AL302

Introduction to Probability & Statistics

Unit - 4

Basic Statistics

Probability distribution - A theoritical probability distribution is a frequency distribution of certain events in which frequencies are obtained by mathematical computation.

Types of distribution :-
i> Binomial distribution
ii> Poisson distribution
iii> Normal distribution.

Binomial distribution → $\boxed{P(r) = {}^nC_r \, p^r \cdot q^{n-r}}$

useful when n is very small

where, n = total no. of times an event takes place
r = happening of an event
★ $\boxed{r \le n}$ $\left[{}^nC_r = \frac{n!}{r! \; n-r!} \right]$

eg → "Coin is tossed 10 times, find probability of getting head 6 times" ⇒ $n = 10$, $r = 6$, $p = \frac{1}{2}$, $q = \frac{1}{2}$.

p = Prob of happening of event in one time.
q = probability of not happening an event
★ $\boxed{p+q=1}$

Q> The probability that a bomb dropped from a plane will strike the target is 1/5, If 6 bombs are dropped, Find the probability :
i> Exactly 2 will strike a target.
ii> Atleast 2 will strike a target.
iii> None will strike a target
iv> Less than 2 will strike a target
v> More than 2 will strike a target.

Sol $p = \frac{1}{5}$, $p + q = 1$, $n = 6$

$\therefore q = \frac{4}{5}$

i> Exactly 2 $\therefore r = 2$

$P(r) = {}^{n}C_{r} \; p^{r} \; q^{n-r}$

$P(2) = {}^{6}C_{2} \left(\frac{1}{5}\right)^{2} \left(\frac{4}{5}\right)^{6-2}$

$P(2) = \frac{6!}{2! \; 4!} \cdot \frac{1}{25} \cdot \frac{4^4}{5^4}$

$P(2) = \frac{6 \times 5 \times \cancel{4!}}{2 \times \cancel{4!}} \cdot \frac{4^4}{5^6}$

$P(2) = \frac{3 \times 4^4}{5^5}$

$\boxed{P(2) = 0.245}$

ii> Atleast 2 , $r \geq 2$

$P(r \geq 2) = P(2)$ or $P(3)$ or $P(4)$ or $P(5)$ or $P(6)$

$P(r \geq 2) =$ ~~0.245~~ $P(2) + P(3) + P(4) + P(5) + P(6)$

$\because P(0) + P(1) + P(2) + - - - + P(n) = 1$

Here, $n = 6$

$P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$

$P(2) + P(3) + P(4) + P(5) + P(6) = 1 - P(0) - P(1)$

$\therefore P(r \geq 2) = 1 - P(0) - P(1)$

$P(r \geq 2) = 1 - \left[{}^{6}C_{0} \left(\frac{1}{5}\right)^{0} \left(\frac{4}{5}\right)^{6-0} \right] - \left[{}^{6}C_{1} \left(\frac{1}{5}\right)^{1} \left(\frac{4}{5}\right)^{6-1} \right]$

$P(r \geq 2) = 1 - \left[\frac{6!}{6!} \cdot 1 \cdot \frac{4^6}{5^6} \right] - \left[\frac{6!}{5!} \cdot \frac{1}{5} \cdot \frac{4^5}{5^5} \right]$

$P(r \geq 2) = 1 - \left[\frac{4^6}{5^6} \right] - \left[\frac{6 \times 4^5}{5^6} \right] = 1 - 0.262 - 0.393$

$= 1 - 0.655$

$\boxed{P(r \geq 2) = 0.345}$

iii> None , $r = 0$

$$P(0) = {}^6C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{6-0}$$

$$\boxed{P(0) = 0.262}$$

iv> Less than 2 , $r < 2$

$$P(r<2) = P(0) + P(1)$$

$$= 0.262 + 0.393$$

$$\boxed{P(r<2) = 0.655}$$

v> More than 2 , $r > 2$

$$P(r>2) = P(3) + P(4) + P(5) + P(6)$$

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

$$= \underbrace{1 - 0.262 - 0.393} - 0.245$$

$$= 0.345 - 0.245$$

$$\boxed{P(r>2) = 0.1}$$

Q2> Out of 800 families with 5 children each, how many would you expect have i> 3 boys <ii> 5 girls iii> Either 2 or 3 boys <iv> Atleast 2 boys. Assume equal probability for boys & girls.

Sol $p = \frac{1}{2}$, $q = \frac{1}{2}$, ~~n = 80~~ $n = 5$, $N = 800 = \Sigma f$

i> 3 boys , $r = 3$

$$P(3) = {}^5C_3 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^2 = \frac{5!}{3!\,2!} \frac{1}{2^5} = \frac{5 \times 4}{2^6} = \frac{5}{2^4}$$

$$P(3) = 0.3125$$

Expected no. of family $= P(3) \times N = 0.3125 \times 800$

$$= \boxed{250 \text{ families}}$$

ii> 5 girls , $r=0$ (none boys)

$$P(0) = {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{5!}{5!} \times 1 \times \frac{1}{2^5}$$

$$P(0) = 0.03125$$

Expected families $= 0.03125 \times 800 = \boxed{25}$

Q3> In a bombing action there is 50% chance that any bomb will strike the target. 2 direct hits are needed to destroy target. How many bombs are required to be dropped to give a 99% chance or better of ~~one~~ completely destroying the target.

Sol $p = 50\% = \frac{50}{100} = \frac{1}{2}$, $\therefore q = \frac{1}{2}$

($n=?$) Let (n) bombs are dropped

According to question,

Total probability $\geq 99\%$

$$P(0) + P(1) + P(2) + \cdots + P(n) \geq 0.99$$

$${}^nC_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{n-2} + {}^nC_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{n-3} + \cdots + {}^nC_n \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{n-n} \geq 0.99$$

$$\cancel{\left(\frac{1}{2}\right)^n} \; {}^nC_2 \left(\frac{1}{2}\right)^n + {}^nC_3 \left(\frac{1}{2}\right)^n + {}^nC_4 \left(\frac{1}{2}\right)^n + \cdots + {}^nC_n \left(\frac{1}{2}\right)^n \geq 0.99$$

$$\left(\frac{1}{2}\right)^n \left[{}^nC_2 + {}^nC_3 + {}^nC_4 + \cdots {}^nC_n \right] \geq 0.99$$

$$\left[{}^nC_0 + {}^nC_1 + {}^nC_2 + \cdots + {}^nC_n \right] = 2^n$$

$$\left(\frac{1}{2}\right)^n \left[2^n - {}^nC_0 - {}^nC_1 \right] \geq 0.99$$

Standard deviation = $\sqrt{npq}$

$$\frac{1}{2^n}\left[2^n - 1 - n\right] \geq 0.99$$

$$1 - \frac{(1+n)}{2^n} \geq 0.99$$

$$1 - 0.99 \geq \frac{1+n}{2^n}$$

$$0.01 \geq \frac{1+n}{2^n}$$

$$\frac{1}{100} \geq \frac{1+n}{2^n}$$

$$2^n \geq 100 + 100n$$

put $n=11$

$$2^{11} > 100 + 100(11)$$

$$2048 > 1200$$

$\therefore$ 11 bombs are required to destroyed the target

Q4) The Sum & product of mean & variance of binomial distribution are 24 & 128, Find the distribution

we know, mean = np

variance = npq

Given, mean + variance = 24

$$np + npq = 24 \quad -①$$

$$np(1+q) = 24 \quad \Rightarrow np = \frac{24}{1+q} \quad -Ⓐ$$

$$np(npq) = 128$$

$$(np)^2 q = 128$$

$$\left(\frac{24}{1+q}\right)^2 q = 128$$

$$576q = 128(1+q)^2$$

$$576q = 128 + 128q^2 + 256q$$

$$128q^2 + 256q - 576q + 128 = 0$$

$$128q^2 - 320q + 128 = 0$$

$$16q^2 - 40q + 16 = 0$$

$$2q^2 - 5q + 2 = 0$$

$$q = \frac{1}{2}, 2$$

$$\because q \neq 2 \quad (\because q \leq 1)$$

$$\therefore \boxed{q = \frac{1}{2}} \qquad \therefore \boxed{p = \frac{1}{2}}$$

Substituting p & q in eq Ⓐ

$$n = \frac{24}{p(1+q)} = \frac{24}{\frac{1}{2}\left(1+\frac{1}{2}\right)} = \frac{24 \times 2 \times 2}{3}$$

$$\boxed{n = 32}$$

$$P(r) = \sum_{r=0}^{n} {}^nC_r \; p^r \; q^{n-r} \qquad \text{or} \qquad P(r) = (q+p)^n$$

$$\boxed{p(r) = \sum_{r=0}^{32} {}^{32}C_r \; p^r \; q^{32-r}} \qquad \boxed{P(r) = \left(\frac{1}{2} + \frac{1}{2}\right)^{32}}$$

Q> Find the mean, variance & standard deviation of binomial distribution.

Sol We know, binomial distribution is $P(r) = {}^nC_r \, p^r q^{n-r}$

→ Mean $= \mu_1' =$ First moment about origin

$$= \sum_{r=0}^{n} r \cdot p(r)$$

$$= \sum_{r=0}^{n} r \cdot {}^nC_r \; p^r \; q^{n-r} \quad \text{—①}$$

$$= \sum_{r=0}^{n} r \cdot \frac{n!}{r!\; n-r!} p^r q^{n-r}$$

$$= \sum_{r=0}^{n} \frac{n\,(n-1)!}{(r-1)!\;(n-1-r+1)!} p^{r-1+1} q^{n-1-r+1}$$

$$= \sum_{r=0}^{n} \frac{n\,(n-1)!}{(r-1)!\;(n-1-r+1)!} p^{r-1} \cdot p \cdot q^{(n-1)-(r-1)}$$

$$= np \cdot \underbrace{\sum_{r=0}^{n} {}^{n-1}C_{r-1}\, p^{r-1} \cdot q^{(n-1)-(r-1)}}$$

↳ total prob always 1

$$= np\,(1) = np$$

$$\boxed{\text{mean} = \bar{x} = np}$$

⇒ Variance $= \mu_2' - (\mu_1')^2$

$\therefore \mu_2' = \sum_{r=0}^{n} r^2\, p(r)$

$$= \sum_{r=0}^{n} r^2 \; {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} (r^2 - r + r)\; {}^nC_r\, p^r \cdot q^{n-r}$$

$$= \sum_{r=0}^{n} [r(r-1) + r]\; {}^nC_r \cdot p^r \cdot q^{n-r}$$

$$= \sum_{r=0}^{n} r(r-1)\; {}^nC_r\, p^r q^{n-r} + \underbrace{\sum_{r=0}^{n} r\; {}^nC_r\, p^r q^{n-r}}_{\text{(from ①)}}$$

$$= \sum_{r=0}^{n} r(r-1) \frac{n!}{r!\,(n-r)!} p^r q^{n-r} + np$$

$$= \sum_{r=0}^{n} \frac{\cancel{r(r-1)}\; n!}{\cancel{r(r-1)}(r-2)!\,(n-r)!} p^r q^{n-r} + np$$

$$= \sum_{r=0}^{n} \frac{n(n-1)(n-2)!}{(r-2)!\;(n-2-r+2)!} \cdot p^{r-2+2} q^{n-2-r+2} + np$$

$$= n(n-1)\,p^2 \sum_{r=0}^{n} \frac{(n-2)!}{(r-2)!\;[(n-2)-(r-2)]!} p^{r-2} q^{(n-2)-(r-2)} + np$$

$$= n(n-1)\,p^2 \underbrace{\sum_{r=0}^{n} {}^{n-2}C_{r-2}\, p^{r-2} q^{(n-2)-(r-2)}} + np$$

$= n(n-1)\, p^2 \,(1) + np \qquad = n^2p^2 - np^2 + np$

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

$$\boxed{\mu_2' = n^2p^2 + npq}$$

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

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

$$\boxed{\text{Variance} = npq}$$

As we know, $\text{standard deviation} = \sqrt{\text{Variance}}$

$$\therefore \boxed{\text{Standard deviation} = \sqrt{npq}}$$

* Poission Distribution

useful for discrete data, very large data, P(E) is very small

$$\boxed{P(r) = \frac{e^{-m} m^r}{r!}} , \quad r \leq n$$

n is very long

P is very small

where m = mean = np = no. of times × prob. of event in one time

r = event happening of event

Q) The probability that an individual suffer a bad reaction from a certain injection is 0.001. Determine the probability that ~~that~~ out of 2000 individuals :

i) Exactly 3 ⟨ii⟩ More than 2 ⟨iii⟩ None (iv) Atleast 2

~~whitt~~ will suffer a bad reaction.

Sol $n = 2000$, $p = 0.001$

$\therefore m = np = 2000 \times 0.001$

$m = 2$

By poission distribution

i) Exactly 3 (i.e $r = 3$)

$$p(r) = \frac{e^{-m} m^r}{r!}$$

$$P(3) = \frac{e^{-2} 2^3}{3!} = \frac{e^{-2} 8}{6} = \frac{4}{3} \times e^{-2}$$

$$\boxed{P(3) = 0.180}$$

ii) More than 2 (i.e $r > 2$)

$$P(r>2) = P(3) + P(4) + \cdots + P(2000)$$

$$= 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[\frac{e^{-2} 2^0}{0!} + \frac{e^{-2} 2^1}{1!} + \frac{e^{-2} . 2^2}{2!} \right]$$

$$= 1 - e^{-2} [1 + 2 + 2] = 1 - e^{-2} \times 5$$

$$\boxed{P(r>2) = 0.325}$$

iii> None (i.e $r=0$)

$$P(0) = \frac{e^{-2}\, 2^0}{0!} = e^{-2}$$

$$\boxed{P(0) = 0.135}$$

iv> Atleast two $(r \geq 2)$

$$P(r \geq 2) = P(2) + P(3) + P(4) + \text{---} \; P(2000)$$

$$= 1 - [P(0) + P(1)]$$

$$= 1 - e^{-2}(3)$$

$$\boxed{P(r \geq 2) = 0.595}$$

Q> A car hire firm hires 2 cars day by day. The number of demands for a car on each day is distributed as poission distribution with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of day on which some demand is refused.

Sol mean i.e $m = 1.5$

By poission distribution, $P(r) = \frac{e^{-m} m^r}{r!}$

i> Neither car is used $(r=0)$

$$P(0) = \frac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.223$$

Proportion of days $= 365 \times 0.223$

$= \boxed{81 \text{ days (approx)}}$

ii>

$$P(r>2) = P(3) + P(4) + P(5) + \text{---} + P(n)$$

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

$$= 1 - 0.223 - \frac{e^{-1.5}(1.5)^1}{1!} - \frac{e^{-1.5}(1.5)^2}{2!} = 1 - 0.223$$

$$-0.3346$$

$$-0.251$$

$$= 0.192$$

Proportion of days $= 365 \times 0.192 = \boxed{70 \text{ days (approx)}}$

Q> If 3% of the electric blub manufactured by a company are defective. Find the probability that in a Sample of 100 bulbs: exactly 5 will be defective.

Sol $P = 3\% = \frac{3}{100} = 0.03$, $n = 100$, $r = 5$

mean i.e $m = np = 100 \times 0.03 = 3$

By poission's distribution, $P(r) = \frac{e^{-m} \times m^r}{r!}$

$$P(5) = \frac{e^{-3}(3)^5}{5!} = \frac{12.098}{5!}$$

$$\boxed{P(5) = 0.1\ 0081}$$

Q> Fit a poission distribution to following data

x	0	1	2	3	4
$y = f$	46	38	22	9	1

Calculate the theoretical (expected) frequency and compare with actual ones.

Sol

$x = r$	$y = f$	$f \cdot x$	$P(r) = \frac{e^{-m} m^r}{r!}$	$N \cdot P(r)$ (theoretical)
0	46	0	$P(0) = \frac{e^{-0.974} \times (0.974)^0}{0!} = 0.377$	$116 \times 0.377 = 43.732$
1	38	38	$P(1) = \frac{e^{-0.974} \times (0.974)^1}{1!} = 0.3677$	$116 \times 0.367 = 42.572$
2	22	44	$P(2) = \frac{e^{-0.974} \times (0.974)^2}{2!} = 0.179$	$116 \times 0.179 = 20.764$
3	9	27	$P(3) = \frac{e^{-0.974} \times (0.974)^3}{3!} = 0.058$	$116 \times 0.058 = 6.728$
4	1	4	$P(4) = \frac{e^{-0.974} \times (0.974)^4}{4!}$ = 0.[illegible]14	$116 \times 0.014 = 1.624$
	$\Sigma f = 116$	$\Sigma fx = 113$		

mean i.e $m = \frac{\Sigma fx}{\Sigma f} = \frac{113}{116} = 0.974$

$N = \Sigma f = 116$

Q> Find mean, variance and standard deviation of poission distribution.

Sol We know that poission distribution

$$p(r) = \frac{e^{-m} \cdot m^r}{r!}$$

1> $\bar{x}$ = mean = μ_1' (μ one prime)

$$= \sum_{r=0}^{\infty} r \cdot p(r)$$

$$= \sum_{r=0}^{\infty} r \cdot \frac{e^{-m} m^r}{r!}$$

$$= \sum_{r=0}^{\infty} \cancel{r} \frac{e^{-m} m^r}{\cancel{r}(r-1)!}$$

$$= e^{-m} \sum_{r=0}^{\infty} \frac{m^r}{(r-1)!}$$

$$= e^{-m} \left[\frac{m^0}{(0-1)!} + \frac{m^1}{(1-1)!} + \frac{m^2}{(2-1)!} + \frac{m^3}{(3-1)!} + \cdots \right]$$

infinity

$$= e^{-m} \left[0 + \frac{m}{0!} + \frac{m^2}{1!} + \frac{m^3}{2!} + \cdots \right]$$

$$= m e^{-m} \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \cdots \right]$$

$$\left[\because e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \right]$$

$$= m e^{-m} \times e^m$$

$$\boxed{\text{mean} = m} = np = \frac{\sum fx}{\sum f} = \mu_1'$$

ii> Variance $= \mu_2' - \mu_1' \rightarrow$ ①

$$\mu_2' = \sum_{r=0}^{\infty} r^2 p(r)$$

$$= \sum_{r=0}^{\infty} (r^2 - r + r)\, p(r)$$

$$= \sum_{r=0}^{\infty} [r(r-1) + r]\, P(r)$$

$$= \sum_{r=0}^{\infty} [r(r-1)\, P(r)] + \underbrace{\sum_{r=0}^{\infty} r P(r)}$$

$$= \sum_{r=0}^{\infty} r(r-1) \frac{e^{-m} . m^r}{r!} + \mu_1'$$

$$= \sum_{r=0}^{\infty} r(r-1) \frac{e^{-m}\, m^r}{r(r-1)(r-2)!} + m$$

$$= e^{-m} \sum_{r=0}^{\infty} \frac{m^r}{(r-2)!} + m$$

$$= e^{-m} \left[0 + 0 + \frac{m^2}{1} + \frac{m^3}{1!} + \frac{m^4}{2!} + --- \right] + m$$

$$= e^{-m} . m^2 \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + - - - - \right] + m$$

$$= e^{-m} . m^2 . e^{m} + m$$

$$\mu_2' = m^2 + m$$

Putting values of μ_1', μ_2' in eq ①, we get

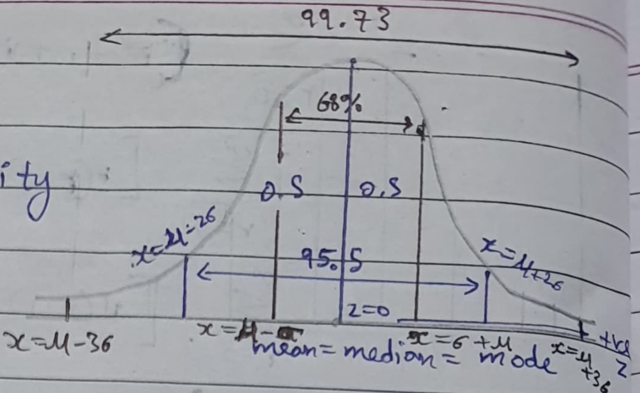
variance $= m^2 + m - m^2$

$\boxed{\text{variance} = m}$

Thus variance & mean of poission distribution are equal

and $\boxed{\text{Standard deviation} = \sqrt{\text{variance}} = \sqrt{m}}$

★ Normal Distribution



when area under curve = Total Probability

area under curve = 1

→ probability at left = 0.5

→ probability at right = 0.5

& $z = 0$ ($\because x = \mu$)

$$Z = \frac{x-\mu}{\sigma}$$

μ = mean

σ = standard deviation

→ right side shows (+ve) value of z

→ The horizontal line is called asymtote because it is a tangent which meets the curve at infinity.

when $Z = 1$

i.e $\frac{x-\mu}{\sigma} = 1$

$\sigma = x - \mu$

$\boxed{x = \sigma + \mu}$

when $z = 2$

$\frac{x-\mu}{\sigma} = 2$

$\boxed{x = 2\sigma + \mu}$

Function of curve, $\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$; $-\infty < x < \infty$

when $\mu = 0$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$ [Probability distribution function]

On integrating with $-\infty < x < \infty$, $f(x) = 1$

Properties of Normal Distribution :-

☆ The normal probability curve with mean (μ) and standard deviation (σ) has following ~~prop~~ property :-

i) The equation of curve :

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

It is bell-shaped

ii) The curve is Symmetrical about line $x = \mu$ and x range from $-\infty$ to $+\infty$.

iii) mean, median and mode ~~are~~ coincide at $x = \mu$ as distribution is Symmetrical.

iv) x-axis is asymtote to the curve.

v) In normal distribution, arithematic $\boxed{\text{mean} = \text{variance}}$

or $\boxed{\mu = \sigma^2}$.

vi) The point of inflation of curve ~~at~~ $x = \mu + \sigma$ & $x = \mu - \sigma$ ~~&~~ the curve changes from concave to convex

vii) The mean deviation from mean in normal distribution is 4/5 of its Standard deviation.

viii) All the odd moment about the mean are zero i.e $\mu_{2r+1} = 0$

ix) The curve of normal distribution has a single peak, it is a uni-model.

x) The two tails of the curve ~~extended~~ extend ~~never~~ Indefinitely & never touch the horizontal line

Q) The mean is 151 cm & standard deviation is 15 cm, assuming that the heights are normally distributed. Find how many students have height b/w 120 & 155 cm.

Sol no. of student, $N = 500$

mean, $\mu = 151$ cm

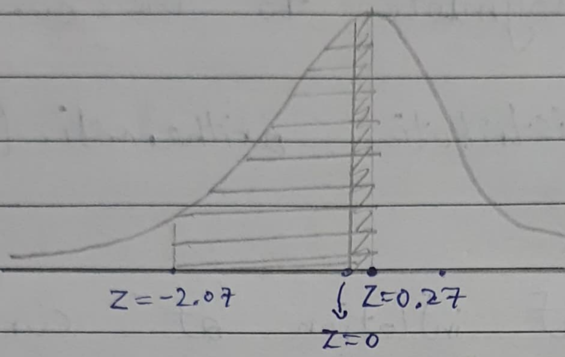
standard deviation, $\sigma = 15$ cm

By standard normal variable, $Z = \frac{x - \mu}{\sigma}$

For $x = 120$ cm, $Z_1 = \frac{120-151}{15} = -2.066 \simeq -2.07$

For $x = 155$ $\quad Z_2 = \frac{155-151}{15} = 0.266 \simeq 0.27$

$\therefore P(120 < x < 155) = P(-2.07 < z < 0.27)$



$P(120 < x < 155) =$ Area b/w $z=0$ & $z=0.27$

$+$ Area b/w $Z=0$ & $z=2.07$

$= (0.0000 + 0.1064) + (0.0000 + 0.4808)$

$\boxed{P(120 < x < 155) = 0.5872}$ $\quad$ no. of students $= 500 \times 0.5872$

$= 294$

Q) A sample of dry 100 battery sales tested to find length of life, produced the following results: mean $\bar{x} = 12$ hrs & standard deviation $SD = 3$ hrs, assuming the data to be normally distributed. What percentage of battery cells are expected to have life

i) More than 15 hrs $\quad$ (iii) ~~&~~ b/w 10 & 14 hrs

ii) Less than 6 sec $\quad$ (iv) Atleast 10 hrs

Sol Given, By standard normal ~~dist~~ variable

$$Z = \frac{x - \mu}{\sigma}$$

For $x = 15$ hrs, $Z_1 = \frac{15-12}{3} = \frac{3}{3}$

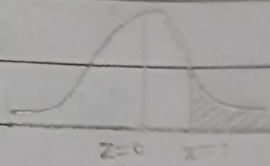
$Z_1 = 1$

i) ∵ $P(x > 15)$

$\therefore = P(z > 1) = 0.5 -$ Area b/w $z = 0$ & $z = 1$

$= 0.5 - 0.3413$

$= 0.1587$

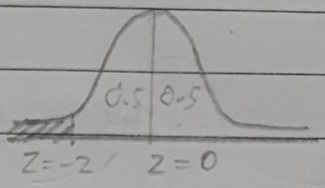


% of life more than 15 $= 0.1587 \times 100 = \boxed{15.87\%}$

∴ 15 out of 100 battery have more than 15 hours.

ii) Less than 6 hrs

$z = \frac{6-12}{3} = \frac{-6}{3} = -2$



$\therefore P(x < 6) = P(z < -2) = 0.5 - ($Area b/w $z=0$ & $z=2)$

$= 0.5 - 0.4772$

% of life less than 6hrs $= 0.0228 \times 100 = \boxed{2.28\%}$

∴ 2 out of 100 battery have less than 6 hrs life

iii) between 10 to 14 hrs

For $x = 10$ hrs

$Z_1 = \frac{10-12}{3} = \frac{-2}{3} = -0.67$

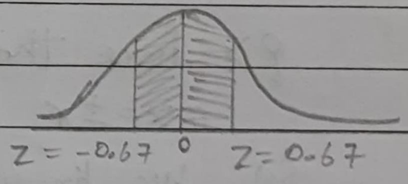
For $x = 14$ hrs

$Z_2 = \frac{14-12}{3} = \frac{2}{3} = 0.67$

$P(10 < x < 14) = P(-0.67 < z < 0.67)$

$= 2($Area b/w $z=0$ & $z=0.67)$

$= 2(0.2486) = 0.4972$



% life b/w 10 to 14 hrs $= 0.4972 \times 100 = \boxed{49.72\%}$

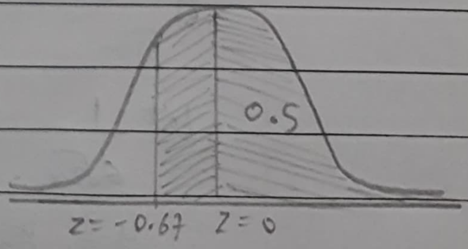
iv) Atleast 10 hrs

$z = \frac{10-12}{3} = -0.67$

$P(x \geq 10) = P(z \geq 0.67)$

$= 0.5 + ($area b/w $z=0$ & $z=0.67)$

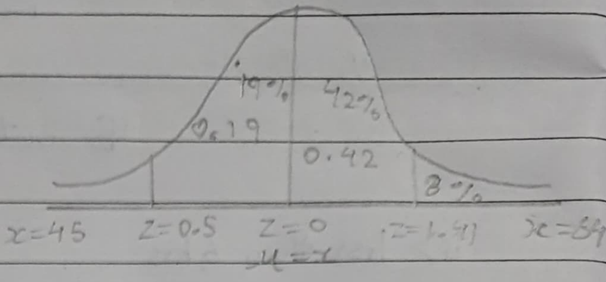
$= 0.5 + 0.2486 = 0.7486$



% life ~~b/w~~ atleast 10 hrs $= \boxed{74.86\%}$

Q> In a normal distribution 31% of item are under 45 & 8% are over 64. Find mean & standard deviation.

Sol Let μ & σ be the mean & standard deviation of normal distribution resp.



$$z = \frac{x-\mu}{\sigma}$$

$$0.5 = \frac{45-\mu}{\sigma} \quad ①$$

$$1.41 = \frac{64-\mu}{\sigma} \quad ②$$

from ① & ②, $\frac{0.5}{1.41} = \frac{45-\mu}{64-\mu}$

$$32 - 0.5\mu = 63.45 - 1.41\mu$$

$$0.91\mu = 31.45$$

mean, $\boxed{\mu = 34.56}$

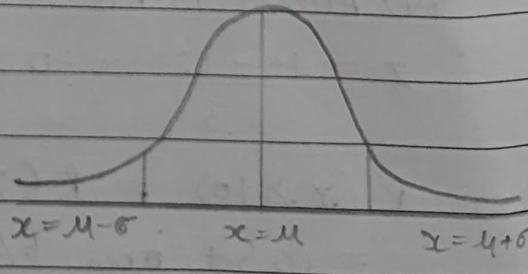
eq ①, $\sigma = \frac{45-\mu}{0.5} = \frac{45-34.56}{0.5} = \frac{10.44}{0.5}$

$\boxed{\sigma = 20.88}$ (standard deviation)

Q> Prove that the point of inflection of normal curve are $x = \pm\sigma$

So) We know that, the PDF with mean zero & standard deviation (σ) is given by:-

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$



Let $\frac{1}{\sigma\sqrt{2\pi}} = y_0$

$$y = y_0 e^{-x^2/2\sigma^2}$$

$$\frac{dy}{dx} = y_0 e^{-x^2/2\sigma^2} \left(\frac{-2x}{2\sigma^2}\right)$$

$$\frac{dy}{dx} = \frac{-y_0}{\sigma^2} x e^{-x^2/2\sigma^2}$$

$$\frac{d^2y}{dx^2} = \frac{-y_0}{\sigma^2}\left[x e^{-x^2/2\sigma^2}\left(\frac{-x}{\sigma^2}\right) + e^{-x^2/2\sigma^2}\right]$$

$$\frac{d^2y}{dx^2} = \frac{-y_0}{\sigma^2} e^{-x^2/2\sigma^2}\left[-\frac{x^2}{\sigma^2} + 1\right]$$

Put $\frac{d^2y}{dx^2} = 0$

$$\frac{-y_0}{\sigma^2} e^{-x^2/2\sigma^2}\left[1 - \frac{x^2}{\sigma^2}\right] = 0$$

$$1 - \frac{x^2}{\sigma^2} = 0$$

$$\sigma^2 = x^2$$

$$\therefore \boxed{x = \pm\sigma}$$

and, $$\frac{d^3y}{dx^3} = \frac{-y_0}{\sigma^2}\left\{\left[1 - \frac{x^2}{\sigma^2}\right] e^{-x^2/2\sigma^2}\left(\frac{-x}{\sigma^2}\right) + e^{-x^2/2\sigma^2}\left(\frac{-2x}{\sigma^2}\right)\right]$$

$$\frac{d^3y}{dx^3} = \frac{-y_0}{\sigma^2}\left[\left(1 - \frac{x^2}{\sigma^2}\right)\left(\frac{-x}{\sigma^2}\right) - \left(\frac{2x}{\sigma^2}\right)\right] e^{-x^2/2\sigma^2}$$

$$\frac{d^3y}{dx^3} = +\frac{y_0 x}{\sigma^2} e^{-x^2/2\sigma^2}\left[\left(1 - \frac{x^2}{\sigma^2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{2x}{\sigma^2}\right]$$

$\therefore$ ~~At~~ $\frac{d^3y}{dx^3} \neq 0$ at $x = +\sigma$ & $x = -\sigma$

$\therefore \boxed{x = \pm\sigma}$ are point of inflection of normal curve.

Q) The mean deviation from the mean of normal distribution is 4/5 times its standard deviation (Prove)

Sol As we know, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ —①

Mean deviation from mean μ,

$$MD = \int_{-\infty}^{\infty} |x - \mu| f(x)\, dx \quad -②$$

On putting value from ① in ②

$$MD = \int_{-\infty}^{\infty} |x-\mu| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $t = \frac{x-\mu}{\sqrt{2}\sigma}$

$x - \mu = \sqrt{2}\sigma t$

$x = \mu + \sqrt{2}\sigma t$

$\boxed{dx = \sqrt{2}\sigma t \, dt}$

$$MD = \int_{-\infty}^{\infty} |\sqrt{2}\sigma t| \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2} \sqrt{2}\sigma \, dt$$

$$= \sqrt{\frac{2\times 2}{2\pi}} \frac{\sigma \times \sigma}{\sigma} \int_{-\infty}^{\infty} t e^{-t^2} dt \qquad \left[\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x) dx\right]$$

$$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} 2\left[\int_{0}^{\infty} |t| e^{-t^2} dt\right.$$

Correlation

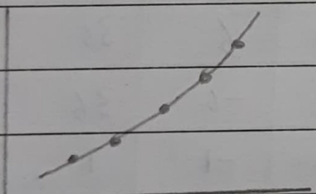
Correlation analysis deals with the association b/w two or more variable.

① i> Positive Correlation is an increase (or decrease) in the value of one variable corresponding to an increase (or decrease) in the others.

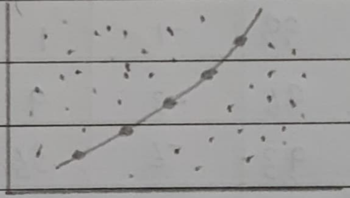
ii> Negative Correlation is a decrease (or increase) in the value of one variable corresponding to an increase (or decrease) in the others.

② Simple Partial / Multiple Correlation

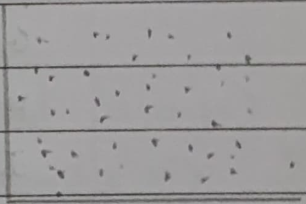
③ Linear / Non linear Correlation



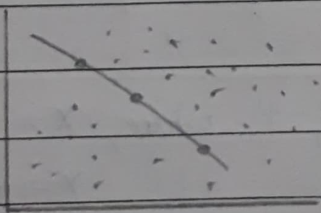
Perfectly linear



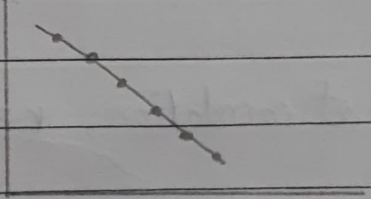
Non - linear



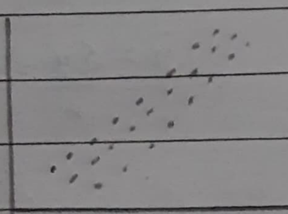
No relation



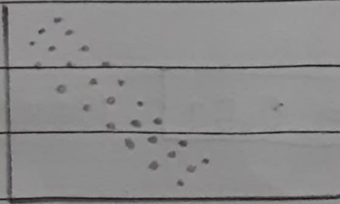
Negative linear



Perfectly negative linear



low degree positive correlation



low degree negative correlation

Q> Find the Karl-Pearson's coefficient of correlation for the following data:

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Sol.

x	y	$X = x - 20$	$Y = y - 21$	XY	X^2	Y^2
10	18	-10	-3	30	100	9
14	12	-6	-9	54	36	81
18	24	-2	3	-6	4	9
22	6	2	-15	-30	4	225
26	30	6	9	54	36	81
30	36	10	15	150	100	225
$\Sigma x = 120$	$\Sigma y = 126$	$\Sigma X = 0$	$\Sigma Y = 0$	$\Sigma XY = 252$	$\Sigma X^2 = 280$	$\Sigma Y^2 = 630$

mean $\bar{x} = \frac{\Sigma x}{n} = \frac{120}{6} = 20$

mean $\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$

Formula, $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$

← (actual mean/exact mean)

Jab exact mean fraction mai aaye toh dusri method lagana

where, $X = x - \bar{x}$

$Y = y - \bar{y}$

$r = \frac{252}{\sqrt{280 \times 630}}$

$\boxed{r = 0.6}$

(assumed mean)

or $r = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \sqrt{n \Sigma y^2 - (\Sigma y)^2}}$

where $X = x - a$

$u = x - a \quad v = y - b$

(table on next page)

Let 18, 24 be the assumed mean of x & y series ~~rep~~ resp.

$X = x - 18$

$Y = y - 24$

$r = \frac{6(216) - (12)(-18)}{\sqrt{6(304) - (12)^2} \sqrt{6(684) - (-18)^2}}$

$r = \frac{1512}{\sqrt{1680} \cdot \sqrt{3780}} = \boxed{0.6}$

x	y	$X = x - 18$	$Y = y - 24$	XY	X^2	Y^2
10	18	-8	-6	48	64	36
14	12	-4	-12	48	16	144
18	24	0	0	0	0	0
22	6	4	-18	-72	16	324
26	30	8	6	48	64	36
30	36	12	12	144	144	144
$\sum x = 120$	$\sum y = 126$	$\sum X = 12$	$\sum Y = -18$	$\sum XY = 216$	$\sum X^2 = 304$	$\sum Y^2 = 684$

Rank Correlation method

(above ques) Coefficient of correlation, $r = 1 - \frac{6 \sum d^2}{n(n^2-1)}$

where, $d = R_x - R_y$

x	y	R_x	R_y	$R_x - R_y = d$	d^2
10	18	1	3	-2	4
14	12	2	2	0	0
18	24	3	4	-1	1
22	6	4	1	3	9
26	30	5	5	0	0
30	36	6	6	0	0
					14

$$r = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 1 - \frac{6(14)}{6(36-1)} = 1 - \frac{84}{6(35)} = 1 - 0.4$$

$$\boxed{r = 0.6}$$

⇒ Rank correlation method when equal rank are given

Coefficient of correlation, ~~$r = 1 - 6[\sum d^2$~~

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12} m_1(m_1^2-1) + \frac{1}{12} m_2(m_2^2-1) + \frac{1}{12} m_3(m_3^2-1) + \dots\right]}{n(n^2-1)}$$

Q) X = 68 64 75 50 64 80 75 40 55 64

Y = 62 58 68 45 81 60 68 48 50 70

Sol Here, in this series X, the value 75 repeated twice and value of 64 repeated thrice

∴ say $m_1 = 2$, $m_2 = 3$

and in the series Y, the value 68 ~~resp~~ repeatedly twice

Say $m_3 = 2$

X	Y	R_x	R_y	$d = R_x - R_y$	d^2
68	62	7	6	1	1
64	58	$\frac{4+5+6}{3} = 5$	4	1	1
75	68	$\frac{8+9}{2} = 8.5$	7.5	1	1
50	45	2	1	1	1
64	81	5	10	-5	25
80	60	10	5	5	25
75	68	8.5	7.5	1	1
40	48	1	2	-1	1
55	50	3	3	0	0
64	70	5	9	4	16
					72

Lines of regressions:-

1) line of regression of x on y $(x = f(y))$

Formula:-

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

where $\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma uv - \frac{\Sigma u \Sigma v}{n}}{\left\{\Sigma v^2 - \frac{(\Sigma v)^2}{n}\right\}}$$

where, $u = x - \bar{x}$, $v = y - \bar{y}$

$$\boxed{SD = \sigma_x = \sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma v}{n}\right)^2}}$$

ii) line of regression of y on x ($y = f(x)$)

Formula $\quad y - \bar{y} = b_{yx} \cdot (x - \bar{x})$

where, $$b_{yx} = \frac{\Sigma uv - \frac{\Sigma u \Sigma v}{n}}{\Sigma u^2 - \frac{(\Sigma u)^2}{n}} = r\,\frac{\sigma_y}{\sigma_x}$$

$$\boxed{\text{Standard deviation} = \sigma_y = \sqrt{\frac{\Sigma v^2}{n} - \left(\frac{\Sigma v}{n}\right)^2}}$$

Coeff of correlation, $r = \sqrt{b_{xy}\ b_{yx}}$

Q) Calculate the coefficient of regression lines & find two lines of regression from the following data:

x :	78	89	97	69	59	79	68	61
y :	125	137	156	112	107	136	123	108

x	y	$u = x - 78$	$v = y - 125$	uv	u^2	v^2
78	125	0	0	0	0	0
89	137	11	12	132	121	144
97	156	19	31	589	361	961
69	112	−9	−13	117	81	169
59	107	−19	−18	342	361	324
79	136	1	11	11	1	121
68	123	−10	−2	20	100	4
61	108	−17	−17	289	289	289
$\Sigma x = 600$	$\Sigma y = 1004$	$\Sigma u = -24$	$\Sigma v = 4$	$\Sigma uv = 1500$	$\Sigma u^2 = 1314$	$\Sigma v = 2012$

$n = 8$

$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{600}{8} = 75$

$\bar{y} = \frac{\Sigma y}{n} = \frac{1000}{8} = 125.5$

Coefficient of regression of x ~~only~~ on y

$$b_{xy} = \frac{\Sigma uv - \frac{\Sigma u \Sigma v}{n}}{\Sigma v^2 - \frac{(\Sigma v)^2}{n}} = \frac{1500 - \frac{(-24)(4)}{8}}{2012 - \frac{(4)^2}{8}} = 0.752$$

Coefficient of regression of y on x

$$b_{yx} = \frac{\Sigma uv - \frac{\Sigma u \Sigma v}{n}}{\Sigma u^2 - \frac{(\Sigma u)^2}{n}} = \frac{1500 - \frac{(-24)(4)}{8}}{1314 - \frac{(-24)^2}{8}} = 1.217$$

The line of regression x on y

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - 75 = 0.752(y - 125.5)$$

$$\Rightarrow x = 0.752y - 0.752(125.5) + 75$$

$$\Rightarrow x = 0.752y - 19.376$$

Q> Find eq^n of line of regression based on given data :-

x : 4 2 3 4 2

y : 2 3 2 4 4

Q> Find two lines of regression from following data :

Age of Husband :	25	22	28	26	35	20	22	40	20	18
Age of wife :	18	15	20	17	22	14	16	21	15	14

Estimate: i> Age of husband when age of wife is 19

ii> Age of wife when age of husband is 30

iii> Correlation coefficient b/w them

Regression coefficient (b_{xy} & b_{yx}) have same sign & are always less than 1

Q> If $2x+3y=7$ and $5x+4y=9$ are two lines of regression. Find:-

1> The mean value of x & y $(\bar{x}, \bar{y})$

2> The regression coefficient (b_{xy}, b_{yx})

3> The correlation coefficient b/w x & y.

Sol As we know, two regression lines x on y & y on x always intersect at the mean $(\bar{x}, \bar{y})$

∴ Given, $2\bar{x}+3\bar{y}=7$ —① (Multiplying by 5)

$5\bar{x}+4\bar{y}=9$ —② (Multiplying by 2)

Solving ① & ②,

$$10\bar{x}+15\bar{y}=35$$
$$10\bar{x}+8\bar{y}=18$$
$$- \quad - \quad -$$
$$7\bar{y}=17$$

$$\boxed{\bar{y}=17/7}$$

Putting $\bar{y}$ in eq ①

$$2\bar{x}+3\left(\frac{17}{7}\right)=7$$

$$2\bar{x}=7-\frac{51}{7}=\frac{49-51}{7}=\frac{-2}{7}$$

$$\boxed{\bar{x}=-\frac{1}{7}}$$

Regression coefficient :-

y on x

$$3y=-2x+7$$
$$y=-\frac{2}{3}x+\frac{7}{3}$$
$$\boxed{b_{yx}=\frac{-2}{3}}$$

x on y

$$5x=-4y+9$$
$$x=-\frac{4}{5}y+\frac{9}{5}$$
$$\boxed{b_{xy}=\frac{-4}{5}}$$

Correlation coefficient, $r=\sqrt{b_{xy}\,b_{yx}}=\sqrt{\left(\frac{-2}{3}\right)\left(\frac{-4}{5}\right)}=\sqrt{\frac{8}{15}}$

$$\boxed{r=0.73}$$

Measures of Central Tendency

⇒ **Mean** → ① $\boxed{\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}}$

② Short cut method → $\boxed{\bar{x} = A + \dfrac{\sum f_i d_i}{\sum f_i}}$

where, $d_i = x_i - A$

A = Assumed mean.

③ Step-deviation method → $\boxed{\bar{x} = A + \dfrac{\sum f_i u_i}{\sum f_i} \times h}$

where, $u = \dfrac{x_i - A}{h}$

A = assumed mean, h = interval

⇒ **Median** → ① If n is odd, $\boxed{\text{median} = \left(\dfrac{n+1}{2}\right)^{th} \text{term}}$

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

③ For grouped data, $\boxed{\text{median} = \left[l + \dfrac{\left(\frac{N}{2}\right) - P.C.F}{f} \right] \times h}$

where, l = lower limit

f = frequency

P.C.F = Previous Cumulative frequency

h = class size

⇒ **Mode** → $\boxed{\text{Mode} = l + \left[\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \right] \times h}$

Q> Find median :-

i> 0, 1, 3, 6, 10, 12, 15

∵ n = 7 (n is odd)

∴ Median = $\left(\frac{n+1}{2}\right)^{th}$ term = $\left(\frac{7+1}{2}\right)^{th}$ term = 4th term

$\boxed{\text{Median} = 6}$

ii> 3, 4, 4, 5, 10, 12, 30, 57

∵ n = 8 (n is even)

∴ Median = $\dfrac{\left(\frac{n}{2}\right)^{th} \text{ term} + \left(\frac{n}{2}+1\right)^{th} \text{ term}}{2} = \dfrac{\left(\frac{8}{2}\right)^{th} + \left(\frac{8}{2}+1\right)^{th}}{2}$

$= \dfrac{4^{th} \text{ term} + 5^{th} \text{ term}}{2} = \dfrac{5+10}{2}$

$\boxed{\text{Median} = 7.5}$

iii>

	f	c.f.
0 − 10	3	3
10 − 20	4	7
20 − 30	3	10
30 − 40	2	12
40 − 50	1	13
h = 10		∴ n = 13

∵ n is odd

$\left(\frac{n+1}{2}\right)^{th}$ term = $\left(\frac{13+1}{2}\right)^{th}$ = 7th term

Median = $l + \left[\dfrac{\frac{n}{2} - p.c.f}{f}\right] \times h$

$= 10 + \left[\dfrac{\frac{13}{2} - 3}{4}\right] \times 10$

$= 10 + \frac{35}{4} = \frac{75}{4}$

$\boxed{\text{Median} = 18.7}$

Q> Find Mode :-

	f	
0 − 10	3	f_0
10 − 20	4	f_1 (max f)
20 − 30	3	f_2
30 − 40	2	
40 − 50	1	

Mode = $l + \dfrac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \times h$

$= 10 + \dfrac{(4+3)}{(8-3-3)} \times 10$

$\boxed{\text{Mode} = 15}$

Q> Find mean :-

	f	mid value (x)	$f.x$	$d_i = x - 25$	$f_i d_i$
0-10	3	5	15	-20	-60
10-20	4	15	60	-10	-40
20-30	3	25	75	0	0
30-40	2	35	70	10	20
40-50	1	45	45	20	30
	$\Sigma f = 13$		$\Sigma fx = 265$		$\Sigma f_i d_i = -60$

Mean i.e $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{265}{13}$ $\boxed{\bar{x} = 20.38}$

OR

Mean i.e $\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i} = 25 + \frac{(-60)}{13} \Rightarrow \boxed{\bar{x} = 20.4}$

OR

Mean i.e $\bar{x} = A + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h$

where, $u_i = \frac{x_i - A}{h}$

$\bar{x} = 25 + \frac{(-6)}{13} \times 10$

$\boxed{\bar{x} = 20.4}$

$u_i = \frac{x_i - 25}{10}$	$f_i u_i$
-2	-6
-1	-4
0	0
1	2
2	2
	$\Sigma f_i u_i = -6$

Q> From the following data of height of 100 people in a commercial concern, determin the mode /modal height.

(S.P. Gupta Pg 233) eg 24

Grouping & analysis method:-

Column 2 → Sum of two freq 2×2

Column 3 → Sum of three freq 3×3

Column 4 → Skipping first freq, Sum of two freq -2×2

Column 5 → Skipping first freq, Sum of three freq -3×3

Column 6 → Skipping two freq, Sum of three freq -3×3

Mark highest grouped frequencies for each Col in analysis table

	Col 1 (f)	Col 2	Col 3	Col 4	Col 5	Col 6
58	4	4+6 =				
		10	4+6+5 =			
60	6		15	6+5 =		
				11	6+5+10 =	
61	5	5+10 =			21	
		15				5+10+20 =
62	10			10+20 =		35
			10+20+22=	30		
63	20	20+22 =	52			
		42			20+22+24=	
64	22			22+24 =	66	
				46		22+24+6 =
65	24	24+6 =				52
		30	24+6+2 =			
66	6		32	6+2 =		
				8	6+2+1 =	
68	2	2+1 =			9	
		3				
70	1					

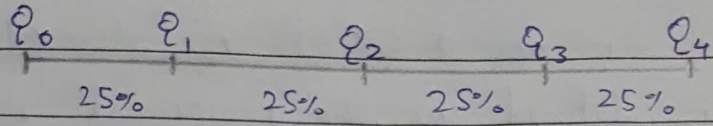
Analysis table :-

	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Total
62			1				1
63		1	1		1		3
64		1	1	1	1	1	5
65	1			1	1	1	4
66						1	1

∴ 65 repeats the maximum number of times

∴ $\boxed{\text{Mode} = 65}$

Quartile: (Percentage)



$Q_1 = l + \left[\dfrac{\frac{N}{4} + P.C.f.}{f}\right] \times h$ $\qquad$ $Q_2 = l + \left[\dfrac{\frac{N}{2} + P.C.f}{f}\right] \times h$

$Q_3 = l + \left[\dfrac{\frac{3N}{4} + P.C.f}{f}\right] \times h$ $\qquad$ Semi intequartile range $= \dfrac{Q_3 - Q_1}{2}$

Interquartile range $= Q_3 - Q_1$

Relation between Mean, Median, Mode

$$\boxed{\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})}$$

OR

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

Geometric Mean [Jab data suddenly increase ho]

If $x_1, x_2, x_3 \;---\; x_n$ are a set of n observations then geometric mean is given by,

$$\boxed{G.M. = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \;---\cdot x_n}}$$ (for Ungrouped data)

or $$\boxed{G.M = \sqrt[n]{(x_1)^{f_1} \cdot (x_2)^{f_2} \cdot (x_3)^{f_3} \;---\; (x_k)^{f_k}}}$$ (for Grouped data)

where, $n = f_1 + f_2 + f_3 + \;---\; + f_k$

Geometric mean is a type of average that is usually used for growth rate like population, interest rate.

It is used only for positive numbers.

Harmonic Mean

If $x_1, x_2, x_3 \dots x_n$ are a set of (n) observations the harmonic mean is defined as the reciprocal of the Arithmatic mean of reciprocal of quantity.

i.e

$$\boxed{HM = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots \frac{1}{x_n}\right)}}$$

Harmonic mean is useful if the data value are in ratio of two variables with different units.

Module - 5

Applied Statistics

Curve fitting by method of Least squares

- * Fitting of Straight line
- * Fitting of Second Degree Parabala
- * General Curves

⇒ For straigtht line

Let equation of straight line be

$$y = a + bx \quad -①$$

The normal eqⁿ is $\Sigma y = na + b\Sigma x \quad -②$

and

$$\Sigma XY = a\Sigma x + b\Sigma x^2 \quad -③$$

Solve eq ② & ③ to get values of a & b then put values of a & b in eq ① and obtain the required eqⁿ of straight line.

⇒ For Second Degree Parabola

Let equation of parabala be ~~$y^2=4ax$~~

$$y = a + bx + cx^2 \quad -①$$

The normal eqⁿ is

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad -②$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad -③$$

$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad -④$$

Solve eq ②, ③, ④ to get values of a, b, c then put values of a, b & c in eq ① to obtain the required equation of parabola

Q>

x	-3	-1	1	3
y	15	5	1	1

Calculate fitting second degree parabola - Curve fitting using least square method

x	y	x^2	xy	x^3	x^2y	x^4
-3	15	9	-45	-27	135	81
-1	5	1	-5	-1	5	1
1	1	1	1	1	1	1
3	1	9	3	27	9	81
$\Sigma x=0$	$\Sigma y=22$	$\Sigma x^2=20$	$\Sigma xy=-46$	$\Sigma x^3=0$	$\Sigma x^2y=150$	$\Sigma x^4=164$

Let the eq^n of parabola be $y = a + bx + cx^2$ — Ⓐ

Normal eq^n of parabola, $\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2$ — ①

$$22 = 4a + b(0) + c(20)$$

$$\boxed{4a + 0b + 20c = 22} \text{ — ②}$$

$$\Sigma xy = a\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3$$

$$-46 = a(0) + b(20) + c(0)$$

$$20b = -46$$

$$\boxed{b = -2.3}$$

$$\Sigma x^2y = a\Sigma x^2 + b\,\Sigma x^3 + c\Sigma x^4$$

$$150 = a(20) + (-2.3)(0) + c(164)$$

$$\boxed{20a + 164c = 150} \text{ — ③}$$

From eq ② & ③, $\boxed{a = 2.375}$, $\boxed{c = 0.625}$

Putting a, b, c in eq $y = a + bx + cx^2$ (eq Ⓐ)

$$\boxed{y = 2.375 - 2.3x + 0.625x^2}$$ This is the required eq^n of Parabola

Q> Fit a second degree parabola:

x :	1	2	3	4	5
y :	1090	1220	1390	1625	1915

x	y	$U = x - 3$	$V = y - 1220$	$\frac{U}{1}$	$\frac{V}{5}$ H.C.F
1	1090	-2	-130	-2	-26
2	1220	-1	0	-1	0
3	1390	0	170	0	34
4	1625	1	405	1	81
5	1915	2	695	2	139
		~~$\Sigma U \neq 0$~~	~~Σx~~	$\Sigma U = 0$	$\Sigma V = 228$

U^2	UV	U^3	U^2V	U^4
4	52	-8	-104	16
1	0	-1	0	1
0	0	0	0	0
1	81	1	81	1
4	278	8	556	16

Let the eq^n of second degree parabola be $\boxed{V = a + bU + CU^2}$ ①

The normal eq^n of parabola, $\Sigma V = na + b\Sigma U + c\Sigma U^2$ ②

$\Sigma UV = a\Sigma U + b\Sigma U^2 + c\Sigma U^3$ ③

$\Sigma U^2 V = a\Sigma U^2 + b\Sigma U^3 + c\Sigma U^4$ ④

Putting values in eq^n ②, ③, ④

$228 = 5a + b(0) + c(10)$

$411 = a(0) + b(10) + c(0)$

$533 = a(10) + b(0) + c(34)$

which gives, $228 = 5a + 10c$ ⑤

$411 = 10b$ ⑥

$\boxed{b = 41.1}$

$533 = 10a + 34c$ ⑦

from ⑤ & ⑦ $a = 34.6$, $c = 5.5$

Putting value of a, b & c in eqⁿ ①

$$v = 34.6 + 41.4u + 5.5u^2$$

$$\frac{y-1220}{5} = 34.6 + 41.1(x-3) + 5.3(x-3)^2$$

$$y - 1220 = (34.6)5 + (41.1)5(x-3) + 5(5.5)(x^2-6x+9)$$

$$y - 1220 = 173 + 205.5x - 616.5 + 27.5x^2 - 165x + 247.5$$

$$\boxed{y = 27.5x^2 + 40.5x + 1024}$$

Q1> Fit a curve $y = ab^x$ to the following data.

x:	2	3	4	5	6
y:	144	172.8	207.4	248.8	298.5

Q2> Fit a curve $y = ax^b$ to the following data

x:	20	16	10	11	14
y:	20	41	120	89	56

Sol 1>

x	y	$Y = \log_{10} y$	xY	x^2
2	144	2.1584	4.3168	4
3	172.8	2.2375	6.7125	9
4	207.4	2.3168	9.2672	16
5	248.8	2.3959	11.9795	25
6	298.5	2.4749	14.8494	36
$\Sigma x^2 = 20$		$\Sigma Y = 11.5835$	$\Sigma xY = 47.1254$	$\Sigma x^2 = 90$

taking $\log_{10}$ both side

$$\log_{10} y = \log_{10}(a \cdot b^x)$$

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

$$Y = A + xB$$

where, $Y = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} b$

The normal eqⁿ is $\Sigma Y = nA + B\Sigma x$

$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

Measures of Dispersion (Module 4)

The measure of central tendency do exhibit one of the important characteristic of a distribution, yet they fail to give any idea as to how the individual values ~~deform~~ differ from the central values i.e whether ~~ty~~ they are closely packed around the central value or widely scattered away from it. Two distribution may have the same mean & same total frequency, for example :-

A	B	C
100	102	1
100	103	2
100	98	3
100	97	490
100	100	4
500	500	500

mean, $\bar{x}_A = 1\cdot00$, $\bar{x}_B = 100$ $\bar{x}_C = 100$

Definition - Dispersion is the measure of variation of items
by Bowley

Definition - The degree of which numerical data tends to spread
by Spiegal about an average value is called dispersion of the (variation) data.

Methods to find measures of dispersion

① Range - Range is given by the difference between the greatest and least value in the distribution

i.e $\boxed{\text{Range} = \text{Max value} - \text{Min value}}$

② Quartile - deviation $\boxed{\dfrac{Q_3 - Q_1}{2} = \text{Quartile deviation}}$

③ **mean deviation** – mean deviation $= \dfrac{\Sigma f_i |(x_i - A_i)|}{\Sigma f_i = N}$

④ **Standard deviation** – (i) $SD = \sigma = \sqrt{\dfrac{\Sigma f_i (x_i - A)^2}{\Sigma f_i}}$

(ii) Short Cut Method

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}$$

(iii) Step deviation

$$\sigma = \sqrt{\frac{\Sigma f_i u_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)^2} \times h$$

where, $u_i = \dfrac{x - a}{h}$

$$\text{variance} = \sigma^2 = \frac{\Sigma f_i (x_i - A)^2}{\Sigma f_i}$$

⑤ **Coefficient of variation** – Coefficient of variation $= \dfrac{\sigma}{\bar{x}} \times 100$ → for %

Q> Calculate mean ~~&~~ standard deviation & coeff of variation, QD, MD of the following.

Size of x items : 6 7 8 9 10 11 12

Frequency : 3 6 9 13 8 5 4

x	f_i	$d_i = x - 9$	$f_i d_i$	d_i^2	$f_i \times d_i^2$
6	3	−3	−9	9	27
7	6	−2	−12	4	24
8	9	−1	−9	1	9
9	13	0	0	0	0
10	8	1	8	1	8
11	5	2	10	4	20
12	4	3	12	9	36
	$\Sigma f_i = 48$		$\Sigma f_i d_i = 0$		124

Applying shortcut method,

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2} = \sqrt{\frac{114}{48} - \left(\frac{0}{48}\right)^2} = 1.61$$

$$\boxed{\sigma = 1.61}$$

Coefficient of variation $= \frac{\sigma}{x} \times 100 = \frac{1.61}{9} \times 100 = 17.8\%$

$QD = \frac{2}{3}(\sigma) = \frac{2}{3} \times 1.61 = 1.07$

$M.D = \frac{4}{5}(\sigma) = \frac{4}{5} \times 1.61 = 1.28$

Q) The following are scores of two batsman A & B in the, who is better ~~st~~ scorer & consistent.

A	$d = x - 50$	d_i^2
12	-38	1444
115	65	4225
6	-44	1936
73	23	529
7	-43	1849
19	-31	961
119	69	4761
36	-14	196
84	34	1156
29	-21	441
$\Sigma A = 500$	0	$\Sigma d_i^2 = 17498$
$\bar{x} = \frac{500}{10} = 50$		

$\sigma = \sqrt{\frac{17498}{10}}$

$\sigma_A = 41.8$

Coeff of variation $= \frac{\sigma_A}{\bar{x}_A} \times 100$

$= \frac{41.8}{50} \times 100$

$= 83.6\%$

$\bar{x} = 27$

$\sigma_B = 18.8$

$CV = \frac{18.8}{27} \times 100 = 69.6\%$

Since, arithematic mean of A = 50 & $\bar{x}$ B = 27, it means A is a

better scorer ~~that~~ than B.
But, since the coefficient of variation of B is less than coefficient of variation of A $\therefore$ B is more consistent than A.

Q> Calculate the mean & standard deviation of following freq distribution:

Weekly Wages in Rupees	No. of man (f)	mid value (x)	$d = \frac{x-A}{8} = \frac{x-32.5}{8}$	$f_i d_i$
4.5−12.5	4	8.5	−3	−12
12.5−20.5	24	16.5	−2	−48
20.5−28.5	21	24.5	−1	−21
28.5−36.5	$\boxed{18}$	Ⓐ $\boxed{32.5}$	0	0
36.5−44.5	5	40.5	1	5
44.5−52.5	3	48.5	2	6
52.5−60.5	5	56.5	3	15
60.5−68.5	8	64.5	4	32
68.5−76.5	2	72.5	5	10
	$\Sigma f_i = 90$			$\Sigma f_i d_i = -13$

d_i^2	$f_i \cdot d_i^2$
9	36
4	96
1	21
0	0
1	5
4	12
9	45
16	128
25	50
	$\Sigma f_i d_i^2 = 393$

$$\text{mean} = A + \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right) \times h$$

$$= 32.5 + \left(\frac{-13}{90}\right) \times 8$$

$$= 32.5 - \frac{13 \times 4}{45}$$

$$\boxed{\bar{x} = 31,35} \text{ Ans}$$

$$SD = \sigma = h \times \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}$$

$$= 8 \times \sqrt{\frac{393}{90} - \left(\frac{-13}{90}\right)^2}$$

$$\boxed{\sigma = 16.64}$$

Q> Calculate median & the lower & upper quartile from the distribution of marks obtained by 49 students in the following frequency table:
Find also the semi~~n~~-interquartile range & the mode.

Marks	Students ~~Stden~~ (f)	C.f
5-10	5	5
10-15	6 f_0	11
15-20	15 f_1	26
20-25	10 f_2	36
25-30	5	41
30-35	4	45
35-40	2	47
40-45	2	49

$n = 49$ ∵ n is odd

$$\text{median} = \left(\frac{n+1}{2}\right)^{th} \text{term} = \left(\frac{49+1}{2}\right)^{th} \text{term}$$

$$= 25^{th} \text{ term}$$

which falls in interval (15-20)

$$Q_2 = \text{Median} = l + \left(\frac{(N/2) - P.C.f}{f}\right) \times 5$$

$$= 15 + \left[\frac{(49/2) - 11}{15}\right] \times 5$$

$$\boxed{Q_2 = 19.5 \text{ marks}}$$

$$Q_1 = l + \frac{(N/4) - P.C.f}{f} \times h.$$

$$\boxed{Q_1 = 15.41 \text{ marks}}$$

$$Q_3 = l + \left(\frac{(3N/4) - P.C.f}{f}\right) \times h$$

$$\boxed{Q_3 = 25.75 \text{ marks}}$$

$$\text{Semi-interquartile range} = \frac{Q_3 - Q_1}{2} = \frac{25.75 - 15.5}{2} = \boxed{5.17}$$

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_2 - f_0}\right) \cancel{\times} h = 15 + \left(\frac{15-6}{30-10-6}\right) \times 5$$

$$= 15 + \frac{9}{14} \times 5 = 15 + \frac{45}{14}$$

$$\boxed{\text{mode} = 18.21} \rightarrow$$ 15 students lies mostly around 18.21 marks

* Coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$ = CQD

Quartile are used to calculate measures of variability around the median.

Moment (mean, variance se jyada Power wala formula)

Moment i.e x is used to represent~~ed~~ the deviation of any item X in a distribution from arithematic mean $\bar{X}$ of that distribution.

If we take the mean of first power of the deviation, we get first moment about mean.

Similarly, the mean of second power or square of deviation, we get second moment about mean.

[We can do it till fourth moment about mean

Thus, when frequency is given, the r^{th} moment about mean is

$$\boxed{\mu_r = \frac{\sum_{r=0}^{n} f_i (x_i - \bar{x})^r}{N}}$$

For assumed mean, ~~μ_r'~~, the r^{th} moment about any point a (it is also called raw moment)

assumed mean

$$\boxed{\mu_r' = \frac{\sum_{r=0}^{n} f_i (x_i - a)^r}{N}}$$

For r = 0, $\mu_0 = \frac{\sum f_i (x - \bar{x})^0}{N} = \frac{\sum f_i}{N} = \frac{N}{N} = 1$

and $\mu_0' = \frac{\sum f_i (x - a)^0}{N} = \frac{\sum f_i}{N} = \frac{N}{N} = 1$

$\therefore \boxed{\mu_0 = \mu_0'}$

For r = 1, $\mu_1' = \frac{\sum f_i (x - a)^1}{N} = \frac{\sum f_i x_i - \sum f a}{N} = \frac{\sum fx}{N} - \frac{a \sum f}{N}$

$\therefore \boxed{\mu_1' = \bar{x} - a}$

$$\mu_1 = \frac{\sum f_i (x_i - \bar{x})^1}{N}$$

$$= \frac{\sum f_i x_i}{N} - \bar{x}\frac{\sum f_i}{N}$$

$$= \bar{x} - \bar{x}$$

$\boxed{\mu_1 = 0}$

Applying binomial theorem,

$$\mu_r = \frac{1}{N} \sum f_i (x_i - \bar{x})^r .$$

$$= \frac{1}{N} \sum f_i (x_i - a - \bar{x} + a)^r$$

$$= 1/N \sum f_i [x_i - a - (\bar{x} - a)]^r$$

$$= 1/N \sum f_i (x_i - d)^r$$

$$= 1/N \sum f_i [{}^nC_0 x_i^r d^0 - {}^nC_1 x^{r-1} d^1 + {}^nC_2 x^{r-2} d^2 + \dots + {}^nC_r x_i^0 d^r + \dots]$$

$$= 1/N \sum f_i [{}^nC_0 x_i^r - {}^nC_1 x_i^{r-1} d + {}^nC_2 x_i^{r-2} d^2 + \dots]$$

$$= \frac{1}{N} \sum f_i (x_i - a)^r - \frac{1}{N} {}^nC_1 \sum f_i (x_1 - a)^{r-1} d + {}^rC_r \frac{1}{N} \sum f_i (x_i - a)^{r-2} d^2 +$$

$$\mu_r = \mu_r' - {}^rC_1 d \mu_{r-1}' + {}^rC_2 d^2 \mu_{r-2}' - \dots \quad ①$$

For $r=2$, put in eqⁿ ①

$$\mu_2 = \mu_2' - {}^2C_1 d \mu_{2-1}' + {}^2C_2 d^2 \mu_{2-2}'$$

$$= \mu_2 - 2d\mu_1' + 1 d^2 \mu_0'$$

$$= \mu_2 - 2\mu_1' \mu_1' + (\mu_1')^2 \cdot 1 \qquad (\mu_1' = (x - a) = d)$$

$$= \mu_2 - 2(\mu_1')^2 + (\mu_1')^2 \qquad \mu_0' = 1$$

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

Similarly, for $r = 3$, $\mu_3 = \mu_3' - {}^3C_1 d \mu_{3-1}' + {}^3C_2 d^2 \mu_{3-2}' - {}^3C_3 d^3 \mu_{3-3}'$

$$= \mu_3' - 3\mu_1'\mu_2' + 3\mu_1'^2\mu_1' - 1\cdot\mu_1'^3\mu_0'$$

$$= \mu - 3\mu_1'\mu_2' + 3\mu_1'^3 - \mu_3'^3\cdot 1$$

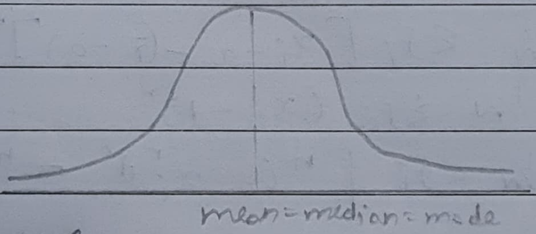
$$\boxed{\mu_3 = \mu_3' - 3\mu_1'\cdot\mu_2' + 2\mu_1'^3}$$

Skewness

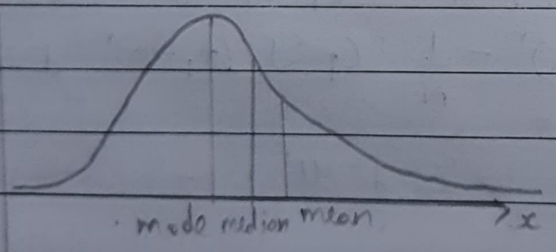
Skewness measures the degree of Asymmetry ~~&~~ or departure from symmetry.

→ If the frequency curve has a longer tail to the right ~~of the made~~ i.e the mean is to the right of the mode, ~~then it is~~ distribution is said to be positive skewness.

→ If the curve is more enlarge to the left then it is said to be negative skewness.

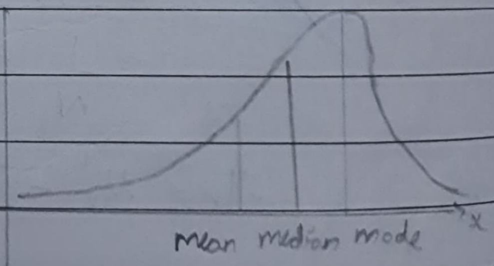


Normal Curve



Positively Skewed

mean median mode



Negative Skewed

mean median mode

Formula:-

1> Karl pearson's coeff of skewness $= \dfrac{\text{mean} - \text{mode}}{\text{S.D}} = S_k$

2> Quartile coeff of skewness $= \dfrac{Q_3 - Q_1 - 2Q_2}{Q_3 - Q_1}$

3> Coefficient of skewness based on third moment,

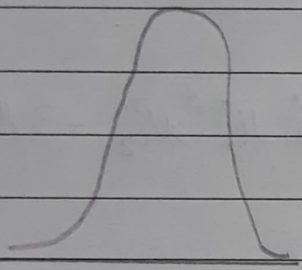
$$\gamma_1 = \sqrt{\beta_1}$$

where, $\beta_1 = \dfrac{\mu_3^2}{\mu_2^2}$

If $\beta_1 = 0$, then distribution is normal.

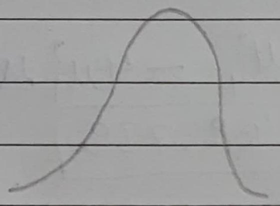
Kurtosis

Kurtosis measures the degree of peakedness of the distribution, It is denoted by β_2.



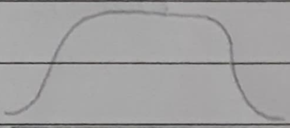
leptokurtic

$(\beta_2 > 3)$



mesokurtic

$(\beta_2 = 3)$



platykurtic

$(\beta_2 < 3)$

where, $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$ & $\gamma_2 = \beta_2 - 3$

Q> The first ~~a~~ four moment about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moment about the mean & also evaluate β_1, β_2 & comment upon the skewness and kurtosis of the distribution.

Sol $\mu_1' = 0.294$, $\mu_2' = 7.144$, $\mu_3' = 42.409$, $\mu_4' = 454.95$,

$a = 28.5$

To find, $\mu_1, \mu_2, \mu_3, \mu_4, \beta_1, \beta_2$

$$\mu_1' = \frac{1}{N} \sum f_i (x_i - a) = \frac{1}{N} \sum f_i (x_i - 28.5)$$

$$0.294 = \frac{\sum f_i x_i}{N} - \frac{28.5 \sum f_i}{N}$$

$0.284 = \bar{x} - 28.5$

$\therefore \bar{x} = 0.294 + 28.5$

$\boxed{\mu_1 = 28.794}$

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

$= 7.144 - (0.294)^2$

$\boxed{\mu_2 = 7.058}$

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

$= 42.409 - 3(0.294)(7.144) + 2(0.294)^3$

$\boxed{\mu_3 = 36.151}$

Now, $\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4$

$\boxed{\mu_4 = 408.738}$

$\beta_1 = \frac{\mu_3^2}{\mu_2^2} = \frac{(36.151)^2}{(7.058)^2} = 3.717$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{408.738}{(7.058)^2} = 8.205$

$\boxed{\beta_2 > 3}$

~~$\beta_1 = \frac{\mu_3^2}{\mu_2^2} = \frac{(36.151)^2}{(7.058)^2} = 3.717$~~

~~$\beta_2 = \frac{\mu_4}{\mu}$~~

$\gamma_1 = \sqrt{\beta_1} = \sqrt{3.717} = \boxed{1.928}$

$\gamma_2 = \beta_2 - 3 = 8.205 - 3 = \boxed{5.205}$

($\therefore$ it is laptokurtic)

Samples ⟶ small sample
(n<30)
↓
large sample
(n>30)

Date
Page

Test of significance (Unit -5)

Topics :-

i> Large sample test for single proportion

ii> difference of proportions

- a> single mean
- b> difference of means
- c> difference of standard deviation.

Sampling - * A small section selected from the population is called a sample and process of drawing a sample is called sampling.

*It is necessary that a sample ~~ment~~ must be a random selection so that each member of population has the same chance of being included in the sample.

* It is also referred as random sampling.

* The statistical constant of the population such as mean (μ) and standard deviation (σ) are called 'parameters'

&

Similarly, constants of the sample drawn from the given population i.e mean

Objective of sampling :-

* Sampling aims at gathering maximum information about population with minimum effort, cost and time.
* The object of sampling studies is to obtain the best possible value of the parameter under specific condition, sampling determines the reliability of the estimates.

Standard Error (S.E.) The standard deviation of sampling distribution are called standard error.

* Similarly, the standard error of the sampling distribution of means is called Standard Error of means.

* In general the standard error is used to find the difference between expected and observed values.

Note → The reciprocal of standard error is called precision.

→ If $n > 30$ a sample is called 'large' otherwise 'small'.

Hypothesis * To reach decision about population on the basis of ~~stan~~ sample information i.e we make certain assumption about the population, such assumption which may or may not be true is called statistical hypothesis.

* By testing a hypothesis it is meant, a process for deciding whether to accept or reject the hypothesis.

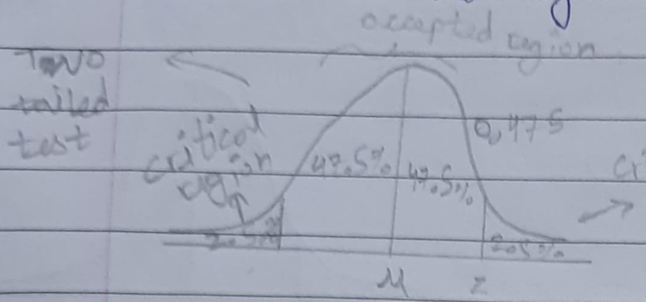
Types of hypothesis :-

i) Null hypothesis

The hypothesis v for the sake of rejecting it under
the |Formulated| assumption that it is true is
called Null hypothesis and is denoted by H_0.

ii> Alternate hypothesis

*The hypothesis when the null hypothesis which is rejected, is called alternate hypothesis.
It is denoted by H_1.



critical region/rejection region →

if $z = \frac{\text{Diff}}{\text{S.E}}$ is in critical region i.e $-1.96 > z$ or $z > 1.96$, null hypothesis H_0 is rejected else accepted

range → $(\mu - 1.96, \mu + 1.96)$ → confidence limit
acceptable region

Large Sample Test

Types of questions

i> Test of number of successes
ii> Test of proportion of successes
iii> Test of difference of proportions

$$\boxed{Z = \frac{\text{Difference}}{\text{Standard error}}}$$

$S.E = \sqrt{npq}$

Q> In 324 throws of a six faced die odd points appeared 181 times. Would you say that the die is fair/unbaised? State carefully the property on which you base your calculation.

Sol $n = 324$, $p = \frac{\text{prob of odd number}}{\text{Total prob}} = \frac{3}{6} = \frac{1}{2}$, $q = 1 - p$
$q = \frac{1}{2}$

Step 1> null hypothesis H_0 = the die is unbaised : $P = \frac{1}{2}$

Step 2> Alternate hypothesis H_1 = the dies is biased

Step 3> Level of significance, $\alpha = 5\% = 0.05$

Step 4> Calculation, $S.E. = \sqrt{npq} = \sqrt{324 \times \frac{1}{2} \times \frac{1}{2}}$

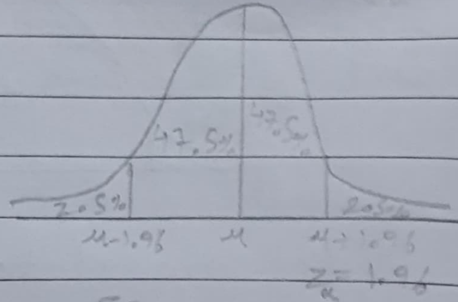
$\boxed{S.E = 9}$

The expected number of odd point in 324 throws $= \frac{324 \times 1}{2} = 162$

but actual value given is 181

Difference $= 181 - 162$

$\boxed{\text{Difference} = 19}$



[fig in Los step]

$$Z = \frac{\text{Difference}}{S.E.} = \frac{19}{9}$$

$\boxed{Z = 2.11}$

Step 5>
Conclusion, since, $|Z| > Z_\alpha$

$2.11 > 1.96$

$\therefore$ Null hypothesis, H_0 is rejected and Alternate hypothesis H_1 is accepted i.e. die is biased

Q> In a sample of 1000 people of Maharastra, 540 are rice eaters & rest are wheat eaters. Can be assume that both rice & wheat are equally popular in this state at 1% level of significance.

Sol Step1
given → $n = 1000$

$X = 540$

$$p = \frac{X}{n} = \frac{540}{1000} = 0.54$$

Step2
null hypothesis, H_0: both rice and wheat are equally popular in the state

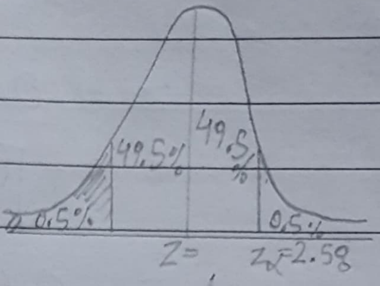
H_0: $P = 1/2$

Alternate hypothesis : H_1 : $P \neq 1/2$

$Q = 1 - P = 1 - \frac{1}{2} = 1/2$

Step 3 Level of signifance, LOS : 1% ⇒ $Z_\alpha = 2.58$
(Tabulated value)

Step 4 Calculation :-



$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$z = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}}$$

$$\boxed{z = 2.532}$$

Step 5 Conclusion :-
Since $|z| < z_\alpha$
$2.532 < 2.58$
i.e H_0 is accepted
Hence, rice & wheat are equally popular in the state

20 people work attacked by a disease and and 18 survived. Will you reject the hypothesis that a survival rate if 80%.

Sol **Step 1** given → $n = 20$, $x = 18$

$$p = \frac{x}{n} = \frac{18}{20} = 0.90$$

Step 2 null hypothesis, H_0 : $P = 85\% = 0.85$
i.e the proportion of peope survived after attack by a disease is 85%.
Alternate hypothesis $H_1 = P > 0.85$

Step 3 LOS : 5% ⇒ $Z_\alpha = 1.65$ (Tabulated Value)

Step 4 Calculation

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$Z = \frac{0.90 - 0.85}{\sqrt{\frac{(0.85)(0.15)}{20}}}$$

$$\boxed{Z = 0.633}$$

Step 5 ~~Calculat~~ Conclusion

Since $|Z| < Z_\alpha$

$0.633 < 1.65$

i.e H_0 is accepted

Hence,

Test of significance for difference of proportions:-

Formula:- $$z = \frac{(p_1 - p_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}$

$Q = 1 - P$, $p_1 = \frac{X_1}{n_1}$, $p_2 = \frac{X_2}{n_2}$

Q> Random sample of 400 men & 600 women were asked ~~to~~ whether they'd like to have a flyover near their residance. 200 men & 325 women were in favour of the proposal. Test the hypothesis that proportion of men & women in favour of proposal are same against that they are not at 5% level.

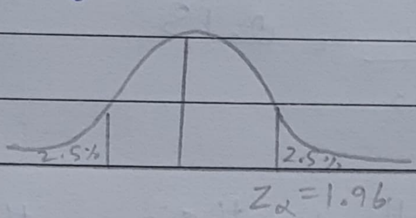
Sol $n_1 = 400, \quad X_1 = 200, \quad n_2 = 600, \quad X_2 = 325$

Proportion $p_1 = \frac{x_1}{n} = \frac{200}{400} = \frac{1}{2} = 0.5, \quad p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.54$

Step 1> null hypothesis, $H_0 : P_1 = P_2$ i.e there is the opinion of men and women as far as proposal of fly over is concerned.

Alternate hypothesis, $H_1 : P_1 \neq P_2$ (two tailed test)

Step 2> Level of significance, $\alpha = 5\%$ $\therefore Z_\alpha = 1.96$



Step 3> $P = \frac{X_1 + X_2}{n_1 + n_2} = \frac{200 + 325}{400 + 600}$

$P = 0.525$

$q = 1 - P = 0.475$

4

~~Step~~ Apply formula, $z = \frac{p_1 - p_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$z = \frac{0.5 - 0.54}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$z = -1.29$

$|z| = 1.29$

Step 4> Since $|z| < Z_\alpha \Rightarrow 1.29 < 1.96$

$\therefore H_0$ is accepted

Hence, we conclude that men and women are in favour of proposal.

Q> A cigarette manufacturer claims that it's brand A of the cigarette outsells its brand B by 8%. It is found that 42 out of a sample of 200 smoker and 18 out of another random sample of 100 smoker prefer brand B. Test whether the 8% difference is a valid claim. Use 5% Level significance.

Sol

$n_1 = 200, \quad n_2 = 100, \quad x_1 = 42, \quad x_2 = 18, \quad p_1 = \frac{x_1}{n_1} = \frac{42}{200} = 0.21$

$p_2 = \frac{x_2}{n_2} = \frac{18}{100} = 0.18$

Step-1 null hypothesis, $H_0 : P_1 - P_2 = 8\% = 0.08$

Alternate hypothesis, $H_1 : P_1 - P_2 \neq 8\%$

Step-2 Level of significance, $\alpha = 5\%$, $Z_\alpha = 1.96$

Step-3 $P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{42 + 18}{200 + 100} = \frac{60}{300}$

$P = 0.2$

$\therefore Q = 0.8$

Step 4

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z = \frac{(0.21 - 0.18) - (0.08)}{\sqrt{(0.2)(0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$z = -1.02 \qquad \therefore |z| = 1.02$

Step5 since $|z| < Z_\alpha \Rightarrow 1.02 < 1.96$

$\therefore$ Ho is accepted

Hence, we conclude that a difference of 8% in the sale of two brands of cigarettes is a valid claim of by company.

Q> A random sample of 500 pineapple was taken from a large consignment and 65 were found to be bad. Show that the standard error of the proportions of bad one's in the sample of size 0.015% and deduce that the percentage of bad pineapple in the consignment almost certainly lies b/w 8.5 and 17.5.

Sol $n = 500$, $x = 65$, $p = \frac{x}{n} = \frac{65}{500} = 0.13 = p$

$q = 1 - p = 1 - 0.13 = 0.87 = q$

$$S.E = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.13)(0.87)}{500}}$$

$S.E = 0.015$

limits for proportion of bad pineapple $= p \pm 3(S.E.)$

$= 0.13 \pm 3(0.015)$

$= 0.13 \pm 0.045$

$= (0.13 - 0.045, 0.13 + 0.045)$

$= (0.085, 0.175)$

$= (8.5\%, 17.5\%)$

Q> A random sample of 500 apple was taken large sample of consignment and 60 was found to be bad, Obtain the 98% confidence limit of bad apples in consignment.

bad apples in consignment

Test of Significance for Single Mean

Formula → $z = \dfrac{\bar{x} - \mu}{\boxed{\sigma/\sqrt{n}}}$ → standard error

Q) A sample of 900 members has a mean 3.4 cm & standard deviation 2.61 cm. Is the sample from a large population of mean 3.25 cm & standard deviation 2.61 cm. If the population is ~~nomal~~ normal and mean is unknown, find the & fiducial limits of two mean

Sol given, $\bar{x} = 3.4$ cm

$\mu = 3.25$ cm

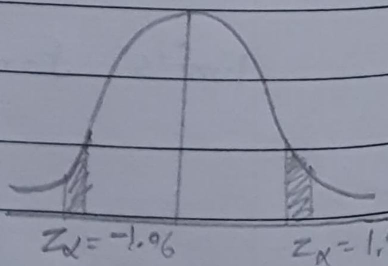
$n = 900$

$\sigma = 2.61$

null hypothesis: $H_0 : \mu = 3.25$ c.m & $\sigma = 2.61$

Alternate hypothesis: $H_1 : \mu \neq 3.25$ (two tailed test)

we know that, $z = \dfrac{x - \mu}{\sigma/\sqrt{n}}$



$$= \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}}$$

$$= \frac{0.15 \times 30}{2.61}$$

$\boxed{Z_{cal} = 1.73}$ → calculated

since, $Z_{cal} < Z_{tab}$ → tabular

$1.73 < 1.96$

∴ we conclude that H_0 is accepted at 5% level of significance.

95% fiducial limits

$\bar{x} \pm 1.96$ (Standard error)

$\Rightarrow \bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$

$\Rightarrow 3.4 \pm 1.96 \left(\frac{2.61}{\sqrt{900}}\right)$

$= 3.4 \pm 0.2027$

For (-ve)

$= 3.4 - 0.27$

$= \boxed{3.1973}$

For (+ve)

$= 3.4 + 0.2027$

$= \boxed{3.6027}$

$\Rightarrow (3.1973, 3.6027)$

Q) A company producing flurescent light bulbs claims that the average life of the bulb is ~~10~~ ~~1500~~ 1570 hrs. The mean lifetime of sample of 200 flurescent bulbs was found to be ~~1600 hrs~~ 1600 hrs. with S.D. of 150 hrs. Test for the company at 1% level of significance whether the claims that the average life of bulb is 1570 hrs is acceptable.

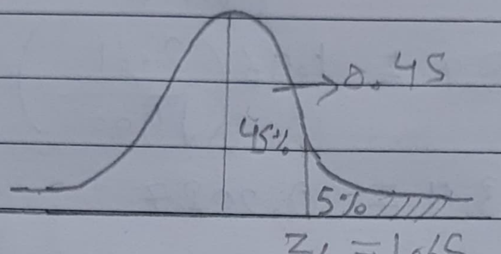
Q) A packaging device is said to fill detergent powder packets with a mean weight of 5 kg with S.D. ~~[illegible]~~ 0.31 kg. The weight of packets can be assumed to be normally distributed. The weight of packets is known to drift upward over a period of time due to machine fault, which is tolerable. A random sample has mean weight of 5.03 kg. Can we conclude that the machine has increased. Use 5% level of Signifance

Sol $n = 100$, $\bar{x} = 5.03$ kg, $\mu = 0.31$ kg

Step 1> null hypothesis, $H_0 : \bar{x} \leq \mu$

$H_1 : \boxed{\bar{x} > \mu}$

Step 2> LOS : 5% (one tailed test)



$Z_\alpha = 1.65$

Step 3> Formula, $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$$z = \frac{5.03 - 5}{0.31/\sqrt{100}} = \frac{0.03 \times 10}{0.31}$$

$$Z_{cal} = 0.96$$

Here, $Z_{cal} < Z_\alpha$

$0.96 < 1.65$

H_0 is accepted.

Step 4> We ~~cot~~ conclude that mean weight produced by machine has not increased.

Test of hypothesis for the difference between two ~~to~~ population mean.

Formulas:-

1) $z = \dfrac{\bar{x}_1 \sim \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1^2} + \dfrac{\sigma^2}{n_2^2}}}$ when the S.D. are given.

2) $$Z = \frac{\bar{x}_1 \sim \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

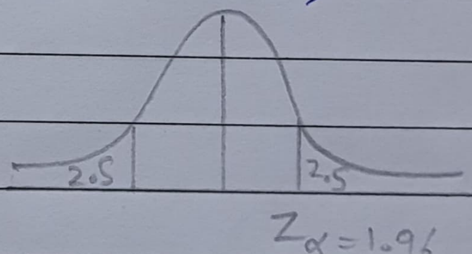
Q) The means of sample of sizes thousand and two thousand are 67.5 & 68.0 respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 cm.

Sol given, $n_1 = 1000$, $\bar{x}_1 = 67.5$ cm, $\sigma = 2.5$ cm
$n_2 = 2000$, $\bar{x}_2 = 68.0$ cm

null hypothesis : H_0 : the sample are drawn from the same population
$\boxed{H_0 : \mu_1 = \mu_2}$

$H_1 : \mu_1 \neq \mu_2$ (two tailed test)

LOS : 5%



$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$= \frac{67.5 \sim 68.0}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}}$$

$Z_{cal} = 5.1$

$\therefore Z_{cal} > Z_\alpha$
$5.1 > 1.96$

H_0 is rejected

We conclude that sample cannot drawn from the same population.

Q> ~~A~~ sample of height of 6400 soldier has a mean of 67.85 inches and a standard deviation of 2.56 inches while a ~~Sample~~ simple sample of height of 1600 sailors has a mean of 68.55 inch and SD. 2.52 inches do the data indicate ~~soldier~~ sailor ~~are~~ are on average taller than soldier.

Sol

$n_1 = 6400$

$\bar{x}_1 = 67.85, \sigma_1 = 2.56$

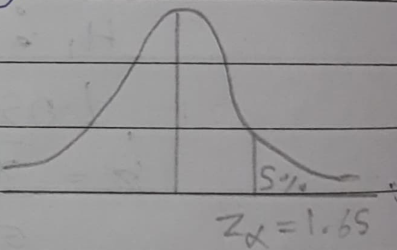
$n_2 = 1600$

$\bar{x}_2 = 68.55, \sigma_2 = 2.52$

null hypothesis, $H_0 : \mu_1 = \mu_2$

(Sailor's and Soldier's height are same)

alternate hypothesis, $H_1 : \mu_2 > \mu_1$



LoS : 5% (one tailed test)

Formula,

$$z = \frac{67.85 \sim 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}}$$

$$Z = \frac{0.9}{0.005}$$

$Z_{cal} = 1.40$

$Z_{cal} > Z_\alpha$

$140 > 1.65$

$\therefore H_0$ is rejected

H_1 is accepted

We conclude that the data indicate the sailor's are on the average taller than the ~~Solid~~ Soldiers.

$a \sim b = |a-b|$

matlab pehle esi value rakho jisse ans +ve aaye

DOF = Degree of Freedom

Test of significance for difference of Standard deviation

If S_1 & S_2 are standard deviation of two independent samples then under null hypothesis i.e sample standard deviation do not differ significantly

$H_o = \sigma_1 = \sigma_2$

$$z = \frac{S_1 \sim S_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

Small Sample

→ DK Jain 3rd Sem book

ⓐ Test of the significance of a mean (small sample)

Null hypothesis: H_o there is no significance difference in between the sample means ($\bar{x}$) and population mean (μ).

Alternative hypothesis: H_1 : There is significant diff b/w --

Calculate t-stastic

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{S}$$

where $S^2 = \frac{1}{n-1} \Sigma (x - \bar{x})^2$

⇒ when variance is known, $t = \frac{(\bar{x} - \mu)\sqrt{n-1}}{\sigma}$

where, $\sigma^2 = \frac{1}{n} \Sigma (x - \bar{x})^2$

LOS:- Find the value of t from the table i.e value of t at α% level of significance & for degree of

DOF = no. of independent values

10
1
9
12 } independent
8 → dependent on above
40

freedom $V = n-1$

if $|t|_{cal} < t_{tab} \Rightarrow H_0$ accepted (inside region)

$|t|_{cal} \not< t_{tab} \Rightarrow H_0$ rejected

Q) A shampoo manufacturing company was distributed a particular brand of shampoo through a large no. of retail shops. Before a heavy advertisement campaign, the mean scale per shampoo was 140 dozens. After the campaign, a sample of 26 shampoo was taken & mean sales figure was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective (given $t_{0.05,25} = 1.708$) → correction → 2.06

LOS DOF

Sol Given, $n = 26$, $\mu = 140$ → population mean, $SD = \sigma = 16$

$\bar{x} = 147$ → sample mean, $DOF = V = n-1 = 25$

Null hypothesis : H_0 : There is no difference in between the sample mean & population mean

Alternative hypothesis, H_1 : There is difference b/w μ & $\bar{x}$ i.e advertisement was effective

$$t = \frac{\bar{x} - \mu}{\sigma}\sqrt{n-1} = \frac{147 - 140}{16} \times \sqrt{26-1}$$

$$t = 2.19$$

$$|t|_{cal} = 2.19$$

LOS : given, $t_{0.05,25} = 2.06$ (tabular)

$\therefore$ $2.19 \not< 2.06$

$\therefore$ $|t|_{cal} \not< t_{tab}$

$\therefore$ H_0 is rejected

Thus, the advertisement was effective.

Q> 10 objects are chosen at random from a population & their heights are found to be in inches:

63, 63, 64, 65, 66, 69, 69, 70, 70, 71

Discuss the suggestion that the mean height in universe is 65 inches given that for 9 degree of freedom, the value of t at 5% level of significance is 2.262.

Sol given, $n = 10$ (small sample)

universe mean i.e $\mu = 65$

sample mean i.e $\bar{x} = \frac{\Sigma x}{n} = \frac{670}{10} = 67$

$\because \sigma$ is not given $\therefore$ applying S formula

$$t = \frac{\bar{x} - \mu}{S} \sqrt{n}$$

where, $$S^2 = \frac{\Sigma(x - \bar{x})^2}{n-1}$$

x	$(x-\bar{x})^2 = (x-67)^2$
63	16
63	16
64	9
65	4
66	1
69	4
69	4
70	9
70	9
71	16
670	$\Sigma(x-\bar{x})^2 = 88$

$$S = \sqrt{\frac{89}{9}} = 3.13$$

$\therefore$ ~~t~~ ⇒ :

Null hypothesis, H_0: $\mu = 65$ i.e there is no significant difference b/w sample mean height & population/ universal mean height.

Alternate hypothesis, H_1: $\mu \neq 65$ i.e there is difference b/w μ & $\bar{x}$

$$t_{cal} = \frac{(67 - 65)\sqrt{10}}{13} = 2.024$$

$|t_{cal}| = 2.024$

$\because$ $2.024 < 2.262$

i.e. $|t_{cal}| < t_{tab}$

$\therefore$ H_0 is accepted

Hence, ~~The~~ Mean height of universe is 65.

⇒ To test the significance of difference between two sample means.

* When standard deviation is not given,

$$t = \frac{(\bar{x}-\bar{y})}{S}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}} = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S^2 = \frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}$$

Degree of freedom, $V = n_1+n_2-2$

* When variance σ_1^2 & σ_2^2 are known or when S.D σ_1 & σ_2 are known/given.

$$t = \frac{(\bar{x}-\bar{y})}{S}\sqrt{\frac{n_1 \cdot n_2}{n_1+n_2}} = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{(Same)}$$

but $$S^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2-2}$$

Note → Find fiducial limits at 95%

when variance is Unknown, $\bar{x} \pm t_{0.05} \times \left(\frac{S}{\sqrt{n}}\right)$

when variance is known, $\bar{x} \pm t_{0.05} \times \frac{S}{\sqrt{n-1}}$

[if 99% was asked then we take $t_{0.01}$ & so on]

Q> The average no. of articles produced by two machines per day are 200 & 250 with standard deviation 20 & 25 resp. On the basis of records of 25 days production, Can you regard both the machines equally efficient at 1% level of significance. $t_{0.01,48} = 2.58$ (given).

Sol Given, $n_1 = n_2 = 25$, $\bar{x} = 200$, $\bar{y} = 250$

$\sigma_1 = 20$, $\sigma_2 = 25$

$t_{0.01,48} = 2.58$

Null hypothesis, H_0 : There is no significant difference between both machines i.e both machines are equally efficient. i.e $\mu_1 = \mu_2$

Alternate hypothesis, H_1 : $\mu_1 \neq \mu_2$ i.e There is Significant difference b/w both machines.

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

where, $$s^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2 - 2}$$

$$\therefore s = \sqrt{\frac{25 (20)^2 + 25 (25)^2}{25 + 25 - 2}}$$

$$\boxed{s = 23.1}$$

DOF = $n_1 + n_2 - 2 = 25 + 25 - 2 = 48$

at 1% level of significance

$$\therefore t = \frac{(200 - 250)}{23.1} \sqrt{\frac{25 (25)}{25 + 25}} = \frac{-50}{23.1} \sqrt{12.5}$$

$t_{cal} = -7.65$

$|t_{cal}| = 7.65$

∴ $7.65 \nless 2.58$

i.e $|t_{cal}| \nless t_{tab}$

∴ H_o is rejected & H_1 is accepted

Thus we ~~calculate~~ conclude Both machines are not equally efficient at 1% level of significance with 48 degree of freedom.

Q> Two horses A & B were tested according to time in seconds to run a particular track with following results:

A : 28 30 32 33 33 29 34

B : 29 30 30 24 27 29

$\bar{x} = \frac{\Sigma A}{n_1}$

$\bar{y} = \frac{\Sigma B}{n_2}$

Test whether you can discriminate b/w two horses. given $t_{0.05,11} = 2.2$

Test for ratio of variances

Method:- Step 1: Null hypothesis, $H_o : \sigma_1^2 = \sigma_2^2$

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

Step 2: DOF, $v_1 = n_1 - 1$, $v_2 = n_2 - 1$

Step 3: Calculate, $F = \frac{S_1^2}{S_2^2}$, $S_1^2 \geq S_2^2$

where $S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$, $S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1}$

Step 4: $F_{cal} < F_{tab}$, H_o is accepted

$F_{cal} \nless F_{tab}$, H_o is rejected

Q) Random Sample are drawn from two populations and the following results were obtained.

Sample X : 20 16 26 27 23 22 18 24 25 19
Sample Y: 27 33 42 35 32 34 38 28 41 43 30 37

Find the variance of two populations & test whether the two samples have same variance (given that $F_{0.05}$ for 11 & 9 DOF is 3.112).

Sol given, $n_1 = 10$, $n_2 = 12$

DOF $V_1 = n_1 - 1 = 10 - 1 = 9$

$V_2 = n_2 - 1 = 12 - 1 = 1$

Null hypothesis : Let $\sigma_1^2 = \sigma_2^2$, the two samples have the same variance

Alternate hypothesis : ~~Let~~ $\sigma_1^2 \neq \sigma_2^2$

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{220}{10} = 22, \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{420}{12} = 35$$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{\Sigma(x-22)^2}{9}$$

$$S_1^2 = \frac{(20-22)^2 + (16-22)^2 + (26-22)^2 + (27-22)^2 + (23-22)^2 + (22-22)^2 + (18-22)^2 + (24-22)^2 + (25-22)^2 + (19-22)^2}{9}$$

$$S_1^2 = \frac{4 + 36 + 16 + 25 + 1 + 0 + 16 + 4 + 9 + 9}{9}$$

$$S_1^2 = \frac{120}{9} = 13.3$$

Now, $S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{\Sigma(y-35)^2}{11} = \frac{314}{11} = 28.$

Since $S_2^2 > S_1^2$

$\therefore \quad F = \frac{S_2^2}{S_1^2} \quad , \because S_2^2 > S_1^2$

$F_{cal} = \frac{28.5}{13.3} = 2.14$

Since, $F_{cal} < F_{Tab}$

$2.14 < 3.112$

at 5% Level of significance will 9 and 11 DoF

$\therefore$ H_0 is accepted

We conclude that, two samples have the Same variance.

Q> In a labaratory experiment, two samples give the following results ~~#~~

Sample	Size	Sample mean	Sum of squre of deviation from mean
I	10	15	90
II	12	14	108

Test the equilty of sample variance at 5% LoS Given that $F_{\underset{v_1}{9},\underset{v_2}{11}}(0.05) = 2.9$.

Sol $S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{90}{10-1} = 10$

$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{108}{12-1} = 9.82$

$\because S_1^2 > S_2^2$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$$

$F_{cal} < F_{tab}$

$1.018 < 2.9$

at 5% level at ~~$~~ significance will 9 and 11 DOF.

$\therefore$ H_0 is ~~accept~~ accepted

we conclude that, two sample have the same variance.

Chi-Square test

↳ Kaaii

observed freq (given)

expected freq

Formula, $$X^2 = \Sigma \left\{ \frac{(f_o - f_e)^2}{f_e} \right\}$$

where, dof v is given by,

i> In Binomial distribution $v = n-1$

ii> In poission distribution $v = n-2$

iii> In Normal distribution $v = n-3$

iv> For $m \times n$ contingency table (matrix), $v = (m-1)(n-1)$

Q> A Survey of ~~300~~ 320 families ~~each~~ with 5 children each as follows:-

No. of boys :	5	4	3	2	1	0
No. of girls :	0	1	2	3	4	5
No. of families :	14	56	110	88	40	12

Is this result consistent with hypothesis that male & female birth are equally probable. Given $\left(X^2_{0.5, v=5} = 10.07\right)$

Sol Null hypothesis, H_0 : male and female births are equally probable.

Alternate hypothesis, H_1 : male & female births are not equally probable.

Here $n = 6$, $\therefore$ dof $v = n - 1 = 6 - 1 = 5$

LOS : 5%

$p = \frac{1}{2}$, $q = 1$

→ observed freq → expected / theoritical freq

r	f_o	$N \cdot P(r) = {}^nC_r\, p^r q^{n-1} = f_e$	$(f_o - f_e)^2 / f_e$
5	14	${}^5C_5 (\frac{1}{2})^5 (\frac{1}{2})^0 \times 320 = 10$	$(14-10)^2/10 = 1.6$
4	56	${}^5C_4 (\frac{1}{2})^4 (\frac{1}{2})^1 \times 320 = 50$	$(56-50)^2/50 = 0.72$
3	110	${}^5C_3 (\frac{1}{2})^3 (\frac{1}{2})^2 \times 320 = 100$	$(110-100)^2/100 = 1.0$
2	88	${}^5C_2 (\frac{1}{2})^2 (\frac{1}{2})^3 \times 320 = 100$	$(88-100)^2/100 = 1.44$
1	40	${}^5C_1 (\frac{1}{2})^1 (\frac{1}{2})^4 \times 320 = 50$	$(40-50)^2/50 = 2.0$
0	12	${}^5C_0 (\frac{1}{2})^0 (\frac{1}{2})^5 \times 320 = 10$	$(12-10)^2/10 = 0.4$
			$\chi^2 = \Sigma \left\{ \frac{(f_o - f_e)^2}{f_e} \right\}$

$\chi^2 = 7.16$

$\therefore \chi^2_{cal} = 7.16$

& $\chi^2_{tab} = 10.07$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is accepted

Hence, we conclude that male & female births are equally probable.

Q) Calculate the expected frequencies for the following data, presuming the two attributes, condition of house & condition of child as independent

Condition of child	Condition of home	
	Clear	Dirty
Clean	70	50
Fairly Clean	80	20
Dirty	35	45

Use chi square test at 5% LOS to state whether the two attributes are independent.

$\chi^2_{0.05,\ v=2} = 5.991$

$\chi^2_{0.05,\ v=3} = 7.815$

$\chi^2_{0.05,\ v=4} = 9.488$

Sol :- DOF, $v = (m-1)(n-1)$

$v = (3-1)(2-1)$

$= 2 \times 1$

$v = 2$

Null hypothesis, Ho : There is no association b/w the attributes.

Condition of child	Condition of home		Total
	Clean	Dirty	
Clean	$\frac{120 \times 185}{300} = 74$	$\frac{120 \times 115}{300} = 46$	120
Fairly Clean	$\frac{185 \times 100}{300} = 61.67$	$\frac{100 \times 115}{300} = 38.33$	100
Dirty	$\frac{80 \times 185}{300} = 19.53$	$\frac{80 \times 115}{300} = 30.67$	80
	185	115	300

$$\chi^2 = \sum \left\{ \frac{(f_o - f_e)^2}{f_e} = \frac{(70-74)^2}{74} + \frac{(80-61.67)^2}{61.67} + \frac{(35-49.33)^2}{49.33} \right.$$

$$+ \frac{(50-46)^2}{46} + \left(\frac{20-38.33}{38.33}\right)^2 + \frac{(45-30.67)^2}{30.67}$$

$$\chi^2 = 0.2162 + 5.4482 + 4.1627 + 0.3478 + 8.9657 + 6.6954$$

$\chi^2_{cal} \not< \chi^2_{tab}$

$25.636 \not< 5.991$

$\therefore$ H_0 is rejected

H_1 is accepted

We conclude that there exist association b/w attributes

Q> A dice is tossed 120 times with the following results

Number turned up :	1	2	3	4	5	6	Total
freq :	30	25	18	10	22	15	120

Test the hypothesis that the dice is unbiased.

(Given that $\chi^2_{0.05, 5} = 11.07$)

$\hookrightarrow \alpha = 5\%$ LOS $\rightarrow$ DOF $= 5 = 6-1 = n-1$

Sol Null hypothesis, H_0 : dice is unbiased

Alternate hypothesis, H_1 : dice is biased

DOF, $v = n-1 = 6-1 = 5$

LOS, $\alpha = 0.05 = 5\%$

$\therefore f_e = N.P(r) = 120 \times 1/6 = 20$

	f_o	f_e	$(f_o - f_e)^2 / f_e$
1	30	20	5.0
2	25	20	1.25
3	18	20	0.2
4	10	20	5.0
5	22	20	0.2
6	15	20	1.25
			$\chi^2 = 12.90$

⚠ don't do this in table directly use χ^2 formula separately

$\because \chi^2_{cal} \not< \chi^2_{tab}$

$12.90 \not< 11.07$

$\therefore$ H_0 is rejected

& H_1 is ~~opped~~ accepted

Hence, the dice is biased

Q) In the accounting department of a bank 100 accounts are selected at random & examined for errors, The following results have been obtained.

No. of Errors :	0	1	2	3	4	5	6	Total
No. of accounts :	36	40	19	2	0	2	1	100

Does this information verify that the errors are distributed according to poision probability law
(Given that $\chi^2_{0.05,5} = 11.07$)

↳ ∵ poission ∴ $n-2 = 7-2 = 5$

Sol Null hypothesis, H_0 : Errors are distributed according to poission probability law

Alternate hypothesis, H_1 : Errors are not distributed according to poission prob law

Dof, $v = n-2 = 7-2 = 5$, $m = \frac{\Sigma fx}{\Sigma f} = \frac{100}{100} = 1$

Los, $\alpha = 0.05 = 5\%$

x	f_o	$f_o x$	$f_e = \frac{e^{-m} \cdot m^r}{r!}$	$(f_o - f_e)^2 / f_e$
0	36	0	37	0.027
1	40	40	37	0.243
2	19	38	19	0.055
3	2	6	6	
4	0	0	2 } 8	1.125
5	2	10	0	
6	1	6	0	
	$\Sigma f_o = 100$	$\Sigma f_o x = 100$		$\chi^2 = 1.45$

∵ $\chi^2_{cal} < \chi^2_{tab}$ i.e $1.45 < 11.07$

∴ H_0 is accepted

Hence, - - - -

z distribution

$$z = \frac{1}{2} \log F$$

$$= \frac{1}{2} \log \left(\frac{s_1^2}{s_2^2} \right), \quad s_1^2 > s_2^2$$

$$v_1 = n_1 - 1, \quad v_2 = n_2 - 1$$

Abhi tak Sab kuch discrete data ke liye padhe thee
Ab continuous data se deal karenge

Unit-2
Continuous Probability Distributions

Random Variable

Let S be the Sample Space associated with a given random experiment then a real valued function X with assigned to each outcomes x, they must belong to S i.e $x \in S$. to be a unique real number $X(x)$

Random variable

Discrete Random variable

Continuous Random variable

A variable when real valued function defined on a discrete sample space is called a discrete random variable.

eg → No. of telephone calls per unit time.

A random variable (X) is said to be continuous random variable if it can take all the possible values b/w certain limits
eg → height

Probability Mass Function (P.M.F):- Let X be a discrete random variable a probability mass function is given by → $f(x) = P(X=x)$ or

* It is a function whose value lies b/w 0 & 1 and whose sum is 1 over all values of x.

eg ⇒ (Q)

Toss a balanced coin, let X denote the number of ~~hed~~ heads, Find PMF of X

⇒ random variable, $X = \{x = 0, 1, 2\}$

Sample Space, $S = \{TT, HT, TH, HH\} = 4$

no. of heads	Events	P.m.f
$x=0$	TT	$1/4$
$x=1$	HT, TH	$2/4 = 1/2$
$x=2$	HH	$1/4$
		Σ Pmf $= 1$

Probability Density Function

Let $f(x)$ be a probability function in the interval the probability of the continuous random variable (X) is given by

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$

The function $f(x)$ is known as Probability Density Function if & only if the total probability is unity and it has following two properties:-

i) $f(x) \ge 0$, $x \in [a,b]$

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

eg ⇒ Q) Check whether the following function is P.D.F:
$f(x) = 6x(1-x)$, $0 \le x \le 1$

$$\Rightarrow \int_0^1 f(x) = \int_0^1 6x(1-x)\,dx$$

$$= \int_0^1 (6x - 6x^2)\,dx$$

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

$$= \frac{6}{2} - \frac{6}{3} = 3-2$$

$$= 1$$

$\therefore f(x)$ is P.d.f

Distribution function

1> Let X be a discrete random variable, then distribution function of X is given by

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i(x)$$

2> If X be a continuous random variable, then distribution function of X is given by

$$F(x) = P(X \leq x) = \cancel{\sum} \ \cancel{\int_{-\infty}^{\infty}} \int_{-\infty}^{x} f(x)\,dx$$

Q> ~~A~~ random variable X has the following prob function:

$X = x$	:	0	1	2	3
$P(x)$	:	0	$1/5$	$2/5$	$2/5$

Determine the D.F. of x

Sol given, X be discrete random variable

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$$

i> $F(0) = P(X \leq 0) = P(0) = 0$

ii> $F(1) = P(X \leq 1) = P(0) + P(1) = 0$

iii> $F(2) = P(X \leq 2) = P(0) + P(1) + P(2)$

$= 0 + 1/5 + 2/5 = 3/5$

$$\begin{aligned} P(3) &= P(X \leq 3) \\ &= P(0) + P(1) + P(2) + P(3) \\ &= 0 + 1/5 + 2/5 + 2/5 \\ &= 1 \end{aligned}$$

Q> A random variable X has the density function

$$f(x) = \begin{cases} \frac{1}{9} x^2 , & 0 \leq x < 3 \\ 0 , & \text{otherwise} \end{cases}$$

Find the D.F.

Sol

Given that, X be continuous random variable

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

$$= \int_{0}^{x} \frac{1}{9} x^2 \, dx$$

$$= \frac{1}{9} \left[\frac{x^3}{3} \right]_0^x$$

$$= \frac{1}{9} \left[\frac{x^3}{3} - 0 \right] = \frac{x^3}{27}$$

The distribution function is

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

Properties of Continuous Random Variable

Let f(x) be the p.d.f of a random variable X where x is define in [a, b] then

i> Arithematic mean expected value of x

i.e $\bar{x} = E(x) = \int_a^b x \cdot f(x)\,dx = \mu_1'$

ii> ~~Arithematic~~ mean
Geometric

$$E(\log x) = \cancel{E} \log G = \int_a^b \log x \cdot f(x)\, dx$$

iii> Harmonic mean

$$\frac{1}{H} = \int_a^b \frac{1}{x} \cdot f(x) = E\left(\frac{1}{x}\right)$$

iv> Median

$$\int_{-\infty}^{M} f(x)\, dx = \frac{1}{2} = \int_{M}^{\infty} f(x)\, dx$$

v> lower quartile (Q_1)

$$\int_{-\infty}^{Q_1} f(x)\, dx = \frac{1}{4} \int_{Q_1}^{\infty} f(x)\, dx$$

vi> Upper quartile (Q_3)

$$\int_{-\infty}^{Q_3} f(x)\, dx = \frac{3}{4} = \int_{Q_3}^{\infty} f(x)\, dx$$

vii> r^{th} moment about origin ⇒ At origin, $\bar{x} = 0$ & $\mu_r = \mu'_r$
$(x - \bar{x})^r = x^r$

$$\mu'_r = \int_a^b x^r f(x)\, dx = E(x^r)$$

viii> r^{th} moment about mean ⇒ for any given mean

$$\mu_r = \int_a^b (x - \bar{x})^r f(x)\, dx$$

ix> Mode : mode is the value of x for which $f(x)$ is maximum ⇒ Same Process of calculating Maxi

method → ① put $f'(x) = 0$

② if $f''(x) < 0$ at $x = 0$

$\therefore x = a$ is called mode.

x> Mean deviation about mean $\bar{x}$

$$M.D. = \int_a^b |x - \bar{x}| \, f(x) \, dx$$

xi> If $F(x)$ is distributive function

$\therefore$ p.d.f is $f(x) = \frac{d}{dx} F(x)$

Q> If $f(x) = kx^2$, $0 < x < 1$ has p.d.f, determine k & find $P\left(\frac{1}{3} < x < \frac{1}{2}\right)$ and find a if $P(X > a) = 0.05$,

Sol If given $f(x)$ is p.d.f then $\int_a^b f(x)\,dx = 1$

$$\int_0^1 kx^2 \, dx = 1$$

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

$$k\left[\frac{1}{3} - 0\right] = 1$$

$$\frac{k}{3} = 1$$

$$\boxed{k = 3}$$

$\therefore f(x) = 3x^2$, $0 < x < 1$

[i> $f(x) \geq 0$, $x \in [a, b]$ if these condⁿ satisfy

ii> $\int_a^b f(x)\,dx = 1$ func is pdf

i> $P(a<x<b) = \int_a^b f(x)\,dx$

$P(1/3 < x < \frac{1}{2}) = \int_{1/3}^{1/2} 3x^2\,dx$

$= \left[3\frac{x^3}{3}\right]_{1/3}^{1/2}$

$= \left(\frac{1}{2}\right)^3 - \left(\frac{1}{3}\right)^3 = \frac{1}{8} - \frac{1}{27}$

$= \frac{27-8}{8\times 27} = \frac{19}{216}$

ii> $P(x>a) = 0.05$

$\int_a^1 f(x)\,dx = 0.05$

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

$[x^3]_a^1 = 0.05$

$1 - a^3 = 0.05$

$a^3 = 0.95$

$\boxed{a = 0.98}$

Q> For the distribution funⁿ $dF = Y_0 e^{-|x|}\,dx$, $-\infty<x<\infty$ prove that $Y_0 = 1/2$, mean $= 0$, $S.D = \sqrt{2}$, variance $= 2$ and mean deviation about mean is 1.

Sol Distributive function, $f(x) = \frac{dF}{dx}$

$f(x) = Y_0 e^{-|x|}, \quad -\infty<x<\infty$

If $f(x)$ is P.d.f then $\int_a^b f(x)\,dx = 1$

$\int_{-\infty}^{\infty} Y_0 e^{-|x|}\,dx = 1$

$$y_0 \int_{-\infty}^{\infty} \underbrace{e^{-|x|}}_{\text{this is even function}} dx = 1$$

$$\left[\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx \quad \text{for even} \right]$$

$$\Rightarrow 2\,y_0 \int_{0}^{\infty} e^{-|x|}\,dx = 1$$

$$\because \quad |x| = \begin{cases} -x, & x<0 \\ x, & x>0 \end{cases}$$

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx$$

$$\Rightarrow 2\,y_0 \int_{a}^{\infty} e^{-x}\,dx = 1$$

$$\Rightarrow 2\,y_0 \left[\frac{e^{-x}}{-1} \right]_0^{\infty} = 1$$

$$2y_0 \left[\frac{e^{-\infty}}{-1} - \frac{e^{-0}}{-1} \right] = 1$$

$$2\,y_0 \,[0+1] = 1$$

$$2y_0 = 1$$

$$\boxed{y_0 = \frac{1}{2}}$$

$$\therefore f(x) = \tfrac{1}{2} e^{-|x|}, \quad -\infty < x < \infty$$

ii) mean $= \mu_1' = \int_{a}^{b} x \cdot f(x)\,dx$

$$\mu_1' = \int_{-\infty}^{\infty} x \cdot \tfrac{1}{2} e^{-|x|}\,dx$$

$$\mu_1' = \tfrac{1}{2} \int_{-\infty}^{\infty} \underbrace{x \cdot e^{-|x|}}_{\text{odd}}\,dx$$

$g(x) = x e^{-|x|}$

$g(-x) = -x e^{-|-x|}$

$= -x e^{-|x|}$

$= -g(x)$

$$\boxed{\mu_1' = 0}$$

$$\mu_2 = \sigma^2 = \mu_2' - (\mu_1')^2 \quad \text{①}$$

Now, $\mu_2' = \int_a^b x^2 f(x)\,dx$

$$= \int_{-\infty}^{\infty} \underbrace{x^2 \cdot \tfrac{1}{2} e^{-|x|}}_{\text{even}} dx$$

$$= \tfrac{1}{2} \cdot 2 \cdot \int_0^{\infty} x^2 \cdot e^{-|x|} dx$$

$$= \int_0^{\infty} x^2 \cdot e^{-x} dx$$

$$= \left[-x^2 e^{-x} - 2x e^{-x} - 2e^{-x}\right]_0^{\infty}$$

$$= (0 - 0 - 0) - (0 - 0 - 2)$$

$$\boxed{\mu_2' = 2}$$

Trick

$$\begin{array}{ll} +x^2 \;\text{mul} & e^{-x} \\ -2x & \searrow -e^{-x} \\ +2 & \searrow e^{-x} \\ -\;0 & \searrow -e^{-x} \end{array}$$

ek func algebraic hona chahiye jiska diff 0 hona chahiye

put values in eq^n ① $\therefore |x| = \begin{cases} -x, & x < 0 \\ x, & x > 0 \end{cases}$

$$\sigma^2 = 2 - (0)^2$$

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

$$\text{S.D.} = \sqrt{\text{variance}} = \sqrt{2}$$

iii) mean deviation about mean,

$$\text{M.D.} = \int_{-\infty}^{\infty} |x - \bar{x}| \cdot f(x)\,dx$$

since, $\bar{x} = 0$

$$= \int_{-\infty}^{\infty} |x - 0| \cdot \tfrac{1}{2} e^{-|x|} dx$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} \underbrace{|x| \cdot e^{-|x|}}_{\text{even}} \, dx$$

$$= \frac{1}{\cancel{2}} \cdot 2 \int_{0}^{\infty} |x| \cdot e^{-|x|} \, dx$$

$$= \int_{0}^{\infty} \underset{I}{x} \cdot \underset{II}{e^{-x}} \, dx$$

$$= \left[-x e^{-x} - e^{-x} \right]_0^{\infty}$$

$$= (0 - 0) - (0 - 1)$$

$$\boxed{M.D. = 1}$$

Q) If $f(x)$ is P.d.f where $f(x) = K \frac{1}{(1+x^2)}$, $-\infty < x < \infty$. Find value of k and the distribution function $F(x)$

Sol Given, $f(x) = K \frac{1}{1+x^2}$

if $f(x)$ is P.d.f then $\int_a^b f(x) \, dx = 1$

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

$$K \int_{-\infty}^{\infty} \underbrace{\frac{1}{1+x^2}}_{\text{even}} \, dx = 1$$

$$2K \int_{0}^{\infty} \frac{1}{1+x^2} \, dx = 1$$

$$2K \left[\tan^{-1} x \right]_0^{\infty} = 1$$

$$2k\left[\frac{\pi}{2}-0\right]=1$$

$$\pi k=1$$

$$\therefore \boxed{k=\frac{1}{\pi}}$$

Hence, $\boxed{f(x)=\frac{1}{\pi}\left(\frac{1}{1+x^2}\right)}$ $-\infty<x<\infty$

Distribution Function

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

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

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

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

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

$$=\boxed{\frac{1}{\pi}\tan^{-1}x+\frac{1}{2}}$$ Ans

Q> Find the median & mode of the freq. of the prob. curve $f(x)=\frac{1}{2}\sin x$, $0<x<\pi$

Sol For P.d.f, $\int_a^b f(x)\,dx=1$

$$\int_0^{\pi} f(x)\,dx=\int_0^{\pi}\frac{1}{2}\sin x\,dx=\frac{1}{2}\left[-\cos x\right]_0^{\pi}$$

$$=\frac{1}{2}\left[-\cos(\pi)-(-\cos 0)\right]$$

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

$$=\frac{1}{2}\left[1+1\right]=1$$ $\therefore$ It is P.d.f

Median (M), $\int_{M}^{\infty} f(x)\, dx = \frac{1}{2}$

$$\int_{m}^{\pi} \frac{1}{2} \sin x \; dx = \frac{1}{2}$$

$$\left[-\cos x\right]_{M}^{\pi} = 1$$

$$-\cos \pi - \cos m = 1$$

$$1 - \cos m = 1$$

$$\cos m = 0$$

$$\cos m = \cos \pi/2$$

$$\boxed{m = \frac{\pi}{2}}$$

for mode, given $f(x) = \frac{1}{2} \sin x$

$$f'(x) = \frac{1}{2} \cos x \quad - ①$$

$$f''(x) = -\frac{1}{2} \sin x \quad - ②$$

Put $f'(x) = 0$

$$\frac{1}{2} \cos x = 0$$

$$\cos x = 0$$

$$\cos x = \cos \pi/2$$

$$\therefore \boxed{x = \pi/2}$$

Put $\boxed{x = \pi/2}$ in eqⁿ ②

$$f''(\pi/2) = -\frac{1}{2} \sin(\pi/2) = -\frac{1}{2} \times 1 = -\frac{1}{2} < 0$$

Normal, exponential & gamma densities

Q> The mean deviation from the mean of a normal distribution in 4/5 times of standard deviation

Q> Prove that in a normal distribution, all the moments of odd order about the mean are vanish i.e zero.

Sol We know that, normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Also, $$\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\, dx$$

Let odd order

$r = 2n+1$

$$\therefore \mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $t = \frac{x-\mu}{\sigma} = t.\sigma = x-\mu$

$\sigma dt = dx - 0$

$$= \int_{-\infty}^{\infty} (t\sigma)^{2n+1} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-t^2/2} \cdot \sigma dt$$

$$= \frac{(\sigma)^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^{2n+1} e^{-t^2/2} dt$$

$$= \frac{(\sigma)^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{t\cdot(t)^{2n}}_{\text{odd}} e^{-t^2/2} dt = 0$$

$$\boxed{\mu_{2n+1} = 0}$$

Q> Prove that in a normal distribution, all the moments or even order about the mean is $\mu_{2n} = (2n-1)(2n-3)(2n-5) \; --- \; 5.3.1\,\sigma^{2n}$

Sol We know that, normal distribution,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Also, $$\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\,dx$$

Let even order, $r = 2n$

$$\therefore \mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $t = \frac{x-\mu}{\sigma}$ $\Rightarrow t\sigma = x-\mu$

$\sigma\,dt = dx - 0$

$$= \int_{-\infty}^{\infty} (t\sigma)^{2n} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2} \cdot \sigma\,dt$$

$$= \frac{(\sigma)^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{t^{2n} \; e^{-t^2/2}}_{\text{even}} dt$$

$$= \frac{2(\sigma)^{2n}}{\sqrt{2\pi}} \int_{0}^{\infty} t^{2n} \cdot e^{(t^2/2)} dt$$

$$= \frac{2(\sigma)^{2n}}{\sqrt{2\pi}} \int_0^\infty (t^2)^n e^{-t^2/2} \, dt$$

Let $\frac{t^2}{2} = y \Rightarrow \frac{2t}{2} \, dt = dy$

$dt = \frac{dy}{t} = \frac{dy}{\sqrt{y}}$

$$= \frac{2(\sigma)^{2n}}{\sqrt{2\pi}} \int_0^\infty (2y)^n e^{-y} \frac{dy}{\sqrt{y}}$$

$$= \frac{2.2^n(\sigma)^{2n}}{\sqrt{2\pi}} \int_0^\infty y^{n-1/2} e^{-y} \, dy$$

$$= \frac{2^{n+1}(\sigma)^{2n}}{\sqrt{2\pi}} \int_0^\infty e^{-y} y^{n+1-1-1/2} \, dy$$

$\Gamma n = \int_0^\infty e^{-x} x^{n-1} \, dx$

$$= \frac{2^{n+1}}{\sqrt{2\pi}} (\sigma)^{2n} \int_0^\infty e^{-y} y^{(n+1/2)-1} \, dy$$

$$\mu_{2n} = \frac{2^{n+1}}{\sqrt{2\pi}} (\sigma)^{2n} \Gamma\left(n+\frac{1}{2}\right) \quad ①$$

replacing n by n-1

$$\mu_{2n-2} = \frac{2^n}{\sqrt{2\pi}} (\sigma)^{2n-2} \Gamma\left(n-\frac{1}{2}\right) \quad ②$$

eq ① ÷ eqⁿ ②

$$\frac{\mu_{2n}}{\mu_{2n-2}} = \frac{\frac{\cancel{2^{n+1}}}{\cancel{\sqrt{2\pi}}} (\sigma)^{2n} \Gamma(n+1/2)}{\frac{\cancel{2^n}}{\cancel{\sqrt{2\pi}}} \cancel{(\sigma)^{2n-2}} \Gamma\left(n-\frac{1}{2}\right)}$$

$$= \frac{2\sigma^2 \Gamma(n+1/2)}{\Gamma\left(n-\frac{1}{2}\right)}$$

$$\left(\because \sqrt{n+1} = n\sqrt{n} \right]$$

$$= \frac{2\sigma^2 \sqrt{n+1-1+\frac{1}{2}}}{\sqrt{n-\frac{1}{2}}}$$

$$= \frac{2\sigma^2 \left(n+\frac{1}{2}\right) \cancel{\sqrt{n-\frac{1}{2}}}}{\cancel{\sqrt{n-\frac{1}{2}}}}$$

$$= \frac{\cancel{2}\sigma^2 (2n-1)}{\cancel{2}}$$

$$\frac{\mu_{2n}}{\mu_{2n-2}} = \sigma^2(2n-1)$$

$$\therefore \mu_{2n} = \sigma^2(2n-1)\underline{\mu_{2n-2}}$$

$$\left[\begin{array}{l} \text{put } n = n-1 \\ \mu_{2n-2} = \sigma^2[2(n-1)-1]\,\mu_{2(n-1)-2} \\ \mu_{2n-2} = \sigma^2(2n-3)\,\mu_{2n-4} \end{array}\right]$$

$$\mu_{2n} = \sigma^2(2n-1)\,\sigma^2(2n-3)\,\mu_{2n-4}$$

$$\mu_{2n} = \sigma^4(2n-1)(2n-3)\,\mu_{2n-4}$$

Similarly, put $n = n-2$

$$\mu_{2n} = \sigma^6(2n-1)(2n-3)(2n-5)\,\mu_{2n-6}$$

$$\vdots$$

$$\boxed{\mu_{2n} = \sigma^{2n}(2n-1)(2n-3)(2n-5)\cdots 5.3.1}$$

Q> For the normal curve:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Find mean and standard deviation.

Sol

Since, $f(x)$ is ~~p.d.f~~ P.d.f

$$\mu_1 = \text{mean} = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $t = \frac{x-\mu}{\sqrt{2}\,\sigma}$

$$\sigma\sqrt{2}\, t = x - \mu$$

$$\sigma\sqrt{2}\, dt = dx \quad - ①$$

$$\therefore \text{mean} = \int_{-\infty}^{\infty} [\sigma\sqrt{2}\, t + \mu] \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2} \sigma\sqrt{2}\, dt$$

$$= \frac{1}{\sqrt{\pi}} \left[\sigma\sqrt{2} \int_{-\infty}^{\infty} \underbrace{t \cdot e^{-t^2}}_{\text{odd}} dt + \mu \right] \int_{-\infty}^{\infty} \underbrace{e^{-t^2}}_{\text{even}} dt$$

$$= \frac{1}{\sqrt{\pi}} \left[\sigma\sqrt{2}\,(0) + \mu\, 2 \int_{0}^{\infty} e^{-t^2} dt \right]$$

$$= \frac{1}{\sqrt{\pi}} \cdot 2\mu \int_{0}^{\infty} e^{-t^2} dt$$

Let $t^2 = y$

$$2t\, dt = dy$$

$$dt = \frac{dy}{2t}$$

$$dt = \frac{dy}{2\sqrt{y}} \qquad = \frac{2\mu}{\sqrt{\pi}} \int_{0}^{\infty} e^{-y} \cdot \frac{dy}{2\sqrt{y}}$$

$$= \frac{\mu}{\sqrt{\pi}} \int_0^\infty e^{-y} \cdot y^{-1/2} \, dy$$

$$= \frac{\mu}{\sqrt{\pi}} \int_0^\infty e^{-y} \cdot y^{1/2 - 1} \, dy$$

$$= \frac{\mu}{\sqrt{\pi}} \sqrt{\frac{1}{2}}$$

$$= \frac{\mu \, \cancel{(\sqrt{\pi})}}{\cancel{\sqrt{\pi}}} \qquad \because \sqrt{\frac{1}{2}} = \sqrt{\pi}$$

$$\boxed{\text{mean} = \mu}$$

Now, we knew that

$$\text{variance } (\mu_2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \, \cancel{(x-\text{mean})^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)}{2\sigma^2}} \, dx$$

Let $t = \frac{x-\mu}{\sqrt{2}\cdot\sigma}$

$\sigma\sqrt{2}\, t = x - \mu \Rightarrow \sigma\sqrt{2}\, t + \mu = x$

$\sigma\sqrt{2}\, dt = dx - 0$

$$\therefore \text{Mean} = \int_{-\infty}^{\infty} (\sigma\sqrt{2}\, t)^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2} \sigma\sqrt{2} \, dt$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \sigma^2 \cdot 2 \cdot t^2 \cdot e^{-t^2} \, dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \underbrace{t^2 \cdot e^{-t^2}}_{\text{even}} \, dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \cdot 2 \int_0^\infty t^2 \cdot e^{-t^2} \, dt$$

Let $t^2 = y$

$$2t\,dt = dy$$

$$dt = \frac{dy}{2t}$$

$$dt = \frac{dy}{2\sqrt{y}}$$

$$\text{variance} = \frac{\overset{2}{\cancel{4}}\sigma^2}{\sqrt{\pi}} \int_0^\infty y \cdot e^{-y} \frac{dy}{\cancel{2}\sqrt{y}}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-y} \cdot y^{1-\frac{1}{2}}\, dy$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-y}\, y^{\frac{1}{2}}\, dy$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-y} \cdot y^{\frac{3}{2}-1}\, dy$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\, \Gamma\frac{3}{2}$$

$$= \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \frac{1}{\cancel{2}} \cdot \cancel{\sqrt{\pi}}$$

$$= \sigma^2$$

$$\therefore \text{SD} = \sqrt{\text{variance}}$$

$$= \sqrt{\sigma^2}$$

$$\boxed{\text{SD} = \sigma}$$

Q> Fit a normal curve to the following data.

Length of line :	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.52
Freq :	2	3	4	9	10	8	4	1	1

Sol

mean = μ

S.D. = σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Exponential Distribution

① A random variable X is said to be an exponential distribution with parameter $\lambda > 0$, if it's P.d.f is given by,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0 \\ 0, & \text{otherwise} \end{cases}$$

② Distributive function of exponential function

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

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

$$= \int_{-\infty}^{\infty} 0 + \int_{0}^{x} \lambda e^{-\lambda x}\,dx$$

$$= \lambda \left[\frac{e^{-\lambda x}}{-\lambda} \right]_{0}^{x}$$

$$= \left[-e^{-\lambda x} - \left(\frac{e^{-\lambda \times 0}}{-1} \right) \right]$$

$$\boxed{F(x) = -e^{-\lambda x} + 1}$$

Q> Find mean, variance & S.D of exponential distribution.

Sol We know that, P.d.f of exponential distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_1' = \text{mean} = \int_{-\infty}^{\infty} x f(x)\,dx = E(x)$$

$$= \int_{0}^{\infty} x \cdot \lambda e^{-\lambda x}\,dx$$

$$= \int_{0}^{\infty} x \cdot e^{-\lambda x}\,dx$$

$+x \quad e^{-\lambda x}$

$-1 \quad \frac{e^{-\lambda x}}{-\lambda}$

$+0 \quad \frac{e^{-\lambda x}}{\lambda^2}$

$$= \lambda\left[x\frac{e^{-\lambda x}}{-\lambda} - \frac{e^{-\lambda x}}{\lambda^2}\right]_0^\infty$$

$$= \lambda\left[(0-0) - \left(0 - \frac{1}{\lambda^2}\right)\right]$$

$$= \lambda\left(\frac{1}{\lambda^2}\right)$$

$$\boxed{\text{mean} = \frac{1}{\lambda}}$$

Now, variance $= \mu_2' - {\mu_1'}^2$ —①

Since, $\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

$$\mu_2' = \int_0^\infty x^2 . \lambda e^{-\lambda x}\,dx$$

$$\mu_2' = \int_0^\infty x^2 . \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_0^\infty x^2 . e^{-\lambda x}\,dx$$

$+ x^2$ $e^{-\lambda x}$

$-2x$ $\frac{e^{-\lambda x}}{-\lambda}$

$+2$ $\frac{e^{-\lambda x}}{\lambda^2}$

-0 $\frac{e^{-\lambda x}}{-\lambda^3}$

$$= \lambda\left[x^2\frac{e^{-\lambda x}}{-\lambda} - 2x\frac{e^{-\lambda x}}{\lambda^2} - 2\frac{e^{-\lambda x}}{\lambda^3}\right]_0^\infty$$

$$= \lambda\left[2(0-0-0) - \left(0-0-\frac{2}{\lambda^3}\right)\right]$$

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

Put value of μ_1', μ_2' in eqn ①

variance $= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \boxed{\frac{1}{\lambda^2}}$ Ans

S.D. $= \sqrt{\text{variance}} = \sqrt{\frac{1}{\lambda^2}}$

$$\boxed{\sigma = \frac{1}{\lambda}}$$ Ans

P.d.f → Prob density fn

Q> Find mean deviation about mean of exponential distribution.

Sol $MD = \int_{-\infty}^{\infty} |x - \text{mean}|\, f(x)\, dx$

$= \int_{0}^{x} \left| x - \frac{1}{\lambda} \right| \lambda e^{-\lambda x}\, dx$

Q> Suppose the life of mobile batteries is exponentially distributed with parameter $\lambda = 0.001$ days. What is the probability last more than 1200 days?

Ans $\lambda = 0.001$ days

we know that the P.d.f of exponential distribution,

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0,\ \lambda \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

P(battery will last more than 1200 days)

$= P(X \geq 1200)$

$= \int_{1200}^{\infty} f(x)\, dx$

put $\lambda = 0.001$

$= 0.001 \int_{1200}^{\infty} e^{-0.001x}\, dx$

$= 0.001 \left[\frac{e^{-0.001x}}{-0.001} \right]_{1200}^{\infty}$ $\therefore \int e^{ax}\, dx = \frac{e^{ax}}{a}$

$= 0.001 \left[\frac{e^{-0.001 \times \infty}}{-0.001} - \frac{e^{-0.001 \times 1200}}{-0.001} \right]$

$= 0.001 \left[0 + \frac{e^{-1.200}}{0.001} \right]$ $= e^{-1.2}$

$P(X \geq 1200) \simeq \boxed{0.301}$

Q> A random variable X has exponential distribution whose p.d.f is given by $f(x) = \begin{cases} 2e^{-2x} , & \text{if } x>0 \\ 0 & \text{if } x \le 0 \end{cases}$

~~Sol~~ Compute the probability that X not less than 3. Also find mean & standard deviation & prove that the coefficient of variation is unity.

Sol

(i) $P(X \ge 3) = \int_3^\infty f(x)\,dx = \int_3^\infty 2e^{-2x}\,dx = e^{-6}$

(ii) mean $= E(x) = \mu_1' = \int_{-\infty}^{\infty} x f(x) = \frac{1}{\lambda}$

$$= \int_0^\infty x \cdot 2e^{-2x}\,dx$$

$$= 2\int_0^\infty x \cdot e^{-2x}\,dx = \boxed{\frac{1}{2}}$$

(iii) $\mu_2' = \int_0^\infty x^2 f(x) = \int_0^\infty x^2 \cdot 2e^{-2x}\,dx = \boxed{\frac{1}{2}}$

Variance $= \mu_2' - (\mu_1')^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \boxed{\frac{1}{4}}$

standard deviation $= \sqrt{\text{variance}} = \sqrt{\frac{1}{4}} = \boxed{\frac{1}{2}}$

iv> Coefficient of variation $= \frac{\text{mean}}{\text{S.D}} = \frac{1/2}{1/2} = \boxed{1}$

Gamma Distribution

A continuous random variable X is said to be Gamma distribution with parameter $\lambda > 0$, if its P.d.f is given by

$$f(x) = \begin{cases} \frac{e^{-x} \cdot x^{\lambda-1}}{\Gamma\lambda} , & 0 < x < \infty , \lambda > 0 \\ 0 , & \text{otherwise} \end{cases}$$

Distribution Function of Gamma Function

we know that

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

$$= \int_{0}^{x} \frac{e^{-x} \cdot x}{x!}\, dx$$

$$F(x) = \frac{1}{\lambda!} \int_{0}^{x} e^{-x} \cdot x^{\lambda - 1}\, dx$$

$$\therefore F(x) = \begin{cases} \frac{1}{\lambda!} \int_{0}^{x} e^{-x} \cdot x^{\lambda-\lambda}\, dx, & 0 < x < \infty, \ \lambda > 0 \\ 0, & \text{otherwise} \end{cases}$$

Unit - 3

Bivariate Distribution

Q) From the following data of marks of students of analysis & statistics of BE examination, max-marks of each subject is 50. Find the bivariate frequency table:

Roll no.: - 1 2 3 4 5 6 7 8 9 10 11 12 13
14 15

marks in Analysis (x) - 28 42 39 41 25 38 45 33 37 34
47 34 39 36 41

marks in statistics (y) - 32 40 35 47 30 43 47 39 38
35 45 36 43 41 43

Ans →

x↓ y→	25 - 30	30 - 35	35 - 40	40-45	45-50
25-30	~~3~~	1+1			
30 - 35					
35-40			1+	1	
40-45				1	1
45 -50					1

Q) For the following probability distribution of X and Y.
Find i) $P(X \leq 2, Y=3)$, ii) $P(X \leq 1)$
iii) $P(Y=4)$, (iv) $P(Y \leq 5)$
v) $P(X<2, Y \leq 3)$

$x\downarrow$ $\vec{Y}$	1	2	3	4	5	6	Sum
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{3}$	$\frac{3}{32}$	$\frac{8}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{10}{16}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$	$\frac{8}{64}$
	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{11}{64}$	$\frac{13}{64}$	$\frac{6}{32}$	$\frac{16}{64}$	1

i> $P(x \leq 2, Y=3) = P(x=0, Y=3) + P(x=1, Y=3) + P(x=2, Y=3)$

$= \frac{1}{32} + \frac{1}{8} + \frac{1}{64} = \frac{11}{32}$

ii> $P(x \leq 1) = P(x=0) + P(x=1)$

$= \frac{8}{32} + \frac{10}{16}$

iii> $P(Y=4) = \frac{13}{64}$

iv> $P(Y \leq 5) = P(Y=1) + P(Y=2) + P(Y=3) + P(Y=4)$
$+ P(Y=5)$

$= \frac{3}{32} + \frac{3}{32} + \frac{11}{64} + \frac{13}{64} + \frac{6}{32} =$

Q> Let U & V be two random variables having the following probability distribution for three variable $-1, 0, +1$

$\downarrow V$ $U \rightarrow$	(U_0) -1	(U_1) 0	(U_2) 1	
$0 \rightarrow (V_0)$ -1	(P_{00}) 0.0	(P_{10}) 0.1	(P_{20}) 0.1	0.2
$1 \rightarrow (V_1)$ 0	(P_{01}) 0.2	(P_{11}) 0.2	(P_{21}) 0.2	0.6
$2 \rightarrow (V_2)$ 1	(P_{02}) 0.0	(P_{12}) 0.1	(P_{22}) 0.1	0.2
	0.2 (P_0) for U	0.4 (P_1) for U	0.4 (P_2) for U	1.0

1> Find $E(U)$, $E(V)$, show that $E(U) \neq E(V)$

2> Prove that U & V are uncorrelated. → covariance = 0

3> Find the conditional ~~probrobat~~ probability distribution ~~for three variable~~ of U when given that $V=0$.

4> ~~Find~~ $Var(X)$ & $Var(Y)$

5> Find the conditional variance $Var[V|U=-1]$

Sol

i> $$E(U) = \sum_{i=0}^{2} P_i U_i$$

$$= P_0U_0 + P_1U_1 + P_2U_2$$

$$= (0.2)(-1) + 0.4(0) + 0.4(1)$$

$$= \cancel{0 + 0 + 0.1} \; -0.2 + 0 + 0.4$$

$$\boxed{E(U) = 0.2} - ①$$

→ expected value or mean of U

$$E(V) = \sum_{i=0}^{2} P_i V_i$$

$$= P_0V_0 + P_1V_1 + P_2V_2$$

$$= (0.2)(-1) + (0.6)(0) + (0.2)(1)$$

$$= -0.2 + 0 + 0.2$$

$$\boxed{E(V) = 0} - ②$$

From ① & ②

$$\therefore \boxed{E(U) \neq E(V)}$$

ii> If U & V are uncorrelated then $cov(U,V) = 0$

~~$cov(U,V) = E(U) + E(V) - E(UV)$~~

~~$= E(U) + E(V) - E(U)E(V)$~~

~~$= 0.2 + 0 - (0.2)(0)$~~

~~$cov(U,V) = 0.2$~~

$$cov(U,V) = E(UV) - E(U)E(V) - ③$$

Here, $$E(UV) = \sum_{j=0}^{2}\sum_{i=0}^{2} \underbrace{P_{ij}U_iV_j}$$

$$E(UV) = \sum_{j=0}^{2} \{P_{0j}U_0V_j + P_{1j}U_1V_j + P_{2j}U_2V_j\}$$

$$= P_{00} U_0 V_0 + P_{01} U_0 V_1 + P_{02} V_0 V_2$$
$$+ P_{10} U_1 V_0 + P_{11} U_1 V_1 + P_{12} U_1 V_2$$
$$+ P_{20} U_2 V_0 + P_{21} U_2 V_1 + P_{22} U_2 V_2$$

$E(UV) = 0$

Putting $E(UV)$ in ③

$$\text{Cov}(U, V) = E(UV) - E(U)\,E(V)$$
$$= 0 - 0$$
$$\text{Cov}(U, V) = 0$$

$\therefore$ U & V are uncorrelated

iii) Conditional prob. distribution of U when given that $v=0$

i.e $P(U|v=0) = P(U=x_i \,|\, v=0) = \dfrac{P(U x_i \cap v=0)}{P(v=0)}$

For $i=0$, $P(U=x_0 \,|\, v=0) = \dfrac{P(U=x_0 \cap v=0)}{P(v=0)}$

$= \dfrac{P(U=-1 \cap v=0)}{P(v=0)} = \dfrac{0.2}{0.6} = \boxed{\dfrac{1}{3}}$

For $i=1$, $P(U=x_1 \,|\, v=0) \cancel{= P(U=x_1 \,|\, v=0)} = \boxed{\dfrac{1}{3}}$

For $i=2$, $P(U=x_2 \,|\, v=0) = \boxed{1/3}$

iv) $\mathrm{var}(X) = \mathrm{Var}(U) = E(U^2) - (E(U))^2$ — ①

Now, $E(U^2) = \sum_{i=0}^{2} P_i x_i^2$

$= P_0 x_0^2 + P_1 x_1^2 + P_2 x_2^2$

$= (0.2)(-1)^2 + (0.4)(0)^2 + (0.4)(1)^2$

$= 0.2 + 0 + 0.4$

$= 0.6$

& $(E(U))^2 = (0.2)^2 = 0.04$

$\therefore \mathrm{Var}(X) = 0.6 - 0.04$

$\boxed{\mathrm{Var}(X) = 0.56}$

Similarly for var(Y) (Do it yourself)

$\text{var}(v|u=-1) = E(v|u=-1)^2 - (E(v|u=-1))^2$ ②

#

Now, $E(V|U=-1) = \sum P(v=y_j|u=-1) \cdot y_j$

$= 0.0(-1) + 0.2(0) + 0.0(1) = 0$

$(E(V|U=-1))^2 = 0$

$E(V|U=-1)^2 = \sum P(V=y_j|U=-1) - y_j^2$

$= 0.0(-1)^2 + 0.2(0)^2 + 0.0(1)^2 = 0$

put value in eq ②

$\text{var}(V|U=-1) = 0 - 0^2 = \boxed{0}$

Q) Find c so that $f(x,y) = cxy$, $1 \leq x \leq y \leq 2$ will be joint probability density function.

Sol

If $f(x,y) = cxy$ is joint p.d.f   1> $f(x) \geq 0$

then $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ 2> $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_{x=1}^{2}\int_{y=x}^{2} cxy\,dx\,dy$$

$$\int_1^2 cx\left[\frac{y^2}{2}\right]_x^2 dx = 1$$

$$= \int_1^2 cx\left[\frac{2^2}{2} - \frac{x^2}{2}\right] dx = 1$$

$$= \int_1^2 cx\left(2 - \frac{x^2}{2}\right) dx = 1$$

$$= c\int_1^2 \left(2x - \frac{x^3}{2}\right) dx = 1$$

$$= c\left[2\frac{x^2}{2} - \frac{x^4}{8}\right]_1^2 = 1$$

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

$$= c\left[2 - \frac{7}{8}\right] = 1 \qquad ; c\cdot\left(\frac{9}{8}\right) = 1$$

$$\therefore \boxed{c = \frac{8}{9}}$$

Q> The joint P.d.f of (u, v) is given by

$$f(x,y) = \begin{cases} 2, & 0 < x, 1, \ 0 < y < x \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that $f(x,y)$ is joint p.d.f
(b) Find marginal density function of u and v.
(c) Find the conditional density direction of V given $v = x$ and that. of u given that $v = y$.
(d) Are u and v independent,

(a) If $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ then $f(x,y)$ is joint P.d.f

L.H.S
$$\int_0^1\int_0^x 2\,dx\,dy = \int_0^1 [2y]_0^x\,dx$$
$$= \int_0^1 [2x - 0]\,dx$$
$$= \int_0^1 2x\,dx = 2\left[\frac{x^2}{2}\right]_0^1$$
$$= 1^2 - 0^2$$
$$= \boxed{1} \text{ RHS}$$

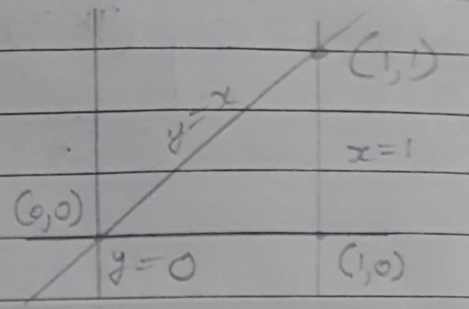
(b) Marginal density function of v is given by
$$f_u(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^x 2\,dy$$

$$= \left[2y\right]_0^x$$

$$= 2x - 0$$

$$f(x) = 2x$$



Similarly

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

$$= \int_0^1 2\,dx = 2\left[x\right]_0^1 = 2[1-0]$$

$$= \boxed{2}$$

(c) The conditional density function v given u=x

$$\text{i.e } \underset{\text{then } v \swarrow \searrow \text{ if } u}{f_{v|u}}(y|x) = \frac{f(x,y)}{f_u(x)} = \frac{2}{2x} = \frac{1}{x}$$

$$0 < x < 1$$

→ marginal density function of u

The conditional density function of U given v=y

$$f_{u|v}(x|y) = \frac{f(x,y)}{f_v(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}, \quad 0 < y < 1$$

(d) u and v are independent if $f(x,y) = f_u(x) \cdot f_v(y)$

Since, $f(x,y) = 2$

& $f_u(x) = 2x$, $f_v(y) = 2(1-y)$

$f_u(x) \cdot f_v(y) = 2x\,(1-y)$

from ① & ②

$f(x,y) \neq f_u(x) \cdot f_v(y)$ Hence u & v are dependent

Q) Let X and Y be two joint P.d.f for continuous random variables.

$$f_{XY}(x,y) = \begin{cases} 2 & , x+y \leq 1, x>0, y>0 \\ 0 & , \text{otherwise} \end{cases}$$

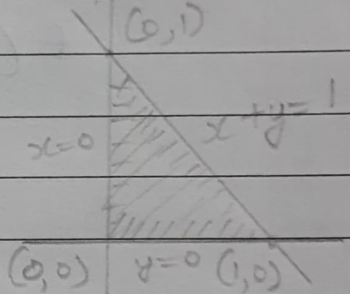
find $\mathrm{cov}(X,Y)$, $\rho(X,Y)$ i.e Find covariance & Correlation

Sol Formula (i). $\mathrm{Cov}(X,Y) = E(XY) - E(X).E(Y)$

(ii) correlation $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{var}(Y)}}$

where, $\mathrm{Var}(X) = E(X^2) - (E(X))^2$

$\mathrm{var}(Y) = E(Y^2) - (E(Y))^2$



$x+y \leq 1$

$\Updownarrow$

$x+y=1$

$\left[\frac{x}{a}+\frac{y}{b}=1\right]$

$\to \frac{x}{1}+\frac{y}{1}=1$

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

$= \int_{y=0}^{1-x} 2\,dy$

$= 2\,[y]_0^{1-x}$

$f_X(x) = 2(1-x)$

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

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

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

and

$$f_Y(y) = \begin{cases} 2(1-y), & 0<y<1 \\ 0, & \text{otherwise} \end{cases}$$

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

limits
trick

eq^n ~~se~~ eq^n
pt se pt

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

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

$$= 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)_0^1$$

$$= 2\left(\frac{1}{2} - \frac{1}{3}\right) = 2\left(\frac{1}{6}\right) = \frac{1}{3}$$

Similarly, $E(Y) = \int_0^1 y\, f_Y(y)\,dx = \int_0^1 y\; 2(1-y)\,dy = \frac{1}{3}$

Now,

$$E(x^2) = \int_0^1 x^2 \cdot f_X(x)\,dx = \int_0^1 x^2 \cdot 2(1-x)\,dx$$

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

$$E(x^2) = 1/6$$

Similarly $E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy = \int_0^1 y^2\; 2(1-y)\,dy$

$$E(Y^2) = \frac{1}{6}$$

Now, $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy$

limits : $y = 0$ to $y = 1-x$

$x = 0$ to $x = 1$

$$= \int_0^1 \int_0^{1-x} xy \cdot 2\,dx\,dy$$

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

$$= \int_0^1 x\left[(1-x)^2 - 0^2\right] dx$$

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

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

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

$$= \frac{1}{2} - \frac{2}{3} + \frac{1}{4}$$

$$= \frac{6-8+3}{12} = \frac{1}{12}$$

i> $\mathrm{Cov}(x,y) = E(XY) - E(X)\,E(Y)$

$$= \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{12} - \frac{1}{9} = \frac{1}{36}$$

ii> Since, $\mathrm{Var}(X) = E(X^2) - \left(E(X)\right)^2$

$$= \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{6} - \frac{1}{9}$$

$$= \boxed{\frac{1}{18}}$$

Similarly, $\mathrm{var}(Y) = E(Y^2) - \left(E(Y)\right)^2$

$$= \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \boxed{\frac{1}{18}}$$

Formula,

Correlation $\rho(X,Y) = \dfrac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{var}(X) \cdot \mathrm{var}(Y)}}$

$$= \frac{-1/36}{\sqrt{\frac{1}{18} \cdot \frac{1}{18}}} = \frac{\frac{-1}{36}}{\frac{1}{18}} = \boxed{\frac{-1}{2}}$$

(Pg 150) Q> Three fair coins are tossed. Let (X) denote the no. of heads on the first two coins & let (Y) denote no. of tails on last two coins.

i> Find the joint distribution of X & Y

ii> Find the conditional distribution of Y given that X = 1

iii> Find the covariance of X & Y.

iv> f (x, y)

Sol Possibilities of 3 coins,

	HHH	HHT	HTH	HTT	THH	TTH	TTT	THT
X	2	2	1	1	1	0	0	1
Y	0	1	1	2	0	1	2	1

$X = \{\underbrace{0, 1, 2}_{x}\}$, $Y = \{\underbrace{0, 1, 2}_{y}\}$

i> Joint probability distribution table for random variable X & Y

	$Y\downarrow$ $X\rightarrow$	0	1	2	Total $f_y(y)$ $\downarrow$	
	0	0	$1/8$	$1/8$	$2/8$	$f_y(0)$
y	1	$1/8$	$2/8$	$1/8$	$4/8$	$f_y(1)$
	2	$1/8$	$1/8$	0	$2/8$	$f_y(2)$
	Total $f_x(x) \rightarrow$	$2/8$ $f_x(0)$	$4/8 = 1/2$ $f_x(1)$	$2/8$ $f_x(2)$	$8/8 = 1$	

ii> $f(Y|X=1) = f(Y=y_i \mid X=1)$

$$= \frac{f(1, y_i)}{f(1)} = \frac{f(1, y_i)}{1/2} = 2f(1, y_i)$$

At $i=1$, $f(Y|X=1) = 2f(1, y_1)$

$$= 2f(1, 0)$$

$$= 2 \times \frac{1}{8} = \frac{1}{4}$$

Baye's Theorem

An event A corresponds to a number of exhaustive events $B_1, B_2, B_3 \text{---} B_n$ and $P(B_i)$ and $P(A/B_i)$ are given, then

$$P(B_i|A) = \frac{P(B_i)\ P(A|B_i)}{\sum P(B_i)\ P(A|B_i)}$$

Q> There are three bags : first containing 1 white, 2 red & 3 green balls, Second 2W, 3R, 1G and third 3W, 1R, 2G. Two balls are drawn from a bag choosen at random. These are found to be 1W & 1R ball. Find the probability that the balls came drawn from the second bag.

Let B_1, B_2, B_3 are three bags.

	B_1	B_2	B_3
	1 W	2W	3W
	2 R	3R	1R
	3 G	1G	2G
Total →	6	6	6

$P(A|B_1)$ = Prob of a white & a red ball drawn from B_1

$$= \frac{{}^1C_1 \times {}^2C_1}{{}^6C_2} = \frac{2}{15}$$

$$P(A/B_2) = \frac{{}^2C_1 \times {}^3C_1}{{}^6C_2} = \frac{2}{5}$$

$$P(A/B_3) = \frac{{}^3C_1 \times {}^1C_1}{{}^6C_2} = \frac{1}{5}$$

Applying Baye's Theorem,

$$P(B_1|A) = \frac{P(B_i)\ P(A|B_i)}{\sum P(B_i) \cdot P(A|B_i)}$$

$$P(B_2 | A) = \frac{P(B_2)\ P(A|B_2)}{\sum_{i=1}^{3} P(B_i) \cdot P(A|B_i)}$$

$$= \frac{P(B_2) \cdot P(A|B_2)}{P(B) \cdot P(A|B_1) + P(B_2)P(A|B_2) + P(B_3) \cdot P(A|B_3)}$$

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

$$\boxed{P(B_2|A) = \frac{6}{11}}$$

Q> Three machines, M_1, M_2 & M_3 produce identical items of their respective output 5%, 4%, 3% of items are faulty on a certain day M_1 has produce 25% of the total output, M_2 has produce 30% & M_3 has produce the remain & item Selected at random is found to be faulty. What are the chance that it was produced by machine with the highest output.

Sol

	M_1	M_2	M_3	Sum
	25%	30%	45%	
$P(B_i)$	0.25	0.30	0.45	1
$P(A\|B_i)$	0.05	0.04	0.03	
$P(B_i) \cdot P(A\|B_i)$	0.0125	0.012	0.0135	$\Sigma P(B_i) \cdot P(A\|B_i) = 0.038$
Baye's Theorem	$\frac{0.0125}{0.038}$	$\frac{0.012}{0.038}$	$\frac{0.0135}{0.038}$	
	= 0.328	0.315	0.355	

$$P(B_i/A) = \frac{P(B_1)\cdot P(A|B_1)}{\sum_{i=1}^{3} P(B_1)\,P(A|B_1)} = \frac{0.0125}{0.38} = 0.32$$

$$P(B_2|A) = \frac{0.012}{0.38} = 0.315$$

$$P(B_3|A) = \frac{0.0135}{0.35} = 0.355$$

The highest output produced by machine is 0.355

Multinomial Theorem

If a dice has (f) faces marked with 1, 2, 3, --- the probability of throwing a total (P) with (N) dice is given by the coefficient of x^n in the expansion,

$$\boxed{\text{Prob} = \frac{\text{coeff of } x^n \text{ in } (x + x^2 + x^3 + \cdots + x^f)^N}{n(S)}}$$

where, $n(S) = f^N$

Q) Find the chance of throwing 10 exactly in one throw with 3 dice

Sol $n = 10$, $N = 3$, $f = 6$ faces (by default) $= 6$

Coeff of x^n in $(x + x^2 + x^3 + \cdots + x^f)^N$

$= \text{Coeff of } x^{10} \text{ in } \underline{(x + x^2 + x^3 + x^4 + x^5 + x^6)}^3$

$\hookrightarrow$ GP Series

$$\text{Sum} = \frac{a(1-r^n)}{1-r}$$

minus

$$= \text{coeff of } x^{10} \text{ in } x^3\left[\frac{1 - x^6}{1-x}\right]^3$$

$$= \text{coeff of } x^{10} \text{ in } x^3 (1-x^6)^3 (1-x)^{-3}$$

$$\left[(1-x)^{-n} = 1 + nx + \frac{n(n+1)x^2}{2!} + \frac{n(n+1)(n+2)}{3!} + \cdots\right]$$

$= \text{Coeff of } x^7 \text{ in } \left[1 - (x^6)^3 - 3(1)^2(x^6) + 3(1)(x^6)^2\right] \times$

$\left[1 + 3x + 6x^2 + 10x^3 + 15x^4 + 21x^5 + \dots\right]$

$= \text{Coeff of } x^7 \text{ in } (36x^7 - 9x^7)$

$= \text{Coeff of } x^7 \text{ in } 27x^7$

$= \boxed{27}$

$n(S) = f^N = 6^3 = \boxed{108}$

$\text{Required Probability} = \frac{27}{108} = \boxed{\frac{1}{8}}$

Q> Calculate the coeff of co-relations from the following data.

Statistics (x)	Maths (Y)	R_x	R_y	$d = R_x - R_y$	d^2
45	40	$\frac{5+6}{2} = 5.5$	7	−1.5	2.25
56	36	$\frac{8+9}{2} = 8.5$	5	3.5	12.25
39	30	3	2	1	1
54	44	7	9	−2	4
45	36	5.5	5	0.5	0.25
40	32	4	3	1	1
56	45	8.5	10	−1.5	2.25
60	42	10	8	2	4
30	20	1	1	0	0
36	36	2	5	−3	9

$\Sigma d^2 = 36$

Formula :-

$$r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}\left\{(m_1^3 - m_1) + (m_2^3 - m_2) + (m_3^3 - m_3)\right\}\right]}{n(n^2-1)}$$

$m_1 = 2$, $m_2 = 2$, $m_3 = 3$, $n = 10$

↳ no. of times a number repeats

$$r = 1 - \frac{6\left[36 + \frac{1}{12}\left\{(8-2) + (8-2) + (27-3)\right\}\right]}{10(100-1)}$$

$$r = 1 - \frac{6(36+3)}{10 \times 99} = \boxed{0.7636}$$

→ always b/w −1 & +1

Q> Calculate μ_1, μ_2, μ_3 from the following series

x:	0	1	2	3	4	5	6	7	8
f:	1	9	26	59	72	52	29	7	1

$$\mu_1 = \frac{1}{N} \Sigma f(x-m)$$

$$= \frac{1}{N}\left(\Sigma fx - \Sigma mf\right)$$

$$= \frac{1}{N}\left[\Sigma fx - m\Sigma f\right]$$

$$\because N = \Sigma f$$

$$= \frac{\Sigma fx}{\Sigma f} - m\cdot\frac{\cancel{\Sigma f}}{\cancel{\Sigma f}}$$

$$= m - m$$

$$\mu_1 = 0$$

$x-A$	$f(x-A)$
-4	-4
-3	-27
-2	-52
-1	-9
0	0
1	52
2	58
3	21
4	4

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

$$\mu_1' = \frac{1}{N} \Sigma f(x-A) = \frac{1}{256}(-7)$$

$$\boxed{\mu_1' = -0.027}$$

$$\mu_2' = \frac{1}{N} \Sigma f(x-A)^2 = \frac{1}{256}(507)$$

$$\boxed{\mu_2' = 1.98}$$

$(x-A)^2$	$(x-A)^3$	$f(x-A)^2$	$f(x-A)^3$
16	-64	16	-64
81	-27	81	-243
104	-8	104	-208
54	-1	59	-59
0	0	0	0
52	1	52	52
116	8	116	232
63	27	63	189
16	64	16	64
		507	-37

Kurtosis $= \beta_2 = \dfrac{\mu_4}{\mu_2^2}$

(i) $\beta = 3$ (mesokurtic)
(ii) $\beta > 2$ (leptokurtic)
(iii) $\beta < 3$ (platykurtic)

γ_1 = Skewness $= \dfrac{\mu_3}{\mu_2^3}$