Unit - 2

SKEWNESS AND KURTOSIS.

★ Moments :- it is used to describe the peculiarities of freq. distribution i.e one can measure the central tendency the despertion, the A-symmetry & peakedness

★ Moments about the mean :- it is represented by 'μ'

case i :- in series of indivial observations

$$\mu_1 = \frac{1}{N} \Sigma(x_i - \bar{x}) \quad , \quad \mu_2 = \frac{1}{N} \Sigma(x_i - \bar{x})^2$$

$$\mu_3 = \frac{1}{N} \Sigma(x_i - \bar{x})^3$$ & so on.

case ii :- in series of freq. distributio $\mu_1 = \frac{1}{N} (\Sigma x_i - \bar{x}) f$

$$\mu_2 = \frac{1}{N} \Sigma f(x_i - \bar{x})^2 \quad \mu_3 = \frac{1}{N} \Sigma f(x_i - \bar{x})^3$$

Note :- 1) the 1st moment is always zero ($\mu_1 = 0$)

2) the 2nd moment indicate the variants

3) Karl pearson suggested a measure of skewness using 3rd & 2nd monent as

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

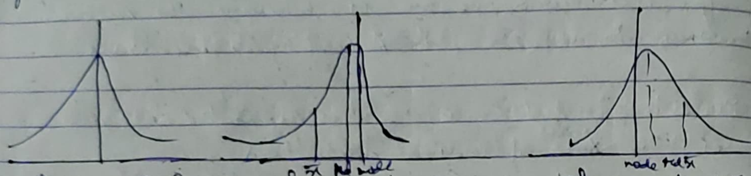
In a symmetrical distribution $\mu_3 = 0$ & ~~that~~ $\beta_1 = 0$.

★ Skewness :- it is a measure that refers to the extent of symmetry or A-symmetry in a distribution i.e it describe the shape of the distribution.

In case of symmetrical distribution the two tail are of equal length. in case of asymmetrical distribution the two tails are of different lengths.

If the left tale is longer than the right tale then the dist. is -vely swekked & the means occurs earier than the more

If the right tale is longer than the left tale then the distribution is +vely skewed & the mode comes first followed mode, meadian & mean.



Skewness is 0 — mean, mode & median coinside

skewness is -ve — mean 1st, median (2nd) & mode (3rd)

skewness is +ve — 1st mode, median, mean

Objectives :- 1) it helps to find out the nature & degree of concentration
2) it helps us in knowing if distribution is normal

★ Measure of skewness :-

* Karl pearson's coeff of skewness :-

$$S_k = \frac{\text{mean} - \text{mode}}{\text{std. devation}}$$

if mode is ill defined then $S_k = \frac{3(\text{mean} - \text{median})}{\text{std. devia}}$

1) consider foll. distribution

	mean	median	st. devation
dist. A	100	90	10
dist. B	90	80	10

check whether both distribution have degree of skewness
⇒

A → $S_k = \frac{3(100-90)}{10}$ B → $S_k = \frac{3(90-80)}{10}$

$S_k = 3$ $S_k = 3$

2) calculate Karl pearson's coeff of skewness for foll. data

value : 10 20 30 40 50 60 70
freq : 1 5 12 22 17 9 4

⇒

mode = ~~22~~ 40

$\text{mean} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{10+100+360+880+850+540+280}{7}$

mean = 43.14

$S.D = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} = 13.26$

$S_k = \frac{\text{mean} - \text{mode}}{\text{st. dev}} = \frac{43.14 - 22}{13.26} = \cancel{1.59}\ 0.2368$

x	f	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10	1	1098.25	
20	5	535.459	
30	12	172.65	
40	22	9.85	
50	17	47.05	
60	9	284.25	
70	4	721.45	

3) calculate coeff of skewness bases on mean & median for the foll.

C.I	f	x	c.f	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
0-10	6	5	6		
10-20	12	15	18		
20-30	22	25	40		
30-40	48	35	88		
40-50	56	45	144		
50-60	32	55	176		
60-70	18	65	194		
70-80	6	75	200		

⇒

$S_k = \frac{3(\text{mean} - \text{median})}{S.D}$ $\text{mean} = \frac{\Sigma fx}{\Sigma f}$ $S.D = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$

median $= l + \frac{h}{f}\left[\frac{N}{2} - c\right]$

mean $= 41.7$, S.D $= 15.43$, ~~median~~

$l = 40$, $h = 10$, $N = 200$, $f = 56$, $c = 81$

median $= l + \frac{h}{f}\left[\frac{N}{2} - c\right] = 40 + \frac{10}{56}\left[\frac{200}{2} - 81\right]$

median $= 42.14$ //

$S_K = \frac{3(\text{mean} - \text{median})}{S.D} = \frac{3(41.7 - 42.14)}{15.43}$

$S_K = -0.085$ //

★ **Bowley coefficient of skewness** :- It is also referred as quartile coeff. of skewness it is useful if the mode is ill-defined & distribution is of unequal class intervals

$$\boxed{S_K = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}}$$

$Q_1 = l + \frac{h}{f}\left[\frac{N}{4} - c\right]$ $Q_2 = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$ $Q_3 = l + \frac{h}{f}\left[\frac{3N}{4} - c\right]$

i) Calculate Bowley measure of skewness for the foll. data

C.I	No. of sales	x	C.f	
1000 - 1200	4	1100	4	
1200 - 1400	10	1300	14	
1400 - 1600	16	1500	30	
1600 - 1800	29	1700	59	
1800 - 2000	52	1900	111	Q_1
2000 - 2200	80	2100	191	Q_2
2200 - 2400	32	2300	223	Q_3
2400 - 2600	23	2500	246	
2600 - 2800	17	2700	263	
2800 - 3000	7	2900	270	

$\frac{270}{4} = 67.5$

$\frac{N}{2} = \frac{270}{2} = 135$

$\frac{3N}{4} = 202.5$

for $Q_1 \rightarrow Q_1 = 1800 + \frac{200}{52}\left[\frac{270}{4} - 59\right]$ $Q_1 = 1832.69$

$Q_2 \rightarrow Q_2 = 2000 + \frac{200}{80}\left[\frac{270}{2} - 111\right]$ $Q_2 = 2060$

$Q_3 \rightarrow Q_3 = 2200 + \frac{200}{32}\left[\frac{270 \times 3}{4} - 191\right] = 2271.875$ //

$S_K = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

2) In a freq. distribution the coeff of skewness based on quartiles is 0.6 if the sum of the lower & upper quartile is 100 & median is 18 find the value of lower & upper quartile

$\Rightarrow$

$S_K = 0.6$

$Q_1 + Q_3 = 100$ $Q_1 = ?$ & $Q_3 = ?$

Md $(Q_2) = 18$

$S_K = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$ $0.6 = \frac{100 - 2(18)}{100 - Q_1 - Q_1}$

$0.6 = \frac{100 - 36}{100 - 2Q_1}$, $60 - 1.2Q_1 = 64$

$1.2Q_1 = 124$

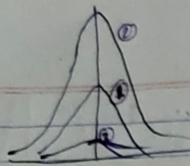
$Q_1 = \frac{124}{1.2}$

$\therefore Q_3 = 103.33$

$Q_1 = 100 - Q_1 = 100 - 103.33$ $Q_1 = -3.33$ //

★ **Kurtosis** :- Kurtosis is an another measure that tells about the form of the distribution.

The degree of Kurtosis is measured relative to the peakedness of normal curve



*The peak curve ② is called leptokurtic it has -ve Kurtosis.
*The intermediately curve ① is called normal or meso kurtic
* flat top curve ③ is called plat kurtic curve it has +ve kurtosis.

★ Measure of Kurtosis

$$\left[\beta_2 = \frac{\mu_4}{\mu_2^2}\right]$$

Also

$$\left[\gamma_2 = \beta_2 - 3\right]$$

Note :- 1) for normal curve $\beta_2 = 3$, $\gamma_2 = 0$
2) for leptokurtic curve $\beta_2 > 3$ or $\gamma_2 > 0$
3) for plat kurtic curve $\beta_2 < 3$ or $\gamma_2 < 0$

q:- 1) give 2, 3, 7, 8, 10
(i) find 1st, 2nd, 3rd & 4th moment about the mean
(ii) find the skewness & kurtosis

⇒

x	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
2	-4	16	-64	256
3	-3	9	-27	81
7	1	1	1	1
8	2	4	8	16
10	4	16	64	256

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{30}{5} = 6$

$\mu_1 = \frac{1}{n} \Sigma (x - \bar{x}) = 0$

$\mu_2 = \frac{1}{n} \Sigma (x - \bar{x})^2 = \frac{46}{5} = 9.2$

$\mu_3 = \frac{1}{n} \Sigma (x - \bar{x})^3 = \frac{-18}{5} = -3.6$

$\mu_4 = \frac{1 \Sigma (x - \bar{x})^2}{n} = \frac{610}{5} = 122$

$S_k = \frac{\mu_3^2}{\mu_2^2} = \frac{-3.6^2}{9.2^2} = 0.15$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{122}{9.2^2} = 1.44$