Lect # 07

Note:- Standard deviation of constant data is also zero bcz if $var(x) = var(a) = 0$.

$S.D = \sqrt{var(a)} = \sqrt{0}$

$\boxed{S.D = 0}$

Properties of Variance:-

Property # 01

Variance of constant data is zero.

Let X is variable which represent a data which are same observation a.

Proof:-

As;

$$Var(x) = \frac{1}{n} \Sigma (x - \bar{x})^2.$$

But here every observation is a

i.e $X = a$ & $\bar{x}$ is also a.

So;

$$var(a) = \frac{1}{n} \Sigma (a - a)^2$$

$$= \frac{1}{n} \Sigma (0) = 0.$$

(Proof)

Property # 2:-

Change of origin will not effect the variance

→ It means if you add a constant to every observation it will not effect the variance.

i.e

$$var(x + a) = var(x).$$

Proof:-

let

$y = x + a$ where a is a fixed number.

than $\bar{y} = \bar{x} + a$.

Note:- Change of origin does not effect S.D.

$$Var(y) = \frac{1}{n}\Sigma(y - \bar{y})^2$$

$$= \frac{1}{n}\Sigma(x + \cancel{a} - \bar{x} - \cancel{a})$$

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

$$= var(x)$$

(Proved)!

Property #3

If you change scale of data i.e you scale the data than variance will change square of variance of x.

i.e $Var(ax) = a^2 var(x)$.

while $S.D(ax) = |a| var(x)$.

Proof:-

$y = ax, \quad \bar{y} = a\bar{x}$.

$$var(y) = \frac{1}{n}\Sigma(y - \bar{y})^2$$

$$= \frac{1}{n}\Sigma(ax - a\bar{x})^2.$$

$= a^2 \frac{1}{n} \Sigma (x - \bar{x})$. "Take a^2 as common."

$= a^2 (var(x))$.

(Proved)

Property # 4:-

If x & y are Independant than
$var(x \pm y) = var(x) \pm var(y)$.

Proof:- Let
$z = x \pm y$ & $\bar{z} = \bar{x} \pm \bar{y}$.

Since
$$var(z) = \frac{1}{n} \Sigma (z - \bar{z})^2.$$

$$= \frac{1}{n} \Sigma ((x \pm y) - (\bar{x} \pm \bar{y}))^2.$$

$$= \frac{1}{n} \Sigma ((x - \bar{x}) \pm (y - \bar{y}))^2.$$

$$= \frac{1}{n} \Sigma [(x - \bar{x})^2 + (y - \bar{y})^2 \pm 2(x - \bar{x})(y - \bar{y})]$$

$$= \frac{1}{n} \Sigma (x - \bar{x})^2 + \frac{1}{n} \Sigma (y - \bar{y})^2 \pm 2 \frac{1}{n} \Sigma (x - \bar{x})(y - \bar{y})$$

$$= Var(x) + Var(y) \pm 2 \, Cov(x, y). \longrightarrow 1$$

Cov variance:-

Cov variance $cov(x,y)$ is used to

the change in data w.r.t change in other type of data.

For e.g if we have one data & other data & we change something (data) in one data & see what is its effect on other data we used covariance.

As Independant

so; $Cov(x,y) = 0$.

$\because Cov(x,y)$ is for two different data.
→ The change in one data w.r.t to other change in data.

eq ① ⇒

$$Var(x \pm y) = Var(x) + Var(y).$$

(Proved)

S.D will be;

$$Var(x,y) = \sqrt{Var(x) + Var(y)}.$$

Standardize Variable:-

If we have a data we change its origin in such a way that its mean become zero & S.D 1.

→ We bring the data to origin.

(*) If we have a data of x_i having mean of $\bar{x}$ & S.D. "s" for sample.

Scanned by CamScanner

We Introduce a new variable.
i.e

$$z = \frac{x_i - \bar{x}}{s}$$

This new variable z is called standardize variable.

Note:- $\bar{x}$ is constant

s is also constant

& mean of x_i again $\bar{x}$. so

$$\bar{z} = \frac{\bar{x} - \bar{x}}{s} = 0.$$

To find its mean.

$$\bar{z} = \frac{\bar{x} - \bar{x}}{s} = 0.$$

$$var(z) = \frac{1}{s^2} var(x).$$

But $var(x) = s^2$.

$$Var(z) = \frac{1}{s^2} var(x).$$

$$= \frac{1}{s^2} (s^2)$$

$$\boxed{var(z) = 1}$$

Standard deviation of z will be also 1

$$\boxed{S.D(z) = 1}$$

Scanned by CamScanner

Trimmed & Winsorized Measure:-

If we have outlayer in date than mean not properly represent the data so we used median.

Trimmed Measure:-

As we know that for outlayer the mean is not properly represent date so we trimmed the data from both end and than find its mean which is approximatly equal to actual of mean, for the date having outlayers. If there is outlayer in date than this new mean (trimmed) mean will give diff value of mean which show There is outlayers in data

→ The data that should be trimmed before Q_1 and after Q_3.

→ Than if we find some mean, variance etc are consider trimmed mean & trimmed variance

(±) Here as we trimmed date so may be there is chance that we also trimmed the something other than outlayer. So we move to other method i.e.

(b) **Winsorized Measure:-**

Winsorized measure tells us that there is no need to cut the data (i.e trimmed)

It tells us that data behind quartile 1 replaced by Q1. & replaced the observation after Q3 replaced by Q3.

→ The mean or variance calculated by this method called Winsorized mean & Winsorized variance respectively.

(*) If there is no out layer than Winsorized & actual has no difference

But if there are out layer than it will effect the data obtained by actual method & Winsorized data.

For example:- Consider the data which has out layer i.e. 1, 3, 7, 9, 12, 15, 20, 29, 100.

Q_1 (7), Q_3 (20)

$\bar{X} = 169$ (Mean).

By trimmed Method:-

$$X_T = \frac{7+9+12+15}{5} = 12.6$$

By Winsorized method

$X_W = 13$ in data.

As we get different values So means there is out layer

→ For Variance:-

If we find variance ~~Kay~~ for actual data and than we trimmed data and find its trimmed variance.

(more value)

(*) → If trimmed variance increase than actual data variance ~~mean~~ so it means due to boundry the variation was less now as trimmed so now variance increase & Intermediate data mean the variation is more in mid.

→ But if trimmed variation decreas (lower value) than actual data variance. So it mean there is more variance when we have end boundryes. So mean less variation is less in between data.

Example # 4.12

Calculate the trimmed & Winsorized means and standard deviation for given exp in 3.11

33, 36, 36, 37, 39, 41, 45, 46, 48.

Solution:- As given data is

33, 36, 36, 37, 39, 41, 45, 46, 48.

$\downarrow Q_1$ $\downarrow Q_2$

$$\bar{X}_T = \frac{36+37+39+41+45}{5} = 39.6.$$

$$\text{Trimmed S.D} = \sqrt{\frac{(36-39.6)^2+(37-39.6)^2+(39-39.6)^2+(41-39.6)^2+(45-39.6)^2}{5}}$$

$$(S.D)_T = \sqrt{\frac{51.2}{5}} \Rightarrow \sqrt{10.24} = 3.2.$$

$$\boxed{(S.D)_T = 3.2}$$

Book answer wrong.

Now By Winsorized method:-

$$\bar{X}_W = \frac{36+36+36+37+39+41+45+45+45}{9}$$

$$\bar{X}_W = \frac{360}{9} \Rightarrow \boxed{\bar{X}_W = 40}$$

$$\text{Winsorized S.D} = \sqrt{\frac{36^2+36^2+36^2+37^2+39^2+41^2+45^2+45^2+45^2}{n} - (40)^2}.$$

$$S.D_W = \sqrt{\frac{14534}{9} - (40)^2}.$$

$$= \sqrt{1614.889 - 1600} \Rightarrow \sqrt{14.889}$$

$$\boxed{S.D_W = 3.86}$$ Ans

Scanned by CamScanner

High Order Dispersion:-

Moment:-

The equation i.e written in form of

$\frac{1}{n}\Sigma(x_i-\bar{x})^r$ is know as moment.

Where $r = 0, 1, 2, 3, \ldots$

In variance $r = 2$, In mean deviation $r = 1$.

Due to $r = 2$ it also makes the -ve value +ve Due to which the value (variance) does not represent data properly.

→ It also called Localization of something i.e to further study the details of data.

(*) The data not easily we seen & known so we localized it i.e consider its different moment.

→ Small variation in data sometimes we not see so we localize that.

Moments are also of two types.

Moment

- Central moment.
- Arbitrary moment.

(a) Central moment.

(i) For Population data

$$\mu_r = \frac{1}{N}\sum(x_i-\mu)^r.$$

(ii) Sample data

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

Where $r = 1, 2, 3, \ldots, n.$

For frequency data

$$m_r = \frac{1}{n}\sum f_i(x_i-\bar{x})^r.$$

Arbitrary Moment.

(i) For population data.

$$\mu_r' = \frac{1}{N}\sum(x_i-a)^r.$$

For Frequency data.

$$\mu_r' = \frac{1}{N}\sum f_i(x_i-a)^r.$$

(ii) For sample data.

$$m_r' = \frac{1}{n}\sum(x_i-a)^r.$$

For FB:-

$$m_r' = \frac{1}{n}\sum f(x_i-a)^r.$$

Where $r = 0, 1, 2, 3, \ldots n.$

In calculation we can also set μ_0', μ_1' & etc.

Cases:-

i- **if $r = 0$.**

$$\mu_0 = m_0 = \mu_0' = m_0' = 0.$$

ii- **if $r = 1$.**

For population and sample

$$\mu_1 = m_1 = \frac{1}{n}\sum \cancel{(x_i - \bar{x})}^{0}$$

$$\boxed{\mu_1 = m_1 = 0.}$$

$$\mu_1' = m_1' = ?$$

As; $\mu_1' = \frac{1}{N}\sum(x_i - a).$

$$= \frac{1}{N}\sum x_i - \frac{1}{N}\sum a$$

$$= \mu_x - \frac{\cancel{N}\, a}{\cancel{N}}$$

$$\boxed{\mu_1' = \mu_x - a}$$

For Sample:-

$$\boxed{m_1' = \bar{x} - a}$$

if $r = 2$

$\mu_2 = \frac{1}{\Sigma}(x - \mu)^2 = \sigma^2$ } variance

$m_2 = 1/\Sigma\,(x - \bar{x})^2 = s^2.$

Relation b/w Central moment and Arbitrary moment :-

Relation to find Arbitrary moment to Central moment :-

i.e m_r in term of m_r'

As we know that;

$$m_r = \frac{1}{n} \Sigma f_i (x_i - \bar{x})^r.$$

Add & Subtract a &

$$= \frac{1}{n} \Sigma f_i (x_i - a + a \mp \bar{x})^r.$$

$$= \frac{1}{n} \Sigma f_i ((x_i - a) - (\bar{x} - a))^r \longrightarrow (1)$$

As $x_i - a = D_i$.

$m_1' = \bar{x} - a$.

eq (1) $\Rightarrow$

$$= \frac{1}{n} \Sigma f_i (D_i - m_1')^r.$$

By using binomial expansion.

$$= \frac{1}{n} \Sigma f_i \left[D_i^r - r D_i^{r-1} m_1' + \frac{r(r-1)}{2!} D_i^{r-2} (m_1')^2 + \cdots + (-1)^r (m_1')^r \right].$$

$\Rightarrow$ As replace $D_i = x_i - a$.

$$= \underbrace{\frac{1}{n}\sum f_i (x_i - a)^r}_{\text{Arbitary moment}} - r m_1' \frac{1}{n}\sum f_i (x_i - a)^{r-1} + \frac{r(r-1)}{2!}(m_1')^2 \frac{1}{n}\sum f_i (x_i - a)^{r-2}$$

$$+ \cdots + r!(m_1')^r.$$

$$= m_r' - r m_1' m_{r-1}' + \frac{r(r-1)}{2!} m_1'^2 m_{r-2}' - \frac{r(r-1)(r-2)}{3!} m_1'^3 \cdot m_{r-3}' + \cdots$$

②.

If $r = 1$, ②

$$m_1 = m_1' - m_1' = 0.$$

if $r = 2$.

$$m_2 = m_2' - 2 m_1' m_1' + \frac{2}{2} m_1'^2 \cancel{m_0'}^{\,1}.$$

Since $m_0' = 1$.

$$m_2 = m_2' - 2m_1'^2 + m_1'^2.$$

$$= m_2' - m_1'^2.$$

If $r = 3$

$$m_3 = m_3' - 3 m_1' m_2' + \frac{3(2)}{\cancel{2}} m_1'^2 m_1' - \frac{3(2)(1)}{\cancel{3!}} m_1'^3 \cdot \cancel{m_0'}^{\,1}.$$

$m_0' = 1$

$m_3 = m_3' - 3m_1' m_2' + (2)(m_1')^3$

Now m_r' in Term of m_r.

$m_1' = 1$

Then

$\boxed{m_1 = \bar{x} - a}$

For $r = 2$.

$\boxed{m_2' = m_2 + (m_1')^2}$.

For $r = 3$

$\boxed{m_3' = m_3 + 3m_1' m_2 + (m_1')^3}$.

Moment ratio :-

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$. : β_1 is square of Third moment & Cube of second moment.

& $\beta_2 = \frac{\mu_4}{\mu_2^2}$. β_2 is fourth moment to square of second moment.

For Sam.

Example # 4.13

Calculate the first four moments about the mean for the following set of marks:

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

Sol:- We have to first find mean of data so arrange the data in accending order.

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{32+36+36+37+39+41+45+46+48}{9}$$

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

to find moments upto 4. So use table.

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
Σ 0		$\Sigma(x_i-\bar{x})^2 = 232$	$186 = \Sigma(x_i-\bar{x})^3$	$\Sigma(x_i-\bar{x})^4 = 10708$

Now;

$$m_1 = \frac{\Sigma(x_i - \bar{x})}{n} = 0.$$

$$m_2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{232}{9} = 25.78.$$

$$m_3 = \frac{\Sigma(x_i - \bar{x})^3}{n} = \frac{186}{9} = 20.67$$

$$m_4 = \frac{\Sigma(x_i - \bar{x})^4}{n} = \frac{10708}{9} = 1189.78.$$

Example # 4.14 Compute the first four moments for the following distribution ~~ages~~. about arbitary a.

Weekly earning (Rs).	5	6	7	8	9	10	11	12	13	14	15
No. of Men	1	2	5	10	20	51	22	11	5	3	1

Solution:- First we have to find (a) so from data we find max freq is 51 so $a = 10$.

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

$$m_1' = \frac{\Sigma f(x_i - a)}{n} = \frac{\Sigma f(x_i - a)}{\Sigma f}$$

$$= \frac{8}{131} \Rightarrow \boxed{m_1' = 0.06}$$

$$m_2' = \frac{\Sigma f(x_i - a)}{\Sigma f} = \frac{346}{131} \Rightarrow \boxed{m_2' = 2.64}$$

$$m_3' = \frac{\Sigma f(x_i - a)^3}{\Sigma f} = \frac{74}{131} \Rightarrow \boxed{m_3' = 0.56}$$

$$m_4' = \frac{\Sigma f(x_i - a)^4}{\Sigma f} = \frac{3718}{131} = \boxed{m_4' = 28.38}$$

Now to find moments ~~mean~~ about mean.

$\boxed{m_1 = 0}$ As $m_1 = \bar{x} - a$

$= 10 - 10$

$= 0$

11
22
22 ×
242

$$m_2 = m_2' - (m_1')^2$$

$$= 2.64 - (0.06)^2$$

$$= 2.64$$

11
12
22
11 ×
132

$$m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3$$

$$m_3 = 0.56 - 3(2.64)(0.06) + 2(0.06)^3 = 0.08$$

$$m_4 = m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m_1')^4$$

$$= 28.38 - 4(0.56)(0.06) + 6(2.64)(0.06) - 3(0.06)^4$$

$\boxed{m_4 = 28.3}$ Ans

Example:- 4.15 Find the first four mean for the following data & Ex # 4.4.

weights	65-84	85-104	105-124	125-144	145-164	165-184	185-204
f	9	10	17	10	5	4	5

Solution:-

x_i	f	u_i	fu_i	fu_i^2	fu_i^3	fu_i^4
74.5	9	-2	-18	36	-72	144
94.5	10	-1	-10	10	-10	10
114.5	17	0	0	0	0	0
134.5	10	1	10	10	10	10
154.5	5	2	10	20	40	80
174.5	4	3	12	36	108	324
194.5	5	4	20	80	320	1280
			$\Sigma fu_i = 24$	$\Sigma fu_i^2 = 192$	$\Sigma fu_i^3 = 396$	$\Sigma fu_i^4 = 1848$

$a = 114.5$

$u = \frac{x_i - a}{h}$

Now to find first moment.

As,
$$m_1' = \frac{\Sigma fu_i}{\Sigma f} = \frac{24}{60} \Rightarrow \boxed{m_1' = 0.4}$$

$$m_2' = \frac{\Sigma fu_i^2}{\Sigma f} = \frac{192}{60} \Rightarrow \boxed{m_2' = 3.2}$$

$$m_3' = \frac{\Sigma fu_i^3}{\Sigma f} = \frac{396}{60} = \boxed{m_3' = 6.6}$$

$$m_4' = \frac{\Sigma fu_i^4}{\Sigma f} = \frac{1848}{60} \rightarrow \boxed{m_4' = 30.8}$$

Scanned by CamScanner

For moment about mean use these formulea.

$m_1 = 0.$

$m_2 = m_2' - (m_1')^2.$

$m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3.$

$m_4 = m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - (3m_1')^4.$

Example # 4.16 ~~Find~~ The First three moments of a distribution about the value 2 of the variable are 1, 16 & −40. Show that the mean is 3, the variance 15 & $m_3 = -86$ Also show that first three moments about $x = 0$ are 3, 24 & 76.

Sol:- As given mean about so.

$a = 2.$

So, As

$m_1' = 1.$

$m_2' = 16$

& $m_3' = -40.$

As we also know that;

$m_1 = \bar{x} - a.$

$\bar{X} = m_1 + a.$

$\bar{X} = 1 + 2.$

$\boxed{\bar{X} = 3}$ //Proved!//

Since

Variance = m_2.

$s^2 = m_2' - (m_1')^2$

$= 16 - (1)^2 \Rightarrow \boxed{15 = s^2}$

$\boxed{\text{proved}}$

Now;

$m_3 = ?$

As;

$m_3 = m_3' - 3m_2'm_1' + 2m_1'^3$.

$= -40 - 3(16)(1) + 2(1)^3 = -40 - 48 + 2$.

$\boxed{m_3 = -86}$

To show m_1 about $a = 0$ is 3.

Since;

$m_1 = \frac{1}{n}\Sigma(x_i - a) = \frac{1}{n}\Sigma(x_i - 0) = \frac{1}{n}\Sigma x_i$

$\frac{1}{n}\Sigma x_i = \bar{x} \longrightarrow (1)$.

And $\bar{x} = 3$.

So; $\boxed{m_1 = \frac{1}{n}\Sigma x_i = 3}$ $\boxed{\text{proved}}$

$m_2 = \frac{1}{n}\Sigma(x_i - a)^2 = \frac{1}{n}\Sigma(x_i - 0)^2 = \frac{1}{\Sigma n}\Sigma x_i^2$.

$\hookrightarrow (2)$.

Scanned by CamScanner

Since;

$$m_2' = \frac{1}{n}\sum(x_i - 2)^2 = 16.$$

$$\Rightarrow \frac{1}{n}\sum x_i^2 - 4\frac{1}{n}\sum x_i + \frac{1}{n}\sum 4 = 16.$$

$$\frac{1}{n}\sum x_i^2 - 4.\bar{x} + \frac{4\cancel{n}}{\cancel{n}} = 16.$$

$$\frac{1}{n}\sum x_i^2 - 4(3) + 4 = 16.$$

$$\frac{1}{n}\sum x_i^2 = 16 + 8.$$

$$\boxed{m_2' = \frac{1}{n}\sum x_i^2 = 24}$$ proved!

$$m_3 = \frac{1}{n}\sum(x_i - 0)^3 = \frac{1}{n}\sum x_i^3 \longrightarrow (3).$$

As

$$m_3' = \frac{1}{n}\sum(x_i - 2)^3 = \frac{1}{n}\sum\left(x_i^3 - 3x_i^2(2) + 3x_i(2)^2 - (2)^3\right)$$

$$= \frac{1}{n}\sum\left(x_i^3 - 6x_i^2 + 12x_i - 8\right) = -40.$$

$$= \frac{1}{n}\sum x_i^3 - \frac{6}{n}\sum x_i^2 + \frac{12}{n}\sum x_i - 8\frac{\sum 8}{n} = -40.$$

$$= +\frac{1}{n}\sum x_i^3 - 6(24) + 12(3) - 8 = -40.$$

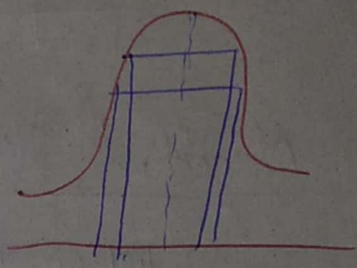
$$\Rightarrow \frac{1}{n}\sum x_i^3 = 116 - 40$$ Proved

$$m_3 = \boxed{\frac{1}{n}\sum x_i^3 = 76}$$ Ans.

Skewness:-

For a distribution or for a data observation equidistant from the mean we have same frequency than we say that is symmatric.

Consider the fig.



If we have a graph which represent a data. Let the mean is at centre so if there is any two observations which are equidistant from mean & at any two points if frequency is same. Than distribution is symmatric.

→ For distribution if frequency of observation equidistant from mean is same the distribution is called symmatric.

And any variation (deviation) from symmatry is called skewness.

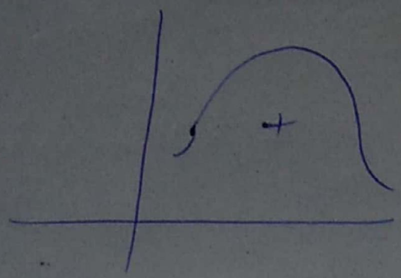
→ If there is no symmatric than skewed distribution.

→ Skewed distribution

- Negative skewed
- Positive skewed

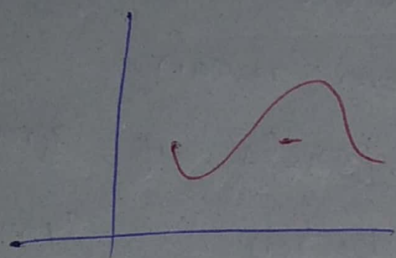
(i) Positive Skewed:-

When mode is ~~greater~~ less than mean than the skewness is positive.



(ii) Negative Skewed:-

When mode is great than mean than negatively skewed.



Formulea for skewness:-

$$Sk = \frac{Mean - Mode}{S.D}$$

or

$$Sk = \frac{3(Mean - Median)}{S.D}$$

or

$$Sk = \sqrt{\beta_1} = \frac{\mu_3}{\delta^3}.$$

For Symmatric distribution.

Note:- Mean = Mode = Median.