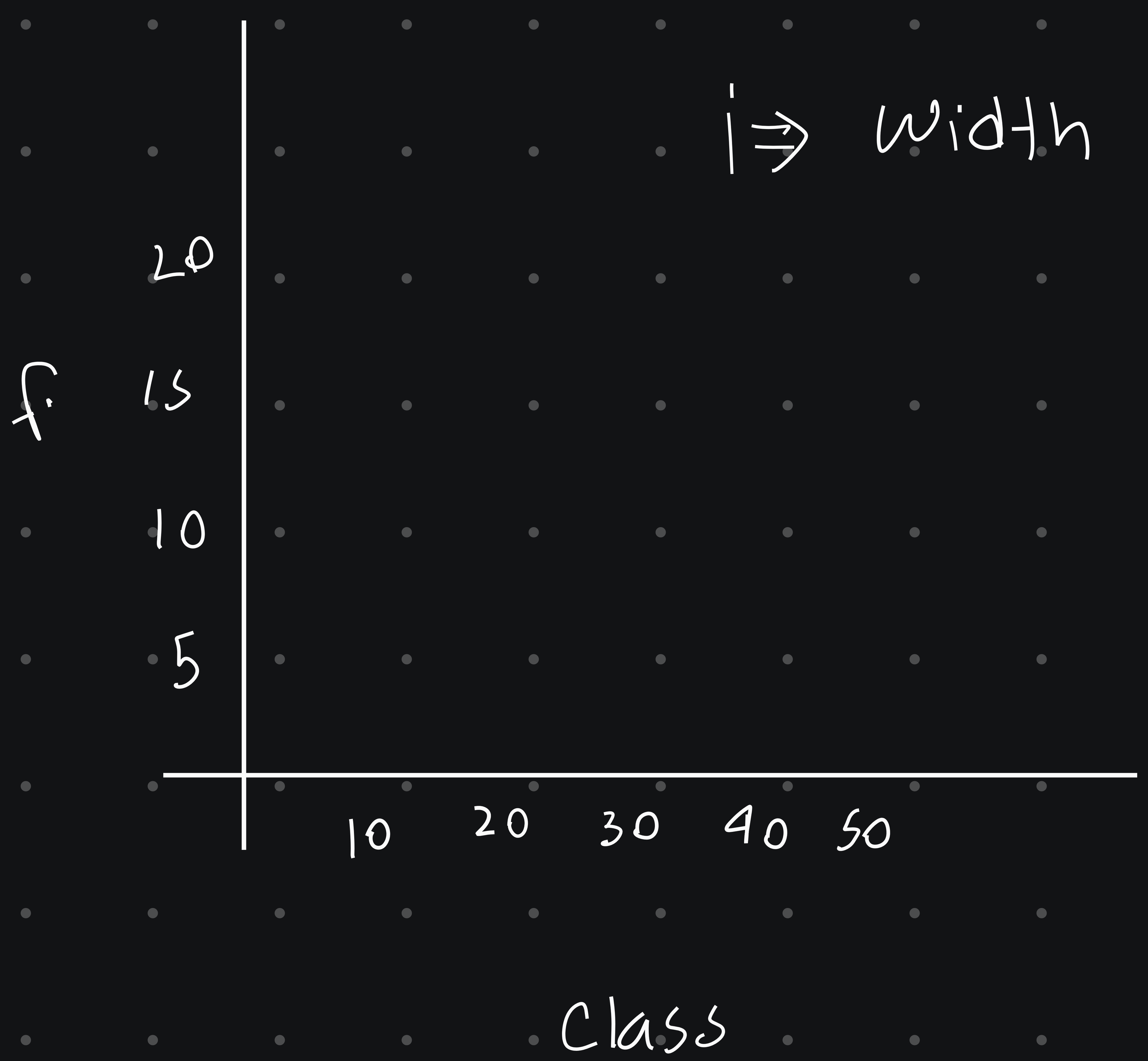
Data — Raw
└ Array

Class interval	frequency	class	class boundaries
0 – 10	7	0 – 9	-0.5 – 9.5
10 – 20	8	10 – 19	9.5 – 19.5
20 - 30	20	20 – 29	19.5 – 29.5
30 – 40	10	30 – 39	29.5 – 39.5
40 – 50	5	40 – 49	39.5 – 49.5

Total 50

$i \Rightarrow$ width of the class interval



Arithmatic mean

(I) Long method

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

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

(II) Short method

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

$d = x_i - A$

where A is assumed mean

(III) Step deviation or Coding method

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

where $u_i = \frac{d_i}{i} = \frac{x - A}{i}$

Class interval	frequency	class	class boundaries	mid Point x	fx	$A=25$ $d_i = x_i - A$	fd
0 − 10	7	0 − 9	−0.5 − 9.5	5	35	−20	−140
10 − 20	8	10 − 19	9.5 − 19.5	15	120	−10	−80
20 − 30	20	20 − 29	19.5 − 29.5	25		0	0
30 − 40	10	30 − 39	29.5 − 39.5	35		10	100
40 − 50	5	40 − 49	39.5 − 49.5	45		20	100
	Total 50						−20

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

$$= 25 + \frac{-20}{50}$$

$$= 24\ 6$$

Class interval	frequency	class	class boundaries	mid point x	fx	$d_i = x_i - A$ ($A = 25$)	fd	u
0–10	7	0–9	-0.5 – 9.5	5	35	-20	-140	
10–20	8	10–19	9.5 – 19.5	15	120	-10	-80	
20–30	20	20–29	19.5 – 29.5	25		0	0	
30–40	10	30–39	29.5 – 39.5	35		10	100	
40–50	5	40–49	39.5 – 49.5	45		20	100	
Total	50						-20	

Median :

$5, 7, \boxed{10}, 13, 16$

$5, 7, \boxed{10, 13}, 16, 20$

11.5

For grouped data

$$\boxed{\text{Median} = L_1 + \left(\frac{\frac{N}{2} - \Sigma f_1}{f}\right) c}$$

$L_1 \Rightarrow$ lower class boundary of the median class

$N \Rightarrow$ total frequency

$f \Rightarrow$ frequency of the mediam

$\Sigma f_1 \Rightarrow$ sum of frequencies of all classes lower than the median class

$c \Rightarrow$ size of the median class interval

Ex) Find median

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
S-10	12	30	34	65	45	25	18

total 229

$N/2 = 114.5$

$L_1 = 40$

$f = 65$

$\Sigma f_1 = 76$

$c = 10$

$$\text{Median} = 40 + \frac{114.5 - 76}{65} \times 10$$

$$= 45.92$$

Mode : (1 টির বেশি হওয়া must

2, 6, 8, 8, 10, 14 → 8

2, 6, 8, 10, 14 → নাই

2, 2, 6, 6, 8, 8, 10, 10, 14, 14 → 2, 6, 8, 10, 14

For grouped data

$$\boxed{\text{Mode} = L_1 + \left(\frac{f - f_1}{2f - f_1 - f_2}\right) c}$$

$L_1 \Rightarrow$

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
S-10	12	30	34	65	45	25	18

$L_1 = 40$ → lower class boundary of m

$f = $ frequency of modal cl (65)

$f_1 = 65 - 34 = 31$

$f_2 = 65 - 45 = 20$

$C = 10$

$$\text{Mode} = 40 + \frac{65 - 31}{2 \times 65 - 31 - 20} \times 10$$

$= 44.303$

Missing frequency

An incomplete distribution is given bellow

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
S-10	10	20	?	40	?	25	15

It is given that total frequency is 170 and the median value is 35.

Find the missing frequencies by using median formula.

Standard daviation!

It is the sequare root of the mean of the square of the deviation from the Arithmatic mean

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

the square of the standard deviation is called the variance σ^2

$$\sigma^2 = \frac{1}{N} \Sigma f (x-\bar{x})^2$$

variance

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

☐ find the s.d for the following distribution:

Mass in kg	60-62	63-65	66-68	69-71	72-74
# of students	5	18	42	27	8

$$\boxed{s.d = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2}}$$

যেকোনো value নেওয়া যাবে

$A = 67$

Soln

mass in kg	# of students	Midpoint (x)	$d = x - A$	fd	fd^2
60-62	5	61	-6	-30	
63-65	18	64	-3		
66-68	42	67	0		
69-71	27	70	3		
72-74	8	73	6		
	100			$\Sigma fd = 45$	$\Sigma fd^2 = 873$

Coefficient of variation (CV)

এটা যার কম তার performance ভালো consistancy

$$\boxed{C.V = \frac{\sigma}{\bar{x}} \times 100}$$

Q. Two cricketers scored the following runs in several innings. Find who is better run-getter and who is more consistent player.

X: 42, 17, 83, 59, 72, 76, 64, 45, 40, 32

Y: 28, 70, 31, 0, 59, 108, 82, 14, 3, 95

Soln

$\bar{x}$

$\bar{y}$

যার average বেশি তার run বেশি

Cricketer X $\bar{X} = 53$

Cricketer Y $\bar{Y} = 49$

X	$(X-\bar{X})$ $= x$	$(X-\bar{X})^2$ $= x^2$
42	-11	121
17		
⋮		
	$\sum x = 530$	$\sum x^2 = 4038$

Y	$(Y-\bar{Y})$ $= y$	$(Y-\bar{Y})^2$ $= y^2$
28	-21	441
70	21	441
	$\sum y = 490$	$\sum y^2 = 1373.4$

$$\sigma_x = \sqrt{\frac{\sum x^2}{N}}$$

$$= 20.09$$

$$C.V_x = \frac{\sigma_x}{\bar{x}_x} \times 100$$

$$= 37.92$$

$\sigma_y = 37.06$

$$C.V_y = \frac{\sigma_y}{\bar{y}} \quad 100$$

$$= 75.63$$

Sample of population

Statistical inference

in the process of using a sample to infer the properties of a population.

Moments:

For the frequency distribution, the first moment about the arithmetic mean is defined as the mean of the deviations of items taken from the A.M.

First moment about the mean

$$\mu_1 = \frac{\Sigma(x-\bar{x})}{N} = \frac{\Sigma d}{N} = 0$$

First moment with frequency

$$\mu_1 = \frac{\Sigma f(x-\bar{x})}{N} = \frac{\Sigma fd}{N}$$

$$\mu_2 = \frac{\Sigma f(x-\bar{x})^2}{N} = \frac{\Sigma fd}{N}$$

$$\mu_2 = \frac{\Sigma(x-\bar{x})^2}{N}$$

$$\mu_3 = \frac{\Sigma(x-\bar{x})^3}{N}$$

$$\mu_4 = \frac{\Sigma(x-\bar{x})^4}{N}$$

Moments about any arbitrary value A

$$\mu_1' = \frac{\Sigma(X-A)}{N}$$

$$\mu_1' = \frac{\Sigma f(X-A)}{N} = \frac{\Sigma fd}{N}$$

$$\mu_2' = \frac{\Sigma(X-A)^2}{N}$$

$$\mu_3' = \frac{\Sigma(X-A)^3}{N}$$

$$\mu_4' = \frac{\Sigma(X-A)^4}{N}$$

Relation between moments about the mean $\overline{X}$ and moments about any arbitrary value A

$$\mu_1 = \mu_1' - \mu_1' = 0$$

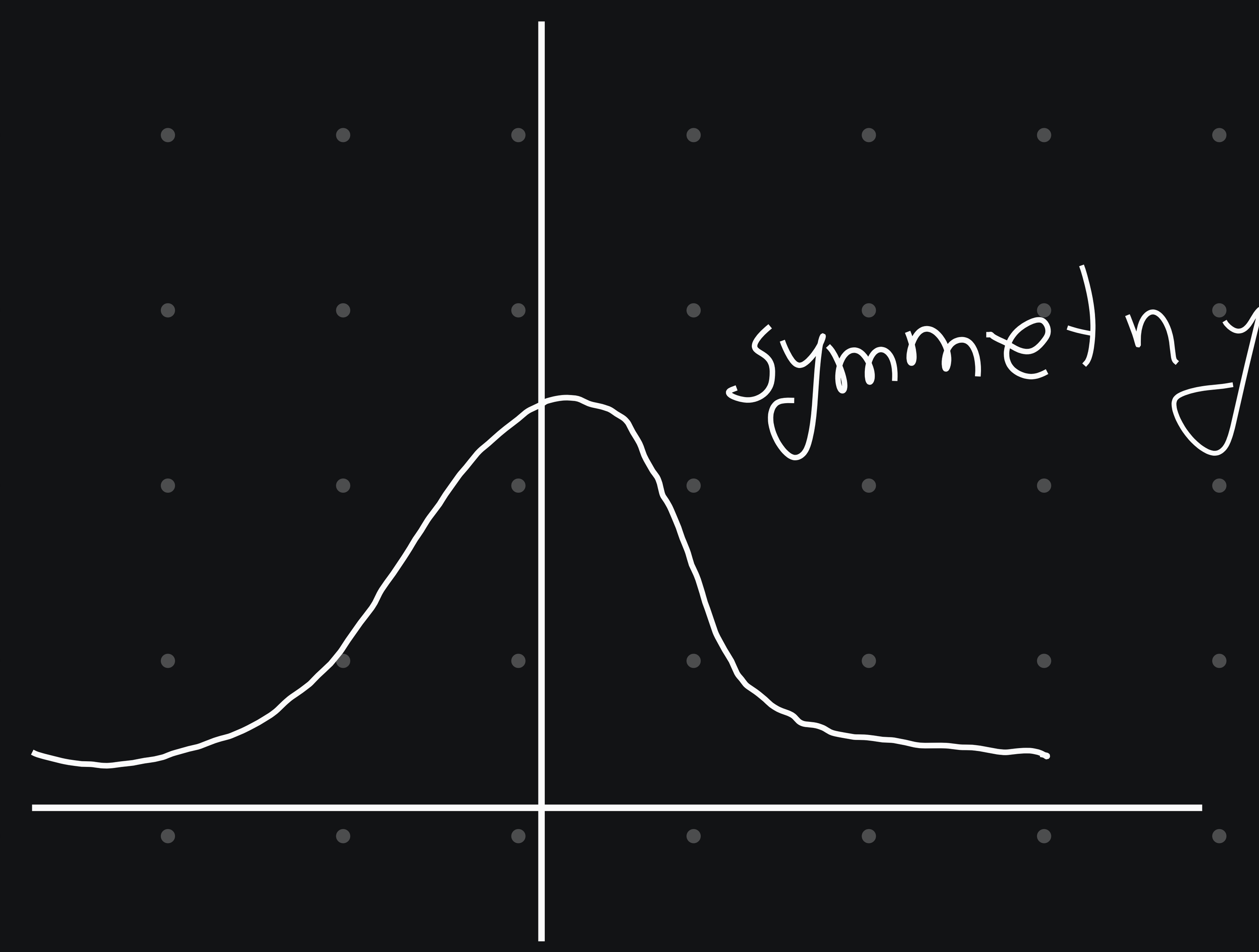
$$\mu_2 = \mu_2' - (\mu_1')^2$$

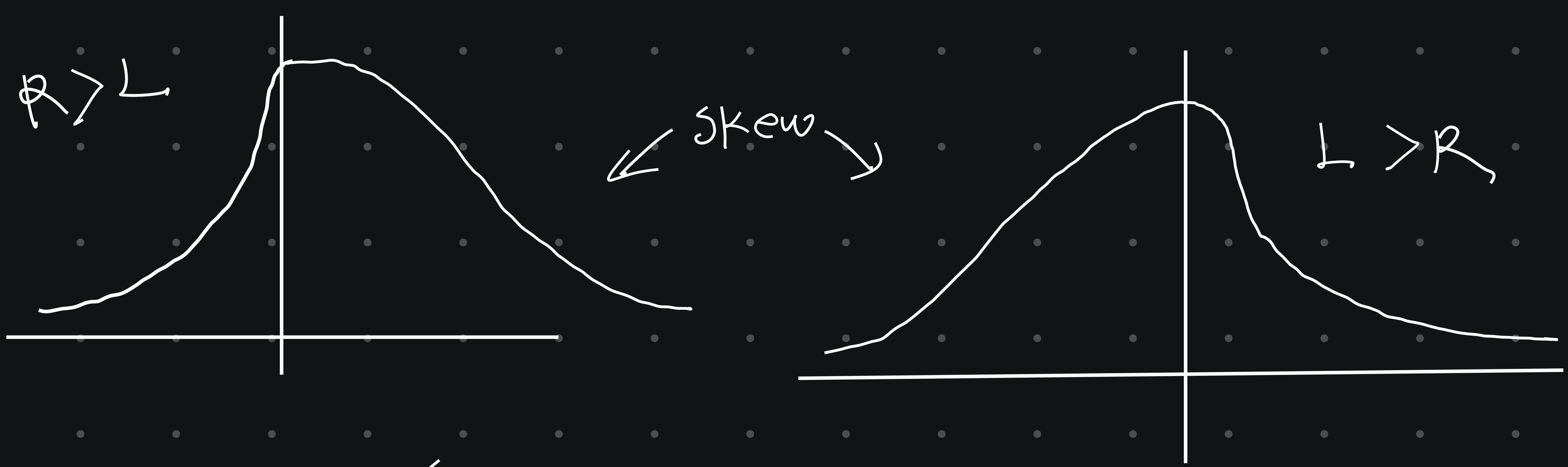
$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3 + 6\mu_2'\mu_1' - 3(\mu_1')^4$$

Skewness

skewness is the lack of symmetry





+ve skewness
on right hand
skewed

-ve skewness
on left hand
skewed

Measures of skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

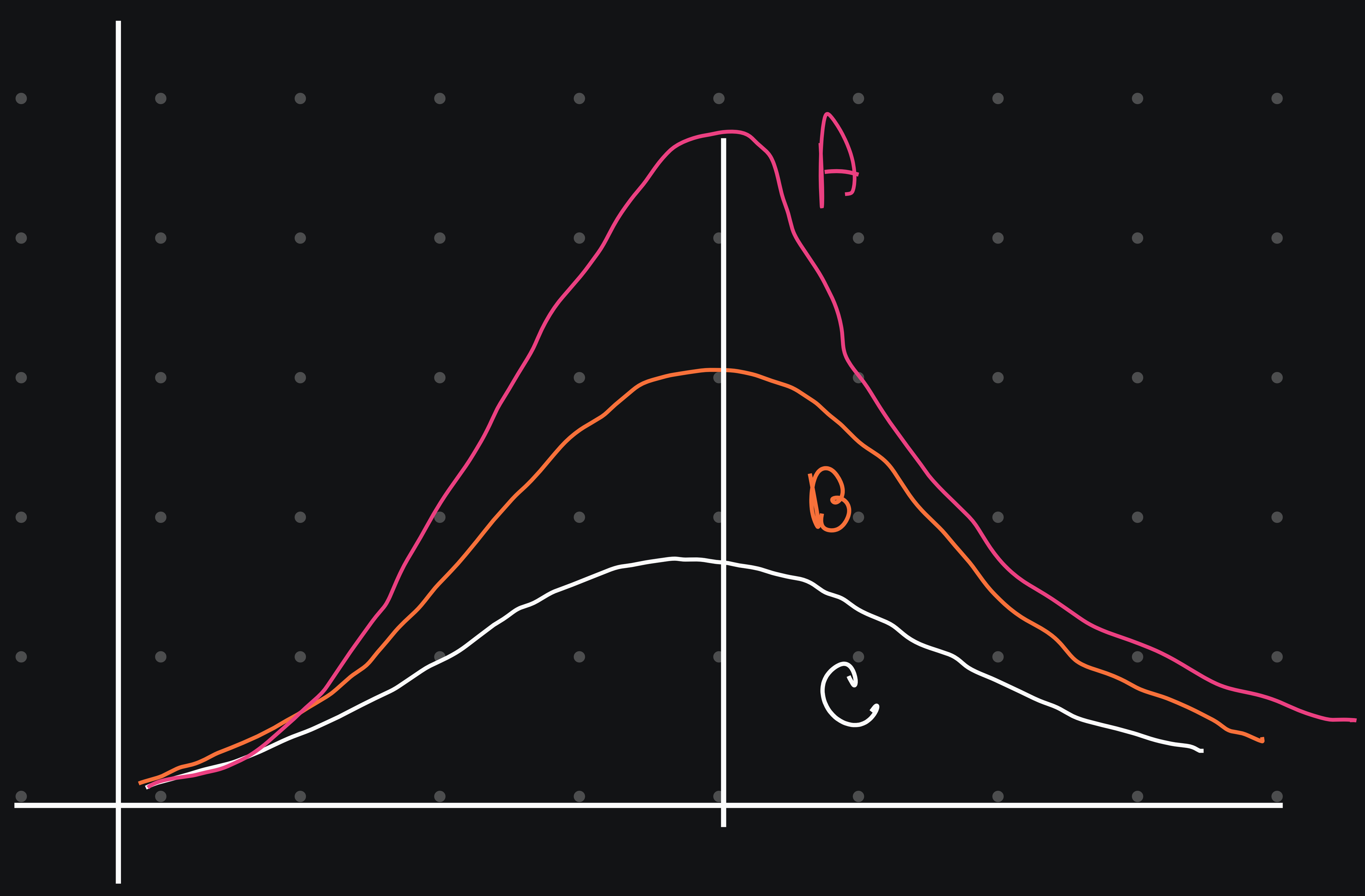
if $\beta_1 = 0 \Rightarrow$ perfectly symmetry

$\beta_1 > 0 \Rightarrow$ +ve skewness

$\beta_1 < 0 \Rightarrow$ -ve skewness

Kurtosis

A measure of Kurtosis indicates the degree to which a curve of a frequency distribution is peaked or flat-topped.



A → Peaked (leptokurti distribution)

B → Normal (mesokurtic distribution)

C → Flat-topped (platykurtic ")

Measures of kurtosis β_2

$$\beta_2 = \frac{\mu_4'}{\mu_2^2}$$

$\beta_2 = 3 +$ mesokurtic

$\beta_2 > 3$, leptokutic

$\beta_2 < 3$, platykurtic distribution

Probability

If an effect can occur in 'a' ways
and fails to " in 'b' "
and these are equally to occur, then
the probability of the even occuring is
$\frac{a}{a+b}$, denoted by p and probability of
not occuring is $\frac{b}{a+b}$, denoted by q

Hence $p+q=1$

$$\text{Probability } p = \frac{\#\text{ of favourable cases}}{\text{total } \#\text{ of equally likely cases}}$$

Conditional probability of B, given A, denoted by $P(B|A)$, is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{if } P(A) \neq 0$$

$$\boxed{\text{A ঘটে গেলে, তারপর B হওয়ার probability}}$$

Example

A study showed that 60% of managers had some business education and some 45% had some engineering education. Furthermore 15% of the mangers had some business education but no engineering education. What is probability that a manger has some business education given that he has some engineering education.

$P(B) \rightarrow 60\%$

$P(E) \rightarrow 45\%$

$P(B \cap E) = 45\% = 60\% - 15\%$

$$P(B|E) = \frac{P(B \cap E)}{P(E)} = 1$$

Dependent events!

from the defination conditional probabilities, if A and B are dependent events,

then

$$P(A \cap B) = P(A|B)\, P(B) \longrightarrow (1)$$

$$P(B \cap A) = P(B|A)\, P(A) \longrightarrow (2)$$

The order is of no significance in the intersection of two events.

since $A \cap B = B \cap A$, we gen an

important property of intersection.

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

$A =$

Independent events:

Two events A and B are said to be independent if and only if

$$P(A \cap B) = P(A).\, P(B)$$

which implies from ① and ②

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

Baye's Theorem:

The applications of the results of the probability theory involves estimating unknown probabilities and making decisions on the basis of new sample information. This concept is referefed to as Baye's theorem.

Probabilities assigned on the basis of personal experience, before observing the outcomes of the experiment are called prior probabilities.

When the probabilities are revised with the use of Baye's rule, they are called posterior probabilities.

Baye's theorem is useful in solving practical business problems in the light of additional information

Ex: A company has two plants to manufacture cars: Plant I manufactures 75% of the cars and Plant II " 25% " " ".

At Plant I, 80 out of 100 cars are rated standard quality.

Plant II, 60 out of 100 cars are rated standard quality.

What is the probability that the car selected at random came from plant I if it is known that the car is of standard quality?

Similarly, from Plant II.

Solve

Let $A_1 \rightarrow$ event that of drawing a car produced by plant I and A_2 be for plant II

$B \rightarrow$ event that of drawing a standard quality car produced by either plant I or plant II.

for the first information $P(A_1) = 0.75$

$P(A_2) = 0.25$

for additional information

$$P\left(B/A_1\right) = \frac{80}{100} = 0.8$$

and $P\left(B/A_2\right) = \frac{60}{100} = 0.6$

The required values are computed in the following table.

Event	Prior Probability	Conditional Probability	Joint Probability	Posterior Probability Reversed
1	2	3	4 = 2 × 3	5 = 4 ÷ P(B)
A_1	0.75	0.8	0.6	$\frac{0.6}{0.75} = 0.8$
A_2	0.25	0.6	0.15	$\frac{0.15}{0.75} = 0.2$
	1	1	$P(B) = 0.75$	1

from the first information we may say that the standard car is drawn from plant I since $P(A_1) = 75\%$ which is greater than $P(A_2) = 25\%$

from the additional information ie plant I, 80 out of 100 and plant II, 60 out of 100 are rated standard quality.

Thus, we may conclude that the standard quality of cans is more likely drawn from the output by plant I.

Population: The group of individuals under study is called population or universe. It may be finite or infinite.

Sampling: A part selected from the population is called a sample. The process of selection of a sample is called sampling. A random sample is one in the which each member of population has an equal chance of being included in it.

There are ^NC_n different samples of size n that can be picked up from a population of size N

Functions of population is called parameter
" " a sample " " statistics.

Standard Error (SE) is the standard deviation of the sampling distribution. For assessing the difference between the expected values and observed values, standard error is used. Reciprocal of standard error is called precision

Testing a Hypothesis:

On the basis of sample information, we make certain decisions about the population. In taking such decisions we make certain assumptions. These assumptions are known as statistical hypothesis.

Null hypothesis: (H_0)

Null hypothesis is based for analysing the problem. Null hypothesis is the hypothesis of no difference. Thus we shall presume that there is no significance difference between the observed value and the expected value sample size $\leq 30 \rightarrow$ T test

n% significance $\rightarrow$ n % ভুল, বাকি (100−n)% ঠিক

calculated value > given value
rejected

otherwise accepted

Errors: In sampling theory to draw valid inferences about the population parameter on the basis of the sample results.

We decide to accept or to reject the results after examining a sample from it. In such we are liable to commit the following two types of errors.

Type I Error: If H_0 is rejected while it should be accepted.

Type II " : If H_0 is accepted while it should be rejected.

Level of significance: There are two critical regions which cover 5% and 1% areas of the normal curve. The shaded portions are