05: Skewness And Kurtosis

(M) → * **Moments :-**

- Moments are the mean of various powers of deviation of observation from any value.
- If the deviations of observations are taken from AM → Central moments (μ_r) as moments about the mean.
- whn ever the deviations are taken from values other than mean, → Raw Moments (μ_r') / Arbitory moments / Non-Central moments.

* **Definition :-**

If $x_1, x_2 \dots x_n$ are the n obsr. in a data & 'A' is any arbitory constant, then the r^{th} moment about A is defined as,

- For raw data –

$$\boxed{\mu_r' = \frac{1}{n} \Sigma (x-A)^r} \qquad r = 1, 2, 3 \dots$$

- For freq distribution –

$$\boxed{\mu_r' = \frac{1}{N} \Sigma f (x-A)^r} \qquad N = \Sigma f.$$

* **Note 1:-**

when $r = 1$, $\mu_r' = \mu_1' = \frac{1}{n} \Sigma (x-A)$

when $r = 2$, $\mu_2' = \frac{1}{n} \Sigma (x-A)^2$ ↳ Mean square deviation

when $r = 3$, $\mu_3' = \frac{1}{n} \Sigma (x-A)^3$

when $r = 4$, $\mu_4' = \frac{1}{n} \Sigma (x-A)^4$

* **Note 2:-**

In particular $A = 0$, we get the Raw moment about the origin 0.

$\mu_1'(0)$.

$\mu_2'(0)$

$$\mu_r'(0) = \frac{1}{n} \Sigma (x-0)^r = \frac{1}{n} \Sigma x^r$$

∴ whn $r = 1$, $\mu_1'(0) = \frac{1}{n} \Sigma x$.

" $r = 2$, $\mu_2'(0) = \frac{1}{n} \Sigma x^2$

" $r = 3$, $\mu_3'(0) = \frac{1}{n} \Sigma x^3$

" $r = 4$, $\mu_4'(0) = \frac{1}{n} \Sigma x^4$.

⇒ **Central Moments :-** (μ_r)

when $A = \bar{x}$, we get C. Moments.

∴ r^{th} central moment are defined as

$$\boxed{\mu_r = \frac{1}{n} \Sigma (x-\bar{x})^r} \longrightarrow \text{for raw data}$$

$$\boxed{\mu_r = \frac{1}{N} \Sigma f (x-\bar{x})^r} \longrightarrow \text{for freq data}$$

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

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

$r = 2$, $\mu_2 = \frac{1}{n} \Sigma (x-\bar{x})^2 = \sigma^2$ (varience)

$r = 3$, $\mu_3 = \frac{1}{n} \Sigma (x-\bar{x})^3$

$r = 4$, $\mu_4 = \frac{1}{n} \Sigma (x-\bar{x})^4$

* **Remarks :-**

1. $\mu_1 = 0$
2. $\mu_1'(A) = \frac{1}{n} \Sigma (x-A)$

$= \frac{1}{n} \Sigma x - \frac{1}{n} \Sigma A.$ $\qquad A = n$ times

$= \bar{x} - \frac{1}{n} \times n$

$$\boxed{\mu_1'(A) = \bar{x} - A}$$

3. $\mu_1'(0) = \frac{1}{n}\sum(x-0)^1 = \frac{1}{n}\sum x = \bar{x}$

4. $\mu_2 = \sigma^2$ (varience)

5. $\mu_0 = \frac{1}{n}\sum(x-\bar{x})^0$ $\quad (x)^0 = 1$

$= \frac{1}{n}\sum 1 = \frac{1}{n}\times n = 1$

$\boxed{\mu_0 = 1}$

⇒ Relation b/w Raw·M & central·M :-

$$\mu_r = \mu_r' - {}^rC_1\,\mu_{r-1}'\,\mu_1' + {}^rC_2\,\mu_{r-2}'\,(\mu_1')^2 \ldots + (-1)^r\,(\mu_1')^r$$

Proof

General form →
$$\boxed{(a-b)^n = a^n - {}^nC_1\,a^{n-1}b + {}^nC_2\,a^{n-2}b^2 - \ldots + (-1)^n b^n}$$

we have $\mu_r' = \frac{1}{n}\sum(x-A)^r$ $\quad r = 1,2,3\ldots$

$\mu_r = \frac{1}{n}\sum(x-\bar{x})^r$

Adding & subtracting A,

$\mu_r = \frac{1}{n}\sum(x - A - \bar{x} + A)^r$

taking -ve

$= \frac{1}{n}\sum\left((x-A) - (\bar{x}-A)\right)^r$ $\quad$ Since $(\bar{x}-A) = \mu_1'(A)$

$= \frac{1}{n}\sum\left[\underset{(a}{(x-A)} - \underset{b)^n}{\mu_1'}\right]^r$

$= \frac{1}{n}\sum\left[(x-A)^r - {}^rC_1(x-A)^{r-1}\times\mu_1' + {}^nC_2(x-A)^{r-2}\times(\mu_1')^2 \ldots + (-1)^r\times(\mu_1')^r\right]$

$= \frac{1}{n}\sum(x-A)^r - {}^rC_1\frac{1}{n}(x-A)^{r-1}\times\mu_1' + {}^rC_2\frac{1}{n}(x-A)^{r-2}\times(\mu_1')^2 - \ldots + (-1)^r\frac{1}{n}\sum(\mu_1')^r$

$\therefore \mu_r = \mu_r' - {}^rC_1\,\mu_{r-1}'\,\mu_1' + {}^rC_2\,\mu_{r-2}'(\mu_1')^2 - \ldots + (-1)^r(\mu_1')^r$

Hence the proof.

In the above result putting $r = 1,2,3,4$ $\quad {}^nC_r$

$(r=3)$

$\mu_1 = \mu_1' - 0 - \mu_1' = 0$ $\quad {}^1C_1$, ${}^1C_0$ 1 always ≤ 1

$\mu_2 = \mu_2' - (\mu_1')^2$ $\quad \times {}^1C_2$ → 1>2

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

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

⇒ Effect of changes of origin & scale on (M):-

let $x_1, x_2 \ldots x_n$ be the set of obs^r.

let $y_i = \frac{x_i - A}{c}$, $i = 1,2\ldots n$.

$A \to$ constant.

$c \to$ width of cls.

$(-A\ x - A\ \ldots)$

$c\,y_i = x_i - A$ —①

$c\,\bar{y} = \bar{x} - A$ —②

①-②. $c y_i - c\bar{y} = x_i - \cancel{A} - \bar{x} + \cancel{A}$.

$c(y_i - \bar{y}) = x_i - \bar{x}$

$c^r(y_i - \bar{y})^r = (x_i - \bar{x})^r$

Apply $\frac{1}{n}\sum$ on both side.

$\frac{1}{n}\sum c^r(y_i-\bar{y})^r = \frac{1}{n}\sum(x_i - \bar{x})^r$

$c^r\frac{1}{n}\sum(y_i-\bar{y})^r = \frac{1}{n}\sum(x_i-\bar{x})^r$

(x's → x values)

$c^r \times \mu_r$ of y values $= \mu_r$ of x values.

$\boxed{r^{th} \text{ moment of } x\text{'s} = c^r \times r^{th} \text{ moment of } y\text{'s}}$

⇒ Sheppard's correction :-

For a freq distr we cannot determine the exact values of (M), bcz in a freq. table we hue assumed that these items who fall in a c/s is having on an avg a value at the mid point of c/s. Hence there exist an error of (μ) & higher orders.

The S. correction —

$$\mu_2 (\text{corrected}) = \mu_2 - \frac{c^2}{12}$$

$$\mu_3 (\text{corrected}) = \mu_3$$

$$\mu_4 (\text{corrected}) = \mu_4 - \frac{c^2}{2} \times \mu_2 + \frac{7c^4}{240}$$

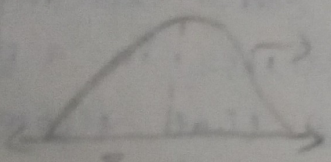
c = c/s interval.

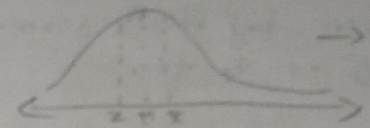
⇒ Skewness :-

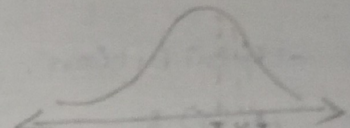
(S) means lack of symmetry or departure from symmetry.

In a freq distribution if more values are > Mode the freq curve has a longer tale to the right than to left then the distribution → +vely skewed.

If more values are < Mode the freq curve has a longer tail to left than to right. Then it is said to be -vely skewed / -ve skewness.


→ Symmetry
$\bar{X} = M = Z$


→ +vely skewed


→ -vely skewed

⇒ Absolute and Relative Measures of Skewness —

* A distri... for which Mean = Median = Mode is → symmetrical distri...
* For a asymmetrical distri... these Measures do not coinside.
* In a distri... AM > M > Z, It is said to be +vly skewed.
* for a -vely skewed distri → AM < M < Z
* Absolute Measure of (S) —

$$\boxed{\begin{aligned} \text{Skewness} &= \text{Mean} - \text{Mode} \\ \text{Skewness} &= 3(\text{Mean} - \text{Median}) \end{aligned}}$$

* Relative Measures of (S) —

1) Karl pearson's coefficient of (S) —

$$\boxed{Sk_p = \frac{\text{Mean} - \text{Mode}}{SD}}$$

$$SK_P = \frac{3(\text{mean} - \text{median})}{SD}.$$

The values given by this formula usually lies -1 to $+1$.

when $SK_P = 0$

=> mean = mode.

=> symmetrical distribution.

2) **Bowley's coefficient of (S) :-**

$$\boxed{SK_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

easy cal. (S) measure

3) **Kelly's coefficient of (S) :-**

$$\boxed{SK_K = \frac{P_{90} + P_{10} - 2M}{P_{90} - P_{10}}}$$

4) **Moment Coefficient of Skewness :-**

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

(gamma) $\gamma = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \frac{\mu_3}{\sigma^3}.$

Note -

$AM = A + \frac{\Sigma fd}{\Sigma f} \times c$

$AM = A + \mu_1'$ (about A)

1) Cal. Bowley's coefficient of skewness.

C/S	F	CF
0-10	8	8
10-20	15	23
20-30	24	47
30-40	21	68
40-50	12	80
	80	

$$SK_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \quad \text{—(1)}$$

$\frac{N}{4} = 20$

Q_1 c/s → 10-20

$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m\right)}{f} \times c$

$= 10 + \frac{20-8}{15} \times 10$

$= 10 + \frac{12}{15} \times 10$

$= 10 + 8 = 18$

$\frac{N}{2} = 40$

$Q_2 = 20-30$

$M = l_2 + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$

$= 20 + \frac{40-23}{24} \times 10$

$= 20 + \frac{17}{24} \times 10 = 20 + 7.08 = 27.08$

$\frac{3N}{4} = \frac{3 \times 80}{4} = 60$

$Q_3 = 30-40$

$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m\right)}{f} \times c$

$= 30 + \frac{60-47}{21} \times 10 = 30 + 6.19$

$= 36.19$

Sub in eq (1)

$$Sk_B = \frac{36.19 + 18 - 27.08 \times 2}{36.19 - 18}$$

$$= \frac{54.19 - 54.16}{18.19} = \underline{\underline{0.0016}}$$

$= 1.64 \times 10^{-3}$
$= \frac{1.64}{1000}$
$= 0.0016$

2) Cal Karl Pearson (C) of (S).

cls	F	x	$d = \frac{x-82}{5}$	fd	fd^2	d^2
65–69	8	67	−3	−24	72	9
70–74	15	72	−2	−30	60	4
75–79	18	77	−1	−18	18	1
79.5–84.5 80–84	25	82	0	0	0	0
85–89	14	87	1	14	14	1
90–94	9	92	2	18	36	4
95–99	6	97	3	18	54	9
100–104	5	102.	4	20	80	16
	100			−2	334	

$$Sk_P = \frac{\text{Mean} - \text{Mode}}{SD.}$$

(sk also possible)

$$\text{Mean} = A + \frac{\Sigma fd}{\Sigma f} \times c$$

$$= 82 + \frac{-2}{100} \times 5$$

$$= 82 + 0.02 = \underline{\underline{81.93}}$$

middle
$A = 82$
$c = 4+1$
$= 5$
[for actual cls].

$$Z = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c.$$

$$= 79.5 + \frac{7}{7+11} \times 5$$

$\Delta_1 = 25 - 18 = 7$
$\Delta_2 = 25 - 14 = 11$

high freq
↓
Modal cls.
↓
80–84

$$= 79.5 + 1.944 = \underline{\underline{81.4}}$$

$$SD = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} \times c$$

$$= \sqrt{\frac{334}{100} - \left(\frac{-2}{100}\right)^2} \times 5$$

$$= \sqrt{3.34 - 0.0004} \times 5$$

$$= \sqrt{3.3396} \times 5$$

$$= 1.827 \times 5 = \underline{\underline{9.137}}$$

$\left(\frac{-2}{100}\right)^2 = 4^{-4}$
$= \frac{-4}{10000}$
$= 0.0004$

$$Sk_P = \frac{0.5}{9.137} = \underline{0.054}$$

3) Cal suitable measure of skewness —

cls	F	cf
0–20	20	20
20–50	50	70
50–100	68	139
100–250	30	169
250–500	25	194
500–1000	19	213
	213	

(c = 30)

$$Q_1 = l + \frac{\left(\frac{N}{4} - m\right)}{f} \times c$$

$\frac{N}{4} = 53.25$

$\boxed{20-50}$

$$= 20 + \frac{53.25 - 20}{50} \times 30$$

$$= 20 + 19.95 = \underline{39.95}$$

$$M = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$$

$\frac{N}{2} = 106$

$= 50 + \frac{(106.60 - 70)}{69} 50$

$= 50 + 26.5 = \underline{\underline{76.5}}$

$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - M\right)}{f} \times c.$ $\quad \frac{3N}{4} = \underline{\underline{159.75}}$

$= 100 + \frac{(159.75 - 139)}{30} \times 150$

$= \underline{\underline{203.75}}$

$SK_p = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$= \frac{203.75 + 39.95 - 2 \times 76.45}{203.75 - 39.95}$

$= \underline{\underline{0.5543}}$

⇒ **Kurtosis :-**

(K) means ~~bulgi~~ bulginess ~~bulb~~ ~~bul~~ ~~bluginess~~.

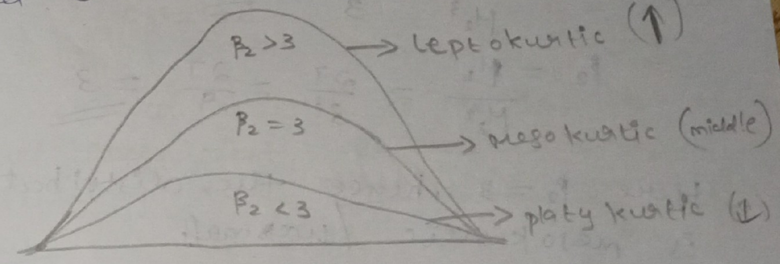
It measures peakedness / flatness of a freq.

The degree of (K) of a distribution is measured related to the flatness of a normal curve.

Meso kurtic :- A bell-shaped / normal curve is meso kurtic, bcz it is kurtic in the centre.

* Leptokurtic :- If a curve is relatively more peaked than the normal curve. → lepto kurtic.

* platy kurtic :- A curve more flat than normal curve → platy kurtic, bcz it is said to lack kurtosis.



The (K) is measured using 4th moment. which is an absolute measure.

The relative measure of (K) —

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2} \quad \text{or} \quad \gamma_2 = \beta_2 - 3.}$$

Q

1) The 1st 4 raw moments of a distribution are 1, 4, 10, 46. compute 1st 4 central moments & β constants. Comment upon the ntre of distribution.

A) $\mu_1' = 1$ $\quad \mu_2' = 4$ $\quad \mu_3' = 10$ $\quad \mu_4' = 46.$

central. m → $\mu_r = ?$

$\mu_1 = 0.$

$\mu_2 = \mu_2' - (\mu_1')^2 = 4 - 1^2 = \underline{\underline{3}}$

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

$= 10 - 3 \times 4 \times 1 + 2 \times 1$

$= 10 - 12 + 2 = 0$ //

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

$= 46 - 4\times 10\times 1 + 6\times 4\times 1 - 3\times 1$

$= 46 - 40 + 24 - 3 = \underline{\underline{27}}$

β constants

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{3} = \underline{\underline{0}}$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{27}{3^3} = \frac{27}{9} = \underline{\underline{3}}$

Here $\beta_2 = 3$, hence the distribution is mesokurtic / normal.

2) In a freq distri.. the coefficient of skewness based on quartiles is 0.5, if the sum of upper & lower quartiles is 28 & median is 11, find the values of upper & lower Q.?

A) Given, $Sk_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = 0.5$

$Q_1 + Q_3 = 28$ — ① $Q_3 - Q_1 = ?$

$M = 11$

$\therefore Sk_B = \frac{28 - 2\times 11}{Q_3 - Q_1} = 0.5$

$= \frac{28-22}{Q_3 - Q_1} = 0.5$

$= \frac{6}{Q_3 - Q_1} = 0.5$

$\Rightarrow \frac{6}{0.5} = Q_3 - Q_1$

$12 = Q_3 - Q_1$ — ②

① + ②

$Q_3 - Q_1 = 12$

$Q_3 + Q_1 = 28$

$2Q_3 = 40$

$Q_3 = \underline{\underline{20}}$

$20 + Q_1 = 28$

~~$20 - Q_1 = 12$~~

~~$Q_1 = 28 - Q_3$~~

$Q_1 = 28 - 20$

$= \underline{\underline{8}}$

3) If 3rd Q. is 178 & Median is 160 find coefficient Q. deviation, assuming the distri.. is symmetrical?

A) Given, $Q_3 = 178$

$M = 160$

(C) of QD = ?

No skewness = 0

(co always)

C of $QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Since the distri.. is symmetrical, then

$Sk_B = 0$

$Sk_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = 0$ (no skewness)

$\frac{178 + Q_1 - 320}{178 - Q_1} = 0$

$\Rightarrow 178 + Q_1 - 320 = 0$

$Q_1 = 320 - 178 = \underline{\underline{142}}$

$\therefore$ (C) of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{178 - 142}{178 + 142}$

$= \underline{\underline{\frac{36}{320}}}$

4) Cal Karl pearson's (C) of Skewness.

mid value	f	fx	$f(x-30.83)^2$
10	8	80	3471.1112
20	12	240	1407.4668
30	20	600	13.778
40	10	400	840.889
50	7	350	2572.4223
60	3	180	2552.6667
	60	1850	10858.334

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

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{1850}{60} = 30.83$$

~~$\text{Mode}(Z) = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$~~

Mode $= 30$ [കൂടുതൽ ഉള്ള freq].

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

$$= \sqrt{\frac{10858.334}{60}} = \sqrt{180.972}$$

$$= 13.45$$

$$Sk_p = \frac{30.83 - 30}{13.45} = \frac{0.83}{13.45}$$

$$= 0.0617$$