DOMS Page No.
Date / /

* Probability - The idea of probability is chance or randomness. Probability theory has been used in genetics, kinetic theory, commⁿ systems, operaⁿs research, actuarial science, etc.

- Sample Space - collection of all possible outcomes of an random experiment is called sample space. It is denoted by Ω (omega). Elements of Ω is denoted by ω (small omega)

eg - The no. of jobs in print queue of a mainframe computer maybe modelled as random. Here the sample space can be taken as $\Omega = \{0, 1, 2, \dots\}$

- The length of ~~type~~ time between successive earthquakes in a particular region that are greater in magnitude. Here Ω is set of non-negative real numbers $\Omega = \{ t \mid t \geq 0 \}$

Event - We're often interested in a particular subset of omega which is called event.

eg- Consider a coin is tossed twice then consider event as atleast 1 head then.
$\Omega = \{HH, HT, TH, TT\} \Rightarrow E = \{HH, HT, TH\}$

* A Probability Measure - on Ω is a function P from subsets of Ω to the real numbers that satisfies the foll. axioms:

Page No.
Date / /

1) $P(\Omega) = 1$

2) If $A \subset \Omega$ [proper subset], then $P(A) \geq 0$ [$0 \leq P(A) \leq 1$]

3) If $A1$ & $A2$ are disjoint [no common element], then $P(A1 \cup A2) = P(A1) + P(A2)$. More generally, if $A1, A2, \ldots, An$ are all mutually disjoint, then $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$

* Properties

1) $P(A^c) = 1 - P(A)$

[$A \cup A^c = \Omega$

$\therefore P(A \cup A^c) = P(\Omega) = 1$

$\therefore P(A) + P(A^c) = 1$

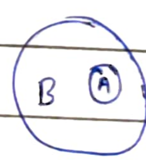
$\therefore P(A^c) = 1 - P(A)$]

2) $P(\phi) = 0$ (nullset)

[~~ϕ^c~~ is a Ω^c is ϕ $\therefore$ $P(\phi) = P(\Omega^c) = 0$]

3) If $A \subset B$ then $P(A) \leq P(B)$

[$B = A \cup (B \cap A^c)$



$\therefore P(B) = P(A \cup (B \cap A^c))$

$P(B) = P(A) + P(B \cap A^c)$ [As they're disjoint]

* This is non rosgen

$\therefore P(B) \geq P(A)$

4) Addⁿ law of probability :) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
let A & B be subset of Ω not necessary disjo..

* Suppose a fair coin is tossed thrice then we get
$\Omega = \{HHH, HTH, THH, TTH, HHT, HTT, THT, TTT\}$ Let A denote the event then atmost 2 heads. $n(\Omega) = 8$

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

$n(A) = 7$ $P(A) = 7/8$

* **Conditional Probability**
$P(A/B)$ Let A & B be 2 events with $P(B) \neq 0$ then conditional probability $P(A/B)$ [A given B]
$$= \frac{P(A \cap B)}{P(B)}$$

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

* **Multiplicaⁿ law** - let A & B be 2 events with $P(B) \neq 0$ then $P(A \cap B) = P(A/B) \cdot P(B)$
$= P(B/A) \cdot P(A)$

* **Law of total probability** - let $B_1, B_2, \ldots B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ & $B_i \cap B_j = \phi$ for $i \neq j$
[then $B_1, B_2 \ldots$ are called partition sets]
with $P(B_i) \neq 0$ for $i = 1$ to n

Page No.
Date / /

then $\boxed{P(A) = \sum_{i=1}^{n} P(A/B_i) \; P(B_i)}$

eg. An urn ~~contains~~ contains 3 red balls & 1 blue ball. What is the prob. that a red ball is selected on the 2nd draw

~~R~~ R R B

1st red ⇒ 2/3 — 3/4 1/4

1st blue ⇒ 3/3 — 2/3 3/3

$\frac{2}{3} + \frac{3}{3} / 2 = 5/6$ — $6/12 + 3/12 = 9/12$

= 3/4

let R_1 & R_2 denote the events that on the 1st trial & on 2nd trial then $P(R_2) = P(R_1)$

$= 3/4$

Using law of total prob. we get $P(R_2) =$

$P(R_2/R_1) \cdot P(R_1) + P(R_2/B_1) \cdot P(B_1)$

$\frac{2}{3} \times \frac{3}{4} + 1 \times \frac{1}{4} = \frac{6}{12} + \frac{1}{4} = \frac{9}{12} = \frac{3}{4}$

~~$\frac{4}{3} \times \frac{3}{3}$~~

Baye's' Rule

~~let A & B be events then prob. of~~

~~$P(B/A) = \frac{P(A/B) \cdot P(B)}{P(A/B) \cdot P(B) + P(B/A) \cdot P(A)}$~~

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

DOMS Page No.
Date / /

Proof: we have $P(A|B) = P(A \cap B) / P(B)$

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

i.e

$P(A \cap B) = P(A/B) \cdot P(B)$ — ①

$P(A \cap B) = P(B/A) \cdot P(A)$ — ②

From ① & ②,

$P(A/B) \cdot P(B) = P(B/A) \cdot P(A)$

$\therefore P(A/B) = \dfrac{P(B/A) \cdot P(A)}{P(B)}$

Remark: In general, let A & $B_1, B_2, \ldots B_n$ be events where B_i's are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$ with $P(B_i) \neq 0$

then $P(A/B) = \dfrac{P(A/B_i)\, P(B_i)}{\sum_{i=1}^{n} P(A/B_i)\, P(B_i)}$

* Independent events: let A & B be the events. Suppose the occurence of B does not depend on occurence of A then we say that A & B are independent. On the other hand, we have

$P(A/B) = P(A)$ &

$P(B/A) = P(B)$

ie $P(A) = P(A/B) \Rightarrow P(A) = \dfrac{P(A \cap B)}{P(B)}$

$\Rightarrow P(A \cap B) = P(A) \cdot P(B)$

In general, $A_1, A_2, \ldots A_n$ are mutually independent if for any subcollection of $A_{i1}, A_{i2}, \ldots A_{in}$

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{in}) = P(A_{i1}) P(A_{i2}) \ldots P(A_{in})$$

* Arithmetic Mean : Summation of all values upon no. of observation = $\Sigma x_i / N$

Q. Find avg. marks obtained by students -
64 + 69 + 72 + 72 + 75 + 65

$$= \frac{64+69+72+72+75+65}{6} = 417/6 = 69.5$$

Q Find AM of foll.

No of day spent	1	2	3	4	5	6	7	8
No. of patients	5	6	5	10	8	4	3	2

$$= \frac{1\times5 + 2\cdot6 + 3\cdot5 + 4\cdot10 + 5\cdot8 + 6\cdot4 + 7\cdot3 + 8\cdot2}{5+6+5+10+8+4+3+2}$$

$$= \frac{173}{43} = 4.02$$

Q Median of 65, 55, 89, 56, 35, 14, 56, 55, 87, 45, 92

Asc = 14, 35, 45, 55, 55, 56, 56, 65, 87, 89, 92

55 56 Mode

Q. Find Median

Monthly Sal	100-120	120-140	140-160	160-180	180-200	200-2[illegible]
Frequency	15	35	50	60	30	10
Cf less than	15	50	100	160	190	200

Median class

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

l_1 = low limit of median class

l_2 = upper " "

cf = cumulative frequency of pre-median class

f = frequency of median class

Total = 15 + 35 + 50 + 60 + 30 + 10 = N = 200

N/2 = 100 → 140 to 160 sal. class

l_1 = 140

l_2 = 160

cf = 50

f = 50

$$\text{Median} = 140 + \frac{(20)(100-50)}{50}$$

$$= 140 + 20 \times 50/50$$

$$= 160$$

Date / /

Exclusive event $A \cap B = \phi$

Exhaustive $A \cup B$ = sample space

Frequency : no. of times a variable takes a value

eg

Class interval	100-120	120-140	140-160	160-180	180-200	200-220
frequency	15	35	50	60	30	10
~~class mark~~	~~129~~	~~129~~	~~149~~	~~169~~	~~189~~	~~209~~

~~class mark $\frac{120-100}{2} = \frac{20}{2} = 10$~~ → at 10^{th} position is 109

Class mark = upper + lower / 2
= 110, 130 . . .

class mark	110	130	150	170	190	210

$$\text{mean} = \frac{(110 \times 15) + (35 \times 130) + (150 \times 50) + (170 \times 60) + (190 \times 30) + (210 \times 10)}{15 + 35 + 50 + 60 + 30 + 10}$$

Q

Age	3	4	5	6	7	8	9	10
No. of children	14	20	40	54	40	18	7	7
cf less than	14	34	74	128	168	186	193	200

$N/2 = 100$

cf just exceeding 100 is 128

∴ Median = 6

DOMS Page No.
Date / /

Q Mode
18, 22, 34, 55, 66, 66, 77, 88, 66
Mode is what occurs most time = 66

Q Mode

Size of pant	60	65	70	75	80	85	90
nos of pa!	11	15	25	40	20	15	12

→ High frequency = 40
∴ mode = 75

Q.

Class interval	100-200	200-300	300-400	400-500	5-6	6-7
Frequen	3	7	8	2	4	6

$$Mode = L1 + \left(\frac{(l_2 - l_1)(f1 - f0)}{2f1 - f0 - f2}\right)$$

l_1 = lower of modal class
l_2 = upper
f1 = frequency of moon
f0 = frequency of pre-mode
f2 = frequency of post

Modal class = having highest frequency
= 300 - 400
8

Mode = $300 + \frac{(400-300)(8-7)}{(2\cdot 8-7-2)}$

$= 300 + \frac{100}{7}$

$= 300 + 14.28$

$= 314.28$

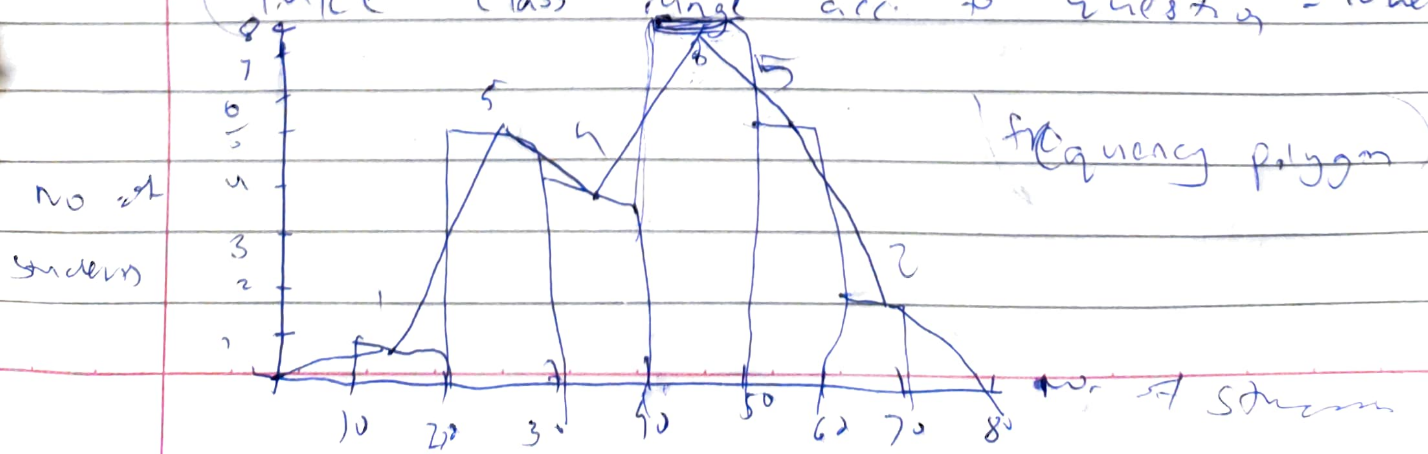
Q Draw histogram

Time (sec)	No. of soc
10-20	1
20-30	5
30-40	4
40-50	8
50-60	5
60-70	2

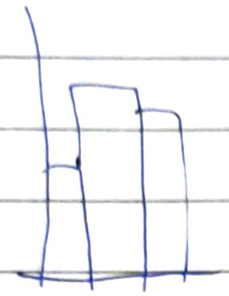
10k

Q 17, 20, 24, 26, 27, 30, 38, 34, 40, 35, 47, 41, 44, 49, 45, 48, 44, 54, 50, 60, 58, 58, 63, 55, 20

→ Make class width acc. to you
(Take class range acc. to question - when)



→ Histogram Bar



(connected)

(not connected)

join with straight line only

- **stem & leaf diagram**

Q. Height of 20 students (cm)

143	176	154	159	172	165	162	171	146	165
160	145	165	182	175	186	160	158	167	172

Q. 1) Display stem & leaf form
2) find range of height
3) Mode of height
4) Median of height

Stem (hundred ten)	leaf (unit)	
14	~~3 6 5~~	3 5 6 (asc order)
15	~~4 9 8~~	4 8 9
16	0 0 2 5 5 5 7	
17	1 2 2 5 6	
18	2 6	

iii) Range = Max - min = 186 - 143 = 43

(leaf)

iii) Mode = unit value repeating most, in single ~~leaf~~ stem

5 in 16 , = 165

ii) Median . N=20 . ~~note~~ middle value = 10th & 11th term avg

3 in 14

3 in 15

4th & 5th term in 16

$\frac{5+5}{2}$ = 5 = median

165 + 165 / 2 = 165

→ Mean - Mode = 3 (Mean - Median)

or

3 Median = Mode + 2 Mean

Q What % of students have height > 175 cm

→ Go to 17 . more than 5 - 16 & all of 18

2, 6

= 3 terms

$\% = \frac{3}{20} \times 100 = 15\%$

Since there are 3 ~~digits~~ leaves ~~after~~ stem after 5 in 15th

Page No.
Date / /

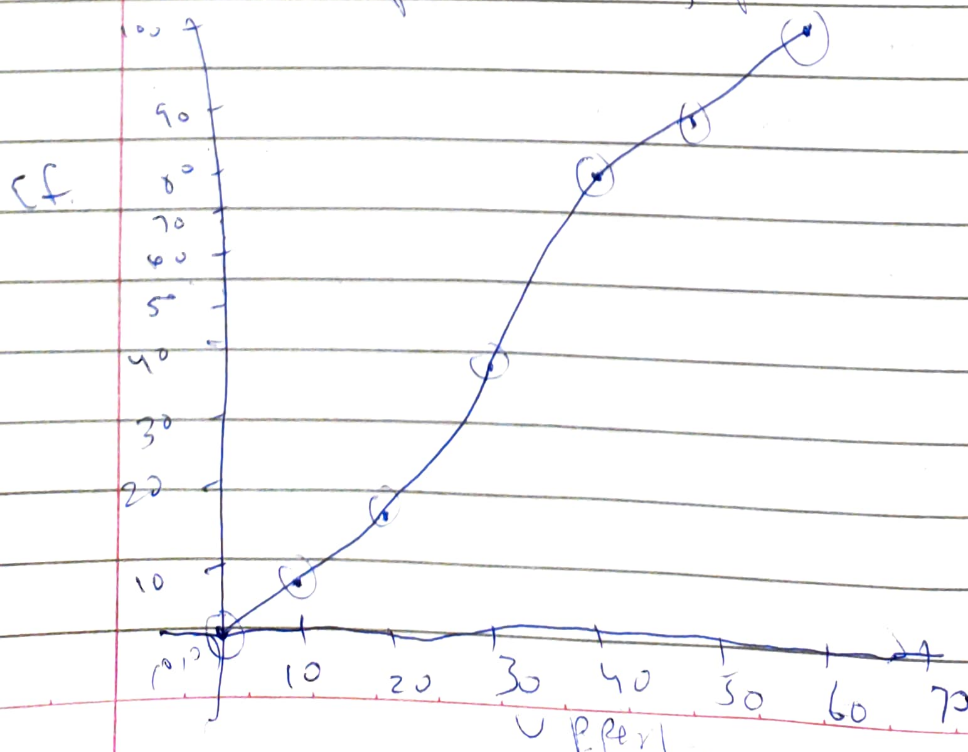
★ Ogives [Cumulative frequency graph]
i) less than method
ii) more than method

i)

Marks	No. of Students	Upper limit of class (x)	Cf (y)	Points
0-10	5	10	5	(10
10-20	10	20	15	(20,
20-30	24	30	39	(30,39
30-40	46	40	85	(40,8
40-50	8	50	93	(50,93
50-60	7	60	100	(60,1
	100			
Take ~~this~~ -10-0	0	0	0	(0,0

~~Plo~~ Draw points on graph

free hand curve



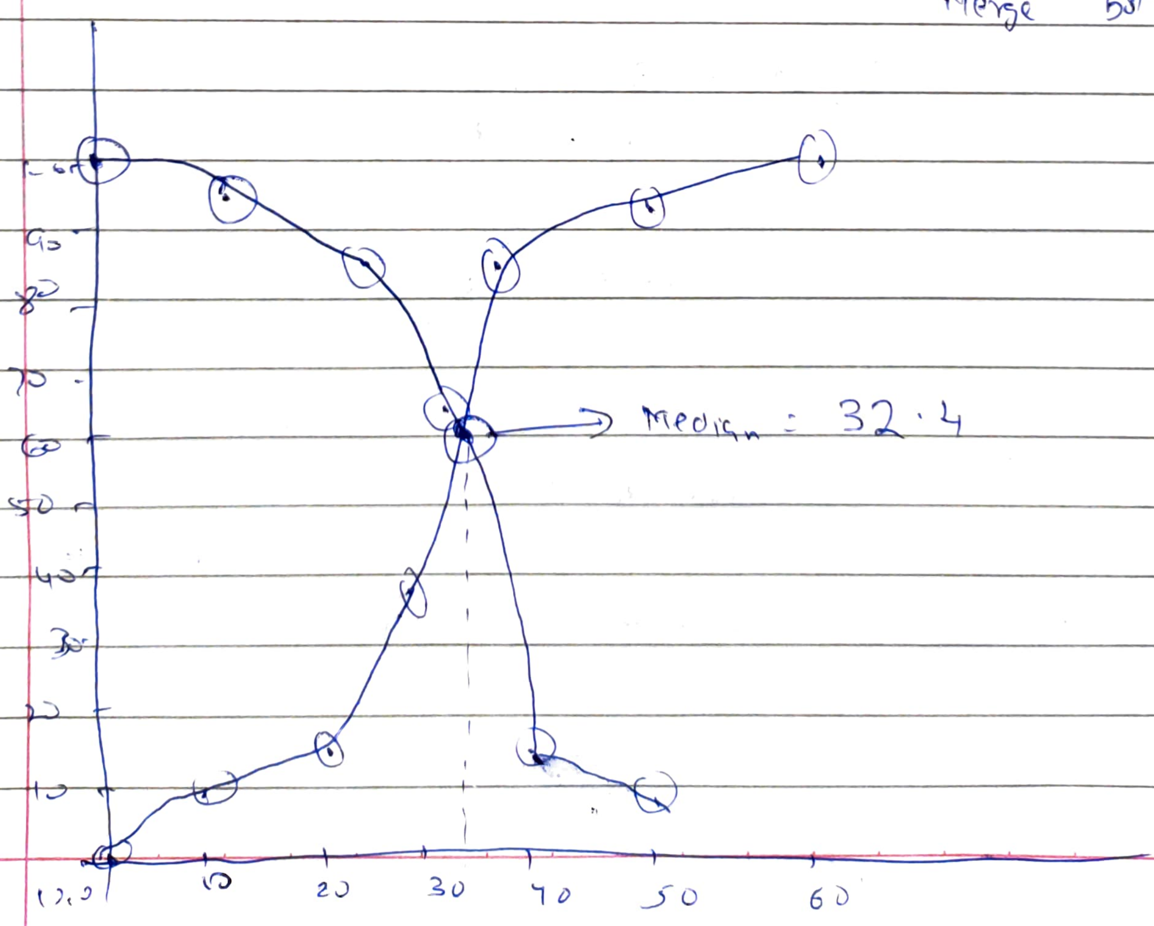
DOMS Page No.
Date / /

ii) more than

Marks	No. of Stu	lower limit	cf (in reverse)	Points
0-10	5	0	100	(0,100)
10-20	10	10	95	(10,95)
20-30	24	20	85	(20,85)
30-40	46	30	61	(30,61)
40-50	8	40	15	(40,15)
50-60	7	50	7	(50,7)

(x,y)

Merge both graph



AM = $\Sigma x_i / N$

GM = $\sqrt[n]{x_1 \cdot x_2 \cdots x_n}$ (Ungrouped data)

$x_1, x_2, \ldots$ are observation

n = total no. of ~~experiment~~ observation

Q. 30, 31, 32, 80, 85

$$GM = \sqrt[5]{30 \times 31 \times 32 \times 80 \times 85} = 45$$

$$GM^n = (x_1 \cdot x_2 \cdots x_n)$$

$$\log(GM^n) = \log(x_1 \cdots x_n)$$

$$n \log GM = \sum_{i=1}^{n} (\log(x_i))$$

$$\therefore \log G = \frac{\sum_{i=1}^{n} \log x_i}{n}$$

$$\therefore G = \text{antilog}\left(\frac{\sum_{i=1}^{n} \log x_i}{n}\right)$$

GM for grouped =

$$G = \text{Antilog}\left(\frac{\sum_{i=1}^{n} f_i \times \log x_i}{n}\right)$$

DOMS Page No.
Date / /

Q Weight of books (gm)

Weight of books (gm)	No. of books (f)	Mid point (x_i)	$\log(x_i)$ ($\log_{10} x_i$)	$x_i f_i$
60-80	22	70	1.84	40.59
80-100	38	90	1.954	74.24
100-120	45	110	2.041	91.85
120-140	35	130	2.114	73.99
140-160	200	150	2.170	43.32
				342

$\sum \log x_i \cdot f = 342$

GM = antilog $(342/n)$ antilog $(342/34$

Harmonic Mean $= \dfrac{n}{(f_1/x_1) + (f_2/x_2) + \ldots (f_n/x_n)}$

Q.

x	f	$1/x_i$	f_i/x_i
1	2	1	2
3	4	0.33	1.332
5	6	0.2	1.2
7	8	0.14	1.144
9	10	0.111	1.111
11	12	0.091	1.092
	42		7.879

$HM = \dfrac{42}{7.879}$

$\simeq 5.33$

Dispersion

Absolute	Relative
i) Scattering of data in terms of distance e.g. range	i) Compare 2 datas coeff. of SD
ii) Variation in terms of avg. deviation of observation (SD)	

- Range = Max - Min
- Deciles : 10 equal data distribution
- Percentile - 100 equal part
- Quartile - 4 equal part

Quartile Deviation

median

min — Q1 — Q2 — Q3 — max

(25% of data) (50%) (75%)

(1st quartile) (2nd) (3rd)

Q3 - Q1 = Inter quartile range [Max. data covered in range]

Quartile deviation (QD) = Q3 - Q1 / 2

DOMS Page No.
Date / /

Q. Find QD of foll [ungrouped data]
35, 45, 53, 42, 39, 34, 40, 51, 57, 52, 47, 62, 55, 63, 50

step 1) Asc order
34, 35, 39, 40, 42, 45, 47, 50, 51, 52, 53, 55, 57, 62, 63

N = 15

step 2) $Q1 = \left(\frac{n+1}{4}\right)^{th}$ observation $= 4^{th} = \boxed{40}$

$Q2$ = Median = $\boxed{50}$ $(n+1/2)$ observation $= 8^{th}$

$Q3 = 3\left(\frac{n+1}{4}\right) = 12^{th} = \boxed{55}$

$QD = Q3 - Q1/2 = 55 - 40/2 = 7.5$

(IQR) Inter quartile range $= Q3 - Q1 = 15$

— Grouped data

Age	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of people	50	70	100	180	150	120	70	60
cf	50	120	220	400	550	670	740	800

N = ~~[illegible]~~ 800

$$Q1 = l_1 + \frac{(l_2 - l_1) \times (N/4 - cf)}{f}$$

Take median class consider N/4

$$= 30 + \frac{(35 - 30)(200 - 120)}{100}$$

$$= 30 + \frac{5 \times 80}{100} = 30 + 4 = \boxed{34}$$

$$Q2 = l_1 + \frac{(l_2 - l_1)(N/2 - cf)}{f}$$

$$= 35 + \frac{(40 - 35)(400 - 220)}{180}$$

$$= 35 + \frac{5 \times 180}{180} = \boxed{40}$$

$$Q3 = l_1 + \frac{(l_2 - l_1)(3N/4 - cf)}{f}$$

$$= 45 + \frac{(50 - 45)(600 - 550)}{120}$$

$$= 45 + \frac{5 \times 5\cancel{0}}{12\cancel{0}}$$

$$= 45 + 25/12)$$

$$= 45 + 2.0833$$

$$= 47.0833$$

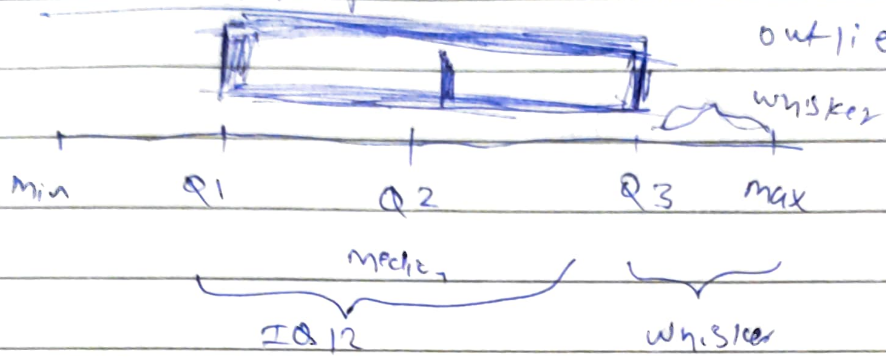
$$IQR = 47.0833 - 34 = 13.0833$$

$$QD = \frac{13.0833}{2} = 6.5\cancel{4}4$$

Box Whisker plot :

[5 number summary]

(max, min
Q1, Q2, Q3)



Q. consider 6.0, 6.7, 3.8, 7.0, 5.8, 9.975, 10.5, 5.99, 20.0

Asc. 3.8, 5.8, 5.99, 6.0, 6.7, 7.0, 9.975, 10.5, 20.0

$N = 9$

$Q2 = \text{Median} = n+1/2$ observation $= 5^{th} = 6.7$

$Q1 = (n+1/4) = 2.5^{th} = 5.8 + 5.99/2 = 5.895$

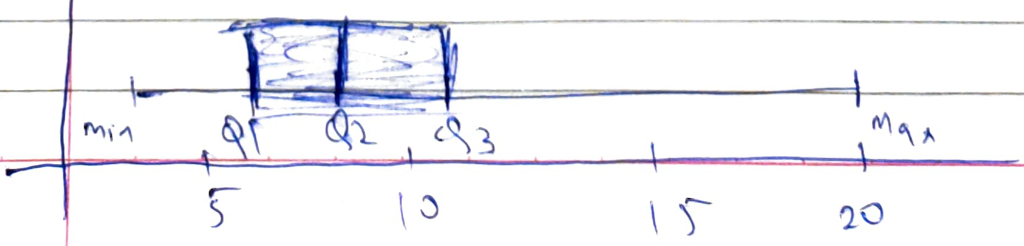
OR consider Median of 1st 4 values

$5.8 + 5.99/2 = \underline{5.895}$

$Q3 = 3(n+1)/4 = 7.5^{th} = \frac{9.975 + 10.5}{2} = \underline{10.2375}$

$IQR = Q3 - Q1 = 10.2375 - 5.895 = 4.3425$

$QD = 2.17125$



- Mean Deviation - Absolute dispersion

A. $MD = \dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$ (ungrouped)

$\dfrac{\sum_{i=1}^{n} f_i \, |x_i - x_1|}{N}$ (grouped)

Q. 5, 6, 9, 11, 12, 13, 14

$\bar{x}$ = mean = $(5+6+9+11+12+13+14)/7 = 10$

MD => $x_1 - \bar{x} = |5-10| = 5$

$\lvert x_2 - \bar{x} \rvert$ =	4
3	1
4	1
5	2
6	3
7	4

$MD = \dfrac{5+4+1+1+2+3+4}{7} = 20/7 = 2.57$

20/7

DOMS Page No.
Date / /

1. Standard Deviation (SD)

population SD

sample SD

PSD $\sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n}}$

ungrouped

SSD $\sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n-1}}$

PSD $\sqrt{\dfrac{\Sigma f_i (x_i - \bar{x})^2}{n}} = \boxed{\sqrt{\dfrac{\Sigma f_i x_i^2}{n} - \bar{x}^2}}$ use this

grouped

SSD

$\sqrt{\dfrac{\Sigma f_i (x_i - \bar{x})^2}{n-1}}$

$= \sqrt{\dfrac{\Sigma f_i x_i^2}{n-1} - \bar{x}^2}$

2 Variance $= (SD)^2$

population variance $= \dfrac{\Sigma f_i (x_i - \bar{x})^2}{n}$ grouped

~~Sample~~ $= \dfrac{\Sigma (x_i - \bar{x})^2}{n}$ - ungrouped

Q.

Class interval	f_i	Class mark x_i	x_i^2	$f_i x_i^2$	$f_i x_i$
0-10	11	5	25	275	55
10-20	15	15	225	3375	225
20-30	25	25	625	15625	625
30-40	12	35	1225	14700	420
40-50	7	45	2025	14175	315
	70			48150	1640

$$\bar{x} = \frac{\Sigma f_i x_i}{n} = \frac{1640}{70} = 23.42$$

$$SD = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{48150}{70} - (23.42)^2}$$

$$= 11.78$$

$$var = SD^2 = (11.78)^2 = 138.76$$

* **Relative Dispersion**

Coeff. of variation (CV)

$= \frac{\text{Sample S.D}}{\text{Sample Mean}}$

dimensionless quantity

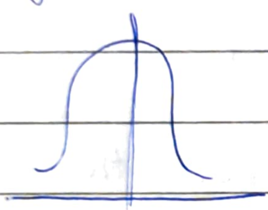
$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \text{variance}$$

$$CV = \frac{s}{\bar{x}} \times 100 \quad (\text{in } \%)$$

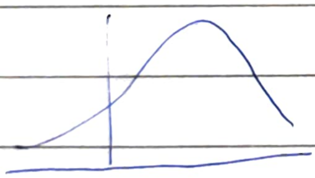
$$\text{Coeff. of variation} = \frac{\left(\sqrt{\frac{\sum_{i=1} (x_i - \bar{x})^2}{n-1}}\right)}{\bar{x}} \times 100$$

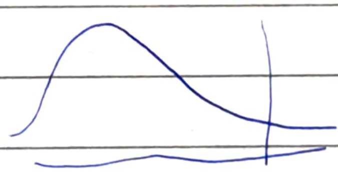
Unit 2

Symmetric graph



mean = median = mode

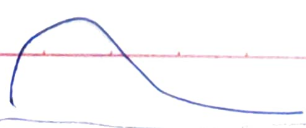
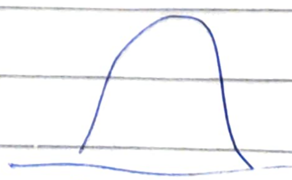
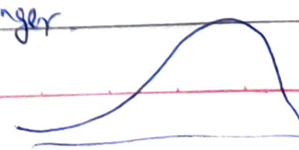




negative skewness (Assymmetric)

positive skewness (Right tailed graph)

Skewness - is the measure of assymmetry in probability distribution (Symmetry about it's mean)

+ve skew	Symmetric	-ve skew
Right tail is longer 	 Both equal	The left tail is longer 

+ve skew	Symmetric	-ve skew
- Mass concentrated on left		- Mass of distribution concentrated on right
- Mean > median	Mean = Median	- Mean < Median

Karl Pearson coefficient to measure skewness (SK_p)

$$\boxed{SK_p = (\text{Mean} - \text{Mode}) / S.D}$$

Q.1. 43, 48, 38, 46, 50, 48, 47, 48, 62, 48

Note: If $SK_p < 0 \Rightarrow$ -ve skew

$SK_p = 0 \Rightarrow$ Symmetric

$SK_p > 0 \Rightarrow$ +ve skew

ANS. mode = 48

mean = 47.8

$$SD = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} \qquad \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}$$

$|x_i - \bar{x}| = 4.8 + 0.2 + 9.8 + 1.8 + 2.2 + 0.2 + 0.8 + 0.2 + 14.2$
$+ 0.2 = 35$

$$SD = \sqrt{\frac{35^2}{10}} \qquad \sqrt{122.5} = 11.067$$

5.74

DOMS Page No.
Date / /

$Sk_p = 47.8 - 48 / 5.74108$
$= -0.03489$
< 0
-very skew...

Calculate Sk_p for soll.

Daily wages	No. of workers	x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
400-500	8	450	3600	202500	1620000
500-600	16	550	8800	302500	4840000
600-700	20	650	13000	422500	8450000
700-800	17	750	12750	562500	9562500
800-900	3	850	2550	722500	2167500
Σ	64		40700		26640000

$\boxed{\frac{\Sigma x_i f_i}{f} = 635.9375}$

$$\text{Mode} = l_1 + \frac{(l_2 - l_1)(f_1 - f_0)}{(2f_1 - f_0 - f_2)}$$

$l_1 = 600$, $l_2 = 700$, $f_1 = 20$, $f_0 = 16$, $f_2 = 17$

$$\text{Mode} = 600 + \frac{100(4)}{2(20) - 16 - 17} = 600 + \frac{400}{7} =$$
$$= 657.1428571$$

$$SD = \sqrt{\frac{\Sigma f_i x_i^2}{n} - (\bar{x})^2}$$

$$= \sqrt{\frac{26640000}{64} - 404416.50}$$

$$= \sqrt{416250 - 404416.50}$$

$$= 11,833.5$$

$$Sk_P = \frac{635 - 657}{11,833} = -1.85 < 0$$

* **Bowley's coeff. of skewness $(Sk)_B$**

$$Sk_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

Q. Calculate Sk_B for foll.

Life in hrs	<10	20-30	30-40	40-50	50-60
No of bulbs	12	18	24	20	6
cf <	12	30	54	74	80

for Q_1 $N/4 = 20$ 20-30 is median class

DOMS Page No.
Date / /

$$Q1 = l_1 + (l_2 - l_1)\frac{(N/4 - cf)}{f}$$

$$= 20 + (30 - 20)\frac{(20 - 12)}{18} = 20 + 10\frac{(8)}{18} = 30$$

$$Q2 = l_1 + (l_2 - l_1)\frac{(N/2 - cf)}{f} = 24.444$$

$$= 30 + 10\frac{(40 - 30)}{24} = 30 + \frac{100}{24} = 34.1666$$

$$Q3 = l_1 + (l_2 - l_1)\frac{(3N/4 - cf)}{f}$$

$$= 40 + 10\left(\frac{60 - 54}{20}\right) = 40 + 10\left(\frac{6}{20}\right) = 43$$

$$SK_B = \frac{(Q3 - Q2) - (Q2 - Q1)}{(Q3 - Q2) + (Q2 - Q1)}$$

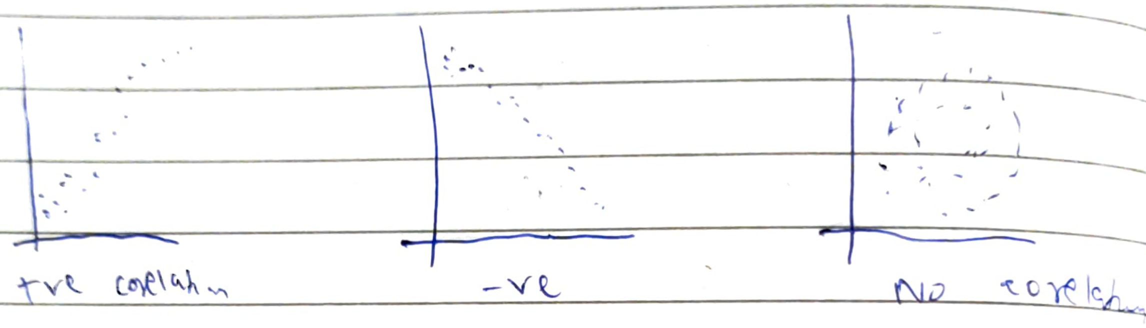
$$= \frac{(43 - 34.1666) - (34.1666 - 24.444)}{(43 - 34.1666) + (34.166 - 24.44)}$$

$$= \frac{8.83334 - 9.72222}{8.83334 + 9.72222}$$

$$= \frac{-0.88888}{18.55554} = -0.0479$$

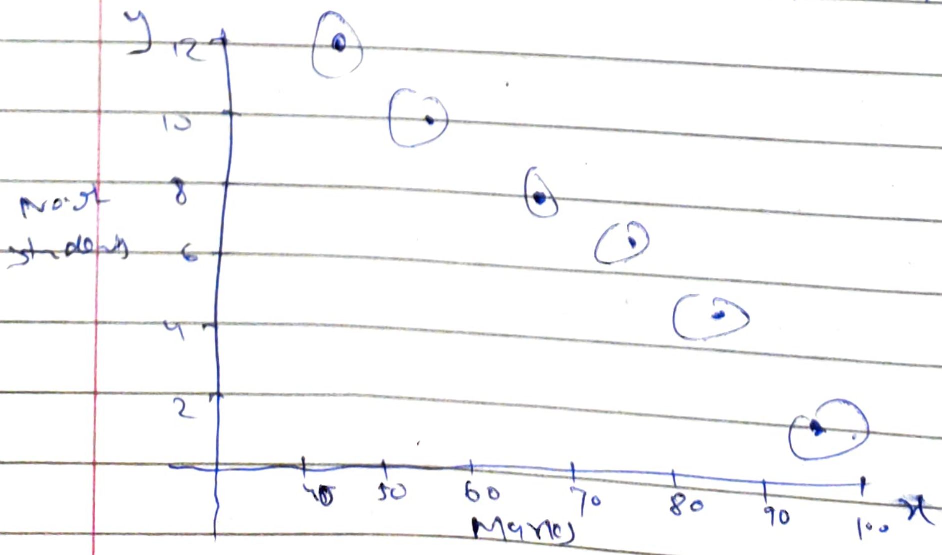
$\therefore$ It is -ve skewed

* Scatterness



Q. Draw scatter diagram of the foll. & understand the type of corelation between them.

Marks (out of 100)	No. of students (y)	Class mark (x)	
40-50	12	45	(4
50-60	10	55	(55, 1
60-70	8	65	(65, 8
70-80	7	75	(75, 7
80-90	5	85	(85, 5
90-100	2	95	(95, 2



-ve corelation

Corelation Coefficient

1) Karl Pearson coeff. of relation

(r) OR (r_{xy})

$= \sqrt{b_{xy}\, b_{yx}}$

2) Spearman Rank coeff.

when $b_{xy} = \dfrac{\Sigma xy - \dfrac{\Sigma x \cdot \Sigma y}{N}}{\Sigma x^2 - \dfrac{(\Sigma x)^2}{N}}$

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

properties:

1) $r \in [-1, 1]$

2) r is unitless quantity

values of r

	+ve	-ve
weak	(0.1) to (0.3)	(-0.1) to (-0.3)
Avg	(0.3) to (0.5)	(-0.3) to (-0.5)
strong	(0.5) to (1.0)	(-0.5) to (-1.0)

eg $r = -0.2 \Rightarrow$ negatively weak.

Q. An experiment of 9 diff smoking subjects resulting in
Calculate correlation coeff b/w [illegible] &
[illegible]

Cigarette smoked per week (x)	No. of years lived (y)	x^2	y^2	xy
25	63	625	3969	1575
35	68	1225	4624	238
10	72	100	5184	720
40	62	1600	3844	2480
85	65	7225	4225	5525
75	46	5625	2116	3450
60	51	3600	2601	3060
45	60	2025	3600	2700
50	55	2500	3025	2750
Σ 425	542	24525	33188	2464

$N = 9$ (No. of x)

$(\Sigma x)^2 = 180625$

$(\Sigma y)^2 = 293764$

$$b_{xy} = \frac{\Sigma xy - \Sigma x \Sigma y / N}{\Sigma x^2 - (\Sigma x)^2 / N} = \frac{24640 - (425 \times 542)/9}{24525 - (180625)/9}$$

$$= \frac{24640 - 25594.444}{24525 - 20069.444} = -\frac{954.444}{4455.551}$$

$$= -0.2142$$

DOMS Page No.
Date / /

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

$$= \frac{-954.444}{33188 - (293764)/9}$$

$$= \frac{-954.444}{33188 - 32640.444}$$

$$= \frac{-954.444}{547.556} = -1.7430$$

$$r = \sqrt{b_{xy} \cdot b_{yx}}$$

$$= \sqrt{-0.2142 \times -1.7430}$$

$$= \sqrt{0.37337}$$

$$= 0.611040$$

= strong +ve.

② Spearman's Rank correlation coeff.

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

Q. The foll. table provides data about % of students who have pre university meals & their cgpa scores. Calc. Spearman's rank correlation b/w the 2 data & interpret the result

University	% of students having Premeals	% of students scoring cgpa			
Pune	14.4	54	3	4	1
Chennai	7.2	64	2	6	16
Delhi	27.5	44	5	3	4
Kanpur	33.8	32	6	1	25
Ahmedabad	38.0	37	7	2	25
Indore	15.9	68	4	7	9
Guwahati	4.9	62	1	5	16
	x	y	dx	dy	dx-d

$\Sigma d^2 = 96$

(Choose asc. or descending rank dx & dy)

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \qquad n = 7$$

$$= 1 - \frac{6 \times 96}{7(49-1)} = 1 - \frac{6 \times \cancel{96}\,2}{7 \times \cancel{48}} = \frac{1-12}{7}$$

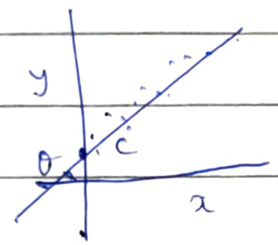
$$= \frac{-5}{7} = -0.714$$

Strong -ve correlation

Linear Regression

$y = mx + c$

$m \to$ slope $\quad c \to$ intercept
($\tan\theta$)



Q. The experimental values relating to centripetal force & radius for a mass travelling at a constant velocity in a circle are as shown below:

(y) Force (N)	5	10	15	20	25	30	35	40
(x) Radius (cm)	55	30	16	12	11	9	7	5

Determine the eq's of regression line of force on radius & hence calculate force at radius 40cm

$y = mx + c \to$ regression line of y on x

y on x

$$\boxed{\begin{aligned} \Sigma y &= a_0 N + a_1 \Sigma x \\ \& \quad \Sigma xy &= a_0 \Sigma x + a_1 \Sigma(x^2) \end{aligned}}$$

$a_0 = m \quad a_1 = c$ from $y = mx + c$

$\boxed{y = a_0 x + a_1}$

$N = 8 \quad \Sigma x = 145$

$\Sigma y = 180 \quad \Sigma y^2 = 5100$

$\Sigma x^2 = 4601 \quad \Sigma xy = 2045$

$\Sigma y = a_0 N + a_1 \Sigma x$

$\Sigma xy = a_0 \Sigma x + a_1 \Sigma(x^2)$

$\therefore$ $180 = 8a_0 + 145 a_1$ — (1)

$2045 = 145 a_0 + 4601 a_1$ — (2)

Solving, $a_0 = 33.7$ $a_1 = -0.617$

$y = a_0 + a_1 x$

$\therefore$ $\boxed{y = 33.7 - 0.617x}$

force at radius 40 =

$y = 33.7 - 0.617 \times 40$

$33.7 - 24.68 = 9.02$

Q In a study between amount of rainfall & quantity of air pollution removed, foll. data were collected

x	Daily rainfall in 0.01 cm	4.3	4.5	5.9	5.6	6.1	5.2	3.8
y	Pollution removed mg/m^3	12.6	12.1	11.6	11.8	~~11.0~~ 11.4	11.8	13.2

Find regression line of y on x

Air pollution depends on rainfall $\therefore$ Take x as daily rainfall & y as pollution.

Advanced Engineering Maths - HK Das

x	y	xy	x^2
4.3	12.6	54.18	18.49
4.5	12.1	54.45	20.25
5.9	11.6	68.44	34.81
5.6	11.8	66.08	31.36
6.1	11.4	69.54	37.21
5.2	11.8	61.36	27.04
3.8	13.2	50.16	14.44
2.1	14.1	29.61	4.41
Σ 37.5	98.6	453.82	188.01

0.037

1000

0.00037

1000

0.00

$\Sigma y = a_0 N + a_1 \Sigma x$

$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2$

$98.6 = 8a + 37.5 a_1$ — (1)

$455.82 = 37.5a + 188.01 a_1$ — (2)

(1) × 4.6875

175.78

$462.1875 = 37.5a + \cancel{17332.03125}$

$- \; 455.82 = 37.5a + 188.01 a_1$

$6.3675 = -12.23$

$6.3675 = 17144.02 a_1$

0.00037

DOMS Page No. Date / /

~~$a_1 = 15.49$~~ $a_0 = -0.675$

~~$y = a_0 x + a_1$~~

~~$y = 15.49x + (-0.675)$~~

~~$y = 15.49 + (-0.675)x$~~

$y = a_0 + a_1 x$

$y = 15.49 - 0.675x$

Probability

Exclusive - $A \cap B = \phi$

Exhaustive - $A \cup B =$ Sample set

Compliment - Exclusive & exhaustive both

$A \cap B = \phi$ & $A \cup B = S$

Exp. Tossing 2 coins

$\Omega = \{HH, HT, TH, TT\}$

$x =$ no. of heads

$X(HH) = 2$; $X(HT) = X(TH) = 1$; $X(TT) = 0$

$X = \{0, 1, 2\}$

$X : \Omega \to R$

mapping X is a Random variable