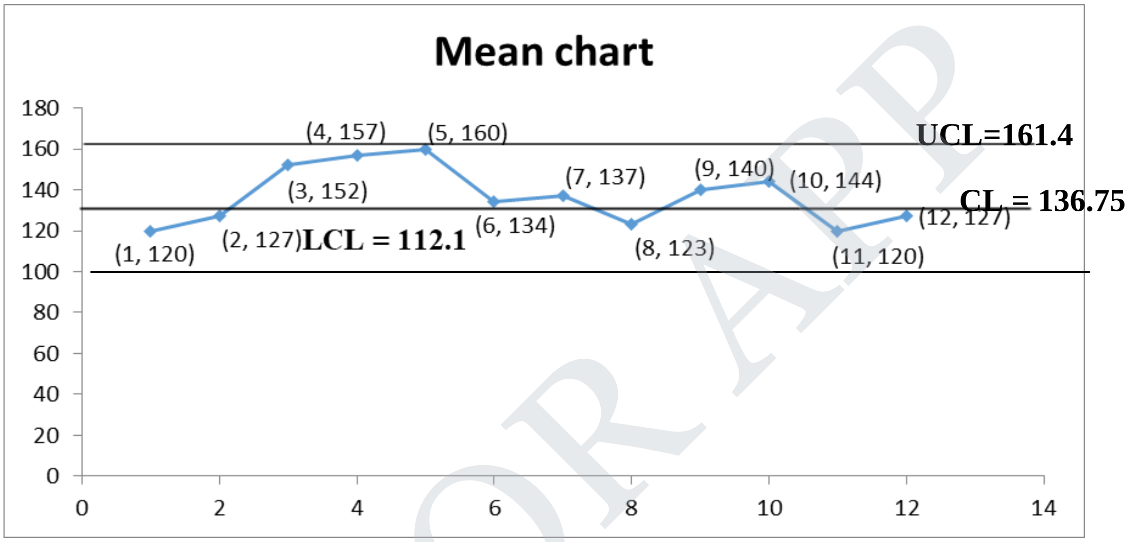
$$\overline{R} = \frac{1}{N}\sum R_i$$
$$= \frac{1}{12}\left[30+44+60+34+38+35+45+62+39+50+35+41\right] = 42.75$$

From the table of control chart for sample size n=5, we have $A_2 = 0.577, D_3 = 0 \,\&\, D_4 = 2.115$

i) Control limits for $\overline{X}$ chart:

CL (central line)= $\overline{\overline{X}} = 136.75$; $LCL = \overline{\overline{X}} - A_2\overline{R} = 136.75 - (0.5775)(42.75) = 112.1$

$UCL = \overline{\overline{X}} + A_2\overline{R} = 136.75 + (0.5775)(42.75) = 161.44$

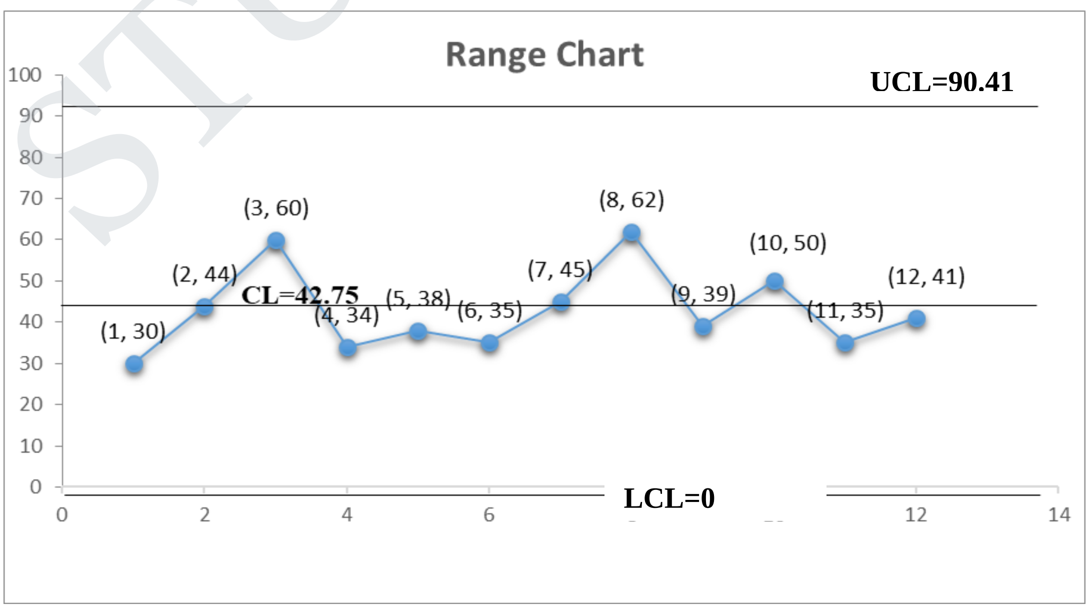


Conclusion:

Since all the sample points lie within the LCL and UCL lines, the process is under control according to $\overline{X}$ *chart*

ii) Control limits for R-Chart:

$$\mathrm{CL} = \overline{\mathrm{R}} = 42.75; \quad \mathrm{LCL} = \mathrm{D_3}\,\overline{\mathrm{R}} = 0; \quad \mathrm{UCL} = \mathrm{D_4}\overline{\mathrm{R}} = (2.115)(42.75) = 90.41$$



Conclusion :

Since all the sample range fall within the control limits the statistical process is under control according to $R\,chart$.

2. **The Values of sample mean $\overline{X}$ and sample standard deviation S for 15 samples, each of size 4, drawn from a production process are given below. Draw the appropriate control charts for the process average and process variability. Comment on the state of control.**

Sample No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Mean	15	10	12.5	13	12.5	13	13.5	11.5	13.5	13	14.5	9.5	12	10.5	11.5
S.D	3.1	2.4	3.6	2.3	5.2	5.4	6.2	4.3	3.4	4.1	3.9	5.1	4.7	3.3	3.3

Solution:

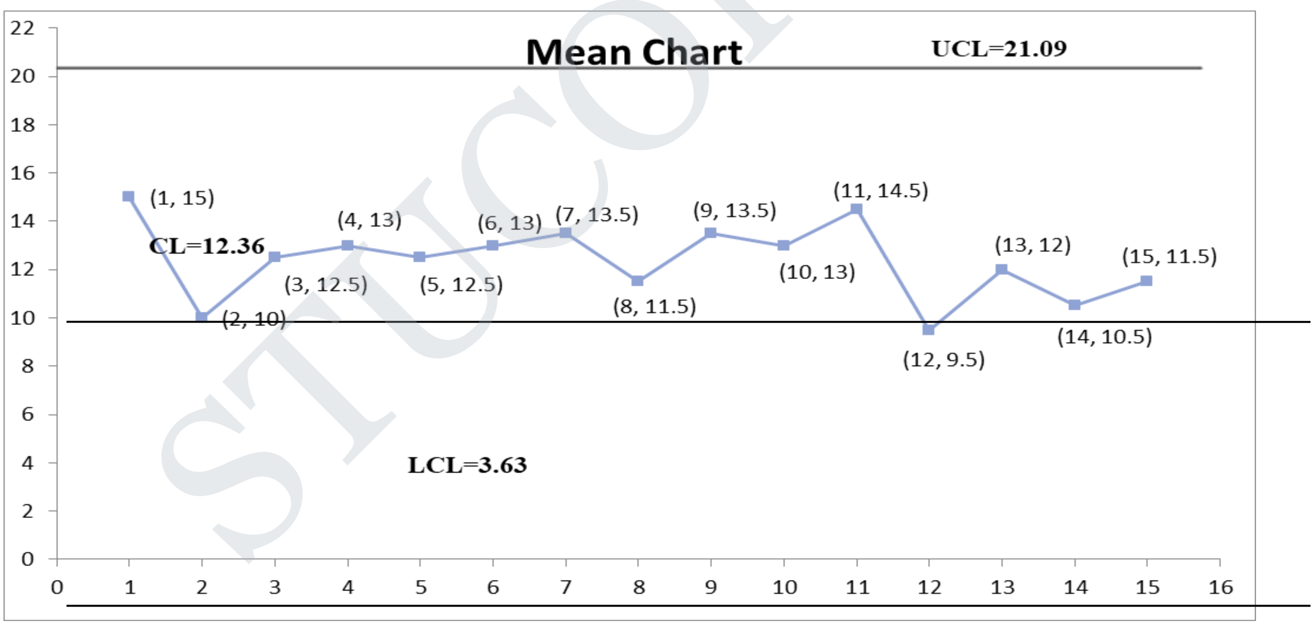
$$\overline{\overline{X}} = \frac{\sum \overline{X}_i}{N} = \frac{185.5}{15} = 12.36; \qquad \overline{s} = \frac{\sum s_i}{N} = \frac{60.3}{15} = 4.02$$

i) Control limits for $\overline{X}$ chart:

From the table of control chart constants, for sample size n = 4, we have $A_1 = 1.880,\ B_3 = 0\ and\ B_4 = 2.266$

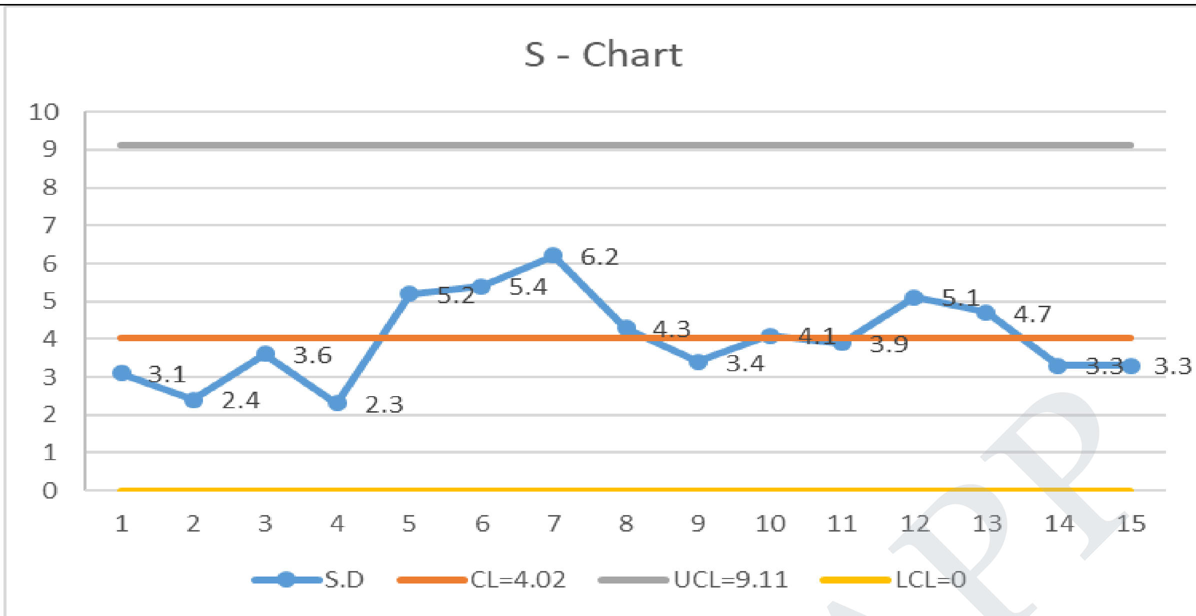
$$CL = \overline{\overline{X}} = 12.36; \qquad LCL = \overline{\overline{X}} - A_1\sqrt{\frac{n}{n-1}}\,\overline{s} = 12.36 - 1.880\left(\sqrt{\frac{4}{3}}\right)(4.02.) = 3.63$$

$$UCL = \overline{\overline{X}} + A_1\sqrt{\frac{n}{n-1}}\,\overline{s} = 12.36 + 1.880\left(\sqrt{\frac{4}{3}}\right)(4.02.) = 21.09$$



ii) Control limits for S-Chart:

$$CL = \overline{s} = 4.02; \quad LCL = B_3\overline{s} = 0; \quad UCL = B_4\overline{s} = (2.266)(4.02) = 9.11$$



Conclusion:

Even before drawing the control chart, we observe that the given sample mean values lie between 3.63 and 21.09 and that the given S.D values fall within 0 and 9.11. Hence the process is under control with respect to average and variability.

3a) **15 tape recorders were examined for quality control test. The number of defects in each tape recorder is recorded below. Draw the appropriate control chart and comment on the state of control.**

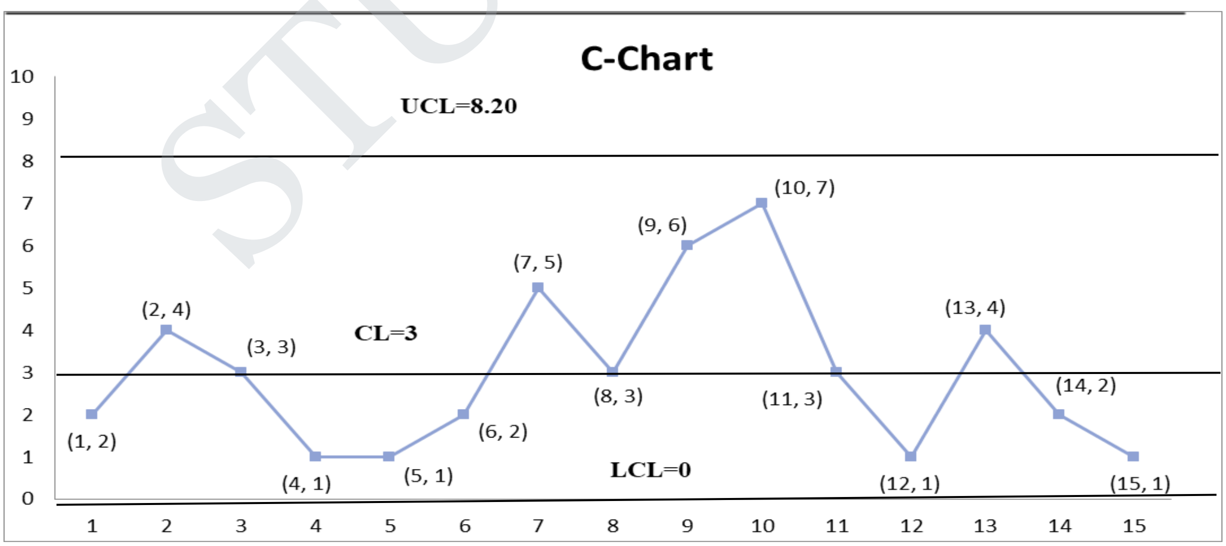
Unit No.(i)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
No. of defects (c)	**2**	**4**	**3**	**1**	**1**	**2**	**5**	**3**	**6**	**7**	**3**	**1**	**4**	**2**	**1**

Solution:

The number of defects per sample containing only one item is given, $\bar{c} = \frac{\sum c_i}{N} = \frac{(2+4+3+\cdots+2+1)}{15} = \frac{45}{15} = 3$

$CL = \bar{c} = 3;\quad LCL = \bar{c} - 3\sqrt{\bar{c}} = 3 - 3\sqrt{3} = -2.20$; $LCL = 0$ (since LCL cannot be negative)

$UCL = \bar{c} + 3\sqrt{\bar{c}} = 3 + 3\sqrt{3} = 8.20$



Since all the sample points lie within the LCL and UCL lines, the process is under control.

3b) **Construct a control chart for defectives for the following data:** **(APRIL / MAY '15) (APR / MAY '19)**

Sample No:	1	2	3	4	5	6	7	8	9	10
No. inspected:	90	65	85	70	80	80	70	95	90	75
No. of defectives:	9	7	3	2	9	5	3	9	6	7

Solution:

We note that the size of the sample varies from sample to sample.

We can construct p-chart , provided $0.75\overline{n} < n_i < 1.25\,\overline{n}$, for all i

Here $\overline{n} = \frac{1}{N}\sum n_i = \frac{1}{10}(90+65+\ldots\ldots\ldots\ldots+90+75) = \frac{1}{10}(800) = 80$

Hence $0.75\,\overline{n} = 60$ and $1.25\,\overline{n} = 100$

The values of n_i be between 60 and 100. Hence p-chart, can be drawn by the method given below.

Now $\overline{p} = \frac{\text{Total no. of defectives}}{\text{Total no. of items inspected}} = \frac{60}{800} = 0.075$

Hence for the p-chart to be constructed,

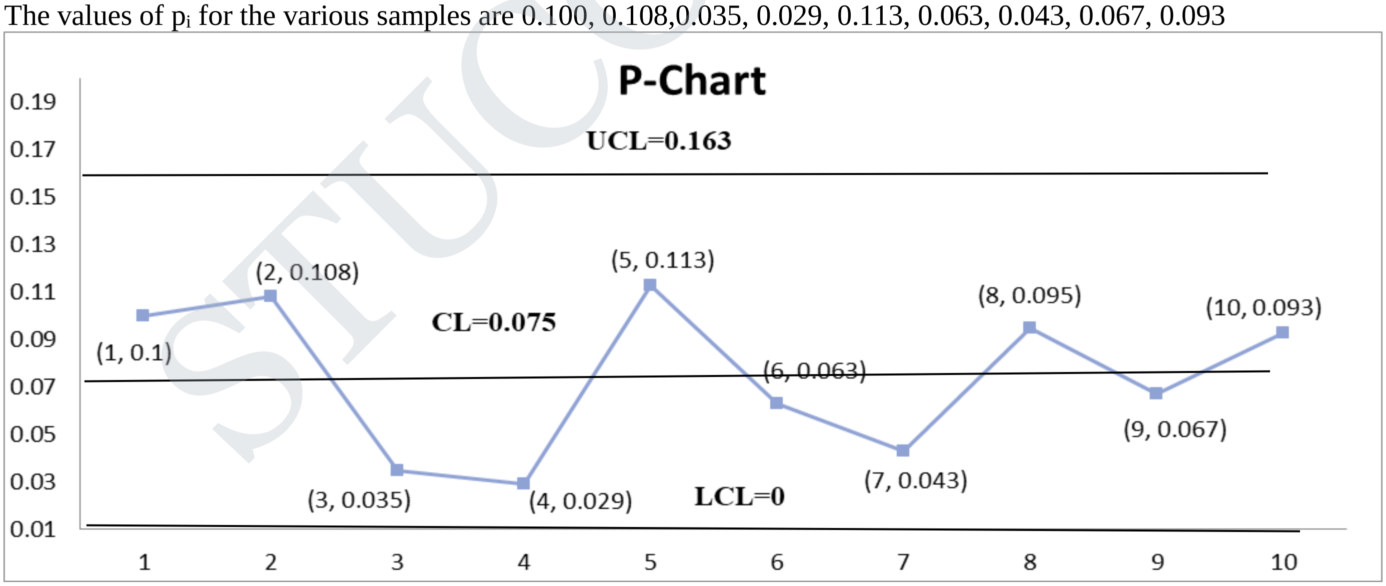
CL= $\overline{p} = 0.075$

LCL= $\overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}} = 0.075 - 3\sqrt{\frac{0.075\text{x}0.925}{80}} = -0.013$

Since LCL cannot be negative, it is taken as 0.

UCL= $\overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}} = 0.075 + 3\sqrt{\frac{0.075\text{x}0.925}{80}} = 0.163$

The values of p_i for the various samples are 0.100, 0.108,0.035, 0.029, 0.113, 0.063, 0.043, 0.067, 0.093



Since all the sample points lie within the control lines, the process is under control.

4. **The data given below are the number of defectives in 10 samples of 100 items each. Construct a p-chart and an np-chart and comment on the results.** **(MAY / JUNE '14) (APRIL / MAY '17)**

Sample No.	1	2	3	4	5	6	7	8	9	10
No. of defectives	6	16	7	3	8	12	7	11	11	4

Solution:

Sample size is constant for all samples, n=100.

Total no. of defectives = 6 + 16+7+3+8+12+7+11+11+4= 85

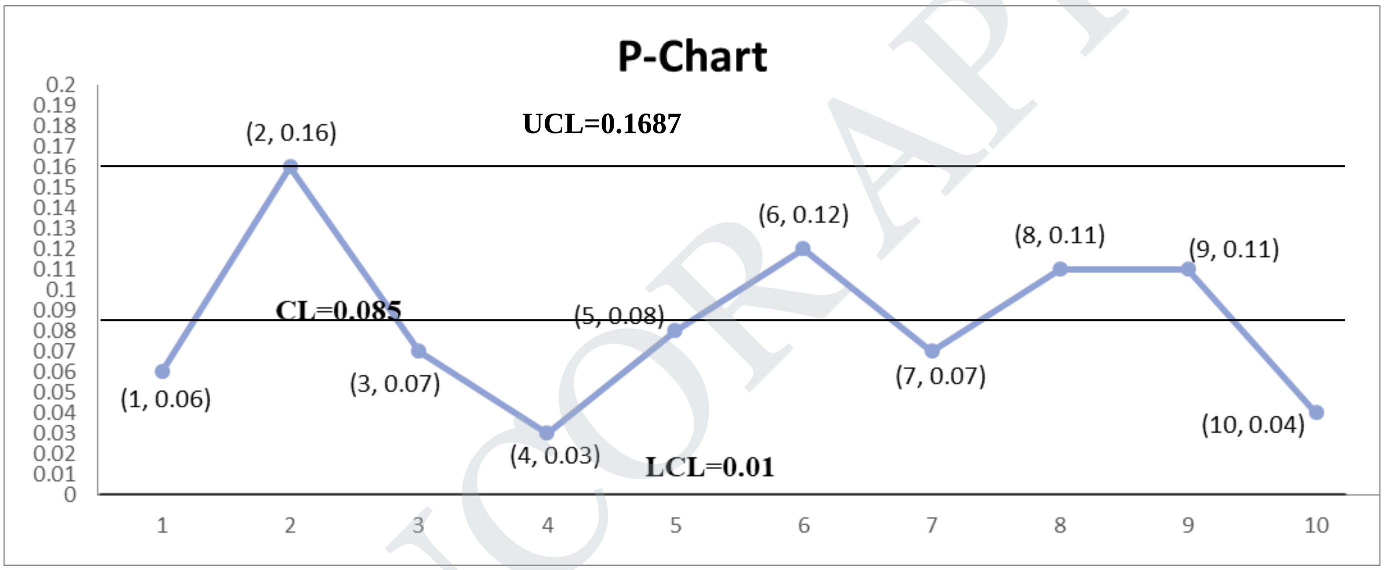
Total no. Inspected= 10 x 100 = 1000

Average fraction defective = $\bar{p} = \dfrac{\text{Total no. of defectives}}{\text{Total no. of items inspected}} = \dfrac{85}{1000} = 0.085$

For p-chart:

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 - 3\sqrt{\frac{(0.085)(0.915)}{100}} = 0.0013$$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 + \left(\sqrt{\frac{(0.085)(0.915)}{3}}\right) = 0.1687$$



Conclusion:

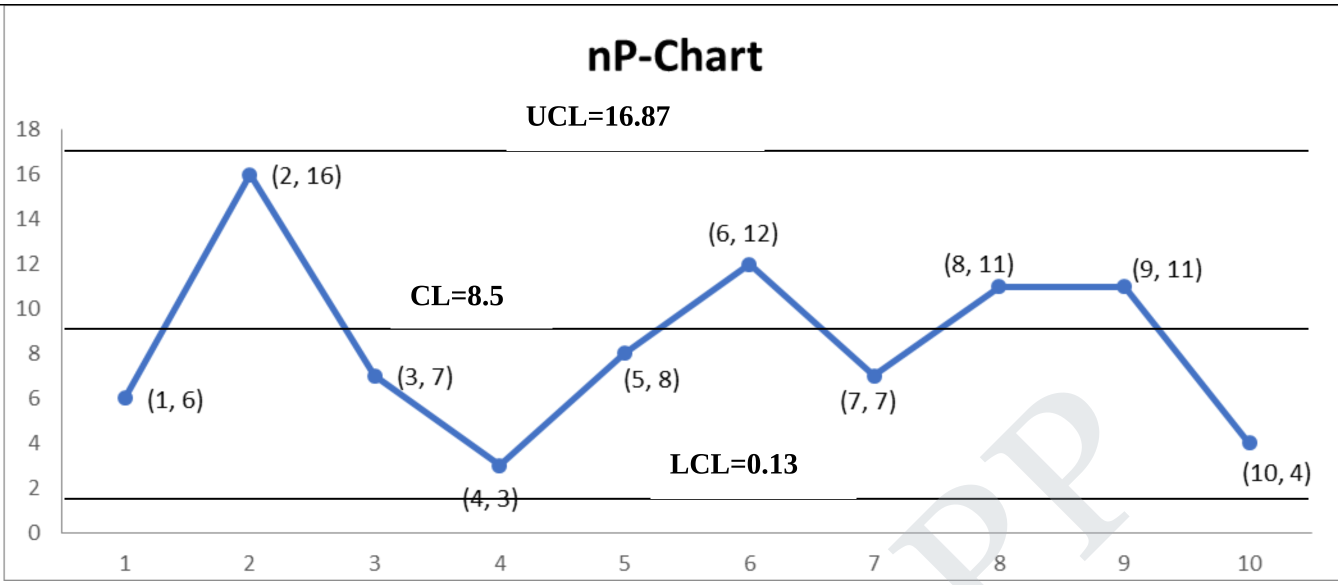
All these values are less than UCL=0.1687 and greater than LCL=0.0013. In the control chart, all sample points lie within the control limits. Hence, the process is under statistical control.

For np-chart:

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} \quad = n\left[\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] \quad = 100(0.1687) = 16.87$$

$$n\bar{p} \quad = 100(0.085) = 8.5$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})} \quad = n\left[\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] \quad = 100(0.0013) = 0.13$$



Conclusion:

All the values of number of defectives in the table lie between 16.87 and 0.13. Hence, the process is under control even in np-chart.

5. **The following data give the measurements of 10 samples each of size 5 in the production process taken in an interval of 2 hours. Calculate the sample means and ranges and draw the control charts for mean and range. (MAY / JUNE '16) (NOV / DEC '18)**

Sample No.	1	2	3	4	5	6	7	8	9	10
Observed measurements X	49	50	50	48	47	52	49	55	53	54
	55	51	53	53	49	55	49	55	50	54
	54	53	48	51	50	47	49	50	54	52
	49	46	52	50	44	56	53	53	47	54
	53	50	47	53	45	50	45	57	51	56

Solution:

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X_i} = \frac{1}{10}[52+50+50+51+47+52+49+54+51+54] = 51.0$$

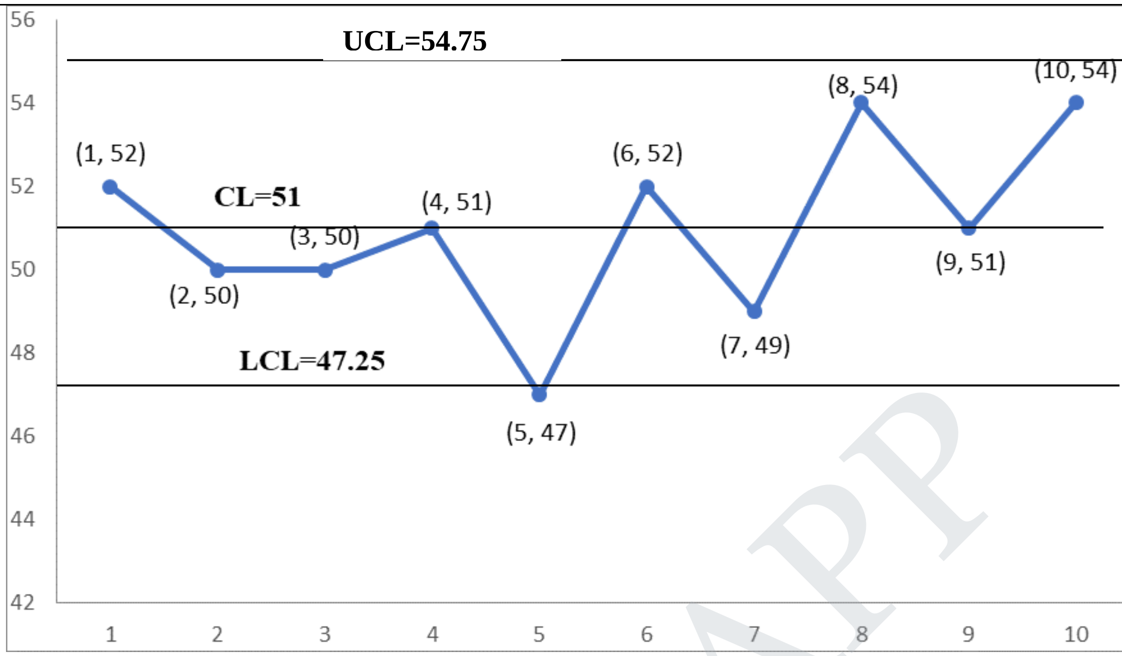
$$\overline{R} = \frac{1}{N}\sum R_i = \frac{1}{10}[6+7+6+5+6+9+8+7+7+4] = 6.5$$

From the table of control chart for sample size n=5, we have $A_2 = 0.577, D_3 = 0 \,\&\, D_4 = 2.115$

i) Control limits for $\overline{X}$ chart:

CL (central line) = $\overline{\overline{X}} = 51.0$; $LCL = \overline{\overline{X}} - A_2\overline{R}_2 = 51.0-(0.577)(6.5) = 47.2495$

$$UCL = \overline{\overline{X}} + A_2\overline{R}_2 = 51.0+(0.577)(6.5) = 54.7505$$



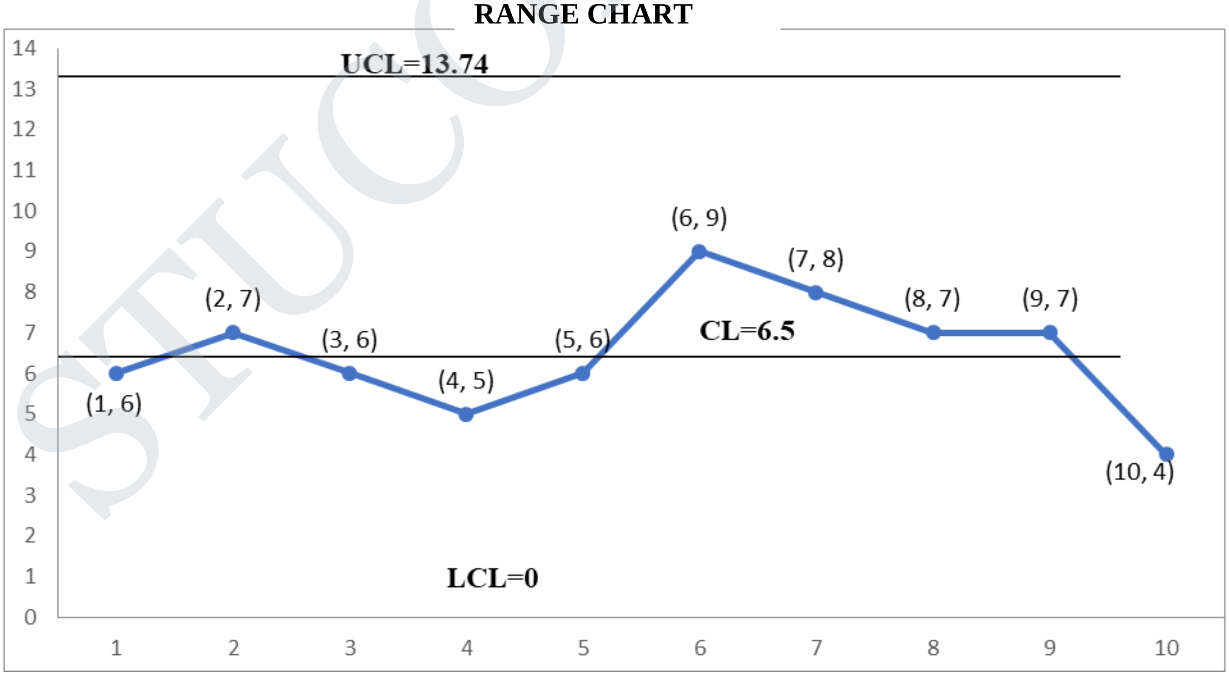
MEAN CHART

Conclusion :Since 5th sample mean fall outside the control limits the statistical process is out of control according to $\overline{X}$ *chart*

Control limits for R-Chart: $CL = \overline{R} = 6.5;\ \ LCL = D_3\ \overline{R} = 0; UCL = D_4\overline{R} = (2.115)(6.5) = 13.7475$

i)

RANGE CHART



Conclusion :
Since all the sample means fall within the control limits the statistical process is under control according to R *chart* .