Probability and Statistics (BSIT-101)
Part-A
1.Introduction to Statistics: Meaning, scope, importance and limitations. Analysis of data: source of data, collection, classification, tabulation, depiction of data. Measures of Central tendency: Arithmetic, weighted, geometric mean, median and mode. Measures of Dispersion: Range, Quartile deviation, Mean deviation, Standard deviation Coefficient of variation, Skewness and Kurtosis. **[7 Hours]**
2.Sampling Distribution & Testing of Hypothesis: Sampling, Distribution of means and variance , Chi – Square distribution, t – distribution , F – distribution . General concepts of hypothesis, Testing a statistical Hypothesis, One and two tailed tests , critical region , Confidence interval estimation . Single and two sample tests on proportion , mean and variance . **[8 Hours]**
3.Correlation Analysis: Significance, types, Methods of correlation analysis: Scatter diagrams, Graphic method, Karl Pearson"s correlation co-efficient, Rank correlation coefficient, Properties of Correlation. Regression analysis: meaning, application of regression analysis, difference between correlation & regression analysis, regression equations, standard error and Regression coefficients. curve fitting. **[7 Hours]**
Part-B
4.Theory of Probability: Definition, basic concepts, events and experiments, random variables, expected value, types of probability, classical approach, relative frequency and subjective approach to probability, theorems of probability, addition, Multiplication and Bays Theorem and its application. **[6 Hours]**
5.Probability Distributions: Difference between frequency and probability distributions, Binomial, Poisson and normal distribution **[6 Hours]**
6.Optimization: Matrix calculus, gradient descent, coordinates descent, introduction to convex optimization. **[6 Hours]**

Measurement of Central tendency

• Measure of central tendency is a single value which represent the whole set of figure and all other individual terms concentrate on average

MCT

Mean | Median | Mode

Mean
- → Arithmetic mean
- → Geometric mean
- → Harmonic mean

(i) Arithmetic mean : AM($\bar{x}$) It is defined as sum of observation divided by total no of items

$$\boxed{\bar{x} = \frac{\Sigma X}{N}}$$

Q 8, 1, 6 = $\bar{x} = \frac{15}{3} = 5$

Q Calculate the weighted mean of from the following data.

Per table (X)	No of Tables sold (f)	fx
36	14	504
40	11	440
44	9	396
48	6	288
		$\Sigma fx = 1628$

$\boxed{\bar{x} = \Sigma fx / N}$

$\boxed{\Sigma f = N}$

$\bar{x} = \frac{1628}{40}$

$= 40.7$

Q the following gives the marks of two candidates find the weighted average mark. by what figure would be the second candidate had to increase his marks in sub 'B' all other marks remaining the same in order both candidates have the same place.

	weight (f)	X	Y	fx	fy
A	1	70	80	70	80
B	2	65	64	130	128
C	3	58	56	174	168
D	4	63	60	252	240

$\bar{x} = ?$

$\bar{y} = ?$

$\bar{x} = \frac{\Sigma fx}{\Sigma f = N}$, $\bar{y} = \frac{\Sigma fy}{\Sigma f = N}$

$\bar{x} = \frac{626}{10}$, $\bar{y} = \frac{616}{10}$

$\boxed{\bar{x} = 62.6}$ $\boxed{\bar{y} = 61.6}$

$2 \times (64 + k) = 128 + 2k = 626$

$2k = 626 - 128$

$2k = 598$

$k = \frac{598}{2}$

$616 + 2k = 626$

$2k = 10$

$\boxed{k = 5}$

Q. Shortcut Method when nothing is given

$$\boxed{\bar{X} = A + \left(\frac{\Sigma d}{N}\right)}$$ $$\boxed{d = X - A}$$

HW

where / A → Assumed mean

Σd → diviation of mid value from the assumed mean

N → no of items / products

Q. the pocket allowance of 10 students is given. Calculate the Arithmetic mean by taking 40 as assumed mean.

S.No	X	X − A = d
1	15 − A	−25
2	20 − A	−20
3	30 − A	−10
4	22	−18
5	25	−15
6	18	−22
7	40	0
8	50	10
9	55	15
10	65	25
F =		$\Sigma d = -60$

given $\boxed{A = 40}$ $N = 10$

$$\bar{X} = A + \left(\frac{\Sigma d}{N}\right)$$

$$\bar{X} = 40 + \left(\frac{-60}{10}\right)$$

$$\bar{X} = 40 - 6$$

$$\boxed{\bar{X} = 34}$$

$$\boxed{\bar{X} = 34}$$

* HW →

individual $\boxed{\bar{X} = \frac{\Sigma x}{N}}$, discrete $\bar{X} = \frac{\Sigma fx}{N}$ weighted mean / average

shortcut $\boxed{\bar{X} = A + \frac{\Sigma fd}{N}}$

Continuous $\bar{X} = \frac{\Sigma fm}{\Sigma f}$

HW → • Inclusive & Exclusive ~~frequency~~ series
• Cummulative frequency

* Inclusive series: An inclusive series includes the upper limit and the lower limit in the class interval.

for ex ⇒ A class Interval of 10 - 19 would include values from 10 to 19 (both 10 and 19 inclusive)

* Exclusive series: A exclusive series in which the upper limit is not included in that class and in included in upcoming class.

for example: 0 - 10
11 - 20
21 - 30

In this series, a value of 10 would be counted in the 11 - 20 class, not the 0 - 10 class. so 10 is exclusive value.

* Cummulative frequency: Cumulative frequency is a concept in Statics that refers to the running total of frequencies within a set of data, grouped by class intervals. It tells you how many data points fall at or below a specific value in your dataset.

Find the arithmetic mean for the following data.

direct Method

Marks	f	m	fm
0 – 10	20	5	100
10 – 20	24	15	360
20 – 30	40	25	1000
30 – 40	36	35	1260
40 – 50	20	45	900

$\Sigma = 3620$

$\bar{x} = \frac{\Sigma fm}{N}$ where $N = \Sigma f$

$\bar{x} = \frac{3620}{140}$

$\boxed{\bar{x} = 25.85}$

Short cut method

Marks	f	m	m − A = d	fd
0 – 10	20	5	−20	−40
10 – 20	24	15	−10	−240
20 – 30	40	25	0	0
30 – 40	36	35	+10	360
40 – 50	20	45	+20	400

Assumed mean $= \frac{125}{5}$ $\Sigma fd = 120$

$\boxed{A = 25}$

$\bar{x} = A + \frac{\Sigma fd}{N}$

where $d = M - A$

$d = 25$

$\bar{x} = 25 + \frac{120}{14}$

$\bar{x} = 25.85$

Step diviation Method (when Nothing is given)

Marks	f	m
0 – 10	20	5
10 – 20	24	15
20 – 30	40	25
30 – 40	36	35
40 – 50	20	45

10 − 20 = 10

$i = 10$

$\bar{x} = A + \frac{\Sigma fd'}{N} \times i$ where

$d' = \frac{d}{i}$

$d = m - A$

$\bar{x} = 25 + \frac{\frac{25}{10}}{14} \times 10$

$\bar{x} = 25 + \frac{25}{14}$ $\bar{x} = 26.78$

* Cumulative frequency

while cumulative frequency, it is necessary to convert the series into simple series.

marks	cf	f
Less than 10	5	5
20	17	12
30	31	14
40	41	10
50	49	8

$\sum f = 49$

$\sum f' = \frac{\sum f}{i} = \frac{49}{10}$

$\boxed{\sum f' = 4.9}$

$\bar{X} = A + \frac{\sum fd'}{N} \times i$

$\bar{X} = 25 + \frac{4.9}{} \times 10$

$\bar{X} =$

* Calculate the missing value when it's mean is given

wages	f	$\sum fx$
110	25	2750
112	17	1904
113	13	1469
117	15	1785
119	14	1400
125	8	1000
128	6	768
130	2	260

$\sum f =$

$\bar{X} = \frac{\sum fX}{\sum f}$

$\bar{X} = 115.86$

$a = 120$ Ans

$115.86 = \frac{9966 + 14a}{100}$

* Sum of deviation of certain no of items measured from 2.5 is 50 and from 3.5 is -50 find N and x, y.

ans ⇒ $\boxed{\bar{X} = A + \frac{\sum d}{N}}$ $A = 2.5$, $d = 50$

$\bar{X} = 2.5 + \frac{50}{N}$ —— ①

$\bar{X} = 3.5 - \frac{50}{N}$ —— ②

$\boxed{N = 100}$

$\bar{X} = 2.5 + \frac{50}{100}$

$\boxed{\bar{X} = 3}$

* correcting incorrect values of the mean

the mean of 100 is 80 by mistake 1 item is miss read 92 instead of 29 find the correct mean.

* Combined arithmetic mean will be X

$\bar{X}_{12} = \frac{N_1 \bar{X_1} + N_2 \bar{X_2}}{N_1 + N_2}$ $\bar{X}_{12} =$

① the mean heigh of 25 male worker 61 and the height of female worker is find the Combined mean of

2Q from home

* Mathematical

① Sum of deviation of the items from the mean is always 0.

$$\Sigma(X-\bar{X})=0$$

X	$\bar{X}$
5	-5
10	0
15	$+5$
$\bar{X}=10$	6

यहाँ से

* **Median:** Median is another measurement of central tendency

Median is defined as the middle of the series when arranged either in ascending or descending order.

* Individual series

$\boxed{\text{for odd} \quad M = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}}$ ∴ (odd)

Calculate median

22, 16, 18, 13, 15, 19, 12, 20, 23

ans → ~~12, 13, 16, 18, 19, 20,~~

12, 13, 15, 16, 18, 19, 20, 22, 23
1 2 3 4 5 6 7 8 9

$M = \left(\frac{9+1}{2}\right)$ $\boxed{M=5}$

$\boxed{\text{Median} = 18}$

* 5th

for even no of

1 200
2 217
3 264
4 282
5 296
6 299
7 316
8 317

4.5th item

$M = \frac{N+1}{2}$ $M = \frac{8+1}{2}$ $M = \frac{9}{2}$

$\boxed{M = 4.5}$ → $\frac{282 + 296}{2} = 289$ $\boxed{M = 289}$

* for even no of item average of $\frac{N}{2}$th, $\left(\frac{N}{2}+1\right)$th term

→ $\frac{8}{2} = 4$ $\left(\frac{8}{2}+1\right) \rightarrow 4+1 = 5$

* Level of knowledge of high = Median

x: 10 12 14 16 18 20 22 calculate the median from the following data.
f: 2 5 12 20 10 7 3

Discrete Series

$\left(\frac{N+1}{2}\right) = M$

X	f	Cf
10	2	2
12	5	7
14	12	19
16	20	39
18	10	49
20	7	56
22	3	59

$\boxed{N = \Sigma f = 59}$

$\boxed{N = 59 \text{ odd}}$

$M = \left(\frac{N+1}{2}\right)$th item

$M = \frac{60}{2}$ $\boxed{M = 30}$

$\boxed{M = 16}$

Continue series

	f	Cf
0 – 5	6	6
5 – 10	12	18
10 – 15	17	35
15 – 20	30 f	65
20 – 25	10	75
25 – 30	10	85
30 – 35	8	93
35 – 40	5	98
40 – 45	2	100

Cf ≈ 50

$\boxed{\Sigma f = N = 100}$

$$\boxed{M = l_1 + \frac{\frac{N}{2} - cf}{f} \times i}$$

l_1 = lower limit of median class
cf = cumulative fre of class
f = frequency of median class
i = size of class interval

Calculate the median from the following data:

Daily wages : 210 160 200 150 170 180 230 220 140 190

No. of employee : 15 20 18 35 27 23 13 12 18 19

Ans ⇒ firstly arrange it into ~~ascending~~ the ascending order

X	f	cf
140	18	18
150	35	53
160	20	73
170	27	100
180	23	123
190	19	142
200	18	160
210	15	175
220	12	187
230	13	200

$N = \Sigma f = 200$

for even

Average of $\left(\frac{N}{2}\right)$th and $\left(\frac{N}{2}+1\right)$th item

$\left(\frac{200}{2}\right) = 100 \quad \left(\frac{200}{2}+1\right) = 101$

$= \frac{170 + 180}{2} =$

$\boxed{M = 175}$

for → continue series

* Calculate the median from the following data:

values	frequency
less than 10	4
" 20	16
" 30	40
" 40	76
" 50	96
" 60	112
" 70	120
" 80	125

ans ⇒

Interval	value	Frequency cf	f
0 – 10	less than 10	4	4
10 – 20	"	16	12
20 – 30	"	40 cf	24
30 – 40	"	76	36 f
40 – 50	"	96	20
50 – 60	"	112	16
60 – 70	"	120	8
70 – 80	"	125	5
			N = 125

$$M = l_1 + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$\frac{N}{2} = \frac{125}{2} = 62.5$

$l_1 = 30$

$cf = 40$

$f = 36$

$i = 10 - 0 = 10$ $\boxed{i = 10}$

$$M = 30 + \left(\frac{62.5 - 40}{36}\right) \times 10$$

$$M = 30 + \left(\frac{62.5 - 40}{36}\right) \times 10 \quad M = 30 + \left(\frac{22.5 \times 10}{36}\right)$$

$$M = 30 + 6.25 \quad \boxed{M = 36.25}$$

* Calculate the median from the following data:

value :	1–10	11–20	21–30	31–40	41–50
Frequency :	4	12	20	9	5

interval	frequency	cf
0.5 – 10.5	4	4
10.5 – 20.5	12	16 cf
20.5 – 30.5	20	36
30.5 – 40.5	9	45
40.5 – 50.5	5	50
	~~N=50~~	N=50

$\frac{N}{2} = 25$

$l_1 = 20.5$

$f = 20$

$cf = 16$

$i = 10.5 - 0.5$

$\boxed{i = 10}$

$$M = l_1 + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$M = 20.5 + \left(\frac{25 - 16}{20}\right) \times 10$$

$$M = 20.5 + \frac{9}{20} \times 10$$

$$M = 20.5 + 4.5$$

$$\boxed{M = 25}$$

1.

meadian class = $\left(\frac{N}{2}\right)$ item = $\frac{100}{2}$ = 50th item

meadian class = 15 – 20 = 5 = i

$cf = 35$

$$M = 15 + \left(\frac{50-35}{30}\right) \times 5 \Rightarrow \boxed{M = 17.5}$$

inclusive series
↓
Mid value series

i = diff b/w 2 middle val = 15 – 5 = 10

M.	F	X	cf
m	15	0 – 10	15
5	7	10 – 20	22
15	11	20 – 30	33
25	10	30 – 40	43
35	13	40 – 50	56
45	8	50 – 60	64
55	20	60 – 70	84
65	16	70 – 80	100
75			

$\boxed{l_1 = m_1 - \frac{i}{2}}$ $\quad l_1 = 5 - \frac{10}{2} \quad i = 0$

$h_1 = m_1 + \frac{i}{2} = 5 + \frac{10}{2} = 10$

$\boxed{M = m + \frac{i}{2}}$ $\quad N = 100 \quad \frac{N}{2} = 50$

$l = 40$

$f = 13$

$cf = 43$

$i = 10$

$$M = 40 + \frac{50-43}{13} \times 10$$

$$M = 40 + \frac{70}{13} \quad \boxed{M = 45.38}$$

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

* find the missing frequency

Marks	f	cf
0 – 10	10	10
10 – 20	f1	10 + f1
20 – 30	25	35 + f1
$\boxed{30 – 40}$	30	65 + f1
40 – 50	f2	65 + f1 + f2
50 – 60	10	75 + f1 + f2
N = 100		

$N = 100$

$M = 30$

$f_1 + f_2 + 75 = 100$

$f_1 + f_2 = 25$

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

$\cancel{80} = 30 + \left(\frac{50 - (35 + f_1)}{\cancel{100}\ 10} \right) \times 10$

given
$M = 30$
$L_1 = 30$
$f =$

$10 = 50 - 35 - f_1$

$10 = 15 - f_1$

$f_1 = 15 - 10$

$\boxed{f_1 = 5}$

$f_1 + f_2 + 75 = 100$

$f_2 = 100 - 75 - f_1$

$f_2 = 100 - 75 - 5$

$\boxed{f_2 = 20}$

Mode

Mode is the ~~impostent measurement of~~ central tendancy. It is the value which occure more freaquently.

8, 8, 12, 8, 12, 8, 4, 5 individual series

X	f
4	1
5	1
8	4
12	2

Discrete series mode: (i) inspection method
(ii) Grouping method (PRE by SKU)

110	120	130	140	150	160
2	4	8	10	5	4

$\boxed{\text{Mode} = 140}$

~~Grouping method~~ Continuous series

$$\boxed{Z = l_1 + \frac{f_1 - f_0}{2f_1 - (f_0 - f_2)} \times i}$$

Z is the mode

l_1 = lower ~~middle~~ limit of the ~~lower~~ Modal class

f_0 → frequency of the pre mode of class

f_1 → frequency of the ~~post~~ ~~in front of mode~~ of class

f_2 → frequency of the post mode of class

$\boxed{\text{Mode} = 3\,\text{median} - 2\bar{x}}$ 2marks

X	f	
0-5	3	
5-10	7	
10-15	15	f_0
15-20	30	f_1
20-25	20	f_2
25-30	10	
30-35	5	

max frequency

Z = $l_1 = 15$

$f_1 = 30$

$f_0 = 15$

$f_2 = 20$

$i = 5$

Calculate the mode from the following data

ans →

$$Z = \left(15 + \frac{30-15}{(2\times 30 - 15 - 20)}\right) \times 5$$

$$Z = 15 + \left(\frac{15}{60-35}\right) \times 5$$

$$Z = 15 + \frac{15}{25} \times 5$$

$\boxed{Z = 18}$

Marks	cf	f	X
10 & 15	4	4	10 – 15
10 & 20	12	8	15 – 20
10 & 25	30	18	20 – 25
10 & 30	60	30	25 – 30
10 & 35	80	20	30 – 35
10 & 40	90	10	35 – 40
10 & 45	95	5	40 – 45
10 & 50	97	2	45 – 50

$i = 5$

$f_1 = 30$
$f_0 = 18$
$f_2 = 20$
$l_1 = 25$

$Z = 25 + \frac{30-18}{2\times30-(18-20)} = \frac{12}{60-38} = \frac{12}{22}$

$= \frac{6}{11} = \underline{0.54}$

27.7
ans ⇒

* Geomitric Mean : It is nth root of the product of all the n value of the variable.

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \ldots x_n}$$

GM of 4 & 9.

$$GM = (x_1 \cdot x_2 \ldots x_n)^{\frac{1}{n}}$$

taking log both side

$$\log GM = \log(x_1 \cdot x_2 \ldots x_n)^{\frac{1}{n}}$$

$$\log GM = \frac{1}{n}\log[x_1 \cdot x_2 \ldots x_n]$$

$$\log GM = \frac{1}{n}[\log x_1 + \log x_2 \ldots \log x_n]$$

$$\boxed{GM = \text{antilog}\left(\frac{\Sigma(\log x)}{n}\right)}$$

this is for individual

Q

x	$\log x$
180	2.2553
190	2.28
240	2.38
386	2.58
492	2.69
662	2.82

sum =

sum $= \frac{15.0053}{GM} = \frac{15}{6}$ after that

$GM = \text{antilog}(2.5)$ = antilog(2.5) =

$$\boxed{GM = 316.22}$$

Geomietric mean

Measure of Dispersion : spread or variation

Conteral tendency : How to find average and it gives Conteral value and How the value revolve around a Center but it donot show the correct picture of variability disperpition of data, dispertion mean distance from the average.

$49, 50, 51 \rightarrow \bar{x_1}$ $\qquad \bar{x_1} = \frac{\Sigma x}{N} = \frac{150}{3} = 50$

$10, 60, 80 \rightarrow \bar{x_2}$ $\qquad \bar{x_2} = \frac{150}{3} = \boxed{\bar{x_2} = 50}$

variability means how the Aborvation of the ~~set~~ are scerated around the average.

dispertion means the extend of uniformity and degree of variation, more the uniformity less is the degree of dispertion.

Absolute value : exact value.

relative value : with respect to another value.

Dispersion

- Range
- QD Quartile Deviation
- Inter Quartile Range
- mean Deviation
- Standard Deviation

Range : Range is the difference b/w ~~lower~~ ~~height~~ Hioghest and lowest value in the series.

Individual series

Q. find the range and coefficient range from the following series

~~3, 8, 2, 5, 6,~~ 3, 8, 2, 5, 6, 9, 10

Range = 10 − 2 = 8

$$\text{coef range} = \frac{10-2}{10+2} = \frac{\cancel{8}}{\cancel{12}} = \frac{2}{3}$$

discrete series

X	f	
10	3	
20	5	
30	2	
40	6	
50	3	

range = 50 − 10 = 40

$$\text{coe range} = \frac{50-10}{50+10} = \frac{\cancel{40}}{\cancel{60}}\ \frac{2}{3}$$

Q. Continue series

X	f	
10 — 20	6	
20 — 30	3	
30 — 40	8	
40 — 50	5	
50 — 60	6	

* IQR → Inter Quartile Range: It is difference b/w upper & lower quartile.

$$\boxed{Q = Q_{up} - Q_{low}}$$

Q_1 $\quad$ Q_2 $\quad$ Q_3 $\quad$ Q_4

* Quartile deviation: ~~Hand~~ Half

$$\boxed{QD = \frac{Q_3 - Q_1}{2}}$$

* Coeficient of the Quartile deviation:

$$\boxed{Coef = \frac{Q_3 - Q_1}{Q_3 + Q_1}}$$

Q find Q_1 & Q_3 of the following data

3, 8, 5, 2, 6, 9, 4, 10

before → arrange ascending order

→ 2, 3, 4, 5, 6, 8, 9, 10

$$\boxed{Q_1 = \left(\frac{N+1}{4}\right)^{th}} \quad \boxed{Q_2 = 2\left(\frac{N+1}{4}\right)^{th}} \quad \boxed{Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{item}}$$

$$\boxed{Q_4 = 4\left(\frac{N+1}{4}\right)^{th} \text{item}}$$

$Q_1 = \left(\frac{8+1}{4}\right) = 2.25$ th item

$\boxed{2\text{th item} + 0.25\ \text{item}\,(3\text{rd item} - 2\text{th item})}$

$2 + 0.25(4 - 3)$

$3 + 0.25(1) = \boxed{3.25}$

$$Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$$

$$Q = 3\left(\frac{8+1}{4}\right) = 3\left(\frac{9}{4}\right) = 3 \times 2.25 = \underline{6.75}$$

ans ⇒ 6th + 0.75 (7th − 6th)

ans → $8 + 0.75(9-8) = 8.75$

$$QD = \frac{Q_3 - Q_1}{2} = \frac{8.75 - 3.25}{2}$$

$$Q\ Coe = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{8.75 - 3.25}{8.75 + 3.25}$$

Assignment:

(1) Define statistics (2) Application of Stat (5) (3) Differe. primary data & secondary data (8 points).

* Mean deviation / ~~standar~~ Average deviation: It is defined as the arithmetic Average of deviation of various items of the series computed from sum measures of central tendency,

Mean deviation from the mean will be

$$\boxed{M.D_M = \frac{\Sigma|X - M|}{N}}$$

Coefficient of mean deviation = .

$$\boxed{M.D\bar{X} = \frac{\Sigma|x - \bar{x}|}{N}}$$

$$\boxed{Coeff\ M.D_{\bar{X}} = \frac{\text{mean Deviation}}{\text{Mean}}}$$

$$\boxed{Coeff\ M.D_M = \frac{\text{Median Deviation}}{\text{Median}}}$$

for Discrete

$$\frac{\Sigma f|X - M|}{N} \text{ and } \frac{\Sigma f|x - \bar{x}|}{N}$$

* Calculate the mean deviation from the mean after ell of median and coffe. of mean deviation from following data.

x	$\lvert x-\bar{x}\rvert$
20	$\lvert 20-45\rvert = 25$
22	$\lvert 22-45\rvert = 23$
25	-20
38	-7
40	-5
50	5
65	15
70	20
75	25

$$M.D_{\bar{x}} = \frac{\Sigma\lvert x-\bar{x}\rvert}{N}$$

$= \Sigma\lvert$ $= 70.22$

$$MD_M = \frac{\Sigma\lvert X - M\rvert}{N}$$

ans =

$\bar{x} = \frac{\Sigma x}{N}$ $\bar{x} = 45$

* Standard deviation (σ): variance $= (\sigma)^2$

$$\boxed{S.D = \sqrt{\text{variance}}}$$

First used by Karl Pearson in 1893, It is also known as root mean squre deviation, It is defined as the saure root of the arithmetic mean of square of deviation of the values taken from the mean.

$$\boxed{\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}}$$ $$\boxed{\text{coefficient of S.d } (\sigma) = \frac{\sigma}{\bar{x}}}$$

Std. deviat^n^
By direct method $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$ $\hookrightarrow \bar{x}$

(1) Difference b/w Mean deviation and standard deviation

+ve only	+ve, -ve

Q calculate the standard deviation from following ~~data~~ series

X	$\|X-\bar{X}\|$	$(X-\bar{X})^2$
10	-20	400
20	-10	100
30	0	0
40	10	100
50	20	400

$\bar{X}=30$ $\qquad \Sigma = 1000$

$$\boxed{\bar{X}=\frac{\Sigma X}{N}}$$

$$\sigma = \sqrt{\frac{1000}{5}} = \sqrt{200} =$$

$$\boxed{\sigma = 14.14}$$

$$\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}} = \sqrt{\frac{1000}{5}}$$

* short cut method: $\boxed{\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}}$

where $\boxed{d = X - A}$

$A = 30$

$d =$

X	$X-A=d$	d^2
10		
20		
30		
40		
50		

$\sigma = \sqrt{200}$

* step deviation method ~~for continuous series~~

$$\sigma = \sqrt{\left(\frac{\Sigma d'^2}{N}\right) - \left(\frac{\Sigma d'}{N}\right)^2} \times i \qquad d' = \frac{d}{i}$$

Discrete ~~mean~~ Series

Calculate the standard deviation ~~S.D.~~ from the following data

X	f	X−A = d	fd^2
3	7	−3	63
4	8	−2	32
5	10	−1	10
6	12	0	0
7	4	1	4
8	3	2	12
9	2	3	18

$$\boxed{\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}}$$

$A = 6$

$\boxed{N = \Sigma f = 46}$

Fd'
−21
−16
−10
0
4
6
6

$\boxed{\Sigma Fd' = -31}$

$\Sigma fd^2 = 139$

$i = 1$

$$\frac{\Sigma fd^2}{N} = \frac{139}{46} = 3.02$$

$$\left(\frac{\Sigma fd}{N}\right)^2 = \left(\frac{-31}{46}\right)^2 = \left(\frac{961}{2116}\right) = 0.45$$

$$\sigma = \sqrt{3.02 - 0.45}$$

$$\sigma = \sqrt{2.570}$$

$$\boxed{\sigma = 1.603}$$

Calculate Standard deviation & Mean deviation & coeff of deviation (σ)

X	f
0 – 10	15
10 – 20	32
20 – 30	51
30 – 40	78
40 – 50	97
50 – 60	109

Question अलग है। तो coef अलग है। don't worry

$$\boxed{\text{Coeff of deviation} = \frac{\sigma}{\bar{X}} \times 100}$$

$\boxed{\text{ans} = 15.48}$

Same

Continuous Series

~~Step deviation~~ method

$$\boxed{\sigma = \sqrt{\left(\frac{\Sigma f d'^2}{N}\right) - \left(\frac{\Sigma f d'}{N}\right)^2} \times i = \boxed{d' = \frac{d}{i}} \quad \text{where } \boxed{d = X - A}}$$

Calculate $\sigma = ?$

$\bar{X} = ?$

coeff of deviation $= ?$

PYQ 4M The mean and standard deviation of 200 items are found to be 60 and 20 respectively. if at the time of calculations, two items were wrongly taken as 3 and 67 instead of 13 and 17, detect the correct mean and standard deviation. verify the correct coefficient of variation.

$$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2} = \sqrt{\frac{\Sigma x^2}{200} - (60)^2} = 20$$

$\frac{\Sigma x}{N} \to \bar{x}$

calculating incorrect Σx^2

$$\Rightarrow \frac{\Sigma x^2}{200} - 3600 = 400$$

$$\Rightarrow \Sigma x^2 = 4000 \times 200$$

$$\Rightarrow \Sigma x^2 = 8 \times 10^5$$

Incorrect	Correct
3	13
67	17

$$\Rightarrow \Sigma x^2 = 8 \times 10^5 - (3)^2 - (67)^2 + (13)^2 + (17)^2 \Rightarrow \text{correct } \Sigma x^2$$

$$\Rightarrow 8 \times 10^5 - 9 - 4489 + 169 + 289$$

$$\Rightarrow 795{,}960$$

$x_1 = 2$

$x_2 = 3$ 7

$x_3 = 4$

$\Sigma x^2 = 2^2 + 3^2 + 4^2 = 50$

$2^2 + 3^2 + 4^2 - 3^2 + 7^2 =$

$$\boxed{\frac{\Sigma x}{n} = \bar{x}}$$

Incorrect $\Sigma x = 60 \times 200$

correct $\Sigma x = 12000 - 3 - 67 + 13 + 17$

$\Sigma x = 11960$

correct $\bar{x} = \frac{11960}{200} = 59.8$

$$CV = \frac{20.09}{59.8} \times 100$$

33.59%

$$\text{correct } \sigma = \sqrt{\frac{795{,}960}{200} - (59.8)^2} = \sqrt{3979.8 - 3576.04}$$

$$\sigma = \sqrt{403.8} = 20.09$$

* Skewness and kurtosis

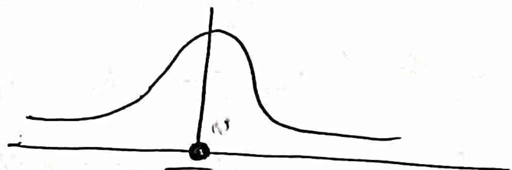
* Skewness: when a series is not symmetrical, it is said to be asymmetrical or skewed. Skewness refers to asymmetry or lack of symmetry in the shape of frequency.

* distribution: A distribution is said to be skewed, when mean, median falls at different points in the distribution.

Distribution → Symmetry
Distribution → Asymmetry

* Symmetry distribution: No skewness is present, Mean, Median and mode are ~~with~~ coincides (equal).

$$\boxed{\bar{X} = M = Z}$$

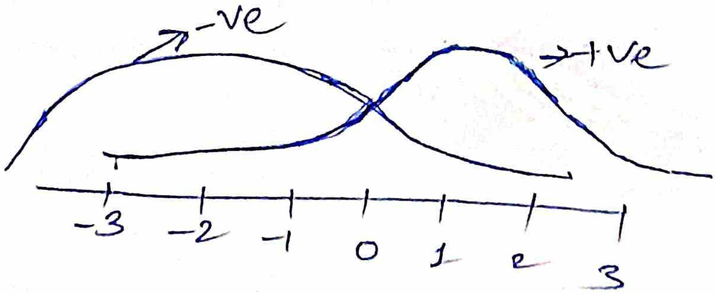


$\bar{X} = M = Z$

* Asymmetry distribution: when a distribution is not symmetrical, then it is skewed or Asymmetry distn.

There are 2 types of skewness.

① +vely skewed ② Negatively skewed.



* diff b/w dispersion and skewness.

dispersion	skewness
(i) Dispersion is concerned with the amount of variation	(ii) skewness is concerned with direction of variation.
(ii) It focus on spread of data.	It focus on symmetry of data
(iii) It helps to judge the reliability of central tendency.	(iii) It helps to identify deviation from normality
(iv) Example: Range, variance, IQR	Ex ⇒ skewness coefficient.

when $\bar{x} > z \rightarrow$ (+vely skewed)

$\bar{x} < z \rightarrow$ (−vely skewed)

There are 4 types of skewness (10 Marks)

(i) karl pearson coefficient of skewness:

$$\boxed{\text{Karl c.o.s} = \frac{\bar{x} - z}{\text{std. deviation}}}$$

$$\boxed{\text{karl c.o.s} = \frac{\bar{x} - z}{\sigma}}$$

(PYQ) Generate Karl Pearson's coefficient of skewness from the following data.

profit RS. L	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
no of cos	12	18	35	42	50	45	30	8

∵ Karl co.s $= \frac{\bar{X} - Z}{\sigma}$, $\boxed{\bar{X} = A + \frac{\Sigma fd}{N}}$ $\boxed{d = X - A}$

$$\boxed{Z = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times i}$$

$$\boxed{\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i}$$

$\boxed{d = X - A,\ d' = \frac{d}{i}}$

profit	f	d=X-A	X	fd	d'	d'^2	fd'	fd'^2
70-80	12	-35	75	-420	-3.5	12.25	-42	159.25
80-90	18	-25	85	-450	-2.5	6.25	-45	112.5
90-100	35	-15	95	-525	-1.5	2.25	-52.5	78.75
100-110	42	-5	105	-210	-0.5	0.25	-21	10.5
110-120	50	5	115	250	0.5	0.25	25	12.5
120-130	45	15	125	675	1.5	2.25	67.5	101.25
130-140	30	25	135	750	2.5	6.25	75	187.5
140-150	8	35	145	280	3.5	12.25	28	98

$\Sigma f = N = 240$ $\boxed{N = 880}$ $\boxed{\Sigma fd = 350}$ $L = 110$, $f_0 = 42$

$\boxed{A = 110}$ $\boxed{i = 10}$ $f_1 = 50$ $f_2 = 45$

$\Sigma fd' = 35$ $\Sigma f \cdot d'^2 = 760.25$

$\bar{X} = 110 + \left(\frac{350}{\cancel{880}\ 240}\right)$

$\cancel{\bar{X} = 110.3977}$

$\boxed{\bar{X} = 111.45}$

$Z = 110 + \left(\frac{50 - 42}{2 \times 50 - 42 - 45}\right) \times 10$

$Z = 110 + \left(\frac{8}{13}\right) \times 10 \Rightarrow Z = 110 + 6.15$ $\boxed{Z = 116.15}$

$$\boxed{\frac{\Sigma fd'}{N} = \frac{35}{880} = 0.039}$$

$$\boxed{\left(\frac{\Sigma fd'}{N}\right)^2 = 0.001521}$$

$$\frac{\Sigma f \cdot d'^2}{N} = \frac{760.25}{880} = 0.8639$$

$$\sigma = \sqrt{0.8639 - 0.001521}$$

$$\sigma = \sqrt{0.86}$$

$$\boxed{\sigma = 0.92}$$

$$\text{karl P.C.O.S} = \frac{\bar{X} - Z}{\sigma} = \frac{110.39 - 116.15}{0.92}$$

$$= \frac{-5.76}{0.92} = -6.26$$

$$\boxed{\text{karl P.C.O.S} = -6.26}$$

* find karl Pearson coefficient of skewness

X	58	59	60	61	62	63	
f	10	18	30	42	35	38	

Mode(z) = 61

$\bar{X} = 61.08$

$\sigma = -0.025$

Ans ⇒

X	f	fX
58	10	580
59	18	1062
60	30	1800
61	42	2562
62	35	2170
63	38	2394

$N = 173$

$\Sigma fx = 10568$

$\boxed{Z = 61}$

$\boxed{\bar{X} = 61.08}$

$$\boxed{\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}}$$

Unit 2 → Sampling Distribution & Testing of Hypothesis

* population: Population means entire field under investigative about which knowledge is sort.

Ex ⇒ if we want to collect information about the monthly expendecture of 2000 students of a college then the entire agricate of two thousand student will be termed as universe or population.

population is of two type

① finite population ② Infinite population.

* sample: sample is a part of population.

(PYQ) (4M) what is sampling? different b/w probability and non-probability sampling.

Sampling is a method of select a sample from a given population is known as sampling.

probability sampling	Non-probability sampling
(i) probability sampling is a method in ~~witch~~ which there is equal opportunity to be selected as a representative.	Non-probability sampling is a method in which there is no equal opportunity to be selected as representative.
(ii) It is Also known as random sampling.	Also known as non-random sampling.
(iii) In this selection is random.	In this selection is arbitary
(iv) It is objective method.	It is subjective method.
(v) Hypothesis is ~~gener~~ tested.	Hypothesis is generated.
(vi) Produces unbiased result.	Produces biased result.

Sampling

Probability sampling

1. Simple Random sampling
2. Straitified " sampling
3. Systematic " sampling
4. Clustar " Sampling

Non-probability Sampling

1. Convenience Sampling
2. Judgement sampling
3. Quota sampling
4. Snowball sampling

① Simple Random sampling : Every member of the population has an equal chance of selection.

Ex ⇒ lottery method / for example 50 people are playing
Ludo method / lottery so 50 lottery ticket are collected
Spin method / in bucket and you are selecting randomly on ticket so any one can selected in b/w 50.

(ii) ~~Stratified random sampling~~

Merit

(i) free from personal buyers.
(ii) Save time money or labour.

Demerit

- Time consuming
- Not suitable for certain populations.

② Stratified Random sampling : In this sampling population is devided into sub groups (called strata) like : (gender, age, range, job role)

Example:
for example you have to select 50 workers for industry so you can make strata and select randomly from each group like 25 from gemeder group, 25 from age group.

Merit

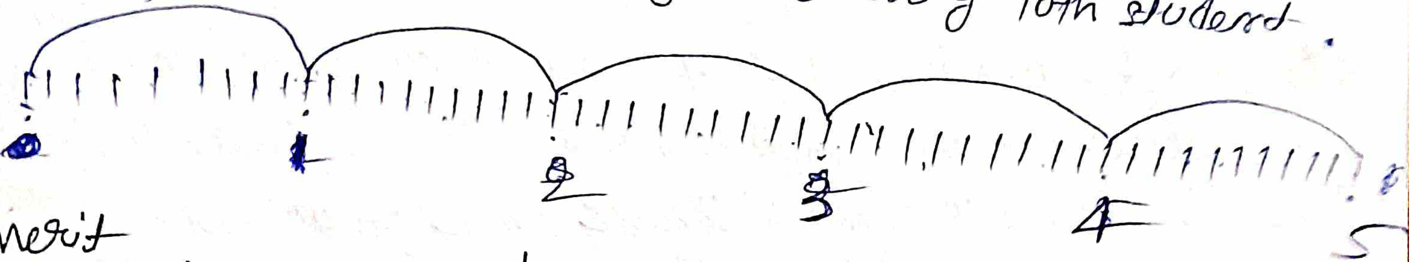
(i) More Accuracy
(ii) Increased precision

Demerit

- Complexity
- Need for information.

③ Systemetic random sampling: In this sampling population is arranged in a particular order either in ascending or a descending.

Example: If we want to select 5 students from 50 the we systemeticly arrange this in either ascending or descending order and since I have to select 5 out of 50 so $\frac{50}{5} = 10$ we can easly select every 10th student.



Merit
- greate Efficiency
- Representativeness

- Not suitable for small population
- if order is not maintained may be biased.

④ Cluster (area) random sampling: In this sampling population is devided into subgroups based upon area like (city blocks)

Example: Consider A city devided Into neighbourhoods, you randomly choose a few neighbourhoods in your sample.

Merit
- Efficiency
- practicality

Demerit
- Analysis complexity
- logistical challenges.

① Convenience sampling: In this sampling researcher selects the ~~most~~ accessible population members.

Example: for example your company in Ludhiana so you will select Labourers in Ludhiana itself.

Merit	Demerit
• More convenient sampling	• Not representative
• cost-effective	• Not accurately when large population
• Saving time and resources	

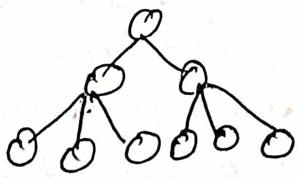
② ~~Jug~~ Judgement sampling: ~~The researcher selects population where~~ good.
In this sampling selecting sample based on the researcher's judgement or expertise.
Example: Imagine a researcher studying customer in a restaurant, the researcher may choose individuals they believe represent a diverse range of opinions.

Merit	Demerit
• Efficiency	• Bias risk
• Expertise	• Difficulty

③ Quota sampling: In this sampling when researchers select participants based on certain quota.
Example: Consider survey on smartphone in a city you might set quotas for age groups.

Merit	Demerit
• Cost-effective	• Bias Risk
• Quick implementation	• Difficulty in Quota Setting.

④ Snowball sampling: In this sampling researcher collect data from a participant and asked for other participants and again collect data from refered ~~participant~~ participants like chain.



Merit
• Cost-effective
• More Accurate

Demerit
(i) Bias
(ii) lack of Control.

diff b/w

(pyq) Sampling error and non sampling error

Sampling error	non- sampling error
① Sampling error due to chance variations in the sample.	non- sampling error due to random variation in the sample.
② It cause → Random variation in the sample.	cause → Bias in the data collection
③ due to wrong method of sampling	③ due to human factor.
④ It can be controlled by ~~reduced~~ increasing sample size	④ It can be controlled by careful study.
⑤ Ex ⇒ The difference between the average height in the sample and the true average height of all students is the sampling error	⑤ Ex ⇒ Imagine a survey conducted over the phone where some individuals do not have access to phone. This introduces a non- sampling error.

* population Mean → μ
* Standard Deviation → σ
* Variance → σ^2
* population proportion → p

sample Mean → $\bar{x}$

Sample SD → s

Sample variance → s^2

sample population proportion → $\hat{p}$

* ~~Test of hypothesis~~: We have to make decision about population on the basis of sample.
Ex → A drug chemist has to decide wheater a new drug is effective ~~or~~ ~~not~~ curing a disease or not
Such decisions are known as ~~Statistical~~ Staftical decision.
Hypothesis : It is a claim that need to be tested.
An attempt to arive at a decision about the population on the basis of sample information
It is necessary to make assumption about the parameter first. Such as assumption are called statical Hypothesis which may or may not be true.
* There are two types of Hypothesis.
(i) Null Hypothesis (ii) Aulternate Hypothesis

(i) Null Hypothesis : ~~It~~ says that there is no diff b/w sample stats ($\bar{x}$) and ~~…~~ population parameter (μ). $[\bar{x} == \mu]$
What are the abosorbed diff due to ~~fluction~~ fluctuation in sampling from same population. Ex = $\bar{x} = \mu$ $\mu = 50$

Aulternate Hy : Any Hypothesis different from null Hy is called aulternet Hypothesis (H_a)
It is denoted by H_1 या H_a Ex → $\mu_1 \Rightarrow \mu \neq 50$
$\mu_1 > 50$ या $\mu < 50$

difference b/w.

Type 1 error and type 2 error

Type 1 error	Type 2 error
It is made when we reject the Null Hypothesis although it was true.	It is made when we accept the null Hypothesis although it was false.
(ii) The probability of making this error is α: $P(E_1) \to \alpha$	(ii) The probability of making this ~~e~~ error is β $P(E_2) \to \beta$
(iii) The probability of making Type 1 correction is $1-\alpha$	(iii) The probability of making Type 2 correction is $1-\beta$
(iv) Type 1 error is called a false positive.	(iv) A ~~type~~ type 2 error is called a false negative.

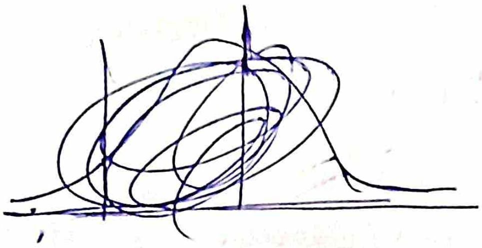
2 Marks

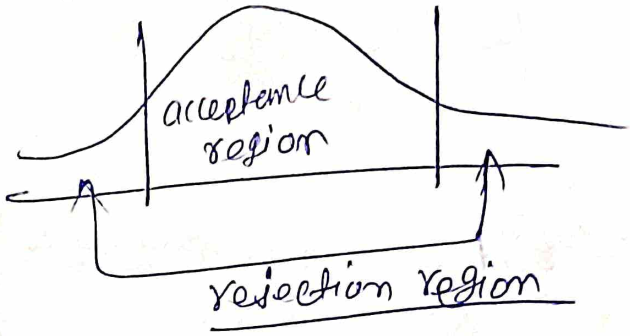
Level of significance : The degree of significance which we accept or reject Hypothesis is known as level of significance.

100% Accuracy is not possible in making a decision so we fix the confidence (error) ~~at~~ 5% ∵ for a greater precision that means there is 5% ~~chance~~ chance of error. If we fix confidence at 1% level the decision is correct to 99%. If no level of significance is given we always take level of confidence 0.05 or 5% (α).

* Critical region/ rejection region:

The state which leads to rejection of Null Hypothesis give us region is known as rejection region.





* Test of Hypothesis

if. $n \geq 30 \rightarrow$ large sample

(i) Test of Hypothesis about population mean.

(ii) " " " diff. b/w two population mean.

(iii) " " " Two population standard deviation

(iv) " " " population proportion.

(v) " " " about difference of two population proportion.

① Test of Hypothesis about population Mean (m)

$$\boxed{Z = \frac{\bar{X} - \mu}{S.E_{\bar{X}}}}$$

Standard error of mean → $S.E_{\bar{X}}$

$$\boxed{S.E_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

$\bar{X}$ = sample mean
μ = population mean
σ = population sd
S = standard deviation

Q the mean height of a sample of 100 student, mean hight of random sample of 100 student is 64 inch and std deviation is is 3 inch. Test the statement the mean height of the population is 67 inch. at 5% level of significance. (1.64)

Ans → $\bar{X} = 64''$ · $S = 3''$ $\mu = 67''$ $\boxed{Z = 1.96}$
$n = 100$

1 tail test of 5% $Z \rightarrow 1.64$
1% $Z \rightarrow 1.33$

① Null Hypothesis → there is no significant diff b/w mean height and mean height of population

$H_0 = \mu_0$ $(\mu = \mu_0)$
$H_A = \mu_a$
$H_0 \rightarrow \bar{X} = \mu_0$
$\mu_0 = 67$ onetailed test

② Alternate Hypothesis (H_1): There is significant diff b/w mean height and mean height of population.
$H_1 : \mu \neq 67$ (Two tailed Test)

$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{100}} = \frac{3}{\sqrt{100}} = 0.3$$

$$Z = \frac{64 - 67}{0.3} \qquad \boxed{Z = 10}$$

at 5% of level of significance, the critical value of z for two tailed test = 1.96 since of significance, we reject the Null Hypothesis and conclude that the population mean cannot be equal 67.

② Test of Hypothesis about diff b/w Two population Mean.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE_{\bar{x}_1 - \bar{x}_2}}$$

when population S.D not given

$$S.E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$\bar{x}_1 \rightarrow$ Mean of 1st sample

$\bar{x}_2 \rightarrow$ mean of 2nd "

$SE.\bar{x}_1 - \bar{x}_2$ = Std err of dif b/w two means when s1 and s2 or given

$$S.E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

① A random sample of 1000 workers from south Ind so that their mean wages are rupes 47 per week with standard deviation 28. A random sample of 1500 worker from North india gives a mean wages of ₹ 49/week with std deviation rupess 40. Is there any significant diff b/w mean level of wages in two places?

Soln: given ⇒

$n_1 = 1000$

$\bar{x}_1 = 47$ ₹

$s_1 = 28$

$n_2 = 1500$

$\bar{x}_2 = 49$

$s_2 = 40$

Ⓢ1

H0 : There is no significant difference b/w the mean wages of north indian and south india. $(\mu_1 = \mu_2)$ or $(\bar{x}_1 = \bar{x}_2)$

H1 : There is significant difference b/w the mean wages of north india and south india. $\mu_1 \neq \mu_2$

Two tailed test

$\boxed{\mu_1 - \mu_2 = 0}$

$$S.E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(28)^2}{1000} + \frac{(40)^2}{1500}} = 1.36$$

$\boxed{S.E_{\bar{x}_1 - \bar{x}_2} = 1.36}$

$$\boxed{|Z| = \frac{47 - 49}{1.36}}$$

$z = 1.47$ ans

tabulated ~~of~~ value of $z = 1.96$

we except the null hypothesis and there is no significant diff b/w mean wages of north india and south india.

Q. The mean of two large sample of size 1000 and 2000 are 168.75 and 170 cm respectively. can we draw sample with same mean and std dev 6.25. value of $\boxed{Z.(5.26)}$

Sol^n: given: $n_1 = 1000$ $\bar{x}_1 = 168.75$ $s_1 = 6.25$
$n_2 = 2000$ $\bar{x}_2 = 170$ $s_2 = 6.25$

~~We cannot draw sample with same mean.~~

~~we can draw sample with same mean.~~

$$\cancel{S.E_{\bar{x}_1 \cdot \bar{x}_2} = \sqrt{\frac{(6.25)^2}{1000} + \frac{(6.25)^2}{2000}} = 0.242}$$

$H_0 \to \mu_1 = \mu_2$ i.e both the samples are drawn from the population with mean and S.D. 6.25.

$H_1 \to \mu_1 \neq \mu_2$; both the samples are not drawn from the population with mean and S.D 6.25.

$\Rightarrow$ (two tailed test)

$$S.E_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{(6.25)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)} = \sqrt{\cancel{0.24}}$$

$$\boxed{S.E_{\bar{x}_1 - \bar{x}_2} = 0.242}$$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{S.E._{\bar{x}_1 - \bar{x}_2}}$$

$$z = \frac{1.28}{0.242} \quad \boxed{z = 5.28}$$

At 5% level, the critical value of z for two tailed test 1.96, since the calculated value of $|z| = 5.28$ is greater than critical value of $z = 1.96$, we reject H_0 and conclude that the both samples are not drawn from population with sam

Q Differennce b/w primary and secondary data

primary data	secondary data
(i) primary data is a type of data researchers directly collect from main source	(i) Secondary data is a type of data researchers collect already existing data produced by previous reachers.
(ii) Researcher has full control over data.	(2) Researcher has limited control
(3) It is more accurate	(3) It depends upon source and collection method.
(4) It is related to real-time data.	(4) It is related to the past.
(5) More Time consuming	(5) less time consuming.

Q. what is statistics and it's importance.

ans ⇒ statistics is the branch of mathematics for collecting, analysing and interpreting data. statistics can be used to predict the future, It determine the ~~probabili~~ probability that a specific event will happen or help answer questions about a survey.

Importance of statistics

1. Data Analysis: It helps in analyzing and interpreting complex data.
2. Research: It helps to researchers to draw conclusions and make decisions.
3. Business ~~and~~ Decision-making: Business use statistics to make informed decisions, understand market trends.
4. Quality control: In manufacturing and industry statistics ~~are~~ helps to ensure quality of products.

Q write properties of Mean.

ans ⇒ ① Sensitivity to changes: The mean is sensitive to extreme value.

② Balancing property: The sum of the deviations of individuals values from the mean is always zero.

③ Affected by skewness ④ Arithmatic Mean formula

⑤ Sample vs. Population mean ⑥ unique value

⑦ Useful in Normal Distributions.

Q prove that $\Sigma(x-\bar{x})=0$

proof →

$\Rightarrow x_1-\bar{x}+x_2-\bar{x}+x_3-\bar{x}+x_4-\bar{x}+ \ldots\ldots\ldots + x_n - \bar{x}$ —— ①

∵ $\bar{x} = \dfrac{\text{sum of all observation}}{\text{no of items}}$ $\bar{x} = \dfrac{x_1+x_2+x_3+x_4+\ldots x_n}{n}$

$\boxed{\bar{x}n = x_1+x_2+x_3+x_4+\ldots\ldots x_n}$ —— ②

from eq^n ①

$x_1+x_2+x_3+x_4+\ldots\ldots x_n - n\bar{x}$

from eq^n ②

$n\bar{x} - n\bar{x} = 0$

hence proved that $\boxed{\Sigma(x-\bar{x})=0}$

x	$x-\bar{x}$
8	3
1	-4
6	1

$\Sigma(x-\bar{x})=0$

$\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{15}{3}$

$\boxed{\bar{x}=5}$

$\Sigma(x-\bar{x}) = 3-4+1$
$= 4-4$

$\boxed{\Sigma(x-\bar{x})=0}$

Q Explain different types of coorelation

① Positive correlation: when two variables move in the same direction so it is called as positive correlation. As one variable increases, the other also increases and vice versa.

Ex → As the hours of study increases, exam scores also increases.

(ii) Negative correlation: Negative correlation when two variables move in opposite directions, As increase in one variable is associated with a decrese in the other and vice versa.

Ex ⇒ As the temperature decreases the number of ice cream sales decreases.

(3) Zero coorelation: Zero correlation indicates no linear relationship between two variables. changes in one variable are not associated with ~~the~~ the other.

Ex ⇒ The amount of rainfall and the number of books sold in a bookstore.

(4) Q) difference b/w population and sample

population	sample
(i) A population is the entire collection of subjects.	A sample is subset of population.
(2) Measurement taken from a whole population are called parameters	Measurements taken from a sample are called statistics.
(3) It provides complete and accurate information	It provides an estimated result of the population.
(4) It always greater than sample.	It always smaller than population.

karl → 4
median → d, 8,

(8) one tailed and two tailed test

one tailed test	two tailed test
(1) A one-tailed test is a statistical hypothesis test in which the alternative hypothesis specifies the direction of a difference b/w two population. $\mu = 50$	A two-taild test is a statistical hypothesis test in which the alternative hypothesis does not specify the direction of a difference between two populations. $\mu \neq 50$ $\mu < 50$ या $\mu > 50$
(2) when the alternative hypothesis specifies a direction then we use a one-tailed test.	(2) if no direction is given then we will use a two-tailed test.
(3) it is used to check whether the one mean is different from another mean or not	(3) It is used to check whether the two mean different from one another or not
(4) It is a uni-directional hypothesis	(4) It a non directional hypothesis

Chi square test

It is an impotent test among several ~~test~~ test of significance, It is used for testing the significance of population variance as ~~known~~ Non parametric test No assumption is taken about the population, It can be used as the test of goodness of fit & as test of independence of Attribute.

(i) Goodness of fit
(ii) Test of independece of Attr

(1) Test of goodness of fit: we try to find out how the absorbed value of given fenomina are significantly different from the expected value.

(2) chi square test determine how well theoritical distibution such as binamial poison. It is always applied if sample size is greater than 50.

(10) the following table give the no of accident that take place in industry during ~~[illegible]~~ various days of the week. test if the accident are uniformally given in week.

Days	Mon	Tue	Wed	Thus	fri	sat
No of Accident	14	18	12	11	15	14

degree of freedom → 2M → Q

Ho : Accident are not uniformly given in week.

Ha : Accident are ~~given~~ uniformly given in week.

$$\frac{14+18+12+11+15+14}{6} = 14$$

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
14	14	0	0	0
18	14	4	16	1.14
12	14	2	4	0.28
11	14	−3	9	0.64
15	14	1	1	0.07
14	14	0	0	0

$\boxed{\sum \frac{(O-E)^2}{E} = 2.14}$

$$\boxed{\chi^2 = \frac{\sum (O-E)^2}{E}} \qquad \chi^2 = \frac{30}{14} \qquad \boxed{\chi^2 = 2.14}$$

$$V = n-1 = 6-1 \qquad \boxed{V = 5}$$

Tabulated val = 11.07

at 0.05 tabulated value is = 11.07

∴ Calculated value is less than tabulated value so we ~~select~~ accept Null hypothesis hence Accident are not uniformly given in week.

Q 2. A sample analysis of examination result of 5000 students ~~were~~ were made it was found that 220 had faild 170 had secured third class, 90 second class, 20 ~~got~~ a first class, are they figure commensurable with the ratio of 4:3:2:1

Ho = No these figure are not commensurable with the ratio of 4:3:2:1

Ha = Yes these figure are commensurable with the ratio of 4:3:2:1

Faild → 220
Th → 170
Sec → 90
Fir → 20

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
200	200	20	400	2
170	150	20	400	2.66
90	100	−10	100	1
20	50	−30	900	18

~~200 + 5000 × 4/18~~

Faild $500 \times \frac{4}{10} = 200$

thir $500 \times \frac{3}{10} = 150$

Sec $500 \times \frac{2}{10} = 100$

Fir $500 \times \frac{1}{10} = 500$

$$\chi^2 = \frac{(O-E)^2}{E} =$$

$$\chi^2 = 2 + 2.66 + 1 + 18$$

$$\chi^2 = 4.66 + 9$$

$$\boxed{\chi^2 = 13.66}$$

~~Tabulated value = 7.81~~

degree of freedom $\vartheta = n-1 = 4-1 = 3$

☆ tabulated value of χ^2 at 5% level of significance 3 = 7.81
since calculated value is greater than tabulated value so we reject the Ho.

Q3. Test of goodness of fit for binomial distribution,

---

(2)

~~Case~~ Test of χ^2 as independence of attribute.

Chi square test enable us to examine wheather or not two attributes are associated or independent of of one another for example. we may be we interesting in knowing wheather a medicine is controlling and a ~~fever~~ fever or not.

Q A sample of 200 person with a particular disease was ~~for~~ selected. out of them 100 were given drug and other were not. The result were observed as follows.

	Drug (c_1)	No – drug (c_2)	Total (c_3)	
Cured	55	65	120	R_1
No Cured	45	35	80	R_2
Total	100	100	200	

Ans → H_0 → drug has not been effective in curing the disease

H_a → drug has been effective in curing the disease.

Let the null hypothesis =

O	E	(O−E)	$(O-E)^2$	$(O-E)^2/E$
55	60	−5	25	0.416
65	60	5	25	0.625
45	40	5	25	0.416
35	40	−5	25	0.625

$\Sigma (O-E)^2/E = 2.082$

$$\chi^2 = \frac{(O-E)^2}{E} = 2.082$$

degree of freedom $= (r-1)(c-1) \Rightarrow (2-1)(2-1)$

$= 1 \times 1 = 1$

The tabulated value of χ^2 at 5% level of frequency 1 = 3.84

Since, the calculated value $\chi^2 = 2.082$ is less than tabulated value 3.84 hence we accept the Null hypothesis and conclude that the drug has not been effective in curing the disease.

Q2. Can vaccination be regarded as preventive measure of small pox as evidence of following data of 1482 people exposed to small pox locality 368 were attacked. of those 1482 persons and 343 were vaccinated and those only 35 are attacked.

χ^2 ans = 48.22

Table are not given in question.

	vactinited	N-vactinited	Total
A1 Attacted	35	333	368
A2 Non-Attacted	308	806	1114
Total	343	1139	1482

O	E	O−E	$(O-E)^2$	$(O-E^2)/E$
35	85.17	−50.17	2517.02	29.55
308	257.82	50.18	2518.03	9.76
333	282.82	50.18	2518.03	8.90
806	856.17	−50.17	2517.02	2.93
				51.14

$\chi^2 = (O-E)^2/E$ $\boxed{\chi^2 = 51.4}$

degree of freedom $= (r-1)(c-1) = (2-1)(2-1) = 1$

$v = 1$

the tabulated value of χ^2 at 5% level of frequency $= 1 = 3.84$

Since, the calculated value i.e. $\chi^2 = 51.4$ is greater than tabulated value hence the null hypothesis is rejected.

Q In a survey of 200 boys, of which 75 were intelligent, 40 had educated father while 85 of the unintelligent boys had uneducated fathers. Do these figures support the hypothesis that educated fathers have intelligent boys use χ^2 test.

Sol:

	intelligent boy	unintelligent boy	Total
Educated father	40	40	80
uneducated father	35	85	120
Total	75	125	200

H_0 = Educated fathers do not have intelligent boys

H_1: Educated fathers do have intelligent boys.

O	E	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
40	30	10	100	3.33
40	50	-10	100	2
35	45	-10	100	2.22
85	75	10	100	1.33
				$\chi^2 = 8.88$

degree of freedom $v = (r-1)(c-1) = 1$

the tabulated value at 5% level of significance 1 = 3.84

since the calculated value $\chi^2 = 8.88$ is greater than tabulated value (3.84) so we reject the Null hypothesis and conclude that Educated fathers do have intelligent boys.

Q/10

Bowley's Coeficient

$$\boxed{B.C = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

Quartile for individual, discrete and Continue series

Individual Series	Discrete	Continue Series	range defin
Q_1 = size of $\left(\frac{N+1}{4}\right)$th item	Q_1 = size of $\left(\frac{N+1}{4}\right)$th item	$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times i$	$Q_1 = \left(\frac{N}{4}th\right)$
Q_3 = size of $3\left(\frac{N+1}{4}\right)$ th item	Q_3 = size of $3\left(\frac{N+1}{4}\right)$ th item	$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times i$	Q_3 = s of $3\left(\frac{N}{4}th\right)$
	M = s of $\left(\frac{N+1}{2}\right)$th item	$M = l_1 + \frac{\left(\frac{N}{2} - cf\right)}{f} \times i$	M = s of $\left(\frac{N}{2}\right)$th item

Q

No of Children (x)	No of famailies (f)	c.f	
0	7	7	
1	10	17	
2	16	33	> 24
3 M	25	58	
4	18	76	> 72
5	11	87	
6	8	95	

$N = \Sigma f = 95$

Q_1 = size of $\left(\frac{N+1}{4}\right)$th item

$Q_1 = \left(\frac{95+1}{4}\right) = \frac{96}{4} = 24$

$Q_1 = 2$

$Q_3 = 3\left(\frac{95+1}{4}\right)$

$Q_3 = 3\left(\frac{96}{4}\right) = 72$

$Q_3 = 4$

Median = $\left(\frac{N+1}{2}\right)$th item $= \left(\frac{95+1}{2}\right) = \left(\frac{96}{2}\right) =$ 48th item

$M = 3$, $Q_1 = 2$, $Q_3 = 4$

$$B.C = \frac{4 + 2 - 2 \times 3}{4 - 2} = \frac{4 + 2 - 6}{2} = \frac{6-6}{2} = 0$$

x	f	cf
0-10	10	10
10-20	25	35
20-30	20	55
30-40	15	70
40-50	10	80
50-60	35	115
60-70	25	140
70-80	10	150
	N = 150	

Q1 (row 20-30), M (row 40-50), Q3 (row 50-60)

$Q_1 = \text{size of } \left(\frac{N}{4}\right)\text{th item}$

$Q_1 = \frac{150}{4} = 37.5$

$Q_1 = 20 + \frac{\left(\frac{150}{4} - 35\right)}{20} \times 10$

$\boxed{Q_1 = 21.25}$

$Q_3 = \left(\frac{3N}{4}\right)\text{th item} = Q_3 = 3 \times \left(\frac{150}{4}\right) = Q_3 = 112.5$

$Q_3 = 50 + \frac{\left(3\left(\frac{150}{4}\right) - 80\right)}{35} \times 10$

$Q_3 = 50 + \left(\frac{112.5 - 80}{35}\right) \times 10$

$Q_3 = 50 + 9.28$ $\boxed{Q_3 = 59.28}$

$M = l_1 + \frac{\left(\frac{N}{2} - cf\right)}{f} \times i$

$M = \text{size of } \left(\frac{N}{2}\right)\text{th item} = \frac{150}{2} = 75$

$M = 40 + \left(\frac{75 - 70}{10}\right) \times 10 = 40 + \frac{5}{\cancel{10}} \times \cancel{10}$

$M = 45$

$B \cdot C = \frac{Q_1 + Q_3 - 2M}{Q_3 - Q_1}$ $B \cdot C = \frac{21.25 + 59.28 - 2 \times 45}{59.28 - 21.25}$

$B \cdot C = \frac{-9.47}{38.03}$ $\boxed{B \cdot C = -0.249}$

3rd test : Test of Hypothesis about difference Two population SD's :

Let s_2, s_1 = be the std deviation of two independent random sample of size n_1, n_2 from two population with std deviation σ_1 & σ_2.

$$Z = \left| \frac{(s_1 - s_2) - (\sigma_1 - \sigma_2)}{S.E_{s_1 - s_2}} \right|$$

$$S.E_{s_1-s_2} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

or $\sigma_1 = s_1$ $\sigma_2 = s_2$

Q. The mean yield of two sets of plot and their variablity are given below examine whether the diff. in variablity in yields is significant at 5% level of significance.

(Null excepted)

Mean yield plot	set of 40 plout 1258 lbs	set of 60 plots 1243 lbs
S.D per plot	34	28

Sol^n: given → $n_1 = 40$, $n_2 = 60$, $\bar{x_1} = 1258$ lbs, $\bar{x_2} = 1243$ lbs,
$s_1 = 34$ lbs $s_2 = 28$ lbs

① Null Hypothesis: $H_0: \boxed{\sigma_1 = \sigma_2}$ there is no significant difference in the variability in the yields b/w two sets at plots $\boxed{\sigma_1 - \sigma_2 = 0}$

② Alternative Hypothesis: $H_1: \sigma_1 \neq \sigma_2$ there is significant difference so on

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \quad Z = \frac{34 - 28}{\sqrt{\frac{(34)^2}{80} + \frac{(28)^2}{120}}} = \frac{6}{\sqrt{20.98}} = \frac{6}{4.58} = 1.31$$

Since calculated value is less than tabulated value 1.96 so we select null hypothesis and conclude that there is no significant difference in the variability in the yields b/w sets at plots.

~~E~~ test 5th test.

Test of Hypothesis about the difference between two population proportions.

$$\boxed{Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{S.E_{p_1 - p_2}}}$$

(a) $S.E_{p_1-p_2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$

when population pro P_1 & P_2 are known

$\boxed{q = 1 - p}$

$\boxed{p = \frac{x_1 + x_2}{n_1 + n_2}}$

(b) $S.E_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

when the population proportion P_1 and P_2 are not known

Q. In a certain district A, 450 persons were considered regular consumers of tea out of a sample of 1000 persons. In another district B, 400 persons were regular consumers of tea out of a sample of 800 persons. Do these data indicate a significant difference between the two districts so far as drinking habit is concerned? (use 5% level).

solⁿ: given $n_2 = 800$, $n_1 = 1000$ $x_1 = 450$, $x_2 = 400$

Null Hypothesis: $H_0: P_1 = P_2$ there is no significant difference in tea drinking habits in two districts $\boxed{P_1 - P_2 = 0}$

Alternative hy: $H_1: P_1 \neq P_2$ there is significant difference in tea drinking habits in two districts.

$p_1 = \frac{450}{100}$ $\boxed{p_1 = 0.45}$ $p_2 = \frac{400}{800}$ $\boxed{p_2 = 0.5}$

$S.E_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ $\because$ $\boxed{q = 1 - p}$ $q = 1 - p$

$p = \frac{450 + 400}{800 + 1000}$ $\boxed{p = 0.47}$

$q = 1 - 0.47$ $\boxed{q = 0.53}$ S.E

$$S.E_{P_1 - P_2} = \sqrt{(0.47)(0.53)\left(\frac{1}{1000} + \frac{1}{800}\right)}$$

$$= 0.0237$$

Now $|z| = \dfrac{0.45 - 0.50}{0.0237} = 2.1097$

At 5% level, the critical value of $z = 1.96$
since, the calculated value is greater than tabulated value so we reject Null hypothesis H_0. and conclude that there is significant difference b/w two districts so far as tea-drinking habit is concerned

(4) Test of hypothesis about population proportion

*Q. A random sample has size sample proportion of the member process ~~population~~ an attribute to test the hypothesis that the population proportion has specified value the test is given as.

$$\boxed{Z = \frac{p - P}{S.E_p}}$$

p - sample proportion

$P \rightarrow$ population proportion

$S.E_p = \sqrt{\frac{PQ}{n}}$ or $\sqrt{\frac{pq}{n}}$ and $q = 1-p$ where P is not known

$S.E_p \rightarrow$ Std error of proportion

Q A coin is tossed 100 times under identical conditions independently yielding 30 heads and 70 tails. Test at 1 % level of significance whether or not the coin is unbaised.

Sol^n: $n = 100$ | sample pro p = proportion to heads in the sample $= \frac{30}{100} = 0.30$

Population pro $(P) = \frac{1}{2} = 0.50 \Rightarrow Q = 1 - 0.50$

$\boxed{Q = 0.50}$

$\boxed{P = .50}$

H_0: coin is unbiased $H: P = 0.5$

H_1: coin is biased; $H_1: P \neq 0.5$

$$\boxed{z = \left|\frac{0.30 - 0.50}{0.05}\right|}$$

$S.E_p = \sqrt{\frac{0.5 \times 0.5}{100}} = 0.05$

$\boxed{S.E_p = 0.05}$

$\boxed{z = 4}$

At 1 % Level of significance the critical value of z for two tailed test $= 2.58$ since, the calculated value of $|z| >$ the critical value of z, we reject the NULL Hypothesis and hence conclude that the coin is baised.

Unit - 3 MST-2

Correlation ~~and~~ Analysis

Correlation : when we come across large no of problem envolving the use of two or more than two variable we make use of correlation. if two quantities varies in such a way that movement in one are accompanied by moveme-ment by ~~other~~ variable if two

the degree of relationship b/w the variable is measure through correlation analysis. It ~~deals with~~ association between the variable. It is a statistical device which help us in analysing the covariation for example → smoking & lung cancer.

We can measure the relationship b/w variables. It doesn't tell us about cause and effect, even a high degree of correlation doesn't necessarly mean that the relationship of cause and effect. It staibalishe only a correlation.

Type of correlation

① possitively correlation ② Negatively correlation,

* Usage of correlation:

① Developing the concept of regression.

② Deriving the degree and dire of relation within the variable

③ Reducing the range of uncertainties in matter of prediction.

④ In feild of bussiness, science

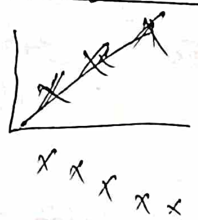
* Simple, partial and multiple correlation

simple corr : when only two vari are studied
partial corr : when more or three variab are confidered for analy but only two and rest not
Multiple corr : Three or more variable are studied
Ex → when we study the rela b/w yield of rice per acer and, both amoun of fertilizer used rainfall
It is a problem of multiple correlation. In partial
Ex → Teacher, doctor, mom dad correlation we recognize more than two variable but consider only two variables to be influencing variable keeping constant.

linear relationship : if amount of one variable

Methods of studying correlation

(i) scattered diagram.



(ii) graphic method

(3) karl pearson coefficient of correlation.

(4) Concorrent

+ & f left by own 12 Marks

* Karl pearson coefficient: denoted by γ.

$$\boxed{\gamma = \frac{(x-\bar{x})(y-\bar{y})}{N\sigma x\sigma y}}$$

where $N \rightarrow$ is the no of pair of observation

$\left.\begin{matrix}\sigma x\\ \sigma y\end{matrix}\right] \Rightarrow$ standerd deviation of xy.

$x = x-\bar{x}$

$y = y-\bar{y}$

$$\sigma x = \sqrt{\frac{\Sigma x^2}{N}} \quad \sigma y = \sqrt{\frac{\Sigma y^2}{N}}$$

$$\boxed{\gamma = \frac{\Sigma xy}{N\sigma x\sigma y}}$$

when mean is whole no

$$\boxed{\gamma = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}}$$

where Σxy is the sum of the product of two variables x and y

$\Sigma x^2, \Sigma y^2$ is the sum of the square of the values of y, x $(x-\bar{x})(y-\bar{y})$

Q find kar. p. c ~~[illegible]~~ marks in math (0.429)

x	y	$x-\bar{x}=x$	$y-\bar{y}=y$	$x*y$	x^2	y^2
48	45	14	10	140	196	100
35	20	1	-15	-15	1	225
17	40	-17	5	-105	289	25
23	25	-11	-10	110	121	100
47	45	13	10	130	169	100
				280	776	550

$\sqrt{\Sigma x^2 \Sigma y^2} = \sqrt{426800}$
$= 653.29931$

$\bar{X} = \frac{\Sigma X}{N} = 34 \quad \frac{170}{5} = 34$

$\bar{Y} = \frac{\Sigma Y}{N} = 35 = \frac{175}{5} = 35$

$\Sigma x^2 = 776$

$\Sigma y^2 = 550$

$\Sigma xy = 280$

$\gamma = \frac{280}{653.29931}$ $\boxed{\gamma = 0.4285}$

thus, there is a high degree of ~~[illegible]~~ correlation b/w variable x and y.

2. when mean is ~~[illegible]~~ → fraction

Assume mean method

$$r = \frac{N \sum dx\,dy - \sum dx \cdot \sum dy}{\sqrt{N \cdot \sum dx^2 - (\sum dx)^2} \cdot \sqrt{N \cdot \sum dy^2 - (\sum dy)^2}}$$

Q.

X	Y	(X − A) = dx	Y − A = dy	dx·dy	dx^2	dy^2
10	30	−6	−12	72	36	144
12	35	−4	−7	28	16	49
18	45	2	3	6	4	9
16	44	0	2	0	0	4
15	42	−1	−2	0	1	0
19	48	3	6	18	9	36
18	47	2	5	10	4	25
17	46	1	4	4	1	16
		$\sum dx = -3$	$\sum dy = 1$	$\sum dx\,dy = 138$	$\sum dx^2 = 71$	$\sum dy^2 = 283$

$\bar{X} = 15.62$

$\bar{Y} = 42.12$

N = 8

A for x = 16

A for y = 44

$$r = \frac{8 \times 138 - (-3)(1)}{\sqrt{(8 \times 71 - 9)}\ \sqrt{8 \times 283 - (1)}}$$

$$r = \frac{1104 + 3}{\sqrt{8 \times 62}\ \sqrt{8 \times 282}} = r = \frac{1107}{\sqrt{8 \times 62}\ \sqrt{8 \times 282}}$$

$$\boxed{r = 0.98}$$

H.W chp-3 → slly,

correlation, properties

partial, simple, multiple

Q.

* Q) $\Sigma xy = 8425$, $\bar{x} = 28.5$, $\bar{y} = 28.0$, $\sigma x = 10.5$

$\Sigma x, \Sigma y, \Sigma x^2, \Sigma y^2 = ?$

$\sigma y = 5.6$

$N = 10$

simple method / direct method

$$r = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$\bar{x} = \frac{\Sigma x}{N}$$

$$28.5 = \frac{\Sigma x}{10}$$

$$\Sigma x = 28.5 \times 10$$

$$\boxed{\Sigma x = 285}$$

$$\bar{y} = \frac{\Sigma y}{N}$$

$$\Sigma y = 28.0 \times 10$$

$$\Sigma y = 280$$

$$\sigma x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}$$

$$10.5 = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}$$

$$\frac{\Sigma x^2}{N} - (\bar{x})^2 = (10.5)^2$$

$$\frac{\Sigma x^2 - 10(28.5)^2}{10} = 110.25$$

$$\frac{\Sigma x^2 - 10(28.8)^2}{10} = 110.25$$

$$\Sigma x^2 - 8122.5 = 110.25 \times 10$$

$$\Sigma x^2 = 1102.5 + 8122.5$$

$$\boxed{\Sigma x^2 = 9225}$$

$$\boxed{\Sigma y^2 = 8153.6}$$

$$r = \frac{10 \times 8425 - (285)(280)}{\sqrt{10 \times 9225 - (285)^2}\sqrt{10 \times 8153.6 - (280)^2}}$$

$$r = \frac{4450}{\sqrt{11025}\sqrt{3136}} \quad r = \frac{4450}{5880} \quad \boxed{r = 0.756}$$

there is positive corre b/w x and y.

Series → Individual Series → only x → 1, 2, 3, 4, 5, 6, 7

① Series → Discrete series → two value x = 1, 2, 3, 4, 5, 6, 7 ; f = 2, 3, 4, 7, 8

Series → Continue series → two value x = 0–10, 10–20, 20–30, 30–40 ; f = 1, 2, 3, 4

Arithmetic Mean

Arithmetic Mean → (simple AM) ; (weighte AM)

Simple AM → individual series

M1 → ① Direct Method

$$\boxed{\bar{x} = \frac{\Sigma x}{N}}$$

$\bar{x}$ = Mean

N = No of values

M2 → ② Short cut Method

$$\boxed{\bar{x} = A + \left(\frac{\Sigma d}{N}\right)}$$

where A → Assumed mean

N = No of items

$\boxed{d = x - A}$

Σd = deviation of mid value

① 52, 40, 70, 43, 75, 40, 48, 35, 56 AM = ? student = 9.

given = $N = 9$

$\bar{x} = \frac{\Sigma x}{N}$ $\bar{x} = \frac{459}{9}$ $\boxed{\bar{x} = 51}$

② The pocket allowances of ten students are given below 15, 20, 30, 22, 25, 18, 40, 50, 55 and 65 Calculate the arithmetic mean by taking 40 as assumed mean.

ans ⇒ given $A = 40$

$N = 10$

SCM ⇒ $\bar{x} = A + \frac{\Sigma d}{N}$ where $\boxed{d = x - A}$

X	d = X − A
15	15 − 40 = −25
20	20 − 40 = −20
30	30 − 40 = −10
22	22 − 40 = −18
25	25 − 40 = −15
18	18 − 40 = −22
40	40 − 40 = 0
50	50 − 40 = 10
55	55 − 40 = 15
65	65 − 40 = 25
	$\Sigma d = -60$

$$\bar{X} = A + \left(\frac{\Sigma d}{N}\right)$$

$$\bar{X} = 40 + \left(\frac{-60}{10}\right)$$

$$\bar{X} = 40 - 6$$

$$\boxed{\bar{X} = 34}$$

* Discrete series

① Direct Method

$$\boxed{\bar{X} = \frac{\Sigma fX}{N}}$$

where $\boxed{N = \Sigma f}$

② Shortcut Method

$$\boxed{\bar{X} = A + \left(\frac{\Sigma fd}{N}\right)}$$

where $\boxed{d = X - A \quad N = \Sigma f}$

Q 10. Calculate the arithmetic mean for the following data.

Marks :	36	42	46	55	63	72
No of Stu :	4	5	9	10	8	4

Marks X	No of stud f	fX
36	4	144
42	5	210
46	9	414
55	10	550
63	8	504
72	4	288
	$\Sigma f = 40$	$\Sigma fx = 1080$

$$\bar{X} = \frac{\Sigma fX}{N}$$ where $N = \Sigma f$

$$\bar{X} = \frac{1080}{40}$$

$$\boxed{\bar{X} = 27}$$

(2) Compute the mean of the following data by short cut method. (2)

Height : 219, 216, 213, 210, 207, 204, 201, 198, 195

No of men : 2 4 6 10 11 7 5 4 1

Height X	No of men f	$d = X - A$	fd
219	2	12	24
216	4	9	36
213	6	6	36
210	10	3	30
207	11	0	0
204	7	−3	−21
201	5	−6	−30
198	4	−9	−36
195	1	−12	−12
			$\Sigma fd = 27$

$$\boxed{\bar{X} = A + \frac{\Sigma fd}{N}} \quad A = ?$$

$d = X - A$

$$A = \frac{\text{sum of } X}{\text{No of } X}$$

$$A = \frac{1863}{9} \quad \boxed{A = 207}$$

$N = \Sigma f$

$N = 50$

$$\bar{X} = 207 + \left(\frac{27}{50}\right)$$

$\bar{X} = 207 + 0.54$

$$\boxed{\bar{X} = 207.54}$$

* Continuous series :

① Direct Method

$$\boxed{\bar{X} = \frac{\Sigma fm}{N}}$$

where m is mid value

Ex = 0 - 10

$$\frac{0 + 10}{2} = 5$$

m = 5

$\boxed{N = \Sigma f}$

② Short cut Method

$$\boxed{\bar{X} = A + \left(\frac{\Sigma fd}{N}\right)}$$

where $\boxed{d = m - A}$

③ Step Deviation method

$$\boxed{\bar{X} = A + \frac{\Sigma fd'}{N} \times i}$$

$N = \Sigma f$

$\boxed{d' = \frac{d}{i}}$

~~i = lower interval~~

i = upper interval − lower interval

(19) Calculate the AM from the following data:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60
No of Stud given	4	7	15	8	3	3

Marks	f	m	fm
0-10	4	5	20
10-20	7	15	105
20-30	15	25	375
30-40	8	35	280
40-50	3	45	135
50-60	3	55	165
	$\Sigma f = N = 40$		$\Sigma fm = 1080$

direct M

$$\boxed{\bar{x} = \frac{\Sigma fm}{N}}$$

$$m = \frac{ll + ul}{2}$$

$$m = \frac{10+20}{2}$$

$$\boxed{m = 15}$$

$$\bar{x} = \frac{1080}{40}$$

$$\boxed{\bar{x} = 27}$$

(20) Calculate the mean of the following distribution from shortcut method.

class:	0-10	10-20	20-30	30-40	40-50	50-60
freque:	12	18	27	20	17	6

~~feature~~ class	freq (f)	m / x	d = m - A	fd
0-10	12	5	-30	-360
10-20	18	15	-20	-360
20-30	27	25	-10	-270
30-40	20	35	0	0
40-50	17	45	10	170
50-60	6	55	20	+120
	N = 100			

~~$\Sigma fd = 1180$~~

$$\boxed{\Sigma fd = -700}$$

$$\bar{x} = A + \left(\frac{\Sigma fd}{N}\right)$$

A = ? Σfd = ? N = ?

N = 10

$$A = \frac{\text{sum of } x}{\text{No of } x} = \frac{180}{6} = 30$$

A = 35

$$\bar{x} = 35 + \left(\frac{-700}{100}\right)$$

$$\bar{x} = 35 + (-7)$$

$$\boxed{\bar{x} = 28}$$

Q3) Calculate the A.M from the following data by (3) using step deviation method.

class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No of stude	12	18	27	20	17	6

class interval	No of stud f	m	d = x − A	d'	fd'
0 – 10	12	5	−30	−3	−36
10 – 20	18	15	−20	−2	−36
20 – 30	27	25	−10	−1	−27
30 – 40	20	35	0	0	0
40 – 50	17	45	10	1	17
50 – 60	6	55	20	2	12
	$\Sigma f = N = 100$				$\Sigma fd' = -70$

$$\boxed{\bar{X} = A + \left(\frac{\Sigma fd'}{N} \times i\right)}$$

$A = \frac{180}{6} = 30$

$\boxed{A = 35}$ upper value

$\boxed{N = 100}$

$i = 10 - 0 = 10$

$\boxed{i = 10}$

$\boxed{d' = \frac{d}{i}}$

$$\bar{X} = 35 + \left(\frac{-70}{100} \times 10\right)$$

$$\bar{X} = 35 + (-7)$$

$$\boxed{\bar{X} = 28}$$

* if in question if we found less than or more than them How to deal with them let's learn.

ans ⇒ Q Calculate the arithmetic mean from the following data:

marks	No of students
less than 10	5
Less than 20	17
less than 30	31
Less than 40	41
Less than 50	49

So create a new table for (X) and (f) for both

Interval	Marks	No of stud	f
0 – 10	less than 10	5	5
10 – 20	" 20	17	17 − 5 = 12
20 – 30	" 30	31	31 − 17 = 14
30 – 40	" 40	41	41 − 31 = 10
40 – 50	" 50	49	49 − 41 = 8

this is continuous series so apply either direct, shortcut or step deviation method to calculate

$\boxed{\bar{X} = \quad}$ practice by your self

Calcula...

(4)

Calculate the arithmetic mean form the following data:

Marks	No of students
More than 0	30
" 2	28
" 4	24
" 6	18
" 8	10

ans ⇒

Interval	Marks	No of stu	f
0 − 2	More than 0	30	30 − 28 = 2
2 − 4	" 2	28	28 − 24 = 4
4 − 6	" 4	24	24 − 18 = 6
6 − 8	" 6	18	18 − 10 = 8
8 − 10	" 8	10	→ 10 ⇒ 10

Now ~~you~~ this is continuous series so you can find $\bar{x}$ using direct, shortcut, step deviation method do by yourself (H.W)

* To find the missing value

* from the following data calculate the missing value when its mean is 115.86

wages	110	112	113	117		125	128	130
No of workers	25	17	13	15	14	8	6	2

wages X	No of workers f	fx
110	25	2750
112	17	1904
113	13	1469
117	15	1755
x	14	$14x$
125	8	1000
128	6	768
130	2	260
	$N = 100$	$\Sigma fx = 9906 + 14x$

given $\bar{X} = 115.86$

this discrete series

$$\boxed{\bar{X} = \frac{\Sigma fx}{N}}$$

$$115.86 = \frac{9906 + 14x}{100}$$

$$11586 = 9906 + 14x$$

$$11586 - 9906 = 14x$$

$$x = \frac{1680}{14}$$

$$\boxed{x = 120}$$

* find the missing frequency in the following frequency distribution table, it being given that the mean of this frequency is 50.

class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	Total
frequency	17	–	32	–	19	120

(5)

Interval	Frequency (f)	m	fm
0 - 20	17	10	170
20 - 40	f1	30	30f1
40 - 60	32	50	1600
60 - 80	f2	70	70f2
80 - 100	19	90	1710

this is continuous series

$$\boxed{\bar{X} = \frac{\sum fm}{N}}$$

given $\bar{x} = 50$

$N = 120$

$\sum fm = 170 + 30f1 + 1600 + 70f2 + 1710$

$$50 = \frac{170 + 30f1 + 1600 + 70f2 + 1710}{120}$$

$6000 = 170 + 1600 + 1710 + 30f1 + 70f2$

$6000 - 3480 = 30f1 + 70f2$

$2520 = 30f1 + 70f2$ — (2)

$\because 17 + f1 + 32 + f2 + 19 = 120$

$f1 + f2 = 120 - 68$

$f1 + f2 = 52$

$\boxed{f2 = 52 - f1}$ — (1)

put eq(1) into (2)

$2520 = 30f1 + 70(52 - f1)$

$2520 = 30f1 + 3640 - 70f1$

$2520 = 3640 - 40f1$

$40f1 = 3640 - 2520$

$40f1 = 1120$

$f1 = \frac{1120}{40}$

$\boxed{f1 = 28}$ so $f2 = 52 - 28$

$\boxed{f2 = 24}$

* Combine topic

$$\boxed{\bar{X} = \frac{N1\,\bar{X1} + N2\,\bar{X2}}{N1 + N2}}$$

① The mean wage of 1000 works in a paper mill, running two shifts of 600 and 400 workers is Rs 100. The mean wage of 600 workers in the ~~1st~~ shift is Rs 900. Find the mean wage of workers working in the 4th shift.

Ans ⇒ $N1 = 600$ $\quad \bar{X} = 1000$

$N2 = 400$ $\quad \bar{X_1} = 900$ $\quad \bar{X_2} = ?$

$$\bar{X} = \frac{N1\bar{X_1} + N2\bar{X_2}}{N1+N2}$$

$$1000 = \frac{600 \times 900 + 400 \times \bar{X_2}}{600+400}$$

$$1000 \times 1000 = \cancel{54}0000 + 400\bar{X_2}$$

$$1000000 - \cancel{54}0000 = 400\bar{X_2}$$

$$460000 = 400\bar{X_2}$$

$$\bar{X_2} = \frac{460000}{400} \quad \boxed{\cancel{\bar{X_2} = 1600}}$$

$$\boxed{\bar{X_2} = 1150}$$

* The average weight of 150 students in a class is 80 kg. The average weight of boys in the class is 85 kg and that of girls is 70 kg. find the no of boys and girls in the class separately.

ans ⇒ total no of student = 150 $\quad \boxed{\bar{X} = 80}$

let $\boxed{N1 = \text{boys}, \; N2 = \text{girls}}$ $\quad \boxed{\bar{X_1} = 85}$ $\quad \boxed{\bar{X_2} = 70}$ $\quad \boxed{N1 + N2 = 150}$ $\quad \boxed{N1 = 150 - N2}$

$$\bar{X} = \frac{N1 \times \bar{X_1} + N2\bar{X_2}}{N1+N2}$$

$$80 = \frac{(150 - N2)85 + N2 \times 70}{150}$$

$$12000 = 12750 - 85N2 + 70N2$$

$$12000 = 12750 - 15N2$$

$$15N2 = 12750 - 12000$$

$$15N2 = 750$$

$$N2 = \frac{750}{15} \quad \boxed{N2 = 50}$$

$$\boxed{\text{girls} = 50}$$

$$N1 + 50 = 150$$

$$N1 = 100$$

$$\boxed{\text{boys} = 100}$$

* ~~Weighted Arithmetic~~ mean

* Mean height of 25 male workers is 61 Cmm mean height of 35 female workers is 58 cm. find combined mean height of 60 workers.

ans ⇒ Male worker $N_1 = 25$

$\bar{X_1} = 61$

female workes $N_2 = 35$

$\bar{X_2} = 58$

$\bar{X} = ?$

$$\bar{X} = \frac{N_1\bar{X_1} + N_2\bar{X_2}}{N_1 + N_2} \qquad \bar{X} = \frac{25 \times 61 + 35 \times 58}{60}$$

$$\bar{X} = \frac{1525 + 2030}{60}$$

$$\bar{X} = \frac{3555}{60} \qquad \boxed{\bar{X} = 59.25}$$

* ~~Weighted A~~ Weighted Arithmetic mean

$$\boxed{\bar{X} = \frac{\Sigma fx}{N}} \quad \text{where } \boxed{N = \Sigma f}$$

items	81	76	74	58	70	73
weight	2	3	6	7	3	7

items (x)	f	fx
81	2	162
76	3	228
74	6	444
58	7	406
70	3	210
73	7	511
	N=28	Σfx=1961

$$\bar{X} = \frac{1961}{28}$$

$$\boxed{\bar{X} = 70.03}$$

(7)

Grm Mean:

correcting, incorrect values of mean

① $\boxed{\bar{X}_{correct} = \frac{\Sigma X_{correct}}{N}}$ ②

② $\boxed{\Sigma x_{correct} = \Sigma x_{incorrect} - \text{incorrect value} + \text{correct value}}$

③ $\boxed{\Sigma x_{incorrect} = \bar{X}_{incorrect} \times N}$ derived from $\boxed{\bar{X}_{in} = \frac{\Sigma x_{in}}{N}}$

* The mean marks of 100 students was found to be 40. later on, it was discovered that a score of 53 was misread as 83 find the correct mean.

ans ⇒ $N = 100$ $\bar{X} = 40$ $\bar{x} = \frac{\Sigma x}{N}$

correct value = 53 incorrect value = 83

③ $\boxed{\bar{X}_{incorrect} = \frac{\Sigma x_{incorrect}}{N}}$ $40 = \frac{\Sigma x_{incorrect}}{100}$

$\boxed{\Sigma x_{incorrect} = 4000}$

① $\boxed{\bar{X}_{correct} = \frac{\Sigma x_{correct}}{N}}$

$\Sigma x_{correct} = \Sigma x_{incorrect} -$

② $\boxed{\Sigma x_{correct} = \Sigma x_{incorrect} - \text{incorrect value} + \text{correct value}}$

$\Sigma x_{correct} = 4000 - 83 + 53$

$\Sigma x_{correct} = 4000 + 53 - 83 = 3970$

$\boxed{\Sigma x_{correct} = 3970}$

$\bar{X}_{correct} = \frac{3970}{100}$ $\boxed{\bar{X}_{correct} = 39.7}$

* Correcting incorrect values of the mean of 100 is 80 by mistake 1 item is misread 92 instead of 29 find the correct mean.

Given: $N = 100$ $\bar{x} = 80$

corrvalue $= 29$ incorrvalue $= 92$

$$\boxed{\bar{x}\text{Correct} = \frac{\Sigma x\,\text{Correct}}{N}}$$

$$\boxed{\Sigma x\,\text{Correct} = \Sigma x\,\text{incorrect} - \text{incorrectvalue} + \text{correctvalue}}$$

$$\bar{x}\text{incorr} = \frac{\Sigma x\,\text{incorr}}{N} \Rightarrow 80 = \frac{\Sigma x\,\text{incorr}}{100}$$

$$\boxed{\Sigma x\,\text{incorr} = 8000}$$

$$\boxed{\Sigma x\,\text{correct} = 8000 - 92 + 29}$$

$$\Sigma x\,\text{correct} = 7937$$

$$\bar{x}\text{correct} = \frac{7937}{100} \qquad \boxed{\bar{x}\text{correct} = 79.37}$$

* The mean of 5 observations is 7. later on, it was found that two observations 4 and 8 were wrongly taken instead of 5 and 9. find the correct mean.

ans ⇒ Given: $N = 5$, $\bar{x}_{in} = 7$

incorrval1 $= 4$, incorrval2 $= 8$, corrval1 $= 5$, corrval2 $= 9$

$$\boxed{\bar{x}\text{correct} = \frac{\Sigma x\,\text{correct}}{N}}$$

$\Sigma x\,\text{corr} = \Sigma x\,\text{incorr} - \text{incorrvalue} + \text{corrvalue}$

$$\therefore \bar{x}\text{incorr} = \frac{\Sigma x\,\text{incorr}}{N} \Rightarrow 7 = \frac{\Sigma x\,\text{incorr}}{5}$$

Σx in corr $= 35$ (8)

Σx corr $= 35 - 1 - 8 + 5 + 9$

$\boxed{\Sigma \text{corr} = 37}$ $\bar{x}$ corr $= \frac{\Sigma x \text{ corr}}{N} = \bar{x}$ corr $= \frac{37}{5}$

$\boxed{\bar{x} \text{ corr} = 7.4}$

PAS

Rank correlation (12 M)

(Spearman Rank correlation)

british ~~pyso~~ psychologist charles edward spearman in 1904 has invented rank correlation when the quantitative measures for certain factor such as leadership ability or female beauty can't be fixed but individual in group can be arranged in order to determined in the order to indicate rank in the group.

$$\boxed{R = 1 - \frac{6\Sigma D^2}{N^3 - N}}$$

where D is the diff b/w two rank (R1 - R2)

$D = R1 - R2$

$N \Rightarrow$ no of pair of observations

Q Judge x and y ranks are given

Judge (x) R1	Judge Y R2	R1 − R2 = D	D^2
8	7	1	1
7	5	2	4
6	4	2	4
3	1	2	4
2	3	−1	1
1	2	−1	1
5	6	−1	1
4	8	−4	16
		$\Sigma D = 0$	$\Sigma D^2 = 32$

$$\boxed{R = 1 - \frac{6\Sigma D^2}{N^3 - N}}$$

$$R = 1 - \frac{6 \times 32}{8^3 - 8}$$

$$R = 1 - \frac{192}{504}$$

$$R = \frac{504 - 192}{504}$$

$$R = \frac{312}{504}$$

$$\boxed{R = 0.619}$$

Q.. two ladies were asked to rank seven diff types of lipstick nilu and ~~mina~~ meena.

	Neelu	Meena	$R_1 - R_2 = D$	D^2
A	2	1	1	1
B	1	3	-2	4
C	4	2	2	4
D	3	4	1	1
E	5	5	0	0
F	7	6	1	1
G	6	7	-1	1

$\Sigma D^2 = 12$

$$\boxed{R = 1 - \frac{6 \times 12}{7^3 - 7}}$$

X and Y are value (not rank)

\# when Ranks are not given:

X	Y	R_1	R_2	$R_1 - R_2 = D$	D^2
15	18	4	1	3	9
17	12	2	2	0	0
14	4	5	8	-3	9
13	6	6	6	0	0
11	7	8	5	3	9
12	9	7	4	3	9
16	3	3	9	-6	36
18	10	1	3	-2	4
10	2	9	10	-1	1
9	5	10	7	3	9

$\Sigma D^2 = 36 + 36 + 9 + 5 = 72 + 14 = 86$

$R = 1 - \frac{6 \times 86}{10^3 - 10}$ $R = 1 - \frac{516}{1000 - 10}$ $R = 1 - \frac{516}{990}$

$R = 1 - \frac{516}{990}$ $R = 1 - 0.52$

$\boxed{R = 0.48}$

When equal ranks are there :

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\ldots\right]}{N^3 - N}$$

$N \rightarrow$ No of Items ~~of repeated element~~

where m is no of items of equal ranks

~~the cor~~ $\boxed{\frac{1}{12}(m^3 - m) \rightarrow \text{the correlation factor}}$

Q. ~~correlation~~ calculate the coff of

X	Y	R_1	R_2	D	D^2
15	16	5	2	3	9
10	14	7.5	4	3.5	12.25
20	10	2	8	-6	36
28	12	1	5.5	-4.5	20.25
12	11	6	7	-1	1
10	15	7.5	3	4.5	20.25
16	18	4	1	3	9
18	12	3	5.5	-2.5	6.25

$\boxed{\Sigma D = 0}$ $\Sigma D^2 = 114$

$$\frac{1}{12}(2^3 - 2) = \frac{1}{12}(8-2) = \frac{1}{12} \times 6 = \frac{1}{2} = 0.5$$

$$R = 1 - \frac{6[114 + 0.5 + 0.5]}{8^3 - 8}$$ ~~$R = 1 - \frac{6[114+1]}{6}$~~

$\boxed{-0.369}$

$$R = 1 - \frac{6(115)}{504} \quad R = \frac{504 - 690}{504} \quad R = \frac{-186}{504}$$ $\boxed{R = -0.369}$

* calculate coeff of ~~correla~~ corelation from following :

X	Y	R1	R2	D=R1−R2	D^2
40	80	8	8	0	0
50	120	6.5	7	−0.5	0.25
60	160	4	4	0	0
80	130	1	5.5	−4.5	20.25
50	200	6.5	2	4.5	20.25
70	210	2	1	1	1
60	130	4	5.5	−1.5	2.25
60	170	4	3	1	1

$\Sigma D = 0$ $\quad \Sigma D^2 = 45$

ans ⇒ 0.429

$\Sigma D^2 = 45, \; m_1 = 3, \; m_2 = 2, \; m_3 = 2 \; N = 8$

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right]}{(N^3 - N)}$$

by substitution

$$R = 1 - \frac{6\left[45 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{8^3 - 8}$$

$$R = 1 - \frac{6[(45 + 2 + 0.5 + 0.5)]}{512 - 8} \Rightarrow 1 - \frac{6(48)}{504}$$

$$R = 1 - \frac{288}{504}$$

$$R = \frac{504 - 288}{504} \qquad R = \frac{216}{504}$$

$$\boxed{R = 0.429}$$ answer

~~Regration~~ regression

- Mutual relationship b/w two series is masured with the help of correlation., direction and magnitude of the relationship between two variables
It tells ~~us~~ best astimate of the value of dependent and independent variable

Ex-) Demand — Dependent; Price — independent

* Applications of regressions (HW).

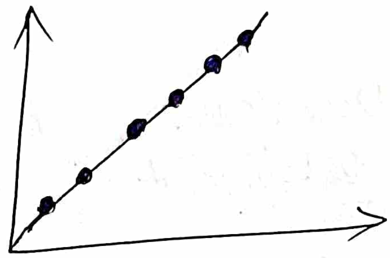
Q diff b/w regression and correlation.

~~Regression~~ correlation	~~correlation~~ Regression

regression coefficient & regression equation

· regression lines shows the avg relationship b/w the two variable this is also known as line of best fit.
on the basis of regression line we can predict the value of a dependent varible on the basis of independent variable.

Line of best fit



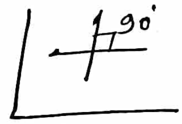
Two types of regression
(i) regression line of x on y
(ii) regression line of y on x.

(i) regression line of x on y:

x on y means given y, find x.

y on x means given x, find y.

The regression lines are coincident or they will be only one regression line when $r = \pm 1$
when $r=0$, regression lines are parallel,



* Regression eqⁿ: regression eqⁿ are the algebric relation of regression

Regression equation:
(i) regression equation y on x:

$y = a + bx$ Here a & b are constants.

~~$y - \bar{y} = r\,\sigma(x - \bar{x})$~~

Regression eqⁿ $\boxed{y - \bar{y} = \frac{r \cdot \sigma y (x - \bar{x})}{\sigma x}}$

Coefficient of regression eqⁿ $\boxed{(y - \bar{y}) = byx\,(x - \bar{x})}$

$\boxed{byx = \frac{r \sigma y}{\sigma x}}$

σy: S.D of y, r → karl pearson's
σx: S.D of x coeff of correlation

(ii) ~~X reg eqn on Y of~~ Regression eqn of x on y:

$x = a + by$

$$\boxed{x - \bar{x} = r\frac{\sigma x}{\sigma y}(y - \bar{y})} \quad \boxed{bxy = r\frac{\sigma x}{\sigma y}}$$

$$\boxed{x - \bar{x} = bxy(y - \bar{y})} \quad \boxed{r = \sqrt{bxy \times byx}}$$

* properties of regression (at least 4 properties)

(i) the coefficient of correlation is geometric mean of regression coefficient

Q (a) Regression equation of x on y from the following data by method of ~~least~~ square.

Sol^n:

x	y	xy	y^2	x^2
1	2	2	4	1
2	5	10	10	4
3	3	9	9	9
4	8	32	64	16
5	7	35	49	25

$x = a + by$

$$\boxed{\begin{aligned}\Sigma x &= Na + b\Sigma y \\ \Sigma xy &= a\Sigma y + b\Sigma y^2\end{aligned}}$$

$\Sigma x = 15$, $N = 5$, $\Sigma y = 25$

$\Sigma xy = 88$, $\Sigma y^2 = 151$

regression eq^n of x on y

$15 = 5a + b25$ —— (1) $\times 5$

$88 = 25a + b151$ —— (2)

$75 = 25a + b125$ —— (3)

$88 = 25a + b151$ —— (2)

subtract the eq^n 3 from (2)

$\rightarrow$ $88 = \cancel{25a} + b151$

$75 = \cancel{25a} + b125$

$13 = b26$

$b = \frac{13}{26}$ $\boxed{b = 0.5}$

put in eq^n (1)

$15 = 5a + 12.5$

$5a = 2.5$

$\boxed{a = 0.5}$

$\boxed{x = \frac{1}{2} + \frac{1}{2}y}$

regression eq^n of y on x

$y = a + bx$

$$\boxed{\begin{aligned}\Sigma y &= Na + b\Sigma x \quad (1) \\ \Sigma yx &= a\Sigma x + b\Sigma x^2 \quad (2)\end{aligned}}$$

$\Sigma x^2 = 55$, $N = 5$

$\Sigma y = 25$

$\Sigma x = 15$

$25 = 5a + 15b$ —— ① $\times 3$

$88 = 15a + 55b$ —— ②

$75 = 15a + 45b$ —— ③

$88 = 15a + 55b$ —— ②'

subtract eq ③ from eq^n ②

$88 = 15a + 55b$

$75 = 15a + 45b$

$13 = 10b$

$\boxed{b = 1.3}$ put in eq^n ①

$25 = 5a + 15 \times 1.3$

$25 = 5a + 19.5$

$5a = 25 - 19.5$

$\boxed{5a = 5.5}$

$a = \frac{5.5}{5}$ $\boxed{a = 1.1}$

$\boxed{y = 1.1 + 1.3x}$

Q ~~Regression~~ $N = 8$, $\Sigma x = 21$, $\Sigma x^2 = 99$, $\Sigma y = 4$, $\Sigma y^2 = 68$

$\Sigma xy = 36$ using the values, find

(i) RE of y on x

(ii) RE of x on y

(3) Most approx val of y when $x=10$

(4) Most approx val of x when $y = 2.5$

Sol^n: ~~RE using Reg Cofficient~~

~~→ y on x~~

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

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

~~$b_{yx} = \frac{N\Sigma xy - (\Sigma x \cdot \Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$~~

~~$b_{yx} = \frac{(5 \times 88) - (15 \times 25)}{[5(55)] - [15]^2}$~~

~~$\boxed{b_{yx} = 1.3}$~~

① Regression equation of y on x:

$y = a + bx$

$$\boxed{b_{yx} = \frac{N \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{N \cdot \Sigma X^2 - (\Sigma X)^2}}$$

$= \frac{8 \times 36 - 4 \times 21}{8 \times 99 - (21)^2}$ $\boxed{b_{yx} = 0.581}$

~~$= \frac{440 - 375}{275 - 225} = \frac{65}{50}$~~

~~A~~ regression equation using regression coefficient

y on x

$\cancel{y-\bar{y} = byx\cdot(x-\bar{x})}$

$\cancel{y = \bar{y} + byx(x-\bar{x})}$

$\cancel{byx = \frac{N\cdot\Sigma XY - \Sigma X\Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}}$

$\cancel{byx = 5\times 88 - 15\times 25}$

$\bar{X} = \frac{\Sigma X}{N} \quad \bar{X} = \frac{21}{8} \quad \boxed{\bar{X} = 2.625}$

$\bar{y} = \frac{\Sigma Y}{N} = \frac{4}{8} \quad \boxed{\bar{y} = 0.5}$

$\boxed{a = \bar{y} - b\bar{x}}$

$a = 0.5 - (0.581)(2.625)$

$\boxed{a = -1.025}$

$\boxed{y = -1.025 + 0.581X}$

(ii) Regression Equation of x on y: $X = a + by$

$\boxed{bxy = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}} = \frac{8\times 36 - (21)(4)}{8(68) - (4)^2} = 0.386$

$a = \bar{x} - b\bar{y}$

$a = 2.625 - (0.386)(0.5)$

$\boxed{a = 2.432}$

$\boxed{X = 2.432 + 0.386y}$

(iii) Prediction for $y \Rightarrow$ when $x = 10$

$\therefore y = -1.025 + 0.581X$

$y = -1.025 + 0.581(10)$

$y = -1.025 + 5.81 \quad \boxed{y = 4.785}$

(iv) Prediction for $x \Rightarrow$ when $y = 2.5$

$\therefore X = 2.432 + 0.386(y)$

$X = 2.432 + 0.386(2.5) \quad \boxed{x = 3.397}$

(ii) Using Deviations taken from Actual Means.

Regression Equation of x on y:

$$\boxed{X - \bar{X} = b_{xy}(Y - \bar{Y})}$$

$$\boxed{b_{xy} = \frac{\Sigma xy}{\Sigma y^2}}$$ → Regression coefficient of x on y

Regression Equation of y on x:

$$\boxed{Y - \bar{Y} = b_{yx}(X - \bar{X})}$$

Regression coefficient of y on x

$$\boxed{b_{yx} = \frac{\Sigma xy}{\Sigma x^2}}$$

Ex: obtain the two regression equations from the following data

X	2	4	6	8	10	12
Y	4	2	5	10	3	6

Solⁿ

X	x $(X-\bar{X})$	x^2	Y	y $(Y-\bar{Y})$	y^2	xy
2	-5	25	4	-1	1	5
4	-3	9	2	-3	9	9
6	-1	1	5	0	0	0
8	+1	1	10	5	25	5
10	+3	9	3	-2	4	-6
12	+5	25	6	1	1	5
$\Sigma X = 42$, $N = 6$	$\Sigma x = 0$	$\Sigma x^2 = 70$	$\Sigma Y = 30$	$\Sigma y = 0$	$\Sigma y^2 = 40$	$\Sigma xy = 18$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{42}{6}$$

$$\boxed{\bar{X} = 7}$$

$$\boxed{\Sigma x = 0}$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{30}{6} = 5$$

$$\boxed{\bar{Y} = 5}$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{18}{70}$$

$$\boxed{b_{yx} = 0.257}$$

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{18}{40} = 0.45$$

$$\boxed{b_{xy} = 0.45}$$

Regression Equation of y on x

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

$$y-5 = 0.257(x-7)$$

$$y-5 = 0.257x - 1.799$$

$$\boxed{y = 0.257x + 3.201}$$

Regression eq^n of x on y

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

$$x-7 = 0.45(y-5)$$

$$x-7 = 0.45y - 2.25$$

$$\boxed{x = 0.45y + 4.75}$$

(pyq) The following are the intermediate results of the two series x and y: $\bar{x} = 90$, $\bar{y} = 70$, $N = 10$ $\Sigma x^2 = 6360$ $\Sigma y^2 = 2860$ $\Sigma xy = 3900$ (where x and y are deviations from the respective means) find two regression equations.

Sol^n: Regression coefficient of y on x:

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{3900}{6360} \quad \boxed{b_{yx} = 0.613}$$

Regression coefficient of x on y

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{3900}{2860} = 1.363$$

Regression Equations of x on y

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

$$x - 90 = 1.363(y-70)$$

$$x = 1.363y - 95.41 + 90$$

$$x = 1.363y - 5.41 \quad \boxed{x = 1.363y - 5.41}$$

Regression eq^n of y on x

$$y-\bar{y} = b_{yx}(x-\bar{x}) \Rightarrow y-70 = 0.613x(x-90) \Rightarrow y-70 = 0.613x - 55.17$$

$$\boxed{y = 0.613x + 14.83}$$

probability

(i) Experiment : when we conduct a trial to obtain some statistical information called experiment.
Ex ⇒ • Tossing of a fair coin is an experiment.
• rolling a fair die is an experiment.

(ii) Event : The possible outcomes of trial/experiment are called events. Ex⇒ if fair coin is tossed, the outcomes : head or tail are called events.

(iii) Exhaustive event : The total no of possible outcome of a trial/experiment is called exhaustive event.
(1, 2, 3, 4, 5, 6) × (1, 2, 3, 4, 5, 6)
outcome = 36 → Exhaustive event.

(iv) Equally-Likely event : The events are said to be Equally likely, happening ~~each~~ each event is equal.
Ex⇒ in coin toss the event H and T are equally-likely events.

(v) Mutually Exclusive event : Mutually exclusive are those events ~~those event~~ when they cannot happen simultaneously in a single trial.
Ex⇒ in toss coin events Head and tail are mutually exclusive because they cannot happen simultaneously in single trial.

(vi) Complementary event : Two events A and B are mutually exclusive and exhaustive event. ce when both are. Ex⇒ in tossing a coin, occurrence of head and tail are complementary events.

(vii) simple and compound event : The probability of happening or not happening of single events.
Ex⇒ If two coins are tossed simultaneously and we shall be finding the probability of getting two heads then we are dealing with compound events.

(viii) Independent event : those who are not dependent on each other. Ex⇒ Coin a flip, tossing a die

(x) dependent event : ~~if a card is drawn~~
when the occurrence of one does affect the probability of the occurrence of the other events.

Ex ⇒ if a card is drawn from a pack of 52 playing cards and is not replaced this will affect the probability of the second card being drawn.

(ii) the probability of drawing a king from a pack of 52 cards is 4/52 or $\frac{1}{13}$ but if the card drawn (king) is not replaced in the pack, the probability of drawing again a king is $\frac{3}{51}$

* Probability: [redacted]

$$\boxed{P(p) = \frac{m}{x}}$$

~~(1) find the probability of tossing a coin~~

~~$n = 2$~~

~~$m = 1$~~

~~$\therefore p = \frac{1}{2}$~~

~~(ii)~~ [redacted] ~~$\frac{52}{2}$~~

Probability : probability is the ratio of the favourable cases to the total number of equally likely cases.

$$P(A) = p = \frac{\text{Number of favourable cases}}{\text{Total no of equally likely case}} = \frac{m}{n}$$

$\boxed{P(A) = \frac{m}{n}}$ $\boxed{P(A) = \frac{m}{n}}$ where m: No of favourable case
where n: Total no of equally likely cases.

Q.
Ex ⇒ find the probability of getting a head in a tossing of a coin.

ans ⇒ when a coin is tossed, there are two possible outcomes Head or Tail.

Total no of equally likely cases ⇒ $n = 2$
Number of ~~cases~~ favourable cases ⇒ $m = 1$

$\therefore P(H) = \frac{m}{n} = \frac{1}{2}$

any other example from this topic refer the book from page 6 (Probab)

MST

PYQ A Man to marry a girl having qualities: white complexion - the probability of getting such a girl is one in twenty; handsome dowry - the probability of getting this one in fifty; westernised manners and etiquettes - the probability here is one in hundred. find out the probability of his getting married to such a girl when the possession of these attributes is independent.

ans: probability of getting a girl with white completion $= P(A) = \frac{1}{20} = 0.05$

probability of getting a girl with handsome dowry

$\Rightarrow P(B) = \frac{1}{50} = 0.02$

probability of getting a girl with westernised manner $\Rightarrow P(C) = \frac{1}{100} = 0.01$

Since the events are independent of the simultaneous occurence of all these qualities is:

$$P(ABC) = P(A) \times P(B) \times P(C)$$

$$= \frac{1}{20} \times \frac{1}{50} \times \frac{1}{100}$$

$$= 0.05 \times 0.02 \times 0.01$$

$$= 0.00001$$

Py) Final paper

Q. A pack of 50 tickets numbered 1 to 50 is shuffled and then two tickets are drawn find the probability that:

(a) Both the tickets drawn have prime numbers
(b) None of ~~them~~ the tickets drawn has prime numbers

(a) No of prime no b/w 1 to 50 : 15
So prob of getting a prime no in the first drawn $= \frac{15}{50}$
The no of remaining prime no are $= 14$
total lottery tickets remaining are $= 49$
then the prob of getting the prime no $= \frac{14}{49}$
So the pro of getting prime no in two drawn $= \frac{15}{50} \times \frac{14}{49} = \frac{3}{35}$

(b) No of non-prime no are $\Rightarrow 50 - 15 = 35$
pro of drawing a non-prime no for first ticket $= \frac{35}{50} = \frac{7}{10}$
After drawing the first non-prime number there are 49 tickets left and 34 non-prime no left
Again pro of drawing a non-prime no for second ticket $= \frac{34}{49}$
There for, pro of drawing non-prime no for both tickets $= \frac{35}{50} \times \frac{34}{49}$

$$= \frac{\cancel{7}}{\cancel{10}_{5}} \times \frac{\cancel{34}^{17}}{\cancel{49}_{7}} = \frac{17}{35}$$

(PYQ) A speaks truth 80% cases, B in 90% cases In what percentage of cases are they likely to contradict each other in starting the same fact.

Soln: $P(A) = \frac{80}{100} = \frac{4}{5}$ $P(B) = \frac{90}{100} = \frac{9}{10}$

P(A) & P(B) prob of speak the truth

$P(\bar{A})$ & $P(\bar{B})$ be the pro of speak lie

$P(\bar{A}) = 1 - P(A) = 1 - \frac{4}{5} = \frac{1}{5}$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{9}{10} = \frac{1}{10}$

They will contradict each other

(i) A speaks truth and B speaks lie

(ii) B speaks truth and A tell a lie

(i) $\frac{4}{5} \times \frac{1}{10} = \frac{4}{50}$ (ii) $\frac{9}{10} \times \frac{1}{5} = \frac{9}{50}$

(PYQ) The chances of Survival after 25 years are 3 out of 10 for a ~~woman~~ man and 4 out of 10 for a woman construct the probability that

(i) Both will be alive after 25 years

(ii) At least one will alive after 25 years.

(i) prob(Both alive) = prob(man alive) × prob(women alive)

$$\Rightarrow \frac{3}{10} \times \frac{4}{10} = \frac{12}{100} = 0.12$$

(ii) prob(at least one alive) = 1 − prob(neither alive)

prob(neither alive) = prob(man dead) × prob(women dead)

$$= \left(1 - \frac{3}{10}\right) \times \left(1 - \frac{4}{10}\right)$$

$$\Rightarrow \frac{7}{10} \times \frac{6}{10} = \frac{42}{100} = 0.42$$

So, prob(at least one alive) = $1 - 0.42 = \underline{\underline{0.58}}$

Bayes' theorem (12 Marks)

Inverse probability theorem

Bayes theorem:

Exampl 103

Q. A manufacturing firm or a bolt factory machine A, B, C manufactures 25%, 35%, and 40% of the total of their output 5, 4, 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine C?

given info

$P(A) = 25\%$

$P(B) = 35\%$

$P(C) = 40\%$

the conditional probability are

$P(D/A) = 5\%$

$P(D/B) = 4\%$

$P(D/C) = 2\%$

Since:

$$\boxed{P(D/A) = \frac{P(AD)}{P(A)}}$$

$$\boxed{P(C/D) = ?}$$

Sol^n: let A, B, C be the events of drawing a bolt produced by machine A, B, C respectively and let D is the event that bolt is defective.

Events (1)	Prior probability (2)	Conditional probability (3)	Joint probability col (2) x (3)
A	$P(A) = \frac{25}{100} = 0.25$	$P(D/A) = \frac{5}{100} = 0.05$	$0.25 \times 0.05 =$
B	$P(B) = \frac{35}{100} = 0.35$	$P(D/B) = 0.04$	$0.35 \times 0.04 =$
C	$P(C) = \frac{40}{100} = 0.40$	$P(D/C) = 0.02$	$0.40 \times 0.02 =$

the probability that a bolt is drawn at random item and is found to be defective that it was manufactured by machine C.

$$P(C/D) = \frac{\text{Joint probability of the machine C}}{\text{sum of Joint probability of all machines}}$$

$$P(C/D) = \frac{0.40 \times 0.02}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = 0.2318 = 23.18\%$$

Q. A company has two plants to manufacture scooters. Plant I manufactures 70% of the scooters and plant II manufactures 30%. At plant I, 80% scooters are rated standard quality and at plant II, 90% of the scooters are rated standard quality. A scooter is picked up at random and is found to be standard quality. What is the chance that is comes from plant I?

$P(A) = 70\%$

$P(B) =$

Let A_1, A_2 be the event that the selected scooters is manufactured by the plant 1 and two and let B be the event that the scooter is of standard quality.

We are given info

$P(A_1) = 70\% = \frac{70}{100}$

$P(A_2) = 30\% = \frac{30}{100}$

Conditional probability are

$P(B/A_1) = 80\% = \frac{80}{100}$

$P(B/A_2) = 90\% = \frac{90}{100}$

Events	Prior Probability (1)	Conditional Probability (2)	Joint probability col (1)(2)
A1	$P(A_1) = \frac{70}{100}$	$P(B/A_1) = \frac{80}{100}$	$\frac{70}{100} \times \frac{80}{100}$
A2	$P(A_2) = \frac{30}{100}$	$P(B/A_2) = \frac{90}{100}$	$\frac{30}{100} \times \frac{90}{100}$

The probability that the standard quality scooter was produced by plant I.

$P(A_1/B) = \frac{\text{joint probability of the 1st plant}}{\text{sum of joint probability of Both plants}}$

$= \frac{\frac{70}{100} \times \frac{80}{100}}{\frac{70}{100} \times \frac{80}{100} + \frac{30}{100} \times \frac{90}{100}} = \frac{70 \times 80}{70 \times 80 + 30 \times 90} = \frac{56}{83} =$

Q An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probability of accident is 0.01, 0.03 and 0.15 respectively. One of the insured person meets an accident. What is the probability that he is a scooter driver?

ans ⇒ Let A_1, A_2 and A_3 be the events that the insured person is a scooter driver, car driver, truck driver respectively and let B the event that is insured person meets an accident.

~~Let A_1, A_2, A_3 be the events that the injured person either scooter driver, car driver or truck driver.~~

$\boxed{0.0191}$

We are given the information

$P(A_1) = \frac{2000}{12000} = \frac{1}{6}$

$P(A_2) = \frac{4000}{12000} = \frac{1}{3}$

$P(A_3) = \frac{6000}{12000} = \frac{1}{2}$

conditional probability

$P(B/A_1) = 0.01 = \frac{1}{100}$

$P(B/A_2) = 0.03 = \frac{3}{100}$

$P(B/A_3) = 0.15 = \frac{15}{100}$

putting the information in the table given below:

Events	prior probability (2)	conditional proba (3)	joint probability col (2) × (3)
A1	$P(A_1) = 1/6$	$P(B/A_1) = \frac{1}{100}$	$\frac{1}{6} \times \frac{1}{100}$
A2	$P(A_2) = 1/3$	$P(B/A_2) = \frac{3}{100}$	$\frac{1}{3} \times \frac{3}{100}$
A3	$P(A_3) = 1/2$	$P(B/A_2) = \frac{15}{100}$	$\frac{1}{2} \times \frac{15}{100}$

the probability of a scooter driver the meets an accident.

$$P(A_1/B) = \frac{\frac{1}{6} \times \frac{1}{100}}{\frac{1}{6} \times \frac{1}{100} + \frac{1}{3} \times \frac{3}{100} + \frac{1}{2} \times \frac{15}{100}} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3} \times 3 + \frac{1}{2} \times 15} = \frac{0.166}{0.166 + 1 + 7.5}$$

$$= \frac{0.166}{8.666} = 0.0191 = 1.91\%$$

Q. there are two urns one contain 4 white, 6 red balls. one of the urn is selected at random and the ball is drawn from each and is found to be white what is probability that is (Ex = 109 · page - 39) must see

(P.20) A, B and C are three candidates for the post of Director in a company. Their respective chances of selection are in the ratio of 4:5:3. The probability that A, if selected will introduce the internet trading in the company. ~~find the probability that the company will introduce internet trading. Also find the probability that Director B introduced the trading in the company.~~ is 0.30. similarly, the probability of B and C are 0.50 and 0.60 respectively find the probability that the company will introduce internet trading. Also find the probability that Director B introduce & the internet trading in the company.

Ans ⇒ let A_1, A_2, and A_3 be the events that the persons A, B and C respectively are of Director of the company and let E be the event of introducing Internet trading in company. ~~Therefore~~ we are given info

$P(A_1) = \frac{4}{12}$

$P(A_2) = \frac{5}{12}$

$P(A_3) = \frac{3}{4+5+3} = \frac{3}{12}$

Conditional probability

$P(E/A_1) = 0.30$

$P(E/A_2) = 0.50$

$P(E/A_3) = 0.60$

Events	prior probability (2)	conditional probability (3)	joint probability cols (2) × (3)
A1	$P(A_1) = \frac{4}{12}$	$P(E/A_1) = 0.30$	$\frac{4}{12} \times 0.30$
A2	$P(A_2) = \frac{5}{12}$	$P(E/A_2) = 0.50$	$\frac{5}{12} \times 0.50$
A3	$P(A_3) = \frac{3}{12}$	$P(E/A_3) = 0.60$	$\frac{3}{12} \times 0.60$

(1) Now, P(internet trading is introduced in the company)

$P(E) = P(A_1E \text{ or } A_2E \text{ or } A_3E) = P(A_1E) + P(A_2E) + P(A_3E)$

$P(E) = P(A_1) \cdot P(E/A_1) + P(A_2) \cdot P(E/A_2) + P(A_3) \cdot P(E/A_3)$

$P(E) = \frac{4}{12} \times 0.30 + \frac{5}{12} \times 0.50 + \frac{3}{12} \times 0.60 = \cancel{\frac{55}{120}} = \frac{11}{24}$

we have to find P(A2/E) or internet trading is introducing by Director B is:

$$P(A_2|E) = \frac{\text{joint Probability of the end}}{\text{sum of the joint probability}}$$

$$= \frac{\frac{5}{12} \times 0.50}{\frac{4}{12} \times 0.30 + \frac{5}{12} \times 0.50 + \frac{3}{12} \times 0.60} = \frac{\frac{5}{12} \times \frac{1}{2}}{\frac{11}{24}} = \frac{5}{11}$$

Q There are two urns. Urn 1 contains 1 white and 6 red balls. Urn II has 4 white and 3 red balls. one of the urn is selected at random and a ball is drawn from it and found to be white, what is the probability that it is drawn from the 1st urn?

Soln: Let A_1, and A_2 stand for the events urn 1 is choosen and urn II is choosen respectively and let W stand for the event that white ball is chosen.

Thus we are given:

$P(A_1) = \frac{1}{2}$

$P(A_2) = \frac{1}{2}$

Conditional prob

$P(W/A_1) = \frac{1}{7}$

$P(W/A_2) = \frac{4}{7}$

Events (1)	prior probab (2)	conditional probab (3)	joint Probability col (2) × (3)
A_1	$P(A_1) = \frac{1}{2}$	$P(W/A_1) = \frac{1}{7}$	$\frac{1}{2} \times \frac{1}{7}$
A_2	$P(A_2) = \frac{1}{2}$	$P(W/A_2) = \frac{4}{7}$	$\frac{1}{2} \times \frac{4}{7}$

the probability that the white ball comes from urn 1 using Baye's Theorem.

$$P(A_1/W) = \frac{\text{joint probability of the 1st urn}}{\text{sum of joint probability of two urns}}$$

$$= \frac{\frac{1}{2} \times \frac{1}{7}}{\frac{1}{2} \times \frac{1}{7} + \frac{1}{2} \times \frac{4}{7}} = \frac{\frac{1}{14}}{\frac{5}{14}} \times \frac{1}{5}$$

Unit -

frequency Distribution

Expected or theoretical | observed, experiment

probability Distribution

- Discrete Pro. Dis
  - Binomial
  - Position
- Continue Pro. Dis
  - Normal

Binomial distribution : It is useful when an

Binomial Dist

- 1 → Successive
- 0 → failure

r = no of r success

2 Marks

$$p(x=r) = {}^{n}C_{r} \cdot \underbrace{p^{r}}_{\text{success}} \cdot \underbrace{q^{n-r}}_{\text{failure}}$$

$$P(x=0) = {}^{n}C_{0} \cdot p^{0} \cdot q^{n-0}$$

$$\frac{n!}{0!\, n!} \cdot q^{n} \cdot \quad \{0! = 1\}$$

$$\boxed{p(x) = q^{n}}$$

when $p(x=n) = {}^{n}C_{x} \cdot p^{x} \cdot q^{a} = \frac{n!}{n!} \cdot p^{x} \cdot 1$

$\cdot P(x=n) =$

Q a fair coin is tossed thrised exactly 2 ~~head~~ head at most 2 ~~[illegible]~~ head.

Sol $= P(2H) = {}^{n}C_{r}\, p^{r} q^{n-r}$

$= {}^{3}C_{2} \left(\frac{1}{2}\right)^{2} \cdot \left(\frac{1}{2}\right)^{3-2}$

$= \frac{3!}{2! \times 1!} \times \frac{1}{2^2} \times \frac{1}{2} = \frac{3}{8}$

$\boxed{\begin{array}{l} \text{at least} \Rightarrow \frac{1}{2} \\ \text{at most } 2 \Rightarrow \frac{7}{8} \end{array}}$

Q. It is observed that 80% of telivision viewes watch "sas bhi kabhi bahu thi". (0.737)

Probability Distribution

Types of prob distribⁿ

- Discrete pro disⁿ
  - Binomial Disⁿ
  - Poisson Distⁿ
- Continuous pro disⁿ
  - Normal Distⁿ

Binomial Distribution:

$$\boxed{P(X=x) = {}^nC_x \cdot q^{n-x} \cdot p^x}$$

where $p \to$ prob of success
q = prob of failure $(1-p)$
n = no of trials

In Binomial Distⁿ, we can obtain the prob $0, 1, 2, \ldots n$ successes as follows:

Number of success (x)	Prob of success $P(X=x)$
0	${}^nC_0 \cdot q^{n-0} \cdot p^0 = q^n$
1	${}^nC_1 \cdot q^{n-1} \cdot p^1 = n \cdot q^{n-1} \cdot p^1$
2	${}^nC_2 \, q^{n-2} \, p^2 = \frac{n(n-1)}{2 \times 1} q^{n-2} \cdot p^2$
x	${}^nC_x \, q^{n-x} \cdot p^x$
⋮	
n	${}^nC_n \cdot q^{n-n} \cdot p^n = p^n$

Q A fair coin is tossed thrice (i) exactly 2 head (ii) at most 2 head (iii) At least 2 Head

D!!

Sol^n: $n = 3$ $\quad$ $P(H) = \frac{1}{2}$ (pro of heads)

$r = 2$ $\quad$ $q(H) = 1 - \frac{1}{2} = \frac{1}{2}$ (pro of tail)

$\because {}^{n}C_{r} = \frac{n!}{r!(n-r)!}$

(i) exactly 2 head

$$p(2H) = {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r} = {}^{3}C_{2}\, p^{2} \cdot q^{3-2}$$

$$= {}^{3}C_{2}\, p^{2} \cdot q^{1}$$

$$= \frac{3!}{2!(3-2)!} \cdot p^{2} \cdot q^{1} \Rightarrow \frac{\cancel{6}\,3}{\cancel{2} \times 1} \times \left(\frac{1}{2}\right)^{2} \cdot \left(\frac{1}{2}\right)^{1}$$

$$= 3 \times \frac{1}{4} \times \frac{1}{2} = \frac{3}{8}$$

(ii) P(at least 2 Heads) = P(2H) + P(3H)

$$= {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r}$$

$$\Rightarrow {}^{3}C_{2} \cdot p^{2} \cdot q^{3-2} + {}^{3}C_{3} \cdot p^{3} \cdot q^{3-3}$$

$$\Rightarrow \frac{3!}{2!(3-2)!} \cdot \left(\frac{1}{2}\right)^{2} + \left(\frac{1}{2}\right)^{3-2} + \frac{3!}{3!(3-3)!} \times \left(\frac{1}{2}\right)^{3} \cdot \left(\frac{1}{2}\right)^{3-3}$$

$$\Rightarrow \frac{3 \times \cancel{2}}{\cancel{2} \times 1} \cdot \frac{1}{4} \times \frac{1}{2} + \frac{\cancel{3 \times 2}}{\cancel{3 \times 2}} \times \frac{1}{8} \times 1$$

$$\frac{3}{8} + \frac{1}{8} = \frac{\cancel{4}}{\cancel{8}} = \frac{1}{2}$$

(iii) P(at most 2 Heads) = P(0H) + P(1H) + P(2H)

$${}^{n}C_{r}\, p^{r} \cdot q^{n-r}$$

$$\Rightarrow \left({}^{3}C_{0}\, p^{0} q^{3-0}\right) + \left({}^{3}C_{1}\, p^{1} q^{3-1}\right) + \left({}^{3}C_{2}\, p^{2} q^{3-2}\right)$$

$$\frac{3!}{0!(3-0)!}\times\left(\frac{1}{2}\right)^0\times\left(\frac{1}{2}\right)^3+\frac{3!}{1!(3-1)!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2+\frac{3!}{2!(1!)}\left(\frac{1}{2}\right)^2\times\left(\frac{1}{2}\right)$$

$$\frac{\cancel{6}}{\cancel{6}}\times 1\times\frac{1}{8}+\frac{\cancel{6}\,3}{\cancel{2}}\times\frac{1}{2}\times\frac{1}{4}+\frac{\cancel{6}\,3}{\cancel{2}}\times\frac{1}{4}+\frac{1}{2}$$

$$\frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8}=\frac{7}{8}$$

One more way to do this

$\because$ P(at most 2 Heads) $\Rightarrow$ ~~P(0H) + P(1H) + P(2H)~~ = 1

$\Rightarrow P(0H)+P(1H)+P(2H)+P(3H)=1$

$P(0H)+P(1H)+P(2H)=1-P(3H)$

$P(0H)+P(1H)+P(2H)=1-{}^3C_3\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3$

$=1-1\times\frac{1}{8}=1-\frac{1}{8}=\frac{7}{8}$

Ex $\Rightarrow$ 2 four coins are tossed simultaneously. what is the probability of getting (i) No head (ii) No tail and (iii) Two heads only?

Soln : Let p = pro of getting head when a coin is drawn $=\frac{1}{2}$

$\therefore$ q = the pro of tail $=1-p=\frac{1}{2}$

$n=4$ $P(X=x)={}^nC_r\, q^{n-r}\, p^r$

(i) $P(0H)={}^4C_0\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0=1\times\frac{1}{16}=\frac{1}{16}$

(ii) $P(0T)=P(4)={}^4C_4\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4=1\times\frac{1}{16}=\frac{1}{16}$

(iii) $P(2H)={}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=6\times\frac{1}{16}=\frac{6}{16}=\frac{3}{8}$

Ex $\Rightarrow$ 3. Eight coins are thrown simultaneously. find the pro of getting at least 6 heads.

Solⁿ: Let p = prob of getting a head, q = prob of getting a tail

$p = \frac{1}{2}$, $q = \frac{1}{2}$, $n = 8$

P(at least 6 heads) = P(6H) + P(7H) + P(8H)

$\Rightarrow {}^{8}C_{6}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{6} + {}^{8}C_{7}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{7} + {}^{8}C_{8}\left(\frac{1}{2}\right)^{8}$

$\Rightarrow 28 \times \frac{1}{256} + 8 \times \frac{1}{256} + 1 \times \frac{1}{256}$

$\Rightarrow \frac{37}{256}$

Ex 8: It is observed that 80% of television viewers watch "Sas Bhi Kabhi Bahu thi" programme. What is prob that at least 80% of the ~~television~~ viewers in a random sample of five watch this program?

Solⁿ: if viewing the programm is a success then

$p = \frac{80}{100} = 0.8$

$q = 1 - 0.8 \quad q = 0.2$

80% of 5 = 4 Therefore, we are to find the prob that 4 or 5 viewers watch the programm.

$= {}^{5}C_{4}(0.2)^{5-4}(0.8)^{4} + {}^{5}C_{5}(0.2)^{0}(0.8)^{5}$

$\Rightarrow 5 \times 0.2 \times (0.8)^{4} + (0.8)^{5}$

$\Rightarrow (0.8)^{4}(1 + 0.8)$

$\Rightarrow 0.4096 \times 1.8$

$\Rightarrow 0.73728$

Ex9: If 8 ships out of 10 ships arrive safely. find the pro that at least one would arrive safely out of 5 ships selected at random;

Sol^n: Let p = pro of safe arrival of a ship

So, $p = \frac{8}{10} = \frac{4}{5}$ and $q = 1 - p = \frac{1}{5}$

$n = 5$

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

pro of at least one would arrive safely means that 1 or 2 or 3 or 4 or 5 arrive safely.

$$P(\text{at least one}) = P(1) + P(2) + P(3) + P(4) + P(5)$$

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

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

$$\Rightarrow 1 - {}^5C_0 \left(\frac{1}{5}\right)^5 \cdot \left(\frac{4}{5}\right)^0$$

$$\Rightarrow 1 - \frac{1}{3125} = \frac{3124}{3125}$$

Ex 16: The mean of a binomial dist^n is 20 and standard deviation is 4 find n, p and q.

Sol^n:

Mean $= np$

$SD = \sqrt{npq}$

$$\bar{x} = np = 20 \quad \text{—— (1)}$$

$$\sigma = \sqrt{npq} = 4$$

squaring both side

$$npq = 16 \quad \text{—— (2)}$$

dividing (ii) by 1

$\frac{\cancel{np}q}{\cancel{np}} = \frac{16}{20}$ $q = \frac{16}{20}$ $\boxed{q = \frac{4}{5}}$

$p = 1 - q$ $\boxed{p = \frac{1}{5}}$

putting the value of p in (i)

$n \times \frac{1}{5} = 20$

$\boxed{n = 100}$ Hence $p = \frac{1}{5}$ $q = \frac{4}{5}$ $n = 100$

Ex = 18: Is there any fallacy in the statement? the mean of a binomial dis^n is 20 and its standard deviation ~~is~~ 7.

Sol^n: Mean = np $\therefore \bar{x} = np = 20$ —— (i)

$S.D = \sqrt{npq}$ $\sigma = \sqrt{npq} = 7$

squaring both sides

$npq = 49$ —— (ii)

dividing eq^n (ii) by (i)

$\frac{\cancel{np}q}{\cancel{np}} = \frac{49}{20}$ $q = \frac{49}{20}$ $\boxed{q = 2.45 > 1}$

which is impossible as $p + q = 1$ Hence statement is wrong

Ex = 19: find the pro of 5 successes in a binomial dis^n whose mean and variance are respectively of 6 and 2.

Sol^n: Mean = $np = 6$ —— (1)

variance $npq = 2$ —— (2) $q = \frac{1}{3}$

dividing eq^n (ii) by (i) $\frac{\cancel{np}q}{\cancel{np}} = \frac{2}{6}$

$p = 1 - q \quad p = \frac{2}{3}$

Substituting the p in eq^n (i)

$\therefore n \times \frac{2}{3} = 6$

$\boxed{n = 9}$ Hence $n = 9, p = \frac{2}{3}\ q = \frac{1}{3}$

Now

$$P(X=5) = {}^9C_5 \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^5 = 126 \times \frac{32}{19683} = \underline{0.2048}$$

fitting of Binomial Distribution : (12 marks)

Ex 22: four coins are tossed 160 times and the following results were obtaind.

No of heads	0	1	2	3	4
frequency	17	52	54	31	6

fit a binomial distribution under the assumption that the coin are unbaised.

sol^n: under the assumption that the coins are unbiased the pro of head (p) and pro of tail q are $\frac{1}{2}, \frac{1}{2}$

in this case $n=4, N=160$

The pro of 0,1,2,3,4 heads will be given by

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

In order to obtain the expected frequencies, we will have to multiply each probability by N.

Number of heads (x)	Expected frequency $N \times {}^nC_r \, q^{n-x} \cdot p^x$
0	$160 \times {}^4C_0 \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^0 = 10$
1	$160 \times {}^4C_1 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^1 = 40$
2	$160 \times {}^4C_2 \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^2 = 60$
3	$160 \times {}^4C_3 \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^3 = 40$
4	$16 \times {}^4C_4 \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^4 = 10$

Ex 24: Four perfect dice were thrown 112 times and the number of times 1, 3 or 5 were thrown as under.

Number of dice showing 1, 3, or 5	0	1	2	3	4
Frequency	10	25	40	30	7

Sol^n: Under the assumption that dice are perfect, the pro of getting 1, 3, or 5 $(p) = \frac{1}{2}$ $q = \frac{1}{2}$

Hence = $n = 4$ $N = 112$

The prob of 0, 1, 2, 3, 4 successes will be given by

$P(X=x) = {}^nC_r \cdot q^{n-x} \cdot p^x$

In order to obtain the expected frequencies we will have to multiply each pro by N.

Number of 1,3 or 5 (x)	Expected frequency $N \times {}^{n}C_x \cdot q^{n-x} \cdot p^{x}$
0	$112 \times {}^{4}C_0 \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^0 = 112 \times \frac{1}{16} = 7$
1	$112 \times {}^{4}C_1 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^1 = 112 \times \frac{4}{16} = 28$
2	$112 \times {}^{4}C_2 \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^2 = 112 \times \frac{6}{16} = 42$
3	$112 \times {}^{4}C_3 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 = 112 \times \frac{4}{16} = 28$
4	$112 \times {}^{4}C_4 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 = 112 \times \frac{1}{16} = 7$

* Poisson Distribution :

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

where $P(X=x)$ = prob of obtaining x number of success

$m = np$ = parameter of the

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

$e = 2.71831$ (base of natural logrithms)

we can obtain the probability of $0, 1, 2 --- x$ successes as follows

Number of success (x)	probability $P(X)$
0	$e^{-m} \cdot \frac{m^0}{0!} = e^{-m}$
1	$e^{-\lambda} \cdot \frac{\lambda^1}{1!} = \lambda e^{-\lambda}$
2	$e^{-\lambda} \cdot \frac{\lambda^2}{2!} = \frac{\lambda^2}{2} \cdot e^{-\lambda}$
3	$e^{-\lambda} \cdot \frac{\lambda^3}{3!} = \frac{\lambda^3}{3!} \cdot e^{-\lambda}$
x	$e^{-\lambda} \cdot \lambda^x / x!$

Exe 8: It is given that 2% of the screws manufactured by a company are defective. Use Poisson Distribution to find the probability that a packet of 100 screws contains (i) No defective screws (ii) one defective and (iii) two or more defective. [Given $e^{-2} = 0.135$]

Sol^n: Let p = probability of a defective screw

$$= 2\% = \frac{2}{100}$$

$$P = \frac{2}{100}, \; n = 100$$

$\therefore$ Mean $= \lambda = np = 100 \times \frac{2}{100} = 2$, $\boxed{\lambda = 2}$

(i) The Poisson Dist^n is given as:

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

$$P(X=0) = \frac{e^{-2} \cdot (2)^0}{0!} = \frac{(0.135) \cdot 1}{1} = 0.135$$

(ii) P(one defective) $= P(X=1) = \frac{e^{-2} \cdot (2)^1}{1!} = \frac{(0.135) \times 2}{1}$

$$= 0.270$$

(iii) P(Two or more defectives) =

$\therefore P(0) + P(1) + P(2) - - - - - + 100 = 1$

$P(2) + P(3) - - - = 1 - P(0) + P(1)$

$$= 1 - [0.135 + 0.270] = 1 - 0.405 = 0.595$$

Ex: 30 → Assume that the probability of a fatal accident in a factory during the year is $\frac{1}{1200}$ calculate the probability that in a factory employing 300 workers, there will be at least 2 fatal accidents in a year (Given $e^{-0.25} = 0.7788$).

Let

p = probability of a fatal accident $= \frac{1}{1200}$

n = no of workers $= 300$

mean $= \lambda = np = \cancel{300} \times \frac{1}{\cancel{1200}_{4}}$ $\left(\lambda = \frac{1}{4}\right)$

The poisson dist^n is given as:

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

p(at least 2 fatal accidents) $= 1 - [p(0) + p(1)]$

p(0 accident) $= P(X=0) = \frac{e^{-\frac{1}{4}} \cdot \lambda^{0}}{0!} = \frac{e^{-0.25}}{1} = 0.7788$

p(1 accident) $= p(X=1) = \frac{e^{-\frac{1}{4}} \cdot \lambda^{1}}{1!} = \frac{e^{-\frac{1}{4}} \cdot \left(\frac{1}{4}\right)^{1}}{1!}$ $= 0.7788$

$= 0.7788 \times 0.25 = 0.1947$

p(at least 2 fatal accident) $= 1 - [0.7788 + 0.1947]$

$\Rightarrow 1 - 0.9735 = 0.0265$

Ex: 33. If x is a poisson variate such that $p(x=1) = p(x=2)$ find the mean and variance of the dist^n.

Sol^n:

Soln: poisson distn:

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

putting $x = 1, 2$

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

$$\boxed{P(X=2) = \frac{e^{-\lambda} \cdot \lambda^{2}}{2!} = \frac{\lambda^{2} \cdot e^{-\lambda}}{2}}$$

By the given condition

$$P(X=1) = P(X=2)$$

$$\cancel{e^{-\lambda}} \cdot \lambda = \frac{\cancel{e^{-\lambda}} \cdot \lambda^{2}}{2}$$

$$2\lambda = \lambda^{2}$$

$$2 = \lambda$$

$$\boxed{\lambda = 2}$$

Hence, mean of the distn $= 2$

as $\bar{X} = \sigma^2$

$$\boxed{(\sigma^2)\ \text{variance} = 2}$$

EX 39: poisson distribution to the following data and calculate the theoretical frequencies:

Deaths:	0	1	2	3	4
Frequency:	109	65	22	3	1

Also find mean and variance of the above distn

Given: $e^{-0.66} = 0.5432$

Sol^n :

Deaths (X)	Frequency	fx	$fe(x) = \frac{N \times e^{-\lambda} \cdot \lambda^x}{x!}$
			$N \times e^{-\lambda} \cdot \lambda^x / x!$
0	109	0	$fe(0) = \frac{200 \times e^{-(0.61)} \cdot (0.61)^0}{0!} = 108.67 \simeq 109$
1	65	65	$fe(1) = \frac{200 \times e^{-(0.61)} \cdot (0.61)^1}{1!} = 66.27 \simeq 66$
2	22	44	$fe(2) = \frac{200 \times e^{-(0.61)} \cdot (0.61)^2}{2!} = 20.21 \simeq 20$
3	3	9	$fe(3) = \frac{200 \times e^{-(0.61)} \cdot (0.61)^3}{3!} = 4.11 \simeq 4$
4	1	4	$fe(4) = \frac{200 \times e^{-(0.61)} \cdot (0.61)^4}{4!} = 0.63 \simeq 1$
	$\Sigma f = 200$	$\Sigma fx = 122$	

$\because \Sigma f = N = 200$

$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{122}{200} = 0.61$

mean → $\boxed{\lambda = 0.61}$

Thus :

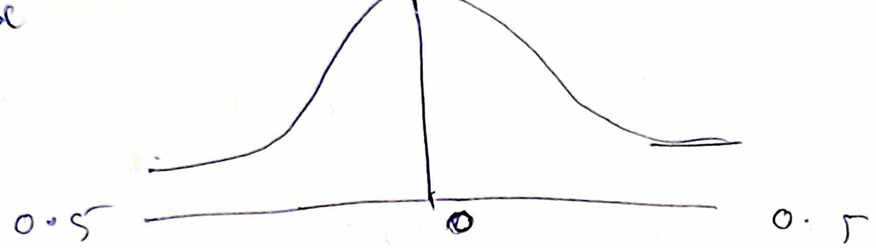
x :	0	1	2	3	4
fe :	109	66	20	4	1

Mean $\bar{x} = 0.61$

$\because$ $\boxed{\text{Mean} = \bar{x} = \text{variance} = \sigma^2 = 0.61}$

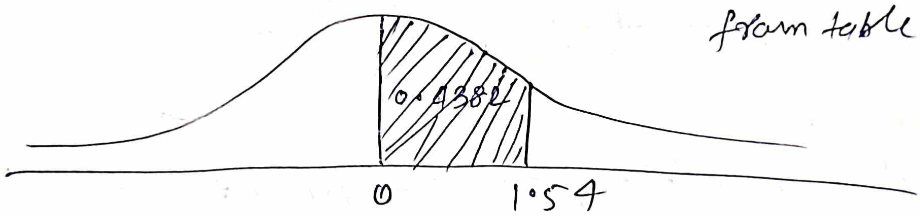
Normal Dist^n

Normal dist^n are developed by laplace and karl Gauss. It is used to study the behavior of continuous random variable like hight, weight and intelligence of a group of student. It is also known as Gaussian dist^n. here value of p and q is very large



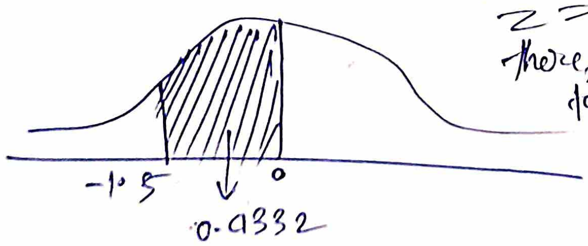
Ex 1: find the area under the normal curve between $z=0$ and $z=1.54$.

Sol^n: if we look to the table given at the end of the book the entry corresponding to $z=1.54$ is 0.0382 and this give the shaded area in the following figure:



Ex 2 find the area b/w $z=-1.5$ and $z=0$

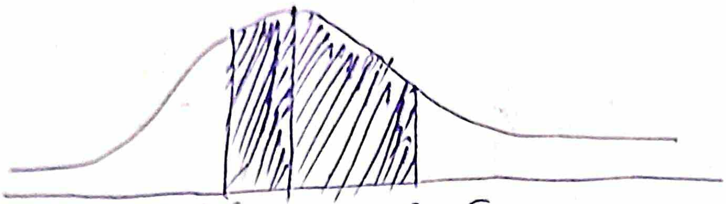
ans: since the curve is symmetrical, we can find the area b/w $z=0$ and $z=-1.5$ by looking the area crossponding to $z=0$ and $z=-1.5$



therefore, the entry crossponding to $z=1.5$ is 0.0332 and it measures the shaded area in the following figure b/w $z=0$ and $z=-1.5$

Ex 3: find the area b/w -0.15 and z = 2.5

Sol^n:



Required Area:

⇒ (Area b/w z = 0.15 and z = 0) + (Area b/w z = 0 and z = 2.5)

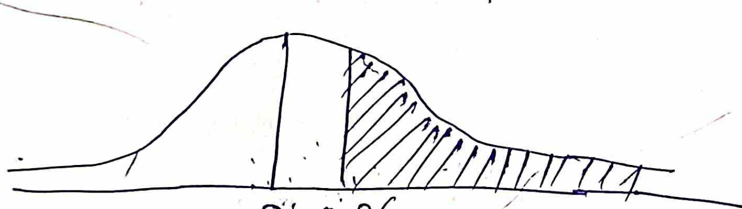
= (Area between z = 0 and z = 0.15) + (Area b/w z = 0 and z = 2.5)

⇒ 0.1736 + 0.4938

⇒ 0.6674

Ex 5: find the area to the righ of z = + 0.36

Sol^n:



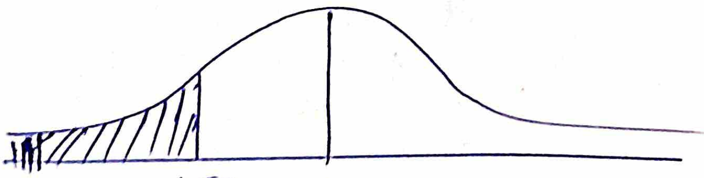
Required Area:

⇒ (Area to the right of z = 0) - (Area b/w z = 0 and z = 0.36)

⇒ 0.5000 - 0.1406 = 0.3594

Ex: 6 find the area to the left of z = -1.52

Sol^n:



Required Area:

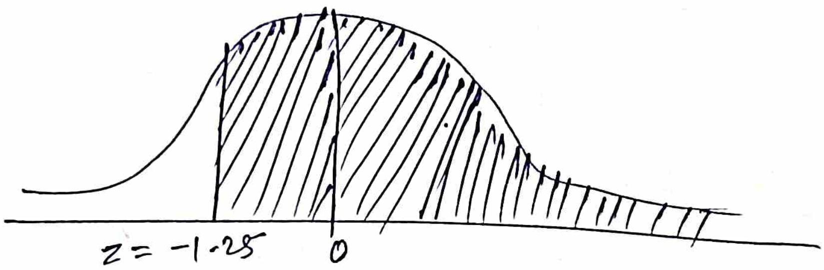
(Area to the left of $z=0$) − (Area b/w $z=-1.52$ & $z=0$)

⇒ (Area to the left of $z=0$) − (Area b/w $z=0$ and $z=1.52$)

⇒ $0.5000 - 0.4357 = \underline{0.0643}$

Ex: 7 Find the area to the right of $z=-1.25$ or greater than $z=-1.25$

ans ⇒



Required Area:

⇒ (Area between $z=-1.25$ and $z=0$) + (Area to the right of $z=0$)

⇒ $0.3944 + 0.5000$

⇒ 0.8944

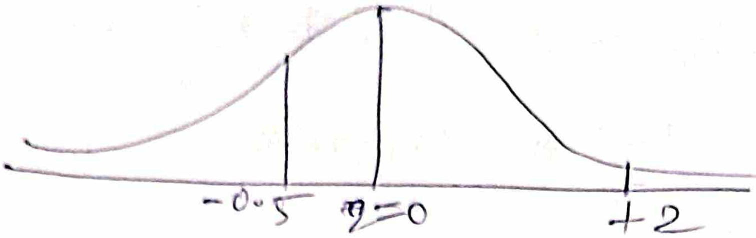
Ex: A normal curve has $\bar{x}=20$ and $\sigma=10$ find the area bet $x_1=15$ and $x_2=40$.

Sol^n: Given $\bar{x}=20$, $\sigma=10$

Between $x_1=15$ and $x_2=40$

z_1 = SNV corresponding to $15 = \frac{x_1-\bar{x}}{\sigma} = \frac{15-20}{10} = -0.5$

z_2 = SNV " " $40 = \frac{x_2-\bar{x}}{\sigma} = \frac{40-20}{10} = 2.0$



Required Area = Area b/w (z = -0.5 and z=0) + Area b/w (z=0 and z=2) ⇒ ∴

⇒ 0.1915 + 0.4772 = 0.6687

2 Marks

(1) Find the Quartile deviation from the given data:
28, 18, 20, 24, 27, 30, 15 ascending order → 15, 18, 20, 24, 27, 28, 30

Soln: $\boxed{\text{Quartile deviation} = \frac{Q_3 - Q_1}{2}}$

Now, $Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item $Q_1 = 1\left(\frac{N+1}{4}\right)^{th}$ item

$Q_3 = 3\left(\frac{7+1}{4}\right) = \frac{24}{4} = 6$th item $= 28$

$Q_1 = 1\left(\frac{7+1}{4}\right) = \frac{8}{4} = 2$th item $= 18$

$Q.d = \frac{28-18}{2} = \frac{10}{2} = 5$

(2) State the relation between the correlation and regression coefficients?

ans ⇒ ∵ Regression coefficient of y on x:

$$\boxed{b_{yx} = r\left(\frac{\sigma_y}{\sigma_x}\right)} \quad \text{— (I)}$$

again Regression coefficient of x on y:

$$b_{xy} = r\left(\frac{\sigma_x}{\sigma_y}\right) \quad \text{— (II)}$$

Multiply both eqn → $b_{xy} \cdot b_{yx} = r\left(\frac{\sigma_y}{\sigma_x}\right) \times r\left(\frac{\sigma_x}{\sigma_y}\right)$

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

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

$$\boxed{r = \sqrt{b_{xy} \times b_{yx}}}$$

(PYQ) Find the mean and the standard deviation of the number of heads in 100 tosses of a fair coin.

ans ⇒ given : No of tosses = 100

$$n = 100$$

· in a fair toss coin the probability is getting success $= \frac{1}{2}$ $\boxed{p = \frac{1}{2}}$

The probability of getting fail $\boxed{q = \frac{1}{2}}$

$\boxed{\text{Mean} = np}$ $\bar{x} = 100 \times \frac{1}{2}$ $\boxed{\bar{x} = 50}$

$\boxed{\text{Standard deviation} = \sqrt{npq}}$ $= \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}}$

$= \sqrt{25}$

$\boxed{\sigma = 5}$

(PYQ) If the Eigen value of a matrix are -1 and 1 · find the third Eigen value when sum of digonal elements of a matrix is given to be -4.

ans ⇒ given sum of digonal elem $= -4$

· let d_1, d_2, d_3 be the eigen values of the matrix

· we know that $d_1 + d_2 + d_3 =$ sum of digonal ele

given that →

$$d_1 + d_2 + d_3 = -4$$

$$\cancel{-1 + 1} + d_3 = -4$$

$$\boxed{d_3 = -4}$$

(PYQ) if the regression coefficient of x on y is 0.8 and that of y on x is 0.2 what is the val of correlation coefficient b/w x and y

∴ relationship b/w correlation and regression coefficient $= r = \sqrt{b_{xy} \times b_{yx}}$

$r = \sqrt{0.8 \times 0.2}$ ~~$= \sqrt{.16} = 0.4$~~

$r = \sqrt{0.8 \times 0.2}$ $r = \sqrt{0.16}$

$\boxed{r = 0.4}$

(PYQ) Two dice are tossed once. find the probability of getting a total of 8.

ans ⇒ combinations ⇒ $(1,2,3,4,5,6) \times (1,2,3,4,5,6) = 36$

$2+6,\ 3+5,\ 4+4,\ 5+3,\ 6+2 = 5$

$\boxed{\text{prob of getting } 8 = \frac{5}{36}}$

(PYQ) check the correlations of the statement "mean of a B.D is 15 and variance is 5"

(PYQ) Average score of two batsman A and B are respectively 54.65, 53.4 and their standard deviation are respectively 1.68, 1.62 which batsman is more consistent.

ans ⇒ To determine which batsman is more consistent we can compare their coefficient of variation.

∴ $\boxed{\text{Coeff variation} = \frac{\sigma}{\bar{x}}}$

for A batsman $CV = \frac{1.68}{54.65} = 0.03074$

for B Batsman $= \frac{1.62}{53.4} = 0.0303$

∴ The coeff variation of Batsman A is greater than batsman B so Batsman A is more consistent

(PYQ) what is difference between skewness and kurtosis?

Skewness	Kurtosis
(i) Measure of the asymmetry of a distribution	(i) Measure of the tail or peakness
(ii) It can be positive, negative or zero	It can be positive, negative & or zero
(iii) positive skewed: right-skewed distribution Negative skewed: left skewed distribution	positive kurtosis: heavy-tailed distribution Negative kurtosis: light-tailed distribution
(iv) Example: stock returns	Exponential distribution

(PYQ) what is the difference b/w correlation and regression?

correlation	regression
(i) It indicates only the nature and extent of linear relationship.	It is the study about the impact of the independent variable on the dependent variable.
(ii) It is symmetric in x and y $r_{xy} = r_{yx}$	It is not symmetric in x and y $b_{xy} \neq b_{yx}$
(iii) It is not used for predictions.	It is used for predictions
(iv) order of variables does not matter	order of variables is important
(v) It is limited to two variables	More than one independent variable is possible.

(PYQ) what is sampling Distribution

ans → It is a probability distribution of a statistic obtained from a large number of samples drawn from a specific population. These distributions help you to understand how a sample statistic varies from sample to sample. It is useful for evaluating the reliability of ~~inter~~ inference based on the statistic

Example: Let's say you are interested in the average height of student in a school, you take multiple random samples of 50 students and calculate the average height for each sample. The distribution of these average heights accross all the samples would be the sampling distribution of the mean.

(PYQ) what is mean and variance of poison distribution

In poisson distribution, the mean represents the average number of events occuring in a fixed interval of time. ~~$\lambda = np$~~ $\lambda = np$. The variance σ^2 is is also equal to the mean λ, so both are equal. which is the average rate of occurrence of the events. ∴ Thus both mean and variance of a poisson distribution are λ ($\sigma^2 = \lambda$)

$$\boxed{(\text{mean})\ \lambda = np = \sigma^2 = \lambda}$$

(PYQ) A bag contains 4 red balls, 3 white balls and 5 green balls. A ball is drawn from A bag at random, what is the probability of getting a non red ball?

ans ⇒
Probability of getting non red ball $= \dfrac{3+5}{12} = \dfrac{8}{12} = \dfrac{2}{3}$

4 Marks

(PYQ) what is the difference between frequency distribution and probability distribution.

Frequency Distribution	Probability distribution
(i) It is used for discrete data	It is used for both discrete and continious data.
(ii) It represents the counts or frequency of each value	It represents the probabilities of each value.
(iii) It contains the actual data or value	It contains the possible outcomes along with their probabilities.
(iv) The sum of frequencies equals to total number of data points	The sum of probabilities is equal to 1.
(v) Commonly use in data analysis.	Utilize in probability theory
(vi) Example: If we roll the die 100 times and record the outcomes a frequency distribution would show how many times each number appears	Example: It would show the probability of rolling each number, since each outcome has an equal chances the Probability of each number would be $\frac{1}{6}$.

12 marks

(9)

(PYQ) what is the difference between probability distribution and sampling Distribution

(b) Explain classical, relative and subjective approaches of probability with example.

probability Distribution	Sampling distribution
(i) It describe the likelihood of each outcome in a population	Describe the distribution of a sample statistic obtaind from multiple samples
(ii) It focous on the population	It focous on samples drawn from the population
(iii) It represents the theoretical distribution of outcomes	It represents the distribution of sample statistics.
(iv) It parameters are fixed and known	It's parameters might vary across samples
(v) Example: Rolling a fair die	Example: calculating the mean of sample means samples drawn from a population.

(b)

(i) classical probability:

Definition: Based on equally likely outcomes

Formula: P(E) = (Number of favorable outcomes) / (Total no of outcomes)

Example: Rolling a fair six-sided die, where the probability of rolling a 3 is 1/6.

(ii) Relative frequency probability:

Definition: Based on observed frequencies

Formula: P(E) = (frequency of event E)/(Total number of observations)

Example: conducting an experiment of flipping a coin 100 times, where the probability of getting heads is the number of times heads occurred devided by 100.

③ Subjective Probability:

Definition: Based on personal judgement or opinion

formula: Based on subjective assessment

Example: Estimating the probability of it raining tommorow based on weather forecasts and personal experiences.

(PY) Q. Describe the different methods of primary data collection.

ans ⇒ Primary data collection methods categorized into two approaches (i) qualitative Methods
(ii) quantitive Methods

(i) Quantitive method

① surveys: This method involves asking a group of peoples a set of questions. Serveys can be conducted online, over the phone, or in person. They can be short and simple or long and detailed depending on the complexity of the information.

② interviews: This method involves asking questions in a more in-depth ~~than~~ servey. They can be one-on-one or conducted with a group of people.

③ observations: This method involves watching and recording behavior. observation can be done in a natural setting or in a controlled environment like a lab.

④ Experiments: Experiments are a more controlled way of collecting data. They involves manipulating variables to see how they effect a particular outcome.

Experiments are often used in ~~business~~ scientific research, but they can also be used in business setting to test ~~[illegible]~~ new products or marketing campaigns.

Qualitative methods :

(i) Case Studies : Detailed examinations of a single entity or phenomena to gain detailed information into complex processes or behaviors.

(2) Document Analysis : Document analysis involves analyzing written documents such as letters, diaries, emails or social media posts. Researchers can use document analysis to learn about the history of a group.

MST-1

Mean

1. Individual Series → (i) Direct Method: $\boxed{\bar{X} = \frac{\Sigma X}{N}}$

(ii) Shortcut method: $\boxed{\bar{X} = A + \left(\frac{\Sigma(X-A)}{N}\right)}$ deviation of midva $d = x - A$

2. Discrete Series → (i) Direct Method: $\boxed{\bar{X} = \frac{\Sigma fX}{N}}$ $\boxed{\Sigma f = N}$

(ii) Shortcut Method: $\boxed{\bar{X} = A + \left(\frac{\Sigma f(X-A)}{N}\right)}$

3. Continuous Series: (i) Direct Method: $\boxed{\bar{X} = \frac{\Sigma fm}{N}}$ the class mid val of $0 \rightarrow 10 = \frac{10+0}{2}$

(ii) Shortcut Method: $\boxed{\bar{X} = A + \left(\frac{\Sigma f(m-A)}{N}\right)}$ $\boxed{d = m - A}$

(iii) Step deviation method: $\boxed{\bar{X} = A + \frac{\Sigma fd'}{N} \times i}$ $d' = \frac{d}{i}$

$\boxed{\bar{X} = A + \frac{\Sigma f\left(\frac{m-A}{i}\right)}{N} \times i}$ $\boxed{\frac{(m-A)}{i} = d'}$

$\boxed{i \rightarrow \text{common gap of the data}}$ $\boxed{i = \text{upper interval} - \text{lower interval}}$

4. Combined Arithmetic mean:

$$\boxed{\bar{X} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}}$$

5. Weighted Arithmetic mean ⇒ $\boxed{\bar{X} = \frac{\Sigma fx}{N}}$

5

Median → sort first

(1) Individual Series: for odd ↓

$$\boxed{M = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}}$$

for even →

$$\boxed{\frac{\left(\frac{N}{2}\right)^{th} \text{ item} + \left(\left(\frac{N}{2}\right)+1\right)^{th} \text{ item}}{2}}$$

(2) Discrete series: for odd ↓ $M = \text{size of } \left(\frac{N+1}{2}\right)^{th}$ item

where N = sum of freq ⇒ Σf

for even → ↓ where $\boxed{N = \Sigma f}$ only all same

(3) Continious Series:

$$\boxed{M = l_1 + \frac{\left(\frac{N}{2} - cf\right)}{f} \times i}$$

l_1 → lower limit of median class

cf → cumulative frequency

f → frequency of median class

i → size of class interval

Note → for Mid value → $\boxed{l_1 = m_1 - \frac{i}{2}}$ $\boxed{h_1 = m + \frac{i}{2}}$

$\boxed{i = \text{diff b/w 2 intervals}}$

$$\boxed{M = l_1 + \frac{\left(\frac{N}{2} - cf\right)}{f} \times i}$$

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

Mode

Continuous Series →

$$\boxed{Z = l_1 + \frac{f_1 - f_0}{2f_1 - (f_0 - f_2)} \times i}$$

l_1 = lower limit of of the mode class

f_0 → freq. of the pre mode of class

f_1 → freq of the class

f_2 → freq. of the post mode class.

i → size of interval class

$\boxed{Z = 3M - 2\bar{X}}$ 4 marks

$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdots x_n}$ geometric mean (8)

$GM = \text{antilog}\left(\frac{\Sigma \log x}{N}\right)$

* Measure of Dispersion

(i) coeff of range $= \frac{H-L}{H+L}$

(1)(a) Inter-quartile Range (IQR) $= Q_3 - Q_1$

(b) Quartile Deviation : $\frac{Q_3 - Q_1}{2}$

(c) Coeff of quartile Deviation : $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

→ individual

$Q_n = \frac{n(N+1)^{th}}{4}$ item → $Q_3 = \frac{3(N+1)^{th}}{4}$ item

Ex → if we get $Q_3 = 6.75$ (for individual ser)

⇒ 6th item + 0.75 (7th item − 6th item)

Note firstly we have to arrange in assending order ↓

→ for discrete $N = \Sigma f$

→ for Continuous → $\boxed{Q_n = l_1 + \frac{n\left(\frac{N}{4}\right)^{th} - cf}{f} \times i}$ → Q_1, Q_2, Q_3

After calculating Q compare with cf colum and find victim row.

* $\boxed{\text{Bowley's coefficient} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$

(1) Standard deviation: Individual Series

(a) $\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$

(b) direct Method

$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$ $\left\{ \bar{X} = \frac{\Sigma X}{N} \right.$

$\boxed{\sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}}$

(c) Shortcut method:

$\boxed{\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}}$ $\sigma = \sqrt{\frac{\Sigma(X-A)^2}{N} - \left(\frac{\Sigma(X-A)}{N}\right)^2}$ $\left\{ d = X - A \right.$

~~for discrete series~~ $\cancel{\sigma = \sqrt{\frac{\Sigma f(X-A)^2}{N} - \left(\frac{\Sigma f(X-A)}{N}\right)^2}}$

(for continuous series $\Rightarrow$ $\sigma = \sqrt{\frac{\Sigma f(m-A)^2}{N} - \left(\frac{\Sigma f(m-A)}{N}\right)^2}$

where m is mid val $\boxed{m = x}$ here

$\boxed{m = \frac{\text{lower interval} + \text{higer interval}}{2}}$

* $\boxed{\text{Coefficient of variation} = \frac{\sigma}{\bar{X}} \times 100}$

* $\boxed{\text{Coefficient of Std variation} = \frac{\sigma}{\bar{X}}}$

for continuous = step deviation method

$\sigma = \sqrt{\frac{\Sigma f(X-A)'^2}{N} - \left(\frac{\Sigma f(X-A)'}{N}\right)^2} \times i$

$\boxed{\sigma = \sqrt{\frac{\Sigma f(d')^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times i}$ $\boxed{d' = \frac{d}{i}}$

$\sigma \sqrt{\frac{\Sigma f(d)^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$ where $\boxed{d = X - A}$ discrete

$\boxed{d = m - A}$ Continuous

Skewness

1. Karl Pearson coeff of Skewness :

$$\text{Karl C.o.s} = \frac{\text{mean} - \text{Mode}}{\sigma} = \frac{\bar{X} - Z}{\sigma}$$

2. Bowley coeff of Skewness :

$$\boxed{B.C.O.S = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

3. Moment of Skewness $\boxed{\sqrt{B_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}}$

4. Skewness $\Rightarrow$ $\boxed{B_1 = \frac{\mu_3^2}{\mu_2^3}}$

5. central moment about mean $\Rightarrow$ $\mu_r =$

$$\boxed{\mu_r = \frac{\sum (x - \bar{x})^r}{N}}$$

Chi square test (χ^2)

for $\rightarrow$ $\boxed{\chi^2 = \sum \left[\frac{(O-E)^2}{E} \right]}$ O $\rightarrow$ observed freq, E $\rightarrow$ expected freq

Step 1 : $H_0 \Rightarrow$ negative Assumption

: S2 : $H_a \rightarrow$ Take same assumptⁿ

S3 : χ^2

S4 : Degree of freedom $= N-1$

Degree of freedom $= (r-1)(c-1)$

S5 : Compare the chi-square Distⁿ table & DoF

1. if calculated val of $\chi^2 <$ Tabulated value $\Rightarrow H_0$ accept
2. if calculated val of $\chi^2 >$ Tabulated value $\Rightarrow H_a$ & reject

MST-2

1 correlation Analysis

* Karl Pearson coefficient of correlation

(i) when mean is whole No :-

$$\boxed{r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}}$$

where $r \rightarrow$ Karl Pea Coef of correlatio

where $\boxed{x = x - \bar{x},\quad y = y - \bar{y}}$

where Σxy is the sum of the product of two variables x and y

$\Sigma x^2, \Sigma y^2$ is the sum of the square of the values of x, y $(x-\bar{x})(y-\bar{y})$

(ii) when mean is fraction (Assume mean metho):

$$\boxed{r = \frac{N \cdot \Sigma dx dy - \Sigma dx \cdot \Sigma dy}{\sqrt{N \cdot \Sigma dx^2 - (\Sigma dx)^2}\sqrt{N \cdot \Sigma dy^2 - (\Sigma dy)^2}}}$$

where $dx, dy \rightarrow$ Deviations of variable x and y

$\Sigma dx, \Sigma dy \rightarrow$ sum of Deviations

$\Sigma dx^2, \Sigma dy^2 \rightarrow$ sum of square of deviations

$N \rightarrow$ No of observations

$dx = (X - A)$ $\boxed{dy = y - A}$

$\boxed{dx^2 = dx \times dx}$, $\boxed{dy^2 = dy \times dy}$

(iii) simple method / direct method:

$$\boxed{r = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}}$$

$$\boxed{\sigma x = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}} \rightarrow S.D$$

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

$$\boxed{r = \frac{\Sigma xy}{N \cdot \sigma x \cdot \sigma y}}$$

$$\boxed{\sigma x = \sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}}$$

(b) Rank correlation (1):

$$\boxed{R = 1 - \frac{6\Sigma D^2}{N^3 - N}}$$

where D is the diff b/w two rank (R1 - R2)

$D = R1 - R2$

$N \rightarrow$ No of pair of observation

(ii) when rank are not given: formula is same but you have to find rank from question. but How catch a victim element and including victim element any greater than that victim elements frequency is the rank of victim element. Regressi

(iii) when ranks are equal in same series:

$$\boxed{R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3) + \text{------}\right]}{N^3 - N}}$$

where m is no of items of equal ranks in series x जितना ele का rank equal होगा उतना बार $\frac{1}{12}(m^3 - m)$ आयेगा जहाँ m no of times repeated.

in series y same

✱ Regression → Regression equation: (~~least square method~~)

(a) (i) regression equation y on x:

$$y = a + bx$$

~~$\Sigma x = Na + b\Sigma$~~

$$\boxed{\Sigma y = Na + b\Sigma x} \text{ ——— (1)}$$

$$\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2} \text{ ——— (2)}$$

(ii) Regression equation x on y:

$$x = a + by$$

$$\boxed{\Sigma x = Na + b\Sigma y} \text{ ——— (1)}$$

$$\boxed{\Sigma xy = a\Sigma y + b\Sigma y^2} \text{ ——— (2)}$$

(b) when direct values are given or when equation is not sloving

(i) Regression Equation of y on x → $\boxed{y = a + bx}$ (least square method)

$$\boxed{b_{yx} = \frac{N \cdot \Sigma xy - \Sigma x \cdot \Sigma y}{N\Sigma(x)^2 - (\Sigma x)^2}}$$

$$\boxed{a = \bar{y} - b\bar{x}}$$

Regression eqn of x on y: ②

$$\boxed{x = a + by}$$

$$\boxed{b_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma y^2 - (\Sigma Y)^2}}$$

$$\boxed{a = \bar{x} - b\bar{y}}$$

Ⓒ Using Deviations taken from Actual Mean (when x, y value is too large)

(i) Regression Equation of X on y:

- regression ~~equation of~~ coefficient of x on y

$$\boxed{b_{xy} = \frac{\Sigma XY}{\Sigma y^2}} \qquad \boxed{X = x - \bar{x}} \quad \boxed{Y = y - \bar{y}}$$

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

(ii) Regression Equation of y on x:

- regression coefficient of y on x

$$\boxed{b_{yx} = \frac{\Sigma XY}{\Sigma x^2}} \qquad \boxed{X = x - \bar{x}} \quad \boxed{Y = y - \bar{y}}$$

- regression equation of y on x

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

Ⓓ Using deviation from ~~Actual~~ Assume mean method

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

$$b_{xy} = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}$$

* Frequency Distn

④ poisson distribution: $\boxed{P(X=x) = \frac{e^{-\lambda} \cdot \lambda^{r}}{r!}}$

where $P(X=x) \rightarrow$ prob of obtaining x no of success.

$\lambda = np \rightarrow$ avg rate of occurrence of the event

$r \rightarrow$ no of events observed.

② Binomial distⁿ: $P(X=x) = {}^{n}C_{x}\, p^{x} \cdot q^{n-x}$

$$\boxed{{}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}}$$

where p = probability of success
q = probability of failure
$n \to$ total no of trials.
$x \to$ no of events observed

$$\boxed{P(X=x) = {}^{n}C_{r}\, q^{n-r} \cdot p^{r}}$$

③ bayes theorem:

$$P(C|D) = \frac{\text{joint probability of the machine/plant etc}}{\text{sum of joint probability of all machines/plant}}$$

* T- test $\to$ $\boxed{|t| = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}}$ $\boxed{S = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n-1}}}$

$\boxed{df = n-1}$

(i) Test of hypothesis about the population mean

(ii) Test of hypothesis about difference two means in case of independent samples

$$\boxed{t = \frac{\bar{X}_1 - \bar{X}_2}{S} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}$$

$$\boxed{S = \sqrt{\frac{\Sigma (X_1 - \bar{X}_1)^2 + \Sigma (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}}$$

$\boxed{\text{degree of freedom} = n_1 + n_2 - 2}$

$\to$ when standard deviations of two samples s_1 and s_2 are given:

$$\boxed{S = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

* F test $\to$ $\boxed{F = \frac{S_1^2}{S_2^2}}$ $\to$ population variance $\boxed{S_1^2 = \frac{\Sigma (X_1 - \bar{X}_1)^2}{n_1 - 1}}$

where $S_1^2 > S_2^2$

$$\boxed{S_2^2 = \frac{\Sigma (X_2 - \bar{X}_2)^2}{n_2 - 1}}$$

$\boxed{V_1 \text{ degree of freedom} : n_1 - 1}$

$\boxed{V_2 \text{ degree of freedom} : n_2 - 1}$

Q) Write properties of Binomial Distribution. (3)

Ans ⇒ ① Theoretical Frequency Distribution: The binomial distribution is a theoretical frequency distribution which is based on Binomial Theorem of algebra. With the help of this distribution, we can obtain the theoretical frequencies by multiplying the probability of success by the ~~to~~ total number (N).

② Discrete Probability Distribution: The binomial distribution is a discrete probability distribution in which the number of successes 0, 1, 2, 3, ---- n are given in whole numbers and not in fractions.

③ Main parameters: The binomial dist^n has two parameters n and p. The entire dist^n can be known from these two parameters.

④ Constants of Binomial Distribution: The constants of Binomial distribution are obtained by using the formula.

$\boxed{\text{Mean} = \bar{x} = np}$ $\boxed{S.D = \sigma = \sqrt{npq}}$ $\boxed{\text{variance} = \sigma^2 = npq}$

$\boxed{\text{Moment coeff. of skewness} = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}}$ $\boxed{\text{Moment coeff of kurtosis} = \beta_2 = 3 + \frac{1-6pq}{npq}}$

⑤ Uses: It has been found useful in those fields where the outcome is classified into success and failure like coin experiment, dice throwing, manufacturing of items by a company.

(PYQ) Write properties of Poisson Distribution.

① Discrete probability Distribution: The Poisson distribution is a discrete probability dist^n in which the number of successes are given in whole numbers such as 0, 1, 2, 3 etc. not in fractions

(ii) Main parameter: It has only one parameter λ and its value is equal to np, $\lambda = np$. The entire distribution can be known from this parameter.

(iii) Constant of poisson Distribution: The constants of the poisson can be obtained from the following formula.

$\boxed{Mean = \bar{x} = \lambda = np}$ $\boxed{S.D = \sigma = \sqrt{\cancel{np}}}$ $\boxed{variance = \sigma^2 = \cancel{npq}}$

$\boxed{\text{Moment coeff of skewness} = \sqrt{\beta_1} = \frac{1}{\sqrt{m}}}$ $\boxed{\text{Moment coeff of kurtosis} = \beta_2 = 3 + \frac{1}{m}}$

(iv) Use: The poisson $dist^n$ is useful in rare events where the probability of success very small and the value of n is very large.

(v) Equality of Mean and variance: An important property of the poisson distribution is that its mean and variance are equal.

$$\bar{x} = \sigma^2 \quad \text{or Mean} = \text{variance}$$

(P99) Write properties of Normal distribution.

(i) Continuous probability Distribution: Normal distribution is a distribution of continuous variables. Therefore it is called Continuous probability distribution.

(ii) Main parameters: The normal $dist^n$ has two parameters namely mean $(\bar{x})$ and standard deviation (σ). The entire distribution can be known from these two parameters.

(iii) Constants: The constants of Normal distribution are denoted by the following symbols.

$\boxed{Mean = \bar{x} \text{ or } \mu \text{ or } m}$ $\boxed{\text{Moment Coeff of skewness} = \sqrt{\beta_1} = 0}$

$\boxed{S.D = \sigma}$ $\boxed{\text{Moment coeff of kurtosis} = \beta_2 = 3}$

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

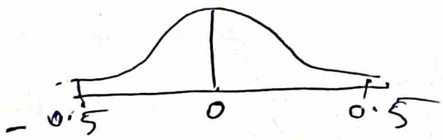
(4)

Equality of Mean, Median and Mode: In a normal distⁿ mean, median and mode are equal.

$$\boxed{\bar{X} = M = Z}$$

(v) **Total Area:** The total area under the normal curve is 1.

(vi) **ordinate:** The ordinate of the normal curve at the mean is maximum.



probability

(1) **Experiment:** when we conduct a trial to obtain some statistical information, it is called an experiment.

Ex ⇒ (i) Tossing of a fair coin is an experiment.

(ii) Rolling a fair die is an experiment.

(iii) drawing a card from a pack of playing cards

(2) **Events:** The possible outcomes of a trial/experiment are called events. denoted by A, B, C etc.

Ex ⇒ (i) if a fair coin is tossed, the outcomes - head or tail are called events.

(ii) if a fair die is rolled, the outcomes 1 or 2 or 3 or 4 or 5 or 6 appearing up are called events.

(3) **Exhaustive Events:** The total number of possible outcomes of a trial/experiment are called exhaustive events.

Ex: (i) In case of tossing a die, the set of six possible outcomes 1, 2, 3, 4, 5 and 6 are exhaustive events.

(ii) In case of tossing a coin, the set of outcomes, H and T are exhaustive events.

(4) **Equally - Likely Events:** The events are said to be equally-likely ~~when one does not occur more than~~ if the chance of happening of each event is equal and same.

Ex ⇒

(i) if a fair coin is tossed, the events H and T are equally likely events.

(5) Mutually Exclusive Events: Two events are said to be mutually exclusive when they cannot happen simultaneously in a single trial.

Ex ⇒ In tossing a coin, the events Head and Tail are mutually exclusive because they cannot happen simultaneously in a single trial. Either head occurs or tail occurs Both cannot occur simultaneously.

(6) Complementary Events: let there be two events A and B. A is called the complementary event of B and B is called complementary event of A if A and B are mutually exclusive and exhaustive.

Ex ⇒ (i) In tossing a coin, occurrence of head (H) and tail (T) are complementary events.

(7) Simple and Compound events: In case of simple events we consider the probability of happening or not happening of single events.

Ex: if a die is rolled once and A be the event that face number 5 is ~~to~~ turned up, then A is called a simple event.

Ex: if two coins are tossed simultaneously and we shall be finding the probability of getting two heads, then we are dealing with compound events.

(8) Independent Events: Two events are said to be independent if the occurrence of one does not affect and not affected by the occurrence of the other.

Ex: in tossing a die twice, the event of getting 4 in the 2nd throw is ~~do~~ independent of getting 5 in the first throw.

Ex ⇒ In tossing a coin twice, the event of getting a head in the 2nd throw is independent of getting head in the 1st throw.

i) Dependent Events: Two events are said to be dependent when the occurrence of one does affect the probability of the occurrence of the other events.

Ex: The Prob of drawing a king from a pack of 52 cards is $\frac{4}{52}$. But if the card drawn (king) is not replaced in the pack, the probability of drawing again a king is $\frac{3}{51}$

(PYQ) Difference between Binomial, Poisson, Normal

Parameter	Binomial	Poisson	Normal
① Nature	Discrete	Discrete	continuous
(ii) Prob fn	$P(X=x) = {}^nC_x \cdot q^{n-x} \cdot p^x$	$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$	$P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\bar{x}}{\sigma}\right]^2}$
(iii) parameter Restriction	n, p $0 < p < 1$	m $m > 0$	$\bar{x}, \sigma$ $-\infty \le x \le \infty$
④ Mean and variance	$\bar{x} = np$ $\sigma^2 = npq$	$\bar{x} = m$ $\sigma^2 = m$	$\bar{x}$ or $m = \sigma^2$
⑤ shape	symmetrical or Asymmetrical	Positively skewed	Perfectly symmetrical
⑥			