Date : / /
Page No.

17/08

Central Tendancy : A measure of central tendancy is a value around which all the observations have a tendancy to cluster.

(i) Arithmatic mean (ii) Geometric mean
(iii) Median (iv) Mode
(v). Harmonic mean

Mean ($\bar{x}$) : $\bar{x} = \sum_{i=1}^{m} \frac{x_i}{m}$

$$\bar{x} = \sum_{i=1}^{k} \frac{f_i x_i}{f_i}$$

Weighted Mean : $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ (weight of x_i is w_i)

Que

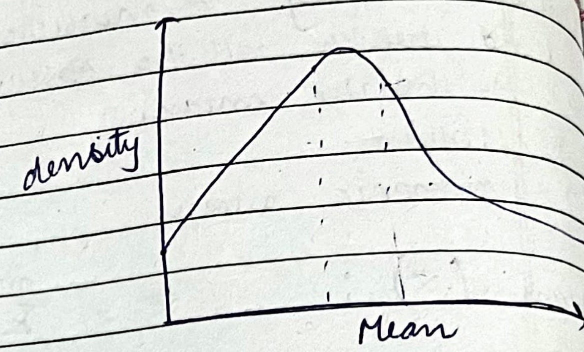
days	Mon	Tue	Wed	Thu	fri	Sat	Sun
Pocket Money	10	15	5	10	15	20	60

- average ~~mean~~ will help in gaussian distribution only
- Median is better indicator than mean if :
  - % of outliers from total sample is higher than expected.
  - distance between mean and outliers is greater than expected
  - distribution of values is still gaussian (normal) but skewed to one side.

variance : $\sum_{i=1}^{m} f \frac{(x_i - \bar{x})^2}{(n-1)}$

- $n-1$ for small samples (for higher no. of samples $n-1 \approx n$)

Skewed



Que: Calculate the mean, variance and standard deviation for the following sample data:

class	0-10	10-20	20-30	30-40	40-50	50-60
freq.	27	10	7	5	4	2

$$\bar{x} = \sum_{i=1}^{6} \frac{f_i x_i}{f_i}$$

$$= \frac{925}{55} = 16.818$$

Measures of dispersion

(i) 5, 5, 5, 5, 5 (mean = 5, median = 5)

(ii) 3, 4, 5, 6, 7 (mean = 5, median = 5)

(iii) 1, 3, 5, 7, 9 (mean = 5, median = 5)

range, mean deviation & standard deviation are some standard measure of dispersion.

Date : / /
Page No.

Median : (i) for odd = $\left(\frac{n+1}{2}\right)^{th}$

(ii) for even = $\frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}\right]$

- Median isn't where items need to be assigned with relative importance and weight.

Mode :

Geometric mean. It is defined as n^{th} root of the product of the values of n items in a given list.

$$G.M. = \sqrt[n]{\pi x_i} = \sqrt[n]{x_1, x_2, \ldots x_n}$$

$$\pi = \text{product} = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

Que Investment which earns 10% first year, 50% second yr and 30% in third year. What is average rate of return?

Sol; Let money be x

$$x \times 1.10 \times 1.50 \times 1.30 = x \times f \times f \times f$$

$$1.10 \times 1.50 \times 1.30 = f^3$$

$$f = (2.145)^{1/3}$$

→ average of change : Application of G.M.

Mean Deviation : Mean deviation is the average of differences of the values of items from some averages of series.

$$MD\ (\delta x) = \frac{\sum |x_i - \bar{x}|}{N}$$

Date : / /
Page No.

Coefficient of mean deviation $= \frac{\delta x}{\bar{x}}$

Mean Deviation : It is less sensitive to extreme scores so it may not be able to identify outliers (non-normal deviation)

standard deviation : $\sigma = \sqrt{\frac{(x_i - \bar{x})^2}{N}} = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}$

Coefficient of SD $= \frac{\sigma}{\bar{x}}$

$$SD(s) = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\bar{x}$; sample mean
n ; no. of items in sample
Incomplete population

Variance : $(\sigma^2) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}$ [exhaustive population universe]

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

variance $(s^2) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$, sample

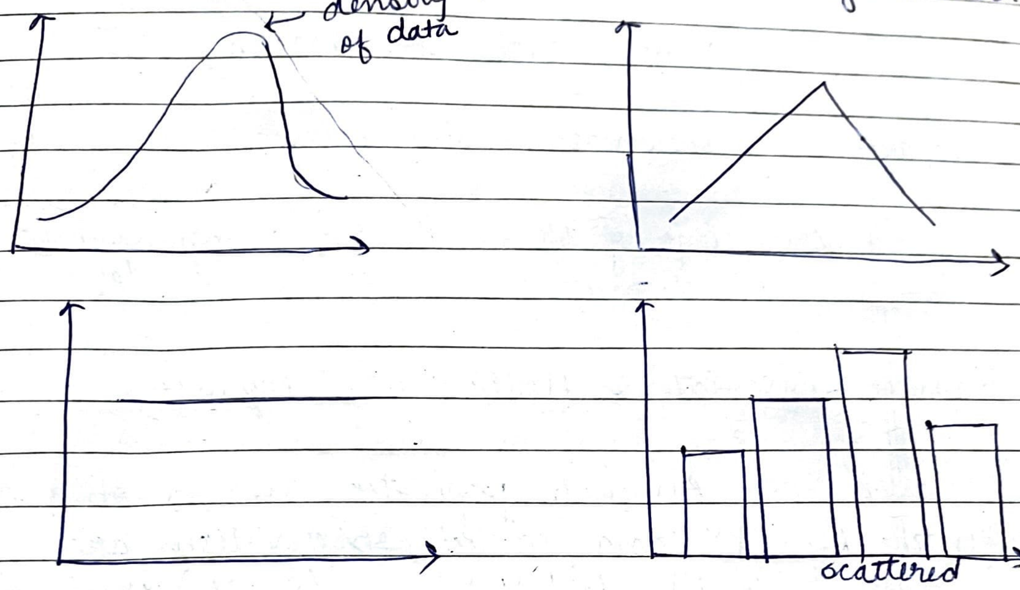
$\bar{x} \rightarrow$ estimated mean.

- Coefficient of Standard deviation
Relative measure & often used for comparison with similar series.

Date : / /
Page No.

Measure of skewness

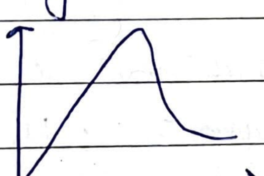
The meaning of skewness is lack of symmetry. skewness gives us the idea of the shape of the distribution of the data.



Skewed distribution : A dataset has a skewed distribution when mean, median, mode (if exist) are not the same.

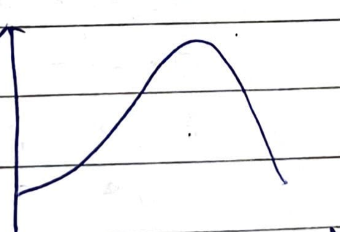
- In such case the plot of distribution is stretched to one side than to the other.

Skewness → +ve (curve is stretched towards right side more)



* [mode < Median < Mean]

Skewness → -ve (curve is stretched towards left side more)



* [mean < median < mode]

→ Measurement of skewness

skewness = mean - mode

• In case mode is ill defined then it can be estimated from mean & median.

Mode = 3 Median - 2 Mean.

→ comparison of skewness of two dataset

$$\text{coefficient of skewness } (S_k) = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

→ Univariate, Bivariate & Multivariate Population

→ Covariance: In Bivariate population two variables are represented by X, Y and paired observations are: $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$ then covariance

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

→ Population/Universe: collection of all items about which the info is ~~decto~~ decided is called Population. The population/Universe can be finite or infinite.

Parameters: Any characteristic or measure of population units is known as a parameter. Mean, population, SD are commonly studied parameters.

$P, \pi \rightarrow$ Population Proportion

$\sigma \rightarrow$ S.D

$\mu \rightarrow$ Mean

Date : / /
Page No.

Statistic : Any characteristic or measure of sample item is known as Statistics.

$\bar{x}$ = Sample mean

s = Sample SD

p = Sample proportion

→ Sampling and non-Sampling Errors

Sampling error arises on account of sampling and it generally happen due to random variations in the sample estimates around the true population values. Sampling error is inversly related to the size of sample.

• Non sampling error may occur during the processing of collecting actual information.

Stem and Leaf Plot

A Stem and leaf plot also k/a stem and leaf diagram is a way to arrange and represent data so that it is simple to see how frequently various data value occur.

→ Digits of data value ~~and~~ are divided into stem (first few digits) and a leaf (usually last digit).

→ Symbol "|" is used to split and illustrate stem and leaf values.

Stem	Leaf
2	0 0 1 2 5 7
3	1 3 8
4	3
5	8 9

Key 2|0 = 20

Date : / /
Page No.

Statistic : Any characteristic or measure of sample item is known as statistics.

$\bar{x}$ = Sample mean
s = sample SD
p = Sample proportion

→ Sampling and non-Sampling Errors

Sampling error arises on account of sampling and it generally happen due to random variations in the sample estimates around the true population values. Sampling error is inversly related to the size of sample.

• Non Sampling error may occur during the processing of collecting actual information.

Stem and Leaf Plot

A Stem and leaf Plot also k/a stem and leaf diagram is a way to arrange and represent data so that it is simple to see how frequently various data value occur.

→ Digits of data value ~~and~~ are divided into Stem (first few digits) and a leaf (usually last digit).

→ Symbol "|" is used to split and illustrate stem and leaf values.

Stem	Leaf
2	0 0 1 2 5 7
3	1 3 8
4	3
5	8 9

key 2|0 = 20

Date : / /
Page No.

Construction of Plot :

Step I : Classify data values in terms of no. of digits.

Step II : Fix the key for stem and leaf plot.

~~Step III~~ : eg : 2|5 = 25 , 3|2 = 3.2

Step III : consider 1st digit as stem and last digit as leaf.

Step IV : Sort the data from lowest to highest.

Step V : Place stem on ~~leaf~~ left & leaf on right.

Step VI : List leaf values and column against the stem from lowest to highest.

Que : A table shows the derivation of calls that Rosy makes each day. Represent using stem and leaf plot.

Date	Minutes
9	56
9	3
9	6
10	14
10	19
10	5
11	23
11	30
11	10
11	2
11	36.
10	23

Date : / /
Page No.

Sol: Step I : Sort the data (minutes)
02, 03, 05, 06, 10, 14, 19, 23, 23, 30, 36, 56

Step II : Choose tens digit for stem & ones digit for leaf.
Step III : Define the key.
Key : 3/6 = 36
Step IV : Write down stem on left and leaf on right

Phone call lengths

Stem	Leaf
0	2, 3, 5, 6
1	0, 4, 9
2	3, 3
3	0, 6
4	
5	6

Que: The stem and leaf plot below shows the scores of student find
(i) no. of students who scores less than 9 points.
(ii) no. of students who scored a min of 9.

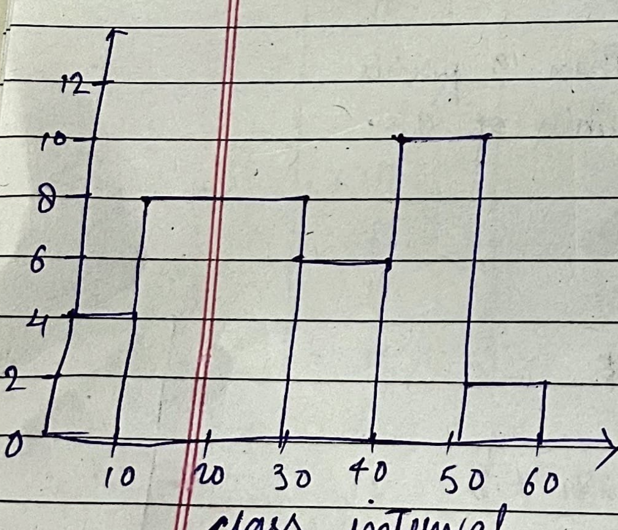
Key : 9/2 → 9.2 points

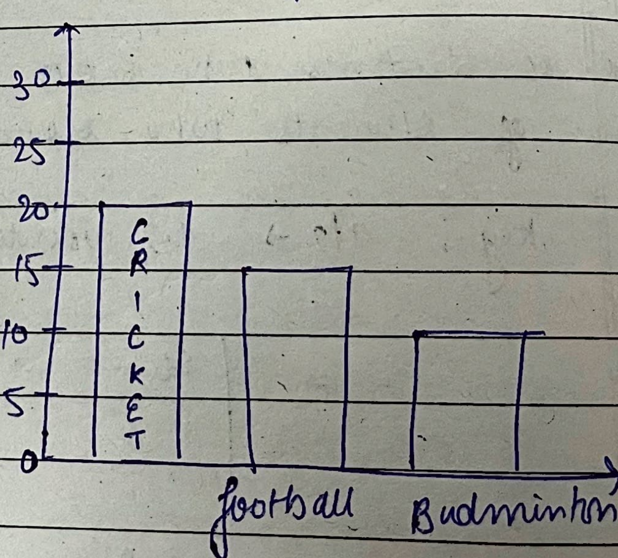
Stem	Leaf
6	6
7	0, 5, 7, 8
8	1, 1, 3, 4, 4, 6, 8, 8, 9
9	0, 2, 9
10	0

Histogram : A histogram is a graphical representation of a grouped frequency distribution with continous classes.

- It is an area diagram and can be defined as a set of rectangles with bases along with the interval class boundaries and with area proportional to frequencies in the corresponding classes.
- All the rectangles are adjacent.
- Heights of rectangles are proportional to corresponding frequency of similar class.

Histogram	Bar Graph
• It is a 2-D figure	• It is a 1-D figure.
• The frequency is shown by the area of each rectangle.	• The height shows the freq. & width has no significance
• It shows rectangle touching each other.	• It consists of rectangles seperated from each other with equal spaces

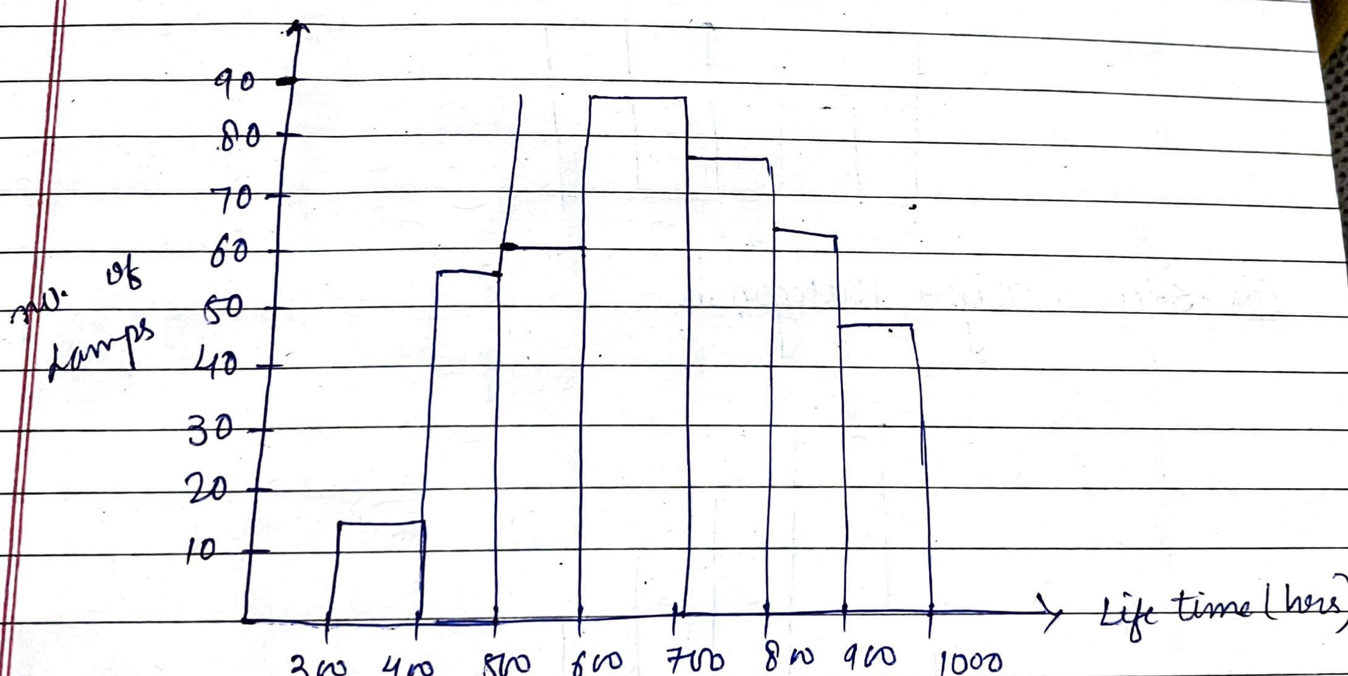




Date : / /
Page No.

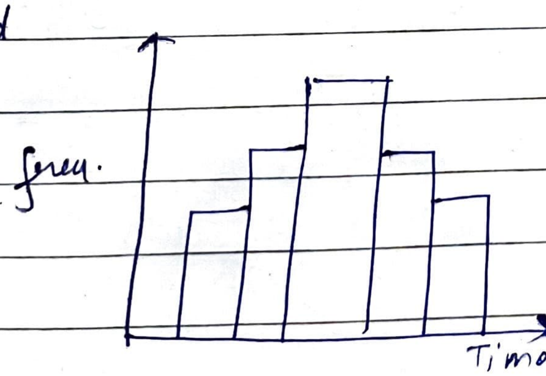
Que: Following table gives lifetime of 400 lamps. Draw the histogram of below data.

Life time (hrs)	No. of Lamps
300 - 400	14
400 - 500	56
500 - 600	60
600 - 700	86
700 - 800	74
800 - 900	62
900 - 1000	48



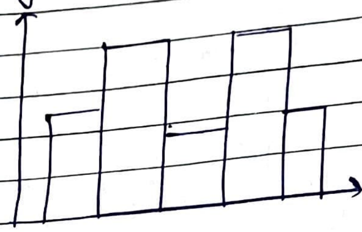
Histogram shapes

1) Bell Shaped

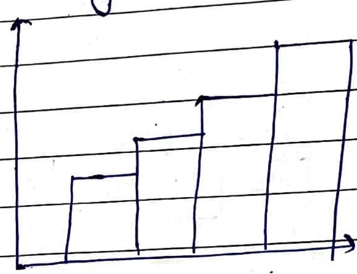


Date : / /
Page No.

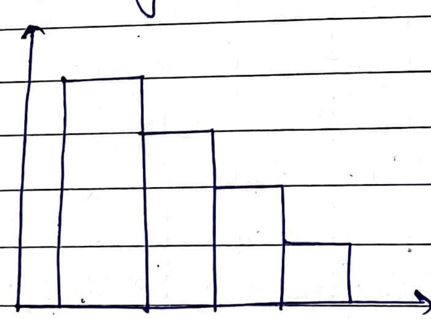
(ii) Bimodal Histogram



(iii) Skewed Left Histogram



(iv) Skewed Right Histogram



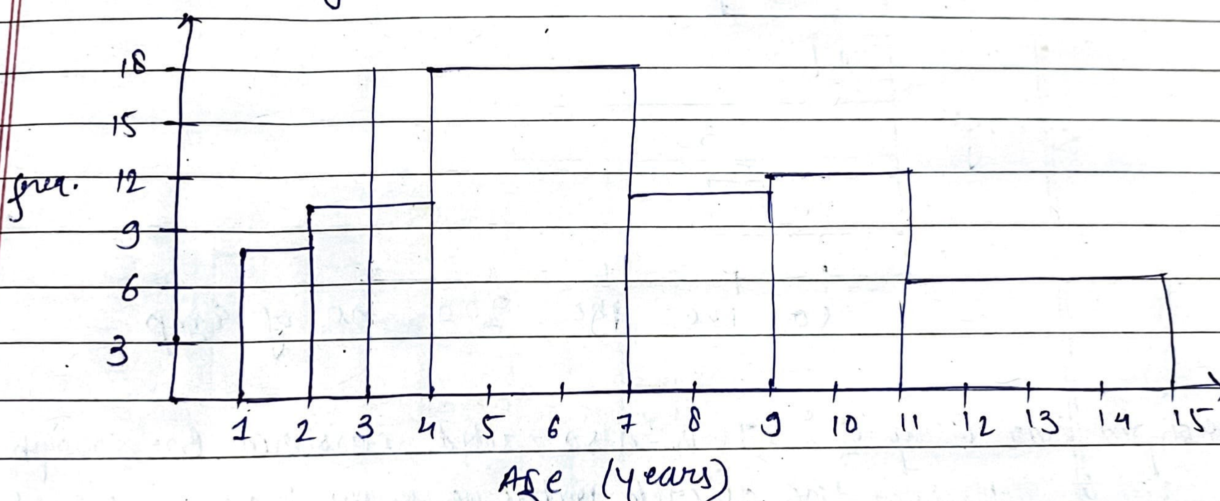
Que: A random survey on the no. of children belonging to diff. age groups who play in govt. parks and the info is tabulated as follow:

Date : / /
Page No.

Age (years)	frequency
1-2	8
2-4	10
4-7	18
7-9	10
9-11	12
11-15	6

(i) Draw a Histogram Representation of data
(ii) Identify no. of children belonging to age group 2,3,4,6,= who play in govt. Parks. −28

(i)

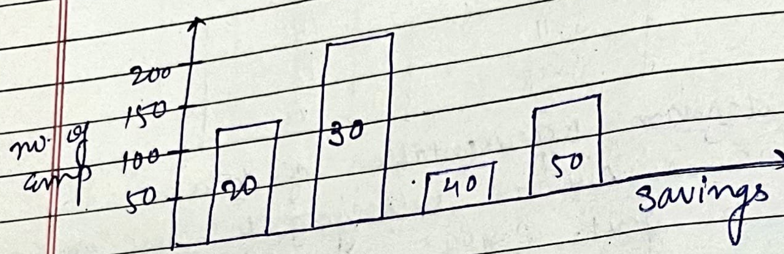


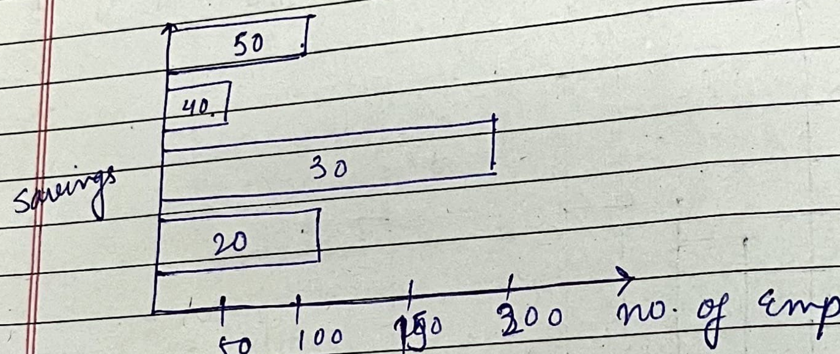
Bar Graph variations
1. Vertical Bar Chart
2. Horizontal Bar Chart } Main
3. Grouped Bar Graph
4. Stacked Bar Graph

Que: In a firm of 400 employees, the percentage of monthly salary saved by each employee is given in the following table. Represent through a bar graph.

Savings (in %)	20	30	40	50	Total
no. of Employees	105	199	29	73	400

Vertical

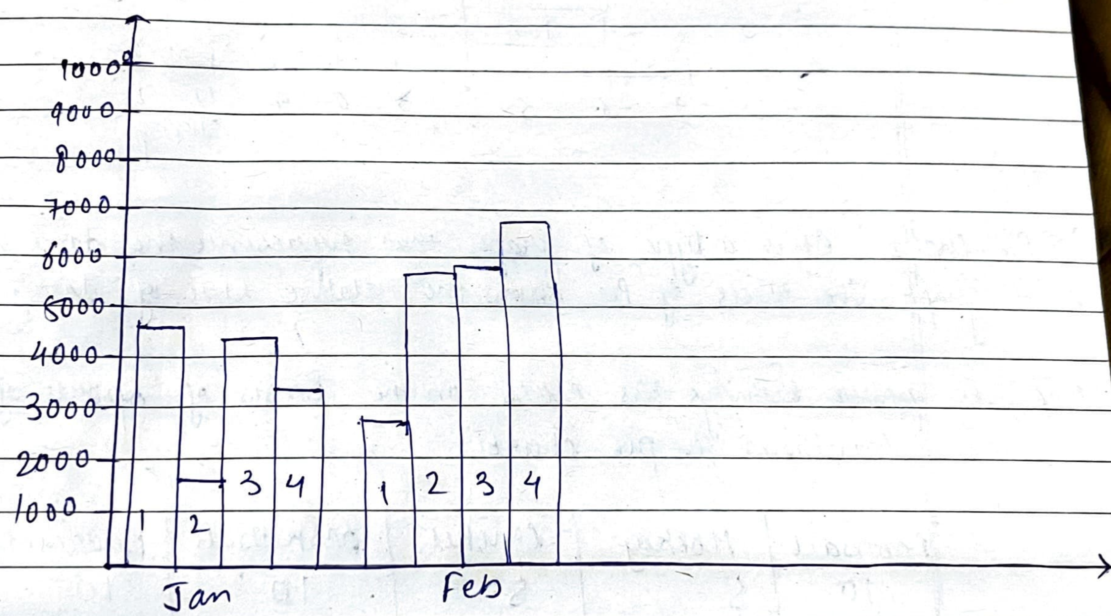




Grouped Bar Graphs : It is also called clustered Bar graph. It is used to represent the discrete value of more than one object that shares the same category.

Que A cosmetic company manufacture 4 different shades of lipsticks. The sale for 6 months is shown in table. Represent using Bar chart.

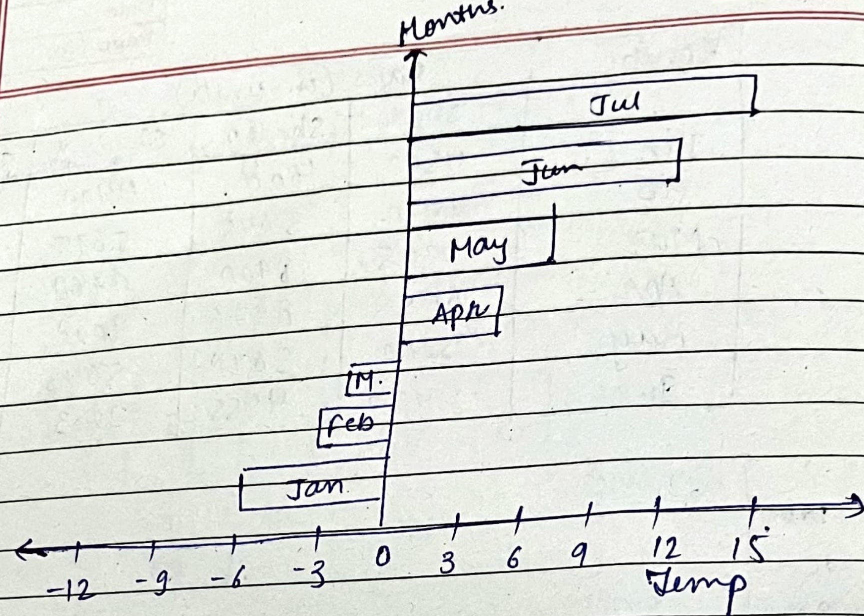
Month	Sales (in units)			
	Shade 1	Shade 2	Shade-3	Shade-4
Jan	4500	1600	4400	3245
Feb	2870	5645	5675	6754
Mar	3985	8900	9768	7786
Apr	6855	8976	9008	8965
May	3200	5678	5643	7865
June	3456	4555	2233	6547



Stacked Bar Graphs : It divides the aggregates into two diff. parts.

Que ; The variation of temperature in a region during a year is given below. Draw the Bar graph :

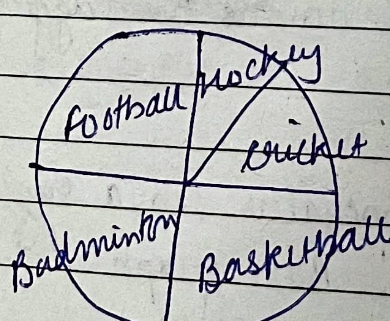
Month	Jan	Feb	March	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
Temp (°C)	-6	-3.5	-2.7	4	6	12	15	8	7.9	6.4	3.1	2.5



Pie Chart: It is a type of graph that represents the data in circle graph. The slices of Pie show the relative size of data.

Que: A teacher surveys his class on the basis of sports of student. Construct a pie chart.

Football	Hockey	Cricket	Basketball	Badminton	
10	5	5	10	10	
1/4	1/8	1/8	1/4	1/4	40



Date : / /
Page No.

Step I : Add values in table to get total (Total no of students = 40)
Step II To get percent, divide each value of sport by total & mul. by 100.

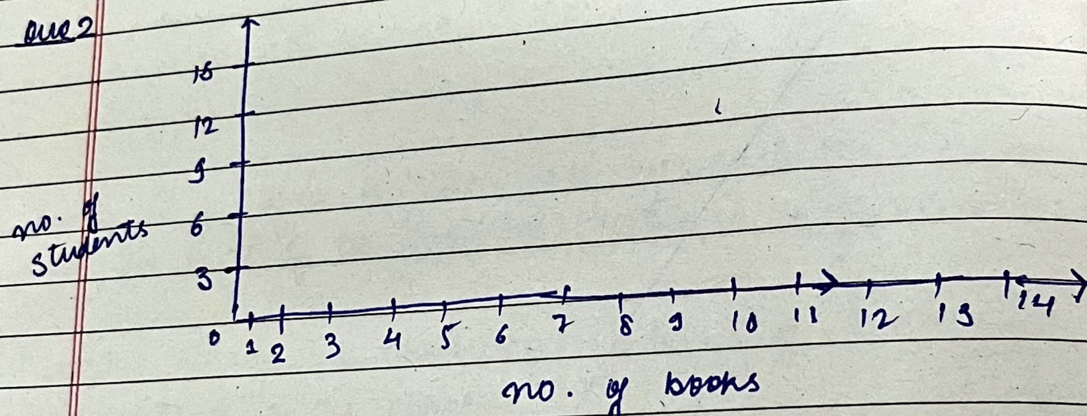
$$\frac{10}{40} \times 360 = 90.$$

Assignment.

Que 1.

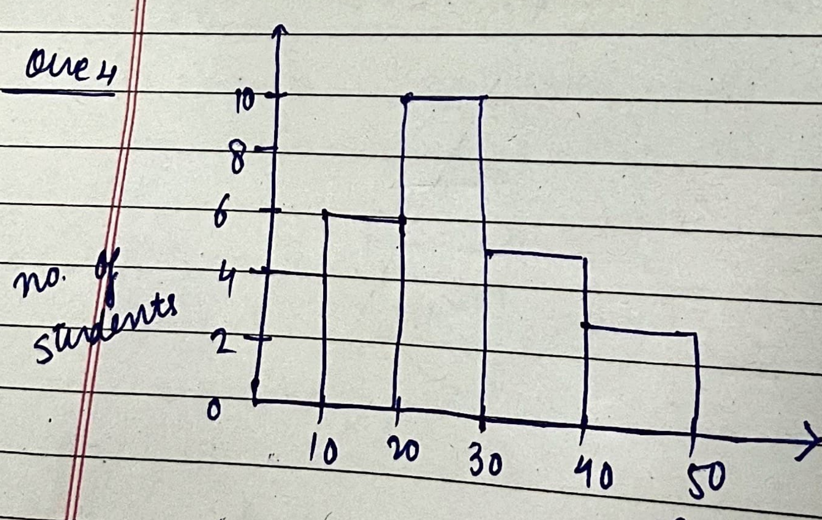
Histogram	Bar Graph
• It is a Two diamensional figure	• It is a one dimensional figure
• The frequency is shown by area of each rect.	• frequency is shown by the axis height and width has no significance.
• The rectangles are touching each other.	• The rectangles are seperated with equal spaces.

Que 2



Que 3. A bar graph is best used to represent categorical (discrete) data.

Que 4

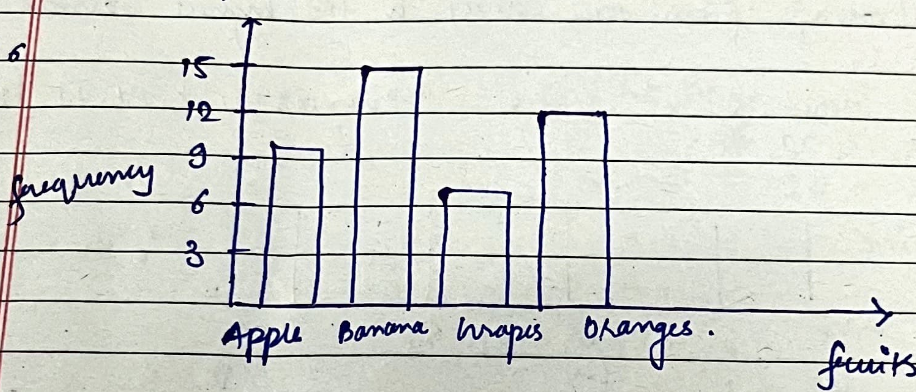


21-30 have highest frequency

Date : / /
Page No.

Que 5 A bar graph shows categorical data like 'cities' or 'months' where each category is seperate and independent from the others. while a histogram shows continuous data. The data is grouped into intervals & they are connected to each other. That's why there are no gaps between bars in a histogram but there are gaps in a bar graph.

Que 6

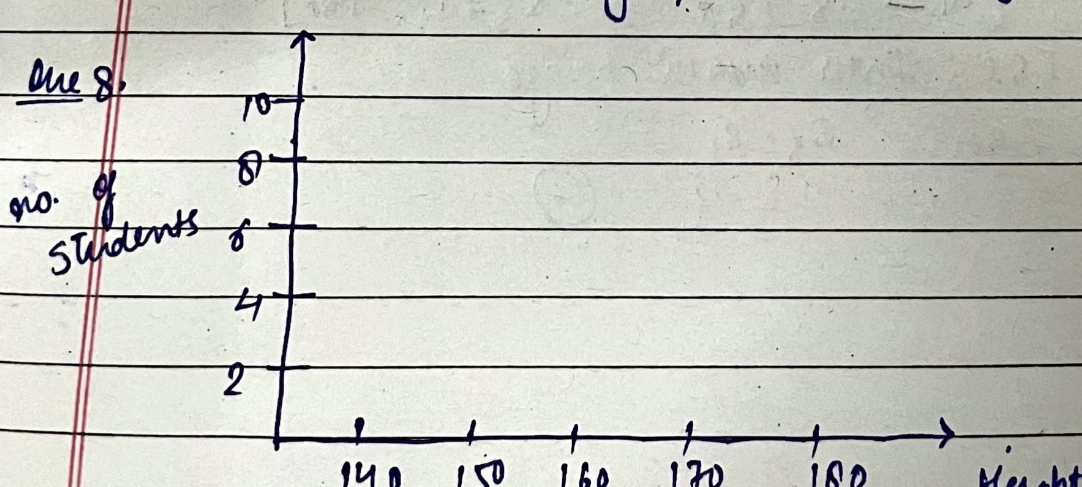


Total no of children surveyed = 44

Que 7. Similarity :

Difference : Histogram represents continuous data while bar graph is used for categorical data.

Que 8.



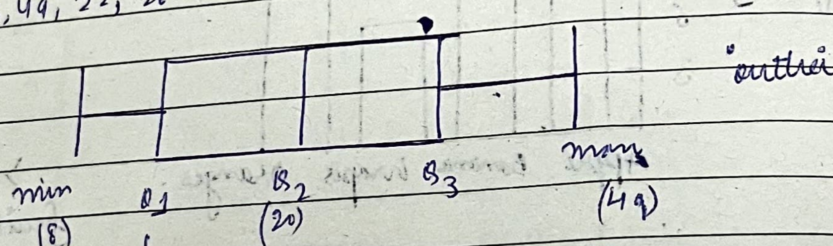
Date : / /
Page No.

BOX PLOT.

Obs.: When we display the data distribution in a Standardized way using 5 summary - minimum, Q_1 (first Quartile), median (Q_2), Q_3 (Third quartile) and maximum, it is called a Box plot.

Definition: A box plot is a special type of diagram that shows the quartiles in a box and the line extending from the lowest to the highest value.

Que: Let there is a series 11, 8, 10, 14, 13, 17, 24, 27, 29, 35, 49, 22, 20



Step 1: Arrange in ascending order

8, 10, 11, 13, 14, 17, 20, 22, 24, 27, 29, 35, 49

12 = Q_1 Q_2 = Mid (Median) Q_3 = 28

→ Check min (8) and max (49) are outliers?

→ Calculate outlier range

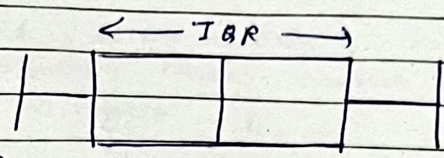
$$[Q_1 - 1.5\ IQR \text{ to } Q_3 + 1.5\ IQR]$$

IQR : Inter Quarter range

$IQR = Q_3 - Q_1$

$= 28 - 12 = 16$

$Q_1 \rightarrow$ 25th percentile, Q_3 - 75th percentile



min $\quad$ Q_1 $\quad$ Q_2 $\quad$ Q_3 $\quad$ max

8 $\quad$ 12 $\quad$ = 20 $\quad$ 28 $\quad$ 44

Range = $Q_1 - 1.5\,IQR$ to $Q_3 + 1.5\,IQR$

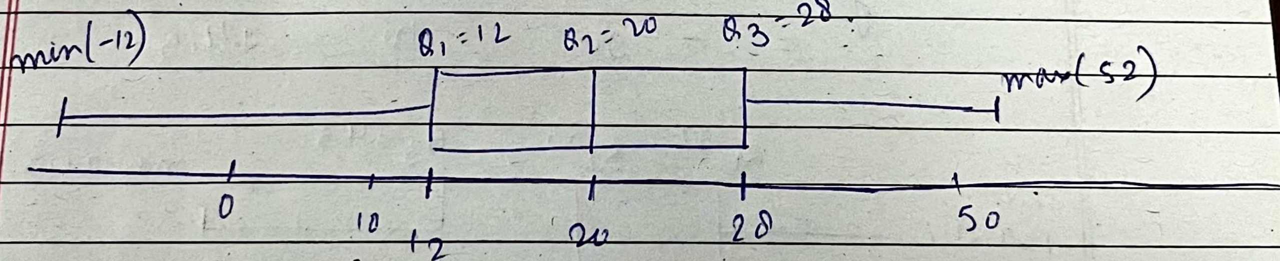
$= 12 - 1.5 \times 16$ $\qquad$ $28 + 1.5 \times 16$

$= 12 - 24 = -12$ $\qquad$ $28 + 24$

min $\qquad$ 52

max.

(−12 to 52)



Que 18, 34, 7.6, 29, 15, 41, 4.6, 25, 5.4, 38, 20, 32, 43, 22

Q_2 = median $= \frac{46+25}{2} = \frac{71}{2} = 35.5$

min

max

15, 18, 20, 22, 25, 27, 32, 34, 38, 41, 43, 46, 54, 76

$$\text{median} = \frac{32+34}{2} = \frac{66}{2} = 33$$

$Q_2 = 33$

min = 15

max = 76. ← outlier

$Q_1 = 22$

$Q_3 = 43.$

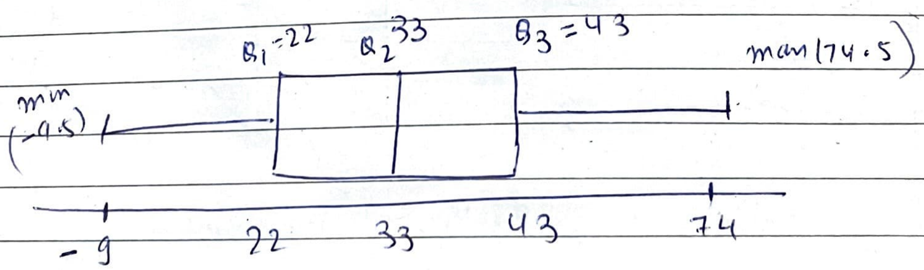
Range = $Q_1 - 1.5\,IQR$ to $Q_3 + 1.5\,IQR$

$IQR = Q_3 - Q_1 = 21$

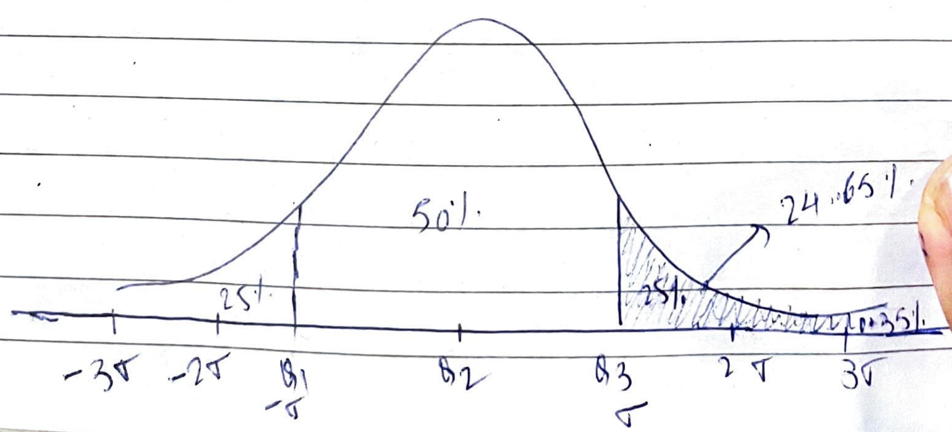
$1.5 \times 21 = 31.5$

Range = $22 - 31.5$ to $43 + 31.5$

$= -9.5$ to 74.5



➤ Box Plot on a Normal distribution



σ : SD

$\boxed{-3\sigma = \mu - 3\sigma}$

Positively sk…
greater than…
is positive…

Applications…
- Outliers…
- Symmetry…
- Tight gr…
- Data

Date: / /
Page No.

$(\mu - 2\sigma$ to $\mu + 2\sigma) \rightarrow 95\%$ of data is covered.

Positively skewed : If the dist. from the median to max is greater than distance from median to min then Box plot is positively skewed.

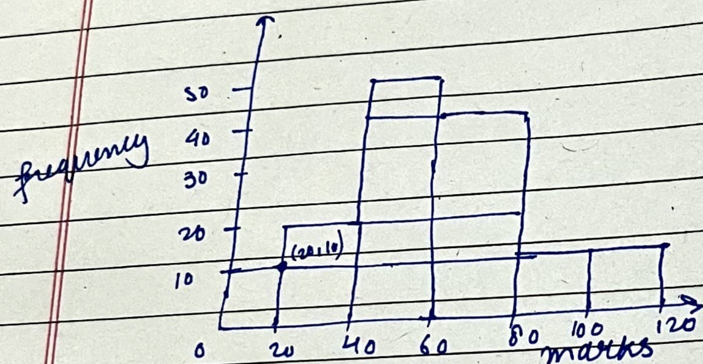
Applications
- Outliers and their values
- Symmetry of data
- Tight grouping of data
- Data Skewness

Date : / /
Page No.

Que Draw a cumulative frequency graph (ogive) from the histogram showing the distribution of marks in a math test for 140 students maximum marks is 120.

(i) Find median, lower quartile and upper quartile.

(ii) Draw box plot



Sol. marks interval $= \frac{120}{6} = 20.$

	≤ 20	≤ 40	≤ 60	≤ 80	≤ 100	≤ 120
Cf	10	30	80	120	130	140
	(20,10)	(40,30)	(60,80)	(80,120)	(100,130)	(120,140)

Q_2: median = 50% of 140

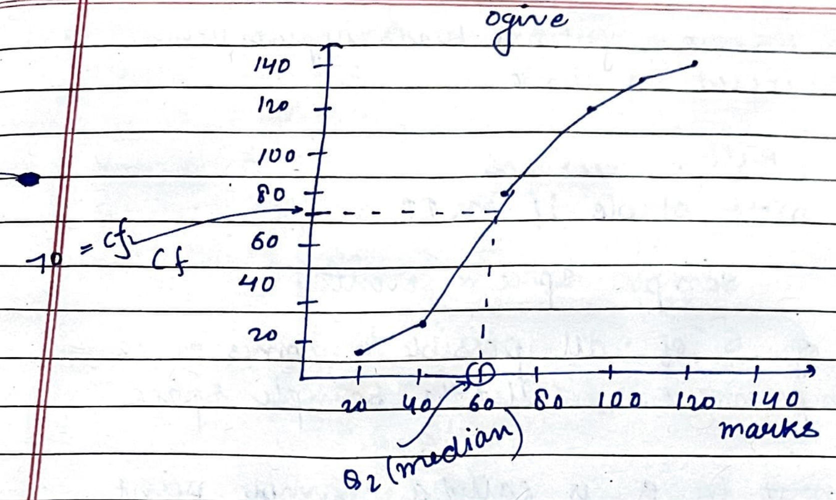
= 70

$70 \leftarrow Cf$

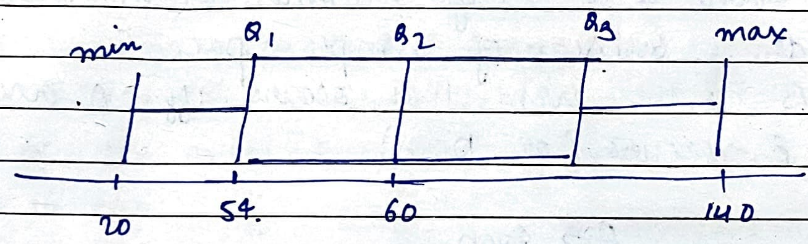
marks =

Rashmi

Date: / /
Page No.



Box Plot.



PROBABILITY

Probability is a mathematical modeling of the phenomenon of chance of randomness.

A probabilistic mathematical model of random phenomenon is defined by assigning probabilities to all the possible outcomes of an experiment.

Let a coin is tossed in a random manner there is 50% chance of getting head & 50% chance of getting tail.

Date : / /
Page No.

Let S be the no. of times heads appear when a coin is tossed n time

Rate

becomes more stable if $n \uparrow\uparrow$

Sample Space & Events

- The set S of all possible outcomes of a given experiment is called the sample space.

- An element in S is called a sample point

- An Event is a set of outcomes or in other words a subset of sample space S.

A→ prime not ∈

(i) $A \cup B$ is the event that occurs iff A occurs or B occurs (or both)

A → Even

B → prime

$A \cup B$ = all elements of A or B − $A \cap B$

(ii) $A \cap B$ is the event that occurs iff A occur & B occur.

(iii) A^c or $\bar{A}$ is the event that occurs iff A doesn't occur.

(iv) if $A \cap B = \phi$ ∴ mutually exclusive events A and B.

Date : / /
Page No.

Que Toss a dice and observe the numbers that appears on top.

$S = \{1, 2, 3, 4, 5, 6\}$

- Let A be the event an even no. occurs
B that odd no. occurs.
C - prime no. occurs

$A = \{2, 4, 6\}$
$B = \{1, 3, 5\}$
$C = \{2, 3, 5\}$
$A \cup C = \{2, 3, 4, 5, 6\}$
$B \cap C = \{3, 5\}$

Que Toss a coin 3 times and observe the sequence of H & T.

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

→ Let A be the event that two or more heads appear consecutively.

$A = \{HHH, HHT, THH\}$

→ B : all tosses are same

$B = \{HHH, TTT\}$

- Let S be a finite sample space

$S = \{a_1, a_2, \ldots, a_n\}$

a finite probability space or probability model is obtained by assigning two each point a_i in S a real no. p_i, called probability of a_i satisfy the following

(i) each P_i is non negative number ; $P_i > 0$

(ii) sum of P_i is 1

$P_1 + P_2 + \cdots P_n = 1$

→ The probability of event A written as P(A) is defined to be the sum of probabilities of the point in A.

$$P(a_i) = \text{Probability of } a_i$$

Que Let 3 coins are tossed and no. of heads are observed.

$$S = \{0, 1, 2, 3\}$$

$$\begin{matrix} \uparrow & \uparrow & \uparrow & \uparrow \\ a_0 & a_1 & a_2 & a_3 \end{matrix}$$

$P(0) = 1/8$

$P(1) = 3/8$

$P(2) = 3/8$

$P(3) = 1/8$

- Let A be the event atleast 1 head appears

  $P(A) = 7/8$

- Let A be an event all heads or all tails

  $P(A) = 2/8 = 1/4$

Date : / /
Page No.

12 face card

Que(2) Let a card be selected from any ordinary deck of 52 cards.

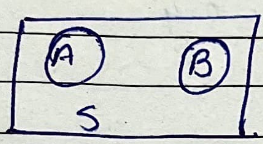
Let $A = \{$the card is a spade$\}$

Let $B = \{$card is a face card$\}$

$$P(A) = \frac{13}{52} = 1/4$$

$$P(B) = 12/52 = 3/13$$

Theorem



(i) for every event A, $0 \le P(A) \le 1$

(ii) $P(S) = 1$

(iii) If A and B are mutually exclusive
$P(A \cup B) = P(A) + P(B)$

(iv) $P(\phi) = 0$

(v) $P(A \setminus B) = P(A) - P(A \cap B)$ - $(A - B)$

(vi) $A \subseteq B$ then $P(A) \le P(B)$

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

Que Let a student is selected at random from 100 students where 30 taking maths, 20 chemistry, 10 taking both maths & chem. find $(A \cup B)$

- find the probab. that student taking mathematics or chemistry.

$P(A \cup B) = P$

Date: / /
Page No.

$P(M) = \frac{30}{100} = \frac{3}{10}$

$P(C) = \frac{20}{100} = \frac{2}{10}$

$P(M \cap C) = \frac{10}{100} = \frac{1}{10}$

$P(M \cup C) = P(M) + P(C) - P(M \cap C)$

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

$= \frac{4}{10} = 0.4$

JOINT PROBABLITY

common space, jointly two events, if mutually exclu-sive, Joint probability $= 0$

→ Joint probability refers statistical methods that calculates the likelihood of two events occuring together and at the same time.

→ Both events must be independent of each other means they are not conditional or don't rely on each other.

Que In Hospital two posts are advertised one for eye specialist and one for ENT total 5 applications are recieved for eye and 6 for ENT.

→ Two posts of MBBS doctors are advertised and total 11 application are received.

what is the probab that both nur

Date : / /
Page No.

Que 1 A box contain three red balls and two blue balls. Two balls are drawn without replace. find the Probab. that first ball is red & second ball is blue.

Que 2 In a class of 30 students 18 are boys & 12 are girls. 10 students are selected at random. what is the probab that 6 of them are boys and 4 are girls. $\simeq 0.3058$

Sol. A card is drawn from a standard deck of 52 cards. What is the probab that the given card is a king given that it is a face cards.

Que In a survey 40% of people like tea, 38% like coffee & 10% like both. what is the probability that a person like coffee given that the person likes tea.

Que Two dice are rolled what is the probab that first dice show an even no & second dice shows a no. greater than 4.

$$= \frac{3}{6} \times \frac{2}{6} = \frac{1}{6}$$

Que A Jar have 5 red balls & 3 green ball. One ball is drawn at random. If it is known that the ball is not green what is the probab that it is red.

Que A coup... point.
(i) If ...
(ii) If ...

Date : / /
Page No.

Que A couple has two children with prob 1/4 for each point. find the prob. that both children are boy

(1) If one of them is boy =

(ii) If older is the boy

Date : / /
Page No.

Que 10. A pair of dice is tossed. The sample space S consists 36 ordered pairs (a,b) where a and b can be any integer from 1 to 6. Find probability that one dice is 2 if sum is 6.

Sol. Find $P(A|E)$ where

$E = \{\text{sum is } 6\}$

$A = \{2 \text{ appears on at least one dice}\}$

$E = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$

$A = \{(2,4), (4,2)\}$

$A \cap E = \{(2,4), (4,2)\}$

$$P(A|E) = \frac{P(A \cap E)}{P(E)}$$

$=$

$S = 6^2 = 36$

$$P(E) = \frac{5}{36}$$

$$P(A) = \frac{2}{36}$$

$$P(E \cap A) = \frac{2}{36}$$

$$P(A|E) = \frac{2}{5}$$

Que 11. ~~A couple has two~~

A lot contains 12 items of which 4 are defective three are drawn at random from the lot one after another. find the probab that all are non-defective.

Date : / /
Page No.

$= \frac{\cancel{8}^{4}}{\cancel{12}} \times \frac{7}{\cancel{12}_{3}} \times \frac{\cancel{6}}{\cancel{12}_{2}}$

$= \frac{\cancel{8}^{\cancel{4}\,2}}{\cancel{12}_{2}} \times \frac{7}{11} \times \frac{\cancel{6}}{\cancel{10}_{5}}$

$= \frac{14}{55} = 0.25$

first item is non defective $= \frac{8}{12}$

Second $= \frac{7}{11}$

Third $= \frac{6}{10}$

Probability that all are non defective $= \frac{8}{12} \times \frac{7}{11} \times \frac{6}{10}$

$= 0.25$

RANDOM VARIABLE

- It may be convinient 1 to "H" and 0 to "T" for tossing of a coin.
- In tossing of a pair of dice, we may assign 1 to sum of two integers to the outcomes.
- Such an assignment of numerical values is called a random variable.

Definition : A random variable X is a root that assigns a numerical value to each outcome in a sample space S.

→ Let R_X denotes the sets of numbers assigned by random variable X and R_X is referred as range space.

Que : 12. A pair of dice is tossed. The sample space S consists of the 36 ordered pairs (a,b) where a and b can be any integers between 1 and 6.

$S = \{(1,1), (1,2), \dots (6,6)\}$

Let X assigned to each point in S the sum of the numbers then X is a random variable with range space $R_X = \{2,3,4,5,6,7,8,9,10,11,12\}$. Let Y assigned to each point the max of the two numbers then Y is a random variable with range space $R_Y = \{1,2,3,4,5,6\}$

Que 13: A Box contains 12 items of which three are defective. A sample of three items is selected from the box. The sample space S consist of the $^{12}C_3 (=220)$ diff. samples of size 3. Let X denote the no. of defective items in the sample. Then X is a random variable with range space $R_X = \{0,1,2,3\}$

PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

Let $R_X = \{x_1, x_2, \dots x_t\}$ be the range space of random variable X defined on a finite sample space S. Then X ~~indicates~~ induces an assignment of possibilities on the range space R_X as

$P_i = P(x_i) = P(X = x_i)$ = sum of probability of point

S whos

P

Que 14:

Th

T

S whose image is x_i.

x_1	x_2	. . .	x_t
P_1	P_2	. . .	P_t

$$P_i = P(x_i) = \frac{\text{No. of Points in S whose image is } x_i}{\text{No. of Points in S}}$$

Que 14:

There is only one item (1,1) whose sum is 2.

$$P(2) = \frac{1}{36}$$

There are two outcomes (1,2) and (2,1) whose sum is 3 so

$$P(3) = \frac{2}{36} = \frac{1}{18}$$

There are three outcomes (2,2), (1,3) and (3,1) whose sum is 4 so

$$P(4) = \frac{3}{36} = \frac{1}{12}$$

$$P(5) = \frac{4}{36} = \frac{1}{9}$$

$$P(6) = \frac{5}{36}$$

$$P(12) = \frac{1}{36}$$

The distribution of X

x_i	2	3	4	5	6	7	8	9	10	11	12
P_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Date: / /
Page No.

refer Que 13.

Que 15. Let X be the random variable

$^9C_3 = 84$ samples of size 3 with no defect.

$P(0) = \frac{84}{220}$

$P(1) = \frac{^9C_2 \times {^3C_1}}{220} = \frac{108}{220}$ samples of size 3 containing one defective item.

$P(2) = \frac{^9C_1 \times {^3C_2}}{220} = \frac{27}{220}$

$P(3) = \frac{^3C_3}{220} = \frac{1}{220}$

Distribution of X

x_i	0	1	2	3
P_i	84/220	108/220	27/220	1/220

~~Expectation of a random variable~~

~~Let X be a random variable~~
~~Let μ or μ_X is the mean of X~~
~~Let σ or σ_X is the standard dev- of X.~~
~~Mean μ is also called the expectation of X., written as $E(X)$.~~
~~Let $S = \{a_1, a_2, \dots a_m\}$, X is the random variable on sample space S.~~

The mean

x_i
P

Que 16: De
expect

The
pr

$\frac{6}{2^6}$ $\quad$ $^6C_2$

Date: / /

Page No.

The mean or expectation of x is defined as -

$$m = E(x) = x(a_1)\cdot P(a_1) + x(a_2)\cdot P(a_2) + \dots x(a_m)\cdot P(a_m) = \sum_{i=1}^{m} x(a_i)\, P(a_i)$$

x_i	x_1	x_2	...	x_m
P_i	P_1	P_2	...	P_m

$$\mu = E(x) = x_1P_1 + x_2P_2 + \dots + x_mP_m$$
$$= \sum x_iP_i$$

Que 16: Let a fair coin is tossed six times. Compute expectation of no. of heads.

Sol.

x_i	0	1	2	3	4	5	6
P_i	$1/2^6$	$6/2^6$	$15/2^6$	$20/2^6$	$15/2^6$	$6/2^6$	$1/2^6$

The no. of heads which can occur with their respective probabilities are as follow:-

So Expectation of no. of heads is

$$E(x) = 0\times\frac{1}{64} + 1\times\frac{6}{64} + 2\times\frac{15}{64} + 3\times\frac{20}{64} + 4\times\frac{15}{64} + 5\times\frac{6}{64} + 6\times\frac{1}{64}$$

$$E(x) = 3$$

Date : / /
Page No.

Que 17. Refer Que. 15.

The expectation of X or in other words, the expected no. of defective items in a sample of size 3 is

$$m = E(X) = 0\left(\frac{84}{220}\right) + 1\left(\frac{108}{220}\right) + 2\left(\frac{27}{220}\right) + 3\left(\frac{1}{220}\right)$$

$$= 0.75.$$

Que 18. There are 3 horses a, b, c are in race. Let their respective winning probabilities are $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$. X denotes the pay off function for the winning horse and let X pays £ 200, £600 or £900 according a, b, c win the race. Find the expected payoff for the race.

$$E(X) = X(A) \cdot P(A) + X(B) \cdot P(B) + X(C) \cdot P(C)$$

$$= 200 \times \frac{1}{2} + 600 \times \frac{1}{3} + \frac{1}{6} \times 900$$

$$= 450.$$

VARIANCE AND STANDARD DEVIATION of a random variable.

x_1	x_2	...	x_n
P_1	P_2	...	P_n

Date: / /
Page No.

Let X be a random variable with mean μ and following probability distribution

The variance $var(x)$ and standard deviation σ of x is

$$var(x) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \dots (x_n - \mu)^2 p_n$$

$$= \Sigma (x_i - \mu)^2 p_i$$

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

(μ: expectation)

$$\sigma = \sqrt{var(x)}$$

$$var(x) = x_1^2 p_1 + x_2^2 p_2 + \dots + x_n^2 p_n - \mu^2$$

$$var(x) = \Sigma x_i^2 p_i - \mu^2$$

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

Que 18 Let X denotes no. of time heads occurs when a ~~paired~~ fair coin is tossed 6 times. The distribution of X appears in question 16. where $\mu = 3$, find variance.

$$var(x) = 0^2 \times \frac{1}{2^6} + 1^2 \times \frac{6}{2^6} + 2^2 \times \frac{15}{2^6} + 3^2 \times \frac{20}{2^6}$$

$$+ 4^2 \times \frac{15}{2^6} + 5^2 \times \frac{6}{2^6} + 6^2 \times \frac{1}{2^6}$$

$$- 9$$

$$= \frac{6 + 60 + 180 + 240 + 150 + 36}{2^6} - 9$$

$$= \frac{672 - 576}{64} = \frac{96}{64} = 1.5$$

$$SD = \sqrt{var(x)} = \sqrt{1.5} = 1.224$$

Date : / /
Page No.

Que 17 Refer Que. 15.

The expectation of X or in other words, the expected no. of defective items in a sample of size 3 is

$$m = E(X) = 0\left(\frac{84}{220}\right) + 1\left(\frac{108}{220}\right) + 2\left(\frac{27}{220}\right) + 3\left(\frac{1}{220}\right)$$

$$= 0.75.$$

Que 18 There are 3 horses a, b, c are in race. Let their respective winning probabilities are $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$.

X denotes the pay off function for the winning horse and let X pays £ 200, £ 600 or £ 900 according a, b, c win the race. Find the expected payoff for the race.

$$E(X) = X(A) \cdot P(A) + X(B) \cdot P(B) + X(C) \cdot P(C)$$

$$= 200 \times \frac{1}{2} + 600 \times \frac{1}{3} + \frac{1}{6} \times 900$$

$$= 450.$$

VARIANCE AND STANDARD DEVIATION of a random variable.

x_1	x_2	...	x_n
P_1	P_2	...	P_n

Date : / /
Page No.

Let X be a random variable with mean μ and following probability distribution

The variance var(X) and standard deviation σ of X is

$$var(X) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \dots (x_n - \mu)^2 p_n$$

$$= \Sigma (x_i - \mu)^2 p_i$$

$$= E\left((X-\mu)^2\right)$$

(μ : expectation)

$$\sigma = \sqrt{var(X)}$$

$$var(X) = x_1^2 p_1 + x_2^2 p_2 + \dots + x_n^2 p_n - \mu^2$$

$$var(X) = \Sigma x_i^2 p_i - \mu^2$$

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

Que 19 Let X denotes no. of time heads occurs when a ~~paired~~ fair coin is tossed 6 times. The distribution of X appears in question 16. where $\mu = 3$, find variance.

$$var(X) = 0^2 \times \frac{1}{2^6} + 1^2 \times \frac{6}{2^6} + 2^2 \times \frac{15}{2^6} + 3^2 \times \frac{20}{2^6} + 4^2 \times \frac{15}{2^6} + 5^2 \times \frac{6}{2^6} + 6^2 \times \frac{1}{2^6} - 9$$

$$= \frac{6 + 60 + 180 + 240 + 150 + 36}{2^6} - 9$$

$$= \frac{672 - 576}{64} = \frac{96}{64} = 1.5$$

$$SD = \sqrt{var(X)} = \sqrt{1.5} = 1.224$$

Date : / /
Page No.

Que 20 Consider random variable x in Que.17, where its mean $\mu = 0.75$ is computed. Compute $var(x)$.

$$var(x) = 1 \times \frac{108}{220} + \frac{4 \times 27}{220} + 9 \times \frac{1}{220} - (0.75)^2$$

$$= \frac{117 + 108}{220} - 0.5625$$

$$= \frac{225 - 123.75}{220}$$

$$= 0.460$$

$$\text{Standard dev} = \sqrt{var(x)}$$
$$= \sqrt{0.460}$$
$$= 0.678.$$

THEORY OF ATTRIBUTES

Let A and B represents the presence of attributes, and α and β represents absence of attributes.

eg: Let A represent males then α would represent females

α = not A

Similarly B represent literates then β would represent illiterate

(AB) : No. of literate males.
$(A\beta)$: —— illiterate ——
(αB) : No. of literate females
$(\alpha\beta)$: —— illiterate ——

1. min = 18
Q_1 = 25
median = 3
Q_3 = 45
max = 8

(b)

(a) SK

Assignment

Date: / /
Page No.

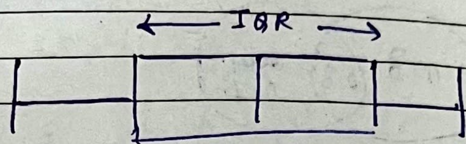
1. min = 18

Q_1 = 25

median = 31

Q_3 = 45

max = 80



(b) min = 18 $Q_1 = 25$ $Q_2 = 31$ $Q_3 = 45$ max = 80

$IQR = Q_3 - Q_1 = 45 - 25 = 20$

Range $= Q_1 - 1.5\,IQR$ to $Q_3 + 1.5\,IQR$

$= 25 - 1.5 \times 20$ to $45 + 1.5 \times 20$

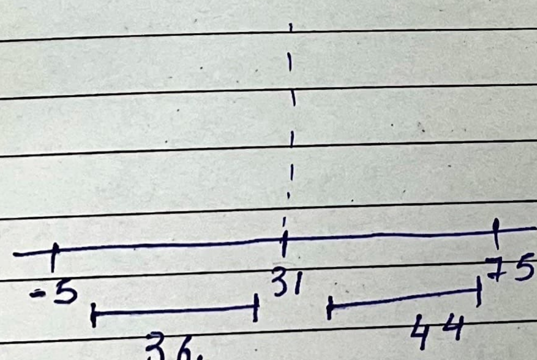
$= 25 - 30$ to $45 + 30$

$= -5$ to 75

Outlier = 80.

(a) Skewness

- vely skewed.



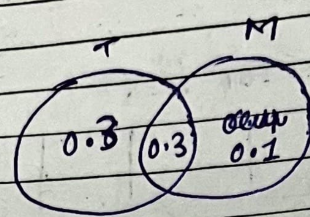
Date : / /
Page No.

Que 4. $T = 60\% = 0.6$

$M = 40\% = 0.4$

$T \cap M = 0.3$

(a)



(b)

(c) $P(T/M) = \dfrac{P(T \cap M)}{P(T)}$

$= \dfrac{\frac{3}{10}}{\frac{3}{5}} = \dfrac{1}{2}$

(d)

$(A) + (\alpha) = N$
$(B) + (\beta) = N$
$(AB) + (A\beta) = (A)$
$(\alpha B) + (\alpha\beta) = (\alpha)$
$(AB) + (\alpha\beta) = (B)$
$(A\beta) + (\alpha\beta) = (\beta)$

	B	β
A	(AB)	$A\beta$
α	αB	αB
Total		

- **Independent Attributes**: Two attributes are independent if proportion of male literate equal proportion of female literate. (ultimate frequency)

$$\frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)}$$

(i) $\because \frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}$

Therefore $\frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)} = \frac{(AB) + (\alpha B)}{(A) + (\alpha)} = \frac{(B)}{N}$

$\therefore \frac{(AB)}{(A)} = \frac{(B)}{N}$ or $(AB) = \frac{(A) \times (B)}{N}.$

Date : / /
Page No.

(2) $\frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)}$ (from table)

$$\frac{(AB)}{(AB)+(A\beta)} = \frac{(\alpha B)}{(\alpha B)+(\alpha\beta)}$$

$$(AB)(\alpha B) + (AB)(\alpha\beta) = (AB)(\alpha B) + (A\beta)(\alpha B)$$

$$\boxed{(AB)(\alpha\beta) = (A\beta)(\alpha B)}$$

→ Positive and negative association b/w attributes

- Two attributes are associated to ~~two~~ each other if they are not independent.
- If the actual frequency of class AB is more than computed frequency then the two attributes are said to be positive associated.

$$(AB) > \frac{(A)\times(B)}{N}$$

Then there is positive association b/w A & B

- If actual frequency of class AB is less than the computed frequency then the two attributes are said to be negatively associated.

$$(AB) < \frac{(A)\times(B)}{N}.$$

$$\left.\begin{array}{l}(AB) = (A) \text{ and } (A\beta) = 0 \\ (AB) = (B) \text{ and } (\alpha B) = 0\end{array}\right\}$$ Then A and B are completely associated

$\frac{100 \times 140}{200}$

Date: / /
Page No.

- $(AB) = 0$ or $(\alpha\beta) = 0$] A and B are completely disassociated

→ Yule's coefficient of Association (Q)

Q quantifies the association between the two attributes. It gives amount of association.

$$(AB)(\alpha\beta) \neq (A\beta)(\alpha B) \quad [\text{If not independent}]$$

Hence

"$(AB)(\alpha\beta) - (A\beta)(\alpha B)$" measures the amt of association

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

Que Given A & B are independent attributes, $N = 200$ $(A) = 100$, $(B) = 140$, find the ultimate class of frequencies

Sol:

$$(AB) = \frac{(A) \times (B)}{N} = \frac{100 \times 140}{200} = 70.$$

	B	β	Total
A	$(AB) = 70$	$(A\beta) = 30$	$(A) = 100$
α	$(\alpha B) = 70$	$(\alpha\beta) = 30$	$(\alpha) = 100$
Total	$(B) = 140$	$(\beta) = 60$	$N = 200$

Que Given $N = 100$, $(A) = 60$, $(B) = 50$, $(AB) = 30$. Examine for attributes A and B indep / +vely associated / -vely associated.

$$(AB) = \frac{(A) \times (B)}{N} = \frac{60 \times 50}{100}$$

$$= 30$$

Independent

Date : / /
Page No.

Que 252 candidates, 140 were boys, 72 candidates were ~~given~~ successful, among them 40 were boys. Obtain coefficient of association.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) - (A\beta)(\alpha B)}$$

$$(AB) = \frac{(A) \times (B)}{N}$$

A → candidate is boy

B → candidate is successful

	B	β	Total
Boy A	$(AB) = 40$	$(A\beta) = 100$	$(A) = 140$
Girl α	$(\alpha B) = 32$	$(\alpha\beta) = 80$	$(\alpha) = 112$
	$(B) = 72$	$(\beta) = 180$	Total = 252

$$Q = \frac{40 \times 80 - 100 \times 32}{40 \times 80 + 100 \times 32}$$

$$Q = \frac{3200 - 3200}{3200 + 3200} = 0$$

i.e Independent.

- <u>Coefficient of Colligation</u>