UNIT – I RANDOM VARIABLES

PART- A

Problem 1 The probability density function of a continuous random variable X is given by $f(x) = Ke^{-|x|}$. Find K and C.D.F of X

Solution: Since it is a probability density function,

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

$$\int_{-\infty}^{\infty} Ke^{-|x|}dx = 1$$

$$2\int_{0}^{\infty} Ke^{-x}dx = 1$$

$$2K\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = 1 \qquad 2K = 1 \qquad K = \frac{1}{2}.$$

Therefore,

$$f(x) = \frac{1}{2}e^{x}, \quad -\infty < x < 0$$
$$= \frac{1}{2}e^{-x}, \quad 0 < x < \infty$$

For x ≤ 0, the C.D.F is $F(x) = \int_{-\infty}^{x} \frac{1}{2}e^{x}dx = \frac{1}{2}e^{x}$

For x > 0, $F(x) = \int_{-\infty}^{0} \frac{1}{2}e^{x}dx + \int_{0}^{x} \frac{1}{2}e^{-x}dx = \frac{1}{2} + \frac{1}{2}(1-e^{-x}) = \frac{1}{2}(2-e^{-x})$

Problem 2 X and Y are independent random variables with variance 2 and 3. Find the variance of $3X + 4Y$.

Solution: $V(3X + 4Y) = 9Var(X) + 16Var(Y) + 24Cov(XY)$

$$= 9\times2 + 16\times3 + 0 \quad (\therefore X \,\&\, Y \textit{ are independent } \mathrm{cov}(XY) = 0)$$

$$= 18 + 48 = 66.$$

Problem 3 A Continuous random variable X has a probability density function $F(x) = 3x^2;\ 0 \le x \le 1$.Find 'a' such that $P(x \le a) = P(x > a)$

Solution: We know that the total probability =1

Given $P(X \le a) = P(X > a) = K(say)$

Then $K + K = 1$

$$K = \frac{1}{2}$$

i.e., $P(X \le a) = \frac{1}{2}$ & $P(X > a) = \frac{1}{2}$

Consider $P(X \le a) = \frac{1}{2}$

i.e., $\int_0^a f(x)\,dx = \frac{1}{2}$

$$\int_0^a 3x^2 dx = \frac{1}{2}$$

$$3\left(\frac{x^3}{3}\right)_0^a = \frac{1}{2} \qquad a^3 = \frac{1}{2} \qquad a = \left(\frac{1}{2}\right)^{1/3}.$$

Problem 4 A random variable X has the p.d.f $f(x)$ given by $f(x) = \begin{cases} Cxe^{-x};\ if\ x > 0 \\ 0 \quad ;\ if\ x \le 0 \end{cases}$ Find the value of C and cumulative density function of X.

<u>Solution:</u>Since $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_0^{\infty} Cxe^{-x} dx = 1$$

$$C\left[x\left(-e^{-x}\right) - \left(e^{-x}\right)\right]_0^{\infty} = 1 \qquad C = 1$$

$$\therefore f(x) = \begin{cases} xe^{-x}; x > 0 \\ 0 \quad ; x \le 0 \end{cases}$$

$$C.D.F\ F(x) = \int_0^x f(x)dt = \int_0^x te^{-t}dt = \left[-te^{-t} - e^{-t}\right]_0^x = -xe^{-x} - e^{-x} + 1$$

$= 1-(1+x)e^{-x}$ for x $\ge$ 0.

Problem 5 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.

<u>Solution:</u> Mean=$\int_{-1}^{1} xf(x)\,dx = \frac{1}{2}\int_{-1}^{1} x(x+1)\,dx = \frac{1}{2}\int_{-1}^{1}\left(x^2 + x\right)dx$

$$=\frac{1}{2}\left(\frac{x^3}{3}+\frac{x^2}{2}\right)_{-1}^{1}=\frac{1}{3}$$

$$\mu_2'=\int_{-1}^{1}x^2 f(x)dx=\frac{1}{2}\int_{-1}^{1}\left(x^3+x^2\right)dx=\frac{1}{2}\left[\frac{x^4}{4}+\frac{x^3}{3}\right]_{-1}^{1}$$

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

$$=\frac{1}{2}.\frac{2}{3}=\frac{1}{3}$$

$$Variance=\mu_2'-\left(\mu_1'\right)^2$$

$$=\frac{1}{3}-\frac{1}{9}=\frac{3-1}{9}=\frac{2}{9}.$$

Problem 6 A random variable X has density function given by $f(x)=\begin{cases}2e^{-2x};x\geq 0\\0\quad ;x<0\end{cases}$. Find the moment generating function.

<u>Solution:</u> $M_X(t)=E\left(e^{tx}\right)=\int_0^{\infty}e^{tx}f(x)\,dx=\int_0^{\infty}e^{tx}2e^{-2x}dx$

$$=2\int_0^{\infty}e^{(t-2)x}dx$$

$$=2\left[\frac{e^{(t-2)x}}{t-2}\right]_0^{\infty}=\frac{2}{2-t},t<2.$$

Problem 7 If X is a Poisson variate such that $P(X=2)=9P(X=4)+90P(X=6)$, find the variance.

<u>Solution:</u>Given $P(X=2)=9P(X=4)+90P(X=6)$

$$\Rightarrow\frac{e^{-\lambda}\lambda^2}{2!}=9\frac{e^{-\lambda}\lambda^4}{4!}+90\frac{e^{-\lambda}\lambda^6}{6!}$$

$$\Rightarrow\lambda^4+3\lambda^2-4=0\Rightarrow\left(\lambda^2+4\right)\left(\lambda^2-1\right)=0$$

$$\Rightarrow\lambda^2=1\ (or)\ \lambda^2=-4$$

$\lambda=$Variance$=1$ ($\because\lambda^2$ cannot be negative)

Problem 8 Comment the following: "The mean of a binomial distribution is 3 and variance is 4

<u>Solution:</u>In a binomial distribution, mean (np) $>$ variance (npq).

Since variance $=4$ & mean $=3$, we have variance $<$ mean. Therefore, the given statement is wrong.

Problem 9 If X and Y are independent binomial variates

$B\left(5,\frac{1}{2}\right)$ and $B\left(7,\frac{1}{2}\right)$ find $P[X+Y=3]$

<u>Solution:</u> $X+Y$ is also a binomial variate with parameters $n_1+n_2=12$ & $p=\frac{1}{2}$

$$\therefore P[X+Y=3]=12C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^9=\frac{55}{2^{10}}$$

Problem 10 If X is uniformly distributed with Mean 1 and Variance $\frac{4}{3}$, find $P[X>0]$

<u>Solution:</u> If X is uniformly distributed over (a,b), then

$$E(X)=\frac{b+a}{2}\ and\ V(X)=\frac{(b-a)^2}{12}$$

$$\therefore \frac{b+a}{2}=1\Rightarrow a+b=2$$

$$\Rightarrow \frac{(b-a)^2}{12}=\frac{4}{3}\Rightarrow (b-a)^2=16$$

$$\Rightarrow a+b=2\ \&\ b-a=4 \text{ We get } b=3, a=-1$$

$\therefore a=-1$ & $b=3$ and probability density function of x is

$$f(x)=\begin{cases}\frac{1}{4};-1<x<3\\ 0\,;Otherwise\end{cases}$$

$$P[x>0]=\int_0^3\frac{1}{4}dx=\frac{1}{4}[x]_0^3=\frac{3}{4}.$$

Problem 11 If X is $N(2,3)$ ind $P\left[Y\geq\frac{3}{2}\right]$ where $Y+1=X$.

<u>Solution:</u> $P\left[Y\geq\frac{3}{2}\right]=P\left[X-1\geq\frac{3}{2}\right]$

$$=P[X\geq 2.5]=P[Z\geq 0.17], \text{ where } Z=\frac{X-2}{3}$$

$$=0.5-P[0\leq Z\leq 0.17]$$

$$=0.5-0.0675=0.4325$$

Problem 12 If the probability is $\frac{1}{4}$ that a man will hit a target, what is the chance that he will hit the target for the first time in the 7th trial?

Solution: The required probability is

$$P[FFFFFFS] = P(F)P(F)P(F)P(F)P(F)P(F)P(S)$$

$$= q^6 p = \left(\frac{3}{4}\right)^6 \cdot \left(\frac{1}{4}\right) = 0.0445 .$$

Here $p =$probability of hitting target and $q = 1 - p$.

13. If the p.d.f of a random variable X is $f(x) = 2x,\ \ 0 < x < 1$ find the C.D.F of the random variable X **[MAY / JUNE 15]**

Sol:

Given $f(x) = 2x,\ \ 0 < x < 1$

If x<0, F(X)=0

If $0 < x < 1, F(x) = \int_{-\infty}^{x} f(x)dx \quad = \int_0^x 2xdx = 2\left[\frac{x^2}{2}\right]_0^x = x^2$

If x>1, F(X)=1

14. The time that , a teacher takes to grade a paper is uniformely distributed between 5 minutes and 10 minutes. Find the mean and variance of the time he takes to grade a paper. **[MAY / JUNE 15]**

$$f(x) = \frac{1}{b-a}, a < x < b$$

$$= \frac{1}{5}, 5 < x < 10$$

$$E(X) = \frac{b+a}{2} = \frac{15}{2} = 7.5$$

$$E(X^2) = \frac{b^2 + ab + a^2}{3} = \frac{(10^2 + 5 * 10 + 5^2)}{3} = \frac{(100 + 50 + 25)}{3}$$

$$= \frac{175}{3} = 58.33$$

$$Var(X) = E(X^2) + \left(E(X)\right)^2 = 58.33 - (7.5)^2 = 58.33 - 56.25 = 2.08$$

15. The p.d.f of a random variable X is $f(x) = \begin{cases} \frac{1}{4}, -2 \le x \le 2 \\ 0,\ \ otherwise \end{cases}$ Find $P(|x| > 1)$ and E(X) **[NOV/DEC 15]**

$$\boldsymbol{P(|X| > 1) = 1 - P(|X| \le 1) = 1 - P(-1 < X < 1)}$$

$$= \boldsymbol{1 - \frac{1}{4}\int_{-1}^{1} dx} = \frac{1}{2}$$

$$E(X) = \int_{-2}^{2} xf(x)dx = \int_{-2}^{2} x\frac{1}{4}dx$$

$$= \frac{1}{8}(4 - 4) = 0$$

16. If the M.G.F of a random variable X is $M_X(t) = (1-7t)^{-20}$, find the p.d.f and mean of X **[NOV/DEC 15]**

Sol:

$$M_X(t) = (1-7t)^{-20}$$

$$E(X) = [140(1-7t)^{-21}]_{t=0} = 140$$

PART-B

1. A random variable X has the following distribution,

X 0 1 2 3 4 5 6 7 P(X) 0

k 2k 2k 3k k^2 2 k^2 7 k^2 +k

Find (i) the value of k (ii) P (1.5 < X < 4.5 | x > 2)

(iii) the smallest value of λ for which P(X <λ) > ½

Find (i) K , (ii) Evaluate $P(X<6), P(X \geq 6)$ and $P(0<X<5)$

(iii). Determine the distribution function of X . **[MAY / JUNE 15]**

Soln: If X is a discrete random variable and p(x) is a pmf of X

$$\sum p(x) = 1$$

$\Rightarrow$ k+2k+2k+3k+k^2+2k^2+7k^2+k = 1

$\Rightarrow$ k = 1/10 or k = -1

$\therefore$ k = 1/10

(ie) the probability distribution of X is given by

Value of X	0	1	2	3	4	5	6	7
P(X = x)	0	1/10	2/10	2/10	3/10	1/100	2/100	17/100

(i) P(1.5 < X < 4.5/X>2) = $\dfrac{P(1.5 < X < 4.5 \cap X > 2)}{P(X>2)} = \dfrac{P(x=3)+P(X=4)}{1-P(X \leq 2)}$

Now $P(x=3)+P(X=4) =$ 5/10

And $1-P(X \leq 2)$ = 7/10

$\therefore$ P(1.5 < X < 4.5/X>2) = 5/7

CDF F(X) = P(X $\leq$ x)

Value of X	0	1	2	3	4	5	6	7
P(X = x)	0	1/10	2/10	2/10	3/10	1/100	2/100	17/100
F(X)	0	1/10	3/10	5/10	8/10	81/100	83/100	1

(ii) if P(X$\leq$ K) > ½ then least value of k =4.

(iii) $P(X<6)=P(X=0)+P(X=1)+...+P(X=5)$

$$=\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+...=\frac{81}{100}$$

Now $P(X\geq 6)=1-P(X<6)$

$$=1-\frac{81}{100}=\frac{19}{100}$$

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

$$=K+2K+2K+3K$$

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

2. Find MGF of a Binomial random variable and hence find its $E(x), Var(x)$**[MAY / JUNE 15]**

The Moment generating function of a random variable X is defined as $E\left[e^{tx}\right]$ for all t$\in(-\infty,\infty)$

It is denoted by $M_X(t)$

If X is a discrete random variables with values $x_{1,}\, x_2$....... and probability function is p(x) is

$$M_X(t)=\sum_{x=0}^{\infty} e^{tx}p(x)$$

(ii)If X is a continuous random variable with pdf f(x) ,x $\in(-\infty,\infty)$ then $M_X(t)=\int_{-\infty}^{\infty} e^{tx}f(x)dx$

Moment generating function (MGF) of Binomial distribution

Binomial distribution is

p(x) = P(X = x) = $nc_x p^x q^{n-x}$, x = 0,1,2….n

$$M_X(t)=E\left[e^{tx}\right]=\sum_{x=0}^{n} e^{tx}p(x)$$

$$=\sum_{x=0}^{n} e^{tx}nc_x\, p^x\, q^{n-x}$$

$$= \sum_{x=0}^{n} nc_x \,(pe^t)^x\, q^{n-x} \;=\; \left(q + pe^t\right)^n \qquad \left[\because (x+a)^n = \sum_{x=0}^{n} nc_r x^{n-r}.a^r\right]$$

Mean and Variance of Binomial distribution

We know that $M_X(t) = \left(q + pe^t\right)^n$

$$M_X{}'(t) = n\left(q + pe^t\right)^{n-1}.pe^t$$

$$M_X{}''(t) = np\left((n-1)(q+pe)^{n-2}\left(pe^t\right)e^t + \left(q+pe^t\right)^{n-1}.e^t\right)$$

$$= np\left(q+pe^t\right)^{n-2}.e^t\left[(n-1)pe^t + q + pe^t\right]$$

Put n =0 we get

$$M_X{}'(0) = n(q+p)^{n-1}.p$$

$$\Rightarrow Mean = \mu_1' = np$$

$$M_X{}''(0) = np((n-1)p + q + p)$$
$$\Rightarrow \mu_2' = np(np+q) = n^2p^2 + npq$$

$$Var(X) = \mu_2{}' - \left(\mu_1{}'\right)^2 = n^2p^2 + npq - n^2p^2 = \text{npq}$$

3. An insurance company has discovered that only about 0.1% of the population is involved in a certain type of accident each year. If its 10,000 policy holders were randomly selected from the population, what is the probability that not more than 5 of its clients are involved in a such an accident next year? **[MAY / JUNE 15]**

Sol:

Given p=0.1% n=10,000

P(X=x)=$\frac{e^{-\lambda}\lambda^x}{x!}$**, x=0, 1, 2,…….**

$$\boldsymbol{P(X \le 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)}$$

$$\boldsymbol{= e^{-10}\left[1 + 10 + \frac{100}{2} + \frac{1000}{6} + \frac{10000}{24} + \frac{100000}{120}\right]}$$

$$\boldsymbol{= 0.0671}$$

4. Memory less property of Exponential distribution. **[MAY / JUNE 15]**

Statement of memory less property:

If X is exponentially distributed with parameter λ then for any two positive integer S and T
P(X >S+T/ X>S) = P(X>T)

Proof: The pdf of X is $f(x) = \begin{cases} \theta e^{-\theta x} & ; x \ge 0 \\ 0\,; otherwise \end{cases}$

P(X>k) = $\int_k^\infty \theta e^{-\theta x} dx$ =θ$\left[\frac{e^{-\theta x}}{-\theta}\right]_k^\infty$ = $e^{-\theta k}$

Consider s , t $\geq$ 0 , $P\left[\frac{X>s+t}{X>s}\right] = \frac{P(X>s+t \cap X>s)}{P(X>s)}$

$$= \frac{P(X>s+t\,)}{P(X>s)} \quad = \frac{e^{-\theta(s+t)}}{e^{-\theta s}} = e^{-\theta t} = P(X > t)$$

5. Suppose the random variable X has Geometric distribution with p=0.5. Compute (i) P(X>2) (ii) MGF of $M_X(t)$ of X (iii) E(X) and Var(X) **[NOV/DEC 15]**

Sol:

$$P(X = x) = q^{x-1}p, \qquad x = 1,2,3, \dots ....$$

$$= (0.5)^x$$

$$(i)\ P(X > 2) = 1 - P(\leq 2)$$

$$= 1 - [\boldsymbol{P(X = 1)} + \boldsymbol{P(X = 2)}]$$

$$= 1 - [0.5 + 0.5^2]$$

$$= 0.25$$

$(ii) M_X(t)$ $=\sum e^{tx}\,(0.5)^x$

$$= 0.5e^t[1 + 0.5e^t + (0.5e^t)^2 + \cdots \dots]$$

$$= \frac{0.5e^t}{1 - 0.5e^t}$$

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

$$= \frac{d}{dt}\left[\frac{0.5e^t}{1 - 0.5e^t}\right]_{t=0}$$

$$= \left[\frac{1}{(1 - 0.5e^t)^2}\right]_{t=0}$$

$$E(X) = 2$$

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

$$= \frac{d}{dt}\left[\frac{0.5e^t}{(1 - 0.5e^t)^2}\right]_{t=0}$$

$$E(X^2) = 6$$

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

6. If a random variable X has the p.d.f $f(x) = \begin{cases} xe^{-x}, x > 0 \\ 0 \quad\ , x \leq 0 \end{cases}$, find (i) cumulative distribution function (ii) P(X>2) (iii) MGF of X (IV) E(X) **[NOV/DEC 15]**

Sol:

(i) If $x > 0,\ F(x) = \int_{-\infty}^{x} f(x)dx \quad = \int_0^x xe^{-x}dx$=$[-xe^{-x} - e^{-x}]_0^x$

$$= 1 - e^{-x}(x+1)$$

(ii) $P(X > 2) = \int_2^{\infty} xe^{-x}dx$=$[-xe^{-x} - e^{-x}]_2^{\infty}$

$$= 3e^{-2} = 0.406$$

(iii) $M_X(t) = \sum_{x=0}^{\infty} xe^{-x}e^{tx} = \sum_{x=0}^{\infty} xe^{-(1-t)x}$

$$= \frac{e^{-(1-t)}}{1 - e^{-(1-t)}}$$

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

$= \left[\frac{e^{-(1-t)}}{\left[1-e^{-(1-t)}\right]^2}\right]_{t=0}$=$\frac{e^{-1}}{(1-e^{-1})^2} = 0.9207$

7. If the p.d.f of the random variable X is $f(x) = \begin{cases} \frac{1}{3}e^{-\frac{1}{3}x}, x > 0 \\ 0, x \leq 0 \end{cases}$, find 1. Cumulative distribution function. 2). $P(\frac{1}{3} < X < \frac{2}{3})$ 3). $P(X > 3/X > 2)$ 4).E(X) **[NOV/DEC 15]**

Sol:

(i) **If** $x > 0,\ F(x) = \int_{-\infty}^{x} f(x)dx \quad = \int_0^x \frac{1}{3}e^{-\frac{1}{3}x}dx = 1 - e^{-\frac{x}{3}}$

(ii) $P\left(\frac{1}{3} < X < \frac{2}{3}\right)$ $=\int_{\frac{1}{3}}^{\frac{2}{3}} f(x)dx = \frac{1}{3}\int_{\frac{1}{3}}^{\frac{2}{3}} e^{-\frac{1}{3}x}dx = -\left(e^{-\frac{2}{9}} - e^{-\frac{1}{9}}\right) = 0.0941$

(iii) **P(X>3/X>2)=**$\frac{\boldsymbol{P(X>3 \cap X>2)}}{\boldsymbol{P(X>2)}} = \frac{\boldsymbol{P(X>3)}}{\boldsymbol{(PX>2)}}$=$\frac{\int_3^{\infty} \boldsymbol{e^{-\frac{x}{3}}dx}}{\int_2^{\infty} \boldsymbol{e^{-\frac{x}{3}}dx}} = \frac{\boldsymbol{e^{-1}}}{\boldsymbol{e^{-\frac{2}{3}}}} = \boldsymbol{e^{-\frac{1}{3}}}$**=0.7165**

(iv) $E(X) = \int_0^{\infty} xf(x)dx = \frac{1}{3}\int_0^{\infty} x\boldsymbol{e}^{-\frac{x}{3}}dx = \frac{1}{3}\left[-3x\boldsymbol{e}^{-\frac{x}{3}} - \boldsymbol{9e}^{-\frac{x}{3}}\right]_0^{\infty} = 3$

8. Let X be a uniform random variable over(-5,5). Find (i) MGF of X (ii) $P(X \leq 0)$ (iii) $P(|X| > 2)$ (iv) $P(-3 < X \leq 3)$ **[NOV/DEC 15]**

$$f(x) = \frac{1}{b-a}, a < x < b$$

$$= \frac{1}{10}, -5 < x < 5$$

(i) $M_X(t) = \frac{1}{10}\int_{-5}^{5} e^{tx}\boldsymbol{dx} = \frac{\boldsymbol{1}}{\boldsymbol{10t}}\left(\boldsymbol{e^{5t} - e^{-5t}}\right)$

(ii) $P(\leq 0) = \frac{1}{10}\int_{-5}^{0} dx = \frac{5}{10} = \frac{1}{2}$

(iii) $P(|X| > 2) = 1 - \int_{-2}^{2} \frac{1}{10}dx = 1 - \frac{1}{10}(2+2) = \frac{6}{10} = \frac{3}{5}$

(iv) $P(-3 < X \leq 3) = \int_{-3}^{3} \frac{1}{10}dx = \frac{1}{10}(3+3) = \frac{6}{10} = \frac{3}{5}$

If the probability distribution of X is given as

X	:	1	2	3	4
$P(X)$	:	0.4	0.3	0.2	0.1

Find $P(1/2 < X < 7/2 / X > 1)$

(ii). If $P(x)=\begin{cases}\frac{x}{15}; x=1,2,3,4,5\\ 0 \quad ; elsewhere\end{cases}$

find (a) $P\{X=1 or 2\}$ *and* (b) $P\{1/2<X<5/2/x>1\}$

Solution:(i) $P\{1/2<X<7/2/X>1\}=\dfrac{P\{(1/2<X<7/2)\cap X>1\}}{P(X>1)}$

$$=\frac{P(X=2 or 3)}{P(X=2,3 or 4)}$$

$$=\frac{P(X=2)+P(X=3)}{P(X=2)+P(X=3)+P(X=4)}$$

$$=\frac{0.3+0.2}{0.3+0.2+0.1}=\frac{0.5}{0.6}=\frac{5}{6}.$$

(ii) (a) $P(X=1 \, or \, 2)=P(X=1)+P(X=2)$

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

(b) $P\left(\frac{1}{2}<X<\frac{5}{2}/x>1\right)=\dfrac{P\left\{\left(\frac{1}{2}<X<\frac{5}{2}\right)\cap(X>1)\right\}}{P(X>1)}$

$$=\frac{P\{(X=1 or 2)\cap(X>1)\}}{P(X>1)}$$

$$=\frac{P(X=2)}{1-P(X=1)}$$

$$=\frac{2/15}{1-(1/15)}=\frac{2/15}{14/15}=\frac{2}{14}=\frac{1}{7}.$$

Problem 15 A random variable X has the following probability distribution

X	:	- 2	- 1	0	1	2	3
$P(X)$	:	0.1	K	0.2	$2K$	0.3	$3K$

a) Find K, b) Evaluate $P(X<2)$ and $P(-2<X<2)$

b) Find the cdf of X and d) Evaluate the mean of X.

Solution:a) Since $\sum P(X)=1$

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

$$6K+0.6=1$$

$$6K=0.4$$

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

b) $P(X < 2) = P(X = -2, -1, 0\, or\, 1)$

$$= P(X = -2) + P(X = -1) + P(X = 0) + P(X = 1)$$

$$= \frac{1}{10} + \frac{1}{15} + \frac{1}{5} + \frac{2}{15}$$

$$= \frac{3+2+6+4}{30} = \frac{15}{30} = \frac{1}{2}$$

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

$$= P(X = -1) + P(X = 0) + P(X = 1)$$

$$= \frac{1}{15} + \frac{1}{5} + \frac{2}{15}$$

$$= \frac{1+3+2}{15} = \frac{6}{15} = \frac{2}{5}$$

c) The distribution function of X is given by $F_X(x)$ defined by

$$F_X(x) = P(X \le x)$$

X	:	- 2	- 1	0	1	2	3
$F_X(x)$	:	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{11}{30}$	$\frac{1}{2}$	$\frac{4}{5}$	1

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

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

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

Problem 16 X is a continuous random variable with pdf given by

$$F(X) = \begin{cases} Kx & in \quad 0 \le x \le 2 \\ 2K & in \quad 2 \le x \le 4 \\ 6K - Kx & in \quad 4 \le x \le 6 \\ 0 & elsewhere \end{cases}$$

Find the value of K and also the cdf $F_X(x)$.

Solution: Since $\int_{\infty}^{\infty} F(x)\,dx = 1$

$$\int_0^2 Kxdx + \int_2^4 2Kdx + \int_4^6 (6k - kx)dx = 1$$

$$K\left[\left(\frac{x^2}{2}\right)_0^2 + (2x)_2^4 + \int_4^6\left(6x - \frac{x^2}{2}\right)_4^6\right] = 1$$

$$K\left[\cancel{2} + \cancel{8} - 4 + 36 - 18 - 24 + 8\right] = 1$$

$$8K = 1$$

$$K = \frac{1}{8}$$

We know that $F_X(x) = \int_{-\infty}^{x} f(x)\,dx$

If $x < 0$, then $F_X(x) = \int_{-\infty}^{x} f(x)\,dx = 0$

If $x \in (0,2)$, then $F_X(x) = \int_{-\infty}^{x} f(x)\,dx$

$$F_X(x) = \int_{-\infty}^{0} f(x)\,dx + \int_0^x f(x)\,dx$$

$$= \int_{-\infty}^{0} 0dx + \int_0^x Kxdx = \int_{-\infty}^{0} 0dx + \frac{1}{8}\int_0^x xdx$$

$$= \left(\frac{x^2}{16}\right)_0^x = \frac{x^2}{16}, 0 \le x \le 2$$

If $x \in (2,4)$, then $F_X(x) = \int_{-\infty}^{x} f(x)\,dx$

$$F_X(x) = \int_{-\infty}^{0} f(x)\,dx + \int_0^2 f(x)\,dx + \int_2^x f(x)\,dx$$

$$= \int_{-\infty}^{0} 0dx + \int_0^2 Kxdx + \int_2^x 2Kdx$$

$$= \int_0^2 \frac{x}{8}dx + \int_2^x \frac{1}{4}dx = \left(\frac{x^2}{16}\right)_0^2 + \left(\frac{x}{4}\right)_2^x$$

$$= \frac{1}{4} + \frac{x}{4} - \frac{1}{2}$$

$$= \frac{x}{4} - \frac{4}{16} = \frac{x-1}{4}, 2 \le x < 4$$

If $x \in (4,6)$, then $F_X(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{2} Kxdx + \int_{2}^{4} 2Kdx + \int_{4}^{x} k(6-x)dx$

$$= \int_{0}^{2} \frac{x}{8}dx + \int_{2}^{4} \frac{1}{4}dx + \int_{4}^{x} \frac{1}{8}(6-x)dx$$

$$= \left(\frac{x^2}{16}\right)_0^2 + \left(\frac{x}{4}\right)_2^4 + \left(\frac{6x}{8} - \frac{x^2}{16}\right)_4^x$$

$$= \frac{1}{4} + 1 - \frac{1}{2} + \frac{6x}{8} - \frac{x^2}{16} - 3 + 1$$

$$= \frac{4+16-8+12x-x^2-48+16}{16}$$

$$= \frac{-x^2+12x-20}{16}, 4 \le x \le 6$$

If $x > 6$, then $F_X(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{2} Kxdx + \int_{2}^{4} 2Kdx + \int_{4}^{6} k(6-x)dx + \int_{6}^{x} 0dx$

$$= 1, x \ge 6$$

$$\therefore F_X(x) = \begin{cases} 0 & ; x \le 0 \\ \frac{x^2}{16} & ; 0 \le x \le 2 \\ \frac{1}{4}(x-1) & ; 2 \le x \le 4 \\ \frac{-1}{16}(20-12x+x^2) & ; 4 \le x \le 6 \\ 1 & ; x \ge 6 \end{cases}$$

Problem 17 A random variable X has density function $f(x) = \begin{cases} \frac{K}{1+x^2}, & -\infty < x < \infty \\ 0, & Otherwise \end{cases}$

Determine K and the distribution function. Evaluate the probability $P(x \ge 0)$.

<u>Solution:</u> Since $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} \frac{K}{1+x^2}dx = 1$$

$$K\int_{\infty}^{\infty} \frac{dx}{1+x^2} = 1$$

$$K\left(\tan^{-1} x\right)_{-\infty}^{\infty} = 1$$

$$K\left(\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right)=1$$

$$K\pi=1$$

$$K=\frac{1}{\pi}$$

$$F_X\left(x\right)=\int_{-\infty}^{x} f\left(x\right)dx=\int_{-\infty}^{x}\frac{K}{1+x^2}dx$$

$$=\frac{1}{\pi}\left(\tan^{-1}x\right)_{-\infty}^{x}$$

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

$$=\frac{1}{\pi}\left[\frac{\pi}{2}+\tan^{-1}x\right],-\infty<x<\infty$$

$$P\left(X\geq 0\right)=\frac{1}{\pi}\int_0^{\infty}\frac{dx}{1+x^2}=\frac{1}{\pi}\left(\tan^{-1}x\right)_0^{\infty}$$

$$=\frac{1}{\pi}\left(\frac{\pi}{2}-\tan^{-1}0\right)=\frac{1}{2}.$$

Problem 18 If X has the probability density function $f\left(x\right)=\begin{cases}Ke^{-3x}, & x>0\\ 0 & ,otherwise\end{cases}$ find K, $P\left[0.5\leq X\leq 1\right]$ and the mean of X.

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

$$\int_0^{\infty}Ke^{-3x}dx=1$$

$$K\left[\frac{e^{-3x}}{-3}\right]_0^{\infty}=1 \qquad \frac{K}{3}=1 \qquad K=3$$

$$P\left(0.5\leq X\leq 1\right)=\int_{0.5}^{1} f\left(x\right)dx=3\int_{0.5}^{1} e^{-3x}dx=\cancel{3}\frac{e^{-3}-e^{-1.5}}{-\cancel{3}}=\left[e^{-1.5}-e^{-3}\right]$$

Mean of $X=E\left(x\right)=\int_0^{\infty}xf\left(x\right)dx=3\int_0^{\infty}xe^{-3x}dx$

$$=3\left[x\left(\frac{-e^{-3x}}{3}\right)-1\left(\frac{e^{-3x}}{9}\right)\right]_0^{\infty}=\frac{3\times 1}{9}=\frac{1}{3}$$

Problem 19 A random variable X has the P.d.f $f(x)=\begin{cases} 2x, 0<x<1 \\ 0 \quad , Otherwise \end{cases}$

Find (i) $P\left(X<\frac{1}{2}\right)$ (ii) $P\left(\frac{1}{4}<x<\frac{1}{2}\right)$ (iii) $P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)$

Solution: (i) $P\left(x<\frac{1}{2}\right)=\int\limits_{0}^{1/2} f(x)\,dx=\int\limits_{0}^{1/2} 2x dx=2\left(\frac{x^2}{2}\right)_0^{1/2}=\frac{2\times 1}{8}=\frac{1}{4}$

$$\text{(ii) } P\left(\frac{1}{4}<x<\frac{1}{2}\right)=\int\limits_{1/4}^{1/2} f(x)\,dx=\int\limits_{1/4}^{1/2} 2x dx=2\left(\frac{x^2}{2}\right)_{1/4}^{1/2}$$

$$=2\left(\frac{1}{8}-\frac{1}{32}\right)=\left(\frac{1}{4}-\frac{1}{16}\right)=\frac{3}{16}.$$

$$\text{(iii) } P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)=\frac{P\left(X>\frac{3}{4}\cap X>\frac{1}{2}\right)}{P\left(X>\frac{1}{2}\right)}=\frac{P\left(X>\frac{3}{4}\right)}{P\left(X>\frac{1}{2}\right)}$$

$$P\left(X>\frac{3}{4}\right)=\int\limits_{3/4}^{1} f(x)\,dx=\int\limits_{3/4}^{1} 2x dx=2\left(\frac{x^2}{2}\right)_{3/4}^{1}=1-\frac{9}{16}=\frac{7}{16}$$

$$P\left(X>\frac{1}{2}\right)=\int\limits_{1/2}^{1} f(x)\,dx=\int\limits_{1/2}^{1} 2x dx=2\left(\frac{x^2}{2}\right)_{1/2}^{1}=1-\frac{1}{4}=\frac{3}{4}$$

$$P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)=\frac{\frac{7}{16}}{\frac{3}{4}}=\frac{7}{16}\times\frac{4}{3}=\frac{7}{12}.$$

Problem 20 A Man drawn 3 balls from an urn containing 5 white and 7 black balls. He gets Rs.10 for each white ball and Rs.5 for each black ball. Find his expectation.

Solution: Let X denotes the amount that he expects to receive

$$\boxed{\begin{array}{cc} W & B \\ 5 & 7 \end{array}}$$

$$\begin{array}{lllll} & 3B & 1W \,\&\, 2B & 2W \,\&\, 1B & 3W \\ X = & Rs15 & Rs20 & Rs25 & Rs30 \end{array}$$

$$P(X=15)=P(3\,Black\ balls)=\frac{7C_3}{12C_3}=\frac{\frac{7\times 6\times 5}{1\times 2\times 3}}{\frac{12\times 11\times 10}{1\times 2\times 3}}=\frac{7}{44}$$

$$P(X=20)=P(2B1W)=\frac{7C_2.5C_1}{12C_3}=\frac{21}{44}$$

$$P(x=25)=P(2W\,1B)=\frac{7c2.5c2}{12c3}=\frac{14}{44}$$

$$P(X=30)=P(3W)=\frac{5C_3}{12C_3}=\frac{2}{44}$$

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

$$=15\times\frac{7}{44}+20\times\frac{21}{44}+25\times\frac{14}{44}+30\times\frac{2}{44}$$

$$=\frac{935}{44}=Rs.21.25$$

Problem 21 From an urn containing 3 red and 2 black balls, a man is to draw 2 balls at random without replacement, being promised Rs.20/- for each red ball he draws and Rs.10/- for each black ball. Find his expectation.

Solution: Let X denotes the amount he receives

R	B
3	2

$$\begin{array}{llll} & 2B & 1R\ 1B & 2R \\ X= & Rs20 & Rs30 & Rs40 \end{array}$$

$$P(X=20)=P(2\ Black\ balls)=\frac{2C_2}{5C_2}=\frac{\frac{2\times1}{1\times2}}{\frac{5\times4}{1\times2}}=\frac{1}{10}$$

$$P(X=30)=P(1\,\mathrm{Re}d\ \ \&\,1\,Black\ ball)=\frac{3C_1\times2C_1}{5C_2}=\frac{6}{10}$$

$$P(X=40)=P(2\ \mathrm{Re}d\ balls)=\frac{3C_2}{5C_2}=\frac{3}{10}$$

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

$$=20\times\frac{1}{10}+30\times\frac{6}{10}+40\times\frac{3}{10}$$

$$=\frac{20+180+120}{10}=\frac{320}{10}$$

$$E(x)=Rs.32/-$$

Problem 22 The elementary probability law of a continuous random variable is $f(x)=y_0e^{-b(x-a)}$, $a\le x<\infty,\ b>0$ where a, b and y_0 are constants. Find y_0, the r^th^ moment about the point $x=a$ and also find the mean and variance.

Solution: Since the total probability is unity,

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

$$y_0 \int_{0}^{\infty} e^{-b(x-a)} dx = 1$$

$$y_0 \left[\frac{e^{-b(x-a)}}{-b} \right]_0^{\infty} = 1$$

$$y_0 \left(\frac{1}{b} \right) = 1$$

$$y_0 = b.$$

μ_r' (rth moment about the point $x = a$)$= \int_{-\infty}^{\infty} (x-a)^r f(x)\,dx$

$$= b \int_{a}^{\infty} (x-a)^r e^{-b(x-a)} dx$$

Put $x - a = t$, $dx = dt$, when $x = a, t = 0$; $x = \infty, t = \infty$

$$= b \int_{0}^{\infty} t^r e^{-bt} dt$$

$$= b \frac{\Gamma(r+1)}{b^{(r+1)}} = \frac{r!}{b^r}$$

In particular $r = 1$

$$\mu_1' = \frac{1}{b}$$

$$\mu_2' = \frac{2}{b^2}$$

Mean$= a + \mu_1' = a + \frac{1}{b}$

Variance$= \mu_2' - \left(\mu_1'\right)^2$

$$= \frac{2}{b^2} - \frac{1}{b^2} = \frac{1}{b^2}.$$

Problem 23 The first four moments of a distribution about $x = 4$ are 1,4,10 and 45 respectively. Show that the mean is 5, variance is 3, $\mu_3 = 0$ and $\mu_4 = 26$.

Solution: Given $\mu_1' = 1, \mu_2' = 4, \mu_3' = 10, \mu_4' = 45$

$\mu_r' = r^{th}$ moment about to value $x = 4$

Here $A = 4$

Hence mean $= A + \mu_1' = 4 + 1 = 5$

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

$$= 4 - 1 = 3 .$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\left(\mu_1'\right)^3$$

$$= 10 - 3(4)(1) + 2(1)^3 = 0$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\left(\mu_1'\right)^2 - 3\left(\mu_1'\right)^4$$

$$= 45 - 4(10)(1) + 6(4)(1)^2 - 3(1)^4$$

$$\mu_4 = 26 .$$

Problem 24 A continuous random variable X has the p.d.f $f(x) = kx^2e^{-x},\ x \geq 0.$ Find the r^th^ moment of X about the origin. Hence find mean and variance of X.

Solution: Since $\int_0^\infty Kx^2e^{-x}dx = 1$

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

$$2K = 1 \qquad K = \frac{1}{2}.$$

$$\mu_r' = \int_0^\infty x^r f(x)dx$$

$$= \frac{1}{2}\int_0^\infty x^{r+2}e^{-x}dx$$

$$= \frac{1}{2}\int_0^\infty e^{-x}x^{(r+3)-1}dx = \frac{(r+2)!}{2}$$

Putting $r = 1$, $\mu_1' = \frac{3!}{2} = 3$

$$r = 2\,, \mu_2' = \frac{4!}{2} = 12$$

$$\therefore \text{ Mean} = \mu_1' = 3$$

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

$$\text{i.e., } \mu_2 = 12 - (3)^2 = 12 - 9 = 3$$

Problem 25 Find the moment generating function of the random variable X, with probability density function $f(x) = \begin{cases} x & for \quad 0 \le x < 1 \\ 2-x & for \quad 1 \le x < 2 \\ 0 & otherwise \end{cases}$. Also find μ_1', μ_2'.

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

$$= \int_0^1 e^{tx} x dx + \int_1^2 e^{tx}(2-x)\,dx$$

$$= \left(\frac{xe^{tx}}{t} - \frac{e^{tx}}{t^2}\right)_0^1 + \left[(2-x)\frac{e^{tx}}{t} - (-1)\frac{e^{tx}}{t^2}\right]_1^2$$

$$= \frac{e^t}{t} - \frac{e^t}{t^2} + \frac{1}{t^2} + \frac{e^{2t}}{t^2} - \frac{e^t}{t} - \frac{e^t}{t^2}$$

$$= \left(\frac{e^t - 1}{t}\right)^2$$

$$= \left[1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots - 1\right]^2$$

$$= \left[1 + \frac{t}{2!} + \frac{t^2}{3!} + \frac{t^3}{4!} + \ldots\right]^2$$

$$\mu_1' = coeff.\ of\ \frac{t}{1!} = 1$$

$$\mu_2' = coeff.\ of\ \frac{t^2}{2!} = \frac{7}{6}.$$

Problem 26 Find the moment generating function and rth moments for the distribution whose p.d.f is $f(x) = Ke^{-x},\ 0 \le x \le \infty$. Find also standard deviation.

Solution: Total probability=1

$$\therefore \int_0^\infty ke^{-x}dx = 1$$

$$k\left[\frac{e^{-x}}{-1}\right]_0^\infty = 1 \qquad k = 1$$

$$M_X(t) = E\left[e^{tx}\right] = \int_0^\infty e^{tx}e^{-x}dx = \int_0^\infty e^{(t-1)x}dx$$

$$= \left[\frac{e^{(t-1)x}}{t-1}\right]_0^\infty = \frac{1}{1-t}, t < 1$$

$$= (1-t)^{-1} = 1 + t + t^2 + \ldots + t^r + \ldots\infty$$

$$\mu_r' = coeff.\ of\ \frac{t^2}{r!} = r!$$

When $r = 1$, $\mu_1' = 1! = 1$

$r = 2, \mu_2' = 2! = 2$

Variance $= \mu_2' - \mu_1' = 2 - 1 = 1$

$\therefore$ Standard deviation = 1.

Problem 27 Find the moment generating function for the distribution whose p.d.f is $f(x) = \lambda e^{-\lambda x}$, $x > 0$ and hence find its mean and variance.

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

$$= \lambda\int_0^\infty e^{-x(\lambda - t)}dx = \lambda\left[\frac{e^{-x(\lambda-t)}}{-(\lambda - t)}\right]_0^\infty = \frac{\lambda}{\lambda - t}$$

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

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

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

Problem 28 Let the random variable X have the p.d.f $f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & ,x>0\\ 0 & ,otherwise.\end{cases}$

Find the moment generating function, mean & variance of X .

Solution:

$$M_X(t)=E\left(e^{tx}\right)=\int_{-\infty}^{\infty}e^{tx}f(x)\,dx=\int_{0}^{\infty}e^{tx}\frac{1}{2}e^{-x/2}dx$$

$$=\frac{1}{2}\int_{0}^{\infty}e^{\left(t-\frac{1}{2}\right)x}dx=\frac{1}{2}\left[\frac{e^{\left(t-\frac{1}{2}\right)x}}{\left(t-\frac{1}{2}\right)}\right]_0^{\infty}=\frac{1}{1-2t},\ if\ t<\frac{1}{2}.$$

$$E(X)=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left[\frac{2}{(1-2t)^2}\right]_{t=0}=2$$

$$E\left(X^2\right)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left[\frac{8}{(1-2t)^3}\right]_{t=0}=8$$

$$Var(X)=E\left(X^2\right)-\left[E(X)\right]^2=8-4=4\,.$$

Problem 29 a) Define Binomial distribution Obtain its m.g.f., mean and variance.

b) Six dice are thrown 729 times. How many times do you expect at least 3 dice show 5 or 6 ?

Solution:a) A random variable X is said to follow binomial distribution if it assumes only non- negative values and its probability mass function is given by $P(X=x)=nC_xp^xq^{n-x},\ x=0,1,2,...,n\ and\ q=1-p.$

M.G.F of Binomial Distribution about origin is

$$M_X(t)=E\left[e^{tx}\right]=\sum_{x=0}^{n}e^{tx}P(X=x)$$

$$=\sum_{x=0}^{n}nC_xx\,P^xq^{n-x}e^{tx}$$

$$=\sum_{x=0}^{n}nC_x\left(pe^t\right)^xq^{n-x}$$

$$M_X(t)=\left(q+pe^t\right)^n$$

Mean of Binomial distribution

$$\text{Mean}=E(X)=M_X{}'(0)$$

$$=\left[n\left(q+pe^t\right)^{n-1}pe^t\right]_{t=0}=np \text{ Since } q+p=1$$

$$E\left(X^2\right)=M_X''(0)$$

$$=\left[n(n-1)\left(q+pe^t\right)^{n-2}\left(pe^t\right)^2+npe^t\left(q+pe^t\right)^{n-1}\right]_{t=0}$$

$$E\left(X^2\right)=n(n-1)p^2+np$$

$$=n^2p^2+np(1-p)=n^2p^2+npq$$

$$\text{Variance}=E\left(X^2\right)-\left[E[X]\right]^2=npq$$

$$Mean=np\,;Variance=npq$$

b) Let X :the number of times the dice shown 5 or 6

$$P[5\;or\;6]=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$$

$$\therefore P=\frac{1}{3} \text{ and } q=\frac{2}{3}$$

Here $n=6$

To evaluate the frequency of $X\geq 3$

By Binomial theorem,

$$P[X=r]=6C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{6-r} \text{ where } r=0,1,2...6.$$

$$P[X\geq 3]=P(3)+P(4)+P(5)+P(6)$$

$$=6C_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3+6C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2+6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)+6C_6\left(\frac{1}{3}\right)^6$$

$$=0.3196$$

$\therefore$ Expected number of times at least 3 dies to show 5 or 6 $=N\times P[X\geq 3]$

$$=729\times 0.3196=233.$$

Problem 30 a) Find the m.g.f. of the geometric distribution and hence find its mean and variance.

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

Solution:a) M.G.F about origin $=M_X(t)=E\left[e^{tx}\right]$

$$= \sum_{x=0}^{\infty} e^{tx} \quad \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!} \quad = e^{-\lambda} e^{\lambda e^t}$$

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

Mean and variance using M.G.F

$$\text{Mean} = E(X) = M_X{}'(0)$$

$$= \left[e^{\lambda\left(e^t-1\right)} \lambda e^t \right]_{t=0} = \lambda$$

$$E\left(X^2\right) = M_X{}''(0) = \left[\left(\lambda e^t\right)^2 e^{\lambda\left(e^t-1\right)} + e^{\lambda\left(e^t-1\right)} \lambda e^t \right]_{t=0}$$

$$= \lambda^2 + \lambda$$

$\therefore$ Variance $= E\left(x^2\right) - \left[E(X)\right]^2 = \lambda$

b) Probability of getting one head with one coin $= \frac{1}{2}$.

$\therefore$ The probability of getting six heads with six coins $\left(\frac{1}{2}\right)^6 = \frac{1}{64}$

$\therefore$ Average number of 6 heads with six coins in 6400 throws $= np = 6400 \times \frac{1}{64} = 100$

$\therefore$ Mean of the Poisson distribution $= \lambda = 100$

By Poisson distribution, the approximate probability of getting six heads x times is given by

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(100)^x e^{-100}}{x!}, \; x = 0, 1, 2, ...$$

Problem 31 a) A die is cast until 6 appears. What is the probability that it must cast more than five times?

b) Suppose that a trainee soldier shoots a target in an independent fashion. If the probability that the target is shot on any one shot is 0.8.

(i) What is the probability that the target would be hit on 6th attempt?

(ii) What is the probability that it takes him less than 5 shots?

Solution: Probability of getting six $= \frac{1}{6}$

$$\therefore p = \frac{1}{6} \;\; \& \;\; q = 1 - \frac{1}{6}$$

Let $x=$ Number of throws for getting the number 6. By geometric distribution $P[X=x]=q^{x-1}p, x=1,2,3....$

Since 6 can be got either in first, second……throws.

To find $P[X>5]=1-P[X\le 5]$

$$=1-\sum_{x=1}^{5}\left(\frac{5}{6}\right)^{x-1}.\frac{1}{6}$$

$$=1-\left[\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)\right]$$

$$=1-\frac{\frac{1}{6}\left[1-\left(\frac{5}{6}\right)^5\right]}{1-\frac{5}{6}}=\left(\frac{5}{6}\right)^5=0.4019$$

b) Here $p=0.8,\ q=1-p=0.2$

$$P[X=r]=q^{r-1}p, r=0,1,2...$$

(i) The probability that the target would be hit on the 6th attempt $=P[X=6]$

$$=(0.2)^5(0.8)=0.00026$$

(ii) The probability that it takes him less than 5 shots

$$=P[X<5]$$

$$=\sum_{r=1}^{4}q^{r-1}p=0.8\sum_{r=1}^{4}(0.2)^{r-1}$$

$$=0.8[1+0.2+0.04+0.008]=0.9984$$

Problem 32 a) If X_1, X_2 are two independent random variables each flowing negative binomial distribution with parameters (r_1, p) and (r_2, p), show that the sum also follows negative binomial distribution

b) If a boy is throwing stones at a target, what is the probability that his 10th throw is his 5th hit, if the probability of hitting the target at any trial is 0.5?

Solution:a) Let X_1 be a negative binomial variate with (r_1, p) and X_2 be another negative binomial variate wit (r_2, p) and let them be independent.

Then $M_{X_1}(t)=(q-pe^t)^{-r_1}$

$$M_{X_2}(t)=(q-pe^t)^{-r_2}$$

Then $M_{X_1+X_2}(t)=M_{X_1}(t)M_{X_2}(t)$

$=\left(q-pe^t\right)^{-(r_1+r_2)}$, which is the m.g.f of a negative binomial variable with r_1+r_2 as parameter. This proves the result.

b) Since the 10th throw should result in the 5th success

(i.e.) the 5th hit, the first 9 throws should have resulted in 4 successes and 5 failures. Hence we have $x=5, r=5, p=q=\frac{1}{2}$

$\therefore$ Required probability $=P[X=5]$

By N.B.D, $P[X=x]=\begin{pmatrix} x+r-1 \\ x \end{pmatrix} p^r q^x$

$$=\begin{pmatrix} 5+5-1 \\ 5 \end{pmatrix} p^5 q^5 = 9C_4\left(\frac{1}{2}\right)^{10} = 0.123.$$

Problem 33 a) State and prove the memoryless property of exponential distribution.

b) 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.

Solution: a) Statement: If X is exponentially distributed with parameters λ, then for any two positive integers s and t, $P[x>s+t/x>s]=P[x>t]$

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

$$\therefore P[X>t]=\int_t^{\infty}\lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_t^{\infty}=e^{-\lambda t}$$

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

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

$$=P[x>t]$$

b) Let X denote the time to failure of the component then X has exponential distribution with $Mean=1000$ hours.

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

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

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

$$= \frac{P[10,000 < X < 15,000]}{P[X > 10,000]}$$

$$= \frac{\int_{10,000}^{15,000} f(x)\,dx}{\int_{10,000}^{\infty} f(x)\,dx} = \frac{e^{-1} - e^{-1.5}}{e^{-1}}$$

$$= \frac{0.3679 - 0.2231}{0.3679} = 0.3936.$$

(ii) Probability that the component will operate for another 5000 hours given that

it is in operation 15,000 hours $= P[X > 20,000 / X > 15,000]$

$$= P[x > 5000] \text{ [By memoryless property]}$$

$$= \int_{5000}^{\infty} f(x)\,dx = e^{-0.5} = 0.6065$$

Problem 34 The Daily consumption of milk in a city in excess of 20,000 gallons is approximately distributed as a Gamma variate with parameters $\alpha = 2$ and $\lambda = \frac{1}{10,000}$. The city has a daily stock of 30,000 gallons. What is the probability that the stock is insufficient on a particular day?

Let X be the r.v denoting the daily consumption of milk (is gallons) in a city

Then $Y = X - 20,000$ has Gamma distribution with p.d.f.

$$f(y) = \frac{1}{(10,000)^2 \Gamma(2)} y^{2-1} e^{-\frac{y}{10,000}}, y \geq 0$$

$$f(y) = \frac{y e^{-\frac{y}{10,000}}}{(10,000)^2}, y \geq 0.$$

$\therefore$ The daily stock of the city is 30,000 gallons; the required probability that the stock is insufficient on a particular day is given by

$$P[X > 30,000] = P[Y > 10,000]$$

$$= \int_{10,000}^{\infty} g(y)\,dy = \int_{10,000}^{\infty} \frac{ye^{-\frac{y}{10,000}}}{(10,000)^2}dy$$

Put $Z = \frac{y}{10,000}$, then $dz = \frac{dy}{10,000}$

$$\therefore P[X > 30,000] = \int_{1}^{\infty} ze^{-z}dz \quad = \left[-ze^{-z} - e^{-z}\right]_1^{\infty} = \frac{2}{e}$$

Problem 37 a) The amount of time that a camera will run without having to be reset is a random variable having exponential distribution with $\theta = 50$ days. Find the probability that such a camera will (i) have to be reset in less than 20 days (ii) not has to be reset in at least 60 days.

b) Subway trains on a certain line run every half an hour between midnight and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least 20 minutes.

Solution:a) Let X be the time that the camera will run without having to be reset.

The X is a random variable with exponentially distributed with

$\theta = 50$ days. The p.d.f of X is given by

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, x \geq 0 \\ 0 \qquad x < 0 \end{cases}$$

(i) P[Camera will have to be reset in less than 20 days]

$$= P[Camera\ will\ run\ for\ less\ than\ 20\ days]$$

$$= P[X < 20] = \int_{0}^{20} \frac{1}{50}e^{-\frac{x}{50}}dx = 1 - e^{-\frac{2}{5}} = 0.3297$$

(ii) P[Camera will not have to be reset in at least 60 days]

$$= P[camera\ will\ run\ for\ atleast\ 60 days]$$

$$= P[X > 60] = \int_{60}^{\infty} \frac{1}{50}e^{-\frac{x}{50}}dx = e^{-\frac{6}{5}} = 0.3012$$

b) Let X denote the waiting time in minutes for the next train. Under the assumption that a man arrives at the station at random time, X is uniformly distributed on (0, 30) with p.d.f. $f(x) = \begin{cases} \frac{1}{30}, 0 < x < 30 \\ 0 \quad , Otherwise \end{cases}$

The probability that he has to wait at least 20 minutes

$$= P[X \geq 20] = \int_{20}^{30} f(x)\,dx = \frac{1}{30}(30 - 20) = \frac{1}{3}$$

UNIT– II - TWO DIMENSIONAL RANDOM VARIABLES

PART -A

Problem 1 Let X and Y have joint density function $f(x,y)=2,\ 0<x<y<1$.Find the marginal density functions and the conditional density function Y given $X=x$.

Solution: Marginal density function of X is given by

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

$$=\int_x^1 f(x,y)\,dy=\int_x^1 2dy=2(y)_x^1$$

$$=2(1-x),0<x<1.$$

Marginal density function of Y is given by

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

$$=\int_0^y 2dx=2y,0<y<1.$$

Conditional distribution function of Y given $X=x$ is $f\left({}^{y}\!/\!_{x}\right)=\dfrac{f(x,y)}{f(x)}=\dfrac{2}{2(1-x)}=\dfrac{1}{1-x}$.

Problem 2 Two random variables X and Y have the joint p.d.f $f(x,y)=Ae^{-(2x+y)},\ x,y\geq 0$. Find A.

Solution: Since $f(x,y)$ is a joint density function

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy=1.$$

$$\Rightarrow \int_0^{\infty}\int_0^{\infty} Ae^{-(2x+y)}dxdy=1$$

$$\Rightarrow A\int_0^{\infty} e^{-y}\left(\frac{e^{-2x}}{-2}\right)_0^{\infty} dy=1$$

$$\Rightarrow A\int_0^{\infty}\frac{1}{2}e^{-y}dy=1$$

$$\Rightarrow A\frac{1}{2}\left[\frac{e^{-y}}{-1}\right]_0^{\infty}=1$$

$$\Rightarrow A\frac{1}{2}[1]=1\Rightarrow A=2$$

Problem 3 Verify whether X and Y are independent if $f(x,y)=kxy,\ 0\le x\le y, 0\le y\le 4$

Solution: Since $f(x,y)$ is a joint density function

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy=1.$$

$$\Rightarrow \int_0^4\int_0^y kxydxdy=1$$

$$\Rightarrow k\int_0^4 y\left(\frac{x^2}{2}\right)_0^y dy=1$$

$$\Rightarrow k\int_0^4 \frac{1}{2}y^3dy=1\Rightarrow k\frac{1}{8}\left[y^4\right]_0^4=1\Rightarrow k[32]=1\Rightarrow k=\frac{1}{32}$$

Marginal density function of X is given by

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

$$=\int_0^4 \frac{1}{32}xydy=\frac{1}{32}x\left[\frac{y^2}{2}\right]_0^4=\frac{x}{4},\quad 0\le x\le y$$

Marginal density function of Y is given by

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

$$=\int_0^y \frac{1}{32}xydx=\frac{1}{32}y\left[\frac{x^2}{2}\right]_0^y=\frac{y^3}{64},\quad 0\le y\le 4$$

$$f(x).f(y)=\frac{x}{4}.\frac{y^3}{64}\ne f(x,y)$$

$\therefore$ X and Y are not independent.

Problem 4 Two random variables X and Y have the joint p.d.f $f(x,y)=x+y,\quad 0\le x\le 1, 0\le y\le 1.$ Determine the marginal distributions of X and Y.

Solution: Marginal density function of X is given by

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

$$=\int_0^1 (x+y)dy=\left[xy+\frac{y^2}{2}\right]_{y=0}^1=x+\frac{1}{2},\quad 0\le x\le 1$$

Marginal density function of Y is given by

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

$$= \int_0^1 (x+y)dx = \left[\frac{x^2}{2} + xy\right]_{x=0}^1 = \frac{1}{2} + y, \quad 0 \le y \le 1$$

Problem 5 Suppose that the joint density function of X and Y is

$$f(x,y) = \begin{cases} Ae^{-x-y}, \; 0 \le x \le y, \; 0 \le y \le \infty \\ 0 \qquad , \; otherwise \end{cases}$$ Determine A.

Solution: Since $f(x,y)$ is a joint density function

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1.$$

$$\Rightarrow \int_0^{\infty}\int_0^{y} Ae^{-x}e^{-y}dxdy = 1$$

$$\Rightarrow A\int_0^{\infty} e^{-y}\left(\frac{e^{-x}}{-1}\right)_0^y dy = 1$$

$$\Rightarrow A\int_0^{\infty}\left[e^{-y} - e^{-2y}\right]dy = 1$$

$$\Rightarrow A\left[\frac{e^{-y}}{-1} - \frac{e^{-2y}}{-2}\right]_0^{\infty} = 1$$

$$\Rightarrow A\left[\frac{1}{2}\right] = 1 \Rightarrow A = 2$$

Problem 6 Examine whether the variables X and Y are independent, whose joint density function is $f(x,y) = xe^{-x(y+1)}, \; 0 < x, y < \infty$.

Solution: The marginal probability function of X is

$$f_X(x) = f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{\infty} xe^{-x(y+1)}dy$$

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

The marginal probability function of Y is

$$f_Y(y) = f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} x\,e^{-x(y+1)}dx$$

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

$$= \frac{1}{(y+1)^2}$$

Here $f(x).f(y) = e^{-x} \times \frac{1}{(1+y)^2} \neq f(x,y)$

$\therefore X$ and Y are not independent.

Problem 7 If X has an exponential distribution with parameter 1. Find the pdf of $y = \sqrt{x}$

<u>Solution:</u>Since $y = \sqrt{x}$, $x = y^2$

Since X has an exponential distribution with parameter 1, the pdf of X is given by

$$f_X(x) = e^{-x},\ x > 0 \qquad \left[\because f(x) = \lambda e^{-\lambda x}, \lambda = 1\right]$$

$$\therefore f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

$$= e^{-x} 2y = 2ye^{-y^2}$$

$$f_Y(y) = 2ye^{-y^2},\ \ y > 0$$

Problem 8 If X is uniformly distributed random variable in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, Find the probability density function of $Y = tanX$.

Solution: Given $Y = tanX \Rightarrow x = tan^{-1} y$

$$\therefore \frac{dx}{dy} = \frac{1}{1+y^2}$$

Since X is uniformly distribution in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$,

$$f_X(x) = \frac{1}{b-a} = \frac{1}{\frac{\pi}{2} + \frac{\pi}{2}}$$

$$f_X(x) = \frac{1}{\pi},\ -\frac{\pi}{2} < x < \frac{\pi}{2}$$

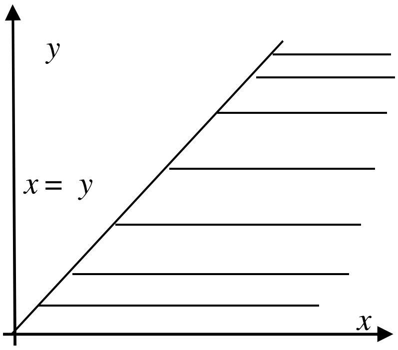
Now $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{\pi}\left(\frac{1}{1+y^2}\right), -\infty < y < \infty \therefore f_Y(y) = \frac{1}{\pi(1+y^2)}, -\infty < y < \infty$

Problem 9 If the Joint probability density function of (x, y) is given by $f(x, y) = 24y(1-x),\ 0 \le y \le x \le 1$ Find $E(XY)$.

Solution: $E(xy) = \int_0^1 \int_y^1 xyf(x, y)\,dxdy$

$$= 24\int_0^1 \int_y^1 xy^2(1-x)\,dxdy$$

$$= 24\int_0^1 y^2\left[\frac{1}{6} - \frac{y^2}{2} + \frac{y^3}{3}\right]dy = \frac{4}{15}.$$



Problem 10 If X and Y are random Variables, Prove that $Cov(X,Y) = E(XY) - E(X)E(Y)$

Solution: $cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]$

$$= E\left(XY - \bar{X}Y - \bar{Y}X + \bar{X}\bar{Y}\right)$$

$$= E(XY) - \bar{X}E(Y) - \bar{Y}E(X) + \bar{X}\bar{Y}$$

$$= E(XY) - \bar{X}\bar{Y} - \bar{X}\bar{Y} + \bar{X}\bar{Y}$$

$$= E(XY) - E(X)E(Y) \qquad \left[\because E(X) = \bar{X}, E(Y) = \bar{Y}\right]$$

Problem 11 If X and Y are independent random variables prove that $cov(x, y) = 0$

Solution: $cov(x, y) = E(xy) - E(x)E(y)$

But if X and Y are independent then $E(xy) = E(x)E(y)$

$cov(x, y) = E(x)E(y) - E(x)E(y)$ Therefore $cov(x, y) = 0.$

Problem 12 Write any two properties of regression coefficients.

Solution: 1. Correction coefficients is the geometric mean of regression coefficients

2. If one of the regression coefficients is greater than unity then the other should be less than 1.

$b_{xy} = r\frac{\sigma_y}{\sigma_x}$ and $b_{yx} = r\frac{\sigma_x}{\sigma_y}$

If $b_{xy} > 1$ then $b_{yx} < 1$

Problem 13 Write the angle between the regression lines.

Solution: The slopes of the regression lines are

$$m_1 = r\frac{\sigma_y}{\sigma_x}, m_2 = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$$

If θ is the angle between the lines, Then

$$\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-r^2}{r}\right]$$

When $r=0$, that is when there is no correlation between x and y, $\tan\theta=\infty$ (or) $\theta=\frac{\pi}{2}$

and so the regression lines are perpendicular

When $r=1$ or $r=-1$, that is when there is a perfect correlation $+ve$ or $-ve$, $\theta=0$ and so the lines coincide.

Problem 14 State central limit theorem

Solution: If $X_1, X_2....X_n$ is a sequence of independent random variable $E(X_i)=\mu_i$ and $Var(X_i)=\sigma_i^2$, $i=1,2,....n$ and if $S_n = X_1+X_2+......+X_n$ then under several conditions S_n follows a normal distribution with mean $\mu=\sum_{i=1}^{n}\mu_i$ and variance $\sigma^2=\sum_{i=1}^{n}\sigma_i^2$ as $n\to\infty$.

Problem 15 Fill up the blanks:

i). Two random variables are said to be orthogonal if correction is ______zero

ii). If $X=Y$ then correlation coefficients between them is _____1

16. ***If X has an expontial distribution with parameter 1, find the pdf of $Y=\sqrt{X}$.***

Given X follows E.D with parameter 1

we know that $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$

$f_Y(y) = 2ye^{-y^2}, y>0$

17. ***Write any two properties of the joint cdf.***

(i) $F(-\infty,-\infty)=0$ (ii) $F(\infty,\infty)=1$

18. ***The joint pmf of (X,Y) is given as $P(1,1)=\frac{1}{8}$, $P(1,2)=\frac{1}{4}$, $P(2,1)=\frac{1}{8}$, $P(2,2)=\frac{1}{2}$. Find $P(X+Y>2)$.***

Given

X |Y	1	2
1	1/8	1/4
2	1/8	1/2

$$P(X+Y>2) = P(1,2)+(2,1)+ P(2,2) = \frac{1}{4}+\frac{1}{8}+\frac{1}{2}=\frac{7}{8}.$$

4. ***Find the value of k if $f(x,y)=k(1-x)(1-y)$ for $0<x,y<1$.***

WKT $\int_{-\propto}^{\propto}\int_{-\propto}^{\propto} f(x,y)dxdy = 1$

$$\Rightarrow \int_0^1 \int_0^1 \mathrm{k}(1-\mathrm{x})(1-\mathrm{y})\mathrm{dxdy} = 1$$

$$\Rightarrow \mathrm{k}\left(\int_0^1 (1-\mathrm{x})\mathrm{dx}\right)\left(\int_0^1 (1-\mathrm{y})\mathrm{dy}\right) = 1$$

$$\Rightarrow \mathrm{k}(1/2)(1/2) = 1$$

$$\Rightarrow \mathrm{k} = 4$$

5. $\boldsymbol{The\ joint\ pdf\ of\ (X,Y)\ is\ given\ by\ f(x,y) = \begin{cases} Kxe^{-y}, 0 \le x \le 2, y > 0 \\ 0\ ,\ otherwise \end{cases}}$.
$\boldsymbol{Find\ the\ value\ of\ K.}$

$$\text{WKT } \int_{-\propto}^{\propto}\int_{-\propto}^{\propto} f(x,y)dxdy = 1$$

$$\Rightarrow \int_0^{\propto}\int_0^2 kx\, e^{-y} dxdy = 1$$

$$\Rightarrow k\left(\int_0^2 x\, dx\right)\left(\int_0^{\propto} e^{-y} dy\right) = 1$$

$$\Rightarrow k(4/2)(1) = 1 \qquad \langle \because \int_0^{\propto} e^{-x^2} dx = 1/2$$

$$\Rightarrow k = \frac{1}{2}$$

6. $\boldsymbol{State\ the\ cental\ limit\ theorem\ for\ i.i.d\ randon\ variables.}$
If $\mathrm{X_1, X_2, \ldots\ X_n}$ be the sequence of i. i. d random variables with $\mathrm{E(X_i) = \mu}$ and $\mathrm{Var((X_i) = \sigma^2}$, $\mathrm{i = 1,2 \ldots n}$ and if $\mathrm{S_n = X_1 + X_2 +, \ldots + X_n}$, then under certain genral conditions, $\mathrm{S_n}$ follows a normal distribution withmean $\mathrm{n\mu}$ and $\mathrm{n\sigma^2}$ as $\mathrm{n \to \infty}$

PART - B

Problem 1 The joint probability density function of a bivariate random variable (X,Y) is $f_{XY}(x,y) = \begin{cases} k(x+y), & 0<x<2,\ 0<y<2 \\ 0 & , otherwise \end{cases}$ where $'k'$ is a constant.

i. Find k.
ii. Find the marginal density function of X and Y.
iii. Are X and Y independent?
iv. Find $f_{Y/X}\left(y/x\right)$ and $f_{X/Y}\left(x/y\right)$.

Solution: (i). Given the joint probability density function of a brivate random variable (X,Y) is

$$f_{XY}(x,y) = \begin{cases} K(x+y), & 0<x<2,\ 0<y<2 \\ 0 & , otherwise \end{cases}$$

Here $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1 \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(x+y)dxdy = 1$

$$\int_0^2\int_0^2 K(x+y)dxdy = 1 \Rightarrow K\int_0^2 \left[\frac{x^2}{2} + xy\right]_0^2 dy = 1$$

$$\Rightarrow K\int_0^2 (2+2y)\,dy = 1$$

$$\Rightarrow K\left[2y+y^2\right]_0^2 = 1$$

$$\Rightarrow K[8-0] = 1$$

$$\Rightarrow K = \frac{1}{8}$$

(ii). The marginal p.d.f of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{8}\int_0^2 (x+y)\,dy$$

$$= \frac{1}{8}\left[xy + \frac{y^2}{2}\right]_0^2 = \frac{1+x}{4}$$

$\therefore$ The marginal p.d.f of X is

$$f_X(x) = \begin{cases} \dfrac{x+1}{4}, & 0 < x < 2 \\ 0 & , otherwise \end{cases}$$

The marginal p.d.f of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{1}{8}\int_0^2 (x+y)\,dx$$

$$= \frac{1}{8}\left[\frac{x^2}{2} + yx\right]_0^2$$

$$= \frac{1}{8}[2+2y] = \frac{y+1}{4}$$

$\therefore$ The marginal p.d.f of Y is

$$f_Y(y) = \begin{cases} \dfrac{y+1}{4}, & 0 < y < 2 \\ 0 & , otherwise \end{cases}$$

(iii). To check whether X and Y are independent or not.

$$f_X(x)\,f_Y(y) = \frac{(x+1)}{4}\frac{(y+1)}{4} \neq f_{XY}(x,y)$$

Hence X and Y are not independent.

(iv). Conditional p.d.f $f_{Y/X}\left(y/x\right)$ is given by

$$f_{Y/X}\left(y/x\right)=\frac{f(x,y)}{f_X(x)}=\frac{\frac{1}{8}(x+y)}{\frac{1}{4}(x+1)}=\frac{1}{2}\frac{(x+y)}{(x+1)}$$

$$f_{Y/X}\left(y/x\right)=\frac{1}{2}\left(\frac{x+y}{x+1}\right),\ 0<x<2,\ 0<y<2$$

(v) $P\left(0<y<\frac{1}{2}\Big/x=1\right)=\int_0^2 f_{Y/X}\left(y/x=1\right)dy$

$$=\frac{1}{2}\int_0^{\frac{1}{2}}\frac{1+y}{2}dy=\frac{5}{32}.$$

Problem 2 a) If X and Y are two random variables having joint probability density function $f(x,y)=\begin{cases}\frac{1}{8}(6-x-y), & 0<x<2,\ 2<y<4\\ 0 & ,\ otherwise\end{cases}$ Find (i) $P(X<1\cap Y<3)$

(ii) $P(X+Y<3)$ (iii) $P\left(X<1/Y<3\right)$.

b). Three balls are drawn at random without replacement from a box containing 2 white, 3 red and 4 black balls. If X denotes the number of white balls drawn and Y denotes the number of red balls drawn find the joint probability distribution of (X,Y).

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

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

$$=\frac{1}{8}\int_2^3\int_0^1(6-x-y)\,dxdy$$

$$=\frac{1}{8}\int_2^3\left[6x-\frac{x^2}{2}-xy\right]_0^1 dy$$

$$=\frac{1}{8}\int_2^3\left[\frac{11}{2}-y\right]dy=\frac{1}{8}\left[\frac{11y}{2}-\frac{y^2}{2}\right]_2^3$$

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

(ii). $P(X+Y<3)=\int_0^1\int_2^{3-x}\frac{1}{8}(6-x-y)\,dydx$

$$=\frac{1}{8}\int_0^1\left[6y-xy-\frac{y^2}{2}\right]_2^{3-x}dx$$

$$=\frac{1}{8}\int_0^1\left[6(3-x)-x(3-x)-\frac{(3-x)^2}{2}-[12-2x-2]\right]dx$$

$$=\frac{1}{8}\int_0^1\left[18-6x-3x+x^2-\frac{(9+x^2-6x)}{2}-(10-2x)\right]dx$$

$$=\frac{1}{8}\int_0^1\left[18-9x+x^2-\frac{9}{2}-\frac{x^2}{2}+\frac{6x}{2}-10+2x\right]dx$$

$$=\frac{1}{8}\int_0^1\left[\frac{7}{2}-4x+\frac{x^2}{2}\right]dx$$

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

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

(iii). $P\left(X<1/_{Y<3}\right)=\dfrac{P(x<1\cap y<3)}{P(y<3)}$

The Marginal density function of Y is $f_Y(y)=\int_0^2 f(x,y)dx$

$$=\int_0^2\frac{1}{8}(6-x-y)dx$$

$$=\frac{1}{8}\left[6x-\frac{x^2}{2}-yx\right]_0^2$$

$$=\frac{1}{8}[12-2-2y]$$

$$=\frac{5-y}{4},\ 2<y<4.$$

$$P\left(X<1/_{Y<3}\right)=\frac{\int_{x=0}^{x=1}\int_{y=2}^{y=3}\frac{1}{8}(6-x-y)\,dxdy}{\int_{y=2}^{y=3}f_Y(y)\,dy}$$

$$= \frac{\frac{3}{8}}{\int_2^3 \left(\frac{5-y}{4}\right) dy} = \frac{\frac{3}{8}}{\frac{1}{4}\left[5y - \frac{y^2}{2}\right]_2^3}$$

$$= \frac{3}{8} \times \frac{8}{5} = \frac{3}{5}.$$

b). Let X takes 0, 1, 2 and Y takes 0, 1, 2 and 3.

$P(X = 0, Y = 0) = P($ drawing 3 balls none of which is white or red)

$= P($ all the 3 balls drawn are black)

$$= \frac{4C_3}{9C_3} = \frac{4 \times 3 \times 2 \times 1}{9 \times 8 \times 7} = \frac{1}{21}.$$

$P(X = 0, Y = 1) = P($ drawing 1 red ball and 2 black balls)

$$= \frac{3C_1 \times 4C_2}{9C_3} = \frac{3}{14}$$

$P(X = 0, Y = 2) = P($ drawing 2 red balls and 1 black ball)

$$= \frac{3C_2 \times 4C_1}{9C_3} = \frac{3 \times 2 \times 4 \times 3}{9 \times 8 \times 7} = \frac{1}{7}.$$

$P(X = 0, Y = 3) = P($ all the three balls drawn are red and no white ball)

$$= \frac{3C_3}{9C_3} = \frac{1}{84}$$

$P(X = 1, Y = 0) = P($ drawing 1White and no red ball)

$$= \frac{2C_1 \times 4C_2}{9C_3} = \frac{\frac{2 \times 4 \times 3}{1 \times 2}}{\frac{9 \times 8 \times 7}{1 \times 2 \times 3}}$$

$$= \frac{12 \times 1 \times 2 \times 3}{9 \times 8 \times 7} = \frac{1}{7}.$$

$P(X = 1, Y = 1) = P($ drawing 1White and 1 red ball)

$$= \frac{2C_1 \times 3C_1}{9C_3} = \frac{2 \times 3}{\frac{9 \times 8 \times 7}{1 \times 2 \times 3}} = \frac{2}{7}$$

$P(X = 1, Y = 2) = P($ drawing 1White and 2 red ball)

$$= \frac{2C_1 \times 3C_2}{9C_3} = \frac{2\times3\times2}{\frac{9\times8\times7}{1\times2\times3}} = \frac{1}{14}$$

$P(X=1, Y=3) = 0$ (Since only three balls are drawn)

$P(X=2, Y=0) = P($ drawing 2 white balls and no red balls)

$$= \frac{2C_2 \times 4C_1}{9C_3} = \frac{1}{21}$$

$P(X=2, Y=1) = P($ drawing 2 white balls and no red balls)

$$= \frac{2C_2 \times 3C_1}{9C_3} = \frac{1}{28}$$

$P(X=2, Y=2) = 0$

$P(X=2, Y=3) = 0$

The joint probability distribution of (X, Y) may be represented as

Y / X	0	1	2	3
0	$\frac{1}{21}$	$\frac{3}{14}$	$\frac{1}{7}$	$\frac{1}{84}$
1	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{14}$	0
2	$\frac{1}{21}$	$\frac{1}{28}$	0	0

Problem 3 a) Two fair dice are tossed simultaneously. Let X denotes the number on the first die and Y denotes the number on the second die. Find the following probabilities.

(i) $P(X+Y)=8$, (ii) $P(X+Y \geq 8)$, (iii) $P(X=Y)$ and (iv) $P\left(X+Y=6 \middle/ Y=4\right)$.

b) The joint probability mass function of a bivariate discrete random variable (X, Y) in given by the table.

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

Find

i. The marginal probability mass function of X and Y.
ii. The conditional distribution of X given $Y=1$.
iii. $P(X+Y<4)$

Solution:a). Two fair dice are thrown simultaneously

$$S=\begin{Bmatrix}(1,1)(1,2)...(1,6)\\(2,1)(2,2)...(2,6)\\.\quad .\quad ...\quad .\\(6,1)(6,2)...(6,6)\end{Bmatrix},\ n(S)=36$$

Let X denotes the number on the first die and Y denotes the number on the second die.

Joint probability density function of (X,Y) is $P(X=x,Y=y)=\frac{1}{36}$ for

$x=1,2,3,4,5,6$ and $y=1,2,3,4,5,6$

(i) $X+Y=\{$ the events that the no is equal to 8 $\}$

$$=\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

$$P(X+Y=8)=P(X=2,Y=6)+P(X=3,Y=5)+P(X=4,Y=4)$$
$$+P(X=5,Y=3)+P(X=6,Y=2)$$

$$=\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{5}{36}$$

(ii) $P(X+Y\geq 8)$

$$X+Y=\begin{Bmatrix}(2,6)\\(3,5),(3,6)\\(4,4),(4,5),(4,6)\\(5,3),(5,4)(5,5),(5,6)\\(6,2),(6,3),(6,4),(6,5)(6,6)\end{Bmatrix}$$

$$\therefore P(X+Y\geq 8)=P(X+Y=8)+P(X+Y=9)+P(X+Y=10)$$
$$+P(X+Y=11)+P(X+Y=12)$$

$$=\frac{5}{36}+\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36}=\frac{15}{36}=\frac{5}{12}$$

(iii) $P(X=Y)$

$$P(X=Y)=P(X=1,Y=1)+P(X=2,Y=2)+......+P(X=6,Y=6)$$

$$=\frac{1}{36}+\frac{1}{36}+..........+\frac{1}{36}=\frac{6}{36}=\frac{1}{6}$$

(iv) $P\left(X+Y=6\big/_{Y=4}\right)=\dfrac{P(X+Y=6\cap Y=4)}{P(Y=4)}$

Now $P(X+Y=6\cap Y=4)=\dfrac{1}{36}$

$P(Y=4)=\dfrac{6}{36}$

$\therefore P\left(X+Y=6\big/_{Y=4}\right)=\dfrac{\frac{1}{36}}{\frac{6}{36}}=\dfrac{1}{6}.$

b). The joint probability mass function of (X,Y) is

$Y \backslash X$	1	2	3	Total
1	0.1	0.1	0.2	0.4
2	0.2	0.3	0.1	0.6
Total	0.3	0.4	0.3	1

From the definition of marginal probability function

$$P_X(x_i)=\sum_{y_j}P_{XY}(x_i,y_j)$$

When $X=1$,

$$P_X(x_i)=P_{XY}(1,1)+P_{XY}(1,2)$$

$$=0.1+0.2=0.3$$

When $X=2$,

$$P_X(x=2)=P_{XY}(2,1)+P_{XY}(2,2)$$

$$=0.2+0.3=0.4$$

When $X=3$,

$$P_X(x=3)=P_{XY}(3,1)+P_{XY}(3,2)$$

$$=0.2+0.1=0.3$$

$\therefore$ The marginal probability mass function of X is

$$P_X(x)=\begin{cases}0.3 & when\ x=1\\ 0.4 & when\ x=2\\ 0.3 & when\ x=3\end{cases}$$

The marginal probability mass function of Y is given by $P_Y\left(y_j\right)=\sum_{x_i} P_{XY}\left(x_i, y_j\right)$

When $Y = 1$, $P_Y\left(y=1\right)=\sum_{x_i=1}^{3} P_{XY}\left(x_i,1\right)$

$$=P_{XY}(1,1)+P_{XY}(2,1)+P_{XY}(3,1)$$

$$=0.1+0.1+0.2=0.4$$

When $Y = 2$, $P_Y\left(y=2\right)=\sum_{x_i=1}^{3} P_{XY}\left(x_i,2\right)$

$$=P_{XY}(1,2)+P_{XY}(2,2)+P_{XY}(3,2)$$

$$=0.2+0.3+0.1=0.6$$

$\therefore$ Marginal probability mass function of Y is

$$P_Y(y)=\begin{cases}0.4 & when\ \ y=1\\ 0.6 & when\ \ y=2\end{cases}$$

(ii) The conditional distribution of X given $Y = 1$ is given by

$$P\left(X=x\big/_{Y=1}\right)=\frac{P\left(X=x\cap Y=1\right)}{P\left(Y=1\right)}$$

From the probability mass function of Y, $P(y = 1) = P_y(1) = 0.4$

When $X = 1$, $P\left(X=1\big/_{Y=1}\right)=\frac{P\left(X=1\cap Y=1\right)}{P\left(Y=1\right)}$

$$=\frac{P_{XY}(1,1)}{P_Y(1)}=\frac{0.1}{0.4}=0.25$$

When $X = 2$, $P\left(X=2\big/_{Y=1}\right)=\frac{P_{XY}(2,1)}{P_Y(1)}=\frac{0.1}{0.4}=0.25$

When $X = 3$, $P\left(X=3\big/_{Y=1}\right)=\frac{P_{XY}(3,1)}{P_Y(1)}=\frac{0.2}{0.4}=0.5$

(iii). $P\left(X+Y<4\right)=P\left\{(x,y)/x+y<4\ \ Where\ \ x=1,2,3;\ \ y=1,2\right\}$

$$=P\left\{(1,1),(1,2),(2,1)\right\}$$

$$=P_{XY}(1,1)+P_{XY}(1,2)+P_{XY}(2,1)$$

$$=0.1+0.1+0.2=0.4$$

Problem 4 a) If X and Y are two random variables having the joint density function $f(x,y)=\frac{1}{27}(x+2y)$ where x and y can assume only integer values 0, 1 and 2, find the conditional distribution of Y for $X = x$.

b) The joint probability density function of (X,Y) is given by $f_{XY}(x,y)=\begin{cases} xy^2+\frac{x^2}{8}, & 0\le x\le 2, \quad 0\le y\le 1 \\ 0 & , \textit{otherwise} \end{cases}$.

Find (i) $P(X>1)$, (ii) $P(X<Y)$ and (iii) $P(X+Y \pounds 1)$

Solution:a). Given X and Y are two random variables having the joint density function

$$f(x,y)=\frac{1}{27}(x+2y)----(1)$$

Where $x=0,1,2$ and $y=0,1,2$

Then the joint probability distribution X and Y becomes as follows

Y X	0	1	2	$f_1(x)$
0	0	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{3}{27}$
1	$\frac{2}{27}$	$\frac{3}{27}$	$\frac{4}{27}$	$\frac{9}{27}$
2	$\frac{4}{27}$	$\frac{5}{27}$	$\frac{6}{27}$	$\frac{15}{27}$

The marginal probability distribution of X is given by $f_1(X)=\sum_j P(x,y)$ and is calculated in the above column of above table.

The conditional distribution of Y for X is given by $f_1\left(Y=y/X=x\right)=\frac{f(x,y)}{f_1(x)}$ and is obtained in the following table.

X Y	0	1	2
0	0	$\frac{1}{3}$	$\frac{2}{3}$
1	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$
2	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

$$P\left(Y=0/X=0\right)=\frac{P(X=0,Y=0)}{P(X=0)}=\frac{0}{\frac{6}{27}}=0$$

$$P\left(Y=1/X=0\right)=\frac{P(X=0,Y=1)}{P(X=0)}=\frac{\frac{2}{27}}{\frac{6}{27}}=\frac{1}{3}$$

$$P\left(Y=2/X=0\right)=\frac{P(X=0,Y=2)}{P(X=0)}=\frac{\frac{4}{27}}{\frac{6}{27}}=\frac{2}{3}$$

$$P\left(Y=0/X=1\right)=\frac{P(X=1,Y=0)}{P(X=1)}=\frac{\frac{1}{27}}{\frac{9}{27}}=\frac{1}{9}$$

$$P\left(Y=1/X=1\right)=\frac{P(X=1,Y=1)}{P(X=1)}=\frac{\frac{3}{27}}{\frac{9}{27}}=\frac{3}{9}=\frac{1}{3}$$

$$P\left(Y=2/X=1\right)=\frac{P(X=1,Y=2)}{P(X=1)}=\frac{\frac{5}{27}}{\frac{9}{27}}=\frac{5}{9}$$

$$P\left(Y=0/X=2\right)=\frac{P(X=2,Y=0)}{P(X=2)}=\frac{\frac{2}{27}}{\frac{12}{27}}=\frac{1}{6}$$

$$P\left(Y=1/X=2\right)=\frac{P(X=2,Y=1)}{P(X=2)}=\frac{\frac{4}{27}}{\frac{12}{27}}=\frac{1}{3}$$

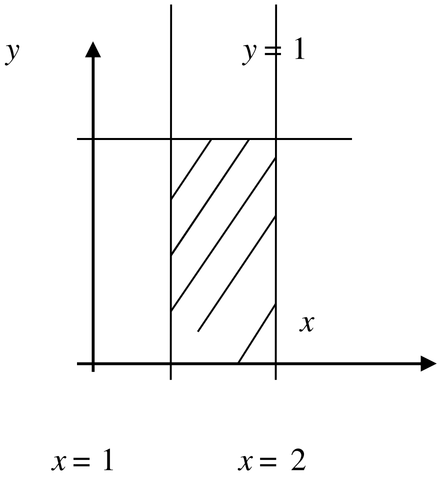
$$P\left(Y=2/X=2\right)=\frac{P(X=2,Y=2)}{P(X=2)}=\frac{\frac{6}{27}}{\frac{12}{27}}=\frac{1}{2}$$

b). Given the joint probability density function of (X,Y) is $f_{XY}\left(x+y\right)=xy^2+\frac{x^2}{8},\ \ 0\le x\le 2,\ 0\le y\le 1$

(i). $P\left(X>1\right)=\int_1^\infty f_X\left(x\right)dx$ The Marginal density function of X is $f_X\left(x\right)=\int_0^1 f\left(x,y\right)dy$

$$f_X(x) = \int_0^1 \left(xy^2 + \frac{x^2}{8} \right) dy$$

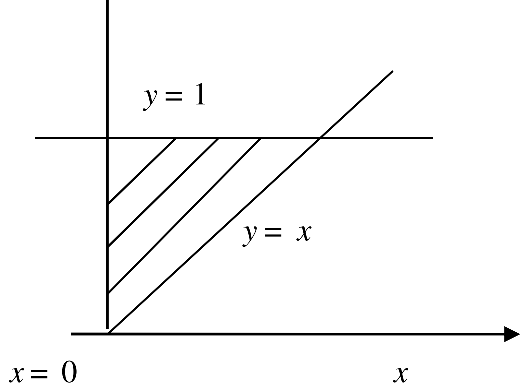
$$= \left[\frac{xy^2}{3} + \frac{x^2 y}{8} \right]_0^1 = \frac{x}{3} + \frac{x^2}{8},\ 1 < x < 2$$



$$P(X > 1) = \int_1^2 \left(\frac{x}{3} + \frac{x^2}{8} \right) dx$$

$$= \left[\frac{x^2}{6} + \frac{x^3}{24} \right]_1^2 = \frac{19}{24}.$$

(ii) $P(X < Y) = \int\limits_{R_2}\int f_{XY}(x, y)\, dxdy$



$$P(X < Y) = \int\limits_{y=0}^{1} \int\limits_{x=0}^{y} \left(xy^2 + \frac{x^2}{8} \right) dxdy$$

$$= \int_0^1 \left[\frac{x^2 y^2}{2} + \frac{x^3}{24} \right]_0^y dy$$

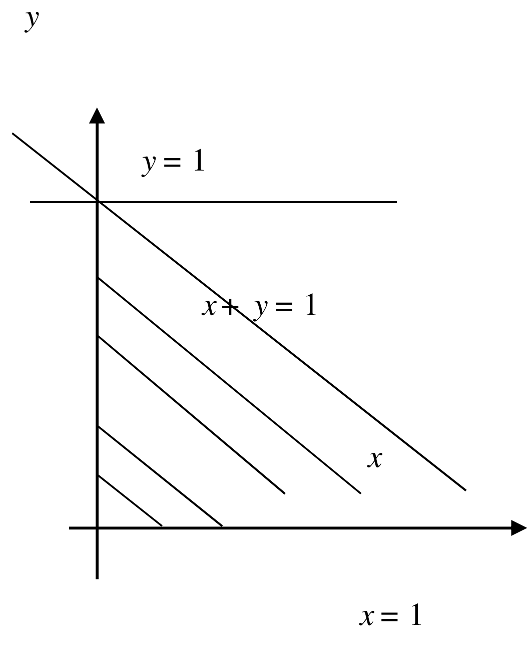
$$= \int_0^1 \left(\frac{y^4}{2} + \frac{y^3}{24} \right) dy = \left[\frac{y^5}{10} + \frac{y^4}{96} \right]_0^1$$

$$= \frac{1}{10} + \frac{1}{96} = \frac{96+10}{960} = \frac{53}{480}$$

(iii) $P(X + Y \le 1) = \iint\limits_{R_3} f_{XY}(x, y)\, dxdy$

Where R_3 is the region

$$P(X + Y \le 1) = \int\limits_{y=0}^{1} \int\limits_{x=0}^{1-y} \left(xy^2 + \frac{x^2}{8} \right) dxdy$$



$$= \int\limits_{y=0}^{1} \left[\left(\frac{x^2 y^2}{2} + \frac{x^3}{24} \right) \right]_0^{1-y} dy$$

$$= \int\limits_{y=0}^{1} \left(\frac{(1-y)^2 y^2}{2} + \frac{(1-y)^3}{24} \right) dy$$

$$= \int_0^1 \left(\frac{(1 + y^2 - 2y) y^2}{2} + \frac{(1-y)^3}{24} \right) dy$$

$$= \left[\left[\frac{y^3}{3} + \frac{y^5}{5} - \frac{2y^2}{4} \right] \frac{1}{2} + \frac{(1-y)^4}{96} \right]_0^1$$

$$=\frac{1}{6}+\frac{1}{10}-\frac{1}{4}+\frac{1}{96}=\frac{13}{480}.$$

Problem 5 a) If the joint distribution functions of X and Y is given by $F(x,y)=\begin{cases}(1-e^{x})(1-e^{-y}), & x>0,\ y>0\\ 0 & , otherwise\end{cases}$

i. Find the marginal density of X and Y.
ii. Are X and Y independent.
iii. $P(1< X < 3,\ 1< Y < 2)$.

b) The joint probability distribution of X and Y is given by $f(x,y)=\begin{cases}\frac{6-x-y}{8}, & 0<x<2,\ 2<y<4\\ 0 & , otherwise\end{cases}$.

Find $P\left(1< Y < 3 \middle/ X = 2\right)$.

Solution:a). Given $F(x,y)=(1-e^{-x})(1-e^{-y})$

$$=1-e^{-x}-e^{-y}+e^{-(x+y)}$$

The joint probability density function is given by

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}$$

$$=\frac{\partial^2}{\partial x\partial y}\left[1-e^{-x}-e^{-y}+e^{-(x+y)}\right]$$

$$=\frac{\partial}{\partial x}\left[e^{-y}-e^{-(x+y)}\right]$$

$$\therefore f(x,y)=\begin{cases}e^{-(x+y)}, & x\geq 0,\ y\geq 0\\ 0 & , otherwise\end{cases}$$

(ii) The marginal probability function of X is given by

$$f(x)=f_X(x)$$

$$=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_{0}^{\infty} e^{-(x+y)}dy$$

$$=\left[\frac{e^{-(x+y)}}{-1}\right]_0^{\infty}$$

$$=\left[-e^{-(x+y)}\right]_0^{\infty}$$

$$=e^{-x},\ x>0$$

The marginal probability function of Y is

$$f(y) = f_Y(y)$$

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

$$= \int_{0}^{\infty} e^{-(x+y)}dx = \left[-e^{-(x+y)}\right]_0^{\infty}$$

$$= e^{-y},\ y > 0$$

$$\therefore f(x)f(y) = e^{-x}e^{-y} = e^{-(x+y)} = f(x,y)$$

$\therefore$ X and Y are independent.

(iii) $P(1 < X < 3, 1 < Y < 2) = P(1 < X < 3) \times P(1 < Y < 2)$ [Since X and Y are independent]

$$= \int_1^3 f(x)dx \times \int_1^2 f(y)\,dy.$$

$$= \int_1^3 e^{-x}dx \times \int_1^2 e^{-y}dy$$

$$= \left[\frac{e^{-x}}{-1}\right]_1^3 \times \left[\frac{e^{-y}}{-1}\right]_1^2$$

$$= \left(-e^{-3} + e^{-1}\right)\left(-e^{-2} + e^{-1}\right)$$

$$= e^{-5} - e^{-4} - e^{-3} + e^{-2}$$

b). $P\left(1 < Y < 3 \big/ X = 2\right) = \int_1^3 f\left({}^{y}\!\big/_{x=2}\right)dy$

$$f_X(x) = \int_2^4 f(x,y)\,dy$$

$$= \int_2^4 \left(\frac{6-x-y}{8}\right)dy$$

$$= \frac{1}{8}\left(6y - xy - \frac{y^2}{2}\right)_2^4$$

$$= \frac{1}{8}(16 - 4x - 10 + 2x)$$

$$f\left({}^{y}\!\big/_{x}\right) = \frac{f(x,y)}{f(x)} = \frac{\frac{6-x-y}{8}}{\frac{6-2x}{8}} = \frac{6-x-y}{6-2x}$$

$$P\left(1<Y<3 \big/ X=2\right)=\int_{1}^{3} f\left(y \big/ x=2\right) dy$$

$$=\int_{2}^{3}\left(\frac{4-y}{2}\right) dy$$

$$=\frac{1}{2}\left[4y-\frac{y^2}{2}\right]_2^3$$

$$=\frac{1}{2}\left[4y-\frac{y^2}{2}\right]_2^3=\frac{1}{2}\left[14-\frac{17}{2}\right]=\frac{11}{4}.$$

Problem 6 a) Two random variables X and Y have the following joint probability density function $f(x,y)=\begin{cases}2-x-y, & 0\le x\le 1,\ 0\le y\le 1\\ 0 & ,\ otherwise\end{cases}$. Find the marginal probability density function of X and Y. Also find the covariance between X and Y.

b) If $f(x,y)=\frac{6-x-y}{8}, 0\le x\le 2, 2\le y\le 4$ for a bivariate (X,Y), find the correlation coefficient

Solution:a) Given the joint probability density function $f(x,y)=\begin{cases}2-x-y, & 0\le x\le 1,\ 0\le y\le 1\\ 0 & ,\ otherwise\end{cases}$

Marginal density function of X is $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$

$$=\int_0^1 (2-x-y)\,dy$$

$$=\left[2y-xy-\frac{y^2}{2}\right]_0^1$$

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

$$f_X(x)=\begin{cases}\frac{3}{2}-x, & 0<x\le 1\\ 0 & ,\ otherwise\end{cases}$$

Marginal density function of Y is $f_Y(y)=\int_0^1 (2-x-y)\,dx$

$$=\left[2x-\frac{x^2}{2}-xy\right]_0^1$$

$$=\frac{3}{2}-y$$

$$f_Y(y)=\begin{cases}\frac{3}{2}-y, & 0\le y\le 1\\ 0 & , \textit{otherwise}\end{cases}$$

Covariance of $(X,Y)=Cov(X,Y)=E(XY)-E(X)E(Y)$

$$E(X)=\int_0^1 xf_X(x)\,dx=\int_0^1 x\left(\frac{3}{2}-x\right)dx=\left[\frac{3}{2}\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1=\frac{5}{12}$$

$$E(Y)=\int_0^1 yf_Y(y)\,dy=\int_0^1 y\left(\frac{3}{2}-y\right)dy=\frac{5}{12}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

$$E(XY)=\int_0^1\int_0^1 xy\,f(x,y)\,dxdy$$

$$=\int_0^1\int_0^1 xy(2-x-y)\,dxdy$$

$$=\int_0^1\int_0^1\left(2xy-x^2y-xy^2\right)dxdy$$

$$=\int_0^1\left[\frac{2x^2y}{2}-\frac{x^3}{3}y-\frac{x^2}{2}y^2\right]_0^1 dy$$

$$=\int_0^1\left(y-\frac{1}{3}-\frac{y^2}{2}\right)dy$$

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

$$Cov(X,Y)=\frac{1}{6}-\frac{5}{12}\times\frac{5}{12}$$

$$=\frac{1}{6}-\frac{25}{144}=-\frac{1}{144}.$$

b). Correlation coefficient $\rho_{XY}=\dfrac{E(XY)-E(X)E(Y)}{\sigma_X\sigma_Y}$

Marginal density function of X is

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_2^4\left(\frac{6-x-y}{8}\right)dy=\frac{6-2x}{8}$$

Marginal density function of Y is

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_0^2\left(\frac{6-x-y}{8}\right)dx=\frac{10-2y}{8}$$

Then $E(X)=\int_0^2 xf_X(x)\,dx=\int_0^2 x\left(\frac{6-2x}{8}\right)dx$

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

$$=\frac{1}{8}\left[12-\frac{16}{13}\right]=\frac{1}{8}\times\frac{20}{3}=\frac{5}{6}$$

$$E(Y)=\int_2^4 y\left(\frac{10-2y}{8}\right)dy=\frac{1}{8}\left[\frac{10y^2}{2}-\frac{2y^3}{3}\right]_2^4\ =\frac{17}{6}$$

$$E\left(X^2\right)=\int_0^2 x^2 f_x(x)\,dx\quad=\int_0^2 x^2\left(\frac{6-2x}{8}\right)dx=\frac{1}{8}\left[\frac{6x^3}{3}-\frac{2x^4}{4}\right]_0^2=1$$

$$E\left(Y^2\right)=\int_2^4 y^2\left(\frac{10-2y}{8}\right)dy=\frac{1}{8}\left[\frac{10y^3}{3}-\frac{2y^4}{4}\right]_2^4=\frac{25}{3}$$

$$Var(X)=\sigma_X^2=E\left(X^2\right)-\left[E(X)\right]^2=1-\left(\frac{5}{6}\right)^2=\frac{11}{36}$$

$$Var(Y)=\sigma_Y^2=E\left(Y^2\right)-\left[E(Y)\right]^2=\frac{25}{3}-\left(\frac{17}{6}\right)^2=\frac{11}{36}$$

$$E(XY)=\int_2^4\int_0^2 xy\left(\frac{6-x-y}{8}\right)dxdy$$

$$=\frac{1}{8}\int_2^4\left[\frac{6x^2y}{2}-\frac{x^3y}{3}-\frac{x^2y^2}{2}\right]_0^2 dy$$

$$=\frac{1}{8}\int_2^4\left(12y-\frac{8}{3}y-2y^2\right)dy=\frac{1}{8}\left[\frac{12y^2}{2}-\frac{8}{3}\frac{y^2}{2}-\frac{2y^3}{3}\right]_2^4$$

$$=\frac{1}{8}\left[96-\frac{64}{3}-\frac{128}{3}-24+\frac{16}{3}+\frac{16}{3}\right]=\frac{1}{8}\left[\frac{56}{3}\right]$$

$$E(XY)=\frac{7}{3}$$

$$\rho_{XY}=\frac{E(XY)-E(X)E(Y)}{\sigma_X\sigma_Y}=\frac{\frac{7}{3}-\left(\frac{5}{6}\right)\left(\frac{17}{6}\right)}{\frac{\sqrt{11}}{6}\frac{\sqrt{11}}{6}}\quad \rho_{XY}=-\frac{1}{11}.$$

Problem 7 Let X_1 and X_2 be two independent random variables with means 5 and 10 and standard devotions 2 and 3 respectively. Obtain the correlation coefficient of UV where $U = 3X_1 + 4X_2$ and $V = 3X_1 - X_2$.

Solution:a). The probability distribution is

X Y	0	1	2	$P(Y)$
0	$\frac{1}{3}$	0	0	$\frac{1}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
0	0	0	$\frac{1}{3}$	$\frac{1}{3}$
$P(X)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

$$E(X) = \sum_i x_i p_i(x_i) = \left(0\times\frac{1}{3}\right) + \left(1\times\frac{1}{3}\right) + \left(2\times\frac{1}{3}\right) = 1$$

$$E(Y) = \sum_j y_i p_j(y_j) = \left(0\times\frac{1}{3}\right) + \left(1\times\frac{1}{3}\right) + \left(0\times\frac{1}{3}\right) = \frac{1}{3}$$

$$E\left(X^2\right) = \sum_i x_i^{\,2} p(x_i) = \left(0\times\frac{1}{3}\right) + \left(1\times\frac{1}{3}\right) + \left(4\times\frac{1}{3}\right) = \frac{5}{3}$$

$$Var(X) = E\left(X^2\right) - \left[E(X)\right]^2 = \frac{5}{3} - 1 = \frac{2}{3}$$

$$E\left(Y^2\right) = \sum_j y_j^{\,2} p(y_j) = \left(0\times\frac{1}{3}\right) + \left(1\times\frac{1}{3}\right) + \left(0\times\frac{1}{3}\right) = \frac{1}{3}$$

$$\therefore V(Y) = E\left(Y^2\right) - \left[E(Y)\right]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Correlation coefficient $\rho_{XY} = \dfrac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$

$$E(XY) = \sum_i \sum_j x_i y_j p\left(x_i, y_j\right)$$

$$= 0.0.\frac{1}{3} + 0.1.0 + 1.0.0 + 1.1.\frac{1}{3} + 1.2.0 + 0.0.0 + 0.1.0 + 0.2.\frac{1}{3} = \frac{1}{3}$$

$$\rho_{XY} = \frac{\frac{1}{3} - (1)\left(\frac{1}{3}\right)}{\sqrt{\frac{2}{3}\times\frac{2}{9}}} = 0$$

Correlation coefficient $= 0$.

b). Given $E\left(X_1\right) = 5,\ \ E\left(X_2\right) = 10$

$$V\left(X_1\right) = 4,\ \ V\left(X_2\right) = 9$$

Since X and Y are independent $E(XY) = E(X)E(Y)$

Correlation coefficient $= \dfrac{E(UV) - E(U)E(V)}{\sqrt{Var(U)Var(V)}}$

$$E(U) = E\left(3X_1 + 4X_2\right) = 3E\left(X_1\right) + 4E\left(X_2\right)$$

$$=(3\times5)+(4\times10)=15+40=55.$$

$$E(V)=E(3X_1-X_2)=3E(X_1)-E(X_2)$$

$$=(3\times5)-10=15-10=5$$

$$E(UV)=E\left[(3X_1+4X_2)(3X_1-X_2)\right]$$

$$=E\left[9X_1^{\;2}-3X_1X_2+12X_1X_2-4X_2^{\;2}\right]$$

$$=9E\left(X_1^{\;2}\right)-3E(X_1X_2)+12E(X_1X_2)-4E\left(X_2^{\;2}\right)$$

$$=9E\left(X_1^{\;2}\right)+9E(X_1X_2)-4E\left(X_2^{\;2}\right)$$

$$=9E\left(X_1^{\;2}\right)+9E(X_1)E(X_2)-4E\left(X_2^{\;2}\right)$$

$$=9E\left(X_1^{\;2}\right)+450-4E\left(X_2^{\;2}\right)$$

$$V(X_1)=E\left(X_1^{\;2}\right)-\left[E(X_1)\right]^2$$

$$E\left(X_1^{\;2}\right)=V(X_1)+\left[E(X_1)\right]^2=4+25=29$$

$$E\left(X_2^{\;2}\right)=V(X_2)+\left[E(X_2)\right]^2=9+100=109$$

$$\therefore E(UV)=(9\times29)+450-(4\times109)$$

$$=261+450-436=275$$

$$Cov(U,V)=E(UV)-E(U)E(V)$$

$$=275-(5\times55)=0$$

Since $Cov(U,V)=0,$ Correlation coefficient$=0$.

Problem 8 a) Let the random variable X has the marginal density function $f(x)=1,\ -\frac{1}{2}<x<\frac{1}{2}$ and let the conditional density of Y be $f\left(y/x\right)=\begin{cases}1,\ x<y<x+1,\ -\frac{1}{2}<x<0\\ 1,\ -x<y<1-x,\ 0<x<\frac{1}{2}\end{cases}$. Prove that the variables X and Y are uncorrelated.

b) Given $f(x,y)=xe^{-x(y+1)},\ x\geq0,\ y\geq0$. Find the regression curve of Y on X.

Solution:a). We have $E\left(X\right)=\int\limits_{-\frac{1}{2}}^{\frac{1}{2}}xf\left(x\right)dx=\int\limits_{-\frac{1}{2}}^{\frac{1}{2}}xdx=\left[\frac{x^2}{2}\right]_{-\frac{1}{2}}^{\frac{1}{2}}=0$

$$E\left(XY\right)=\int\limits_{-\frac{1}{2}}^{0}\int\limits_{x}^{x+1}xydxdy+\int\limits_{0}^{\frac{1}{2}}\int\limits_{-x}^{1-x}xydxdy$$

$$=\int\limits_{-\frac{1}{2}}^{0}x\left[\int\limits_{x}^{x+1}ydy\right]dx+\int\limits_{0}^{\frac{1}{2}}x\left[\int\limits_{-x}^{1-x}ydy\right]dx$$

$$=\frac{1}{2}\int\limits_{-\frac{1}{2}}^{0}x\left(2x+1\right)dx+\frac{1}{2}\int\limits_{0}^{\frac{1}{2}}x\left(1-2x\right)dx$$

$$=\frac{1}{2}\left[\frac{2x^3}{3}+\frac{x^2}{2}\right]_{-\frac{1}{2}}^{0}+\frac{1}{2}\left[\frac{x^2}{2}-\frac{2x^3}{3}\right]_{0}^{\frac{1}{2}}=0$$

Since $Cov\left(X,Y\right)=E\left(XY\right)-E\left(X\right)E\left(Y\right)=0$, the variables X and Y are uncorrelated.

b). Regression curve of Y on X is $E\left({}^{y}\!/_{x}\right)$

$$E\left({}^{y}\!/_{x}\right)=\int\limits_{-\infty}^{\infty}yf\left({}^{y}\!/_{x}\right)dy$$

$$f\left(y/x\right)=\frac{f\left(x,y\right)}{f_X\left(X\right)}$$

Marginal density function $f_X\left(x\right)=\int\limits_{0}^{\infty}f\left(x,y\right)dy$

$$=x\int\limits_{0}^{\infty}e^{-x\left(y+1\right)}dy$$

$$=x\left[\frac{e^{-x\left(y+1\right)}}{-x}\right]_{0}^{\infty}=e^{-x},\ \ x\geq 0$$

Conditional pdf of Y on X is $f\left({}^{y}\!/_{x}\right)=\frac{f\left(x,y\right)}{f_X\left(x\right)}=\frac{xe^{-xy-x}}{e^{-x}}=xe^{-xy}$

The regression curve of Y on X is given by

$$E\left({}^{y}\!/_{x}\right)=\int\limits_{0}^{\infty}yxe^{-xy}dy$$

$$= x\left[y\frac{e^{-xy}}{-x} - \frac{e^{-xy}}{x^2} \right]_0^{\infty}$$

$E\left(y/_x\right) = \frac{1}{x} \Rightarrow y = \frac{1}{x}$ and hence $xy = 1$.

Problem 9 a) Given $f(x,y) = \begin{cases} \frac{x+y}{3}, & 0<x<1,\ 0<y<2 \\ 0 & , otherwise \end{cases}$, obtain the regression of Y on X and X on Y.

b) Distinguish between correlation and regression Analysis

Solution: Regression of Y on X is $E\left(Y/_X\right)$

$$E\left(Y/_X\right) = \int_{-\alpha}^{\alpha} yf\left(y/_x\right)dy$$

$$f\left(Y/_X\right) = \frac{f(x,y)}{f_X(x)}$$

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

$$= \int_0^2 \left(\frac{x+y}{3}\right)dy = \frac{1}{3}\left[xy + \frac{y^2}{2} \right]_0^2$$

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

$$f\left(Y/_X\right) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{2(x+1)}$$

Regression of Y on X = $E\left(Y/_X\right) = \int_0^2 \frac{y(x+y)}{2(x+1)}dy$

$$= \frac{1}{2(x+1)}\left[\frac{xy^2}{2} + \frac{y^3}{3} \right]_0^2$$

$$= \frac{1}{2(x+1)}\left[2x + \frac{8}{3} \right] = \frac{3x+4}{3(x+1)}$$

$$E\left(X/_Y\right) = \int_{-\infty}^{\infty} xf\left(x/_y\right)dx$$

$$f\left(x/_y\right) = \frac{f(x,y)}{f_Y(y)}$$

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

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

$$= \frac{1}{3}\left[\frac{1}{2} + y\right]$$

$$f\left(x\!\!\diagup_{y}\right) = \frac{2(x+y)}{2y+1}$$

Regression of X on $Y = E\left(X\!\!\diagup_{Y}\right) = \int_0^1 \frac{x+y}{2y+1} dx$

$$= \frac{1}{2y+1}\left[\frac{x^2}{2} + xy\right]_0^1$$

$$= \frac{\frac{1}{2} + y}{2y+1} = \frac{1}{2}.$$

1. Correlation means relationship between two variables and Regression is a Mathematical Measure of expressing the average relationship between the two variables.

2. Correlation need not imply cause and effect relationship between the variables. Regression analysis clearly indicates the cause and effect relationship between Variables.

3. Correlation coefficient is symmetric i.e. $r_{xy} = r_{yx}$ where regression coefficient is not symmetric

4. Correlation coefficient is the measure of the direction and degree of linear relationship between two variables. In regression using the relationship between two variables we can predict the dependent variable value for any given independent variable value.

Problem 10 a) X any Y are two random variables with variances σ_x^2 and σ_y^2 respectively and r is the coefficient of correlation between them. If $U = X + KY$ and $V = X + \frac{y\sigma_x}{\sigma_y}$, find the value of k so that U and V are uncorrelated.

b) Find the regression lines:

X	6	8	10	18	20	23
Y	40	36	20	14	10	2

Solution:

Given $U = X + KY$

$$E(U) = E(X) + KE(Y)$$

$$V = X + \frac{\sigma_X}{\sigma_Y} Y$$

$$E(V) = E(X) + \frac{\sigma_X}{\sigma_Y} E(Y)$$

If U and V are uncorrelated, $Cov(U,V) = 0$

$$E\left[\left(U - E(U)\right)\left(V - E(V)\right)\right] = 0$$

$$\Rightarrow E\left[\left(X + KY - E(X) - KE(Y)\right) \times \left(X + \frac{\sigma_X}{\sigma_Y} Y - E(X) - \frac{\sigma_X}{\sigma_Y} E(Y)\right)\right] = 0$$

$$\Rightarrow E\left\{\left[\left(X - E(X)\right) + K\left(Y - E(Y)\right)\right] \times \left[\left(X - E(X)\right) + \frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)\right]\right\} = 0$$

$$\Rightarrow E\left\{\left(X - E(X)\right)^2 + \frac{\sigma_X}{\sigma_Y}\left(X - E(X)\right)\left(Y - E(Y)\right) + K\left(Y - E(Y)\right)\left(X - E(X)\right) + K\frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)^2\right\} = 0$$

$$\Rightarrow V(X) + \frac{\sigma_X}{\sigma_Y} Cov(X,Y) + KCov(X,Y) + K\frac{\sigma_X}{\sigma_Y} V(Y) = 0$$

$$K\left[Cov(X,Y) + \frac{\sigma_X}{\sigma_Y} V(Y)\right] = -V(X) - \frac{\sigma_X}{\sigma_Y} Cov(x,y)$$

$$K = \frac{-V(X) - \frac{\sigma_X}{\sigma_Y} r\sigma_X\sigma_Y}{r\sigma_X\sigma_Y + \frac{\sigma_X}{\sigma_Y} V(Y)} = \frac{-\sigma_X^2 - r\sigma_X^2}{r\sigma_X\sigma_Y + \sigma_X\sigma_Y}$$

$$= \frac{-\sigma_X^2(1+r)}{\sigma_X\sigma_Y(1+r)} = -\frac{\sigma_X}{\sigma_Y}.$$

b).

X	Y	X^2	Y^2	XY
6	40	36	1600	240
8	36	64	1296	288
10	20	100	400	200
18	14	324	196	252
20	10	400	100	200
23	2	529	4	46
$\sum X = 85$	$\sum Y = 122$	$\sum X^2 = 1453$	$\sum Y^2 = 3596$	$\sum XY = 1226$

,

$$\overline{X} = \frac{\sum x}{n} = \frac{85}{6} = 14.17$$

$$\overline{Y} = \frac{\sum y}{n} = \frac{122}{6} = 20.33$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{1453}{6} - \left(\frac{85}{6}\right)^2} = 6.44$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{3596}{6} - \left(\frac{122}{6}\right)^2} = 13.63$$

$$r = \frac{\frac{\sum xy}{n} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \frac{\frac{1226}{6} - (14.17)(20.33)}{(6.44)(13.63)} = -0.95$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = -0.95 \times \frac{6.44}{13.63} = -0.45$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = -0.95 \times \frac{13.63}{6.44} = -2.01$$

The regression line X on Y is

$$x - \overline{x} = b_{xy}\left(y - \overline{y}\right) \Rightarrow x - 14.17 = -0.45\left(y - \overline{y}\right)$$

$$\Rightarrow x = -0.45y + 23.32$$

The regression line Y on X is

$$y - \overline{y} = b_{yx}\left(x - \overline{x}\right) \Rightarrow y - 20.33 = -2.01(x - 14.17)$$

$$\Rightarrow y = -2.01x + 48.81$$

Problem 11 a) Using the given information given below compute $\overline{x}, \overline{y}$ and r. Also compute σ_y when $\sigma_x = 2,\ 2x + 3y = 8$ and $4x + y = 10$.

b) The joint pdf of X and Y is

Y \ X	-1	1
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$

Find the correlation coefficient of X and Y.

Solution:a). When the regression equation are Known the arithmetic means are computed by solving the equation.

$$2x+3y=8 \text{------------ (1)}$$

$$4x+y=10 \text{------------ (2)}$$

$(1)\times 2 \Rightarrow 4x+6y=16$ ------- (3)

$(2)-(3) \Rightarrow -5y=-6$

$$\Rightarrow y=\frac{6}{5}$$

Equation $(1) \Rightarrow 2x+3\left(\frac{6}{5}\right)=8$

$$\Rightarrow 2x=8-\frac{18}{5}$$

$$\Rightarrow x=\frac{11}{5}$$

i.e. $\overline{x}=\frac{11}{5}$ & $\overline{y}=\frac{6}{5}$

To find r, Let $2x+3y=8$ be the regression equation of X on Y.

$$2x=8-3y \Rightarrow x=4-\frac{3}{2}y$$

$\Rightarrow b_{xy}=$Coefficient of Y in the equation of X on $Y=-\frac{3}{2}$

Let $4x+y=10$ be the regression equation of Y on X

$\Rightarrow y=10-4x$

$\Rightarrow b_{yx}=$coefficient of X in the equation of Y on $X=-4$.

$$r=\pm\sqrt{b_{xy}b_{yx}}$$

$$=-\sqrt{\left(-\frac{3}{2}\right)(-4)} \qquad (\because b_{xy} \text{ \& } b_{yx} \text{ are negative })$$

$$=-2.45$$

Since r is not in the range of $(-1\le r\le 1)$ the assumption is wrong.

Now let equation (1) be the equation of Y on X

$$\Rightarrow y = \frac{8}{3} - \frac{2x}{3}$$

$\Rightarrow b_{yx} =$ Coefficient of X in the equation of Y on X

$$b_{yx} = -\frac{2}{3}$$

from equation (2) be the equation of X on Y

$$b_{xy} = -\frac{1}{4}$$

$$r = \pm\sqrt{b_{xy} b_{yx}} \quad = \sqrt{-\frac{2}{3} \times -\frac{1}{4}} = 0.4081$$

To compute σ_y from equation (4) $b_{yx} = -\frac{2}{3}$

But we know that $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$$\Rightarrow -\frac{2}{3} = 0.4081 \times \frac{\sigma_y}{2}$$

$$\Rightarrow \sigma_y = -3.26$$

b). Marginal probability mass function of X is

When $X = 0, \quad P(X) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$

$X = 1, \quad P(X) = \frac{2}{8} + \frac{2}{8} = \frac{4}{8}$

Marginal probability mass function of Y is

When $Y = -1, \quad P(Y) = \frac{1}{8} + \frac{2}{8} = \frac{3}{8}$

$Y = 1, \quad P(Y) = \frac{3}{8} + \frac{2}{8} = \frac{5}{8}$

$$E(X) = \sum_x x\, p(x) = 0 \times \frac{4}{8} + 1 \times \frac{4}{8} = \frac{4}{8}$$

$$E(Y) = \sum_y y\, p(y) = -1 \times \frac{3}{8} + 1 \times \frac{5}{8} = -\frac{3}{8} + \frac{5}{8} = \frac{2}{8}$$

$$E(X^2) = \sum_x x^2\, p(x) = 0^2 \times \frac{4}{8} + 1^2 \times \frac{4}{8} = \frac{4}{8}$$

$$E\left(Y^2\right)=\sum_y y^2\, p(y)=(-1)^2\times\frac{3}{8}+1^2\times\frac{5}{8}=\frac{3}{8}+\frac{5}{8}=1$$

$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2$$

$$=\frac{4}{8}-\left(\frac{4}{8}\right)^2=\frac{1}{4}$$

$$V(Y)=E\left(Y^2\right)-\left(E(Y)\right)^2$$

$$=1-\left(\frac{1}{4}\right)^2=\frac{15}{16}$$

$$E(XY)=\sum_x\sum_y xy\, p(x,y)$$

$$=0\times\frac{1}{8}+0\times\frac{3}{8}+(-1)\frac{2}{8}+1\times\left(\frac{2}{8}\right)=0$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0-\frac{1}{2}\times\frac{1}{4}=-\frac{1}{8}$$

$$r=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}=\frac{-\frac{1}{8}}{\sqrt{\frac{1}{4}}\sqrt{\frac{15}{16}}}=-0.26\,.$$

Problem 12 a) Calculate the correlation coefficient for the following heights (in inches) of fathers X and their sons Y.

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

b) If X and Y are independent exponential variates with parameters 1, find the pdf of $U = X - Y$.

Solution:

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4359	4624
67	65	4355	4489	4285
68	72	4896	4624	5184
69	72	4968	4761	5184

70	69	4830	4900	4761
72	71	5112	5184	5041
å $X = 544$	å $Y = 552$	å $XY = 37560$	å $X^2 = 37028$	å $Y^2 = 38132$

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

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

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

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

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

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

$$= 4695 - 4692 = 3$$

The correlation coefficient of X and Y is given by

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{3}{(2.121)(2.345)}$$

$$= \frac{3}{4.973} = 0.6032\,.$$

b). Given that X and Y are exponential variates with parameters 1

$$f_X(x) = e^{-x},\ x \ge 0,\ f_Y(y) = e^{-y},\ y \ge 0$$

Also $f_{XY}(x,y) = f_X(x) f_y(y)$ since X and Y are independent

$$= e^{-x} e^{-y}$$

$$= e^{-(x+y)};\ x \ge 0,\ y \ge 0$$

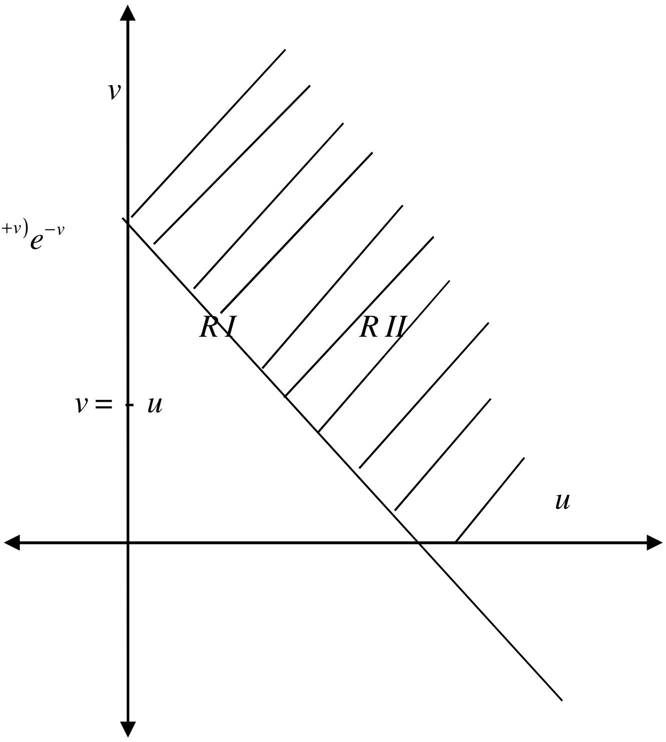
Consider the transformations $u = x - y$ and $v = y$

$$\Rightarrow x = u + v\,,\ y = v$$

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

$$f_{UV}(u,v) = f_{XY}(x,y)|J| = e^{-x}e^{-y} = e^{-(u+v)}e^{-v}$$

$$= e^{-(u+2v)},\ u+v \geq 0,\ v \geq 0$$



In Region I when $u < 0$

$$f(u) = \int_{-u}^{\infty} f(u,v)\,dv = \int_{-u}^{\infty} e^{-u}.e^{-2v}\,dv$$

$$= e^{-u}\left[\frac{e^{-2v}}{-2}\right]_{-u}^{\infty}$$

$$= \frac{e^{-u}}{-2}\left[0 - e^{2u}\right] = \frac{e^{u}}{2}$$

In Region II when $u > 0$

$$f(u) = \int_{0}^{\infty} f(u,v)\,dv$$

$$= \int_{0}^{\infty} e^{-(u+2v)}\,dv = \frac{e^{-u}}{2}$$

$$\therefore f(u) = \begin{cases} \frac{e^{u}}{2}, & u < 0 \\ \frac{e^{-u}}{2}, & u > 0 \end{cases}$$

Problem 13 The joint pdf of X and Y is given by $f(x,y) = e^{-(x+y)}, x > 0, y > 0$. Find the pdf of $U = \frac{X+Y}{2}$.

b) If X and Y are independent random variables each following $N(0,2)$, find the pdf of $Z = 2X + 3Y$. If X and Y are independent rectangular variates on $(0,1)$ find the distribution of $\frac{X}{Y}$.

Solution:a). Consider the transformation $u = \frac{x+y}{2}$ & $v = y$

$$\Rightarrow x = 2u - v \text{ and } y = v$$

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ 0 & 1 \end{vmatrix} = 2$$

$$f_{UV}(u,v) = f_{XY}(x,y)|J|$$

$$= e^{-(x+y)}2 = 2e^{-(x+y)} = 2e^{-(2u-v+v)}$$

$$= 2e^{-2u},\ 2u - v \ge 0,\ v \ge 0$$

$$f_{UV}(u,v) = 2e^{-2u},\ u \ge 0,\ 0 \le v \le \frac{u}{2}$$

$$f(u) = \int_0^{\frac{u}{2}} f_{UV}(u,v)\,dv = \int_0^{\frac{u}{2}} 2e^{-2u}\,dv$$

$$= \left[2e^{-2u}v\right]_0^{\frac{u}{2}}$$

$$f(u) = \begin{cases} 2\frac{u}{2}e^{-2u}, & u \ge 0 \\ 0 & , \textit{otherwise} \end{cases}$$

b).(i) Consider the transformations $w = y$,

i.e. $z = 2x + 3y$ and $w = y$

i.e. $x = \frac{1}{2}(z - 3w)$, $y = w$

$$|J| = \frac{\partial(x,y)}{\partial(z,w)} = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & -\frac{3}{2} \\ 0 & 1 \end{vmatrix} = \frac{1}{2}.$$

Given that X and Y are independent random variables following $N(0,2)$

$$\therefore f_{XY}(x,y) = \frac{1}{8\pi}e^{\frac{-(x^2+y^2)}{8}}, -\infty < x, y < \infty$$

The joint pdf of (z,w) is given by

$$f_{ZW}(z,w) = |J|\,f_{XY}(x,y)$$

$$= \frac{1}{2}.\frac{1}{8\pi}e^{\frac{-\left[\frac{1}{4}(z-3w)^2+w^2\right]}{8}}$$

$$= \frac{1}{16\pi}e^{-\frac{1}{32}\left[(z-3w)^2+4w^2\right]}, -\infty < z, w < \infty\ .$$

The pdf of z is the marginal pdf obtained by interchanging $f_{ZW}(z,w)$ w.r.to w over the range of w.

$$\therefore f_Z(z) = \frac{1}{16\pi}\int_{-\infty}^{\infty}\left(e^{-\frac{1}{32}\left(z^2-6wz+13w^2\right)}\right)dw$$

$$= \frac{1}{16\pi}e^{-\frac{z^2}{32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w^2-\frac{6wz}{13}+\left(\frac{3z}{13}\right)^2-\left(\frac{3z}{13}\right)^2\right)}\right)dw$$

$$= \frac{1}{16\pi}e^{-\frac{z^2}{32}+\frac{9z^2}{13\times 32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w-\frac{3z}{13}\right)^2}\right)dw$$

$$= \frac{1}{16\pi}e^{-\frac{z^2}{8\times 13}}\int_{-\infty}^{\infty}e^{-\frac{13}{32}t^2}\,dt$$

$$r = \frac{13}{32}t^2 \Rightarrow dr = \frac{13}{16}tdt \Rightarrow \frac{16}{13t}dr = dt \Rightarrow \sqrt{\frac{r32}{13}}dr = dt$$

$$\frac{16}{13}\sqrt{\frac{13}{r32}}dr = dt \Rightarrow \frac{4}{\sqrt{13}\times\sqrt{2}}r^{-\frac{1}{2}}dr = dt$$

$$= \frac{2}{16\pi}\frac{4}{\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times 13}}\int_{0}^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$= \frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times 13}}\int_{0}^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$= \frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times 13}}\sqrt{\pi} = \frac{1}{2\sqrt{13}\sqrt{2\pi}}e^{-\frac{z^2}{2\left(2\sqrt{13}\right)^2}}$$

i.e. $Z \sim N\left(0, 2\sqrt{13}\right)$

b).(ii) Given that X and Y are uniform Variants over $(0,1)$

$$\therefore f_X(x) = \begin{cases} 1, \ 0 < x < 1 \\ 0, \ otherwise \end{cases} \text{ and } f_Y(y) = \begin{cases} 1, \ 0 < y < 1 \\ 0, \ otherwise \end{cases}$$

Since X and Y are independent,

$$f_{XY}(x,y) = f_X(x)f_y(y)\begin{cases} 1, \ 0 < x, y < 1 \\ 0, \ otherwise \end{cases}$$

Consider the transformation $u = \frac{x}{y}$ and $v = y$

i.e. $x = uv$ and $y = v$

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = v$$

$$\therefore f_{UV}(u,v) = f_{XY}(x,y)|J|$$

$$= v,\ 0 < u < \infty,\ 0 < v < \infty$$

The range for u and v are identified as follows.

$0 < x < 1$ and $0 < y < 1. \Rightarrow 0 < uv < 1$ and $0 < v < 1$

$\Rightarrow uv > 0,\ uv < 1,\ v > 0$ and $v < 1$

$\Rightarrow uv > 0$ and $v > 0 \Rightarrow u > 0$

Now $f(u) = \int f_{UV}(u,v)\,dv$

The range for v differs in two regions

$$f(u) = \int_0^1 f_{UV}(u,v)dv$$

$$= \int_0^1 v dv = \left[\frac{v^2}{2}\right]_0^1 = \frac{1}{2},\ 0 < u < 1$$

$$f(u) = \int_0^{\frac{1}{u}} f_{UV}(u,v)dv$$

$$= \int_0^{\frac{1}{u}} v dv = \left[\frac{v^2}{2}\right]_0^{\frac{1}{u}} = \frac{1}{2u^2},\ 1 \le u \le \infty$$

$$\therefore f(u) = \begin{cases} \frac{1}{2} & ,\ 0 \le u \le 1 \\ \frac{1}{2u^2}, & u > 1 \end{cases}$$

Problem 14 a) If $X_1, X_2, ...X_n$ are Poisson variates with parameter $\lambda = 2$. Use the central limit theorem to estimate $P(120 < S_n < 160)$ where $s_n = X_1 + X_2 + ... + X_n$ and $n = 75$.

b) A random sample of size 100 is taken from a population whose mean is 60 and variance is 400. Using central limit theorem, with what probability can we assent that the mean of the sample will not differ from $\mu = 60$ by more than 4.

Solution: a). Given that $E(X_i) = \lambda = 2$ and $Var(X_i) = \lambda = 2$

[Since in Poisson distribution mean and variance are equal to λ]

i.e. $\mu = 2$ and $\sigma^2 = 2$

By central limit theorem, $S_n \sim N\left(n\mu, n\sigma^2\right)$

$$S_n \sim N(150,150)$$

$$\therefore P(120 < S_n < 160) = P\left(\frac{120-150}{\sqrt{150}} < z < \frac{160-150}{\sqrt{150}}\right)$$

$$= P(-2.45 < z < 0.85)$$

$$= P(-2.45 < z < 0) + P(0 < z < 0.85)$$

$$= P(0 < z < 2.45) + P(0 < z < 0.85)$$

$$= 0.4927 + 0.2939 = 0.7866$$

b). Given that $n = 100$, $\mu = 60$, $\sigma^2 = 400$

Since the probability statement is with respect to mean, we use the Linderberg-levy form of central limit Theorem.

$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ i.e. $\overline{X}$ follows normal distribution with mean 'μ' and variance $\frac{\sigma^2}{n}$.

$$\text{i.e. } \overline{X} \sim N\left(60, \frac{400}{100}\right)$$

$$\overline{X} \sim N(60,4)$$

$$P\begin{bmatrix} \textit{mean of the sample will not} \\ \textit{differ from } 60 \textit{ by more than } 4 \end{bmatrix} = P\begin{bmatrix} \overline{X} \textit{ will not differ from} \\ \mu = 60 \textit{ by more than } 4 \end{bmatrix}$$

$$= P\left(\overline{X} - \mu\right) \le 4$$

$$= P\left[-4 \le \overline{X} - \mu \le 4\right]$$

$$= P\left[-4 \le \overline{X} - 60 \le 4\right]$$

$$= P\left[56 \le \overline{X} \le 64\right] = P\left[\frac{56-60}{2} \le z \le \frac{64-60}{2}\right]$$

$$= P[-2 \le Z \le 2]$$

$$= 2P[0 \le Z \le 2] = 2 \times 0.4773 = 0.9446$$

Problem 15 a) If the variable X_1, X_2, X_3, X_4 are independent uniform variates in the interval $(450, 550)$, find $P(1900 \le X_1 + X_2 + X_3 + X_4 \le 2100)$ using central limit theorem.

b) A distribution with unknown mean μ has a variance equal to 1.5. Use central limit theorem to find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample mean will be within 0.5 of the population mean.

Solutiona). Given that X follows a uniform distribution in the interval $(450, 550)$

$$\text{Mean} = \frac{b+a}{2} = \frac{450+550}{2} = 500$$

$$\text{Variance} = \frac{(b-a)^2}{12} = \frac{(550-450)^2}{12} = 833.33$$

By CLT, $S_n = X_1 + X_2 + X_3 + X_4$ follows a normal distribution with $N\left(n\mu, n\sigma^2\right)$

The standard normal variable is given by $Z = \dfrac{S_n - n\mu}{n\sigma^2}$

when $S_n = 1900$, $Z = \dfrac{1900 - 4\times500}{\sqrt{4\times833.33}} = -\dfrac{100}{57.73} = -1.732$

when $S_n = 2100$, $Z = \dfrac{2100 - 2000}{\sqrt{4\times833.33}} = \dfrac{100}{57.73} = 1.732$

$$\therefore P(1900 \le S_n \le 2100) = P(-1.732 < z < 1.732)$$

$$= 2\times P(0 < z < 1.732)$$

$$= 2\times0.4582 = 0.9164.$$

b). Given $E(X_i) = \mu$ and $Var(X_i) = 1.5$

Let $\overline{X}$ denote the sample mean.

By C.L.T, $\overline{X}$ follows $N\left(\mu, \dfrac{\sqrt{1.5}}{\sqrt{n}}\right)$

We have to find 'n' such that $P\left(\mu - 0.5 < \overline{X} < \mu + 0.5\right) \ge 0.95$

i.e. $P\left(-0.5 < \overline{X} - \mu < 0.5\right) \ge .95$

$$P\left(\left|\overline{X} - \mu\right| < 0.5\right) \ge .95$$

$$P\left[\left|z\frac{\sigma}{\sqrt{n}}\right| < 0.5\right] \ge 0.95$$

$$P\left[|z| < 0.5\frac{\sqrt{n}}{\sigma}\right] \geq 0.95$$

$$P\left[|z| < 0.5\frac{\sqrt{n}}{\sqrt{1.5}}\right] \geq 0.95$$

ie $P\left(|Z| < 0.4082\sqrt{n}\right) \geq 0.95$

Where $'Z'$ is the standard normal variable.

The Last value of $'n'$ is obtained from $P\left(|Z| < 0.4082\sqrt{n}\right) = 0.95$

$$2P\left(0 < z < 0.4082\sqrt{n}\right) = 0.95$$

$$\Rightarrow 0.4082\sqrt{n} = 1.96$$

$$\Rightarrow n = 23.05$$

$\therefore$ The size of the sample must be at least 24.

16.*The joint probability function of the R.V's (X,Y) is $f(x,y) = c(2x+y)$,*

where x & y can assume all integers such that $0 \leq x \leq 2$ & $0 \leq y \leq 3$ and

$f(x,y) = 0$ otherwise. Find the value of c and $P(X > 1, Y \leq 2)$. **[APRIL/MAY 2015]**

Solution: Given $f(x,y) = c(2x+y)$,

where x & y can assume all integers such that $0 \leq x \leq 2$ & $0 \leq y \leq 3$

X ∣ Y	0	1	2	3	$P(Y=y)$
0	0	c	2c	3c	6c
1	2c	3c	4c	5c	14c
2	4c	5c	6c	7c	22c
$P(X=x)$	6c	9c	12c	15c	42c

WKT $\sum\sum P(x,y) = 1 \Rightarrow \quad 42c = 1 \Rightarrow \quad c = \frac{1}{42}$

$$P(X > 1, Y \leq 2) = P(X = 2, Y = 0,1,2)$$

$$= P(2,0) + P(2,1) + P(2,2)$$

$$= 4c + 5c + 6c = 15c = \frac{15}{42} = \frac{5}{14}$$

17.*Suppose that the RV's X & Y have the joint pdf $f(x,y) = \begin{cases} e^{-(x+y)}, x > 0, y > 0 \\ 0 \qquad , Otherwise \end{cases}$.*

Find (i)$P(X > 2, Y < 4)$ (ii)$P(X > Y)$. **[NOV/DEC 2015]**

Solution: Given $f(x,y) = \begin{cases} e^{-(x+y)}, x > 0, y > 0 \\ 0 \qquad , \text{Otherwise} \end{cases}$

$$(i) P(X > 2, Y < 4) = \int_0^4 \int_2^{\propto} e^{-(x+y)} dxdy$$

$$= \left(\int_2^{\propto} e^{-x}dx\right)\left(\int_0^4 e^{-y}dy\right)$$

$$= [-e^{-x}]_2^{\infty}\,.\,[-e^{-y}]_0^4 = e^{-2}(1-e^{-4})$$

$$(ii)\ P(X>Y) = \int_0^{\propto}\int_y^{\propto} e^{-(x+y)}dxdy$$

$$= \int_0^{\propto}\int_y^{\propto} e^{-x}e^{-y}dxdy$$

$$= \int_0^{\infty} e^{-y}[-e^{-x}]_y^{\infty}\,dy$$

$$= \int_0^{\infty} e^{-y}\,[0+e^{-y}]dy$$

$$= \int_0^{\infty} e^{-2y}\,dy = \left[\frac{e^{-2y}}{-2}\right]_0^{\infty} = \left[0+\frac{1}{2}\right] = \frac{1}{2}$$

18. $\boldsymbol{The\ joint\ probability\ function\ of\ the\ R.V's\ (X,Y)\ is\ f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, x>0, y>0 \\ 0 \qquad\quad, Otherwise \end{cases}}$

$\boldsymbol{Compute\ (1)\ P(X>1, Y<1) \quad (2) P(X<Y) \quad (3) P(X<a).}$ **[APR/MAY 2015]**

Solution: Given $\mathrm{f(x,y)} = \begin{cases} \mathrm{2e^{-x}e^{-2y}, x>0, y>0} \\ 0 \qquad\quad \mathrm{, Otherwise} \end{cases}$

$$(1)\ P(X>1, Y<1) = \int_0^1\int_1^{\propto} 2e^{-x}e^{-2y}dxdy$$

$$= 2\left(\int_1^{\propto} e^{-x}dx\right).\left(\int_0^1 e^{-2y}dy\right)$$

$$= 2\left[\frac{e^{-x}}{-1}\right]_1^{\infty}\ \left[\frac{e^{-2y}}{-2}\right]_0^1 = e^{-1}(1-e^{-2})$$

$$(2) P(X<Y) = \int_0^{\infty}\int_0^y 2e^{-x}e^{-2y}dxdy = 2\int_0^{\propto} e^{-2y}\left[\frac{e^{-x}}{-1}\right]_0^y dxdy$$

$$= 2\int_0^{\infty}(e^{-2y}-e^{-3y})dy = 2\left[\frac{e^{-2y}}{-2}+\frac{e^{-3y}}{-3}\right]_0^{\infty} = 2\left[\frac{1}{2}-\frac{1}{3}\right] = \frac{1}{3}$$

$$(3)\mathrm{P(X<a)} = \int_0^a f(x)\,dx = \int_0^a\left(\int_{-\propto}^{\propto} f(x,y)dy\right)dx$$

$$= \int_0^a\left(\int_0^{\propto} \mathrm{2e^{-x}e^{-2y}dy}\right)dx$$

$$= 2\left(\int_0^a e^{-x}dx\right).\left(\int_0^{\infty} e^{-2y}dy\right)$$

$$= 2\left[\frac{e^{-x}}{-1}\right]_0^a\ \left[\frac{e^{-2y}}{-2}\right]_0^{\infty} = 1-e^{-a}$$

19. $\boldsymbol{Let\ the\ joint\ pdf\ of\ the\ random\ variables\ X\ and\ Y\ is\ given\ by}$

$\boldsymbol{f(x,y) = \begin{cases} Cxy^2, 0\le x\le y\le 1 \\ 0 \qquad\quad, Otherwise \end{cases} (1) Find\ C\ (2) Find\ the\ marginal\ pdf'sof\ X\ and\ Y}$

$\boldsymbol{(3) Find\ f(x|y).}$ **[NOV/DEC 2015]**

Solution: Given $\mathrm{f(x,y)} = \begin{cases} \mathrm{Cxy^2, 0\le x\le y\le 1} \\ 0 \qquad\quad \mathrm{, Otherwise} \end{cases}$

(i) WKT $\int_{-\propto}^{\propto}\int_{-\propto}^{\propto} f(x,y)dxdy = 1$

$$\int_0^1\int_0^y \mathrm{Cxy^2}dxdy = 1 \Rightarrow C\int_0^1\left[\frac{x^2}{2}y^2\right]_0^y dy = 1$$

$$\Rightarrow \frac{C}{2}\int_0^1 y^4 dy = \frac{C}{2}\left[\frac{y^5}{5}\right]_0^1 \Rightarrow \frac{C}{2}.\frac{1}{5} = 1 \Rightarrow C = 10$$

(ii)To find marginal df's of X & Y

WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy$

$$= \int_x^1 \mathrm{Cxy}^2 dy = Cx\left[\frac{y^3}{3}\right]_x^1 = \frac{Cx(1-x^3)}{3}, 0 \le \mathrm{x} \le 1$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx$

$$f(y) = \int_0^y \mathrm{Cxy}^2 dx = \mathrm{Cy}^2\left[\frac{\mathrm{x}^2}{2}\right]_0^{\mathrm{y}} = 5\mathrm{y}^4, 0 \le \mathrm{y} \le 1$$

(iii)WKT the conditional density function of X given Y=y is $f(x/y) = \frac{f(x,y)}{f(Y=y)}$

$$f(x/y) = \frac{10xy^2}{5y^4} = \frac{2x}{y^2}, \le \mathrm{x} \le \mathrm{y} \le 1$$

20. $\boldsymbol{The\ joint\ pdf\ of\ a\ \ two-dimensional\ random\ variables\ (X,Y) is\ given\ by}$

$$\boldsymbol{f(x,y) = \begin{cases} 2 & ,0 < x < 1, 0 < y < x, \\ 0 & ,Otherwise \end{cases}\ Find\ the\ marginal\ density\ functions\ of\ X\ \&\ Y.}$$

[APR/MAY 2015]

Solution: WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy$

$$= \int_0^x 2dy = 2x\ ,0 < \mathrm{x} < 1$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx$

$$= \int_y^1 2dx = 2(1-y)\ ,0 < \mathrm{y} < 1$$

21. $\boldsymbol{The\ joint\ pdf\ of\ the\ R.V's\ (X,Y) is\ given\ by\ f(x,y) = 4xy\ e^{-(x^2+y^2)}, x,y \ge 0.}$

$\boldsymbol{Are\ X\ and\ Y\ independent?\ Find\ conditional\ density\ function\ of\ X\ given\ Y = y.}$

[APR/MAY 2015]

Solution: Verify X & Y are independent ?

i.e Verify $f(x,y) = f(x).f(y)$

To find marginal df's of X & Y

WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy$

$$f(x) = \int_0^{\propto} 4xy\ e^{-(x^2+y^2)}dy$$

$$= 4xe^{-x^2}\int_0^{\propto} ye^{-y^2}dy$$

$$= 4xe^{-x^2}\left(\frac{1}{2}\right) = 2xe^{-x^2}\ , x > 0$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx$

$$f(y) = \int_0^{\propto} 4xy\ e^{-(x^2+y^2)}dx$$

$$= 4ye^{-y^2}\int_0^{\propto} xe^{-x^2}dx$$

$$= 4ye^{-y^2}\left(\frac{1}{2}\right) = 2ye^{-y^2}\ , y > 0$$

$$\therefore f(x).f(y) = 2xe^{-x^2}.2ye^{-y^2} = 4xy\ e^{-(x^2+y^2)} = f(x,y)$$

Hence X and Y are independent

WKT the conditional density function of X given Y=y is $f(x/y) = \frac{f(x,y)}{f(Y=y)}$

$$f(x/y) = \frac{4xy\ e^{-(x^2+y^2)}}{2ye^{-y^2}} = 2xe^{-x^2}, \text{x, y} \geq 0$$

22. $\boldsymbol{The\ joint\ pdf\ of\ the\ R.V's\ (X,Y) is\ given\ by\ \ f(x,y) = kxy\ e^{-(x^2+y^2)}, x,y \geq 0.}$

$\boldsymbol{Are\ X\ and\ Y\ independent?Justify.}$ **[APRIL/MAY 2015]**

Solution: WKT $\int_{-\propto}^{\propto}\int_{-\propto}^{\propto} f(x,y)dxdy = 1$

$$\Rightarrow \int_0^{\propto}\int_0^{\propto} kxy\ e^{-(x^2+y^2)}dxdy = 1$$

$$\Rightarrow\ k\left(\int_0^{\propto} e^{-x^2}dx\right)\left(\int_0^{\propto} e^{-y^2}dy\right) = 1$$

$$\Rightarrow\ k(1/2)(1/2) = 1 \qquad \langle \because\ \int_0^{\propto} e^{-x^2}dx = 1/2$$

$$\Rightarrow k = 4$$

Verify X & Y are independent ?

i.e Verify $f(x,y) = f(x).f(y)$

To find marginal df's of X & Y

WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy$

$$f(x) = \int_0^{\propto} 4xy\ e^{-(x^2+y^2)}dy$$

$$= 4xe^{-x^2}\int_0^{\propto} ye^{-y^2}dy$$

$$= 4xe^{-x^2}\left(\frac{1}{2}\right) = 2xe^{-x^2}\ , x > 0$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx$

$$f(y) = \int_0^{\propto} 4xy\ e^{-(x^2+y^2)}dx$$

$$= 4ye^{-y^2}\int_0^{\propto} xe^{-x^2}dx$$

$$= 4ye^{-y^2}\left(\frac{1}{2}\right) = 2ye^{-y^2}\ , y > 0$$

$$\therefore f(x).f(y) = 2xe^{-x^2}.2ye^{-y^2} = 4xy\ e^{-(x^2+y^2)} = f(x,y)$$

Hence X and Y are independent

23. $\boldsymbol{Let\ the\ joint\ pdf\ of\ the\ random\ variables\ X\ and\ Y\ is\ given\ by}$

$$\boldsymbol{f(x,y) = \begin{cases} 4xy & ,0 \le x,y \le 1 \\ 0 & ,Otherwise \end{cases} \quad Find\ (1)E(X) \quad (2)E(Y) \quad (3)E(X+Y) \quad (4)E(XY).}$$

[APR/MAY 2015]

Solution: Given $f(x,y) = \begin{cases} 4xy & ,0 \le x,y \le 1 \\ 0 & ,Otherwise \end{cases}$

To find marginal df's of X & Y

WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy = \int_0^1 4xy\ dy$

$$= 4x\int_0^1 y\ dy = 4x\ \left[\frac{y^2}{2}\right]_0^1 = 2x\ ,0 < x < 1$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx = \int_0^1 4xy\ dx$

$$= 4y\int_0^1 x\ dx = 4y\ \left[\frac{x^2}{2}\right]_0^1 = 2y\ ,0 < y < 1$$

$$(i)E(X) = \int_{-\propto}^{\propto} xf(x)dx = \int_0^1 x.2x\ dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$

$$(ii)E(Y) = \int_{-\propto}^{\propto} yf(y)dy = \int_0^1 y.2y\ dy = \left[\frac{2y^3}{3}\right]_0^1 = \frac{2}{3}$$

$$(iii)\ E(X+Y) = E(X) + E(Y) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

$$(iv)E(XY) = E(X).E(Y) = \frac{2}{3}.\frac{2}{3} = \frac{4}{9} \qquad \langle \because\ X\ \&\ Y\ are\ independent$$

24. $\boldsymbol{Suppose\ that\ the\ R.V's\ X\ \&\ Y\ have\ the\ joint\ pdf}$

$$\boldsymbol{f(x,y) = \begin{cases} 2\ ,x > 0, y > 0, x + y < 1 \\ 0 \qquad\qquad ,Otherwise \end{cases} \quad Compute\ \rho_{xy}.}$$

[NOV/DEC 2015]

Solution: Given $\mathrm{f(x,y)} = \begin{cases} 2\ ,\mathrm{x} > 0, \mathrm{y} > 0, \mathrm{x} + \mathrm{y} < 1 \\ 0 \qquad\qquad ,\mathrm{Otherwise} \end{cases}$

To find marginal df's of X & Y

WKT Marginal df of X is $f(x) = \int_{-\propto}^{\propto} f(x,y)dy$

$$= \int_0^{1-x} 2dy = 2(1-x), 0 < x < 1$$

WKT Marginal df of Y is $f(y) = \int_{-\propto}^{\propto} f(x,y)dx$

$$= \int_0^{1-y} 2dx = 2(1-y), 0 < y < 1$$

$$E(X) = \int_{-\propto}^{\propto} xf(x)dx = \int_0^1 x.2(1-x)\ dx = \frac{1}{3}$$

$$\mathrm{E(X^2)} = \int_{-\propto}^{\propto} \mathrm{x^2f(x)dx} = \int_0^1 \mathrm{x.2(1-x)dx} = \frac{1}{6}$$

$$\mathrm{Var(X)} = \mathrm{E(X^2)} - [\mathrm{E(X)}]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

Similarly $\mathrm{E(X)} = \frac{1}{3}$, $\mathrm{E(X^2)} = \frac{1}{6}$ and $\mathrm{Var(Y)} = \frac{1}{18}$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy = \int_0^1\int_0^{1-x} xy.2\,dxdy = \frac{1}{12}$$

$$\mathrm{Cov(X,Y) = E(XY) - E(X).E(Y)} = \frac{1}{12} - \frac{1}{9} = -\frac{1}{36}$$

$$\therefore\ \rho_{\,xy} = \frac{\mathrm{Cov(X,Y)}}{\sigma_x\sigma_y} = -\frac{1}{2}$$

25. $\boldsymbol{Suppose\ that\ the\ R.V's\ X\ \&\ Y\ have\ the\ joint\ pdf\ f(x,y) = \begin{cases} e^{-(x+y)}, x > 0, y > 0 \\ 0 \qquad\quad , Otherwise \end{cases}}$

$\boldsymbol{Find\ the\ joint\ pdf\ of\ the\ random\ variables\ U = \frac{x}{y}\, and\ V = y\ and\ hence\ obtain\ the}$

$\boldsymbol{marginal\ pdf\ of\ U}$. **[NOV/DEC 2015]**

Solution: Given $\mathrm{f(x,y)} = \begin{cases} \mathrm{e^{-(x+y)}, x > 0, y > 0} \\ 0 \qquad\quad \mathrm{, Otherwise} \end{cases}$

(i)Find the joint pdf of the random variables $\mathrm{U} = \frac{\mathrm{x}}{\mathrm{y}}$ and $\mathrm{V = y}$

WKT $\mathrm{h(U,V) = |\,J\,|\,f(x,y)}$ Where $\mathrm{J} = \frac{\partial\mathrm{(x,y)}}{\partial\mathrm{(U,V)}}$

Given $\mathrm{U} = \frac{\mathrm{x}}{\mathrm{y}}$ and $\mathrm{V = y \Rightarrow x = UV}$ and $\mathrm{y = V}$

$$J = \begin{vmatrix} \frac{\partial U}{\partial x} & \frac{\partial U}{\partial y} \\ \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} \end{vmatrix} = \begin{vmatrix} V & U \\ 0 & 1 \end{vmatrix} = V$$

$$\therefore \mathrm{h(U,V) = Ve^{-(UV+V)} = Ve^{-V(U+1)}}\ \mathrm{, U, V \geq 0}$$

(ii) The marginal pdf of U

WKT Marginal df of U is $f(U) = \int_{-\propto}^{\propto} f(U,V)dV$

$$= \int_0^{\infty} \mathrm{Ve^{-V(U+1)}}\, dV$$

$$= \left\{ V\left[\frac{\mathrm{e^{-V(U+1)}}}{-(U+1)}\right] - 1.\left[\frac{\mathrm{e^{-V(U+1)}}}{(U+1)^2}\right]\right\}_0^{\infty}$$

$$= \frac{1}{(U+1)^2}\ , U > 0$$

26. $\boldsymbol{A\ random\ sample\ of\ size\ 100\ is\ taken\ from\ a\ population\ whose\ mean\ is\ 60\ and}$ $\boldsymbol{variance\ 400. Using\ CLT, with\ what\ probability\ can\ we\ assert\ that\ the\ mean\ of\ the}$ $\boldsymbol{sample\ will\ not\ differ\ from\ \mu = 60\ by\ more\ than\ 4}$? **[APRIL/MAY 2015]**

Solution: Given $\mu = E(X)$=60, σ^2= Var (X) = 400 & n=100

$\bar{X}$ is sample mean $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ by CLT => $z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Here μ =60, $\frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2$ i.e., $\overline{X} \sim N(60,2)$ => $z = \frac{\bar{X}-60}{2}$

we have to find $P[|\bar{X} - \mu| \leq 4] = P[|\bar{X} - 60| \leq 4]$

= P (-4$\leq \bar{X} - 60 \leq 4$)= P (56$\leq \bar{X} \leq 64$)

=$P\left[\frac{56-60}{2} \leq \frac{\bar{X}-60}{2} \leq \frac{64-60}{2}\right]$ =P (-2$\leq z \leq 2$)

= P (-2$\leq z \leq 0$)+ P (0$\leq z \leq 2$)

=2 P (0$\leq z \leq 2$)=2(0.4772)

$P[|\bar{X} - \mu| \leq 4] = 0.9544$

27.*Compute the coefficient of correlation for the following data* :

X	10	11	13	15	18
Y	60	52	48	40	30

[APRIL/MAY 2015]

UNIT – 3 TESTING OF HYPOTHESIS

- ***LARGE SAMPLE n > 30***

- **Testing of Hypothesis about a population p & P**

  Null hypothesis H_0: P= Population proportion

  p = sample proportion

  Test statistic for z-test about p:z= $\frac{p-P}{\sqrt{\frac{PQ}{n}}}$ where Q = 1-P

  The confidence Interval is $p \pm z_\alpha \sqrt{\frac{PQ}{n}}$

  1) If the sample is from a finite population of size N then

     SE (p) = $\sqrt{\frac{PQ}{n}\left(\frac{N-n}{N-1}\right)}$

  2) If P is not knows assume p=P and SE(p) = $\sqrt{\frac{pq}{n}}$

- **Testing of Hypothesis about the difference between two population P_1 & P_2**

  Null hypothesis H_0: $P_1 = P_2$

  $p_1 =$ sample proportion1

  $p_2 =$ sample proportion2

  z= $\frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where P = $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ is the pooled sample proportion.

- **Testing of Hypothesis about population mean $\bar{x}$ & μ**

  Null hypothesis H_0: $\bar{x} = \mu$

  Sample mean = $\bar{x}$

  Population mean = μ $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

  If the population S.D. “σ” is not given then the sample S.D. “s” is used when n is large

  The confidence Interval is $\bar{x} \pm z_\alpha \left(\frac{s}{\sqrt{n}}\right)$

❖ **Testing of Hypothesis about the difference between two means μ_1& μ_2**

Null hypothesis H_0: $\mu_1 = \mu_2$

	Sample 1	Sample 2
Mean	$\bar{x}_1$	$\bar{x}_2$
Standard Deviation	s_1	s_2

Population variances as σ_1 and σ_2

$$z= \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

1) If the samples are drawn from the same population with common S.D. $\sigma = \sigma_1=\sigma_2$

$$z=\frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

2) If σ_1 & σ_2 are equal and not known the $\sigma^2 = \frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2}$

3) If If $\sigma_1 \neq \sigma_2$ and If σ_1 & σ_2 are not known $\sigma_1=s_1$ & $\sigma_2=s_2$ and $z= \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$

4) The confidence interval for difference of two population is $(\bar{x}_1-\bar{x}_2)\pm z_\alpha \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

❖ **Testing of Hypothesis about the difference between two standard deviations s_1 & s_2**

Null hypothesis H_0: $\sigma_1 = \sigma_2$

Population variances as σ_1 and σ_2

$$z= \frac{s_1-s_2}{\sqrt{\frac{\sigma_1^2}{2n_1}+\frac{\sigma_2^2}{2n_2}}}$$

If σ_1 & σ_2 are known $\sigma_1=s_1$ & $\sigma_2=s_2$ and $z= \frac{s_1-s_2}{\sqrt{\frac{s_1^2}{2n_1}+\frac{s_2^2}{2n_2}}}$

- **Degree of freedom** If n is the number of observation and k is the number of independent constrains (the number of constants that have to be estimated from the original data) then n-k id the degree of freedom

- ***Small SAMPLE n < 30***
- ***t- test***
  - ***Properties***
    - t- distribution ranges from (-∞,∞)
    - t- distribution is bell shaped and symmetrical around mean zero
    - The shape of the t- distribution changes as of degree of freedom changes. Hence the degree of freedom ν is the parameter
    - The variance of the t- distribution is always greater than one and it is defined only when ν ≥ 3 and is $\frac{\nu}{\nu-2}$

- **Testing of Hypothesis about the population mean**
  - Assumptions for t- distribution for population mean
    - The parent population from which the sample is drawn is normal
    - The sample observation are independent
    - The population S.D. σ is not known.
  - **Null hypothesis** H_0: μ = population mean

Sample mean = $\bar{x}$

Population mean = μ

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \text{ with df (n-1) } \quad \bar{x} = \frac{\sum x_i}{n} \quad S = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

1. The confidence interval for population mean is $\bar{x} \pm t_\alpha \left(\frac{S}{\sqrt{n}}\right)$ where t is based on $n-1$ d.f.
2. Is s is known then $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

- **Testing of Hypothesis about the difference between two means μ_1& μ_2**

Null hypothesis H_0: $\mu_1 = \mu_2$

	Sample 1	Sample 2
Mean	$\bar{x}$	$\bar{y}$
Standard Deviation	s_1	s_2
Sample size	n_1	n_2

$$\bar{x} = \frac{\sum x_i}{n_1} \qquad \bar{y} = \frac{\sum y_i}{n_2}$$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2\right]$$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[n_1 s_1{}^2 + n_2 s_2{}^2\right]$$

Degree of freedom is n_1+n_2-2

- Assumptions for t- test for difference of means
  1. Parent population from which the samples have been drawn are normally distributed
  2. The population variance are equal and unknown
  3. The two samples are random and independent of each other.

❖ **Paired t-Test for difference of Means μ1& μ2**

Null hypothesis H_0: **$\mu_1 = \mu_2$**

The samples are drawn from the same population under two different time instants

The test statistics t = $t = \frac{\bar{d}}{S/\sqrt{n}}$ with df n-1where $\bar{d} = \sum \frac{d_i}{n}$ and $S = \sqrt{\frac{1}{n-1}\sum (d_i - \bar{d})^2}$

➢ **F –Test**

The F-distribution is formed by the ratio of two independent chi-square variables divided by their respective degrees of freedom.

❖ Properties of F – Distribution

1. The F – Distribution is positively skewed and its skewness decreases with increase in ν_1 & ν_2
2. The value of F is always be positive or 0
3. The mean of F – Distribution is $\frac{\nu_2}{\nu_2 - 2}$ ν_2>2
4. The shape of F – Distribution is depends upon the number of degree of freedom.

❖ **Testing of Hypothesis for Equality of Two Variances**

Null hypothesis H_0: **$\sigma_1^2 = \sigma_2^2$**

To test the equality of variances in two independently selected random samples drawn from two normal populations with H_0: **$\sigma_1^2 = \sigma_2^2$**

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$

when $S_1 > S_2$

$$S_1 = \sqrt{\frac{1}{n_1 - 1}\sum (x_i - \bar{x}_1)^2}$$

$$S_2 = \sqrt{\frac{1}{n_2 - 1}\sum (x_j - \bar{x}_2)^2}$$

$\nu_1 = n_1 - 1$& $\nu_2 = n_2 - 1$ $S_1 > S_2$

Note To test whether two independent samples have been drawn from same normal population

(i) Equality of population means using t –test

(ii) Equality of population variances using F-test

First apply F test then t-test

If sample variances are given as s_1 and $s_2$the S_1 and S_2 are calculated as

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} \text{ and } S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

- Chi-square distribution
  Definition: A distribution obtained from the multiplying the ratio of sample variance to population variance by the degrees of freedom when random samples are selected from a normally distributed population
- Contingency Table
  Data arranged in table form for the chi-square independence test
- Expected Frequency
  The frequencies obtained by calculation.
- Goodness-of-fit Test
  A test to see if a sample comes from a population with the given distribution.
- Independence Test
  A test to see if the row and column variables are independent.
- Observed Frequency
  The frequencies obtained by observation. These are the sample frequencies.

Goodness-of-Fit Chi-Square Test

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

df = # categories – 1

Chi-Square Test for Independence

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ df = (rows - 1)(columns – 1) Expected value = $\frac{\text{row sum} \times \text{column sum}}{\text{grand total}}$

Probability and Statistics
Formula Sheet

I. Descriptive Statistics

Sample Mean	**Sample Mean**	**Depth of Median**	**Midrange**	**Range**
(List of Data)	**(Freq. Distrib.)**			
$\bar{x}=\frac{\sum x}{n}$	$\bar{x}=\frac{\sum xf}{\sum f}$	$d(x)=\frac{n+1}{2}$	$Midrange=\frac{H+L}{2}$	$R=H-L$

Sample Variance	**Pop. Variance**	**Sample St. Deviation**	**Pop. St. Deviation**
(List of Data)	**(List of Data)**	**(List of Data)**	**(List of Data)**
$s^2=\frac{\sum(x-\bar{x})^2}{n-1}$	$\sigma^2=\frac{\sum(x-\mu)^2}{N}$	$s=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$	$\sigma=\sqrt{\frac{\sum(x-\mu)^2}{N}}$

Variance	**St. Deviation**
(Grouped Data)	**(Grouped Data)**
$s^2=\frac{\sum f\cdot x_m^2-\left(\frac{(\sum f\cdot x_m)^2}{n}\right)}{n-1}$	$s=\sqrt{\frac{\sum f\cdot x_m^2-\left(\frac{(\sum f\cdot x_m)^2}{n}\right)}{n-1}}$

Chebyshev's	**Standard Score**	**Percentile of a Piece of Data**		***k*th Percentile**
	Interquartile			
Theorem	***z*-score**			**Range**
$at\,least\,1-\frac{1}{k^2}$	$z=\frac{x-\mu}{\sigma}$	$\text{percentile}=\frac{\text{number of values below}+0.5}{\text{total number of values}}\cdot 100$	$P_k=\frac{nk}{100}$	$IQR=Q_3-Q_1$

II. Probability

Permutation Rule: The arrangement of *n* objects in a specific order using *r* objects at a time is called a *permutation of n objects taken r objects at a time*. It is written $_nP_r$ and the formula is

$$_nP_r=\frac{n!}{(n-r)!}$$

Combination Rule: The number of combinations of *r* objects selected from *n* objects is denoted by $_nC_r$, or $\binom{n}{r}$, and is given by the formula $_nC_r=\frac{n!}{r!(n-r)!}$.

Empirical Rule	Theoretical	Complement	General	Special Addition
Probability	*Probability*	*Rule*	*Addition Rule*	*Mutually Exclusive Events*
$P'(A)=\frac{n(A)}{n}$	$P(A)=\frac{n(A)}{n(S)}$	$P(\bar{A})=1-P(A)$	$P(A\,or\,B)=P(A)+P(B)-P(A\,and\,B)$	$P(A\,or\,B\,or\ldots D)=P(A)+P(B)+\ldots+P(D)$

General Multiplication Rule

$P(A\,and\,B)=P(A)\cdot P(B|A)$

Special Multiplication Rule for Independent Events

$P(A\,and\,B\,and\,...\,and\,D)=P(A)\cdot P(B)\cdot ...\cdot P(D)$

Conditional Probability

$P(A|B)=\frac{P(A\,and\,B)}{P(B)}$

Mean of a Prob. Distribution

$\mu=\sum[x\cdot P(x)]$

Variance of a Prob. Distribution

$\sigma^2=\sum\left[x^2\cdot P(x)\right]-\mu^2$

St. Deviation of a Prob. Distribution

$\sigma=\sqrt{\sum\left[x^2\cdot P(x)\right]-\mu^2}$

Binomial Prob. Function

$P(x)=\binom{n}{x}\cdot p^x\cdot q^{n-x}\ for\ x=0,1,2,...$

Mean of a Binomial Random Variable

$\mu=np$

Variance of a Binomial Random Variable

$\sigma^2=npq$

Standard Deviation of a Binomial Random Variable

$\sigma=\sqrt{npq}$

PART – A

1. **Define Sample**
   A finite subset of statistical individuals in a population is called sample.
2. **Define Sample Size**:
   The number of individuals in a sample is called the sample size.
3. **Define Parameter**
   Any statistical constant which is computed by considering each and every observation of the population is called as a parameter.
4. **Define Statistic**
   Any statistical constant which is computed by considering a part of the information from the population is called as a statistic.
5. **What is sampling distribution of the statistic?**
   The Probability distribution of a statistic is called a sampling distribution.
6. **Define Standard Error**
   The standard deviation of sampling distribution of a statistic is known as its standard error and it is denoted by (S.E)
7. **Explain Test of Significance**
   A very important aspect of the sampling theory is the study of tests of significance which enable us to decide on the basis of the sample results, if
   (i) The deviation between the observed sample statistic and the hypothetical parameter value is significant.
   (ii) The deviation between two sample statistics is significant.
8. **Define Null Hypothesis**
   A null hypothesis is denoted as H_0 and it is assumed to be true by carrying out the test procedure. The null hypothesis always assumes that there is no significant difference between the things that are compared.
9. **Define Alternative Hypothesis**
   It is a contradiction to the null hypothesis. It is represented as H_1. It may be an inequality (or) greater than (or) less than type. The alternate hypothesis decides the nature of test under consideration.

10. Define Type I error and Type II error.

Type I error: Rejecting H_0, when H_0 is true

Type II error: Accepting H_0, when H_0 is false.

11. What is Producer's risk and Consumer's risk?

The sizes of Type I error and Type II errors are called Producer's risk and Consumer's risk respectively.

12. Explain Level of Significance

Level of significance is the probability of type I error. It is denoted by α. It is the size of the rejection region. The significance of α is that it gives the percentage of cases in which the conclusion derived on the basis of sample will go wrong.

13. Define Critical Region

In tests of significance, the total area of the probability curve is divided into two regions (i) Acceptance and (ii) Rejection. The rejection region is also called as the critical region and its size is equal to α. Consequently the size of the acceptance region will be 1-α.

14. Define Critical Value

The value of the test statistic which divides the rejection region and the acceptance region is called the critical value. The Critical values is usually extracted from the statistical tables.

15. Define Student's t-Distribution

A random variable T is said to follow student's t-Distribution or simply t-distribution, if its pdf is given by

$f(t) = \frac{1}{\sqrt{v}\beta\left(\frac{v}{2},\frac{1}{2}\right)}\left(1+\frac{t^2}{v}\right)^{\frac{-(v+1)}{2}} - \infty < t < \infty$, where v denotes the number of degrees of freedom of the t-distribution.

16. Uses of t-distribution

The t-distribution is used to test the significance of the difference between

(i) The mean of a small sample and the mean of the population

(ii) The means of two small samples and

(iii) The coefficient of correlation in the small sample and that in the population, assumed zero.

17. State the properties of t – distribution

(i) The probability curve of the t-distribution is similar to the standard normal curve, and is symmetric about t=0, bell-shaped and asymptotic to the t-axis

(ii) For sufficiently large value of v, the t-distribution tends to the standard normal distribution.

(iii) The mean of the t-distribution is zero.

(iv) The variance of the t-distribution is $\frac{v}{v-2}$, if v>2 and is greater than 1, but it tends to 1 as v→∞

18. Define Snedecor's F-Distribution

A random variable F is said to follow snedecor's F-distribution or simply F-distribution if its pdf is given by

$f(F) = \frac{(v_1|v_2)^{v_1/2} F^{v_1/2-1}}{\beta\left(\frac{v_1}{2},\frac{v_2}{2}\right)\left(1+\frac{v_1F}{v_2}\right)^{(v_1+v_2)/2}}$, F>0.

19. Write down the value of χ^2 for 2 X 2 contingency table with cell frequencies a, b, c and d.

The 2 X 2 contingency table for χ^2 test is $\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$

PART-B

1. **An automatic machine fills tea in sealed tins with mean weight of tea 1 kg and standard deviation of 1gm. A random sample of 50 tins was examined, and it was found that their mean weight was 999.50g. State whether the machine is working properly or not.**

Solution: $\bar{x}$= sample mean = 999.50 μ = population mean = 1000 σ = population standard deviation = 1 n= 50 large sample

H$_0$= the machine work properly i.e $\bar{x}$ = μ

H$_1$= the machine not work properly i.e $\bar{x}$ ≠μ

LOS = 5 %

Test statistics: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{999.50-1000}{\frac{1}{\sqrt{50}}}$ = =-3.53535 The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie the machine not work properly as Z cal is greater than Z tab

2. **The weights of fish in a certain pond that is regularly stocked are considered to be normally distributed with a mean of 3.1 kg and a standard deviation of 1.1kg. A random sample of size 30 is selected from the pond and the sample mean is found to be 2.4kg. Is these sufficient evidence to indicate that the mean weight of the fish differs from 3.1kg? Use a 10% significance level.**

Solution: $\bar{x}$= sample mean = 2.4 μ = population mean = 3.1 σ = population standard deviation = 1.1 n= 30 large sample

H$_0$= The mean weight of the fish does not differ significantly i.e $\bar{x}$ = μ

H$_1$= The mean weight of the fish differ significantly i.e $\bar{x}$ ≠μ

LOS = 10 %

Test statistics: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.4-3.1}{\frac{1.1}{\sqrt{30}}}$ = =-3.48

The critical value of z =1.645

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie The mean weight of the fish differ significantly as Z cal is greater than Z tab

3. **A Stenographer claims that she can type at a rate of 120 words per minute. Can use reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with a standard deviation of 15 words? Use 5% level of significance**.

Solution: $\bar{x}$= sample mean = 116 μ = population mean = 120 σ = population standard deviation = not given

n= 100 large sample s= Sample standard deviation = 15

H$_0$= The mean typing speed =120

H₁= The mean typing speed not 120

LOS = 5 % Test statistics: $Z = \frac{\bar{x}- \mu}{\frac{s}{\sqrt{n}}} = \frac{116-120}{\frac{15}{\sqrt{100}}} = = -2.66$

The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie the typing speed is not 120 as Z cal is greater than Z tab

4. **In a certain factory there are two different processes manufacturing the same item. The average weight in sample of 250 items produced from one process is found to be 120 gms with a s.d of 12 gms; the corresponding figures in a sample of 400 items from the other process are 124 and 14. In there any significant difference between the averages of two samples at 1% level of significance?**

Solution n1= 250 $\bar{x}1 = 120$ s1 =12

n2= 400 $\bar{x}2 = 124$ s2 =14

H_0= there is no significant difference between the average of two samples $\mu_1=\mu_2$

H_1= there is a significant difference between the average of two samples $\mu_1 \neq \mu_2$

LOS = 5 %

Test statistics: $z= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{120-124}{\sqrt{\frac{12^2}{250} + \frac{14^2}{124}}} = -3.87$

The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis ie the typing speed is not 120 as Z cal is greater than Z tab

5. **The average annual pay in 1989 was Rs 21,128 in the state of Tamil Nadu and Rs. 25,233 in the state of Maharastra. There is a difference of Rs 4,105. Suppose that a statistician believes that the difference is much less for employees in the manufacturing industry and takes an independent random sample of employees in the manufacturing industry in each state. The results are as follows:**

At the 0.05 significance level, do the data support the statistician's belief that for employees in the manufacturing industry, the mean annual salary in Tamil Nadu differs from the mean annual salary in Maharastra by less than Rs 4105?

State	$\bar{x}$	s	**n**
TamilNadu	**21,900**	**3,700**	**150**
Maharashtra	**24,800**	**3,100**	**190**

Solution

n1= 150 $\bar{x}1 = 21900$ s1 =3700

n2= 190 $\bar{x}2 = 24800$ s2 =3100

H_0= there is no significant difference between the average of two samples μ_1-μ_2=4105

H_1= there is a significant difference between the average of two samples μ_1-$\mu_2 \neq$4105

LOS = 5 %

Test statistics: $z= \frac{(x_1 - x_2) - \mu 1 - \mu 2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(21900 - 24800) - 4105}{\sqrt{\frac{3700^2}{150} + \frac{3100^2}{190}}}$ = -18.5995

The critical value of z =1.96

Conclusion: reject the null hypothesis and accept the alternative hypothesis that there is a significant difference between the averages of two samples

6. **A college conducts both day and night classes intended to be identical. A sample of 100 day students' field estimation results as below: $\overline{x_1} = 72.4 \quad and \quad \sigma_1 = 14.8$ A sample of 200 night students' field examination results as below: $\overline{x_2} = 73.9 \quad and \quad \sigma_2 = 17.9$ Are the two means statistically equal at 10% level of significance?**

Solution n_1=100 $\overline{x_1} = 72.4 \quad and \quad \sigma_1 = 14.8$

n_2=200 $\overline{x_2} = 73.9 \quad and \quad \sigma_2 = 17.9$

H_0= there is no significant difference between day and night classes μ_1=μ_2

H_1= there is a significant difference between day and night classes $\mu_1 \neq \mu_2$

LOS = 5 %

Test statistics: $z= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{12^2}{250} + \frac{14^2}{124}}}$ = -0.77025

The critical value of z =1.96

Conclusion: Z cal is less than Z critical hence accept the null hypothesis that there is no significant difference between day and night classes

7. **A soap manufacturing company was distributing a particular brand of soap through a large no of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop were 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with s.d =16. Can you consider the advertisement effective?**

Solution: <u>Null hypothesis</u> $H_0:\mu=140$ there is no effectiveness in advertisement

$H_1:\mu>140$ there is an effectiveness in advertisement

n = 26 s = 16, $\bar{x}$ =147

$$t=\frac{\bar{x}-\mu}{s/\sqrt{n-1}}=\frac{147-140}{16/\sqrt{25}}=2.19$$

$t_{tab(25)}$= 1.708 Conclusion: $t_{cal}>t_{tab}$ reject the H_0 that is there is an effectiveness in advertisement

8. **A delivery service is considering delivering Arun's ice-cream if the average order in a suburban area is greater than 1.5 liters of ice-cream. A random sample of 23 household order yield a mean of 1.7liters with a standard deviation of 0.5 liters. Test at 0.05 significance level that the mean household order is greater than 1.5 liters.**

Solution: <u>Null hypothesis</u> $H_0:\mu=1.5$ average order in a suburban area is 1.5 liters of ice-cream

$H_1: \mu > 1.5$ average order in a suburban area is >1.5 liters of ice-cream

n = 23 s = 0.5, $\bar{x}$ =1.7

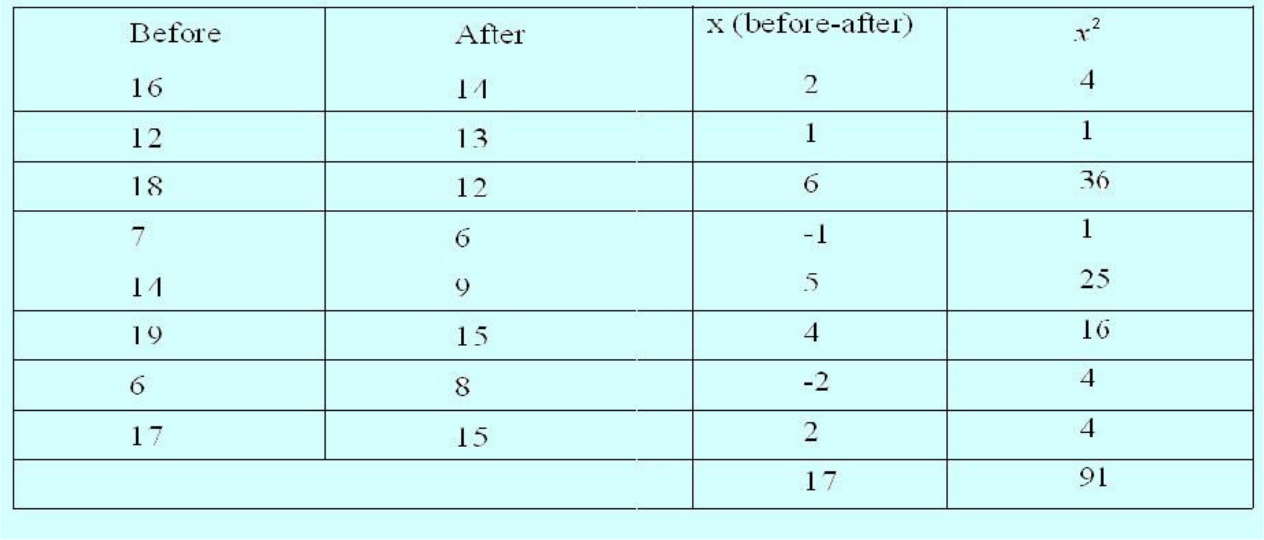
$$t=\frac{\bar{x}-\mu}{s/\sqrt{n-1}}=\frac{1.7-1.5}{0.5/\sqrt{23}}=1.918$$

$t_{tab(22)}$= 1.717

Conclusion: $t_{cal}>t_{tab}$ reject the H_0 that is average order in a suburban area is >1.5 liters of ice-cream

9. **A chemical products distributor is faced with the continuing problem of broken glassware and flasks. He has determined some additional shipping precaution and asked the purchasing director to inform the supplies of these precautions. Data for 8 suppliers are given below in terms of average number of broken items per shipment. Do the data indicates, at α=0.05, that the new measures have lowered the average number of broken item?**

Supplier	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Before	**16**	**12**	**18**	**7**	**14**	**19**	**6**
After	**14**	**13**	**12**	**6**	**9**	**15**	**8**

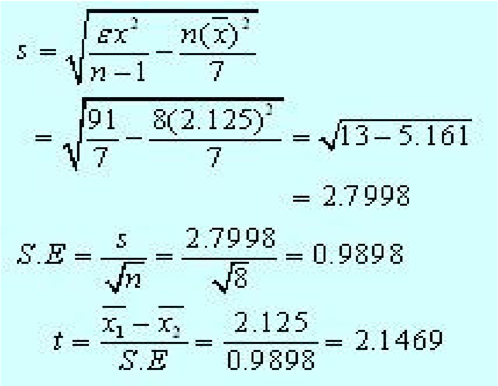


Before	After	x (before-after)	x^2
16	14	2	4
12	13	1	1
18	12	6	36
7	6	-1	1
14	9	5	25
19	15	4	16
6	8	-2	4
17	15	2	4
		17	91

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$ $\alpha = 5\%$

$$X = \frac{\varepsilon x^2}{n} = \frac{17}{8} = 2.125$$



$$s = \sqrt{\frac{\varepsilon x^2}{n-1} - \frac{n(\bar{x})^2}{7}}$$

$$= \sqrt{\frac{91}{7} - \frac{8(2.125)^2}{7}} = \sqrt{13 - 5.161}$$

$$= 2.7998$$

$$S.E = \frac{s}{\sqrt{n}} = \frac{2.7998}{\sqrt{8}} = 0.9898$$

$$t = \frac{\overline{x_1} - \overline{x_2}}{S.E} = \frac{2.125}{0.9898} = 2.1469$$

H_0 is rejected. The new measures have not lowered the average no. of broken items.

10. A lawn-equipment shop is considering adding a brand of lawn movers to its merchandise. The manager of the shop believes that the highest quality lawn movers are Trooper, lawn eater and Nipper, and he needs to decide whether it makes a difference which of these three shop adds to its existing merchandise. Twenty owners of each of these three types of lawn movers are randomly sampled and asked how satisfied they are with their lawn movers

Lawn mover	**Very satisfied**	**Satisfied**	**Not satisfied**	**Total**
Trooper	**11**	**6**	**3**	**20**
Lawn eater	**13**	**4**	**3**	**20**
Nipper	**13**	**6**	**1**	**20**

Are the owners of the lawn movers homogeneous in their response of the survey? Use a 5% significance level.

Solution :

H_0: Owners of lawn movers homogenous in this response to the server.

H_1: Owners of the lawn movers are not homogenous in the response to the server.

We combine satisfied and not satisfied category as not satisfied category has value less than S.

Observed	Expected	$(\text{Obse-Exp})^2/\text{Exp}$
11	12	0.0833
9	8	0.1250
13	12	0.0833
7	8	0.1250
13	12	0.0833
7	8	0.1250
Total		0.0625

Eater from X^2 table for

df = (r - 1) (C - 1) =(3 - 1)(2 - 1) = 2

$\alpha = 0.05$

$x_u^2 = 5.991$

$(x^2 = 0.625) < (x_u^2 = 5.991)$

H_0 is accepted. Hence the owners of the Lawn Movers are homogeneous in their response to the survey.

11. In an industry, 200 workers, employed for a specific job, were classified according their performance and training received /not received to test independence of a specific training and performance. The dates is Performance

	Good	**Not Good**	**Total**
Trained	**100**	**50**	**150**
Untrained	**20**	**30**	**50**
Total	**120**	**80**	**200**

5% level of significance

Solution :H_0: Worker are homogenous

H_1: Worker are not homogenous in.

We combine satisfied and not satisfied category as not satisfied category has value less than S.

Observed	Expected	(Obse-Exp)^2/Exp
100	90	1.111111111
50	60	1.666666667
20	30	3.333333333
30	20	5
Total		11.11111111

Eater from X^2 table for

df = (r - 1) (C - 1) =(2 - 1)(2 - 1) = 1 α= 0.05

$x_u^2 = (x^2 = 11.11111111) > (x_u^2 =)$

H_0 is Rejected. Hence the the workers are not homogeneous.

12. To see whether silicon chip sales are independent of where the U.S economy is in the business cycle, data have been collected on the weekly sales of Zippy Chippy, a silicon Valley firm , and on whether the U.S economy was rising to a cycle peak, falling to a cycle through, of at a cycle through the results are Economically weakly chip sales:

	High	**Medium**	**Low**	**Total**
At Peak	**8**	**3**	**7**	**18**
Rising	**4**	**8**	**5**	**17**
Falling	**8**	**4**	**3**	**15**
Total	**20**	**15**	**15**	**50**

Soln:

O_i	E_i	$(O_i - E_i)^2/ E_i$
8	7.2	0.0889
4	6.8	1.153
8	6.0	0.6667
3	5.4	1.0667
8	5.1	0.056
4	4.5	0.474
7	5.4	0.0020
5	.1	0.5
3	4.5	5.657

d.f = (r - 1) (c - 1) = 2 x 2 = 4

Table Value = 7.779 for = 0.10

13. An automatic machine fills tea in sealed tins with mean weight of tea 1 kg and standard deviation of 1gm. A random sample of 50 tins was examined, and it was found that their mean weight was 999.50g. State whether the machine is working properly or not.

Solution:

$\bar{x}$= sample mean = 999.50; μ = population mean = 1000;

σ= populationstandard deviation = 1; n= 50 large sample

i. H_0= The machine work properly i.e$\bar{x}$ = μ
ii. H_1= The machine not work properly i.e$\bar{x} \neq$μ
iii. LOS = 5 %
iv. Test statistics:Z = $\frac{\bar{x}- \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{999.50- 1000}{\frac{1}{\sqrt{50}}}$ = =-3.53535

The critical value of z =1.96

v. Conclusion: reject the null hypothesis and accept the alternative hypothesis

i.e the machine not work properly as Z cal is greater than Z tab.

14. .Test significance of the difference between the means of the samples, drawn from two normal populations with the same SD usingthe following data :

	Size	***Mean***	***SD***
Sample-1	***100***	***61***	***4***
Sample-2	***200***	***63***	***6***

Solution:

Given

	Size	Mean	SD
Sample-1	100	61	4
Sample-2	200	63	6

We want test that the samples are drawn from a population with same SD.

i. $H_{0:}\mu_1 = \mu_2$
ii. $H_{1:}\mu_1 \neq \mu_2$
iii. LOS = 5 %
iv. Test statistics :Z = $\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{61-63}{\sqrt{\frac{16}{100}+\frac{36}{200}}} = -3.43 \Rightarrow |Z| = 3.43$

The table value of Z at 5% level is 1.96

Calculated value >The table value.

v. Conclusion: reject the null hypothesis H_0.

15. Two random samples of sizes 400 and 500 have mean 10.9 and 11.5 respectively. Can the samples be regarded as drawn from the same population with variance 25?

Solution: Given

	Size	Mean	Variance
Sample-1	400	10.9	25
Sample-2	500	11.5	25

We want test that the samples are drawn from a population with variance 25

i. H0:$\mu_1 = \mu_2$

ii. H1:$\mu_1 \neq \mu_2$

iii. LOS = 5 %

iv. Test statistics :Z = $\frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{10.9-11.5}{5\sqrt{\frac{1}{400}+\frac{1}{500}}} = -\frac{12}{\sqrt{45}} = -1.78$

$$|Z| = 1.78$$

The table value of Z at 5% level is 1.96

Calculated value < The table value.

v. Conclusion: Accept the null hypothesis H0.

16.A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.Does the data support the assumption of a population mean IQ of 100?

Solution: Given n=10 and $\mu = 100$

i. H0:$\mu = 100$

ii. H1:$\mu = 100$

iii. LOS = 5 %

x	70	120	110	101	88	83	95	98	107	100	$\sum x$=972
x^2	70^2	120^2	110^2	101^2	88^2	83^2	95^2	98^2	107^2	100^2	$\sum x^2$ =96312

$\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$; $s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = 183.96 \Rightarrow s = 13.5$

iv. Test statistics : $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{97.2-100}{\frac{13.5}{\sqrt{9}}} = -\frac{2.8}{4.5} = -0.62$

The table value of t at 5% level is 2.262

Calculated value<The table value.

v. Conclusion: Accept the null hypothesis H0.

17. In a sample of 8 observations, the sum of the squared deviations of items from the mean was 94.5. In another sample of 10 observations, the value was found to be 101.7. Test whether the difference in the variances is significant at 5% level.

Solution: Given

$$n_1 = 8 \; and \; n_2 = 10 \quad ; \quad \sum(x-\bar{x})^2 = 94.5 \;\; and \;\; \sum(y-\bar{y})^2 = 101.7$$

The sample variances are $s_1^2 = \frac{\sum(x-\bar{x})^2}{n_1} = \frac{94.5}{8}$ and $s_2^2 = \frac{\sum(y-\bar{y})^2}{n_2} = \frac{101.7}{10}$

i. H₀: $\sigma_1^2 = \sigma_2^2$

ii. H₁: $\sigma_1^2 \neq \sigma_2^2$

iii. LOS = 5 % ; d.f is $F_{(7,9)}$

iv. Calculate the population variances

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = 13.5 \quad and \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = 11.3$$

$$\text{here } S_1^2 > S_2^2 \qquad \therefore F = \frac{S_1^2}{S_2^2} = 1.174$$

The table value of $F_{(7,9)}$ =3.68 at 5% LOS

Hence calculated value < The table value

v. Conclusion: Accept the null hypothesis H_0.

18. The following random sample are measurement of heat-producing capability(in millions of calories per ton) of specimens of coal from two mines:

Mine I : 8260 8130 8350 8070 8340
Mine II: 7950 7890 7900 8140 7920 7840

Use the a=0.01 level of significance to test whether the difference between the means of these two sample is significant?

Solution:

i. H₀: $\mu_1 = \mu_2$

ii. H₁: $\mu_1 \neq \mu_2$

iii. LOS = 1 % d.f=$n_1+n_2-2=9$

$$\sum x_1 = 41150 \ and \sum x_2 = 47650$$

$$\sum x_1^2 = 338727500 \ and \sum x_2^2 = 378316200$$

$$\bar{x}_1 = 8230 \ and \ \bar{x}_2 = 7940$$

$$s_1^2 = \frac{\sum x_1^2}{n_1} - \left(\frac{\sum x_1}{n_1}\right)^2 = 12600 \ and \ s_2^2 = \frac{\sum x_2^2}{n_2} - \left(\frac{\sum x_2}{n_2}\right)^2 = 9100$$

$$S^2 = \frac{n_1 s_1^2 + + n_2 s_2^2}{n_1 + n_2 - 2} = 13066.67$$

iv. Test statistics : $t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{8230-7940}{114.31\sqrt{\frac{1}{5}+\frac{1}{6}}} = 4.19$

Calculated value < The table value.

v. Conclusion: Accept the null hypothesis H_0.

19. Two scientists, Dr.X and Dr. Y, find a previously unknown type of fish in a remote river. They both trap some Fish, from different location some distance apart. The weights of their fish are as follows(in kg):

Dr.X : 0.15 0.18 0.25 0.36 0.42 0.44

Dr.Y: 0.25 0.26 0.26 0.30 0.32 0.33 0.37 0.37

Do the figure support, at the5% significance level, the theory that the fish came from populations with the same variation?

Solution:

vi. H0:$\sigma_1^2 = \sigma_2^2$

vii. H1:$\sigma_1^2 \neq \sigma_2^2$

viii. LOS = 5 % ; d.f is $F_{(5,7)}$

ix. Calculate the mean and variance

Dr. X		Dr. Y	
x	x*x	y	y*y
0.15	0.023	0.25	0.063
0.18	0.032	0.25	0.063
0.25	0.063	0.26	0.068
0.36	0.130	0.30	0.090
0.42	0.176	0.32	0.102
0.44	0.194	0.33	0.109
		0.37	0.137
		0.37	0.137
1.8	0.617	1.71	0.4939

$$n_1 = 6 \; and \; n_2 = 8 \qquad ; \qquad \bar{x} = 0.3 \; and \; \bar{y} = 0.308$$

$$s_1^2 = \frac{\sum x^2}{n_1} - \left(\frac{\sum x}{n_1}\right)^2 = 0.193 \quad and \quad s_2^2 = \frac{\sum y^2}{n_2} - \left(\frac{\sum y}{n_2}\right)^2 = 0.0024$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = 0.23 \quad and \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = 0.003$$

$$\text{here } S_1^2 > S_2^2 \qquad \therefore F = \frac{S_1^2}{S_2^2} = 76.67$$

The table value of $F_{(5,7)}$ =4.87 at 5% LOS

Hence calculated value > The table value

x. Conclusion: Reject the null hypothesis H0.

20. *The following is the distribution of the number of trucks arriving at a Company s warehouse.*

Tracks arrival /hour:	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
Frequency	***: 52***	***151***	***130***	***102***	***45***	***12***	***5***	***1***	***2***

Find the mean of this distribution and using it fit a Poisson distribution. Test for goodness of fit at the 0.05 level of significance.

Solution:

i. H_0= Poisson fit is a good fit

ii. H_1= Poisson fit is not a good fit

iii. LOS = 5 % ; d.f=9-1-1-2=5

iv. Test statistics

$$\text{Mean } \lambda = \frac{\sum fx}{\sum f} = \frac{1040}{500} = 2.1$$

$$P(r) = Ne^{-\lambda}\frac{(\lambda^r)}{r!} \;;\quad N=500$$

P(0)=61, P(1)=129, P(2)=135, P(3)=94, P(4)=50,

P(5)=21, P(6)=7, P(7)=2, P(8)=1

Hence the frequency table is

x	0	1	2	3	4	5	6	7	8	total
O_i	52	151	130	102	45	12	5	1	2	500
E_i	61	129	135	94	50	21	7	2	1	500

The Test statistics $\chi^2 = \sum\frac{(O-E)^2}{E}$

O_i	E_i	$(O_i - E_i)^2/E_i$
52	61	1.33
151	129	3.75
130	135	0.19
102	94	0.68
45	50	0.5
12	21	3.86
5	7	0.4
1	2	
2	1	

$$\chi^2 = 10.705$$

From the χ^2 table at 5% level for 5 d.f =11.07

Calculated value < The table value.

v. Conclusion: Accept the null hypothesis H_0

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

Solution:

i. H_0= The proportion of beans in the four groups should be 9:3:3:1

ii. H_1= The proportion of beans in the four groups should not be 9:3:3:1

iii. LOS = 5 %

iv. Test statistics: Given that the observed frequencies are respectively 882, 313, 287, and 118. The total observed frequencies=882+313+287+118=1600.

Under $H_0$the expected frequencies are 900, 300, 300 and 100

The Test statistics$\boldsymbol{\chi^2} = \sum \frac{(\boldsymbol{O}-\boldsymbol{E})^2}{\boldsymbol{E}}$

O_i	E_i	$(O_i - E_i)^2/E_i$
882	900	0.360
313	300	0.563
287	300	0.563
118	100	3.240

$$\chi^2 = 4.727$$

Degrees of freedom *v*=n-1=3

From the χ^2 table at 5% level for 3 d.f =7.81

Calculated value < The table value.

iv. Conclusion: Accept the null hypothesis H_0

22. In a survey on class mobility, a research student interviews 200 men, asking them about their jobs and those of their fathers. He wishes to know whether the employment class of a man is related to that of his father. He classifies the results, using his own scale, as the following contingency tale:

Father

		Upper	Middle	Lower
Son	Upper	16	25	19
	Middle	15	33	22
	Lower	19	22	29

Use the a=0.01 level of significance to test whether there is relationship (dependence) between the employment class of a man and his father?

Solution: Given

	Upper	Middle	Lower	Total
Upper	16	25	19	60
Middle	15	33	22	70
Lower	19	22	29	70
Total	50	80	70	200

i. H_0: There is a relationship between the employment class of a man and his father.

ii. H_1: There is no relationship between the employment class of a man and his father.

iii. LOS=5%

O_i	E_i	$(O_i - E_i)^2/E_i$
16	15	0.07
25	24	0.04
19	21	0.19
15	17.5	0.36
33	28	0.89
22	24.5	0.26
19	17.5	0.13
22	28	1.29
29	24.5	0.83

$$X^2 = 4.06$$

d.f = (r - 1) (C - 1) =(3 - 1)(3 - 1) = 4

iv. From the χ^2 table at 5% level for 4 d.f =2.776

Calculated value >The table value.

v. Conclusion: Reject the null hypothesis H_0

UNIT -4 DESIGNS OF EXPERIMENT

Define Anova?

Anova is separation of variance ascribable to one group of causes from the variance ascribable to other group

Some Important Abbreviations:

- SSC- Between sum of squares (Column)
- TSS- Total sum of squares
- SST- Sum of squares due to Treatments
- MSS- Mean Sum of squares
- SSE- Error Sum of squares (or) Within Sum of squares
- RSS- Row Sum of squares
- CF- Correction Factor
- CD- Critical Difference
- SSR- Sum of squares between Rows
- MSC- Mean Sum of squares (Between Columns)
- MSE- Mean Sum of squares (within Columns)
- MSR-Mean Sum of squares (Between Rows)
- N1- Number of Elements in each Column
- N2- Number of Elements in each Row

Anova Table For One way Classification (C.R.D):

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio
Between Columns	SSC	C-1	$MSC = \frac{SSC}{C-1}$	$F = \frac{MSC}{MSE}$ (OR) $F = \frac{MSE}{MSC}$
Within Columns	SSE	N-C	$MSE = \frac{SSE}{N-C}$	
Total	TSS	N-1	Condition Always F >1	

Anova Table For Two way Classification :

Randomized Block Design (R.B.D)

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio
Column Treatment	SSC	C-1	$MSC = \frac{SSC}{C-1}$	$F_c = \frac{MSC}{MSE}$
Row Treatment	SSR	R-1	$MSR = \frac{SSR}{R-1}$	$F_R = \frac{MSR}{MSE}$
Error	SSE	N-C-R+1	$MSE = \frac{SSE}{(R-1)(C-1)}$	
Total	TSS		Condition Always F >1	

Anova Table For Three way Classification :

Latin Square Design (L.S.D)

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio
Between Columns	SSC	K-1	$MSC = \frac{SSC}{K-1}$	$F_c = \frac{MSC}{MSE}$ $F_R = \frac{MSR}{MSE}$ $F_T = \frac{MST}{MSE}$
Between Rows	SSR	K-1	$MSR = \frac{SSR}{K-1}$	
Between Treatments	SST	K-1	$MST = \frac{SST}{K-1}$	
Error	SSE	(K-1)(K-2)	$MSE = \frac{SSE}{(K-1)(K-2)}$	
Total	TSS	$K^2 - 1$	Condition Always F >1	

Merits & Demerits of C.R.D:

- Merits
  - It has a simple layout
  - The analysis the design as it results in a one way classification analysis of variance
  - There is complete flexibility as the number of replication is not fixed
  - Analysis can be performed if some observations are missing
- Demerits:
  - The Experimental error is large as compare to other designs because homogeneity of the units is not taken into consideration

Merits & Demerits of R.B.D:

- Merits
  - It has a simple layout but it is more efficient than CRD because of reduction of experimental error
  - Analysis is possible even if some observations are missing
  - It is flexible and so any number of treatments and any number of replication may be used
  - The analysis of design is simple as its results in a two way classification of analysis of variance
  - This is more popular design with experiments because of its simplicity, flexibility and validity
- Demerits:
  - If the number of treatments is large, then the size of the block will increase this may causes heterogeneity within the blocks
  - The shape of experimental material should be Rectangle
  - If the interation are large, the experiment may yields misreading results

Merits & Demerits of L.S.D:

- Merits
  - The analysis of design is simple as its results in a three way classification of analysis of variance
  - LSD controls variations in two directions of the experimental materials as row and column resulting in the reduction of experimental error

  - The analysis of remains relatively simple even with missing data

- Demerits:
  - The number of treatments should be equal to the number of rows and columns as the area should be in square form
  - It is suitable only for smaller number of treatments say between 5 to 12
  - 2x2 Latin square is not possible
  - The process of randomization is not as simple as RBD

Some Important Formulas:

- $Correction\ Factor\ (CF) = \frac{T^2}{N}$, Where N=Number of Data's given in the problem, T = Total
- $TSS = \left[\left(\sum_{i=1,2,3\ldots} X_i\right) - \frac{T^2}{N}\right], i\ ranges\ from\ number\ of\ colums\ given\ in\ the\ Pb$
- $SSC = \left[\left(\frac{\left(\sum_{i=1,2,3\ldots} X_i\right)^2}{N_1}\right) - \frac{T^2}{N}\right], i\ ranges\ from\ number\ of\ colums\ given\ in\ the\ Pb$
- $SSR = \left[\left(\frac{\left(\sum_{i=1,2,3\ldots} Y_i\right)^2}{N_2}\right) - \frac{T^2}{N}\right], i\ ranges\ from\ number\ of\ rows\ given\ in\ the\ Pb$
- SSE = TSS – SSC (OR) SSE = TSS – SSR, Based on the Problem (For CRD)
- SSE = TSS – SSR- SSC (For RBD)
- SSE = TSS – SSR-SSC-SST (For LSD)

  Note:
  - ✓ $N_1 = number\ of\ elements\ in\ each\ Column$
  - ✓ $N_2 = number\ of\ elements\ in\ each\ Row$

- **Compare RBD and LSD:**

S.No.	LSD	RBD
1	The number of replication of each treatment is equal to the number of treatments in LSD	There are no such restrictions on treatments and replication in RBD.
2	LSD can be performed on a square field.	While RBD can be performed either on a square field or a rectangle field.
3	LSD is known to be suitable for the case when the number of treatments is between 5 and 12	RBD can be used for any number of treatments.
4	The main advantage of LSD is that it controls the effect of two extraneous variables.	RBD controls the effect of only one extraneous variable. Hence the experimental error is reduced to a larger extent in LSD than in RBD.

- **Name the basic principles of experimental design.**

  There are three basic principles of experimental design. They are:

  (i) Randomization

  (ii) Replication

  (iii) Local Control

- **Define Randomization:**

  It ensures that each treatment gets an equal chance of being allocated. Consequently randomization eliminates the bias of any form.

- **Define Replication:**
  By replication we mean, the repetition of the treatments under investigation. Due to replication more reliable estimates can be made available. To be more precise as the replication increases the experimental error decreases.
- **Define Local Control:**
  It is a process of reducing the experimental error by dividing the heterogeneous experimental area into homogeneous blocks.
- **Define "Analysis of Variance" (or) ANOVA.**
  According to R.A. Fisher, Analysis Of Variance (ANOVA) is the separation of variance ascribable to one group of causes from the variance ascribable to other groups.
- **Define "experimental error".**
  The estimation of the amount of variation due to each of the independent factors separately and then comparing these estimates due to assignable factors with the estimate due to the chance factor is known as experimental error or simple error.
- **What do you mean by one-way classification in analysis of variance?**
  In one-way classification the data are classified according to only one criterion (or) factor.
- **Explain the meaning and use of Analysis of Variance?**
  Analysis of variance to test the homogeneity several means.
  Uses:
  (i) It helps to find out the F-test
  (ii) Between the samples we can find the variances.
- **Define the term Completely Randomized Design.**
  The completely randomized design is the simplest of all the designs, based on principles of randomization and replication. In this design, treatments are allocated at random to the experimental units over the entire experimental.
- **What is a randomized block design?**
  Let us consider an agricultural experiment using which we wish to test the effect of 'k' fertilizing treatments on the yield of crops. We assume that we know some information about the soil fertility of the plots. Then we divide the plots into 'h' blocks. Thus the plots in each block will be of homogeneous fertility as far as possible within each block, the 'k' treatments are given to the 'k' plots in a perfectly random manner, such that each treatment occurs only once in any block. But the same k treatments are repeated from block to block. This design is called Randomized Block Design
- **State the differences between CRD and RBD.**

S.No.	CRD	RBD
1	This design provides a one-way classified data according to levels of a single factor namely 'treatment'	The analysis of the design is simple and straight forward as in the case of two-way classification.
2	It has a simple layout	The analysis of this decision is not as simple as a completely randomized design.
3	Grouping of the experimental site so as to allocate the treatments at random to the experimental units is not done	Treatments are allocated at random within the units of each stratum.

- **Define Latin Square Design:**
  Here for k treatments we should have k^2 experimental units arranged in a square. So that each row as well as each column contains k units. Such a layout is known as k x k latin square design. The treatments should be allocated in a random manner in such a way that each treatment occurs in each row and each column.
- **What are the basic principal of experimental design? (Apr/May 2015 (R2013 & R2008))**

Soln: There are three basic principles of experimental design. They are:

(iv) Randomization
(v) Replication
(vi) Local Control

- **Is 2×2 Latin square is possible? Why? (Apr/May, & Nov/Dec 2015)**

Soln: No because SSE degree of freedom is (k-1)(k-2) where k is number of rows and columns
In 2X2 Latin square k=2 so SSE degree of freedom is zero, If error is zero we are not able to find calculated value of F

- **Define (a) Mean Square (b) Complete randomized design (Nov/Dec 2015)**

Soln: (a) mean squares are used to determine whether factors (treatments) are significant. The treatment mean square is obtained by dividing the treatment sum of squares by the degrees of freedom. The treatment mean square represents the variation between the sample means.

(b) The completely randomized design is the simplest of all the designs, based on principles of randomization and replication. In this design, treatments are allocated at random to the experimental units over the entire experimental.

- **What is the aim of designs of experiments? (Apr/May 2015)**

Soln: The design of experiments is the design of any task that aims to describe or explain the variation of information under conditions that are hypothesized to reflect the variation. The term is generally associated with true experiments in which the design introduces conditions that directly affect the variation, but may also refer to the design of quasi-experiments in which natural conditions that influence the variation are selected for observation.

PART-B

1) **The following are the numbers of mistakes made in 5 successive days of 4 Technicians working a photographic laboratory**

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

Test the level of signification α = 0.01 whether the difference among the 4 sample can be attributed to chance?

Soln: $H_0: There\ is\ no\ significance\ difference\ betwee\ the\ Technicians$

$$H_1: There\ is\ a\ significance\ difference\ betwee\ the\ Technicians$$

We shifted our origin to 10

X_1 (X_1-10)	X_2 (X_2-10)	X_3 (X_3-10)	X_4 (X_4-10)	Total	X_1^2	X_2^2	X_3^2	X_4^2
-4	4	0	-1	-1	16	16	0	1
4	-1	2	2	7	16	1	4	4
0	2	-3	-2	-3	0	4	9	4
-2	0	5	0	3	4	0	25	0
1	4	1	1	7	1	16	1	1
ΣX_1	**ΣX_2**	**ΣX_3**	**ΣX_4**	**Σy(Total)**	**ΣX_1^2**	**ΣX_2^2**	**ΣX_3^2**	**ΣX_4^2**
-1	**9**	**5**	**0**	**T=13**	**37**	**37**	**39**	**10**

Step-1: N→ Number of Data given in the Problem (N=20)

Step 2: T=**13** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(13)^2}{20} = \mathbf{8.45}(\ Correction\ Factor)$

Step4 : TSS = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N} = (37 + 37 + 39 + 10 - 8.45) = \mathbf{114.55}$

Step 5: SSC = $\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} - \frac{T^2}{N} = \left(\frac{(-1)^2}{5} + \frac{9^2}{5} + \frac{5^2}{5} + 0 - 8.45\right) = \mathbf{12.95}$

Step 6: SSE = TSS – SSC = 114.55-12.95 = **101.6**

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 1% level**
Between Columns	SSC = 12.95	C-1 = 4-1 = 3	$MSC = \frac{SSC}{C-1} = 4.317$	$F = \cancel{\frac{MSC}{MSE}} < 1$ $F = \frac{MSE}{MSC} = \mathbf{1.471}$	$F_c(16,3) = 26.87$
Error	SSE = 101.6	N-C = 20-4 =16	$MSE = \frac{SSE}{N-C} = 6.35$		
Total	TSS = 114.55	N-1= 20-1 =19	**Condition Always F >1**		

Conclusion→ Calculate $F_c(1.471) < table\ value\ of\ F_c(26.87)\ \ so\ we\ accept\ H_0$

$(ie)\boldsymbol{There\ is\ no\ significance\ difference\ betwee\ the\ Technicians}$

---

2) There are three main brands of a certain powder, A set of 120 sample values is examined and found to be allocated among four groups (A,B,C,D) and three brands are (I,II,III) are shown under

Brands	Groups			
	A	B	C	D
I	0	4	8	15
II	5	8	13	6
III	8	19	11	13

Is there any significance difference in brands preference answer at $\mathbf{5\%}$Level?

$Soln$: $H_0: There\ is\ no\ significance\ difference\ in\ Brands$

$H_1: There\ is\ a\ significance\ difference\ in\ Brands$

Brand	**Groups**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1 (A)	X_2 (B)	X_3 (C)	X_4 (D)					
I (Y_1)	0	4	8	15	$\boldsymbol{\Sigma y_1}$ =27	0	16	64	225
II (Y_2)	5	8	13	6	$\boldsymbol{\Sigma y_2}$ =32	25	64	169	36
III (Y_3)	8	19	11	13	$\boldsymbol{\Sigma y_3}$ =51	64	361	121	169
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$
	13	**31**	**32**	**34**	**T=110**	**89**	**441**	**354**	**430**

Step-1: N→ Number of Data given in the Problem (N=12)

Step 2: T=**110** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(110)^2}{12} = \mathbf{1008.3}(\ Correction\ Factor)$

Step4 : TSS = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N} = (89 + 441 + 354 + 430 - 1008.3) = \mathbf{305.7}$

Step 5: SSR = $\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} - \frac{T^2}{N} = \left(\frac{(27)^2}{4} + \frac{32^2}{4} + \frac{51^2}{4} - 1008.3\right) = \mathbf{80.2}$

Step 6: SSE = TSS – SSR = 305.7 – 80.2 =**225.5**

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio	Table Value at 5% level
Between Rows	SSR = 80.2	R-1 = 3-1 = 2	$MSC = \frac{SSR}{R-1}$ $= 40.1$	$F_R = \frac{MSR}{MSE}$ $= \mathbf{1.999}$ $F = \cancel{\frac{MSE}{MSR}} < 1$	$F_R(2,9)$ $= 4.26$
Error	SSE = 225	N-C = 12-3 =9	$MSE = \frac{SSE}{N-R}$ $= 20.06$		
Total	TSS = 305.7	N-1= 12-1 =11	**Condition Always F >1**		

Conclusion→ Calculate $F_R\ (1.999) < table\ value\ of\ F_R\ (4.26)\ \ so\ we\ accept\ H_0$

$(ie)\boldsymbol{There\ is\ no\ significance\ difference\ in\ Brands}$

3) **Three different machines are used for production ,on the basis of the outputs ,setup One – Way ANOVA table and test whether the machines are equally effective.**

MACHINE I	**MACHINE II**	**MACHINE III**
10	**9**	**20**
15	**7**	**16**
11	**5**	**10**
10	**6**	**14**

Given that the value of F at 5% level of significance for (2,9) d.f is 4.26

Solution:

MACHINES		
X_1	X_2	X_3
10	9	20
15	7	16
11	5	10
10	6	14
$\Sigma X_1 = 56$	$\Sigma X_2 = 27$	$\Sigma X_3 = 60$

Null hypothesis H_0 : The machines are equally effective

Alternate Hypothesis H_1 : The machines are not equally effective

Step:1

Grand Total (G) = 56 + 27 + 60

T = 143

Step:2

Correction factor (C.F) = $\frac{T^2}{N}$

$$= \frac{(143)^2}{12}$$

C.F = 1704.08

Step:3

TSS = Total sum of squares.

$= 10^2 + 15^2 + 11^2 + 10^2 + 9^2 + ... - C.F$

= 1866.25 – 1704.08

TSS = 284.92

Step :4

SSC = Sum of squares between samples.

$$= \frac{(\Sigma X_1)^2}{n} + \frac{(\Sigma X_2)^2}{n} + \frac{(\Sigma X_3)^2}{n} - C.F$$

$$= \frac{(56)^2}{4} + \frac{(27)^2}{4} + \frac{(60)^2}{4} - 1704.08$$

$$= \frac{3136}{4} + \frac{729}{4} + \frac{3600}{4} - 1704.08$$

$$= 784 + 182.25 + 400 - 1704.08$$

$$= 1866.25 - 1704.08$$

SSC $= 162.17$

Step: 5

SSE = TSS – SSC

= 284.92 – 162.17

SSE = 122.75

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Between samples	SSC =162.17	$v_1 = C-1$=3-1 = 2	$MSC = \frac{162.17}{2} = 81.08$	$F_C = \frac{MSC}{MSE} (or) = \frac{MSE}{MSC}$
Within samples	SSE =122.75	$v_2 = n-C =$ 12- 3 = 9	$MSE = \frac{122.75}{9} = 13.63$	$F_C = \frac{81.085}{13.638} = 5.945$
Total	TSS = 284.92	$n-1$=12 -1 =11		

RESULT:

F calculated value = 5.945

T tab(2,9) df at 5% level = 4.26

Fcal > F tab

5.945 > 4.26

$\therefore H_0$ is rejected. Hence we conclude that the machines are not equally effective.

4) **Three samples below have been obtained from normal population with equal variances .test the hypothesis that the samples means are equal.**

Samples.		
8	**7**	**12**
10	**5**	**19**
7	**10**	**13**
14	**9**	**12**
11	**9**	**14**

The value of F at 5% level of significance is 3.88

Samples.		
8	7	12
10	5	19
7	10	13
14	9	12
11	9	14
$\Sigma X_1 = 50$	$\Sigma X_2 = 40$	$\Sigma X_3 = 70$

Soln:

Null hypothesis H_0 : The sample means are equal

Alternate Hypothesis H_1 : The sample means are not equal

Step: 1

Grand total (G) = 50 + 40 + 70

T = 160

Step: 2

Correction factor (C.F) = $\frac{T^2}{N} = \frac{(160)^2}{15}$

C.F = 1706.7

Step : 3

TSS = Total sum squares

= $8^2 + 10^2 + 7^2 + 14^2 + \ldots\ldots\ldots\ldots\ldots\ldots.C.F$

= 1880 – 1706.7

= 173.3

Step: 4

SSC = Sum of squares between samples

$$= \frac{(\Sigma X_1)^2}{n} + \frac{(\Sigma X_2)^2}{n} + \frac{(\Sigma X_3)^2}{n} - C.F$$

$$= \frac{(50)^2}{5} + \frac{(40)^2}{5} + \frac{(70)^2}{5} - 1706.7$$

$$= 500 + 320 + 980 \ - 1706.7$$

= 1800 – 1706.7

SSC = 93.3

Step : 5

SSE = TSS – SSC

= 173.3 – 93.3

SSE = 80

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Between samples	SSC =93.3	$v_1 = C - 1$ = 3-1 = 2	$MSC = \frac{93.3}{2} = 46.65$	$F_C = \frac{MSC}{MSE} (or) = \frac{MSE}{MSC}$
Within samples	SSE =80	$v_2 = n - C$ = 15-3 = 12	$MSE = \frac{80}{12} = 6.7$	$F_C = \frac{46.65}{6.7} = 6.97$
Total	TSS = 173.3	$n - 1$=15-1 =14		

F calculated value = 6.97

T tab(2,12) df at 5% level = 3.88

Fcal > F tab

6.97 > 3.88 $\therefore H_0$ is rejected. Hence we conclude that the sample means are not equal (i.e. There is a significant difference between the means of the three samples.

5) **An Experiment was designed to study the performance of 4 different detergents for cleaning fuel injectors the following cleanness reading were obtained with specially Designed equipments for 12 tanks of Gas distributed over 3 different models of engine**

Detergent	**Engine-I**	**Engine-II**	**Engine-I**	**Total**
A	45	43	51	**139**
B	47	46	52	**145**
C	48	50	55	**153**
D	42	37	49	**128**
Total	**182**	**176**	**207**	**565**

Perform the ANOVA & Test at 0.01 level of significance whether there are differences in detergent or in engines

$$H_0: \begin{cases} (i)\ There\ is\ no\ significance\ difference\ in\ Engines \\ (ii) There\ is\ no\ significance\ difference\ in\ Detergents \end{cases}$$

$$H_1 : \begin{cases} (i)\ There\ is\ a\ significance\ difference\ in\ Engines \\ (ii) There\ is\ a\ significance\ difference\ in\ Detergents \end{cases}$$

We shifted our origin to 50

Brand	**Groups**			Total	X_1^2	X_2^2	X_3^2
	X_1 (I)	X_2 (II)	X_3 (III)				
A(Y_1)	-5	-7	1	$\mathbf{\Sigma y_1 = 11}$	25	49	1
B(Y_2)	-3	-4	2	$\mathbf{\Sigma y_2 = -5}$	9	16	4
C(Y_3)	-2	0	5	$\mathbf{\Sigma y_3 = 3}$	4	0	25
D(Y_4)	-8	-13	-1	$\mathbf{\Sigma y_4 = -22}$	64	169	1
Total	$\mathbf{\Sigma X_1}$	$\mathbf{\Sigma X_2}$	$\mathbf{\Sigma X_3}$	**Σy(Total)**	$\mathbf{\Sigma X_1^2}$	$\mathbf{\Sigma X_2^2}$	$\mathbf{\Sigma X_3^2}$
	-18	**-24**	**7**	**T= -35**	**102**	**234**	**31**

Step-1: N→ Number of Data given in the Problem (N=12)

Step 2: T = **-35** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(-35)^2}{12} = \mathbf{102.08}(\ Correction\ Factor)$

Step4 : TSS = $(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2) - \frac{T^2}{N} = (102 + 234 + 31 - 102.08) = \mathbf{264.92}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2}\right) - \frac{T^2}{N} = \left(\frac{-11^2}{3} + \frac{(-5)^2}{3} + \frac{3^2}{3} + \frac{(-22)^2}{3} - 102.08\right) = \mathbf{110.91}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\mathbf{N_1}}\right) - \frac{T^2}{N} = \left(\frac{18^2}{4} + \frac{(-24)^2}{4} + \frac{7^2}{4} - 102.08\right) = \mathbf{135.17}$

Step 7: SSE = TSS –SSC- SSR = $264.92 - 135.17 - 110.91 = \mathbf{18.84}$

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 1% level**
Column Treatment	SSC = 135.17	C-1=2	$MSC = \frac{SSC}{C-1} = \frac{135.17}{2} = 67.58$	$F_c = \frac{MSC}{MSE} = \frac{67.585}{3.14} = \mathbf{21.52}$	$F_C(2{,}6) = 10.92$
Row Treatment	SSR = 110.91	R-1=3	$MSR = \frac{SSR}{R-1} = \frac{110.91}{3} = 36.97$	$F_R = \frac{MSR}{MSE} = \frac{36.97}{3.14} = \mathbf{11.77}$	$F_R(3{,}6) = 9.78$
Error	SSE = 18.84	N-C-R+1= 6	$MSE = \frac{SSE}{N-C-R+1} = \frac{18.84}{6} = 3.14$		
Total	TSS = 264.92	**Condition Always F >1**			

Conclusion→ Calculate $Value\ > table\ value\ in\ both\ the\ Cases\ \ so\ we\ Reject\ H_0$

$(ie)\mathbf{There\ is\ a\ \ significance\ difference\ between\ Detergents}\ \&\ Engines$

6) Five Doctors each test five treatments for a certain disease and observe the number of days each patients recover the results are (Anna University --Dec-13)

Doctors	**Treatments**				
	I	**II**	**III**	**IV**	**V**
A	10	14	23	19	20
B	11	15	24	17	21
C	9	11	20	16	19
D	8	13	17	17	20
E	12	15	19	15	22

Discuss the Difference between Doctors and Treatments?

Soln:

$$H_0:\begin{cases} (i)\ There\ is\ no\ significance\ difference\ between\ Doctors \\ (ii) There\ is\ no\ significance\ difference\ between\ Treatments \end{cases}$$

$$H_1:\begin{cases} (i)\ There\ is\ a\ significance\ difference\ between\ Doctors \\ (ii) There\ is\ a\ significance\ difference\ between\ Treatments \end{cases}$$

We shifted our origin to 16

Doctors	**Treatments**					Total	X_1^2	X_2^2	X_3^2	X_4^2	X_5^2
	X_1	X_2	X_3	X_4	X_5						
A (Y_1)	-6	-2	7	3	4	$\boldsymbol{\Sigma y_1 = 6}$	36	4	49	9	16
B (Y_2)	-5	-1	8	1	5	$\boldsymbol{\Sigma y_2 = 8}$	25	1	64	1	25
C (Y_3)	-7	-5	4	0	3	$\boldsymbol{\Sigma y_3 = -4}$	49	16	16	0	9
D (Y_4)	-8	-3	1	1	4	$\boldsymbol{\Sigma y_4 = -5}$	64	9	9	1	16
E (Y_5)	-4	-1	3	-1	6	$\boldsymbol{\Sigma y_5 = 3}$	16	1	1	1	36
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	$\boldsymbol{\Sigma X_5}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$	$\boldsymbol{\Sigma X_5^2}$
	-30	**-12**	**23**	**4**	**22**	**T=8**	**190**	**31**	**139**	**12**	**102**

Step-1: N→ Number of Data given in the Problem (N=25)

Step 2: T = **8** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(8)^2}{25} = \mathbf{2.56}(\,Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = 474 - 2.56 = \mathbf{471.44}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2} + \frac{\Sigma y_5^2}{N_2}\right) - \frac{T^2}{N}$

SSR= $\left(\frac{(6)^2}{5} + \frac{(8)^2}{5} + \frac{(-4)^2}{5} + \frac{(-5)^2}{5} + \frac{(3)^2}{5} - 2.56\right) = 30 - 2.56 = \mathbf{27.64}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} + \frac{\Sigma X_5^2}{\boldsymbol{N_1}}\right) - \frac{T^2}{N}$

$$SSC = \left(\frac{(-30)^2}{5} + \frac{(-12)^2}{5} + \frac{23^2}{5} + \frac{(4)^2}{5} + \frac{(22)^2}{5} - 2.56\right) = 410 - 2.56 = \mathbf{407.44}$$

Step 7: SSE = TSS –SSC- SSR = $471.44 - 407.44 - 27.64 = \mathbf{36.56}$

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio	Table Value at 1% level
Column Treatment	SSC = 407.44	C-1=4	$MSC = \frac{SSC}{C-1} = \frac{407.44}{4} = 101.86$	$F_c = \frac{MSC}{MSE} = \frac{101.86}{2.28} = \mathbf{44.67}$	$F_C(4,6) = 3.01$
Row Treatment	SSR = 27.44	R-1=4	$MSR = \frac{SSR}{R-1} = \frac{27.44}{4} = 6.86$	$F_R = \frac{MSR}{MSE} = \frac{6.86}{2.28} = \mathbf{3.01}$	$F_R(4,6) = 3.01$
Error	SSE = 36.56	N-C-R+1= 16	$MSE = \frac{SSE}{N-C-R+1} = \frac{36.56}{16} = 2.28$		
Total	TSS = 471.44	**Condition Always F >1**			

$$\text{Calculated } Value\ F_R(3.01) \le table\ value\ F_R(3.01)\ so\ we\ accept\ \ H_0$$

$$\text{Calculated } Value\ F_C(44.67) > table\ value\ F_C(3.01)\ so\ we\ Reject\ \ H_0$$

Conclusion:

$$There\ is\ no\ significance\ difference\ between\ Doctors\ but,$$

$$There\ is\ a\ significance\ difference\ between\ Trearments$$

7) The following table gives monthly sales (in thousand rupees) of a certain firm in three states by its four salesman.

	Salesman			
States	**I**	**II**	**III**	**IV**
A	**6**	**5**	**3**	**8**
B	**8**	**9**	**6**	**5**
C	**10**	**7**	**8**	**7**

Setup the analysis of variance table and test whether there is any significant difference (i) between the sales by the firm salesman ,(ii) between sales in the three states.

Solution:

H_0 :There is no significant difference between the sales by the firm's salesman

H_1 : There is significant difference between the sales by the firm's salesman

H_0 :There is no significant difference between the three states.

H_1 : There is significant difference between the three states

States	Saleman				Total
	I	II	III	IV	
A	6	5	3	8	22
B	8	9	6	5	28
C	10	7	8	7	32
Total	24	21	17	20	T = 82

Step : 1

Correction factor (C.F) = $\frac{T^2}{N} = \frac{(82)^2}{12}$

C.F = 560.333

Step;2

TSS = Sum of squares of each values – C.F

$= 6^2 + 8^2 + 10^2 + 5^2 + 9^2 + ... - 560.333$

= 602 – 560.333

TSS = 41.667

Step : 3

SSC = Sum of squares between columns, (salesman)

$= \frac{1}{3}\left[24^2 + 21^2 + 17^2 + 20^2\right]$ - C.F

$= \frac{1}{3}\left[24^2 + 21^2 + 17^2 + 20^2\right]$ - 560.333

= 568.667 - 560.333

SSC = 8.334

Step :4

SSR = Row sum of squares.(states)

$= \frac{1}{4}\left[22^2 + 28^2 + 32^2\right]$ - C.F

$= \frac{1}{4}\left[22^2 + 28^2 + 32^2\right]$ -560.333

= 573 - 560.333

SSR = 12.667

Step :5

SSE = Error sum of squares.

= TSS – SSC – SSR

= 41.667 – 8.334 – 12.667

SSE = 20.666

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Between columns	SSC = 8.334	k-1= 4-1=3	$MSC = \frac{8.334}{3} = 2.778$	$F_C = \frac{3.444}{2.778} = 1.239$
Between rows	SSR=12.667	r-1=3 -1 =2	$MSR = \frac{12.667}{2} = 6.334$	$F_R = \frac{6.334}{3.444} = 1.84$
Residual error	SSE=20.667	(k-1)(r-1)= (3)(2)=6	$MSE = \frac{20.667}{6} = 3.444$	
Total	TSS=41.667	rk-1		

RESULT :1

F Calculated value = 1.239

F tab(3,6) df at 5% level = 4.75

F Cal < F tab

1.239 < 4.75

H_0 is accepted.

Hence we conclude that there is no significant difference between the sales by the firm's salesman.

RESULT :2

F Calculated value = 1.84

F tab(2,6) df at 5% level = 5.14

F Cal < F tab

1.84 < 4.75

H_0 is accepted. Hence we conclude that there is no significant difference between the sales in the three states.

$\therefore$ There is no significant difference in the states as far as sales are concerned at 5% level of significance

8) A tea company appoints four saleman A,B,C,D and observes their sales in three seasons summer ,winter , monsoon. The figures (in lakhs) are given in the following table.

Seasons	**Salesman**				**Season's Total**
	A	**B**	**C**	**D**	
Summer	**36**	**36**	**21**	**35**	**128**
Winter	**28**	**29**	**31**	**32**	**120**
Monsoon	**26**	**28**	**29**	**29**	**112**
Salesman's Total	**90**	**93**	**81**	**96**	**360**

(i) Does the salesman significantly differ in performance ?

(ii) Is there significant difference between the seasons?

Solution:

H_0 :The salesman does not differ significantly differ in performance.

H_1 : The salesman differs significantly differ in their performance.

H_0 : There is no significant difference between the three seasons.

H_1 : There is significant difference between the three seasons

Step:1

Correction factor (C.F) = $\frac{T^2}{N} = \frac{(360)^2}{12}$

C.F = 10800

Step :2

TSS = Sum of squares of each values – C.F

$= 36^2 + 36^2 + 21^2 + 35^2 + 28^2 + \ldots\ldots\ldots\ldots\ldots\ldots - 10800$

= 11010 – 10800

TSS = 210

Step :3

SSC = Sum of squares between columns, (salesman)

$= \frac{1}{3}\left[90^2 + 93^2 + 81^2 + 96^2\right]$ - C.F

$= \frac{1}{3}\left[90^2 + 93^2 + 81^2 + 96^2\right]$-10800

= 10842 -10800

SSC = 42

Step :4

SSR = Sum of squares between rows.(seasons)

$= \frac{1}{4}\left[128^2 + 120^2 + 112^2\right]$ - C.F = $\frac{1}{4}\left[128^2 + 120^2 + 112^2\right]$ -10800 = 10832 – 10800 SSR = 32

SSE = Error sum of squares. = TSS – SSC – SSR = 210 – 42 -32 SSE = 136

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Between columns	SSC =42	k-1= 4 – 1 =3	$MSC = \frac{42}{3} = 14$	$F_C = \frac{22.67}{14} = 1.619$
Between rows	SSR=32	r-1 = 3 -1 =2	$MSR = \frac{32}{2} = 16$	$F_R = \frac{22.67}{16} = 1.41$
Residual error	SSE=136	(k-1)(r-1) $3 \times 2 = 6$	$MSE = \frac{136}{6} = 22.67$	
Total	TSS=210	rk-1=11		

RESULT :1

F Calculated value = 1.619

F tab(3,6) df at 5% level = 4.75

F Cal < F tab

1.619 < 4.75

H_0 is accepted.

Hence we conclude that the salesman do not differ significantly in their performance .

RESULT :2

F Calculated value = 1.41

F tab(2,6) df at 5% level = 5.14

F Cal < F tab

1.41 < 5.14

H_0 is accepted. Hence we conclude that there is no significant difference between the three seasons.

9) Preform a Two – way ANOVA on the data given below.

Plots of land	**Treatments**			
	A	**B**	**C**	**D**
I	**38**	**40**	**41**	**39**
II	**45**	**42**	**49**	**36**
III	**40**	**38**	**42**	**42**

Solution:

H_0 :There is no significant difference between Treatments

H_1 : There is significant difference between Treatments

H_0 :There is no significant difference between Plots

H_1 : There is significant difference between Plots

Plots of land	Treatments				Total
	A	B	C	D	
I	38	40	41	39	158
II	45	42	49	36	172
III	40	38	42	42	162
Total	123	120	132	117	492

Step :1

Correction factor (C.F) = $\frac{T^2}{N} = \frac{(492)^2}{12}$

C.F = 20172

Step :2

TSS = Sum of squares of each values – C.F

$= 38^2 + 45^2 + 40^2 + 40^2 + 42^2 + \ldots\ldots\ldots\ldots\ldots\ldots - 20172$

= 20304 - 20172

TSS = 132

Step :3

SSC = Sum of squares between columns, (Treatments)

$= \frac{1}{3}\left[123^2 + 120^2 + 132^2 + 117^2\right]$ - C.F

$$= \frac{1}{3}\left[123^2 + 120^2 + 132^2 + 117^2\right]\text{-20172}$$

= 20214-20172

SSC = 42

Step :4

SSR = Sum of squares between rows.(plots)

$$= \frac{1}{4}\left[158^2 + 172^2 + 162^2\right] \text{ - C.F}$$

$$= \frac{1}{4}\left[158^2 + 172^2 + 162^2\right] \text{ - 20172}$$

= 20198 - 210172

SSR = 26

Step :5

SSE = Error sum of squares.

= TSS – SSC – SSR

= 132 -42 -26

SSE = 64

ANOVA TABLE

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Between columns	SSC =42	k-1= 4 -1 =3	$MSC = \frac{42}{3} = 14$	$F_C = \frac{14}{10.67} = 1.312$
Between rows	SSR=26	r-1=3 -1 =2	$MSR = \frac{26}{2} = 13$	$F_R = \frac{13}{10.67} = 1.218$
Residual error	SSE=64	(k-1)(r-1) $3 \times 2 = 6$	$MSR = \frac{64}{6} = 10.67$	
Total	TSS=132	rk-1=11		

F Calculated value = 1.312

F tab(3,6) df at 5% level = 4.75

F Cal < F tab

1.312 < 4.75

H_0 is accepted. Hence we conclude there is no significant difference between treatments

F Calculated value = 1.218

F tab(2,6) df at 5% level = 5.14

F Cal < F tab

1.218 < 5.14

H_0 is accepted. Hence we conclude that there is no significant difference between the Plots.

10) The Following Latin Square of a design when four varieties of seeds are being tested set up the analysis table and state your conclusion you may carry out suitable change of origin and scale

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

Soln: $Let \quad H_0 = \begin{cases} (i)\ There\ \ is\, no\ significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\ \ There\ \ is\, no\ significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\ There\ \ is\, no\ significance\ \ Difference\ \ in\ \ Lands \end{cases}$

$$H_1 = \begin{cases} (i)\ \ There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\ There\ \ is\, a\ significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\ There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Lands \end{cases}$$

Shifted Origin to 100 and divided by 5,

Lands	**Seeds**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1 (A)	X_2 (B)	X_3 (C)	X_4 (D)					
I (Y_1)	1	-1	5	3	$\boldsymbol{\Sigma y_1 = 8}$	1	1	25	9
II (Y_2)	3	5	1	1	$\boldsymbol{\Sigma y_2 = 10}$	9	25	1	1
III (Y_3)	3	-1	1	3	$\boldsymbol{\Sigma y_3 = 6}$	9	1	1	9
IV (Y_4)	-1	7	-1	3	$\boldsymbol{\Sigma y_4 = 8}$	1	49	1	9
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$
	6	**10**	**6**	**10**	**T= 32**	**20**	**76**	**28**	**28**

Step-1: N→ Number of Data given in the Problem (N=16)

Step 2: T = **32** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(32)^2}{16} = \mathbf{64}(\ Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = (20 + 76 + 28 + 28) - 64 = \mathbf{88}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2}\right) - \frac{T^2}{N} = \left(\frac{(8)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(8)^2}{4} - 64\right) = \mathbf{2}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} + \right) - \frac{T^2}{N} = \left(\frac{(6)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 64\right) = \mathbf{4}$

SST					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**

Step 7: SST = $\left(\frac{(12)^2}{4} + \frac{(0)^2}{4} + \frac{(10)^2}{4} + \frac{(10)^2}{4} - 64\right) = \mathbf{22}$

Step 8: SSE = TSS –SSC- SSR- SST = $88-2-4-22=\mathbf{60}$

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio	Table Value at 5% level
Between Columns	SSC=4	K-1=3	$MSC=\frac{SSC}{K-1}=\frac{4}{3}=1.33$	$F_c=\frac{MSE}{MSC}=\frac{10}{1.33}=\mathbf{7.52}$ $F_R=\frac{MSE}{MSR}=\frac{10}{0.67}=\mathbf{14.9}$ $F_T=\frac{MSE}{MST}=\frac{10}{7.33}=\mathbf{1.36}$	$F_c(6,3)=8.94$ $F_R(6,3)=8.94$ $F_T(6,3)=8.94$
Between Rows	SSR=2	K-1=3	$MSR=\frac{SSR}{K-1}=\frac{2}{3}=0.67$		
Between Treatments	SST=22	K-1=3	$MST=\frac{SST}{K-1}=\frac{22}{3}=7.33$		
Error	SSE=60	(K-1)(K-2) =6	$MSE=\frac{SSE}{(K-1)(K-2)}=\frac{60}{6}=10$		
Total	TSS=88	$K^2-1=15$	Condition Always F >1		

Calculated $Value\ F_C(7.52) < table\ value\ F_C(8.94)\ so\ we\ accept\ H_0$

Calculated $Value\ F_R(14.9) > table\ value\ F_R(8.94)\ so\ we\ Reject\ H_0$

Calculated $Value\ F_T(1.36) < table\ value\ F_T(8.94)\ so\ we\ accept\ H_0$

Conclusion:

$There\ is\ no\ significance\ difference\ between\ Seeds\ \&\ Treatments, But$

$There\ is\ a\ significance\ difference\ between\ Lands$

11) Analyze the following Latin Square experiment at 1% level

A (12)	**D (20)**	**C (16)**	**B (10)**
D (18)	**A (14)**	**B (11)**	**C (14)**
B (12)	**C (15)**	**D (19)**	**A (13)**
C (16)	**B (11)**	**A (15)**	**D (20)**

The Letters (A,B,C,D) denotes the treatments & the figures in brackets denotes the observation

Soln: We Shifted our origin to 12

$$Let\quad H_0=\begin{cases}(i)\ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Seeds\\(ii)\ \ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Treatments\\(iii)\ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Lands\end{cases}$$

$$H_1 = \begin{cases} (i)\ There\ \ is\ a\ significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\ There\ \ is\ a\ significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\ There\ \ is\ a\ significance\ \ Difference\ \ in\ \ Lands \end{cases}$$

Lands	**Seeds**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1 (A)	X_2 (B)	X_3 (C)	X_4 (D)					
I (Y_1)	0	8	4	2	$\boldsymbol{\Sigma y_1 = 10}$	0	64	16	4
II (Y_2)	6	2	-1	2	$\boldsymbol{\Sigma y_2 = 9}$	36	4	1	4
III (Y_3)	0	3	7	1	$\boldsymbol{\Sigma y_3 = 11}$	0	9	49	1
IV (Y_4)	4	-1	3	8	$\boldsymbol{\Sigma y_4 = 14}$	16	1	9	64
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$
	10	**12**	**13**	**9**	**T= 44**	**52**	**78**	**75**	**73**

Step-1: N→ Number of Data given in the Problem (N=16)

Step 2: T = **44** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(44)^2}{16} = \mathbf{121}(\ Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = (52 + 78 + 75 + 73) - 121 = \mathbf{157}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2}\right) - \frac{T^2}{N} = \left(\frac{(10)^2}{4} + \frac{(9)^2}{4} + \frac{(11)^2}{4} + \frac{(14)^2}{4} - 121\right) = \mathbf{3.5}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} +\right) - \frac{T^2}{N} = \left(\frac{(10)^2}{4} + \frac{(12)^2}{4} + \frac{(13)^2}{4} + \frac{(9)^2}{4} - 121\right) = \mathbf{2.5}$

SST					Total
A	0	2	3	1	**6**
B	0	-1	-1	-2	**-4**
C	4	3	4	2	**13**
D	6	8	7	8	**29**

Step 7: SST = $\left(\frac{(6)^2}{4} + \frac{(-4)^2}{4} + \frac{(13)^2}{4} + \frac{(29)^2}{4} - 121\right) = \mathbf{144.5}$

Step 8: SSE = TSS –SSC- SSR- SST = $157 - 2.5 - 3.5 - 144.5 = \mathbf{6.5}$

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 1% level**
Between Columns	SSC=2.5	K-1=3	$MSC = \frac{SSC}{K-1} = \frac{2.5}{3} = 0.83$	$F_c = \frac{MSE}{MSC} = \frac{1.08}{0.83} = \mathbf{1.301}$	$F_c(6,3) = 27.91$
Between Rows	SSR=3.5	K-1=3	$MSR = \frac{SSR}{K-1} = \frac{3.5}{3} = 1.17$		$F_R(3,6) = 9.78$

Between Treatments	SST=144.5	K-1=3	$MST = \frac{SST}{K-1} = \frac{144.5}{3} = 48.17$	$F_R = \frac{MSR}{MSE} = \frac{1.17}{1.08} = \mathbf{1.08}$ $F_T = \frac{MST}{MSE} = \frac{48.17}{1.08} = \mathbf{44.6}$	$F_T(3,6) = 9.78$
Error	SSE=6.5	(K-1)(K-2) =6	$MSE = \frac{SSE}{(K-1)(K-2)} = \frac{65}{6} = 1.08$		
Total	TSS=157	$K^2 - 1 = 15$	Condition Always F >1		

Calculated $Value\ F_C(1.301) < table\ value\ F_C(27.91)\ so\ we\ accept\ H_0$

Calculated $Value\ F_R(1.08) < table\ value\ F_R(9.78)\ so\ we\ accept\ H_0$

Calculated $Value\ F_T(44.6) > table\ value\ F_T(9.78)\ so\ we\ Reject\ H_0$

Conclusion: $There\ is\ no\ significance\ difference\ between\ Seeds\ \&\ Lands\ , But$

$There\ is\ a\ significance\ difference\ between\ Treatments$

12) Three varieties of a crop are tested in the Randomized block design with four replications, the layout being has given below: The yields are given in kilograms Analyse for significance

(Apr/May 2015 (R13 &R08))

C48	**A51**	**B52**	**A49**
A47	**B49**	**C52**	**C51**
B49	**C53**	**A49**	**B50**

Soln: $H_0: \begin{cases} (i)\ There\ is\ no\ significance\ difference\ between\ Yields \\ (ii) There\ is\ no\ significance\ difference\ between\ Crops \end{cases}$

$H_1: \begin{cases} (i)\ There\ is\ a\ significance\ difference\ between\ Yields \\ (ii) There\ is\ a\ significance\ difference\ between\ Crops \end{cases}$

Yields	**Treatments**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1	X_2	X_3	X_4					
A (Y_1)	48	51	52	49	$\mathbf{\Sigma y_1 = 200}$	2304	2601	2704	2401
B (Y_2)	47	49	52	51	$\mathbf{\Sigma y_2 = 199}$	2209	2401	2704	2601
C (Y_3)	49	53	49	50	$\mathbf{\Sigma y_3 = 201}$	2401	2809	2401	2500
Total	$\mathbf{\Sigma X_1}$	$\mathbf{\Sigma X_2}$	$\mathbf{\Sigma X_3}$	$\mathbf{\Sigma X_4}$	**Σy(Total)**	$\mathbf{\Sigma X_1^2}$	$\mathbf{\Sigma X_2^2}$	$\mathbf{\Sigma X_3^2}$	$\mathbf{\Sigma X_4^2}$
	144	**153**	**153**	**150**	**T= 600**	**6914**	**7811**	**7809**	**7502**

Step-1: N→ Number of Data given in the Problem (N=12)

Step 2: T = **600** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(600)^2}{12} = \mathbf{30000}(\ Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = \mathbf{36}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2} + \frac{\Sigma y_5^2}{N_2}\right) - \frac{T^2}{N}$

SSR= $\left(\frac{(200)^2}{4} + \frac{(199)^2}{4} + \frac{(201)^2}{4} - 30000\right) = \mathbf{0.5}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} + \frac{\Sigma X_5^2}{\boldsymbol{N_1}}\right) - \frac{T^2}{N}$

$$SSC = \left(\frac{(144)^2}{3} + \frac{(153)^2}{3} + \frac{(153)^2}{3} + \frac{(150)^2}{3} - 30000\right) = \mathbf{18}$$

Step 7: SSE = TSS –SSC- SSR = $36 - 0.5 - 18 = \mathbf{17.5}$

Source of Variance	Sum of Squares	Degree of Freedom	Mean sum of Squares	Variance Ratio	Table Value at 5% level
Column Treatment	SSC = 18	C-1=3	$MSC = \frac{SSC}{C-1} = \frac{184}{3} = 6$	$F_c = \frac{MSC}{MSE} = \frac{6}{2.91} = \mathbf{2.057}$	$F_C(3,6) = 4.75$
Row Treatment	SSR = 0.5	R-1=2	$MSR = \frac{SSR}{R-1} = \frac{0.5}{2} = 0.25$	$F_R = \frac{MSE}{MSR} = \frac{2.91}{0.25} = \mathbf{11.667}$	$F_R(6,2) = 19.32$
Error	SSE = 17.5	N-C-R+1= 6	$MSE = \frac{SSE}{N-C-R+1} = \frac{17.5}{6} = 2.91$		
Total	TSS = 36	**Condition Always F >1**			

Calculated $Value\ F_R(11.66) < table\ value\ F_R(19.32)\ so\ we\ Accept\ H_0$

Calculated $Value\ F_C(2.057) < table\ value\ F_C(4.75)\ so\ we\ Accept\ H_0$

<u>Conclusion: </u>*There is no significance difference between Yields and crops*

13) Analyse the variance in the latin square of yields in (Kgs) of paddy where A,B,C,D denote the different method of cultivation. Examine whether the different method of cultivation have given significantly different yields (Apr/May 2015(R13 &R08))

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**

Soln: $$Let\ \ H_0 = \begin{cases} (i)\ There\ is\ no\ significance\ Difference\ in\ Seeds \\ (ii)\ There\ is\ no\ significance\ Difference\ in\ Treatments \\ (iii)\ There\ is\ no\ significance\ Difference\ in\ Lands \end{cases}$$

$$H_1 = \begin{cases} (i)\ There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\, There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\, There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Lands \end{cases}$$

Lands	**Seeds**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1 (A)	X_2 (B)	X_3 (C)	X_4 (D)					
I (Y_1)	122	121	123	122	$\boldsymbol{\Sigma y_1 = 488}$	14884	14641	15129	14884
II (Y_2)	124	123	122	125	$\boldsymbol{\Sigma y_2 = 494}$	15376	15129	14884	15625
III (Y_3)	120	119	120	121	$\boldsymbol{\Sigma y_3 = 480}$	14400	14161	14400	14641
IV (Y_4)	122	123	121	122	$\boldsymbol{\Sigma y_4 = 488}$	14884	15129	14641	14884
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$
	488	**486**	**486**	**490**	**T= 1950**	**59544**	**59060**	**59054**	**60034**

Step-1: N→ Number of Data given in the Problem (N=16)

Step 2: T = **1950** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(1950)^2}{16} = \mathbf{237656.3}$ $(\,Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = \mathbf{35.75}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2}\right) - \frac{T^2}{N} = \left(\frac{(488)^2}{4} + \frac{(494)^2}{4} + \frac{(480)^2}{4} + \frac{(488)^2}{4} - 237656.3\right) = \mathbf{24.75}$

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 5% level**
Between Columns	SSC=2.75	K-1=3	$MSC = \frac{SSC}{K-1} = \frac{2.75}{3} = 0.91$	$F_c = \frac{MSC}{MSE} = \frac{0.91}{0.66} = \mathbf{1.375}$	$F_c(3,6) = 4.75$
Between Rows	SSR=24.75	K-1=3	$MSR = \frac{SSR}{K-1} = \frac{24.75}{3} = 8.25$	$F_R = \frac{MSR}{MSE} = \frac{8.25}{0.66} = \mathbf{12.375}$	$F_R(3,6) = 4.75$
Between Treatments	SST=4.25	K-1=3	$MST = \frac{SST}{K-1} = \frac{4.25}{3} = 1.41$	$F_T = \frac{MST}{MSE} = \frac{1.41}{0.66} = \mathbf{2.125}$	$F_T(3,6) = 4.75$
Error	SSE=4	(K-1)(K-2) =6	$MSE = \frac{SSE}{(K-1)(K-2)} = \frac{4}{6} = 0.66$		
Total	TSS=35.75	$K^2 - 1 = 15$	Condition Always F >1		

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1}+\frac{\Sigma X_2^2}{N_1}+\frac{\Sigma X_3^2}{N_1}+\frac{\Sigma X_4^2}{\boldsymbol{N_1}}+\right)-\frac{T^2}{N}=\left(\frac{(488)^2}{4}+\frac{(486)^2}{4}+\frac{(486)^2}{4}+\frac{(490)^2}{4}-237656.3\right)=\mathbf{2.75}$

SST					Total
A	121	122	120	122	**485**
B	122	124	119	121	**486**
C	123	123	121	122	**489**
D	122	125	120	123	**490**

Step 7: SST = $\left(\frac{(485)^2}{4}+\frac{(486)^2}{4}+\frac{(489)^2}{4}+\frac{(490)^2}{4}-237656.3\right)=\mathbf{4.25}$

Step 8: SSE = TSS –SSC- SSR- SST = $35.75-24.75-2.75-4.25=\mathbf{4}$

Calculated $Value\ F_C(1.375) < table\ value\ F_C(4.75)\ so\ we\ Accept\ \ H_0$

Calculated $Value\ F_R(12.375) > table\ value\ F_R(4.75)\ so\ we\ \ Reject\ \ H_0$

Calculated $Value\ F_T(2.125) < table\ value\ F_T(4.75)\ so\ we\ Accept\ \ H_0$

<u>Conclusion:</u> $There\ is\ no\ significance\ difference\ between\ Seeds\ \&Treatments\ \ ,But$

$There\ is\ a\ significance\ difference\ between\ Lands$

14) Four different, through supposed by equivalent, forms of a standardized reading achievements test where give to each of five students and the followings are the scores which they obtained (Nov/Dec 2015)

	Student-1	**Student-2**	**Student-3**	**Student-4**	**Student-5**
Form A	**75**	**73**	**59**	**69**	**84**
Form B	**83**	**72**	**56**	**70**	**92**
Form C	**86**	**61**	**53**	**72**	**88**
Form D	**73**	**67**	**62**	**79**	**95**

Perform two way analysis of variance to test at the level of significance α = 0.01 whether it is reasonable to treat the four form are equivalent, Are the scores of the students significantly difference α = 0.01 level?

Soln: $H_0\colon\begin{cases}(i)\ There\ is\ no\ significance\ difference\ between\ Students\\ (ii) There\ is\ no\ significance\ difference\ between\ Forms\end{cases}$

$H_1\colon\begin{cases}(i)\ There\ is\ a\ significance\ difference\ between\ Students\\ (ii) There\ is\ a\ significance\ difference\ between\ Forms\end{cases}$

Forms	**Students**					Total	X_1^2	X_2^2	X_3^2	X_4^2	X_5^2
	X_1	X_2	X_3	X_4	X_5						
A (Y_1)	75	73	59	69	84	$\boldsymbol{\Sigma y_1 = 360}$	5625	5329	3481	4761	7056
B (Y_2)	83	72	56	70	92	$\boldsymbol{\Sigma y_2 = 373}$	6889	5184	3136	4900	8464
C (Y_3)	86	61	53	72	88	$\boldsymbol{\Sigma y_3 = 360}$	7396	3721	2809	5184	7744
D (Y_4)	73	67	62	79	95	$\boldsymbol{\Sigma y_4 = 376}$	5329	4489	3844	6241	9025
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	$\boldsymbol{\Sigma X_5}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$	$\boldsymbol{\Sigma X_5^2}$
	317	**273**	**230**	**290**	**359**	**T=1469**	**25239**	**18723**	**13270**	**21086**	**32289**

Step-1: N→ Number of Data given in the Problem (N=20)

Step 2: T = **1469** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(1469)^2}{20} = \mathbf{107898.1}$ $\ (Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = \mathbf{2708.95}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2} + \frac{\Sigma y_5^2}{N_2}\right) - \frac{T^2}{N}$

SSR= $\left(\frac{(360)^2}{5} + \frac{(373)^2}{5} + \frac{(360)^2}{5} + \frac{(376)^2}{5} - 107898.1\right) = \mathbf{42.95}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} + \frac{\Sigma X_5^2}{\boldsymbol{N_1}}\right) - \frac{T^2}{N}$

$$SSC = \left(\frac{(317)^2}{4} + \frac{(273)^2}{4} + \frac{(230)^2}{4} + \frac{(290)^2}{4} + \frac{(359)^2}{4} - 107898.1\right) = \mathbf{2326.7}$$

Step 7: SSE = TSS –SSC- SSR = $2708.95 - 42.95 - 2326.7 = \mathbf{339.3}$

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 5% level**
Column Treatment	SSC = 2326.7	C-1=4	$MSC = \frac{SSC}{C-1} = \frac{2326.7}{4} = 581.67$	$F_c = \frac{MSC}{MSE} = \frac{581.67}{28.27} = \mathbf{20.57}$	$F_C(4,12) = 3.25$
Row Treatment	SSR = 42.95	R-1=3	$MSR = \frac{SSR}{R-1} = \frac{42.95}{3} = 14.31$	$F_R = \frac{MSE}{MSR} = \frac{28.27}{14.31} = \mathbf{1.975}$	$F_R(12,3) = 8.74$
Error	SSE = 339.3	N-C-R+1= 12	$MSE = \frac{SSE}{N-C-R+1} = \frac{339.3}{12} = 28.27$		
Total	TSS = 2708.95	**Condition Always F >1**			

Calculated $Value\ F_R(1.97) < table\ value\ F_R(8.74)\ so\ we\ Accept\ \ H_0$

Calculated $Value\ F_C(20.57) > table\ value\ F_C(3.25)\ so\ we\ Reject\ \ H_0$

Conclusion:

There is a significance difference between Students, but not in Forms,

-The following data related to the Latin square experiment on four varieties of paddy A,B,C & D

(Nov/Dec 2015)

18 A	**21 C**	**25 D**	**11 B**
22 D	**12 B**	**15 A**	**19 C**
15 B	**20 A**	**23 C**	**24 D**
22 C	**21 D**	**10 B**	**17 A**

Analyse the result and offer your comments of α = 0.05 level of significance

Soln: $Let\quad H_0 = \begin{cases} (i)\ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\ \ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\ There\ \ is\ no\ significance\ \ Difference\ \ in\ \ Lands \end{cases}$

$$H_1 = \begin{cases} (i)\ There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Seeds \\ (ii)\, There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Treatments \\ (iii)\, There\ \ is\, a\, significance\ \ Difference\ \ in\ \ Lands \end{cases}$$

Lands	**Seeds**				Total	X_1^2	X_2^2	X_3^2	X_4^2
	X_1 (A)	X_2 (B)	X_3 (C)	X_4 (D)					
I (Y_1)	18	21	25	11	$\boldsymbol{\Sigma y_1 = 75}$	324	441	625	121
II (Y_2)	22	12	15	19	$\boldsymbol{\Sigma y_2}$ **=68**	484	144	225	361
III (Y_3)	15	20	23	24	$\boldsymbol{\Sigma y_3 = 82}$	225	400	529	576
IV (Y_4)	22	21	10	17	$\boldsymbol{\Sigma y_4 = 70}$	484	441	100	289
Total	$\boldsymbol{\Sigma X_1}$	$\boldsymbol{\Sigma X_2}$	$\boldsymbol{\Sigma X_3}$	$\boldsymbol{\Sigma X_4}$	**Σy(Total)**	$\boldsymbol{\Sigma X_1^2}$	$\boldsymbol{\Sigma X_2^2}$	$\boldsymbol{\Sigma X_3^2}$	$\boldsymbol{\Sigma X_4^2}$
	77	**74**	**73**	**71**	**T= 295**	**1517**	**1426**	**1479**	**1347**

Step-1: N→ Number of Data given in the Problem (N=16)

Step 2: T = **295** (From above Table)

Step 3: $\frac{T^2}{N} = \frac{(295)^2}{16} = \mathbf{5439.06}(\ Correction\ Factor)$

Step4 : TSS = $\left(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}\right) = \mathbf{329.93}$

Step 5: SSR= $\left(\frac{\Sigma y_1^2}{N_2} + \frac{\Sigma y_2^2}{N_2} + \frac{\Sigma y_3^2}{N_2} + + \frac{\Sigma y_4^2}{N_2}\right) - \frac{T^2}{N} = \left(\frac{(75)^2}{4} + \frac{(68)^2}{4} + \frac{(82)^2}{4} + \frac{(70)^2}{4} - 5439.06\right) = \mathbf{29.18}$

Step 6: SSC = $\left(\frac{\Sigma X_1^2}{N_1} + \frac{\Sigma X_2^2}{N_1} + \frac{\Sigma X_3^2}{N_1} + \frac{\Sigma X_4^2}{\boldsymbol{N_1}} +\right) - \frac{T^2}{N} = \left(\frac{(77)^2}{4} + \frac{(74)^2}{4} + \frac{(73)^2}{4} + \frac{(71)^2}{4} - 5439.06\right) = \mathbf{4.68}$

SST					Total
A	18	15	20	17	**70**
B	11	12	15	10	**48**
C	21	19	23	22	**85**
D	25	22	24	21	**92**

Step 7: SST = $\left(\frac{(70)^2}{4} + \frac{(48)^2}{4} + \frac{(85)^2}{4} + \frac{(92)^2}{4} - 5439.06\right) = \mathbf{284.18}$

Step 8: SSE = TSS –SSC- SSR- SST = $329.93 - 29.18 - 4.68 - 284.18 = \mathbf{11.87}$

Source of Variance	**Sum of Squares**	**Degree of Freedom**	**Mean sum of Squares**	**Variance Ratio**	**Table Value at 5% level**
Between Columns	SSC=4.68	K-1=3	$MSC = \frac{SSC}{K-1} = \frac{4.68}{3} = 1.56$	$F_c = \frac{MSE}{MSC} = \frac{1.97}{1.56} = \mathbf{1.26}$	$F_c(6{,}3) = 8.94$
Between Rows	SSR=29.18	K-1=3	$MSR = \frac{SSR}{K-1} = \frac{29.19}{3} = 9.72$		$F_R(3{,}6) = 4.75$

Between Treatments	SST=284.18	K-1=3	$MST = \frac{SST}{K-1} = \frac{284.18}{3} = 94.72$	$F_R = \frac{MSR}{MSE} = \frac{9.72}{1.97} = \mathbf{4.91}$ $F_T = \frac{MST}{MSE} = \frac{94.72}{1.97} = \mathbf{47.86}$	$F_T(3,6) = 4.75$
Error	SSE=11.875	(K-1)(K-2) =6	$MSE = \frac{SSE}{(K-1)(K-2)} = \frac{11.875}{6} = 1.97$		
Total	TSS=329.93	$K^2 - 1 = 15$	Condition Always F >1		

Calculated $Value\ F_C(1.26) < table\ value\ F_C(8.94)\ so\ we\ Accept\ \ H_0$

Calculated $Value\ F_R(4.91) > table\ value\ F_R(4.75)\ so\ we\ Reject\ \ H_0$

Calculated $Value\ F_T(47.86) > table\ value\ F_T(4.75)\ so\ we\ Reject\ \ H_0$

Conclusion:

There is no significance difference between Seeds, But there is a significance difference between Treatments & Lands

Describe the Latin Square Layout.

Latin Square Designs are probably not used as much as they should be - they are very efficient designs. Latin square designs allow for two blocking factors. In other words, these designs are used to simultaneously control (or eliminate) **two sources of nuisance variability**. For instance, if you had a plot of land the fertility of this land might change in both directions, North -- South and East -- West due to soil or moisture gradients. So, both rows and columns can be used as blocking factors. However, you can use Latin squares in lots of other settings. As we shall see, Latin squares can be used as much as the RCBD in industrial experimentation as well as other experiments.

Whenever, you have more than one blocking factor a Latin square design will allow you to remove the variation for these two sources from the error variation. So, consider we had a plot of land, we might have blocked it in columns and rows, i.e. each row is a level of the row factor, and each column is a level of the column factor. We can remove the variation from our measured response in both directions if we consider both rows and columns as factors in our design.

The Latin Square Design gets its name from the fact that we can write it as a square with Latin letters to correspond to the treatments. The treatment factor levels are the Latin letters in the Latin square design. The number of rows and columns has to correspond to the number of treatment levels. So, if we have four treatments then we would need to have four rows and four columns in order to create a Latin square. This gives us a design where we have each of the treatments and in each row and in each column.

This is just one of many 4×4 squares that you could create. In fact, you can make any size square you want, for any number of treatments - it just needs to have the following property associated with it - that each treatment occurs only once in each row and once in each column.

Consider another example in an industrial setting: the rows are the batch of raw material, the columns are the operator of the equipment, and the treatments (A, B, C and D) are an industrial process or protocol for producing a particular product.

What is the model? We let:

$$y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + e_{ijk}$$

$$\boldsymbol{i} = 1,\ldots, \boldsymbol{t}$$
$$\boldsymbol{j} = 1,\ldots, \boldsymbol{t}$$

$[\boldsymbol{k} = 1,\ldots, \boldsymbol{t}]\ where - \boldsymbol{k} = \boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})$ and the total number of observations

N = t^2 (the number of rows times the number of columns) and *t* is the number of treatments.

Note that a Latin Square is an incomplete design, which means that it does not include observations for all possible combinations of i, j and k. This is why we use notation $\boldsymbol{k} = \boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})$. Once we know the row and column of the design, then the treatment is specified. In other words, if we know *i* and *j*, then *k* is specified by the Latin Square design.

This property has an impact on how we calculate means and sums of squares, and for this reason we can not use the balanced ANOVA command in Minitab even though it looks perfectly balanced. We will see later that although it has the property of orthogonality, you still cannot use the balanced ANOVA command in Minitab because it is not complete.

An assumption that we make when using a Latin square design is that the three factors (treatments, and two nuisance factors) **do not interact**. If this assumption is violated, the Latin Square design error term will be inflated.

The randomization procedure for assigning treatments that you would like to use when you actually apply a Latin Square, is somewhat restricted to preserve the structure of the Latin Square. The ideal randomization would be to select a square from the set of all possible Latin squares of the specified size. However, a more practical randomization scheme would be to select a standardized Latin square at random (these are tabulated) and then:

1. randomly permute the columns,
2. randomly permute the rows, and then
3. assign the treatments to the Latin letters in a random fashion.

Consider a factory setting where you are producing a product with 4 operators and 4 machines. We call the columns the operators and the rows the machines. Then you can randomly assign the specific operators to a row and the specific machines to a column. The treatment is one of four protocols for producing the product and our interest is in the average time needed to produce each product. If both the machine and the operator have an effect on the time to produce, then by using a Latin Square Design this variation due to machine or operators will be effectively removed from the analysis.

The following table gives the degrees of freedom for the terms in the model.

AOV	***df***	***df* for the example**
Rows	*t*-1	3
Cols	*t*-1	3
Treatments	*t*-1	3
Error	(*t*-1)(*t*-2)	6
Total	(t^2 - 1)	15

A Latin Square design is actually easy to analyze. Because of the restricted layout, one observation per treatment in each row and column, the model is orthogonal.

If the row, ρ_i, and column, β_j, effects are random with expectations zero, the expected value of Y_{ijk} is $\mu + \tau_k$. In other words, the treatment effects and treatment means are orthogonal to the row and column effects. We can also write the sums of squares, as seen in Table 4.10 in the text.

We can test for row and column effects, but our focus of interest in a Latin square design is on the treatments. Just as in RCBD, the row and column factors are included to reduce the error variation but are not typically of interest. And, depending on how we've conducted the experiment they often haven't been randomized in a way that allows us to make any reliable inference from those tests.

Note: if you have missing data then you need to use the general linear model and test the effect of treatment after fitting the model that would account for the row and column effects.

In general, the General Linear Model tests the hypothesis that:

$$\boldsymbol{H}_0 : \tau_i = 0 \quad vs. \quad \boldsymbol{H}_a : \tau_i \neq 0$$

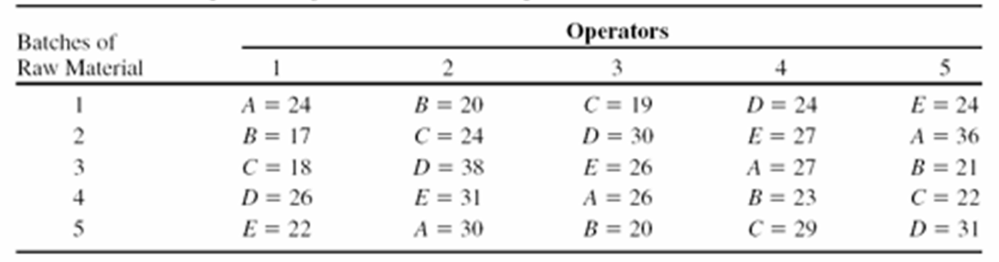
To test this hypothesis we will look at the *F*-ratio which is written as:

$$F = \frac{MS(\tau k|\mu, \rho i, \beta j)}{MSE(\mu, \rho i, \beta j, \tau k)} \sim F((t-1),(t-1)(t-2))$$

To get this in Minitab you would use GLM and fit the three terms: rows, columns and treatments.
The *F* statistic is based on the adjusted MS for treatment.

The Rocket Propellant Problem – A Latin Square Design

Table 4-8 Latin Square Design for the Rocket Propellant Problem



Batches of Raw Material	Operators				
	1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

Table 4-13 (4-12 in 7th ed) shows some other Latin Squares from *t* = 3 to *t* = 7 and states the number of different arrangements available.

Statistical Analysis of the Latin Square Design

The statistical (effects) model is:

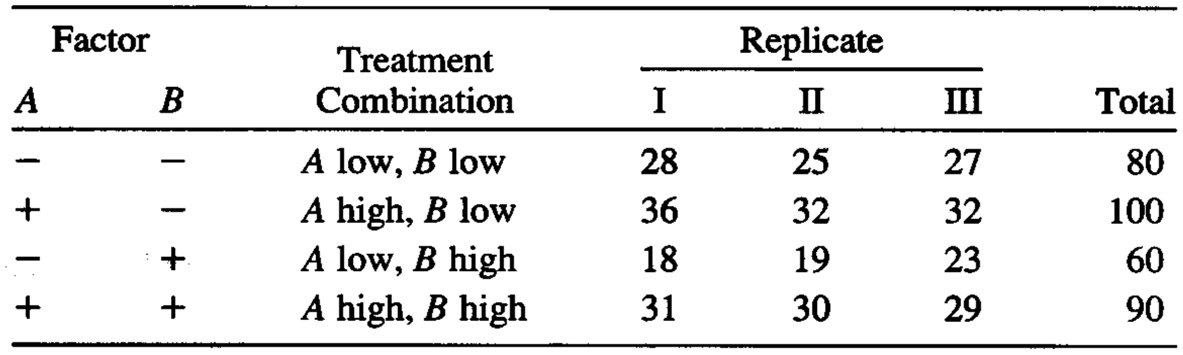
$$Y_{ijk} = \mu + \rho_i + \beta_j + \tau_k + \varepsilon_{ijk} \begin{cases} i = 1,2,\dots,p \\ j = 1,2,\dots,p \\ k = 1,2,\dots,p \end{cases}$$

but $\boldsymbol{k = d(i, j)}$ shows the dependence of *k* in the cell *i, j* on the design layout, and *p* = *t* the number of treatment levels.
The statistical analysis (ANOVA) is much like the analysis for the RCBD.

The 2^2 Factorial Design

Two factors, A and B, and each factor has two levels, low and high

The concentration of reactant v.s. the amount of the catalyst



Factor		Treatment Combination	Replicate			Total
A	*B*		I	II	III	
−	−	*A* low, *B* low	28	25	27	80
+	−	*A* high, *B* low	36	32	32	100
−	+	*A* low, *B* high	18	19	23	60
+	+	*A* high, *B* high	31	30	29	90

- "-" And "+" denote the low and high levels of a factor, respectively
- Low and high are arbitrary terms
- Geometrically, the four runs form the corners of a square
- Factors can be quantitative or qualitative, although their treatment in the final model will be different

-

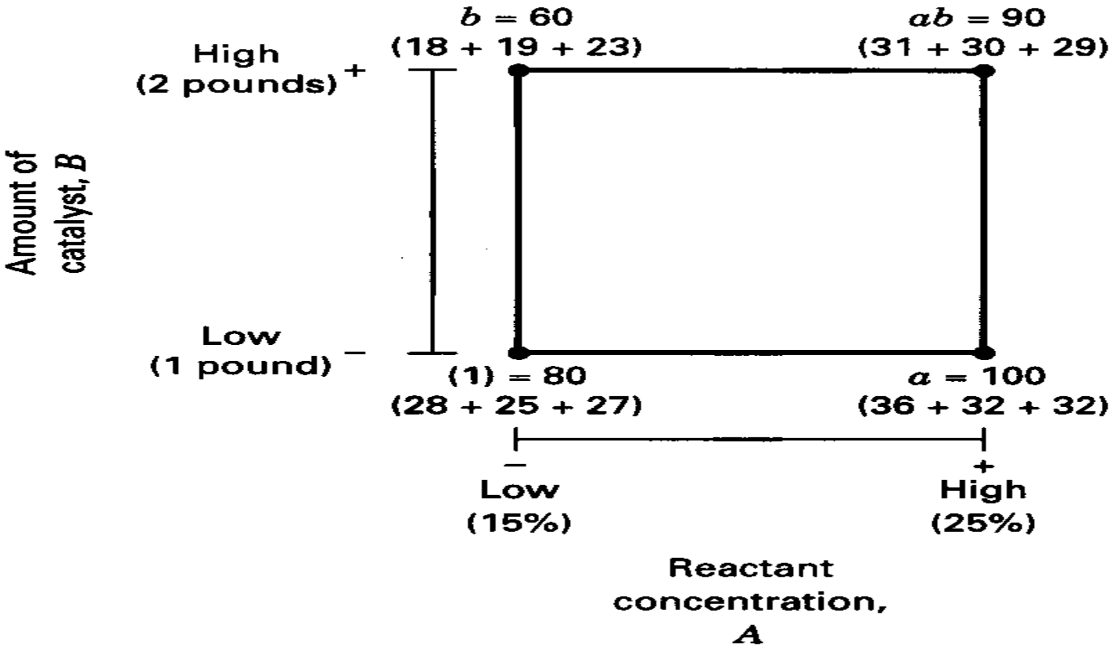


Figure 6-1 **Treatment combinations in the 2^2 design.**

Average effect of a factor = the change in response produced by a change in the level of that factor averaged over the levels if the other factors.

- (1), *a, b* and *ab*: the total of *n* replicates taken at the treatment combination.

The main effects

$$A = \frac{1}{2n}\{[ab-b]+[a-(1)]\} = \frac{1}{2n}[ab+a-b-(1)]$$

$$= \frac{ab+a}{2n} - \frac{b+(1)}{2n} = \bar{y}_{A^+} - \bar{y}_{A^-}$$

$$B = \frac{1}{2n}\{[ab-a]+[b-(1)]\} = \frac{1}{2n}[ab+b-a-(1)]$$

$$= \frac{ab+b}{2n} - \frac{a+(1)}{2n} = \bar{y}_{B^+} - \bar{y}_{B^-}$$

The interaction effect

$$AB = \frac{1}{2n}\{[ab-b]-[a-(1)]\} = \frac{1}{2n}[ab+(1)-a-b]$$

$$= \frac{ab+(1)}{2n} - \frac{b+a}{2n}$$

The total effects

$$Contrast_A = ab+a-b-(1)$$

$$Contrast_B = ab+b-a-(1)$$

$$Contrast_{AB} = ab+(1)-a-b$$

- Sum of squares:

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n}$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n}$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

$$SS_{AB} = \frac{[ab + (1) - b - a]^2}{4n}$$

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

The Analysis of Variance is completed by computing the total sum of squares $\boldsymbol{SS_T}\ \boldsymbol{with}\ (\boldsymbol{4n-1})\boldsymbol{d.f}$ as usual and the Error Sum of Squares $\boldsymbol{SS_E}\ \boldsymbol{with}\ (\boldsymbol{4(n-1)d.f})$

Table of plus and minus signs

	I	A	B	AB
(1)	+	–	–	+
a	+	+	–	–
b	+	–	+	–
ab	+	+	+	+

- The regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

  - x_1 and x_2 are coded variables that represent the two factors, i.e. x_1 (or x_2) only take values on –1 and 1.
  - Use least square method to get the estimations of the coefficients
  - For that example,

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

  - Model adequacy: residuals and normal probability plot

Degree of Freedom as follows

RBD		LSD	
Treatments	$4-1=3$	Treatments	$4-1=3$
Blocks	$b-1$	Row	$4-1=3$
Error	$3(b-1)$	Columns	$4-1=3$
Total	$4(b-1)$	Error	$(4-1)(4-2)=6$
		Total	15

Problem-1

14.037	14.165	13.972	13.907
14.821	14.757	14.843	14.878
13.880	13.860	14.032	13.914
14.888	14.921	14.415	14.932

The above table presents the results of a 2^2 factorial design with n = 4 replicates, using the factor

A= deposition time and B= Arsenic Flow rate

The two Level of deposition time are - = short & + = long ,

The two Level of Arsenic Flow rate - = 55 % & + = 59% , The response variable epitaxial layer thickness (μ_m)

2^2 design for Epitaxial layer thickness (μ_m)

Treatment Combinations	**Design Factors**			**Thickness (μ_m)**				**Thickness**	
	A	**B**	**AB**					**Total**	**Average**
(1)	-	-	+	14.037	14.165	13.972	13.907	56.081	14.020
a	+	-	-	14.821	14.757	14.843	14.878	59.299	14.825
b	-	+	-	13.880	13.860	14.032	13.914	55.686	13.922
ab	+	+	+	14.888	14.921	14.415	14.932	59.156	14.789

$$A = \frac{1}{2n}[a + ab - b - (1)] = \frac{1}{2(4)}[59.299 + 59.156 - 55.686 - 56.081] = \frac{1}{8}[6.688] = 0.836$$

$$B = \frac{1}{2n}[b + ab - a - (1)] = \frac{1}{2(4)}[55.686 + 59.156 - 59.299 - 56.081] = \frac{1}{8}[-0.536] = -0.067$$

$$AB = \frac{1}{2n}[(1) + ab - a - b] = \frac{1}{2(4)}[56.081 + 59.156 - 59.299 - 55.686] = \frac{1}{8}[0.256] = 0.032$$

$$SS_A = \frac{[a + ab - b - (1)]^2}{4n} = \frac{[6.688]^2}{16} = 2.7956$$

$$SS_B = \frac{[b + ab - a - (1)]^2}{4n} = \frac{[-0.536]^2}{16} = 0.0181$$

$$SS_{AB} = \frac{[(1) + ab - a - b]^2}{4n} = \frac{[0.252]^2}{16} = 0.0040$$

Source of variations	Sum of squares	Degree of freedom	Mean Square	Variance Ratio	Table Value 5% Level	Table Value 1% Level
A	$SS_A = 2.7956$	1	$MS_A = \frac{SS_A}{d.f} = 2.7956$	$F_A = \frac{MS_A}{MS_E} = 134.4$	$F_A(1,12) = 4.75$	$F_A(1,12) = 9.33$
B	$SS_B = 0.0181$	1	$MS_B = \frac{SS_B}{d.f} = 0.0181$	$F_B = \frac{MS_B}{MS_E} = 0.87$	$F_B(1,12) = 4.75$	$F_B(1,12) = 9.33$
AB	$SS_{AB} = 0.0040$	1	$MS_{AB} = \frac{SS_{AB}}{d.f} = 0.0040$	$F_{AB} = \frac{MS_{AB}}{MS_E} = 0.19$	$F_{AB}(1,12) = 4.75$	$F_{AB}(1,12) = 9.33$
Error	$SS_E = 0.2495$	12	$MS_E = \frac{SS_E}{d.f} = 0.0208$	Always F > 1		

$$SS_T = \{(14.037)^2 + (14.165)^2 + \cdots + (14.415)^2 + (14.932)^2\}$$

$$-\left\{\frac{56.081 + 59.299 + 55.686 + 59.156}{16}\right\}^2$$

$$SS_T = 3.0672$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} = 3.0672 - 2.7956 - 0.0181 - 0.004 = 0.2495$$

Analysis of Variance Epitaxial Process Experiment

Here Cal $F_A >$ Table F_A

Cal $F_B <$ Table F_B

Cal $F_{AB} <$ Table F_{AB}

---

2) Find out the main effects and interactions in the following 2^2 Factorial experiment and write down the analysis of variance Table

	(I) 00	a 10	b 01	ab 11
BLOCK-I	64	25	30	6
BLOCK-II	75	14	50	33
BLOCK-III	76	12	41	17
BLOCK-IV	75	33	25	10

Soln:

Taking Deviation y = 37 , We get

Treatment Combinations	**BLOCKS**				**TOTAL**	$\boldsymbol{X_1^2}$	$\boldsymbol{X_2^2}$	$\boldsymbol{X_3^2}$	$\boldsymbol{X_4^2}$
	I X_1	II X_2	III X_3	IV X_4					
(y_1) (1)	27	38	39	38	142	729	1444	1521	1444
(y_2) a	-12	-23	-25	-4	-64	144	529	625	16
(y_3) b	-7	13	4	-12	-2	49	169	16	144
(y_4) ab	-31	-4	-20	-27	-82	961	16	400	729
Total	**-23**	**24**	**-2**	**-5**	**-6**	**1883**	**2158**	**2562**	**2333**

Step:1 N = 16

Step:2 T = -6

Step:3 Correction Factor (CF) = $\frac{T^2}{N} = \frac{(-6)^2}{16} = \frac{36}{16} = \mathbf{2.25}$

Step:4 TSS = $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - \frac{T^2}{N} = 1883 + 2158 + 2562 + 2333 - 2.25 = \mathbf{8933.75}$

Step:5 SSC = $\frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} + \frac{(\sum X_4)^2}{N_1} - \frac{T^2}{N} = \frac{(-23)^2}{4} + \frac{(24)^2}{4} + \frac{(-2)^2}{4} + \frac{(-5)^2}{4} - 2.25$

SSC = $\mathbf{281.25}$

N_1 → Number of elements in each row

Step:6 SSR = $\frac{(\sum Y_1)^2}{N_2} + \frac{(\sum Y_2)^2}{N_2} + \frac{(\sum Y_3)^2}{N_2} + \frac{(\sum Y_4)^2}{N_2} - \frac{T^2}{N} = \frac{(142)^2}{4} + \frac{(-64)^2}{4} + \frac{(-2)^2}{4} + \frac{(-82)^2}{4} - 2.25$

SSR = $\mathbf{7745.75}$

Step:7: SSE = TSS – SSC – SSR = 8933.75 – 281.25 – 7744.75 = **907.75**

For 2^2 Experiment

Contrast A = $a + ab - b - (1) = [-64 - 82 + 2 - 142] = -286$

Contrast B = $b + ab - a - (1) = [-2 - 82 + 64 - 142] = -162$

Contrast AB = $(1) + ab - b - a = [142 - 82 + 2 + 64] = 126$

Main Effects:

$$\text{A} = \frac{1}{2}\big(\, a + ab - b - (1)\big) = -\frac{286}{2} = -143$$

$$\text{B} = \frac{1}{2}\big(b + ab - a - (1)\big) = -\frac{162}{2} = -81$$

$$\text{AB} = \frac{1}{2}(\,(1) + ab - b - a) = \frac{126}{2} = 63$$

$$SS_A = \frac{\big(\, a + ab - b - (1)\big)^2}{16} = \frac{(-286)^2}{16} = 5112.25$$

$$SS_B = \frac{\big(\, b + ab - a - (1)\big)^2}{16} = \frac{(-162)^2}{16} = 1640.25$$

$$SS_{AB} = \frac{(\,(1) + ab - b - a)^2}{16} = \frac{(126)^2}{16} = 992.25$$

Source of variance	Degree of freedom	Sum of Squares	Mean Square	Variance Ratio	Table Value	
					5% Level	1% Level
Blocks	3	281.5	93.83	$\frac{100.86}{93.83} = 1.075$	$F(9,3) = 8.81$	27.35
Treatments	3	7744.75	2581.58	$\frac{2581.88}{100.86} = 25.6$	$F(3,9) = 3.86$	6.99
A	1	5112.25	5112.25	$\frac{5112.25}{100.86} = 50.69$	$F(1,9) = 5.12$	6.99
B	1	1640.25	1640.25	$\frac{1640.25}{100.86} = 16.26$	$F(1,9) = 5.12$	6.99
AB	1	992.25	992.25	$\frac{992.25}{100.86} = 9.84$	$F(1,9) = 5.12$	6.99
Error	9	907.75	100.86	Always F > 1		

$$Error\ (d.f) = N - c - r + 1 = 16 - 4 - 4 + 1 = 9$$

$$Cal\ F_A > Tab\ F_A$$

$$Cal\ F_B > Tab\ F_B$$

$$Cal\ F_{AB} > Tab\ F_{AB}$$

Unit -5 STATISTICAL QUALITY CONTROL

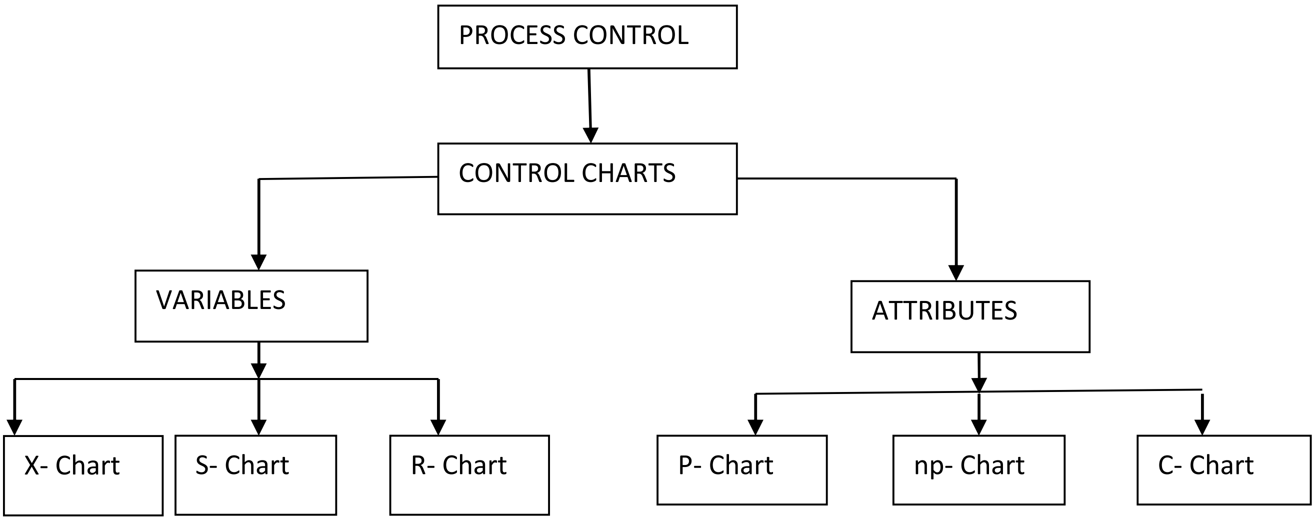
Control Charts for Measurements

Control Charts:

A Control charts provides a basis for deciding whether the variation in the output is due to random causes or due to assignable causes, It will assist us in making decision whether to adjust the process or not

A control charts is designed to display successive measurements of process with a centre line and control limits, The control limits are above and below the centre line, The control limits are equidistance from the centre line are known as UPPER CONTROL LIMIT (UCL) & LOWER CONTROL LIMIT(LCL)

It has two types (i) control charts for Variables, (ii) control charts for attributes



- **Construction of X- Chart:**

Draw Independent samples each of size "n" from a largest production process,
Let $\overline{x_1}, \overline{x_2}, \overline{x_3} \dots \dots \dots .. \overline{x_n}$ be the means of these samples

The Mean of all these mean is $\boxed{\boldsymbol{X} = \frac{\overline{x_1}+\overline{x_2}+\overline{x_3}+\ldots\ldots\ldots+\overline{x_n}}{n}}$
The Control limits are given by

$$\boldsymbol{UCL = \bar{x} + 3 * SE(\bar{x}), \qquad LCL = \bar{x} - 3 * SE(\bar{x}),\ CL = \bar{x}}$$

Where $\boldsymbol{SE(\bar{x})} = \frac{\sigma}{\sqrt{n}}, \sigma$ being the Standard Deviation (SD) of the production, Ifσ is not available the SD of sampling distribution of the mean can be taken as the best estimate of σ,
In the case of small sample the estimate of Standard Error(SE) of $\bar{x}\ is\ \frac{\sigma}{n-1}$, Alternatively in the case of small sample of size less than 20, the following results for control limits are available

$$\boldsymbol{UCL = \bar{x} + A_2\bar{R}, \qquad LCL = \bar{x} - A_2\bar{R},\ CL = \bar{x}}$$

Where $\bar{R}$is the mean of sample range $R_1, R_2, R_3, \dots \dots .. R_n$ obtained from the "k" samples, The factor A_2 has to be determined from statistical tables when the sample size "n" is Known

- **Range Chart[R-Chart]:**

For the sample size less than 20 the range provides a good estimate of , Hence to measure the variance in the variable, Range chart is used

- **Construction of R- Chart:**

Let $R_1, R_2, R_3, \dots \dots .. R_k$ be the values of the range in "k" samples the mean of all these range is

$$\overline{\boldsymbol{R}} = \frac{\boldsymbol{R_1, R_2, R_3, \ldots \ldots . . R_k}}{\boldsymbol{k}}$$

The Control limits are given by $\boldsymbol{UCL = D_4\overline{R}}$, $\boldsymbol{LCL = D_3\overline{R}}$, $The\ factor\ D_3 \& D_4$ are determine from the statistical table for known sample size

In this case of production process if the characteristics to be tested in non-measurable control charts for attributes used to find out whether a production process is under control or not, The most common control charts under this category are

(i) Control charts for number of Defectives

(ii) Control charts for Fraction Defectives

1) **Distinguish between variables and attributes in connection with the quality characteristic of a product (A/M 15)**

Soln: In simple terms attributes control is at the limits, variables control within the limits. Concerning the data that is generated by each concept, attributes data is discreet whereas variables data is continuous.

2) **Write the formula for control chart values (Central line, UCL & LCL) of a C-Chart (N/D 2015)**

Soln:

$$CL = \bar{c}, \quad UCL = \left[\bar{c} + 3\left(\sqrt{\bar{c}}\right)\right], \quad LCL = \left[\bar{c} - 3\left(\sqrt{\bar{c}}\right)\right]$$

3) **Control charts for $\overline{X}$ and R to be set up for an important quality characteristic the sample size is n = 5 and $\overline{X}$ and R are computed for each of 35 preliminary samples. The summary data are $\sum_{i=1}^{35} \boldsymbol{X_i} = \mathbf{7805}$ $\sum_{i=1}^{35} \boldsymbol{R_i} = \mathbf{1200}$ Find the control limit for $\overline{X}$ and R (N/D 2015)**

Soln: $\overline{X} = \frac{\sum x_i}{N} = \frac{7805}{35} = 202.42,\ \ \overline{R} = \frac{\sum R_i}{N} = \frac{1200}{35} = 34.28$

<u>Control Limit For $\overline{X}$:</u>

$$CL = \overline{X} = \mathbf{202.42}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 202.42 + (0.129)34.28 = \mathbf{206.84}, \quad A_2 = 0.129\ from\ table$$

$$LCL = \overline{X} - (A_2)\overline{R} = 202.42 - (0.129)34.28 = \mathbf{198}, \quad A_2 = 0.129\ from\ table$$

<u>Control Limit For $\overline{R}$:</u>

$$CL = \overline{R} = \mathbf{34.28}$$

$$UCL = (D_4)\overline{R} = (1.541)(34.28) = \mathbf{52.83}, \qquad D_4 = 1.541\ from\ \ table$$

$$LCL = (D_3)\overline{R} = (0.459)(34.28) = \mathbf{15.73}, \qquad D_3 = 0.459\ from\ \ table$$

4) **What is the purpose of using control channel? (A/M 15)**

Soln: The purpose of drawing a control chart is to detect any changes in the process that would be evident by any abnormal points listed on the graph from the data collected. If these points are plotted in "real time", the operator will immediately see that the point is exceeding one of the control limits, or is heading in that direction, and can make an immediate adjustment. The operator should also record on the chart the cause of the drift, and what was done to correct the problem bringing the process back into a "state of control".

5) **A Garments was sampled on 10 consecutive hours of production, The number of defects found per garments is given below Defects: 5,1,7,0,2,3,4,0,3,2 compute the Upper and lower control limits for monitoring the number of Defects (A/M 15)**

Soln:

$$CL = \bar{c} = 2.7,$$

$$UCL = \left[\bar{c} + 3\left(\sqrt{\bar{c}}\right)\right] = \left[2.7 + 3\left(\sqrt{2.7}\right)\right] = 7.63$$

$$LCL = \left[\bar{c} - 3\left(\sqrt{\bar{c}}\right)\right] = \left[2.7 - 3\left(\sqrt{2.7}\right)\right] = 0$$

PART-B

1) **Given below are the values of sample mean $\overline{x}$and sample range R for 10 samples each of size 5, Draw the appropriate mean & range charts and comment on the state of control of the process**

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean($\overline{x_i}$):	**43**	**49**	**37**	**44**	**45**	**37**	**51**	**46**	**43**	**47**
Range ($\overline{R_i}$)	**5**	**6**	**5**	**7**	**7**	**4**	**8**	**6**	**4**	**6**

Soln: N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{43+49+37+44+45+37+51+46+43+47}{10} = \mathbf{44.2}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{5+6+5+7+7+4+8+6+4+6}{10} = \mathbf{5.8}$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{44.2}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 44.2 + (0.577)5.8 = \mathbf{47.55}, \quad A_2 = 0.577 \ from\ table$$

$$LCL = \overline{X} - (A_2)\overline{R} = 44.2 - (0.577)5.8 = \mathbf{40.85}, \quad A_2 = 0.577 \ from\ table$$

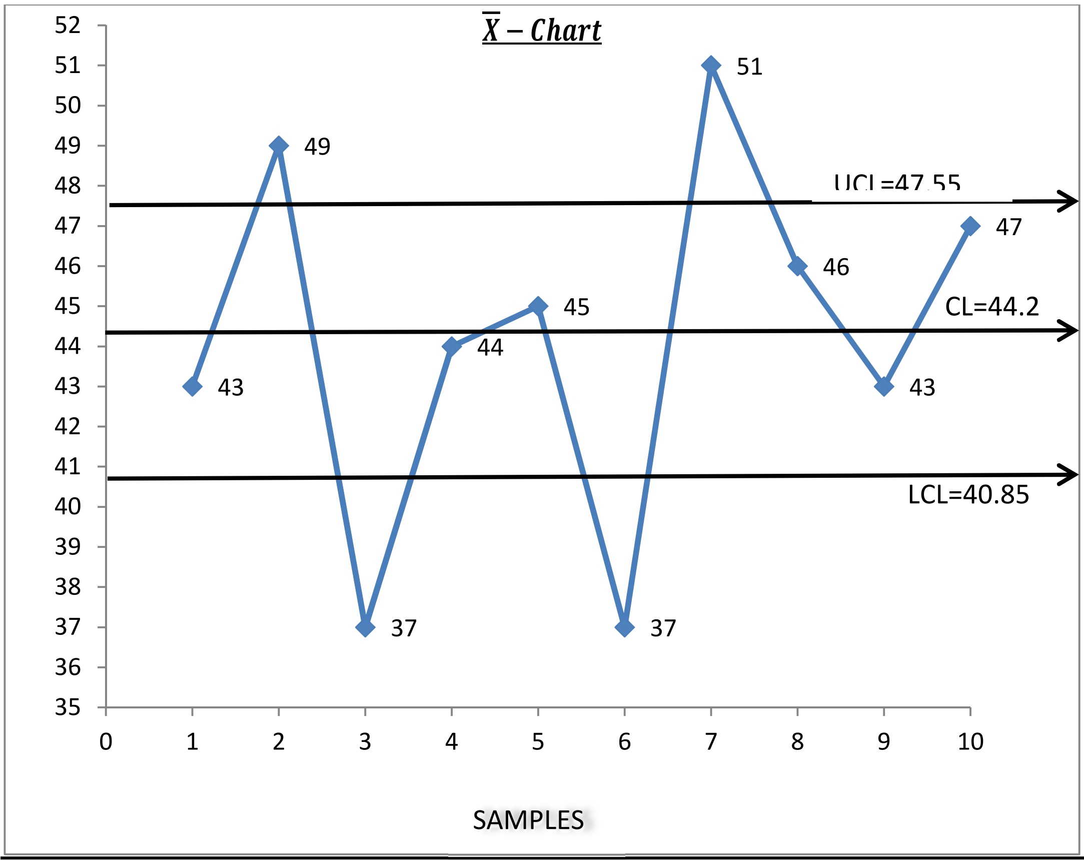
Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{5.8}$$

$$UCL = (D_4)\overline{R} = (2.115)(5.8) = \mathbf{12.27}, \qquad D_4 = 2.115\, from\ table$$

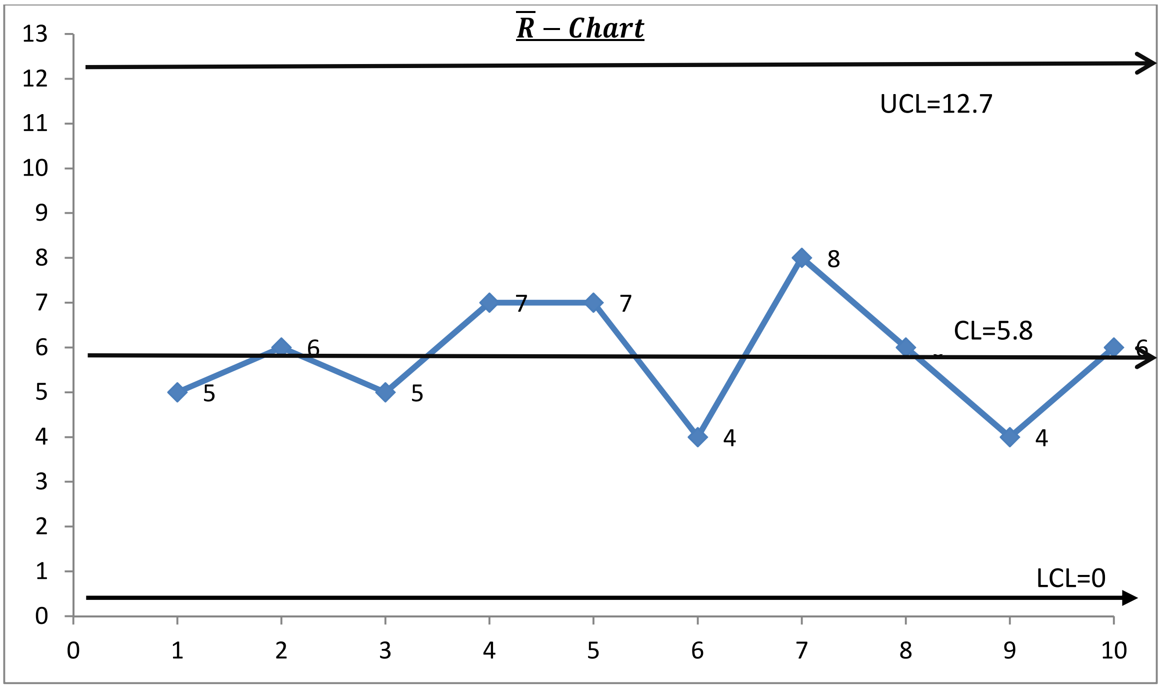
$$LCL = (D_3)\overline{R} = (0)(5.8) = \mathbf{0}, \qquad D_3 = 0\, from\ table$$

Control Limit For $\overline{X} - Chart$:



Since sample number 2,3,6&7 falls outside the control limits the statistical process is out of control according to $\overline{X} - Chart$

Control Limit For $\overline{R} - Chart$:



Since all the sample fall within the CL, The Statistical Process Under control for R-chart

2) **The following are the sample mean and ranges for 10 samples, Each of size 5, Construct the control chart for Mean & Range and comment on the nature of control**

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean:	**12.8**	**13.1**	**13.5**	**12.9**	**13.2**	**14.1**	**12.1**	**15.5**	**13.9**	**14.2**
Range:	**2.1**	**3.1**	**3.9**	**2.1**	**1.9**	**3**	**2.5**	**2.8**	**2.5**	**2**

Soln:

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{12.8 + 13.1 + 13.5 + 12.9 + 13.2 + 14.1 + 12.1 + 15.5 + 13.9 + 14.2}{10}$$

$$\boldsymbol{\overline{X} = 13.53}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{2.1 + 3.1 + 3.9 + 2.1 + 1.9 + 3 + 2.5 + 2.8 + 2.5 + 2}{10} = \mathbf{2.59}$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{13.53}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 13.53 + (0.577)2.59 = \mathbf{15.02}, \quad A_2 = 0.577 \ from\ table$$

$$LCL = \overline{X} - (A_2)\overline{R} = 13.53 - (0.577)2.59 = \mathbf{12.04}, \quad A_2 = 0.577 \ from\ table$$

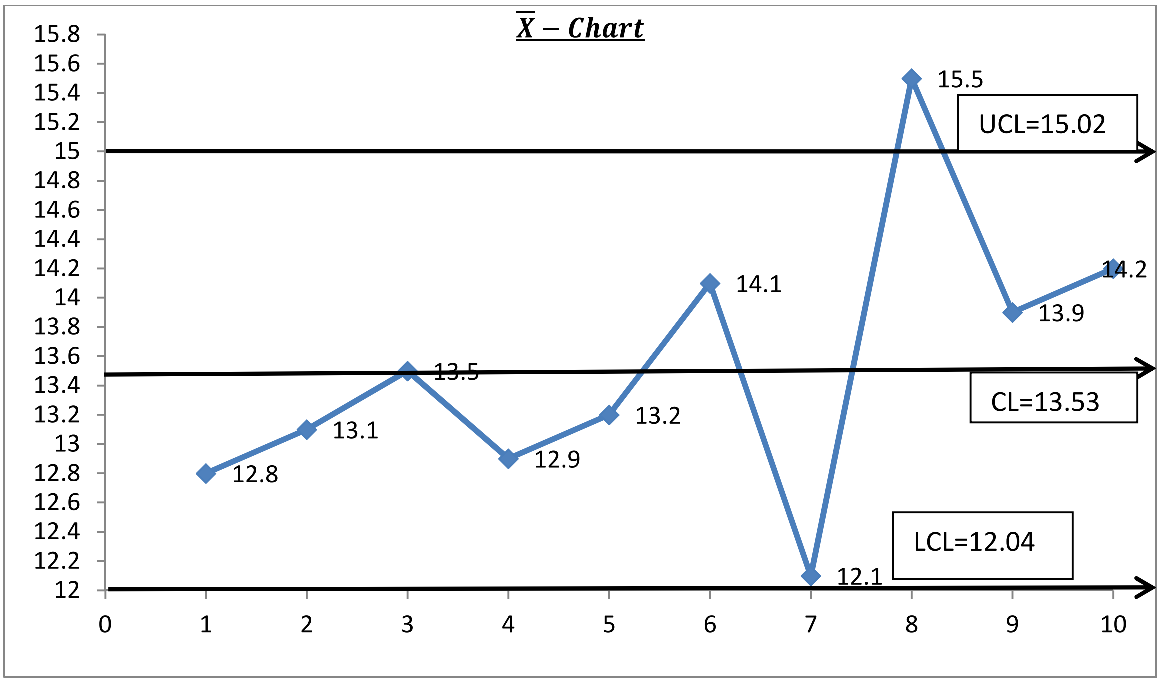
Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{2.59}$$

$$UCL = (D_4)\overline{R} = (2.115)(2.59) = \mathbf{5.48}, \qquad D_4 = 2.115\, from\ table$$

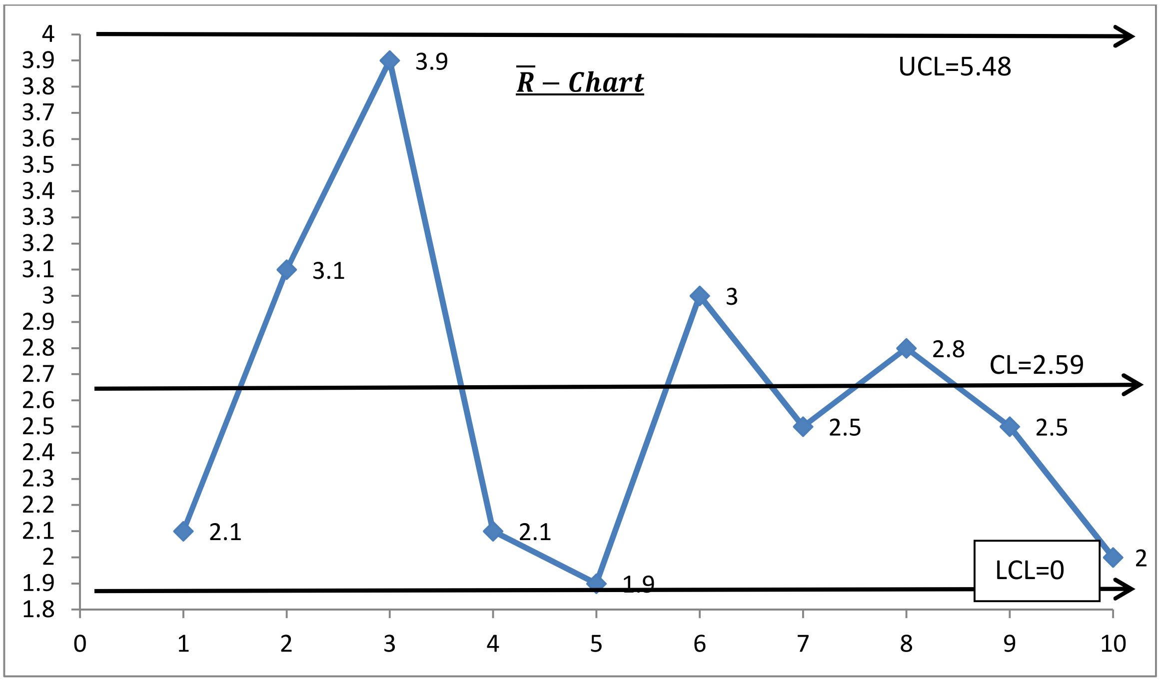
$$LCL = (D_3)\overline{R} = (0)(2.59) = \mathbf{0}, \qquad D_3 = 0\, from\ table$$

Control Limit For $\overline{X} - Chart$:



Since sample number 8 falls outside the control limits the statistical process is out of control according to $\overline{X} - Chart$

Control Limit For $\overline{R} - Chart$:



Since all the sample fall within the CL, The Statistical Process Under control for R-chart

3) **The Following gives the sample mean & Range for 10 samples each of size 6 in the production of certain component, Construct the control chart for mean & range, Comment on the nature of control**

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean:	**37.3**	**49.8**	**51.5**	**59.2**	**54.7**	**34.7**	**51.4**	**61.4**	**70.7**	**75.3**
Range:	**9.5**	**12.8**	**10**	**9.1**	**7.8**	**5.8**	**14.5**	**2.8**	**3.7**	**8**

Soln:

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{37.3 + 49.8 + 51.5 + 59.2 + 54.7 + 34.7 + 51.4 + 61.4 + 70.7 + 75.3}{10}$$

$$\overline{\boldsymbol{X}} = \mathbf{54.6}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{9.5 + 12.8 + 10.9.1 + 7.8 + 5.8 + 14.5 + 2.8 + 3.7 + 8}{10} = \mathbf{8.4}$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{54.6}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 54.6 + (0.4829)8.4 = \mathbf{58.65}, \quad A_2 = 0.4829 \ \textit{from table}$$

$$LCL = \overline{X} - (A_2)\overline{R} = 54.6 - (0.4829)8.4 = \mathbf{50.54}, \quad A_2 = 0.4829 \ \textit{from table}$$

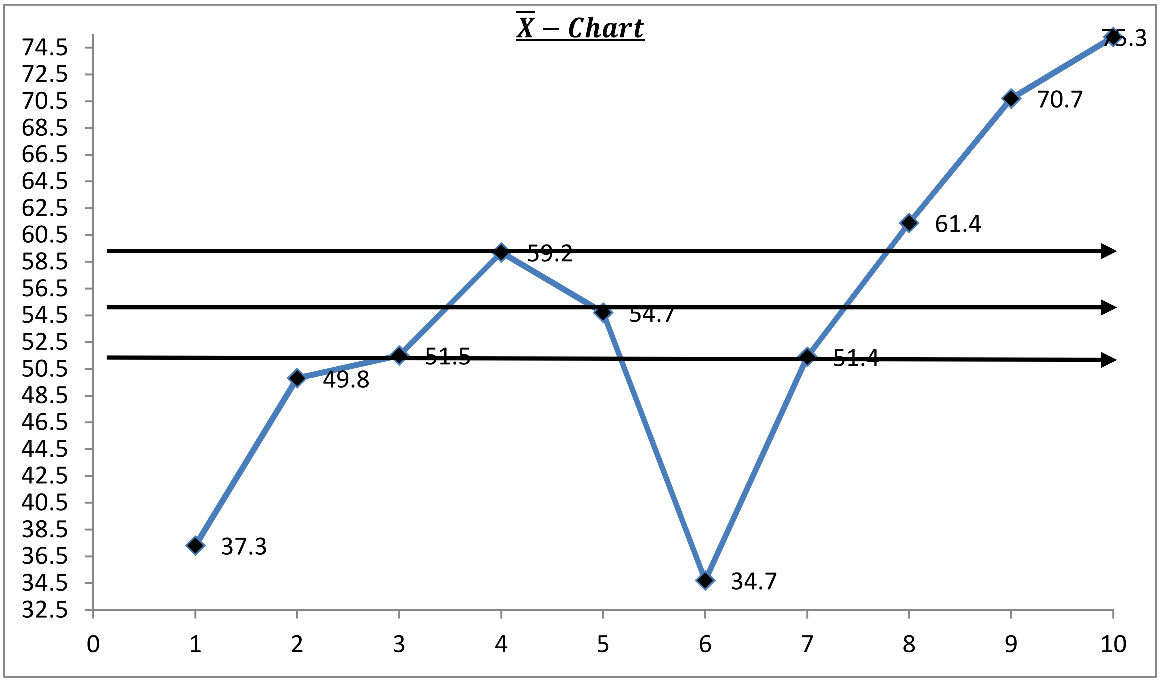
Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{8.4}$$

$$UCL = (D_4)\overline{R} = (2.004)(8.4) = \mathbf{16.834}, \qquad D_4 = 2.115 \ \textit{from table}$$

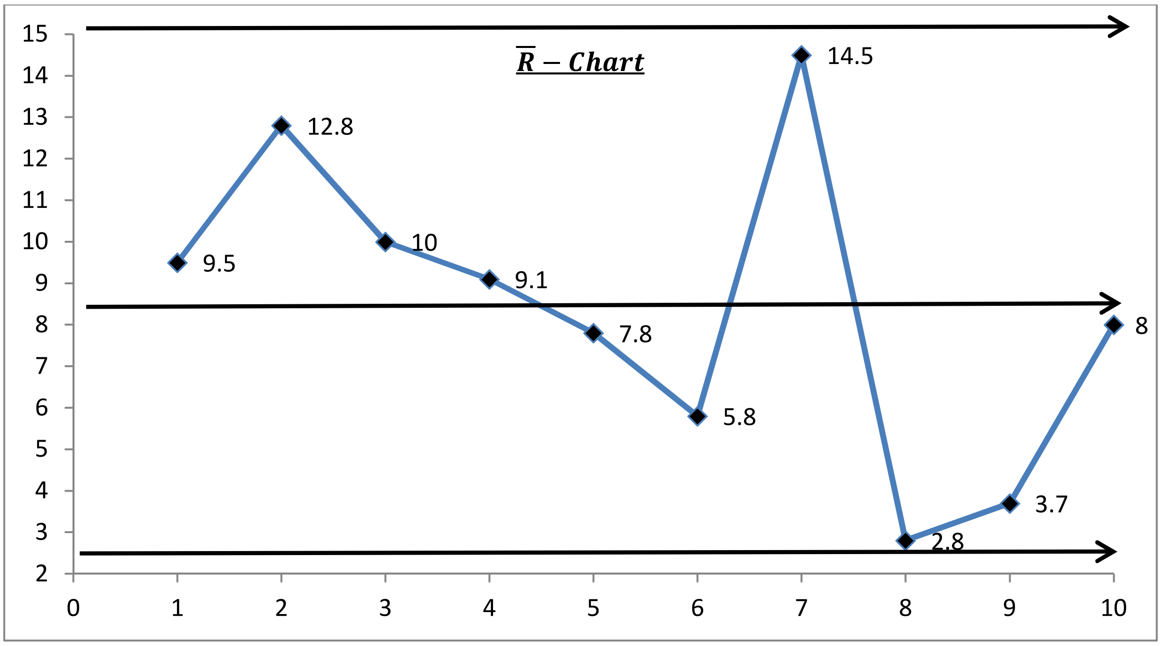
$$LCL = (D_3)\overline{R} = (0)(8.4) = \mathbf{0}, \qquad D_3 = 0 \ \textit{from table}$$

Control Limit For $\overline{X} - \boldsymbol{Chart}$:



Since sample number 1,2,4,6,8,9&10 falls outside the control limits the statistical process is out of control according to $\overline{X} - Chart$

Control Limit For $\overline{R} - \boldsymbol{Chart}$:



Since all the sample fall within the CL, The Statistical Process Under control for R-chart

4) **The following Data gives the measurement of 10 samples each of size 5 in the production process taken in an interval of two hours, Calculate the sample mean and ranges and drawn the control chart for mean and range**

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Observed Measurement of 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**	**46**	**53**	**53**	**47**	**54**
	53	**50**	**47**	**53**	**45**	**50**	**45**	**57**	**51**	**56**

Soln:

Sample	1	2	3	4	5	6	7	8	9	10
ΣX	260	250	250	255	235	260	245	270	255	270
$\overline{X}$= ΣX/5	52	50	50	51	47	50	49	54	51	54
Range	6	7	6	5	6	9	8	7	7	4

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{52 + 50 + 50 + 51 + 47 + 50 + 49 + 54 + 51 + 54}{10} = \mathbf{50.8}$$

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

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{50.8}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 50.8 + (0.577)6.5 = \mathbf{54.55}, \quad A_2 = 0.577 \ from\ table$$

$$LCL = \overline{X} - (A_2)\overline{R} = 50.8 - (0.577)6.5 = \mathbf{47.04}, \quad A_2 = 0.577 \ from\ table$$

Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{6.5}$$

$$UCL = (D_4)\overline{R} = (2.115)(6.5) = \mathbf{13.74} D_4 = 2.115 \, from\ table$$

$$LCL = (D_3)\overline{R} = (0)(6.5) = \mathbf{0}, \qquad D_3 = 0 \, from\ table$$

Control Limit For $\overline{X} - Chart$:

Since sample number 5 falls outside the control limits the statistical process is out of control according to $\overline{X} - Chart$

Control Limit For $\overline{R} - Chart$:

Since all the sample fall within the CL, The Statistical Process Under control for R-chart

5) **The Table given below gives the measurement obtained in 10 samples, Construct the control chart for mean & Range, Discuss the nature of control**

	1	2	3	4	5	6	7	8	9	10
Observed Measurement of X	62	50	67	64	49	63	61	63	48	70
	68	58	70	62	98	75	71	72	79	52
	66	52	68	57	65	62	66	61	53	62
	68	58	56	62	64	58	69	53	61	50
	73	65	61	63	66	68	77	55	49	66
	68	66	66	74	64	55	53	57	56	75

Soln:

Sample	1	2	3	4	5	6	7	8	9	10
ΣX	405	349	388	382	406	381	397	361	346	395
$\overline{X}$= ΣX/6	67.5	58.2	64.7	63.7	67.7	63.5	66.2	60.1	57.7	62.5
Range	11	16	14	17	49	20	24	19	31	25

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{67.5 + 58.2 + 64.7 + 63.7 + 67.7 + 63.5 + 66.2 + 60.1 + 57.7 + 62.5}{10}$$

$$\overline{X} = \mathbf{63.12}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{11 + 16 + 14 + 17 + 49 + 20 + 24 + 19 + 31 + 25}{10} = \mathbf{22.6}$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{63.12}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 63.12 + (0.4829)22.6 = \mathbf{74.03}, \quad A_2 = 0.577 \ from\ table$$

$$LCL = \overline{X} - (A_2)\overline{R} = 63.12 - (0.4829)22.6 = \mathbf{52.20}, \quad A_2 = 0.577 \ from\ table$$

Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{22.6}$$

$$UCL = (D_4)\overline{R} = (2.004)(22.6) = \mathbf{45.29} D_4 = 2.115\ from\ \ table$$

$$LCL = (D_3)\overline{R} = (0)(22.6) = \mathbf{0}, \qquad D_3 = 0\ from\ \ table$$

Control Limit For $\overline{X} - Chart$:

Since all sample falls inside the control limits the statistical process is under control according to $\overline{X} - Chart$

Control Limit For $\overline{R} - Chart$:

Since sample number 5 fall outside the CL, The Statistical Process is out of control according to R-chart

- **Control Charts for Standard Deviation (S.D) [OR] S- Chart:**

The SD is an ideal measure of dispersion a combination of control charts for the sample mean $\overline{X}$ and the sample SD

Control Limit for$\overline{X}$:

$$\boxed{CL = \overline{X}, \qquad UCL = \overline{X} + A_1\left(\sqrt{\frac{n}{n-1}}\right)\overline{S}, \qquad LCL = \overline{X} - A_1\left(\sqrt{\frac{n}{n-1}}\right)\overline{S}}$$

Control Limit for$\overline{S}$:

$$\boxed{CL = \overline{S}, \qquad UCL = B_4\overline{S}\,, \qquad LCL = B_3\overline{S}}$$

❖ Problem

1) **The Following data give the coded measurement of 10 samples each of size 5 from a production process of one hour calculate the sample mean & SD`s and draw the control chart for $\overline{X}$& S**

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Coded Measurements (X)	**9**	**10**	**10**	**8**	**7**	**12**	**9**	**15**	**10**	**16**
	15	**11**	**13**	**13**	**9**	**15**	**9**	**15**	**13**	**14**
	14	**13**	**8**	**11**	**10**	**7**	**9**	**10**	**14**	**12**
	9	**6**	**12**	**10**	**4**	**16**	**13**	**13**	**7**	**14**
	13	**10**	**7**	**13**	**15**	**10**	**5**	**17**	**11**	**14**

Soln:

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Coded Measurements (X)	9	10	10	8	7	12	9	15	10	16
	15	11	13	13	9	15	9	15	13	14
	14	13	8	11	10	7	9	10	14	12
	9	6	12	10	4	16	13	13	7	14
	13	10	7	13	15	10	5	17	11	14
Σx	60	50	50	55	45	60	45	70	55	70
$\boldsymbol{X} = \frac{\Sigma x}{5}$	12	10	10	11	9	12	9	14	11	14

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{12 + 10 + 10 + 11 + 9 + 12 + 9 + 14 + 11 + 14}{10} = \frac{110}{10} = 11$$

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$(x - X)^2$	9	0	0	9	0	0	0	1	1	4
	9	1	9	4	4	9	0	1	4	0
	4	9	4	0	9	25	0	16	9	4
	9	16	4	1	9	16	16	1	16	0
	1	0	9	4	4	4	16	9	0	0
$\Sigma(x - X)^2$	32	26	26	18	26	54	32	28	30	8
$\boldsymbol{S} = \sqrt{\frac{\Sigma(x - X)^2}{n}}$	$\sqrt{\frac{32}{5}}$ =2.5	$\sqrt{\frac{26}{5}}$ =2.3	$\sqrt{\frac{26}{5}}$ =2.3	$\sqrt{\frac{18}{5}}$ =1.9	$\sqrt{\frac{26}{5}}$ =2.3	$\sqrt{\frac{54}{5}}$ =3.3	$\sqrt{\frac{32}{5}}$ =2.5	$\sqrt{\frac{28}{5}}$ =2.4	$\sqrt{\frac{30}{5}}$ =2.4	$\sqrt{\frac{08}{5}}$ =1.3

$$\overline{S} = \frac{2.5 + 2.3 + 2.3 + 1.9 + 2.3 + 3.3 + 2.5 + 2.4 + 2.4 + 1.3}{10} = \frac{23.2}{10} = 2.32$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{11}$$

$$UCL = \overline{X} + (A_1)\overline{S} = 11 + (1.596)\left(\sqrt{\frac{5}{4}}\right)2.32 = \mathbf{15.14}, \quad A_1 = 1.596 \ \mathit{from\ table}$$

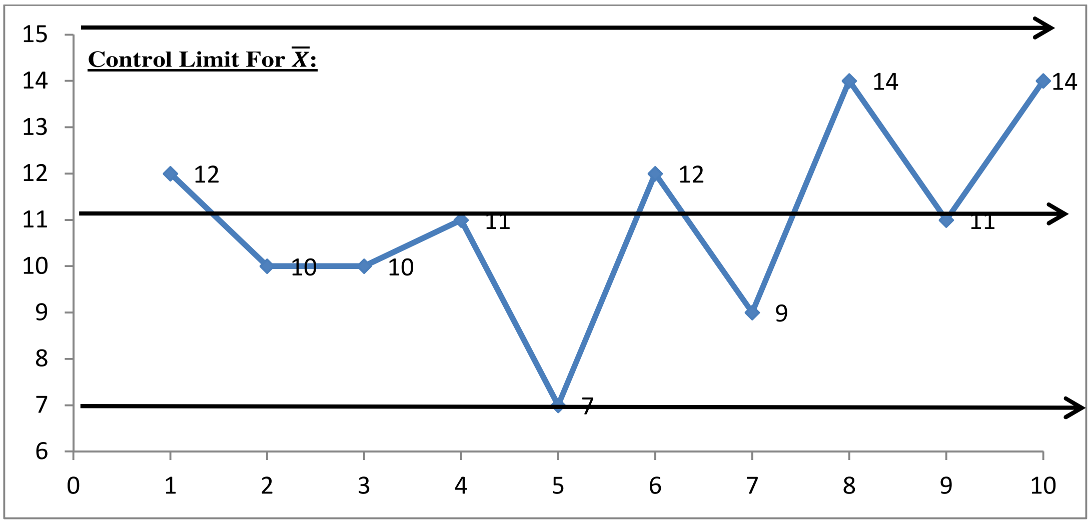
$$LCL = \overline{X} - (A_1)\overline{S} = 11 - (1.596)\left(\sqrt{\frac{5}{4}}\right)2.32 = \mathbf{6.86}, \quad A_1 = 1.596 \ from\,table$$

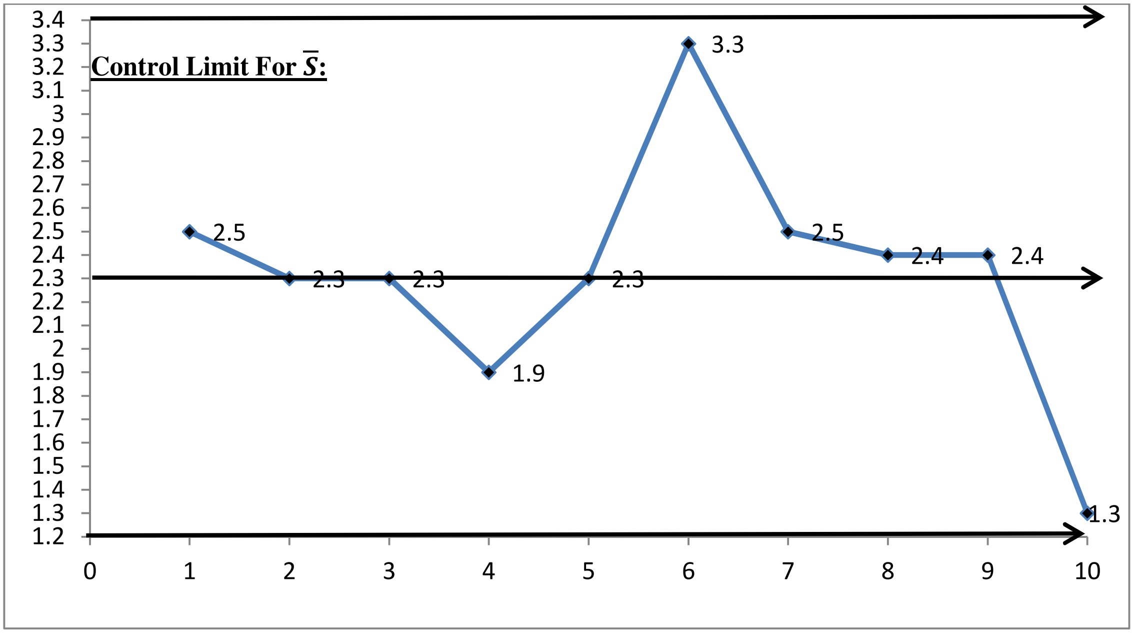
Control Limit For $\overline{S}$:

$$CL = \overline{S} = \mathbf{2.32}$$

$$UCL = (B_4)\overline{S} = (2.004)(2.32) = \mathbf{4.85} \qquad B_4 = 2.089\,from\ table$$

$$LCL = (B_3)\overline{S} = (0)(2.32) = \mathbf{0}, \qquad B_3 = 0 \ \ from\ table$$





Control Limit For $\overline{X} - Chart$:

Since all sample falls inside the control limits the statistical process is under control according to $\overline{X} - Chart$

Control Limit For $\overline{S} - Chart$:

Since all sample falls inside the control limits the statistical process is under control according to $\overline{S} - Chart$

2) **The values of sample mean & SD for 15 samples each of size 4 drawn from a production process are given below, Draw the appropriate control charts for the process avg and process variability comment on the state**

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	13.5	14.5	9.5	12	10.5	11.5
SD	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

Soln:

$$\overline{X} = \frac{\sum x_i}{N} = \frac{185.5}{15} = 12.36, \qquad \overline{S} = \frac{\sum s_i}{N} = \frac{60.3}{15} = 4.02$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{12.36}$$

$$UCL = \overline{X} + (A_1)\overline{S} = 12.36 + (1.880)\left(\sqrt{\frac{4}{3}}\right)4.02 = \mathbf{21.09}, \quad A_1 = 1.880 \ \mathit{from\ table}$$

$$LCL = \overline{X} - (A_1)\overline{S} = 12.36 - (1.880)\left(\sqrt{\frac{4}{3}}\right)4.02 = \mathbf{3.63}, \quad A_1 = 1.880 \ \mathit{from\ table}$$

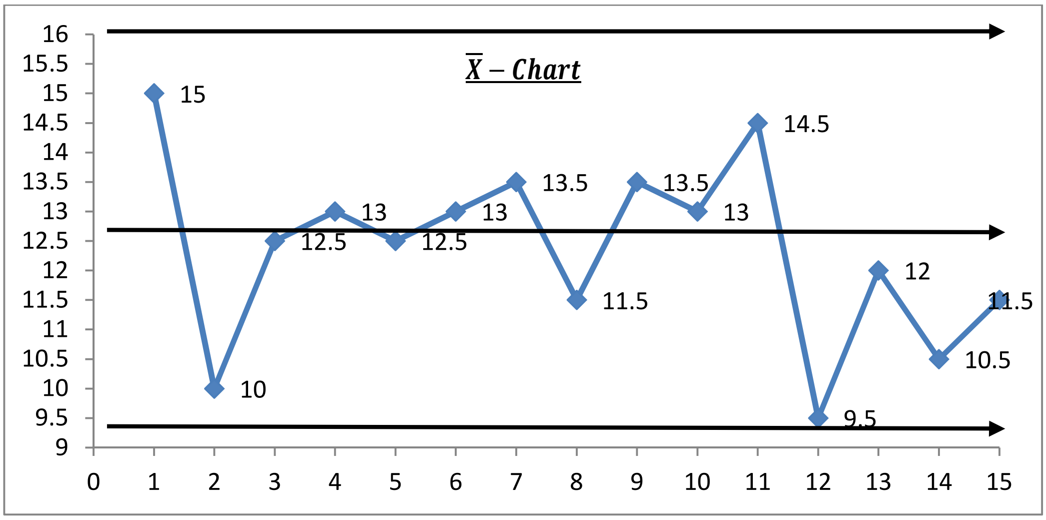
Control Limit For $\overline{S}$:

$$CL = \overline{S} = \mathbf{4.02}$$

$$UCL = (B_4)\overline{S} = (2.266)(4.02) = \mathbf{9.11} \qquad B_4 = 2.266\ \mathit{from\ table}$$

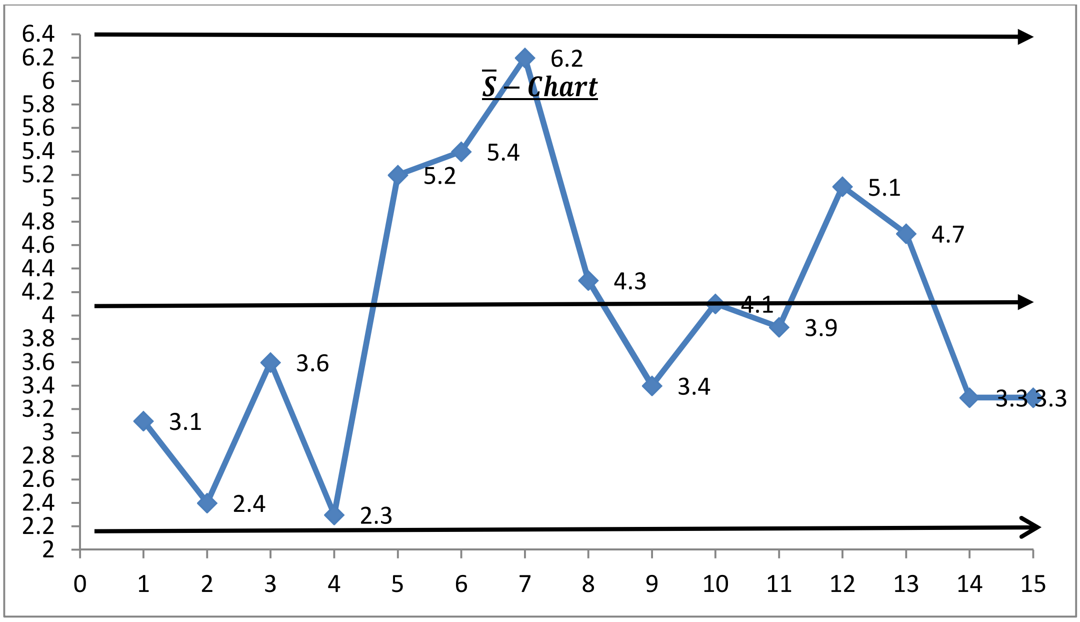
$$LCL = (B_3)\overline{S} = (0)(4.02) = \mathbf{0}, \qquad B_3 = 0 \ \mathit{from\ table}$$

Control Limit For $\overline{X} - Chart$:



Since all sample falls inside the control limits the statistical process is under control according to $\overline{X} - Chart$

Control Limit For $\overline{S} - Chart$:



Since all sample falls inside the control limits the statistical process is under control according to $\overline{S} - Chart$

✓ Control Charts for Attributes

To control the quality of certain products whose attribute are available, the following control charts are used

- np-Chart (Number of Defectives)
- p-chart (For proportion of Defectives)
- c-Chart (For the number of Defectives in a Unit)

np-chart

$$CL = n\overline{p}, \quad UCL = n\left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right)\right], \quad LCL = n\left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right)\right]$$

p-chart

$$CL = \overline{p}, \quad UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right)\right], \quad LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right)\right]$$

C-chart

$$CL = \overline{c}, \quad UCL = \left[\overline{c} + 3\left(\sqrt{\overline{c}}\right)\right], \quad LCL = \left[\overline{c} - 3\left(\sqrt{\overline{c}}\right)\right]$$

❖ c-chart

1) **20 pieces of cloth out of different rolls contained respectively 1,4,3,2,4,5,6,7,3,3,2,5,7, 6,4,5,2,1,3&8 imperfection, As certain whether the process is in state of statistical control**

Soln:

Let c denotes the number of imperfection per unit

$$\overline{c} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected} = \frac{\sum c}{n}$$

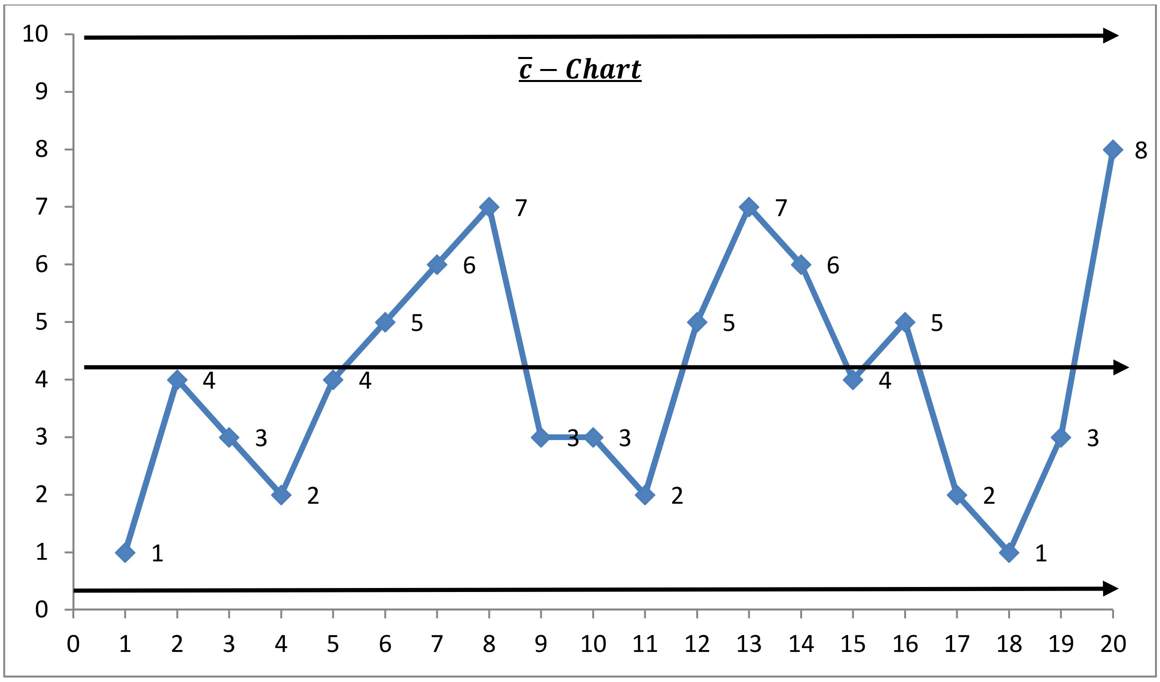
$$\overline{c} = \frac{1+4+3+2+4+5+6+7+2+3+2+5+7+6+4+5+2+1+3+8}{20} = \frac{80}{20} = \mathbf{4}$$

$$\sqrt{\overline{c}} = \sqrt{4} = \mathbf{2}$$

$$UCL = \overline{c} + 3\sqrt{\overline{c}} = 4 + 3(2) = \mathbf{10}, \quad LCL = \overline{c} - 3\sqrt{\overline{c}} = 4 - 3(2) = -\mathbf{2}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case

Control Limit For $\overline{c} - Chart$



Conclusion:

Since all the values of c lies between LCL & UCL so the process is under control

2) **15 Tape Recorder were examined for quality control test, The number of defect in each tape recorder is recorded below, Draw the appropriate control charts and comment on the state of control**

Sample No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
No of Defective	**2**	**4**	**3**	**1**	**1**	**2**	**5**	**3**	**6**	**7**	**3**	**1**	**4**	**2**	**1**

Soln:

Let c denotes the number of imperfection per unit

$$\overline{c} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected} = \frac{\sum c}{n}$$

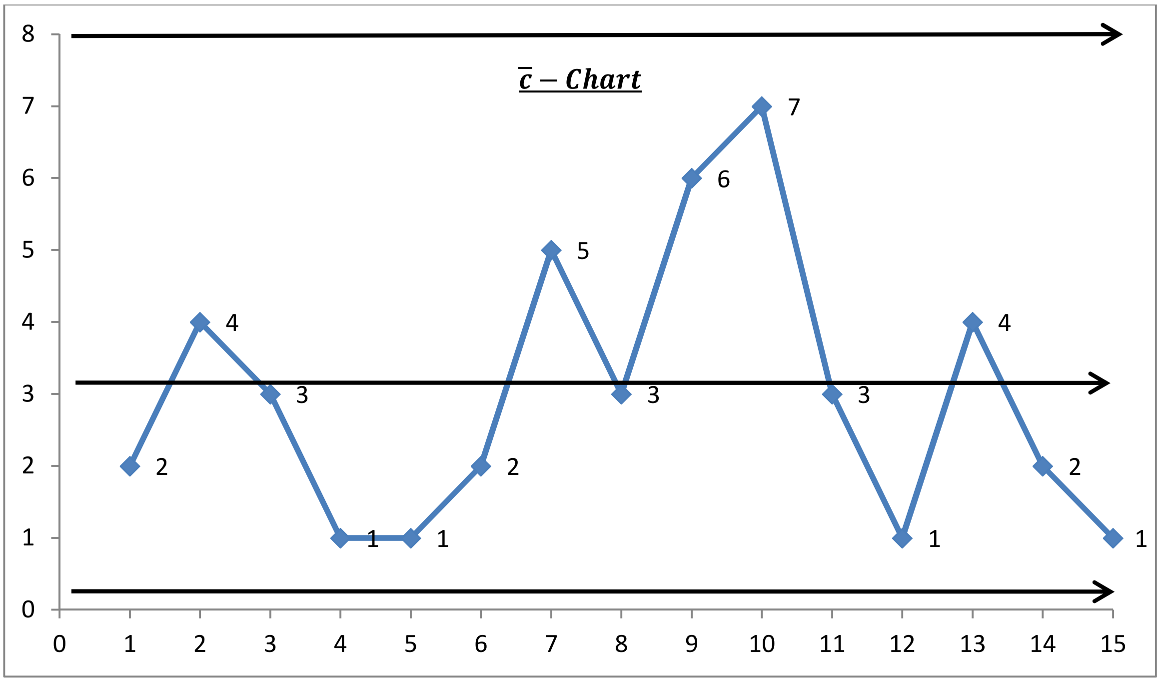
$$\overline{c} = \frac{2+4+3+1+1+2+5+3+6+7+3+1+4+2+1}{15} = \frac{45}{15} = \mathbf{3}$$

$$\sqrt{\overline{c}} = \sqrt{3} = \mathbf{1.732}$$

$$UCL = \overline{c} + 3\sqrt{\overline{c}} = 3 + 3(1.732) = \mathbf{8.20}, \quad LCL = \overline{c} - 3\sqrt{\overline{c}} = 3 - 3(1.732) = -\mathbf{2.20}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case

Control Limit For $\overline{c} - Chart$



Conclusion:

Since all the values of c lies between LCL & UCL so the process is under control

3) **Construct the control chart for the number of defects from the following data given the number of defects in 15 pieces of cloth of equal length when inspected in a textile mill and find the nature of the process, Number of defects are 3,4,2,7,9,6,5,4,8,10,5,8,7,7,5**

Soln:

Let c denotes the number of imperfection per unit

$$\bar{c} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected} = \frac{\sum c}{n}$$

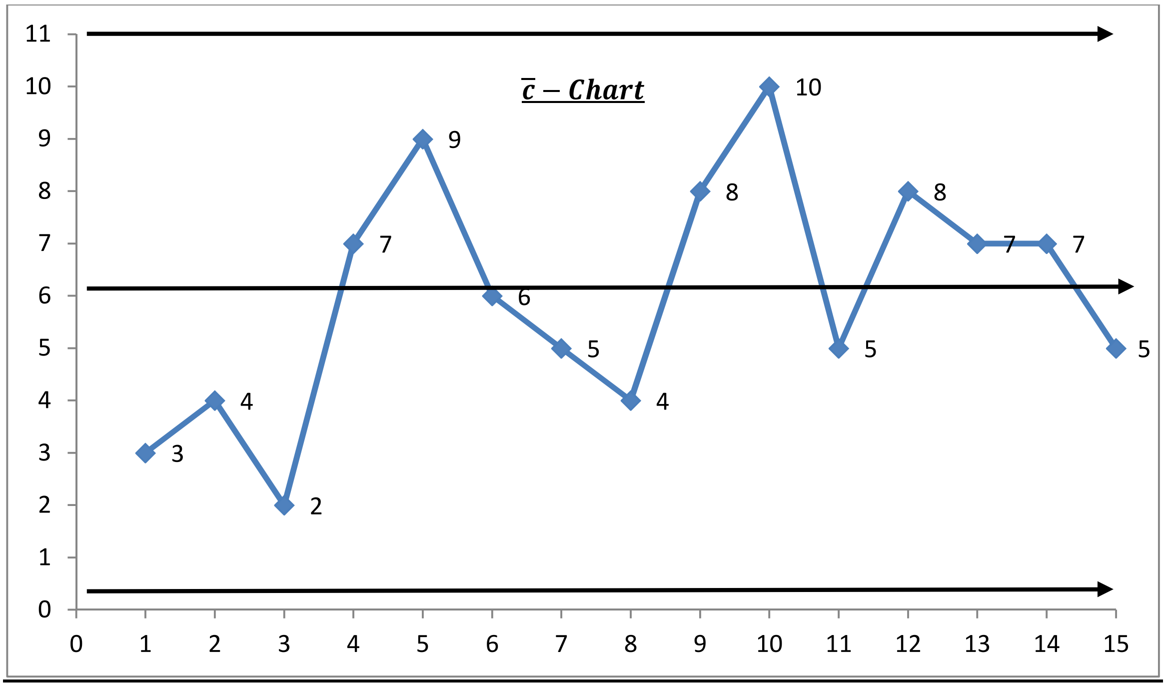
$$\bar{c} = \frac{3+4+2+7+9+6+5+4+8+10+5+8+7+7+5}{15} = \frac{90}{15} = \mathbf{6}$$

$$\sqrt{\bar{c}} = \sqrt{6} = \mathbf{2.4494}$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 6 + 3(2.4494) = \mathbf{13.35}, \quad LCL = \bar{c} - 3\sqrt{\bar{c}} = 6 - 3(2.4494) = \mathbf{-1}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case

Control Limit For $\bar{c} - Chart$



Conclusion:

Since all the values of c lies between LCL & UCL so the process is under control

❖ p-chart &np-chart

1) Construct a control charts for defective for the following data

Sample No	1	2	3	4	5	6	7	8	9	10
No of Inspected	90	65	85	70	80	80	70	95	90	75
No of Defective	9	7	3	2	9	5	3	9	6	7

Soln:

We note that the size of the sample varies from sample to sample, so we construct the p-chart provided

$$\mathbf{0.75}\,\overline{\boldsymbol{n}} < \boldsymbol{n_i} < 1.25\,\overline{\boldsymbol{n}}$$

$$\overline{p} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected}$$

$$\overline{p} = \frac{9+7+3+2+9+5+3+9+6+7}{90+65+85+70+80+80+70+95+90+75} = \frac{60}{800} = 0.075$$

$$\overline{n} = \frac{90+65+85+70+80+80+70+95+90+75}{10} = \frac{800}{10} = 80$$

$$\mathbf{0.75}\ (\mathbf{80}) < \boldsymbol{n_i} < 1.25\ (\mathbf{80}) \Rightarrow \mathbf{60} < \boldsymbol{n_i} < 100$$

$$CL = \overline{p} = \mathbf{0.075}$$

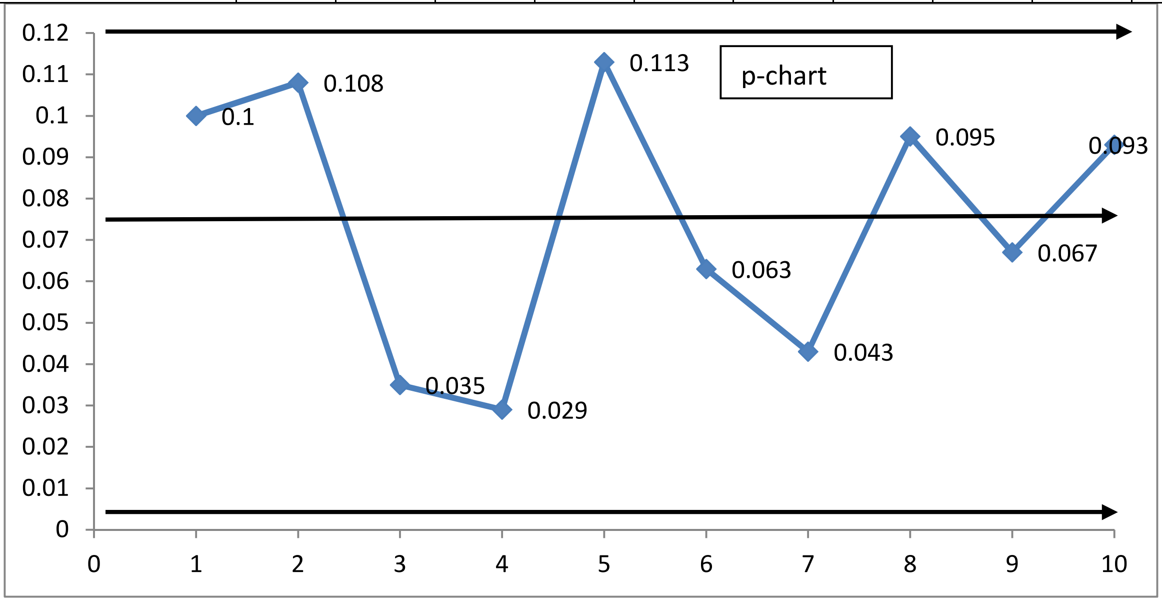
$$UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.075 + 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right) = \mathbf{0.16332}$$

$$LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.075 - 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right) = -\mathbf{0.01332}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case

P-Ratio

Sample No	1	2	3	4	5	6	7	8	9	10
No of Inspected	90	65	85	70	80	80	70	95	90	75
No of Defective	9	7	3	2	9	5	3	9	6	7
P-Ratio	9/90= 0.1	7/65= 0.108	3/85= 0.035	2/70= 0.029	9/80= 0.113	5/80= 0.063	3/70= 0.043	9/95= 0.095	6/90= 0.067	7/75= 0.093



Conclusion: Since all the values of p-chart lies between LCL & UCL so the process is under control

np-chart:

$$CL = (\overline{n})\overline{p} = 80 * 0.075 = \mathbf{6}$$

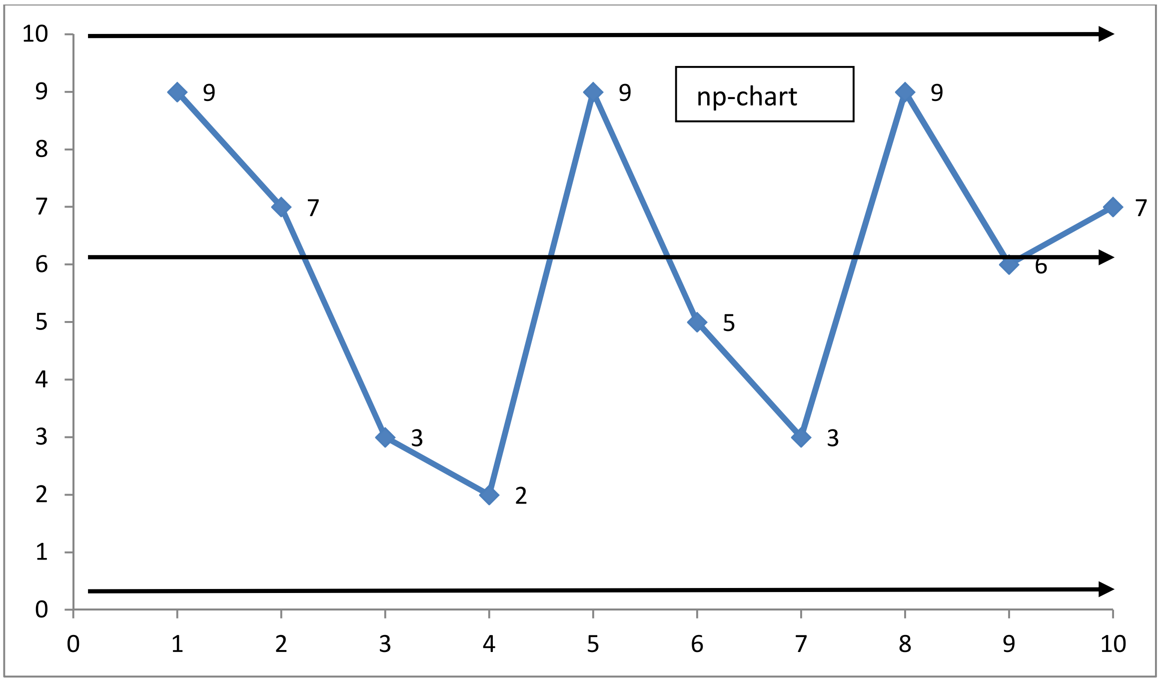
$$UCL = \overline{n}\left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 80\left[0.075 + 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right)\right] = 80 * 0.16332$$

$$\boldsymbol{UCL = 13.07}$$

$$LCL = \overline{n}\left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 80\left[0.075 - 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right)\right] = 80 * (-0.01332)$$

$$\boldsymbol{LCL = -1.066}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case



Conclusion: Since all the values of np-chart lies between LCL & UCL so the process is under control

2) **The following are the figures for the number of defective of 10 samples, each containing 100 items, 8,10,9,8,10,11,7,9,6,12, Draw the control chart for fraction defective and comments on the state of control of the process**

Soln:

Given sample size all samples are equal n=100 (given)

$$\overline{p} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected}$$

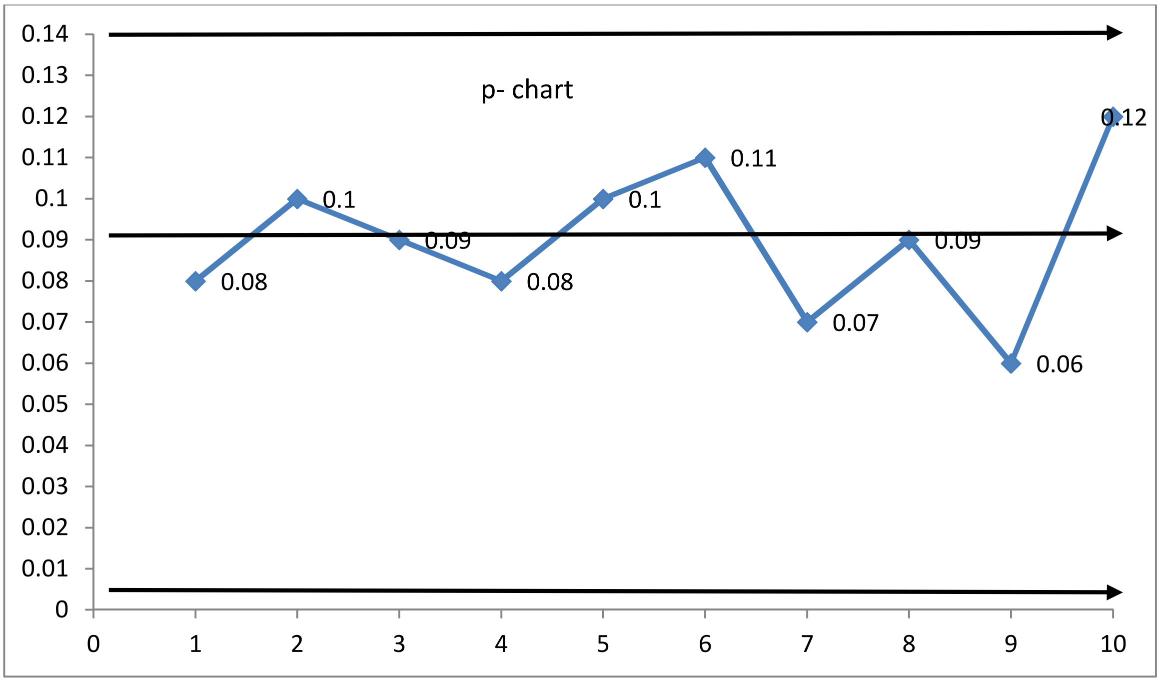
Sample No	1	2	3	4	5	6	7	8	9	10
No of Inspected	100	100	100	100	100	100	100	100	100	100
No of Defective	8	10	9	8	10	11	7	9	6	12
P-Ratio	8/100= 0.08	10/100= 0.1	9/100= 0.09	8/100= 0.08	10/100= 0.1	11/100= 0.11	7/100= 0.07	9/100= 0.09	6/100= 0.06	12/100= 0.12

$$\overline{p} = \frac{0.08 + 0.1 + 0.09 + 0.08 + 0.1 + 0.11 + 0.07 + 0.09 + 0.06 + 0.12}{10} = \mathbf{0.09}$$

$$CL = \overline{p} = \mathbf{0.09}$$

$$UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.09 + 3\left(\sqrt{\frac{0.09(1-0.09)}{100}}\right) = \mathbf{0.177}$$

$$LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.09 - 3\left(\sqrt{\frac{0.09(1-0.09)}{100}}\right) = \mathbf{0.004}$$



Conclusion: Since all the values of p-chart lies between LCL & UCL so the process is under control

np-chart:

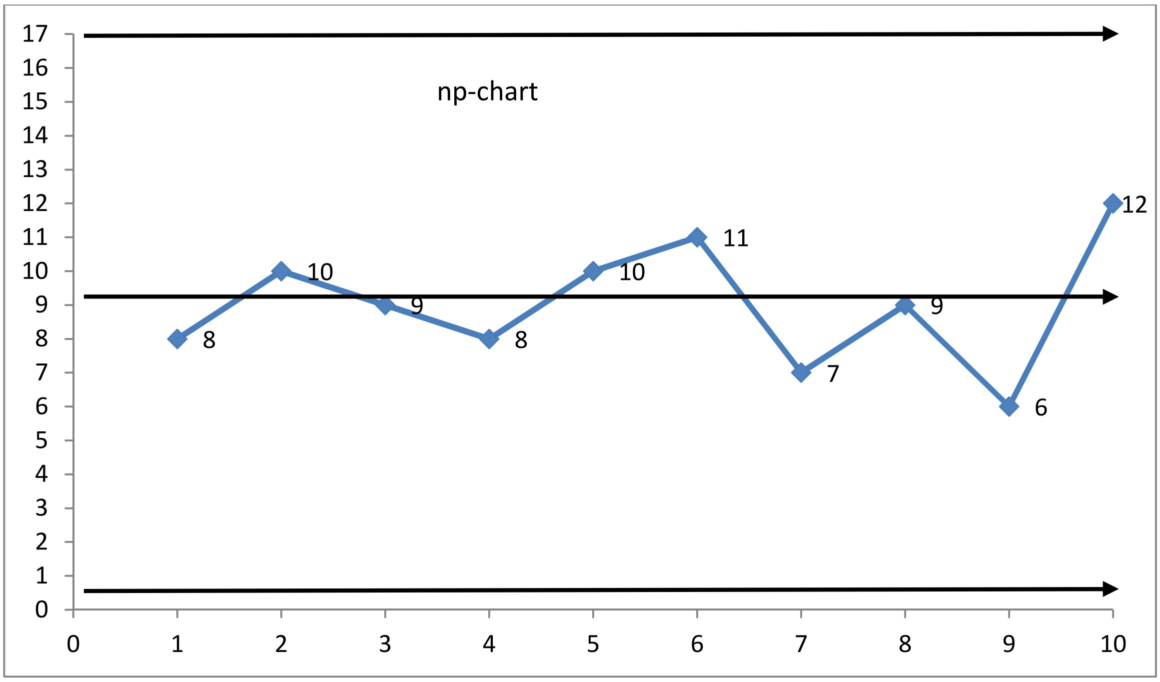
$$\boldsymbol{CL} = n\overline{p} = 100 * 0.09 = \mathbf{9}$$

$$UCL = n\left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 100\left[0.09 + 3\left(\sqrt{\frac{0.09(1-0.09)}{100}}\right)\right] = 100 * 0.177$$

$$\boldsymbol{UCL = 17.7}$$

$$LCL = n\left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 100\left[0.09 - 3\left(\sqrt{\frac{0.09(1-0.09)}{100}}\right)\right] = 100 * 0.004$$

$$\boldsymbol{LCL = 0.4 \cong 0}$$



Conclusion: Since all the values of np-chart lies between LCL & UCL so the process is under control

3) The data given below are the number of defective in 10 samples of 100 items each, construct a p-chart &np-chart and comments on the result

Sample No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of Defectives	**6**	**16**	**7**	**3**	**8**	**12**	**7**	**11**	**11**	**4**

Soln:

Given sample size all samples are equal n=100 (given)

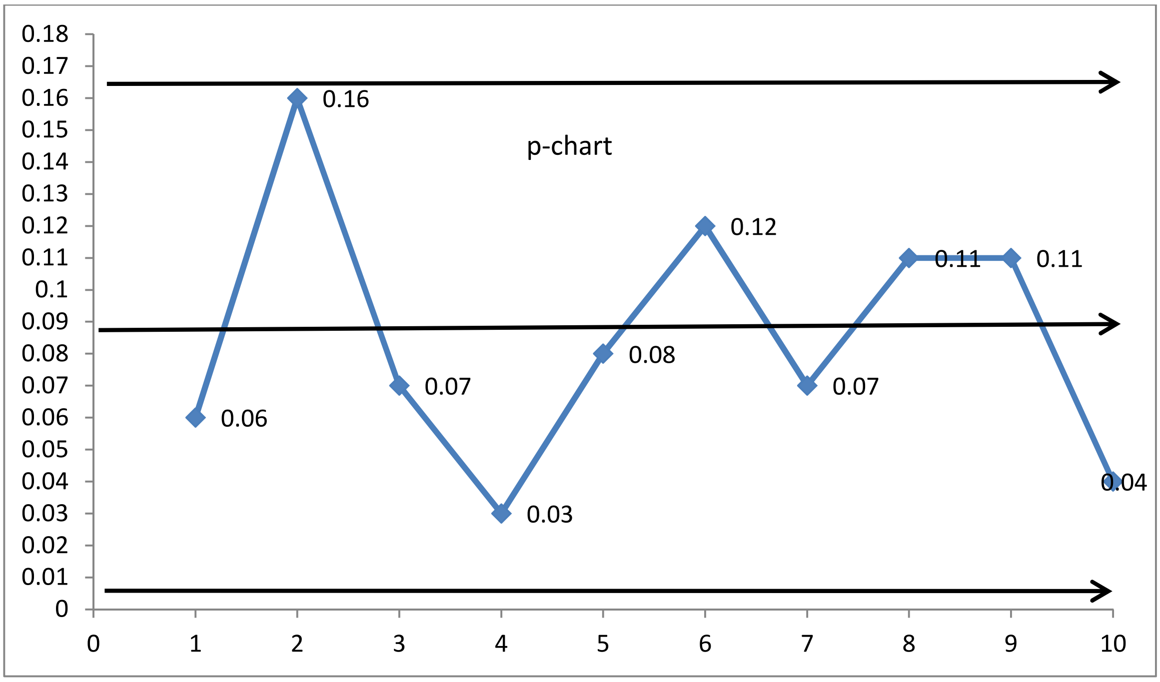
Sample No:	1	2	3	4	5	6	7	8	9	10
Number of Defectives	6	16	7	3	8	12	7	11	11	4
P-Ratio	0.06	0.16	0.07	0.03	0.08	0.12	0.07	0.11	0.11	0.04

$$\overline{p} = \frac{0.06 + 0.16 + 0.07 + 0.03 + 0.08 + 0.12 + 0.07 + 0.11 + 0.11 + 0.04}{10} = \mathbf{0.085}$$

$$CL = \overline{p} = \mathbf{0.085}$$

$$UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.085 + 3\left(\sqrt{\frac{0.085(1-0.085)}{100}}\right) = \mathbf{0.1687}$$

$$LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.082 - 3\left(\sqrt{\frac{0.085(1-0.085)}{100}}\right) = \mathbf{0.0013}$$



<u>Conclusion:</u> Since all the values of p-chart lies between LCL & UCL so the process is under control

<u>np-chart:</u>

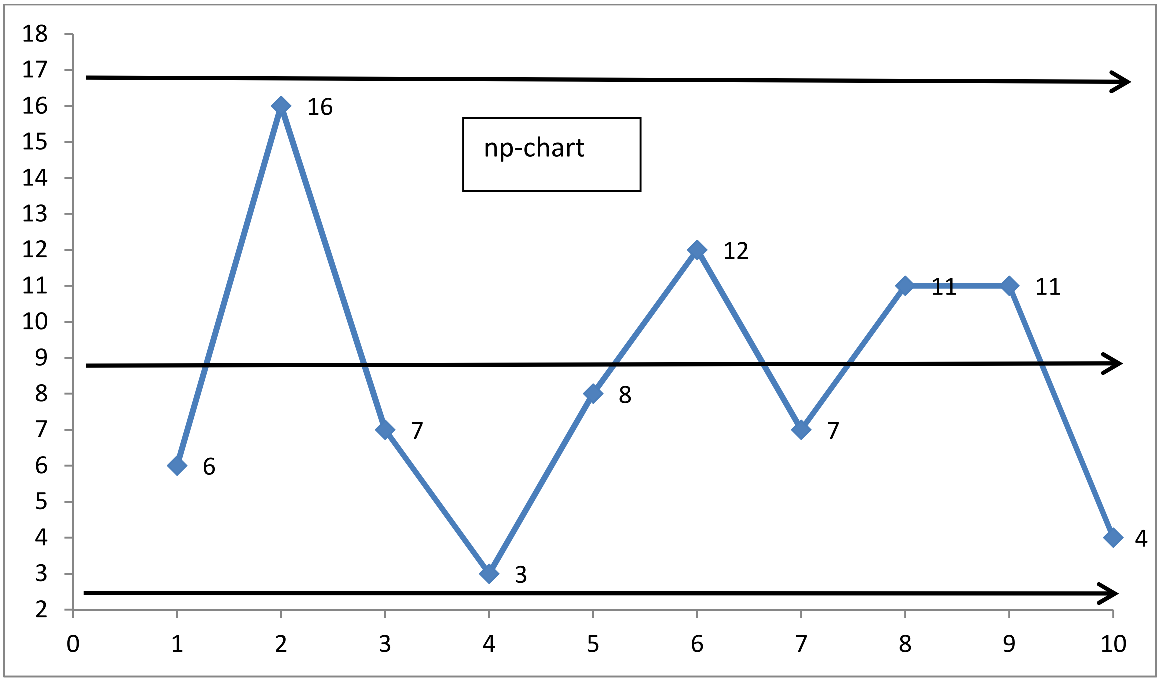
$$\boldsymbol{CL} = n\overline{p} = 100 * 0.085 = \mathbf{8.5}$$

$$UCL = n\left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 100\left[0.085 + 3\left(\sqrt{\frac{0.085(1-0.085)}{100}}\right)\right] = 100 * 0.1687$$

$$\boldsymbol{UCL} = \mathbf{16.87}$$

$$LCL = n\left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 100\left[0.085 - 3\left(\sqrt{\frac{0.085(1-0.085)}{100}}\right)\right] = 100 * 0.0013$$

$$\boldsymbol{LCL = 0.13 \cong\ 0}$$



Conclusion: Since all the values of np-chart lies between LCL & UCL so the process is under control

4) Construct a Control Chart for defectives for the following data **(Apr/May-2015)**

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

Soln: We note that the size of the sample varies from sample to sample, so we construct the p-chart provided

$$\boldsymbol{0.75\,\overline{n} < n_i <} 1.25\,\boldsymbol{\overline{n}}$$

$$\overline{p} = \frac{Total\ number\ of\ defects}{Total\ sample\ Inspected}$$

$$\overline{p} = \frac{9+7+3+2+9+5+3+9+6+7}{90+65+85+70+80+80+70+95+90+75} = \frac{60}{800} = 0.075$$

$$\overline{n} = \frac{90+65+85+70+80+80+70+95+90+75}{10} = \frac{800}{10} = 80$$

$$\boldsymbol{0.75\,(80) < n_i <} 1.25\,\boldsymbol{(80) \Rightarrow 60 < n_i <} 100$$

$$CL = \overline{p} = \boldsymbol{0.075}$$

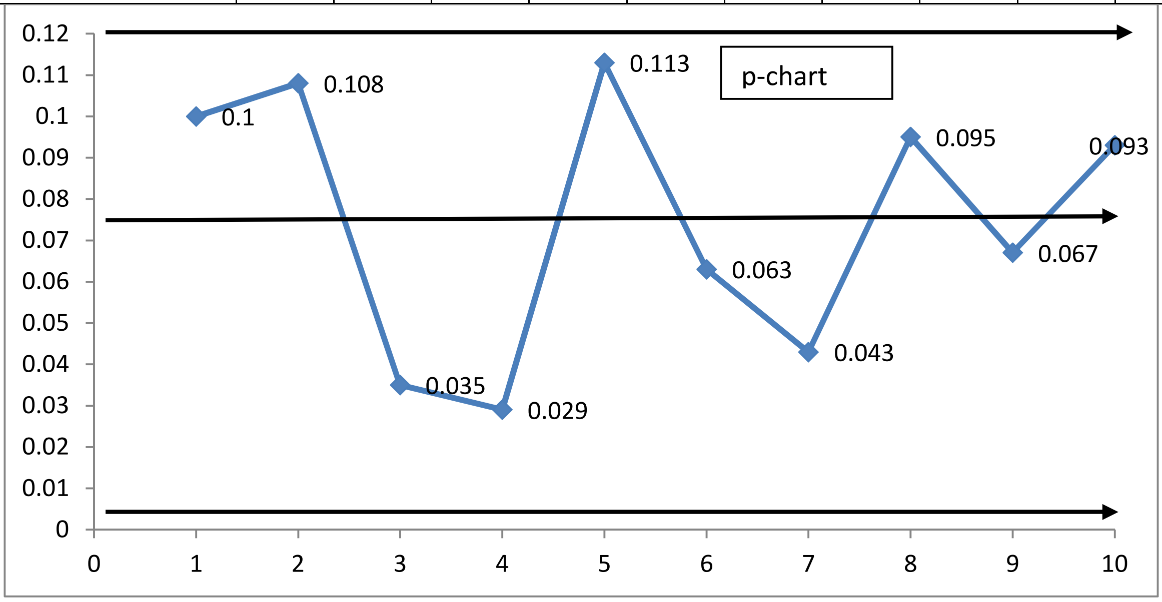
$$UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.075 + 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right) = \boldsymbol{0.16332}$$

$$LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.075 - 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right) = -\mathbf{0.01332}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case

P-Ratio

Sample No	1	2	3	4	5	6	7	8	9	10
No of Inspected	90	65	85	70	80	80	70	95	90	75
No of Defective	9	7	3	2	9	5	3	9	6	7
P-Ratio	9/90= 0.1	7/65= 0.108	3/85= 0.035	2/70= 0.029	9/80= 0.113	5/80= 0.063	3/70= 0.043	9/95= 0.095	6/90= 0.067	7/75= 0.093



np-chart:

$$CL = (\overline{n})\overline{p} = 80 * 0.075 = \mathbf{6}$$

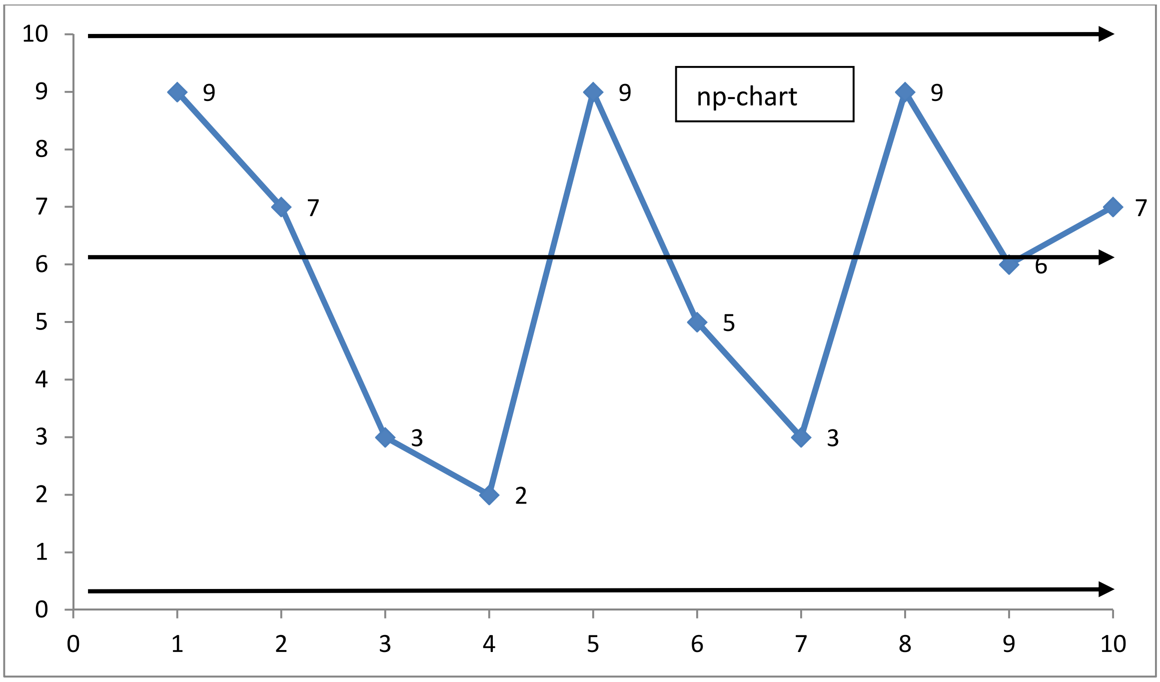
$$UCL = \overline{n}\left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 80\left[0.075 + 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right)\right] = 80 * 0.16332$$

$$\boldsymbol{UCL = 13.07}$$

$$LCL = \overline{n}\left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 80\left[0.075 - 3\left(\sqrt{\frac{0.075(1-0.075)}{80}}\right)\right] = 80 * (-0.01332)$$

$$\boldsymbol{LCL = -1.066}$$

Since LCL is negative, We take LCL=0 as negative of imperfection is not valid in practical case



Conclusion: Since all the values of p-chart lies between LCL & UCL so the process is under control

Since all the values of np-chart lies between LCL & UCL so the process is under control

---

5) Construct R chart for the following data comment on the state of control (Apr/May-2015)

Sample No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Observations	**1.7**	**0.8**	**1.0**	**0.4**	**1.4**	**1.8**	**1.6**	**2.5**	**2.9**
	2.2	**1.5**	**1.4**	**0.6**	**2.3**	**2.0**	**1.0**	**1.6**	**2.0**
	1.9	**2.1**	**1.0**	**0.7**	**2.8**	**1.1**	**1.5**	**1.8**	**0.5**
	1.2	**0.9**	**1.3**	**0.2**	**2.7**	**0.1**	**2.0**	**1.2**	**2.2**

Soln:

Sample	1	2	3	4	5	6	7	8	9
Range	1	1.3	0.4	0.5	1.4	1.9	1	1.3	2.4

N = Number of Samples = 9

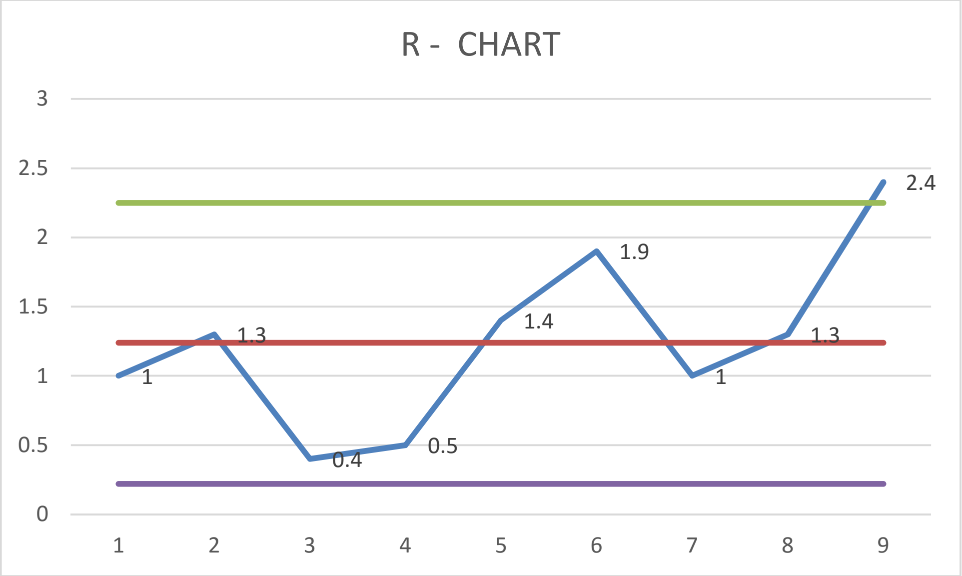
$$\overline{R} = \frac{\sum R_i}{N} = \frac{1 + 1.3 + 0.4 + 0.5 + 1.4 + 1.9 + 1 + 1.3 + 2.4}{9} = \mathbf{1.244}$$

Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{1.244}$$

$$UCL = (D_4)\overline{R} = (1.816)(1.244) = \mathbf{2.25} \qquad D_4 = 1.816\, from\ table$$

$$LCL = (D_3)\overline{R} = (0.184)(1.244) = \mathbf{0.22}, \qquad D_3 = 0.184\, from\ table$$



Control Limit For $\overline{R} - Chart$:

Since sample number 9 fall outside the CL, The Statistical Process is out of control according to R-chart

----

6) Define (a) Acceptance Quality Level (AQL) (Apr/May-2015)

Soln: Acceptable quality level is the poorest level of quality from a supplier's process that would be considered acceptable as a process average. You want to design a sampling plan that accepts a particular lot of product at the AQL frequently. (OR) An acceptable quality level is a test and/or inspection standard that prescribes the range of the number of defective components that is considered acceptable when random sampling those components during an inspection. The defects found during an electronic or electrical test, or during a physical (mechanical) inspection, are sometimes classified into three levels: critical, major and minor. Critical defects are those that render the product unsafe or hazardous for the end user or that contravene mandatory regulations. Major defects can result in the product's failure, reducing its marketability, usability or saleability. Lastly, minor defects do not affect the product's marketability or usability, but represent workmanship defects that make the product fall short of defined quality standards. Different companies maintain different interpretations of each defect type. In order to avoid argument, buyers and sellers agree on an AQL standard, chosen according to the level of risk each party assumes, which they use as a reference during pre-shipment inspection.

For example, you receive a shipment of microchips and your acceptable quality level (AQL) is 1.5%. Realizing that you won't always make the correct decision (sampling risk), you set the producer's risk (alpha) at 0.05. This means that approximately 95% of the time you will correctly accept a lot with a quality level of 1.5% or better and 5% of the time you will incorrectly reject the lot with a quality level of 1.5% or better.

The AQL describes what the sampling plan will accept, whereas the reject able quality level (RQL) describes what the sampling plan will reject.

(b) Lot Tolerance Proportion Defective (LTPD)

Soln: The LTPD of a sampling plan is a level of quality routinely rejected by the sampling plan. It is generally defined as that level of quality (percent defective, defects per hundred units, etc.) that the sampling plan will accept 10% of the time. This means lots at or worse than the LTPD are accepted at most 10% of the time. In other words, they are rejected at least 90% of the time. The LTPD can be determined using the OC curve by finding that quality level on the bottom axis that corresponds to a probability of acceptance of 0.10 (10%) on the left axis.

Associated with the LTPD is a confidence statement one can make. If the lot passes the sampling plan, one can state with 90% confidence that the quality level (defective rate, etc.) is below the LTPD. In other words, passing the sampling plan demonstrates that the LTPD has been meet.

The probability of acceptance at the LTPD can be reset using the *Definitions of AQL and LTPD* dialog box. The associated confidence statements are also displayed in this dialog box.

The LTPD is used to help describe the protection provided a sampling plan. But it only provides half the answer. It describes what the sampling plan will reject. We would also like to know what the sampling plan will accept. The answer to this second question is provided by the AQL.

Equivalent terms commonly used instead of LTPD are: (i) LQ (ii) RQL (iii) UQL

(c) Single sampling plan

Two numbers specify a single sampling plan: They are the number of items to be sampled (n) and a prespecified acceptable number of defects (c). If there are fewer or equal defects in the lot than the acceptance number, c, and then the whole batch will be accepted. If there are more than c defects, the whole lot will be rejected or subjected to 100% screening.

---

7) A Machine fills boxes with dry cereal. 15 sample of 4 boxes are drawn randomly. The weights of the sampled boxes are shown as follows. Draw the control charts for the sample mean and sample range and determine whether the process is in a state of control (Apr/May-2015)

Sample No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Weight of Boxes (X):	**10.0**	**10.3**	**11.5**	**11.0**	**11.3**	**10.7**	**11.3**	**12.3**	**11.0**	**11.3**	**12.5**	**11.9**	**12.1**	**11.9**	**10.6**
	10.2	**10.9**	**10.7**	**11.1**	**11.6**	**11.4**	**11.4**	**12.1**	**13.1**	**12.1**	**11.9**	**12.1**	**11.1**	**12.1**	**11.9**
	11.3	**10.7**	**11.4**	**10.7**	**11.9**	**10.7**	**11.1**	**12.7**	**13.1**	**10.7**	**11.8**	**11.6**	**12.1**	**13.1**	**11.7**
	12.4	**11.7**	**12.4**	**11.4**	**12.1**	**11.0**	**10.3**	**10.7**	**12.4**	**11.5**	**11.3**	**11.4**	**11.7**	**12.0**	**12.1**

Soln:

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ΣX	43.9	43.6	46	44.2	46.9	43.8	44.1	47.8	49.6	45.6	47.5	47	47	49.1	46.3
$\overline{X}$= ΣX/4	10.975	10.9	11.5	11.05	11.725	10.95	11.025	11.95	12.4	11.4	11.875	11.75	11.75	12.275	11.575
Range	2.4	1.4	1.7	0.7	0.8	0.7	1.1	2	2.1	1.4	1.2	0.7	1	1.2	1.5

N = Number of Samples = 15

$$\overline{X} = \frac{\sum x_i}{N} = \frac{10.97 + 10.9 + 11.5 + 11.05 + \cdots + 11.75 + 11.75 + 12.275 + 11.575}{15} = \mathbf{11.54}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{2.4 + 1.4 + 1.7 + 0.7 + \cdots + 0.7 + 1 + 1.2 + 1.5}{15} = \mathbf{1.32}$$

<u>Control Limit For $\overline{X}$:</u>

$$CL = \overline{X} = \mathbf{11.54}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 11.54 + (0.223)1.32 = \mathbf{11.83}, \quad A_2 = 0.223 \ from\ table$$

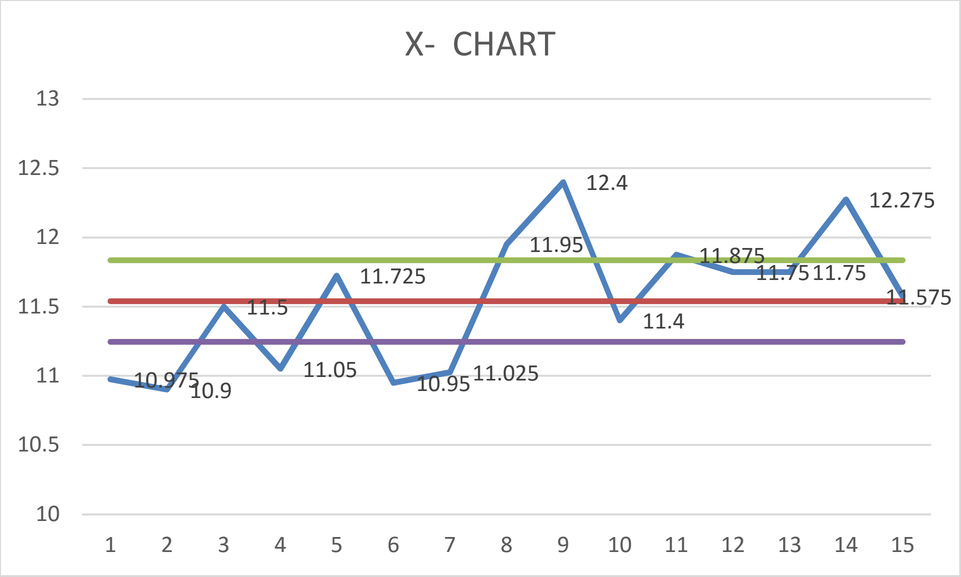
$$LCL = \overline{X} - (A_2)\overline{R} = 11.54 - (0.223)1.32 = \mathbf{11.24}, \quad A_2 = 0.223 \ from\ table$$

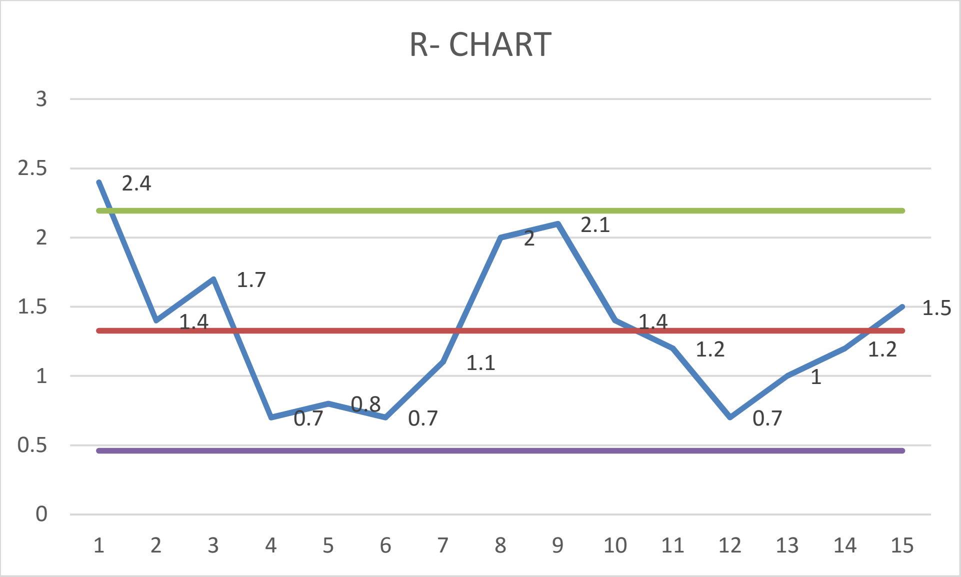
<u>Control Limit For $\overline{R}$:</u>

$$CL = \overline{R} = \mathbf{1.32}$$

$$UCL = (D_4)\overline{R} = (1.653)(1.32) = \mathbf{2.19} \qquad D_4 = 1.653\, from\ \ table$$

$$LCL = (D_3)\overline{R} = (0.347)(1.32) = \mathbf{0.46}, \qquad D_3 = 0.347\ from\ table$$





<u>Control Limit For $\overline{X} - Chart$:</u>

Since sample number 1,2,4,6,7,9,11&14 fall outside the CL, The Statistical Process is out of control according to $\overline{X} - Chart$

<u>Control Limit For $\overline{R} - Chart$:</u>

Since sample number 1 fall outside the CL, The Statistical Process is out of control according to R-chart

----

8) Ten inspection sample lots of five amplifiers each are drawn from production the following table lists the average life and range of the power output obtained for each amplifier construct $\overline{X}$-Chart and R Chart comment on the state of control

(Nov/Dec-2015)

Sample No	1	2	3	4	5	6	7	8	9	10
$\overline{X}$	11	12	12.8	14	13.6	12.8	11.8	12.9	13.0	11.8
R	4	4	6	4	6	5	5	6	4	6

Soln:

N = Number of Samples = 10

$$\overline{X} = \frac{\sum x_i}{N} = \frac{11 + 12 + 12.8 + 14 + 13.6 + 12.8 + 11.8 + 12.9 + 13 + 11.8}{10}$$

$$\overline{\boldsymbol{X}} = \mathbf{12.57}$$

$$\overline{R} = \frac{\sum R_i}{N} = \frac{4+4+6+4+6+5+5+6+4+6}{10} = \mathbf{5}$$

Control Limit For $\overline{X}$:

$$CL = \overline{X} = \mathbf{5}$$

$$UCL = \overline{X} + (A_2)\overline{R} = 12.57 + (0.308)5 = \mathbf{14.11}, \quad A_2 = 0.308 \ from\, table$$

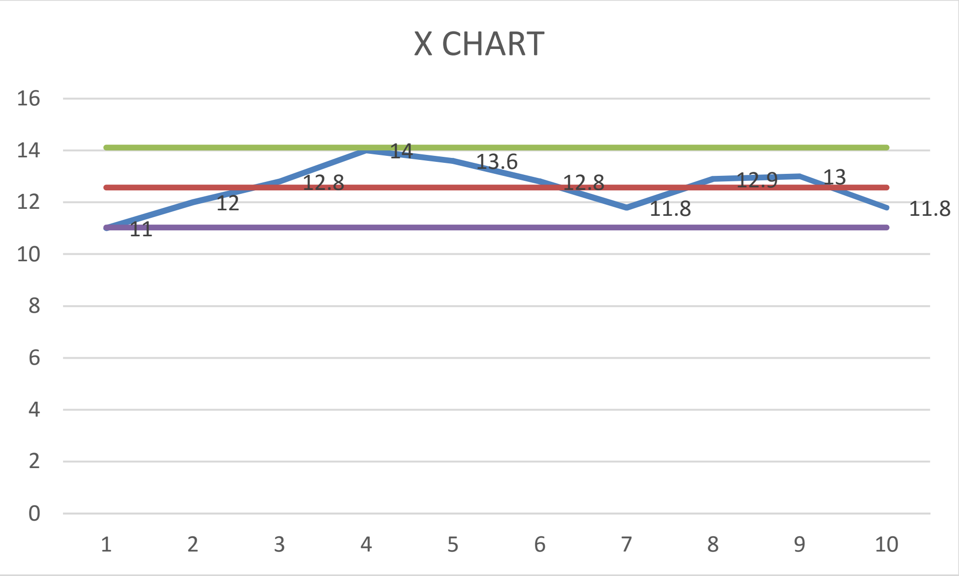
$$LCL = \overline{X} - (A_2)\overline{R} = 12.57 - (0.308)5 = \mathbf{11.03}, \quad A_2 = 0.308 \ from\, table$$

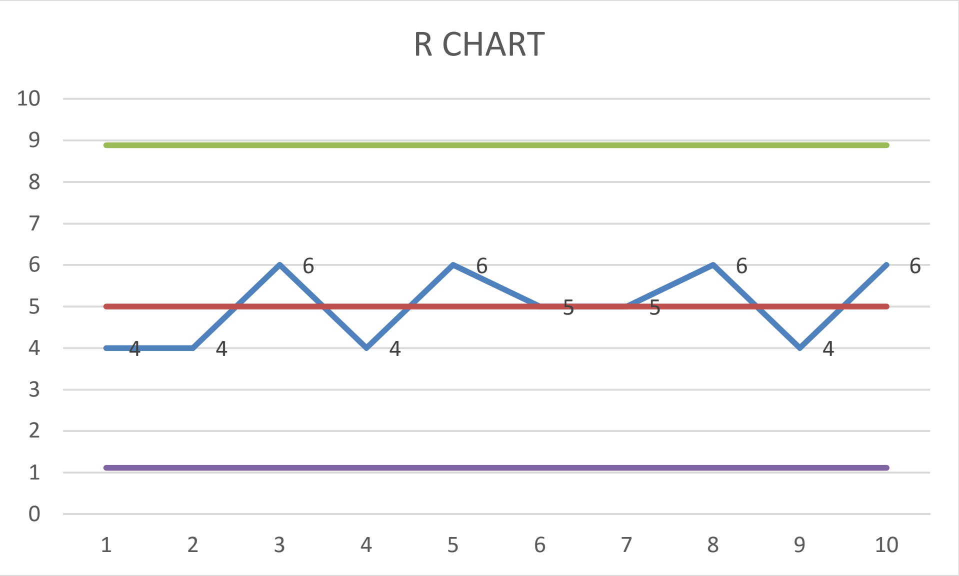
Control Limit For $\overline{R}$:

$$CL = \overline{R} = \mathbf{5}$$

$$UCL = (D_4)\overline{R} = (0.223)(2.59) = \mathbf{8.88}, \qquad D_4 = 0.223\, from\ table$$

$$LCL = (D_3)\overline{R} = (1.777)(2.59) = \mathbf{1.11}, \qquad D_3 = 1.777\, from\ table$$





Control Limit For$\overline{X} - Chart$:

Since all sample falls inside the control limits the statistical process is under control according to $\overline{X} - Chart$

Control Limit For$\overline{R} - Chart$:

Since all sample falls inside the control limits the statistical process is under control according to R-chart

9) **Thirty five successive sample of 100 castings each taken from a production line, contained respectively 3,3,5,3,0 5,3,2,3,5,6,5,9,1,2,4,5,2,0,10,3,6,3,2,5,6,3,3,2,5,1,0,7,4 & 3 defectives if the fraction defective is to be maintained at 0.02, construct the P-Chart for these data and state whether or not this standard is being met** **(Nov/Dec-2015)**

Soln: Since the size of the sample are equal

$$p\ for\ sample = \frac{No\ of\ defectives\ in\ the\ sample}{No\ of\ items\ in\ the\ sample} \qquad p\ (for\ sample\ \ No:1) = \frac{3}{100} = 0.03$$

Similarly calculate p for each sample and tabulate, Divide the number of defectives by 100 to get the fraction defectives

Sample No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No of Defectives	3	3	5	3	0	5	3	2	3	5	6	5	9	1
P = Fraction Defectives	0.03	0.03	0.05	0.03	0	0.05	0.03	0.02	0.03	0.05	0.06	0.05	0.09	0.01
Sample No	15	16	17	18	19	20	21	22	23	24	25	26	27	28
No of Defectives	2	4	5	2	0	10	3	6	3	2	5	6	3	3
P = Fraction Defectives	0.02	0.04	0.05	0.02	0	0.1	0.03	0.06	0.03	0.02	0.05	0.06	0.03	0.03
Sample No	29	30	31	32	33	34	35							
No of Defectives	2	5	1	0	7	4	3							
P = Fraction Defectives	0.02	0.05	0.01	0	0.07	0.04	0.03							

$$\overline{p} = \frac{0.03 + 0.03 + 0.05 + \cdots + 0.04 + 0.03}{35}$$

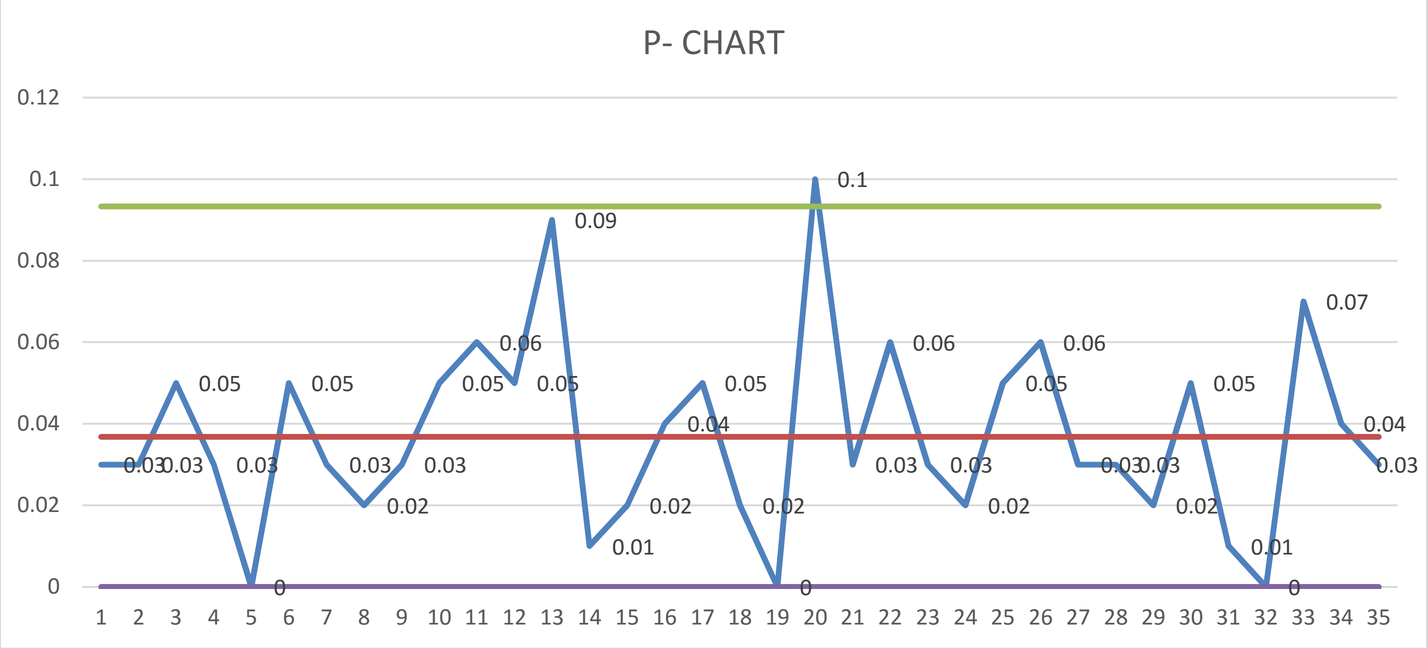
$$\boldsymbol{\overline{p} = 0.0368}$$

$$CL = \overline{p} = \mathbf{0.0368}$$

$$UCL = \left[\overline{p} + 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.0368 + 3\left(\sqrt{\frac{0.0368(1-0.0368)}{100}}\right) = \mathbf{0.09338}$$

$$LCL = \left[\overline{p} - 3\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}\right)\right] = 0.0368 - 3\left(\sqrt{\frac{0.0368(1-0368)}{100}}\right) = -\mathbf{0.01967}$$

Since LCL is negative, We take LCL = 0 as negative of imperfection is not valid in practical case



Conclusion: Since sample number 20 fall outside the CL, The Statistical Process is out of control according to p-chart

16.4 TOLERANCE LIMITS

In Chap. 8 we discussed confidence intervals for the population mean when we assume that the sample observations came from a normally distributed population. Theorem 8.2.2 defined the confidence interval on the mean when the population variance must be estimated. We note here that this confidence interval relates to an interval within which we are highly confident that the true mean μ lies. Often, particularly in engineering applications, we are interested in statements about individual observations. For example, we may need to know the proportion of individual values in the population that lie in some specified interval. Or, there might be specification limits and we may wish to estimate what proportion of items lie within the specification limits. We will consider two methods for computing tolerance intervals. The first method assumes a normal distribution for the population. In the second approach, we do not assume any specific distribution (nonparametric).

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 at least δ proportion of the population is included between these values.

Assumed Normal Distribution

When normality is assumed, we have seen that the interval

$$(\mu - 1.96\sigma, \mu + 1.96\sigma)$$

contains 95% of the population. In practice, μ and σ are usually unknown and must be estimated by $\overline{X}$ and S, the sample mean and standard deviation, Thus, the interval

$$(\overline{X} - 1.96S, \overline{X} + 1.96S)$$

is a random interval, and hence will no longer cover exactly 95% of the population. However, it can be shown that the interval

$$(\overline{X} - KS, \overline{X} + KS)$$

covers δ of the population with confidence $1 - \alpha$. Values of the constant K are given in Table XIV for various values of δ and $1 - \alpha$.

Example 16.4.1. A certain machine was made to dispense twelve ounces of cereal per box. To check on the precision of the machine, a team sampled 25 boxes and measured the weight of their contents. The sample average weight was 11.959 oz., and the sample standard deviation was 0.228. Assuming a normal distribution, calculate an interval so that we can claim, with 95% confidence, that 99% of the population lies between the smallest and largest sample observation.

From Table XIV we find $K = 3.457$ for $1 - \alpha = .95$ and $\delta = .99$. Thus the tolerance interval is given by

$$(\overline{X} - KS, \overline{X} + KS)$$

which becomes

$$[11.959 - (3.457)(.228), 11.959 + (3.457)(.228)]$$

or

$$(11.171, 12.747)$$

For some problems we need one-sided tolerance limits. That is, determine the sample size needed so that a specified proportion δ of the population is above the smallest value or below the largest value in the sample.

One-Sided Tolerance Limits

A one-sided tolerance limit is a minimum (or maximum) value determined from a sample of size n, chosen so that one can claim with $(1 - \alpha)\%$ confidence that at least δ proportion of the population will exceed this minimum (is less than this maximum) value.

Table XV can be used for this purpose, as demonstrated in Example 16.4.2.

Example 16.4.2. A manufacturer of automotive batteries wishes to establish a warranty so that they can be 95% confident that 99% of the batteries will last as long as the warranty period. Assume that battery life time follows a normal distribution. The research team randomly selected $n = 50$ batteries and ran a test on the life of each battery. They found the average life for the sample to be 39 months with sample standard deviation 3.0 months.

The lower (one-sided) tolerance limit is given by

$$\overline{X} - KS$$

From Table XV we find $K = 2.863$, which gives a lower tolerance limit of 30.411. Hence, a warranty period of 30 months seems reasonable.

Nonparametric Tolerance Interval

The presentation of tolerance intervals given requires the assumption of normality for the population from which the sample is taken. Often that is not a reasonable assumption. For example, the distribution may be skewed to the right or left. There does exist a nonparametric method (independent of the distribution). These intervals will usually be wider and/or require larger sample sizes for specified δ and $1 - \alpha$.

It can be shown that

$$P[(Y_{(1)}, Y_{(n)}) \text{ covers at least } \delta \text{ of the population}] \\ = 1 - n\delta^{n-1} + (n-1)\delta^{n} \tag{16.1}$$

where $Y_{(1)}$ and $Y_{(n)}$ denote the minimum and maximum value in the sample of size n, respectively. Table XVI gives the values for the sample size needed so that δ proportion of the population is between $Y_{(1)}$ and $Y_{(n)}$ with $1 - \alpha$ percent confidence.

Example 16.4.3. If we do not assume a normal distribution, what size sample is needed so that we can claim, with 90% confidence, that at least 95% of the population will be included between the smallest and largest observation (i.e., $\alpha = .10$ and $\delta = .95$)?

From Table XVI we see that $n = 77$ observations are required.

Nonparametric intervals can be used in various ways. One obvious approach is to find the sample size needed so that the tolerance interval covers δ percent of the population with confidence $1 - \alpha$. Another approach would be, for a given sample size, to find the confidence level for a specified δ proportion of the population to be included within the tolerance interval. This can be done by solving equation 16.1 for various values of n with δ fixed until an acceptable confidence is obtained.

16.5 ACCEPTANCE SAMPLING

Although modern quality control techniques tend to emphasize process control so that defective items are not produced, another important area of statistical quality control is acceptance sampling. When a batch or lot of items has been received by the buyer, he or she must decide whether to accept the items. Usually, inspection of every item in the lot is impractical. This may be due to the time or cost required to do such an inspection; it may be due to the fact that inspection is destructive in the sense that inspecting an item thoroughly can be done only by cutting the item open or by testing it in some other way that renders it useless. Thus the decision to reject a lot must be made based on testing only a sample of items drawn from the lot. The sampling plans that we shall consider are called *attribute* plans. In these plans each item is classified as being either defective or acceptable. We make our decision as to whether or not to reject the lot based on the number of defectives found in the sample. As you will see, acceptance sampling is just an adaptation of classical hypothesis testing.

To begin, let us denote the number of items in the lot or batch by N. The true but unknown proportion of defective items in the lot is denoted by Π. We agree that the entire lot is acceptable if the proportion of defectives Π is less than or equal to some specified value Π_0. Since our job is to detect unacceptable lots, we want to test the hypothesis

$$H_0\colon \Pi \le \Pi_0 \qquad \text{(lot is acceptable)}$$

$$H_1\colon \Pi > \Pi_0 \qquad \text{(lot is unacceptable)}$$

Usually, to decide whether to reject H_0, we determine what is called an *acceptance number,* which we denote by c. If the number of defective items sampled exceeds c, we reject the lot; otherwise we accept it. As you know, two kinds of errors may be committed when testing a hypothesis. We might reject a lot that is, in fact, acceptable, thus committing a Type I error; we might fail to reject an unacceptable lot, thus committing a Type II error. Alpha, the probability of committing a Type I error in this context, is called the *producer's risk.* Beta, the probability of committing a Type II error, is called the *consumer's risk.*

As in the past, we shall be able to compute the value of α. In this case it will depend on the specific value of Π_0, the sample size n, and the lot size N. Thus in a particular case we shall always know the risk to the producer. To see how to compute α, consider a lot of size N of which the proportion Π_0 is defective. Let $r = N\Pi_0$ denote the number of defective items. We select a random sample of size n from the lot and consider the random variable D, the number of defective items found in the sample. This random variable follows a hypergeometric distribution. From Sec. 3.6 we know that its probability density function is given by

$$f(d) = \frac{\binom{r}{d}\binom{N-r}{n-d}}{\binom{N}{n}}$$

where d is an integer lying between $\max[0, n - (N - r)]$ and $\min[n, r]$. For a preset acceptance number c the producer's risk is given by

$$\begin{aligned}\alpha &= P[\text{reject } H_0 | \Pi = \Pi_0]\\ &= P[D > c | \Pi = \Pi_0]\\ &= \sum_{d>c} \frac{\binom{r}{d}\binom{N-r}{n-d}}{\binom{N}{n}}\end{aligned}$$

For relatively small samples this probability can be calculated directly. However, in practice we usually approximate it using either the binomial density or the Poisson density. In the binomial approximation the probability of "success," obtaining a defective part, is assumed to be r/N; in the Poisson approximation the parameter k is given by $k = nr/N$. These ideas are illustrated in the next example. We show you all three calculations. In practice, we would use the hypergeometric probability and would only turn to the approximations when the hypergeometric computations become too cumbersome to be practical.

Example 16.5.1. A construction firm receives a shipment of $N = 20$ steel rods to be used in the construction of a bridge. The lot must be checked to ensure that the breaking strength of the rods meets specifications. The lot will be rejected if it appears that more than 10% of the rods fail to meet specifications. We are testing

$$H_0\colon \Pi \leq .1 \qquad \text{(lot is acceptable)}$$
$$H_1\colon \Pi > .1 \qquad \text{(lot is unacceptable)}$$

We compute α under the assumption that the null value is correct. That is, we compute α under the assumption that the lot actually contains $r = N\Pi_0 = 20(.1) = 2$ defective rods. Since testing a rod requires that it be broken, we cannot test each rod. Let us assume that a sample of size $n = 5$ is selected for testing. Let us agree to reject the lot if more than one rod is found to be defective. In this way we are setting our acceptance number at $c = 1$. Note that D can assume only the values 0, 1, or 2. The producer's risk is given by

$$\begin{aligned}\alpha &= P[\text{reject } H_0|\Pi = .10]\\ &= P[D > 1|\Pi = .10]\\ &= P[D = 2|\Pi = .10]\\ &= \frac{\binom{2}{2}\binom{18}{5-2}}{\binom{20}{5}}\end{aligned}$$

Using the combination formula given in Chap. 1 to evaluate the terms shown above, we see that

$$\alpha = 816/15504 = .0526$$

That is, there is about a 5% chance that our sampling technique will lead us to reject an acceptable lot that contains only two defective items; there is about a 95% chance that we shall not reject such a lot. Since the numbers used in this example are small, the calculation based on the hypergeometric distribution is not difficult. For comparative purposes we approximate the value of α by using a binomial random variable X with $n = 5$ and $p = .1$. Since we want to find the probability associated with the right-tail region of the hypergeometric distribution, we approximate α by finding the probability associated with the right-tail region of the appropriate binomial distribution. In this case

$$\begin{aligned}\alpha = P[D = 2] &\doteq P[X \geq 2]\\ &= 1 - P[X < 2]\\ &= 1 - P[X \leq 1]\end{aligned}$$

From Table I of App. A, $\alpha \doteq 1 - .9185 = .0815$. We can also approximate α by using a Poisson random variable Y with parameter $k = nr/N = 5(2)/20 = .5$. From Table II of App. A,

$$\begin{aligned}\alpha = P[D = 2] &\doteq P[Y \geq 2] \\ &= 1 - P[Y < 2] \\ &= 1 - P[Y \leq 1] \\ &= 1 - .910 \\ &= .09\end{aligned}$$

These approximations overestimate α, but considering the small numbers involved, they are not bad!

For a set sample size, a set lot size, and a set acceptance number, the probability of accepting a lot depends only on $\Pi = r/N$, the proportion of defectives actually in the lot. The hypergeometric distribution can be used to compute this probability for $r = 0, 1, 2, 3, \ldots, N$. The graph of this acceptance probability as a function of Π is called the *operating characteristic* or OC curve. In the next example we demonstrate how to construct and read an OC curve.

Example 16.5.2. Consider the problem described in Example 16.5.1 in which $N = 20$, $n = 5$, and $c = 1$. The probability of accepting this lot depends only on the proportion of defectives in the lot. We calculate the probability for various values of r and Π by using the equation

$$P\left[\text{accept lot}|\Pi = \frac{r}{N}\right] = \sum_{d \leq 1} \frac{\binom{r}{d}\binom{N-r}{n-d}}{\binom{N}{n}}$$

For example, the probability of accepting a lot that contains no defective items is given by

$$\frac{\binom{0}{0}\binom{20}{5}}{\binom{20}{5}} = 1$$

The probability of accepting a lot that contains exactly one defective item is

$$\frac{\binom{1}{0}\binom{19}{5}}{\binom{20}{5}} + \frac{\binom{1}{1}\binom{19}{4}}{\binom{20}{5}} = \frac{11628 + 3876}{15504} = 1$$

We have already seen that the probability of accepting a lot that contains exactly two defective items is $1 - .0526 = .9474$. Similar calculations can be done for $r = 3, 4, 5, \ldots, 20$. The results of these calculations for selected values of r are shown in Table 16.5. Using Table 16.5, we can make a quick sketch of the OC curve for this sampling

TABLE 16.5

r	Π	Probability of acceptance
0	0	1
1	.05	1
2	.10	.9474
5	.25	.6339
10	.50	.1517
15	.75	.0049
20	1.00	0

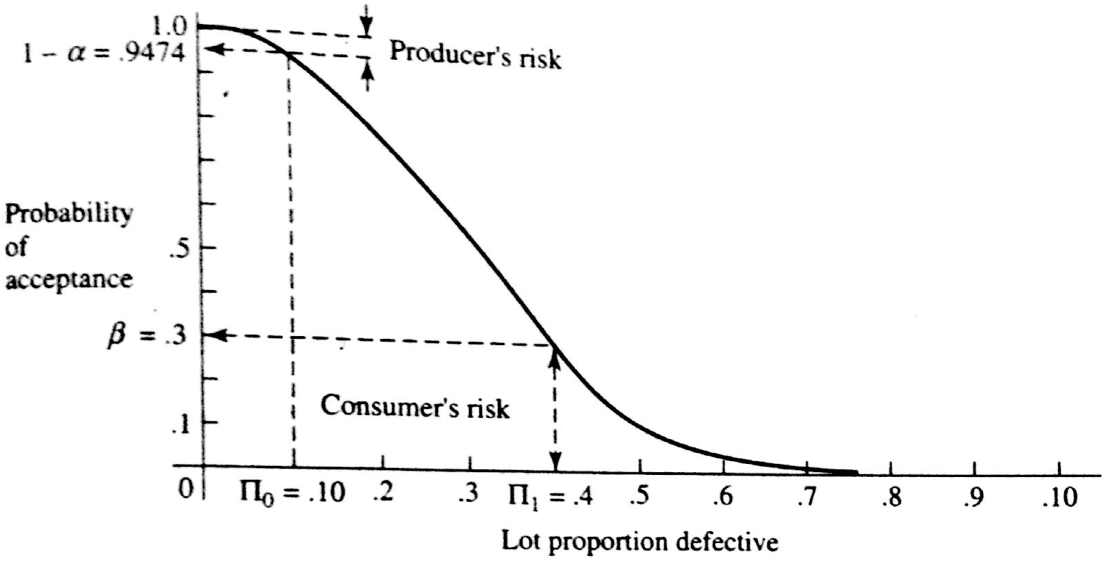


FIGURE 16.7
An OC curve with $N = 20$, $\Pi_0 = .10$, $c = 1$, $n = 5$; the producer's risk is $\alpha = .0526$; the consumer's risk when $\Pi = A$, β, is approximately .3.

plan by plotting the lot proportion defective versus the probability of acceptance for these selected values and then by joining the points with a smooth curve. The resulting sketch is shown in Fig. 16.7. The producer's risk is found by projecting a vertical line up from the point $\Pi = \Pi_0$ until it intersects the OC curve. A horizontal line is then projected over to the vertical axis. It intersects this axis at the point $1 - \alpha$. The producer's risk (a) is the length of the line segment from this intersection point to 1, as shown in Fig. 16.7. The consumer's risk (β) for a specified alternative $\Pi_1 > \Pi_0$ can also be read from the OC curve. For example, suppose that we want to determine the probability of accepting a lot in which the true proportion of defectives is $\Pi_1 = .4$. We use the projection method to see that this probability is approximately .3 as shown in Fig. 16.7. Note that as the difference in Π_0 and Π_1 increases, β decreases. That is, as the proportion of defectives increases, we are less likely to accept an unacceptable lot.

As we have seen in earlier discussions on hypothesis testing, the typical approach is to specify a value for α and then to determine the appropriate rejection region. Here we would specify α and then determine the acceptance number that gives us this approximate α value. In this way we control the producer's risk. However, if samples are small, this might result in an unacceptably large risk to the consumer.

In practice, efforts are made to obtain a balance between the producer's risk (α) and the consumer's risk (β). To do so, we specify a value $\Pi_1 > \Pi_0$ that represents to us a "barely acceptable" lot. For example, if we really want $\Pi \leq .10$, we might agree that a defective rate of .12, while not ideal, is at least barely acceptable. When N, Π_0, Π_1, α, and β are specified, it is possible to find a combination of n and C that meets the targets for α and β. That is, it is possible to find an OC curve such that at Π_0 the probability of accepting the lot is $1 - \alpha$ and at Π_1 the probability of accepting the lot is β. There are many sources available that give OC curves for specified values of N, n, α, and β. One of the most popular sources is Military Standard 105D [33].

16.6 TWO-STAGE ACCEPTANCE SAMPLING

Sometimes multiple-stage acceptance sampling plans are used. These plans can lead to smaller average sample sizes required to produce the same or similar OC curves as those that result in single-stage sampling. This is important when sampling is expensive, as in the case of destructive sampling. In this section we consider two-stage sampling in which lot sizes are large enough so that the binomial or normal distributions yield a good approximation to the hypergeometric distribution.

In a two-stage sampling scheme a single sample is drawn. If the number of defective items in the sample is large, the lot is rejected immediately and sampling ceases. If the number of defective items is very small, then the lot is accepted immediately and sampling also ceases. However, if the number of defective items is deemed to be moderate in size so that no clear decision is obvious, then a second sample is drawn. The decision to accept or reject the lot is made based on the total number of defective items in the two samples combined. The terms "very small," "large," and "moderate" are defined relative to the probability of obtaining various numbers of defective items.

The next example illustrates the computation of an OC chart in a two-stage sampling design. Recall that to compute an OC chart, we must find

$$P[\text{accept lot}|\Pi] = P[\text{accept lot on first or second sample}|\Pi]$$

That is, the OC chart is a graph of the probability of accepting a lot as a function of the true proportion of defectives in the lot.

Example 16.6.1 Consider the following two-stage sampling scheme. We draw a sample of size $n_1 = 50$ and decide to reject the lot if the number of defective items is four or more, to accept the lot if the number is 0 or 1, and to take a second sample otherwise. Figure 16.8(*a*) illustrates this first stage of sampling. If a second sample of size $n_2 = 50$ is needed, then we reject the lot if the total number of defective items in the two samples combined is five or more; otherwise the lot is accepted. The second-stage acceptance rule is shown in Fig. 16.8(*b*). Notice that the rejection rule can change from four to five because the sample size has increased from 50 to 100. Figure 16.9 summarizes the entire sampling procedure. Let D denote the total number of defective items obtained while sampling. The probability of acceptance when the true proportion of defectives is Π is given by

$$P[\text{accept}] = P[\text{accept on first sample}] + P[\text{take a second sample and accept}]$$

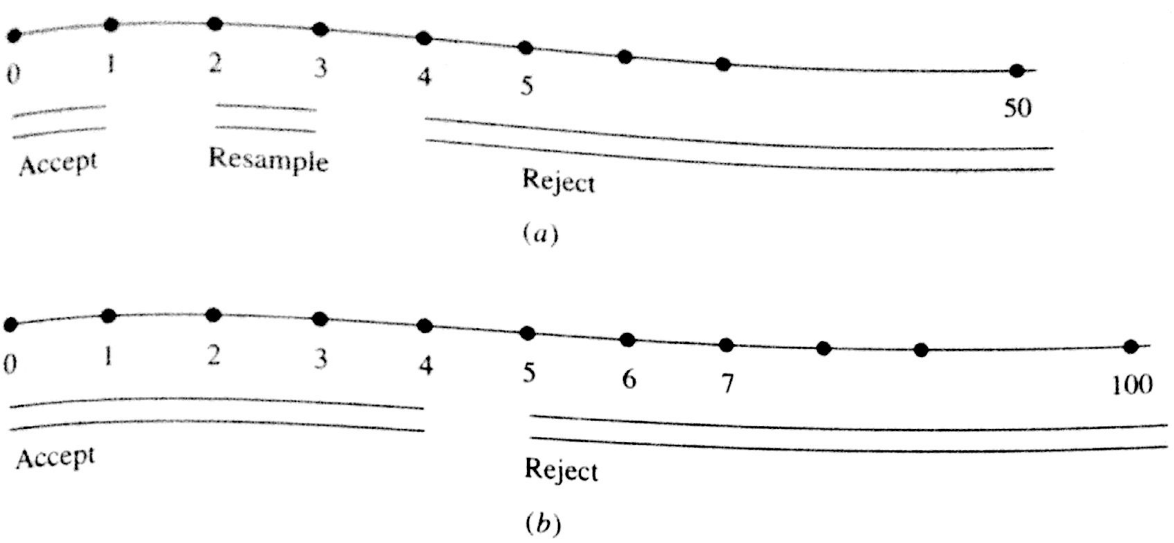


FIGURE 16.8

(*a*) A three-way decision rule is used in the first stage of a two-stage sampling scheme; (*b*) a two-way decision rule is used on the combined sample of size 100 in the second stage of sampling.

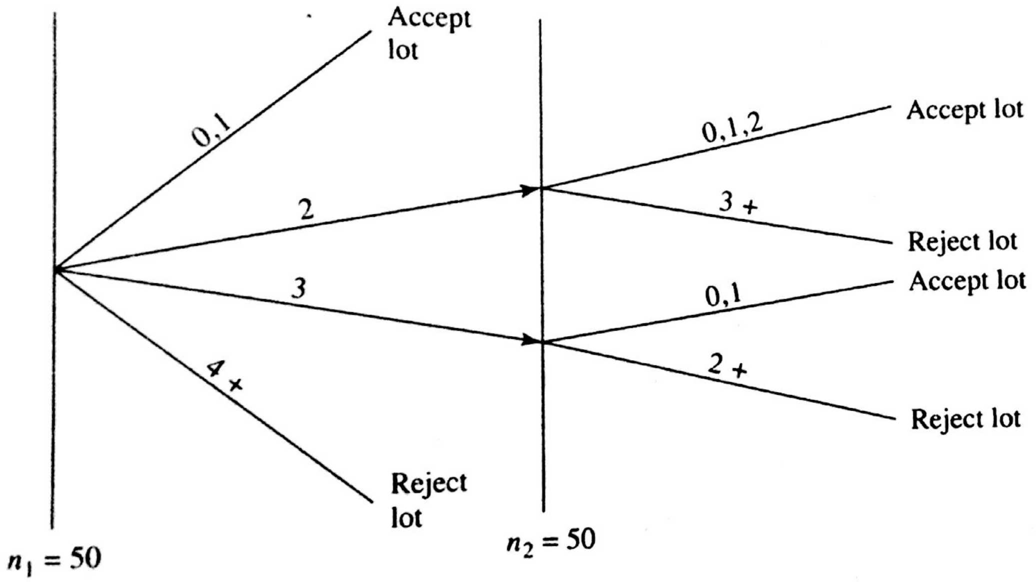


FIGURE 16.9

A two-stage lot acceptance plan.

By using the multiplication rule discussed in Sec. 2.3, we can express the latter probability as

$$
\begin{aligned}
P[\text{take a second sample and accept}] &= P[\text{accept}|\text{take a second sample}] \\
&\quad \times P[\text{take a second sample}]
\end{aligned}
$$

In this example,

$$
\begin{aligned}
P[\text{accept}] = P[D \le 1|n_1 = 50, \Pi] &+ P[D = 2|n_j = 50, \Pi]P[D \le 2|n_2 = 50, \Pi] \\
&+ P[D = 3|n_1 = 50, \Pi]P[D \le 1|n_2 = 50, \Pi]
\end{aligned}
$$

To illustrate, let us calculate the probability of acceptance when $\Pi = .10$. The probabilities can be found by using the binomial distribution with $n = 5$. Those probabilities given in Table 16.6 were found by using an extended binomial table that lists probabilities for $n = 50$ and $p = .01$ through $p = .10$. Since our binomial table does

TABLE 16.6

Π	Probability of acceptance
0	1
.01	.996
.02	.952
.03	.833
.04	.661
.05	.482
.06	.328
.07	.212
.08	.132
.09	.080
.10	.047

not list these values, we can either calculate the desired probabilities from the binomial density or approximate them by using the normal curve. The normal approximation technique is demonstrated below. In this approximation it is assumed that D is approximately normally distributed with $\mu = 50(.1) \quad 5$, $\sigma^2 = 50(.1)(.9) = 4.5$, and $\sigma = 2.12$:

$$\begin{aligned}
P[D \leq 1 | n_1 = 50, \Pi = .1] &\doteq P\left[Z \leq \frac{1.5 - 5}{2.12}\right] \\
&= P[Z \leq -1.65] \\
&= .0495 \\
P[D = 2 | n_1 = 50, \Pi = .1] &\doteq P[-1.65 \leq Z \leq -1.18] \\
&= .0695 \\
P[D \leq 2 | n_2 = 50, \Pi = .1] &= .1190 \\
P[D = 3 | n_1 = 50, \Pi = .1] &= .1199
\end{aligned}$$

Substitution yields

$$\begin{aligned}
P[\text{accept}] &\doteq .0495 + .0695(.1190) + .1199(.0495) \\
&= .0637
\end{aligned}$$

Notice that this approximation is fairly close to the binomial value given in Table 16.6.

The ideas illustrated here for two-stage acceptance sampling can be extended to multiple-stage sampling.

TABLE XIV
Factors for two-sided tolerance limits

	$1-\alpha=0.95$			$1-\alpha=0.99$		
δ / n	0.90	0.95	0.99	0.90	0.95	0.99
2	32.019	37.674	48.430	160.193	188.491	242.300
3	8.380	9.916	12.861	18.930	22.401	29.055
4	5.369	6.370	8.299	9.398	11.150	14.527
5	4.275	5.079	6.634	6.612	7.855	10.260
6	3.712	4.414	5.775	5.337	6.345	8.301
7	3.369	4.007	5.248	4.613	5.488	7.187
8	3.136	3.732	4.891	4.147	4.936	6.468
9	2.967	3.532	4.631	3.822	4.550	5.966
10	2.839	3.379	4.433	3.582	4.265	5.594
11	2.737	3.259	4.277	3.397	4.045	5.308
12	2.655	3.162	4.150	3.250	3.870	5.079
13	2.587	3.081	4.044	3.130	3.727	4.893
14	2.529	3.012	3.955	3.029	3.608	4.737
15	2.480	2.954	3.878	2.945	3.507	4.605
16	2.437	2.903	3.812	2.872	3.421	4.492
17	2.400	2.858	3.754	2.808	3.345	4.393
18	2.366	2.819	3.702	2.753	3.279	4.307
19	2.337	2.784	3.656	2.703	3.221	4.230
20	2.310	2.752	3.615	2.659	3.168	4.161
25	2.208	2.631	3.457	2.494	2.972	3.904
30	2.140	2.549	3.350	2.385	2.841	3.733
35	2.090	2.490	3.272	2.306	2.748	3.611
40	2.052	2.445	3.213	2.247	2.677	3.518
45	2.021	2.408	3.165	2.200	2.621	3.444
50	1.996	2.379	3.126	2.162	2.576	3.385
55	1.976	2.354	3.094	2.130	2.538	3.335
60	1.958	2.333	3.066	2.103	2.506	3.293
65	1.943	2.315	3.042	2.080	2.478	3.257
70	1.929	2.299	3.021	2.060	2.454	3.225
75	1.917	2.285	3.002	2.042	2.433	3.197
80	1.907	2.272	2.986	2.026	2.414	3.173
85	1.897	2.261	2.971	2.012	2.397	3.150
90	1.889	2.251	2.958	1.999	2.382	3.130
95	1.881	2.241	2.945	1.987	2.368	3.112
100	1.874	2.233	2.934	1.977	2.355	3.096
150	1.825	2.175	2.859	1.905	2.270	2.983
200	1.798	2.143	2.816	1.865	2.222	2.921
250	1.780	2.121	2.788	1.839	2.191	2.880
300	1.767	2.106	2.767	1.820	2.169	2.850
400	1.749	2.084	2.739	1.794	2.138	2.809
500	1.737	2.070	2.721	1.777	2.117	2.783
600	1.729	2.060	2.707	1.764	2.102	2.763
700	1.722	2.052	2.697	1.755	2.091	2.748
800	1.717	2.046	2.688	1.747	2.082	2.736
900	1.712	2.040	2.682	1.741	2.075	2.726
1000	1.709	2.036	2.676	1.736	2.068	2.718
∞	1.645	1.960	2.576	1.645	1.960	2.576

TABLE XV

Factors for one-sided tolerance limits

	$1-\alpha = 0.95$			$1-\alpha = 0.99$		
n \ δ	0.90	0.95	0.99	0.90	0.95	0.99
2	20.581	26.260	37.094	103.029	131.426	185.617
3	6.156	7.656	10.553	13.995	17.370	23.896
4	4.162	5.144	7.042	7.380	9.083	12.387
5	3.407	4.203	5.741	5.362	6.578	8.939
6	3.006	3.708	5.062	4.411	5.406	7.335
7	2.756	3.400	4.642	3.859	4.728	6.412
8	2.582	3.187	4.354	3.497	4.285	5.812
9	2.454	3.031	4.143	3.241	3.972	5.389
10	2.355	2.911	3.981	3.048	3.738	5.074
11	2.275	2.815	3.852	2.898	3.556	4.829
12	2.210	2.736	3.747	2.777	3.410	4.633
13	2.155	2.671	3.659	2.677	3.290	4.472
14	2.109	2.615	3.585	2.593	3.189	4.337
15	2.068	2.566	3.520	2.522	3.102	4.222
16	2.033	2.524	3.464	2.460	3.028	4.123
17	2.002	2.486	3.414	2.405	2.963	4.037
18	1.974	2.453	3.370	2.357	2.905	3.960
19	1.949	2.423	3.331	2.314	2.854	3.892
20	1.926	2.396	3.295	2.276	2.808	3.832
25	1.838	2.292	3.158	2.129	2.633	3.601
30	1.777	2.220	3.064	2.030	2.516	3.447
35	1.732	2.167	2.995	1.957	2.430	3.334
40	1.697	2.126	2.941	1.902	2.364	3.249
45	1.669	2.092	2.898	1.857	2.312	3.180
50	1.646	2.065	2.863	1.821	2.269	3.125
55	1.626	2.042	2.833	1.790	2.233	3.078
60	1.609	2.022	2.807	1.764	2.202	3.038
65	1.594	2.005	2.785	1.741	2.176	3.004
70	1.581	1.990	2.765	1.722	2.153	2.974
75	1.570	1.976	2.748	1.704	2.132	2.947
80	1.559	1.965	2.733	1.688	2.114	2.924
85	1.550	1.954	2.719	1.674	2.097	2.902
90	1.542	1.944	2.706	1.661	2.082	2.883
95	1.534	1.935	2.695	1.650	2.069	2.866
100	1.527	1.927	2.684	1.639	2.056	2.850
150	1.478	1.870	2.611	1.566	1.971	2.741
200	1.450	1.837	2.570	1.524	1.923	2.679
250	1.431	1.815	2.542	1.496	1.891	2.638
300	1.417	1.800	2.522	1.476	1.868	2.608
∞	1.282	1.645	2.326	1.282	1.645	2.326

TABLE XVI

Sample size for two-sided nonparametric tolerance limits

δ \ $1-\alpha$	0.50	0.70	0.90	0.95	0.99	0.995
0.995	336	488	777	947	1,325	1,483
0.99	168	244	388	473	662	740
0.95	34	49	77	93	130	146
0.90	17	24	38	46	64	72
0.85	11	16	25	30	42	47
0.80	9	12	18	22	31	34
0.75	7	10	15	18	24	27
0.70	6	8	12	14	20	22
0.60	4	6	9	10	14	16
0.50	3	5	7	8	11	12