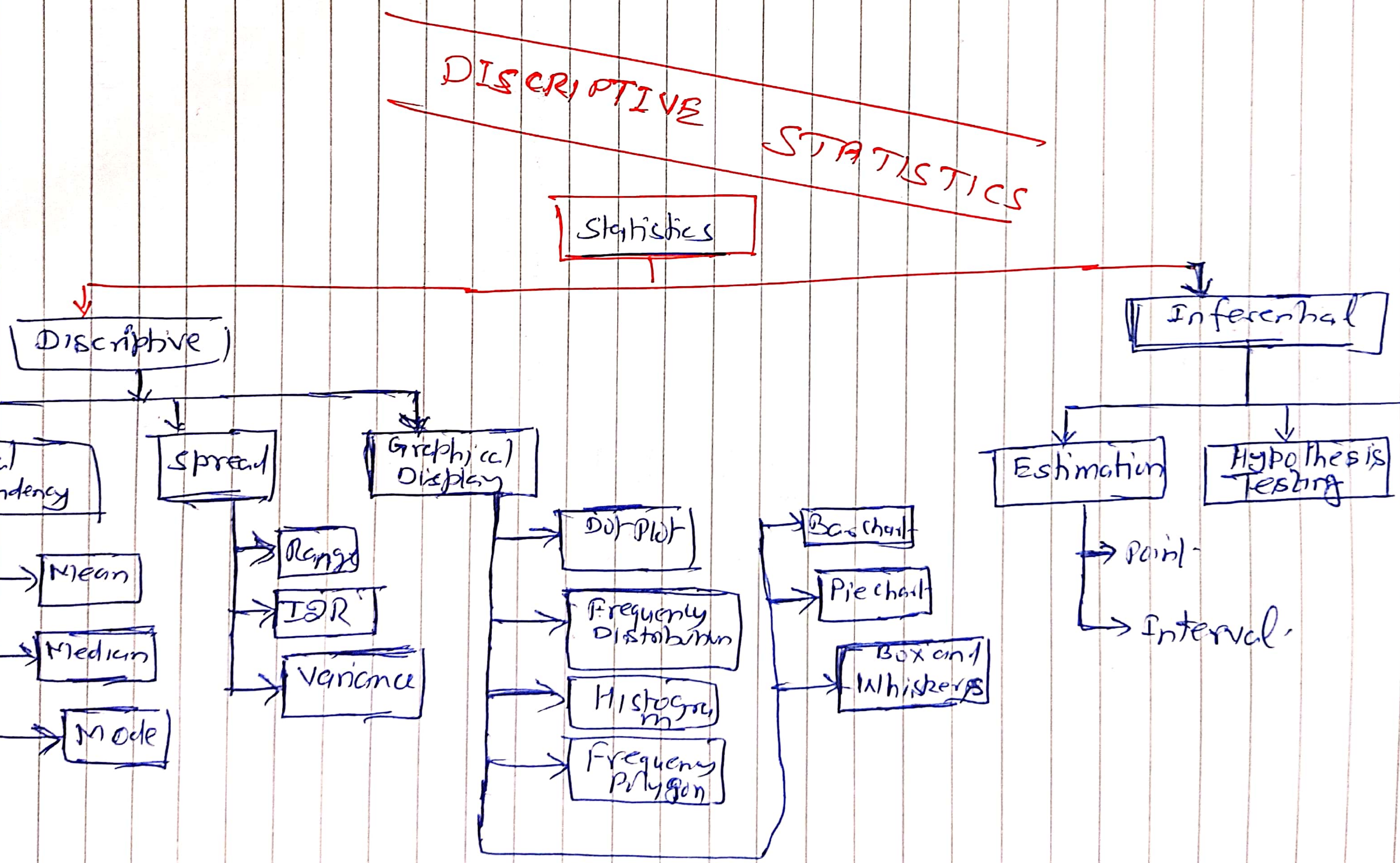
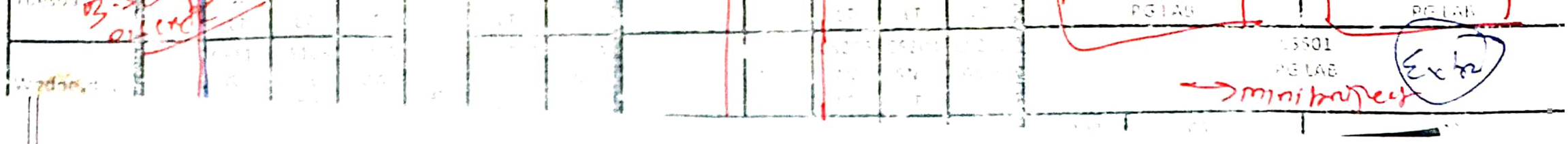
# DISCRIPTIVE STATISTICS






# Discriptive statistics

Enroll
Page No: 2
Date: / /

Statistics ⟹ used as singular or plural sense

Singular :— In on singular case statistics refers to the procedures used to organize and interpret observed data.

↳ In this statistics is defined as branch of mathematics that deals with collecting, organizing and summarizing data and drawing conclusions about the environment from which the data was collected.

Plural Sense :—

Statistics are quantitative values that are used to describe a set of observed data.

Thus, some time "Statistics is" and "statistics are" are used depending on context.

Enroll
Page No: 22
Date: / /

## Statistics Vs. Probability

**Probability :-**

— Fully defined probability problems have unique and precise solutions.

— Probability laws apply across an entire population of interest

**Statistics :-**

Statistics is concerned with the ~~relationship~~ relationship between an observed segment of a population and the entire population.

↳ In statistics we are interested in understanding an ~~observa~~ observation that is based on a segment of a population of interest and how the observation can apply to the entire population

Enroll
Page No: 23
Date: / /

Statician works by formatting the data in way that make sense ↓↓

~~Stats~~

Postulate the Probabilistic model.

for system under investigation based on the physical mechanisms involved in the system and on personal experience

↓↓

& Statistician expects that the model to exhibit a probabilistic behavior that is similar to that of the physical system.

## Branches of statistics

- General branches
  - Discriptive statistics
  - Inferential (inductive) statistics.

**Discriptive statistics :—** ~~Discrip.~~

↳ concerned with collecting, organizing, and summarizing the raw scores in ~~the~~ more meaningful ways.

Data ≡ raw score are measurement or Observed ~~Values~~ Values

Enroll
Page No: 24
Date: / /

## Inferential Statistics :-

Inferential statistics deals with procedures or techniques that can be used to study a segment of the population called a sample, and make generalizations about the population from which the sample was obtained with the help of probability.

Enroll
Page No: 25
Date: / /

## Discreptive statistics:

↳ This deals with collecting, grouping, and presenting data in a way that can be easily understood.

⇓

make sense of observed data.

⇓

present data the data in a more meaningful way, which allows simpler interpretation of ~~data~~ the data.

Eg Grade of two students → i) distribution of grade
ii) spread of grade.

organizing the data by a set of graphs, bar charts, tables, or frequency distributions.

⇓

Discriptive statistics does not, however, allow us to make conclusions beyond the data we have analyzed; it is simply a way to describe a set of observed data by providing simple summaries about the sample and measures.

Methods that are used to describe the set of observed data.

a) Measures of central tendency :—
=> Ways of describing the central position of a frequency distribution for a group of data.

b) Measures of Spread :—
Ways to summarize a group of data by describing how spread out the data values

c) Graphical Displays :— Ways to visualize the data to see how it is distributed and if any patterns emerge from the data.

Enroll
Page No: 27
Date: / /

## Measures of Central Tendency:

describe the "center" of the data set.

↳ central tendency describes the tendency of the observations to bunch around a particular value.

↳ Measure of central tendency are numerical summaries that are used to summarize a data set with a single "typical" number.

Three main measure of central tendency:

(i) Mean
(ii) Median
(iii) Mode.

} measures of the "average" of the distribution of the data set.

**Mean:-** Mean ⟺ Average

Let $x_1, x_2, x_3, \ldots, x_N$ be data set, the mean is given by:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Enroll
Page No: 27
Date: / /

eg. 66, 54, 08, 56, 34, 12, 40, 50, 00, 50, 90, 65

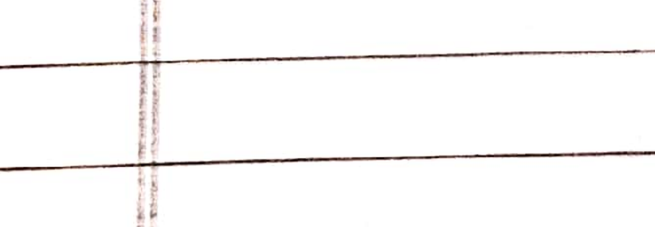
D1

Then, the mean is given by:

$$\bar{x} = \frac{66+54+08+56+34+12+40+50+00+50+90+65}{12}$$

$$\boxed{\bar{x} = 57.75}$$

Note:- Mean may not be a member of the data set.



Page No: 29
Date: / /

## Median

Divide the data into two equal halves where each half contains 50% of the data.

- The numerical value where the data set is divided is called the median.

For median computation:

(i) ordered rank by arranging them in increasing order of magnitude.

(ii) Median data position is calculated

- Even no of data values
- odd number of data values

$R_M$ : rank-position of the median in the rank-ordered data set.

$N$ : Number of data values.

(iii) If $N$ is odd:

$$R_M = \frac{N+1}{2}$$

And median is data value at the position $R_M$.

Enroll

Page No: 30

Date: / /

If N is even then the value of the median is given by

~~$M = \frac{d_{\frac{N}{2}} + d_{\frac{N}{2}+1}}{2}$~~

$$M = \frac{d_{\frac{N}{2}} + d_{\frac{N}{2}+1}}{2}$$

$d_k$ : data value at position k.

Ex:-

Arrange the data (D) in average section in increasing order

1 2 3 4 5 6 7 8 9 10 11 12

12, 34, 40, 50, 50, 54, 56, 65, 66, 80, 88, 90

(R D)

N = 12 data values. ⇒ Even ⇒ median is nd member of data set

⇓

It is average of data values at $\frac{12}{2} = 6^{th}$ and $\frac{12}{2}+1 = 7^{th}$ position of sorted data.

Therefore, median:

$$M = \frac{54 + 56}{2} = 55$$

Note:- Append 95 to the rank data set (RD).

N = 13 = odd.

$\therefore R_M = \frac{13+1}{2} = 7$

⇒ m = 56 is the member of the ~~data~~ data set.

Enroll

Page No: 31

Date: / /

## "Mode"

→ Mode is data value that occurs the most frequently in the data set.

12, 34, 40, 50, 50, 54, 56, 65, 66, 80, 88, 90, 95 — RD2

In RD1 and RD2 50 occurs twice while every other value occurs once.

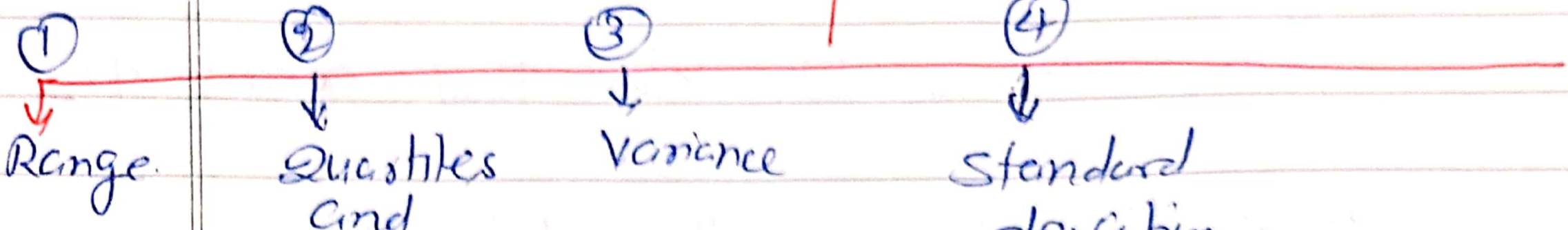
Thus mode of RD1 and RD2 is 50.

and the data sets are said to be unimodal.

→ Some times data set can have more than one mode, in this case it is said to be a multi-modal set.

Enroll
Page No: 32
Date: / /

## → Measure of Dispersion




### — Range: —

↳ measure of statistical dispersion or spread.

Range = Maximum data value — Minimum Data Value

Range = Largest data value — Smallest data value.

### → Quartiles And Percentiles:—

↳ Divide the ordered ~~dat~~ data sets into quarters.

↳ $100p^{th}$ percentile of an ordered data set is value such that at least $100p\%$ of the observations are at or below this value and atleast $100(1-p)\%$ are at or above this value, where $0<p<1$.

Enroll
Page No: 33
Date: / /

(i) First Quartile $Q_1$ = 25th percentile and $p = 0.25$

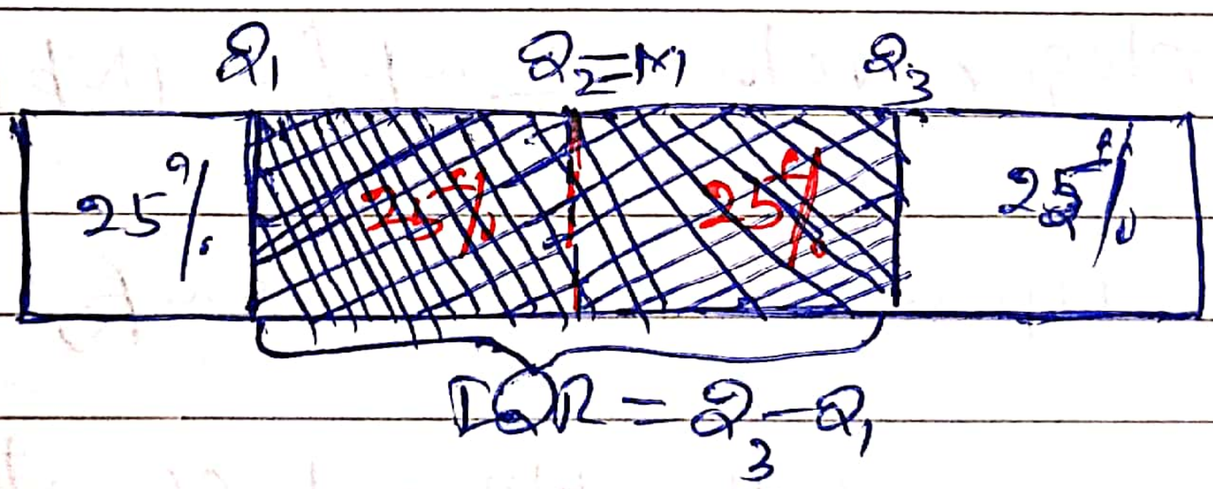
(ii) Second Quartile $Q_2$ = 50th percentile and $p = 0.50$

(iii) Third Quartile $Q_3$ = 75th percentile and $p = 0.75$

Second quartile is the median

Interquartile range (IQR) : This is difference between the third quartile and the first quartile.

$$IQR = Q_3 - Q_1$$




IQR = middle one-half (or 50%) of an ordered data set.

Quartile computation depends on the ordered data set i.e. data position in a rank-ordered data set and not on the data value itself.

Procedure for computing the kth percentile is as follows:

(i) order the N observations from the smallest to the largest.

(ii) Compute the product Np. If Np is not integer, round it to next integer and find the corresponding ordered value i.e. ~~[e [Np]~~ i.e. compute $\lfloor Np \rfloor$ integer value $\lfloor Np \rfloor$ or $\lfloor Np \rfloor + 1$

(iii) If Np is an integer, say k, then calculate the mean of the $k^{th}$ and $(k+1)^{th}$ ordered value.

e.g.:- 1 2 3 4 5 6 7 8 9 10 11 12
66, 54, 80, 56, 34, 12, 40, 50, 80, 55, 90, 65

↳ ordered

1 2 3 4 5 6 7 8 9 10 11 12
12, 34, 40, 50, 50, 54, 56, 65, 66, 80, 80, 90

N=12, For first quartile p=0.25.

Therefor first quartile Np = 12×0.25 = 3

∴ First quartile $Q_1$ = mean of 3rd and 4th element of the ordered data.

First quartile $Q_1 = \frac{40+50}{2} = 49$

For third quartile $p = 0.75$

$Np = 12 \times 0.75 = 9$

∴ Third quartile $Q_3$ = mean of the 9th and 10th element of the ordered data

Third quartile $Q_3 = \frac{66+80}{2} = \frac{146}{2}$

Third quartile $Q_3 = 73$.

Inter-quartile range (IQR) $= Q_3 - Q_1$

$= 73 - 49$

$\boxed{IQR = 24}$

For second quartile $p = 0.50$

$Np = 12 \times 0.50 = 6$

Median = mean of 6th and 7th data value in ordered data set

Median $= \frac{54+56}{2} = 55$

$\boxed{\text{Median} = 55}$

Enroll

Page No: 36

Date: / /

## Alternative Method For computing $Q_2$:-

(i) Find the median: It devides the data ie. $Q_2$ into two halves.

lower half < Median

Upper half > Median

(ii) - Find the median of lower half, which is $Q_1$

= Find the median of upper half, which is $Q_3$

## In previous example

Median $Q_2 = \frac{54+56}{2} = 55$

lower half of ordered data:

12, 34, 40, 50, 50, 54.

In this total number of data values = 6

Which even therefore median of lower half = average of 3rd and 4th element

$= \frac{40+50}{2}$

$= \frac{90}{2}$

$= 49$

Thus first quartile $Q_1 = 49$.

Upper half of the data value 56, 65, 66, 90, 98, 90

Here $N = 6$, ~~Ther~~ which is even. ~~Therefore~~ Therefore third quartiles $Q_3$ is the average of 3rd and 4th

$$Q_3 = \frac{66 + 80}{2} = 73$$

$$\therefore \quad IQR = 73 - 49$$

$$\boxed{IQR = 24}$$

Enroll
Page No: 30
Date: / /

## : Variance :

**Deviation Score :-** Subtract each data value from the mean, the obtained value called a deviation score.

↳ deviation score gives the numerical distance between data value and data sets "typical" value.

↳ The sum of all the deviation scores equals zero. as shown in table below.

| Data value $x$ | Deviation $x-\bar{x}$ ($\bar{x}=59.75$) | Squared deviation $(x-\bar{x})^2$ |
|---|---|---|
| 66 | 8.25 | 68.0625 |
| 54 | -3.75 | 14.0625 |
| 88 | 30.25 | 915.0625 |
| 56 | -1.75 | 3.0625 |
| 34 | -23.75 | 564.0625 |
| 12 | -45.75 | 2093.0625 |
| 48 | -9.75 | 95.0625 |
| 50 | -7.75 | 60.0625 |
| 80 | 22.25 | 495.0625 |
| 50 | -7.75 | 60.0625 |
| 90 | 32.25 | 1040.0625 |
| 65 | 7.25 | 52.5625 |
| Sum | $\Sigma(x-\bar{x})=0$ | $\frac{1}{12}\Sigma(x-\bar{x})^2=455.24$ |

Enroll
Page No: 39
Date: / /

→ The data values above and below the mean have positive and negative ~~devi~~ deviation scores, respectively, that cancel each other out.



↳ Remove the negative score ⟹ square the deviation scores ⟹ Sum of squared deviation scores

⇒ The average of square deviation ~~score~~ score gives the value of variance.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

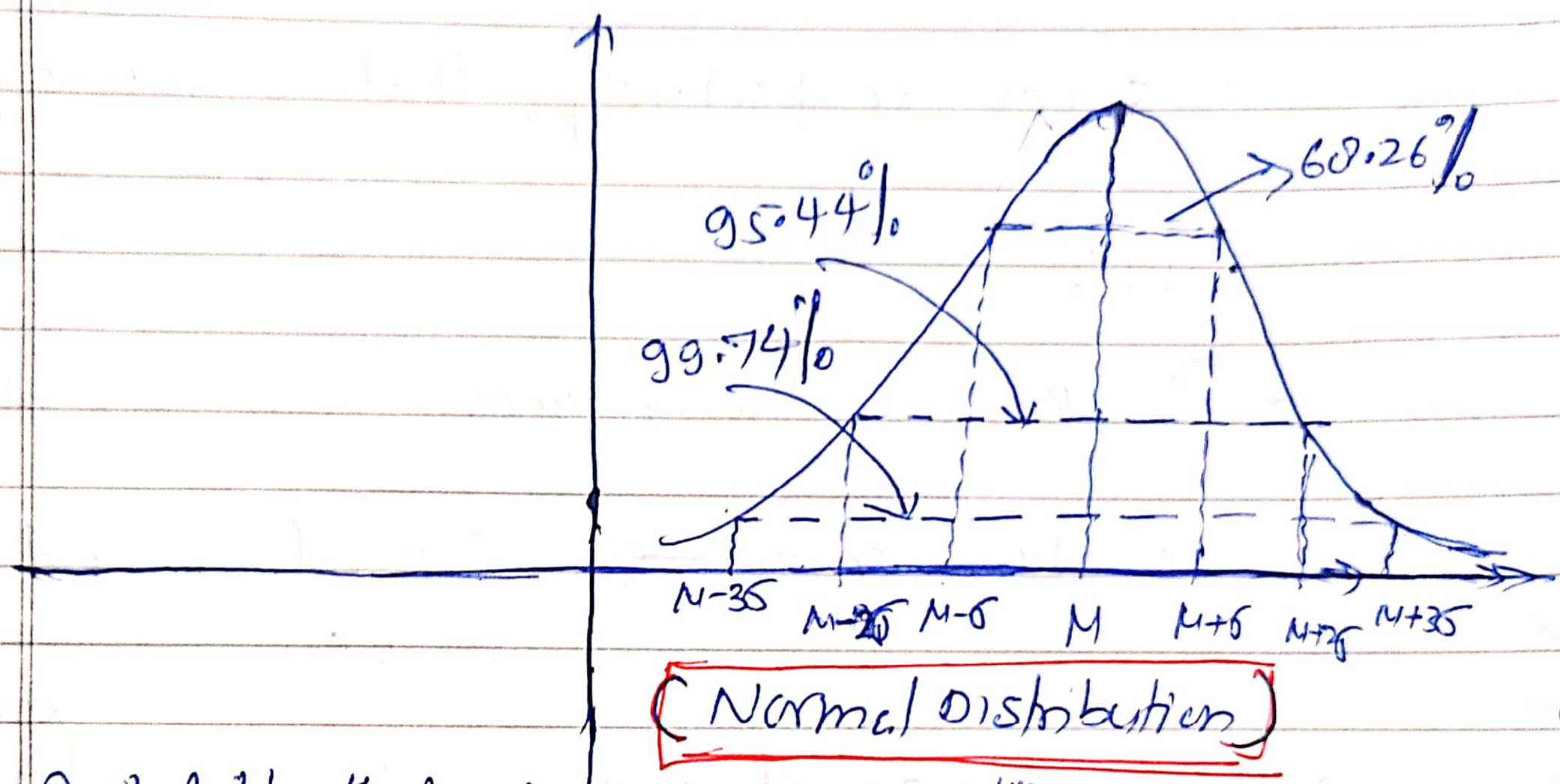
For example ~~given~~ discussed above.

$$\sigma^2 = \frac{1}{12} \sum_{i=1}^{12} (x_i - \bar{x})^2 = 455.021$$

Enroll
Page No: 40
Date: / /

## :- Standard deviation :-

Standard deviation ($\sigma$) = square root of variance.




(Normal Distribution)

Probability that observation lies within :-

$P[M_x - \sigma_x \leq X \leq M_x + \sigma_x] = 0.6826$

$P[M_x - 2\sigma_x \leq X \leq M + 2\sigma_x] = 0.9544$

$P[M_x - 3\sigma_x \leq X \leq M + 3\sigma_x] = 0.9974$

Enroll
Page No: 41
Date: / /

## Graphical And Tabular ~~Displays~~?

organize the data into a graphical or tabular form so that ~~a trend~~ a trend, if any, emerging out of the data can seen easily

(i) Dot plots (ii) frequency distn'bution (iii) Bar charts (iv) Histograms, (v) frequency Polygon.

(v) Pie charts (vi) Box plots and ~~(viii) whisker plot~~

(vii) Whiskers plot

## Dot plots:

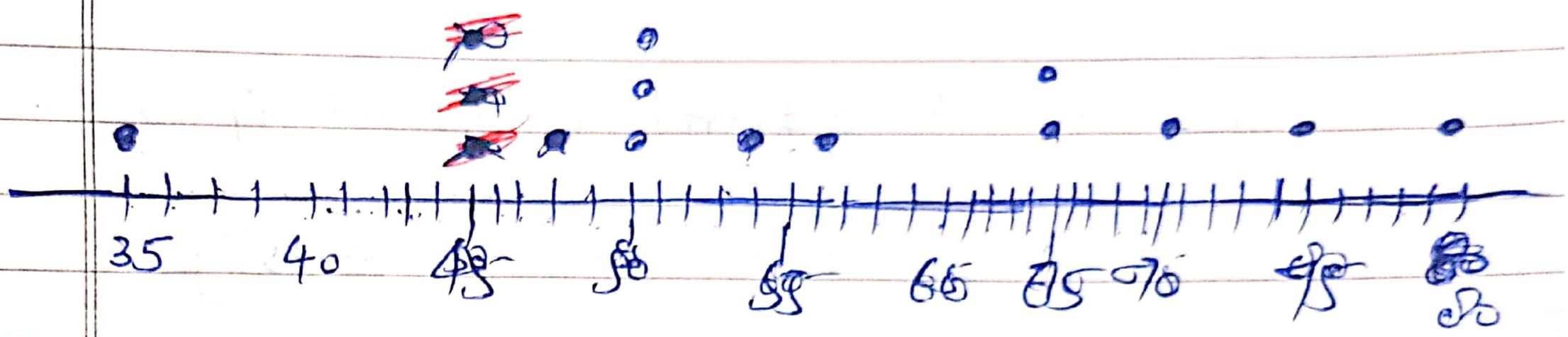
Dot plots ≡ dot chart

— A dot plot uses dots to show where the data values (or scores) in a distribution.

— ~~uses dot~~ Dots are plotted against their actual data values that are on the horizontal scale.

↳ If there are identical data values, the dots are "piled" on top of each other.

35, 40, 50, 50, 50, 54, 56, 65, 65, 70, 75, 80




Dot plot is useful when number of data is small in size.

## Frequency distribution

→ Table that lists the set of values in a dataset and their ~~frequency~~ frequencies

— frequencies ≡ number of times each occurs in the data sets

eg.

35, 40, 50, 50, 50, 54, 56, 65, 65, 70, 75, 80.

Enroll
Page No: 43
Date: / /

Table ~~Fit~~.

| X | Frequency |
|---|---|
| 36 | 1 |
| 48 | 1 |
| 50 | 3 |
| 54 | 1 |
| 56 | 1 |
| 65 | 2 |
| 70 | 1 |
| 75 | 1 |
| 80 | 1 |

→ When data covers ~~large~~ wide range of values that a list of all the X values would be too long to be a "simple" presentation of ~~data~~ data.

↳ In this case grouped frequency distribution table in which the X column lists groups of data values, called class intervals, rather than individual data values.

↳ The width of each class can be determined by dividing the range of observations by the number of classes.

↳ Have equal class widths, and the class interval should be mutually exclusive and non-overlapping

Enroll
Page No: 44
Date: / /

Class width = difference between the lower limits of two consecutive classes
= difference between the upper ~~limits~~ limits of two consecutive classes

Class mark = is the ~~middle~~ number in the middle of the class
= Average of lower and upper limit

Group Frequency Distribution

| class | Frequency | class mark |
|---|---|---|
| 30 – 40 | 1 | 37 |
| 41 – 47 | 0 | 44 |
| 48 – 54 | 5 | 51 |
| 55 – 61 | 1 | 58 |
| 62 – 68 | 2 | 65 |
| 69 – 75 | 2 | 72 |
| 76 – 82 | 1 | 79 |

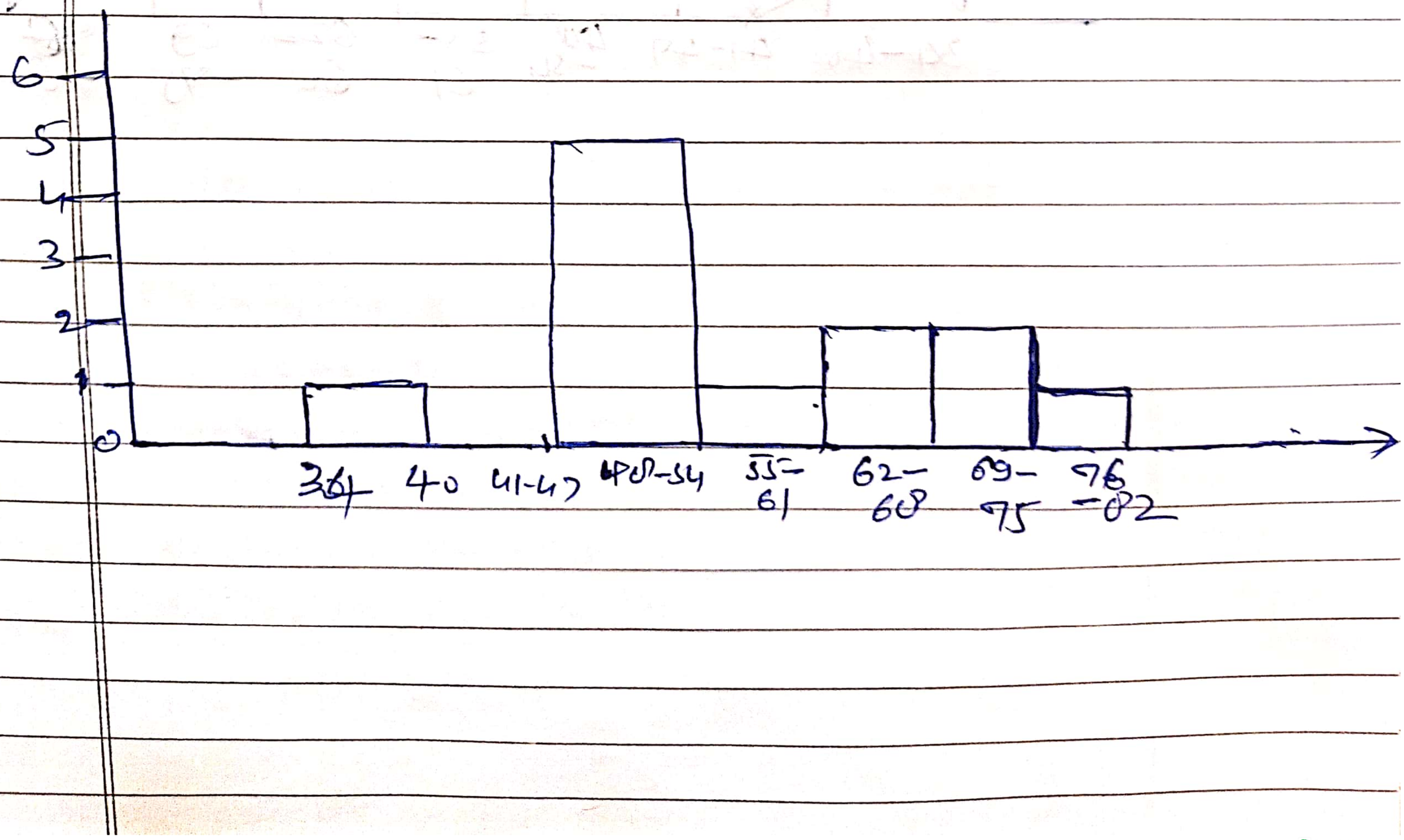
Enroll
Page No: 45
Date: / /

## → Histogram :—

Frequency histogram (or simply histogram) is used to graphically display the grouped frequency ~~distri~~ distribution.

→ Draw vertical bars above the classes so ~~that~~ that the height of a bar corresponds to the frequency of the class that it represents

— Width of the bar ~~exte~~ extends to the real limits of the score class. Thus, ~~the~~ colums are of equal widths and there are no spaces between column.



Enroll

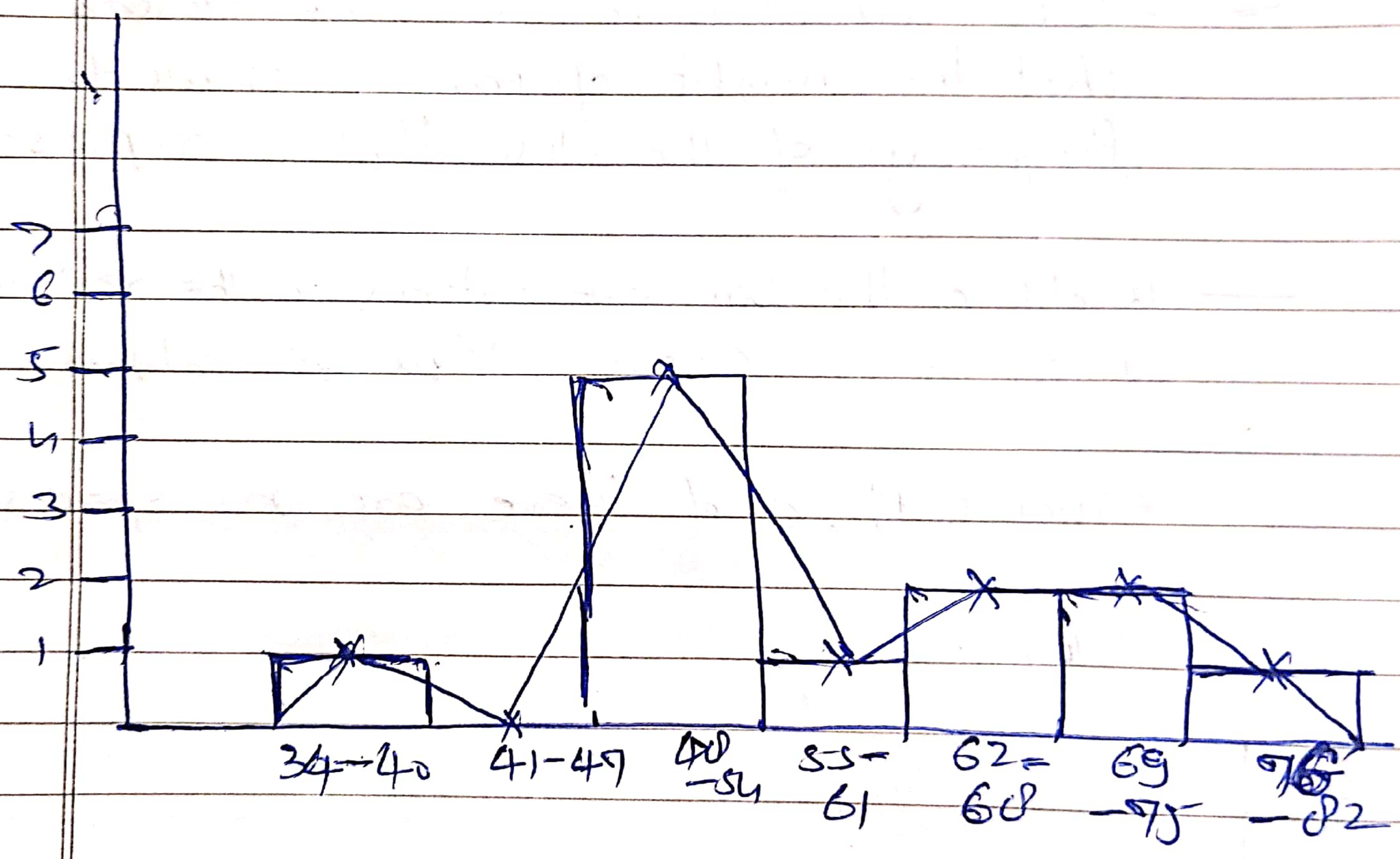
Page No: 46

Date: / /

## Frequency Polygons :

This gives idea about shape of the ~~distribution~~ distribution.

Obtained by joining the class marks of a histogram with the two end points lying on the horizontal axis.



Enroll
Page No: 47
Date: / /

## Bar Graphs

Used for ~~pictorial~~ graphical representation of the categorical variable

↳ the variable that has two or more categories with no intrinsic ordering to the categories.

↳ e.g. geneler is categorical variable with two categories : Male - female.

⇒ length (horizontal), height (vertical) shows the frequency for each category or characteristics.

Eg. survey of 100 computer scienc students to determine that their interest in subjects

(Prob) Probability : 30

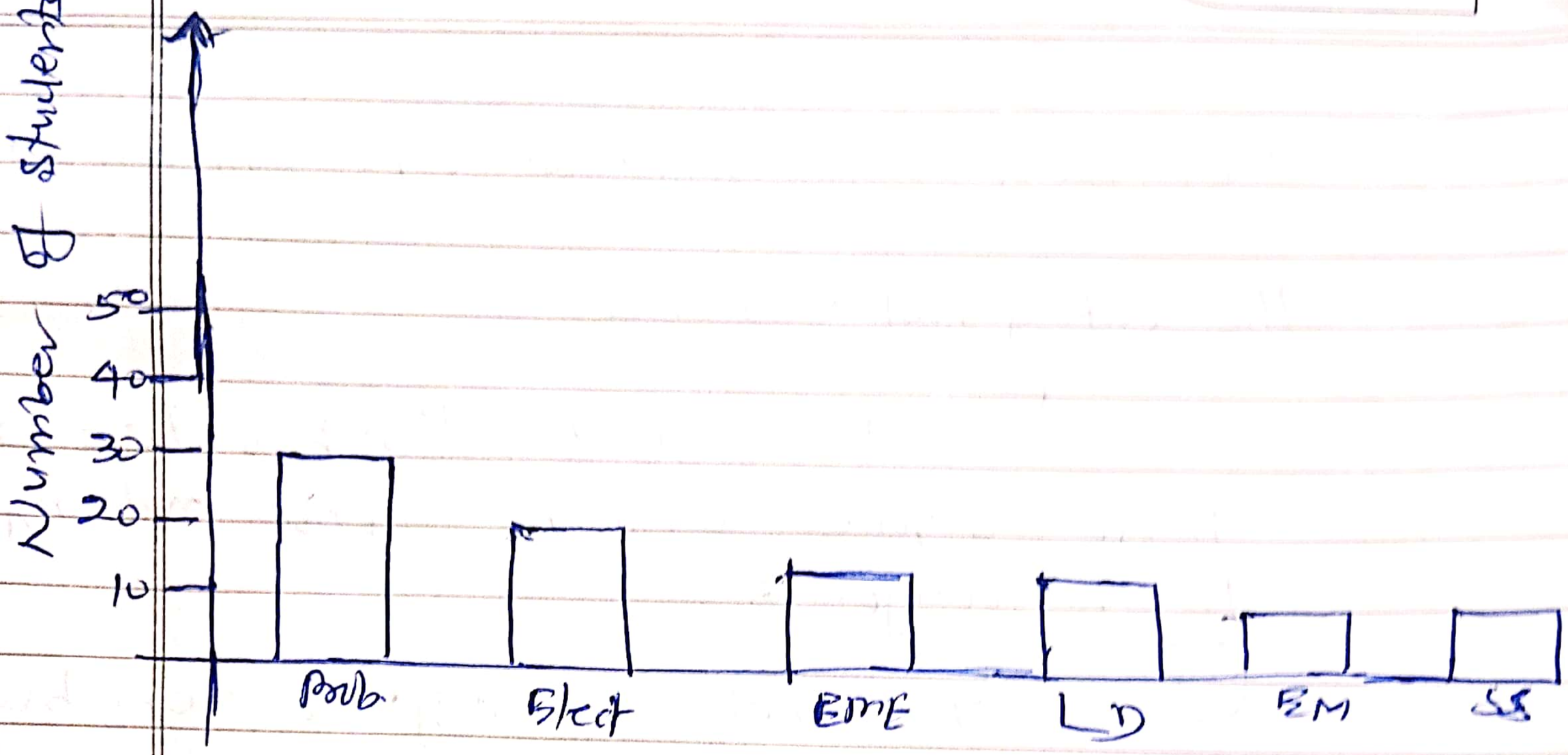
(Elect) Electronics : 20

(EM) Electromechanics : 15

(LD) Logic Design : 15

(Em) Electromagnetics : 10

(SS) signal and system : 10




Note:-

(i) Gaps are included between the bars of each category.

(ii) Bars can be arranged in any order without affecting the ~~data~~ ~~data~~ data.

Enroll
Page No: 49
Date: / /

## Pie chart

— "Pie slices" are used to show relative sizes of data.

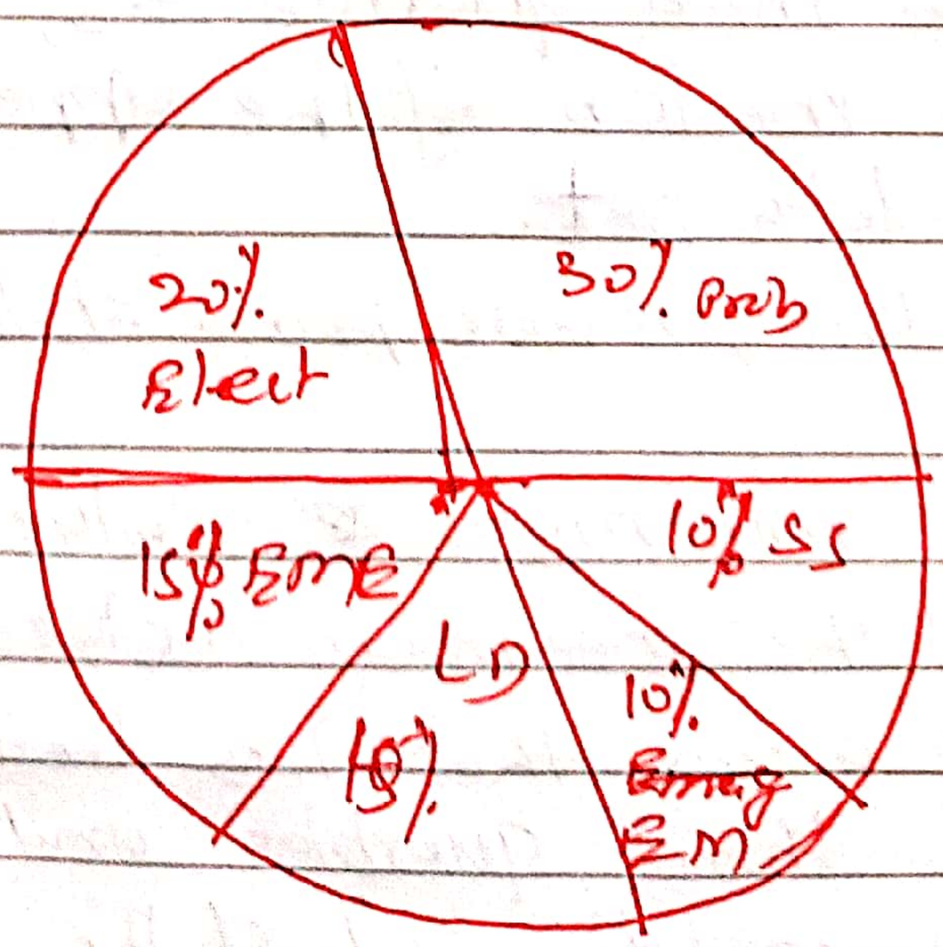
— The size of each slice is proportional to the probability of the event that the slice represents.

Example the survey of ECE student to find out their favorite subjects.

(+) i. Probability (Prob) : 30% ii. Electronics : 20% (Elect)

iii. Electromechanics (EMB) : 15% ; iv. Logic design 15% v. Electromagnetics (EM) : 10%

vi: Signal and systems : 10%



Enroll

Page No: 50

Date: / /

## :- Box and Whiskers Plot :-

↳ Way to ~~visualize~~ visually organize data into fourths or ~~quartiles~~ quartiles

↳ ~~Made~~ Draw box including first and third quartiles. and straight lines( Whiskers) extending from the ends of the box to the minimum and maximum data values.

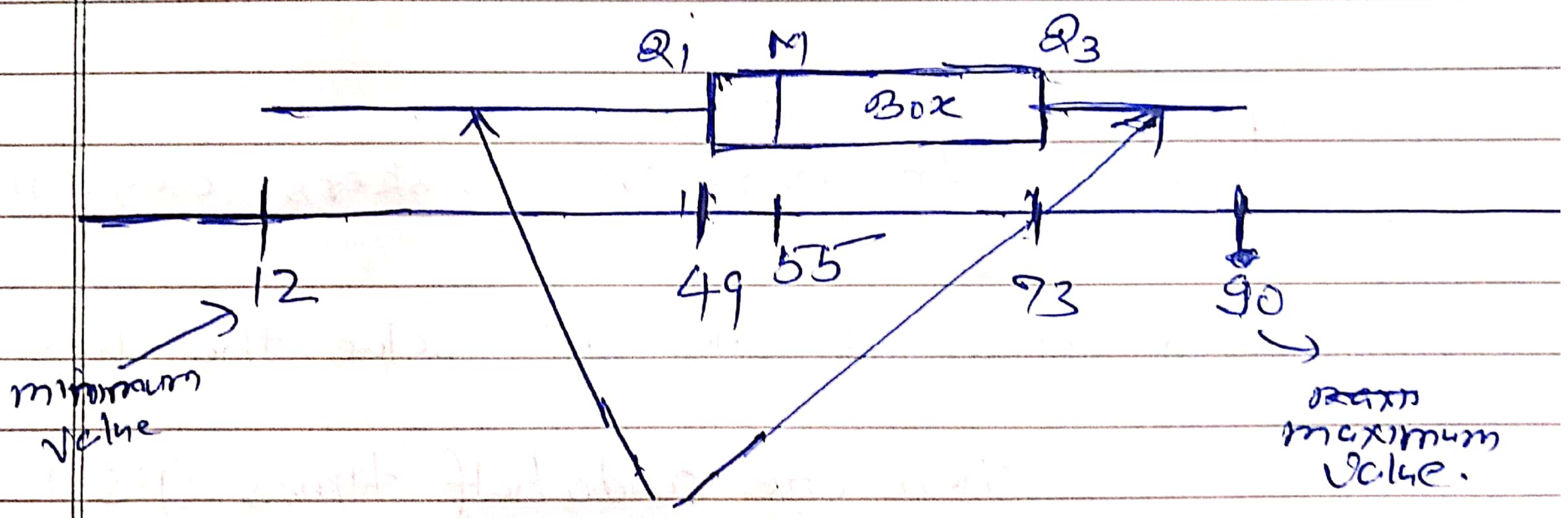
### Procedure :-

1. Arrange the data in increasing order
2. Find the median
3. Find the first quartile $Q_1$, which is the median of the lower half of the data set, and third quartile, $Q_3$, which is the median of the upper half of the data set.
4. on a line, mark points at the median, $Q_1$, $Q_3$ the minimum & value of the data set and the maximum value of the data set
5. Draw a box that lies between the first and third quartiles and thus represents the middle 50% of the data.

Enroll
Page No: 57
Date: / /

6. Draw a line from the first quartile to the minimum data value, and another line from the third quartile to the maximum data value. These lines are whisbers of the plot.

e.g :- 12, 34, 40, 50, 50, 54, 56, 65, 66, 80, 88, 90




~~outl~~ outliers :- Values much higher or much lower than all of the other values.

↳ such values are known as outlier

↳ These outliers are usually excluded from the whisher portion of the ~~the~~ box and whiskers diagram. They are plotted individually and labeled as outliers.

Enroll
Page No: 52
Date: / /

## Interquartile range (IQR) and outlier:-

IQR = $Q_3 - Q_1$ = width of box in whiskers ~~plot~~ diagram

IQR ⇒ is measure of dispersion

⇓

Statistics assumes that data values are clustered around some central value.

Outlier :- In box-and whiskers diagram, an outlier is any data value that lies more than one and a half times (1.5) the length of box from either end of box.

i.e. data point below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$ is viewed as being too far from the central values to be ~~re~~ reasonable.

↳ Thus $Q_1 - 1.5 \times IQR$ and $Q_1 + 1.5 \times IQR$ are the "fences" that merks of the

" reasonable" values from the outlier values.

Lower fence : $Q_1 - 1.5 \times IQR$

Upper fence : $Q_3 + 1.5 \times IQR$

In previous example :-

$IQR = Q_3 - Q_1 = 73 - 49$

$IQR = 24$ $\quad IQR \times 1.5 = 36$

Lower fence $= Q_1 - IQR \times 1.5 = 49 - 36 = 13$

Lower fence $= 13$

Upper fence $= Q_3 + IQR \times 1.5$

$= 73 + 36 = 109$

Upper fence $= 109$.

The only data value outside the fences is 12, and all other values are within two fences

$\Rightarrow$ Thus 12 is only outlier in the data set.

Enroll

Page No: 54

Date: / /

Example :- Find outlier with the help of

## Box plot — Whiskers plot

10, 12, 8, 1, 10, 13, 24, 15, 15, 24

Rank this in increasing order :

Position → 1 2 3 4 5 6 7 8 9 10

1, 8, 10, 10, 12, 13, 15, 15, 24, 24

Total No. of data (N) = 10 = even

→ Median is average of 5th and 6th position element in sorted sequence

$$M = \frac{12+13}{2} =$$

$$\boxed{M = 12.5}$$

Lower half data set : 1, 8, 10, 10, 12

Whose median $Q_1 = 10$

Upper half data set : 13, 15, 15, 24, 24

Whose median $Q_3 = 15$

Thus $IQR = Q_3 - Q_1 = 15 - 10 = 5$

Lower fence $= Q_1 - 1.5 \times IQR$

$= 10 - 1.5 \times 5$

$= 10 - 7.5$

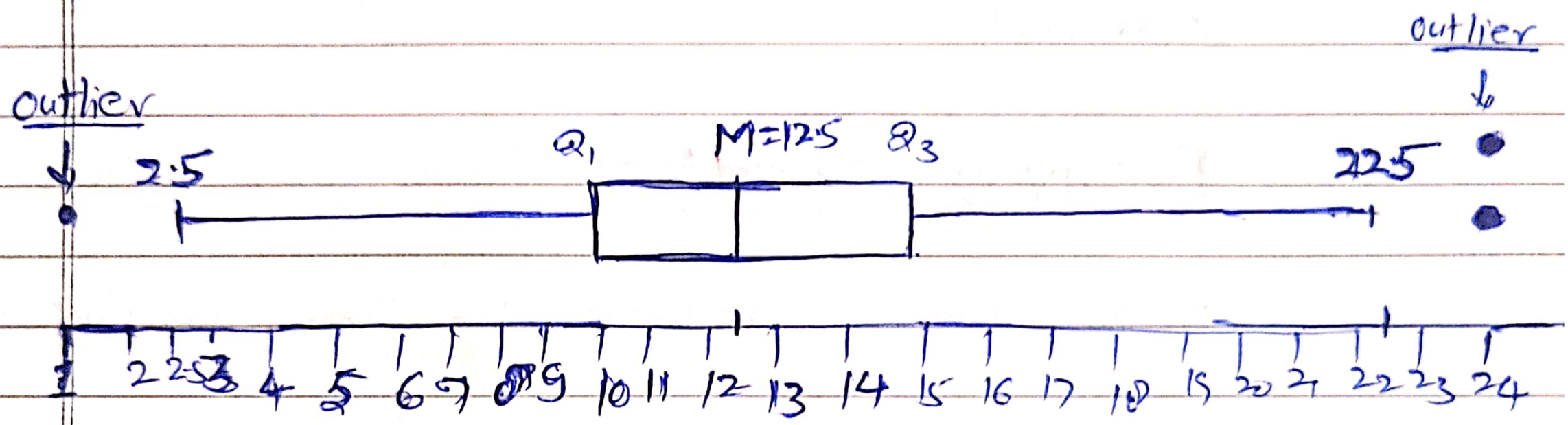
$\boxed{\text{Lower fence} = 2.5}$

Enroll
Page No: 55
Date: / /

Upper fence $= Q_3 + IQR \times 1.5$

Upper fence $= 15 + 7.5$

Upper fence $= 22.5$



Enroll
Page No: 56
Date: 2 + 1

## → : Shape of Frequency Distribution

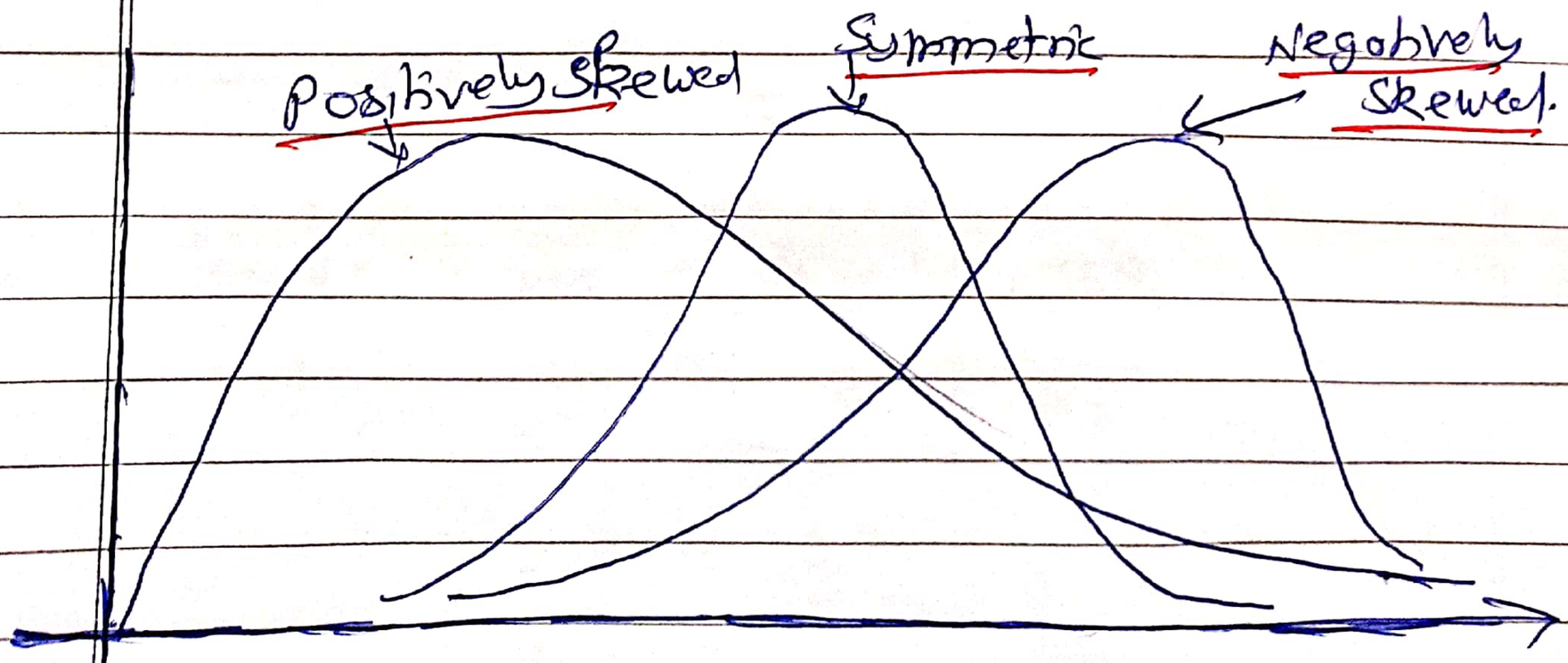
Skew ≡ तिरछा / तिरछा

## ( SKEWNESS )

Any distribution/curve which is not symmetrical about longitude is said to be skewed.

**Positively skewed :—** lower data values have higher frequencies. The positively skewed distribution is right-tailed distribution

**Negatively skewed :—** Higher data values have higher frequencies, the distribution is negatively skewed. The negatively skewed distribution is a left-tailed distribution.



Enroll
Page No: 57
Date: / /

## Quantitative measure of skewness

Quantitatively, the skewness of a random variable X is denoted by $\gamma_1$ and defined as

$$\gamma_1 = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right]$$

$$\gamma_1 = \frac{1}{\sigma_X^3} E[(X-\mu_X)^3] \quad \text{---(1)}$$

where $\mu_X$ : mean

$\sigma_X$ : standard deviation.

### (1) Skewness computation based on population:-

Let data set size (N) (size of population)

Then skewness is calculated as follows

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}} \quad \text{---(2)}$$

### (2) Skewness Computation Based on sample size (n):

The skewness is given by

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 \quad \text{---(3)}$$

Enroll

Page No: 58

Date: / /

## Skewness Measure Based on Pearson's Coefficient of Skewness

### I^st Version:

$x_{med}$ : median of a data set

$x_{mode}$ : mode

The Pearson's mode / first skewness coefficient is defined by:

$$\gamma_1 = \frac{\mu_x - x_{mode}}{\sigma_x} \quad \text{---(4)}$$

### II^nd version

The Pearson's median / second skewness coefficient is used when the mode of a distribution is not known and is given by:

$$\gamma_1 = \frac{3(\mu_x - x_{med})}{\sigma_x} \quad \text{---(5)}$$

Note:- The factor 3 in the equation (5) is due to fact that empirical results indicate that

Mean − Mode ≅ 3 (Mean − Median) ⇒

$$\mu_x - x_{mode} \cong 3(\mu_x - x_{med})$$

Enroll
Page No: 59
Date: / /

~~Note:~~ Note :- ① Pearson Median Coefficient is more popularly used.

② ②

(a) Positively skewed distribution ($\gamma_1 > 0$) &

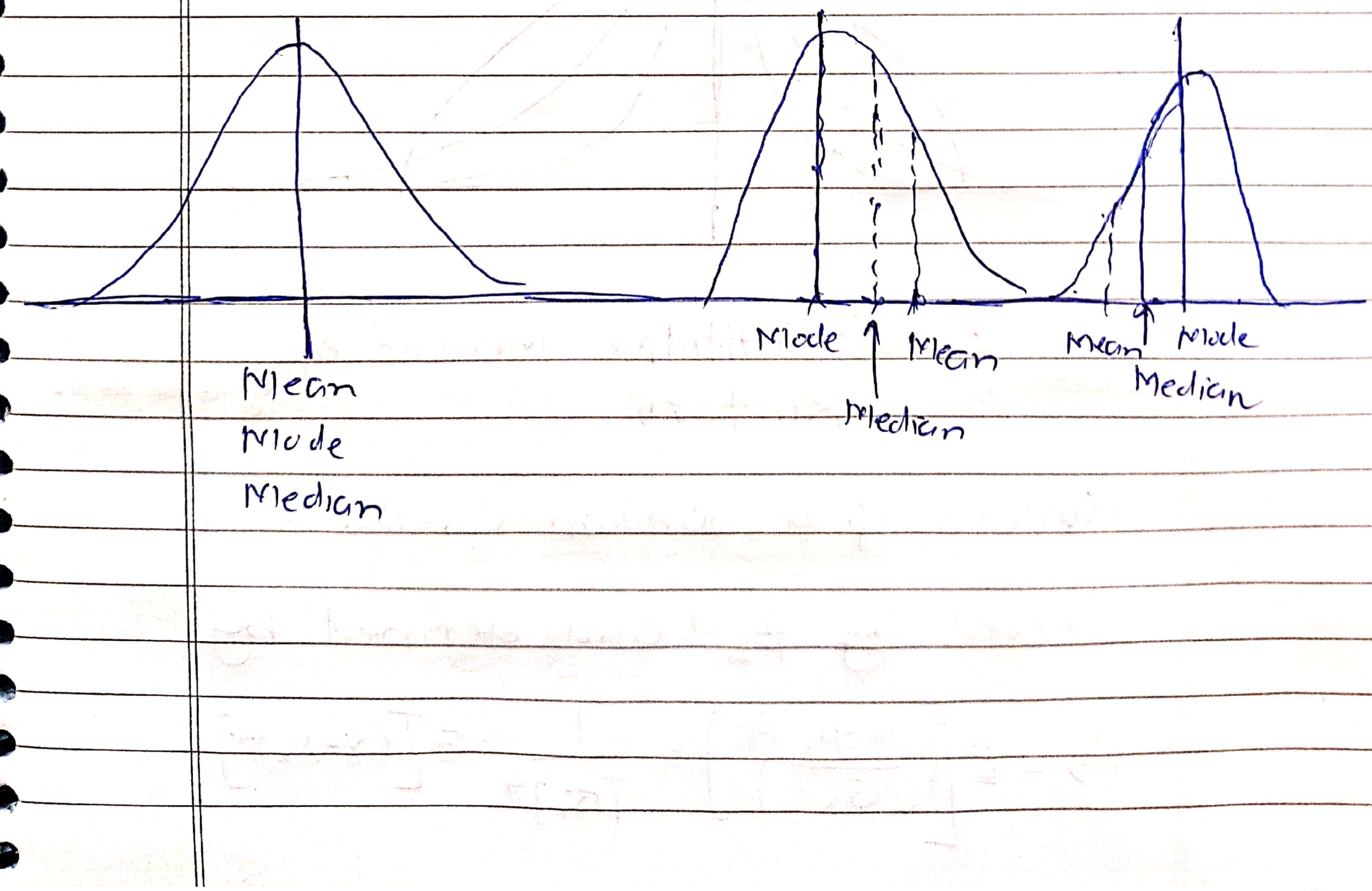
Mean > Median > Mode

(b) For negatively skewed distribution ($\gamma_1 < 0$):

Mean < Median < Mode

(c) For symmetric distribution ($\gamma_1 = 0$):

Mean = Median = Mode.



Enroll

Page No: 60

Date: / /

## Shape of frequency distribution
## ( Peakedness )

Kurtosis (कुकुदता) ≡ Refers to their degree of peakedness.

⇒ Reference is normal distribution

(i) more peaked than the normal distribution

"Lepto ≡ slender ≡ skinny"

↓↓

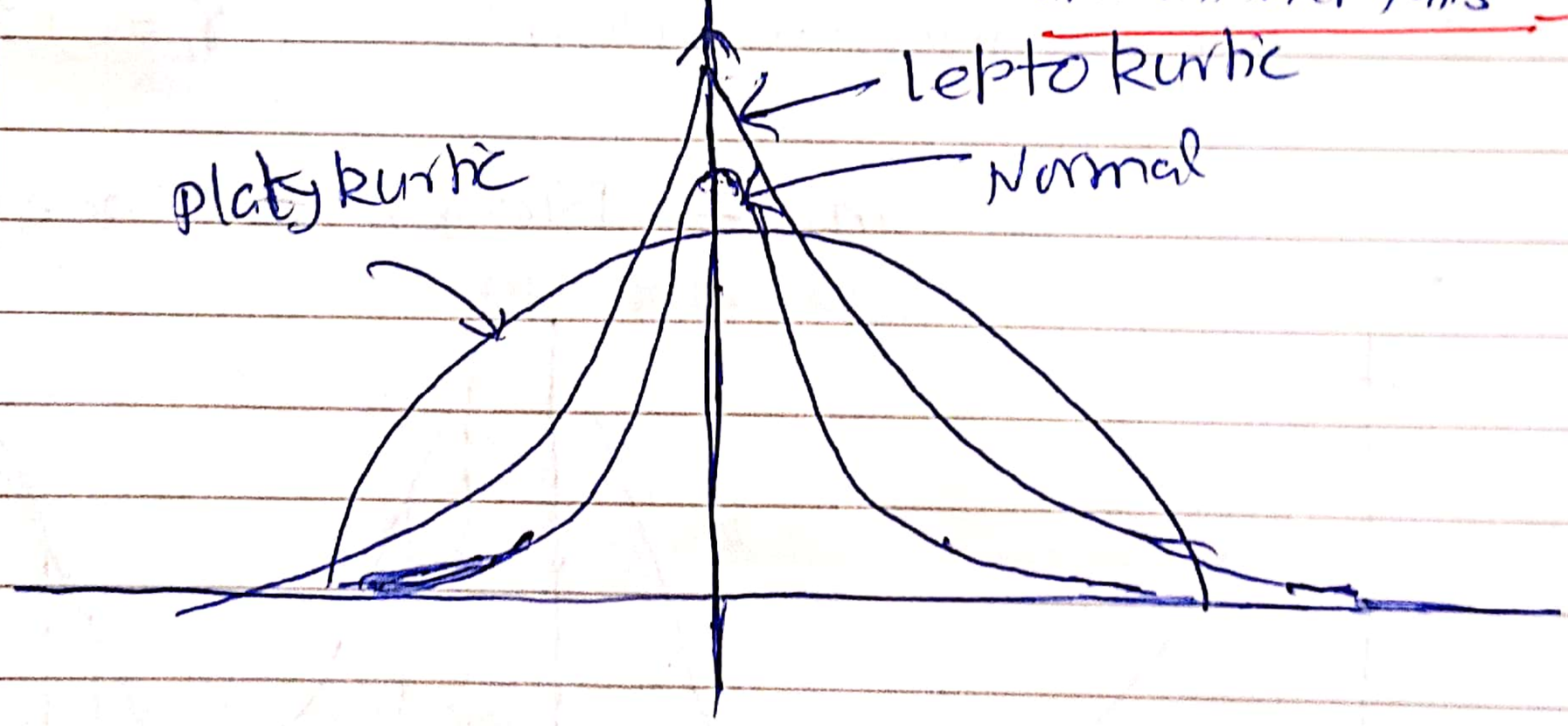
It is said to be leptokurtic ; (more acute peak) | more acute peak around the mean and fatter tails.

(ii) less peaked than the normal distribution

↓↓

platy ≡ "broad"

platykurtic ; | wider peak around mean and thinner tails.




Quantitative Measure of Kurtosis

Kurtosis of a random variable X is denoted by $\beta_2$ and defined by:

$$\beta_2 = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right] = \frac{1}{[\sigma_X]^4} E\left[(X-\mu_X)^4\right]$$

$$\beta_2 = \frac{1}{(\sigma_x^2)^2} E[(x-\mu_x)^4] \quad \text{——— (6)}$$

**Excess Kurtosis** ——— Excess Kurtosis, $\gamma_2$, of the frequency distribution is given by:

$$\gamma_2 = \beta_2 - 3$$

↳ Three (3) is subtracted from $\beta_2$ is used as a correction to make the kurtosis of the ~~normal~~ normal distribution equal to zero.

**Leptokurtic** ⇒ distribution: Have positive excess kurtosis

**Platykurtic** ——— distribution :— Have negative excess kurtosis.

Normal distribution has excess kurtosis of zero.

**Mesokurtic distribution** ← The distribution that is peaked the same way as the normal distribution is called mesokurtic.

Enroll

Page No: 62

Date: / /

## Computation of Kurtosis :-

Data ≡ Population of size N, the Kurtosis is given by

$$b_2 = \frac{\frac{1}{N}\sum_{i=1}^{n}(x_i-\bar{x})^4}{s^4} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2} \quad \text{— (7)}$$

As in the case of the skewness parameter, this computation applies to a population of N members.

Data ≡ Sample of size n, the excess Kurtosis is given by:

$$g_2 = \frac{n-1}{(n-2)(n-3)}\{(n+1)b_2+6\} \quad \text{— (8)}$$