Drawback → Effected by extreme values

$\bar{x} = \Sigma 1$

Measures of Central Tendency

AM → $\bar{x}$

both x & f given

$\frac{\Sigma f x}{f (n)}$

$\bar{x} = \frac{\Sigma x}{n}$ → Best, Accurate, reliable method

Note → If class lengths are equal or unequal the formula is same.

Find the AM of 5, 10, 15, 20, - - - 75 ?

$\bar{x} = \frac{\text{First term} + \text{last term}}{2}$

Note - If the given data is in AP

AP is difference is constant

Find the AP of 1, 2, 3 . . . n?

$\bar{x} = \frac{\Sigma x}{n} = \frac{1+2+3+\cdots n}{n}$

∴ Sum of natural numbers

$= \frac{n+1}{2}$ //

If frequency of variable increase or decreases by same variable the value of ~~xen~~ AM will be remain constant

Series of Data

a. $1, 2, 3, 4, 5 \ldots n = \frac{n+1}{2}$

b. $1^2, 2^2, 3^2 \ldots n^2 = \frac{(n+1)(2n+1)}{6}$

c. $1^3, 2^3, 3^3 \ldots n^3 = \frac{n(n+1)^2}{4}$

Weighted Average → $\bar{x}_w = \frac{\Sigma wx}{\Sigma w}$ $\left[\bar{x} = \frac{2n+1}{3}\right]$

1) $\bar{x}$ (AM) is most effected by extreme observation

Note → AM cannot be calculated for open end frequency distribution. → Start from - less & end with above.

Properties of AM →

Missing freq / value → Q → $2x$ x x $3x$ $4x+3$ $x+1$ $x+2$. AM = 10

$\bar{x} = \frac{\Sigma x}{n}$ /

Properties of AM →

1) If data values are constant the AM values constant
for ex- 20, 20 . . . . ans = 20 //

2) AM shows linear relationship. for ex- $y = 5x+3$
find $\bar{y}$, $\bar{x} = 10$ $y = ax+b$

3) All averages $\overset{mean}{(\bar{x})}$, median(M), mode(Z), GM, HM
shows linear relation property.
for ex - $5x + 3y = 10$ $\bar{x} = 5$ find y?
Ans = -5 //

4) AM can be pooled (combined) $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$
in case $n_1 = n_2$, $\bar{x} = \dfrac{\bar{x}_1 + \bar{x}_2}{2}$

5) Sum of deviation from $\bar{x}$ is always 0
$(\Sigma x - \bar{x} = 0)$

6) If frequency of variable increases or decrease
by same variable of AM will be remain
constant.

7)

find →

7) $\frac{\Sigma fx}{f}$ If f is high then opposite class frequency in that distribution there will be ans

GM

Application –

1. It is less affected by extreme values
2. Extensively used in the constructuretion of index no's
3. GM cannot be calculated of any variable assumes 0 or negative value.
4. Less affected by sampling fluctuating compare to Z & M.
5. GM is most different avg to cal^n & understanding (It involve log)
6. GM can be used in accounts to calculate dep^n in WDV method.
7. It is used to calculate Growth rate

Prop –

8. If data values are constant 'k' then GM = k
9. GM shows linear relation
10. GM can be combined

$$G_{12} = \left(G_1^{n_1} \times G_2^{n_2}\right)^{\frac{1}{n_1+n_2}}$$

GM

Ex- $n_1 = 3$ $n_2 = 5$ $G_{12} = (4^3 \times 8\ 3^5)$

$G_1 = 4$ $G_2 = 3$

$= \sqrt[8]{64 + 243}$

$= \sqrt[8]{15552}$ → Pinto

$= 1.82$

* GM = 0, when data values are zeros.
* GM is only for +ve data values
* When there is 0 or −ve we should not calculate GM.

$$GM = \sqrt{\text{Product of values}}$$

$$GM = \sqrt{x_1^f \times x_2^f \times x_3^f}$$

HM

$$\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}\cdots\frac{1}{x_n}}$$

If in case

HM cannot calculated when if any variable assumes 0

HM can be calculated for +ve or -ve data

If all observavation is same ans is same

$$HM = \frac{\Sigma f}{\Sigma \frac{f}{x}} \rightarrow \text{Grouped frequency}$$

~~Application~~ ~~of~~ ~~HM~~

HM can be combined (Pooled HM)

$$H_{12} = \frac{n_1+n_2}{\frac{n_1}{h_1}+\frac{n_2}{h_2}}$$

for equal data value

$$n_1 = n_2$$

$$H_{12} = \frac{2H_1 H_2}{H_1+H_2}$$

Relation

AM ≥ GM ≥ HM

AM = GM = HM → when data is constant

AM > GM > HM → When data is distict

relation

Mathematical = $GM = \sqrt{(AM \times HM)}$

Median → Median is second best average.

↳ It can be calculated for open end as well as close end frequency.

- Quartile
- Decile
- Percentile

} Fractile family.

Median is dependent of scale ~~off~~ origin

Ungrouped → step 1 → Make ascending order

odd → middle value is median

even → middle two values ~~do~~ divide by 2

Grouped

$$X \left\{ l + \frac{\left(\frac{N}{2} - f\right)}{fm} ; \right.$$

Total freq ← $M = \frac{N}{2}$

$$l + \frac{\left(\frac{n}{2} - f\right)}{fm} ;$$

or

$$m = l_1 + \left(\frac{m - cf}{f}\right)(l_2 - l_1)$$

Mode → It is most repeated value.
↳ It is least used avg method.
↳ It is ill defined
↳ It can be calculated for open end freq.
↳ It is dependent on change of origin & scale. It is not affected by extreme values.
↳ It is affected by sampling flactuation.

x →	5	4	8	30	18	5
f →	5	10	20	90	18	90

Ans → 30, 5

Mode can located graphically using histogram or area digram or freq diagram.

Unimodel → one number repeated multiple times

bi → two number's repeated multiple

$$m_0 = l_1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) l_2 - l_1$$

upper f_0

highest freq is f_1

Relation b/w mean, median & mode

Mean = Median = mode. → Normal (symetrical)

Mean ≤ Median ≤ mode → left symetry

mean ≥ median ≥ mode → right symetry

↳ If it is not =, ≤, ≥ → Assmetrical.

Mode = 3 median − 2 mean

mean − mode = 3 (mean − median)

↳ logic → Memo = 3 meme.

upper quartile

$Q_1 = 25\%$. $Q_2 = 50$ $Q_3 = 75$ $Q_4 = 100$

lower quartile Q_2 → median of data.

Measures of Dispersion

Dispersion → Deviation or Scatterness Of values from their central values (mean & median)

→ Variability in uniformity.

→ If values are equal then MOD is always zero.

→ All MOD are independent of change of origin but dependent on change of scale.

- Has 2 types

1> Absolute Dispersion	Relative Dispersion
Same units	Expressed as ratios or percentage & unit free.
- Related to distinct itself Range, QD, MD, SD	- To compare variability b/w 2/more series. - To check relative accuracy of data.

Range - • Quickest method. Easy to calⁿ. less time consuming.

• It does not depend in all observation
• Most unrealiable.
• Unaffected by presence of frequency.
• Indepence of change of origin, dependent on change of scale.

$$\text{Range} = HV - LV$$

$$\text{Co. of Range} = \frac{HV - LV}{HV + LV} \times 100$$

• It is not a accurate measure.
• It can not be calⁿ for open end.

Quartile Deviation -

- Half of the Range b/w the quartiles.
- Based on upper & lower quartile & covers 50% of observation.
- Does not depend on all observation.
- QD is the best & only MOD for open end.
- QD is not accurate measure.

$$QD = \frac{Q_3 - Q_1}{2} \qquad \text{Coeff. of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

For grouped freq dist

$$Q = L_1 + \frac{m - cf}{f} \times (L_2 - L_1) \qquad Q_1 = m = \frac{N}{4} \qquad Q_3 = m = \frac{3N}{4}$$

Mean Deviation -

- MD better than Range & QD as MOD as it depends on 100+ data.
- MD is not accurate MOD (becoz of module) - as it lesser absolute deviation.
- Cannot be used for open-end dist.

$$MD = \frac{\Sigma|x - \bar{x}|}{n} \qquad \bar{x} = \frac{\Sigma x}{n}$$

$$\text{Coeft} = \frac{MD}{\text{mean}} \times 100$$

Standard Deviation –

→ Best, Commonly used, Accurate, Reliable method

→ Unduly effected by presence of extreme values

→ It is also know as Root - Mean - Square Deviā (RMSD)

→ It is denoted by σ.

→ SD^2 = variance = σ^2

→ If all observations are equal variance = SD = 0

→ SD – Defined as the +ve square root of AM of the square deviations of the value from their AM

Simple series/ w/o freq ungrouped

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

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left[\frac{\Sigma x}{n}\right]^2}$$

$$\sigma^2 = \sqrt{\frac{\Sigma d^2}{n} - \left[\frac{\Sigma d}{n}\right]^2} \times i$$

Simple / Grouped

Not imp
Just go through

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

$$\sigma = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left[\frac{\Sigma f d}{\Sigma f}\right]^2}$$

$$\sigma_x = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left[\frac{\Sigma f d}{\Sigma f}\right]^2} \times i$$

where $d = \frac{x-A}{i}$, x = mid value, i = class width

A = assumed mean

coeff of variation $= \frac{SD}{mean} \times 100 = \frac{\sigma}{\bar{x}} \times 100$

- Coeff of variation is the best measure of disp.
- It is used to compare variability or consistency b/w 2/more series.
- More CV implies more ~~variability~~ or reliability indicating less stability or consistency & vice versa
- Regarding choice - always choose item which has less CV becoz, item with lower CV is more stable

SD of 2 no's a, b. SD $= \left|\frac{a-b}{2}\right|$

variance of first 'n' natural no's $= \frac{n^2-1}{12}$

Sum of squares of observations $\Sigma x^2 = n(\sigma^2 + \bar{x}^2)$.

Linear R/s prop

$R_y = |b| R_x$

$QD_y = |b| QD_x$

$MD_y = |b| MD_x$

$SD_y = |b| SD_x$

coefficients

$\frac{R_y}{y} \times 100 \cdot \frac{MD_y}{\bar{y}} \times 100$

Relation b/w QD, MD & SD

$SD \geq MD \geq QD$

$SD = MD = QD = 0$

$QD = \frac{2}{3} SD, MD = \frac{4}{3} SD.$