Events :

Any subset A of a sample space "S" is called an <u>event</u>.

$S = \{HH, HT, TH, TT\}$ (sample space)

$A = \{HH\}$ $B = \{TH, TT\}$

$C = \{HH, HT, TH, TT\}$ (A, B, C are events)

<u>Elementary Event</u> : An event consisting of a single element of sample space is called a simple / elementary event.

<u>Mutually Exclusive Events</u> : If events have no elements in common, then they are called as mutually exclusive events.

<u>Ex</u> : $S = \{H, T\}$ $\underbrace{A = \{H\} \quad B = \{T\}}_{\text{mutually exclusive events}}$

$A \cap B = \phi$ (Empty).

<u>Probability</u> : If an event can occur in "m"
(classical approach)

different ways out of a total number of "n" possible ways, all of which are equally likely, then the probability of the event is : $\boxed{P = \frac{m}{n}}$.

Ex: $S = \{H, T\}$ $A = \{H\}$ $B = \{T\}$

$P(A) = \frac{1}{2}$ $P(B) = \frac{1}{2}$

(Q). what is the probability of drawing an ace from a well-shuffled deck of 52 playing cards?

ANS). $P = \frac{4 \text{ aces}}{52 \text{ cards}} = \frac{4}{52} \left(= \frac{^4C_1}{^{52}C_1} \right)$

$$\boxed{P = 1/13}$$

Axioms of Probability: Let S be a finite sample space and A is an event of S, then P(A), i.e, probability of event "A" satisfies & compli

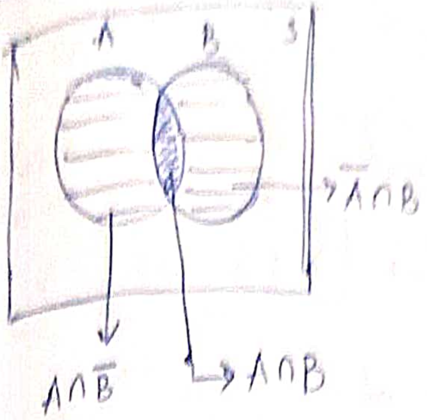
(i) $0 \leq P(A) \leq 1$ (ii) $P(S) = 1$

(iii) If A & B are mutually exclusive events in 'S' then $\boxed{P(A \cup B) = P(A) + P(B)}$

General Addition Rule: If A and B are any events in sample space S, then probability of $P(A \cup B)$ =:

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

*** Probability of impossible event is zero.



$$P(A\cup B) = P\left[(A\cap\bar{B})\cup(A\cap B)\cup(\bar{A}\cap B)\right]$$

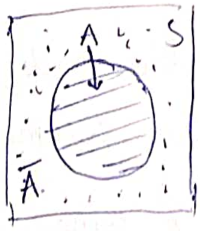
$$P(A\cup B) = P(A\cap\bar{B}) + P(\bar{A}\cap B) + P(A\cap B)$$

$$P(A\cup B) = \left[P(A\cap\bar{B}) + P(A\cap B)\right] + \left[P(\bar{A}\cap B) + P(A\cap B)\right] - P(A\cap B)$$

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

NOTE: When A & B are ME, then $P(A\cap B) = 0$, & Therefore, it reduces to (iii) axiom of probability.

→ If A is any event in S, then : $\boxed{P(\bar{A}) = 1 - P(A)}$



$$P(S) = P[A\cup\bar{A}]$$

$$P(S) = P(A) + P(\bar{A})$$

$$\boxed{P(\bar{A}) = 1 - P(A)}$$

Note: From the above result, we can write,

$P(\phi) = 1 - P(S)$ [ϕ, S are complements]

$P(\phi) = 1 - 1$

→ $\boxed{P(\phi) = 0}$

Random Variables :

A random variable, usually denoted by "X", on a sample space "S", is a function $\boxed{X : S \to R}$, i.e., from sample space to set of real numbers "R".

RV's are classified according to the number of values which taken assumed.

(i) Discrete RV - "X" is said to be discrete if "S" is discrete.

(ii) Continuous RV - "X" is said to be continuous if "S" is continuous.

Probability function (or) discrete probability distribution:

Let 'X' be a discrete random variable and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \ldots$ arranged in some order. Suppose also that these values are assumed with probabilities given by $P[X = x_k] = f(x_k)$, $(k = 1, 2, 3, \ldots)$

then the probability function is given by $\boxed{P[X = x] = f(x)}$. In general, $f(x)$ is a probability function if :

(i) $\boxed{f(x) \geq 0}$ (ii) $\boxed{\sum_{x} f(x) = 1}$

Distribution Function : The distribution function for a discrete RV "X" with prob. function, "$f(x)$" is given by :

$$\underline{F(x)} = P[X \leq x] = \sum_{u \leq x} f(u) \quad , \quad -\infty < x < \infty$$

(distribution function)

→ If "X" takes on only a finite number of values $x_1, x_2, x_3, \ldots x_n$, then the distribution function is given by :

$$F(x) = \begin{cases} 0 & , \quad -\infty < x < x_1 \\ f(x_1) & , \quad x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & , \quad x_2 \leq x < x_3 \\ \vdots & \\ f(x_1) + f(x_2) + \cdots + f(x_n) & , \quad x_n \leq x < \infty \end{cases}$$

---

① Find the probability func. corresponding to the random variable "X", where X represents the number of heads that can come up in tossing two coins simultaneously. Also, find the distribution function for the RV "X" and also obtaine its graph.

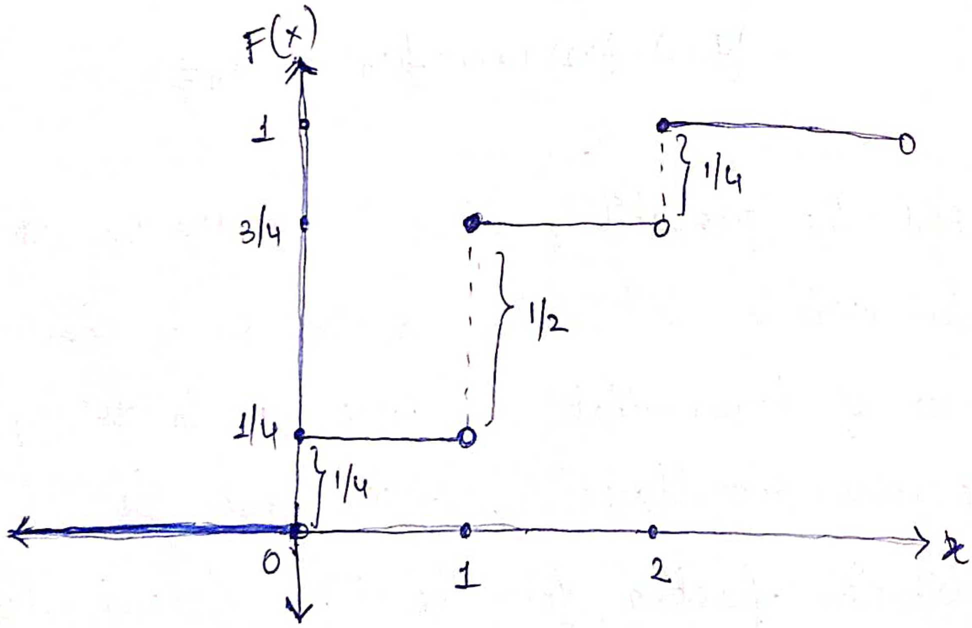
Sol). Let $S = \{HH, HT, TH, TT\}$ be the sample space. Let X be the no. of heads coming up when tossing 2 coins. Then the probability function is :

X	0	1	2
$P[X=x] = f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

* The distribution function is :

$$F(x) = \begin{cases} 0 & , -\infty < x < 0 \\ f(x_1) = f(0) = \frac{1}{4} & , 0 \le x < 1 \\ f(x_1) + f(x_2) = \frac{3}{4} & , 1 \le x < 2 \\ 1 & , 2 \le x < \infty \end{cases}$$

Graph of $F(x)$ v/s X ;



(i) "$F(x)$" is non-decreasing function.

(ii) $\lim_{x \to -\infty} F(x) = 0$ (iii) $\lim_{x \to \infty} F(x) = 1$

Q) A random var. "x" has the following probability function :

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

(i) Find "k"

(ii) Evaluate : $P[X<4]$

(iii) Evaluate: $P[X \geq 5]$

(iv) Evaluate: $P[3<X \leq 6]$

(v) What is the smallest value of "x" for which $P[X \leq x] > \frac{1}{2}$.

Sol).

(i) We know that $\sum_{x} P(x) = 1$

$$k+9k+3k+5k+7k+11k+13k = 1$$

$$\boxed{k = 1/49}$$

(ii) $P[X<4] = P[X=0] + P[X=1] + P[X=2] + P[X=3]$

$= k+3k+5k+7k$

$= 16k = 16/49$

(iii) $P[X \geq 5] = P[X=5] + P[X=6] = 11k + 13k$

$= 24k = 24/49$

(iv) $P[3<X \leq 6] = P[X=4] + P[X=5] + P[X=6]$

$= 33k = 33/49$

(v) when $x = 0$; $P[X \leq 0] = k = \frac{1}{49}$

$x = 1$; $P[X \leq 1] = 4k = \frac{4}{49}$

$x = 2$; $P[X \leq 2] = 9k = \frac{9}{49}$

$x = 3$; $P[X \leq 3] = 16k = \frac{16}{49}$

$x = 4$; $P[X \leq 4] = 25k = \frac{25}{49} > \frac{1}{2}$

$\therefore$ Smallest values of "x" is 4.

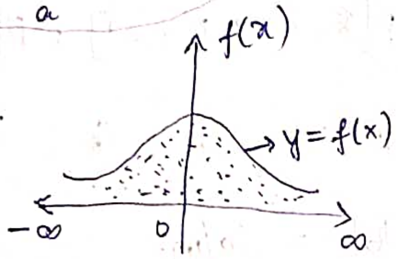
Probability density function: (PDF)

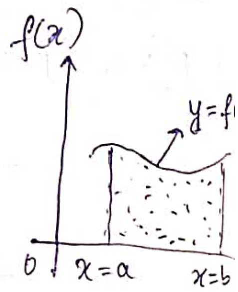
Let "X" be a continuous RV. Then the function $f(x)$ satisfy the following properties:

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

*** (iii) $P[a < X < b] = \int_{a}^{b} f(x)\,dx$

is known as PDF.





NOTE:

$P[a<x<b] = P[a \leq x < b] = P[a < x \leq b] = P[a \leq x \leq b]$

Distribution Function (D.F): The D.F for a continuous random variable "X" with PDF $f(x)$ is given by: $F(x) = P[X \leq x] = \int_{-\infty}^{x} f(u)\,du$ $(-\infty < x <$

Note : (i) $\lim_{x \to -\infty} F(x) = 0$ (ii) $\lim_{x \to \infty} F(x) = 1$

(iii) $f(x) = \frac{d}{dx}[F(x)] = F'(x) \geq 0$

(iv) $P[a < X < b] = F(b) - F(a)$

---

(3) Is the function defined as follows a density function?

$$f(x) = \begin{cases} 0 & , \quad x < 2 \\ \frac{1}{18}(3+2x) & , \quad 2 < x \leq 4 \\ 0 & , \quad x > 4 \end{cases}$$

Sol). In order for $f(x)$ to be a PDF,

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

$$\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{2} f(x)\,dx + \int_{2}^{4} f(x)\,dx + \int_{4}^{\infty} f(x)\,dx$$

$$= 0 + 0 + \int_{2}^{4} f(x)\,dx$$

$$= \frac{1}{18}\int_{2}^{4}(3+2x)\,dx$$

$$= \frac{1}{18}\left[3(4-2) + (16-4)\right]$$

$$= 1$$

$\therefore$ Given function is a PDF.

(4) If the probability density function of a continuous RV is given by, :

$$f(x) = \begin{cases} k(1-x^2), & \text{for } 0<x<1 \\ 0, & \text{elsewhere} \end{cases}$$

find the value of "k" and the probabilities of RV having this probability density will take on a value

(i) ~~...~~ b/w 0.1 and 0.2 (ii) > than 0.5

Sol).

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

$$\Rightarrow \int_{0}^{1} f(x)\,dx = 1$$

$$k\int_{0}^{1}(1-x^2)\,dx = 1$$

$$k\left[x-\frac{x^3}{3}\right]_0^1 = 1$$

$$\boxed{k = 3/2}$$

(i) $P[0.1<x<0.2] = \int_{0.1}^{0.2} f(x)\,dx = \int_{0.1}^{0.2} \frac{3}{2}(1-x^2)\,dx$

$$= \frac{3}{2}\left[x-\frac{x^3}{3}\right]_{0.1}^{0.2} = \frac{3}{2}\left[\left(0.2-\frac{1}{375}\right)-\left(0.1-\frac{1}{3000}\right)\right]$$

$$= \frac{3}{2}\left(0.1-\frac{7}{3000}\right) = 0.1465$$

(ii) $P[X > 0.5] = P[0.5 < x < \infty] = \int_{0.5}^{\infty} f(x)\,dx$

$$= \int_{0.5}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$$

$$= \frac{3}{2}\int_{0.5}^{1} (1-x^2)\,dx + 0$$

$$= \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0.5}^{1} = \frac{3}{2}\left[\left(1-\frac{1}{3}\right) - \left(0.5 - \frac{1}{24}\right)\right]$$

$$= \frac{3}{2}\left(\frac{1}{2} + \frac{1}{24} - \frac{1}{3}\right) = \underline{\underline{0.3125}}$$

---

(5) A RV "X" has density function

$$f(x) = \begin{cases} 1/4, & -2 < x < 2 \\ 0, & \text{elsewhere}, \end{cases}$$

then find

(i) $P[X < 1]$ (ii) $P[|X| > 1]$ (iii) $P[2X+3 > 5]$

Sol:

(i) $P[X < 1] = \int_{-\infty}^{1} f(x)\,dx = \int_{-\infty}^{-2} f(x)\,dx + \int_{-2}^{1} f(x)\,dx$

$$= 0 + \frac{1}{4}\int_{-2}^{1} dx = \frac{3}{4}$$

$\cancel{+\int_{1}^{\infty} f(x)\,dx}$

~~(ii) $P[|X| > 1] = \int_{-\infty}^{-1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$~~

~~$= \int_{-\infty}^{-2} + \int_{-2}^{1} + \int_{1}^{4}$~~

(ii) $P[|X|>1] = \int_{-\infty}^{-1} + \int_{1}^{\infty} = \int_{-\infty}^{-2} + \int_{-2}^{-1} + \int_{1}^{2} + \int_{2}^{\infty}$

$$\cancel{= 0 + \int_{-2}^{-1} + \int_{1}^{4} + 0}$$

$$\cancel{= \int_{-2}^{-1} \frac{dx}{4} + \int_{1}^{4} \frac{dx}{4} = \frac{1}{4}(1) + \frac{1}{4}(3)}$$

$$= 0 + \int_{-2}^{-1} + \int_{1}^{2} + 0$$

$$= \frac{1}{4}(1) + \frac{1}{4}(1) = \frac{1}{2}$$

(iii) $P[2X+3>5] = P[X>1]$

$$= \int_{1}^{2} + \int_{2}^{\infty} = \int_{1}^{2} f(x)\,dx + 0 = \frac{1}{4}$$

---

(6) The length of time that a certain lady speaks on the telephone is found to be random phenomena with a PDF specified by:

$$f(x) = \begin{cases} ke^{-x/5}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Determine the value of "k".

(ii) What is the prob. that the number of minutes that she will talk over the phone is

(a) More than 10 mins (b) Less than 5 mins

(c) Between 5 and 10 mins.

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

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

$$k\int_{0}^{\infty} e^{-x/5}\,dx = 1$$

$$\cancel{k\left[e^{-x/5}\cdot(-5)\right] = 1}$$

$$\cancel{-5k = e^{x/5}}$$

$$\cancel{\boxed{k = -\frac{e^{x/5}}{5}}}$$

$$-5k\left[e^{-x/5}\right]_0^{\infty} = 1$$

$$-5k\left[\frac{1}{e^{x/5}}\right]_0^{\infty} = 1$$

$$-5k\left[\frac{1}{\infty} - 1\right] = 1$$

$$\boxed{k = 1/5}$$

(ii)

(a) $P[X > 10] = \int_{10}^{\infty} f(x)\,dx = \frac{1}{5}\int_{10}^{\infty} e^{-x/5}\,dx$

$$= -1\,\cancel{5}\left[\frac{1}{e^{x/5}}\right]_{10}^{\infty} = -\left(0 - \frac{1}{e^2}\right) = \underline{\underline{e^{-2}}}$$

(b) $P[X<5] = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{5} f(x)\,dx$

$$= \frac{1}{5}\int_{0}^{5} e^{-x/5}\,dx = \frac{1}{5}(-5)\left[\frac{1}{e^{x/5}}\right]_{0}^{5}$$

$$= -\left(\frac{1}{e} - 1\right) = \boxed{1 - \frac{1}{e}}$$

(c) $P[5<X<10] = \int_{5}^{10} f(x)\,dx = \frac{1}{5}\int_{5}^{10} e^{-x/5}\,dx$

$$= -\left(\frac{1}{e^{x/5}}\right)_{5}^{10} = -\left(\frac{1}{e^2} - \frac{1}{e}\right) = \boxed{\frac{1}{e} - \frac{1}{e^2}}$$

---

(7) A continuous RV "X" has the PDF: $f(x) = 3x^2,\ 0 \le x < 1$. Find "a" and "b" such that: (i) Prob. of $[x \le a] = 0.5$ (ii) $P[x \ge b] = 0.05$

Sol).

(i) $P[x \le a] = 0.5$

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

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx = 0.5$$

$$3\int_{0}^{a} x^2\,dx = 0.5$$

$$[a^3] = 0.5$$

$$\boxed{a = 0.793700}$$

(ii) $P[x \geq b] = 0.05$

$$\int_b^\infty f(x)\,dx = 0.05$$

$$\int_b^1 f(x)\,dx + \int_1^\infty \cancel{f(x)\,dx}^{\,\to 0} = 0.05$$

$$3\int_b^1 x^2\,dx = 0.05$$

$$[1 - b^3] = 0.05$$

$$b^3 = 0.95$$

$$\boxed{b = 0.983047}$$

7/10/22

---

① Find the DF for the continuous RV "x"

$$f(x) = \begin{cases} x^2/9, & 0 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$$ and also find $P[1 < X \leq 2]$

Sol).

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(u)\,du \qquad (-\infty < x < \infty)$$

If $x \leq 0$; $F(x) = \int_{-\infty}^{x} 0\,du = 0$

If $0 < x < 3$; $F(x) = \int_{-\infty}^{0} \cancel{f(u)\,du}^{\,\to 0} + \int_0^x f(u)\,du$

$$= \int_0^x \frac{u^2}{9}\,du = \frac{x^2}{27}$$

If $x \geq 3$; $F(x) = \int_{-\infty}^{3} f(u)\,du + \int_3^x f(u)\,du$

$$F(x) = \int_{-\infty}^{0} f(u)\,du + \int_0^3 f(u)\,du + \int_3^x f(u)\,du$$

$$F(x) = 0 + \frac{1}{9}\left(\frac{x^3}{3}\right)_0^3 + 0 = \frac{1}{9}(9)$$

$$F(x) = 1$$

$$\therefore F(x) = \begin{cases} 0, & x \le 0 \\ x^3/27, & 0 < x < 3 \\ 1, & x \ge 3 \end{cases}$$

$$P[1 < X \le 2] = \int_1^2 \frac{x^2}{9} dx = \frac{1}{27}(x^3)_1^2 = \frac{7}{27}$$

---

(2) $F(x) = \begin{cases} 0, & x \le a \\ \frac{1}{2}\left(\frac{x}{a}+1\right), & -a \le x \le a \\ 1, & x > a \end{cases}$ Verify that the following is a distribution of a CRV.

Sol). $F(x)$ is DF only when $f(x)$ is a PDF.

If $x \le a$, $f(x) = \frac{d}{dx}[F(x)] = 0$

If $-a \le x \le a$, $f(x) = \frac{d}{dx}\left[\frac{1}{2}\left(\frac{x}{a}+1\right)\right] = \frac{1}{2a}$

If $x > a$, $f(x) = \frac{d}{dx}(1) = 0$

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

$$= 0 + \int_{-a}^{a} \frac{dx}{2a} + 0$$

$$= \frac{1}{2a}(2a) = 1$$

$\therefore$ Given $F(x)$ is a distribution function.

A discrete RV assumes the values $-3, -2, -1, 0, 1, 2, 3$

d satisfies :

) $P[X>0] = P[X=0] = P[X<0]$

) $P[X=-3] = P[X=-2] = P[X=-1]$

:) $P[X=1] = P[X=2] = P[X=3]$, then construct

obability mass function (MF) & determine the dist. func. also.

l). We know $\sum_x f(x) = 1$

$$P[X>0] + P[X=0] + P[X>0] = 1$$

$$P[X=0] = 1/3$$

ts $P[X \nless 0] = 1/3$

$$P[X=-3] + P[X=-2] + P[X=-1] = 1/3$$

$$P[X=-2] = 1/9$$

s $P[X>0] = 1/3$

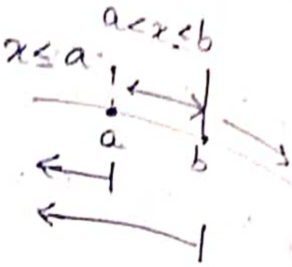
$$P[X=1] + P[X=2] + P[X=3] = 1/3$$

$$P[X=2] = 1/9$$

∴ The probability MF :

x	-3	-2	-1	0	1	2	3
$P[X=x] = f(x)$	$1/9$	$1/9$	$1/9$	$1/3$	$1/9$	$1/9$	$1/9$
$F(x)$	$1/9$	$2/9$	$1/3$	$2/3$	$7/9$	$8/9$	1

* If "F" is the CDF, of a RV "x", $a<b$,
then $P[a<x\leq b] = F(b)-F(a)$



Proof: We know that,

$$P\{[x\leq a]\cup[a<x\leq b]\} = P[x\leq b]$$

$$P[X\leq a]+P[a<x\leq b] = P[x\leq b]$$

$$P[a<x\leq b] = P[x\leq b]-P[X\leq a]$$

$$= F(b)-F(a)$$

Similarly, $P[a\leq x\leq b] = F(b)-F(a)+P[X=a]$

$$P[a<x<b] = F(b)-F(a)-P[X=b]$$

$$P[a\leq x<b] = P[X=a]+F(b)-F(a)-P[X=b]$$

Note: If $P[X=a]=P[X=b]=0$, then,

$$P[a<x\leq b] = P[a\leq x<b] = P[a<x<b] = P[a\leq x\leq b]$$

$$= F(b)-F(a)$$

---

* Constants of Random variables: The behaviour of a random variable (cont. or discr.) is completely characterized by the distribution function. Instead of a function, a more compact discretion can be made by a single numbers such as mean, variance, moments, etc.

(i) Mean (or) Expectation:

→ For a discrete RV "X" having the possible values $x_1, x_2, x_3, \ldots, x_n$, the mean or exp., is denoted by $E(X) / \mu_X / \mu$ and is defined as:

$$E_x = \sum_{i=1}^{n} x_i f(x_i) = \sum x \cdot f(x)$$

→ For a continuous RV "x" having the probability DF $f(x)$, the mean : $\boxed{\int_{-\infty}^{\infty} x f(x)\,dx}$.

Note: For a discrete RV ("x"), expectation of constant = constant.

(i) $E(k) = k$ (ii) $E(kx) = k \cdot E(x)$

(iii) $E(x \pm y) = E(x) \pm E(y)$ (iv) $E(xy) = E(x)E(y)$

(v) $E(kx \pm b) = k E(x) \pm b$ (vi) $E(x+k) = E(x) + k$

---

1. A discrete RV has a PDF, $f(x) = \frac{1}{2^x}$, where $x = 1, 2, 3, \ldots$, then find mean.

Ans). $E(x) = \sum_{i=1}^{n} x f(x)$

$$= 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + \ldots = S \text{ (let)}$$

$$S = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \ldots$$

$$\frac{S}{2} = \frac{1}{4} + \frac{2}{8} + \frac{3}{16} + \frac{4}{32} + \ldots$$

$$S - \frac{S}{2} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots$$

$$\frac{S}{2} = \frac{1}{2}\left[1 + \frac{S}{2}\right]$$

$$S = 1 + S/2$$

$$\boxed{S = 2}$$

(mean)

(2) $f(x) = \begin{cases} \frac{4}{81} x^2(9-x^2) & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$

$$\text{Mean} = E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= \int_{-\infty}^{0} x f(x)\, dx \rightarrow 0 + \int_{0}^{3} x f(x)\, dx + \int_{3}^{\infty} x f(x)\, dx \rightarrow 0$$

$$= \frac{4}{81} \int_{0}^{3} x^2(9-x^2)\, dx = \frac{4}{81} \int_{0}^{3} (9x^2 - x^4)\, dx$$

$$= \frac{4}{81} \left[\frac{9}{3}(27) - \frac{1}{5}(243) \right]$$

$$= \frac{4}{81} (81 - 243/5) = 4 - \frac{12}{5} = \frac{8}{5}$$

Variance : It characterises the variability in the distributions. Since two distributions with the same mean can still have different dispersion of data about the mean.

(i) If "X" is a D. R. variable having possible values $x_1, x_2, \ldots, x_n$ and having probability func. $f(x)$, then the variance is denoted by $var(x)$ or σ_x^2 or σ^2

$$\sigma^2 = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = \sum (x-\mu)^2 f(x)$$

(ii) If "X" is a C.R. variable having PDF $f(x)$, then the variance is given by $= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

$\frac{11}{16} + \frac{1}{16} + \frac{7}{144} + \frac{5}{16} + \frac{25}{48} + \frac{49}{144}$

Standard Deviation (σ): Positive square root of variance.

1) Determine the disc. prob. distr., expectation, variance and S.D of a DRV "X" which denotes the min of the 2 numbers that appear when a pair of fair dice ~~are~~ is thrown once.

Sol). $S = \{(1,1), (1,2), \dots, (6,6)\}$ (Sample Space)

Given that "X" takes the min. of 2 numbers that appear when a pair of fair dice is thrown:

x	1	2	3	4	5	6
$P[X=x] = f(x)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

$$\text{Expectation}(\mu) = \sum_{i=1}^{n} x_i f(x_i) \quad (n=6)$$

$$= \frac{11}{36} + 2\left(\frac{9}{36}\right) + 3\left(\frac{7}{36}\right) + 4\left(\frac{5}{36}\right) + 5\left(\frac{3}{36}\right) + 6\left(\frac{1}{36}\right)$$

$$= \frac{11}{36} + \frac{18}{36} + \frac{21}{36} + \frac{20}{36} + \frac{15}{36} + \frac{6}{36} = \frac{91}{36} = \underline{\underline{2.5277}}$$

$$\text{Variance}(\sigma^2) = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) \quad (n=6)$$

$$\sigma^2 = \left(1-\frac{91}{36}\right)^2\frac{11}{36} + \left(2-\frac{91}{36}\right)^2\frac{9}{36} + \left(3-\frac{91}{36}\right)^2\frac{7}{36} + \left(4-\frac{91}{36}\right)^2\frac{5}{36} + \left(5-\frac{91}{36}\right)^2\frac{3}{36} + \left(6-\frac{91}{36}\right)^2\frac{1}{36}$$

$$= (1-2.5)^2\frac{11}{36} + (2-2.5)^2\frac{9}{36} + (3-2.5)^2\frac{7}{36} + (4-2.5)^2\frac{5}{36} + (5-2.5)^2\frac{3}{36} + (6-2.5)^2\frac{1}{36} = \underline{\underline{1.972}}$$

$S.D(r) = \sqrt{r^2} = \underline{1.404}$

(2) Suppose a CRV 'X' has the PDF

$$f(x) = \begin{cases} k[1-x^2], & 0<x<1 \\ 0, & \text{elsewhere} \end{cases}$$ Determine

(i) Mean & variance of PDF (ii) Using distr. func., determine $P[0.4<x<0.6]$.

~~Sol).~~

$$\cancel{\text{Mean} = \int_{-\infty}^{\infty} x f(x)\,dx}$$

$$\cancel{= k\int_{-\infty}^{\infty}}$$

~~As $f(x)$ is PDF,~~

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

$$\cancel{k\int_{-\infty}^{\infty} (1-x^2)\,dx = 1}$$

$$\cancel{k\,[\infty}$$

Sol). As $f(x)$ is PDF, $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_{-\infty}^{0} + \int_{0}^{1} + \int_{1}^{\infty} = 1$$

$$k\int_{0}^{1} (1-x^2)\,dx = 1$$

$$\boxed{k = 3/2}$$

(i) Mean = $\cancel{\frac{3}{2}\int_{-\infty}^{\infty} (x - x^2)\,dx}$ $\int_{-\infty}^{\infty} x f(x)\,dx$

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

$$= \frac{3}{2}\left[\frac{1}{2}(1) - \frac{1}{4}(1)\right] = \frac{3}{8}$$

$$\text{Variance } (\sigma^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= \frac{3}{2}\int_0^1 \left(x-\frac{3}{8}\right)^2 (1-x^2)\,dx$$

$$= \frac{3}{2}\int_0^1 \left(x^2+\frac{9}{64}-\frac{6x}{8}\right)(1-x^2)\,dx$$

$$= \frac{3}{2}\int_0^1 \left(x^2-x^4+\frac{9}{64}-\frac{9x^2}{64}-\frac{6x}{8}+\frac{6x^3}{8}\right)dx$$

$$= \frac{3}{2}\left[\frac{1}{3}-\frac{1}{5}+\frac{9}{64}-\frac{9(1)}{64(3)}-\frac{6}{16}+\frac{6}{32}\right]$$

$$= \underline{\underline{0.059375}}$$

(ii) We know that, $F(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)\,dx$

$$F(x) = \int_{-\infty}^{0} + \int_0^x = \int_0^x f(x)\,dx \qquad (0<x<1)$$

$$F(x) = \frac{3}{2}\left[x-\frac{x^3}{3}\right] \qquad (0<x<1)$$

$$P(0.4<X<0.6) = F(0.6)-F(0.4)$$

$$= \frac{3}{2}\left[0.6-\frac{(0.6)^3}{3}\right] - \frac{3}{2}\left[0.4-\frac{(0.4)^3}{3}\right]$$

$$= \frac{99}{125}-\frac{71}{125}$$

$$= \underline{\underline{0.224}}$$

For a DRV "X", $\boxed{\sigma_X^2 = E(X^2)-[E(X)]^2}$

(var(x))

Proof:

We know $var(X) = E(X-\mu)^2$

$$var(x) = E[x^2 + \mu^2 - 2\mu x]$$

$$= E(x^2) + E(\mu^2) - E(2\mu x)$$

$$= E(x^2) + \mu^2 - 2\mu E(x) \qquad (\because E(k) = k)$$

$$= E(x^2) + \mu^2 - 2\mu^2 \qquad (\because E(x) = \mu)$$

$$= E(x^2) - \mu^2$$

$$= E(x^2) - [E(x)]^2$$

Note: Similar result follows for CRV too!

(*) For a discrete random variable "X", prove tha

(i) $Var(x+k) = var(x)$ (ii) $Var(kx) = k^2 . var(x)$

Proof: We know that: $var(x) = E(x^2) - [E(x)]^2$

(i)

$$var(x+k) = E(x+k)^2 - [E(x+k)]^2$$

$$= E(x^2) + E(k^2) + E(2kx) - [E(x+k)]^2$$

$$= E(x^2) + E(k^2) + E(2kx) - [E(x)+k]^2$$

$$= E(x^2) + E(k^2) + E(2kx) - [E(x)]^2 - k^2 - 2kE(x$$

$$= E(x^2) + \cancel{k^2} - \cancel{k^2} + \cancel{2kE(x)} - \cancel{2kE(x)} - [E(x)]^2$$

$$= E(x^2) - [E(x)]^2$$

$$= var(x)$$

$$\therefore var(x+k) = var(x)$$

Proof:

(ii) $var(kx) = E(k^2x^2) - (E[kx])^2$

$$var(kx) = k^2E(x^2) - [kE(x)]^2$$

$$= k^2[E(x^2) - [E(x)]^2]$$

$$= k^2 \cdot var(x)$$

Note: $\rightarrow var(k) = E(k^2) - [E(k)]^2$

$$= k^2 - k.k \qquad \therefore \boxed{var(k) = 0}$$

$$= k^2 - k^2 = 0$$

$\rightarrow \boxed{var[ax \pm b] = a^2.var(x)}$

---

(3) The daily consumption of electric power (in $kwh \times 10^6$) is a RV having the PDF:

$$f(x) = \begin{cases} \frac{1}{9}x e^{-x/3}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

If the total production is 12×10^6 kWH, then determine the probability that there is power cut on any given day.

Sol). Power cut occurs when (requirement > production).
or
consumption
(power supply is inadequate)

$$P[X > 12] = 1 - P[X \leq 12]$$

$$= 1 - \int_{-\infty}^{12} f(x)dx$$

$$= 1 - \left[\int_{-\infty}^{0} + \int_{0}^{12} f(x)dx\right]$$

$$= 1 - \int_{0}^{12} f(x)dx$$

$$P[X \geq 12] = 1 - \frac{1}{9}\int_0^{12} x e^{-x/3}\,dx$$

$$= 1 - \frac{1}{9}\left[x\int_0^{12} e^{-x/3}\,dx - \int_0^{12}\!\!\int e^{-x/3}\,dx\right]$$

$$= 1 - \frac{1}{9}\left[-3x\left(e^{-x/3}\right)_0^{12} + \int_0^{12} 3e^{-x/3}\,dx\right]$$

$$= 1 - \frac{1}{9}\left[-3x\left(e^{-4}-1\right) + 3\int_0^{12} e^{-x/3}\,dx\right]$$

$$= 1 - \frac{1}{9}\left[3x\left(1-e^{-4}\right) - 9\left(e^{-x/3}\right)_0^{12}\right]$$

$$= 1 - \frac{1}{9}\left[3x\left(1-e^{-4}\right) - 9\left(e^{-4}-1\right)\right]$$

$$\cancel{= 1 - \frac{1}{9}\left(e^{-4}-1\right)(9+3x)}$$

$$P[x \geq 12] = 1 - \frac{1}{9}\int_0^{12} x e^{-x/3}\,dx$$

$$= 1 - \frac{1}{9}\left[x\left(-3e^{-x/3}\right) - 9e^{-x/3}\right]_0^{12}$$

(4) A box contains 8 items of which 2 are defective. A person draws 3 items from the box. Determine the expected number of defective items he has drawn.

Sol). Let "X" denote the number of defectives.

$X = x$	0	1	2
$P[X=x] = f(x)$	$\frac{^6C_3}{^8C_3}$	$\frac{^2C_1 \cdot {^6C_2}}{^8C_3}$	$\frac{^2C_2 \cdot {^6C_1}}{^8C_3}$

$x = 0 \longrightarrow$

$x = 1 \longrightarrow$

$x = 2 \longrightarrow$

Expected number of defective items $= 0 \times \frac{^6C_3}{^8C_3} + 1 \times \frac{^2C_1 \cdot {^6C_2}}{^8C_3} + 2 \times \frac{^2C_2 \cdot {^6C_1}}{^8C_3}$

$$= 0 + \frac{30}{56} + 2\left(\frac{6}{56}\right) = \underline{\underline{0.75}}$$

(5) Determine the expected number of families to have
(i) 2 boys and 2 girls
(ii) ~~Atlea~~ Atleast 1 boy
(iii) No girls
(iv) Atmost 2 girls
out of 800 families with 4 children each. Assume equal probabilities for boys and girls.

Sol). Let "X" be number of boys (denote [RV]) in a family

X	0	1	2	3	4
~~$P[X=x] = f(x)$~~					

P.T.O →

Let "S" be the sample space.

S = {BBBB, BBBG, BBGB, BGBB, GBBB, GGGG, GGGB, GBGG, GGBG, BGGG, BBGG, GGBB, BGBG, GBGB, BGGB, GBBG}

X = x	0	1	2	3	4
P[X = x] = f(x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

(i) $\frac{6}{16} \times 800 = 300$ families

(ii) $\frac{15}{16} \times 800 = 750$ families

(iii) $\frac{1}{16} \times 800 = 50$ families

(iv) $\frac{11}{16} \times 800 = 550$ families

---

(6) A person ~~being~~ wins ₹80 if 3 heads occur, ₹30 if 2 heads occur, ₹10 if only 1 head occurs, in a single toss of 3 fair coins. If the game is to be fair, how much should he lose if no heads occur?

***** NOTE : In a gambling game, the expected value "E" of the game is considered to be value of the game to the player. Game is favourable to the player if $\boxed{\text{"E"} > 0}$, unfavourable if $\boxed{\text{"E"} < 0}$, fair if $\boxed{\text{"E"} = 0}$.

Sol). Let "X" denote the number of heads.

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

$X = x$	0	1	2	3
$P[X=x] = f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E = \sum_{i=0}^{3} x_i f(x_i) = 1\times\frac{3}{8} + 2\times\frac{3}{8} + 3\times\frac{1}{8} = \cancel{\frac{4}{8}}\ \frac{3}{2}$$

As game is fair, $E = 0$

$$E = \frac{-1}{8}\times(x) + \frac{3}{8}(30) + \frac{3}{8}(10) + \frac{1}{8}(80)$$

$$\frac{1}{8}x = \frac{90}{8} + \frac{30}{8} + \frac{80}{8}$$

$$\boxed{x = 200}$$

---

(7) A player tosses "3" fair coins. He wins ₹500 if 3 heads occur, ₹300 if 2 heads occur, ₹100 if 1 head occurs. On the other hand, he loses ₹1500 if 3 tails occur. Find the expected value of the game to the player. Is it favourable?

Sol). X = no. of heads

$X = x$	0	1	2	3
$P[X=x] = f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E = \frac{-1}{8}(1500) + \frac{3}{8}\times(100) + \frac{3}{8}(300) + \frac{1}{8}(500)$$

$$E = \frac{300 + 900 + 500 - 1500}{8} \Rightarrow \boxed{E > 0} \therefore \text{favourable.}$$

(8) The PDF of a CRV is given by

$f(x) = ce^{-|x|}, -\infty < x < \infty$. Then :

(i) Find "c" value. (ii) Find mean, variance?

(iii) Evaluate $P[0 \le x \le 4]$. ~~The densi~~

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

$$c\int_{-\infty}^{\infty} e^{-|x|} = 1 \Rightarrow c\left[\int_{-\infty}^{0} e^{x} + \int_{0}^{\infty} e^{-x}\right] = 1$$

$$c\left[(e^{0} - e^{-\infty}) + (-e^{-x})_{0}^{\infty}\right] = 1$$

$$c\left[1 - 1/e^{\infty} - (e^{-\infty} - 1)\right] = 1$$

$$2c(1-0) = 1$$

$$\boxed{c = 1/2}$$

(ii) mean $= E(x) = \int_{-\infty}^{\infty} x\, f(x)$

$$E(x) = \frac{1}{2}\int_{-\infty}^{\infty} x e^{-|x|}\,dx \cancel{= \frac{1}{2}\left[\int_{-\infty}^{0} x e^{x}\,dx + \int_{0}^{\infty} x e^{-x}\,dx\right]}$$

$$\cancel{E(x) = \frac{1}{2}\left[(xe^{x} - e^{x})_{-\infty}^{0}\right]}$$

$$\boxed{\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx \quad \text{(even)} \qquad \int_{-a}^{a} f(x)\,dx = 0 \quad \text{(odd)}}$$

$$E(x) = \frac{1}{2}(0) = 0$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{0}^{\infty} x^2 e^{-|x|}\,dx = \int_{0}^{\infty} x^2 e^{-x}\,dx$$

$$E(x^2) = 2 \qquad var(x) = E(x^2) - [E(x)]^2 = 2$$

(iii) $P[0 \le x \le 4] = \int_0^4 f(x)\,dx = \frac{1}{2}\int_0^4 e^{-x}\,dx$

$P[0 \le x \le 4] = -\frac{1}{2}(e^{-4} - 1) = \frac{1}{2}\left(1 - \frac{1}{e^4}\right)$

Measures of Central Tendency (CT):

① Mean : It is the measure of calc CT of (one of) a data. Also known as avg.
(arithmetic mean)

Ex: Calculate arithmetic mean of 7, 68, 41, 3, 5

$$\text{mean} = \frac{68+41+3+5+7}{5} = \frac{124}{5} = 24.8$$

① Calculate the A.M for the following data:

Student	A	B	C	D	E	F	G	H	I	J
Marks	2	7	10	8	6	3	5	4	5	6

$$A.M = \frac{\Sigma X}{N} = \frac{2+7+10+8+6+3+5+4+5+0}{10} = \frac{50}{10} = 5$$

$$\boxed{A.M = A + \frac{\Sigma\, dx}{N}} \qquad dx = X - A$$

$\downarrow$ (values)

(A = assumed mean)

② Calculate A.M for :

Marks	5	2	3	4
Frequency	4	6	8	6

Sol :

$$A.M = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{20+12+24+24}{24} = \frac{10}{3}$$

$$\boxed{A.M = A + \frac{\sum f_i dx_i}{\sum f_i}}$$ (A = assumed mean) ($dx_i = x_i - A$)

(2) The following table shows wages of workers. Cal. the arithmetic mean.

wages	10-20	20-30	30-40	40-50	50-60
workers	8	9	12	11	6

mid (x) = 15, 25, 35, 45, 55 , $\sum f = 46$

$\sum fx = 15\times8 + 25\times9 + 35\times12 + 45\times11 + 55\times6$
$= 1590$

mean ($\bar{x}$) $= \frac{1590}{46} = 34.56521739$

Median: Middle most or central value of variable in set of observations when arranged in order. odd $= \left(\frac{N+1}{2}\right)^{th}$ even $= \left(\frac{N}{2}\right)^{th} + \left(\frac{N+1}{2}\right)^{th}$

C.I	10-20	20-30	30-40	40-50	50-60	60-70
f	5	6	10	5	3	4

Sol). $N = \sum f = 33$

Median $= l + \left(\frac{\frac{N}{2} - cf}{f}\right) \times$

Median: The middle most or central value of the variable in a set of observations are arranged & either in asc/desc. order. It's a positional average. (mean is ~~calculated~~ known as calculated average).

After arranging "n" observations in asc./desc.

(i) n = odd, median = $\left(\frac{n+1}{2}\right)^{th}$ term

(ii) n = even, median = $\dfrac{\left(\frac{n}{2}\right)^{th} \text{ term} + \left(\frac{n}{2}+1\right)^{th} \text{ term}}{2}$

For grouped data:

$$\boxed{\text{median} = l + \frac{\frac{n}{2} - cf}{f} \cdot c}$$

l = lower limit of median class
n = total freq.
f = freq. of med. class

cf = cumulative freq. of preceding class to median class
c = width of class interval [median class = class interval whose c.f is greater than or nearest to n/2]

(Q). Calculate median of:

C.I	10-20	20-30	30-40	40-50	50-60	60-70
f	5	6	10	5	3	4

Sol).

	f	c.f
10-20	5	5
20-30	6	11
30-40	10	21
40-50	5	26
50-60	3	29
60-70	4	33

(N = 33)

l = 30
n = 33
f = 10
c.f = 11
c = 10

median = $30 + \left(\frac{16.5 - 11}{\cancel{10}}\right)\cancel{10} = 35.5$

MODE: The mode is the most commonly observed value in a set of data.

For grouped data : $\boxed{\text{mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} * c}$

l = lower limit of modal class

f_1 = freq. of modal class (M.C)

f_0 = freq. of preceding class of M.C

f_2 = freq. of succeeding class of M.C

c = width of class interval

[M.C = class having highest frequency]

(Q). Calculate mode

C.I	10-20	20-30	30-40	40-50	50-60	60-70
f	5	4	8	3	2	3

Sol).

$l = 30;\ f_1 = 8;\ f_0 = 4;\ f_2 = 3;\ c = 10$

$$\text{mode} = 30 + \left(\frac{4}{16-4-3}\right)10 = 30 + \frac{40}{9} = 34.4$$

Relation : $\boxed{\text{Mode} = 3\text{Median} - 2\text{Mean}}$

Mean : Average ; calculated to reach a single value which represents the entire data.

MOMENTS : Constants of data which help in deciding the characteristics of the population. Help in finding AM, SD and variance of the population directly, and they help in knowing the graphic stages of the population (data). Moment can be defined as the average of deviations of observations taken from a point that are raised to a certain power. Broadly classified as : ① Central moments ② Raw moments

① μ_r = r^{th} central moment $= \frac{1}{n} \Sigma (x - \bar{x})^r$ (ungrouped)

($\bar{x}$ = mean) $= \frac{1}{N} \Sigma f \cdot (x - \bar{x})^r$ (grouped)

② μ_r' = r^{th} raw moment about a point "a". $= \frac{1}{n} \Sigma (x - a)^r$ (individual)

$= \frac{1}{N} \Sigma f \cdot (x - a)^r$ (grouped)

Note: $\boxed{\mu_1 = 0}$; $\boxed{\mu_2 = \sigma^2 = \text{variance}}$;

If $a = 0$, $\boxed{\mu_1' = \text{mean}}$

(Q). Find 1st four moments about "4" for:

2, 8, 7, 4, 4

Sol).

X	x−4	$(x-4)^2$	$(x-4)^3$	$(x-4)^4$
2	−2	4	−8	16
8	4	16	64	256
7	3	9	27	81
4	0	0	0	0
4	0	0	0	0

n = 5 elements

$\mu_1' = \frac{1}{n} \Sigma (x - a)^1 = \frac{5}{5} = 1$

$\mu_2' = \frac{1}{n} \Sigma (x - a)^2 = \frac{29}{5} = 5.8$

$\mu_3' = \frac{1}{n} \Sigma (x - a)^3 = \frac{83}{5} = 16.6$

$\mu_4' = \frac{1}{n} \Sigma (x - a)^4 = \frac{353}{5} = 70.6$

(Q). Find out first four moments about origin for:

Sol).

C.I	f	X (mid-point)	$x-0$	$(x-0)^2$	$(x-0)^3$	$(x-0)^4$
0-10	12	5	5	25	125	625
10-20	23	15	15	225	3375	50625
20-30	24	25	25	625	15625	390625
30-40	32	35	35	1225	42875	1500625
40-50	15	45	45	2025	91125	4100625

(= 106)

$$\mu_1' = \frac{1}{n}\Sigma f(x-0)^1 = \frac{2800}{106} = 26.42$$

$$\mu_2' = \frac{1}{n}\Sigma f(x-0)^2 = \frac{90050}{106} = 849.53$$

$$\mu_3' = \frac{1}{n}\Sigma f(x-0)^3 = \frac{3193000}{106} = 30122.64$$

$$\mu_4' = \frac{1}{n}\Sigma f(x-0)^4 = \frac{120076250}{106} = 1132794.811$$

<u>Absolute Measure of Dispersion</u>	<u>Relative Measure of Dispersion</u>
1) Range	1) Co-eff. of Range
2) Quartile Deviation	2) Co-eff. of Q.D
3) Mean Deviation	3) Co-eff. of M.D
4) Standard Deviation	4) Co-eff. of S.D

1) Range : $\boxed{\text{Largest value} - \text{smallest value}} = (L-S)$

$$\text{Co. eff of Range} = \boxed{\frac{L-S}{L+S}}$$

2) $\boxed{Q.D = \frac{Q_3 - Q_1}{2}}$ (semi-inter quartile range)

($Q_3 - Q_1$ = inter quartile range)

$$\text{Co. eff of QD} = \boxed{\frac{Q_3 - Q_1}{Q_3 + Q_1}}$$ (N = no. of observations)

$Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ term $\qquad Q_1 = \left(\frac{n+1}{4}\right)^{th}$ term

(after arranging in asc. order)

Properties of Moments :

① $\boxed{\mu_1 = 0}$ ② $\boxed{\sigma^2 = \mu_2}$ ③ ($\mu_3 > 0$) (+vely skewed) ($\mu_3 < 0$) (-vely skewed) ($\mu_3 = 0$) (symm.)

3) μ_4 gives "kurtosis".

$$\text{skewness} = \beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad — ①$$

$$\text{kurtosis} = \beta_2 = \frac{\mu_4}{\mu_2^2} \quad — ②$$

$$\boxed{\sqrt{\beta_1} = \pm \frac{\mu_3}{\mu_2^{3/2}}}$$

4) $\mu_1 = \mu_1' - \mu_1' = 0$

$\mu_2 = \mu_2' - (\mu_1')^2$

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

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

(5) Mean $= A + \mu_1'$

(6) Coeff. of skewness (sk) $= \pm \dfrac{\mu_3}{\mu_2^{3/2}}$

Mean deviation: (ungrouped data) $M.D = \dfrac{\sum_{i=1}^{n} |x_i - a|}{n}$ (about a)

For grouped data - $M.D = \dfrac{\sum_{i=1}^{n} f_i |x_i - a|}{n}$ (about a)

Variance $(\sigma^2) = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ (ungrouped)

$= \dfrac{1}{n} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$ (grouped)

1) Calculate the first four moments of the following distribution about the mean. Also evaluate β_1 and β_2.

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

Sol). Let us calculate moments about $A = 4$; $d = x - A$

x	f	$d = x-4$	fd	fd^2	fd^3	fd^4
0	1	-4	-4	16	-64	256
1	8	-3	-24	72	-216	648
2	28	-2	-56	112	-224	448
3	56	-1	-56	56	-56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256

$\mu_1' = \frac{1}{n}\Sigma fd = 0$ $\qquad$ $\mu_3' = \frac{1}{n}\Sigma fd^3 = 0$

$\mu_2' = \frac{1}{n}\Sigma fd^2 = 2$ $\qquad$ $\mu_4' = \frac{1}{n}\Sigma fd^4 = 11$

Moments about actual mean are :

$\boxed{\mu_1 = 0}$; $\mu_2 = \mu_2' - (\mu_1')^2 = 2$

$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 = 0$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = 11$

Nature of Distribution :

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ (skewness)

Distribution is (symmetrical)

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75 \ (<3) \longrightarrow \text{kurtosis}$$

The distribution is (platykurtic)

Joint Probability Distributions:

In general, if X, Y, ..., P are n random variables, the probability distribution that defines their simultaneous behaviour is called a JPD. When 2 variables are there: bivariate prob. distribution.

Types: JPD for DRV, JPD for CRV.

* A function $f(x,y)$ is said to be joint pdf of two RV's X & Y, $(-\infty \le X \le \infty, -\infty \le Y \le \infty)$, if

(i) $f(x,y) \ge 0$ and ≤ 1 (ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$

Marginal probability density func's of X & Y:

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

Joint cumulative DF:

$$F(x,y) = P[X \le x, Y \le y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\,dy\,dx$$

* Two RV's are said to be independent if the joint density is equal to product of marginal densities: $\boxed{f(x,y) = f(x) * f(y)}$

The conditional PDF for X given Y is $f(x/y)$ is given as:

$$\boxed{f(x/y) = \frac{f(x,y)}{f_y(y)}}$$

Similarly,

$$\boxed{f(y/x) = \frac{f(x,y)}{f_x(x)}}$$

1) Show that JPF

$$f_{xy}(x,y) = \begin{cases} 9e^{-3x}e^{-3y}, & x,y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

is a valid density function. Also check whether the random variables are mutually independent or not?

ANS). Condition: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$$= \int_0^\infty \int_0^\infty 9e^{-3x} \cdot e^{-3y}\,dy\,dx$$

$$= 9\int_0^\infty e^{-3x} \left[\frac{-e^{-3y}}{3}\right]_0^\infty dx$$

$$= 9\int_0^\infty e^{-3x}\left(\frac{1}{3}\right)dx = 3\int_0^\infty e^{-3x}\,dx$$

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

2) Suppose (X,Y) have the joint density function defined by

$$f(x,y) = \begin{cases} c(2x+y) & 2<x<6;\ 0<y<5, \\ 0, & \text{otherwise} \end{cases}$$

then find c, and find $P(x>3, y>2)$, $P(x>3)$ and $P(x+y<4)$

(i) (ii) (iii)

ANS). It's a valid density func. if $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$$= c\int_{2}^{6}\int_{0}^{5}(2x+y)\,dy\,dx$$

$$= c\int_{2}^{6} 2x(5)+\frac{1}{2}(25)\,dx$$

$$= c\int_{2}^{6}\left(10x+\frac{25}{2}\right)dx = c\left[5(32)+\frac{25}{2}(4)\right]$$

$$\Rightarrow \cancel{c\left(105+75/2\right) = 1}$$

$$c(160+50) = 1 \Rightarrow \boxed{c = 1/210}$$

(i) $P(x>3, y>2) = \int_{3}^{6}\int_{2}^{5}\frac{1}{210}(2x+y)\,dy\,dx$

$$= \int_{3}^{6}\frac{1}{210}\left[2x(3)+\frac{1}{2}(21)\right]dx = \frac{1}{210}\int_{3}^{6}\left(6x+\frac{21}{2}\right)dx$$

$$= \frac{1}{210}\left[3(36-9)+\frac{21}{2}(3)\right] = \frac{1}{210}\left(81+\frac{63}{2}\right) = \frac{15}{28}$$

(ii) $P(x>3) = \int_{3}^{6}\int_{0}^{5}\frac{1}{210}(2x+y)\,dy\,dx$

$$= \int_{3}^{6}\frac{1}{210}\left[2x(5)+\frac{1}{2}(25)\right]dx = \frac{1}{210}\int_{3}^{6}\left(10x+\frac{25}{2}\right)dx$$

$$= \frac{1}{210}\left[5(36-9)+\frac{25}{2}(6-3)\right] = \frac{23}{28}$$

(iii) $P((x+y)<4)$

$$= \int_{0}^{2}\int_{2}^{4-y}\frac{1}{210}(2x+y)\,dx\,dy$$

$$= \frac{1}{210}\int_0^2 (x^2 + xy)_2^{4-y}\, dy$$

$$= \frac{1}{210}\int_0^2 (12 - 6y)\, dy = \frac{6}{210}\left(2y - \frac{y^2}{2}\right)_0^2 = \frac{2}{35}$$

NOTE : The random variable X & Y are said to be mutually independent if $\boxed{f_{xy}(x,y) = f_x(x) \cdot f_y(y)}$

Bivarient Random Variables are of two types :

(i) Two dimensional discrete RV.

(ii) Two dimensional continuous RV. (already wrote ←)

Joint probability mass function : If an ordered pair (x, y) is a 2-dimensional DRV, then $P(X=x, Y=y) = P(x,y)$ is called the joint probability mass func. if :

(i) $P(x,y) \geq 0$ (ii) $\sum_x \sum_y P(x,y) = 1$

x \ y	y_1	y_2	y_3	...	y_j	y_m	
X_1	P_{11}	P_{12}	P_{13}	...	P_{1j}	P_{1m}	$P_X(X_1)$
X_2	P_{21}	P_{22}	P_{23}	...	P_{2j}	P_{2m}	$P_X(X_2)$
⋮	⋮	⋮	⋮		⋮	⋮	⋮
X_i	P_{i1}	P_{i2}	P_{i3}	...	P_{ij}	P_{im}	$P_X(X_i)$
X_n	P_{n1}	P_{n2}	P_{n3}	...	P_{nj}	P_{nm}	$P_X(X_n)$
	$P_y(y_1)$	$P_y(y_2)$	$P_y(y_3)$	...	$P_y(y_j)$	$P_y(y_m)$	

Marginal probability mass functions :

$$\boxed{\begin{aligned} P_X(x) &= \sum_y P(x,y) \\ P_Y(y) &= \sum_x P(x,y) \end{aligned}}$$

$P(x,y) = P_x(x) \cdot P_y(y) \longrightarrow$ "independent"

---

① Joint distribution of two random variables X & Y are given as follows :

X \ Y	−4	2	7
1	1/8	1/4	1/8
5	1/4	1/8	1/8

Compute :- (a) $E(x)$ and $E(y)$ (b) $E(xy)$ (c) σ_x & σ_y

(d) $\mathrm{cov}(x,y)$ e) $\rho(x,y)$

ANS).

x_i	1	5
$f(x_i)$	1/2	1/2

y_i	−4	2	7
$f(y_i)$	3/8	3/8	1/4

(a) $E(x) = \sum x_i f(x_i) = 1 \times \frac{1}{2} + 5 \times \frac{1}{2} = 3$

$E(y) = \sum y_i f(y_i) = -4 \times \frac{3}{8} + 2 \times \frac{3}{8} + 7 \times \frac{1}{4} = 1$

(b) $E(xy) = \sum x_i y_j J_{ij}$

$E(xy) = 1(-4) \times \frac{1}{8} + 1(2) \times \frac{1}{4} + 1(7) \times \frac{1}{8} + 5(-4) \times \frac{1}{4}$

$+ 5(2) \times \frac{1}{8} + 5(7) \times \frac{1}{8}$

$$\boxed{E(xy) = 3/2}$$

(c) $\sigma_x^2 = E(x^2) - [E(x)]^2$ $\sigma_y^2 = E(y^2) - [E(y)]^2$

$\sigma_x^2 = \sum x_i^2 f(x_i) - [E(x)]^2$ $\sigma_y^2 = \sum y_i^2 f(y_i) - [E(y)]^2$

$\sigma_x^2 = (1)^2 \times \frac{1}{2} + (5)^2 \times \frac{1}{2} - (3)^2$ $\sigma_y^2 = (-4)^2 \times \frac{3}{8} + (2)^2 \times \frac{3}{8} + (7)^2 \times \frac{1}{4} -$

$\sigma_x^2 = 4$; $\sigma_y^2 = \frac{75}{4}$ $\Rightarrow$ $\boxed{\sigma_x = 2}$

$$\boxed{\sigma_y = \frac{\sqrt{75}}{2}}$$

d) $cov(xy) = E(xy) - E(x).E(y)$

$$= \frac{3}{2} - (3)(1) = \boxed{-3/2}$$

e) $\rho(x,y)$ (corelation) $= \dfrac{cov(x,y)}{\sigma_x . \sigma_y} = \dfrac{(-3/2)}{2 . \frac{\sqrt{75}}{2}} = \underline{\underline{-0.1732}}$

---

2) Given the following joint distribution of the RV x & y, find the corresponding marginal distribution. Also compute the covariance & corralation of the RV X & Y.

x \ y	1	3	9
2	1/8	1/24	1/12
4	1/4	1/4	0
6	1/8	1/24	1/12

ANS).

x_i	2	4	6
$f(x_i)$	1/4	1/2	1/4

y_i	1	3	9
$f(y_i)$	1/2	1/3	1/6

$cov(x,y) = E(xy) - E(x).E(y)$

$$E(xy) = \sum_{ij} x_i y_j J_{ij} = 2\times1\times\frac{1}{8} + 2\times3\times\frac{1}{24} + 2\times9\times\frac{1}{12} + 4\times\frac{1}{4}$$
$$+ \frac{12}{4} + 36\times0 + \frac{6}{8} + \frac{18}{24} + \frac{54}{12}$$

$E(xy) = 12$

~~$E(x) = E(x^2) - [E(x)]^2$~~ ~~$E(y) = E(y^2) - [E(y)]^2$~~

$E(x) = \sum x_i f(x_i)$ $E(y) = \sum y_i f(y_i)$

$E(x) = 4$ $E(y) = 3$

$cov(x, y) = 12 - (4 \cdot 3) = 0$

$$\rho(x, y) = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y} = 0$$

Correlation :-

Bivariate distribution – A data in which each item may assume the values of two variables is called a bivariate distribution.

Ex :- Data showing heights & weights of students of a class.

Data showing income & the expenditure of individuals.

The data showing the rainfall and crop yield.

Correlation – If a change in one variable causes a change in another variable, then the two vars are correlated.

Ex :- volume of a cube, i.e., $\boxed{V = l^3}$ is correlated.

Note : Two coins being tossed simultaneously is uncorrelated.

Methods of Studying Correlation :

To study how to measure the degree or extent of these interdependence b/w two vars, there are 3 methods

(i) Scatter diagram

(ii) Karl Pearson's coefficient of correlation

(iii) Spearman's rank correlation

~~Karl Pearson~~ coefficient of correlation : In correlation analysis, degree (or) strength of relationship b/w two variables say x & y is measured by a single number

Karl Pearson's [illegible] of correlation: It is defi[illegible]

$$\boxed{r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}}$$

where $x = X - \bar{X}$; $y = Y -$ [illegible]
the deviations measured [illegible]
respective means.

(1) Find coeff of corr. b/w the variables "X" and [illegible] for the following data:

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

sol).

Mean $(\bar{X}) = \frac{56}{8} = 7$ Mean $(\bar{Y}) = \frac{40}{8} = 5$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
1	1	-6	-4	24	36	16
3	2	-4	-3	12	16	9
4	4	-3	-1	3	9	1
6	4	-1	-1	1	1	1
8	5	1	0	0	1	0
9	7	2	2	4	4	4
11	8	4	3	12	16	9
14	9	7	4	28	49	16

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{84}{\sqrt{(132)(56)}} \Rightarrow \boxed{r = 0.977}$$

(2) 10 students got the following percentage of marks in economics & statistics:

Roll no	1	2	3	4	5	6	7	8	9	10
Eco.	78	36	98	25	75	82	90	62	65	39
Statistics	84	51	91	60	68	62	86	58	53	47

Calculate the coefficient of correlation.

ANS)

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
78	84	13	18	234	169	324
36	51	−29	−15	435	841	225
98	91	33	25	825	1089	625
25	60	−40	−6	240	1600	36
75	68	10	2	20	100	4
82	62	17	−4	−68	289	16
90	86	25	20	500	625	400
62	58	−3	−8	24	9	64
65	53	0	−13	0	0	169
39	47	−26	−19	494	676	361

Mean $(\bar{X}) = 65$ Mean $(\bar{Y}) = 66$

$\Sigma xy = 2704$ $\Sigma x^2 = 5398$ $\Sigma y^2 = 2224$

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{2704}{3464.845162} \Rightarrow \boxed{r = 0.7804}$$

Properties of Correlation Coefficient :

(i) $\boxed{-1 \leq r \leq 1}$; always lies b/w -1 and $+1$.

(ii) If $\boxed{r = 0}$, the variables are non-correlated (or) there is no linear correlation b/w the variables X & Y but a non-linear correlation may exist.

(iii) If $\boxed{r = 1}$, then we say that the variables are positively perfectly correlated. If $\boxed{r = -1}$, then we say that they are negatively perfectly correlated.

(iv) If $\boxed{r > -1 \;\&\; r < 0}$, then the variables are negatively partially correlated.

(iv) If $0 < r < 1$, then the variables are said to be positively partially correlated.

(v) It is independent of units.

* Although data is measured numerically (quantitative), in several cases, the data ~~comes~~ turns out to be non-numeric (qualitative).

Ex :- Appearance - beautiful, ugly.

* In such cases, the data is ranked according to that particular character, ~~the~~ instead of taking numeric measurements on them. Therefore, the usual Karl Pearson's coefficient (r) cannot be calculated.

* Instead of that, we use Spearman's rank correlation.

Spearman's Rank Correlation :

* Suppose $(x_1, x_2, \ldots, x_n)$ & $(y_1, y_2, \ldots, y_n)$ are ranks of two variables X and Y in the order of merit w.r.t some property, correlation b/w these n pairs of ranks is called rank correlation and denoted by r_s (or) r_{rank}.

* $$\boxed{1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}}$$

d_i = difference b/w ranks assigned to x_i & y_i.
n = no. of pairs of values (x, y) in the data.

① 10 recruits were subjected to a selection test to ascertain their suitability for a certain course of training. At the end of training, they were given a proficiency test. The marks secured by recruits in the selection test (X) and proficiency test (Y) are given as follows :

Calculate the rank corellation coefficient.

S.NO	1	2	3	4	5	6	7	8	9	10
X	10	15	12	17	13	16	24	14	22	20
Y	30	42	45	46	33	34	40	35	39	38

ANS).

X	Y	x = Rank of X	y = Rank of Y	$d_i = x - y$	d_i^2
10	30	10	10	0	0
15	42	6	3	3	9
12	45	9	2	7	49
17	46	4	1	3	9
13	33	8	9	−1	1
16	34	5	8	−3	9
24	40	1	4	−3	9
14	35	7	7	0	0
22	39	2	5	−3	9
20	38	3	6	−3	9

$$\therefore \; r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(104)}{10(99)}$$

$$\boxed{r_s = 0.369}$$

<u>Repeated Ranks</u>: If individuals are repeated, add the factor $\boxed{\frac{m(m^2-1)}{12}}$ (C.F) to $\sum d_i^2$, where "m" is the number of times an item is repeated. This correlation factor is to be added for each repeated value in both the "X" series and "Y" series.

(2) Obtain the [illegible] correlation for the following data:

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Ans).

X	Y	x	y	$d_i = x - y$	d_i^2
68	62	4	5	−1	1
64	58	6	7	−1	1
75	68	2.5	3.5	−1	1
50	45	9	10	−1	1
64	81	6	1	5	25
80	60	1	6	−5	25
75	68	2.5	3.5	−1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16

For X-series, the factor $\left[\frac{m(m^2-1)}{12}\right]$ $= \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} = 5/2$

For Y-series, the factor $= \frac{2(2^2-1)}{12} = \frac{1}{2}$

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \left(\begin{array}{r}\sum d_i^2 = 72 \\ n = 10\end{array}\right)$$

$$r_s = 1 - \frac{6\left[72 + \frac{5}{2} + \frac{1}{2}\right]}{10(99)} = 1 - \frac{(75)\times 6}{990}$$

$$\boxed{r_s = 0.5454}$$

(3) A sample of 12 fathers and their eldest sons gave the following data about their height (in inches). Calc. the coefficient of rank correlation:

Fathers | 65 63 67 64 68 62 70 66 68

Sons | 66 65 68 65 69 66 68 ~~71~~ 67 68 70

Fathers Heights	65	63	67	64	68	62	70	66	68	67	69	71
	68	66	68	65	69	66	68	65	71	67	68	70

x	y	rank (x)	rank (y)	$d_i = x - y$	$(d_i)^2$
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	−1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	−3.5	12.25
66	65	8	11.5	3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	−1.5	2.25
69	68	3	5.5	−2.5	6.25
71	70	1	2	−1	1

(b) For X-series, $\frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12} = \frac{1}{2} + \frac{1}{2} = 1$

For Y-series, $\frac{4(4^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12} = 5 + \frac{1}{2} + \frac{1}{2} = ($

(c) $r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = \frac{1 - 6(72.5+1+6)}{12(144-1)}$

$r_s = 1 - \frac{795}{286}$

$r_s = 0.722$

1

(*) <u>Regression</u> — Means "stepping back" towards the average. Regression analysis is mathematical measure of the avg. relationship b/w 2 or more variables in terms of original units of data. In regression analysis, there are 2 types of variables. The variable whose value is to be predicted is called dependent variable and variable used for prediction is called independent variable.

<u>Lines of Regression</u> : If the variables in a bivariate distribut" are related, we will find that the points the scatter diagram will cluster around some curve called the curve of regression. If the curve is a straight line, it's called line of regression, and is said to "linear regression" within variables, otherwise regression is said to be curve linear.

① Line of regression y on x passing through the point $(\bar{x}, \bar{y})$ is $(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}[x-\bar{x}]$

↓ (means)

② LOR x on y passing through the point $(\bar{x}, \bar{y})$

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$r \times \frac{\sigma_y}{\sigma_x} \longrightarrow$ regression coefficient of y on x

$r \times \frac{\sigma_x}{\sigma_y} \longrightarrow$ " " x " y

① Two regression lines are having their means and SD as 31.6, 38 and 3.72, 6.31 respectively $r = -0.36$. Find two regression lines.

Sol). Regression line of y on x is,

$$\rightarrow (y-38) = -0.36 \times \frac{6.31}{3.72}(x-31.6)$$

$$y + (0.611)x + 57.31 = 0$$

Regression line of x on y is,

$$\rightarrow (x-31.6) = -0.36 \times \frac{3.72}{6.31}(y-38)$$

$$x + (0.213)y = 39.7$$

② Equations of two regression lines obtained in a correlation analysis are $3x + 12y = 19$, $3y + 9x = 46$. Find (i) coefficient of correlation (ii) mean of x & y.

Sol). $12y = 19 - 3x \Rightarrow 12y = -(3x - 19)$

$$y = -\frac{3}{12}(x - 19/3)$$

$$y = -\frac{1}{4}\left(x - \frac{19}{3}\right)$$

$$r \times \frac{\sigma_y}{\sigma_x} = -\frac{1}{4} \quad \text{—— ①}$$

$$3y + 9x = 46$$

$$9x = -(3y - 46)$$

$$x = -\frac{1}{3}\left(y - \frac{46}{3}\right)$$

$$r \times \frac{\sigma_x}{\sigma_y} = -\frac{1}{3} \quad \text{—— ②}$$

Divide ① & ②;

$$\left(\frac{\sigma_y}{\sigma_x}\right)^2 = \frac{3}{4}$$

$$\frac{\sigma_y}{\sigma_x} = \pm\frac{\sqrt{3}}{2}$$

$$\rightarrow \quad r \times \frac{\sigma_y}{\sigma_x} = -\frac{1}{4} \qquad (\text{eqn} — ①)$$

$$r \times \frac{\sqrt{3}}{2} = -\frac{1}{4}$$

$$\boxed{r = \pm 0.2886}$$

(ii) Point of intersection of the two lines of regression is mean value of x & y

$$3 \times (3x + 12y = 19)$$

$$9x + 3y = 46$$

$$\rightarrow \quad 9x + 36y = 57$$

$$\underset{(-)}{9x} + \underset{(-)}{3y} = \underset{(-)}{46}$$

$$33y = 11$$

$$\boxed{\bar{x} = 5; \quad \bar{y} = \frac{1}{3}}$$

---

③ Calculate coeff. of corell. and the two lines of regression for the following data:

Price (x) :	14	16	17	18	19	20	21	22	23
Demand (y) :	84	78	70	75	66	67	62	58	60

$\bar{X} = \frac{170}{9}$ $\quad$ $\bar{Y} = \frac{620}{9}$

$\bar{X} = 18.89$ (19) $\quad$ $\bar{Y} = 68.89$ (69)

When actual mean is not a whole number but a fraction, the calc. by direct met. will involve lot of time. To avoid such type of calcs, we take assumed mean method.

$$\therefore \boxed{r = \frac{\Sigma xy - \frac{\Sigma x . \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}}$$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
14	84	-5	15	-75	25	225
16	78	-3	9	-27	9	81
17	70	-2	1	-2	4	1
18	75	-1	6	-6	1	36
19	66	0	-3	0	0	9
20	67	1	-2	-2	1	4
21	62	2	-7	-14	4	49
22	58	3	-11	-33	9	121
23	60	4	-9	-36	16	81

$\Sigma xy = -195$; $\Sigma x = -1$; $\Sigma y = -1$; $\Sigma x^2 = 69$; $\Sigma y^2 = 607$

$$r = \frac{-195 - \frac{(1)}{9}}{\sqrt{69 - \frac{1}{9}}\sqrt{607 - \frac{1}{9}}} = \frac{(-1756/9)}{\sqrt{\frac{620}{9}} . \sqrt{\frac{5462}{9}}} = \frac{-1756}{\sqrt{620} . \sqrt{5462}}$$

$$r = \frac{-1756}{1840.228} \Rightarrow \boxed{r = -0.9542}$$

Regression line of y on x :

$$y - \bar{y} = r \times \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$(\bar{x}, \bar{y}$ → assumed mean

$$\sigma_x = \sqrt{\frac{(\Sigma x)^2}{n}} \; ; \; \sigma_y = \sqrt{\frac{(\Sigma y)^2}{n}}$$

$$\sigma_x = \frac{1}{3} \; ; \; \sigma_y = \frac{1}{3}$$

$$(y - 69) = (-0.95) \times 1 (x - 19)$$

$$(y - 69) = -0.95x + 18.05$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n}} \qquad \sigma_y = \sqrt{\frac{\Sigma y^2}{n}}$$

$$= \sqrt{\frac{69}{9}} = 2.76 \qquad = \sqrt{\frac{607}{9}} = 8.21$$

$$y - 69 = -0.95 \times \frac{8.21}{2.76}(x - 19)$$

$$-(y - 69) = 2.825(x - 19)$$

$$2.825(x) - 69 + y - 53.675 = 0$$

$$\boxed{2.825(x) + (y) + 122.6 = 0}$$

Regression line of x on y :

$$x - \bar{x} = r \times \frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$(\bar{x}, \bar{y}$ → assumed mean

$$x - 19 = -0.95 \times \frac{2.76}{8.21}(y - 69)$$

Regression Equation: It is an algebraic equation of the regression line. The standard form of the regression equation is: $\boxed{Y = a + bX}$ (a, b are constants)

→ "a" is called Y-intercept.

→ "b" is called slope of regression line.

The values of "a" and "b" can be obtained with the help of normal equations.

Normal Equations:

The normal equations of the regression line of y on x, i.e., $\boxed{Y = a + bX}$ are $\boxed{\Sigma Y = na + b\Sigma X}$

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

① Determine the equation of a straight line which best fits for the following data:

X	10	12	13	16	17	20	25
Y	10	22	24	27	29	33	37

Ans).
Let the eqn of straight line which fits best to the given data is: $\boxed{Y = a + bX}$. Then corresponding normal eqn's are: $(\Sigma Y = na + b\Sigma X)$

$(\Sigma xy = a\Sigma X + b\Sigma x^2)$

X	Y	xy	x^2
10	10	100	100
12	22	264	144
13	24	312	169
16	27	432	256
17	29	493	289
20	33	660	400
25	37	925	625

$\Sigma x^2 = 1983$

$\Sigma xy = 3186$

$\Sigma X = 113$

$\Sigma y = 182$

Normal Equations

$$182 = 7a + 113b$$
$$3186 = 113a + 1983b$$

$a = 0.798$; $b = 1.56$

$\Rightarrow \therefore Y = a + bX$ (line)

$$\boxed{Y = 0.798 + X(1.56)}$$

Angle b/w 2 regression lines:

Let the regression lines of Y on X and X on Y

are $\left[Y - \bar{Y} = r \times \frac{\sigma_y}{\sigma_x}(X - \bar{X})\right]$ & $\left[X - \bar{X} = r \times \frac{\sigma_x}{\sigma_y}(Y - \bar{Y})\right]$.

The angle b/w these lines is given by :

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

(i) If $r = 0$; $\theta = \pi/2$; lines are perpendicular to each other. (variables are uncorrelated).

(ii) If $r = \pm 1$; $\theta = 0$; lines are parallel to each other, or they coincide.

(Q). If "θ" is the angle b/w two regression lines, and σ_y is twice of σ_x, $r = 0.25$, then find $\tan\theta$.

Ans).

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

$$\tan\theta = \left[\frac{1 - (0.25)^2}{0.25}\right]\left[\frac{\sigma_x(2\sigma_x)}{\sigma_x^2 + (2\sigma_x)^2}\right]$$

$$\tan\theta = \frac{15}{4}\left(\frac{2\cancel{\sigma_x^2}}{5\cancel{\sigma_x^2}}\right)$$

$$\boxed{\tan\theta = 1.5}$$

Q). If $\sigma_X = \sigma_Y = \sigma$ and angle b/w regression lines is $\tan^{-1}\left(\frac{4}{3}\right)$, then find "r".

Ans).

$$\frac{4}{3} = \frac{1-r^2}{r}\left\{\frac{\sigma^2}{2\sigma^2}\right\}$$

$$\frac{4}{3} = \frac{1-r^2}{2r}$$

$$8r = 3-3r^2$$

$$3r^2+8r-3=0$$

$$r=-3 \;/\; r=1/3$$

But $-1 \le r \le 1 \Rightarrow \boxed{r = 1/3}$

Multiple Regression : 2 or more independent variables are used to estimate the values of a dependent variable. The regression equation of Y on X_1 and X_2 is : $Y = a + bX_1 + cX_2$. Corresponding normal eqn's :

(i) $\Sigma Y = na + b\Sigma X_1 + c\Sigma X_2$

(ii) $\Sigma X_1 Y = a\Sigma X_1 + b\Sigma X_1^2 + c\Sigma X_1 X_2$

(iii) $\Sigma X_2 Y = a\Sigma X_2 + b\Sigma X_1 X_2 + c\Sigma X_2^2$

---

① Find Y when $X_1 = 10$; $X_2 = 6$, from the least squares regression equation of Y on X_1 and X_2 for the following data:

X_1	3	5	6	8	12	14
X_2	16	10	7	4	3	2
Y	90	72	54	42	30	12

Sol). Regression line : $Y = a + bX_1 + cX_2$

Normal eqn's are :

(i) $\Sigma Y = na + b\Sigma X_1 + c\Sigma X_2$

(ii) $\Sigma X_1 Y = a\Sigma X_1 + c\Sigma X_1 X_2 + b\Sigma X_1^2$

(iii) $\Sigma X_2 Y = a\Sigma X_2 + b\Sigma X_1 X_2 + c\Sigma X_2^2$

X_1	X_2	Y	X_1Y	X_1^2	X_1X_2	X_2Y	X_2^2
3	16	90	270	9	48	1440	256
5	10	72	360	25	50	720	100
6	7	54	324	36	42	378	49
8	4	42	336	64	32	168	16
12	3	30	360	144	36	90	9
14	2	12	168	196	28	24	4

$\Sigma X_1 = 48$; $\Sigma X_2 = 42$; $\Sigma Y = 300$; $\Sigma X_1Y = 1818$; $\Sigma X_1^2 = 4$

$\Sigma X_1X_2 = 236$; $\Sigma X_2Y = 2820$; $\Sigma X_2^2 = 434$

$$\left.\begin{aligned} &(i)\ 300 = 6a + 48b + 42c \\ &(ii)\ 1818 = 48a + 474b + 236c \\ &(iii)\ 2820 = 42a + 236b + 434c \end{aligned}\right\}\text{Solving}$$

$a = 61.4$; $b = -3.65$; $c = 2.54$

$$\boxed{Y = 61.4 + (-3.65)X_1 + (2.54)X_2}$$

$X_1 = 10,\ X_2 = 6 \Rightarrow \underline{Y(10,6) = 40.14}$

---

② Fit a multiple linear regression line eq^n to the following data and predict Y when $X_1 = 50$; $X_2 = 8$.

Y	64	71	53	67	55	58	77	57	56	51	76	6
X_1	57	59	49	62	51	50	55	48	52	42	61	5
X_2	8	10	6	11	8	7	10	9	10	6	12	

Sol).

Let the line be $Y = a + bX_1 + cX_2$

X_1	X_2	Y	X_1Y	X_1X_2	X_1^2	X_2Y	X_2^2
57	8	64	3648	456	3249	512	64
57	10	71	4189	590	3481	710	100
49	6	53	2597	294	2401	318	36
62	11	67	4154	682	3844	737	121
51	8	55	2805	408	2601	440	64
50	7	58	2900	350	2500	406	49
55	10	77	4235	550	3025	770	100
48	9	57	2736	432	2304	513	81
52	10	56	2912	520	2704	560	100
42	6	51	2142	252	1764	306	36
61	12	76	4636	732	3721	912	144
57	9	68	3876	513	3249	612	81

$\Sigma X_1 = 643$; $\Sigma X_2 = 106$; $\Sigma Y = 753$; $\Sigma X_1 Y = 40830$

$\Sigma X_1 X_2 = 5779$; $\Sigma X_1^2 = 34843$; $\Sigma X_2 Y = 6796$; $\Sigma X_2^2 = 976$

(i) $753 = 12a + 643b + 106c$;

(ii) $40830 = 643a + 34843b + 5779c$;

(iii) $6796 = 106a + 5779b + 976c$;

$\boxed{a = 3.65}$ $\boxed{b = 0.854}$ $\boxed{c = 1.506}$

Regression line : $Y = 3.65 + (0.854)X_1 + (1.506)X_2$

when $X_1 = 50$, $X_2 = 8 \Rightarrow \boxed{Y = 58.39}$

(3) The following is the data on the number of twists required to break a certain kind of forged alloy bar & the percentages of two alloying elements present in the metal:

(8) No. of traits (Y): 41 49 69 65 40 50 58 57 31 36 44 57 19

% of element A (X_1): 1 2 3 4 1 2 3 4 1 2 3 4 1

% of element B (X_2): 5 5 5 5 10 10 10 10 15 15 15 15 20

Find the least squares regression line.

Ans).

X_1	X_2	Y	X_1Y	X_1^2	X_2^2	X_1X_2	X_2Y
1	5	41	41	1	25	5	205
2	5	49	98	4	25	10	245
3	5	69	207	9	25	15	345
4	5	65	260	16	25	20	325
1	10	40	40	1	100	10	400
2	10	50	100	4	100	20	500
3	10	58	174	9	100	30	580
4	10	57	228	16	100	40	570
1	15	31	31	1	225	15	465
2	15	36	72	4	225	30	540
3	15	44	132	9	225	45	660
4	15	57	228	16	225	60	855
1	20	19	19	1	400	20	380
2	20	31	62	4	400	40	620
3	20	33	99	9	400	60	660
4	20	43	172	16	400	80	860

$\Sigma X_1 = 40$; $\Sigma X_2 = 200$; $\Sigma Y = 723$; $\Sigma X_1Y = 1963$

$\Sigma X_1^2 = 120$; $\Sigma X_2^2 = 3{,}000$; $\Sigma X_1X_2 = 500$; $\Sigma X_2Y = 8210$

Normal equations are :

(i) $\Sigma Y = na + b\,\Sigma X_1 + c\,\Sigma X_2 \Rightarrow 723 = 16a + 40$

(ii) $1963 = 40a + 120b + 500c$;

(iii) $8210 = 200a + 500b + 3000c$;

Solving above eq^n's; $\boxed{a = 46.36}$; $\boxed{b = 7.78}$; $\boxed{c = -}$

$\therefore$ Regression line is :

$$\boxed{Y = 46.36 + (x_1)7.78 \cancel{+} - 1.65(x_2)}$$

(Q). In a paired data ..., with $n = 25$, $\sum X = 127$, $\sum Y = 100$, $\sum X^2 = 760$, $\sum Y^2 = 449$, $\sum XY = 500$, it was found later that two pairs of correct values $(X, Y) = (8, 12)$ and $(6, 8)$ were copied down as $(8, 14)$ and $(8, 6)$. Determine correlation coefficient for the correct data.

10

4

20

ANS).

$$\boxed{r = \frac{\left(\frac{1}{n}\sum XY\right) - \bar{X}\bar{Y}}{\sqrt{\frac{1}{n}\sum X^2 - (\bar{X})^2}\sqrt{\frac{1}{n}\sum Y^2 - (\bar{Y})^2}}}$$

Karl Pearson's product moment formula.

In order to get correct data, subtract incorrect & add correct.

$$\sum X = 127 - 8 - 8 + 8 + 6 = 125$$

$$\sum Y = 100 + 12 + 8 - 14 - 6 = 100$$

$$\sum X^2 = 760 - 64 - 64 + 64 + 36 = 732$$

$$\sum Y^2 = 449 - 196 - 36 + 144 + 64 = 425$$

$$\sum XY = 500 - 112 - 48 + 96 + 48 = 484$$

$$\bar{X} = \frac{\sum X}{n} = \frac{125}{5} = 5$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{100}{25} = 4$$

$$\therefore r = \frac{\frac{1}{25}(484) - 20}{\sqrt{\frac{1}{25}(732) - 25}\sqrt{\frac{1}{25}(425) - 16}} \Rightarrow \boxed{r = -0.309}$$

+2000 **Note:** Coefficient of correlation "r" is the G.M between regression coefficients.

65

Unit-03 :

Binomial Distribution : Suppose that we have an experiment such as tossing a coin or a die repeatedly or choosing a marble from an urn repeatedly. Each toss/selection is called a trial. In any single trial, there'll be a probability associated with a particular event such as head on the coin, 4 on a die or a selection of a red marble. In some cases, the probability will not change from one trial to the next. Such trials are said to be independent, and are often called Bernoulli Trials.

Result : If the prob. of success of an event in a single trial is P, so that prob. of failure $(Q) = 1-P$, then P of exactly "x" successes of event in "n" independent trials is given by :

$$P[x; n, P] = {}^nC_x P^x q^{n-x} \qquad (x = 0, 1, 2, 3, \dots n)$$

(or)

$$P[x] \text{ (or) } P[X = x]$$

Mean of Binomial Distribution (***) :

$$\text{Mean} = \mu = \sum_{x=0}^{n} x P(x) = \sum_{x=0}^{n} x\left[{}^nC_x P^x q^{n-x}\right]$$

$$\mu = npq^{n-1} + 2\,{}^nC_2 p^2 q^{n-2} + \dots + np^n$$

$$= npq^{n-1} + \cancel{2} \cdot \frac{n!}{(n-2)!\,\cancel{2}} p^2 q^{n-2} + \dots + np^n$$

$$= npq^{n-1} + n(n-1)p^2 q^{n-2} + \dots + np^n$$

$$= np\left[q^{n-1} + (n-1)pq^{n-2} + \dots + p^{n-1}\right]$$

$$= np\left[{}^{n-1}C_0 q^{n-1} + {}^{n-1}C_1 pq^{n-2} + \dots + {}^{n-1}C_{n-1} p^{n-1}\right]$$

$\mu = np(1)^{n-1}$ $\qquad [\because q+p=1]$

$\boxed{\mu = np}$

Variance of Binomial Distribution (***):

$$\sigma^2 = E(x^2) - [E(x)]^2$$

$$E(x^2) = \sum_{x=0}^{n} x^2 p(x) = \sum_{x=0}^{n} [x(x-1)+x]\, {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\, {}^nC_x p^x q^{n-x} + \underbrace{\sum_{x=0}^{n} x\, {}^nC_x p^x q^{n-x}}$$

$$= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \underbrace{np}$$

$$= \sum_{x=2}^{n} \frac{\cancel{x(x-1)}\, n(n-1)(n-2)! \times p^2 p^{x-2} \cdot q^{(n-2)-(x-2)}}{\cancel{x(x-1)}(x-2)!\,[(n-2)-(x-2)]!} + np$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!\, p^{x-2} \cdot q^{(n-2)-(x-2)}}{(x-2)!\,[(n-2)-(x-2)]!} + np$$

$$= n(n-1)p^2 \left[\sum_{x=2}^{n} {}^{(n-2)}C_{(x-2)}\, p^{x-2} \cdot q^{(n-2)-(x-2)}\right] + np$$

$$= n(n-1)p^2 [\cancel{q+p}]^{n-2} + np$$

$$= n^2p^2 - np^2 + np = np[np + (1-p)] = np[np+q]$$

$$= n^2p^2 + npq$$

$$\Rightarrow \sigma^2 = E(x^2) - [E(x)]^2 = n^2p^2 + npq - (np)^2$$

$$\sigma^2 = \cancel{n^2p^2} + npq - \cancel{n^2p^2}$$

$$\boxed{\sigma^2 = npq}$$

(Q). 10 coins are tossed ... find the probability of getting atleast 7 heads.

sol). $n = 10,\ p = \frac{1}{2},\ q = 1-p \Rightarrow q = \frac{1}{2}$

$$P[X \geq 7] = P[x=7] + P[x=8] + P[x=9] + P[x=10]$$

$$P[X \geq 7] = {}^{10}C_7 \left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + {}^{10}C_8 \left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + {}^{10}C_9 \left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right) + {}^{10}C_{10} \left(\frac{1}{2}\right)^{10}$$

$$= \frac{1}{2^{10}}\left[{}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right]$$

$$= \frac{120+45+10+1}{2^{10}} = \boxed{0.172}$$

(Q). Team A has probability $\frac{2}{3}$ of winning, whenever it plays the game. If A plays 4 games, find the probability that A wins (i) exactly 2 games, (ii) atleast 1 game, (iii) more than half of the games.

sol). $n=4,\ p = \frac{2}{3},\ q = \frac{1}{3}$

(i) $P[X=2] = {}^4C_2 \left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 = 6 \times \frac{4}{9} \times \frac{1}{9} = \underline{0.296}$

(ii) $P[X \geq 1] = 1 - P[X<1] = 1 - P[X=0]$

$$= 1 - {}^4C_0 \left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^4$$

$$= \underline{0.987}$$

(iii) $P[X \geq 3]$

$$= {}^4C_3 \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^1 + {}^4C_4 \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^0$$

$$= \left(\frac{4 \times 8}{27} \times \frac{1}{3}\right) + \left(\frac{16}{81}\right)$$

$$= \underline{\underline{0.592}}$$

(Q). A man hits a target with probability 1/4.

(i) Determine prob. of hitting atleast twice when he fires 7 times.

(ii) How many times must he fire so that the prob. of him hitting the target atleast once is $> \frac{2}{3}$?

Sol). $p = \frac{1}{4}$; $q = 1 - \frac{1}{4} = \frac{3}{4}$

(i) $n = 7$, $x \geq 2$

$$P[X \geq 2] = 1 - P[X < 2] = 1 - (P[X=1] + P[X=0])$$

$$P[X \geq 2] = 1 - \left[{}^7C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^6 + {}^7C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^7\right)$$

$$= 1 - \left[7\left(\frac{1}{4}\right)\left(\frac{3^6}{4^6}\right) + \frac{3^7}{4^7}\right]$$

$$= \underline{\underline{0.555}}$$

(ii) $n = ?$, $P[X \geq 1] > 2/3$

$$P[X \geq 1] = 1 - P[X < 1] = 1 - [P[X=0])$$

$$= 1 - {}^nC_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^n$$

$$= 1 - \frac{3^n}{4^n} > \frac{2}{3}$$

$$\frac{3^n}{4^n} - 1 < -\frac{2}{3}$$

$$\frac{3^n}{4^n} < \frac{1}{3}$$

$$\boxed{n = 4}$$

(Q). Determine the binomial distribution for which the mean is 4 and variance is 3. Also find its mode.

Sol). $np = 4$ $\qquad npq = 3$

$$\boxed{q = 3/4}$$

$p = 1 - q$

$\boxed{p = \frac{1}{4}} \Rightarrow \boxed{n = 16}$

$\therefore$ Binomial distribution is $(q+p)^n = \left(\frac{3}{4} + \frac{1}{4}\right)^{16}$

To find mode:

$(n+1)p = (16+1)\frac{1}{4} = 4.25$

$\therefore \boxed{\text{mode} = 4}$

***

If $(n+1)p$ is decim[al]
mode = $[p]$, else i[f]
$(n+1)p$ is integer,
mode = int & (int-1)

***

(Q). The incidence of occupational disease in an indus[try] is such that the workers have a 20% chance of suffering from it. What is the probability that out of 6 workers chosen at random, 4 (or) more will suffer from dise[ase]

Sol). $p = 20\% = 0.2$; $q = 0.8$; $n = 6$

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

$$= {}^6C_4 (0.2)^4 (0.8)^2 + {}^6C_5 (0.2)^5 (0.8)^1 + {}^6C_6 (0.2)^6 (\ldots$$

$P[X \geq 4] = $

$$= \frac{240}{15625} + \frac{48}{31250} + \frac{1}{15625}$$

$$= \underline{\underline{0.01696}}$$

(Q). A fair die is tossed 7 times, determine the probability that a 5 (or) 6 appears (i) exactly 3 times (ii) never occurs

Sol). (i) $P[X=3] = {}^7C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^4$

$$p = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

(5) (6)

$$q = 1 - p$$

$$q = 1 - \frac{1}{3}$$

$$q = \frac{2}{3}$$

$$P[X=3] = \frac{35 \times 16}{2187}$$

$$P[X=3] = \underline{0.256}$$

(ii) $P[X=0] = {}^7C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^7$

$$= 1 \times 1 \times \left(\frac{2}{3}\right)^7$$

$$= \underline{\underline{0.0585}}$$

(Q). In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain atleast 3 defective parts?

Sol). Given that; $np = 2$

$$20p = 2$$

$$p = 0.1$$

$$q = 1 - p = 0.9$$

$$P[X \geq 3] = 1 - P[X < 3]$$

$$= 1 - P[X=0] - P[X=1] - P[X=2]$$

$$= 1 - \left[{}^{20}C_0 (0.1)^0 (0.9)^{20} + {}^{20}C_1 (0.1)^1 (0.9)^{19} + {}^{20}C_2 (0.1)^2 \ldots\right.$$

$$= 1 - [0.1215 + 20(0.1)(0.1350) + 190(0.01) \ldots$$

$$= \cancel{1 - [0.1215 + 0.7748 + 0.285]}$$

$$= 1 - [0.1215 + 0.27 + 0.285]$$

$$= \underline{\underline{0.3235}}$$

Expected Number of samples $= 1000 \times 0.323$

$\simeq$ 323 samples

(Q). In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads & 4 tails?

Sol).

Recurrence Relation for Binomial Expression :

we know that : $b(x\,;\,n,p) = {}^nC_x\, p^x q^{n-x}$

$$b(x+1\,;\,n,p) = {}^nC_{x+1}\, p^{x+1} q^{n-x-1}$$

$$\frac{b(x+1\,;\,n,p)}{b(x\,;\,n,p)} = \frac{{}^nC_{x+1}\, p^{x+1} q^{n-x-1}}{{}^nC_x\, p^x q^{n-x}}$$

$$= \frac{\dfrac{n!}{(x+1)!\,(n-x-1)!}}{\dfrac{n!}{x!\,(n-x)!}}\;\frac{p}{q}$$

$$= \frac{\cancel{x!}\,(n-x)\,\cancel{(n-x-1)!}}{(x+1)\,\cancel{x!}\,\cancel{(n-x-1)!}} \cdot \frac{p}{q}$$

$$\therefore \boxed{b(x+1\,;\,n,p) = \left\{\frac{n-x}{x+1}\cdot\frac{p}{q}\right\} b(x\,;\,n,p)}$$

(Q). Fit a binomial distribution for the following data:

x :	0	1	2	3	4	5	6
f :	5	18	28	12	7	6	4

Sol). n = no. of trials = 6

$N = \Sigma f = 80$

$$\text{Mean}(\mu) = \frac{\Sigma fx}{N}$$

$$\mu = \frac{0+18+56+36+28+30+24}{80}$$

$$\boxed{\mu = 2.4}$$

$np = 2.4$

$6p = 2.4$

$\boxed{p = 0.4}$

$\boxed{q = 0.6}$ $(1-p)$

x	$P(x)$	Expected frequency $N \times P(x)$
0	$P(0) = {}^6C_0 (0.4)^0 (0.6)^6 = 0.0466$	$0.0466 \times 80 = 3.72 \approx$
1	$P(1) = {}^6C_1 (0.4)^1 (0.6)^5 = 0.1866$	$0.1866 \times 80 = 15$
2	$P(2) = {}^6C_2 (0.4)^2 (0.6)^4 = 0.31104$	$0.31104 \times 80 = 25$
3	$P(3) = {}^6C_3 (0.4)^3 (0.6)^3 = 0.27648$	$0.27648 \times 80 \approx 22$
4	$P(4) = {}^6C_4 (0.4)^4 (0.6)^2 = 0.13824$	$0.13824 \times 80 \approx 11$
5	$P(5) = {}^6C_5 (0.4)^5 (0.6)^1 = 0.0368$	$0.036 \times 80 \approx 3$
6	$P(6) = {}^6C_6 (0.4)^6 (0.6)^0 = 0.004096$	$0.004096 \times 80 \approx 0$

(Q). 7 coins are tossed and number of heads noted. The experiment is repeated 128 times with the following data:

No. of heads	0	1	2	3	4	5	6
Frequencies	7	6	19	35	30	23	7

Fit a binomial distribution assuming: (i) coin is unbiased. (ii) nature of coin is unknown.

Sol).

Poisson Distribution : [n >>>, p <<<]

* A RV "X" is said to follow Poisson distribution if its probability DF is given by: ~~$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$~~

$$f(x,\lambda) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{elsewhere} \end{cases}$$

($\lambda \rightarrow$ non zero, finite constant)

Variance of Poisson Distribution :

$$\sigma^2 = E(x^2) - [E(x)]^2$$

$$E(x^2) = \sum_{x=0}^{\infty} x^2 f(x,\lambda) = \sum_{x=0}^{\infty} [x(x-1) + x] f(x,\lambda)$$

$$= \sum_{x=0}^{\infty} x f(x,\lambda) + \sum_{x=0}^{\infty} x(x-1) f(x,\lambda)$$

$$= \lambda + \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \lambda + \sum_{x=0}^{\infty} \frac{\cancel{x(x-1)}\, \lambda^x e^{-\lambda}}{\cancel{x(x-1)}(x-2)!}$$

$$= \lambda + \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-2)!}$$

$$= \lambda + \lambda^{+2} \sum_{x=0}^{\infty} \frac{\lambda^{x-2} . e^{-\lambda}}{(x-2)!}$$

$$= \lambda + e^{\lambda}(e^{-\lambda}) . \lambda^{+2}$$

$$= \lambda + \lambda^{+2}$$

$$\left[\because e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \qquad e^{\lambda} = \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right]$$

$$\sigma^2 = E(x^2) - [E(x)]^2 = \lambda + \lambda^2 - \lambda^2$$

$$\boxed{\sigma^2 = \lambda}$$

Mean of ...

$$\text{Mean} = \mu = \sum_{x=0}^{\infty} x f(x,\lambda) = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \sum_{x=1}^{\infty} \frac{\lambda^x \cdot e^{-\lambda}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \qquad \left[\because e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \qquad e^{\lambda} = \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right]$$

$$= \lambda e^{-\lambda} \cdot (e^{\lambda})$$

$$\boxed{\mu = \lambda}$$

(Q). If the probability that an individual suffers a bad reaction due to a certain injection is 0.001, then determine the probability that out of 2000 individuals, (i) exactly 3, (ii) more than 2 individuals, will suffer a bad reaction?

Sol). $n = 2000$, $p = 0.001$

we know, $\text{mean}(x) = \lambda$

$$np = \lambda$$

$$\boxed{\lambda = 2}$$

(i) $P[x=3] = \dfrac{2^3 e^{-2}}{3!} = \underline{\underline{0.18044}}$

(ii) $P[x>2] = 1 - P[x=0] - P[x=1] - P[x=2]$

$$= 1 - \left[\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!}\right]$$

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

$$= 1 - 5e^{-2} = \underline{\underline{0.323}}$$

(Q). The probability of getting no misprint in a page of a book is e^{-4}. Determine the probability that a page of a book contains more than 2 misprints. (0.761)

Q). In a factory producing ~~...~~ blades, the prob. of any blade being defective is 0.002. If the blades are supplied in packets of 10, then determine the number of packets in a consignment of 10,000 packets containing (i) no defective (ii) 1 defective (iii) 2 defects

ANS). Let p = probability of a blade being defective.

$$p = 0.002$$

$$n = 10$$

$$\lambda = np = 10(0.002) = 0.02 \qquad \left(\because \lambda = \mu,\ \mu = np\right)$$

(i) $P[X=0] = \dfrac{\lambda^0 e^{-\lambda}}{0!} = \dfrac{e^{-0.02}}{1} = 0.98020$

Number of packets $= 0.98020 \times 10000 \approx$ 9802 packets

(ii) $P[X=1] = \dfrac{\lambda^1 e^{-\lambda}}{1!} = (0.02)e^{-0.02} = 0.01960$

(iii) $P[x=2] = \frac{\lambda^2 e^{-\lambda}}{2!} = 0.00020$

Number of packets $= 0.00020 \times 10000$

$\approx$ 2 packets

(Q). In a Poisson's distribution, ~~probability~~ $P[x=0] = 2P[x=1]$. Then find $P[x=2]$

Ans).
$$\frac{\lambda^0 e^{-\lambda}}{0!} = 2 \times \frac{\lambda^1 e^{-\lambda}}{1!}$$

$$1 = 2\lambda$$

$$\boxed{\lambda = 0.5}$$

$$P[x=2] = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{\frac{1}{4} \times e^{-0.5}}{2} = 0.0758$$

(Q). A manufacturer of cotterpins notes that 5% of his product is defective. If he sells cotterpins in boxes of 100, and guarantees that not more than 10 pins will be defective, then what is the probability that a box will fail to meet the guaranteed quality?

Ans).
$p = 0.05$, $n = 100$

$\lambda = np = 100 \times 0.05 = 5$

$$P[x > 10] = 1 - P[x \le 10] = 1 - \sum_{x=0}^{10} \frac{5^x e^{-5}}{x!}$$

$$= 1 - e^{-5}\left[\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \dots + \frac{5^{10}}{10!}\right]$$

$$= 1 - e^{-5}\left[1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} + \frac{3125}{120} + \frac{15625}{720} + \frac{78125}{5040} + 9.68812 + 5.38229 + 2.69114\right]$$

$$= 1 - e^{-5}(146.36) = 0.0137$$

Q) A car hire firm has 2 cars which it hires from day to day. No. of demands for a car on each day is distributed as a Poisson variate with mean 1.5. Calculate the proportion of days in which : (i) neither car is used, (ii) some demand is refused.

Ans). given : $\lambda = 1.5$

let "x" be the number of demands for a car.

(i) $P[x=0] = \dfrac{\lambda^0 e^{-\lambda}}{0!} = \dfrac{e^{-1.5}}{1} = \underline{\underline{0.223}}$

(ii) As demand is refused, it means $x > 2$. (greater than available cars)

$$P[x>2] = 1 - (P[x=2] + P[x=1] + P[x=0])$$

$$= 1 - \left[\frac{\lambda^2 e^{-\lambda}}{2} + \frac{\lambda^1 e^{-\lambda}}{1} + \frac{\lambda^0 e^{-\lambda}}{1}\right]$$

$$= 1 - \left[\frac{(2.25)e^{-1.5}}{2} + \frac{1.5e^{-1.5}}{1} + e^{-1.5}\right]$$

$$= \underline{\underline{0.19115}}$$

Q). A hospital switchboard receives on an average of 4 emergency calls in a 10 minutes interval. What is the probability that : (i) Atmost 2 emergency calls. (ii) Exactly 3 emergency calls, in a 10-minutes interval.

Ans). $\boxed{\lambda = 4}$

(i) $P[X \leq 2] = \lambda^0 e^{-4} + \lambda^1 e^{-4} + \dfrac{\lambda^2 e^{-4}}{2} = e^{-4} + 4e^{-4} + 8e^{-4}$

$= \underline{\underline{0.238}}$

(ii) $P[x=3] = \dfrac{\lambda^3 e^{-\lambda}}{3!} = \dfrac{64e^{-4}}{6}$

$= \underline{\underline{0.195}}$

Recurrence Relation for Poisson D.:

$$P[x] = \frac{\lambda^x e^{-\lambda}}{x!} \qquad P[x+1] = \frac{\lambda^{x+1} \cdot e^{-\lambda}}{(x+1)!}$$

$$\frac{P[x+1]}{P[x]} = \frac{\lambda^{x+1} \cancel{e^{-\lambda}}}{(x+1)!} \times \frac{x!}{\lambda^x \cancel{e^{-\lambda}}} = \frac{\lambda x!}{(x+1)!}$$

$$\boxed{P[x+1] = \left(\frac{\lambda}{(x+1)}\right)\left[P(x)\right]}$$

(5). Fit a P.D for given data:

ans)

x:	0	1	2	3	4
f:	122	60	15	2	1

ans). $N = \Sigma f = 200$ $\qquad \lambda = \underset{(\mu)}{\text{mean}} = \frac{\Sigma f x}{\Sigma f}$

$$\lambda = \frac{60+30+6+4}{200} = \frac{100}{200}$$

$\rightarrow \boxed{\lambda = 0.5}$

x	$P(x)$	Expected Frequency $= N \times P(x)$
0	$P(0) = \frac{\lambda^0 e^{-0.5}}{0!} = 0.606$	$0.606 \times 200 \sim 121$
1	$P(1) = (0.5)^1 e^{-0.5} = 0.303$	$0.303 \times 200 \sim 61$
2	$P(2) = (0.5)^2 e^{-0.5}/2 = 0.0758$	$0.0758 \times 200 \sim 15$
3	$P(3) = (0.5)^3 e^{-0.5}/6 = 0.01264$	$0.01264 \times 200 \sim 3$
4	$P(4) = (0.5)^4 e^{-0.5}/24 = 0.00158$	$0.00158 \times 200 \sim 0$

(6). Determine the no. of pages expected with 0,1,2,3,4 errors in 1000 pages, if on avg. 2 errors are found in 5 pages.

ans). $\lambda = \frac{2}{5} = 0.4$

"x" $\longrightarrow$ no. of errors

$P[X=0] = \frac{(0.4)^0 e^{-0.4}}{0!} = 0.67032$

No. of pages $= 0.67032 \times 1000 \simeq$ 670 pages

$P[X=1] = \frac{(0.4)^1 e^{-0.4}}{1!} = 0.26813$

No. of pages $= 0.26813 \times 1000 \approx$ 268 pages

$P[X=2] = \frac{(0.4)^2 e^{-0.4}}{2} = 0.05363$

No. of pages $= 0.05363 \times 1000 \simeq$ 54 pages

$P[X=3] = \frac{(0.4)^3 e^{-0.4}}{3!} = 0.00715$

No. of pages $= 0.00715 \times 1000 \simeq$ 7 pages

$P[X=4] = \frac{(0.4)^4 e^{-0.4}}{4!} = 0.00072$

No. of pages $= 0.00072 \times 1000 \simeq$ 1 page

---

* Normal Distribution: A CRV "X" is said to have normal distribution $f(x)$, if its prob. DFunc. is given by:

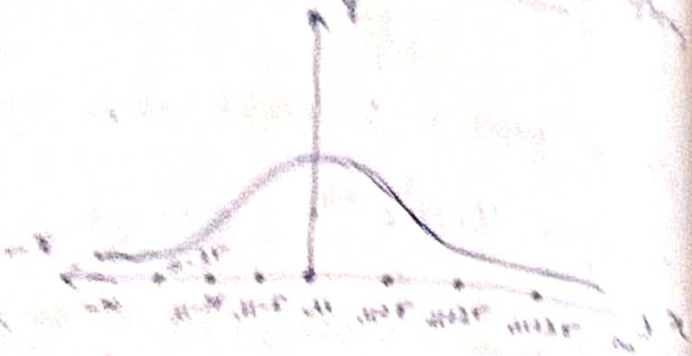
$$\underset{\substack{\downarrow \\ \text{mean}}}{f(x;\mu},\underset{\substack{\downarrow \\ \text{SD}}}{\sigma)} \text{ or } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \ \sigma > 0$$

$$= 0 \quad , \text{ otherwise.}$$

* Graph of normal distribution $y = f(x)$ in xy-plane is known as "normal" curve.
* It is a bell-shaped curve extending from $-\infty$ to $+\infty$ with its peak at $\boxed{x = \text{"}\mu\text{"}}$.
* It is symmetrical about y-axis.
* For normal distribution, $\boxed{\text{mean} = \text{median} = \text{mode}}$. Thus,

Normal curve is "unimodal".

- Normal curve is asymptotic to both sides of $x=-\infty$, $x=\infty$.
- Normal curve has inflection points at $\boxed{x = \mu \pm \sigma}$. The inflection points of the normal curve are equidistant from the mean.



Mean of Normal Distribution :

$$\text{Mean} = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

Let $t = \dfrac{x-\mu}{\sqrt{2}\,\sigma}$ $\Rightarrow$ As $x : -\infty$ to $+\infty$, $t : -\infty$ to $+\infty$

$$\text{Mean} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \sqrt{2}\,\sigma \left[\sqrt{2}\,\sigma t + \mu\right] dt$$

$$= \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{\infty} \sqrt{2}\,\sigma t e^{-t^2} dt + \int_{-\infty}^{\infty} \mu e^{-t^2} dt\right]$$

$$= \frac{1}{\sqrt{\pi}} \left[\sqrt{2}\,\sigma(0) + \mu \cdot 2\int_{0}^{\infty} e^{-t^2} dt\right]$$

$$= \frac{2\mu}{\sqrt{\pi}} \left[\int_{0}^{\infty} e^{-t^2} dt\right]$$

$$= \frac{2\mu}{\sqrt{\pi}} \left[\frac{\sqrt{\pi}}{2}\right]$$

$$= \mu .$$

Area Under The Normal Curve:

Area under normal curve is unity (1). We know that

$$\text{Area} = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$A = \frac{1}{\cancel{\sigma}\sqrt{\cancel{2}\pi}} \int_{-\infty}^{\infty} e^{-t^2}\,dt\,(\cancel{\sqrt{2}}\cancel{\sigma})$$

Put $t = \frac{x-\mu}{\sqrt{2}\sigma}$

$dx = \sqrt{2}\sigma\,dt$

$x : -\infty \longrightarrow \infty$

$t : -\infty \longrightarrow \infty$

$$A = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}\,dt$$

$$A = \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t^2}\,dt \quad (\because e^{-t^2} \text{ is even function})$$

$$A = \frac{A^2}{\sqrt{\pi}}\left(\frac{\sqrt{\pi}}{A}\right) \Rightarrow \boxed{\text{Area} = 1}$$

---

* The mean deviation about the mean is $\boxed{\frac{4}{5}(\sigma)}$

<u>Change of Scale from x-axis to z-axis</u> :

Probability that a CRV "X" lies between X_1 and X_2 is ~~defined by~~ denoted by : $P[X_1 \le X \le X_2]$ and is defined as :

$$P[X_1 \le X \le X_2] = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \quad \text{—— ①}$$

* By introducing standard normal variable / variate,

$\boxed{z = \frac{x-\mu}{\sigma}}$, eqn—① becomes :

$$P[z_1 \le z \le z_2] = \int_{z_1}^{z_2} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \cdot e^{-z^2/2} \, dz \cdot \cancel{\sigma}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} \, dz \quad \text{—— ②}$$

$$\left[\begin{array}{l} z_1 = \frac{X_1-\mu}{\sigma} \\ z_2 = \frac{X_2-\mu}{\sigma} \end{array}\right]$$

* We know that the <u>error function</u> / <u>probability integral</u> is

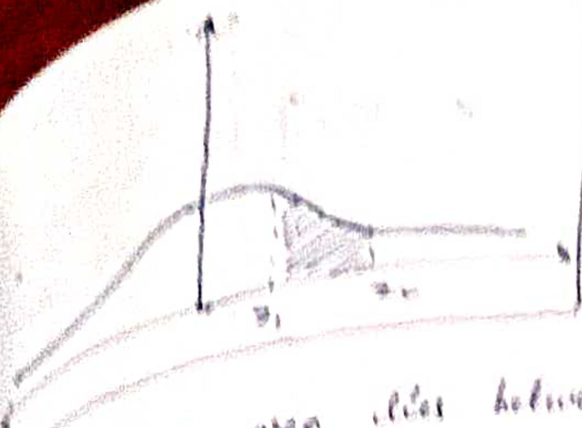
$$P[z] = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} \, dz . \quad \text{—— ③}$$

Now, eqn—② ~~[illegible]~~ can be rewritten as :

$$P[z_1 \le z \le z_2] = \frac{1}{\sqrt{2\pi}}\left[\int_{z_1}^{0} e^{-z^2/2} dz + \int_{0}^{z_2} e^{-z^2/2} dz\right]$$

$$= \boxed{P(z_2) - P(z_1)} \qquad [\text{from equation ③}]$$

(area from $z=0$ to $z=z_2$) (area from $z=0$ to $z=z_1$)



Area under normal curve is distributed as follows:

(i) 68.27% area lies between "$\mu - \sigma$" to "$\mu + \sigma$", i.e., $-1 \le z \le 1$

(ii) 95.45% area lies between "$\mu - 2\sigma$" to "$\mu + 2\sigma$", i.e., $-2 \le z \le 2$

(iii) 99.73% area lies between "$\mu - 3\sigma$" to "$\mu + 3\sigma$", i.e., $-3 \le z \le 3$.

Q) Determine the area under the normal curve for the following:

(i) Between $z = -1.2$ and $z = 2.4$

(ii) Between $z = 1.23$ and $z = 1.87$

(iii) Between $z = -2.35$ and $z = -0.5$

(iv) To the left of $z = -1.90$

(v) To the ~~right of~~ left of $z = 1.0$

(vi) To the right of $z = -2.40$

(vii) To the left of $z = -2.0$ and right of $z = 2.0$

Ans).

(i) Total area = (Area from $z=0$ to $z=2.4$) + (Area $z=0$ to $z=1.2$)

$= 0.4918 + 0.3849$

$= \underline{0.8767}$

(ii) Total area = (Area $z=0 \to 1.87$) − (Area $z=0 \to 1.23$)

$= 0.4693 - 0.3907$

$= \underline{0.0786}$

(iii) Area = $(z=0 \rightarrow z=2.35)$

$- (z=0 \rightarrow z=0.5)$

Area = $0.4906 - 0.1915$

Area = $\underline{\underline{0.2991}}$

(iv) Area = $0.5 - (z=0 \rightarrow 1.90)$

$= 0.5 - 0.4713$

$= \underline{\underline{0.0287}}$

(v) Area = $0.5 + (z=0 \rightarrow 1)$

$= 0.5 + 0.3413$

$= \underline{\underline{0.8413}}$

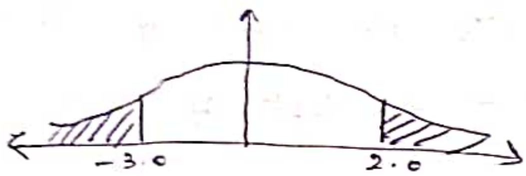
(vi) Area = $0.5 + (z=0 \rightarrow z=2.4)$

$= 0.5 + \cancel{0.4918}\ 0.4918$

$= \underline{\underline{0.9918}}$

(vii) Area = $0.5 - (z=0 \rightarrow z=3)$

$+ 0.5 - (z=0 \rightarrow z=2)$



Area = $0.5 - 0.49865 + 0.5 - 0.4778$

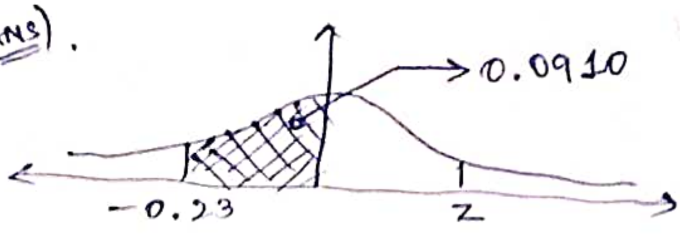
$= \underline{\underline{0.02355}}$

(Q). Find value of "z" such that :

(i) area between -0.23 and z is 0.5772.

(ii) area between $-z$ and $+z$ is 0.9.

ANS).



Total area $= 0.5772$

remaining area $= 0.4862$

$\therefore \boxed{z = 2.20}$

(ii) $z = 1.65$

area from $-z \rightarrow +z = 0.9$

$\therefore \frac{0.9}{2} = 0.45$ which is area for $\boxed{z = 1.65}$

Q). If masses of 300 students are normally distributed with mean 68kg & S.D = 3kg, how many students have masses (i) > 72kg (ii) <= 64kg (iii) b/w 65kg and 71kg (inclusive)

Ans). $\mu = 68kg$, $\sigma = 3kg$

(i) $P[x > 72]$. We know that $z = \frac{x - \mu}{\sigma}$

$z = \frac{72 - 68}{3} = 1.33 \Rightarrow \therefore P[z > 1.33]$

$P[z > 1.33] = 0.5 - (0.4082) = 0.0918$

$\therefore$ No. of students $= 0.0918 \times 300 \approx$ 28 students

(ii) $P[x <= 64]$ $z = \frac{64 - 68}{3} = -1.33$

$P[z <= -1.33] = 0.0918$

$\therefore$ No. of students $= 300 \times 0.0918 \approx 28$ students

(iii) $P[65 \le z \le 71]$

$z_1 = \frac{65 - 68}{3} = -1$ For $-1 \le z \le 1$

$z_2 = \frac{71 - 68}{3} = 1$ $\rightarrow 0.6827$

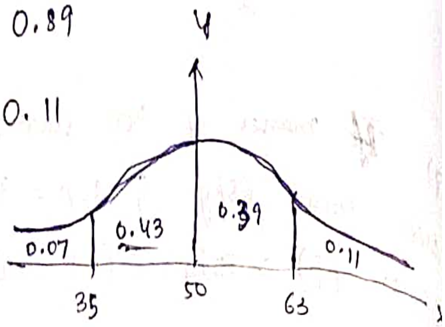
No. of students $= 300 \times 0.6827 \approx$ 205 students

*****
(Q). In a normal distribution, 7% of the items are under 35, and 89% are under 63. Find the mean and S.D of the distribution.

Ans). Given that, $\cancel{P[X]} =$ $P[X < 35] = 0.07$

$$P[X < 63] = 0.89$$

$$P[X > 63] = 0.11$$



area $= 0.43$, $z = -1.48$

we know that, $z = \frac{X - \mu}{\sigma}$

area $= 0.43$, $\quad -1.48 = \frac{35 - \mu}{\sigma}$ — (1)

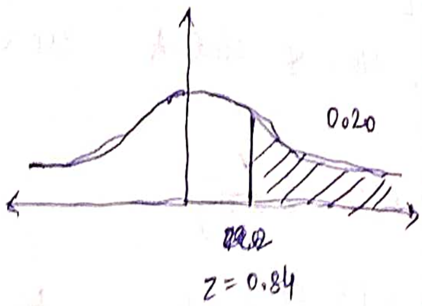
when area $= 0.39$, then $z = 1.23$

$$1.23 = \frac{63 - \mu}{\sigma} \quad — (2)$$

Solving (1) & (2); $\boxed{\mu = 50.29}$; $\boxed{\sigma = 10.33}$

(Q). A RV has normal dist. with $\mu = 62.4$. Find its standard deviation if the probability is 0.20 that it will take on a value > 79.2.

ANS).
$$P[X > 79.2] = 0.20$$



$$z = \frac{X - \mu}{\sigma}$$

$$0.84\,\sigma = 79.2 - 62.4$$

$$\boxed{\sigma = 20}$$

*****
(Q). Fit a normal distribution to the following data :

class :	60-62	63-65	66-68	69-71	72-74
frequency :	5	18	42	27	8

C.I	f					$f(x-\mu)^2$
60-62	5	61	305	-6	36	180
63-65	18	64	1152	-3	9	162
66-68	42	67	2814	0	0	0
69-71	27	70	1890	3	9	243
72-74	8	73	584	6	36	288

C.I	f	x	fx	$x-\mu$	$(x-\mu)^2$	$f(x-\mu)^2$
60-62	5	61	305	-6.45		208.0125
63-65	18	64	1152	-3.45		214.245
66-68	42	67	2814	-0.45		8.505
69-71	27	70	1890	2.55		175.5675
72-74	8	73	584	5.55		246.42

$\Sigma fx = 6745$; $\Sigma f = 100$; $\mu = \dfrac{\Sigma fx}{\Sigma f} = \underline{\underline{67.45}}$

∴ Normal distribution that fits to the given data:

$$\sigma = \sqrt{\frac{\Sigma f(x-\mu)^2}{N}} = \sqrt{\frac{852.75}{100}} = \underline{\underline{2.92}}$$

Class	(f)	True lower class limit (x_i)	$z_i = \frac{x_i - \mu}{\sigma}$	Area from $z=0$ to $z=z_i$	Area for class P	Exp. freq. NP
		59.5	-2.72	0.4967		
60-62	5				0.0413	~4
		62.5	-1.70	0.4554		
63-65	18				0.2068	~21
		65.5	-0.67	0.2486		
66-68	42				0.3892	~39
		68.5	0.36	0.1406		
69-71	27				0.2771	~28
		71.5	1.39	0.4177		
72-74	8				0.0743	~7
		74.5	2.41	0.4920		

Exponential Distribution : A crv ... is said to have an exponential distribution with parameter "λ" (>0) if the probability density F is :

$$f(x\,;\,\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

(λ = rate parameter)

Note : The exponential distribution is used as a model for the distribution of times between the occurrence of successive events.

- $$\int_{-\infty}^{\infty} f(x,\lambda)\,dx = 1 \qquad (\because f(x) \text{ is a PDF})$$

$$\lambda \int_{-\infty}^{\infty} e^{-\lambda x}\,dx = 1$$

$$\cancel{\lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_{-\infty}^{\infty} = 1}$$

$$\lambda\left[\int_{-\infty}^{0} e^{-\lambda x}\,dx + \int_{0}^{\infty} e^{-\lambda x}\,dx\right.$$

$$\cancel{\lambda}\left[\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx\right] = 1$$

$$\lambda \int_{0}^{\infty} e^{-\lambda x}\,dx = 1 \qquad (\because f(x) = 0 \text{ when } x \leq 0)$$

$$\cancel{\lambda}\left[\frac{e^{-\lambda x}}{-\cancel{\lambda}}\right]_{0}^{\infty} = 1$$

$$-\left(\frac{1}{e^{\lambda x}}\right)_{0}^{\infty} = 1$$

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

$$1 = 1 \qquad \text{(Hence, proved)}$$

Mean of Exponential Distribution :

$$\text{Mean} = \mu = \int_{-\infty}^{\infty} x\, f(x,\lambda)\,dx$$

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

$$\mu = \lambda \int_{0}^{\infty} x e^{-\lambda x}\,dx$$

$$\mu = \lambda \left[x\left(\frac{e^{-\lambda x}}{-\lambda}\right) - \left(\frac{e^{-\lambda x}}{\lambda^2}\right) \right]_{x=0}^{x=\infty} \qquad \left(\lim_{x\to\infty} \frac{x}{e^{\lambda x}} = \lim_{x\to\infty} \frac{1}{\lambda e^{\lambda x}} \right)$$

$$\mu = \lambda \left[-\frac{1}{\lambda}\left(\frac{x}{e^{\lambda x}}\right) - \frac{1}{\lambda^2}\left(\frac{1}{e^{\lambda x}}\right) \right]_{x=0}^{x=\infty}$$

$$\mu = \lambda \left[-\frac{1}{\lambda}(0) - \frac{1}{\lambda^2}(0) + \frac{1}{\lambda}(0) + \frac{1}{\lambda^2}(1) \right]$$

$$\boxed{\mu = 1/\lambda}$$

Variance of Exponential Distribution:

$$\sigma^2 = E(x^2) - [E(x)]^2 \quad \text{—— (1)}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x}\,dx$$

$$E(x^2) = \lambda \left[x^2\left(\frac{e^{-\lambda x}}{-\lambda}\right) - 2x\left(\frac{e^{-\lambda x}}{\lambda^2}\right) + 2\left(\frac{e^{-\lambda x}}{-\lambda^3}\right) \right]_{x=0}^{x=\infty} \quad \text{(using limits)}$$

$$E(x^2) = \lambda \left[x^2(0) - 2x(0) + 2(0) - \left[0 + 0 - 2/\lambda^3\right] \right]$$

$$E(x^2) = \lambda\left(\frac{2}{\lambda^3}\right) \Rightarrow E(x^2) = \frac{2}{\lambda^2}$$

$$\sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \Rightarrow \boxed{\sigma^2 = 1/\lambda^2}$$

$\therefore$ "λ" is inverse of expected duration (μ).

Cumulative Distribution Function

$$F(x;\lambda) \text{ (or) } P[X \le x] = \begin{cases} 0 & , x < 0 \\ 1-e^{-\lambda x} & , x \ge 0 \end{cases}$$

(Q). Assume that the length of a phone call in minutes an exponential RV "x" with parameter $\boxed{\lambda = 1/10}$. If someone arrives at a phone booth just before you arriv then find the probability that you will have to wait

(i) < 5 mins (ii) > 10 mins (iii) between 5 & 10 mins

(iv) Compute expected value & variance. (length of phone call in mins)

ANS). $\mu = \frac{1}{\lambda} = \underline{10 \text{ mins}}$ (expected)

(i) $P[X<5] = 1-e^{-\lambda(5)}$, $x \ge 0$ ["x" denotes no. of minutes]

$P[X<5] = 1-e^{-0.5} = \underline{0.3934}$

~~(ii) $P[X>10] = 1-e^{-\lambda(10)}$~~

(ii) $P[X>10] = 1 - P[X<10]$

$= 1 - (1-e^{-\lambda(10)})$

$= e^{-1} = \underline{0.3678}$

(iii) $P[5<X<10]$

$= F[10; 0.1] - F[5; 0.1]$

$= P[X<10] - P[X<5]$

$= 0.632 - 0.3934$

$= \underline{0.23865}$

★★★★

$$\boxed{\begin{aligned} &P[a \le X \le b] \\ &= F(b;\lambda) - F(a;\lambda) \end{aligned}}$$

(iv) $\mu = \frac{1}{\lambda} = \frac{1}{0.1} = 10$ $\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{(0.1)^2} = 100$

(Q) Laptops produced by a company ABC last on an average for 5 years. The lifespan of each laptop follows an exponential distribution:

(i) Calculate rate parameter (ii) Write PDF

(iii) what's the prob. that the laptop will last < 3 years?

(iv) what's " " " the laptop " " > 10 years?

(v) what's " " that a " " last b/w 4 & 7 years?

ANS). (i) $\lambda = \frac{1}{\mu} = \frac{1}{5} \Rightarrow \boxed{\lambda = 0.2}$

(ii) $f(x;\lambda) = \begin{cases} 0.2e^{-0.2x} & , x \geq 0 \\ 0 & , \text{otherwise} \end{cases}$

(iii) $P[X<3] = 1 - e^{-\lambda x} = 1 - e^{-0.2(3)} = \underline{\underline{0.4511}}$

(iv) $P[X>10] = 1 - P[X<10] = 1 - (1 - e^{-0.2(10)}) = e^{-2} = \underline{\underline{0.1353}}$

(v) $P[4<X<7] = F(7;0.2) - F(4;0.2)$

$= P[X<7] - P[X<4]$

$= 1 - e^{-0.2(7)} - 1 + e^{-0.2(4)}$

$= e^{-0.8} - e^{-1.4} = \underline{\underline{0.2027}}$

(Q). If "X" is an exp. var. with mean "5", then evaluate

(i) $P[0 \leq X \leq 1]$ (ii) $P[-\infty \leq X \leq 10]$

ANS). $\mu = 1/\lambda \Rightarrow \lambda = 1/5 \Rightarrow \boxed{\lambda = 0.2}$

(i) $P[0 \leq X \leq 1] = F[1;0.2] - F[0;0.2]$

$= e^{-0.2(0)} - e^{-0.2(1)} = 1 - e^{-0.2}$

$= 0.1813$

(ii) $P[-\infty < X < 10]$ ~~$= P[X < 10] = 1 - e$~~

$$= P[0 < X < 10] = F(10;\lambda) - F(0;\lambda)$$

$$= 1 - e^{-\lambda 10} - 1 + e^{0\lambda} = 1 - e^{-2} = \underline{0.8647}$$

Gamma Distribution: A continuous RV "X" is said to [have] a gamma distribution if the PDF of "X" is:

$$f(x;k,\lambda) = \begin{cases} \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $k > 0$

Mean of Gamma Distribution: $\boxed{\mu = k/\lambda}$

Variance of Gamma Distribution: $\boxed{\sigma^2 = k/\lambda^2}$

(Q). The time to repair a machine is gamma-distribut[ed] with mean & variance 2 and 1 respectively. Find the probability that the repair time of a machine exceeds

ANS).

$$\frac{k}{\lambda} = 2 \qquad \frac{k}{\lambda^2} = 1$$

$$k = 2\lambda \qquad k = \lambda^2$$

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

$$\boxed{\lambda = 2} \; ; \; \boxed{k = 4}$$

Let "X" denote the repair time of a machine in "hrs".

$$P[X > 2] = \int_2^\infty \frac{2^4 x^{4-1} e^{-2x}}{\Gamma(k)} = 16 \int_2^\infty \frac{x^3 e^{-2x}}{\Gamma(k)} = 16 \int_2^\infty \frac{x^{3-2x}}{\Gamma(4)}$$

$$= \frac{8}{3} e^3 \left[\frac{e^{-2x}}{-2}\right]_2^\infty$$

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

$\Gamma(n) = (n-1)!$

$(n = \text{integer})$

$$P[X>2] = \int_2^{\infty} \frac{2^4 x^3 e^{-2x}}{\Gamma(4)} = \frac{8}{3}\int_2^{\infty} x^3 e^{-2x}\,dx$$

$$= \frac{8}{3}\left[x^3\left(\frac{e^{-2x}}{-2}\right) - 3x^2\left(\frac{e^{-2x}}{4}\right) + 6x\left(\frac{e^{-2x}}{-8}\right) - 6\left(\frac{e^{-2x}}{16}\right)\right]_{x=2}^{x=\infty}$$

$$= \frac{8}{3}\left[x^3(0) - 3x^2(0) + 6x(0) - 6(0) - \left[-4e^{-4} - \frac{3}{4}(4)e^{-4} - \frac{3}{4}(2)e^{-4} - \frac{3}{8}e^{-4}\right]\right]$$

$$= \frac{8}{3}\left[4e^{-4} + 3e^{-4} + \frac{3}{2}e^{-4} + \frac{3}{8}e^{-4}\right]$$

$$= \underline{\underline{0.4334}}$$

(Q). The daily consumption of milk in a city in excess of 30,000 gallons is appr. distributed as a gamma variate, with parameters $k=2$, $\lambda = 1/20000$. The city has a daily stock of 40,000 gallons. What is the probability of stock is not sufficient on a particular day? (0.6065)

ANS).

(Q). It has been claimed that, in 60% of all solar heat installations, the utility bill is reduced by atleast 1/3rd. Accordingly, what are the probabilities that the utility will be reduced by atleast 1/3rd in (i) 4 of 5 installations

(ii) Atleast 4 of 5 installations.

(Q). 10% of bolts produced by a certain machine turns out to be defective. Find the prob. that in a sample of 10 bolts selected at random, exactly 2 will be defective, using

(i) Binomial dist. (ii) Using Poisson's and Comment upon the result.

(Q). In a normal dist., 31% of the items are under 45, and 8% are over 64. Find mean & standard deviation of distr.

(Q). The mean weight of 500 male students at a certain clg is 75kg & S.D is 7kg. Assuming that the weights are normally distributed, find how many students weigh (i) between 60 and 78kg, (ii) more than 92kg.

ANSWERS:

1) $n = 5$; $p = 0.6$; $q = 0.4$

(i) $P[X = 4] = {}^5C_4 (p)^4 (q)^1 = {}^5C_4 (0.6)^4 (0.4)^1$

$= 5 \times 0.4 \times 0.6^4$

$= \underline{0.2592}$

(ii) $P[X=4] + P[X=5] = {}^5C_4 (0.6)^4 (0.4)^1 + {}^5C_5 (0.6)^5 (0.4)^0$

$= 0.2592 + (0.6)^5$

$= \underline{0.33696}$

2) (i) $n = 10$; $p = 0.1$; $q = 0.9$

$P[X = 2] = {}^{10}C_2 (p)^2 (q)^8$

$P[X=2] = {}^{10}C_2 (0.1)^2 (0.9)^8 = \underline{0.1937}$

(ii) $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

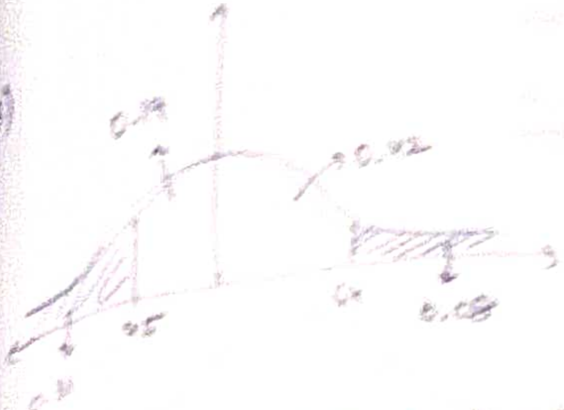
$\lambda = np = 10 \times 0.1 = 1$

$P[X=2] = \dfrac{(1)^2 e^{-1}}{2!} = \dfrac{e^{-1}}{2} = \underline{0.1839}$

$P[X<45] = 0.31$

$P[X>64] = 0.08$

3)



when area $= 0.19$; $z = 0.5$

$$-0.5 = \frac{45-\mu}{\sigma} \quad \text{— (1)}$$

when area $= 0.42$; $z = 1.41$

$$1.41 = \frac{64-\mu}{\sigma} \quad \text{— (2)}$$

$-0.5\sigma = 45 - \mu$

$\boxed{\mu - 0.5\sigma = 45}$

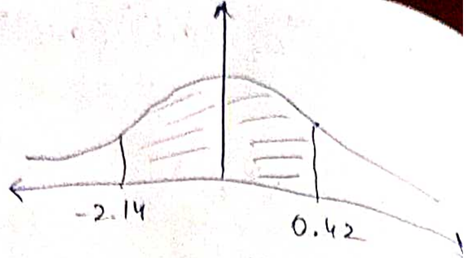
$1.41\sigma = 64 - \mu$

$\boxed{\mu + 1.41\sigma = 64}$

Solving (1) & (2) ; $\boxed{\begin{array}{l} \mu = 49.9 \approx 50 \\ \sigma = 9.94 \approx 10 \end{array}}$

4) $\mu = 75$; $\sigma = 7$

(i)



$$z_1 = \frac{x_1 - \mu}{\sigma}$$

$$z_1 = \frac{78-75}{7} = \frac{3}{7} = 0.4285 \simeq 0.43$$

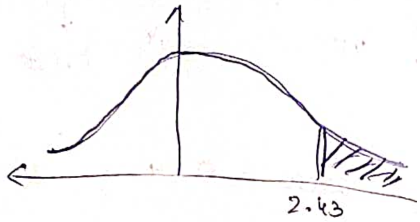
$$z_2 = \frac{60_2 - \mu}{\sigma} = \frac{-15}{7} = -2.1428 = -2.14$$

$$P[60 < z < 78] = 0.1664 + 0.4888 = 0.6502$$

No. of students $= 0.6502 \times 500$

$\simeq$ 325 students

(ii) $z = \frac{92-\mu}{\sigma} = 2.43$



$$\therefore P[z > 2.43] = 0.5 - 0.4925$$

$$= 7.5 \times 10^{-3}$$

No. of students $= 500 \times 7.5 \times 10^{-3}$

$\simeq$ 4 students