502

Moment, Skewness, Kurtosis

Moments: Raw moment

If $x_1, x_2, \ldots, x_n$ be n observations of a variate 'x' then the rth raw moment defined by

$$\mu_r' = \frac{\sum_{i=1}^{n} (x_i - A)^r}{n} \; ; \text{ where } A \text{ is any arbitrary value,}$$

$r = 1, 2, 3, \ldots$

* helps in describing the shape of the distribution

(1) Central moment

The rth central moment is defined by

$$\mu_r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^r}{n} \; ; \; \bar{x} = \text{mean},$$

$r = 1, 2, 3, 4, \ldots$

$$\mu_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \sigma^2 \text{ i.e. variance}$$

$\sigma = \sqrt{\mu_2}$ = standard deviation

If $x_1, x_2, \ldots, x_k$ occur with frequencies $f_1, f_2, \ldots, f_k$ respectively then the raw moment is

$$\mu_r' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^r}{N}, \quad N = \sum_{i=1}^{k} f_i,$$

A = Any arbitrary value

and rth central moment is

$$\mu_r = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^r}{N}, \quad N = \sum_{i=1}^{n} f_i,$$

$\bar{x}$ = mean

502

161002051 - 163002007
183002080 - 183002087

Relation between central moments and raw moments

Expression of 1st four central moments in terms of raw moments:

The rth raw moment is defined by $\mu_r' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^r}{n}$

$$\mu_1' = \frac{\sum_{i=1}^{n} f_i (x_i - A)}{n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{n} - \frac{\sum_{i=1}^{n} f_i A}{n}$$

$$= \bar{x} - \frac{nA}{n} = \bar{x} - A$$

$$\mu_2' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^2}{n}; \quad \mu_3' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^3}{n}; \quad \mu_4' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^4}{n}$$

Now 1st central moment is $\mu_1 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})}{n}$

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

$$= \bar{x} - \frac{\cancel{n}\bar{x}}{\cancel{n}}$$

$$= 0$$

2nd central moment $\mu_2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{n}$

$$= \frac{\sum_{i=1}^{n} f_i (x_i - A + A - \bar{x})^2}{n}$$

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

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

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

$$= \mu_2' - 2\mu_1'\mu_1' + \mu_1'^2 = \mu_2' - \mu_1'^2$$

3rd central moment $\mu_3 = \dfrac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^3}{n}$

$$= \frac{\sum_{i=1}^{n} f_i (x_i - A + A - \bar{x})^3}{n}$$

$$= \frac{\sum f_i \{(x_i - A) - (\bar{x} - A)\}^3}{n}$$

$$= \frac{\sum f_i \{(x_i - A)^3 - 3(x_i - A)^2(\bar{x} - A) + 3(x_i - A)(\bar{x} - A)^2 - (\bar{x} - A)^3\}}{n}$$

$$= \frac{\sum f_i (x_i - A)^3}{n} - 3\,\frac{(\bar{x} - A)\sum f_i (x_i - A)^2}{n} + 3\,\frac{(\bar{x} - A)^2 \sum f_i (x_i - A)}{n} - \frac{\sum f_i (\bar{x} - A)^3}{n}$$

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

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

Scanned with CamScanner

4^{th} central moment $\mu_4 = \frac{\Sigma f_i (x_i - \bar{x})^4}{n}$

$$= \frac{\Sigma f_i (x_i - A + A - \bar{x})^4}{n}$$

$$= \frac{\Sigma f_i \{(x_i - A) - (\bar{x} - A)\}^4}{n}$$

$$= \frac{\Sigma f_i \{(x_i - A)^4 - 4(x_i - A)^3(\bar{x} - A) + 6(x_i - A)^2(\bar{x} - A)^2 - 4(x_i - A)(\bar{x} - A)^3 + (\bar{x} - A)^4\}}{n}$$

$$= \frac{\Sigma f_i (x_i - A)^4}{n} - \frac{4(\bar{x} - A)\Sigma f_i (x_i - A)^3}{n} + \frac{6(\bar{x} - A)^2 \Sigma f_i (x_i - A)^2}{n} - \frac{4(\bar{x} - A)^3 \Sigma f_i (x_i - A)}{n} + \frac{\Sigma f_i (\bar{x} - A)^4}{n}$$

$$= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 4\mu_1'\mu_1'^3 + \mu_1'^4$$

$$= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

pearson's β and γ coefficients:

Kearl pearson defined the following four coefficients, based upon the first four moments about mean:

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\gamma_1 = +\sqrt{\beta_1}$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$,

$\gamma_2 = \beta_2 - 3$

Scanned with CamScanner

Symmetrical distribution:

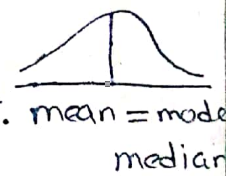
A distribution in which equidistant values on both sides of the center have equal frequencies is called a symmetrical distribution.

In fact, a distribution is symmetrical if its curve when folded at the center, the two halves of the curve coincide.

A symmetrical distribution has the following important properties:

① In a symmetrical distribution mean = median = mode.

② In a symmetrical distribution $\cancel{Q_1} - Me = Q_3 - Me$



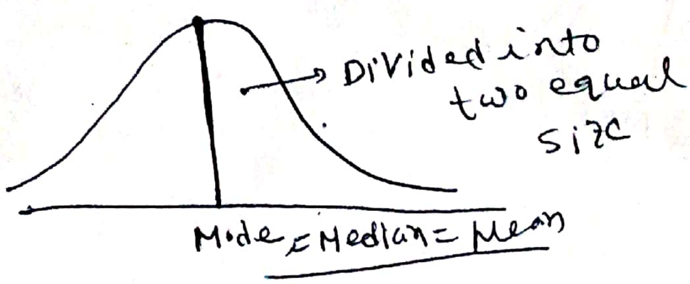
$$Q_3 - Me = Me - Q_1$$

Q_3 → 3nd quartile; Me → median; Q_1 → 1st quartile

③ In a symmetrical distribution all central moments of odd orders vanish that is $\mu_3 = \mu_5 = \mu_7 = \cdots = 0$

④ A symmetrical distribution has equal tail on both sides of central value.

In a symmetrical distribution $\beta_1 = 0$



Skewness:

Skewness means "Lack of symmetry" or departure from symmetry.

A distribution which is not symmetrical is called asymmetrical or skew distribution.

There are two types of skewness namely

(i) positive skewness

(ii) Negative skewness

Positive skewness:

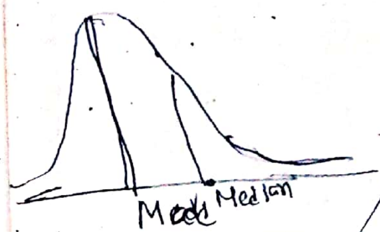
A distribution which has a longer tail on the right side of the mode than the left side is called a positively skew or right skew distribution.

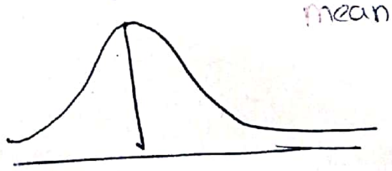
mean > median > mode

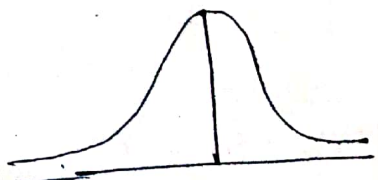
Negative skewness:

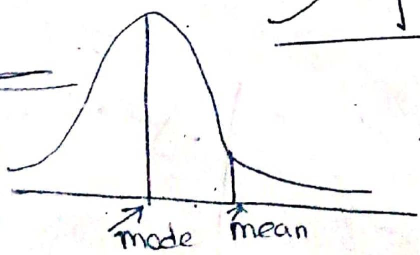
A distribution which has a longer tail on the left side of the mode than the right side, is called a negatively skew or left skew distribution.

mean < median < mode

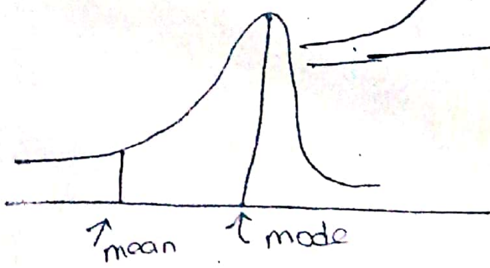








positive skewness



Negative skewness

Measures of skewness:

Pearson's 1st measure of skewness:

$$\text{skewness} = \frac{\text{Mean} - \text{mode}}{\text{standard deviation}}$$

Pearson's 2nd measure of skewness:

$$\text{skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

Bowley's ~~smeas~~ measure of skewness:

$$\text{skewness} = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

Measure of skewness based on moments:

coefficient of skewness:

$$\text{Skewness}\ \gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\mu_3}{\sqrt{(\sigma^2)^3}} = \frac{\mu_3}{\sigma^3}$$

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

$\gamma_1 < 0 \rightarrow$ negatively skewed

$\gamma_1 > 0 \rightarrow$ positively skewed

If $\beta_1 = 0$, ~~then~~ the distribution is symmetric

Kurtosis:

Kurtosis means to the peakedness or flatness of the curve of a distribution, relative to the normal curve.

A distribution whose curve is highly peaked is called a Leptokurtic distribution.

A distribution whose curve is flat topped is called a platykurtic distribution.

Keare ~~Pearson's coefficient~~ of

A distribution whose curve is neither peaked nor flat topped is called a mesokurtic distribution.

Measure of Kurtosis:

Karl pearson's measure of kurtosis

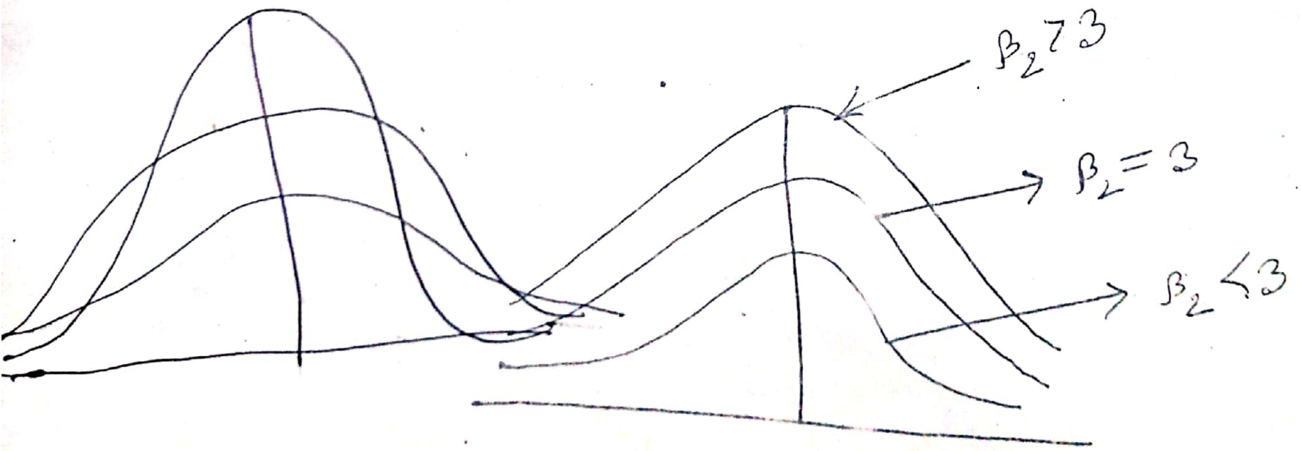
Coefficient of kurtosis $\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$

Mesokurtic distribution $\beta_2 = 3$.

Leptokurtic. " $\beta_2 > 3$

Platykurtic " $\beta_2 < 3$

Skewness and kurtosis are collectively known as shape, characteristic of a distribution, since they show the light on the shape of a distribution.



Different types of kurtosis

2020

Problem: The first four moments about the value 4 are $-1.5, 17, -30$ and 108, find the moments about mean and also find β_1 and β_2.

Problem: The 1st four central moments of a distribution are $0, 16, -36$ and 120. Comment on the skewness and kurtosis of the distribution.

Solution: Given,

$\mu_1 = 0, \mu_2 = 16, \mu_3 = -36$ and $\mu_4 = 120$.

For commenting on the skewness we calculate γ_1

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{-36}{4^3} = -0.5625 \quad [\because \sigma = \sqrt{\mu_2} = \sqrt{16} = 4$$

Since μ_3 is negative and $\gamma_1 < 0$

so the distribution is negatively skewed.

For commenting on the kurtosis we calculate β_2

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

$$= \frac{120}{(16)^2}$$

$$= 0.469 < 3$$

So the distribution is platykurtic.

Scanned with CamScanner

Problem: The first four moments of a distribution about 4 are 1.5, 17, -30 and 108. Find the moments about mean and also find β_1 and β_2.

Soln: The 1st four moments about 4 are

$\mu_1' = -1.5$, $\mu_2' = 17$, $\mu_3' = -30$ and $\mu_4' = 108$

central moments,

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

$$= 17 - (-1.5)^2$$

$$= 14.75$$

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

$$= -30 - 3 \times 17 \times (-1.5) + 2 \times (-1.5)^3$$

$$= 39.75$$

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

$$- 3\mu_1'^4$$

$$= 108 - 4 \times (-30) \times (-1.5)$$

$$+ 6 \times 17 \times (-1.5)^2 - 3 \times (-1.5)^4$$

$$= 142.31$$

Now $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(39.75)^2}{(14.75)^3} = 0.4926$

Scanned with CamScanner

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

$= \frac{142.31}{(14.75)^2}$

$= 0.6543$

problems: For a frequency distribution mean, median and coefficient of variance 25, 20 and 50% respectively. Find the mode and variance of the distribution.

Soln: Given that

Mean, $\bar{x} = 25$

Mode, $M_0 = ?$

$C.V = 50\%$

Median $= 20$

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

$\Rightarrow 50 = \frac{\sigma}{25} \times 100$

$\therefore \sigma = 12.5$

Variance $= \sigma^2$

$= (12.5)^2$

Mode $=$ 3 Median $-$ 2 Mode

$= 3 \times 20 - 2 \times 25$

$= 10$

Problem: For a skew distribution mean = 100, C.V = 35%, Karl Pearson's skewness is 0.2. Find median and mode.

Soln: Given that,

Arithmetic mean, $\bar{x} = 100$

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

$$\Rightarrow 35 = \frac{\sigma}{\bar{x}} \times 100$$

$$\therefore \sigma = 35$$

$$\text{Karl Pearson's skewness} = \frac{\bar{x} - M_0}{\sigma}$$

$$\Rightarrow 0.2 = \frac{100 - M_0}{35}$$

$$\therefore M_0 = 93$$

We know, Mode = 3 Median − 2 Mean

$$\Rightarrow 93 = 3 \times \text{Median} - 2 \times 100$$

$$\therefore \text{Median} = 97.65$$

Problem: For a mesokurtic distribution Skewness = − 0.03 and $\mu_3 = -4$. Find 4th central moment.

Soln: We know,

$$\text{Skewness} = \sqrt{\beta_1}$$

$$\therefore \sqrt{\beta_1} = -0.03$$

$$\therefore \beta_1 = 0.0009$$

Again, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.0009$

$$\Rightarrow \mu_3^2 = 0.0009 \times \mu_2^3$$

$$\Rightarrow \mu_2^3 = \frac{\mu_3^2}{0.0009}$$

$$= \frac{(-4)^2}{0.0009}$$

$$\therefore \mu_2 = 26.09$$

Kurtosis,

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

$$\Rightarrow \mu_4 = \mu_2^2 \times \beta_2$$

$$= (26.09)^2 \times 3$$ [For mesokurtic distribution $\beta_2 = 3$]

Problem: The 1st three moments of a distribution about the value 3 are -1, 5 and 9. Find C.V and 3rd central moment.

Soln: The 1st three central moments about the value 3 are

$\mu_1' = -1$, $\mu_2' = 5$ and $\mu_3' = 9$

2nd central moment,

$$\mu_2 = \mu_2' - \mu_1'^2$$
$$= 5 - 1 = 4$$

3rd central moment;

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$= 9 - 3 \times 5(-1) + 2 \times (-1)^3$$
$$= 22$$

we know,

$$\mu_1' = \bar{x} - A$$
$$\Rightarrow 1 = \bar{x} - 3$$
$$\therefore \bar{x} = 2$$

and $\sigma^2 = \mu_2 = 4$

$$\therefore \sigma = 2$$

Coefficient of variance,

$$C.V = \frac{\sigma}{\bar{x}} \times 100$$
$$= \frac{2}{2} \times 100$$
$$= 100\%$$

Problem: From the following data, calculate the Skewness and comment on the result:

Marks Obtained	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of Students	3	5	7	9	11	13	15

Soln:

Table for Calculation of Skewness

Classes	Mid value (x_i)	Frequency f_i	$d_i = \frac{x_i - A}{c}$	$f_i d_i$	$f_i d_i^2$	cf
20-30	25	3	-3	-9	27	3
30-40	35	5	-2	-10	20	8
40-50	45	7	-1	-7	7	15
50-60	55 A	9	0	0	0	24
60-70	65	11	1	11	11	35
70-80	75	13	2	26	52	48
80-90	85	15	3	45	135	63
		N = 63		$\Sigma f_i d_i = 56$	$\Sigma f_i d_i^2 = 252$	

Determination of Skewness:

$$Sk = \frac{3(\bar{x} - Me)}{SD}$$

Here, Mean, $\bar{x} = A + \frac{\Sigma f_i d_i}{N} \times c$

$= 55 + \frac{56}{63} \times 10$

$= 63.89$

Standard deviation,

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2} \times h$$

$$= \sqrt{\frac{252}{63} - \left(\frac{56}{63}\right)^2} \times 10$$

$$= 17.89$$

$$\text{Median} = L_1 + \frac{\frac{N}{2} - f_c}{f_m} \times c$$

$$= 60 + \frac{31.5 - 24}{11} \times 10$$

$$= 66.82$$

Skewness, $$Sk = \frac{3(\bar{x} - Me)}{SD}$$

$$= \frac{3(63.89 - 66.82)}{17.89}$$

$$= -0.49$$

Comments: Negative Skewed.

Problem: Calculate the skewness and kurtosis from the following distribution:

Allowance (TK)	110-115	115-120	120-125	125-130	130-135	135-140	140-145	145-150	150-155	155-160
Number	17	20	26	49	72	90	92	33	17	7

Soln:

Calculation table for Moments, Skewness and kurtosis

Classes	f_i	Mid value x_i	d	$f_i d_i$	$f_i d_i^2$	$f_i d_i^3$	$f_i d_i^4$
110-115	17	112.5	-5	-85	425	-2125	10625
115-120	20	117.5	-4	-80	320	-1280	5120
120-125	26	122.5	-3	-78	234	-702	2706
126-130	49	127.5	-2	-98	196	-392	784
130-135	72	132.5	-1	-72	72	-72	72
135-140	90	137.5	0	0	0	0	0
140-145	92	142.5	1	92	92	92	92
145-150	33	147.5	2	66	132	164	528
150-155	17	152.5	3	51	153	459	1377
155-160	7	157.5	4	28	112	448	1792
	$N=423$			$\Sigma f_i d_i = -176$	$\Sigma f_i d_i^2 = 1736$	$\Sigma f_i d_i^3 = 3308$	

$\Sigma f_i d_i^4 = 2249$

1st Raw moment,

$$\mu_1' = \frac{\Sigma f_i d_i}{N} \times C$$

$$= \frac{-176}{423} \times 5$$

$$= -2.08$$

2nd Raw moment.

$$\mu_2' = \frac{\Sigma f_i d_i^2}{N} \times C^2$$

$$= \frac{1736}{423} \times 5^2$$

$$= 102.6$$

3rd Raw moment

$$\mu_3' = \frac{\Sigma f_i d_i^3}{N} \times C^3$$

$$= \frac{-3308}{423} \times 5^3$$

$$= -977.54$$

4th raw moment.

$$\mu_4' = \frac{\Sigma f_i d_i^4}{N} \times C^4$$

$$= \frac{22496}{423} \times 5^4$$

$$= 33238.77$$

We know

1st central moment, $\mu_1 = 0$

2nd central moment, $\mu_2 = \mu_2' - (\mu_1')^2$

$= 102.6 - (-2.08)^2$

$= 98.27$

3rd central moment,

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

$= -977.54 - 3 \times 102.6 \times (-2.08) + 2 \times (-2.08)^3$

$= -355.314$

4th central moment,

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$= 27712.8$

We know

$\sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-355.314}{\sqrt{(98.27)^3}} =$

Skewness, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-355.314)^2}{(98.27)^3}$

$= 0.133$

Kurtosis $= \frac{\mu_4}{(\mu_2)^2} = \frac{27712.8}{(98.27)^2} = 2.87$

Comment:

(i) The nature of distribution is positive because Sk. is positive

(ii) Since the kurtosis < 3, it is platykurtic.

Problem: Calculate the skewness from the following data and comment on the result:

Income (Tk)	200-300	300-400	400-500	500-600	600-700	700-800	800-900
No. of worker	3	10	25	18	12	7	4

Calculation Table for Skewness

Classes	Frequency f_i	Mid value x_i	d_i	$f_i d_i$	$f_i d_i^2$
200-300	3	250	-2	-6	12
300-400	10	350	-1	-10	10
400-500	25	450 A	0	0	0
500-600	18	550	1	18	18
600-700	12	650	2	24	48
700-800	7	750	3	21	63
800-900	4	850	4	16	64
	N=19			$\Sigma f_i d_i = 63$	$\Sigma f_i d_i^2 = 215$

We know

$$\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{SD}$$

Here Mean $(\bar{x}) = A + \frac{\sum f_i d_i}{N} \times c$

$$= 450 + \frac{63}{79} \times 100$$

$$= 529.75$$

$$\text{Mode} = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

$$= 400 + \frac{15}{15+7} \times 100$$

$$= 468.18$$

$$SD = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2} \times c$$

$$= \sqrt{\frac{215}{79} - \left(\frac{63}{79}\right)^2} \times 100$$

$$= 144.415$$

$$\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{SD}$$

$$= \frac{529.75 - 468.18}{144.415}$$

$$= .426$$

Comment: The result shows that the distribution is positive, because the coefficient of skewness is positive.

H.W Calculate the Skewness from the following data:

Daily Income	1-50	51-100	101-150	151-200	201-250	251-300	301-400
No. of family	5	8	15	25	13	6	3

Ans: $Sk = -0.08$

Problem: The information of the payment of income tax of some factories are given below:

Income Tax (TK)	100-120	120-140	140-160	160-180	180-200	200-220	220-240
No. of factory	73	53	199	194	327	208	2

Calculate the co-efficient of skewness and kurtosis and comment on the result.

Ans: Skewness $= 0.38$

kurtosis $= 2.57$

H.W

From the following information calculate the coefficient of skewness and comment on the nature of this distribution:

Age	21-25	26-30	31-35	36-40	41-45
People	8	15	25	12	10