Sub :- Probability & Statistics

1. Probability

* Imp points

(1) Sample space (S) :- The set of all possible outcomes is called as Sample space

(2) Event (E) :- The set of desired ~~outco~~ outcomes is called as an Event. I

(3) Probability of an Event (P(E)) :- If S is a sample space and A is its event, then probability of A denoted by P(A) is defined as

$$\boxed{P(A) = \frac{n(A)}{n(S)}}$$

* Axioms of Probability

(1) Probability of an event lies between 0 and 1

i.e $0 \leq P(A) \leq 1$

(2) Addition Theorem:

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

(3) Addition Theorem for mutually exclusive (disjoint) events.

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

(4) Conditional Probability

The probability of A such that B has already occured, is called as "Conditional Probability of A given B"

It is denoted by $P(A/B)$ and it is given by

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Similarly, $P(B/A) = \frac{P(A \cap B)}{P(A)}$

(5) Multiplication Theorem

$$P(A \cap B) = P(A/B) \times P(B)$$

$$P(A \cap B) = P(B/A) \times P(A)$$

(6) Independent Events

Events which are not dependent on the action of another event, are called as independent event.

If A & B are independent events

- $P(A/B) = P(A)$ & • $P(B/A) = P(B)$

- $\therefore \boxed{P(A \cap B) = P(A) \times P(B)}$

(7) Probability of not occuring an event

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

(8) De-Morgan's Laws

(i) In sets

- $(\overline{A \cup B}) = \bar{A} \cap \bar{B}$
- $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$

(ii) In probability

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

(9) $P(A \text{ only}) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$

$P(B \text{ only}) = P(\bar{A} \cap B) = P(B) - P(A \cap B)$

$P(A \,\&\, B \text{ happening once}) = P(A \cup B) - P(A \cap B)$

(10) Baye's Theorem

There are 'n' ~~AA~~ disjoint events say $E_1, E_2, E_3, \ldots\ldots, E_n$

Event A is subset of $\bigcup_{i=1}^{n} E_i$

Then,

$$\boxed{P(E_i/A) = \frac{P(A/E_i) \times P(E_i)}{\sum P(A/E_i) \times P(E_i)}}$$

NOTE:- $P(A) = \sum P(A/E_i) \times P(E_i)$

* Exercise

(1) Given, $P(A) = \cancel{0.03}\ 0.23$, $P(B) = 0.5$, $P(A \cap B) = 0.20$
(i) Find $P(A \cup B)$; $P(A/B)$ (iii) Are A & B independent.

⇒ Solution,

Given,
$P(A) = 0.23$; $P(B) = 0.5$; $P(A \cap B) = 0.20$

(i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .... (Using Addition Theorem)
$= 0.23 + 0.5 - 0.20$
$= 0.23 + 0.3 = \boxed{0.53}$

(ii) $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.20}{0.5} = \frac{2}{5} = \boxed{0.4}$
.... (using conditional probability)

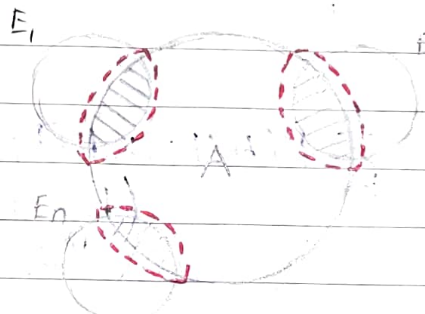
(iii) R.H.S $= P(A) \times P(B) = 0.23 \times 0.5 = 0.115$
L.H.S $= P(A \cap B) = 0.20$
∵ $\boxed{LHS \neq RHS}$
∴ A & B are not independent.

State and Prove Baye's Theorem :

Given : $E_1, E_2, \ldots\ldots, E_n$ are disjoint events.
A is subset of $\bigcup_{i=1}^{n} E_i$

Then,

$$P(E_i/A) = \frac{P(A/E_i) \times P(E_i)}{\sum P(A/E_i) \times P(E_i)}$$



We have to find the ratio of Probability of individual red dotted region to the complete region of A (i.e Summation of all the $P(A \cap E_i)$ terms)

Proof

A is a subset of $\bigcup_{i=1}^{n} E_i$

$$\therefore A \subseteq \bigcup_{i=1}^{n} E_i$$

We know that,

$$A = A \cap \left(\bigcup_{i=1}^{n} E_i \right)$$

$$= A \cap (E_1 \cup E_2 \cup E_3 \cup \ldots\ldots \cup E_n)$$

$$= (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \ldots\ldots \cup (A \cap E_n)$$

But $A \cap E_1$, $A \cap E_2$; $A \cap E_3, \ldots\ldots, A \cap E_n$ are also disjoint in nature

∴ Therefore,
Using addition theorem for disjoints events we get,

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + \ldots + P(A \cap E_n)$$

$$\therefore P(A) = \sum_{i=1}^{n} P(A \cap E_i)$$

$$\therefore P(A) = \sum_{i=1}^{n} P(A/E_i) \times P(E_i) \quad \text{—— (1)}$$

.... (By using multiplication theorem)

Consider ~~RHS~~ L.H.S,

$$\therefore LHS = P(E_i/A)$$

$$= \cancel{P(A/E_i)}$$

$$= \frac{P(A \cap E_i)}{P(A)}$$ .... (By using conditional probability theorem)

$$= \frac{P(A/E_i) \times P(E_i)}{\sum_{i=1}^{n} P(A/E_i) \times P(E_i)}$$ .... (By using multiplication theorem and ~~A~~ from (1))

$= R.H.S$

= Hence proved

* Examples

Q1 In the year 2022, there were 3 candidates for the position of principal. Their chances of getting the appointment are in the ratio 2 : 3 : 5

Info 1 The names of 3 candidates are :- Mr. Chatterji, Mr Iyanger and Mr. Wagh respectively.

Info 2 If Mr. Chatterji is selected as principal, he will introduce computer education in the college. The probability of doing so is 0.3

Info 3 The probability of Mr. Iyanger & Mr. Wagh doing the same are 0.5 and 0.8 respectively.

Find the probability that there was computer ~~educa~~ education in the college in 2023.

⇒ **Solution**

Let E_1 = Mr Chatterji is selected as Principal.
E_2 = Mr Iyangar is ——— " ———
E_3 = Mr Wagh is ——— " ———
A = Computer Education is Introduced.

Given,

$$P(E_1) = \frac{2}{10} \; ; \quad P(E_2) = \frac{3}{10} \; ; \quad P(E_3) = \frac{5}{10}$$

$$P(A/E_1) = 0.3 \; ; \quad P(A/E_2) = 0.5 \; ; \quad P(A/E_3) = 0.8$$

To find: $P(A)$

* Formula :- $P(A) = \sum_{i=1}^{n} P(A/E_i) \times P(E_i)$

$\therefore P(A) = P(A/E_1) \times P(E_1) + P(A/E_2) \times P(E_2) +$
$P(A/E_3) \times P(E_3)$

$= 0.3 \times \frac{2}{10} + 0.5 \times \frac{3}{10} + 0.8 \times \frac{5}{10}$

$= 0.06 + 0.15 + 0.40$

$= \boxed{0.61}$

Date: 05/04/2023

Q2 Find the probability that a 2-digit number constructed using the digits 1, 2, 3, 4, 5 is even if (i) The digits are not allowed to repeat
(ii) Repetition of digits is allowed.

$\Rightarrow$ Solution,

(i) No Repetition is allowed.

~~(i) No Repetition is allowed~~

Sample space = Construct a two-digit number using 1, 2, 3, 4, 5

Ten's place	Unit's place
	1, 2, 3, 4, 5
(4 choices)	(5 choices)

$\therefore n(S) = 4 \times 5 = 20.$

~~or~~ or

$n(S) = {}^5P_2 = \frac{5!}{3!} = \frac{5 \times 4 \times \cancel{3!}}{\cancel{3!}} = 20$

Let A be the event of getting no ~~repit~~ repeated numbers and also a even a number

Ten's place	Unit's place
	2, 4
(4 choices)	(2 choices)

$\therefore n(A) = 4 \times 2 = 8$

$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{8}{20} = 0.4$

(ii) Repetition is allowed.

Sample space (S) = Construct a two-digit number using 1, 2, 3, 4, 5

Ten's place	Unit's place
1, 2, 3, 4, 5	1, 2, 3, 4, 5
(5 choices)	(5 choices)

$\therefore n(S) = 5 \times 5 = \cancel{20}\,25$

or

~~$n(S) = {}^5P_5 = \frac{5!}{0!} =$~~

Let A be the event of getting a even number

Event (A) = The 2 digit n'o is

Ten's place	Unit's place
1, 2, 3, 4, 5	2, 4
(5 choices)	(2 choices)

$\therefore n(A) = 5 \times 2 = 10$

$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{10}{25} = \frac{2}{5}$

Q3 Given $P(A \cap B) = 0.07$, $P(A) = 0.03$, $P(B) = 0.17$
Find (i) $P(A \cup B)$ (ii) $P(A/B)$ (iii) To check for independence

⇒ Solution,
(i) $P(A \cup B)$

Using Addition Theorem,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.03 + 0.17 - 0.07$
$= 0.20 - 0.07 = \underline{\underline{0.13}}$

(ii) $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.07}{0.17} = \frac{7}{17}$

(iii) To check for independence.

We know that,

$$P(A \cap B) = P(A) \times P(B)$$

$\therefore$ RHS $= P(A) \times P(B)$

$= 0.23 \times 0.17$

$= 391 \times 10^{-4} = \underline{0.0391}$

23
17
161
23
391

LHS $= \underline{0.07}$

$\therefore$ L.H.S $\neq$ R.H.S

$\therefore$ Both A & B are not independent events

Date 06/04/2023

* Probability of Random variables

It is a variable where exact value is not known before the experiment but all the values and probabilities are known before the experiment.

* Types of Random variables

1. Discrete Random variable
2. Continuous Random variable

Difference between Discrete R.V & Continuous R.V.

(Single) Discrete Random variable	(Single) Continuous Random variable
1. It takes only discrete values eg:- 0, 1, 2, 3, ...	1. It take all the values in an interval eg:- $0 \leq x \leq 5$
2. Probability ~~of~~ function or ~~of~~ D.R.V is called as Probability Mass Function (PMF).	2. Probability function of C.R.V is called as Probability Density Function (PDF)
3. The probability mass function is generally denoted by $P(X)$ (Rarely denoted by $f(X)$)	3. The probability density function is generally denoted by $f(x)$ (Rarely denoted by $P(X)$)
4. The cummulative distribution function (C.D.F) ~~(cdf)~~ is denoted by $F(X=x)$. It is defined as $F(X=x) = \sum_{X=-\infty}^{x} P(X)$	4. The cummulative distribution function (C.D.F) is denoted by $F(X=x)$. It is defined as $F(X=x) = \int_{-\infty}^{x} f(x) \cdot dx$
5. Total probability	

5. Total probability = 1

$$\therefore \sum_{x=-\infty}^{\infty} P(X) = 1$$

5. Total Probability = 1

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

Examples

(1) A coin is tossed 3 times. Find the p.m.f & c.d.f of R.V.X, where X = "Number of heads".

⇒ ~~Sop~~ Solution

Sample space (S) = A coin is tossed 3 times

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

∴ R.V.X = N.o of heads = {3, 2, 2, 1, 2, 1, 1, 0}

The ~~p~~ P.M.F, P(x) and CDF F(x) are shown in given table

X	P(x)	F(x)
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

(2) A random variable 'x' takes values 0, 1, 2, 3 & 4 with the probabilities ~~P(x=0) = P 2P(x=1) = 3~~

$$P(X=0) = 2P(X=1) = 3P(X=2) = 4P(X=3) = P(X=4)$$

Find the P.M.F of X.

⇒ Solution,

Let,

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

$\therefore P(X=0) = K$

$2P(X=1) = K \quad \therefore P(X=1) = \frac{K}{2}$

$3P(X=2) = K \quad \therefore P(X=2) = \frac{K}{3}$

$4P(X=3) = K \quad \therefore P(X=3) = \frac{K}{4}$

$P(X=4) = K.$

We know that,
Sum of all probabilities is 1.

$\therefore P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 1$

$\therefore\ K + \frac{K}{2} + \frac{K}{3} + \frac{K}{4} + K = 1$

$\therefore\ 2K + \frac{2K}{4} + \frac{K}{4} + \frac{K}{3}\ \cancel{K} = 1$

$\therefore\ 2K + \frac{3K}{4} + \frac{K}{3} = 1$

$\therefore\ 24K + 9K + 4K = 12$

$\therefore\ 37K = 12$ $\quad\therefore\ \boxed{K = \frac{12}{37}}$

The P.M.F is shown below

X	$P(x)$
0	$K = 12/37$
1	$K/2 = 12/37 \times 1/2 = 6/37$
2	$K/3 = 12/37 \times 1/3 = 4/37$
3	$K/4 = 12/37 \times 1/4 = 3/37$
4	$K = 12/37$

* Joint Probability

The probability ~~of~~ depending on more than one variable is called as joint probability.

Joint Probability of...

Discrete R.V	Continuous R.V
(1) It depends upon ~~more~~ more than one discrete random variable	(1) It depends upon more than one continuous ~~rat~~ random variable.
(2) For bivariate (Two variables): It is denoted by $P(X, Y)$	(2) For bivariate: It is denoted by $f(X, Y)$
(3) The probability ~~of~~ of ~~on~~ one variable out of two is called as Marginal probability	(3) The probability of one variable out of two is called as Marginal probability
(4) The marginal probability of X is denoted by $P(x)$ or $P_x(x)$	The marginal probability of X is denoted by ~~P~~ $f(x)$ or $f_x(x)$

~~It is given by:~~

~~P(X) = P~~

It is given by:

$$P(X) = \sum_{Y=-\infty}^{\infty} P(x,y)$$

Similarly,

$$P(Y) = \sum_{x=-\infty}^{\infty} P(x,y)$$

It is given by:

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

lly,

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

* Examples

① Given the following bivariate probability distribution for (x, y)

y \ x	-1	0	1
1	1/27	3/27	4/27
2	3/27	2/27	6/27
3	2/27	5/27	1/27

NOTE:- Even if the question doesn't asks to find marginal probability find it for every such question

(i) Find the marginal probabilities of X & Y
(ii) Find probability when $x \leq 0$, i.e $P(X \leq 0)$, $P(X \geq 0)$, $P(Y=2)$
(iii) Find ~~margi~~ $P(X=0) + P(X=1) + P(X=-1)$
(iv) Find $P(X \leq 0, Y \leq 2)$
(v) Find $P(X \geq 0, Y=3)$

⇒ Solution

(i) The marginal probability $P(X)$ & $P(Y)$ are shown below.

Y \ X	-1	0	1	$P(Y)$
1	$1/27$	$3/27$	$4/27$	$8/27$
2	$3/27$	$2/27$	$6/27$	$11/27$
3	$2/27$	$5/27$	$1/27$	$8/27$
$P(X)$	$6/27$	$10/27$	$11/27$	$\sum\sum P(x,y) = 1$

(ii) To find $P(X \leq 0)$

$$\therefore P(X \leq 0) = P(X=-1) + P(X=0)$$

$$= \frac{6}{27} + \frac{10}{27} = \cancel{\frac{17}{27}}\ \frac{16}{27}$$

To find $P(X \geq 0)$

$$\therefore P(X \geq 0) = P(X=0) + P(X=1)$$

$$= \frac{10}{27} + \frac{11}{27} = \frac{21}{27}$$

$$P(X \geq 0) = 1 - P(X<0)$$
$$= 1 - P(X=-1)$$
$$= 1 - \frac{6}{27}$$
$$= \frac{27-6}{27} = \frac{21}{27}$$

To find $P(Y=2)$

$$\therefore P(Y=2) = \frac{11}{27}$$

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

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

$$= \frac{10}{27} + \frac{11}{27} + \frac{6}{27} = 1$$

(iv) Find $P(X \leq 0, Y \leq 2)$

$$P(X \leq 0, Y \leq 2) = P(X=-1, Y=1) + P(X=-1, Y=2) + P(X=0, Y=1) + P(X=0, Y=2)$$

$$= \frac{1}{27} + \frac{3}{27} + \frac{3}{27} + \frac{2}{27} = \frac{9}{27}$$

(v) Find $P(X \geq 0, Y=3)$

$$\therefore P(X \geq 0, Y=3) = P(X=0, Y=3) + P(X=1, Y=3)$$

$$= \frac{5}{27} + \frac{1}{27} = \frac{6}{27}$$

Date 12/04/2023

Stochastic Independence $[P(A \cap B) = P(A) \times P(B)]$

Random variable's X and Y are said to be Stochastically independent if.

$$P(X, Y) = P(X) \cdot P(Y) \Rightarrow \text{For discrete R.V}$$

$$f(x, y) = f(x) \cdot f(y) \Rightarrow \text{For Continuous R.V}$$

* Conditional Probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Discrete R.V

$$P(X/Y) = \frac{P(X, Y)}{P(Y)}$$

OR

$$P(Y/X) = \frac{P(X, Y)}{P(X)}$$

Continuous R.V

$$f(X/Y) = \frac{f(X, Y)}{f(Y)}$$

$$f(Y/X) = \frac{f(X, Y)}{f(X)}$$

* Examples

① Two ~~r~~ Discrete Random variable. Following joint probability mass function.

X \ Y	1	2	3
1	2/16	2/16	1/16
2	3/16	2/16	1/16
3	2/16	1/16	2/16

(i) Find $P(X \leq 1\frac{1}{2})$,

(ii) ~~com~~ conditional probability distribution of X given $Y=2$

(iii) Are X & Y independent?

⇒ Solution

x \ y	1	2	3	P(X)
1	2/16	2/16	1/16	5/16
2	3/16	2/16	1/16	6/16
3	2/16	1/16	2/16	5/16
P(Y)	7/16	5/16	4/16	ΣΣ 1

1 ⇒ Find $P(X \leq 1\frac{1}{2})$

$\therefore P(X \leq 1\frac{1}{2}) = P(X=1) = 5/16$

2 To Find $P(X/Y=2)$ ⇐ X not specified ∴ Find for all values

$\therefore P(X/Y=2) = \{P(X=1/Y=2);\ P(X=2/Y=2),\ P(X=3, Y=2)\}$

$\therefore$ (i) $P(X=1/Y=2) = \dfrac{P(X=1, Y=2)}{P(Y=2)}$

$= \dfrac{2/16}{5/16} = \dfrac{2}{5}$

(ii) $P(X=2/Y=2) = \dfrac{P(X=2, Y=2)}{P(Y=2)}$

$= \dfrac{2/16}{5/16} = \dfrac{2}{5}$

(iii) $P(X=3/Y=2) = \frac{P(X=3, Y=2)}{P(Y=2)}$

$= \frac{1/16}{5/16} = \frac{1}{5}$

(3) For Independence

$\therefore P(X,Y) = P(X) \cdot P(Y)$

* for $x=1$ & $y=1$

$L.H.S = P(X=1, Y=1) = \frac{2}{16}$

$R.H.S = P(X=1) \times P(Y=2) = \frac{5}{16} \times \frac{7}{16} = \frac{35}{254}$

Since $L.H.S \neq R.H.S$

$\therefore$ X & Y are not independent

NOTE

(1) Given ~~PDF~~ CDF, ~~To find~~ How to find ~~PDF?~~ PDF?

$$\boxed{f(x) = \frac{d}{dx} F(x)}$$

(2) Given PDF, How to find ~~PDF~~ CDF?

$$\boxed{F(X) = \int_{x=0}^{x} f(x)}$$

(3) Given C.D.F of 2 Discrete Random Variable (x, y), How to find PDF ?

$$\boxed{\text{PDF of } f(x,y) = \frac{\delta}{\delta x}\frac{\delta}{\delta y} F(x,y)}$$

(4) Joint Distribution Function (CDF ← Marginal Density) of x and y is given by

$$\boxed{\begin{aligned} F(x,y) &= 1-e^{-x}-e^{-y}+e^{-(x+y)}, & x\geq 0 \ \& \ y \geq 0 \\ &= 0, & \text{Otherwise} \end{aligned}}$$

Questions

(1) Find (i) PDF of x, y
(ii) Marginal density functions of x and y
(iii) Conditional density functions of x/y & y/x
(iv) Whether x & y are dependent or not.

⇒ Solution:

Given CDF,

$$F(x,y) = 1-e^{-x}-e^{-y}+e^{-(x+y)}, \quad x\geq 0 \ \& \ y\geq 0$$

$$= 0, \quad \text{Otherwise}$$

(i) To find PDF ⇒ $f(x,y)$

$$f(x,y) = \frac{\delta}{\delta x} \cdot \frac{\delta}{\delta y} F(x,y)$$

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

$$= \frac{\delta}{\delta x} \left[0 - 0 + e^{-y} - e^{-(x+y)}\right]$$

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

$$= \cancel{\delta} 0 + e^{-(x+y)}$$

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

(ii) To find Marginal Density of x i.e $f(x)$

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

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

$$= \int_{y=0}^{\infty} e^{-x} \cdot e^{-y} \cdot dy$$

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

$$= e^{-x}\left[\frac{e^{-\infty}}{-1} - \frac{e^{0}}{-1}\right]$$

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

$$= e^{-x}[1] = \underline{\underline{e^{-x}}}$$

Similarly, $\boxed{f(y) = e^{-y}}$

(iii) To find conditional Density function of x/y

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

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

$$= \frac{e^{-x}\cdot \cancel{e^{-y}}}{\cancel{e^{-y}}} = \boxed{e^{-x}}$$

Similarly, $\boxed{f(y/x) = e^{-y}}$

(iv) To check whether x & y are independent or not.

$$f(x,y) = f(x)\cdot f(y)$$

$$L.H.S = f(x,y) = e^{-(x+y)}$$

$$R.H.S = f(x)\cdot f(y) = e^{-y}\cdot e^{-x} = e^{-(x+y)} = \cancel{R.H.}\ L.H.S$$

$\because$ LHS = RHS

$\therefore$ x & y are independent of each other.

~~Q2~~

(v) Find (i) $P(X \le 1, Y \le 1)$ & (ii) $P(X+Y \le 1)$

$\Rightarrow$ (i) $P(X \le 1, Y \le 1)$

$$P(X \le x, Y \le y) = \int_{x=0}^{x} \int_{y=0}^{y} f(x,y) \cdot dx \cdot dy$$

$$= \int_{x=0}^{1} \int_{y=0}^{1} e^{-x} \cdot e^{-y} \, dx \cdot dy$$

$$= \int_{x=0}^{1} e^{-x} \cdot dx \cdot \int_{y=0}^{1} e^{-y} \cdot dy$$

$$= \left[\frac{e^{-x}}{-1}\right]_0^1 \cdot \left[\frac{e^{-y}}{-1}\right]_0^1$$

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

$$= \left[-e^{-1} + 1\right] \cdot \left[-e^{-1} + 1\right]$$

$$= \boxed{(1 - e^{-1})^2}$$

(ii) $P(X+Y \le 1)$

$$P(X+Y \le 1) = \int_{x=0}^{1} \int_{y=0}^{1-x} e^{-x} \cdot e^{-y} \cdot dx \cdot dy$$

NOTE :- $\because x+y=1$

x	0	0.2	0.5	0.7	~~x~~ 0.4	~~x~~
y	1	0.8	0.5	0.3	~~x~~ 0.6	$1-x$

$$= \int_{x=0}^{1} \int_{y=0}^{1-x} e^{-x} \cdot e^{-y} \cdot dy \cdot dx$$

$$= \int_{x=0}^{1} e^{-x} \cdot dx \int_{y=0}^{1-x} e^{-y} \cdot dy$$

$$= \int_{x=0}^{1} e^{-x} \cdot dx \left[\frac{e^{-y}}{-1} \right]_{0}^{1-x}$$

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

$$= \int_{x=0}^{1} e^{-x} \cdot dx \left[-e^{-(1-x)} + 1 \right]$$

$$= \int_{x=0}^{1} e^{-x} \cdot \left[\cancel{e} - e^{-(1-x)} + 1 \right] \cdot dx$$

$$= \int_{x=0}^{1} -e^{-x} \cdot e^{-(1-x)} + e^{-x} \cdot dx$$

$$= \int_{x=0}^{1} -e^{-\cancel{x}-1+x} + e^{-x} \cdot dx$$

$$= \int_{x=0}^{1} [-e^{-1} + e^{-x}] \cdot dx$$

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

$$= \left[-e^{-1}x - e^{-x}\right]_0^1$$

$$\cancel{= [(-ex^{1} - [(-e^{-1}(1) - (-e^{-1}(0))]}$$

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

$$= \left[-e^{-1} - e^{-1} + 1\right]$$

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

* Expectation & Variance of a Random Variable

* Expectation

If X is a R.V, then its expectation $E(X)$ is defined as

$$\boxed{\begin{aligned} E(X) &= \sum X \cdot P(X) && \text{for discrete R.V} \\ &= \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx && \text{for continuous R.V} \end{aligned}}$$

* Variance

If X is a R.V, then its variance V(X) is defined as

$$\boxed{V(X) = E(X^2) - [E(X)]^2}$$

[where $E(X^2) = \sum X^2 \cdot P(X)$]

* Properties of E(X) and V(X)

1.) $E(X \pm Y) = E(X) \pm E(Y)$

2.) $E(aX \pm bY) = aE(X) \pm bE(Y)$ — Expectations

3.) $E(aX \pm b) = aE(X) \pm \cancel{E(b)}\, b$

4.) $V(X \pm Y) = V(X) \pm V(Y)$

5.) $V(aX \pm bY) = a^2 V(X) \pm b^2 V(Y)$ — Variance

6.) $V(aX \pm b) = a^2 V(X)$

Example

Eg 1 A R.V X takes values −3, −1, 2 and 5 with the probabilities

X	−3	−1	2	5
P(X)	$\frac{2k-3}{10}$	$\frac{k-2}{10}$	$\frac{k-1}{10}$	$\frac{k+1}{10}$

(i) Determine k and PMF of X
(ii) Determine the distribution function of X
(iii) Find the expected value of X
(iv) Find the variance of X.

⇒ **Solution**

(i) To find k

We know $\sum P(X) = 1$

$$\frac{2k-3}{10} + \frac{k-2}{10} + \frac{k-1}{10} + \frac{k+1}{10} = 1$$

$$\frac{5k-5}{10} = 1$$

$$\therefore k-1 = 2 \qquad \therefore \boxed{k=3}$$

(ii) To find CDF (CDF = Distribution Function)

X	$P(X)$	$F(X)$
-3	$\frac{2k-3}{10} = \frac{3}{10}$	$\frac{3}{10}$
-1	$\frac{k-2}{10} = \frac{1}{10}$	$\frac{4}{10}$
2	$\frac{k-1}{10} = \frac{2}{10}$	$\frac{6}{10}$
5	$\frac{k+1}{10} = \frac{4}{10}$	$\frac{10}{10} = 1$

(iii) To find $E(x)$

$$E(x) = \sum x \cdot P(x)$$

$$= -3\left(\frac{3}{10}\right) - 1\left(\frac{1}{10}\right) + 2\left(\frac{2}{10}\right) + 5\left(\frac{4}{10}\right)$$

$$= \frac{-9}{10} - \frac{1}{10} + \frac{4}{10} + \frac{20}{10}$$

$$= \frac{-10+4+20}{10} = \frac{14}{10} = \boxed{1.4}$$

(iv) <u>For variance</u> : To find $V(X)$

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

$$= (-3)^2\left(\frac{3}{10}\right) + (-1)^2\left(\frac{1}{10}\right) + (2)^2\left(\frac{2}{10}\right) + (5)^2\left(\frac{4}{10}\right)$$

$$= 9\left(\frac{3}{10}\right) + 1\left(\frac{1}{10}\right) + \left(\frac{8}{10}\right) + \left(\frac{100}{10}\right)$$

$$= \frac{27+1+8+100}{10} = \frac{136}{10} = \boxed{13.6}$$

$$V(X) = E(X^2) - [E(X)^2]^2$$

$$= 13.6 - 1.4$$

$$= \boxed{11.64}$$

Eg 2 X is a R.V for which $E(X)=10$, $V(X)=25$
Find the values of 'a' & 'b' such that $Y = aX - b$ has expectation zero and variance 1

$\Rightarrow$ Given, $E(X)=10$; $V(X)=25$; $Y = aX - b$ — (1)

We know that,

$E(Y) = 0$ & $V(Y) = 1$ .... (From given)

Consider $E(Y) = 0$

$\therefore E(aX-b) = 0$

$\therefore aE(X) - b = 0$ ..... (By property)

$\therefore 10a - b = 0$ ..... (from (1))

$\therefore \boxed{b = 10a}$ — (2)

Consider $V(Y) = 1$

$V(aX-b) = 1$

$\therefore a^2 V(X) = 1$

$a^2 \cdot 25 = 1$

$\therefore a^2 = \frac{1}{25}$

$\therefore \boxed{a = \pm \frac{1}{5}}$ — (3)

~~But~~,

Substitute value of a in (2)

$b = 10\left(\pm \frac{1}{5}\right)$

$\therefore \boxed{b = \pm 2}$

* Probability Distribution

(1) Bernaulli's Distribution :

A R.V X is said to follow Bernaulli's Distribution if its PMF is given by

X	1	0
P(X)	p	q

(1) Mean of Bernaulli's Distribution = P

(2) Variance of Bernaulli's Distribution = pq

Proof

$$\text{Mean}(X) = \sum X \cdot P(X)$$
$$= 1 \times P + 0 \times q = p$$

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

$$\because E(X^2) = \sum x^2 \cdot p(x)$$
$$= (1)^2 \cdot p + 0^2 \times q$$
$$= p$$

$$\therefore V(X) = E(X^2) - [E(X)]^2$$
$$= p - p^2$$
$$= p(1-p) = pq \quad \ldots\ldots (\because p+q=1)$$

(2) Binomial Distribution

A R.V 'x' is said to be following Binomial Distribution if it takes values 0, 1, 2, ..., n with the probability mass functions.

$$\boxed{\begin{aligned} P(X=x) &= {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}, \quad x = 0, 1, 2, 3, \ldots n \\ &= 0 \quad , \text{ otherwise} \end{aligned}}$$

Here p = Probability that a trial is successful
$q = (1-p)$ Probability of a failure

a) Mean of Binomial Distribution $= np$
b) Variance of Binomial Distribution $= npq$

Proof

$$\text{Mean} = \sum x \cdot p(x)$$

$$= \sum x \cdot {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x} \quad , \quad x = 0, 1, 2, \ldots n$$

$$= \cancel{\sum_{x=0}^{n} x \cdot \frac{n}{x} \cdot C^{n-1}_{x-1}} \quad \sum_{x=0}^{n} x \cdot \frac{n}{x} \cdot C^{n-1}_{x-1} \cdot q^{n-x}$$

$$= \sum_{x=1}^{n} \cancel{x} \cdot \frac{n}{\cancel{x}} \cdot C^{n-1}_{x-1} \cdot q^{n-x} \cdot p^{x} \ldots \text{(Since } x=0 \text{ cannot be divided)}$$

$$= n \sum_{x=0}^{n} {}^{n-1}C_{x-1} \cdot p^{x} \cdot q^{n-x}$$

$$= np \sum_{x=1}^{n} {}^{n-1}C_{x-1} \cdot p^{x-1} \cdot q^{(n-1)-(x-1)}$$

But $\sum_{x=1}^{n} {}^{n-1}C_{x-1} \cdot p^{x-1} \cdot q^{(n-1)-(x-1)} = 1$, since it is the

PMF of a binomial distributions ~~with parameters n-1 and~~

$\therefore$ The sum of the term is 1

$$= np(1)$$

$$= \boxed{np}$$

Eg ① A coin is tossed 10 times. Find the probability of getting (i) Exactly 2 heads
(ii) At least 8 heads
(iii) At most 2 heads

$\Rightarrow$ Solution

Let P = Probability of getting head
(Probability of success)

$p = \frac{1}{2} = 0.5$; $q = 1-p = 1-0.5 = 0.5$

Total n.o of trials, $n=10$

B.D $\Rightarrow P(X=x) = {}^nC_x \cdot p^x \cdot q^{n-x}$

(i) To find $P(x=2)$

$$\therefore P(x=2) = {}^{10}C_2 \cdot p^2 \cdot q^8$$

$$= \frac{10!}{2! \cdot 8!} \cdot (0.5)^2 \cdot (0.5)^8$$

$$= \frac{\cancel{10}^{5} \times 9}{\cancel{2} \times 1} \cdot \left(\frac{1}{2}\right)^{10}$$

$$= 45 \times \frac{1}{1024} = \frac{45}{1024}$$

$$\therefore \boxed{P(x=2) = \frac{45}{1024}}$$

(ii) At-least 8 heads

$\therefore$ To find $P(X>=8)$

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

$$= {}^{10}C_8 \cdot (1/2)^8 \cdot (1/2)^2 + {}^{10}C_9 (1/2)^1 \cdot (1/2)^9 +$$

$${}^{10}C_{10} (1/2)^0 \cdot (1/2)^{10}$$

$$= \frac{45}{1024} + \frac{10}{1024} + \frac{1}{1024} = \frac{56}{1024}$$

$$\therefore \boxed{P(X \geq 8) = \frac{56}{1024}}$$

(iii) At-most 2 heads

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= {}^{10}C_0 (1/2)^0 (1/2)^{10} + {}^{10}C_1 (1/2)^1 (1/2)^9 + {}^{10}C_2 (1/2)^2 \cdot (1/2)^8$$

$$= [{}^{10}C_0 + {}^{10}C_1 + {}^{10}C_2] \cdot (1/2)^{10}$$

$$= [1 + 10 + 45] \cdot (1/2)^{10}$$

$$= \boxed{\frac{56}{1024}}$$

③ Poisson's Distribution

Limiting case of Binomial Distribution

$$\boxed{\text{Mean} = \lambda = np}$$

Probability that there are "x" numbers of success trials is :

$$\boxed{\begin{aligned} P(X=x) &= \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0, 1, 2, \ldots \\ &= 0 \qquad \text{otherwise} \end{aligned}}$$

NOTE :- When there are only 2 events (Either success or failure) and the sample size is large use poisson's distribution method.

Q1 A company manufactures bulbs. It is observed that 2% of electric bulbs manufactured are ~~deffec~~ defective. Find the probability that, in a sample of 200 bulbs
(i) Exactly 4 are defective.
(ii) Atleast 2 defective
(iii) Atmost 3 defective

⇒ Solution,
let p = probability that a bulb is defective
$\therefore p = 0.02$

Sample size $n = 200$
Mean $= \lambda = np = 200 \times 0.02 = 4$

We know that,
According to Possion's Distribution

$$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

(i) Exactly 4 are defective

$$P(X=4) = \frac{e^{-4} \cdot 4^{4}}{4!} = \boxed{0.1954}$$

(ii) To find $P(X \geq 2)$

$$P(X \geq 2) = 1 - P(X \leq 2)$$

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

$$P(X=0) = \frac{e^{-4} \cdot 4^{0}}{0!} =$$

$$P(X=1) = \cancel{e^{-4}\lambda^{1}} \frac{e^{-4}\cdot 4^{1}}{1!} =$$

$$P(X\geq 2) = 1 - \left[\frac{e^{-4}\cdot 1}{0!} + \frac{e^{-4}\cdot 4^{1}}{1!}\right]$$

$$= 1 - [1+4]\cdot e^{-4}$$

$$= 1 - (5)\times 0.0183$$

$$= 1 - 0.0915 = \boxed{0.9085}$$

(iii) Atmost 3 defectives $\Rightarrow P(X\leq 3)$

$$P(X\leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= \frac{e^{-4}4^{0}}{0!} + \frac{e^{-4}\cdot 4^{1}}{1!} + \frac{e^{-4}\cdot 4^{2}}{2!} + \frac{e^{-4}\cdot 4^{3}}{3!}$$

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

$$= e^{-4}\left[1+4+8+\frac{32}{3}\right]$$

$$\frac{4\times 4\times 4}{3\times 2\times 1}$$

$$= 0.0183\times[13+10.6667]$$

$$= 0.0183\times 23.6667$$

$$= \boxed{0.4331}$$

Bernaulli's Distribution
Binomial Distribution
Poisson's Distribution

$\Big\}$ ⇒ Discrete Random Variable

Normal Distribution ⇒ Continuous Random Variable

* Normal Distribution

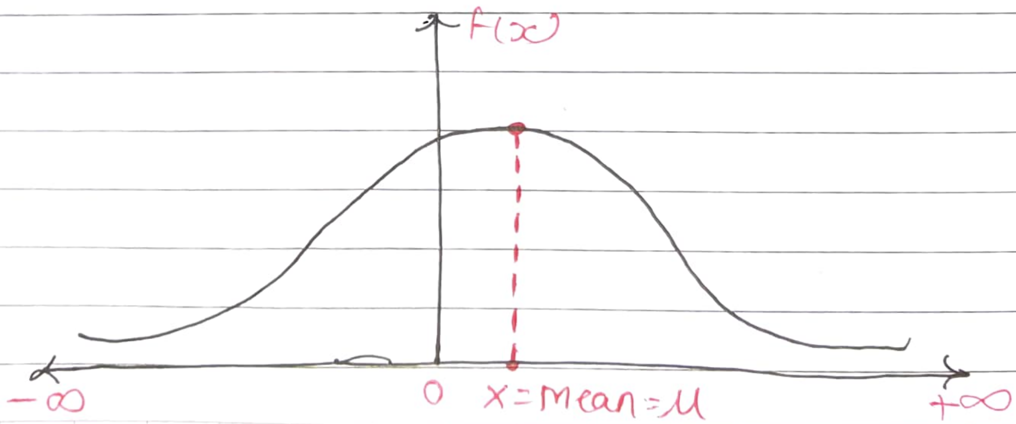
A continuous R.V is said to follow normal distribution if its P.D.F is given by

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

~~wh~~ Where σ = Standard deviation ~~[illegible]~~
μ = Mean of ~~st~~ distribution

Properties of Normal Distribution

(1)



The graph of $f(x)$ is bell-shaped having peak at $x = \mu$

(2) The graph of $f(x)$ is symmetric about the line $x = \mu$

(3) I.E

$$P(x \leq \mu) = P(x \geq \mu) = 0.5$$

(4) If we substitute, $Z = \frac{X - \mu}{\sigma}$ then we get,

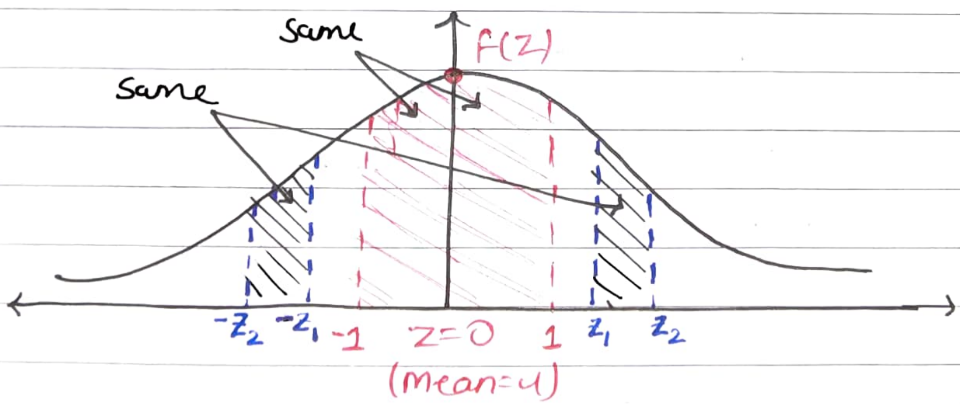
$$\boxed{f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(Z)^2}}$$

This variable Z is called as standard normal variable.

(5) (i) Mean of Standard Normal Variable is 0

(ii) Standard Deviation of Standard normal variable is 1

(6)



The graph of $f(Z)$ is bell shaped having peak at $Z = 0$

(7) The graph of $f(z)$ is symmetric about the vertical axis (i.e $z=0$)

(8) $P(Z \leq 0) = P(Z \geq 0) = 0.5$

(9) $P(-z \leq Z \leq 0) = P(0 \leq Z \leq z)$

(10) $P(-z_2 \leq Z \leq -z_1) = P(z_1 \leq Z \leq z_2)$

NOTE :-

We cannot solve the problems on Normal variable as it is.

Therefore, we transform given normal variable to Standard normal variable $Z = \frac{X-\mu}{\sigma}$

Use of probabilities of SNV 'Z' from statistical tables & solve the problem.

Ex 1 Marks of a competitive exam follows normal distribution with mean 60 and S.D 5. Find the probability that marks scored by a student are (i) More than 65 (ii) Between 55 & 65
[Given $P(0 \leq Z \leq 1) = 0.3413$]

⇒ Solution

Let X = Marks scored

∴ $X \sim$ Normal Distribution ... (Given)

Given,

Mean$(\mu) = 60$, & S.D$(\sigma) = 5$

(i) To find $P(X > 65)$

$$P(X > 65) = P\left(\frac{X-\mu}{\sigma} > \frac{65-\mu}{\sigma}\right)$$

CONVERSION FROM X to Z

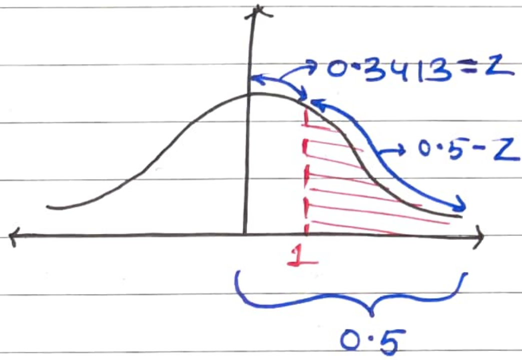
$$= P\left(Z > \frac{65-60}{5}\right)$$

$$= P(Z > 1)$$

$$= 0.5 - P(0 \leq Z \leq 1)$$

$$= 0.5 - 0.3413$$

$$= \underline{\underline{0.1587}}$$



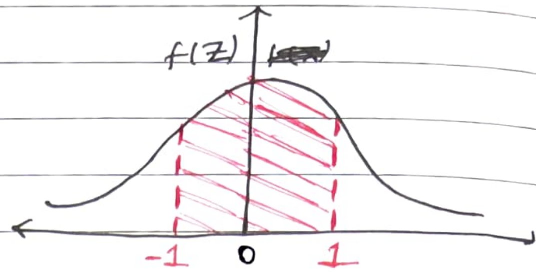
(ii) To find $P(55 < X < 65)$

$$P(55 < X < 65) = P\left(\frac{55-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{65-\mu}{\sigma}\right)$$

$$= P\left(\frac{55-60}{5} < Z < \frac{65-60}{\cancel{6}5}\right)$$

$$= P(-1 < Z < 1)$$

$= 2 \cdot P(0<Z<1)$
$= 2 \times 0.3413$
$= 0.6826$



Let $N = \cancel{100}$ 1000 students
$E(X) = N \times P(55 < X < 65)$
$= 1000 \times 0.6826$
$= 682.6 \approx \boxed{683}$ Expectation

Ex2 Marks of students follows normal distribution with mean 60 and standard deviation 5. How many students in a ~~sam~~ group of 1000 students will score marks
(i) Less than 55 (ii) Between 55 & 65
[Given $P(0 \le Z \le 1) = 0.3413$]

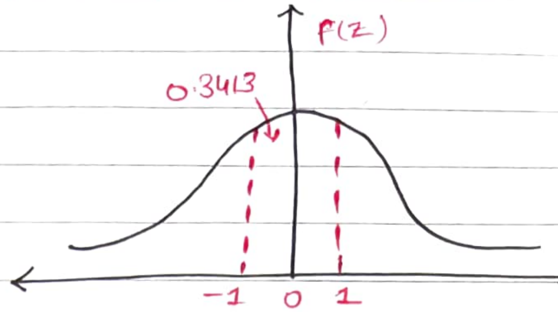
⇒ Solution
Let X = ~~Marks~~ Marks scored by students

Given X ~ Normal Distribution
Mean $= \mu = 60$
S.D $= \sigma = 5$

(i) To find $P(X < 55)$

$$P(X<55) = P\left(\frac{X-\mu}{\sigma} < \frac{55-\mu}{\sigma}\right)$$

$$= P\left(Z < \frac{35-60}{5}\right)$$



$$= P(Z < -1)$$

$$= 0.5 - P(-1 < Z < 0)$$

$= r$

i.e. $= 0.5 - P(0 < Z < 1)$

$$= 0.5 - 0.3413$$

$$= \boxed{0.1587}$$

Expected NO $= N \times P(x < 55)$

$$= 1000 \times 0.1587 = 158.7 = \boxed{159}$$

(ii) To find $P(55 < x < 65)$

$$P(55 < x < 65) = P\left(\frac{55-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{65-\mu}{\sigma}\right)$$

$$= P\left(\frac{55-60}{5} < Z < \frac{65-60}{5}\right)$$

$$= P(-1 < Z < 1)$$

$$= 2 \cdot P(0 < Z < 1)$$

$$= 2 \times \cancel{0.6826}\, 0.3413 = \boxed{0.6826}$$

Expected NO $= N \times P(55 < x < 65)$

$$= 1000 \times 0.6826$$

$$= 682.6 = \boxed{683}$$

Question Pattern (MSE)

Q1.) Basic Properties of Probability

Q2.) Baye's Theorem

Q3.) ~~R.D~~ PMF of 2-D random variable

Q4.) Joint Distribution Function

Q5.) X is R.V mean = 20, Variance = 4. then Y = AX ± b
Find a & b

Q6.) A coin is tossed 10 times find the probability of getting 4 heads (Binomial Distribution)

~~Q7.) A poultry keeper has 'x'~~ Q7.) (Expectation)

Q8.) A poultry keeper has 5 hens and they gives eggs on 5 days out of 7 days. Find the probability that they give eggs on atleast 4 days

Q8.) Poisson's Distribution

11/05/2023

* Measures of Central Tendancy

1 2 3   1 2 3 4
↑
Middle   Middle = ?

* The 3 measures of central tendency are :-

① Mean (6 formulae)
② Median (3 formulae)
③ Mode (1 formulae)

* Types of Data

There are 3 types of data

① Simple Data (Raw Data):

(i) Mean

Given $x_1, x_2, x_3, \ldots \ldots x_n$

$$\boxed{\text{Mean}(\bar{x}) = \frac{\Sigma x_i}{n}}$$

(ii) Median

When data is arranged in the increasing or decreasing order then the central most observations is called as Median

For eg:-

(i) Given data is

$x_1, x_2, \ldots, x_n$

(a) If n is odd

(i) Arrange in the order

$$\boxed{\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{value}}$$

(b) If n is even

Arrange the data in the order

$$\boxed{\text{Median} = \text{Average of } \left(\frac{n}{2}\right)^{th} \,\&\, \left(\frac{n}{2}+1\right)^{th} \text{ values}}$$

(iii) Mode

Most frequently repeated observation is called as Mode

Eg:- Given,

$x_1, x_2, \ldots, x_n$

NO FORMULA ONLY OBSERVATION.

1 Examples

(i) Find Mean, Median & Mode of the given data

5, 6, 0, 7, 7, 8

⇒ Solution.

(i) To find mean

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5+6+0+7+7+8}{6} = \underline{\underline{5.5}}$$

(ii) To find median

Arrange in ~~order~~ increasing order.

0, 5, 6, 7, 7, 8

∴ n = 6

∴ Since n is a ~~odd~~ even we will be using formula of median for even n.o of numbers.

$$\therefore \text{Median} = \text{Avg} \left(\frac{n}{2}\right)^{th} \& \left(\frac{n}{2}+1\right)^{th} \text{ value}$$

$$= \text{Avg of} \left(\frac{6}{2}\right)^{th} \& \left(\frac{6}{2}+1\right)^{th} \text{ value}$$

$$= \text{Avg of } 3^{rd} \& 4^{th} \text{ value}$$

$$= \frac{6+7}{2} = \underline{\underline{6.5}}$$

(iii) To find mode

$\because$ 7 is the repetiting data

$\therefore$ $\boxed{\text{Mode} = 7}$

② Discrete Frequency Data

(i) Mean

Given,

$x_i = x_1, x_2, x_3, \ldots\ldots, x_n$

$f_i = f_1, f_2, f_3, \ldots\ldots, f_n$

$\therefore$ Mean $(\bar{x}) = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$ ~~where~~ $\cancel{\Sigma f_i = N}$

$= \dfrac{\Sigma f_i x_i}{N}$ where $N = \Sigma f_i$

(ii) Median

SAME AS RAW DATA

NOTE : We prepare cummulative frequency table

(iii) Mode

NO FORMULA. ONLY OBSERVATIONS

Example

Find mean, median & mode

Marks	0	1	3	5	6
N.O of students	3	5	10	5	2

⇒ Solution,

(i) To Find Mean

$$\text{Mean } (\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

x_i	f_i	$x_i f_i$
0	3	0
1	5	5
3	10	30
5	5	25
6	2	12

$\boxed{\sum f_i = 25}$ $\boxed{\sum x_i f_i = 72}$

$\therefore$ Mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{72}{25} = \boxed{2.88}$

(ii) To find median

x_i	f_i	C.F
0	3	3
1	5	8
3	10	18
5	5	23
6	2	25

$\Sigma f_i = 25$

~~$\Sigma f_i = N$~~ $N = \Sigma f_i = 25$

$\therefore$ Since N is odd

$\therefore$ Median $= \left(\dfrac{N+1}{2}\right)^{th}$ value

$= \left(\dfrac{25+1}{2}\right) = 13^{th}$ value

$\therefore$ The 13^{th} value lies in the value 3

$\therefore \boxed{\text{Median} = 3}$

(iii) To find mode

(Having most ~~fre~~ frequencies)

$\therefore \boxed{\text{Mode} = 3}$

③ Class Interval & Frequency Type of Data

⇓

(Grouped Frequency Data)

(i) Mean of Grouped Frequency Data

Given,

Class Interval (C.I)	:	$l_1 - u_1$	$l_2 - u_2$	$l_n - u_n$
Frequency (f_i)	:	f_1	f_2	f_n

Therfore,

$$\underset{(\bar{X})}{\text{Mean}} = \frac{\sum f_i x_i}{\sum f_i} \quad \text{where } x_i = \text{class mark} \quad \text{i.e } x_i = \frac{l_i + u_i}{2}$$

~~Formula~~ Example.

C.I	0-10	10-20	~~30-40~~ 20-30	30-40
F	2	7	10	8

Find mean

C.I	class mark (x_i)	f_i	$x_i f_i$
0-10	5	2	10
10-20	15	7	105
20-30	25	10	250
30-40	35	8	280
40-50	45	8	360
		$\Sigma f_i = 35$	$\Sigma f_i x_i = 1005$

$\therefore$ Mean $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1005}{35} = \boxed{28.71}$

* SHORT CUT METHOD TO FIND MEAN OF GROUPED FREQUENCY DATA.

~~Assumed~~
Let,
(i) Assumed Mean $= a$ (i.e Approximate Middle value of x_i)

(ii) Class width $= C =$ Difference between 2 class limits of a class

(iii) Define new variable $\Rightarrow u_i = \frac{x_i - a}{c}$

$$\boxed{\text{Mean}(\bar{x}) = a + C \times \frac{\Sigma f_i \cdot u_i}{\Sigma f_i}}$$

Same ~~x~~ example.

$\therefore$ Assumed mean $=$ ~~25~~ $a = 25$
Class width $= c = 10$

We prepare the following table

C.I	x_i	f_i	$u_i = \frac{x_i - a}{c}$	$f_i \cdot u_i$
0-10	5	2	-2	-4
10-20	15	7	-1	-7
20-30	25	10	0	0
30-40	35	8	1	8
40-50	45	8	2	16
		$\Sigma f_i = 35$		$\Sigma f_i \cdot u_i = 13$

$$\therefore \text{Mean} = a + c \times \frac{\Sigma f_i u_i}{\Sigma f_i}$$

$$= 25 + 10 \times \frac{13}{35}$$

$$= 25 + 10 \times 0.37$$

$$= 25 + 3.7$$

$$= \boxed{28.7}$$

(ii) Median of Grouped Frequency Data

(a) Prepare the C.F table

$N = \Sigma f_i$

(b) Median class ~~$(l_i - u_i)$~~ $(l_1 - l_2)$

= C.I in which $(N/2)^{th}$ value is present

(c) Frequency of median class = f

(d) Cummulative Frequency of Pre-Median class = F

$$\boxed{\text{Median} = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{2} - F\right)}$$

* Example

C.I:	0-10	10-20	20-30	30-40	40-50	50-60
f	10	15	30	25	25	20

⇒ Find median

⇒ Solution

C.I	f_i	C.F
0-10	10	10
10-20	15	25
20-30	30	55

30-40	25	80
40-50	25	105
50-60	20	125
	$\Sigma f_i = 125$	

$N = \Sigma f_i = 125$ $\therefore \frac{N}{2} = \frac{125}{2} = 62.5$

$\therefore$ Median class ~~$(t_i - t_i)$~~ $(l_1 - l_2)$ =

C.I in which $(N/2)^{th}$ value is present

$= (30-40)$

Frequency of median class $(f) = 25$

C.F of pre-Median class $(F) = 55$

$\therefore$ Median $= l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{2} - F\right)$

$= 30 + \left(\frac{40-30}{25}\right)(62.5 - 55)$

$= 33$

$\therefore \boxed{\text{Median} = 33}$

(iii) Mode of Grouped Frequency Distribution

(a) Modal class $(l_1 - l_2)$ ⇒
(Class Having maximum frequency)

(b) Frequency of Modal class ⇒ f

(c) Frequency of Pre-modal class ⇒ f_1

(d) Frequency of Post modal class ⇒ f_2

(e) Find $d_1 = f - f_1$ & $d_2 = f - f_2$

$$\boxed{\text{Mode} = l_1 + \left(\frac{d_1}{d_1 + d_2}\right)(l_2 - l_1)}$$

Example ((Previous Example)

⇒ Solution

Modal class = $(l_1 - l_2) = (20 - 30)$
Frequency of modal class = $(f) = 30$
Frequency of Pre-modal class $(f_1) = 15$
Frequency of Post-modal class $(f_2) = 25$

$d_1 = f - f_1 = 30 - 15 = 15$
$d_2 = f - f_2 = 30 - 25 = 5$

$$\text{Mode} = l_1 + \left(\frac{d_1}{d_1+d_2}\right)(l_2 - l_1)$$

$$= 20 + \left(\frac{15}{15+5}\right)(30-20)$$

$$= 20 + \frac{15}{20} \times 10$$

$$= 20 + 7.5 = \boxed{27.5}$$

* Quartiles

- Data is arranged in the increasing order.
- Divide the data into 4 equal parts
- We need 3 points to divide into 4 equal parts i.e Q_1, Q_2, Q_3

$$Q_1 = \left(\frac{N}{4}\right)^{th} \text{value} = 1^{st} \text{ Quartile}$$

$$Q_2 = \left(\frac{N}{2}\right)^{th} \text{value} = 2^{nd} \text{ Quartile}$$

$$Q_3 = \left(\frac{3N}{4}\right)^{th} \text{value} = 3^{rd} \text{ Quartile}$$

PAGE No.
DATE / /

* Grouped Frequency Data

(I) To find Q_1

(a) First Quartile class $(l_1 - l_2)$ = C.I in which $(N/4)^{th}$ value of C.F is present

(b) Frequency of Q_1 class = f

(c) C.F of pre-Q_1 class = F

$$\boxed{Q_1 = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{4} - F\right)}$$

(II) To find Q_3

(a) Third Quartile class = C.I in which $(3N/4)^{th}$ value is present

(b) Frequency of Q_3 class = f

(c) C.F of pre-Q_3 class = F

$$\boxed{Q_3 = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{3N}{4} - F\right)}$$

NOTE :- 2^{nd} Quartile is same as Median value

$\therefore \boxed{\text{Median} = Q_2}$

* Measures of Dispersion

* Absolute Measures

① Range: High - Low = Max value - Min value

② Mean Deviation from M : ($\because$ M = Mean, Median, Mode)

(i) Raw Data : $x_1, x_2, \ldots\ldots, x_n$

(a) Mean Deviation from Mean

$$\boxed{MD = \frac{\sum(|x_i - \bar{x}|)}{n}}$$ where $\bar{x}$ = Mean

(b) ~~Median~~ Mean Deviation from median

$$\boxed{MD = \frac{\sum(|x_i - M|)}{n}}$$ where M = Median

(c) ~~Mode~~ Mean Deviation from Mode

$$\boxed{MD = \frac{\sum(|x_i - M|)}{n}}$$ where M = Mode

PAGE No.
DATE / /

(ii) Discrete Frequency Data

Given:

$x_i = x_1, x_2, \ldots, x_n$

$f_i = f_1, f_2, \ldots, f_n$

(a) Mean deviation ~~from mean~~

$$\boxed{MD = \frac{\sum f_i \cdot (|x_i - M|)}{\sum f_i}}$$

where M = Mean, Median, Mode

(iii) Grouped Frequency Data

~~Given~~: Given: C.I:

f_i =

Mean deviation ~~fr~~:

$$\boxed{MD = \frac{\sum f_i (|x_i - M|)}{\sum f_i}}$$

where M = Median, Mean, Mode

x_i = class mark of i

③ Quartile Range = $Q_3 - Q_1$

8. Semi-Inter Quartile Range = $\frac{Q_3 - Q_1}{2}$

(also known as Quartile Deviation)

PAGE No.	
DATE	/ /

④ Standard Deviation (σ)

(i) Raw Data:

Given, $x_1, x_2, \ldots\ldots, x_n$

$$\boxed{\text{Standard Deviation} = \sqrt{\frac{\Sigma(x_i - x)^2}{n}}}$$

Also known as: Root Mean Square (RMS)

$$\boxed{\text{Standard Deviation} = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}}$$

(ii) Discrete Frequency Data

Given: $x_i : x_1, x_2, \ldots\ldots, x_n$
$f_i : f_1, f_2, \ldots\ldots, f_n$

$$\text{Standard Deviation} = \sqrt{\frac{\Sigma f_i(x_i - \bar{X})^2}{\Sigma f_i}}$$

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

(iii) Grouped Frequency Data

Given : C.I : . . . .

f_i : . . . .

$$\text{Standard Deviation} = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}$$

$$= \sqrt{\frac{\sum f_i \cdot x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i}\right)^2}$$

* Relative Measures

(1) Coefficient of Range $= \dfrac{H - L}{H + L}$

(2) Coefficient of M.D $= \dfrac{M.D}{M}$

(3) Coefficient of Q.D $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

(4) Coefficient of Variance $= \dfrac{S.D}{Mean}$

(Standard Deviation)

Examples

(1) The runs scored by 2 players A & B are given below:

A: 101 27 0 36 82 45 7 13 65 14
B: 97 12 40 96 13 8 85 8 56 15

(a) Who is more consistent?
(b) Who is ~~bethe~~ better run getter?

⇒ **Solution**

Player A

~~x_i : 101 27 0 36 82 45 7 13 65 14~~

~~x_i^2~~

x_i	x_i^2
101	
27	
0	
36	
82	
45	
7	
13	
65	
14	

~~Σx_i~~

$\Sigma x_i = 390$ $\Sigma x_i^2 = 25614$

PAGE No.
DATE / /

$\therefore$ Mean $\bar{x} = \frac{\Sigma x_i}{n} = \frac{390}{10} = 39$

$\therefore$ S.D $(\sigma_x) = \sqrt{\frac{\Sigma x_i^2}{\cancel{\Sigma}\,n} - \left(\frac{\Sigma x_i}{n}\right)^2}$

$= \sqrt{\frac{25614}{10} - \left(\frac{390}{10}\right)^2}$

$= \sqrt{\frac{25614}{10} - (39)^2}$

$= \sqrt{2561.4 - 1521}$

$= \sqrt{1040.4} = \boxed{32.2552}$

$C.V_{(x)} = \frac{S.D}{Mean} = \frac{32.2552}{39} = 8\ 0.8271 \quad \therefore 82.71\%$

Player B

y_i	y_i^2
97	
12	
40	
96	
13	
8	
85	

8

56

15

$\Sigma x_i = 430$ $\Sigma x_i^2 = 31252$

$$\text{Mean } \bar{y} = \frac{\Sigma y_i}{n} = \frac{430}{10} = 43$$

$$\therefore SD(\sigma_y) = \sqrt{\frac{\Sigma y_i^2}{n} - \left(\frac{\Sigma y_i}{n}\right)^2}$$

$$= \sqrt{\frac{31252}{10} - \left(\frac{\Sigma 430}{10}\right)^2}$$

$$= \sqrt{3125.2 - (43)^2}$$

$$= \sqrt{3125.2 - 1849}$$

$$= \sqrt{1276.2} = 35.7239$$

$$\therefore C.V_{(\bar{y})} = \frac{S.D_{(y)}}{\text{Mean}_{(y)}} = \frac{35.7239}{43} = 0.8307 \text{ i.e } 83.07\%$$

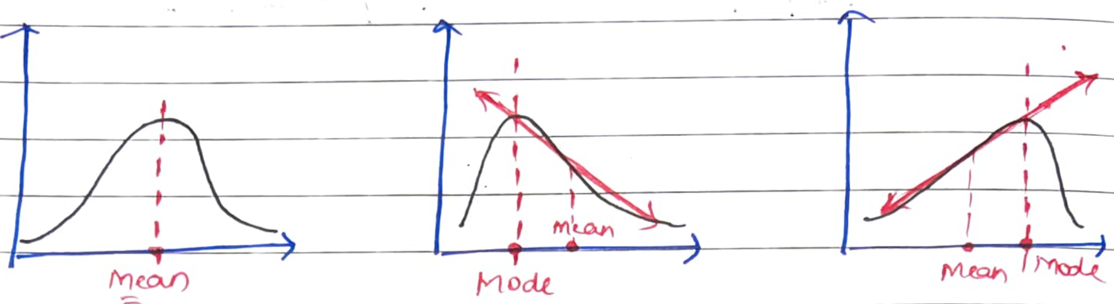
(i) Since C.V of A < C.V of B,
$\therefore$ Player A is consistent

(ii) Since Mean of B > Mean of A
$\therefore$ Player B is better run getter.

PAGE No.
DATE / /

19/05/2023

SKEWNESS



No-Skewed graph

-ve skewed graph

+ve skewed graph

* KARL PEARSON'S COEFFICIENT OF SKEWNESS.

Karl Pearson's coefficient of skewness is defined as

$$\boxed{\frac{\text{Mean} - \text{Mode}}{S.D}}$$

* Bowley's Coefficient of skewness

~~Karl~~
Bowley's coefficient of skewness is defined as

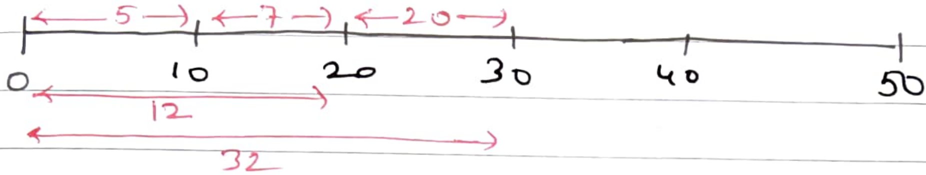
$$\boxed{\cancel{\frac{2Q_2 - Q_1 - Q_3}{Q_3 - Q_1}}} \quad \boxed{\frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

where, $Q_1 = 1^{st}$ Quartile $(N/4)^{th}$ value; $Q_3 = 3^{rd}$ Quartile $(3N/4)^{th}$ value.
$Q_2 = 2^{nd}$ Quartile $(N/2)^{th}$ value;
(Median)

Example

(1) Find karl pearson's coefficient of skewness by following data.

Marks Scored less than :	10	20	30	40	50
N.o of Students :	5	12	32	44	50



⇒ Solution,

(i) To find Mean

C.I	f_i	C.F	x_i	$u_i = \frac{x_i - a}{c}$	$f_i u_i$	$f_i u_i^2$
0-10	5	5	5	-2	-10	20
10-20	7	12	15	-1	-7	7
20-30	20	32	25	0	0	0
30-40	12	44.	35	1	12	12
40-50	6	50	45	2	12	24
	$\Sigma f_i = 50$				$\Sigma f_i u_i = 7$	$\Sigma f_i u_i^2 = 63$

Let us assume Assumed mean $(a) = 25$

$\therefore$ Class width $= 10$

$\therefore$ We know that,

$$\text{Mean } (\bar{x}) = \cancel{\Sigma f u}\ a + c \times \frac{\Sigma f_i u_i}{\Sigma f_i}$$

$$= = 25 + \overset{1}{\cancel{10}} \times \frac{7}{\underset{5}{\cancel{50}}}$$

PAGE No.
DATE / /

$= 25 + 1.4 = \underline{\underline{26.4}}$

$\therefore \boxed{\text{Mean} = 26.4}$

$\therefore$ ~~Sta~~ We know that,

$$S.D = \sqrt{\frac{\Sigma f_i u_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)^2} \times C$$

$$= \sqrt{\cancel{\frac{63}{50} - (26.4)^2}}$$

$$= \sqrt{\frac{63}{50} - \left(\frac{7}{50}\right)^2} \times 10$$

$= 1.1137 \times 10 = \underline{\underline{11.137}}$

$\therefore \boxed{SD = 11.137}$

Highest frequency = 20

$\therefore$ Modal class = 20 − 30

$\therefore l_1 = 20,\ l_2 = 30,\ f = 20,\ f_1 = 7,\ f_2 = 12$

$\therefore d_1 = f - f_1 = 20 - 7 = 13$

$\therefore d_2 = f - f_2 = 20 - 12 = 8$

We know that,

$$\text{Mode} = l_1 + \left(\frac{d_1}{d_1 + d_2}\right) \times (l_2 - l_1)$$

PAGE No.

DATE / /

$$= 20 + \left(\frac{13}{13+8}\right) \times (30-20)$$

$$= 20 + \left(\frac{13}{21}\right) \times 10$$

$$= 20 + (0.6190) \times 10$$

$$= 20 + 6.190$$

$$= 26.190$$

$$\therefore \boxed{\text{Mode} = 26.190}$$

$\therefore$ According to Karl Pearson's Coefficient of Skewness we get

$$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{S.D}$$

$$= \frac{26.4 - 26.190}{11.137}$$

$$= \frac{0.21}{11.137}$$

$$= \boxed{0.01886}$$

PAGE No
DATE / /

(2) Find the C.V of a frequency distribution, given that its Mean is 120, Mode is 123 & K.P coeff of Skewness is -0.3

⇒ ~~Solution~~,

Given,

Mean = 120

Mode = 123

K.P = -0.3

To find: CV = ?

Solution.

$$K.P = \frac{Mean - Mode}{S.D}$$

$$-0.3 = \frac{120 - 123}{S.D}$$

$$\boxed{KP = \frac{3(Mean - Median)}{S.D}}$$

$$SD = \frac{\cancel{-}3}{\cancel{+}0.3} = 10$$

∴ We know that,

$$C.V = \frac{S.D}{Mean} = \frac{1\cancel{0}}{12\cancel{0}} = \boxed{0.0833}$$

NOTE :- If mode is not known, but Mean & Median are known ~~to~~ then use ~~to~~ the formula

$$Mean - Mode = 3(Mean - Median)$$

PAGE No.
DATE / /

(3) Given, Mean = 20, CV = 20%, Median = 17,
Find K·P coefficient of Skewness

→ Given,
Mean = 20, C·V = 20% = 0·2, Median = 17

To find
K·P's coefficient of skewness.

Solution,
∴ We know that,

$$C.V = \frac{S.D}{Mean}$$

$$\therefore 0.2 = \frac{S.D}{20} \qquad \therefore \boxed{SD = 20 \times 0.2 = 4}$$

∴ According to KP's coefficient of Skewness,

$$\therefore KP = \frac{3(Mean - Median)}{SD}$$

$$= \frac{3(20-17)}{4}$$

$$= \frac{3 \times 3}{4} = \frac{9}{4} = \boxed{2.25}$$

PAGE No
DATE / /

(3) Find Bowley's coefficient of Skewness

C.I :	30-35	35-40	40-45	45-50	~~55-60~~ 50-55	55-60
f :	5	10	30	35	15	5

⇒ Solution

~~C.I~~	C.I	(f) Frequency	C.F
~~30-35~~	30-35	5	5
	35-40	10	15
	40-45	30	45
	45-50	35	80
	50-55	15	95
	55-60	5	100
		$\Sigma f_i = 100$	

$N = 100$.

$\therefore \frac{N}{2} = \frac{100}{2} = 50$; $\frac{N}{4} = \frac{100}{4} = 25$; $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$

To find Q_1

$\frac{N}{4} = 25$

The class having $(N/4)^{th}$ value present is ~~30-35~~ 40-45

$\therefore$ Q_1 class $(l_1 - l_2) = 40-45$

Frequency of Q_1 class = 30

C.F of pre-Q_1 class = 15

$\therefore Q_1 = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{4} - F\right)$

$= 40 + \left(\frac{45-40}{30}\right)(25-15)$

$= 40 + \left(\frac{5}{\cancel{30}\,3}\right) \times \cancel{10}\,1$

$= 40 + 1.6667 = \boxed{41.6667}$ $\boxed{Q_1 = 41.6667}$

To find Q_2/Median

$\frac{N}{2} = 50$

The class having $(N/2)^{th}$ value present is 45-50

$\therefore Q_2$ class $(l_1 - l_2) = 45 - 50$

Frequency of Q_2 class = 35

C.F of pre-Q_2 class = 45

$\therefore Q_2 = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{2} - F\right)$

$= 45 + \left(\frac{50-45}{35}\right)(50-45)$

$= 45 + \left(\frac{5}{\cancel{35}\,7}\right) \times \cancel{5}\,1$

$= 45 + 0.7142 = \boxed{45.7142}$ $\boxed{Q_2 = 45.7142}$

PAGE No
DATE / /

To find Q_3

$$\frac{3N}{4} = 75$$

$\therefore$ The class having the $(3N/4)^{th}$ value present is 45-50

$\therefore$ Q_3 class $(l_1 - l_2) = 45-50$
Frequency of Q_3 class $= 35$
C.F of pre-Q_3 class $= 45$

$$\therefore Q_3 = l_1 + \left(\frac{l_2 - l_1}{f}\right) \cdot \left(\frac{3N}{4} - F\right)$$

$$= 45 + \left(\frac{50-45}{35}\right) \cdot (75-45)$$

$$= 45 + \left(\frac{5}{\cancel{35}_{7}}\right) \cdot \cancel{30}^{6}$$

$$= 45 + 4.2857 = \boxed{49.2857}$$

$\therefore$ $\boxed{Q_3 = 49.2857}$

According to Bowley's Coefficient of Skewness

$$\frac{-2Q_2 + Q_1 + Q_3}{Q_3 - Q_1}$$

$$\therefore \frac{-2(45.7142) + 41.6667 + 49.2857}{49.2857 - 41.6667}$$

$$\therefore \frac{-91.4284 + 90.9524}{7.619} = \boxed{-0.06247}$$

01/06/2023

CORRELATION & REGRESSION

x (books)	y (price)
1	5
5	25
3	15

+5, −2; +20, −10

(+ve correlation)

x (No of per day chocolates)	y (days)
1	50
50	1
5	10

(−ve correlation)

x (cement bags)	y (sugar)
50	2
0	2
2	13
100	10

(No correlation)

* Correlation

The change in one variable w.r.t the change in other variable is called as correlation

Types of Correlation

Correlation is classified as:

(a) Positive correlation
(b) Negative correlation
(c) No correlation

} Broad classification

PAGE No.
DATE / /

Positive correlation ~~change~~ $\Rightarrow$ Change of the same type.
Negative correlation ~~change~~ $\Rightarrow$ Change of the opposite type
No ~~change~~ correlation $\Rightarrow$ can't say

NOTE :- The correlation in two variables can be studied using K.P coefficient of Correlation (r)

Formula :

Given n pairs,

x	x_1	x_2	......	x_n
y	y_1	y_2	......	y_n

$$\therefore \boxed{r = \frac{Cov(x,y)}{\sigma_x \times \sigma_y} = \frac{E(XY) - E(X)\cdot E(Y)}{\sigma_x \cdot \sigma_y} = \frac{\frac{\sum x_i y_i}{n} - \left(\frac{\sum x_i}{n} \times \frac{\sum y_i}{n}\right)}{\sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} \cdot \sqrt{\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2}}}$$

($r = -1$) ($r = -0.5$) ($r = 0$) ($r = 0.5$) ($r = 1$)

(Perfect −ve correlation) | Strong −ve correlation | Weak −ve correlation | (No correlation) | Weak +ve correlation | Strong +ve correlation | (Perfect +ve correlation)

PAGE No.
DATE / /

Examples

Find K.P coefficient of correlation from the following data

City :	A	B	C	D	E	F	G
Robbery: (x)	4	6	10	5	1	2	3
Murder: (y)	16	29	43	20	3	4	6

⇒ Solution,

$n = 7$

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
4	16	16	256	64
6	29	36	841	174
10	43	100	1849	430
5	20	25	400	100
1	3	1	9	3
2	4	4	16	8
3	6	9	36	18

$\Sigma x_i = 31$ $\Sigma y_i = 121$ $\Sigma x_i^2 = 191$ $\Sigma y_i^2 = 3407$ $\Sigma x_i y_i = 797$

$$r = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

PAGE No.
DATE / /

$$= \frac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2} \cdot \sqrt{\frac{\Sigma y_i^2}{n} - \left(\frac{\Sigma y_i}{n}\right)^2}}$$

$$= \frac{\frac{797}{7} - \left(\frac{31}{7} \times \frac{121}{7}\right)}{\sqrt{\frac{191}{7} - \left(\frac{31}{7}\right)^2} \cdot \sqrt{\frac{3407}{7} - \left(\frac{121}{7}\right)^2}}$$

$$= \frac{37.3061}{2.7701 \cdot 13.7083} = \frac{37.3061}{37.97336} = \boxed{0.9824}$$

* Spearman's Rank Correlation (R)

Spearman's Rank correlation (R) is given by

$$\boxed{R = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}}$$

where $d_i = R_1 - R_2$ or $R_2 - R_1$

n = n.o of observations

Examples

(1) Find Spearman's Rank Correlation coefficient

(Example with different ranks)

Marks in DM:	64	80	44	42	56	65	59
Marks in C.O.A:	80	60	37	51	30	75	44

⇒ Solution,

$n = 7$

Marks in D.M(x_i)	C.O.A(y_i)	Rank in DM(R_1)	COA(R_2)	$d_i = R_1 - R_2$	d_i^2
64	80	2	1	1	1
50	60	5	3	2	4
44	87	6	6	0	0
42	51	7	4	3	9
56	80	4	7	-3	9
65	75	1	2	-1	1
59	44	3	5	-2	4

$\Sigma d_i^2 = 28$

According to spearman's rank correlation,

$$R = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(28)}{7(7^2-1)}$$

$$= 1 - \frac{168}{7 \times 48}$$

$$= 1 - \frac{168}{336} = 1 - 0.5 = 0.5$$

$\therefore \boxed{R = 0.5}$

PAGE No.
DATE / /

① Ranks are given
② Ranks are not given
③ Ranks are repetitive
④ ~~Ranks are~~ More than 2 observations

} Types of questions

* **Type I** (Ranks are given)

Given,

Rank in x	1	4	5	6	7	3	2
Rank in y	2	1	3	4	5	7	6

⇒ Solution

x (R_1)	y (R_2)	$d_i = R_1 - R_2$	d_i^2
1	2	−1	1
4	1	3	9
5	3	2	4
6	4	2	4
7	5	2	4
3	7	−4	16
2	6	−4	16
			$\Sigma d_i^2 = 54$

∴ Spearman's Rank coefficient

$$= 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 54}{7(7^2-1)} = \boxed{0.0357}$$

~~Type II~~

* Type II (Ranks are not given)

Example

Previously done

* Type III (Ranks ~~may or may not be~~ are given but repeated)

(i) ~~Given~~ If ranks are repeated then find the average rank of the ~~total~~ n·o of occurence ~~of the~~ for that data set

$$\boxed{\text{Avg Rank} = \frac{\text{Sum of all the ranks of repeated data}}{\text{N·o of occurences of the data}}}$$

(ii) ~~find the~~ If there is repetition, then find correction factor for each repetition of x & y dataset

If rank is repeated "m" times then correction factor is given by

$$\boxed{\text{Correction Factor} = \frac{m(m^2-1)}{12}}$$

(iii) Find the corrected Σd_i^2

$$\therefore \underset{(\text{corrected})}{\Sigma d_i^2} = \Sigma d_i^2 + \Sigma \text{ All the corrected factors}$$

NEW FORMULA $= 1 - \dfrac{6\,\Sigma d_i^2 (\text{corrected})}{n(n^2-1)}$

Example

In a sample of 12 fathers and their eldest sons gave following data about their heights (in inches)

Father (X)	65	63	67	64	68	62	70	66	68	67	69	71
Eldest son (Y)	68	66	68	65	69	66	68	65	71	67	68	70

Calculate Spearman's Rank correlation between X & Y

⇒ **Solution**

$n = 12$

X (R_1)	Y (R_2)	$R_1(x)$	$R_2(Y)$	$d_i = R_1 - R_2$	d_i^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	~~9~~ 10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1

$$\boxed{\Sigma d_i^2 = 72.5}$$

PAGE No.
DATE / /

(1) Rank 4.5 is repeated $m=2$ times

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = 0.5$$

(2) Rank 6.5 is repeated $m=2$ times

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = 0.5$$

(3) Rank ~~4.5~~ 5.5 is repeated $m=4$ times

$$C.F = \frac{m(m^2-1)}{12} = \frac{4(4^2-1)}{12} = \cancel{6.5}\ 5$$

(4) Rank 9.5 is repeated $m=2$ times

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = 0.5$$

(8) Rank 11.5 is repeated $m=2$ times

$$C.F = \frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = 0.5$$

$\therefore$ Corrected $\Sigma d_i^2 = \Sigma d_i^2 + \Sigma$ All the correction factors

$$= 72.5 + 0.5 + 0.5 + 5 + 0.5 + 0.5$$
$$= \underline{\underline{79.5}}$$

PAGE No	
DATE	/ /

∴ Spearman's Rank correlation coefficient

$$= 1 - \frac{6\sum d_i^2 (\text{corrected})}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 79.5}{12(12^2-1)}$$

$$= 1 - 0.2779$$

$$= \boxed{0.7221}$$

Type IV (More than 2 observations (MAX LIMIT = 3))

~~Example~~

Given, There are 3 Judges say x, y, z. They give ranks to 'n' participants in competition. We want to find which pair of Judges has common ~~opi~~ opinion about the rest.

$X(R_1)$	$Y(R_2)$	$Z(R_3)$	$d_{12}(R_1-R_2)$	$d_{23}(R_1-R_3)$	$d_{23}(R_2-R_3)$

	d_{12}^2	d_{13}^2	d_{23}^2
Σ			

~~Find~~

Find = $R_{xy} = 1 - \frac{6\sum d_{12}^2}{n(n^2-1)}$; $R_{yz} = 1 - \frac{6\sum d_{23}^2}{n(n^2-1)}$; $R_{xz} = 1 - \frac{6\sum d_{13}^2}{n(n^2-1)}$

Conclusion ⇒ Highest value of $\{R_{12}, R_{13}, R_{23}\}$

Examples

Ten competitors in a beauty contest are ranked by 3 Judges as follows

Judges	Competitors									
	1	2	3	4	5	6	7	8	9	10
A	6	5	3	10	2	4	9	7	8	1
B	5	8	4	7	10	2	1	6	9	3
C	4	9	8	1	2	3	10	5	7	6

Which pair of Judges has the nearest approach to common tests of beauty contest?

⇒ **Solution**

$n = 10$

$X(R_1)$	$Y(R_2)$	$Z(R_3)$	d_{12} (R_1-R_2)	d_{13} (R_1-R_3)	d_{23} (R_2-R_3)	d_{12}^2	d_{13}^2	d_{23}^2
6	5	4	1	2	1	1	4	1
5	8	9	-3	-4	-1	9	16	1
3	4	8	-1	-5	-4	1	25	16
10	7	1	3	9	6	9	81	36
2	10	2	-8	0	8	64	0	64
4	2	3	2	1	-1	4	1	1
9	1	10	8	-1	-9	64	1	81
7	6	5	1	2	1	1	4	1
8	9	7	-1	1	2	1	1	4
1	3	6	-2	-5	-3	4	25	9
					Σ	158	158	214

PAGE No.
DATE / /

$$R_{xy} = 1 - \frac{6\Sigma d_{12}^2}{n(n^2-1)} = 1 - \frac{6\times158}{10(10^2-1)}$$

$$= 1 - 0.9575 = \boxed{0.0425}$$

$$R_{xz} = 1 - \frac{6\Sigma d_{13}^2}{n(n^2-1)} = 1 - \frac{6\times188}{10(10^2-1)} = \boxed{0.0425}$$

$$R_{yz} = 1 - \frac{6\Sigma d_{23}^2}{n(n^2-1)} = 1 - \frac{6\times214}{10(10^2-1)} = 1 - 1.2969$$

$$= \boxed{-0.2969}$$

$\therefore \boxed{R_{xy} = 0.0425}$ $\boxed{R_{xz} = 0.0425}$ $\boxed{R_{yz} = -0.2969}$

i.e. From the above values we conclude that Pairs (A, B) & (A, C) have a common opinion

* REGRESSION

Given 'n' pairs

x_i	x_1	x_2	x_3	...	x_n
y_i	y_1	y_2	y_3	...	y_n

(i) Regression of y on x

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

(i) Regression of y on x

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

Where $\bar{x} = \frac{\Sigma x_i}{n}$; $\bar{y} = \frac{\Sigma y_i}{n}$;

NOTE :-
byx & bxy are the regression coefficient

$$b_{yx} = \frac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

(ii) Regression of x on y

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

Where,

$$b_{xy} = \frac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\frac{\Sigma y_i^2}{n} - \left(\frac{\Sigma y_i}{n}\right)^2}$$

NOTE :- $b_{xy} \times b_{yx} = r^2$

Where r is k.P coefficient of ~~coor~~ correlation

PAGE No.

DATE / /

Eg ① Find regression of y on x.

Given,

x_i	1	3	7	8	10
y_i	-1	3	11	13	17

Estimate y for $x=4$

⇒ Solution,

$n=5$

	x_i	y_i	x_i^2	$x_i y_i$
	1	-1	1	-1
	3	3	9	9
	7	11	49	77
	8	13	64	104
	10	17	100	170
Σ	$x_i = 29$	$y_i = 43$ ~~$y_i=43$~~	$x_i^2 = 223$	$x_i y_i = 359$

$\therefore \bar{x} = \frac{\Sigma x_i}{n} = \frac{29}{5} = 5.8$

$\bar{y} = \frac{\Sigma y_i}{n} = \frac{43}{5} = 8.6$

$\cancel{b_{xy}} = b_{yx} = \dfrac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$

PAGE No.

DATE / /

$$= \frac{\frac{359}{5} - \{(5.8 \times 8.6)}{\frac{223}{5} - (5.8)^2}$$

$$= \frac{71.8 - 49.88}{44.6 - 33.64}$$

$$= \frac{21.92}{10.96} = 2$$

$\therefore$ $\boxed{b_{yx} = 2}$

$\therefore$ We know that,

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

Substitute values we get,

$$\boxed{y - 8.6 = 2(x - 5.8)}$$

$\therefore$ General Formula for above question is

$$\boxed{y - 8.6 = 2(x - 5.8)}$$

To estimate y for $x = 4$, Substitute $x = 4$ we get

$$y - 8.6 = 2(4 - 5.8)$$

$$y = -3.6 + 8.6$$

$$\therefore \boxed{y = 5}$$

Eg ② Given,

Year: (x)	1991	1993	1997	1998	2000
(y) Rainfall:	13	33	73	81	103

Find Regression of ~~Rain.~~ Rainfall (y) on year (x) and estimate the rainfall in year 2025

⇒ Solution:- $n = 5$

x_i $(x_i = x_i - 1990)$	y_i	x_i^2	$x_i y_i$
~~1991~~ 1	13	1	13
~~1993~~ 3	33	9	99
~~1997~~ 7	73	49	511
~~1998~~ 8	81	64	648
~~2000~~ 10	103	100	1030

Σ $x_i = 29$ $\quad y_i = 303$ $\quad x_i^2 = 223$ $\quad x_i y_i = \cancel{26}2301$

$$\therefore \bar{x} = \frac{\Sigma x_i}{n} = \frac{29}{5} = 5.8$$

$$\therefore \bar{y} = \frac{\Sigma y_i}{n} = \frac{303}{5} = 60.6$$

$$\therefore b_{yx} = \frac{\frac{\Sigma x_i y_i}{n} - \left(\frac{\Sigma x_i}{n} \times \frac{\Sigma y_i}{n}\right)}{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

PAGE No. DATE / /

$$= \frac{\left(\frac{2301}{5}\right) - (5.8 \times 60.6)}{\frac{223}{5} - (5.8)^2}$$

$$= \frac{460.2 - 351.48}{44.6 - 33.64}$$

$$= \frac{108.72}{10.96} = \cancel{9.92}\ 9.9197$$

$$\therefore \boxed{b_{yx} = 9.9197}$$

We know that,

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

$$\boxed{y - 60.6 = 9.9197(x - 5.8)}$$

Estimation of rainfall for year 2025

$$\therefore x = x - 1990$$

$$\therefore x = 2025 - 1990 = 35$$

$$\therefore y - 60.6 = 9.9197(35 - 5.8)$$

$$\therefore y - 60.6 = 289.6552$$

$$y = 289.6552 + 60.6$$

$$\boxed{y = 350.2552}$$

PAGE No.
DATE / /

Type I :-

Type II :- Given Regression Lines (Equations)

To Find

(i) Mean values of $\bar{x}$ & $\bar{y}$

[Solve the two linear equations & find values of x & y. These are the values for $\bar{x}$ & $\bar{y}$]

Type III :- Relation between Regression coefficient, Correlation coefficient and standard deviation. (σ_x, σ_y)

~~$b_{yx} = r.$~~ $\boxed{b_{xy} = r.\frac{\sigma_x}{\sigma_y} \quad \& \quad b_{yx} = r.\frac{\sigma_y}{\sigma_x}}$

Eg (1) The regression line of y on x for a certain bivariate data is $5y+3x=52$ and the line of regression of x on y is $2x+y=30$.

Find: (i) Arithmetic mean of x & y

(ii) The coefficient of correlation between x & y

(iii) The most probable value of y when $x=12$

⇒ Solution

Given two regression lines

'y on x' ⇒ $5y+3x=52$ i.e $3x+5y=52$ — (1)

PAGE No.
DATE / /

'x' on 'y' $\Rightarrow$ $2x+y=30$ — (2)

(i) To find, $\bar{x}$ & $\bar{y}$

Multiply eq (2) by 5

$\therefore$ $10x+5y=150$ — (3)

Subtract eq (3) from (1)

$$\begin{array}{rcl} 10x + \cancel{5y} & = & 150 \\ -\ 3x + \cancel{5y} & = & 52 \\ (-)\quad (-) & & (-) \\ \hline 7x & = & 98 \end{array}$$

$\therefore x = \frac{98}{7} = 14$ $\therefore$ $\boxed{x=14}$

&

Substitute $x=14$ in eq (2)

$\therefore 2(14)+y=30$

$\therefore y=30-28$ $\therefore$ $\boxed{y=2}$

$\therefore$ $\boxed{\bar{x}=14 \ \& \ \bar{y}=2}$

(ii) To find coefficient of correlation between x & y i.e (r)

We know regression of y on x is

$5y+3x=52$ i.e $5y=52-3x$

$\therefore y = \frac{52-3x}{5}$

PAGE No.

DATE / /

$$\therefore y = \left(\frac{-3}{5}\right)x + \frac{52}{5} \quad — (1)$$

~~Comparin~~ $\boxed{b_{yx} = \frac{-3}{5}}$

We know regression of x on y is

$2x + y = 30 \quad \therefore 2x = 30 - y$

$\therefore x = \frac{30 - y}{2}$

$$\therefore x = \left(\frac{-1}{2}\right)y + 15 \quad — (2)$$

$\therefore \boxed{b_{xy} = \frac{-1}{2}}$

$\therefore$ ~~bxy~~ We know that,

$b_{xy} \times b_{yx} = r^2$

$\therefore r^2 = \frac{-3}{5} \times \frac{-1}{2}$

$r^2 = 0.3$

$\therefore r = \sqrt{0.3} \quad \therefore \boxed{r = 0.5477}$

PAGE No.
DATE / /

Since b_{xy} & b_{yx} are negative

$\therefore \boxed{r = -0.5477}$

(iii) The most probable value of y when $x = 10$

~~Subs~~ Use Regression of y on x

Substitute $x = 10$ in eq (1)

$\therefore \; y = \left(\frac{-3}{\cancel{5}_1}\right) \times \cancel{10}^2 + \frac{52}{5}$

$\therefore \; y = -6 + 10.4$

$\boxed{y = 4.4}$

Eg (2) If the two regression lines of the regression are $4x - 5y + 30 = 0$ and $20x - 9y - 106 = 0$.

Find: (i) Mean values of x & y

(ii) The regression coefficient of x ~~&~~ on y

(iii) The coefficient of correlation between x & y

(iv) The most ~~possible~~ probable value of y when $x = 30$.

$\Rightarrow$ Solution

(i) Consider regression of x on y as

$4x - 5y + 30 = 0$

PAGE No.

DATE / /

$4x = 5y - 30$

$\therefore x = \left(\frac{5}{4}\right)y - \frac{30}{4}$ —— (1)

$\therefore \boxed{bxy = \frac{5}{4}}$

Consider regression of y on x as

$20x - 9y - 106 = 0$.

$\therefore -9y = \cancel{2}\,106 - 20x$

$\therefore y = \frac{106 - 20x}{-9}$

$\therefore y = \left(\frac{20}{9}\right)x - \frac{106}{9}$

$\therefore \boxed{byx = \frac{20}{9}}$

We know that,

$bxy \times b\cancel{x}yx = r^2$

$\therefore r^2 = \left(\frac{5}{4}\right) \times \frac{20}{9}$

$\boxed{r^2 = 2.7777}$

But $-1 \leq r \leq 1$ OR $0 \leq r^2 \leq 1$

Hence,

Regression of y on x is $4x - 5y + 30 = 0$

$$5y = 4x + 30$$

$$\therefore y = \left(\frac{4}{5}\right)x + 6$$

$$\therefore \boxed{b_{yx} = \frac{4}{5}}$$

Regression of x on y is $20x - 9y - 106 = 0$

$$\therefore 20x = 9y + 106$$

$$x = \left(\frac{9}{20}\right)y + \frac{106}{20}$$

$$\therefore \boxed{b_{xy} = \frac{9}{20}}$$

$\therefore$ We know that

Coefficient of correlation (r^2) = $b_{xy} \times b_{yx}$

$$\therefore r^2 = \frac{9}{20} \times \frac{4}{5}$$

PAGE No.
DATE / /

$\therefore r^2 = 0.36$

$\boxed{r = 0.6}$

Since bxy & byx ~~and~~ are positive

$\therefore \boxed{r = 0.6}$

Eg 3 Two regression lines are given:
$x+2y-5=0$ & $2x+3y-8=0$
and also given $\sigma_x^2 = 12$. Find the values of
$\bar{x}$, $\bar{y}$ and σ_y^2

⇒ Solution,
$x+2y-5=0$ — (1) $2x+3y-8=0$ — (2)
(i) To find $\bar{x}$ & $\bar{y}$

$\therefore$ Multiply eq (1) by 2

$\therefore 2x+4y-10=0$ — (3)

Subtract eq (3) from (2)

$$\begin{array}{l} \cancel{2x}+4y-10=0 \\ \cancel{2x}+3y-8=0 \\ (-)\quad(-)\quad(+) \\ \hline y-2=0 \end{array}$$

$\therefore \boxed{y=2}$

PAGE No.
DATE / /

$\therefore$ Substitute $y=2$ in eq ①

$$x + 2(2) - 5 = 0$$
$$x - 1 = 0$$
$$\therefore \boxed{x = 1}$$

$$\therefore \boxed{\bar{x} = 1 \ \& \ \bar{y} = 2}$$

(ii)

Consider regression on y on x

$$x + 2y - 5 = 0$$
$$\therefore 2y = 5 - x$$

$$\therefore y = \left(\frac{-1}{2}\right)x + \frac{5}{2}$$

$$\therefore \boxed{b_{yx} = \frac{-1}{2}}$$

Consider regression on x on y

$$2x + 3y - 8 = 0$$
$$2x = 8 - 3y$$

$$\therefore x = \left(\frac{-3}{2}\right)y + 4$$

$$\therefore \boxed{b_{xy} = \frac{-3}{2}}$$

PAGE No.
DATE / /

We know that,

$$r^2 = b_{xy} \times b_{yx}$$

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

$$r^2 = 0.75$$

$$\therefore r = \sqrt{0.75} = 0.8660$$

$$\therefore \boxed{r = 0.8660}$$

Since b_{xy} & b_{yx} are negative

$$\therefore \boxed{r = -0.8660}$$

We know that,

$$b_{yx} = \cancel{r} = r \cdot \frac{\sigma_y}{\sigma_x}$$

Do not consider this

$$\frac{-1}{2} = (-0.8660) \cdot \frac{\sigma_y}{\sqrt{\sigma_x^2}}$$

$$\frac{\cancel{-}1}{2} = (\cancel{-}0.8660) \cdot \frac{\sigma_y}{\sqrt{12}}$$

$$\therefore \frac{1}{2} = 0.8660 \cdot \frac{\sigma_y}{3.4641}$$

PAGE No.
DATE / /

$$\therefore \sigma_y = \frac{3.4641}{2 \times 0.8666}$$

$$\sigma_y = 2.0000577 \approx 2.0001$$

$$\therefore \sigma_y^2 = (2.0001)^2$$

$$= 4.0$$

We know that

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$$

$$b_{yx}^2 = r^2 \cdot \frac{\sigma_y^2}{\sigma_x^2}$$

$$\left(\frac{-1}{2}\right)^2 = \frac{3}{4} \times \frac{\sigma_y^2}{12}$$

$$\frac{1}{4} = \frac{3}{4} \times \frac{\sigma_y^2}{12}$$

$$\therefore \sigma_y^2 = \frac{1}{\cancel{4}} \times \frac{\cancel{4}}{\cancel{3}} \times \cancel{12}\,4$$

$$\therefore \boxed{\sigma_y^2 = 4}$$

PAGE No.
DATE / /

Eg (4) Find two regression lines.

Given,

Variable	Mean	S.D
x	1	$2\sqrt{3}$
y	2	2

and $r = \sqrt{3}/2$

⇒ Solution,

Given,

$\bar{x} = 1,\ \bar{y} = 2,\ \sigma_x = 2\sqrt{3},\ \sigma_y = 2,\ r = \sqrt{3}/2$

$\therefore\ b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = \frac{\sqrt{3}}{2} \cdot \frac{2\sqrt{3}}{2} = \frac{3}{2}$

$\therefore\ b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = \frac{\sqrt{3}}{2} \cdot \frac{2}{2\sqrt{3}} = \frac{1}{2}$

∴ We know that,

Regression of y on x

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

$y - 2 = \frac{1}{2}(x - 1)$

$\therefore\ 2y - 4 = x - 1$

$\therefore\ \boxed{x - 2y = -3}$

PAGE No.
DATE / /

Regression on x on y

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

$$x - 1 = \frac{3}{2}(y - 2)$$

$$2x - 2 = 3y - 6$$

$$\therefore \boxed{2x - 3y = -4}$$

* TESTING OF HYPOTHESIS

Assuming the statement to be true and perform some tests on the statement to proof that whether our assumptions were ~~corrected~~ correct or not.

Data members

Null Hypothesis (H_0): Population mean is $\bar{x} = \mu$

Alternate Hypothesis (H_1/ H_a): Population mean is $\neq \mu$.

* Types of ~~Students~~ Tests

(1) Small Sample tests: (Student's Test / t-test)

Algorithm

* Step 1 :- Given a sample of size n, say $x_1, x_2, \ldots, x_n$ ~~where~~ where n is small i.e $n < 30$.

PAGE No.
DATE / /

Find Sample mean $(\bar{x}) = \frac{\Sigma x_i}{n}$

Find ~~Sample~~ Standard deviation of sample $(S) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

* **Step 2 :-** Find the test statistics "t" using formula.

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$$

Where $\bar{x}$ = mean
S = Standard deviation
μ = Assumed ~~mea~~ mean
n = ~~No~~ Degree of freedom / No of observation

* **Step 3 :-** Decide the level of significance (LOS) i.e α
$\Downarrow$
(~~How~~ How much error is allowed)

Find the critical value t_α for LOS α and degree of ~~fr~~ freedom $n-1$ from the statistical table.
(NOTE :- This value will be given in the question itself)

* **Step 4 :-** Compare calculated test statistics "t" with the critical value "t_α"

PAGE No.
DATE / /

If $|t| \leq t_\alpha$, then null hypothesis is accepted ~~either~~ otherwise rejected

Eg:- The lengths in cm of 10 nails produced by a certain machine are 5·10, 4·98, 5·03, 4·99, 5·00, ~~5·9~~ 5·07, 5·04, 5·03, 4·91, 4·97. Can it be concluded that average length of the nail is ~~b~~ produced by machine is 5cm.
[Given : t_α at 5% LOS for 9 degree of freedom is 1·8333]

$\Rightarrow$ Solution,
Null Hypothesis (H_0): Population mean is (μ) = 5cm
Alternate Hypothesis (H_a): Population mean is (μ) $\neq$ 5cm

$n = 10$

	x	x^2
	5·10	26·01
	4·98	24·8004
	5·03	25·3009
	4·99	24·9001
	5·00	25
	5·07	25·7049
	5·04	25·4016
	5·03	25·3009
	4·91	24·1081
	4·97	24·7009
$\Sigma =$	50·12	251·2278

PAGE No
DATE / /

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50.12}{10} = 5.012 \quad \therefore \boxed{\bar{x} = 5.012}$$

$$S = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{251.2278}{10} - \left(\frac{50.12}{10}\right)^2}$$

$$= \sqrt{25.12278 - 25.12014}$$

$$= \sqrt{2.64 \times 10^{-3}} = \sqrt{0.00264}$$

$$\boxed{S = 0.0513}$$

We know that,

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$$

$$= \frac{5.012 - 5}{0.0513/\sqrt{9-1}}$$

$$= \frac{0.012}{0.0513/\sqrt{8}} =$$

$$t = \frac{0.012}{0.0181} = 0.6629 \quad \therefore \boxed{t = 0.6629}$$

Since, $|t| \leq t_\alpha$ i.e $0.6629 \leq 1.8333$

$\therefore$ We will accept the null hypothesis i.e The ~~aveg~~ average length of nail produced ~~it~~ by the machine is 5 cm.

PAGE No.
DATE / /

② Large Sample tests (Z-test)

Algorithm

Null Hypothesis & Alternate Hypothesis same as ~~ex~~ ~~a~~ earlier.

* Step 1 :- Given a sample of size 'n' say, $x_1, x_2, x_3, \ldots\ldots, x_n$ (where $n \geq 30$)

Find Mean ($\bar{x}$) and Standard deviation (s)

* Step 2 :- Find test statistics "Z" using formula

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

* Step 3 :- ~~B~~. Decide the level of significance

Find the critical value Z_α for ~~te~~ LOS.

* Step 4 :- Compare calculated ~~Z~~. test statistics "Z" with the critical value "Z_α"

If $|Z| \leq Z_\alpha$, then null hypothesis is accepted or rejected

PAGE No.
DATE / /

Eg: The specified diameter of cylindrical part of a machine is 3cm. A sample of 900 such parts shows ~~as~~ an average diameter of 2.99cm with standard deviation of 0.01cm. Does the product differ the specification?
[Given at 1% LOS $Z_\alpha = 2.58$]

⇒ Solution,
Null Hypothesis (H_0) = Population mean is (μ) = 3cm
Alternate Hypothesis (H_a) = Population mean is (μ) $\neq$ 3cm

$n = 900,\ \bar{x} = 2.99,\ s = 0.01,\ \mu = 3cm$

We know that,

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{2.99 - 3}{0.01/\sqrt{900}}$$

$$= \frac{-0.01}{0.01/\sqrt{900}}$$

$$= \frac{-\cancel{0.01}}{\cancel{0.01}} \times 30$$

$$\boxed{Z = -30}$$

Since $|Z| \not< Z_\alpha$ i.e $|-30| \not< 2.58$ i.e $30 \not< 2.58$
∴ We will reject null hypothesis
i.e The product differs from the given specification

PAGE No.
DATE / /

(2) A random sample of 100 Students given mean weight of 58kg with a SD of 4kg. Test the hypothesis that mean weight in the population is 60 kg. Use 1% LOS [~~Use~~ Given at 1% LOS $Z_\alpha = 2.58$]

=) Solution,
Null Hypothesis = Population mean = (μ) = 60kg
Alternate Hypothesis = Population mean (μ) $\neq$ 60kg

$n = 100$, $\bar{x} = 58$, $s = 4$, $\mu = 60$

We know that

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{58 - 60}{4/\sqrt{100}}$$

$$= \frac{-2}{4/10} =$$

$$= \frac{-2}{\cancel{4}2} \times 10 = -5$$

$\therefore \boxed{Z = -5}$

Since $|Z| \not< Z_\alpha$ i.e $|-5| \not< 2.58$ i.e $5 \not< 2.58$

$\therefore$ We will reject the null hypothesis
i.e The mean weight of the population is not 60kg

PAGE No
DATE / /

③ Chi-square tests

Algorithm

Null Hypothesis : (H_0) : Given data is uniformly ~~distribut~~ distributed

Alternate Hypothesis (H_a) : Given data is not uniformly distributed

* **Step 1** : Given a sample size "N" to classify it into "n" class intervals. Find the observed frequencies "O_i".
  Find the expected frequencies "E_i" using the formula

$$E_i = \frac{N}{n}$$

* **Step 2** :- Find the test statistics χ_0^2 using the formula

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

* **Step 3** :- Decide the level of significance α

Find the critical value $\chi^2_{\alpha, n-1}$ for LOS α and degree of freedom $n-1$ from the statistical table

PAGE No.
DATE / /

Step 4 :- Compare the calculated test statistics χ^2 with critical value $\chi^2_{\alpha,n-1}$

If $\chi^2 \leq \chi^2_{\alpha,n-1}$ then null hypothesis is accepted otherwise rejected

Eg:- The following table gives the number of accidents in a city during 10 days of time. Find whether the accidents are uniformly distributed over that period.

Day	1	2	3	4	5	6	7	8	9	10
No. of accidents	8	8	10	9	12	8	10	14	10	11

(Given for 9 degree of freedom at 5% level of significance, ~~that~~ the table value of χ^2 is 16.9)

⇒ Solution

Null hypothesis (H_0): Accidents are uniformly ~~distribute~~ distributed

Alternate Hypothesis (H_a): Accidents are not uniformly distributed

~~N=100~~ Here, $N=100$, $n=10$

$\therefore$ Expected Frequency $(E_i) = \frac{N}{n} = \frac{100}{10} = 10.$

PAGE No.
DATE / /

Table

Day	O_i	E_i	$O-E_i$	$(O-E_i)^2$	$(O-E_i)^2/E_i$
1	8	10	-2	4	0.4
2	8	10	-2	4	0.4
3	10	10	0	0	0.0
4	9	10	-1	1	0.1
5	12	10	2	4	0.4
6	8	10	-2	4	0.4
7	10	10	0	0	0.0
8	14	10	4	16	1.6
9	10	10	0	0	0.0
10	11	10	1	1	0.1
Σ Total	100	100	0	34	3.4

We know that,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{34}{10} = 3.4$$

Since $\chi^2 \leq \chi^2_{\alpha, n-1}$ i.e $3.4 \leq 16.9$,

$\therefore$ We will accept the null hypothesis

$\therefore$ The accidents are uniformly distributed

PAGE No

DATE / /

Eg 2 The following table gives the n.o of participants for a certain program during 10 days D1 to D10. Find whether they are uniformly distributed?

Day	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
N.o of participants	17	20	24	13	25	25	16	12	10	11

(Given for 9 degree of freedom at 5% LOS, the critical value of $\chi^2 = 16.9$)

⇒ Solution

Null Hypothesis: The participants are uniformly distributed

Alternate Hypothesis: Participants are not uniformly distributed

~~Given~~

~~N = 100,~~

Given sample of size N=100. Classify it into n=10 days.

$\therefore$ Expected Frequencies $(E_i) = \frac{N}{n} = \frac{\cancel{100}\ 173}{10} = \cancel{10}\ 17.3$

PAGE No
DATE / /

Day	O_i	E_i	$O-E_i$	$(O-E_i)^2$	$(O-E_i)^2/E_i$
1	17	17.3	0.3	0.09	0.0052
2	20	17.3	2.7	7.29	0.4213
3	24	17.3	6.7	44.89	2.5947
4	13	17.3	-4.3	18.49	1.0687
5	25	17.3	7.7	59.29	3.4271
6	25	17.3	7.7	59.29	3.4271
7	16	17.3	-1.3	1.69	0.0976
8	12	17.3	-5.3	28.09	1.6236
9	10	17.3	-7.3	53.29	3.0803
10	11	17.3	-6.3	39.69	2.2942
				~~312.1~~	~~18.0404~~
Σ Total	173	173		312.1	18.0398

We know that

$$\chi^2 = \sum_{i=1}^{n} \frac{(O-E_i)^2}{E_i}$$

$$= \frac{312.1}{17.3} = \boxed{18.0404}$$

$$\boxed{\chi^2 = 18.0404}$$

Since $\chi^2 \nless \chi^2_{\alpha, n-1}$ i.e $18.0404 \nless 16.9$

$\therefore$ We will reject null hypothesis

i.e Participants are not uniformly distributed

PAGE No.
DATE / /

Problem

① Given $\Sigma x^2 = 650$, $\Sigma y^2 = 460$, $\Sigma x = 125$, $\Sigma y = 100$, $\Sigma xy = 508$

Wrong values	Correct value
(6, 14)	(8, 12)
(9, 6)	(6, 8)

Calculate the correct value of r.

⇒ Solution

Corrected values:

$\Sigma x^2 = 650 - 6^2 - 9^2 + 8^2 + 6^2 = 633$

$\Sigma y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$

$\Sigma x = 125 - 6 - 9 + 8 + 6 = 124$

$\Sigma y = 100 - 14 - 6 + 12 + 8 = 100$

$\Sigma xy = 508 - (6 \times 14) - (9 \times 6) + (8 \times 12) + (6 \times 8)$

$= 508 - 84 - 54 + 96 + 48$

$= 514$

$14 \times 6 = 84$

We know that,

$$r = \frac{\frac{\Sigma xy}{n} - \left(\frac{\Sigma x}{n} \times \frac{\Sigma y}{n}\right)}{\sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} \cdot \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}}$$

$$= \frac{\frac{514}{25} - \left(\frac{124}{25} \times \frac{100}{25}\right)}{\sqrt{\frac{633}{25} - \left(\frac{124}{25}\right)^2} \cdot \sqrt{\frac{436}{25} - \left(\frac{100}{25}\right)^2}}$$

PAGE No.
DATE / /

(2) Find missing frequency given mode = 136

x	120–125	125–130	130–135	135–140	140–145	145–150
f	7	10	18	?	12	7

⇒ Solution,
Let the missing frequency be f

We have,
Mode = 136
∴ Modal Class $(l_1 - l_2) = (135 - 140)$
Frequency of modal class $= (f) = f$
Frequency of pre-modal class $(f_1) = 18$
Frequency of post-modal class $(f_2) = 12$

~~We know~~ ∴ $d_1 = f - f_1 = f - 18$
$d_2 = f - f_2 = f - 12$

We know that,

$$\text{Mode} = l_1 + \left(\frac{d_1}{d_1 + d_2}\right) \cdot (l_2 - l_1)$$

$$136 = 135 + \left(\frac{f - 18}{f - 18 + f - 12}\right) \cdot (140 - 135)$$

$$136 = 135 + \left(\frac{f - 18}{2f - 30}\right) \cdot 5$$

$$\frac{f - 18}{2f - 30} = \frac{1}{5}$$

PAGE No.
DATE / /

$$5f - 90 = 2f - 30$$
$$3f = 60$$
$$\boxed{f = 20}$$

(3) In a class there are 100 students. The mean marks of the class is 72. The mean marks of boys is 75 and n.o of boys is 70. Find mean marks of girls

⇒

Total = 100

$n_1 = 70$	$n_2 = 30$
$\bar{x}_1 = 75$	$\bar{x}_2 = ?$

Combined mean $\bar{x} = 72$

Total $n = n_1 + n_2 = 100$

Formula

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$72 = \frac{70 \times 75 + 30 \times \bar{x}_2}{100}$$

$$7200 = 5250 + 30\bar{x}_2$$

$$\therefore 30\bar{x}_2 = 1950$$

$$\bar{x}_2 = \frac{1950}{30} = 65$$

$$\therefore \boxed{\bar{x}_2 = 65}$$

Q3 Find Median

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	5	8	7	12	28	20	10	10

⇒

Solution.

C.I	f	C.F
0-10	5	5
10-20	8	13
20-30	7	20
30-40	12	32
40-50	28	60
50-60	20	80
60-70	10	90
70-80	10	100

$N = \Sigma f_i = 100$

Median $= \left(\frac{N}{2}\right)^{th}$ value $= \left(\frac{100}{2}\right)^{th}$ value

$= 50^{th}$ value

50^{th} value lies in the class 40-50.

∴ Median class $(l_1 - l_2)$ = (40-50)

Frequency of median class = 28

C.F of pre-median class = 32

PAGE No
DATE / /

$$\text{Median} = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{N}{2} - F\right)$$

$$= 40 + \left(\frac{50-40}{28}\right)(50-32)$$

$$= 40 + \left(\frac{10}{28}\right) \times 18$$

$$= 40 + (0.3571) \times 18$$

$$= 40 + 6.4278$$

$$= \underline{\underline{46.4278}}$$

Q4 Find C.V of the following
101, 27, 0, 36, 82, 45, 07, 13, 65, 14

⇒ Solution,
$n = 10$

x_i	x_i^2
101	10201
27	729
0	0
36	1296
82	6724
45	2025
7	49
13	169
65	4225
14	196
$\Sigma x_i = 390$	$\Sigma x_i^2 = 25614$

$$\text{Mean} = \frac{\sum x_i}{n} = \frac{390}{10} = 39$$

$$S.D = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

$$= \sqrt{\frac{25614}{10} - \left(\frac{390}{10}\right)^2} = \sqrt{2561.4 - 1521} = 32.2552$$

We know that,

$$C.V = \frac{S.D}{\text{Mean}} = \frac{32.2552}{39} = \underline{\underline{0.8271}}$$

Q5 Find Q3

C.I	0-10	10-20	20-30	30-40	40-50	50-60
f	14	15	27	23	15	6

⇒ Solution

C.I	f	C.F
0-10	14	14
10-20	15	29
20-30	27	56
30-40	23	79
40-50	15	94
50-60	6	100

$N = \sum f_i = 100$

PAGE No.
DATE / /

Q_3 class $= \left(\frac{3N}{4}\right)^{th}$ value

$= 3 \times \frac{\cancel{100}^{25}}{\cancel{4}_{1}} = 75^{th}$ value

Q_3 class $(l_1 - l_2) = (30-40)$
Frequency of Q_3 class $= 23$
C.F of premodal Q_3 class $= 56$

We know that,

$$Q_3 = l_1 + \left(\frac{l_2 - l_1}{f}\right)\left(\frac{3N}{4} - F\right)$$

$$= 30 + \left(\frac{40-30}{23}\right)(75-56)$$

$$= 30 + \left(\frac{10}{23}\right) \times 19$$

$$= 30 + (0.43) \times 19$$
$$= 30 + 8.17$$
$$= \underline{\underline{38.17}}$$

PAGE No
DATE / /

* Possible Questions for ESE

- Missing Frequencies
- Mean, median, Mode
- Quartiles
- C.V (Coefficient of Variance)
- Regression lines → (i) Find lines of regression (using linear eqn method)
  → (ii) Given Regression Equation find b_{xy}, b_{yx}, r
- Spearman's Rank correlation coefficient.

PAGE No.
DATE / /

/06/2023

Q. The following table gives no. of aircraft accidents that have occured during various days of week.

Mon	Tue	Wed	Thurs	Fri	Sat
15	19	13	12	16	15

Test if the accidents are uniformly distributed over the week (Chi-square value at 5% LOS at 5 d.f is 11.07)

PAGE No. | DATE / /

Eg. An IQ test was conducted. The results are

Person	1	2	3	4	5
IQ before Training	110	120	123	132	125
IQ after Training	120	118	125	136	121

Check for the t-test at 1% LOS

⇒ Solution,

Null Hypothesis (H_0) : The average change in IQ = $\mu = 0$

Alternate Hypothesis (H_1): The average change in IQ = $\mu \neq 0$

Person	IQ Before	IQ After	d	d^2
1	110	120	-10	100
2	120	118	2	4
3	123	125	-2	4
4	132	136	-4	16
5	125	121	4	16
			-10	140

$n = 5$

$$\text{Mean} = \bar{d} = \frac{\Sigma d}{n} = \frac{-10}{5} = -2$$

$$S.D = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{140}{5} - \left(\frac{-10}{5}\right)^2} = \sqrt{28 - 4}$$

$$= \sqrt{24} = 4.899$$

Step 2 :- Find the test statistics.

$$t = \frac{\bar{d} - \mu}{S/\sqrt{n-1}} = \frac{-2 - 0}{4.899/\sqrt{5-1}}$$

$$= \frac{-2}{4.899/2}$$

$$= \frac{-4}{4.899} = -0.817$$

$$\therefore \boxed{t = -0.817}$$

Step 3 :-

Critical value at 1% LOS $= 4.604$

Since,

$|t| \leq t_{\alpha, n-1}$ i.e

i.e $|-0.817| \leq 4.604$

$0.817 \leq 4.604$

$\therefore$ Null Hypothesis is accepted

$\therefore$ There is no change in the IQ after the training

REVISION

PAGE No.
DATE / /

(I) Binomial Distribution

The probability of a man ~~gettir~~ hitting a target is 0.25. If he fires 7 times, what is the probability of his hitting the target at least twice? [How many times must ~~be th~~ he fire so that the probability of his hitting the target atleast once is greater than 2/3?] ⇒ Extra question

⇒ Solution,

Total n.o of trials $(n) = 7$

$p =$ probability of hitting a target $= 0.25$

$q = 1 - p = 1 - 0.25 = 0.75$

To find: $P(X \geq 2)$

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

$$= 1 - [{}^7C_0 (0.25)^0 \cdot (0.75)^7 + {}^7C_1 (0.25)^1 \cdot (0.75)^6]$$

$$= 1 - (0.75)^6 [1(0.75) + 7(0.25)]$$

$$= 1 - (0.75)^6 [0.75 + 1.75]$$

$$= 1 - (0.75)^6 \cdot (2.5)$$

$$= 1 - 0.4449$$

$$= \underline{\underline{0.5551}}$$

$\therefore$ ~~Pa~~ Probability of hitting the target at least twice is $\underline{\underline{0.5551}}$

PAGE No

DATE / /

~~② Probability of a bomb hitting~~

② The mean & variance of a binomial are 3 & 2 respectively. Find the probability that the variate takes the values
(i) less than or equal to 2
(ii) greater than or equal to 7

⇒ Solution,
We know that,

Mean = np Variance = npq

$\therefore \frac{npq}{np} = \frac{2}{3}$ $\therefore \boxed{q = \frac{2}{3}}$

$\therefore p = 1 - q = 1 - \frac{2}{3} = \frac{1}{3}$ $\therefore \boxed{p = \frac{1}{3}}$

We have,
Mean = np

$\therefore np = 3$
$\therefore n \times \frac{1}{3} = 3$ $\therefore \boxed{n = 9}$

To find: $P(X \leq 2)$

$\therefore P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

$= {}^9C_0 (1/3)^0 (2/3)^9 + {}^9C_1 (1/3)^1 (2/3)^8 + {}^9C_2 (1/3)^2 (2/3)^{7}$

$$= \left(\tfrac{2}{3}\right)^7 \cdot \left[{}^9C_0 \left(\tfrac{2}{3}\right)^2 + {}^9C_1 \left(\tfrac{1}{8}\right)\left(\tfrac{2}{3}\right) + {}^9C_2 \left(\tfrac{1}{3}\right)^2 \right]$$

$$= \left(\tfrac{2}{3}\right)^7 \cdot \left[1\left(\tfrac{2}{3}\right)^2 + 9(0.0833) + 36(0.1111) \right]$$

$$= \left(\tfrac{2}{3}\right)^7 \cdot \left[0.4444 + 0.7497 + 3.9996 \right]$$

$$= \cancel{(0.000} \left(\tfrac{2}{3}\right)^7 \cdot (5.1937)$$

$$= 0.0585 \times 5.1937 = \underline{\underline{0.3038}}$$

To find $P(x \geq 7)$