Class-09 Date: 03.10.2023

Imp

Moments:

Arithmatic mean of rth power of deviations taken either from zero, mean or any arbitrary origin are called moments.
describe the characteristic of dist^n / central tendency

1) When deviations are computed from arithmatic mean then, it's called mean moment / central moment.

Notation (moment): μ_r or m.

For ungrouped data:

$$\mu_r = \frac{\sum (x_i - \mu)^r}{N} \quad ; \quad r = (1, 2, 3, 4,) \ldots\ldots$$

For Grouped data:

$$\mu_r = \frac{\sum f_i (x_i - \mu)^r}{N} \quad ; \quad r = (1, 2, 3, 4,) \ldots\ldots$$

Central (moment)

2) When deviations are computed from origin zero, such moments are called moments about origin

Raw moment = μ'_r

For ungrouped data: $\mu'_r = \frac{\sum x_i^r}{N}$

For grouped data: $\mu'_r = \frac{\sum f_i x_i^r}{N}$

3) When deviations of the values are computed from any arbitrary value say A, then it's known as moments about provisional mean.

For ungrouped data:

$$\mu'_r = \frac{\sum D^r}{N} \quad ; r = 1, 2, 3, 4, \ldots$$

$$D = x_i - A$$

For grouped data:

$$\mu'_r = \frac{\sum f_i D^r}{N} \quad ; r = 1, 2, 3, 4, \ldots$$

and $D = x_i - A$

* Relation between Central and Raw moments:

All the raw moments can then be converted into Central moments, mean moments $, A = \bar{x}$

First moment will always be zero

$$\mu_1 = 0 \quad [\text{Always will be zero}] \quad \mu'_1 - \mu'_1 \qquad \frac{\sum (x_i - \bar{x})^1}{N} = \bar{x} - \bar{x} = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_1)^2\mu'_2 - 3(\mu'_1)^4$$

General formula for moment about mean are given below:

$$m_r = \frac{\sum (x_i - \bar{x})^r}{n} \quad ; \quad r = 1, 2, 3, 4$$

If we put, $r = 1, 2, 3, 4$

$$m_1 = \frac{\sum (x_i - \bar{x})}{n} = \text{mean} \qquad m_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$

Skewness

$$m_2 = \frac{\sum (x_i - \bar{x})^2}{n} = \text{Variance} \qquad m_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

(2nd moment = variance)

Kurtosis

(1)

$\boxed{\text{Problem:}}$ Observe the following data set:

32, 36, 36, 37, 39, 41, 45, 46, 48

Find first $\boxed{\text{four}}$ Raw moments. about mean.

$\boxed{\text{Soln:}}$ Mean, $\bar{x} = \dfrac{\sum x_i}{n}$

$$= \frac{360}{9}$$

$$\therefore \bar{x} = 40$$

Sol.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
32	-8	64	512	4096
36	-4	16	64	256
36	-4	16	64	256
37	-3	9	27	81
39	-1	1	1	1
41	1	1	1	1
45	5	25	125	625
46	6	36	216	1296
48	8	64	512	4096
	$\Sigma x_i - \bar{x} = 0$	$\Sigma(x_i - \bar{x})^2 = 232$	$\Sigma(x_i - \bar{x})^3 = 186$	$\Sigma(x_i - \bar{x})^4 = 10708$

First moment, $m'_1 = \frac{\sum(x_i - \bar{x})}{n} = 0$

2nd ", $m'_2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{232}{9}$
≈ 25.78

3rd ", $m'_3 = \frac{\sum(x_i - \bar{x})^3}{n} = 20.67$

4th ", $m'_4 = \frac{\sum(x_i - \bar{x})^4}{n} = 1189.78$

[Ans.]

(2)

$\boxed{\text{Problem:}}$ Calculate first four central moments for the following distribution of wages. ~~using~~ from Arbitrary value, $A = 10$.

Weakly earnings (mid point)	Frequency
	1
5	2
6	5
7	10
8	20
9	51
10	22
11	11
12	5
13	3
14	1
15	

Soln:

Arbitrary value, $A = 10$

Earning (x_i)	Frequency (f_i)	$D_i = x_i - A$ ($A=10$)	f_iD_i	$f_iD_i^2$	$f_iD_i^3$	$f_iD_i^4$
5	1	-5	-5	25	-125	625
6	2	-4	-8	32	-128	512
7	5	-3	-15	45	-135	405
8	10	-2	-20	40	-80	160
9	20	-1	-20	20	-20	20
10	51	0	0	0	0	0
11	22	1	22	22	22	22
12	11	2	22	44	88	176
13	5	3	15	45	135	405
14	3	4	12	48	192	768
15	1	5	5	25	125	625
	$\sum f_i = 131$		$\sum f_iD_i = 8$	$\sum f_iD_i^2 = 346$	$\sum f_iD_i^3 = 74$	$\sum f_iD_i^4 = 3718$

First moment:

$$\mu'_1 = \frac{\sum f_i D_i}{\sum f_i} = \frac{8}{131} = 0.06$$

2nd moment:

$$\mu'_2 = \frac{\sum f_i D_i^2}{\sum f_i} = \frac{346}{131} = 2.64$$

3rd moment:

$$\mu'_3 = \frac{\sum f_i D_i^3}{\sum f_i} = \frac{74}{131} = 0.56$$

4th moment:

$$\mu'_4 = \frac{\sum f_i D_i^4}{\sum f_i} = \frac{3718}{131} = 28.38$$

Raw moments

Then,

$\mu'_1 - \mu'_1$

Central moments

$$\mu_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = 2.64 - (0.06)^2 = 2.64$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$$

$$= 0.56 - (3 \times 0 \times 2.64) + 2(0)^3$$

$$= 0.56$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_1)^2\mu'_2 - 3(\mu'_1)^4$$

$$= 28.38 \text{ [Ans.]}$$

Date: ....

* Moments are some of the constant values in a given data distribution that help the statisticians to confirm the nature and type of data $dist^n$.

⇒ shape of any $dist^n$ can be described by it's various moments, first four are ⇒

⇒ Mean (First moment about mean)

⇒ Variance (2nd moment)

⇒ Third moment is skewness which indicates symmetric, either left / right.

⇒ Fourth moment is Kurtosis which indicates peakness / flatness of $dist^n$.

Date: ....

Skewness and Kurtosis: → degree of peakness

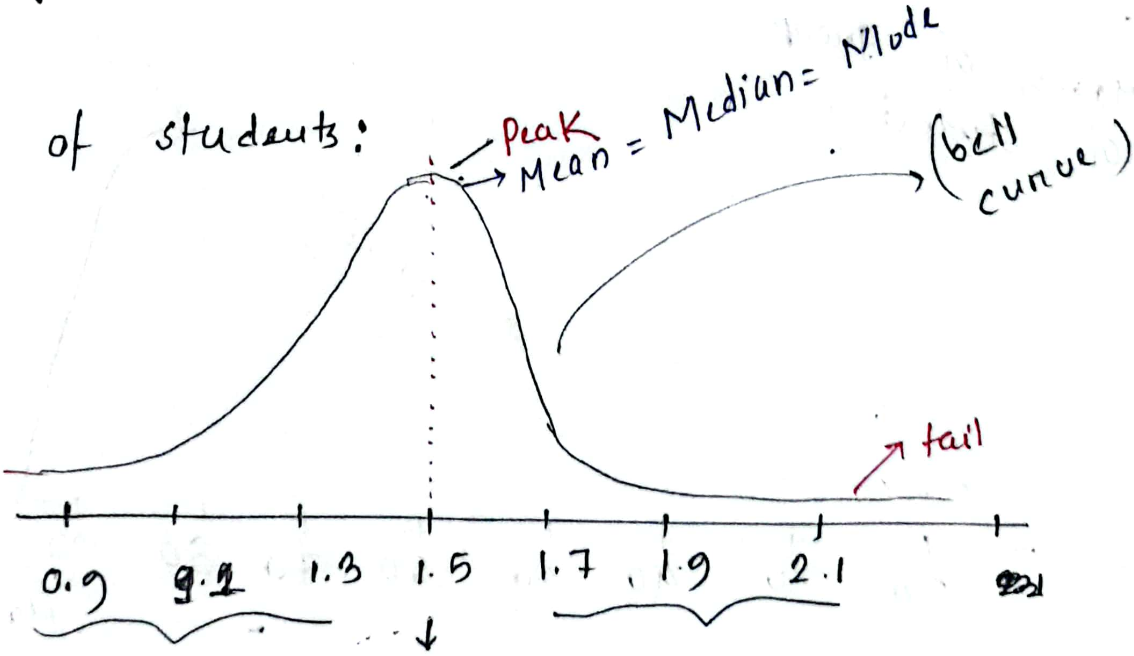
gree

Measurement of a frequency distribution.

⇒ Skewness focuses on the spread of normal distribution. while kurtosis focuses more on the height.

mmetry

Height of students:



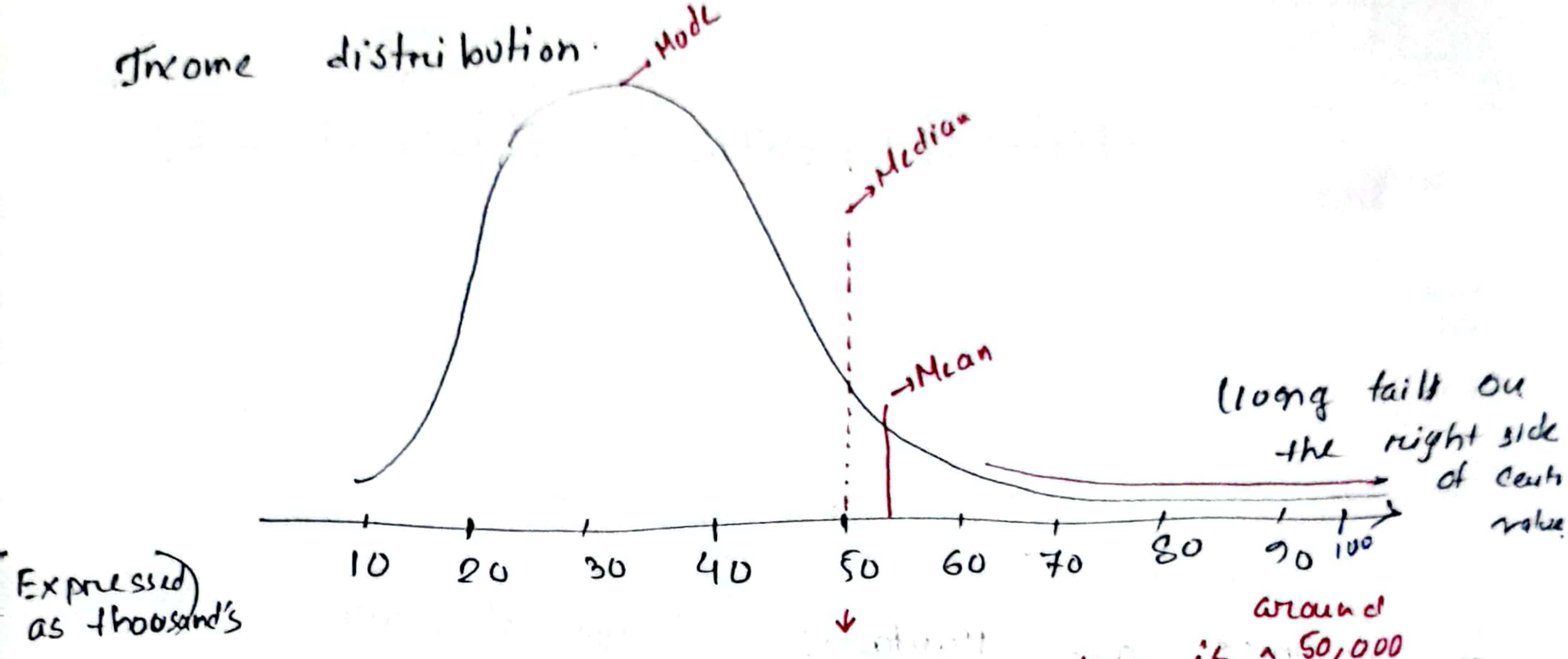
central value (Median)

→ Heights

(Zero skewness)
data set is
..n, symmetrical distributed)

* data set is evenly distributed on both sides. there is no skewness to either the left or the right.

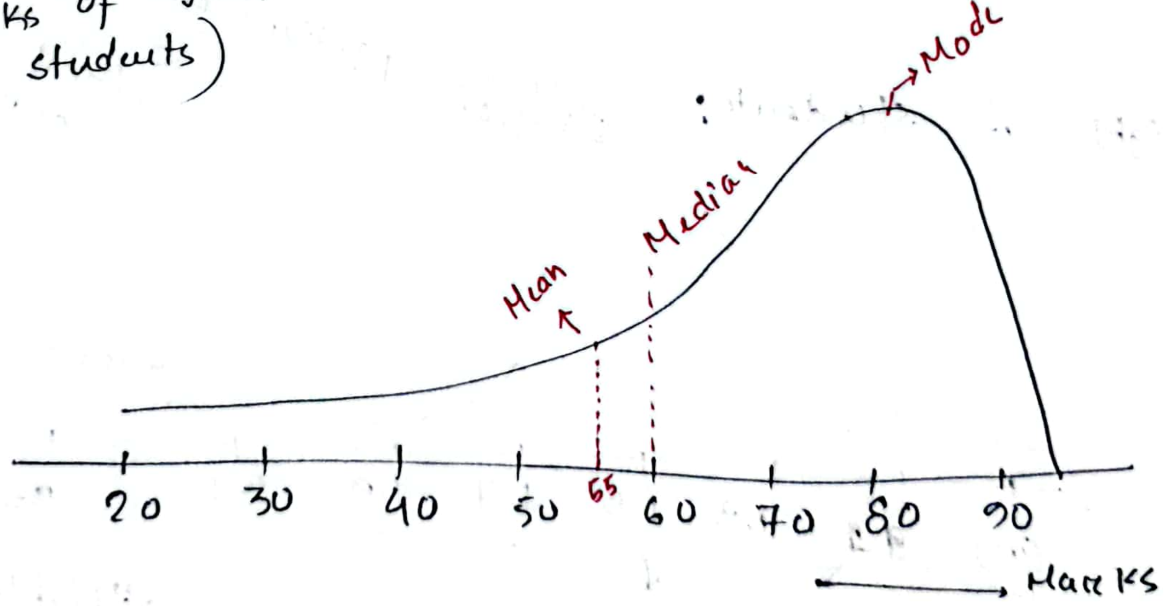
Income distribution



Central values is around 50,000 dollar

(Mean > Median > Mode) (Positive skewness) (Right skewness)

(Marks of group of students)



Mean < Median < Mode

(Negative skewness)

Example:

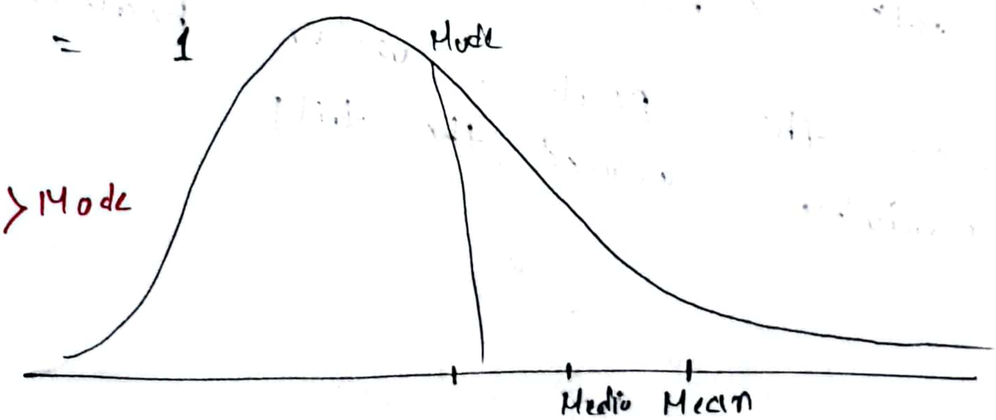
1, 1, 1, $\boxed{2}$ 3, 5, 100 . Describe the shape of statistical distribution.

Soln: $n = 7$

Median $= \frac{7+1}{2}$ th term $=$ 4th $= 2$

Mean $=$ 16.14

Mode $=$ 1



Mean > Median > Mode

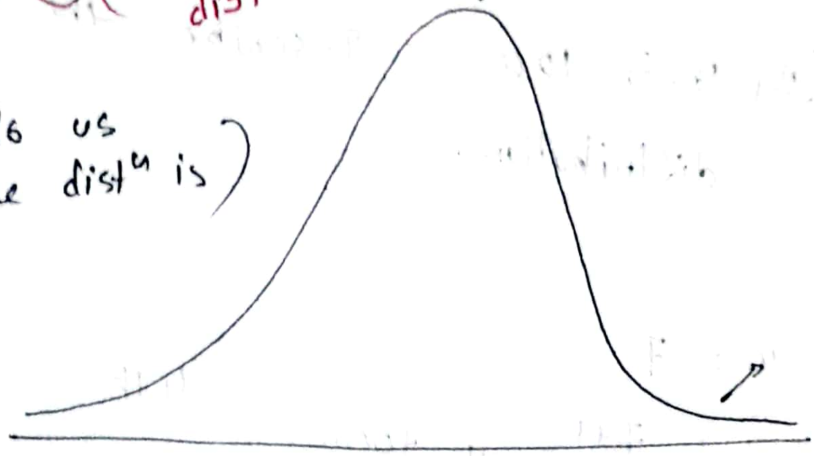
(Positive skewness)

Real life example: (skewness)

Temperature of a particular region.
It could be high/Low depends on climate

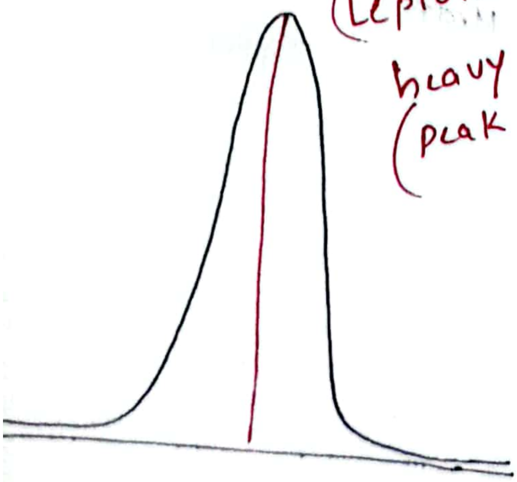
[*] Kurtosis: (center of distribution)

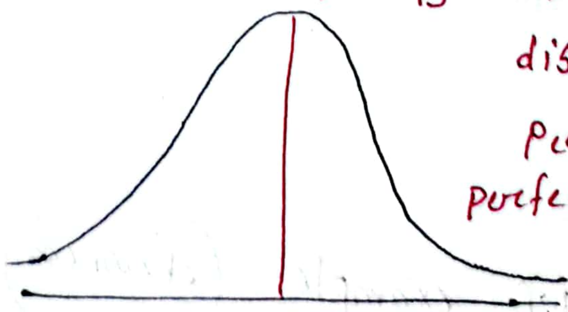
(Peak tells us how tall the dist^n is)

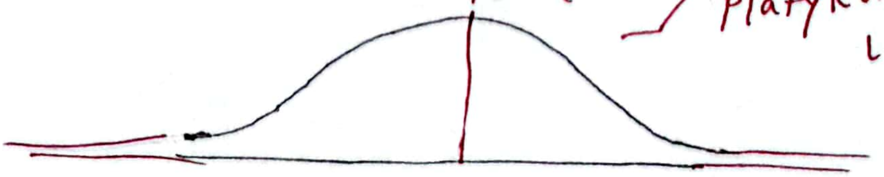


(measurement of how tall are the peak of a dist^n / how heavy / Light tails are)

If we increase the peak, we're heavier tails / more variables around the tail)







*Measure of skewness and kurtosis:

$$\text{skewness} \Leftarrow \beta_1 / b_1 = \frac{\mu_3^2}{\mu_2^3}$$ [moment Ratio]

$$\text{Kurtosis} \Leftarrow \beta_2 / b_2 = \frac{\mu_4}{\mu_2^2}$$ [moment Ratio]

*Karl Pearson's measures of skewness:

$$S_{KP} = \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$$

Or, $$S_{KP} = \frac{3(\text{Mean} - \text{Median})}{\text{Standard deviation}}$$

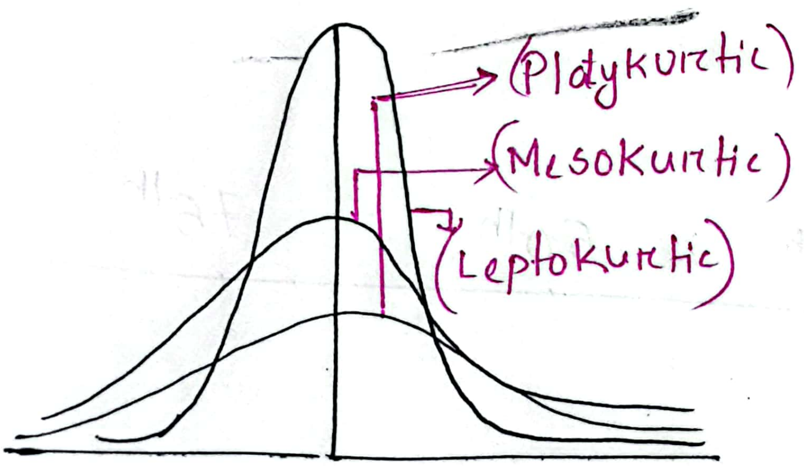
*Bowley's Measure of skewness:

$$S_{KB} = \frac{Q_3 + Q_1 - 2(\text{Median})}{Q_3 - Q_1}$$

Co-efficient of skewness:

Kurtosis:

Degree of peakness or flatness of a frequency distribution.



co-efficient of kurtosis

$$\left[b_2 = \frac{m_4}{m_2^2}\right]$$

1) If $b_2 > 3$; Leptokurtic.

2) If $b_2 = 3$; mesokurtic.

3) If $b_2 < 3$; platykurtic

Class: (05+06) Date: 07.10.2023

Problem: Observe the following:

Age (in years)	Frequency
24.5 – 29.5	3
29.5 – 34.5	9
34.5 – 39.5	15
39.5 – 44.5	12
44.5 – 49.5	7
49.5 – 54.5	4

a) Compute the first four raw moments about A=

b) Find measures of skewness and kurtosis.

c) Examine the shape characteristics of the age distribution by all possible measures.

Soln: a)

Age	Frequency (f_i)	cf	Midpoint (x_i)	$D_i = x_i - A$ ($A = 42$)	f_iD_i	$f_iD_i^2$	$f_iD_i^3$	$f_iD_i^4$
24.5-29.5	3	3	27	-15	-45	675	-10125	151875
29.5 - 34.5	9	12	32	-10	-90	900	-9000	90000
34.5 - 39.5	15	27	37	-5	-75	375	-1875	9375
39.5 - 44.5	12	39	42	0	0	0	0	0
44.5 - 49.5	7	46	47	5	35	175	875	4375
49.5 - 54.5	4	50	52	10	40	400	4000	40000
	$\Sigma f_i = 50$				-135	2525	-16125	295625

$\therefore m_1' = \frac{-135}{50} = -2.7$

$m_2' = \frac{2525}{50} = 50.5$

$m_3' = \frac{-16125}{50} = -322.5$

$m_4' = \frac{295625}{50} = 5912.5$

Now, If we want to find central moments,

$m_1 = m'_1 - m'_1 = 0$

$m_2 = m'_2 - (m'_1)^2 = 43.21$

$m_3 = m'_3 - 3m'_2 m'_1 + 2(m'_1)^3 = 47.18$

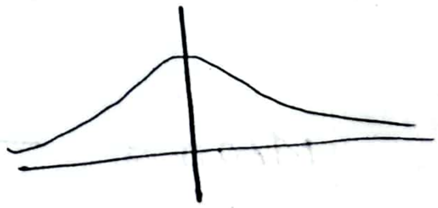
$m_4 = m'_4 - 4m'_3 m'_1 + 6 m'_1 m_1'^2 - 3m_1'^4$

$= 4478.94$

b) Measures of skewness:

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(47.18)^2}{(43.21)^3} = 0.03$$

Measures of kurtosis:



$$b_2 = \frac{m_4}{m_2^2} = \frac{4478.94}{(43.21)^2} = 2.40$$

As $b_2 < 3$, it is platykurtic.

c) First Quartile, $Q_1 = l + \left(\frac{\frac{KN}{4} - C_f}{f}\right) i$

$\frac{KN}{4} = \frac{1 \times 50}{4} = 12.5$ $\quad = 34.5 + \left(\frac{12.5 - 12}{15}\right) 5$

$= 34.67$

Q_2 = Median

$$= l + \frac{h}{f}\left(\frac{n}{2} - cf\right) \qquad \frac{n}{2} = \frac{50}{2}$$

$$= 34.5 + \frac{5}{15}(25 - 12)$$

$$= 38.83$$

$$Q_3 = l + \left(\frac{\frac{KN}{4} - C_f}{f}\right)i$$

$$= 39.5 + \left(\frac{37.5 - 27}{12}\right)5$$

$$= 43.87$$

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1965}{50} = 39.3$$

$$\text{Variance}, = \frac{\Sigma f_i (x_i - \bar{x})^2}{n} = \frac{2160.5}{50} = 43.21$$

Standard deviation, $S = \sqrt{43.21} = 6.6$

Karl Pearson's measure of skewness:

$$S_{kP} = \frac{3(\text{Mean} - \text{Median})}{\text{Standard deviation}}$$

$$= \frac{3(39.3 - 38.83)}{6.6} = 0.21$$

Here, measure of skewness is greater than zero. So, positive skewness.

Bowley's M of Ske

$$S_{kB} = \frac{Q_3 + Q_1 - 2(Me}{Q_3 - Q_1}$$

$$= 0.10$$

[Ans.]